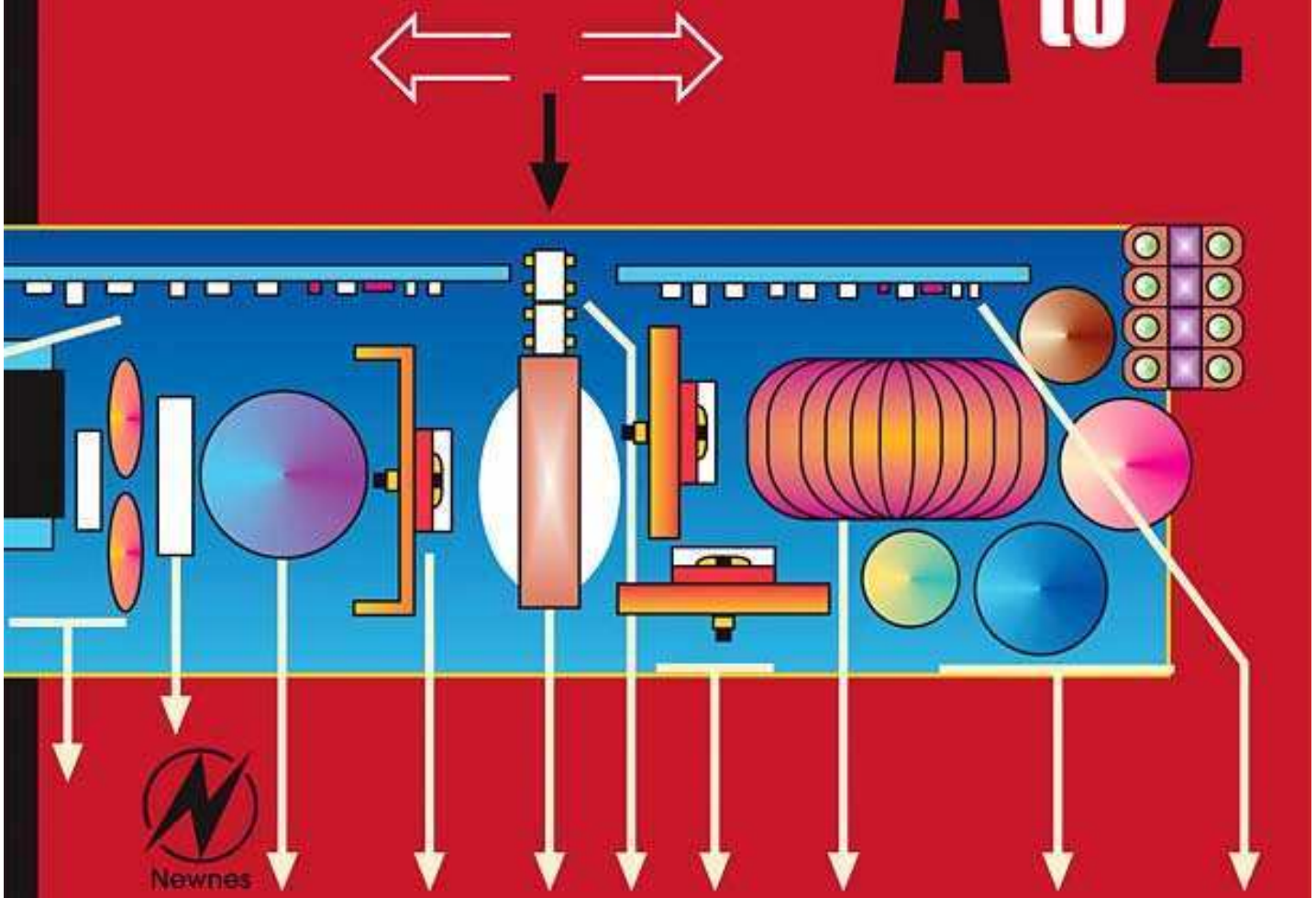
Switching Power Supplies A to Z

SANJAYA MANIKTALA



Newnes

Switching Power Supplies A to Z

This Page Intentionally Left Blank

Switching Power Supplies A to Z

Sanjaya Maniktala

ELSEVIER

AMSTERDAM • BOSTON • HEIDELBERG • LONDON
NEW YORK • OXFORD • PARIS • SAN DIEGO
SAN FRANCISCO • SINGAPORE • SYDNEY • TOKYO

Newnes is an imprint of Elsevier

Newnes is an imprint of Elsevier
30 Corporate Drive, Suite 400, Burlington, MA 01803, USA
Linacre House, Jordan Hill, Oxford OX2 8DP, UK

Library of Congress Cataloging-in-Publication Data
Application Submitted

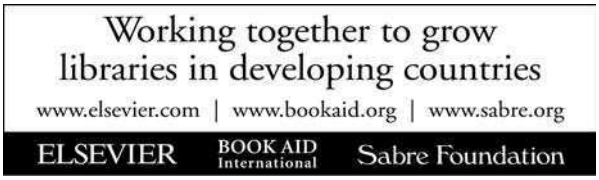
British Library Cataloguing-in-Publication Data
A catalogue record for this book is available from the British Library.

ISBN 13: 978-0-7506-7970-1
ISBN 10: 0-7506-7970-0

For information on all Newnes publications
visit our Web site at www.books.elsevier.com

06 07 08 09 10 10 9 8 7 6 5 4 3 2 1

Printed in the United States of America



Contents

This Page Intentionally Left Blank

Preface

Looking back, I realize there was actually never an exact moment when I said to myself "I have had it with Physics — let me do Electronics now"! My near mid-life career change was a rather gradual process. Faced with an exponentially declining interest in pi-mesons, Lagrangian multipliers, quantum electrodynamics, and so on, my grades had started scraping the bottom of the barrel. It didn't help that I perceived my last bunch of professors to be largely apathetic to students in general — it seemed that teaching just happened to be what they *needed* to keep doing, to be *eligible* for research grants, which is what they *really* enjoyed doing. And Physics itself, for all its initial undergraduate allure, had at the post-graduate level, turned suddenly very mathematical and abstruse, contradicting my inherent desire to be firmly embedded in reality (not virtual reality). Unfortunately, the disenchantment reached a culmination only during my second Masters degree program, in Chicago. Too late! So though I eventually did part company with Physics (as good friends, I may add), there was a slight problem — I didn't have a clue what to do next. I call that my *Problem Number 1*.

Back in hot, bustling and dusty India, it took me several years to figure things out. But finally, I did! The small bags of transistors, capacitors, resistors and inductors that I had lately started tinkering with, held the answer to all my problems. And hope for the future. This was my long-awaited lifeboat. I could now feel, touch, build and test whatever I did. No deceptive sense of comfort lolling around in lush minefields of equations and algebraic abstractions. This was the real world, the one that we live in every day.

Problem Number 2: I still didn't know the ABCs (or the NPNs) of electronics. So I had to teach myself very gradually, working days and often very late into the nights, barely stopping only to ask the elderly local components dealer, daring questions like — what is a transistor?! This act went on for a pretty long time — in fact I became the Rocky Horror Midnight show — you got to see me mostly at midnight for several years in succession. But it would have still been impossible if I had not met a few very remarkable men along the way (see Acknowledgements). Finally, with all the help at my disposal (most of it mine), I think I made it into the exciting world of electronics. And into power electronics. Aha! I could start rolling down the shutters now. Or could I?

The above chain of hair-raising events is the one and only reason this book ever got written in the first place!

But wait! I have explained "how" this book got written, but did I explain *why*? Actually I haven't yet. Because that has something to do with the last major problem I faced. I call it *Problem Number 3* — encountering people who knowingly or unknowingly thwart the growth of the engineering discipline that gives us growth. Now, I had personally been through a rather life-changing process (of being rescued by Electronics). So perhaps it was more natural for me to always think I owed Electronics my best, in return for favors received. But I realize not everyone thinks along those lines, at least not *all* the time. Maybe they had affluent fathers paying for their shiny EE degrees from MIT or Yale. But I didn't have an affluent father nor an EE degree. However, at some stage, we all have to realize that we share the same forces of nature, and a common stake in its existence and further development. Our destiny is eventually common, and therefore we have a common responsibility to uphold it too. Anybody who has learnt enough about the immense mystical forces of nature realizes that he or she has really learnt *nothing* at all. It will therefore be very surprising if they don't end up imbibing the sense of humility that Newton once expressed in the following words:

"I know not what I may appear to the world, but to myself I seem to have been only like a boy playing on the sea-shore, and diverting myself in now and then finding a smoother pebble or a prettier shell, whilst the great ocean of truth lay all undiscovered before me."

Power Electronics too, is just a small part of that infinitesimal part of the universe that we have just begun to understand. There is much, much more, just waiting to be discovered. Should we be the ones to encourage that onward natural process, or thwart it (even momentarily) with our petty office-space personal agendas?

Finally, when I had seen too much and heard too much, I wrote the following paragraphs somewhere on the web, in what is now a rather controversial opinion piece for some people who obviously don't understand the logic or the motivation for it (see last page of Appendix 1)

'Technology may never gain a foothold in a "king's court," where you are either rewarded with largesse for being vehemently agreeable, or unceremoniously sentenced to the dark dungeons for the rest of your life. Engineers like to speak out - but usually only when they were sure of their facts and have incontrovertible data to back themselves up. They therefore deserve and need a "peer environment," where they are judged (primarily) by the respect received from their peers — the king be damned (on occasion)!

It must be kept in mind that this can really bother the king sometimes! So managers who supervise engineers, should be fairly competent at a technical level themselves and respect data and facts equally. They can't attempt to win a technical argument by throwing rank on

their subordinates. Nor should they ever go around, God forbid, trying to subsequently shoot the "emotional and/or disrespectful" engineer down ("that'll teach him"). Surprisingly that does happen more than we dare admit. Not only does the good engineer pay the price, but so does technology in the long run.'

The only reason that this piece (based largely on the old wisdom of my dear long-time mentor and former-former Boss Dr GT Murthy) turned controversial was I suspect, because it had hit closer to home than even I had imagined. It is always amusing that whenever someone has one-too-many skeletons in their closets, the very sound of a distant siren triggers off their worst fears. I was told to stop talking about things I didn't understand, and stick to my (humble) circuits. I was also refused the normal official Author Encouragement Program payment that I thought was due to me as per their guidelines — for this article and even for my other popular power electronics book, which they had already used freely to promote their products. Finally the best thing I did was to quit as soon as I could! Without notice. Then surprise, surprise! Just after I resigned, they went and restructured exactly as I had been preaching all along — by re-amalgamating their two erstwhile groups "Portable Power" and "Power Management" into one, saying privately that there would be "*more sharing among the engineers finally*". My words exactly (read the article)! Weighing all these events in mind, I found some peace knowing the net result of my article was that a few better-designed, more peer-reviewed products would ultimately emerge from the very same company in question (whether they cared to admit it or not). For sure, the winner wasn't me, certainly not some insecure small-minded manager in a hopelessly high position. It was electronics that had had its day. And that was enough for me.

Till a while ago, I had naively thought large corporations, especially those showcasing themselves in glitzy facilities headquartered in Silicon Valley, had woken up to the times and become more *professionally managed*. To me that meant things like *not* allowing race-related slurs to demoralize struggling engineers, *not* allowing chilling war-rhetoric indiscriminately sent via company E-mail (making employees wary of their own supervisor's basic sanity at times), and simply, simply, just *rewarding all efforts* fairly and without discrimination. Too much to ask! I wasn't too sure anymore that the *field* of electronics, the one that I was trying so hard to nurture, was getting even close to what it deserved. Sure, they had now started declaring "record gross margins" and so on. But behind this benumbing onslaught of pure PR, you have to remember that that their new-found exhilaration was *a)* borne mainly on the shoulders of a new breed of extremely talented, friendly and pro-active *engineers* and *b)* basically, they just *stopped* loss-making operations, in areas that were outside their "core competence" (in reality: those business units that had been so badly managed from start to finish, *that even the engineers couldn't make a difference anymore*). Further, almost without further thought, they kept laying-off several talented or promising engineers, some that I knew personally — often because their own managers had screwed up so bad they needed alibis to present before their equally bad supervisors, who needed alibis to present to

theirs …. and so on. Of course the last man standing was apparently just too busy counting the millions of dollars cash bonus he had just received for meeting the company's (short-term?) "targets". End of story. Not a tear for those engineers that were walked out one fine day a) without the slightest warning b) without even being given an opportunity to present their side of the story — quite unlike even a normal court of law anywhere in the world. I asked myself — what if Newton or Einstein had been similarly dogged by incompetent dishonest supervisors? Would the world have been a better place today? And come to think of it, how many potential Newtons and Einsteins had these companies already banned into the hell of dark obscurity, and possibly premature retirement, while chanting that their analog ICs were *nature incarnate* ("the sight and sound of information")? We never know the real casualty toll ever, do we?

As you can see, I can honestly say I have not found the solution to *Problem Number 3* yet. But I am still trying! And this book is an effort to do just that.

So now, it is time to tell you *what* exactly I have tried to achieve with this book. One unique aspect about designing power supplies is that the "devil is in the details". In other words, I, as a technical writer, can either put in everything, including the supporting Math, and come up with a book that (only) professionals would like. Or I could try making it very simple and straightforward for the beginner. But then the chances are very high I would miss out on the very essence of what power supply design is all about — the optimization, and design trade-offs. To strike a meaningful compromise between simplicity and depth requires a very carefully considered structure of presentation, one that I have really tried hard to achieve in this book. For example, several books out there, try to give a step-by-step detailed design procedure for DC-DC converters. However, they seem to routinely miss out on the important fact that the input is rarely, if ever, a "single-point" input voltage level. It is usually a "wide-range input" and we need to be very clear which converter stresses are at their worst at the highest end of the input, and which ones at the lowest input end. We also need to know which stresses we need to give priority to during a particular design step within that procedure. Clearly, designing a good power supply is not a trivial task! In Chapter 2 I have presented a universal design procedure for DC-DC converters that hopefully fulfills the simultaneous demands of rigorous detail as well as simplicity.

So what did we do in Chapter 1? That to most readers is just an introduction that they can readily skim over. Wrong! Let's take a step (and page) back. This particular introduction actually starts at the *component* level, not at the topology level as most other books do. The hope is that now, even a beginner, can understand the mysteries of a capacitor and inductor, then tie them up synergistically, to derive a switching converter topology. In fact, it will become clear that *all* topologies evolve out of a basic understanding of how, in particular, the *inductor* works. Here, advanced readers should beware. Because, while interviewing even

senior engineers for applications engineer positions, I found that many of them are still quite uncomfortable with the very concept of an inductor. Therefore, I think it is a good idea for every reader, novice or advanced, to read the book in the order of chapters presented, *starting from the very first chapter*. Just don't be caught reading it (by your *perception*-driven supervisor!). The temptation of jumping straight into an advanced chapter to "save time" may just end up slowing things down even more in the long run (besides causing avoidable bruising of self-confidence for some, along the way). *Basic concepts always need to be brought in at the right time, exemplified, and then firmed up to last a lifetime*.

In Chapter 3 I have tried to start at a fairly basic level again, but then ramped up steeply to provide one of the most detailed step-by-step procedures available for designing off-line converters and their associated magnetics. This includes the dreaded Proximity Effect analysis. I have broken up the basic procedure into two separate iterative strategies — one for foil windings and another for round windings, because their respective optimization procedures are really very different. There are also generous amounts of curves and plots thrown in to quickly help the engineer visualize and design the magnetics optimally.

I have included a chapter devoted largely to switching losses in MOSFETs, since this topic has become increasingly vital as switching frequencies are increased. But it has been presented with some of the most carefully prepared and detailed graphics probably seen in related literature — highlighting each phase of the turn-on and turn-off individually. Common simplifying assumptions have also been made whenever appropriate, and the user should thereafter have no trouble anymore practicing this rather poorly understood area of power conversion. There is also some interesting parameterized graphical information available that can come in handy either for an applications engineer selecting external MOSFETs, or an IC designer trying to optimize the driver stages of the chip.

The chapter on loop stability is likewise presented from scratch to finish, with very detailed accompanying graphics. My hope is that for the first time the reader will have easy access to almost all the equations required for loop compensation. Now, even a novice, can very quickly get very deep into this area (as I once did).

There are also seven chapters on EMI, starting from the very basics and moving up to a full mathematical treatment. This is again a topic that has been almost studiously avoided in most related literature, and yet is needed so badly today. It needs much more elaboration I thought.

To cap it all, there is an "interview-friendly" FAQ, several Mathcad files, and various design spreadsheets thrown in.

As you can see, the book has been designed to try to live up to its name "A to Z". Of course that is never really going to be possible, least of all in an all-encompassing area such as Power Conversion. But hey, I did give it a shot! The stage is now set. I hope you like

this book, *even if it is A to Z with some of the alphabets missing along the way*, and go on to make a *small* but noticeable difference, using it. Though I do strongly suggest you choose *where* you attempt to do it, because that that makes a *big* difference in the long run — to technology and to its committed practitioners: you the *engineers*. And of course, it is to *you* that this book is solely dedicated.

—Sanjaya Maniktala

Acknowledgements

It all started rather innocuously. I walked into Dr GT Murthy's office one fine day, and changed my life. "Doc" was then the General Manager, Central R&D, of a very large electrical company headquartered in Bombay. In his new state-of-the-art electronics center, he had hand-picked some of India's best engineers (over a hundred already) ever assembled under one roof. Luckily, he too was originally a Physicist, and that certainly helped me gain some empathy. Nowadays he is in retirement, but I will always remember him as a thoroughly fair, honest and facts-oriented person, who led by example. There were several things I absorbed from him that are very much part of my basic engineering persona today. You can certainly look upon this book as an extension of what Doc started many years ago in India … because that's what it really is! I certainly wouldn't be here today if I hadn't met Doc. And in fact, several of the brash, high-flying managers I've met in recent years, desperately need some sort of crash course in technology and human values from Doc!

Several people appeared a little later, aiding me along, especially when it mattered most. There was Stephan Ohr, the affable former Editor of Planet Analog who really believed in me, and gave me all my initial confidence as a writer. His personal writings too, inspired me a lot. Then there was Mike He, the former clever, pushy, and straight-talking PR person from National Semiconductor. Though Mike left rather quickly for another company, he still manages to keep in touch with me and encourage me along whenever necessary. There was also the likeable Charles Glaser ("Chuck") from Elsevier, who suddenly appeared on the scene, and believed right from the start that I could do it.

Among the technical people who reviewed major parts of this book before publication, and helped me improve it a lot is Harry Holt, a very bright senior *engineer* from National Semiconductor. Harry's eagle eye and absolutely frank technical feedback saved the day for me several times.

Other engineers I want to especially remember from my previous company, for adding immeasurably in one way or another to my peace and positive energy while writing this book, are: Linh Truong, Anne Lu, Michele Sclocchi, Thomas Mathews, Iain Mosley, Ricardo Capetillo, Maurice Eaglin, Shantha Natarajan, Jerry Zheng, Faruk Nome and Wallace Ly.
At my present company, Freescale Semiconductor, I would like to thank our brilliant product

definition team consisting of Tim Herklots and Tony Allen, besides of course my new straight-shooting (and completely sane for a change!) supervisor, Ken Lenks.

Thanks are also due to my old Marketing buddy at National Semiconductor, Ajithkumar Jain, who besides helping out with a few chapters here, had actually also helped me make up my mind years ago by describing unforgettable scenes he witnessed while on a business tour in South-Asia with the still-incumbent topmost brass — high-powered executives celebrating a successful deal, with overflowing country liquor laced with freshly poured snake blood, administered by extremely hospitable hostesses on their laps. And a multi-course dinner that all but Ajith consumed without batting an eyelid — including an entree consisting of braised puppies, their tiny tails still intact. None of this was ever intended to see the light of day. I suspect Doc wouldn't have approved of this new method of propelling analog chip technology into the future. In any case, Ajith made me realize that my future at least could never be intertwined with theirs for long.

Thanks are also due to Carl M. Soares of Elsevier, for picking up this project rather late from a suddenly departing Production Manager, and then very quietly and professionally turning it around, despite my occasional impatience.

Of course this book was completely impossible, certainly not within the time frame, were it not for the unending support and patience of my wonderful wife Disha and daughter Aartika. Not to forget the newest member of our family "Munchi". Nor the memory of dear Chippy and Monty, who gave me the strength to keep going, many many years ago. Once again, they all created the right circumstances at home for me to be able to pull this off through sleepless nights, without looking like some bleary-eyed economic refugee.

CHAPTER

1

The Principles of Switching Power Conversion

This Page Intentionally Left Blank

CHAPTER 1

The Principles of Switching Power Conversion

Introduction

Imagine we are at some busy "metro" terminus one evening at peak hour. Almost instantly, thousands of commuters swarm the station trying to make their way home. Of course there is no train big enough to carry all of them *simultaneously*. So, what do we do? Simple! We *split* this sea of humanity into several *trainloads* — and move them out in rapid succession. Many of these outbound passengers will later transfer to alternative forms of transport. So for example, trainloads may turn into bus-loads or taxi-loads, and so on. But eventually, all these "packets" will merge once again, and a throng will be seen, exiting at the destination.

Switching power conversion is remarkably similar to a mass transit system. The difference is that instead of people, it is *energy* that gets transferred from one level to another. So we draw energy continuously from an "input source," *chop* this incoming stream into packets by means of a 'switch' (a transistor), and then transfer it with the help of components (inductors and capacitors), that are able to *accommodate* these energy packets and exchange them among themselves as required. Finally, we make all these packets merge again, and thereby get a smooth and steady flow of energy into the output.

So, in either of the cases above (energy or people), from the viewpoint of an *observer*, a stream will be seen entering, and a similar one exiting. But at an *intermediate* stage, the transference is accomplished by breaking up this stream into more *manageable packets*.

Looking more closely at the train station analogy, we also realize that to be able to transfer a given number of passengers *in a given time* (note that in electrical engineering, energy transferred in unit time is 'power') — either we need bigger trains with departure times spaced relatively far apart OR several *smaller* trains leaving in *rapid* succession. Therefore, it should come as no surprise, that in switching power conversion, we always try to *switch at high frequencies*. The primary purpose for that is to *reduce the size* of the energy packets, and thereby also the *size of the components* required to store and transport them.

Power supplies that use this principle are called 'switching power supplies' or 'switching power converters.'

'Dc-dc converters' are the basic building blocks of modern high-frequency switching power supplies. As their name suggests, they 'convert' an available *dc* (direct current) input voltage rail 'V_{IN},' to another more *desirable or usable* dc output voltage level 'V_O.' 'Ac-dc

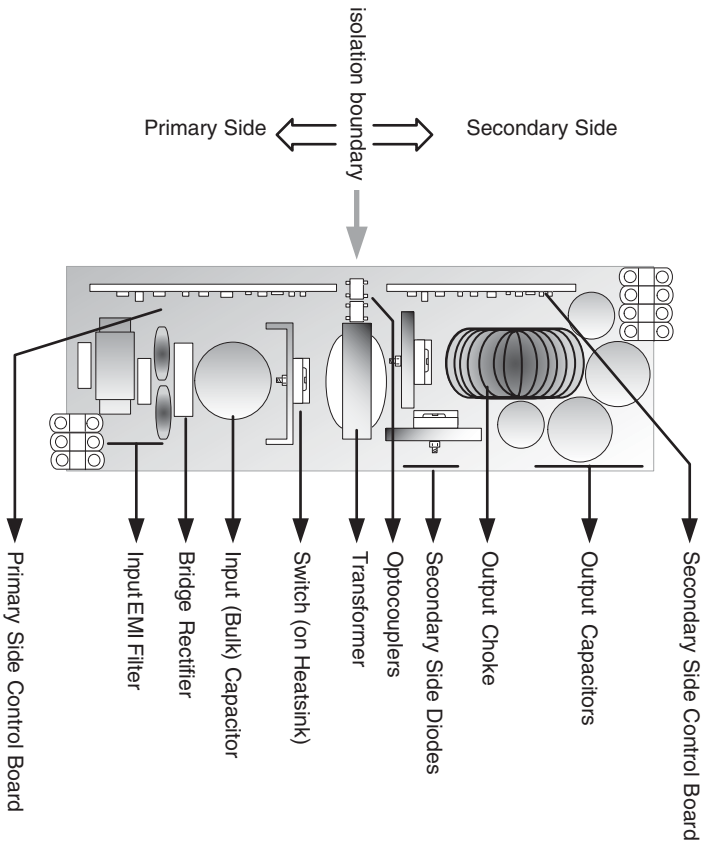


Figure 1-1: Typical Off-line Power Supply

converters' (see *Figure 1-1*), also called 'off-line power supplies,' typically run off the mains input (or 'line input'). But they first *rectify* the incoming sinusoidal ac (alternating current) voltage 'V_{AC}' to a dc voltage level (often called the 'HVDC' rail, or 'high voltage dc rail') — which then gets applied at the input of what is essentially just *another dc-dc converter stage* (or derivative thereof). We thus see that power conversion is, in essence, almost always a *dc-dc voltage conversion process*.

But it is also equally important to create a *steady* dc output voltage level, from what can often be a widely *varying* and different dc input voltage level. Therefore, a 'control circuit' is used in all power converters to constantly monitor and compare the output voltage against an internal 'reference voltage.' Corrective action is taken if the output drifts from its set value. This process is called 'output regulation' or simply 'regulation.' Hence the generic term 'voltage regulator' for supplies which can achieve this function, switching or otherwise.

In a practical implementation, 'application conditions' are considered to be the applied input voltage V_{IN} (also called the 'line voltage'), the current being drawn from the output,

that is, I_O (the 'load current') and the set output voltage V_O. Temperature is also an application condition, but we will ignore it for now, since its effect on the system is usually not so dramatic. Therefore, for a given output voltage, there are *two* specific application conditions whose variations can cause the output voltage to be immediately impacted (were it not for the control circuit). Maintaining the output voltage steady when V_{IN} varies over its stated operating range V_{INMIN} to V_{INMAX} (minimum input to maximum input), is called 'line regulation.' Whereas maintaining regulation when I_O varies over its operating range I_{OMIN} to I_{OMAX} (minimum to maximum load), is referred to as 'load regulation.' Of course, nothing is ever "perfect," so nor is the regulation. Therefore, despite the correction, there is a small but measurable change in the output voltage, which we call "ΔV_O" here. Note that mathematically, line regulation is expressed as "$\Delta V_O/V_O \times 100\%$ (implicitly implying it is over V_{INMIN} to V_{INMAX})." Load regulation is similarly "$\Delta V_O/V_O \times 100\%$" (from I_{OMIN} to I_{OMAX}).

However, the *rate* at which the output can be corrected by the power supply (under *sudden* changes in line and load) is also important — since no physical process is "instantaneous" either. So the property of any converter to provide quick regulation (correction) under external disturbances is referred to as its 'loop response.' Clearly, the loop response is as before, a combination of its 'step-load response' and its 'line transient response.'

As we move on, we will first introduce the reader to some of the most basic terminology of power conversion and its key concerns. Later, we will progress toward understanding the behavior of the most vital component of power conversion — the *inductor*. It is this component that even some relatively experienced power designers still have trouble with! Clearly, real progress in any area cannot occur without a clear understanding of the components and basic concepts involved. Therefore, only after understanding the *inductor* well, will we go on to demonstrate that switching converters themselves are not all that mysterious either — in fact they evolve quite naturally out of our newly acquired understanding of the inductor.

Overview and Basic Terminology

Efficiency

Any regulator carries out the process of power conversion with an 'efficiency,' defined as

$$\eta = \frac{P_O}{P_{IN}}$$

where P_O is the 'output power,' equal to

$$P_O = V_O \times I_O$$

and P_{IN} is the 'input power,' equal to

$$P_{IN} = V_{IN} \times I_{IN}$$

Here, I_{IN} is the *average* or dc current being drawn from the source.

Ideally we want $\eta = 1$, and that would represent a "perfect" conversion efficiency of 100%. But in a *real* converter, that is with $\eta < 1$, the *difference* '$P_{IN} - P_O$' is simply the wasted power "P_{loss}," or 'dissipation' (occurring within the converter itself). By simple manipulation we get

$$P_{loss} = P_{IN} - P_O$$

$$P_{loss} = \frac{P_O}{\eta} - P_O$$

$$P_{loss} = P_O \times \left(\frac{1-\eta}{\eta}\right)$$

This is the loss expressed in terms of the output power. In terms of the input power we would similarly get

$$P_{loss} = P_{IN} \times (1-\eta)$$

The loss manifests itself as *heat* in the converter, which in turn causes a certain measurable 'temperature rise' ΔT over the surrounding 'room temperature' (or 'ambient temperature'). Note that high temperatures affect the *reliability* of all systems — the rule-of-thumb being that every 10°C rise causes the failure rate to double. Therefore, part of our skill as designers is to reduce this temperature rise, and thereby also achieve higher efficiencies.

Coming to the *input current* (drawn by the converter), for the hypothetical case of 100% efficiency, we get

$$I_{IN\_ideal} = I_O \times \left(\frac{V_O}{V_{IN}}\right)$$

So, in a real converter, the input current increases from its "ideal" value by the factor $1/\eta$.

$$I_{IN\_measured} = \frac{1}{\eta} \times I_{IN\_ideal}$$

Therefore, if we can achieve a high efficiency, the current drawn from the input (keeping application conditions unchanged) will decrease — *but only up to a point*. The input current clearly cannot fall below the "brickwall" that is "I_{IN_ideal}," because this current is

equal to P_O/V_{IN} — that is, related only to the 'useful power' P_O, delivered by the power supply, which we are assuming has not changed.

Further, since

$$V_O \times I_O = V_{IN} \times I_{IN\_ideal}$$

by simple algebra, the dissipation in the power supply (energy lost per second as heat) can also be written as

$$P_{loss} = V_{IN} \times \left(I_{IN\_measured} - I_{IN\_ideal}\right)$$

This form of the dissipation equation indicates a little more explicitly how *additional* energy (more input current for a given input voltage) is pushed into the input terminals of the power supply by the applied dc source — to compensate for the wasted energy inside the power supply — even as the converter continues to provide the useful energy P_O being constantly demanded by the load.

A modern switching power supply's efficiency can typically range from 65 to 95% — that figure being considered attractive enough to have taken switchers to the level of interest they arouse today, and their consequent wide application. *Traditional regulators* (like the 'linear regulator') provide much poorer efficiencies — and that is the main reason why they are slowly but surely getting replaced by switching regulators.

Linear Regulators

'Linear regulators,' equivalently called 'series-pass regulators,' or simply 'series regulators,' also produce a regulated dc output rail from an input rail. But they do this by placing a transistor in series between the input and output. Further, this 'series-pass transistor' (or 'pass-transistor') is operated in the *linear* region of its voltage-current characteristics — thus acting like a variable *resistance* of sorts. As shown in the uppermost schematic of *Figure 1-2*, this transistor is made to literally "drop" (abandon) the unwanted or "excess" voltage across itself.

The excess voltage is clearly just the difference '$V_{IN} - V_O$' — and this term is commonly called the 'headroom' of the linear regulator. We can see that the headroom needs to be a *positive* number always, thus implying $V_O < V_{IN}$. Therefore, linear regulators are, in principle, always 'step-down' in nature — that being their most obvious limitation.

In some applications (e.g. battery powered portable electronic equipment), we may want the output rail to remain well-regulated even if the input voltage dips very low — say down to within *0.6 V or less* of the set output level V_O. In such cases, the *minimum possible headroom* (or 'dropout') achievable by the linear regulator stage may become an issue.

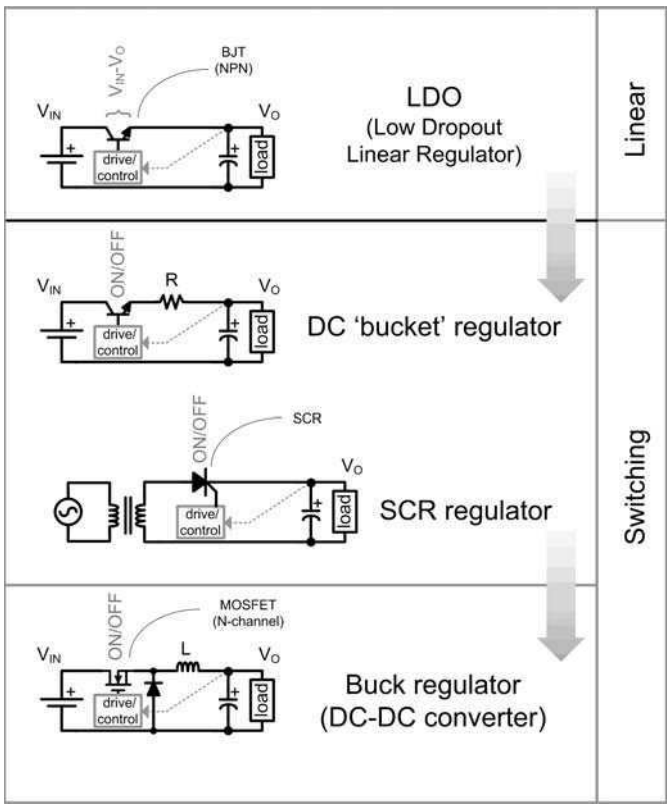


Figure 1-2: Basic Types of Linear and Switching Regulators

No switch is perfect, and even if held fully conducting, it does have some voltage drop across it. So the dropout is simply the minimum achievable 'forward-drop' across the switch. Regulators which can continue to work (i.e. regulate their output), with V_{IN} barely exceeding V_O, are called '*low* dropout' regulators, or 'LDOs.' But note that there is really no precise voltage drop at which a linear regulator "officially" becomes an LDO. So the term is sometimes applied rather loosely to linear regulators in general. However, the rule-of thumb is that a dropout of about 200 mV or lower qualifies as an LDO, whereas older devices (conventional linear regulators) have a typical dropout voltage of around 2 V. There is also an intermediate category, called 'quasi-LDOs' that have a dropout of about 1 V, that is, somewhere in between the two.

Besides being step-down in principle, linear regulators have another limitation — poor efficiency. Let us understand why that is so. The instantaneous power dissipated in any device is by definition the cross-product V × I, where V is the instantaneous voltage drop across it and I the instantaneous current through it. In the case of the series-pass transistor, under steady application conditions, both V and I are actually constant with respect to

time — V in this case being the headroom $V_{IN} - V_O$, and I the load current I_O (since the transistor is always in *series* with the load). So we see that the V × I dissipation term for linear regulators can, under certain conditions, become a significant proportion of the useful output power P_O. And that simply spells "*poor efficiency*"! Further, if we stare hard at the equations, we will realize there is also nothing we can do about it — how can we possibly argue against something as basic as V × I? For example, if the input is 12 V, and the output is 5 V, then at a load current of 100 mA, the dissipation in the regulator is necessarily $\Delta V \times I_O = (12 - 5)$ V × 100 mA = 700 mW. The useful (output) power is however $V_O \times I_O = 5$ V × 100 mA = 500 mW. Therefore, the efficiency is $P_O/P_{IN} = 500/(700 + 500) = 41.6\%$. What can we do about that?!

On the positive side, linear regulators are very "quiet" — exhibiting none of the noise and EMI (electromagnetic interference) that have unfortunately become a "signature" or "trademark" of modern switching regulators. Switching regulators need *filters* — usually both at the input and the output, to quell some of this noise, which can interfere with other gadgets in the vicinity, possibly causing them to malfunction. Note that sometimes, the usual input/output capacitors of the converter may themselves serve the purpose, especially when we are dealing with 'low-power' (and 'low-voltage') applications. But in general, we may require filter stages containing *both* inductors and capacitors. Sometimes these stages may need to be cascaded to provide even greater noise attenuation.

Achieving High Efficiency through Switching

Why are switchers so much more efficient than "linears"?

As their name indicates, in a *switching* regulator, the series transistor is not held in a perpetual *partially conducting* (and therefore dissipative) mode — but is instead *switched* repetitively. So there are only two *states* possible — either the switch is held 'ON' (fully conducting) or it is 'OFF' (fully non-conducting) — there is no "middle ground" (at least not in principle). When the transistor is ON, there is (ideally) zero *voltage* across it (V = 0), and when it is OFF we have zero *current* through it (I = 0). So it is clear that the cross-product 'V × I' is also zero for *either* of the two states. And that simply implies zero 'switch dissipation' at all times. Of course this too represents an impractical or "ideal" case. Real switches do dissipate. One reason for that they are never either *fully* ON nor *fully* OFF. Even when they are supposedly ON, they have a small voltage drop across them, and when they are supposedly "OFF," a small current still flows through them. Further, no device switches "instantly" either — there is a always definable period in which the device is *transiting between states*. During this interval too, V × I is not zero, and some additional dissipation occurs.

We may have noticed that in most introductory texts on switching power conversion, the switch is shown as a *mechanical* device — with contacts that simply open ("switch OFF")

or close ("switch ON"). So a mechanical device comes very close to our definition of a "perfect switch" — and that is the reason why it is often the vehicle of choice to present the most basic principles of power conversion. But one obvious problem with actually *using* a mechanical switch in any practical converter is that such switches can wear out and fail over a relatively short period of time. So in practice, we always prefer to use a *semiconductor device* (e.g. a transistor) as the switching element. As expected, that greatly enhances the life and reliability of the converter. But the most important advantage is that since a semiconductor switch has none of the mechanical "inertia" associated with a mechanical device, it gives us the ability to switch *repetitively* between the ON and OFF states — and do so *very fast*. We have already realized that that will lead to *smaller* components in general.

We should be clear that the phrase "switching fast," or "high switching speed," has slightly varying connotations, even within the area of switching power conversion. When it is applied to the overall *circuit*, it refers to the frequency at which we are repeatedly switching — ON OFF ON OFF and so on. This is the converter's basic *switching frequency* 'f' (in Hz). But when the same term is applied specifically to the *switching element* or device, it refers to the *time* spent transiting *between* its two states (i.e. from ON to OFF and OFF to ON), and is typically expressed in 'ns' (nanoseconds). Of course this transition interval is then rather *implicitly* and intuitively being compared to the total 'time period' T (where $T = 1/f$), and therefore to the switching frequency — though we should be clear there is no *direct* relationship between the transition time and the switching frequency.

We will learn shortly that the ability to *crossover* (i.e. transit) quickly between switching states is in fact rather crucial. Yes, up to a point, the switching speed is almost completely determined by how "strong" and effective we can make our external 'drive circuit.' But ultimately, the speed becomes limited purely by the device and its technology — an "inertia" of sorts at an electrical level.

Basic Types of Semiconductor Switches

Historically, most power supplies used the 'bjt' (bipolar junction transistor) shown in *Figure 1-2*. It is admittedly a rather *slow* device by modern standards. But it is still relatively cheap! In fact its 'npn' version is even cheaper, and therefore more popular than its 'pnp' version. Modern switching supplies prefer to use a 'mosfet' (metal oxide semiconductor field effect transistor), often simply called a 'fet' (see *Figure 1-2* again). This modern high-speed switching device also comes in several "flavors" — the most commonly used ones being the *n-channel* and *p-channel* types (both usually being the 'enhancement mode' variety). The *n-channel mosfet* happens to be the favorite in terms of cost-effectiveness and performance, for most applications. However, sometimes, p-channel devices may be preferred for various reasons — mainly because they usually require simpler drive circuits.

Despite the steady course of history in favor of mosfets in general, there still remain some arguments for continuing to prefer bjts in certain applications. Some points to consider and debate here are:

a) It is often said that it is *easier to drive a mosfet than a bjt*. In a bjt we do need a large drive current (injected into its 'base' terminal) — to turn it ON. We also need to *keep* injecting base current to *keep it* in that state. On the other hand, a mosfet is considered easier to drive. In theory, we just have to apply a certain voltage at its 'gate' terminal to turn it ON, and also keep it that way. Therefore, a mosfet is called a 'voltage-controlled' device, whereas a bjt is considered a 'current controlled' device. However, in reality, a modern mosfet needs a certain amount of gate current *during* the time it is *in transit* (ON to OFF and OFF to ON). Further, to make it change state *fast*, we may in fact need to push in (or pull) out a *lot* of current (typically 1 to 2 A).

b) The *drive requirements of a bjt may actually turn out easier to implement* in many cases. The reason for that is, to turn an npn bjt ON for example, its gate has to be taken only about 0.8 V above its emitter (and can even be tied directly to its collector on occasion). Whereas, in an n-channel mosfet, its gate has to be taken *several volts* higher than its source. Therefore, in certain types of dc-dc converters, when using an n-channel mosfet, it can be shown that we need a 'drive rail' that is significantly *higher* than the (available) input rail V_{IN}. And how else can we hope to have such a rail except by a circuit that can somehow manage to "push" or "pump" the input voltage to a higher level? When thus implemented, such a rail is called the 'bootstrap' rail.

   Note: The most obvious implementation of a 'bootstrap circuit' may just consist of a small capacitor that gets charged by the input source (through a small signal diode) whenever the switch turns OFF. Thereafter, when the switch turns ON, we know that certain voltage nodes in the power supply suddenly "flip" whenever the switch changes state. But since the 'bootstrap capacitor' continues to hold on to its acquired voltage (and charge), it automatically pumps the bootstrap rail to a level higher than the input rail, as desired. This rail then helps drive the mosfet properly under all conditions.

c) The main advantage of bjts is that they are known to generate *significantly less EMI and 'noise and ripple'* than mosfets. That ironically is a positive outcome of their *slower* switching speed!

d) Bjts are also often better suited for *high-current* applications — because their 'forward drop' (*on-state voltage drop*) is *relatively constant*, even for very high switch currents. This leads to significantly lower 'switch dissipation,' more so when the switching frequencies are not too high. On the contrary, in a mosfet, the forward drop is almost proportional to the current passing through it — so its dissipation can become significant at high loads. Luckily, since it also switches faster (lower transition times), it usually more than makes up, and so in fact becomes much better in terms of the *overall* loss — more so when operated at very high switching frequencies.

Note: In an effort to combine the "best of both worlds," a "combo" device called the 'IGBT' (insulated gate bipolar transistor) is also often used nowadays. It is driven like a mosfet (voltage-controlled), but behaves like a bjt in other ways (the forward drop and switching speed). It too is therefore suited mainly for low-frequency and high-current applications, but is considered easier to drive than a bjt.

Semiconductor Switches Are Not "Perfect"

We mentioned that all semiconductor switches suffer losses. Despite their advantages, they are certainly not the perfect or ideal switches we may have imagined them to be at first sight.

So for example, unlike a mechanical switch, in the case of a semiconductor device, we may have to account for the small but measurable 'leakage current' flowing through it when it is considered "fully OFF" (i.e. non-conducting). This gives us a dissipation term called the 'leakage loss.' This term is usually not very significant and can be ignored. However, there is a small but *significant* voltage drop ('forward drop') across the semiconductor when it is considered "fully ON" (i.e. conducting) — and that gives us a significant 'conduction loss' term. In addition, there is also a brief moment as we transition *between* the two switching states, when the current and voltage in the switch need to *slew up or down almost simultaneously* to their new respective levels. So, during this 'transition time' or 'crossover time,' we *neither have* $V = 0$ *nor* $I = 0$ instantaneously, and therefore nor is $V \times I = 0$. This therefore leads to some additional dissipation, and is called the 'crossover loss' (or sometimes just 'switching loss'). Eventually, we need to learn to minimize all such loss terms if we want to improve the efficiency of our power supply.

However, we must remember that power supply design is by its very nature full of *design tradeoffs* and subtle compromises. For example, if we look around for a transistor with a very low forward voltage drop, possibly with the intent of minimizing the conduction loss, we usually end up with a device that also happens to transition more slowly — thus leading to a higher crossover loss. There is also an overriding concern for cost that needs to be constantly looked into, particularly in the commercial power supply arena. So, we should not underestimate the importance of having an astute and seasoned engineer at the helm of affairs, one who can really grapple with the finer details of power supply design. As a corollary, neither can we probably ever hope to replace him or her (at least not entirely), by some smart automatic test system, nor by any "expert design software" that we may have been dreaming of.

Achieving High Efficiency through the Use of Reactive Components

We have seen that one reason why switching regulators have such a high efficiency is because they use a *switch* (rather than a transistor that "thinks" it is a resistor, as in an LDO). Another root cause of the high efficiency of modern switching power supplies is their effective use of *both capacitors and inductors*.

Capacitors and inductors are categorized as 'reactive' components because they have the unique ability of being able to *store energy*. However, that is why they cannot ever be made to *dissipate* any energy either (at least not *within* themselves) — they just store any energy "thrown at them"! On the other hand, we know that 'resistive' components dissipate energy, but unfortunately, can't store any!

A capacitor's stored energy is called *electrostatic*, equal to $\frac{1}{2} \times C \times V^2$ where C is the 'capacitance' (in Farads), and V the voltage across the capacitor. Whereas an inductor stores *magnetic* energy, equal to $\frac{1}{2} \times L \times I^2$, L being the 'inductance' (in Henries) and I the current passing through it (at any given moment).

But we may well ask — despite the obvious efficiency concerns, do we really *need* reactive components *in principle*? For example, we may have realized we don't really need an input or output capacitor for implementing a *linear regulator* — because the series-pass element is all that is required to block any excess voltage. For switching regulators however, the reasoning is rather different. This leads us to the general "*logic of switching power conversion*" summarized below.

- A transistor is needed to establish control on the output voltage, and thereby bring it into regulation. The reason we *switch* it is as follows — dissipation in this control element is related to the product of the voltage across the control device and the current through it, that is $V \times I$. So if we make either V or I zero (or very small), we will get zero (or very small) dissipation. By switching *constantly* between ON and OFF states, we can keep the switch dissipation down, but at the same time, by controlling the *ratio* of the ON and OFF intervals, we can *regulate* the output, based on average energy flow considerations.

- But whenever we switch the transistor, we effectively *disconnect the input from the output* (during either the ON or OFF state). However, the output (load) always demands a *continuous* flow of energy. Therefore we need to introduce *energy storage* elements somewhere inside the converter. In particular, we use output *capacitors* to "hold" the voltage steady across the load during the above-mentioned input-to-output "disconnect" interval.

- But as soon as we put in a capacitor, we now also need to limit the *inrush current* into it — all capacitors connected directly across a dc source, will exhibit this uncontrolled inrush — and that can't be good either for noise, EMI, or for efficiency. Of course we could simply opt for a *resistor* to subdue this inrush, and that in fact was the approach behind the early "bucket regulators" (*Figure 1-2*).

- But unfortunately a resistor always *dissipates* — so what we may have saved in terms of switch dissipation, may ultimately end up in the resistor! To maximize the overall efficiency, we therefore need to *use only reactive* elements in the conversion

process. Reactive elements can *store* energy but do not dissipate any (in principle). Therefore, an *inductor becomes our final choice* (along with the capacitor), based on its ability to *non-dissipatively limit the (rate of rise) of current*, as is desired for the purpose of limiting the capacitor inrush current.

Some of the finer points in this summary will become clearer as we go on. We will also learn that *once the inductor has stored some energy, we just can't* ***wish*** *this stored energy away at the drop of a hat"*. We need to do something about it! And that in fact gives us an actual working converter down the road.

Early RC-based Switching Regulators

As indicated above, a possible way out of the "input-to-output disconnect" problem is to use *only an output capacitor*. This can store some extra energy when the switch connects the load to the input, and then provide this energy to the load when the switch disconnects the load.

But we still need to limit the capacitor *charging current* ('inrush current'). And as indicated, we could use a resistor. That was in fact the basic principle behind some early linear-to-switcher "crossover products" like the 'bucket regulator' shown in *Figure 1-2*.

The bucket regulator uses a transistor driven like a *switch* (as in modern switching regulators), a small *series resistor* to limit the current (not entirely unlike a linear regulator), and an *output capacitor* (the "bucket") to store and then provide energy when the switch is OFF. Whenever the output voltage falls below a certain threshold, the switch turns ON, "tops up" the bucket, and then turns OFF. Another version of the bucket regulator uses a cheap low-frequency switch called an SCR ('semiconductor controlled rectifier') that works off the secondary windings of a step-down transformer connected to an ac mains supply, as also shown in *Figure 1-2*. Note that in this case, the resistance of the windings (usually) serves as the (only) effective limiting resistance.

Note also that in either of these RC-based bucket regulator implementations, the switch ultimately ends up being *toggled repetitively* at a certain rate — and in the process, a rather crudely regulated stepped down output dc rail is created. By definition, that makes these regulators *switching regulators* too!

But we realize that the very use of a resistor in any power conversion process always bodes ill for efficiency. So, we may have just succeeded in *shifting* the dissipation away from the transistor — into the resistor! If we really want to maximize overall efficiency, we need to do away with *any intervening resistance* altogether.

So we attempt to use an inductor instead of a resistor for the purpose — we don't really have many other component choices left in our bag! In fact, if we manage to do that, we get our first modern *LC-based switching regulator* — the 'buck regulator' (i.e. step-down converter), as also presented in *Figure 1-2*.

LC-based Switching Regulators

Though the detailed functioning of the modern buck regulator of *Figure 1-2* will be explained a little later, we note that besides the obvious replacement of R with an L, it looks very similar to the bucket regulator — *except for a "mysterious" diode*. The basic principles of power conversion will in fact become clear only *when we realize the purpose of this diode*. This component goes by several names — 'catch diode,' 'freewheeling diode,' 'commutation diode,' and 'output diode,' to name a few! But its basic purpose is always the same — a purpose we will soon learn is intricately related to the behavior of the *inductor* itself.

Aside from the buck regulator, there are *two* other ways to implement the basic goal of switching power conversion (using *both inductors and capacitors*). Each of these leads to a distinct *'topology.'* So besides the *buck* (step-down), we also have the *'boost'* (step-up), and the *'buck-boost'* (step-up or step-down). We will see that though all these are based on the same *underlying* principles, they are set up to look and behave quite differently. As a prospective power supply designer, we really do need to learn and master each of them almost on an *individual basis*. We must also keep in mind that in the process, our *mental picture will usually need a drastic change as we go from one topology to another*.

> **Note:** There are some other capacitor-based possibilities — in particular 'charge pumps' — also called 'inductor-less switching regulators.' These are usually restricted to rather low powers and produce output rails that are rather crudely regulated *multiples* of the input rail. In this book, we are going to ignore these types altogether.
>
> Then there are also some other types of LC-based possibilities — in particular the 'resonant topologies.' Like conventional dc-dc converters, these also use both types of reactive components (L and C) along with a switch. However, their basic principle of operation is very different. Without getting into their actual details, we note that these topologies do *not* maintain *a constant switching frequency, which is something we usually rather strongly desire*. From a practical standpoint, any switching topology with a variable switching frequency, can lead to an *unpredictable* and varying EMI spectrum and noise signature. To mitigate these effects, we may require rather complicated filters. For such reasons, resonant topologies have not really found widespread acceptance in commercial designs, and so we too will largely ignore them from this point on.

The Role of Parasitics

In using conventional LC-based switching regulators, we may have noticed that their constituent inductors and capacitors do get fairly *hot* in most applications. But if, as we said, these components are *reactive*, why at all are they getting hot? We need to know why, because any source of heat impacts the overall efficiency! And *efficiency* is what modern switching regulators are all about!

The heat arising from *real-world* reactive components can invariably be traced back to dissipation occurring within the small 'parasitic' *resistive* elements, which always accompany any such (reactive) component.

For example, a real inductor has the basic property of inductance L, but it also has a certain non-zero *dc resistance* ('DCR') term, mainly associated with the copper windings used. Similarly, any real capacitor has a capacitance C, but it also has a small *equivalent series resistance* ('ESR'). Each of these terms produces 'ohmic' losses — that can all add up and become fairly significant.

As indicated previously, a real-world semiconductor switch can also be considered as having a parasitic resistance "strapped" *across* it. This parallel resistor in effect "models" the leakage current path, and thus the 'leakage loss' term. Similarly, the forward drop across the device can also, in a sense, be thought of as a *series* parasitic resistance — leading to a conduction loss term.

But any real-world component also comes along with various *reactive* parasitics. For example an inductor can have a significant parasitic capacitance across its terminals — associated with electrostatic effects between the layers of its windings. A capacitor can also have an *equivalent series inductance* ('ESL') — coming from the small inductances associated with its leads, foil, and terminations. Similarly, a mosfet also has various parasitics — for example the "unseen" capacitances present *between* each of its terminals (within the package). In fact, these mosfet parasitics play a major part in determining the limits of its switching speed (transition times).

In terms of dissipation, we understand that reactive parasitics certainly cannot dissipate heat — at least not *within the parasitic element itself*. But more often than not, these reactive parasitics do manage to "dump" their stored energy (at specific moments during the switching cycle) into a nearby *resistive* element — thus increasing the overall losses indirectly.

Therefore we see that to improve efficiency, we generally need to go about minimizing all such parasitics — *resistive or reactive*. We should not forget they are the very reason we are not getting 100% efficiency from our converter in the first place. Of course, we have to learn to be able to do this optimization to within *reasonable* and *cost-effective* bounds, as dictated by market compulsions and similar constraints.

But we should also bear in mind that *nothing is so straightforward in power*! So these parasitic elements should not be considered entirely "useless" either. In fact they do play a rather helpful and stabilizing role on occasion.

- For example, if we short the outputs of a dc-dc converter, we know it is unable to regulate, however hard it tries. In this 'fault condition' ('open-loop'), the momentary 'overload current' within the circuit can be "tamed" (or mitigated) a great deal by the very presence of certain identifiably "friendly" parasitics.
- We will also learn that the so-called 'voltage-mode control' switching regulators actually *rely on the ESR* of the output capacitor for ensuring 'loop stability' — even under normal operation. As indicated previously, loop stability refers to the ability of

a power supply to regulate its output quickly, when faced with sudden changes in line and load, without undue oscillations or ringing.

Certain other parasitics however may just prove to be a nuisance and some others a sheer bane. But their actual *roles too may keep shifting*, depending upon the prevailing conditions in the converter. For example

- A certain parasitic *inductance* may be quite *helpful* during the *turn-on* transition of the switch — by acting to limit any current spike trying to pass through the switch. But it can be *harmful* due to the high *voltage spike* it creates across the switch at *turn-off* (as it tries to release its stored magnetic energy).
- On the other hand, a parasitic *capacitance* present across the switch for example, can be *helpful* at *turn-off* — but *unhelpful* at *turn-on*, as it tries to dump its stored electrostatic energy inside the switch.

  Note: We will find that during turn-off, the parasitic capacitance mentioned above helps limit or 'clamp' any potentially destructive voltage spikes appearing across the switch, by *absorbing* the energy residing in that spike. It also helps *decrease the crossover loss* by slowing down the rising ramp of voltage, and thereby reducing the V-I "overlap" (between the transiting V and I waveforms of the switch). However at *turn-on*, the same parasitic capacitance now has to discharge whatever energy it acquired during the preceding turn-off transition — and that leads to a *current spike* inside the switch. Note that this spike is externally "invisible" — apparent only by the higher-than-expected switch dissipation, and the resulting higher-than-expected temperature.

Therefore, generally speaking, *all parasitics constitute a somewhat* "*double-edged sword*," one that we just can't afford to overlook for very long in *practical* power supply design. However, as we too will do in some of our discussions that follow, sometimes we can consciously and *selectively* decide to ignore some of these second-order influences initially, just to build up *basic concepts* in power first. Because the truth is if we don't do that, we just run the risk of feeling quite overwhelmed, too early in the game!

Switching at High Frequencies

In attempting to generally reduce parasitics and their associated losses, we may notice that these are often dependent on various external factors — *temperature* for one. Some losses increase with temperature — for example the conduction loss in a mosfet. And some may decrease — for example the conduction loss in a bjt *(when operated with low currents)*. Another example of the latter type is the ESR-related loss of a typical aluminum electrolytic capacitor, which also decreases with temperature. On the other hand, some losses may have rather "strange" shapes. For example, we could have an inverted "bell-shaped" curve — representing an optimum operating point somewhere *between the two extremes*. This is what the 'core loss' term of many modern 'ferrite' materials (used for inductor cores) looks like — it is at its minimum at around 80 to 90°C, increasing on either side.

From an overall perspective, it is hard to predict how all these variations with respect to temperature add up — and how the efficiency of the power supply is thereby affected by changes in temperature.

Coming to the dependency of parasitics and related loss terms on *frequency*, we do find a somewhat clearer trend. In fact it is rather rare to find any loss term that *decreases* at higher frequencies (though a notable exception to this is the loss in an aluminum electrolytic capacitor — because its ESR decreases with frequency). Some of the loss terms are virtually *independent* of frequency (e.g. conduction loss). And the remaining losses actually *increase* almost *proportionally* to the switching frequency — for example, the crossover loss. So in general, we realize that *lowering, not increasing, the switching frequency would almost invariably help improve efficiency*.

There are other frequency-related issues too, besides efficiency. For example, we know that switching power supplies are inherently noisy, and generate a lot of EMI. By going to higher switching frequencies, we may just be making matters worse. We can mentally visualize that even the small connecting wires and 'printed circuit board' (PCB) traces become very effective antennas at high frequencies, and will likely spew out *radiated EMI* in every direction.

This therefore begs the question: ***why at all are we face to face with a modern trend of ever-increasing switching frequencies?*** Why should we not *decrease* the switching frequency?

The first motivation toward higher switching frequencies was to simply take "the action" beyond audible human hearing range. Reactive components are prone to creating sound pressure waves for various reasons. So, the early LC-based switching power supplies switched at around 15–20 kHz, and were therefore barely audible, if at all.

The next impetus toward even higher switching frequencies came with the realization that the bulkiest component of a power supply, that is, *the inductor, could be almost proportionately reduced in size if the switching frequency was increased* (everybody does seem to want smaller products, after all!). Therefore, successive generations of power converters moved upward in almost arbitrary steps, typically 20 kHz, 50 kHz, 70 kHz, 100 kHz, 150 kHz, 250 kHz, 300 kHz, 500 kHz, 1 MHz, 2 MHz, and often even higher today. This actually helped simultaneously *reduce* the size of the *conducted EMI* and input/output filtering components — including the capacitors! High switching frequencies can also almost proportionately enhance the *loop response* of a power supply.

Therefore, we realize that the only thing holding us back at any moment of time from going to even higher frequencies are the "*switching losses*." This term is in fact rather broad —- encompassing all the losses that occur *at the moment* when we actually switch the transistor (i.e. from ON to OFF and/or OFF to ON). Clearly, the *crossover loss* mentioned earlier is

just one of several possible switching loss terms. Note that it is easy to visualize why such losses are (usually) exactly proportional to the switching frequency — since energy is lost *only whenever we actually switch* — therefore, the greater the number of times we do that (in a second), the more energy is lost (dissipation).

Finally, we also do need to learn how to *manage* whatever dissipation is still remaining in the power supply. This is called 'thermal management,' and that is one of the most important goals in any good power supply design. Let us look at that now.

Reliability, Life, and Thermal Management

Thermal management basically just means trying to get the heat out from the power supply and into the surroundings — thereby lowering the local temperatures at various points inside it. The most basic and obvious reason for doing this is to keep all the components to within their maximum rated operating temperatures. But in fact, that is rarely enough. We always strive to reduce the temperatures even further, and every couple of *degrees Celsius* (°C) may well be worth fighting for.

The reliability 'R' of a power supply at any given moment of time is defined as $R(t) = e^{-\lambda t}$. So at time $t = 0$ (start of operational life), the reliability is considered to be at its maximum value of 1. Thereafter it decreases exponentially as time elapses. 'λ' is the failure rate of a power supply, that is, the number of supplies failing over a specified period of time. Another commonly used term is 'MTBF,' or *mean time between failures*. This is the reciprocal of the overall failure rate, that is, $\lambda = 1/\text{MTBF}$. A typical commercial power supply will have an MTBF of between 100,000 hours to 500,000 hours — *assuming it is being operated at a fairly typical and benign 'ambient temperature' of around 25°C*.

Looking now at the variation of failure rate with respect to temperature, we come across the well-known rule-of-thumb — *failure rate doubles every 10°C rise in temperature*. If we apply this admittedly loose rule-of-thumb to each and every component used in the power supply, we see it must also hold for the entire power supply too — since the overall failure rate of the power supply is simply the sum of the failure rates of each component comprising it ($\lambda = \lambda_1 + \lambda_2 + \lambda_3 + \ldots$). All this clearly gives us a good reason to try and reduce temperatures of *all* the components even further.

But aside from failure rate, which clearly applies to *every* component used in a power supply, there are also certain 'lifetime' considerations that apply to *specific* components. The 'life' of a component is stated to be the duration it can work for continuously, without *degrading* beyond certain specified limits. At the end of this 'useful life,' it is considered to have become a 'wearout failure' — or simply put — it is "worn-out." Note that this need not imply the component has failed "catastrophically" — more often than not, it may be just "out of spec." The latter phrase simply means the component no longer provides the

expected performance — as specified by the limits published in the electrical tables of its datasheet.

> **Note:** Of course a datasheet can always be "massaged" to make the part look good in one way or another — and that is the origin of a rather shady but widespread industry practice called "specmanship." A good designer will therefore keep in mind that not all vendors' datasheets are equal — even for what may seem to be the same or equivalent part number at first sight.

As designers, it is important that we not only do our best to extend the 'useful life' of any such component, but also account *upfront* for its slow degradation over time. In effect, that implies that the power supply may *initially* perform better than its minimum specifications. Ultimately however, the worn-out component, especially if it is present at a critical location, could cause the entire power supply to "go out of spec," and even fail catastrophically.

Luckily, most of the components used in a power supply have no meaningful or definable lifetime — at least not within the usual 5 to 10 years of useful life expected from most electronic products. We therefore usually don't, for example, talk in terms of an inductor or transistor "degrading" (over a period of time) — though of course either of these components can certainly *fail* at any given moment, even under normal operation, as evidenced by their non-zero failure rates.

> **Note:** Lifetime issues related to the *materials* used in the construction of a component can affect the life of the component indirectly. For example, if a semiconductor device is operated *well beyond its usual maximum rating of 150°C*, its *plastic package* can exhibit wearout or degradation — even though nothing happens to the semiconductor itself up to a much higher temperature. Subsequently, over a period of time, this degraded package can cause the junction to get severely affected by environmental factors, causing the device to fail catastrophically — usually taking the power supply (and system) with it too! In a similar manner, inductors made of a 'powdered iron' type of core material are also known to degrade under extended periods of high temperatures — and this can produce not only a failed inductor, but a failed power supply too.

A common example of lifetime considerations in a commercial power supply design comes from its use of aluminum electrolytic capacitors. Despite their great affordability and respectable performance in many applications, such capacitors are a victim of wearout due to the steady evaporation of their enclosed electrolyte over time. Extensive calculations are needed to predict their *internal* temperature ('core temperature') and thereby estimate the true rate of evaporation and thereby extend the capacitor's useful life. The rule recommended for doing this life calculation is — *the useful life of an aluminum electrolytic capacitor halves every 10°C rise in temperature.* We can see that this relatively hard-and-fast rule is uncannily similar to the rule-of-thumb of failure rate. But that again is just a coincidence, since life and failure rate are really two different issues altogether.

In either case, we can now clearly see that the way to extend life *and* improve reliability is to *lower the temperatures of all the components in a power supply* and also the *ambient temperature inside the enclosure of the power supply.* This may also call out for a better-ventilated enclosure (more air vents), more exposed copper on the PCB (printed

circuit board), or say, even a built-in fan to push the hot air out. Though in the latter case, we now have to start worrying about both the failure rate and life of the fan itself!

Stress Derating

Temperature can ultimately be viewed as a 'thermal stress' — one that causes an increase in failure rate (and life if applicable). But how severe a stress really is, must naturally be judged *relative to the 'ratings' of the device*. For example, most semiconductors are rated for a 'maximum junction temperature' of 150°C. Therefore, keeping the junction no higher than 105°C in a given application represents a *stress reduction factor*, or alternately — a 'temperature derating' factor equal to 105/150 = 70%.

In general, 'stress derating' is the established technique used by good designers to diminish internal stresses and thereby reduce the failure rate. Besides temperature, the failure rate (and life) of any component can also depend on the applied *electrical* stresses — voltage and current. For example, a typical 'voltage derating' of 80% as applied to semiconductors means that the worst-case operating voltage across the component never exceeds 80% of the maximum specified voltage rating of the device. Similarly, we can usually apply a typical 'current derating' of 70–80% to most semiconductors.

The practice of derating also implies that we need to select our components judiciously *during the design phase itself* — with well-considered and built-in operating *margins*. And though, as we know, some loss terms decrease with temperature, contemplating raising the temperatures just to achieve better efficiency or performance is clearly not the preferred direction, because of the obvious impact on system reliability.

A good designer eventually learns to *weigh reliability and life concerns against cost, performance, size, and so on*.

Advances in Technology

But despite the best efforts of many a good power supply designer, certain sought after improvements may still have remained merely on our annual Christmas wish list! Luckily, there have been significant accompanying advances in the technology of the *components* available, to help enact our goals. For example, the burning desire to reduce resistive losses and simultaneously make designs suitable for high frequency operation has ushered in significant improvements in terms of a whole new generation of high-frequency, low-ESR ceramic and other specialty capacitors. We also have diodes with very low forward voltage drops and 'ultra-fast recovery,' much faster switches like the mosfet, and several new low-loss ferrite material types for making the transformers and inductors.

> **Note:** 'Recovery' refers to the ability of a diode to quickly change from a conducting state to a non-conducting (i.e. 'blocking') state as soon as the voltage across it reverses. Diodes which do this well are

called 'ultrafast diodes.' Note that the 'Schottky diode' is preferred in certain applications, because of its low forward drop (~0.5 V). *In principle*, it is also supposed to have zero recovery time. But unfortunately, it also has a comparatively higher parasitic 'body capacitance' (across itself), that in some ways tends to mimic conventional recovery phenomena. Note that it also has a higher leakage current and is typically limited to blocking voltages of less than 100 V.

However we observe that the actual *topologies* used in power conversion have not really changed significantly over the years. We still have just *three* basic topologies: the buck, the boost, and the buck-boost. Admittedly, there have been significant improvements like 'ZVS' (zero voltage switching), 'current-fed converters,' and 'composite topologies' like the 'Cuk converter' and the 'SEPIC' (single ended primary inductance converter), but all these are perhaps best viewed as icing on a three-layer cake. The basic building blocks (or topologies) of power conversion have themselves proven to be quite fundamental. And that is borne out by the fact that they have stood the test of time and remained virtually unchallenged to date.

So, *finally, we can get on with the task of really getting to understand these topologies well. We will soon realize that the best way to do so is via the route that takes us past that rather enigmatic component —* ***the inductor****. And that's where we begin our journey now. . .*

Understanding the Inductor

Capacitors/Inductors and Voltage/Current

In power conversion, we may have noticed that we always talk rather instinctively of *voltage* rails. That is why we also have dc-dc *voltage* converters forming the subject of this book. But why not *current* rails, or *current* converters for example?

We should realize that the world we live in, keenly interact with, and are thus comfortable with, is ultimately one of voltage, not current. So for example, every electrical gadget or appliance we use runs off a specified *voltage* source, the currents drawn from which being largely ours to determine. So for example, we may have 110 V-ac or 115 V-ac 'mains input' available in many countries. Many other places may have 220 V-ac or 240 V-ac. So if for example, an electric room heater is connected to the 'mains outlet,' it would draw a very large *current* (~10–20 amperes), but the line voltage itself would hardly change in the process. Similarly, a clock radio would typically draw only a few hundred milliamperes of current, the line voltage again remaining fixed. That is by definition a voltage source. On the other hand, imagine for a moment that we had a 20 A *current source* outlet available in our wall. By definition, this would try to push out 20 A, *come what may* — even adjusting the voltage if necessary to bring that about. So, even if don't connect any appliance to it, it would still attempt to arc over, just to keep 20 A flowing. No wonder we hate current sources!

We may have also observed that *capacitors* have a rather more direct relationship with voltage, rather than current. So C = Q/V, where C is the capacitance, Q is the charge on either plate of the capacitor, and V is the voltage across it. This gives capacitors a somewhat imperceptible, but natural association with our more "comfortable" world of voltages. It's perhaps no wonder we tend to understand their behavior so readily.

Unfortunately, capacitors are not the only power-handling component in a switching power supply! Let us now take a closer look at the main circuit blocks and components of a typical off-line power supply as shown in *Figure 1-1*. Knowing what we now know about capacitors and their natural relationship to voltage, we are not surprised to find there are capacitors present at both the input and output ends of the supply. But we also find an *inductor* (or 'choke') — in fact a rather bulky one at that too! We will learn that this behaves like a *current source*, and therefore, quite naturally, we don't relate too well to it! However, to gain mastery in the field of power conversion, we need to understand *both* the key components involved in the process: capacitors *and inductors*.

Coming in from a more seemingly natural world of voltages and capacitances, it may require a certain degree of mental re-adjustment to understand inductors well enough. Sure, most power supply engineers, novice or experienced, are able to faithfully reproduce the buck converter *duty cycle equation* for example (i.e. the relationship between input and output voltage). Perhaps they can even derive it too on a good day! But scratch the surface, and we can surprisingly often find a noticeable lack of "feel" for inductors. We would do well to recognize this early on and remedy it. With that intention, we are going to start at the very basics. . . .

The Inductor and Capacitor Charging/Discharging Circuits

Let's start by a simple question, one that is sometimes asked of a prospective power supply hire (read "nervous interviewee"). This is shown in *Figure 1-3*.

Note that here we are using a mechanical switch for the sake of simplicity, thus also assuming it has none of the parasitics we talked about earlier. At time t = 0, we close the switch (ON), and thus apply the dc voltage supply (V_{IN}) across the capacitor (C) through the small series limiting resistor (R). What happens?

Most people get this right. The capacitor *voltage* increases according to the well-known exponential curve $V_{IN} \times (1 - e^{-t/\tau})$, with a 'time constant' of $\tau = RC$. The capacitor *current*, on the other hand, starts from a high initial value of V_{IN}/R and then decays exponentially according to $(V_{IN}/R) \times e^{-t/\tau}$. Yes, if we wait "a very long time," the capacitor would get charged up almost fully to the applied voltage V_{IN}, and the current would correspondingly fall (almost) to zero. Let us now *open* the switch (OFF), though not necessarily having waited a very long time. In doing so we are essentially attempting to force the *current* to

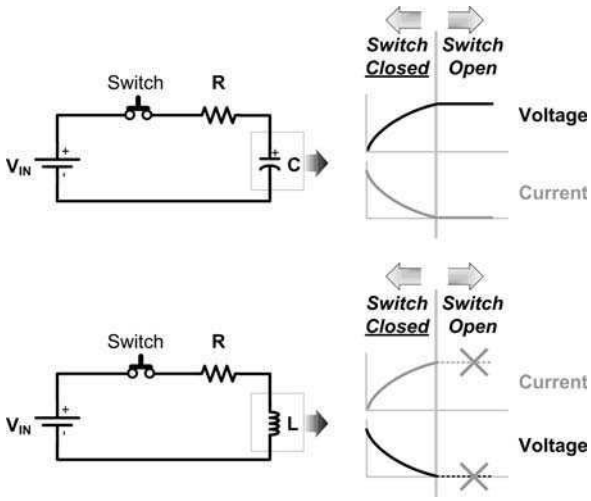


Figure 1-3: Basic Charging/Discharging Circuits for Capacitor and Inductor

zero (that is what a *series* switch is always supposed to do). What happens? The capacitor remains charged to whatever voltage it had already reached, and its current goes down immediately to zero (if not already there).

Now let us repeat the same experiment, but *with the capacitor replaced by an inductor* (L), as also shown in *Figure 1-3*. Interviewees usually get the "charging" part (switch-closed phase) of this question right too. They are quick to point out that the current in the inductor behaves just as the voltage across the capacitor did during the charging phase. And the voltage across the inductor decays exponentially, just as the capacitor current did. They also seem to know that the time constant here is $\tau = L/R$, not RC.

This is actually quite encouraging, as it seems we have, after all, heard of the '*duality principle*.' In simple terms this principle says that *a capacitor can be considered as an inverse (or 'mirror') of an inductor, because the voltage-current equations of the two devices can be transformed into one another by exchanging the voltage and current terms*. So, in essence, capacitors are analogous to inductors, and voltage to current.

But wait! Why are we even interested in this exotic-sounding new principle? Don't we have enough on our hands already? Well, it so happens, that by using the duality principle we can often *derive a lot of clues about any L-based circuit from a C-based circuit, and vice versa* — right off the bat — *without* having to plunge headlong into a web of hopelessly non-intuitive equations. So in fact, we would do well to try and use the duality principle to our advantage if possible.

With the duality principle in mind, let us attempt to *open* the switch in the inductor circuit and try to predict the outcome. What happens? No! Unfortunately, things don't remain almost "unchanged" as they did for a capacitor. In fact, the behavior of the inductor during the off-phase is really no replica of the off-phase of the capacitor circuit.

So does that mean we need to jettison our precious duality principle altogether? Actually we don't. The problem here is that the two circuits in *Figure 1-3*, despite being deceptively similar, are *really not duals of each other*. And for that reason, we really can't use them to derive any clues either. A little later, we will construct *proper* dual circuits. But for now we may have already started to suspect that we really don't understand inductors as well as we thought, nor in fact the duality principle we were perhaps counting on to do so.

The Law of Conservation of Energy

If a nervous interviewee hazards the guess that the current in the inductor simply "goes to zero immediately" on opening the switch, a gentle reminder of what we all learnt in high school is probably due. The stored energy in a capacitor is $CV^2/2$, and so there is really no problem opening the switch in the capacitor circuit — the capacitor just continues to hold its stored energy (and voltage). But in an inductor, the stored energy is $LI^2/2$. Therefore, if we speculate that the current in the inductor circuit is at some finite value before the switch is opened and zero *immediately* afterward, the question arises: to *where did all the stored inductor energy suddenly disappear?* ***Hint:*** we have all heard of the *law of conservation of energy* — energy can change its form, but it just cannot be wished away!

Yes, sometimes a particularly intrepid interviewee will suggest that the inductor current *"decays exponentially to zero"* on opening the switch. So the question arises — where is the current in the inductor flowing to and from? We always need a *closed* galvanic path for current to flow (from Kirchhoff's laws)!

But, wait! Do we even fully understand the *charging phase* of the inductor well enough? Now this is getting really troubling! Let's find out for ourselves!

The Charging Phase and the Concept of Induced Voltage

From an intuitive viewpoint, most engineers are quite comfortable with the mental picture they have acquired over time *of a capacitor being charged* — the accumulated charge keeps trying to repel any charge trying to climb aboard the capacitor plates, till finally a balance is reached and the incoming charge (current) gets reduced to near-zero. This picture is also intuitively reassuring, because at the back of our minds, we realize it corresponds closely with our understanding of *real-life* situations — like that of an over-crowded bus during rush hour, where the number of commuters that manage to get on board at a stop depends on the

capacity of the bus (double-decker or otherwise), and also on the sheer desperation of the commuters (the applied voltage).

But coming to the inductor charging circuit (i.e. switch closed), we can't seem to connect this too readily to any of our immediate real-life experiences. Our basic question here is — why does the charging *current* in the inductor circuit actually *increase* with time. Or equivalently, what prevents the current from being high to start with? We know there is no mutually repelling "charge" here, as in the case of the capacitor. So why?

We can also ask an even more basic question — why is there *any voltage* even present across the inductor? We always accept a voltage across a *resistor* without argument — because we know *Ohm's law* ($V = I \times R$) all too well. But an inductor has (almost) *no resistance* — it is basically just a length of solid conducting copper wire (wound on a certain core). So how does it manage to "hold-off" any voltage across it? In fact, we are comfortable about the fact that a capacitor can hold voltage across it. But for the inductor — we are not very clear! Further, if what we have learnt in school is true — that electric field by definition is the *voltage gradient* dV/dx ("x" being the distance), we are now faced with having to explain a mysterious *electric field* somewhere inside the inductor! Where did that come from?

It turns out, that according to Lenz and/or Faraday, the current takes time to build up in an inductor only because of '*induced voltage*.' This voltage, by definition, opposes any external effort to *change* the existing flux (or current) in an inductor. So if the current is fixed, yes, there is no voltage present across the inductor — it then behaves just as a piece of conducting wire. But the moment we try to *change* the current, we get an induced voltage across it. By definition, the *voltage measured across an inductor at any moment* (whether the switch is open or closed, as in *Figure 1-3*) is the 'induced voltage.'

> **Note:** We also observe that the analogy between a capacitor/inductor and voltage/current, as invoked by the duality principle, doesn't stop right there! For example, it was considered equally puzzling at some point in history, how at all any current was apparently managing to flow through a capacitor — when the applied voltage across it was changed. Keeping in mind that a capacitor is basically two metal plates with an interposing (non-conducting) *insulator*, it seemed contrary to the very understanding of what an "insulator" was supposed to be. This phenomenon was ultimately explained in terms of a 'displacement current,' that flows (or rather *seems* to flow) through the plates of the capacitor, when the voltage changes. In fact, this current is completely analogous to the concept of 'induced voltage' — introduced much later to explain the fact that a voltage was being observed across an inductor, when the current through it was changing.

So let us now try to figure out exactly how the induced voltage behaves when the switch is closed. Looking at the inductor charging phase in *Figure 1-3*, the inductor current is initially zero. Thereafter, by closing the switch, we are attempting to cause a sudden change in the current. The induced voltage now steps in to try to keep the current down to its initial value (zero). So we apply 'Kirchhoff's voltage law' to the closed loop in question. Therefore, at the moment the switch closes, the induced voltage must be exactly equal to the applied voltage, since the voltage drop across the series resistance R is initially zero (by Ohm's law).

As time progresses, we can think intuitively in terms of the applied voltage "winning." This causes the current to rise up progressively. But that also causes the voltage drop across R to increase, and so the induced voltage must fall *by the same amount* (to remain faithful to Kirchhoff's voltage law). That tells us exactly what the *induced voltage* (voltage across inductor) is during the entire switch-closed phase.

Why does the applied voltage "win"? For a moment, let's suppose it didn't. That would mean the applied voltage and the induced voltage have managed to completely counter-balance each other — and the current would then *remain* at zero. However, that cannot be, because zero rate of change in current implies no induced voltage either! In other words, the very existence of induced voltage depends on the fact that current changes, and it *must* change.

We also observe rather thankfully, that all the laws of nature bear each other out. There is no contradiction whichever way we look at the situation. For example, even though the current in the inductor is subsequently higher, its *rate of change* is less, and therefore, so is the induced voltage (on the basis of Faraday's/Lenz's law). And this "allows" for the additional drop appearing across the resistor, as per Kirchhoff's voltage law!

But we still don't know how the induced voltage behaves when the switch turns OFF! To unravel this part of the puzzle, we actually need some more analysis.

The Effect of the Series Resistance on the Time Constant

Let us ask — what are the *final* levels at the end of the charging phase in *Figure 1-3* — that is, of the current in the inductor and the voltage across the capacitor. This requires us to focus on the exact role being played by *R*. Intuitively we expect that for the capacitor circuit, increasing the R will increase the charging time constant τ. This is borne out by the equation $\tau = RC$ too, and is what happens in reality too. But for the inductor charging circuit, we are again up against another seemingly counter-intuitive behavior — *increasing R actually decreases the charging time constant*. That is in fact indicated by $\tau = L/R$ too.

Let us attempt to explain all this. Looking at *Figure 1-4* which shows the inductor charging current, we can see that the $R = 1\ \Omega$ current curve does indeed rise faster than the $R = 2\ \Omega$ curve (as intuitively expected). But the *final value* of the $R = 1\ \Omega$ curve is *twice* as high. Since by definition, the time constant is "the time to get to 63% of the *final value*," therefore the $R = 1\ \Omega$ curve has a *larger* time constant, despite the fact that it did rise much faster from the get-go. So that explains the inductor *current* waveforms.

But looking at the inductor *voltage* waveforms in *Figure 1-5*, we see there is still some explaining to do. Note that for a decaying exponential curve, the time constant is defined as the time it takes to get to 37% of the *initial* value. So in this case we see that though the initial values of all the curves are the same, *yet for example, the* $R = 1\ \Omega$ *curve has a slower decay (larger time constant) than the* $R = 2\ \Omega$ *curve*! There is actually no mystery involved

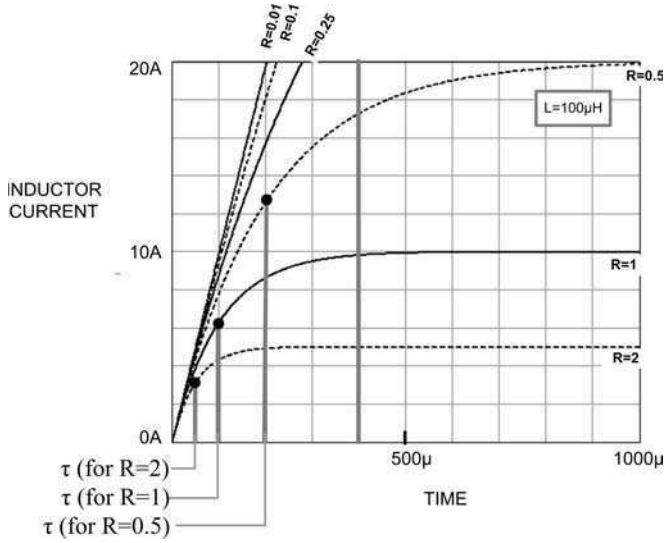


Figure 1-4: Inductor Current during Charging Phase for Different R (in ohms), for an Applied Input of 10 V

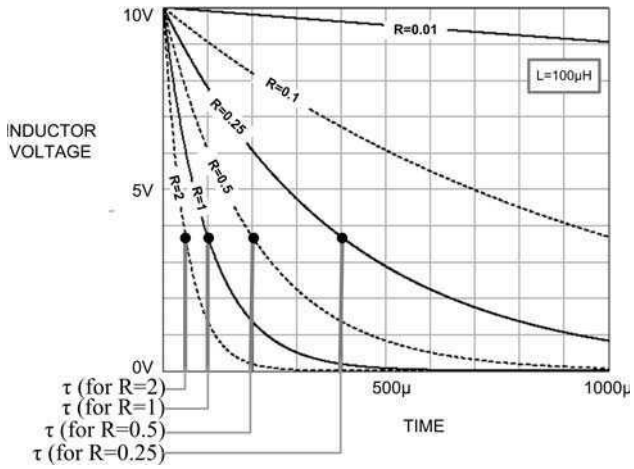


Figure 1-5: Inductor Voltage during Charging Phase for Different R (in ohms), for an Applied Input of 10 V

here, since we already know what the current is doing during this time (*Figure 1-4*), and therefore the voltage curves follow automatically from Kirchhoff's laws.

The conclusion is that if, in general, we ever make the mistake of looking *only* at an inductor *voltage* waveform, we may find ourselves continually baffled by an inductor! *For an inductor, we should always try to see what the* ***current*** *in it is trying to do*. That is why, as we

just found out, the voltage during the off-time is determined entirely by the current. The voltage just follows the dictates of the current, not the other way around. In fact, in Chapter 5, we will see how this particular behavioral aspect of an inductor determines the exact shape of the voltage and current waveforms during a switch transition, and thereby determines the crossover (transition) loss too.

The Inductor Charging Circuit with R = 0, and the "Inductor Equation"

What happens if R is made to decrease to zero?

From *Figure 1-5* we can correctly guess that the only reason that the voltage across the inductor during the on-time changes at all from its initial value V_{IN} is the presence of R! So if R is 0, we can expect that the voltage across the inductor *never changes* during the on-time! The induced voltage must then be equal to the applied dc voltage. That is not strange at all — if we look at it from the point of view of Kirchhoff's voltage law, there is *no* voltage drop present across the resistor — simply because there is no resistor! So in this case, all the applied voltage appears across the inductor. And we know it can "hold-off" this applied voltage, provided the current in it is changing. Alternatively, *if any voltage is present across an inductor, the current through it* ***must*** *be changing*!

So now, as suggested by the low-R curves of *Figure 1-4* and *Figure 1-5*, we expect that the *inductor current will keep ramping up with a constant slope during the on-time*. Eventually, it will reach an *infinite* value (in theory). In fact, this can be mathematically proven to ourselves by differentiating the inductor charging current equation with respect to time, and then putting R = 0 as follows

$$I(t) = \frac{V_{IN}}{R}\left(1 - e^{-tR/L}\right)$$

$$\frac{dI(t)}{dt} = \frac{V_{IN}}{R}\left(\frac{R}{L}e^{-tR/L}\right)$$

$$\left.\frac{dI(t)}{dt}\right|_{R=0} = \frac{V_{IN}}{L}$$

So we see that when the inductor is connected directly across a voltage source V_{IN}, the slope of the line representing the inductor current is constant, and equal to V_{IN}/L (the current *rising* constantly).

Note that in the above derivation, the voltage across the inductor happened to be equal to V_{IN}, because R was 0. But in general, if we call "V" the voltage actually present *across* the inductor (at any given moment), I being the current through it, we get the general

"inductor equation"

$$\frac{dI}{dt} = \frac{V}{L} \quad \textit{(inductor equation)}$$

This equation applies to an ideal inductor (R = 0), in any circuit, under any condition. For example, it not only applies to the "charging" phase of the inductor, but also its "discharging" phase!

Note: When working with the inductor equation, for simplicity, we usually plug in only the *magnitudes* of all the quantities involved (though we do mentally keep track of what is really happening — i.e. current rising or falling).

The Duality Principle

We now know how the voltage and current (rather its rate of change), are mutually related in an inductor, during both the charging and discharging phases. Let us use this information, along with a more complete statement of the *duality principle*, to finally understand what really happens when we try to interrupt the current in an inductor.

The principle of duality concerns the transformation between two apparently different circuits, which have similar properties when current and voltage are interchanged. Duality transformations are applicable to planar circuits only, and involve a topological conversion: capacitor and inductor interchange, resistance and conductance interchange, and ***voltage source and current source*** *interchange.*

We can thus spot our "mistakes" in *Figure 1-3*. First, we were using an input *voltage* source applied to *both* circuits — whereas we should have used a *current* source for the "other" circuit. Second, we used a *series* switch in both the circuits. We note that the primary function of a series switch is only to interrupt the flow of *current* — not to change the voltage (though that may happen as a result). So if we really want to create proper mirror (dual) circuits, then forcing the current to zero in the inductor is the dual of forcing the *voltage* across the capacitor to zero. And to implement that, we obviously need to place a switch in *parallel* to the capacitor (not in series with it). With these changes in mind, we have finally created *true* dual circuits as shown in *Figure 1-6* (*both* are actually *equally* impractical in reality!).

The "Capacitor Equation"

To analyze what happens in *Figure 1-6* we must first learn the "capacitor equation" — analogous to the "inductor equation" derived previously. If the duality principle is correct,

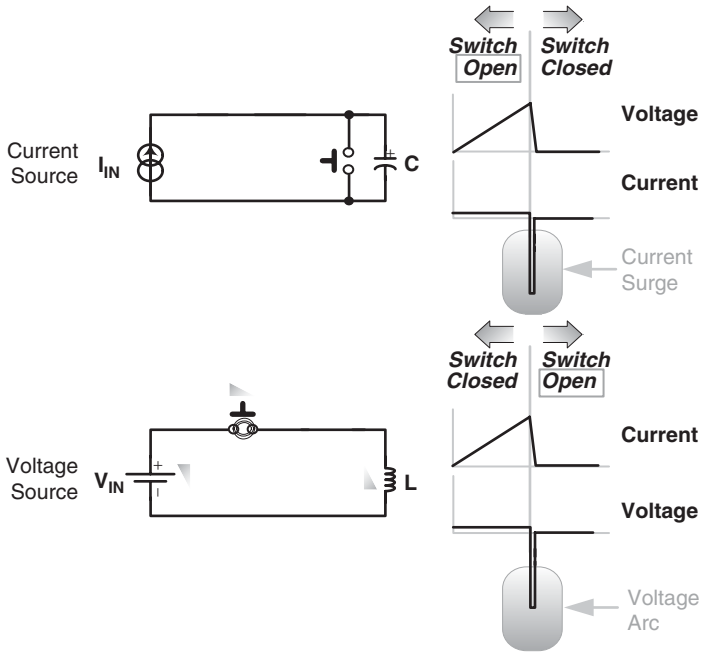


Figure 1-6: Mirror Circuits for Understanding Inductor Discharge

both the following two equations must be valid

$$V = L\frac{dI}{dt} \quad \textit{(inductor equation)}$$

$$I = C\frac{dV}{dt} \quad \textit{(capacitor equation)}$$

Further, if we are dealing with "straight-line segments" (constant V for an inductor and constant I for a capacitor), we can write the above equations in terms of the corresponding *increments* or *decrements* during the given time segment.

$$V = L\frac{\Delta I}{\Delta t} \quad \textit{(inductor equation for constant voltage)}$$

$$I = C\frac{\Delta V}{\Delta t} \quad \textit{(capacitor equation for constant current)}$$

It is interesting to observe that the duality principle is actually helping us understand how the *capacitor* behaves when being charged (or discharged) by a *current source*! We can guess that the *voltage* across the capacitor will then ramp up in a straight line — to near infinite

values — just as the inductor *current* does with an applied *voltage* source. And in both cases, the final values reached (of the voltage across the capacitor and the current through the inductor) *are dictated only by various parasitics* that we have not considered here — mainly the ESR of the capacitor and the DCR of the inductor respectively.

The Inductor Discharge Phase

We now analyze the mirror circuits of *Figure 1-6* in more detail.

We know intuitively (and also from the capacitor equation) what happens to a *capacitor* when we attempt to suddenly discharge it (by means of the parallel switch). Therefore, we can now easily guess what happens when we suddenly try to "discharge" the inductor (i.e. force its current to zero by means of the series switch).

We know that if a "short" is applied across any capacitor terminals, we get an extremely high current surge for a brief moment — during which time the capacitor discharges, and the voltage across it ramps down steeply to zero. So we can correctly infer that if we try to interrupt the current through an inductor, we will get *a very high voltage across it — with the current simultaneously ramping down steeply to zero*. So the mystery of the inductor "discharge" phase is solved — *with the help of the duality principle!*

But we still don't know *exactly* what the actual *magnitude* of the voltage spike appearing across the switch/inductor is. That is simple — as we said previously, during the off-time, the voltage will take on any value to force *current continuity*. So a brief arc will appear across the contacts as we try to pull them apart (see *Figure 1-6*). If the contacts are separated by a greater distance, the voltage will increase automatically to maintain the spark. And during this time, the current will ramp down steeply. The arcing will last for *as long* as there is *any* remaining inductor stored energy — that is, till the current *completely* ramps down to zero. The rate of fall of current is simply V/L, from the inductor equation. So eventually, *all the stored energy in the inductor is completely dissipated* in the resulting flash of heat and light, and the current returns to zero simultaneously. At this moment, the induced voltage collapses to zero too, its purpose complete. This is in fact the basic principle behind the automotive spark plug, and the camera flash too (occurring in a more controlled fashion).

But wait — we have stated above that the rate of fall of current in the inductor circuit was "V/L." What is V? V is the voltage *across the inductor*, *not* the voltage across the contacts. In the following sections, we will learn that the voltage *across an inductor* (almost always) reverses when we try to interrupt its current. If that is true, then by Kirchhoff's voltage law, since the algebraic sum of all the voltage drops in any closed circuit must add up to zero, the voltage *across the contacts* will be equal to the sum of the magnitudes of the induced voltage and the applied dc rail — however, the sign of the voltage across the contacts (i.e. its direction) will necessarily be *opposite* to the other voltages (see the gray triangles in the

lower schematic of *Figure 1-6*). Therefore, we conclude that the magnitude of the voltage spike across the inductor is equal to the magnitude of the voltage across the contacts, minus the magnitude of the input dc voltage.

Finally, we know everything about the puzzling inductor discharge phase!

Flyback Energy and Freewheeling Current

The energy that "must get out" of the inductor when we try to open the switch is called the 'flyback' energy. The current that continues to force its way through is called the 'freewheeling' current. Note that this not only sounds, but in fact is, very similar to another real-world situation — that of a *mechanical spinning wheel*, or a 'flywheel.' In fact, understanding the flywheel can help greatly in gaining an intuitive insight into the behavior of an inductor.

Just as the inductor has stored energy related to the current flowing through it, the flywheel stores energy related to its spinning action. And neither of these energy terms can be wished away in an instant. In the case of the flywheel, we can apply "brakes" to dissipate its rotational energy (as heat in the brake linings) — and we know this will produce a ***progressive*** reduction in the spinning. Further, if the brakes are applied more emphatically, the time that will elapse till the spinning stops entirely gets proportionately decreased. That is very similar to an inductor — with the *induced voltage (during the off-time) playing the part of the "brakes" and the current being akin to the spinning.* So, the induced voltage causes a *progressive* reduction in the current. If we have a higher induced voltage, this will cause a steeper fall in the current. In fact, that is also indicated by the inductor equation $V = LdI/dt$!

However, we have also learned something more fundamental about the behavior of an inductor, as described next.

Current Must Be Continuous, Its Slope Need Not Be

The key word in the previous section was *progressive*. From a completely mathematical/geometrical point of view now, we should understand that any curve representing inductor *current* cannot be *discontinuous* (no *sudden* jumps allowed) — because that will in effect cause *energy* to be discontinuous, which we know is impossible. But we can certainly cause the *slope* of the current (i.e. its dI/dt) to have "jumps." So we can, for example, change the slope of current (dI/dt) *in an instant* — from one representing a rising ramp (increasing stored energy), to one representing a falling ramp (opposite sign, i.e. decreasing energy). However, the current itself must always be continuous. This is shown in *Figure 1-7*, under the choices marked "possible."

Note that there are *two* options in the figure that are "possible." Both are so, simply because they do not violate any known physical laws. However, one of these choices is considered

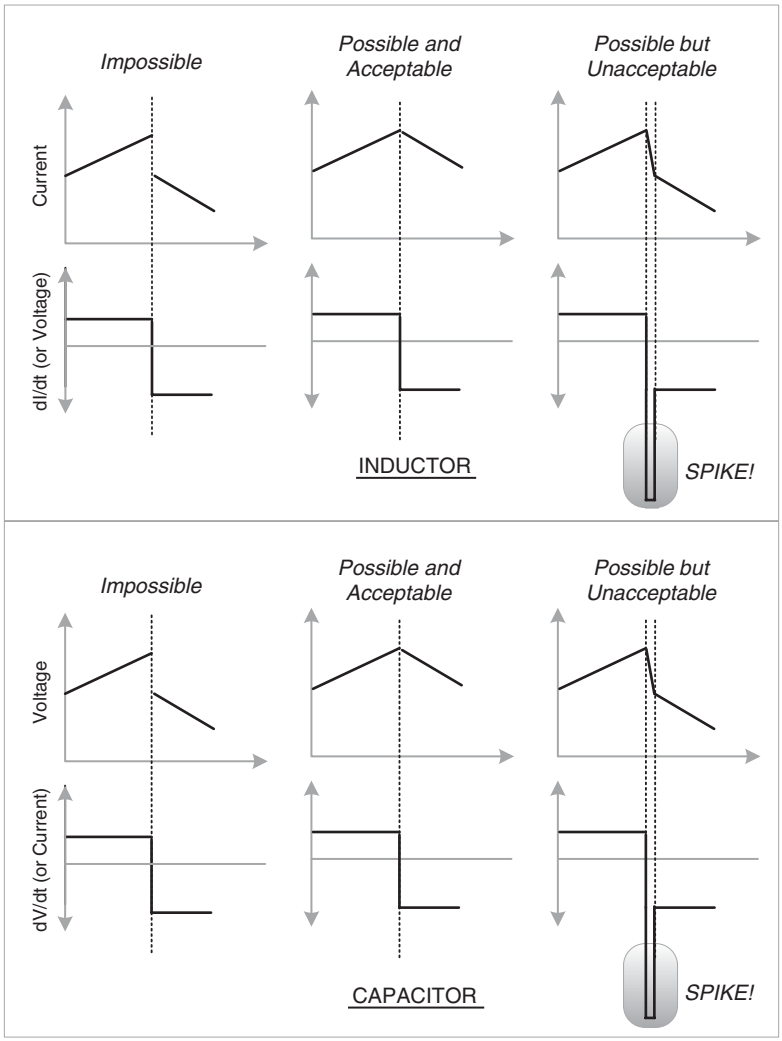


Figure 1-7: Inductor Current Must Be Continuous, But Its Slope Need Not Be. Capacitor Voltage Must Be Continuous, But Its Slope Need Not Be.

"unacceptable," because of the *huge spike* — which we know can damage the switch. The other choice, marked "acceptable," is in fact what really happens in any switching converter topology, as we will soon see.

The Voltage Reversal Phenomenon

We mentioned there is a *voltage reversal* across the switch, when we try to interrupt its current. Let us try to understand this better now.

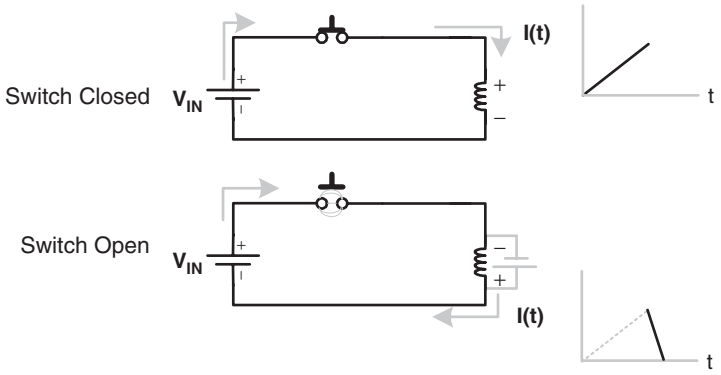


Figure 1-8: How the Voltage Reverses on Attempted Interruption of Inductor Current

An intuitive (but not necessarily rigorous) way to visualize it is shown in *Figure 1-8*. Here we note that when the switch is closed (upper schematic), the *current is shown leaving the positive terminal of the applied dc voltage source* — that being the normal convention for describing the direction of current flow. During this on-time, the *upper* end of the inductor gets set to a higher voltage than its lower end. Subsequently, when the switch opens, the input dc source gets disconnected from the inductor. But we have just learnt that the current demands to keep flowing (at least for a while) — *in the same direction as previously flowing*. So during the switch off-time, we can mentally visualize the inductor as becoming a sort of "voltage source," forcing the current to keep flowing. For that reason, we have placed an imaginary (gray) voltage source (battery symbol) across the inductor in the lower half of the figure — its polarity in accordance with the convention that current must leave by the positive terminal of any voltage source. Thus we can see that this causes the lower end of the inductor to now be at a higher voltage than its upper end. Clearly, voltage reversal has occurred — simply by the need to maintain current continuity.

The phenomena of voltage reversal can be traced back to the fact that induced voltage always opposes any change (in current). However, in fact, voltage reversal does *not always* occur. For example, voltage reversal does *not* occur during the *initial* startup ('power-up') phase of a boost converter. That is because the primary requirement is only that the inductor *current* somehow needs to keep flowing — voltage takes a backseat. So hypothetically, if a circuit is wired in a certain way, and the conditions are "right," it is certainly possible that voltage reversal won't occur, so long as current continuity can still be maintained.

However, we must be clear that *if and when a converter reaches a 'steady state,' voltage reversal will necessarily occur at every switch transition.*

For that we now have to understand what a "steady state" is.

A Steady State in Power Conversion, and the Different Operating Modes

A *steady state* is, as the name indicates — *stable*. So it is in essence the opposite of a runaway or unstable condition. But we can easily visualize that we will in fact get an unstable condition if at the end of every cycle, we ***don't*** *return to the current we started the cycle with* — because then, every successive cycle, we will accumulate a net increase or decrease of current, and the situation will keep changing forever (in theory).

From $V = L\Delta I/\Delta t$, it is clear that if the current is ramping *up* for a positive (i.e. applied) voltage, the current must ramp *down* if the voltage reverses. So the following equations must apply (magnitudes only)

$$V_{ON} = L\frac{\Delta I_{ON}}{\Delta t_{ON}}$$

$$V_{OFF} = L\frac{\Delta I_{OFF}}{\Delta t_{OFF}}$$

Here the subscript "ON" refers to the switch being closed, and "OFF" refers to the switch being open. V_{ON} and V_{OFF} are the respective voltages *across the inductor* during the durations Δt_{ON} and Δt_{OFF}. Note that very often, Δt_{ON} is written simply as "t_{ON}," the switch on-time. And similarly, Δt_{OFF} is simply "t_{OFF}," the switch off-time.

Now suppose we are able to create a circuit in which the *amount* the current ramps *up* by in the on-time (ΔI_{ON}) is *exactly equal* to the amount the current ramps *down* by during the off-time (ΔI_{OFF}). If that happens, we would have reached a steady state. Now we could repeat the same sequence an innumerable amount of times, and get the *same* result each and every time. In other words, *every "switching cycle" would then be an exact replica of the previous cycle*. Further, we could also perhaps get our circuit to deliver a steady stream of (identical) energy packets continuously to an output capacitor and load. If we could do that, by definition, we would have created a *power converter*!

Achieving a steady state is luckily not as hard as it may sound. Nature *automatically* tries to help every natural process move towards a stable state (without "user intervention"). So in our case, all we need to do on our part is to *provide* a circuit that *allows* these conditions to develop *naturally* (over several cycles). And if we have created the right conditions, a steady state *will* ultimately result. Further, this would be self-sustaining thereafter. Such a circuit would then be called a switching "*topology*"!

Conversely, any valid topology must be able to reach a state described by the following key equation $\Delta I_{ON} = \Delta I_{OFF} \equiv \Delta I$. If it can't get this to happen, it is *not* a topology. Therefore, this simple current increment/decrement equation forms the litmus test for validating any new switching topology.

Note that the inductor equation, and thereby the definition of 'steady state,' refers only to the *increase/decrease* in current — it says nothing about the actual (absolute) value of the current at the start (and end) of every cycle. So there are in fact several possibilities. We could have a steady state in which the current returns to *zero* every cycle, and this is called a 'discontinuous conduction mode' (DCM). However, if the current stays pegged at some non-zero value throughout, we will have 'continuous conduction mode' (CCM). The latter mode is the most common mode of operation encountered in power conversion. In *Figure 1-9* we have graphically shown these operating modes (all in steady state). We also have some other modes that we will talk about very soon. Note that in the figure, the "square" waveform is the voltage across the inductor, and the slowly ramping waveform is the inductor current. Let us make some related observations:

a) We see that the voltage across the inductor *always* reverses at every switching event (as expected in steady state).

b) We note that since the inductor equation relates voltage to the *slope of the current*, not to the actual current, therefore, for a given V_{ON} and V_{OFF}, several current waveforms are possible (all having the *same dI/dt* for corresponding segments). Each of these possibilities has a name — CCM, DCM, BCM (boundary conduction mode, also called critical conduction mode), and so on. Which of these operating modes actually occurs depends on the specific circuit (i.e. the topology) and also the application conditions (how much output power we are demanding and what the input and output voltages are).

c) The inductor voltages, V_{ON} and V_{OFF} shown in the figure, are related to the application conditions V_{IN} and/or V_O. Their exact relationship will become known a little later, and we will also learn that it depends on the specific topology.

d) A key question is — what is the exact relationship between the average *inductor* current and the *load* current? We will soon see that that too depends on the specific *topology*. However, in all cases, the average inductor current ("I_{AVG}" or "I_L") is proportional to the load current ("I_O"). So if for example I_O is 2 A and I_{AVG} is 10 A, then if I_O is decreased to 1 A, I_{AVG} will fall to 5 A. Therefore on decreasing the load current, we can get I_{AVG} to decrease, as indicated in *Figure 1-9*.

e) Typically, we transit *automatically* from CCM to DCM, just by reducing the load current of the converter. But note that we will necessarily have to pass through BCM along the way.

f) "BCM" is just that — a '*boundary* conduction mode' — situated exactly between CCM and DCM. It is therefore a purely philosophical question to ask whether BCM should be viewed as CCM or DCM (at their respective extremes) — it really doesn't matter.

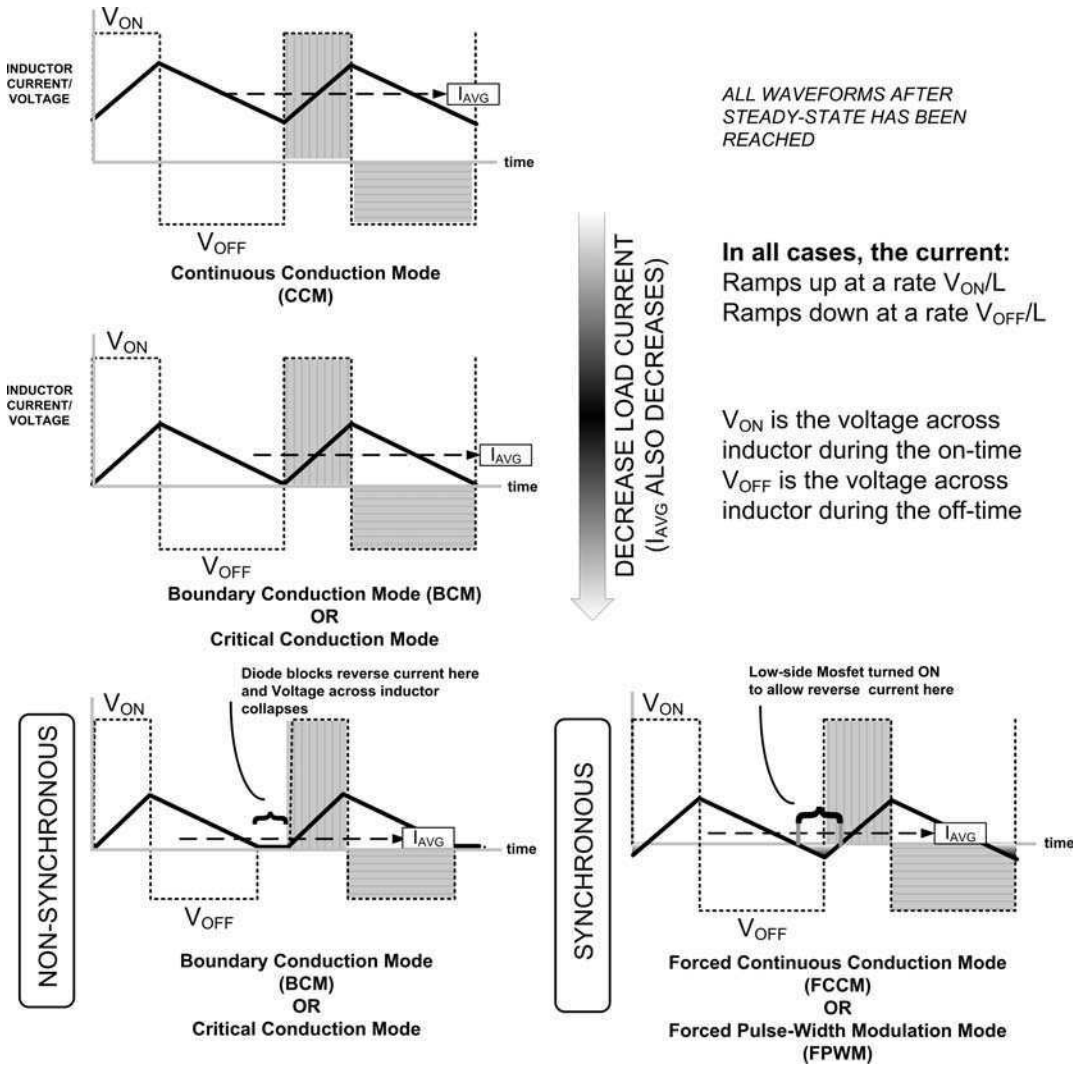


Figure 1-9: Different Operating Modes of Switching Regulators

g) Note that in all the cases shown in *Figure 1-9*, with the exception of DCM, the average inductor current I_{AVG} is just the *geometrical center of the ramp* part of the current waveform. In DCM however, we have an additional interval in which there is *no current* passing for a while. So, to find the average value of the inductor current, a rather more detailed calculation is required. In fact that is the primary reason why *DCM equations turn out looking so complicated* — to the point that many engineers seem to rather instinctively ignore DCM altogether, despite some advantages of operating a converter in DCM instead of CCM.

Note: Expectedly, all DCM equations lead to exactly the same *numerical* results as the CCM equations — *when the converter is in BCM*. Practically speaking, we can freely pick and choose whether to use the CCM equations, or the more formidable looking DCM equations, for evaluating a converter in BCM. Of course, there is no reason why we would ever want to struggle through complicated equations, when we can use much simpler equations to get the same results!

h) What really is the average inductor current "I_{AVG}" as shown in *Figure 1-9*? A nice way to understand this parameter is through the "car analogy." Suppose we press the gas pedal of a car. The car responds by increasing its speed. In an analogous fashion, when we apply a voltage across an inductor (the on-time voltage "V_{ON}"), the current ramps up. Subsequently, suppose we press on the brakes of the car. The car will then respond by decreasing its speed. Similarly, when the applied voltage is removed from the inductor, voltage reversal occurs, and an induced voltage (the "brakes") appears across the inductor, "V_{OFF}." Since it is in the opposite direction as V_{ON}, it causes the current to ramp down. So now, if we press the gas pedal (V_{ON}), followed by the brakes (V_{OFF}), in *quick succession*, and with the right timing, we could still make the car continue to move forward despite the constant lurching. It would then have a certain *average speed* — depending on the ratio of the gas pedal duration and the subsequent braking duration. In power conversion, this "lurching" is analogous to the 'current ripple' $\Delta I = \Delta I_{ON} = \Delta I_{OFF}$. And quite similarly, we have an 'average inductor current' I_{AVG} too, as shown in *Figure 1-9*. However, we do understand that in a power converter, the output capacitor eventually absorbs (or smoothens) this "lurching," and thus manages to deliver a steady dc current to the load as desired.

i) Some control ICs manage to *maintain* the converter in BCM mode under all application conditions. Examples of these are certain types of 'hysteretic controllers' and self-oscillating types called 'ringing choke converters' (RCCs). However, we know that the current ramps down at a rate V/L. And since V depends on the input/output voltages, the time to get to zero current depends on the specific application conditions. Therefore, in any BCM implementation, we always lose the advantage of fixed switching frequency operation.

j) Most conventional topologies are nowadays labeled 'non-synchronous' — to distinguish them from more recent 'synchronous' topologies. In the former, a diode is always present (the catch diode), that *prevents the inductor current from reversing direction*, any time during the switching cycle. That is why, on reducing output power and/or increasing input voltage, we automatically transit from CCM to DCM. However, in synchronous topologies, the catch diode is either supplanted or completely replaced by a 'low-drop' mosfet across it. So whenever the diode is supposed to conduct, we force this extra mosfet into conduction for that duration. Since the drop across this mosfet is much lower than across a diode, not only do we

manage to significantly reduce the conduction loss in the freewheeling path, but we can also now allow *reverse inductor current* — that is, current moving instantaneously *away* from the load. However note that the average inductor current could still be positive — see *Figure 1-9*. Further, with negative currents now being "allowed," we no longer get DCM on reducing output power, but rather enter FPWM/FCCM as described in the figure.

Note: It is fortunate that almost all the standard CCM design equations (for non-synchronous topologies) apply equally to FCCM. So from the viewpoint of a harried designer, one of the "advantages" of using synchronous topologies is that the complicated DCM equations are a thing of the past! Though there are some new complications and nuances of synchronous topologies that we need to understand eventually.

The Voltseconds Law, Inductor Rest, and Converter Duty Cycle

There is another way to describe a *steady state*, by bringing in the inductor equation $V = L\Delta I/\Delta t$.

We know that during a steady state $\Delta I_{ON} = \Delta I_{OFF} \equiv \Delta I$. So what we are also saying is that

In steady state, the product of the voltage applied across the inductor, multiplied by the duration we apply it for (i.e. the on-time), must be equal to the voltage that appears across the inductor during the off-time, multiplied by the duration that lasts for.

Therefore we get

$$V_{ON} \times t_{ON} = V_{OFF} \times t_{OFF}$$

The product of the voltage, and the time for which it appears across the inductor, is called the 'voltseconds' across the inductor. So equivalently, what we are also saying is that

If we have an inductor in a steady state, the voltseconds present across it during the on-time (i.e. current ramp-up phase) must be exactly equal in magnitude, though opposite in sign, to the voltseconds present across it during the off-time (i.e. during the current ramp-down phase).

That also means that if we plot the inductor voltage versus time, *the area under the voltage curve during the on-time must be equal to the area under the voltage curve during the off-time*. But we also know that *voltage reversal* always occurs in steady state. So clearly, these two areas must have the opposite sign. See the vertically and horizontally hatched segments in *Figure 1-9*.

Therefore, we can also say that *the **net** area under the voltage curve of an inductor must be equal to zero (in any switching cycle under steady state operation).*

Note that since the typical times involved in modern switching power conversion are so small, "voltseconds" turns out to be a very small number. Therefore, to make numbers more

manageable, we usually prefer to talk in terms of 'Et' or the '*voltμseconds*.' Et is clearly just the voltage applied across the inductor multiplied by the time in *microseconds* (not seconds). Further, we know that typical inductance values used in power conversion are also better expressed in terms of "μH" (microhenries), not H. So from V = LdI/dt we can write

$$\Delta I_{ON} = \frac{V_{ON} \times t_{ON}}{L} = \frac{V_{ON} \times t_{ON\_\mu H}}{L_{\mu H}} = \frac{Et}{L_{\mu H}}$$

or simply

$$\Delta I = \frac{Et}{L} \quad \textit{(steady state, L in } \mu H\textit{)}$$

Note: If in any given equation Et and L appear *together*, it should be generally assumed that L is in μH. Similarly, if we are using voltseconds, that would usually imply L is in H (unless otherwise indicated).

Another term often used in power, one that tells us that we have managed to return to the same inductor current (and energy) that we started off with, is called inductor '*reset*.' Reset occurs at the very moment when the equality $\Delta I_{OFF} = \Delta I_{ON}$ is established. Of course, we could also have a *non-repetitive* (or 'single-shot') event, where the current starts at *zero* and then returns to *zero* — and that too would be inductor "reset."

The corollary is that in a repetitive switching scenario (steady state), *an inductor must be able to reset every cycle*. Reversing the argument — *any circuit configuration that makes inductor reset an impossibility, is not a viable switching topology*.

When we switch repetitively at a switching frequency "f," the 'time period' (T) is equal to 1/f. We can also define the 'duty cycle' (D) of a power converter as the ratio of the on-time of the switch to the time period. So

$$D = \frac{t_{ON}}{T} \quad \textit{(duty cycle definition)}$$

Note that we can also write this as

$$D = \frac{t_{ON}}{t_{ON} + (T - t_{ON})} \quad \textit{(duty cycle definition)}$$

At this point we should be very clear how we are defining "t_{OFF}" in particular. While applying the voltseconds law, we had implicitly assumed that t_{OFF} was *the time for which the induced voltage* V_{OFF} *lasts*, not necessarily the time *for which the switch is OFF* (i.e. $T - t_{ON}$). In DCM they are ***not*** the same (see *Figure 1-9*)! Only in CCM do we get

$$t_{OFF} = T - t_{ON} \quad \textit{(duty cycle in CCM)}$$

and therefore

$$D = \frac{t_{ON}}{t_{ON} + t_{OFF}} \quad \textit{(duty cycle in CCM)}$$

If working in DCM, we should stick to the more general definition of duty cycle given initially.

Using and Protecting Semiconductor Switches

We realize that all topologies exist only because they can achieve a steady state. In an "experimental" topology in which we can't make $\Delta I_{ON} = \Delta I_{OFF}$ happen, the inductor may see a *net increase of current every cycle*, and this can eventually escalate to a very large, almost uncontrolled value of current in just a few cycles. The name given to this progressive ramping-up (or down) of current (or inductor energy), one that is ultimately limited only by parasitics like the ESR and DCR, is called 'staircasing.' The switch will also turn ON into the same current, and can thus be destroyed — that is if the induced voltage spike hasn't already done so (which can happen, if the situation is anything similar to the "unacceptable" plots shown in *Figure 1-7*!).

> **Note:** The very use of the inductor equation $V = LdI/dt$ actually implies we are ignoring its parasitic resistance, DCR. The inductor equation is an idealization, applying only to a "perfect" inductor. That is why we had to put $R = 0$ when we derived it previously.

In an actual power supply, the "mechanical switch" is replaced with a modern semiconductor device (like the mosfet) — largely because then the switching action can be implemented reliably and also at a *very high repetition rate.* But semiconductor devices have certain electrical *ratings* that we need to be well aware of.

Every semiconductor device has an 'absolute maximum *voltage* rating' that, unlike any typical mechanical relay, cannot be exceeded *even momentarily* — without possibly causing its *immediate* destruction. So most mosfets do *not* allow any "latitude" whatsoever, in terms of their *voltage* ratings.

> **Note:** There are some 'avalanche-rated' mosfets available, which can internally 'clamp' the excess voltage appearing across them to some extent. In doing so, they are basically dissipating the excess energy associated with the voltage spike, within their internal clamp. Therefore, they can survive a certain amount of excess voltage (and energy), but only for a short duration (since the device heats up quickly).

There is also a maximum semiconductor device '*current* rating,' but that is usually more *long-term* in nature, dictated by the comparatively slower process of internal heat build-up inside the device. So hypothetically speaking, we could perhaps exceed the current rating somewhat, though only for a *short* time. Of course we don't want to run a device constantly in this excess-current condition. However, under 'abnormal conditions,' like an "overload" on the output of the converter (or the extreme case of a shorted output), we may *judiciously*

allow for a certain amount of "abuse" with regard to the current rating — but certainly not with the voltage rating!

In a practical implementation, we have to design the converter, select the switch, and then lay it all out on a printed circuit board (PCB) with great care — to ensure in particular that there is no *voltage* spike that can "kill" the switch (nor any other semiconductor devices present on the board). Occasionally, we may therefore need to add an external 'snubber' or 'clamp' across the switch, so as to truncate any remnant spikes to within the voltage ratings of the switch.

To protect the switch (and converter) from excess currents, a 'current limit' is usually required. In this case, the current in the inductor, or in the switch, is sensed, and then compared against a set threshold. If and when that is momentarily exceeded, the control circuitry forces the switch to turn OFF *immediately for the remainder of the switching cycle*, so as to protect itself. In the next cycle, no "memory" is usually retained of what may have happened in the preceding cycle. Therefore, every switching cycle is started "afresh," with the current being continuously monitored to ensure it is at a "safe" level. If not, protective action is again initiated, and can be repeated every cycle for several cycles if necessary, until the "overcurrent" condition ceases.

> **Note:** One of the best-known examples of the perils of "previous-cycle memory" in implementing current limit occurs in the popular "Simple Switcher®" family of parts (at www.national.com). In the "third generation" LM267x family, the control circuit rather surprisingly *reduces the duty cycle to about 45% for several cycles after any single current limit event*. It then tries to progressively allow the duty cycle to increase over several successive cycles back to its required value. But this causes severe output 'foldback' and consequent inability to regulate up to full rated load, particularly in applications that require a duty cycle greater than 50%. This condition is further exacerbated with large output capacitances, because the higher currents required to charge the output capacitor after the removal of an abnormal condition (e.g. output short), can lead to another current limit event (and consequent foldback for several cycles again) — *before the duty cycle has been able to return to its desired value*. In effect the converter goes into a continuous "motorboating" condition on removal of the output short, and so the output never recovers. This is rather obliquely "revealed" only deep within the product datasheets (liability cover?).

With the introduction to power conversion now complete, we turn our attention to how switching topologies develop naturally out of the behavior of an inductor.

Evolution of Switching Topologies

Controlling the Induced Voltage Spike by Diversion through a Diode

We realize that our "problem" with using an inductor is *two-fold*: either we are going to end up with *near-infinite induced voltage* spikes, as shown in *Figure 1-6* and *Figure 1-7*, or if we do somehow manage to control the induced voltage to some finite level, the equation $V = LdI/dt$ tells us we could very well end up with *near-infinite currents* (staircasing).

And further, coming to think of it, our basic purpose is still not close to being fulfilled — we still don't know how to derive any *useful power* from our circuit!

Luckily, *all* the above problems can be solved in one stroke! And in doing so, we will arrive at our very first 'switching topology.' Let's now see how that comes about.

We recollect from *Figure 1-6* that the spike of induced voltage at switch turn-off occurs only because the current (previously flowing in the inductor) was still *demanding* a path along which to flow — *and somehow unknowingly, we had failed to provide any*. Therefore nature, in search of the "weakest link," *found this in the switch itself* — and produced an arc across it, to try and move the current across anyway.

But suppose we consciously provide a "diversionary path"? Then there would be no problem turning the switch OFF and stopping the inductor current flowing through the switch — because it could continue to flow via this alternate route. The inductor would then no longer "complain" in the form of a dangerous voltage spike. Thereafter, perhaps we can even re-route the current back into the switch when it turns ON again. Finally, we can perhaps even repeat the ON-OFF-ON-OFF process indefinitely, at a certain switching frequency.

In *Figure 1-10* we have created such an *alternate path*. We will see that the way the diode is pointed, this path can come into play automatically, and *only* when the switch turns OFF.

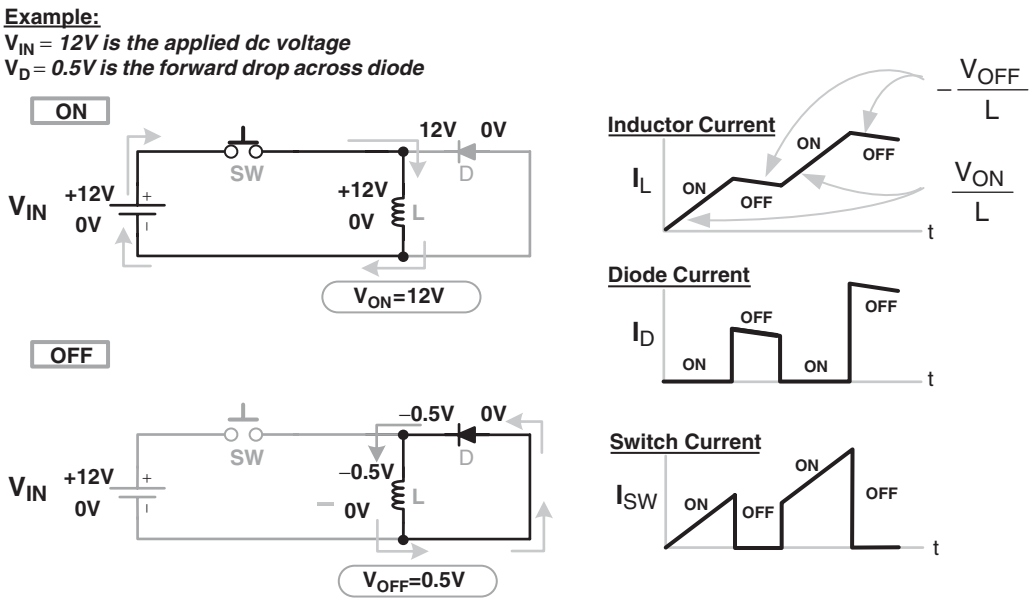


Figure 1-10: Providing a "Diversion" for the Inductor Current through a Diode

Just to make things clearer, we have used some sample numbers in *Figure 1-10*. We have taken the applied input voltage to be 12 V and assumed a typical Schottky diode forward drop of 0.5 V. Note that we are assuming a "perfect" switch here (no forward drop), for the sake of simplicity. We make the following observations:

- When the switch is ON (closed), the voltage at the upper end of the inductor L is at 12 V and the lower end is at 0 V ('ground'). So the diode is reverse-biased and does not conduct. Energy is then being built up in the inductor by the applied dc voltage source.

  The magnitude of the voltage applied across the inductor during the on-time of the switch (i.e. 'V_{ON}') is equal to 12 V.

- When the switch turns OFF (open), an alternate path is *available* for the inductor current to flow — through the diode. And we can be sure that "nature" (in our case the "induced voltage") will attempt to exploit this path — by forcing the diode to conduct. But for that, the diode must get '*forward-biased*,' that is, its anode must get to a voltage 0.5 V higher than the cathode. But the anode is being held at ground (0 V rail). Therefore, the cathode must fall to −0.5 V.

  The magnitude of the voltage applied across the inductor during the off-time of the switch (i.e. 'V_{OFF}') is equal to 0.5 V.

- Note that the induced voltage during the switch off-time has had its polarity reversed.

- The rate of rise of the current (in the inductor and switch) during the on-time is equal to V_{ON}/L. And during the off-time, the current ramps down (much more slowly), at a rate of V_{OFF}/L (in the inductor and diode).

- Yes, if we *wait long enough* the inductor current will finally ramp down to zero (inductor 'reset'). But if we don't wait, and turn the switch back ON again, the current will again start to ramp up (staircasing), as shown in *Figure 1-10*.

- Note that *both* the switch and diode currents have a "choppy" waveform — since one takes over where the other left off. This is in fact always true for any switching power converter (or topology).

Summarizing: We see that having provided a diversionary path for the current, the inductor isn't 'complaining' anymore, and there is no uncontrolled induced voltage spike anymore. But we certainly have now ended up with a possible problem of escalating currents. And come to think of it, neither do we have a *useful* output rail yet, which is what we are basically looking to do finally. In fact, *all that we are accomplishing in Figure 1-10 is dissipating some of the stored energy built-up in the inductor during the on-time, within the diode during the off-time.*

Achieving a Steady State and Deriving Useful Energy

We realize, that to prevent staircasing, we need to somehow induce *voltseconds balance*. Yes, as mentioned, we could perhaps wait long enough before turning the switch ON again. But that still won't give us a useful output rail.

To finally solve all our problems in one go, let us take a hint from our "natural world of voltages." Since we realize we are looking for an *output dc voltage rail*, isn't it natural to use a *capacitor* somewhere in the circuit of *Figure 1-10*? Let us therefore now interpose a capacitor in series with the diode, as shown in *Figure 1-11*. If we do that, the diode (freewheeling) current would charge the capacitor up — and hopefully the capacitor voltage would eventually reach a steady level 'V_O'! Further, *since that would increase the voltage drop appearing across the inductor during the off-time (V_{OFF}), it would increase the rate at which the inductor current can ramp down* — which we recognize was the basic problem with the circuit in *Figure 1-10*. So we are finally seeing light at the end of the tunnel — by making V_{OFF} comparable to V_{ON}, we are hoping to achieve *voltseconds balance* expressed by $V_{ON} \times t_{ON} = V_{OFF} \times t_{OFF}$.

In *Figure 1-11* the current escalates initially, but then after several cycles, *automatically* levels out, in what is clearly a steady state. That is because every cycle the capacitor charges up, it progressively increases the slope of the down-ramp, eventually allowing the converter to settle down *naturally* into the basic condition $\Delta I_{ON} = \Delta I_{OFF} \equiv \Delta I$. And once that is achieved, it is self-sustaining!

We also have a *useful* rail now — available across the output capacitor, from which we can draw some stored energy. So we have shown a dc current passing through to the load by the dashed arrows in *Figure 1-11*.

In fact, *this is our very first switching topology — the* ***buck-boost topology***.

> **Note:** Under the abnormal condition of an output short for example, *Figure 1-11* effectively reduces to *Figure 1-10*! Therefore, to protect the converter under such conditions, a *current limit* is required.

The Buck-boost Converter

To understand *Figure 1-11* better, we are actually going to work *backward* from here. So let us *assume* we have achieved a steady state — and therefore the output capacitor too has reached a steady value of say, 5 V. Let us now find the *conditions* needed to make that a reality.

In *Figure 1-11* the slope of the rising ramp is unchanged every cycle, being equal to V_{IN}/L. The slope of the falling ramp is initially V_D/L, where "V_D" is the drop across the diode. So from the inductor equation, initially $\Delta I_{ON} > \Delta I_{OFF}$. Thus the current starts to staircase. But the magnitude of the slope of the falling ramp, and therefore ΔI_{OFF}, keeps getting larger and

Example:
V_{IN} = *2V is the applied dc voltage*
V_D = *0.5V is the forward drop across diode*
V_O = *5V is the final value of voltage across output capacitor*

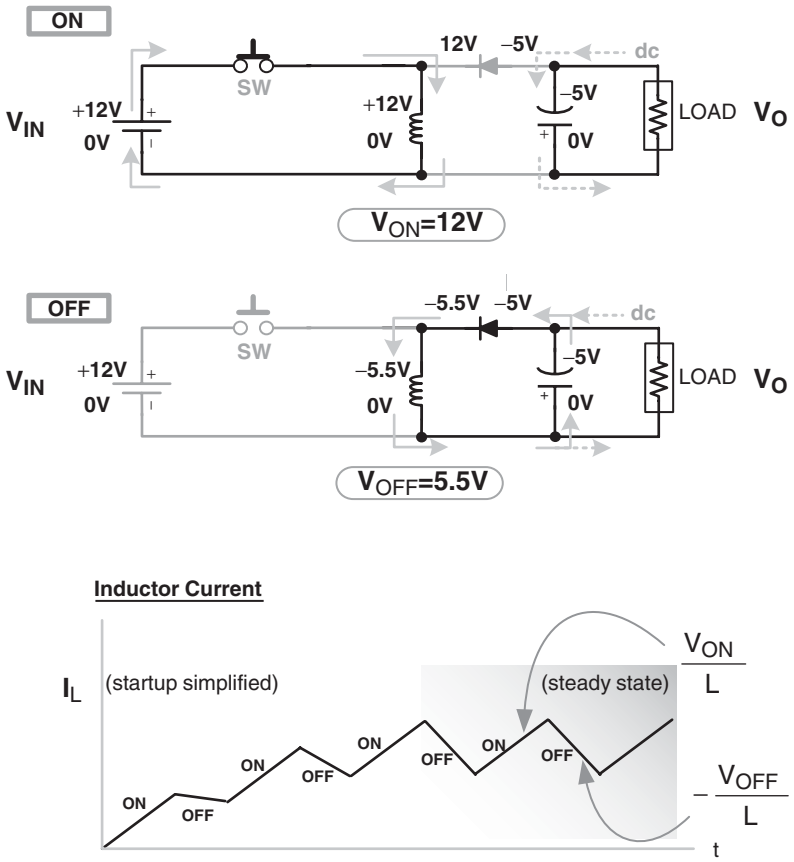


Figure 1-11: Evolution of the Buck-boost Topology

larger as the capacitor charges up. Eventually we will reach a steady state defined by $\Delta I_{OFF} = \Delta I_{ON}$. At that moment, the voltseconds law applies.

$$V_{ON} \times t_{ON} = V_{OFF} \times t_{OFF}$$

Using the numbers of the example, we get

$$12 \times t_{ON} = 5.5 \times t_{OFF}$$

We see that a 5 V output is possible only if we have been switching with a *constant ratio* between the switch ON and switch OFF time, as given by

$$\frac{t_{OFF}}{t_{ON}} = \frac{12}{5.5} = 2.18$$

So to get the voltseconds to balance out for this case (5 V output and a 12 V input), we have to make the *off-time* 2.18 times larger than the on-time. Why so? Simply because the voltage during the *on-time* (across the inductor) is larger by *exactly the same proportion*: 12 V during the on-time as compared to 5.5 V during the off-time. Check: 12/5.5 = 2.18.

The duty cycle (assuming CCM) is therefore equal to

$$D = \frac{t_{ON}}{t_{ON} + t_{OFF}} = \frac{1}{1 + \frac{t_{OFF}}{t_{ON}}} = \frac{1}{1 + 2.18} = 0.314$$

Now, had we taken a semiconductor switch instead of a mechanical one, we would have had a non-zero forward voltage drop, of say "V_{SW}." This forward drop effectively just subtracts from the applied dc input during the on-time. So, had we done the above calculations symbolically, we would get

$$V_{ON} = V_{IN} - V_{SW} \quad \textit{(Buck-boost)}$$

and

$$V_{OFF} = V_O + V_D \quad \textit{(Buck-boost)}$$

Then, from the voltseconds law

$$\frac{t_{OFF}}{t_{ON}} = \frac{V_{IN} - V_{SW}}{V_O + V_D} \quad \textit{(Buck-boost)}$$

We thus get the duty cycle

$$D = \frac{V_O + V_D}{V_{IN} - V_{SW} + V_O + V_D} \quad \textit{(Buck-boost)}$$

If the switch and diode drops are small as compared to the input and output rails, we can simply write

$$D \approx \frac{V_O}{V_{IN} + V_O} \quad \textit{(Buck-boost)}$$

We can also write the relationship between the input and output as follows

$$V_O = V_{IN} \times \frac{D}{1 - D} \quad \textit{(Buck-boost)}$$

Note that some other easily derivable, and convenient relationships to remember are

$$\frac{t_{ON}}{t_{OFF}} = \frac{D}{1 - D} \quad \textit{(any topology)}$$

$$t_{ON} = \frac{D}{f} \quad \textit{(any topology)}$$

$$t_{OFF} = \frac{1 - D}{f} \equiv \frac{D'}{f} \quad \textit{(any topology)}$$

where we have defined $D' = 1 - D$ as the 'duty cycle of the diode,' since the diode is conducting for the remainder of the switching cycle duration (in CCM).

Ground-referencing Our Circuits

We need to clearly establish what is referred to as the 'ground' rail in any dc-dc switching topology. We know that there are two rails by which we apply the dc input voltage (current goes in from one and returns from the other). Similarly, there are also two output rails. But all practical topologies generally have *one rail that is common to both the input and the output*. It is this *common* rail, that by convention, is called the system 'ground' in dc-dc converter applications.

However, there is yet another convention in place — the ground is also considered to be "0 V" (zero volts).

The Buck-boost Configurations

In *Figure 1-12* the common (ground) rails have been highlighted in bold gray background.

We now realize that the buck-boost we presented in *Figure 1-11* is actually a 'positive (input) to negative (output)' buck-boost. There is another possibility, as shown in the lower half of *Figure 1-12*. We have re-labeled its ground in accordance with the normal convention. Therefore this is a 'negative to positive buck-boost.'

For either configuration, we see that whatever polarity is present at the input, it gets reversed at the output. Therefore, *the buck-boost is often simply called an 'inverting' topology* (though we should keep in mind that that allows for *two different configurations*).

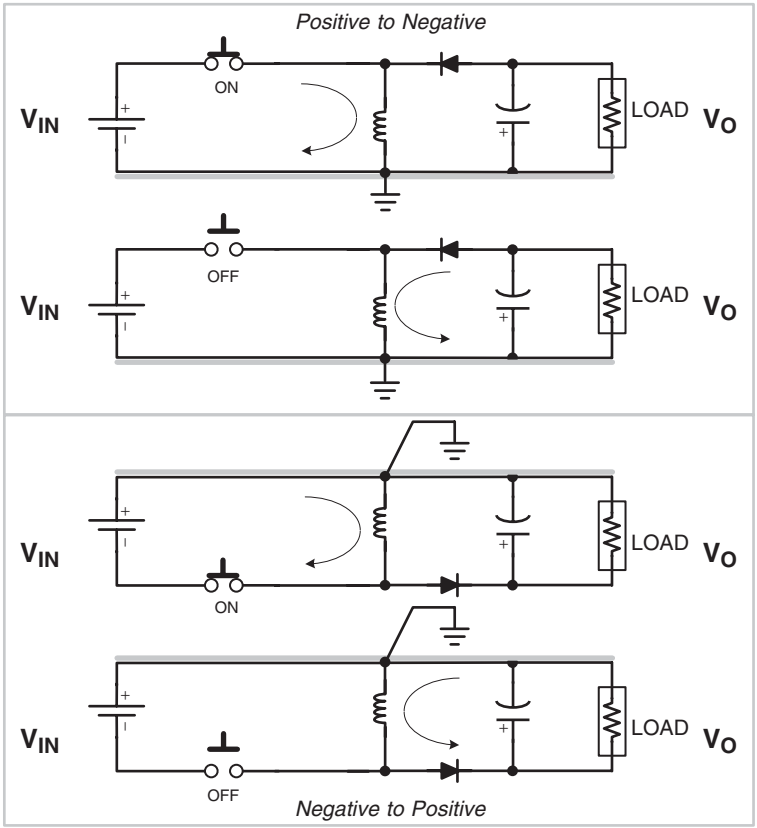


Figure 1-12: The Two Configurations of the Buck-boost (inverting) Topology

The Switching Node

Very simply put — the "point of detour" for the inductor current, that is, between the switch and the diode, is called the 'switching node.' Current coming into this node from the inductor, can go either into the diode or the switch, depending upon the state of the switch. Every dc-dc switching topology has this node (without it we would get the huge voltage spike we talked about!).

Since the current at this node needs to alternate between the diode and the switch, it needs to alternately force the diode to change state too (i.e. reverse-biased when the switch turns ON and forward-biased when the switch is OFF). So, the voltage at this node must necessarily be 'swinging.' An oscilloscope probe connected here (with its ground clip connected to the power supply ground, i.e. 0 V), will always see a voltage waveform with "square edges." This is in fact very similar to the voltage across the inductor, except that it is dc level-shifted by a certain amount, depending on the topology.

On a practical level, while designing the PCB (printed circuit board), we have to be cautious in not putting too much copper at the switching node. Otherwise it becomes an effective electric-field antenna, spewing radiated radio frequency interference all around. The output cables can thereafter pick up the radiated noise and transmit it directly to the load.

Analyzing the Buck-boost

In *Figure 1-13* we have drawn a line 'I_L' through the *geometric center* of the ramp portion of the steady-state inductor current waveform. This is defined as the *average inductor current*. The switch current also has an average value of I_L during the interval t_{ON}. Similarly the average of the diode current is also I_L, during t_{OFF}. However, the switch and diode currents when *averaged over the entire cycle* (i.e. over both the ON and OFF durations) are by simple

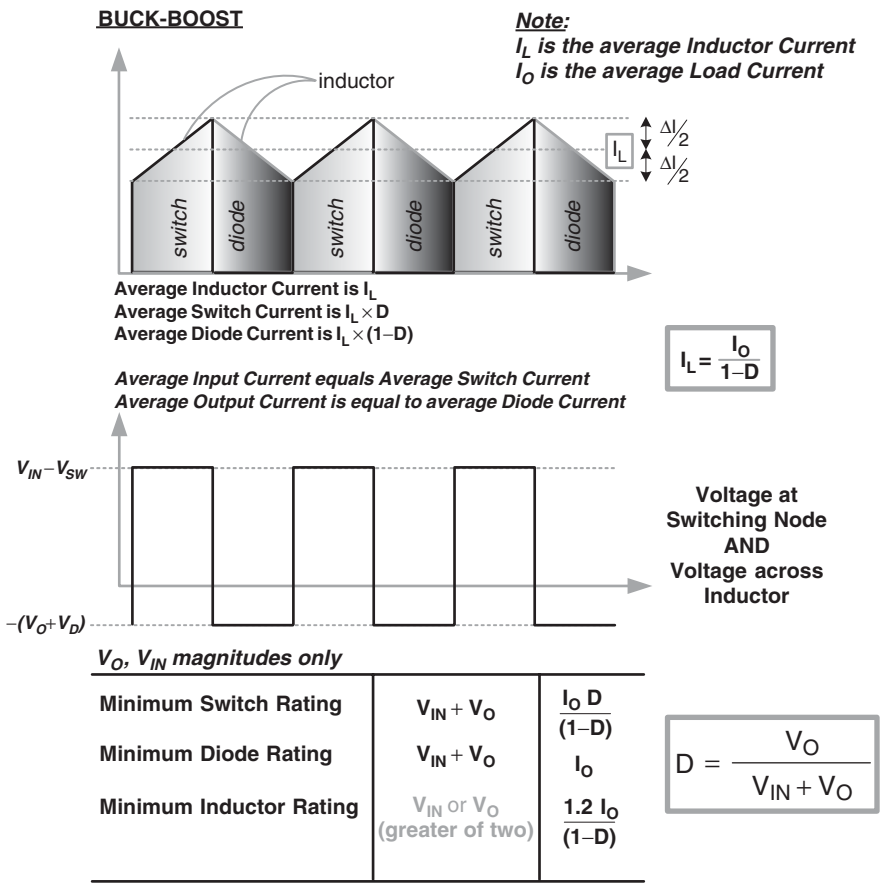


Figure 1-13: Analyzing the Buck-boost

mathematics their respective *weighted* averages.

$$I_{SW\_AVG} = I_L \times \frac{t_{ON}}{T} = I_L \times D \quad \textit{(Buck-boost)}$$

$$I_{D\_AVG} = I_L \times \frac{t_{OFF}}{T} = I_L \times D' = I_L \times \left(1 - D\right) \quad \textit{(Buck-boost)}$$

where D' is the duty cycle of the diode, that is, $1 - D$. It is also easy to visualize that for this particular topology, the *average input current is equal to the average switch current*. Further, as we will see below, the *average diode current* is equal to the *load current*. This is what makes the buck-boost topology quite different from the buck topology.

Properties of the Buck-boost

We now make some observations based on *Figure 1-11*, *Figure 1-12*, and *Figure 1-13*

- For example, a "positive to negative" buck-boost can convert 12 V to −5 V (step down) or 12 V to −15 V (step up). A "negative to positive" buck-boost can convert say −12 V to 5 V or 5 V to 15 V, and so on. The *magnitude* of the output voltage can thus be either smaller or larger (or equal to) the *magnitude* of the input voltage.
- When the switch is ON, energy is delivered *only* into the inductor by the input dc source (via the *switch*), and *none* of it passes through to the output.
- When the switch is OFF, *only* the stored energy of the inductor is pushed into the output (through the *diode*), and *none* comes directly from the input dc source.
- The above two observations make the buck-boost topology *the only "pure flyback" topology around, in the sense that* ***all*** *the energy transferred from the input to the output, must have been previously stored in the inductor*. No other topology shares this unique property.
- The current coming from the *input capacitor* (dc source) is "choppy," that is, pulsating. That is because, this current, combined with the steady dc current (I_{IN}) coming in from the dc source, basically forms the *switch* current waveform (which we know is always choppy for any topology) (see *Figure 1-9*).
- Similarly, the current into the *output capacitor* is also choppy, because combined with the steady dc current into the load (I_{OUT}), it forms the diode current (which we know is always choppy for any topology) (see *Figure 1-9*).
- We know that heat dissipation is proportional to the square of the RMS current. And since choppy waveforms have high RMS values, the efficiency of a buck-boost is not very good. Also, there is generally a relatively high level of noise and ripple across

the board. Therefore, the buck-boost may also demand much better filtering at its input, and often at its output too.

- Though current enters the output capacitor to charge it up when the switch turns ON, and leaves it to go into the load when the switch is OFF, the *average capacitor current is always zero*. In fact, any capacitor in 'steady state' must, by definition, have *zero average current* passing through it — otherwise it would keep charging or discharging until it too reaches a steady state, just like the inductor current.

  Since the average current from the output capacitor is zero, therefore, for *the buck-boost, the average diode current must be equal to the load current* (where else can the current come from?). Therefore

$$I_{D\_AVG} = I_O = I_L \times (1 - D)$$

  So,

$$I_L = \frac{I_O}{1 - D} \quad \textit{(Buck-boost)}$$

This is the relationship between the average inductor current and the load current. Note that in *Figure 1-13*, in the embedded table, we have asked for an inductor rated for $1.2 \times I_O/(1 - D)$. The factor "1.2" comes from the fact, that by typical design criteria, the peak of the inductor current waveform is about 20% higher than its average. So we need to look for an inductor rated at least for a current of $1.2 \times I_L$.

Why Three Basic Topologies Only?

There are certainly several ways to set up circuits using an inductor, which provide a "freewheeling path" too, for the inductor current. But some of these are usually disqualified simply because the input and output do not share a common rail, and thus there is no proper *ground reference* available for the converter and the rest of the system. Two examples of such "working-but-unacceptable" converters are the buck-boost configurations shown in *Figure 1-14*. Compare these with *Figure 1-12* to see what the problem is! However, note that if these were "front-end converters," the system ground could be established starting at the output of this converter itself, and may thus be acceptable.

Of the remaining ways, several are just "configurations" of a *basic topology* (like the two configurations in *Figure 1-12*). Among the basic topologies, we actually have just three — the buck, the boost, and the buck-boost. Why only three? That is because of *the way the inductor is connected*. Note that with proper ground-referencing in place, there are only three distinct rails possible — the input, the output, and the (common) ground. So if one end of the inductor is connected to the ground, it becomes a buck-boost! On the other hand, if it is

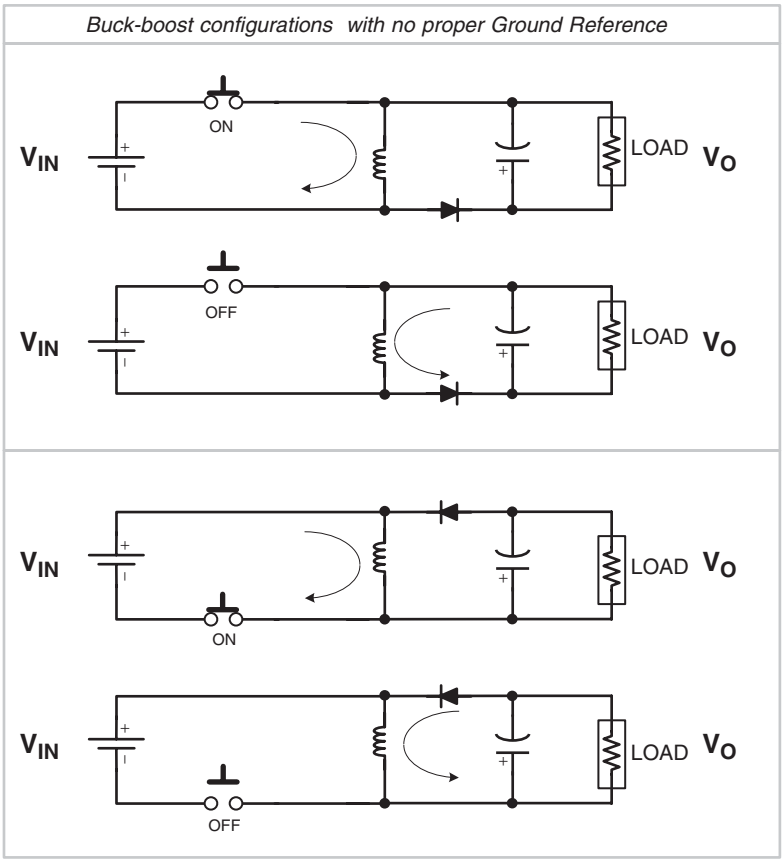


Figure 1-14: Improperly Referenced Buck-boost Configurations

connected to the input, it becomes a boost. And if connected to the output, it becomes a buck. See *Figure 1-15*.

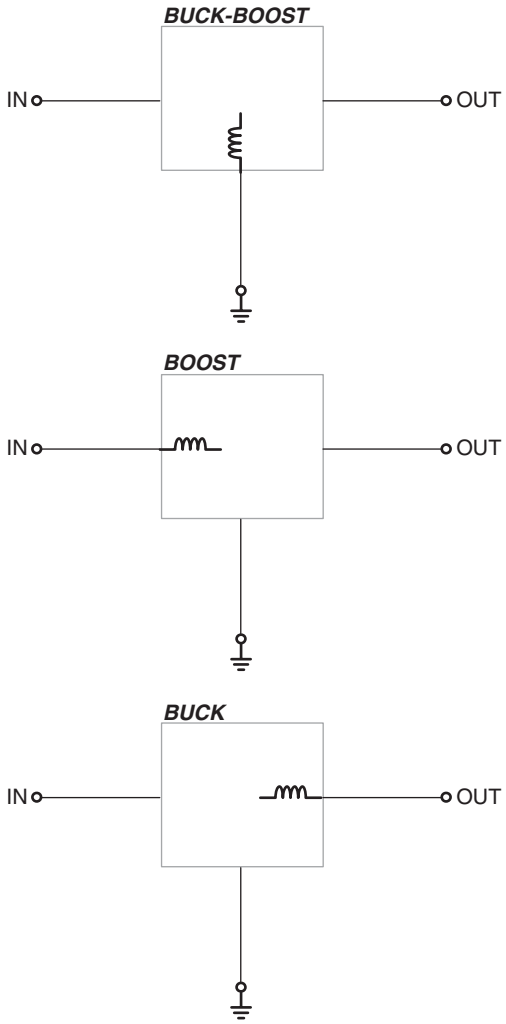
The Boost Topology

In *Figure 1-16* we have presented the schematic of the boost topology. The direct and the freewheeling paths are indicated therein. In *Figure 1-17*, we have the corresponding analysis, including the key waveforms.

We now make some observations

- For example, a "positive to positive" boost can convert 12 V to 50 V. A "negative to negative" boost would be able to convert say −12 V to −50 V. The *magnitude* of the output voltage must therefore always be larger than the *magnitude* of the input voltage. So a boost converter only steps-up, and also does not change the polarity.

Figure 1-15: Three Basic Topologies Possible Only



- When the switch is ON, energy is delivered *only* into the inductor by the input dc source (via the *switch*), and *none* of it passes through to the output.
- When the switch is OFF, the stored energy of the inductor is pushed into the output (through the *diode*). ***But some of it also comes from the input dc source***.
- The current coming from the *input capacitor* (dc source) is "smooth," since it is in series with the inductor (which prevents sudden jumps in current).
- However, the current into the *output capacitor* is "choppy," because combined with the steady dc current into the load (I_{OUT}), it forms the diode current (which we know is always choppy for any topology) (see *Figure 1-9*).

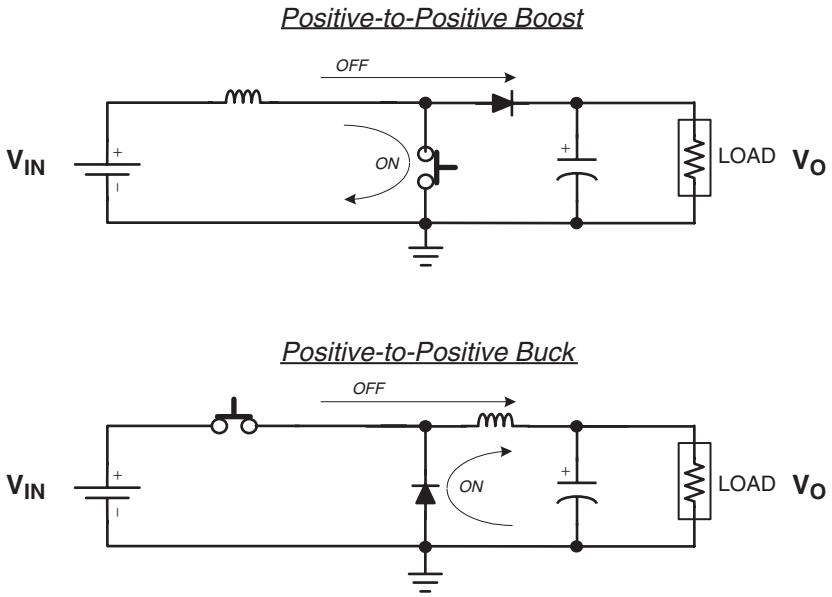


Figure 1-16: The (positive) Boost and Buck Topologies

- Since the average current from the output capacitor is zero, therefore, for *the boost, the average diode current must be equal to the load current* (where else can the current come from?). Therefore

$$I_{D\_AVG} = I_O = I_L \times (1 - D)$$

So,

$$I_L = \frac{I_O}{1 - D} \quad \textit{(Boost)}$$

This is the relationship between the average inductor current and the load current. Note that in *Figure 1-17*, in the embedded table, we have asked for an inductor rated for $1.2 \times I_O/(1 - D)$. The factor "1.2" comes from the fact, that by typical design criteria, the peak of the inductor current waveform is about 20% higher than its average. So we need to look for an inductor rated at least for a current of $1.2 \times I_L$.

Let us analyze the boost topology in terms of the voltseconds in steady state. We have

$$V_{ON} = V_{IN} - V_{SW} \quad \textit{(Boost)}$$

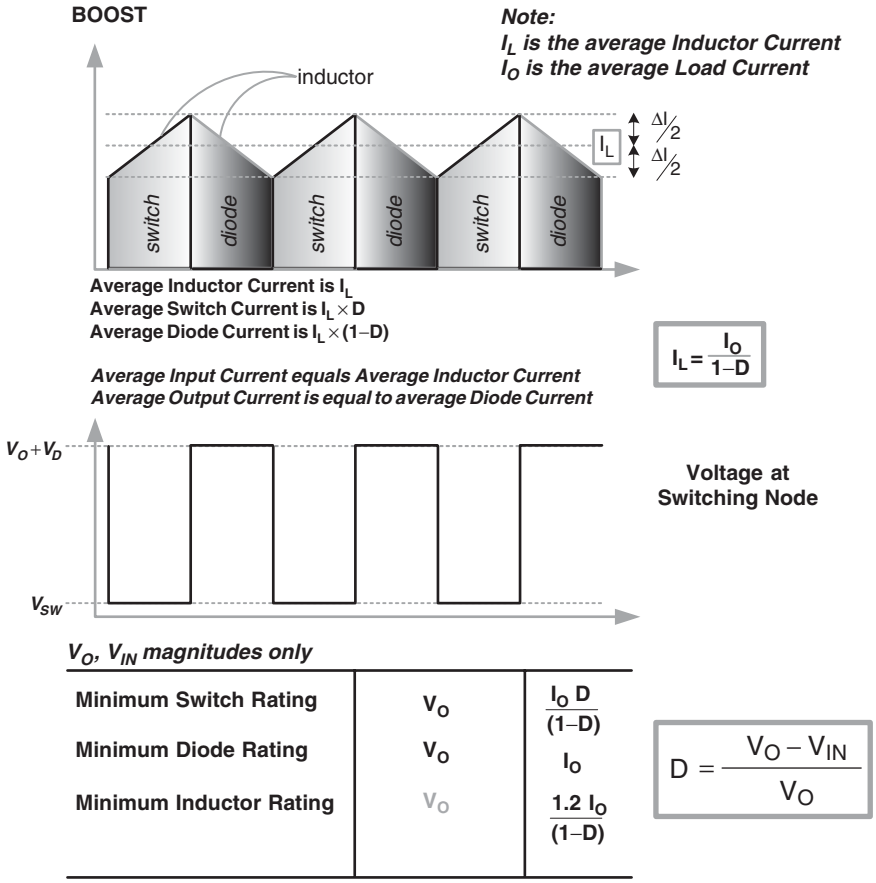


Figure 1-17: Analyzing the Boost

and

$$V_{OFF} = V_O + V_D - V_{IN} \quad \textit{(Boost)}$$

So, from the voltseconds law

$$\frac{t_{OFF}}{t_{ON}} = \frac{V_{IN} - V_{SW}}{V_O + V_D - V_{IN}} \quad \textit{(Boost)}$$

Performing some algebra on this to eliminate t_{OFF}

$$\frac{t_{OFF}}{t_{ON}} + 1 = \frac{V_{IN} - V_{SW}}{V_O + V_D - V_{IN}} + 1$$

$$\frac{t_{OFF} + t_{ON}}{t_{ON}} = \frac{V_{IN} - V_{SW} + V_O + V_D - V_{IN}}{V_O + V_D - V_{IN}}$$

Finally, the 'duty cycle' of the converter D, which is defined as

$$D = \frac{t_{ON}}{T} \quad \textit{(any topology)}$$

is the reciprocal of the preceding equation. So

$$D = \frac{V_O + V_D - V_{IN}}{V_O + V_D - V_{SW}} \quad \textit{(Boost)}$$

We have thus derived the classical dc transfer function of a *boost converter.*

If the switch and diode drops are small as compared to the input and output rails, we can just write

$$D \approx \frac{V_O - V_{IN}}{V_O} \quad \textit{(Boost)}$$

We can also write the relationship between the input and output as follows

$$V_O = V_{IN} \times \frac{1}{1 - D} \quad \textit{(Boost)}$$

The Buck Topology

In *Figure 1-16* we had also presented the schematic of the buck topology. The direct and the freewheeling paths are indicated therein. In *Figure 1-18*, we have the corresponding analysis, including the key waveforms.

We now make some observations

- For example, a "positive to positive" buck can convert 12 V to 5 V. A "negative to negative" buck would be able to convert say −12 V to −5 V. The *magnitude* of the output voltage must therefore always be smaller than the *magnitude* of the input voltage. So a buck converter only steps-down, and also does not change the polarity.
- When the switch is ON, energy is delivered to the inductor by the input dc source (via the *switch*). ***But some of it also passes through to the output***.
- When the switch is OFF, the stored energy of the inductor is pushed into the output (through the *diode*). And *none* of it now comes from the input dc source.
- The current coming from the *input capacitor* (dc source) is "choppy." That is because, this current, combined with the steady dc current (I_{IN}) coming in from the

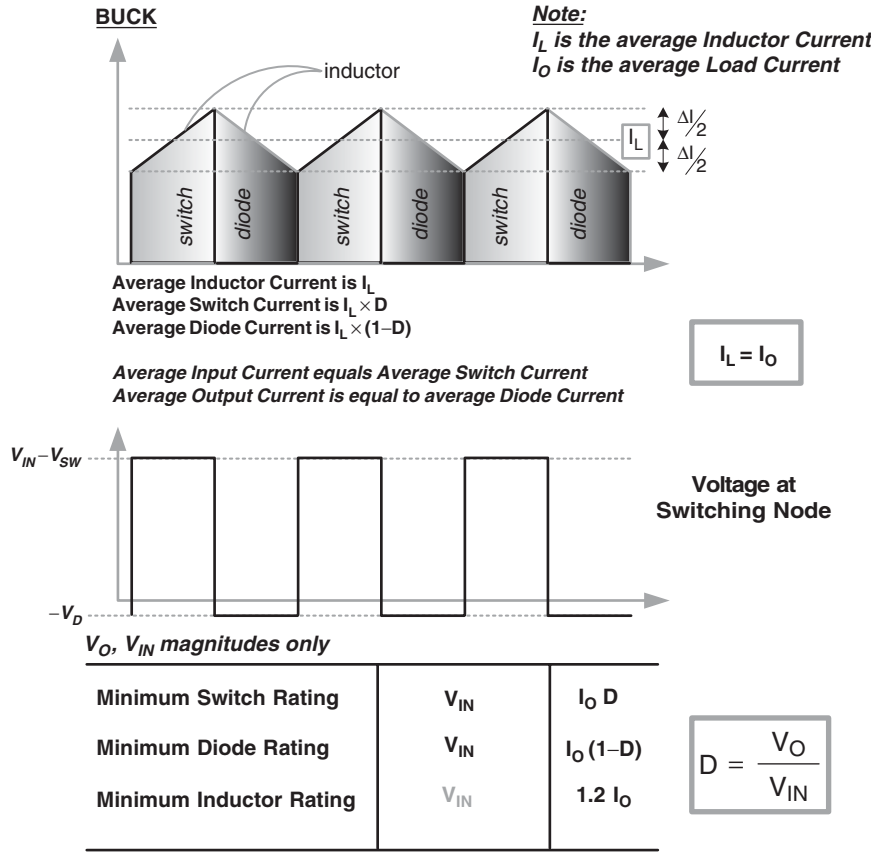


Figure 1-18: Analyzing the Buck

dc source, basically forms the *switch* current waveform (which we know is always choppy for any topology).

- However, the current into the *output capacitor* is "smooth," because it is in series with the inductor (which prevents sudden jumps in current).
- Since the average current from the output capacitor is zero, therefore, for *the buck, the average inductor current must be equal to the load current* (where else can the current come from?). Therefore

$$I_L = I_O \quad \textit{(Buck)}$$

This is the relationship between the average inductor current and the load current. Note that in *Figure 1-18*, in the embedded table, we have asked for an inductor rated for $1.2 \times I_O$. The factor "1.2" comes from the fact, that by typical design criteria, the peak of the inductor

current waveform is about 20% higher than its average. So we need to look for an inductor rated at least for a current of $1.2 \times I_L$.

Let us analyze the buck topology in terms of the voltseconds in steady state. We have

$$V_{ON} = V_{IN} - V_{SW} - V_O \quad \textit{(Buck)}$$

and

$$V_{OFF} = V_O - (-V_D) = V_O + V_D \quad \textit{(Buck)}$$

As before, using the voltseconds law and simplifying, we get the 'duty cycle' of the converter

$$D = \frac{V_O + V_D}{V_{IN} + V_D - V_{SW}} \quad \textit{(Buck)}$$

We have thus derived the classical dc transfer function of a *buck converter.* If the switch and diode drops are small as compared to the input and output rails, we can just write

$$D \approx \frac{V_O}{V_{IN}} \quad \textit{(Buck)}$$

We can also write the relationship between the input and output as follows

$$V_O = V_{IN} \times D \quad \textit{(Buck)}$$

Advanced Converter Design

This should serve as an introduction to understanding and designing switching power converters. More details and worked examples can be found in the next **chapter** (titled "DC-DC Converter Design and Magnetics"). The reader can also at this point briefly scan *Chapter 4* for some finer nuances of design. A full design table is also available in *Appendix 2* for future reference.

CHAPTER

2

DC-DC Converter Design and Magnetics

This Page Intentionally Left Blank

CHAPTER 2

DC-DC Converter Design and Magnetics

The reader is strongly advised to read *Chapter 1* before attempting this chapter.

The magnetic components of any switching power supply are an integral part of its topology. The design and/or selection of the magnetics can affect the selection and cost of all the other associated power components, besides dictating the overall performance and size of the converter itself. Therefore, we really should not try to design a converter, without looking closely at its magnetics, and vice versa. With that in mind, in this chapter, we will be introducing the basic concepts of magnetics — in parallel with a formal dc-dc converter design procedure.

Note that in the area of *dc-dc converters*, we have only a single magnetic component to consider — its inductor. Further, in this particular area of power conversion, it is customary to just pick an *off-the-shelf* inductor for most applications. Of course there cannot possibly be enough "standard" inductors going around to cover all possible application scenarios. But the good news is that, given a certain inductor, and knowing its performance under a *stated set of conditions*, we can easily calculate how it will perform under our specific *application conditions*. And thereby, we can either validate or invalidate our initial selection. It may take more than one iteration or attempt, but moving in this direction, we can almost always find a standard inductor that fits our application.

In the next chapter we will take up *"off-line"* power supply design. Such converters usually work off an ac (mains) input that ranges from 90 to 270 Volts. To protect users from the high voltage, these converters almost invariably use an isolating *transformer* — in addition to or in place of the inductor. But though these topologies are really just derivatives of standard dc-dc topologies, in terms of their *magnetics*, they are quite *different*. For example, we encounter significant (non-negligible) *high-frequency* effects within the transformer — like skin depth and proximity effects — the analysis of which can be quite challenging.
In addition, we find that there are definitely not enough general-purpose (off-the-shelf) parts going around, that can meet all possible permutations and combinations of requirements, as can arise in off-line applications. So in these applications, we usually always end up having to *custom-design* the magnetics. And as mentioned, this is not a mean task. But by trying to first understand *dc-dc converter* design, and the selection of *off-the-shelf* inductors, we are in a much better position to tackle off-line power supplies. We can thereby build up basic concepts and skills, while garnering a much-needed "feel" for magnetics.

Off-line converters and dc-dc converters are also relatively quite different in terms of some rather implicit (often completely unstated) differences in basic design strategy — like the issue relating to the size of the magnetics vis-à-vis the current limit of the converter, as we will soon learn. With regard to their *similarities*, we should remember that both can have a *wide-input* voltage range, *not a single-value input voltage*, as is often assumed in related literature. Having a wide-input raises the following question — what voltage point within the prescribed input range is the "worst-case" (or maximum) for a given stress parameter? Note that in selecting a power component we often need to consider the *worst-case* stress it is going to endure in our application. And then, provided that that particular stress parameter happens to be a relevant and decisive factor in its selection, we usually add an additional amount of safety margin, for the sake of reliability. However, the problem is that *different stress parameters do not attain their worst-case values at the same input voltage point.* We therefore realize that the design of a wide-input converter is necessarily going to be "tricky." For sure, designing a *functional* switching converter may be considered "easy," but designing it *well* certainly isn't.

Toward the end of this chapter, we will finally present the detailed dc-dc converter design procedure. But to account for a wide-input range, we will proceed in two distinct steps:

- A *"general inductor design procedure,"* for choosing and validating an off-the-shelf inductor for our application. We will see that depending on the topology at hand, this is to be carried out at a certain, specified voltage end — one that we will identify as being the "worst-case" *from the viewpoint of the inductor*.
- Then we will consider the other power components. We will point out which particular stress parameters are important in each case, and also the input voltage at which they reach their maximum, and how to ultimately select the component.

Note that, although the design procedure may be seen to specifically address only the buck topology, the accompanying annotations clearly indicate how a particular step or equation may need to change if the procedure were being carried out for a boost or a buck-boost topology.

DC Transfer Functions

When the switch turns ON, the current ramps up in the inductor according to the inductor equation $V_{ON} = L \times \Delta I_{ON}/t_{ON}$. The current *increment* during the on-time is $\Delta I_{ON} = (V_{ON} \times t_{ON})/L$. When the switch turns OFF, the inductor equation $V_{OFF} = L \times \Delta I_{ON}/t_{OFF}$ leads to a current *decrement* $\Delta I_{OFF} = (V_{OFF} \times t_{OFF})/L$.

Table 2-1: Derivation of dc transfer functions of the three topologies

Applying Voltseconds Law and $D = t_{ON}/(t_{ON} + t_{OFF})$			
Steps	$V_{ON} \times t_{ON} = V_{OFF} \times t_{OFF}$ $\frac{t_{ON}}{t_{OFF}} = \frac{V_{OFF}}{V_{ON}}$ $\frac{t_{ON}}{t_{ON} + t_{OFF}} = \frac{V_{OFF}}{V_{OFF} + V_{ON}}$ *Therefore,* $D = \frac{V_{OFF}}{V_{ON} + V_{OFF}}$ *(duty cycle equation for all topologies)*		
	Buck	***Boost***	***Buck-Boost***
V_{ON}	$V_{IN} - V_O$	V_{IN}	V_{IN}
V_{OFF}	V_O	$V_O - V_{IN}$	V_O
DC Transfer Functions	$D = \frac{V_O}{V_{IN}}$	$D = \frac{V_O - V_{IN}}{V_O}$	$D = \frac{V_O}{V_{IN} + V_O}$

The current increment ΔI_{ON} must be equal to the decrement ΔI_{OFF}, so that the current at the end of the switching cycle returns to the *exact* value it had at the start of the cycle — otherwise we wouldn't be in a repeatable (steady) state. Using this argument, we can derive the input-output (dc) transfer functions of the three topologies, as shown in *Table 2-1*. It is interesting to note that the reason the transfer functions turn out different in each of the three cases can be traced back to the fact that the *expressions for V_{ON} and V_{OFF} are different.* Other than that, the derivation and its underlying principles remain the same *for all topologies*.

The DC Level and the "Swing" of the Inductor Current Waveform

From $V = LdI/dt$, we get $\Delta I = V\Delta t/L$. So the *"swinging"* component of the inductor current "ΔI" is *completely* determined by the applied voltseconds and the inductance. Voltseconds is the *applied voltage multiplied by the time that it is applied for*. To calculate it, we can either use V_{ON} times t_{ON} (where $t_{ON} = D/f$), or V_{OFF} times t_{OFF} (where $t_{OFF} = (1 - D)/f$) — and we will get the same result (for that is how D gets defined in the first place!). But note also, that if we apply 10 V across a given inductor for 2 μs, we will get the same current swing ΔI, if we apply say, 20 V for 1 μs, or 5 V for 4 μs, and so on. So, *for a given inductor*, either talking about the voltseconds or about ΔI is effectively one and the same thing.

Table 2-2: How varying the inductance, frequency, load current, and duty cycle influence ΔI and I_{DC}

		Action:											
		L ↑ (increasing)			I_O ↑ (increasing)			D ↑ (increasing)			f ↑ (increasing)		
		Buck	Boost	Buck-Boost	Buck	Boost	Buck-Boost	Buck	Boost	Buck-Boost	Buck	Boost	Buck-Boost
Response:	**Δ I = ?**	↓	↓	↓	×	×	×	↓	↑ ↓*	↓	↓	↓	↓
	I_{DC} = ?	×	×	×	↑(=)	↑	↑	×	↑	↑	×	×	×

↑↓ indicates it increases and decreases over the range

** maximum at D = 0.5*

'×' indicates no change

↑ (=) indicates — I_{DC} is increasing and is equal to I_O

What does the voltseconds depend on? It depends on the input/output voltages (duty cycle) and the switching frequency. ***Therefore, only by changing L, f, or D can we affect ΔI***. Nothing else! See *Table 2-2*. In particular, *changing the load current I_O does nothing to* ΔI. I_O is therefore, in effect, an altogether *independent* influence on the inductor current waveform. But what part of the inductor current does it specifically influence/determine? We will see that *I_O is proportional to the average inductor current*.

The inductor current waveform is considered to have another (independent) component besides its swing ΔI — this is the *dc* (average) level "I_{DC}," defined as the level *around* which the swing ΔI takes place symmetrically — that is, ΔI/2 above it, and ΔI/2 below it. See *Figure 2-1*. Geometrically speaking, this is the "center of the ramp." It is sometimes also called the "platform" or "pedestal" of the inductor current. The important point to note is that I_{DC} is based only on *energy flow requirements* — that is, the need to maintain an *average* rate of energy flow consistent with the input/output voltages and desired output power. So if the "application conditions," that is, the output power and the input/output voltages, do not change, there is in fact *nothing* we can do to alter this dc level — in that sense, I_{DC} is rather "stubborn" (see *Figure 2-1*). In particular

- Changing the inductance L doesn't affect I_{DC}.
- Changing the frequency f doesn't affect I_{DC}.
- Changing the duty cycle D *does* affect I_{DC} — for the boost and buck-boost.

To understand the last bullet above, we should note the following equations that we will derive a little later

$$I_{DC} = I_O \quad \textit{(buck)}$$

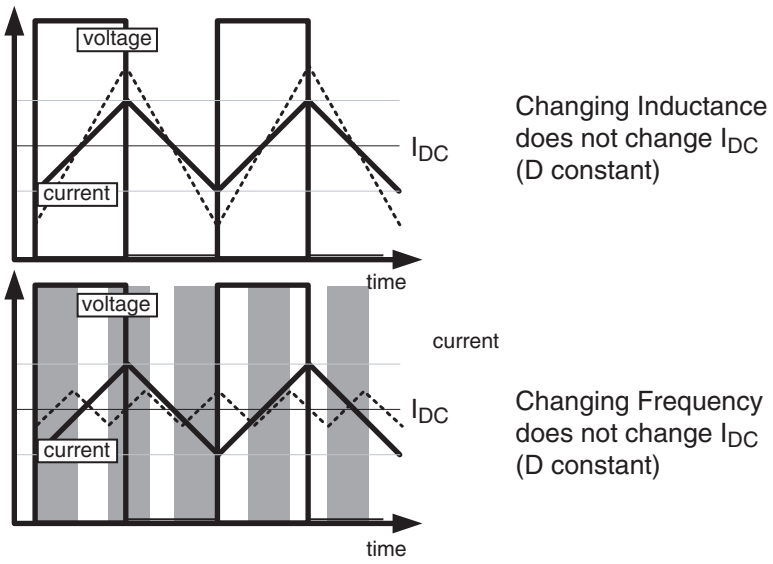


Figure 2-1: If *D* and I_O Are Fixed, I_{DC} Cannot Change

$$I_{DC} = \frac{I_O}{1 - D} \quad \textit{(boost and buck-boost)}$$

The intuitive reason why the above relations are different is that in a buck, the output is in series with the inductor (from the standpoint of the dc currents — the output capacitor contributing nothing to the dc current distribution), and therefore the average inductor current must at all times be equal to the load current. Whereas, in a boost and buck-boost, the output is likewise in series with the diode, and so the average diode current is equal to the load current.

Therefore, if we keep the load current constant, and change only the input/output voltages (duty cycle), we can affect I_{DC} — *in all cases* ***except*** *for the buck*. In fact, the *only* way to change the dc inductor current level for a buck is to change the load current. Nothing else will work!

In the buck, I_{DC} and I_O are equal. But in the boost and buck-boost, I_{DC} depends also on the duty cycle. That makes the design/selection of magnetics for these two topologies rather different from a buck. For example, if the duty cycle is 0.5, their average inductor current is twice the load current. Therefore, using a 5 A inductor for a 5 A load current may be a recipe for disaster.

One thing we can be sure of is that in the boost and buck-boost, I_{DC} is *always greater* than the load current. We may be able to cause this dc level to fall and even approach the load current value if we reduce the *duty cycle* close to 0 (i.e. a very small *difference between the input and output voltages*). But then, on increasing the duty cycle toward 1, the dc level of the inductor current will climb steeply. It is important we recognize this clearly and early on.

Another thing we can conclude with certainty is that in *all* the topologies, the dc level of the inductor current is *proportional* to the load current. So doubling the load current for example (keeping everything else the same), doubles the dc level of the inductor current (whatever it was to start with). So in a boost with a duty cycle of 0.5 for example, if we have a 5 A load, then the I_{DC} is 10 A. And if I_O is increased to 10 A, I_{DC} will become 20 A.

Summarizing, changing the input/output voltages (duty cycle) does affect the dc level of the inductor current for the boost and the buck-boost. But changing D affects the *swing* ΔI *in all three topologies*, because it changes the duration of the applied voltage and thereby changes the voltseconds.

- Changing the duty cycle affects I_{DC} for the boost and the buck-boost.
- Changing the duty cycle affects ΔI for all topologies.

Note: The off-line forward converter transformer is probably the only known exception to the above logic. We will learn that if we for example double the duty cycle (i.e double t_{ON}), then almost *coincidentally*, V_{ON} halves, and therefore the voltseconds does not change (and nor does ΔI). In effect, ΔI is then independent of duty cycle.

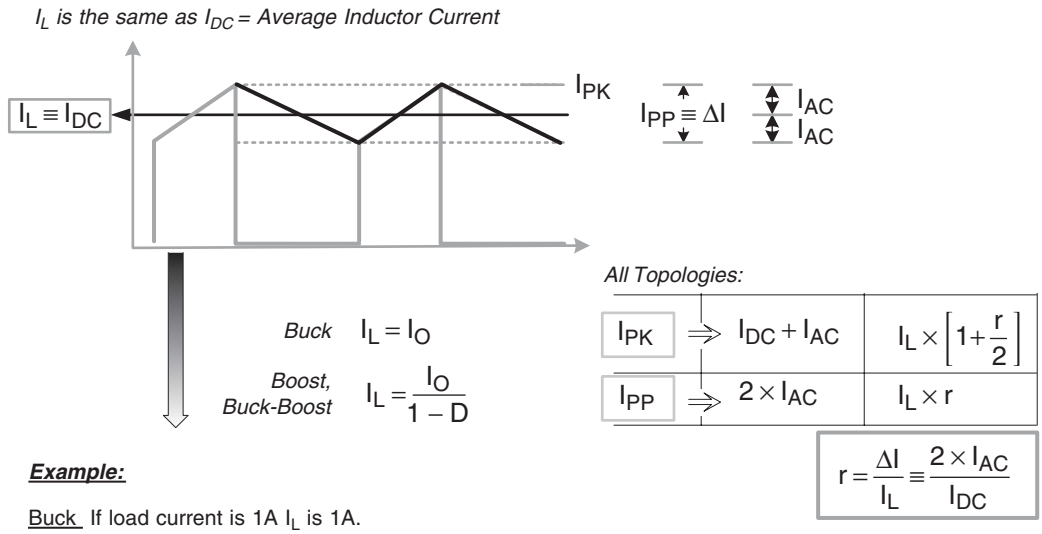
Based on the discussions above, and also the detailed design equations, we have summarized these "variations" in *Table 2-2*. This table should hopefully help the reader eventually develop a more intuitive and analytical "feel" for converter and magnetics design, one which can come in handy at a later stage. We will continue to discuss certain aspects of this table, in more detail, a little later.

Defining the AC, DC, and Peak Currents

In *Figure 2-2*, we see how the ac, dc, peak-to-peak, and peak values of the inductor current waveform are defined. In particular we note that the ac value of the current waveform is defined as

$$I_{AC} = \frac{\Delta I}{2}$$

We should also note from *Figure 2-2* that $I_L \equiv I_{DC}$. ***Therefore, sometimes in our discussions that follow, we may refer to the dc level of the inductor current as "I_{DC}," and sometimes as the average inductor current "I_L," but they are actually synonymous***. In particular we



Example:

Buck If load current is 1A I_L is 1A.
So if r=0.4 peak-to-peak current ('ΔI') is 0.4A and the peak current is 1.2A

Boost/Buck-Boost If load current is 1A D=0.5 I_L is 2A.
So if r=0.4, peak-to-peak current ('ΔI') is 0.8A and the peak current is 2.4A

Figure 2-2: The AC, DC, Peak, and Peak-to-Peak Currents, and the Current Ripple Ratio 'r' Defined

should not get confused by the subscript "L" in "I_L." The "L" stands for *inductor*, not *load*. The load current is always designated as "I_O." Of course, we do realize that $I_L = I_O$ for a *buck*, but that is just happenstance.

In *Figure 2-2* we have also defined another key parameter called 'r,' or the 'current ripple ratio.' This *connects the two independent current components I_{DC} and ΔI*. We will explore this particular parameter in much greater detail a little later. Here, it suffices to mention that *r* needs to be set to an "optimum" value in any converter — usually around 0.3 to 0.5, *irrespective of the specific application conditions, the switching frequency, and even the topology itself*. That therefore becomes a universal design *rule-of-thumb*. We will also learn that the choice of *r* affects the current stresses and dissipation in all the power components, and thereby impacts their selection. Therefore, **setting *r* should be the *first step* when commencing any power converter design**.

The dc level of the inductor current (largely) determines the I^2R losses in the copper windings ('copper loss'). However, the final temperature of the inductor is also affected by another term — the '*core loss*' — that occurs *inside* the magnetic material (core) of the inductor. Core loss is, to a first approximation, determined *only by the* ac (swinging)

component of the inductor current (ΔI), and is therefore virtually independent of the dc level (I_{DC} or "dc bias").

We must pay the closest attention to the *peak current*. Note that in any converter, the terms 'peak inductor current', 'peak switch current' and 'peak diode' current are all *synonymous*. Therefore, in general, we just refer to all of them as simply the 'peak current' I_{PK} where

$$I_{PK} = I_{DC} + I_{AC}$$

The peak current is in fact the *most critical current component* of all — because it is not just a source of *long-term* heat buildup and consequent temperature rise, but a potential cause of *immediate* destruction of the switch. We will show later that the *inductor current is instantaneously proportional to the magnetic field* inside the core. So at the exact moment when the current reaches its peak value, so does this field. We also know that real-world inductors can 'saturate' (start losing their inductance) if the field inside them exceeds a certain "safe" level — that value being dependent on the actual *material* used for the core (not on the geometry, or number of turns or even the air-gap, for example). Once saturation occurs, we may get an almost *uncontrolled* surge of current passing through the switch — because, the ability to limit current (which is one of the reasons the inductor is used in switching power supplies in the first place), depends on the inductor *behaving* like one. Therefore, losing inductance is certainly not going to help! In fact, we *usually* cannot afford to allow the inductor to 'saturate' *even momentarily*. And for this reason, *we need to monitor the peak current closely* (usually on a cycle-by-cycle basis). As indicated, the peak is the likeliest point of the inductor current waveform where saturation can start to occur.

Note: A slight amount of core saturation *may turn out to be acceptable* on occasion, especially if it occurs only under temporary conditions, like power-up for example. This will be discussed in more detail later.

Understanding the AC, DC and Peak Currents

We have seen that the ac component ($I_{AC} = \Delta I/2$) is derivable from the voltseconds law. From the basic inductor equation $V = L dI/dt$, we get

$$2 \times I_{AC} = \Delta I = \frac{\text{voltseconds}}{\text{inductance}}$$

So the current swing $I_{PP} \equiv \Delta I$, can be *intuitively visualized* as "*voltseconds per unit inductance"*. If the applied voltseconds doubles, so does the current swing (and ac component). And if the inductance doubles, the swing (and ac component) is halved.

Let us now consider the dc level again. Note that *any capacitor has zero average (dc) current through it in steady-state*, so all capacitors can be considered to be missing altogether when

calculating dc current distributions. Therefore, for a buck, since energy flows into the output during *both* the on-time and off-time, and *via the inductor*, therefore the average *inductor* current must always be equal to the load current. So

$$I_L = I_O \quad \textit{(buck)}$$

On the other hand, in both the boost and the buck-boost, energy flows into the output only during the off-time, and *via the diode*. Therefore, in this case, the average *diode* current must be equal to the load current. Note that the diode current has an average value equal to I_L *when it is conducting* (see the dashed line passing through the center of the down-ramp in the upper half of *Figure 2-3*). If we calculate the average of this diode current over the *entire switching cycle*, we need to weight it by *its* duty cycle, that is, $1 - D$. Therefore, calling 'I_D' the average diode current, we get

$$I_D = I_L \times (1 - D) \equiv I_O$$

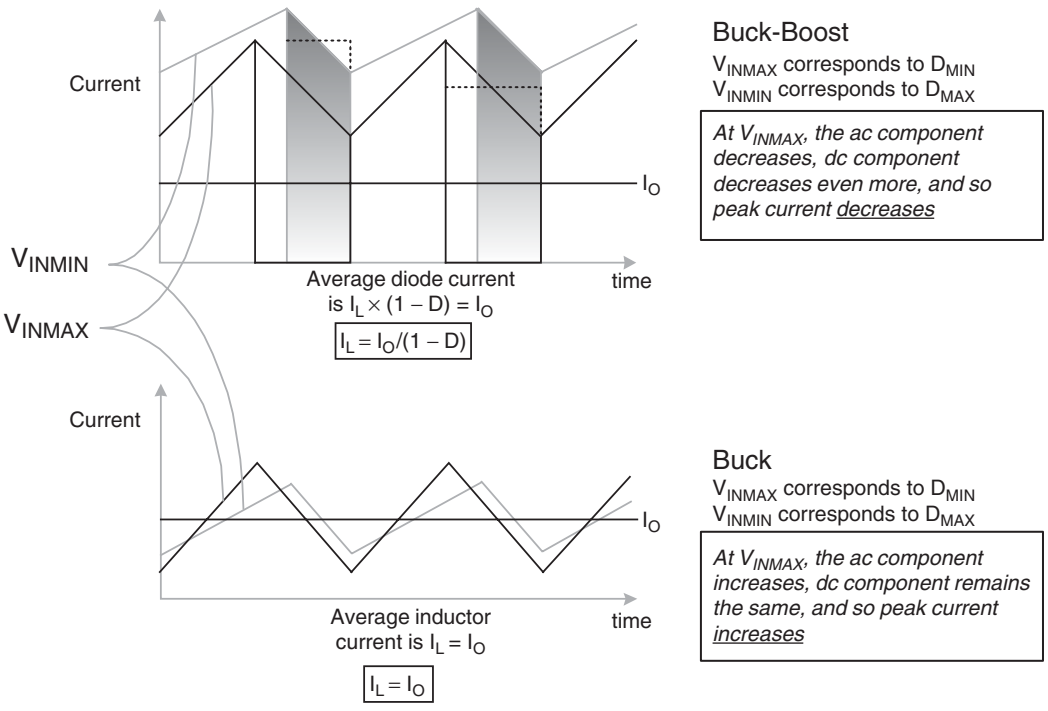


Figure 2-3: Visualizing the AC and DC Components of the Inductor Current as Input Voltage Varies

solving

$$I_L = \frac{I_O}{1 - D} \quad \textit{(boost and buck-boost)}$$

Note also, that ***for any topology, a high duty cycle corresponds to a low input voltage, and a low duty cycle is equivalent to a high input.*** So increasing D amounts to decreasing the input voltage (its magnitude) in all cases. Therefore, *in a boost or buck-boost, if the difference between the input and output voltages is large, we get the highest dc inductor current*.

Finally, with the dc and ac components known, we can calculate the peak current using

$$I_{PK} = I_{AC} + I_{DC} \equiv \frac{\Delta I}{2} + I_L$$

Defining the "Worst-case" Input Voltage

So far, we have been implicitly assuming a *fixed* input voltage. In reality, in most practical applications, the input voltage is a certain *range*, say from 'V_{INMIN}' to 'V_{INMAX}'. We therefore also need to know how the ac, dc, and peak current components *change as we vary the input voltage*. Most importantly, we need to know at what specific voltage within this range we get the maximum *peak* current. As mentioned, ***the peak is critical from the standpoint of ensuring there is no inductor saturation***. Therefore, defining the "worst-case" voltage (for inductor design) as the point of the input voltage range where the peak current is at its maximum, we need to design/select our inductor *at this particular point always*. This is in fact the underlying basis of the "*general inductor design procedure"* that we will be presenting soon.

We will now try to understand where and *why* we get the highest peak currents for each topology. In *Figure 2-3*, we have drawn various inductor current waveforms to help us better visualize what really happens as the input is varied. We have chosen two topologies here, the buck and the buck-boost, for which we display two waveforms each, corresponding to two different input voltages. Finally, in *Figure 2-4* we have plotted out the ac, dc, and peak values. Note that these plots are based on the actual design equations, which are also presented within the same figure. While interpreting the plots, we should again keep in mind that for all topologies, a high D corresponds to a low input. The following analysis will also explain certain cells of the previously provided *Table 2-2*, where the variations of ΔI and I_{DC}, with respect to D, were summarized.

a) ***For the buck***, the situation can be analyzed as follows:

- As the input *increases*, the duty cycle decreases in an effort to maintain regulation. But ***the slope of the down-ramp $\Delta I/t_{OFF}$ cannot change***, because it

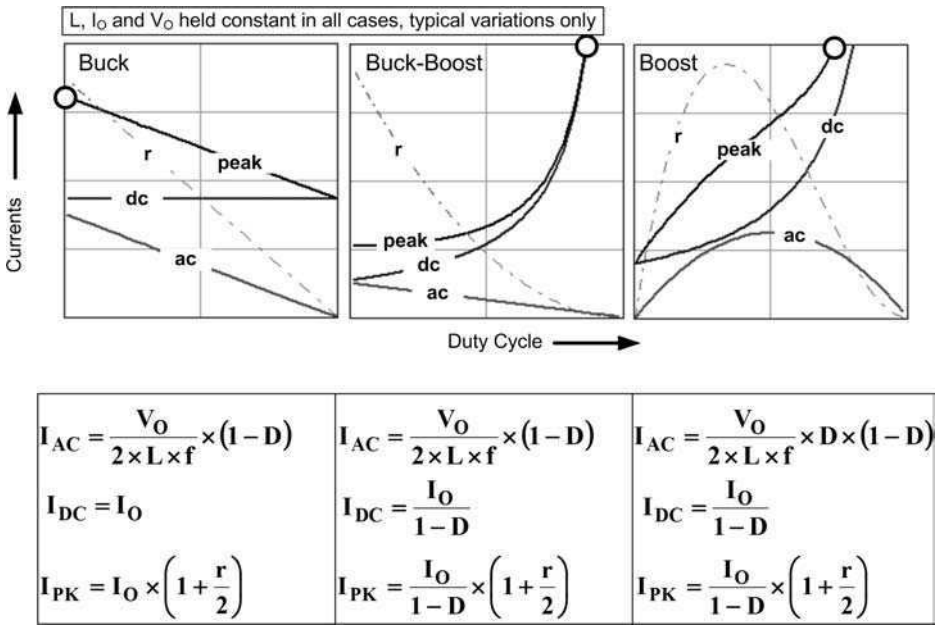


Figure 2-4: Plotting How the AC, DC, and Peak Currents Change with Duty Cycle

is equal to V_{OFF}/L, that is, V_O/L, and we are assuming V_O is fixed. But now, since t_{OFF} has increased, but the slope $\Delta I/t_{OFF}$ has not changed, the only possibility is that ΔI must have increased (proportionally). So we conclude — that the *ac component* of the buck inductor current actually *increases as the input increases* (even though the duty cycle decreased in the process).

- On the other hand, the center of the ramp I_L is fixed at I_O, so we know the *dc level does not change*.
- So finally, since the peak current is the *sum* of the ac and dc components, we realize it also *increases at high input voltages* (see relevant plot in *Figure 2-4*).

Therefore, ***for a buck, it is always preferable to start the inductor design at V_{INMAX} (i.e. at D_{MIN}).***

b) ***For the buck-boost,*** the situation can be analyzed as follows:

- As the input *increases*, the duty cycle decreases. But ***the slope of the down-ramp $\Delta I/t_{OFF}$ cannot change***, because it is equal to V_{OFF}/L, that is, V_O/L, and V_O is fixed (same situation as for the buck). But since t_{OFF} has increased, ΔI must also increase to keep the slope $\Delta I/t_{OFF}$ unchanged. So we see that the *ac component* ($\Delta I/2$) *increases as the input increases* (duty cycle

decreasing). Note that up till this point, the analysis is the same as for the buck — traced back to the fact that in both these topologies $V_{OFF} = V_O$.

- But now coming to the dc level I_L of the buck-boost, we will find it *must change* for this topology (though it remained fixed for the buck). Note that the shaded portion of the waveform in the upper half of *Figure 2-3* represents the diode current. The average value of this *during the off-time* is the square dashed line passing through its center, that is, I_L. So the average diode current, calculated *over the entire switching cycle*, is $I_L \times (1 - D)$. And we know this must equal the load current I_O. So, as the input *increases* and duty cycle decreases, the term $(1 - D)$ increases. So the only way $I_L \times (1 - D)$ can remain constant at the value I_O is if I_L *decreases correspondingly*. We therefore realize that the *dc level decreases as the input increases* (duty cycle decreasing).
- Further, since the peak current is the sum of the ac and dc components, it also *decreases at high input voltages* (see relevant plot in *Figure 2-4*).

Therefore, for a *buck-boost, we should always start the inductor design at* V_{INMIN} *(i.e. at* D_{MAX}*).*

c) ***For the boost***, the situation is a little *trickier* to understand. On the face of it, it is quite similar to the buck-boost, *but there is a notable difference* — and that is the reason why we did not even try to include it in *Figure 2-3*.

- Once again, as the input *increases*, the duty cycle decreases. But the difference here is that **the slope of the down-ramp $\Delta I/t_{OFF}$ must *decrease*** — because it is equal to V_{OFF}/L, that is, $(V_O - V_{IN})/L$ (magnitudes only) — and we know that $V_O - V_{IN}$ *is decreasing*. Further, the required decrease in the slope $\Delta I/t_{OFF}$ can come about in *two* ways — either from an increase in t_{OFF} (which is already occurring as the duty cycle decreases), *or* from a decrease in ΔI. But in fact, ΔI may *actually increase rather than decrease* (as we increase the input). For example, if t_{OFF} is increasing more the increase in ΔI — then $\Delta I/t_{OFF}$ will still decrease as required. And in practice, that is what actually does happen in the case of the boost. With some detailed math, we can show that ΔI increases as D *approaches 0.5*, but decreases on either side (see *Table 2-2* and *Figure 2-4*).
- It is therefore also clear that in either case above, the increase/decrease in the ac level *does not dominate*, and therefore, the peak current ends up being dictated *only by the dc component*. But we already know that the dc level of a boost changes in exactly the same way as for the buck-boost (discussed above) — *it decreases as the input increases* (duty cycle decreasing).

- So we conclude that the peak current for the boost also *decreases at high input voltages* (see relevant plot in *Figure 2-4*).

Therefore, for a *boost, we should always start the inductor design at* V_{INMIN} *(i.e. at* D_{MAX}*).*

The Current Ripple Ratio 'r'

In *Figure 2-2* we first introduced the most basic, yet far-reaching design parameter of the power supply itself — its *current ripple ratio* '*r*.' This is a geometrical ratio that compares and connects the ac value of the inductor current to its associated dc value. So

$$r = \frac{\Delta I}{I_L} \equiv 2 \times \frac{I_{AC}}{I_{DC}}$$

Here we have used $\Delta I = 2 \times I_{AC}$, as defined earlier in *Figure 2-2*. ***Once r is set by the designer (at maximum load current and worst-case input), almost everything else is pre-ordained*** — like the currents in the input and output capacitors, the 'RMS' (root mean square) current in the switch, and so on. Therefore, *the choice of r affects component selection and cost, and it must be understood clearly, and picked carefully.*

Note that the ratio *r* is defined for CCM (*continuous conduction mode*) operation only. Its valid range is from 0 to 2. When *r* is 0, ΔI must be 0, and the inductor equation then implies a very large (infinite) inductance. Clearly, $r = 0$ is not a practical value! If *r* equals 2, the converter is operating at the *boundary* of continuous and discontinuous conduction modes (boundary conduction mode or 'BCM'). See *Figure 2-5*. In this so-called boundary (or "critical") conduction mode, $I_{AC} = I_{DC}$ by definition. ***Note that readers can refer back to Chapter 1, in which CCM, DCM, and BCM were all initially introduced and explained.***

Note that an exception to the "valid" range of *r* from 0 to 2 occurs in 'forced CCM' mode, discussed in more detail later.

Relating r to the Inductance

We know that current swing is voltseconds per unit inductance. So we can also write

$$\Delta I = Et/L_{\mu H} \quad \textit{(any topology)}$$

Here *'Et'* is defined as the (magnitude of the) *volt*μ*seconds* across the inductor (either during the on-time or off-time — both being necessarily equal in steady state), and $L_{\mu H}$ is the inductance in μH. The reason for defining Et is that this number is simply easier to

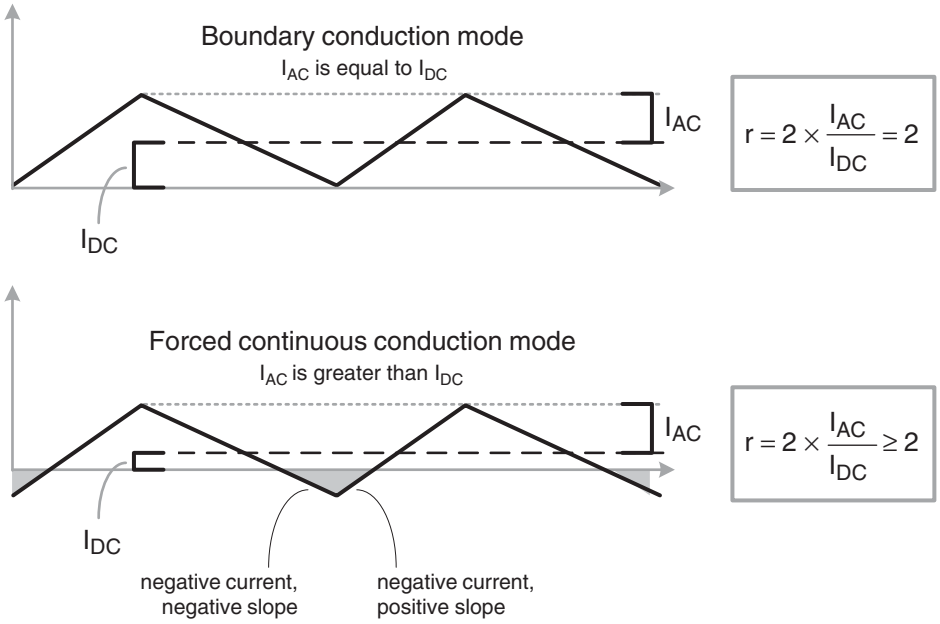


Figure 2-5: BCM and Forced CCM Operating Modes

manipulate than voltseconds because of the very small time intervals involved in modern power conversion.

Therefore, the current ripple ratio is

$$r = \frac{\Delta I}{I_L} = \frac{Et}{L_{\mu H} I_L} \quad \textit{(any topology)}$$

Note also that from now on, *whenever L is paired up with Et in any given equation, we will drop the subscript of L, that is, "μH." It will then be "understood" that L is in μH.*

Finally, we have the following key relationships between *r* and L

$$r = \frac{Et}{(L \times I_L)} \equiv \frac{V_{ON} \times D}{(L \times I_L) \times f} \equiv \frac{V_{OFF} \times (1 - D)}{(L \times I_L) \times f} \quad \textit{(any topology)}$$

Incidentally, the preceding equation, that is, the one involving V_{OFF}, assumes CCM, because it assumes that t_{OFF} (the time for which V_{OFF} is applied) is equal to the full available off-time $(1 - D)/f$.

Conversely, L as a function of *r* is

$$L = \frac{V_{ON} \times D}{r \times I_L \times f} \quad \textit{(any topology)}$$

In subsequent sections we will often use the following *easy-to-remember* form of the previous equations. We are going to nickname this the ***"L × I" equation*** (or rule)

$$L \times I_L = \frac{Et}{r} \quad \textit{(any topology)}$$

But perhaps we are still wondering — why do we even need to talk in terms of *r* — why not talk *directly* in terms of *L*? We do realize from the above equations that L and *r* are related. However, the "desirable" value of inductance depends on the *specific application conditions, the switching frequency, and even the topology*. So it is just not possible to give a general design rule for picking L. But there is in fact such a general design rule-of-thumb for selecting *r* — one that applies almost universally. We mentioned that it should be around 0.3 to 0.5 in all cases. And that is why ***it makes sense to calculate L by first setting the value of r***. Of course, once we pick *r*, L gets automatically determined — but only *for a given set of application conditions and switching frequency*.

The Optimum Value of r

It can be shown that, in terms of overall stresses in a converter and size, $r \approx 0.4$ represents an "optimum" of sorts. We will now try to understand why this is so, *and later we will try to point out exceptions to this reasoning*.

The size of an inductor can be thought of as being virtually proportional to its *energy-handling capability* (the effect of air-gap on size will be studied later). So for example, we probably already know intuitively that we need bigger cores to handle higher powers. The energy-handling capability of the selected core must, at a bare minimum, match the energy we need to store in it in our application — that is, $\frac{1}{2} \times L \times {I_{PK}}^2$. Otherwise the inductor will saturate.

In *Figure 2-6*, we have plotted the energy, $E = \frac{1}{2} \times L \times {I_{PK}}^2$, as a function of *r*. We see that it has a "knee" at around 0.4. This tells us that if we try to reduce *r* much lower than 0.4, we will certainly need a *very large inductor*. On the other hand, if we increase *r*, there isn't much greater reduction in the size of the inductor. In fact, we will see that beyond $r \sim 0.4$, we enter a *region of diminishing returns*.

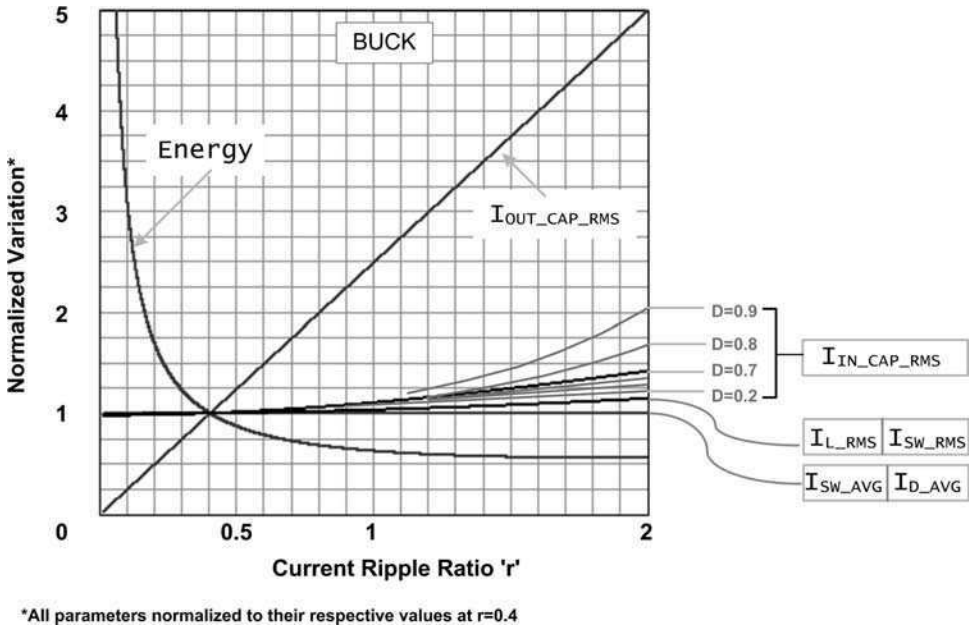


Figure 2-6: How Varying the Current Ripple Ratio *r* Affects All the Components

In *Figure 2-6*, we have also plotted the *capacitor* RMS currents for a buck converter. We see that if *r* is increased beyond 0.4, the currents will increase significantly. This will lead to increased heat generation inside the capacitors (and other related components too). Eventually, we may be forced to pick a capacitor with a lower ESR and/or lower case-to-air thermal resistance (more expensive/bigger).

Note: The RMS value of the current through any component is the current component responsible for the *heat* developed in it — via the equation $P = I_{RMS}^2 \times R$, where P is the dissipation, and R is the series resistance term associated with the particular component (e.g. the DCR of an inductor, or the ESR of a capacitor). However, it can be shown that *the switch, diode, and inductor RMS current values are not very "shape-dependent."* Therefore, the heat developed in them does *not* depend much on *r*, but mainly on the average value of the current. On the other hand, the RMS of the capacitor current waveforms can increase significantly, if *r* is increased. So *capacitor currents are very "shape-dependent,"* and therefore depend strongly on *r*. The reason for that is fairly obvious — any capacitor in a steady state has *zero average (dc) current* through it. So since a capacitor effectively subtracts out the dc level of the accompanying current waveform, we are left with a capacitor current waveform that has a large "ramp portion" built-in into it. Therefore, changing *r* changes this ramp portion, thereby impacting the capacitor current greatly.

Note that in *Figure 2-6*, though we have used the buck topology as an example, the *energy curve in particular is exactly the same for any topology*. The capacitor current curves though, may not be identical to those of the buck, but are *similar*, and so the conclusions above still apply.

Therefore, in general, ***a current ripple ratio of around 0.4 is a good design target for any topology, any application, and any switching frequency***.

Later, we will discuss some reasons/considerations for *not* adhering to this r ~ 0.4 rule-of-thumb (under certain conditions).

Do We Mean Inductor? Or Inductance?

Note that in the previous section, we said nothing explicitly about what the *inductance* was — we just talked about the *size* of the *inductor*. We know that in theory, we can put almost any number of turns on a given core, and get almost any inductance. So *inductance* and *size* of inductor are not *necessarily* related. However, we will now see that in power conversion they often do turn out to be so, though rather indirectly.

Looking at *Figure 2-6*, we can see that a smaller *r* will require a higher energy-handling capability, and thus a *larger* inductor. Let us now formally go through *all* the possible ways of reducing *r*.

Since *we are assuming our application conditions are fixed*, the load current and input/output voltages are also fixed. Therefore, I_{DC} is fixed too. The only way we can cause *r* to decrease under these circumstances is to make ΔI smaller. However, ΔI is

$$\Delta I = \frac{\text{voltseconds}}{\text{inductance}} \quad (V\text{-}s/H)$$

But we know the applied voltseconds is fixed too (input and output voltages being fixed). So the *only way to decrease r (for a given set of application conditions) is to increase the inductance*. We can therefore conclude that if we choose a *high inductance*, we *will invariably require a bigger inductor*. It is therefore no surprise that when power supply designers instinctively ask for a "large inductance," they might well mean a "large inductor." Therefore ***the designer is cautioned against being too "ripple-phobic" in their designs. A certain amount of ripple is certainly "healthy."***

However, we must not forget that if, for example, we *increase the load current* (i.e. a change in application conditions), we will clearly need to move to a *larger inductor* (with greater energy-handling capability). *But simultaneously, we will need to* ***decrease*** *the inductance*. That's because I_{DC} will increase, and so to keep to the "optimum" value of *r*, we will need to *increase* ΔI in the same proportion as the increase in I_{DC}. And to do this, we have to decrease, rather than increase, L.

Therefore the general statement that a "large inductance is equivalent to a large inductor" only applies to a *given* application.

How Inductance and Inductor Size Depend on Frequency

The following discussion applies to *all* the topologies.

If keeping everything else fixed (including D) we *double the frequency*, the voltseconds will halve, because the durations t_{ON} and t_{OFF} have halved. But since ΔI is "voltseconds per unit inductance," this too will halve. Further, since I_{DC} has not changed, $r = \Delta I/I_{DC}$ will also halve. So if we started off with $r = 0.4$, we now have $r = 0.2$.

If we want to return the converter to the optimum value of $r = 0.4$, we will now need to somehow double the ΔI we were left with at the end of the last step. The way to do that is to *halve the inductance*.

- Therefore, we can generally state that *inductance is inversely proportional to frequency*.

Finally, having restored *r* to 0.4, the peak will still be 20% higher than the dc level. But the dc level has not changed. So the peak value is also unchanged (since *r* hasn't changed either, eventually). However, the energy-handling requirement (size of inductor) is $\frac{1}{2} \times L \times {I_{PK}}^2$. So since L has halved, and I_{PK} is unchanged, the required *size of the inductor has halved*.

- Therefore, we can generally state that *the size of the inductor is inversely proportional to frequency*.
- Note also that *the required current rating of the inductor is independent of the frequency (since peak is unchanged)*.

How Inductance and Inductor Size Depend on Load Current

For all topologies, if we *double the load current* (keeping input/output voltages and D fixed), *r* will tend to halve since ΔI has not changed but I_{DC} has doubled. Therefore to restore *r* to its optimum value of 0.4, we need to get ΔI to double too. But we know that ΔI is simply "voltseconds per unit inductance," and in this case the voltseconds has not changed. So the only way to get ΔI to double is to *halve the inductance*.

- Therefore, we can generally state that *inductance is inversely proportional to the load current*.

What about the size? Since we doubled the load current, but still kept *r* at 0.4, the peak current $I_{DC}(1 + r/2)$ has also doubled. But the inductance has halved. So the energy-handling requirement (size of inductor), $\frac{1}{2} \times L \times {I_{PK}}^2$, will double.

- Therefore, we can generally state that *the size of the inductor is proportional to the load current*.

How Vendors Specify the Current Rating of an Off-the-shelf Inductor and How to Select It

The "energy-handling capability" of an inductor, $1/2 \times LI^2$, is one way of picking the size of the inductor. But most vendors do not provide this number upfront. However, they do provide one or more "current ratings." And if we interpret these current rating(s) correctly, that serves the purpose too.

The current rating may be expressed by the vendor either as a maximum rated I_{DC}, or a maximum rated I_{RMS}, or/and a maximum I_{SAT}. The first two are usually considered synonymous, since the RMS and dc values of a typical inductor current waveform are almost equal (we had indicated previously that the RMS of the inductor current is not very "shape-dependent"). So the dc/RMS rating of an inductor is by defintion basically the direct current we can pass through it, such that we get a specified temperature rise (typically 40 to 55°C depending on the vendor). The last rating, that is, the I_{SAT}, is the maximum current we can pass, just before the *core starts saturating*. At that point, the inductor is considered close to the useful limit of its energy-storing capability.

We will also find that many, if not most, vendors have chosen the wire gauge in such a manner that the I_{DC} and I_{SAT} ratings of any inductor are also virtually the same. And by doing this, they can publish one (*single*) current rating — for example, "the inductor is rated for 5 A." Basically, having determined the I_{SAT} of the inductor, the vendor has then consciously tweaked the wire gauge (at the saturation current level), so as to also get the specified temperature rise too.

The rationale for wanting to set $I_{DC} = I_{SAT}$ is as follows — suppose the inductor had a dc rating of 3 A and an I_{SAT} of 5 A. The 5-A rating is then likely to be *superfluous*, because users would probably never select this inductor for an application that required more than 3 A anyway. Therefore, the excessive I_{SAT} rating in this case essentially amounts to an *unnecessarily over-sized core.* Of course, if we do find an inductor with different I_{DC} and I_{SAT} ratings, it is also possible the vendor may have (unsuccessfully) tried to exploit the larger size of the chosen core (by increasing the wire thickness), but the stumbling block was that the selected core geometry was somehow not conducive to doing so — maybe it just did not have enough *window space* for accommodating the thicker windings.

In general, an inductor with a "single" current rating is usually the most optimum/ cost-effective too.

However, in some rare off-the-shelf inductors, we may even find I_{SAT} stated to be less than I_{DC}. But what use is that? We can't operate beyond I_{SAT} in any case! So the only advantage, if any, that can be gleaned from such an inductor is that the temperature rise in a real application will be less than the maximum specified. Automotive applications?

*In general, for most practical purposes, the current rating of the inductor that we need to consider is the **lowest** rating of all the published current ratings. We can usually simply ignore all the rest.*

There are some subtle considerations and exceptions to the argument for always preferring an inductor with $I_{DC} \approx I_{SAT}$. For example, under transient/temporary conditions, the *momentary* current may exceed the *normal steady operating current by a wide margin.* So for example, suppose we are we using a switcher with an internally *fixed* current limit 'I_{CLIM}' of 5 A — in *a 3-A application*. Then under startup (or sudden line/load steps), the current is very likely to hit the limiting value of 5 A for several cycles in succession as the control circuitry struggles to bring up the output rail into regulation. We will discuss this issue in greater detail below — in particular, *whether this is even a concern to start with!* However, assuming for now that it is, it then seems that it may actually make sense to use an inductor rated for 3-A continuous current, with an I_{SAT} rating of 5 A (provided such an inductor is freely available, and cheap). Of course, alternatively, we could just pick a standard "5 A inductor" (for the 3-A application), and thereby we would certainly avoid inductor saturation under all conditions (and the consequent likelihood of switch destruction). But we realize that in doing so, our inductor may be considered slightly *over-designed* from the viewpoint of its copper/temperature-rise — the wire being unnecessarily thicker. However, we should keep in mind that larger cores certainly affect cost, but a little more copper rarely does!

What Is the Inductor Current Rating We Need to Consider for a Given Application?

Whenever we start-up, or subject the converter to sudden line/load transients, the current no longer stays at the steady value it has under *normal* operation (i.e. when delivering the required maximum rated load current). For example, if we suddenly short the output the control circuitry in an effort to regulate the output may momentarily expand the duty cycle to the *highest permissible value* (as set by the controller). We then are no longer in steady state, and so under the increased on-time voltseconds, the current ramps up progressively, and can reach the set current limit.

But then, the inductor would probably be saturating! For example, if we are using a 5-A fixed current limit *buck* switcher IC for a 3-A application, we have probably picked an inductor rated for only around 3 A. But when we short the output, the current momentarily hits the current limit (which may be around 5.3 A for a "5-A buck switcher").

So the question is — *should we select an inductor with a rating based on the current limit threshold (that it may encounter under severe transients), or simply on the basis of the maximum continuous normal operating current (under steady state operation in our*

application)? In fact, this question is not as philosophical as it may seem — it virtually separates standard industry off-line design procedures from those of dc-dc converters. To answer it effectively, a lot of factors may need to be considered, often on an individual or case-by-case basis. Let us address some of these concerns next.

Luckily, in most *low-voltage applications*, a certain amount of core saturation doesn't cause any problem. The reason for that is that if in the above example, the switch is rated for 5 A, and the current limiting circuit in the IC is known to act ***fast enough*** to prevent the current from ever rising beyond 5 A, then even if the inductor has started saturating as it gets to 5 A, there is no cause for concern — *after all, if the switch doesn't break, we don't have a problem! And since the current doesn't exceed 5 A, the switch cannot break.* So in this case, we could certainly pick a cost-effective "3 A inductor" for our application, knowing well in advance that it would saturate somewhat under various non-steady conditions. Of course we don't want to operate a switching converter constantly (under its rated maximum load conditions) with a saturating inductor — we just "allow" it to do so under abnormal and temporary conditions, so long as we are sure that the switch can never be damaged.

However, the above logic begs another key question to be answered — *what exactly constitutes "fast enough"* — that is, which factors affect our ability to turn the switch OFF fast enough to protect it from the consequences of a saturating inductor? Since this consideration may eventually end up dictating the size and cost of the inductor, it is important to understand this *response-time issue* well.

a) All current limit circuitry takes some finite time to respond. There are inherent (internal) "propagation delays" as we move the overcurrent signal through the internal comparators of the IC, its op-amps, level-shifters, driver, and so on to the IC pin driving the switch.

b) If we are using a controller IC (as opposed to an 'integrated switcher,' i.e. with an internal switch), the switch will necessarily be at a certain physical distance from its driver (which is usually inside the IC). In that case, the parasitic inductances of the intervening PCB traces (roughly ***20 nH per inch of trace***) will resist any sudden change in current, thereby creating an additional delay before the turn-off command issued by the IC actually reaches the gate/base of the switch.

c) Theoretically speaking, even if the current limiting circuitry had responded *immediately* to the overcurrent condition, *and* if the intervening traces had truly negligible inductance, the switch may still take a little time before it really turns itself OFF. During this delay, if the inductor is saturating, it will not be able to effectively prevent or limit the current spike that can be pushed through the transistor by the applied input dc source — well beyond the "safe" current limit threshold.

Bipolar junction transistors (bjts) are inherently slow, as compared to more modern devices like mosfets. But large mosfets (e.g. high-current, high-voltage devices) also produce delays because of their higher internal parasitic gate resistance and inductance and significant *inter-electrode parasitic capacitances* (that demand to be either discharged or charged as the case may be, before they allow the switch to change its state). Matters can get worse if we parallel several such mosfets together, as say for a very high-current application.

d) Many controllers and ICs incorporate an internal "blanking time" — during which they deliberately "do not look" at the current waveform. The basic purpose is to avoid false triggering of the current limit circuitry by the noise generated at the turn-on transition. But this delay time could prove fatal to the switch, especially if the inductor has already started saturating, because the current limit circuitry won't even "know" if there is any overcurrent condition during this blanking interval. Further, in current-mode control ICs, the ramp to the PWM (pulse-width modulator) comparator stage is usually derived from the (noisy) switch current. So the blanking time is typically set even higher — typically about 100 ns for low-voltage applications and up to 300 ns for off-line applications.

e) Integrated high-frequency switchers (i.e. with the mosfet or bjt switch contained in the *same package* as the control and driver) are usually the best-protected and most reliable, because the intervening inductances are minimized. Also, the blanking times can be set more accurately and optimally, since there is not going to be much variation in terms of different switches with widely varying characteristics. Therefore, integrated switchers can usually survive momentarily saturating inductors with almost no problem —*unless the input voltage is very high* (typically above 40–60 V), and *plus*, the inductor is sized very small.

f) If the input voltage is high, the *rate of rise of the saturating inductor current ramp* can become very large ("steep"). This follows from the basic equation $V = LdI/dt$. Here, if $L \rightarrow 0$, since V is fixed, the dI/dt must increase dramatically (see *Figure 2-7*). So now, even a small delay can prove fatal because a *large* ΔI can take place during a very small interval. The current can therefore overshoot the set current limit threshold by a very large amount, thereby endangering the switch. That is why, especially when we come to off-line applications, it is actually *customary* to select a core *large enough to avoid saturation at the current limit threshold*. And that usually gives enough time for the current limit circuitry to act — *before* the slope of the current has gone completely out of control.

Note however, that the copper windings still only need to be proportioned to handle the continuous current (i.e. based on the maximum operating load).

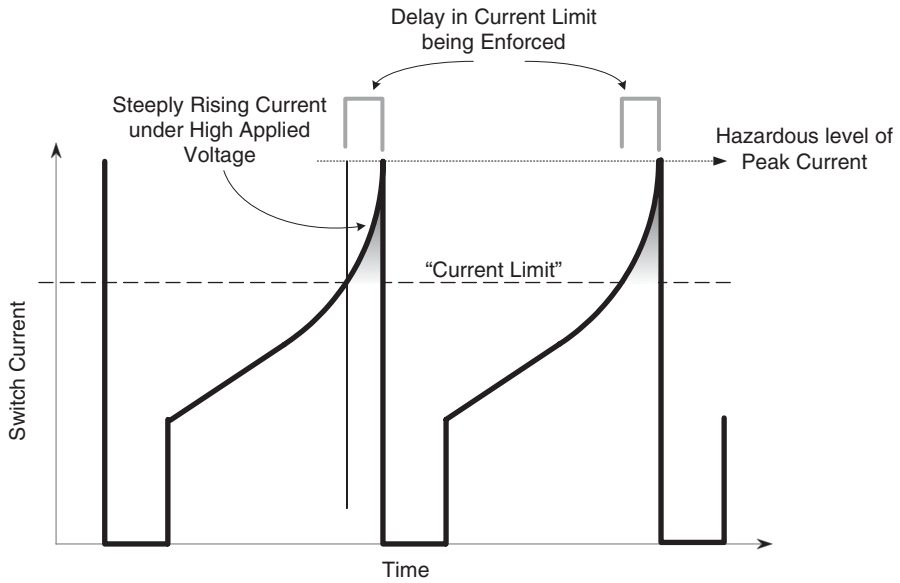


Figure 2-7: How Higher Voltages Combined with Inherent Response-time Delays Can Cause Overstress in the Switch When the Inductor Starts Saturating

In effect, what we are therefore always implicitly doing in off-line applications is *setting the I_{SAT} of the transformer higher than its I_{DC} rating.* That is clearly not what we usually do in low-voltage dc-dc converter design.

g) Generally speaking, in most *low-voltage applications* (i.e. V_{IN} typically less than about 40 V), the inductors are selected based only *on the maximum operating load current*. The current limit is therefore, in effect, virtually ignored! This is the usual industry practice for dc-dc converter design, though it is probably not clearly spelled out in this way most of the time. But luckily, it seems to have worked!

The Spread and Tolerance of the Current Limit

Any specification, including the current limit, either set by the user or fixed internally in the IC, will have a certain inherent *tolerance band* — that includes spreads over process variations and over temperature. All these variations are combined together inside the electrical tables of the datasheet of the device, under its "MIN" and "MAX" limits. In a practical converter design, a good designer learns to pay heed to such spreads.

But let us first summarize the general procedure for selecting the *inductance* for a switching power converter. Then we will look at the practical issues concerning spreads/tolerance.

The normal procedure is to determine the inductance by requiring that the current ripple ratio is about 0.4 — because we know that that represents an optimum of sorts for the entire converter. But there may be another possible limitation when dealing with switcher ICs, especially those with internally set (fixed) current limits — ***if our normal operating peak currents are close to the set current limit of the device*** (i.e. we are operating close to the maximum current capability of the switcher IC), ***we need to ensure that the inductance is large enough not to cause the calculated operating peak current (within any given cycle) to exceed the current limit*** — otherwise foldback will obviously occur at the current limit threshold, and so the desired maximum output power cannot be guaranteed.

For example, if we have a "5-A buck switcher IC," being operated at 5-A load, with an r of 0.4, then the normal operating peak current is $5 \times (1 + 0.4/2) = 5 \times 1.2 = 6$ A. So *ideally*, we would want the current limit of the device to be at least 6 A. ***Unfortunately, when we come to such integrated switchers, that much of "margin" is rarely available*** — manufacturers always like to "bolster" the advertised ratings of their parts to be close to the maximum stress limits. So yes, if this particular part was declared to be a "4-A IC" instead of a "5-A IC," we would have been just fine. But as things stand, manufacturers usually pay scant regard as to what may constitute an *optimum* rating for the device, in relationship to its associated components and the overall design strategy. Therefore, for example, a certain commercial "5-A switcher IC" may have a published (set) current limit of only 5.3 A. But on analysis, we see that that that allows only 0.3 A above, and 0.3 A below, the average level of 5 A. Therefore, the maximum allowed ΔI at 5 A load is only 0.6 A. And the maximum r is $0.6/5 = 0.12$ (when operated at a load current of 5 A). We can see that that is clearly much less than the optimum r of 0.4. And no doubt, this lowered r will adversely impact the size of the inductor (and converter).

Now we take up the issue of the *spread* in current limit. So I_{CLIM} is actually two limits — I_{CLIM_MIN} and I_{CLIM_MAX} (i.e. the MIN and MAX of the current limit respectively). The question is — *which of these limits should we consider for designing the inductor*?

- To *guarantee output power*, we need to look at the *MIN of the current limit only*. In most low-voltage dc-dc converter applications, the MIN limit is the only threshold that really counts — we can usually completely ignore the MAX (and of course the TYP value). The basic criterion for guaranteeing output power is — we must ensure that the calculated normal operating peak current in our application is always *less than the MIN* value of the current limit. Of course, if we are not operating close to the current limit of the device, this condition will be met without any struggle, and so we can then just focus on setting r to about 0.4.

- But like all components, inductors also have a typical tolerance — usually about ±10%. So if we are operating very close to the limits of the device, and thereby *r* is being effectively dictated by the MIN of the current limit (rather than by its optimum or desirable value), then the (nominal) value of inductance we ultimately choose should be about *10% higher than the calculated value*. That will guarantee output power unconditionally — under all possible variations in current limit *and inductance*.
- Note that ideally, we would also like to leave at least 20% additional margin *(headroom)* between the peak current of our application and the MIN of the current limit. This is usually necessary for getting a quick *response (correction) to a sudden increase in load.* So in general, if we somehow manage to curtail the ability of the converter to respond quickly (for example, by not providing sufficient headroom in the current limit and/or maximum duty cycle), the inductor will not be able to ramp up current quickly enough to meet the sudden increase in energy demand. Therefore, the output will droop rather severely for several cycles, before it eventually recovers.

But unfortunately, once again, when dealing with fixed current limit (integrated) switchers, we will find that this "nice-to-have transient headroom" may be a *luxury we just can't afford* — because in most cases, the MIN current limit is set only slightly higher than the declared "rating" of the device. So in fact even a 20% headroom may not be available! And further, even assuming it is, this may demand a very *large (impractical) inductance*. And we know that that by itself is fairly *counterproductive* — a large inductance takes even more time for its current to ramp up, and that thereby effectively slows down the transient (loop) response — incidentally, just opposite to what we were hoping to derive here! Therefore, in general, we almost always end up ignoring this 20% or so *step-response headroom/margin* completely, especially when dealing with integrated switcher ICs.

As for the MAX of the current limit, *whenever we deem that inductor saturation is of real concern to us* (as in high-voltage applications), *we must look at the MAX of the current limit* to decide upon the size of the inductor — that being the worst-case in terms of peak current under overloads, inductor energy storage, and its possible saturation.

Therefore in general, *in high-voltage dc-dc (or off-line) applications, the MIN of the current limit may sometimes need to be considered when selecting inductance (as when operating close to current limit), but the MAX of the current limit will certainly always be used to determine the size of the inductor.*

As a corollary, manufacturers of (low-voltage) *dc-dc converter ICs* actually need *not* (and probably justifiably do not) struggle too hard to minimize the spreads and tolerances of current limit (provided of course the MIN of the current limit is at least set high enough not

to intrude on the declared power-handling capability of the IC). And for *low-voltage dc-dc converter applications*, the current limit is typically ignored altogether — the final selection of inductor current rating (and size) is simply based on the cycle-by-cycle peak inductor current under *normal* (steady) operation (i.e. the maximum load of the application, at the worst-case input voltage end).

On the other hand, manufacturers of *off-line* switcher ICs *do* need to maintain a *tight tolerance* on the current limit. In their case, the maximum power-handling capability of their particular device is in effect dependent only on the 'MIN' (minimum limit) of the current limit specification, whereas, the transformer size is determined entirely by the 'MAX' of the current limit specification. So in this case, a "loose" current limit specification effectively amounts to requiring *bigger components* (transformer) for the same maximum power-handling capability.

> **Note:** Some makers of off-line integrated switcher ICs (e.g. the "Topswitch" from Power Integrations) often tout their "precise" current limit — thus suggesting that we get the best power-to-size ratio (i.e. converter power density) when using their products. However, we should remember that in most cases, their product families have a *discrete* set of fixed current limits. And that is a problem! For example, we may have devices available with current limits in steps of 2 A, 3 A, 4 A, and so on. So yes, we may indeed get a higher power density *when operating at the maximum rated output power* of a particular IC. But when operating at a power level *between* available current limits, we are not going to get an optimum solution. For example, in an application where the peak current is 2.2 A, then we would need to select the 3 A current limit part, and we will need to design our magnetics to avoid core saturation at 3 A. So in effect, we have a very imprecise current limit now! The best solution is to look for a part (integrated switcher or controller plus mosfet solution) where we can precisely set the current limit *externally*, depending upon our application.

With all these subtle considerations in mind, a designer can hopefully pick a more appropriate inductor current rating for his or her application. Clearly, there are no hard and fast rules. Engineering judgment needs to be applied as usual, and perhaps some further bench-testing may also be needed to validate the final choice of inductor.

In the worked examples that follow, the general approach and design procedure will become clearer.

Worked Example (1)

A boost converter has an input range of 12 V to 15 V, a regulated output of 24 V, and a maximum load current of 2 A. What would be a reasonable goal for its inductance, if the switching frequency is a) 100 kHz, b) 200 kHz, and c) 1 MHz? What is the peak current in each case? And what is the energy-handling requirement?

The first thing we have to remember is that for this topology (as for the buck-boost), the *worst-case* is the *lowest end* of the input range, since that corresponds to the highest duty cycle and thus the highest average current $I_L = I_O/(1 - D)$. So for all practical purposes,

we can completely disregard V_{INMAX} here — in fact it was a red herring to start with, for this particular analysis!

From *Table 2-1*, the duty cycle is

$$D = \frac{V_O - V_{IN}}{V_O} = \frac{24 - 12}{24} = 0.5$$

Therefore

$$I_L = \frac{I_O}{1 - D} = \frac{2}{1 - 0.5} = 4 \text{ Amperes}$$

Let us target a current ripple ratio of 0.4. So

$$I_{PK} = I_L \left(1 + \frac{r}{2}\right) = 4 \times \left(1 + \frac{0.4}{2}\right) = 4.8 \text{ Amperes}$$

- We should remember that *r = 0.4 always implies that the peak is 20% higher than the average*. So we realize that in effect, the peak current does *not* depend on the frequency. The inductor must be able to handle this above peak current without saturating. So in this example, we would be fine just picking an inductor rated for 4.8 A (or more), ***irrespective of frequency***. In fact we had previously learned in the section "How Inductance and Inductor Size Depend on Frequency" that *the required current rating of an inductor is independent of the frequency (since the peak is unchanged)*. However, the size does change with frequency, because size is $1/2 \times L \times I_{PK}{}^2$, and L changes as follows.

To calculate the inductance corresponding to the chosen value of *r*, we can use the following equation (presented previously). We also note from *Table 2-1* that $V_{ON} = V_{IN}$ for the boost. Therefore for f = 100 kHz

$$L = \frac{V_{ON} \times D}{r \times I_L \times f} = \frac{12 \times 0.5}{0.4 \times 4 \times 100 \times 10^3} \Rightarrow 37.5 \ \mu H$$

For f = 200 kHz, we would get *half* of this, that is, 18.75 μH. And for f = 1 MHz, we get 3.75 μH. We clearly see that *high frequencies lead to smaller inductances*.

We have previously observed that *for a given application*, small inductances invariably lead to small inductors. Therefore we conclude that *on increasing the switching frequency, we will get smaller-sized inductors too.* And that is the basic reason for hiking up switching frequencies in general.

The energy-handling requirement, if desired, can be explicitly calculated in each case, by using $E = \frac{1}{2} \times L \times {I_{PK}}^2$.

So far, we have been generally targeting r = 0.4 as an optimum value. Let us now understand all the reasons why this may not be a good choice on occasion.

Current Limit Considerations in Setting r

We had indicated previously that the current limit may be too low to allow *r* from being set to its optimum. Now we will also include the impact of the *spread* in the current limit.

So for example, in *Table 2-3* we have the published specifications for the current limit of an integrated "5-A" switcher, the LM2679. To be able to guarantee the specified power output (or load current in this case) unconditionally, we need to guarantee that the peak current in our application never reaches even the *lower limit* ("MIN") of the published current limit specification. So in fact, in *Table 2-3*, we need to disregard all the numbers except for the 'MIN' value — given as *5.3 A*.

Now, if we are trying to get 5 A out of our converter with an *r* of 0.4, the estimated peak current will be $1.2 \times 5 = 6$ A. Clearly, as mentioned earlier, we are *not going to get there* with the LM2679! Unless we *lower* the value of *r (increase inductance).* Maximum value of *r* is

$$I_{PK} = I_O \times \left(1 + \frac{r}{2}\right) \leq I_{CLIM\_MIN}$$

Solving, with $I_O = 5$ A, and $I_{CLIM\_MIN} = 5.3$ A, we get

$$r \leq 2\left(\frac{I_{CLIM\_MIN}}{I_O} - 1\right) = 2\left(\frac{5.3}{5} - 1\right) = 0.12$$

We can see from Figure 2-6, that this calls for an energy-handling capability (size of inductor) almost 3× the optimum!

Actually, it turns out this part is just *specified* inappropriately. This part is in reality one with an *adjustable* current limit. And so we could have probably adjusted the current limit adjust

Table 2-3: Published current limit specs for the LM2679

	Conditions		TYP	MIN	MAX	Units
Current Limit 'I_{CLIM}'	$R_{CLIM} = 5.6$ kΩ	Room Temperature	6.3	5.5	7.6	A
		Full Operating Temperature Range		5.3	8.1	

resistor quoted in the electrical tables to allow for a "better" value of current limit, and thereby a better value of *r* (at maximum rated load). But unfortunately, that is not clarified in the tables.

We should always remember that the minimum and maximum limits of the electrical tables are the only parts of a datasheet really guaranteed by any vendor (certainly not the *typical* values!). So as a matter of fact, *any* other information in a datasheet just amounts to general design "guidance" — and that includes any 'typical performance curves' provided. A prudent designer would never second-guess the vendor — in this case as to whether the current limit resistor can indeed be adjusted to give us a smaller inductor, or not. Therefore, as it stands, if we are using the LM2679 for a 5-A load current application, we *do need an inductor three times larger than the optimum*. Note that if the current limit can indeed be adjusted higher, the vendor should have picked the appropriate value for the current limit adjust resistor in the "conditions" column of the electrical table (and stated the limits accordingly).

Note also that when we talk of a "5 A buck IC," that implies the part is supposed to deliver 5-A *load current*. The current limit of course needs to be set (and stated) correctly for the rated load, as discussed above. However, we should be very clear that **when we are talking of boost or buck-boost switcher ICs, a "5 A" part for example, does *not* give us a 5-A *load current***. That is because the dc inductor current is not equal to I_O, but $I_O/(1 - D)$ for these topologies. So a "5-A" rating in this case only refers to the *current limit* of the device. What load current we can derive from a "non-buck" IC depends on our specific application — in particular on the D_{MAX} (duty cycle at V_{INMIN}). For example, if the desired load current is 5 A, and the (maximum) duty cycle in our application is 0.5, then the average inductor current is actually $I_O/(1 - D) = 10$ A. Further, with an *r* of 0.4, the peak would be 20% higher, that is, $1.2 \times 10 = 12$ A. So, for an optimum case, we would need to actually look for a device whose *minimum* current limit is 12 A or more in this case. At the bare minimum, we need a device with a current limit higher than 10 A, just to guarantee output power.

Continuous Conduction Mode Considerations in Fixing r

As discussed previously, under various conditions, we may enter discontinuous conduction mode (DCM). From *Figure 2-5* we can see that just as DCM starts to occur, the current ripple ratio is 2. However we can pose the question in the following manner — what if we have set the current ripple ratio to a certain value ***r′*** (i.e. the current ripple ratio *at the maximum load current*, I_{O_MAX}). And then we decrease the load current slowly — at what load does the converter enter DCM?

By simple geometry it can be shown that the transition to DCM will occur at r′/2 times the maximum load. For example suppose we set r' to 0.4 at 3 A load, the converter will transition into DCM at $(0.4/2) \times 3 = 0.6$ A.

But designers know that when DCM is entered, a lot of things within the converter change suddenly! The duty cycle, for one, will now starting pinching off toward zero as we decrease the load current further. In addition, the loop response of the converter (its ability to correct quickly for disturbances in line and load) also usually gets degraded in DCM. The noise and EMI profile can change suddenly too, and so on. Of course there are some advantages of operating in DCM too, but let us for now assume that for various reasons, the designer wishes to avoid DCM altogether, if possible.

Maintaining the converter in CCM, down to the minimum load of our application, enforces a certain *maximum* value for r'. For example, if the minimum load is $I_{O\_MIN} = 0.5$ A, then to maintain the converter in CCM at 0.5 A, the set current ripple ratio (r' at 3 A) needs to be lowered. Back calculating, we get the required condition for this

$$I_O \times \frac{r'}{2} = I_{O\_MIN}$$

So

$$r' = \frac{2 \times I_{O\_MIN}}{I_{O\_MAX}}$$

In our case we get

$$r' = \frac{2 \times 0.5}{3} = 0.333$$

We therefore need to set the current ripple ratio to less than 0.333 at maximum load, to ensure CCM at I_{O_MIN}.

Note that generally speaking, we can make the converter operate in boundary conduction mode (BCM), or in full DCM, in three ways — a) by decreasing the load, b) choosing a small inductance, or c) ***increasing the input voltage***.

We realize that decreasing the load will proportionally decrease I_{DC} to virtually any value, and so the condition $r \geq 2$ (BCM to DCM) will certainly occur sooner or later — below a certain load current. Similarly, decreasing L will necessarily increase ΔI, and so at some point we can expect the ratio $\Delta I/I_{DC}$ (i.e. r) to try and become greater than 2 (implying DCM).

However, as far as the *third* method of entering DCM mentioned above, we should realize that **solely** *increasing the input voltage just might not do the trick! DCM or BCM can only happen under an input (line) variation, provided the load current is simultaneously below a certain value to start with* (the value being dependent on L).

It is instructive to study the three topologies separately in this regard. Note that the general equation for *r* is

$$r = \frac{V_{ON} \times D}{I_L \times L \times f} \quad \textit{(any topology, any mode)}$$

Applying the voltseconds law in CCM (or BCM), we also get

$$r = \frac{V_{OFF} \times (1 - D)}{I_L \times L \times f} \quad \textit{(any topology, CCM or BCM only)}$$

a) From the plots of *r* in *Figure 2-4*, we see that both the **buck and the buck-boost** have the highest value of *r* when D approaches zero, i.e. at maximum input voltage. For these topologies, the equation for *r* (derivable from the more general equation for *r* just given immediately above) is

$$r = \frac{V_O}{I_O \times L \times f}(1 - D) \quad \textit{(Buck)}$$

$$r = \frac{V_O}{I_O \times L \times f}(1 - D)^2 \quad \textit{(Buck-Boost)}$$

So putting $r = 2$ and $D = 0$ (i.e. highest input voltage plus BCM), we get the limiting condition

$$I_O = \frac{1}{2} \times \frac{V_O}{L \times f} \quad \textit{(Buck and Buck-Boost)}$$

Therefore, for these two topologies, if I_O is *greater* than the above limiting value, we will *always remain in CCM, no matter how high we increase the input voltage*.

b) Coming to the **boost**, the situation is not so obvious. From *Figure 2-4*, we see that *r* peaks at $D = 0.33$ (corresponding to the input being exactly two-thirds of the output). *So the boost is most likely to enter DCM at* $D = 0.33$ — not say, at $D = 0$ or $D = 1$. We can derive the following (exact) equation for *r*

$$r = \frac{V_O}{I_O \times L \times f} D \times (1 - D)^2 \quad \textit{(Boost)}$$

So putting D = 0.33, and r = 2 in this equation, we get the following limiting condition

$$I_O = \frac{2}{27} \times \frac{V_O}{Lf} \quad \textit{(Boost)}$$

Therefore, for the boost topology, if I_O is *greater* than this value, we will *always remain in CCM*, no matter how high we increase the input voltage.

Note that, if we do manage to enter DCM, the most likely input point for this to happen is an *input of 0.67 times the output*. In other words, if we are not in DCM at this particular input voltage, we can be sure we will be in CCM throughout the entire input range (whatever it may be).

Setting r to Values Higher Than 0.4 When Using Low-ESR Capacitors

Nowadays, with improvements in capacitor technology, we are seeing a new generation of very 'low-ESR' capacitors — like monolithic multilayer ceramic capacitors ('MLCs' or 'MLCCs'), polymer capacitors, and so on. Due to their extremely low ESRs, these capacitors usually have very high ripple (RMS) current ratings. Therefore, the required size of such capacitors in any application *is no longer dictated by their ripple current handling capability*. In addition, these capacitors also have almost no *ageing* characteristics (or lifetime issues) that we need to account for beforehand in the design (as we customarily do for electrolytic capacitors — that can "dry out" over time). Further, due to their very high dielectric constant, these new capacitors have also become very *small* in size. So in fact nowadays, *increasing r may not necessarily cause a noticeable increase in the space occupied by the capacitors (or size of converter)*. On the other hand, increasing *r* may still lead to a relatively significant reduction in size of the inductor.

Summing up, with modern capacitors to the rescue, it may start making perfect sense to increase *r* from its traditional "optimum" of 0.4, to say around 0.6 to 1 on occasion (provided other considerations do not restrict this). If we do so, *Figure 2-6* tells us, we can still get an additional 30 to 50% reduction in the size of the inductor. And that is certainly not insignificant, provided of course that that advantage is not offset by having to use larger capacitors in the bargain!

Setting r to Avoid Device "Eccentricities"

Surprisingly, device *eccentricities* may on occasion play a part in defining the limits of *r* too. For example in *Figure 2-8* we have presented the current limit plot of an integrated high-voltage flyback switcher IC called the "Topswitch®." On it we have superimposed a typical switch current waveform, just to make things a little clearer.

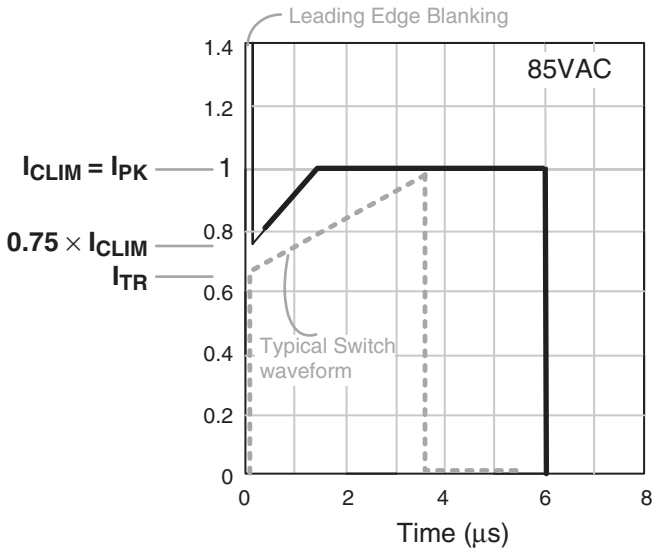


Figure 2-8: The 'Initial Current Limit' of the Topswitch®

We see that surprisingly, the current limit of this device is *time-dependent* for about 1.5 µs after the turn-on transition — something we don't intuitively ever expect. This 'initial current limit' of the device occurs just as its internal current limit comparator starts to come out of its (valid) 'leading edge blanking' time. As mentioned, during this blanking time the IC is just "not looking" at the current at all to avoid spurious triggering on the noise edge of the turn-on transition. But the problem is that once the current limit circuit gets down to monitoring the switch current again, it takes a *certain time* for the current limit threshold to settle down — and during this time it can be triggered at only about 75% of the supposed current limit!

Looking at the switch (or inductor) current waveform, we know that the current at the moment the switch turns ON is always *less* than the average value by the amount $\Delta I/2$. In other words this *trough* (*valley*) current 'I_{TR}' is related to *r* according to the equation

$$I_{TR} = I_L \times \left(1 - \frac{r}{2}\right)$$

We realize that to avoid hitting the initial current limit of the device, we need to ensure that the trough falls below $0.75 \times I_{CLIM}$. So

$$I_{TR} = I_L \times \left(1 - \frac{r}{2}\right) \leq 0.75 \times I_{CLIM}$$

Now, we are assuming the power supply is at maximum load in this analysis. Therefore, the peak current is set equal to the current limit I_{CLIM}

$$I_{PK} = I_L \times \left(1 + \frac{r}{2}\right) = I_{CLIM}$$

Therefore equating the two equations above, we get the limiting condition for *r*

$$\left(1 - \frac{r}{2}\right) \leq 0.75 \times \left(1 + \frac{r}{2}\right)$$

or

$$r \geq 0.286$$

Since *r* in any case is typically set to about 0.4, we should normally have no trouble with this "initial current limit" issue. However, note that on finer examination of the electrical tables of the datasheet, this 0.75 × factor is specified *only at 25°C*. Unfortunately, very few power devices stay at 25°C for long! So, the bottom line is that, we, as designers, *do not really know* the value of the current limit as the device *heats up*. Yes, we can certainly make an educated guess, possibly leave an additional safety margin when fixing *r*, and certainly, we may face no problem whatsoever. But the truth is we are *on our own now* — the vendor has *not* provided the requisite data (in the form of *guaranteed* limits within the electrical tables).

Setting r to Avoid Subharmonic Oscillations

Looking at *Figure 2-9*, we see that in any converter, the output voltage is first compared against an internal reference voltage. Then, the difference between the two (the 'error') is filtered, amplified, and inverted by an 'error amplifier,' the output of which (the '*control*

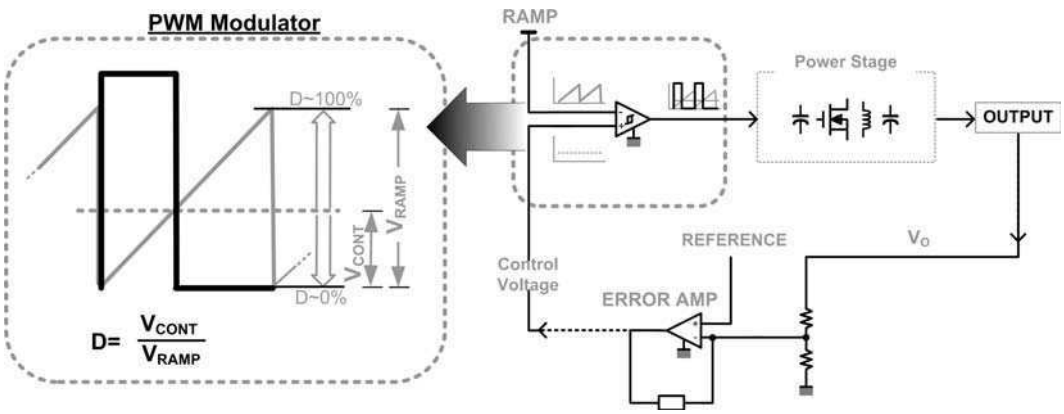


Figure 2-9: The Pulse Width Modulator Section of a Power Converter

voltage') is fed to one of the two inputs of a 'pulse width modulator' (PWM) comparator. On the other input of this PWM comparator, a *ramp* is applied, and this produces the switching pulses. So for example, if the error at the output increases, the control voltage will decrease, and the duty cycle will thus decrease in an effort to reduce the output voltage. That is how regulation usually works.

In voltage mode control, the ramp applied to the PWM comparator is derived from an internal (fixed) clock. However in current mode control, it is derived from the *inductor current (or switch current)*. And the latter leads to a rather odd situation where even a slight disturbance in the inductor current waveform can become *worse in the next cycle* (see upper half of *Figure 2-10*).

Eventually, the converter may lapse into a strange "one pulse wide, one pulse narrow" switching waveform. This represents an operating mode that is definitely not "legitimate" or

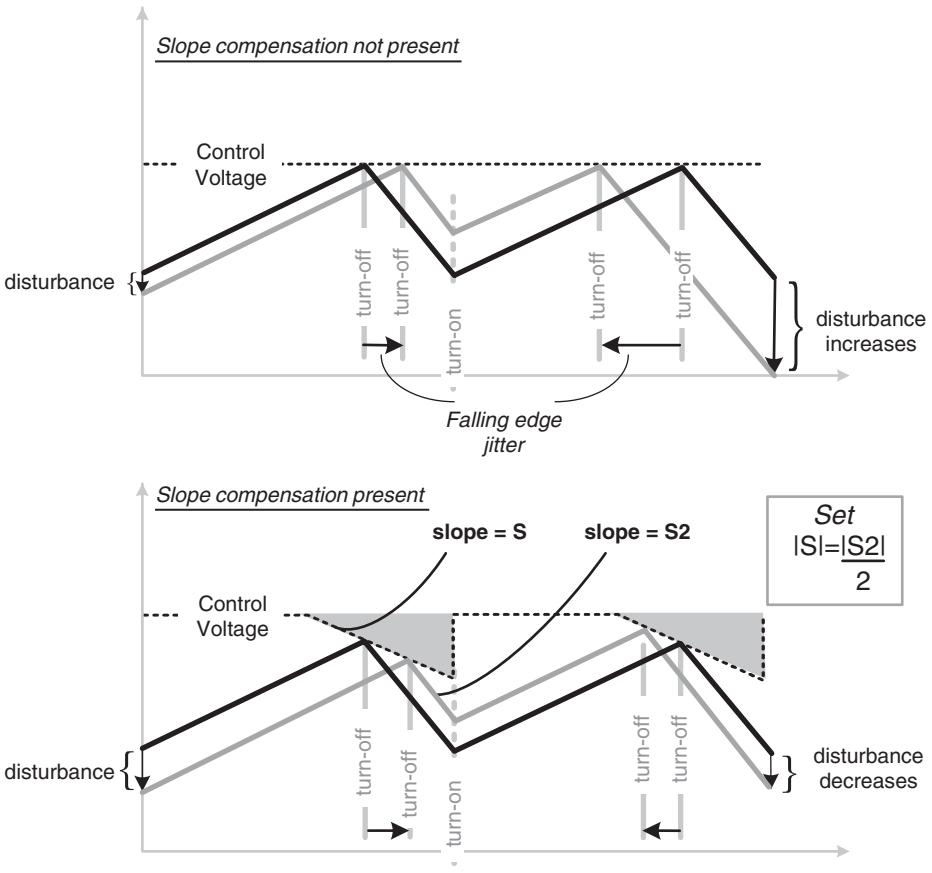


Figure 2-10: Subharmonic Instability in Current Mode Control, and Avoiding It By Slope Compensation

desirable for several reasons — in particular, the output voltage ripple is now much higher, and the loop response is severely degraded.

To get the disturbance to *decrease* every cycle and eventually die out, it can be shown that we need to do one of two things. Actually, both methods effectively amount to mixing a *little voltage-mode control into current-mode control*. So

a) Either we add a small fixed (clock-derived) voltage ramp to the sensed voltage ramp (derived from the inductor/switch)

b) Or we subtract the same fixed voltage ramp from the control voltage (output of error amplifier)

As we can see from *Figure 2-11*, both are equivalent. That is in fact not surprising at all, considering that both the ramp and the control voltage go to the pins of a *comparator*. So if we compare a signal A + B with a signal C, that is exactly equivalent to comparing A to C − B. And in both cases, equality at the input pins is established when A + B = C.

This technique is called 'slope compensation,' and is the most recognized way of quenching the alternate wide and narrow pulsing (or 'subharmonic instability') associated with current-mode control (see lower half of *Figure 2-10*).

It can be shown that to avoid subharmonic instability, we need to ensure that the amount of slope compensation (expressed in A/s) is equal to *half* the slope of the falling inductor current ramp, *or more*. Note that in principle, subharmonic instability can occur only if D is (close to or) greater than 50%. So slope compensation can be applied either over the full duty cycle range, or just for $D \geq 0.5$ as shown in *Figure 2-10*. Note that subharmonic instability can also occur only if we are operating in continuous conduction mode (CCM). So one way of avoiding it altogether is to operate in DCM.

If the amount of slope compensation is *fixed* by the controller, then as designers, we need to personally ensure that the slope of the falling inductor current ramp is equal to *twice* the slope compensation — or *less* (note that we are talking in terms of the magnitudes of the slopes only). This will in effect dictate a certain *minimum value of inductance*. And in terms of *r*, this tells us that we could have a situation where we may need to set *r* to *less* than the optimum of 0.4 — for example if the control IC has an inadequate amount of built-in slope compensation.

As a result of more detailed modeling of current-mode control, optimum relationships for the minimum inductance required (to avoid subharmonic instability) have been generated as follows

$$L \geq \frac{D - 0.34}{\text{slopecomp}} \times V_{IN}\ \mu H \quad \textit{(buck)}$$

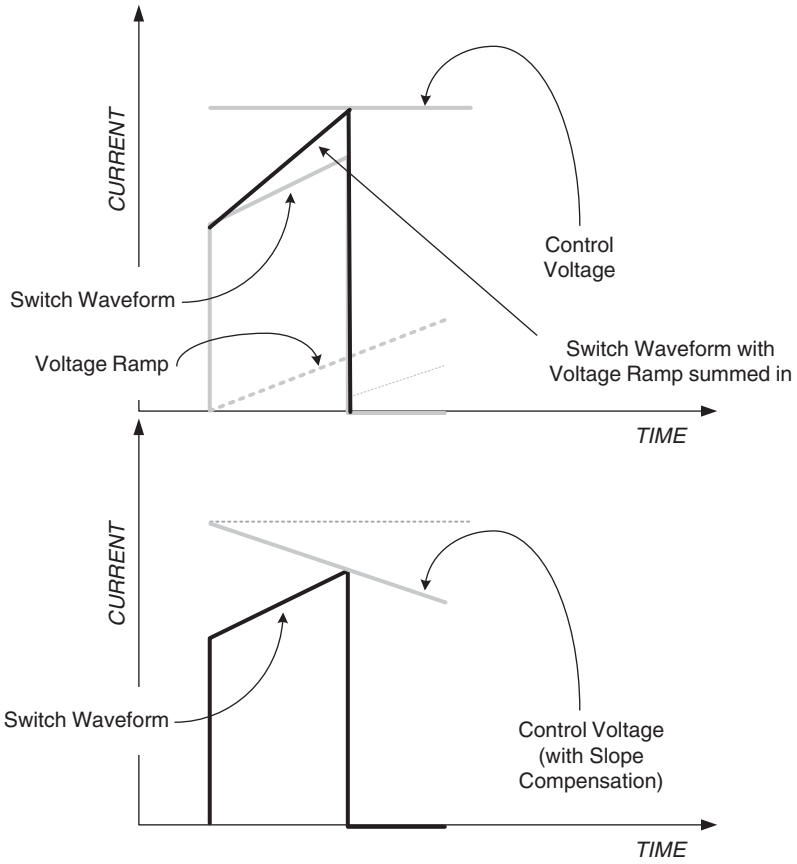


Figure 2-11: Adding a Fixed Ramp to the Sensed Signal, or Modifying the Control Voltage, Are Equivalent Methods of Slope Compensation in Current-mode Control

$$L \geq \frac{D - 0.34}{\text{slopecomp}} \times V_O \ \mu H \quad \textit{(boost)}$$

$$L \geq \frac{D - 0.34}{\text{slopecomp}} \times \left(V_{IN} + V_O\right) \ \mu H \quad \textit{(buck-boost)}$$

where the slope compensation is in A/μs.

Note that for all these topologies, we have to do the preceding calculation at the *maximum input voltage point at which the duty cycle is greater than 50%, AND we are also simultaneously in CCM.*

More details on subharmonic instability and slope compensation can be found in Chapter 7.

Quick-selection of Inductors Using "L × I" and "Load Scaling" Rules

Finally, having decided upon the value of *r* based on all the considerations outlined so far, we first present a *quick* method of picking an inductor for a given application. After that we will proceed to a more detailed analysis and worked example.

As mentioned previously, from the inductor equation $V = LdI/dt$, we can derive another useful relationship that we are calling the "L × I" equation

$$(L \times I_L) = \frac{Et}{r} \quad \textit{(any topology)}$$

Symbolically

$$L \times I = {}^{\text{voltseconds}}/_{\text{current ripple ratio}} \quad \textit{(any topology)}$$

So if we know the voltseconds (from our application conditions), and have a target value for *r*, we can calculate "L × I." Then knowing I, we can calculate L.

Note that *L × I can be visualized as a sort of* ***inductance "per" ampere*** — except that the relationship is *inverse* — that is, if we *increase* the current, we need to *decrease* the inductance (by the same amount). So for example, if we get an inductance of 100 μH for a 2-A application, then for a 1 A application, the inductance must be 200 μH, and for a 4-A application, the inductance would be 50 μH, and so on.

Note that because the L × I equation doesn't depend on topology, switching frequency, or on the specific input/output voltages, we can graph it out universally, as in *Figure 2-12*. That helps quickly pick an inductance for any application. Let us now exemplify the L × I graphical selection method for each topology.

Worked Examples (2, 3, and 4)

Buck: *Suppose we have an input of 15–20 V, an output of 5 V, and a maximum load current of 5 A. What is the recommended inductance if the switching frequency is 200 kHz?*

a) We need to start the inductor design at V_{INMAX} (20 V) for a buck.

b) The duty cycle from *Table 2-1* is $V_O/V_{IN} = 5/20 = 0.25$.

c) The time period is $1/f = 1/200\ \text{kHz} = 5\ \mu\text{s}$.

d) The off-time t_{OFF} is $(1 - D) \times T = (1 - 0.25) \times 5 = 3.75\ \mu\text{s}$.

e) The voltseconds (calculated using the *off-time*) is $V_O \times t_{OFF} = 5 \times 3.75 = 18.75\ \mu\text{s}$.

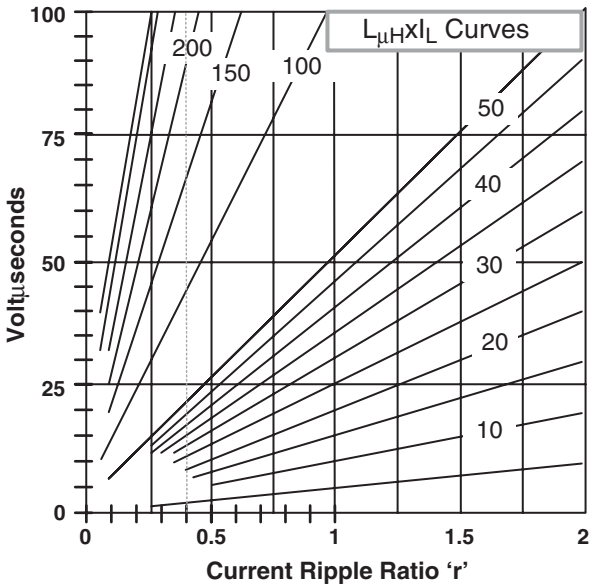


Figure 2-12: The "L x I" Curves for Quick Selection of Inductance

f) From *Figure 2-12*, with $r = 0.4$, and Et =18.75 μs, we get L × I = 45 μHA.

g) For a 5 A load, $I_L = Io = 5$ A.

h) Therefore, we need L = 45/5 = 9 μH.

i) The inductor must be rated for at least $(1 + r/2) \times I_L = 1.2 \times 5 = 6$ A.

Summarizing, we need a 9 μH/6 A inductor (or closest available).

Boost: *Suppose we have an input of 5 to 10 V, an output of 25 V, and a maximum load current of 2 A. What is the recommended inductance if the switching frequency is 200 kHz?*

a) We need to start the inductor design at V_{INMIN} (5 V) for a boost.

b) The duty cycle from *Table 2-1* is $(V_O - V_{IN})/V_O = (25 - 5)/25 = 0.8$.

c) The time period is 1/f = 1/200 kHz = 5 μs.

d) The on-time t_{ON}, is D × T = 0.8 × 5 = 4 μs.

d) The voltseconds (calculated using the *on-time*) is $V_{IN} \times t_{ON} = 5 \times 4 = 20$ μs.

e) From *Figure 2-12*, with $r = 0.4$, and Et = 20 μs, we get L × I = 47 μHA.

f) For a 2 A load, $I_L = Io/(1 - D) = 2/(1 - 0.8) = 10$ A.

g) Therefore, we need L = 47/10 = 4.7 μH.

h) The inductor must be rated for at least $(1 + r/2) \times I_L = 1.2 \times 10 = 12$ A.

Summarizing, we need a 4.7 μH/12 A inductor (or closest available).

Buck-boost: *Suppose we have an input of 5 to 10 V, an output of –25 V output, and a maximum load current of 2 A. What is the recommended inductance if the switching frequency is 200 kHz?*

a) We need to start the inductor design at V_{INMIN} (5 V) for a buck-boost.

b) The duty cycle from *Table 2-1* is $V_O/(V_{IN} + V_O) = 25/(5 + 25) = 0.833$.

c) The time period is 1/f = 1/200 kHz = 5 μs.

d) The on-time t_{ON} is D × T = 0.833 × 5 = 4.17 μs.

e) The voltseconds (calculated using the *on-time*) is $V_{IN} \times t_{ON} = 5 \times 4.17 = 20.83$ μs.

f) From *Figure 2-12*, with $r = 0.4$, and Et = 20.83 μs, we get L × I = 52 μHA.

g) For a 2 A load, $I_L = Io/(1 - D) = 2/(1 - 0.833) = 12$ A.

h) Therefore, we need L = 52/12 = 4.3 μH.

i) The inductor must be rated for at least $(1 + r/2) \times I_L = 1.2 \times 12 = 14.4$ A.

Summarizing, we need a 4.3 μH/14.4 A inductor (or closest available).

The Current Ripple Ratio r in Forced Continuous Conduction Mode ('FCCM')

Finally, before we move on to magnetic fields, we make some closing remarks on designing with forced continuous conduction mode ('FCCM').

We had said previously, that by definition, *r* is defined only for CCM, and therefore cannot exceed 2 (since that marks the boundary between CCM and DCM). However in *synchronous regulators* (with diode replaced or supplanted by a low-drop mosfet across it), we actually never enter DCM (unless the IC is deliberately designed to mimic that mode on demand). So now, on decreasing the load, we actually continue to remain in CCM. That is because for DCM to ever occur, the inductor current must be forced to *stay* at least for some part of the switching cycle at zero. And to get that to happen, we need to have a reverse-biased diode that prevents the inductor current from "going the other way." But in synchronous regulators, the mosfet across the diode allows reverse-conduction even if the diode is reverse-biased, so we do not get DCM.

The CCM-type mode that replaces the DCM mode in synchronous regulators is distinguished from the usual (normal) CCM mode, by calling it the 'forced continuous conduction mode' (FCCM). The main switch is usually identified as the *top* (or "high-side") mosfet, whereas

the mosfet across the diode is called the *bottom* (or "low-side") mosfet. Further, in FCCM, *r is legitimately allowed to exceed 2* (see *Figure 2-5*).

We can visualize FCCM as starting to occur when the load current is decreased sufficiently to cause part of the inductor current waveform to become "submerged" below ground — that is, with parts of it having a negative value (inductor current flowing momentarily away from the load). But note that as long as we are still drawing some load current *out* of the output terminals of the converter, the *average* value of the waveform, I_{DC} (center of ramp), is still positive — that is, going towards the load — *on an average*. Further, because I_{DC} is always proportional to the load current, it can be made to decrease all the way *down to zero* while still maintaining CCM. Since the *swing* in current, ΔI, depends *only* on the input and output voltages, which we have assumed have not changed, the ratio $r = \Delta I/I_L$ not only exceeds 2, but can in fact become extremely large.

All the basic design equations we can write for the RMS, dc, ac, or peak currents in the input/output capacitors and the switch, when operating in conventional CCM, *apply to the converter in FCCM too* (though there may be some *additional* losses, as for example when the current flows through the body diode of the top mosfet). This, despite the fact that *r* can now exceed 2. In other words, the CCM equations do *not* get invalidated in FCCM. However, a specific computational problem can arise in some cases, because if *r* is infinite (zero load current), we can get a singularity — a "0" in the denominator. At first sight, that seems to make the CCM equations (presented the way we have been doing), unusable. But one trick we can employ to avoid the singularity is to *assume* a few milliamperes of minimum load, however small. Alternatively, we can substitute $r = \Delta I/I_{DC}$ back into the equations, and we will then see that I_{DC} cancels out (does not appear in the denominator anywhere). Either way, the equations of CCM (see Appendix 2), apply to FCCM too.

Basic Magnetic Definitions

Having understood basic concepts like voltseconds, current components, worst-case voltage, and also how to do an initial (quick) selection of an off-the-shelf inductor, we will now try to go *inside* the magnetic component, so as to learn what happens in terms of the *magnetic fields* present inside its core. We will then use this information to do a more complete validation of a selected off-the-shelf inductor. Then we will find the remaining (worst-case) stresses of the converter.

At the outset, we should note that in magnetics, there are several different *systems of units* in use. This can become very confusing, since even the basic equations look different depending on the system in use. It is therefore a wise policy to stick to one system of units all the way through — converting to a different system, if required, *only at the very end*, that is, only at the level of *numerical* results (not at the level of the *equations*).

Further, unless otherwise stated, the reader can safely assume we are using the *meter-kilogram-seconds* system of units — that is, 'MKS' system, also called the 'SI' system (for System International).

Here are the basic definitions:

- H-field: Also called 'field strength,' 'field intensity,' 'magnetizing force,' 'applied field,' and so on. Its units are A/m.
- B-field: Also called 'flux density' or 'magnetic induction.' Units of B are tesla ('T') or webers per square meter (Wb/m^2).
- Flux: This is the integral of B over a given surface area. That is,

$$\phi = \int_S BdS \text{ Wb}$$

  If B is constant over the surface, we get the more common form $\phi = BA$ where A is the area of the surface.

  Note: The integral of B over a closed surface is zero since flux lines do not start or end at any given point but are continuous.

- B is related to H *at any given point* by the equation $B = \mu H$ where μ is the *permeability* of the material. Note that later we will use the symbol μ for 'relative permeability,' that is, the *ratio* of the permeability of the material to that of air. So in MKS units we should actually preferably write $B = \mu_c H$, where μ_c is the permeability of the core (magnetic material). By definition, $\mu_c = \mu\mu_0$.
- The permeability of air, denoted by μ_0, is equal to $4\pi \times 10^{-7}$ Henries/m in MKS units. In CGS units it is equal to 1. That is why in CGS units $\mu_c = \mu$, where μ is also automatically the relative permeability of the material (though units are different).
- Faraday's law of induction (also called Lenz's law) relates the induced voltage V that is developed across the ends of a coil (N turns), to the (time varying) B-field passing through it. So

$$V = N\frac{d\phi}{dt} = NA\frac{dB}{dt}$$

- The "inertia" of a coil to a change in flux through it due to a time varying current through it is its 'inductance' L, defined as

$$L = \frac{N\phi}{I} \text{ Henries}$$

- Since it can be shown that the flux is proportional to the number of turns N, the inductance L is proportional to the *square* of the number of turns. This proportionality constant is called the 'inductance index' and is denoted by 'A_L.' It is usually expressed as nH/turns^2 (though sometimes it is considered to be mH/1000 turns^2, both being *numerically* the same). So

$$L = A_L \times N^2 \times 10^{-9} \text{ Henries}$$

- When H is integrated over a *closed loop*, we get the current enclosed by the loop

$$\oint Hdl = I \text{ Amperes}$$

  where the integration symbol above reflects the fact that it is being performed over a closed loop. This is also called 'Ampere's circuital law.'

- Combining Lenz's law with the inductor equation $V = LdI/dt$, we get

$$V = N\frac{d\phi}{dt} = NA\frac{dB}{dt} = L\frac{dI}{dt}$$

- From this we get the *two key equations* used in power conversion

$$\Delta B = \frac{L\Delta I}{NA} \quad \textit{("voltage independent equation")}$$

$$\Delta B = \frac{V\Delta t}{NA} \quad \textit{("voltage dependent equation")}$$

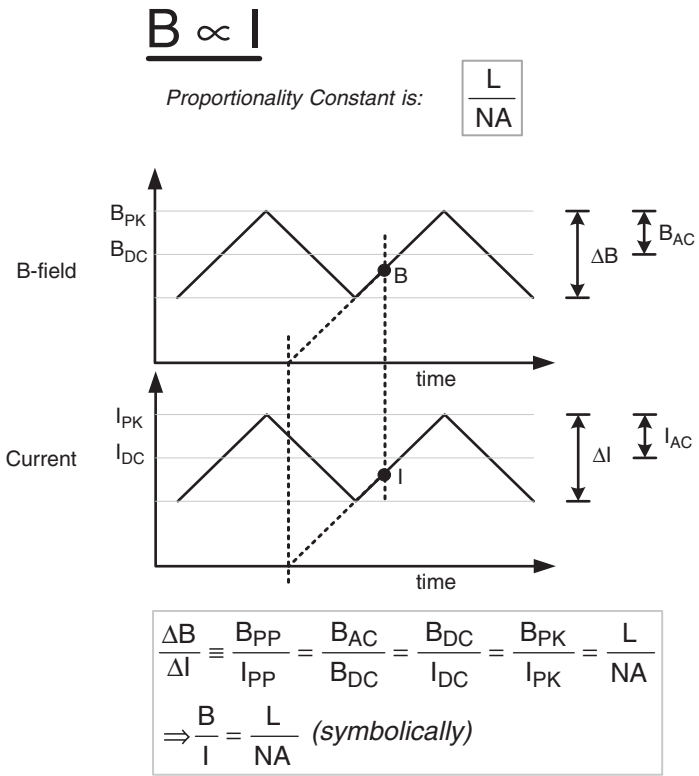
  The first equation can be written symbolically as

$$B = \frac{LI}{NA} \quad \textit{(voltage independent equation)}$$

  And the latter equation can be written in a more "power-conversion-friendly" form as follows

$$B_{AC} = \frac{V_{ON}\, D}{2 \times NAf} \quad \textit{(voltage dependent equation)}$$

For most inductors used in power conversion, if we reduce the current to zero, the field inside the core also goes to zero. Therefore, an implicit assumption of complete linearity is also usually made — that is, *B and I are considered proportional to each other* as shown in *Figure 2-13* (unless of course the core starts saturating, at which point, all bets are off!). The *voltage independent equation* can then be expressed as any of the equations shown in the figure — in other words, *this proportionality applies to the peak values of current and field,*



$$\frac{\Delta B}{\Delta I} \equiv \frac{B_{PP}}{I_{PP}} = \frac{B_{AC}}{B_{DC}} = \frac{B_{DC}}{I_{DC}} = \frac{B_{PK}}{I_{PK}} = \frac{L}{NA}$$

$$\Rightarrow \frac{B}{I} = \frac{L}{NA} \quad \textit{(symbolically)}$$

Figure 2-13: B and I Can Be Usually Considered to Be Proportional to Each Other

their average values, their ac values, their dc values, and so on. The constant of proportionality is equal to

$$\frac{L}{NA} \quad \textit{(Proportionality constant linking B and I)}$$

where N is the number of turns and A the actual geometrical cross-sectional area of the core (its center limb usually, or simply the 'effective area' A_e given in the datasheet of the core).

Worked Example (5) — When Not to Increase the Number of Turns

Note that the voltage-independent equation is useful if for example we want to do a *quick check to see if our core may be saturating*. Suppose we are custom-designing our inductor. We have wound 40 turns on a core with an area of A = 2 cm^2. Its *measured* inductance

is 200 μH, and the peak inductor current in our given application is 10 A. Then the peak flux density can be calculated as follows

$$B_{PK} = \frac{L}{NA} I_{PK} = \frac{200 \times 10^{-6}}{40 \times \frac{2}{10^4}} = 0.25 \text{ teslas}$$

Note that we have converted the area to m^2 in the above equation, because we are using the MKS version of the equation.

For most ferrites, an operating flux density of 0.25 T is acceptable, since the saturation flux density is typically around 0.3 T.

Based on the B and I linearity, we can also *linearly extrapolate* and thus conclude that the peak current in our application should under no condition be allowed to exceed $(0.3/0.25) \times 10 = 12$ A, because at 12 A, the field will be 0.3 T, and the core will then start to saturate.

But note that nor *should the number of turns be increased any further* (at 12 A). Looking at the B_{PK} equation above, it seems at first sight that increasing the number of turns will reduce the B-field. However, *inductance increases as* N^2 (from the A_L equation given previously), so the numerator will increase much faster than the denominator. Therefore, in reality, the B-field will *increase*, rather than decrease if we increase the number of turns, and we know we can't afford to exceed 0.3 T.

In other words — we usually tend to instinctively rely on the current-limiting properties of an inductor. And in general, increasing the inductance will certainly help increase the inductance and therefore help limit the current. However, *if we are already close to the energy-storage limits of the material of the core*, *we have to be very careful* — a few extra turns could take us "over the edge" (saturation), and then in fact, the inductance will start collapsing rather than increasing.

We should also not forget our basic premise of inductors in power conversion — for a given application, *a large inductance does usually end up requiring a large inductor!* So, increasing the number of turns, without increasing the size, may naturally turn out to be a recipe for disaster.

The "Field Ripple Ratio"

Since I and B are proportional to each other, and *r* happens to be a ratio, we realize that *r* must apply equally to the field components as it does to the current components. So, in that sense, *r* can be looked at as a "field ripple ratio" too. We can therefore extend the definition

of *r* as follows

$$r = 2\frac{I_{AC}}{I_{DC}} = 2\frac{B_{AC}}{B_{DC}}$$

Therefore, *r* can also be used to relate the peak, ac, and dc values of both the current and field according to the equations

$$B_{DC} = \frac{2 \times B_{PK}}{r + 2} \quad or \quad I_{DC} = \frac{2 \times I_{PK}}{r + 2}$$
$$B_{AC} = \frac{r \times B_{PK}}{r + 2} \quad or \quad I_{AC} = \frac{r \times I_{PK}}{r + 2}$$

We can relate the peak to the swing too as follows:

$$B_{PK} = \frac{r + 2}{2 \times r} \times \Delta B \quad or \quad I_{PK} = \frac{r + 2}{2 \times r} \times \Delta I$$

The latter form will in fact be used later by us in the worked example that will follow.

The Voltage Dependent Equation in Terms of Voltseconds (MKS units)

When discussing the current swing ΔI, we related it to the voltseconds. Now we can do the same for the B-field

$$\Delta B = \frac{L \times \Delta I}{N \times A} = \frac{Et}{N \times A} \quad \text{teslas}$$

So as for current, the voltseconds in our application also determines the *swing* of the magnetic field — though not its dc level.

CGS Units

We may personally prefer to use the more broadly accepted MKS units, but we have to deal with the ground reality of the situation — that certain vendors (especially North American ones) still use 'CGS' (centimeter-gram-seconds) units. Since we would certainly be evaluating and looking at their datasheets too, we will need to use the conversions in *Table 2-4*.

In particular, we should remember that the saturation flux density B_{SAT}, which is around 0.3 T (300 mT) for most ferrites, is 3000 gauss ('G') in CGS units. Also note that permeability of a material in MKS units needs to be divided by $4\pi \times 10^{-7}$ to get the permeability in CGS units. The reason for that is that permeability of air is set to 1 in CGS units, but in MKS units it is (numerically) equal to $4\pi \times 10^{-7}$.

Table 2-4: Magnetic systems of units and their conversions

	CGS units	MKS Units	Conversions
Magnetic Flux	Line (or Maxwell)	Weber	1 Weber = 10^6 Lines
Flux Density (B)	Gauss	Tesla (or Wb/m^2)	1 Tesla = 10^4 Gauss
Magnetomotive force	Gilbert	Ampere-turn	1 Gilbert = 0.796 Ampere-turn
Magnetizing Force Field (H)	Oersted	Ampere-turn/meter	1 Oersted = $1000/4\pi$ = 79.577 Ampere/meter
Permeability	Gauss/Oersted	Weber/m-Ampere-turn	$\mu_{MKS} = \mu_{CGS} \times (4\pi \times 10^{-7})$

The Voltage Dependent Equation in Terms of Voltseconds (CGS units)

It is also therefore helpful to know how to write the voltage dependent equation, (expressed in terms of Et), in CGS units instead.

So, converting A in m^2 to A in cm^2, we get from the previous equation

$$\Delta B = \frac{100 \times E_t}{N \times A} \text{ gauss} \quad (A \text{ in } cm^2)$$

Core Loss

The core loss depends on various factors — the flux swing ΔB, the (switching) frequency *f*, and the temperature (though we usually ignore this latter dependency for most estimates). Note however, that **when vendors of magnetic materials express the dependency of core loss on a certain "B," what they are really talking about is ΔB/2, that is, B_{AC}**. This happens to be the usual industry convention, but it is often quite confusing to power supply designers. In fact, there is more confusion caused by the fact that "B" may be expressed by the vendor, either in terms of gauss or in teslas. In fact, the dissipation also (due to the core loss) may be expressed either as mW or as W.

First let us look at the general form of core loss.

Core Loss = (Core Loss per unit volume) × Volume where 'core loss per unit volume' is expressed generally as

$$\text{constant } t_1 \times B^{\text{constant } t_2} \times f^{\text{constant } t_3}$$

In *Table 2-5* we have indicated the three main systems of units in use for describing the core loss per unit volume, and also provided the rules for converting between them. Note we are using 'V_e' (effective volume) here — this can usually be considered to be simply the actual physical volume of the core, or we can just look it up in the datasheet of the core.

Table 2-5: The different systems in use for describing core loss (and their conversions)

	Constant	exponent of B	exponent of f	B	f	Ve	Units
System A	Cc	Cb	Cf	Tesla	Hz	cm^3	W/cm^3
	$= \frac{C \times 10^{4\times p}}{10^3}$	= p	= d				
System B	C	p	d	Gauss	Hz	cm^3	mW/cm^3
	$= \frac{Cc \times 10^3}{10^{4\times Cb}}$	= Cb	= Cf				
System C	Kp	n	m	Gauss	Hz	cm^3	W/cm^3
	$= \frac{C}{10^3}$	= p	= d				

In *Table 2-6* we have provided values for the constants in the core loss equation in one of these systems of units, besides some other operating limits. The reader is however advised to confirm these values from the respective vendors.

Worked Example (6) — Characterizing an Off-the-shelf Inductor in a Specific Application

Now we will present the ***"general inductor design procedure"*** we have been talking about. We will be considering a *wide-input* voltage range here. The procedure is to be carried out at the "*worst-case input voltage end"* ***with respect to the peak current***. The basic purpose is to ensure that we are avoiding inductor saturation under normal operation. So for the ***buck***, we will work at V_{INMAX}, since that is the point at which the peak current is at its maximum. For a ***boost or a buck-boost***, we need to conduct this procedure at V_{INMIN}, not V_{INMAX}, since that is the worst-case input voltage end with regard to the peak current, for these topologies.

The procedure will be illustrated by means of a step-by-step worked example. Though it is carried out for a buck, throughout the calculation, we will indicate precisely how the procedure and equations may need to change, were this a boost or a buck-boost. So for example, to the right of any equation presented below, we have indicated in brackets, which topology it is valid for.

A buck converter has an input of 18–24 V, an output of 12 V, and a maximum load of 1 A. We desire a current ripple ratio of 0.3 (at maximum load). We assume $V_{SW} = 1.5$ V, $V_D = 0.5$ V, *and* $f = 150{,}000$ Hz. *An off-the-shelf inductor is to be selected and characterized for this application.*

Table 2-6: Typical core loss coefficients of common materials

Material (Vendor)	Grade	C	p (B^p)	d (f^d)	μ	≈ B_{SAT} (gauss)	≈ f_{MAX} (MHz)
Powdered Iron (Micrometals)	8	4.3E−10	2.41	1.13	35	12500	100
	18	6.4E−10	2.27	1.18	55	10300	10
	26	7E−10	2.03	1.36	75	13800	0.5
	52	9.1E−10	2.11	1.26	75	14000	1
Ferrite (Magnetics Inc)	F	1.8E−14	2.57	1.62	3000	3000	1.3
	K	2.2E−18	3.1	2	1500	3000	2
	P	2.9E−17	2.7	2.06	2500	3000	1.2
	R	1.1E−16	2.63	1.98	2300	3000	1.5
Ferrite (Ferroxcube)	3C81	6.8E−14	2.5	1.6	2700	3600	0.2
	3F3	1.3E−16	2.5	2	2000	3700	0.5
	3F4	1.4E−14	2.7	1.5	900	3500	2
Ferrite (TDK)	PC40	4.5E−14	2.5	1.55	2300	3900	1
	PC50	1.2E−17	3.1	1.9	1400	3800	2
Ferrite (Fair-Rite)	77	1.7E−12	2.3	1.5	2000	3700	1
Note: (a)E−(b) is the same as (a)×$10^{-(b)}$							

As mentioned, all the steps involved in the "general inductor design procedure" below are being carried out at a certain "V_{IN}" — which is the *maximum input voltage* for a ***buck***, and *minimum input voltage* for a ***boost or a buck-boost***.

Estimating Requirements

For a buck regulator, the duty cycle is (now including the switch and diode forward drops)

$$D = \frac{V_O + V_D}{V_{IN} - V_{SW} + V_D} \quad \textit{(buck)}$$

So

$$D = \frac{12 + 0.5}{24 - 1.5 + 0.5} = 0.543$$

(For boost, use $D = \frac{V_O - V_{IN} + V_D}{V_O - V_{SW} + V_D}$*, for buck-boost, use* $D = \frac{V_O + V_D}{V_{IN} + V_O - V_{SW} + V_D}$*).*

The switch on-time is therefore

$$t_{ON} = \frac{D}{f} \Rightarrow \frac{0.543}{150000} \Rightarrow 3.62\ \mu s \quad \textit{(any topology)}$$

$$t_{ON} = 3.62\ \mu s$$

The voltage across the inductor when the switch is ON is

$$V_{ON} = V_{IN} - V_{SW} - V_O = 24 - 1.5 - 12 = 10.5 \text{ V} \quad \textit{(buck)}$$

(For boost and buck-boost, use $V_{ON} = V_{IN} - V_{SW}$*).*

So the voltsμseconds is

$$Et = V_{ON} \times t_{ON} = 10.5 \times 3.62 = 38.0 \text{ V}\mu\text{s} \quad \textit{(any topology)}$$

Using the "L × I" equation

$$(L \times I_L) = \frac{Et}{r} \quad \textit{(any topology)}$$

we get

$$(L \times I_L) = \frac{38}{0.3} = 127 \ \mu\text{H-A}$$

But the average inductor current is

$$I_L = I_O \quad \textit{(buck)}$$

(For a boost and buck-boost, use $I_L = \frac{I_O}{1 - D}$*).*

Therefore

$$L = \frac{(L \times I_L)}{I_L} \equiv \frac{(L \times I_O)}{I_O} = \frac{127}{1} = 127 \ \mu\text{H} \quad \textit{(any topology)}$$

The peak current will be 15% higher than I_L for r = 0.3. This follows from

$$I_{PK} = \left(1 + \frac{r}{2}\right) \times I_L = 1.15 \times 1 = 1.15 \text{ A} \quad \textit{(any topology)}$$

We now pick a promising off-the-shelf inductor — the PO150 from Pulse Engineering. Its inductance is 137 μH, which is close to our requirement of 127 μH, and it is rated for a continuous dc of 0.99 A, which is very close to our requirement of 1 A. Its datasheet is reproduced in *Table 2-7*. Note that the other conditions mentioned by the vendor *do not match our application* (but that is not unexpected — what are the chances of an off-the-shelf inductor that precisely matches a given application?). Nevertheless we can perform a full analysis, and thus either validate, or invalidate our choice of component.

Table 2-7: Specifications of a selected inductor (the PO150)

I_{DC} (A)	L_{DC} (μH)	Et (Vμs)	DCR (mΩ)	Et_{100} (Vμs)
0.99	137	59.4	387	10.12

- The inductor is such that 380 mW dissipation corresponds to 50°C rise in temperature.
- The core loss equation for the core is $6.11 \times 10^{-18} \times B^{2.7} \times f^{2.04}$ mW where f is in Hz and B is in gauss.
- Et_{100} is the Vμsecs at which "B" is 100 gauss.
- "B" is B_{AC}, i.e. $\Delta B/2$.
- Rated frequency of operation is 250 kHz.

Current Ripple Ratio

We use the "L × I" rule

$$(L \times I_L) = \frac{Et}{r} \quad \textit{(any topology)}$$

So

$$r = \frac{Et}{L \times I_L} \quad \textit{(any topology)}$$

The inductor has been designed by its vendor, for an *r* of

$$r = \frac{59.4}{137 \times 0.99} = 0.438$$

In our application we will get

$$r = \frac{38}{137 \times 1} = 0.277$$

This is very close to (and less than) our target of r = 0.3, and is therefore acceptable.

Peak Current

The inductor has been designed for a peak current of

$$I_{PK} = \left(1 + \frac{r}{2}\right) \times I_L = \left(1 + \frac{0.438}{2}\right) \times 0.99 = 1.21 \text{ A} \quad \textit{(any topology)}$$

In our application we will get

$$I_{PK} = \left(1 + \frac{r}{2}\right) \times I_L = \left(1 + \frac{0.277}{2}\right) \times 1 = 1.14 \text{ A} \quad \textit{(any topology)}$$

The peak current in our application is considered "safe," being *less* than what the inductor was originally designed for. Therefore, we can safely assume that the peak B-field of our application also must be within the design limits of the inductor. However it is instructive to confirm that directly, as we will do next.

*Note that the frequency has not even entered the picture directly so far, since volt*μ*seconds is all that really matters to an inductor*. ***Different applications, with the same dc level of current, and the same voltseconds, are essentially the same application from the viewpoint of the inductor***. It just "doesn't care" for example, what topology this is, or what is the duty cycle. It doesn't even care about the frequency directly (though the exception to this is the core loss term, because that depends not only on the voltseconds, i.e. the current swing, but on the frequency too). However, we will also see that the core loss term is much smaller anyway, compared to the copper loss. So for all practical purposes, ***if the rated voltseconds of a given inductor (current swing), and its dc current rating correspond to the voltseconds and dc current of our application, we are almost certainly going to be fine right off-the-bat***. ***However, even if the rated voltseconds and dc level are quite different, as long as the peak flux density is close to or less than the rated value, we are OK from the saturation point of view. That's a good start, and we can then proceed to do a full validation analysis*** — of the temperature rise and so on under our specific application conditions.

Flux Density

The vendor provides the following information (see *Table 2-7*):

$$Et_{100} = 10.12 \text{ V}\mu\text{s}$$

This means that the voltμseconds that produces a B_{AC} of 100 gauss is 10.12. Since $B_{AC} = \Delta B/2$, the corresponding ΔB is 200 gauss (for every 10.12 Vμs).

We had previously presented the following relationship between ΔB and Et:

$$\Delta B = \frac{100 \times Et}{N \times A} \text{ gauss} \quad \textit{(any topology)}$$

Since ΔB and Et are proportional to each other (for a given inductor), we can conclude that the inductor has been designed for a flux density swing of

$$\Delta B = \frac{Et}{Et_{100}} \times 200 = \frac{59.4}{10.12} \times 200 = 1174 \text{ gauss} \quad \textit{(any topology)}$$

and a peak flux density of

$$B_{PK} = \frac{r+2}{2\times r} \times \Delta B = \frac{0.438+2}{2\times 0.438} \times 1174 = 3267 \text{ gauss} \quad \textit{(any topology)}$$

In our application this will give us a swing of

$$\Delta B = \frac{Et}{Et_{100}} \times 200 = \frac{38}{10.12} \times 200 = 751 \text{ gauss} \quad \textit{(any topology)}$$

and a peak of

$$B_{PK} = \frac{r+2}{2\times r} \times \Delta B = \frac{0.277+2}{2\times 0.277} \times 751 = 3087 \text{ gauss} \quad \textit{(any topology)}$$

We see that the peak field in our application is within the design limits of the inductor, as expected, so we need not worry about core saturation. This is a basic qualification the inductor must pass before we can proceed with the rest of the analysis.

Note that the proportionality constant connecting B and I is

$$\frac{L}{NA} = \frac{B_{PK}}{I_{PK}} = \frac{3087}{1.14} = 2708 \text{ gauss/A} \quad \textit{(any topology)}$$

Note: If we break open the inductor and measure the number of turns, and also estimate/measure the cross-sectional area of the central limb of its core, we can verify this number.

Copper Loss

From the equations contained in *Figure 2-14*, we can calculate the RMS of the inductor current waveform. The inductor was designed for an RMS *squared* of

$${I_{RMS}}^2 = \frac{\Delta I^2}{12} + {I_{DC}}^2 = {I_{DC}}^2\left(1+\frac{r^2}{12}\right) = 0.99^2\left(1+\frac{0.438^2}{12}\right) = 0.996 \text{ A}^2 \quad \textit{(any topology)}$$

and a copper loss of

$$P_{CU} = {I_{RMS}}^2 \times DCR = 0.996 \times 387 = 385 \text{ mW} \quad \textit{(any topology)}$$

Whereas in our application we will get

$${I_{RMS}}^2 = {I_L}^2\left(1+\frac{r^2}{12}\right) = 1^2\left(1+\frac{0.277^2}{12}\right) = 1.006 \text{ A}^2 \quad \textit{(any topology)}$$

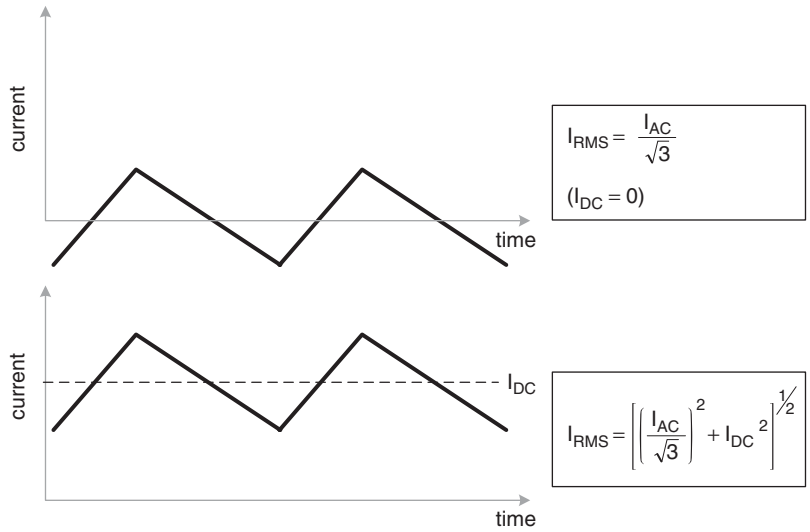


Figure 2-14: RMS Value of an Inductor Current Waveform

and a copper loss of

$$P_{CU} = I_{RMS}{}^2 \times DCR = 1.006 \times 387 = 389 \text{ mW} \quad \textit{(any topology)}$$

Core Loss

Note that the vendor has already factored in the *volume of the core* and thus provided the following overall equation for the core loss of the inductor:

$$P_{CORE} = 6.11 \times 10^{-18} \times B^{2.7} \times f^{2.04} \text{ mW} \quad \textit{(any topology)}$$

where f is in Hz and B is in gauss. Note that "B" is $\Delta B/2$ here as per convention.

So the core loss that the inductor was originally designed for is

$$P_{CORE} = 6.11 \times 10^{-18} \times \left(\frac{1174}{2}\right)^{2.7} \times \left(250 \times 10^3\right)^{2.04} = 18.8 \text{ mW}$$

Whereas in our application

$$P_{CORE} = 6.11 \times 10^{-18} \times \left(\frac{751}{2}\right)^{2.7} \times \left(150 \times 10^3\right)^{2.04} = 2 \text{ mW}$$

In general, we will find that in most ferrite-based off-the-shelf inductors, the designed core loss is only 5 to 10% of the total inductor loss (copper-plus-core loss). However, if the inductor uses a 'powdered iron' core, this number may rise to about 20 to 30%.

Note: Powdered iron cores tend to saturate more "softly" than ferrites, and that usually enhances their ability to withstand severe abnormal currents without leading to immediate switch destruction. On the other hand, powdered iron cores may have "lifetime" issues caused by slow degradation of the organic binder that holds their iron particles together. The vendor must be consulted about this possibility, and the steps necessary to avoid a premature end to our converter!

Temperature Rise

The vendor has stated that the inductor is such that 380 mW dissipation corresponds to 50°C rise in temperature. In effect this tells us that the thermal resistance of the core 'Rth' is

$$\text{Rth} = \frac{\Delta T}{W} = \frac{50}{0.38} = 131.6^{\circ}\text{C/W} \quad \textit{(any topology)}$$

The inductor was originally designed for a total loss of

$$P = P_{CORE} + P_{CU} = 385 + 18.8 = 403.8 \text{ mW} \quad \textit{(any topology)}$$

This would have given a temperature rise of

$$\Delta T = \text{Rth} \times P = 131.6 \times 0.404 = 53^{\circ}\text{C} \quad \textit{(any topology)}$$

In our application

$$P = P_{CORE} + P_{CU} = 389 + 2 = 391 \text{ mW}$$

This will give a temperature rise of

$$\Delta T = \text{Rth} \times P = 131.6 \times 0.391 = 51^{\circ}\text{C}$$

Provided we accept this temperature rise in our application (that will depend on our maximum operating ambient temperature), we can validate the chosen inductor. We have already confirmed it does not saturate in our application, and further, the current ripple ratio it provides is acceptable too.

This completes the general inductor design procedure.

Calculating the "Other" Worst-case Stresses

Having validated our choice of inductor, we can look a little more closely at the important issue of how the wide-input range impacts the ***other*** *key parameters and stresses in our proposed converter. This also helps in correctly selecting the other power components.*

Worst-case Core Loss

In the above so-called "general inductor design procedure," we have actually been working at V_{INMAX} for a buck, and at V_{INMIN} for a boost or buck-boost. The reason was that the *inductor* sees the highest peak current, at this voltage end, so we have to "insure" the magnetics design at this particular point. But this point *may not be the worst-case point for the other stresses in the power supply*, and we need to start understanding that clearly now.

Let us first focus on the inductor itself. The point at which we are doing the inductor design will usually always give us the worst-case temperature rise too. *But that is because the I_{DC} component of the inductor current is the dominant term.* If for any reason, we are interested in knowing what the maximum *core loss* component of the total loss is, we should realize, looking back at *Figure 2-4,* that though the dc level may be going up, the ac component (on which the core loss term depends) may be decreasing (or even having an odd-shaped profile, as for the boost).

From *Figure 2-4*, we see that I_{AC} increases at high input voltages for both the buck and the buck-boost. For a ***buck***, the general inductor design calculation above was carried out at V_{INMAX} and that just happens to be the point at which the core loss is a maximum too. Therefore, calculating the core loss at V_{INMAX} as we did in the previous example does coincidentally also give us the worst-case core loss.

However, if we were doing the calculation for a ***buck-boost***, our general inductor design calculation would be being carried out at V_{INMIN}. But the core loss is a maximum at V_{INMAX}. Similarly, for a ***boost***, we would also be carrying out the general inductor design calculation at V_{INMIN}. But the worst-case core loss for this topology occurs at $D = 0.5$ (see I_{AC} curve for boost in *Figure 2-4*). From the duty cycle equation of the boost, $D = 0.5$ corresponds to an input voltage equal to half the output.

> **Note:** If for the boost, the input range of the given application does *not* include the $D = 0.5$ point, we need to identify which voltage end of the range provides a duty cycle closest to $D = 0.5$. And we need to then do the worst-case core loss calculation at that end (if so desired).

Generally, the core loss term, being such a small component of the total loss, is of no great concern to us, so we won't even bother to do a numerical calculation here. But the general procedure to handle such cases will become apparent as we study the other worst-case loss terms of the converter below.

But first let us now start *annotating (or subscripting)* some of the terms derived so far, just for gaining clarity in the discussion to follow. We should be clear that

- ***For a buck*:** The general inductor design procedure was carried out at V_{INMAX}, that is, $D_{MIN.}$ So for example, the *r* we have set to 0.3–0.4 (and possibly re-calculated with the selected inductor) is actually 'r_{DMIN}' to be precise. Similarly, the voltseconds, Et, we have calculated so far is actually 'Et_{DMIN}.'
- ***For a boost and buck-boost*:** If a similar general inductor design procedure were carried for these topologies, it would be done at V_{INMIN}, that is, D_{MAX}. So for example, the *r* we would have set to 0.3–0.4 (and possibly re-calculated with the selected inductor) would actually be 'r_{DMAX}.' Similarly, the voltseconds, Et, we would have calculated so far is actually 'Et_{DMAX}.'

We need to keep these distinctions in mind, otherwise the following discussion can become confusing to no end!

Worst-case Diode Dissipation

The general equation for the average diode current is

$$I_D = I_L \times (1 - D) \quad \textit{(any topology)}$$

or equivalently

$$I_D = I_O \times (1 - D) \quad \textit{(buck)}$$
$$I_D = I_O \quad \textit{(boost and buck-boost)}$$

This leads to a diode dissipation of

$$P_D = V_D \times I_D = V_D \times I_O \times \big(1 - D\big) \quad \textit{(buck)}$$
$$P_D = V_D \times I_D = V_D \times I_O \quad \textit{(boost and buck-boost)}$$

For the buck, as the input voltage is raised, the duty cycle falls, and because the average inductor current I_L remains fixed at I_O, the average diode current *increases*. That means we get the worst-case diode current (and dissipation) at V_{INMAX} for a buck. *So we can just use the numbers we already have derived from carrying out the general inductor design procedure* (at V_{INMAX}).

For the boost and the buck-boost, as the input is raised, D decreases, but the average inductor current also falls, thereby keeping I_D always fixed at I_O. (We should remember that the boost and the buck-boost are unique in the sense that *all* the output current must pass

through the diode when it conducts, so I_D must necessarily be equal to I_O at all times). That means the diode dissipation is independent of input voltage for these topologies. So we can, if we want, just *use the numbers we already have derived from carrying out the general inductor design procedure* (at V_{INMIN}).

Finally, for the ongoing buck converter design example, the calculation is as follows:

$$P_D = V_D \times I_O \times (1 - D_{MIN}) = 0.5 \times 1 \times (1 - 0.543) = 0.23 \text{ W} \quad \textit{(buck)}$$

Note that the general diode selection procedure is as follows:

The rule-of-thumb is to pick a diode with a current rating at least equal to, but preferably at least *twice* the worst-case average diode current given below (for low losses, since the diode forward drop decreases substantially if its current rating is increased):

- For a ***buck*** — maximum diode current is $I_O \times (1 - D_{MIN})$.
- For a ***boost*** — maximum diode current is I_O.
- For a ***buck-boost*** — maximum diode current is I_O.

Its voltage rating is usually picked to be at least 20% higher (~ "80% derating" — i.e. safety margin) than the worst-case diode voltage given below:

- For a ***buck*** — maximum diode voltage is V_{INMAX}.
- For a ***boost*** — maximum diode voltage is V_O.
- For a ***buck-boost*** — maximum diode voltage is $V_O + V_{INMAX}$.

Worst-case Switch Dissipation

For all topologies the average input current (and therefore switch current) must increase as the input voltage decreases, so as to continue to satisfy the basic power requirement expressed by $P_{IN} = I_{IN} \times V_{IN} = P_O/\eta$ (where η is the efficiency, assumed fixed). Therefore, the switch RMS current is a maximum at V_{INMIN} (i.e. D_{MAX}) *for all topologies*.

For the boost and buck-boost, the general inductor design procedure is at D_{MAX} in any case. So *we can directly use the numbers derived from that*, to find the switch RMS current using the equation below:

$$I_{RMS\_SW} = I_{L\_DMAX} \times \sqrt{D_{MAX} \times \left(1 + \frac{r_{DMAX}{}^2}{12}\right)} \quad \textit{(any topology)}$$

where I_{L_DMAX} and r_{DMAX} are, respectively, the average inductor current and current ripple ratio at D_{MAX} (i.e. at V_{INMIN}). D_{MAX} can be calculated using

$$D_{MAX} = \frac{V_O - V_{INMIN} + V_D}{V_O - V_{SW} + V_D} \quad \textit{(boost)}$$

$$D_{MAX} = \frac{V_O + V_D}{V_{INMIN} + V_O - V_{SW} + V_D} \quad \textit{(buck-boost)}$$

and we should remember that

$$I_{L\_DMAX} = \frac{I_O}{1 - D_{MAX}} \quad \textit{(boost and buck-boost)}$$

For the buck, the general inductor design procedure is at D_{MIN}. So *we cannot directly use the numbers derived* from that to find the switch RMS current (by the previously given equation). We need to calculate r_{DMAX}, but we only know r_{DMIN} so far. Let us proceed with the required steps.

$$r_{DMAX} = \frac{Et_{DMAX}}{L \times I_L} \quad \textit{(any topology)}$$

In other words, if we know the voltseconds at V_{INMIN}, we will know the corresponding current ripple ratio r_{DMAX} for the chosen inductor. But first we have to calculate D_{MAX}:

$$D_{MAX} = \frac{V_O + V_D}{V_{INMIN} - V_{SW} + V_D} = \frac{12 + 0.5}{18 - 1.5 + 0.5} = 0.735 \quad \textit{(buck)}$$

The switch on-time is therefore

$$t_{ON\_DMAX} = \frac{D_{MAX}}{f} \Rightarrow \frac{0.735 \times 10^6}{150{,}000} = 4.9\ \mu s \quad \textit{(any topology)}$$

The voltage across the inductor when the switch is ON is

$$V_{ON\_DMAX} = V_{INMIN} - V_{SW} - V_O = 18 - 1.5 - 12 = 4.5\ V \quad \textit{(buck)}$$

So the voltsμseconds is

$$Et_{DMAX} = V_{ON\_DMAX} \times t_{ON\_DMAX} = 4.5 \times 4.9 = 22\ V\mu s \quad \textit{(any topology)}$$

Therefore

$$r_{DMAX} = \frac{Et_{DMAX}}{L \times I_O} = \frac{22}{137 \times 1} = 0.16 \quad \textit{(buck)}$$

Finally, we are in a position to calculate the switch dissipation

$$I_{RMS\_SW} = I_O \times \sqrt{D_{MAX} \times \left(1 + \frac{{r_{DMAX}}^2}{12}\right)} = 1 \times \sqrt{0.735 \times \left(1 + \frac{0.16^2}{12}\right)} = 0.86 \text{ A} \quad \textit{(buck)}$$

If for example, if the drain-to-source resistance is 0.5 Ω, the dissipation in the mosfet is

$$P_{SW} = {I_{RMS\_SW}}^2 \times R_{DS} = 0.86^2 \times 0.5 = 0.37 \text{ W} \quad \textit{(any topology)}$$

Note that the general switch selection procedure is as follows:

The rule-of-thumb is to pick a switch with a current rating at least equal to, but preferably at least *twice* the worst-case RMS switch current calculated above (for low losses, since the switch forward drop will decrease substantially if its current rating is increased)

Its voltage rating is usually picked to be at least 20% higher (~ "80% derating" — i.e. safety margin) than the worst-case switch voltage given below

- For a ***buck*** — maximum switch voltage is V_{INMAX}.
- For a ***boost*** — maximum switch voltage is V_O.
- For a ***buck-boost*** — maximum switch voltage is $V_O + V_{INMAX}$.

Worst-case Output Capacitor Dissipation

Coincidentally, *the worst-case output capacitor RMS current for all three topologies occurs at the same point at which the general inductor design procedure for each of them is carried out*. In other words, this point is V_{INMAX} for the buck, and V_{INMIN} for the boost and buck-boost. So we should have no trouble, *directly using the numbers derived from the general inductor design procedure*, to find the worst-case RMS current of the output capacitor, using the equations below.

For the buck, we get

$$I_{RMS\_OUT} = I_O \times \frac{r_{DMIN}}{\sqrt{12}} = 1 \times \frac{0.277}{\sqrt{12}} = 0.08 \text{ A} \quad \textit{(buck)}$$

So for example, if the ESR of the output capacitor is 10 Ω, we get the dissipation

$$P_{SW} = I_{RMS\_OUT}{}^2 \times ESR = 0.08^2 \times 10 = 0.064\ W \quad \textit{(any topology)}$$

For the boost and the buck-boost, we need to use

$$I_{RMS\_OUT} = I_O \times \sqrt{\frac{D_{MAX} + \frac{r_{DMAX}{}^2}{12}}{1 - D_{DMAX}}} \quad \textit{(boost and buck-boost)}$$

Note that the general output capacitor selection procedure is as follows:

The rule-of-thumb is to pick an output capacitor with a ripple current rating equal to or greater than the worst-case RMS capacitor current calculated above. Its voltage rating is usually picked to be at least 20 to 50% higher than what it will see in the application (i.e. V_O for all topologies). The output voltage ripple of the converter is also usually a concern. The total peak to peak output voltage ripple produced by the output capacitor is equal to its ESR multiplied by the worst-case peak to peak output current given below (ignoring the ESL of the capacitor)

- For a ***buck*** — peak to peak capacitor current is $I_O \times r_{DMIN}$. This is the same point at which the general inductor design procedure would have been carried out, and so r_{DMIN} is already known.
- For a ***boost*** — peak to peak capacitor current is $I_O \times (1 + r_{DMAX}/2)/(1 - D_{MAX})$. This is the same point at which the general inductor design procedure would have been carried out for this topology, so r_{DMAX} and D_{MAX} are already known.
- For a ***buck-boost*** — peak to peak capacitor current is $I_O \times (1 + r_{DMAX}/2)/(1 - D_{MAX})$. This is the same point at which the general inductor design procedure would have been carried out for this topology, so r_{DMAX} and D_{MAX} are already known.

Worst-case Input Capacitor Dissipation

For the buck-boost, things are much simpler, since the worst-case input capacitor RMS current occurs at D_{MAX}, which is also the point at which we carry out the general inductor design procedure. So all the numbers available from that procedure can be used directly in the equation below

$$I_{RMS\_IN} = I_{L\_DMAX} \times \sqrt{D_{MAX} \times \left(1 - D_{MAX} + \frac{r_{DMAX}{}^2}{12}\right)} \quad \textit{(buck-boost)}$$

For the buck and the boost, the worst-case input RMS capacitor current occurs at D = 0.5. So we have to calculate 'r_{50},' that is, the *current ripple ratio at* D = 50% (or whatever voltage within the specified input range of our application range is closest to this point).

Let us do the numerical calculation for the buck, and the procedure will become clearer.

The input voltage at which D = 50% occurs for the buck is

$$V_{IN\_50} = 2 \times V_O + V_{SW} + V_D = 2 \times 12 + 1.5 + 0.5 = 26 \text{ V} \quad \textit{(buck)}$$

(For the boost use $V_{IN\_50} = \dfrac{V_O + V_{SW} + V_D}{2} \approx \dfrac{V_O}{2}$*).*

We see that our input range does not include this point. But the closest to it is V_{INMAX}. However, coincidentally, this is already the point at which the general inductor design procedure was carried out. So we can use all the numbers derived from that procedure to calculate the input capacitor RMS current, using the equation below

$$I_{RMS\_IN} = I_O \times \sqrt{D \times \left(1 - D + \frac{r^2}{12}\right)} = 1 \times \sqrt{0.543 \times \left(1 - 0.543 + \frac{0.277^2}{12}\right)} \quad \textit{(buck)}$$

(For the boost use $I_{RMS\_IN} = \dfrac{I_O}{1 - D} \times \dfrac{r}{\sqrt{12}}$*).*

So finally

$$I_{RMS\_IN} = 0.502 \text{ A}$$

> **Note:** If for our worked buck example, the input range was not 18–24 V but say 30–45 V, then the general inductor design procedure would clearly be carried out at 45 V. However, the input capacitor current would be a maximum at 30 V. So we can use the above equation for the RMS current, but we would now need to use r_{DMIN} and D_{MAX}. Therefore, knowing only r_{DMAX} so far, we would need to calculate r_{DMIN} by the same procedure presented earlier — that is, by recalculating the voltseconds, and so on.

Note that the general input capacitor selection procedure is as follows:

The rule-of-thumb is to pick an output capacitor with a ripple current rating equal to or greater than the worst-case RMS capacitor current calculated above. Its voltage rating is usually picked to be at least 20 to 50% higher than what it will see in the application (i.e. V_{IN_MAX} for all topologies). The input voltage ripple of the converter is also usually a concern because a small part of it does get transmitted to the output. There can also be EMI considerations involved. In addition, every control IC has a certain (usually unspecified) amount of input noise and ripple rejection, and it may misbehave if the ripple is too much. Typically, the input ripple needs to be kept down to less than ±5% to ±10% of the input

voltage. The total peak to peak input voltage ripple produced by the input capacitor is equal to its ESR multiplied by the worst-case peak to peak input current given below (ignoring the ESL of the capacitor):

- For a ***buck*** — peak to peak capacitor current is $I_O \times (1 + r_{DMIN}/2)$. This is the same point at which the general inductor design procedure would have been carried out, and so r_{DMIN} is already known.
- For a ***buck-boost*** — peak to peak capacitor current is $I_O \times (1 + r_{DMAX}/2)/(1 - D_{MAX})$. This is the same point at which the general inductor design procedure would have been carried out for this topology, so r_{DMAX} and D_{MAX} are already known.
- For a ***boost*** — peak to peak capacitor current at the worst-case point for this parameter (i.e. $D = 0.5$) is equal to $2 \times I_O \times r_{50}$ where

$$r_{50} = \frac{V_{IN\_50}}{4 \times f \times L \times I_O} \qquad \text{and} \qquad V_{IN\_50} = \frac{V_O + V_{SW} + V_D}{2} \approx \frac{V_O}{2}$$

Note that if the input range does not include the $D = 0.5$ point, we need to look for the input voltage end closest to $D = 0.5$. Then we can use the general equation for the peak to peak input capacitor current

$$I_{PK\_PK} = \frac{I_O \times r}{1 - D}$$

where *r* and D correspond to this particular worst-case input voltage end. To find *r* we can use

$$r = \frac{V_O - V_{SW} + V_D}{I_O \times L \times f} \times D \times (1 - D)^2$$

where L is in H, and f is in Hz.

That completes the converter and magnetics design procedure. Next we will move on to off-line converters.

This Page Intentionally Left Blank

CHAPTER

3

Off-line Converter Design and Magnetics

This Page Intentionally Left Blank

CHAPTER 3

Off-line Converter Design and Magnetics

Off-line converters are *derivatives* of standard dc-dc converter topologies. For example, the flyback topology, popular for low-power applications (typically <100 W), is really a buck-boost, with its usual single-winding inductor replaced by an inductor with multiple windings. Similarly, the forward converter, popular for medium to high powers, is a buck-derived topology, with the usual inductor ("choke") supplemented by a transformer. The flyback inductor actually behaves both as an inductor and a transformer. It stores magnetic energy as any inductor would, but it also provides "mains isolation" (mandated for safety reasons), just like any transformer would. In the forward converter, the energy storage function is fulfilled by the choke, whereas its transformer provides the necessary mains isolation.

Because of the similarities between dc-dc converters and off-line converters, most of the spadework for this chapter is in fact contained in the preceding chapter ("DC-DC Converter Design and Magnetics"). The basic magnetic definitions have also been presented therein. Therefore, the reader should read that chapter before attempting this one.

Note that in both the flyback and the forward converters, the transformer, besides providing the necessary mains isolation, also provides another very important function — that of a *fixed-ratio down-conversion step*, determined by the "turns ratio" of the transformer. The turns ratio is the number of turns of the *input* ("primary") winding, divided by the number of turns of the *output* ("secondary") winding. The question arises — why do we even feel the need for a transformer-based step-down conversion stage, when in principle, a switching converter should by itself have been able to up-convert or down-convert at will? The reason will become obvious if we carry out a sample calculation — we will then find that without any additional "help," the converter would require *impractically low values of duty cycle* — to down-convert from such a high input voltage to such a low output. Note that the worst-case ac mains input (somewhere in the world) can be as high as 270 V. So when this ac voltage is rectified by a conventional bridge-rectifier stage, it becomes a dc rail of almost $\sqrt{2} \times 270 = 382$ V, which is fed to the input of the switching converter stage that follows. But the corresponding output voltage can be very low (5 V, 3.3 V, or 1.8 V, and so on), so the required dc transfer ratio (*conversion ratio*) is extremely hard to meet, given the minimum on-time limitations of any typical converter, especially when switching at high frequencies. Therefore, in both the flyback and forward converters, we can intuitively think of the transformer as performing a rather coarse fixed-ratio step-down of the input to a more

amenable (lower) value, from which point onwards the converter does the rest (including the regulation function).

Flyback Converter Magnetics

Polarity of Windings in a Transformer

In *Figure 3-1*, the *turns ratio* is $n = n_P/n_S$, where n_P is the number of turns of the primary winding, and n_S is the number of turns of the secondary winding.

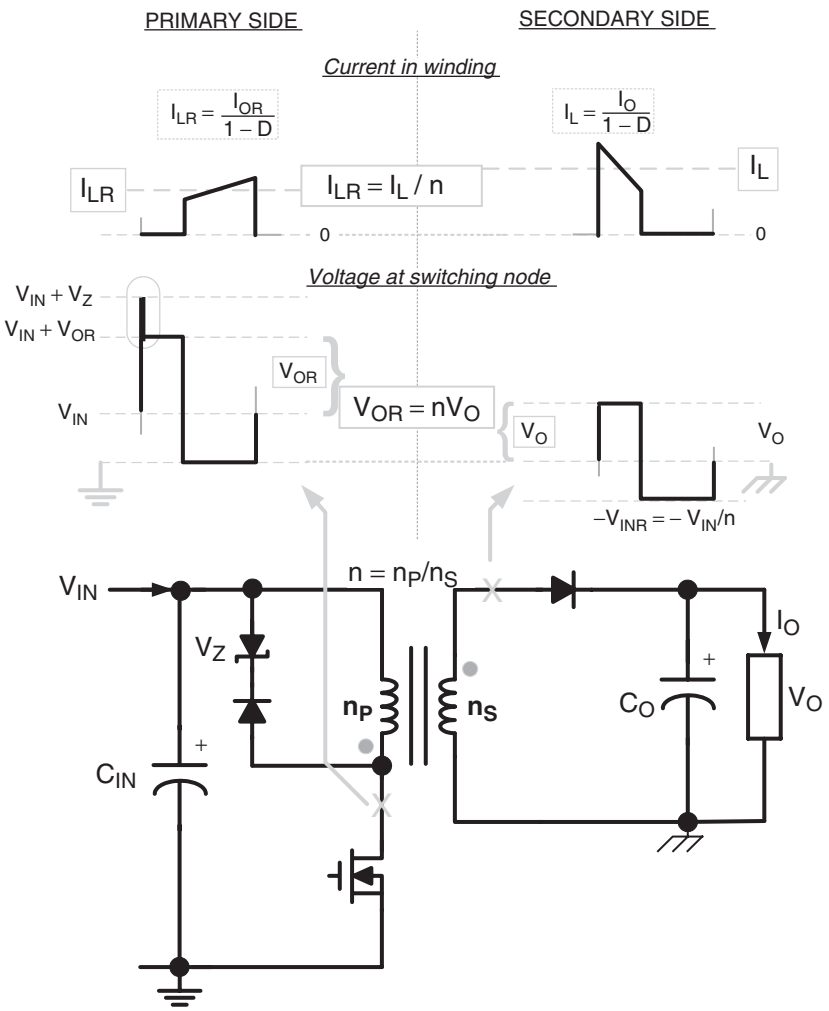


Figure 3-1: Voltage and Currents in a Flyback

We have also placed a *dot* on one end of *each* of the windings. All dotted ends of a transformer are considered to be mutually "equivalent." All non-dotted ends are also obviously *mutually* equivalent. That means that when the voltage on a given dotted end goes "high" (to whatever value), so does the voltage on the dotted ends of all other windings. That happens because all windings share the *same magnetic core*, despite the fact that they are not physically (galvanically) connected to each other. Similarly, all the dotted ends also go "low" at the same time. Clearly, the dots are only an indication of *relative* polarity. Therefore, in any given schematic, we can always swap the dotted and non-dotted ends of the transformer, without changing the schematic in the slightest way.

In the flyback, the polarity of the windings is deliberately arranged such that when the primary winding conducts, the secondary winding is *not* allowed to do so. So when the switch conducts, the dotted end at the drain of the mosfet in *Figure 3-1* goes low. And therefore, the anode of the output diode also goes low, thereby reverse-biasing the diode. We should recall that the basic purpose of a buck-boost (which this in fact also is) is to allow incoming energy from the source during the switch on-time to build up in the inductor (*only*), and then later, during the off-time, to "collect" all this energy (and no more) at the output. Note that this is the unique property that distinguishes the buck-boost (and the flyback) from the buck and the boost. For example, in a buck, energy from the input source gets delivered to the inductor *and* the output (during the on-time). Whereas, in a boost, stored energy from the inductor *and* the input source gets delivered to the output (during the off-time). Only in a buck-boost do we have *complete separation* between the energy storage and the collection process, during the on-time and the off-time. So, now we understand why the flyback is considered to be just a buck-boost derivative.

We know that every dc-dc topology has a so-called "switching node." This node represents the point of *diversion* of the inductor current — from its *main* path (i.e. in which the inductor *receives* energy from the input) to its *freewheelin*g path (i.e. in which the inductor *provides* stored energy to the output). So clearly, the switching node is necessarily the node *common* to the switch, the inductor, and the diode. Further, we will find that the voltage at this node is always "swinging" — because that is what is required to get the diode to alternately forward and reverse-bias, as the switch toggles. But looking at *Figure 3-1*, we see that with a transformer replacing the traditional dc-dc inductor, there are now, in effect, *two* "switching nodes" — one on each side of the transformer, as indicated by the "X" markings in *Figure 3-1* — one "X" is at the drain of the mosfet, and the other "X" is at the anode of the output diode. These two nodes are clearly "equivalent" because of the dots, as explained above. And since at both these nodes, the voltage is swinging, therefore both are considered to be "switching nodes" (of the transformer-based topology). Note that if we had, say, three windings (e.g. an additional output winding), we would have had three (equivalent) switching nodes.

Transformer Action in a Flyback, and Its Duty Cycle

Classic "transformer action" implies that the voltages *across* the windings of the transformer, and the currents *through* each of them, *scale* according to the turns ratio, as described in *Figure 3-1*. But it is perhaps not immediately apparent why the flyback inductor exhibits transformer action.

When the switch turns ON, a voltage V_{IN} (the rectified ac input) gets impressed across the primary winding of the transformer. And at the same time, a voltage equal to $V_{INR} = V_{IN}/n$ ("R" for reflected) gets impressed across the secondary winding (in a direction that causes the output diode to get reverse-biased). Therefore, there is no current in the secondary winding when the primary winding is conducting.

Let us calculate what V_{INR} is. This voltage translation across the isolation boundary follows from the induced voltage equation applied to each winding

$$V_P = -n_P \frac{d\phi}{dt} \quad \text{and} \quad V_S = -n_S \frac{d\phi}{dt}$$

Note that both windings enclose the same magnetic core, so the flux ϕ is the same for both, and so is the rate of change of flux $d\phi/dt$ for each winding. Therefore

$$V_S = -n_S \times \left(\frac{V_P}{-n_P} \right)$$

or

$$V_S = n_S \times \left(\frac{V_{IN}}{n_P} \right) = \frac{V_{IN}}{n} \equiv V_{INR}$$

Also,

$$\frac{V_P}{n_P} = \frac{V_S}{n_S}$$
$$\frac{V_P}{V_S} = n$$

This above equation represents classic "transformer action" with respect to the voltages involved. But we also learn from the preceding equation that the *Volts/turn* for *any winding* (at *any given instant*) is the same for all the windings present on a given magnetic core — and this is what eventually leads to the observed *voltage scaling*.

Note also that voltage scaling in any transformer occurs *irrespective of whether a given winding is passing current or not*. That is because, whether a given winding is contributing to

the net flux ϕ present in the core or not, each winding encloses this entire flux, and so the basic equation $V = -N \times d\phi/dt$ applies to all windings, and so does voltage scaling.

We know that energy is built up in the transformer during the on-time. When the switch turns OFF, this stored energy (and its associated current) needs to flyback/freewheel. We also know that the voltages will automatically try to adjust themselves in any possible way, so as to make that happen. So we can safely assume the diode will somehow conduct during the switch off-time. Now, assuming we have reached a "steady state," the voltage on the output capacitor has stabilized at some fixed value V_O. Therefore, the voltage at the secondary-side switching node gets clamped at V_O (ignoring the diode drop). Further, since one end of the secondary winding is tied to ground, the voltage *across* this winding is now equal to V_O. By transformer action, this reflects a voltage *across* the primary winding, equal to $V_{OR} = V_O \times n$. But the switch is OFF during this time. Therefore, under normal circumstances, the voltage at the primary-side switching node would have settled at V_{IN}. However, now this reflected output voltage V_{OR}, coming through the transformer, adds to that. Therefore, the voltage at the primary-side switching node eventually goes up to $V_{IN} + V_{OR}$ (for now, we are ignoring the turn-off spike encircled in *Figure 3-1*).

> **Note:** During the on-time, the primary side is the one determining the voltages across *all* the windings. And during the off-time, it is the secondary winding that gets to "call the shots"!

We can calculate duty cycle from the most basic equation (from voltseconds law)

$$D = \frac{V_{OFF}}{V_{OFF} + V_{IN}}$$

We have the option of performing this calculation, either on the primary winding, or the secondary winding. Either way, we get the same result, as shown in *Table 3-1*.

Table 3-1: Derivation of dc transfer function of flyback

	Primary Winding	*Secondary Winding*
V_{ON}	V_{IN}	$V_{INR} \equiv V_{IN}/n$
V_{OFF}	$V_{OR} \equiv V_O \times n$	V_O
DC Transfer Function	$D = \frac{V_{OFF}}{V_{ON}+V_{OFF}}$	
	$D = \frac{V_{OR}}{V_{IN}+V_{OR}}$	$D = \frac{V_O}{V_{INR}+V_O}$
	$D = \frac{nV_O}{V_{IN}+nV_O}$	

We should be always very clear that transformer action applies only to the voltages *across* windings. And "voltage *across*" is not necessarily "voltage *at*"! To measure the voltage *at* a given point, we have to consider what the reference level (i.e. "ground" by definition) is, *with respect to which* its voltage needs to be measured, or stated. In fact, the reference level (i.e. by definition, "ground") is called the "primary ground" on the primary side and the "secondary ground" on the secondary side. Note that these are indicated by different ground symbols in *Figure 3-1*.

To find out the (absolute) voltage *at* the swinging end of any winding, we can use the following level-shifting rule:

To get the absolute value of the voltage at the swinging end of any winding, we must add to the voltage across the winding, the dc voltage present at its "non-swinging" end.

So, for example, to get the voltage at the drain of the mosfet (swinging end of primary winding), we need to add V_{IN} (voltage at other end of winding), to the voltage waveform that represents the voltage across the primary winding. That is how we got the voltage waveforms shown in *Figure 3-1*.

Coming to the question of how *currents* actually reflect from one side of the transformer to the other, it must be pointed out that *even though the final current scaling equations of a flyback transformer are exactly the same as in the case of an actual transformer*, this is not strictly "classic transformer action." The difference from a conventional transformer is, that in the flyback, the primary and secondary windings do *not* conduct at the same time. So in fact, it is a mystery why their currents are related to each other at all!

The current scaling that occurs in a flyback actually follows from *energy considerations*. The energy in a core is in general written as

$$E = \frac{1}{2}LI^2$$

We know the windings of our flyback conduct at different times, but the energy associated with each of the current flows must be equal to the energy in the core, and must therefore be equal to each other (we are ignoring the ramp portion of the current here for simplicity). Therefore,

$$E = \frac{1}{2}L_P{I_P}^2 = \frac{1}{2}L_S{I_S}^2$$

where L_P is the inductance measured across the primary winding — with the secondary winding floating (no current), and L_S the inductance measured across the secondary winding — with the primary winding floating. But we also know that

$$L = N^2 \times A_L \times 10^{-9} \text{ Henry}$$

where A_L is the *inductance index*, defined previously. Therefore in our case we get

$$L_P = n_P{}^2 \times A_L \times 10^{-9}$$
$$L_S = n_S{}^2 \times A_L \times 10^{-9}$$

Substituting in the energy equation, we get the well-known current scaling equations

$$n_P I_P = n_S I_S$$

or

$$\frac{I_P}{I_S} = \frac{1}{n}$$

We see that analogous to the Volts/turns rule, the *ampere-turns* also need to be preserved at all times. In fact, the core itself doesn't really "care" which particular winding is passing current at any given moment, so long as there is no sudden change in the *net* ampere-turns of the transformer. This becomes the "transformer-version" of the basic rule we learned in *Chapter 1* — that the current through an inductor cannot change discontinuously. Now we see that *the ampere-turns of a transformer cannot change discontinuously*.

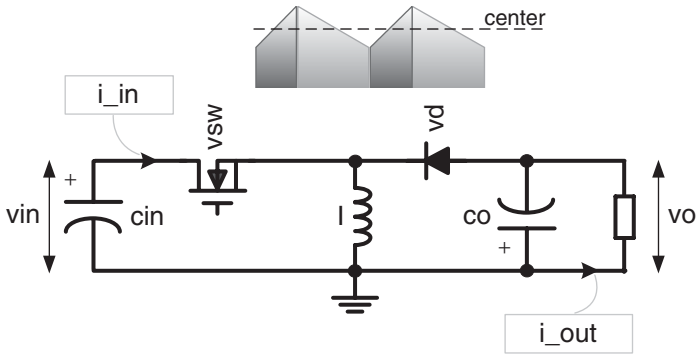
Summarizing, transformer action works as follows — *when reflecting a voltage from primary to secondary side, we need to divide by the turns ratio. When going from the secondary to the primary side, we need to multiply by the turns ratio. The rule reverses for currents — so we multiply by the turns ratio when going from primary to secondary, and divide in the opposite direction.*

The Equivalent Buck-boost Models

Because of the many similarities, and also because of the way voltages scale in the transformer, it becomes very convenient (most of the time) to study the flyback as an equivalent dc-dc (inductor-based) buck-boost. In other words, we separate out the coarse fixed-ratio step-down ratio and incorporate it into equivalent (reflected) voltages and currents. We thereby manage to reduce the flyback transformer into a simple energy-storage medium, just like any conventional dc-dc buck-boost inductor. In other words, for most practical purposes, the transformer goes "out of the picture." The advantage is that almost all the equations and design procedure we can write for a conventional buck-boost now apply to this *equivalent* buck-boost model. One exception to this is the leakage inductance issue (and everything related to it — the clamp, the loss in efficiency due to it, the turn-off voltage spike on the switch, and so on). We will discuss this exception later. But other than that,

all other parameters — like the capacitor, diode, and switch currents for example, can be more readily visualized and calculated if we use this dc-dc model approach.

The *equivalent dc-dc model* is created essentially by reflecting the voltages and currents across the isolation boundary of the transformer to one side. But again, as in the case of the duty cycle calculation (see *Table 3-1*), we have two options here — we can either reflect everything to the primary side, *or* everything to the secondary side. We thus get the *two* equivalent buck-boost models as shown in *Figure 3-2*. We can use the primary-side equivalent model to calculate all the voltages and currents on the primary side of the original flyback and the secondary-side equivalent model for calculating all the currents and voltages on the secondary side of the original flyback.



	Primary-side equivalent model	Secondary-side equivalent model
vin	V_{IN}	$V_{INR}=V_{IN}/n$
i_in	I_{IN}	$I_{INR}=I_{IN}xn$
cin	C_{IN}	n^2xC_{IN}
l	L_p	$L_S=L_P/n^2$
vsw	V_{SW}	V_{SW}/n
vo	$V_{OR}=V_Oxn$	V_O
i_out	$I_{OR}=I_O/n$	I_O
center	$I_{OR}/(1–D)=I_O/[nx(1–D)]$	$I_O/(1–D)$
co	C_O/n^2	C_O
vd	V_Dxn	V_D
Duty Cycle	D	D
Current Ripple Ratio	r	r

Figure 3-2: The Equivalent Buck-Boost Models of the Flyback

We know that voltages and currents reflect across the boundary by getting either multiplied or divided by the turns ratio. In fact, the 'reflected output voltage,' V_{OR}, is one of the most important parameters of a flyback, as we will see. As the name indicates, ***V_{OR} is effectively the output voltage as seen by the primary side***. In fact, if we compare the switch waveform of the flyback in *Figure 3-1* with that of a buck-boost, we will realize that to the switch, it seems as if the *output voltage is really* V_{OR}.

As an example, suppose we have a 50 W converter with an output of 5 V at 10 A, and a turns ratio of 20. The V_{OR} is therefore $5 \times 20 = 100$ V. Now, if we change the set output to say 10 V and reduce the turns ratio to 10, the V_{OR} is still 100 V. We will find that none of the primary-side *voltage* waveforms change in the process (assuming efficiency doesn't change). Further, if we have also kept the output *power* constant in the process, that is, by changing the load to 5 A for an output at 10 V, all the *currents* on the primary side will also be unaffected. Therefore, *the switch will never know the difference*. In other words, the switch virtually "thinks" that it is a simple dc-dc buck-boost — delivering an output voltage of V_{OR} at a load current of I_{OR}.

As mentioned, the only difference between a transformer-based flyback that "thinks" it is providing an output of V_{OR} at the rate of I_{OR}, and an inductor-based version that *really* is providing an output of V_{OR} at the rate of I_{OR}, is the 'leakage inductance' of the flyback transformer. This is that part of the primary side inductance that is *not coupled* to the secondary side and therefore cannot partake in the transfer of useful energy from the input to the output. We can confirm from *Figure 3-1*, that the only portion of the primary-side (switch) voltage waveform that "doesn't make it" to the secondary side is the *spike* occurring just after the turn-off transition. This spike comes from the uncoupled leakage inductance, as we will soon see.

Note that in the equivalent buck-boost models, the reactive component values also get reflected — though as the *square* of the turns ratio. We can understand this fact easily from *energy considerations*. For example, the output capacitor C_O in the original flyback was charged up to a value of V_O. So its stored energy was $1/2\, C_O V_O^2$. In the primary-side buck-boost model, the output of the converter is V_{OR}, that is, $V_O \times n$. Therefore, to keep the energy stored in this capacitor invariant (in the dc-dc model, as in the flyback), the output capacitance must get reflected to the primary side according to C_O/n^2. Note also from *Figure 3-2* how the inductance reflects. This is consistent with the fact that $L \propto N^2$.

The Current Ripple Ratio for the Flyback

Looking at the equivalent buck-boost models in *Figure 3-2*, the *center of the ramp* on the secondary side (average inductor current, "I_L") must be equal to $I_O/(1 - D)$, as for a buck-boost (because the average diode current must equal the load current). This secondary-side "inductor" current gets reflected to the primary side, and so the center of the primary-side inductor current ramp is "I_{LR}," where $I_{LR} = I_L/n$. Equivalently, it is

equal to $I_{OR}/(1 - D)$, where I_{OR} is the *reflected load current*, that is, $I_{OR} = I_O/n$. Similarly, the current swings on the primary and secondary sides are also related, via scaling (turns ratio n). Therefore, we see that the *ratio* of the swing to the center of the ramp is *identical* on both sides (primary- and secondary-side dc-dc models). We are thus in a position to define a *current ripple ratio r* for the flyback topology too — just as we did for a dc-dc converter. We just need to visualize *r* in a slightly different manner this time ***— in terms of the center of the ramp (switch or diode), rather than the dc inductor level*** (because there is no inductor present really). And as for dc-dc converters, we should normally try to set it to around 0.4.

The value of r for a flyback is the same for either the primary and secondary dc-dc equivalent models.

The Leakage Inductance

The leakage inductance can be thought of as a *parasitic inductance* in series with the primary-side inductance of the transformer. So just at the moment the switch turns OFF, the current flowing through both these inductances is "I_{PKP}," that is, the peak current on the primary side. However, when the switch turns OFF, the energy in the primary inductance has an available freewheeling path (through the output diode), but the leakage inductance energy has nowhere to go. So it expectedly "complains" in the form of a huge voltage spike (see *Figure 3-1*). This spike (or a scaled version of it) is *not* seen on the secondary side, simply because this is not a coupled inductance, like the primary inductance.

If we don't make any effort to collect this energy, the induced spike can be very large, causing switch destruction. Since we certainly can't get this energy to transfer to the secondary side, we have just two options — either we can try to *recover* it and cycle it back into the input capacitor, or we can simply burn it (dissipation). The latter approach is usually preferred for the sake of simplicity. It is commonly accomplished by means of a straightforward 'zener diode clamp,' as shown in *Figure 3-1*. Of course the zener voltage must be chosen according to the maximum voltage the switch can tolerate. Note that for several reasons, in particular that of efficiency, it is usually considered preferable to connect this zener *across* the primary winding (via a blocking diode in series with it). The alternative is to connect it from the switching node to primary ground.

We can ask — where does the leakage inductance really reside? *Most* of it is inside the primary winding of the transformer, though some of it lies in the PCB trace sections and transformer terminations, especially with those associated with the *secondary* winding, as we will see below.

Zener Clamp Dissipation

If we burn the energy in the leakage, it is important to know how this affects the efficiency. It is sometimes intuitively felt the energy dissipated every cycle is just $1/2 \times L_{LKP}I_{PK}{}^2$, where

I_{PK} is the peak switch current, and L_{LKP} is the primary-side leakage. That certainly is the energy residing *in the leakage inductance* (at the moment the switch turns OFF), but it is *not* the *entire* energy that eventually gets dissipated in the zener clamp on account of the leakage.

The primary winding is in series with the leakage, so during the small interval that the leakage inductance is trying, in effect, to *reset*, by freewheeling into the zener, the primary winding is *forced to follow suit* and continue to provide this series current. Though the primary winding is certainly trying (and managing partly) to freewheel into the secondary side, a part of its energy also gets diverted into the zener clamp — until the leakage inductance achieves full reset (zero clamp current). In other words, some energy from the primary inductance gets literally "snatched away" by the series leakage inductance, and this also finds its way into the zener, along with the energy residing in the leakage itself. A detailed calculation reveals that the zener dissipation actually is

$$P_Z = \frac{1}{2} \times L_{LK} \times {I_{PK}}^2 \times \frac{V_Z}{V_Z - V_{OR}} \text{ W}$$

So the energy in the leakage $1/2 \times L_{LK} \times {I_{PK}}^2$ gets *multiplied* by the term $V_Z/(V_Z - V_{OR})$ (this additional term from the primary inductance).

Note that if the zener voltage is too close to the chosen V_{OR}, the dissipation in the clamp goes up steeply. ***V_{OR} therefore always needs to be picked with great care.*** *That simply means that* ***the turns ratio has to be chosen carefully!***

Secondary-side Leakages also Affect the Primary Side

Why did we use the symbol "L_{LK}" in the dissipation equation above? Why didn't we identify it as the *primary-side* leakage ("L_{LKP}")? The reason is that L_{LK} represents the *overall* leakage inductance as seen by the switch. So, it is partly L_{LKP} — but it also is influenced by the *secondary-side* leakage inductance. This is a little hard to visualize, since by definition, the secondary-side leakage inductance is not supposed to be coupled to the primary side (and vice versa). So how could it be affecting anything on the primary side?

The reason is that just as the primary-side leakage *prevents* the primary-side current from freewheeling into the output immediately (and thereby causes an increase in the zener dissipation), any secondary-side inductance also prevents the freewheeling path from *becoming available* immediately (following switch turn-off). Basically, the secondary-side inductance *insists* that we ("politely" and) slowly build up the current through it — respecting the fact that it is an inductance after all! However, until the current in the bona fide freewheeling path can build up to the required level, the primary-side current still needs to freewheel *somewhere* (because the switch is turning OFF)! The path the inductor current therefore seeks out is the one containing the zener clamp (being the only path available). The zener can therefore see significant dissipation, even assuming *zero* primary-side leakage.

In brief, the secondary-side leakage has created much the same effect as a primary-side leakage.

When both primary- and secondary-side leakages are present, we can find the effective primary-side leakage (as seen by the switch and zener clamp) as

$$L_{LK} = L_{LKP} + n^2 L_{LKS}$$

So, like any other reactive element, the secondary-side leakage also reflects onto the primary side according to the *square of the turns ratio*, where it adds up in series with any primary-side leakage present.

For a given V_{OR}, if the output voltage is "low" (for example 5 V or 3.3 V), the turns ratio is much greater. Therefore, *if the chosen V_{OR} is very high, the reflected secondary-side leakage can become even greater than any primary-side leakage*. This can become quite devastating from the efficiency standpoint.

Measuring the Effective Primary-side Leakage Inductance

The best way to know what L_{LK} really is, is by measuring it! Commonly, a leakage inductance measurement is done by shorting the secondary winding pins and then measuring the inductance across the ends of the (open) primary winding. By shorting, we virtually cancel out all coupled inductance. And so what we measure is just the primary-side leakage inductance in this case.

However, the best method to measure leakage is actually an *in-circuit* measurement — so that we include the secondary-side PCB traces in the measurement. The recommended procedure is as follows.

On the given application board, a thick piece of copper foil (or a thick section of braided copper strands), with as short a length as possible, is placed directly across the diode solder pads on the PCB. A similar piece of conductor is placed across the output capacitor solder pads. Then, if we measure the inductance across the (open) primary winding pins, we will measure the effective leakage inductance L_{LK} (not just L_{LKP}).

We will find that the contribution from the secondary-sides traces can in fact make L_{LK} several times larger than L_{LKP}. L_{LKP} can of course be measured, if desired, by placing a thick conductor across the secondary *pins* of the transformer.

The PCB used in the above procedure can be just a bare board with no components mounted on it, other than the transformer. Or it can even be a fully assembled board (though sometimes, we may need to cut the trace connecting the drain of the mosfet to the transformer).

If we want to *mathematically* estimate the inductance of the secondary-side traces, the rule-of-thumb we can use is **20 nH per inch**. But here, we need to include the *full* electrical path of the high-frequency output current — starting from one end of the secondary winding, returning to its other end, *through the diode and output capacitor*(s). We will be surprised to calculate or measure, that even an inch or two of trace length can dramatically decrease the efficiency by 5 to 10% in low output voltage applications.

Worked Example (7) — Designing the Flyback Transformer

A 74 W universal input (90 VAC to 270 VAC) flyback is to be designed for an output of 5 V @ 10 A and 12 V @ 2 A. Design a suitable transformer for it, assuming a switching frequency of 150 kHz. Also, try to use a cost-effective 600 V-rated mosfet.

Fixing the V_{OR} and V_Z

At maximum input voltage, the rectified dc to the converter is

$$V_{INMAX} = \sqrt{2} \times VAC_{MAX} = \sqrt{2} \times 270 = 382 \text{ V}$$

With a 600 V mosfet, we must leave at least 30 V safety margin when at V_{INMAX}. So in our case, we do not want to exceed 570 V on the drain. But from *Figure 3-1*, the voltage on the drain is $V_{IN} + V_Z$. Therefore

$$V_{IN} + V_Z = 382 + V_Z \leq 570$$
$$V_Z \leq 570 - 382 = 188 \text{ V}$$

We pick a standard 180 V zener.

Note that if we plot the zener dissipation equation presented earlier, as a function of V_Z/V_{OR}, we will discover that in all cases, ***we get a "knee" in the dissipation curve at around $V_Z / V_{OR} = 1.4$.*** So here too, we pick this value as an optimum ratio that we would like to target. Therefore

$$V_{OR} = \frac{V_Z}{1.4} = 0.7 \times V_Z = 0.71 \times 180 = 128 \text{ V}$$

Turns Ratio

Assuming the 5 V output diode has a forward drop of 0.6 V, the turns ratio is

$$n = \frac{V_{OR}}{V_O + V_D} = \frac{128}{5.6} = 22.86$$

Note that the 12 V output may sometimes be regulated by a post-linear-regulator. In that case, we may have to make the transformer provide an output 3 to 5 V higher (than the final expected 12 V) — to provide the necessary "headroom" for the linear regulator to operate properly. This additional headroom not only caters to the dropout limits of the linear regulator, but in general also helps achieve a regulated 12 V *under all load conditions*. However, there are also some clever *cross-regulation* techniques available that allow us to omit the 12 V linear regulator, particularly if the regulation requirements of the 12 V rail are not too "tight," and also if there is some minimum load assured on the outputs. In our example, we are assuming there is no 12 V post regulator present. Therefore, the required turns ratio for the 12 V output is $128/(12 + 1) = 9.85$, where we have assumed the diode drop is 1 V in this case.

Maximum Duty Cycle (Theoretical)

Having verified the selection of V_Z and V_{OR} at highest input, now we need to *get back to the lowest input voltage*, because we know from the previous discussions about the buck-boost (see the "general inductor design procedure" in the previous chapter) that V_{INMIN} is the worst-case point we need to consider for a buck-boost inductor/transformer design.

The minimum rectified dc voltage to the converter is

$$V_{INMIN} = \sqrt{2} \times VAC_{MIN} = \sqrt{2} \times 90 = 127 \text{ V}$$

We are ignoring the voltage ripple on the input terminals of the converter, and therefore we will take this as the dc input to the converter stage. So the duty cycle at minimum input voltage is

$$D = \frac{V_{OR}}{V_{OR} + V_{INMIN}} = \frac{128}{128 + 127} = 0.5 \quad \textit{(flyback)}$$

This is clearly a "theoretical" estimate — implying 100% efficiency. We will in fact ignore this value ultimately, as we ***will be estimating D more accurately by another trick***.

Note however, that this is the *operating* D_{MAX}.When we "*power down*" our converter for example*, the duty cycle will actually increase further in an effort to maintain regulation* (unless current limit and/or duty cycle limit is encountered along the way). Then depending upon the number of *missing ac cycles* for which we may need to ensure regulation (the 'holdup time' specification), we will need to select a suitable input capacitance and also the maximum duty cycle limit, D_{LIM}, of our controller. Typically, D_{LIM} is set around 70%, and the capacitance is selected on the basis of the ***3 µF/W rule-of-thumb***. For example, for our 74 W supply with an estimated 70% efficiency at low line, we will draw an input power of $74/0.7 = 106$ W. Therefore we should use a $106 \times 3 = 318$ µF (standard value 330 µF)

input capacitor. However, note that the ripple current rating of this capacitor (and its life expectancy) must be verified.

Effective Load Current on Primary and Secondary Sides

Let us *lump* all the 74 W output power into an equivalent *single* output of 5 V. So the load current for a 5 V output is

$$I_O = \frac{74}{5} \approx 15 \text{ A}$$

On the primary side, the switch "thinks" its output is V_{OR} and the load current is I_{OR}, where

$$I_{OR} = \frac{I_O}{n} = \frac{15}{22.86} = 0.656 \text{ A}$$

Duty Cycle

The actual duty cycle is important because a slight increase in it (from the theoretical 100% efficiency value) may lead to a significant increase in the operating peak current and the corresponding magnetic fields.

The input power is

$$P_{IN} = \frac{P_O}{\text{Efficiency}} = \frac{74}{0.7} = 105.7 \text{ W}$$

The average input current is therefore

$$I_{IN} = \frac{P_{IN}}{V_{IN}} = \frac{105.7}{127} = 0.832 \text{ A}$$

The average input current tells us what the actual duty cycle 'D' is, because I_{IN}/D is also the center of the primary-side current ramp, and must equal I_{LR}, that is,

$$\frac{I_{IN}}{D} = \frac{I_{OR}}{1 - D}$$

solving,

$$D = \frac{I_{IN}}{I_{IN} + I_{OR}} = \frac{0.832}{0.832 + 0.656} = 0.559$$

We thus have a more accurate estimate of duty cycle.

Actual Center of Primary and Secondary Current Ramps

The center of the secondary-side current ramp (lumped)

$$I_L = \frac{I_O}{1 - D} = \frac{15}{1 - 0.559} = 34.01 \text{ A}$$

The center of the primary-side current ramp is

$$I_{LR} = \frac{I_L}{n} = \frac{34.01}{22.86} = 1.488 \text{ A}$$

Peak Switch Current

Knowing I_{LR}, we know the peak current for our selected current ripple ratio

$$I_{PK} = \left(1 + \frac{r}{2}\right) \times I_{LR} = 1.25 \times 1.488 = 1.86 \text{ A}$$

We may need to set the current limit of the controller, for example, based on this estimate.

Voltseconds

We have at V_{INMIN}

$$V_{ON} = V_{IN} = 127 \text{ V}$$

The on-time is

$$t_{ON} = \frac{D}{f} = \frac{0.559}{150 \times 10^3} \Rightarrow 3.727 \text{ } \mu\text{s}$$

So the voltsμseconds is

$$Et = V_{ON} \times t_{ON} = 127 \times 3.727 = 473 \text{ V}\mu\text{s}$$

Primary-side Inductance

Note that when we come to designing off-line transformers, for various reasons like reducing high-frequency copper loss, reducing size of transformer, and so on, it is more common to set *r at around 0.5*. So the primary-side inductance must then be (from the "L × I" rule)

$$L_P = \frac{1}{I_{LR}} \times \frac{Et}{r} = \frac{473}{1.488 \times 0.5} = 636 \text{ } \mu\text{H}$$

Selecting the Core

Unlike made-to-order or off-the-shelf inductors, when designing our own magnetic components, we should not forget that adding an air gap dramatically improves the energy storage capability of a core. Without the air gap, the core could saturate even with very little stored energy.

Of course, we still need to maintain the desired L, corresponding to the desired *r*! So if we add too much of a gap, we will also need to add many more turns — thus increasing the copper loss in the windings. *At one point, we will also run out of window space to accommodate these windings.* So a practical compromise must be made here, one that the following equation actually takes into account (applicable to *ferrites* in general, for any topology):

$$V_e = 0.7 \times \frac{(2 + r)^2}{r} \times \frac{P_{IN}}{f} \text{ cm}^3$$

where f is in kHz.

In our case we get

$$V_e = 0.7 \times \frac{(2.5)^2}{0.5} \times \frac{105.7}{150} = 6.17 \text{ cm}^3$$

We start looking for a core of this volume (or higher). We find a candidate in the *EI-30*. Its effective length and area are given in its datasheet as

$$A_e = 1.11 \text{ cm}^2$$

$$l_e = 5.8 \text{ cm}$$

So its volume is

$$V_e = A_e \times l_e = 5.8 \times 1.11 = 6.438 \text{ cm}^3$$

which is a little larger than we need, but close enough.

Number of Turns

The voltage dependent equation

$$B = \frac{LI}{NA} \text{ tesla}$$

connects B to L. However we also know that a statement about *r* is equivalent to a statement about *L* — for a given frequency (the "L × I equation"). So combining these equations, and

also connecting the swing in the B-field to its peak (through *r*), we get a very useful form of the voltage dependent equation, in terms of *r* (expressed in MKS units):

$$N = \left(1 + \frac{2}{r}\right) \times \frac{V_{ON} \times D}{2 \times B_{PK} \times A_e \times f} \quad \textit{(voltage dependent equation, any topology)}$$

So even with *no* information about the permeability of the material, air gap, and so on, *we already know the number of turns required on a core with area* A_e *that will produce a certain B-field*. We also know that with or without an air gap, the B-field should not exceed 0.3 T for most ferrites. So solving the equation for N (N is n_P here, number of primary turns),

$$n_P = \left(1 + \frac{2}{0.5}\right) \times \frac{127 \times 0.559}{2 \times 0.3 \times 1.11 \times 10^{-4} \times 150 \times 10^3} = 35.5 \text{ Turns}$$

We will have to verify that this can be accommodated in the window of the core — along with the bobbin, tape insulation, margin tape, secondary windings, sleeving, and so on. Usually, that is no problem for a flyback.

Note that if we want to reduce N, the only possible ways are — to allow for a larger r, or decrease the duty cycle (i.e. pick a lower V_{OR}***), or allow for a higher B (new material!?), or increase the area of the core*** — the latter, hopefully, without increasing the volume, because that would amount to over-design. But certainly, just playing with the permeability and air gap is *not* going to help!

The number of secondary turns (5 V output) is

$$n_S = \frac{n_P}{n} = \frac{35.5}{22.86} = 1.55 \text{ Turns}$$

But we want an *integral* number of turns. Further, approximating this to one full turn is not a good idea since there will be more leakage. We therefore prefer to set

$$n_S = 2 \text{ Turns}$$

So, with the same turns ratio (i.e. V_{OR} unchanged)

$$n_P = n_S \times n = 2 \times 22.86 \approx 46 \text{ Turns}$$

The number of turns for the 12 V output is obtained by the scaling rule

$$n_{S\_AUX} = \frac{12 + 1}{5 + 0.6} \times 2 = 4.64 \approx 5 \text{ Turns}$$

where we have assumed the 5 V diode has a drop of 0.6 V and the 12 V diode has a drop of 1 V.

Actual B-field

So now we can use the voltage dependent equation again, to solve for B

$$B_{PK} = \left(1 + \frac{2}{r}\right) \times \frac{V_{ON} \times D}{2 \times n_P \times A_e \times f} \text{ teslas}$$

But in fact we don't have to use this equation anymore! We realize that B_{PK} is inversely proportional to the number of turns. So if, with a calculated 35.5 turns, we had a peak field of 0.3 T, then with 46 turns we will have (***keeping L and r unchanged!***)

$$B_{PK} = \frac{35.5}{46} \times 0.3 = 0.2315 \text{ T}$$

The swing is related to the peak by

$$\Delta B \equiv 2 \times B_{AC} = \frac{2r}{r+2} \times B_{PK} = \frac{1}{2.5} \times 0.2315 = 0.0926 \text{ T}$$

Note that in CGS units, the peak is now 2315 gauss, and the ac component is half the swing, that is, 463 gauss (since $r = 0.5$).

> **Note:** If we start with a B-field target of 0.3 T, we are likely to reach a ***lesser*** B-field after rounding up the secondary turns to the nearest higher integer, as we did above. That of course is not only expected, but acceptable. However note that on power-up or power-down, for example, the B-field will increase further, as the converter tries to continue regulating. That is why we need to set the maximum duty cycle limit and/or current limit accurately, or the switch can be destroyed due to inductor/transformer saturation. Cost-effective flyback designs with fast-acting current limit and fast switches (especially those with an *integrated mosfet*) generally ***allow for a peak B-field of up to 0.42 T, so long as the operating field is 0.3 T or less***.

Air Gap

Finally, we need to consider the *permeability* of the material! L is related to permeability by the equation

$$L = \frac{1}{z} \times \left(\frac{\mu\mu_0 A_e}{l_e}\right) \times N^2 \text{ H}$$

Here z is the 'gap factor'

$$z = \frac{l_e + \mu l_g}{l_e}$$

Note that z can range from 1 (no gap) to virtually any value. A z of 10, for example, increases the energy-handling capability of an ungapped core set by a factor of 10 (its A_L value falls by the same factor, and so does its *effective permeability* — $\mu e = \mu\mu_0/z$). So large gaps certainly help, but since we are still interested in maintaining L to a certain value

based on our choice of *r*, we will have to increase the number of turns substantially. As mentioned, at some point, we just may not be able to accommodate these windings in the available window, and further, the copper loss will also increase greatly. So ***z in the range of 10 to 20 is a good compromise for gapped transformers made out of ferrite material.*** Let us see what it comes out to be, based on our requirements

$$z = \frac{1}{L} \times \left(\frac{\mu \mu_0 A_e}{l_e} \right) \times N^2 = \frac{1}{636 \times 10^{-6}} \times \left(\frac{2000 \times 4\pi \times 10^{-7} \times 1.11 \times 10^{-4}}{5.8 \times 10^{-2}} \right) \times 46^2$$

So,

$$z = 16$$

Finally, solving for the length of air gap,

$$z = 16 = 2 = \frac{5.8 + (2000)\, l_g}{5.8} \Rightarrow l_g = 0.435 \text{ mm}$$

Note: In general, if we use a center-gapped transformer, the *total* gap in the center must be equal to the above calculated value, whether *each* center limb has been ground or not. But if spacers are being inserted on both side limbs (say on an EE or EI type of core), the thickness of the spacer on each outer limb must be *half* of the above-calculated value, because the total air gap is then as desired.

Selecting the Wire Gauge and Foil Thickness

In an inductor, the current undulates relatively smoothly. However, in a transformer, the current in one winding stops completely, to let the other winding take over. Yes, the core doesn't care (and doesn't even know) which of its windings is passing current at a given moment, *as long as the ampere-turns is maintained* — because only the net ampere-turns determine the field (and energy) inside the core. But as far as the windings themselves are concerned, the current is now *pulsed* — with *sharp* edges, and therefore with significant high-frequency content. Because of this, 'skin depth' considerations are necessary for choosing the appropriate wire thickness of the windings of a flyback transformer.

Note: We had ignored this for dc-dc inductors, but in high-frequency dc-dc designs too (or with high *r*), we may need to apply these concepts there too.

At high frequencies, the electric fields between the electrons become strong enough to cause them to repel each other rather decisively, and thereby cause the current to crowd on the *exterior* (surface) of the conductor (see exponential curve in *Figure 3-3*). This crowding worsens with frequency as per $\sqrt{f}$. There is thus the possibility that though we may be using thick wire in an effort to reduce the copper loss, a good part of the cross-section of the wire (its "innards") just may not be *available* to the current. The resistance presented to the current flow is inversely proportional to the area through which the current is flowing, or is

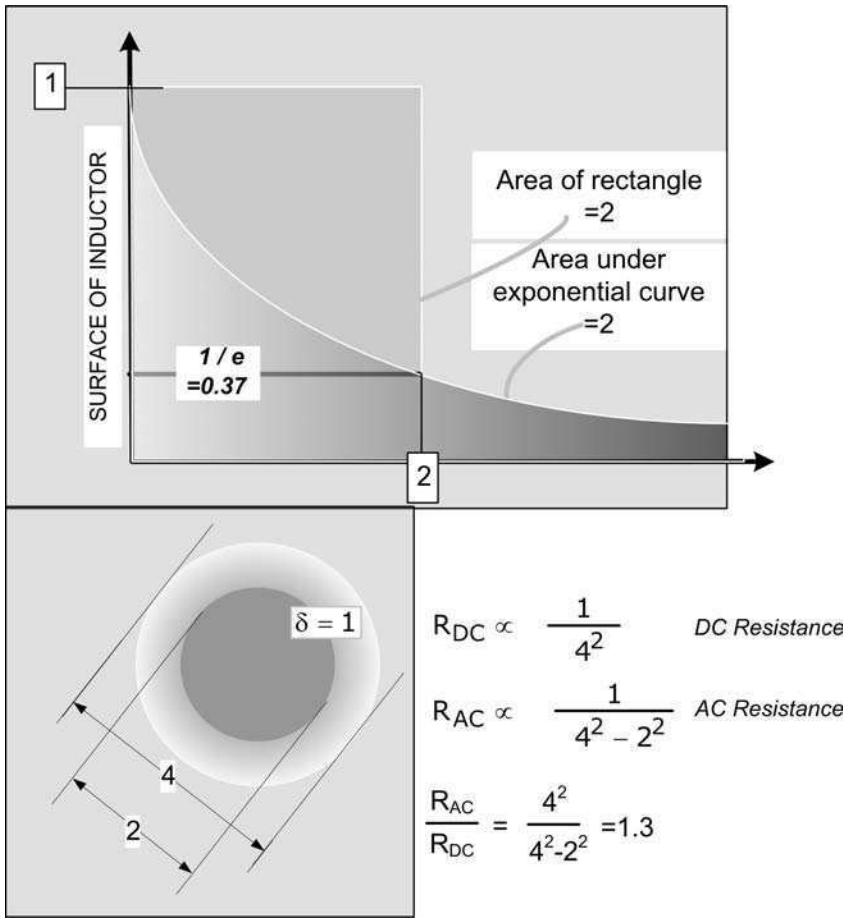


Figure 3-3: Skin Depth and AC Resistance Explained

able to flow. So this *current crowding* causes an increase in the effective resistance of the copper (as compared to its dc value). The resistance now presented to the current is called the '*ac resistance*' (see lower half of *Figure 3-3*). This is a function of frequency, because so is the skin depth. Instead of thus wasting precious space inside the transformer and losing efficiency too, we must try to use more *optimum diameters* of wire — in which the cross-sectional area is better-utilized. Thereafter, if we need to pass more current than the chosen cross-sectional area can handle, we need to *parallel* several such strands.

So how much current can a given wire strand handle? That depends purely on the heat buildup and the need to keep the overall transformer with an acceptable temperature rise. For this a good guideline/rule-of-thumb for the current density of flyback transformers is 400 *circular mils ('cmils') per ampere*, and that is our goal too in the analysis that follows.

Note: Expressing "current density" in the North American way of cmils/A needs a little getting used to. It is actually *area per unit ampere*, not *ampere per unit area* (as we would normally expect a "current density" to be)! So a higher cmils/A value actually is a *lower* current density (and vice versa) — and will produce a *lower* temperature rise.

We define the *skin depth* 'δ' as the distance from the surface of a conductor at which the current density falls to 1/e times the value at the surface. Note that the current density at the surface is the same as the value it would have had all through the copper, were there no high-frequency effects. As a good approximation to the exponential curve, we can also imagine the current density *remaining unchanged* from the value at the surface, until the skin depth is reached, *falling abruptly to zero thereafter*. This follows from an interesting property of the exponential curve that the area under it from 0 to ∞ is equal to the area of a rectangle passing through its 1/e point (see *Figure 3-3*).

Therefore, when using round wires, if we choose the diameter as *twice* the skin depth, no point inside the conductor will be more than one skin depth away from the surface. So no part of the conductor is unutilized. In that case, we can consider this wire as having an ac resistance equal to its dc resistance — *there is no need to continue to account for high-frequency effects so long as the wire thickness is chosen in this manner*.

If we use copper foil, its *thickness* too needs to be about twice the skin depth.

In *Figure 3-4* we have a simple nomogram for selecting the wire gauge and thickness. The *upper* half of this is based on the current carrying capability as per the usual requirement of 400 cmil/A. But the readings can obviously be linearly scaled for any other desired current density. The vertical grid on the nomogram represents wire gauges. An example based on a switching frequency of 70 kHz is presented within the figure. In a similar manner, for our previous worked example, we see that for 150 kHz operation, we should use AWG 27. But its current carrying capacity is only 0.5 A at 400 cmils/A (and only 0.25 A at a *lower* current density of 800 cmils/A!). Therefore, since the center of the primary current ramp was iterated and estimated to be 1.488 A, we need *three* strands of AWG 27 (twisted together) to give a combined current capability of 1.5 A (which is slightly better than what we need).

Coming to the secondary side of the worked example, we remember we had *lumped* all the current as a 5 V equivalent load of 15 A. But in reality it is only 10 A, two-thirds of that. So the center of its current ramp, which we had calculated was about 34 A, is actually $(2/3) \times 34 = 22.7$ A. The balance of this, that is, $34 - 22.7 = 11.3$ A, reflects as $(5.6/13) \times 11.3 = 4.87$ A into the 12 V winding. So the center of the 12 V output's current ramp is 4.87 A. We can choose the 12 V winding arrangement using the same arguments we present below for the 5 V winding.

For the 5 V winding, we can consider using copper foil, since we have only two turns, and we need a high current capability. The center of the 5 V secondary-side current ramp

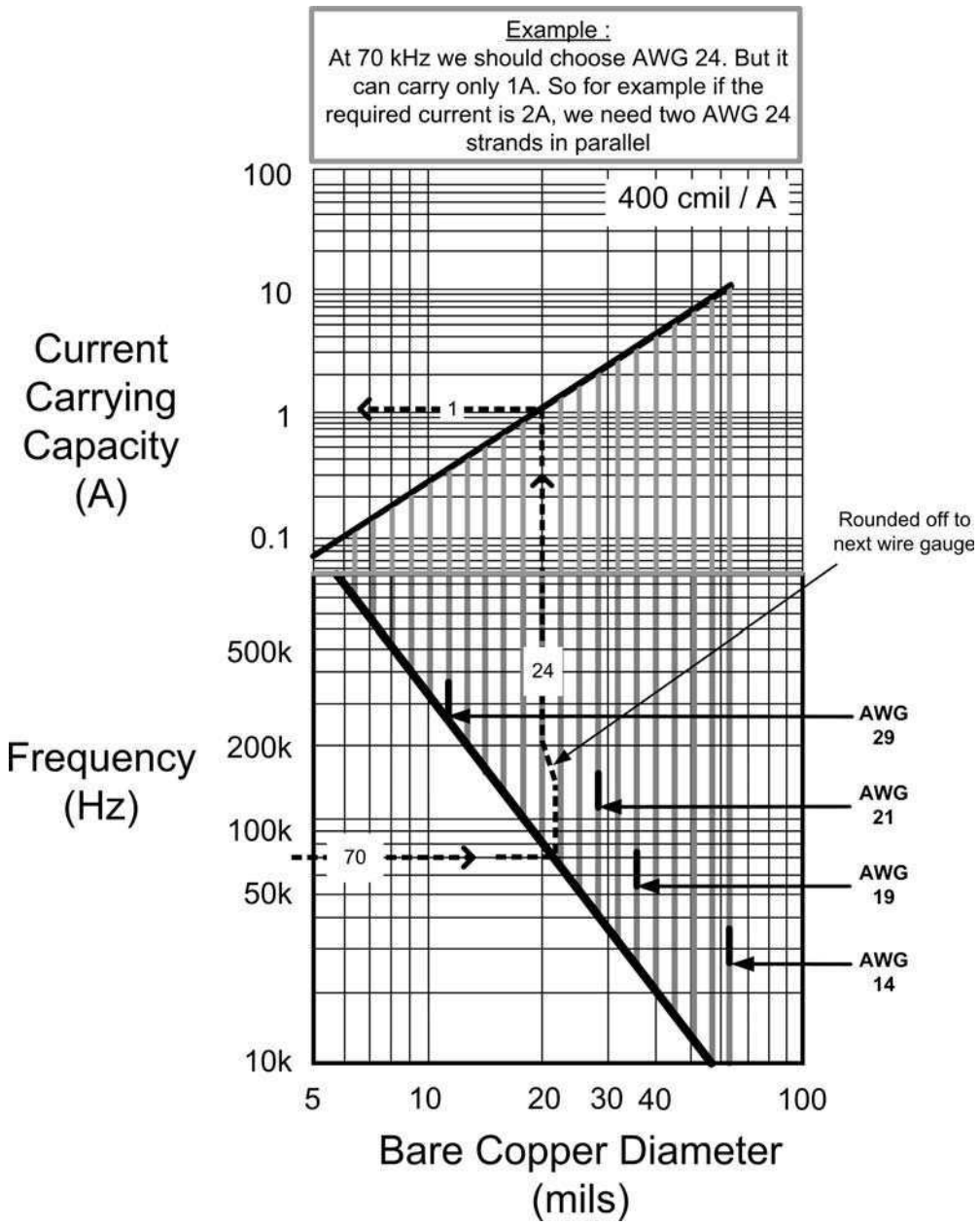


Figure 3-4: Nomogram for Selecting Wires and Foils Thicknesses, Based on Skin Depth Considerations

is about 23 A. The appropriate thickness (2δ) at this frequency is found by projecting downward along the AWG 27 vertical line. We get about 14 mils thickness. But we still don't know if the current through it will follow our guideline of 400 cmil/A, since it is a foil. We need to check this out further.

One circular mil ('cmil') is equal to 0.7854 square mils. Therefore 400 cmils is 400 × 0.7854 = 314 sq. mils (note $\pi/4 = 0.7854$). So for 23 A we need 23 × 314 = 7222 sq.mils. But the thickness of the foil is 14 mils. Therefore we need the copper foil to be 7222/14 = 515 mils wide, that is, about half an inch. Looking at a bobbin for the EI-30 in *Figure 3-5*, we see it can accommodate a foil 530 mils wide. So this is just about acceptable. Note that if the available width is insufficient, we would need to look for another core altogether — one with a "longer" (stretched out) profile. Cores like that are available as American "EER" cores. Or we can again consider using several paralleled strands of round wire. The problem is that a bunch of 46 twisted strands (of AWG 27) is going to be bulky, difficult to wind, and will also increase the leakage inductance. So we may like to use say 11 or 12 strands of AWG 27 twisted together into one *bunch*, and then take four of these bunches (all electrically in parallel), laid out side by side to form one layer of the transformer. For a two-turns secondary, therefore, we would wind two layers of this.

Forward Converter Magnetics

The procedure presented in this section applies explicitly to the *single-switch* forward converter. However, the general procedure remains unchanged for the *two-switch* forward converter as well.

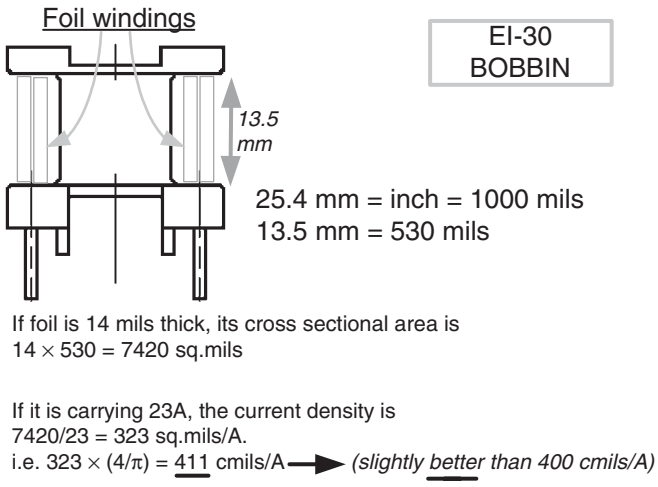


Figure 3-5: Checking to See if a 23 A Foil Can Be Accommodated on an EI-30 Bobbin

Duty Cycle

The duty cycle of a forward converter is

$$V_O = V_{IN} \times D \times \frac{n_S}{n_P}$$

Comparing this with the duty cycle of a buck, we see that the only difference is the term n_S/n_P. As mentioned, this is the coarse fixed-ratio step-down function available due to transformer action. We can therefore visualize that the input voltage V_{IN} gets reflected to the secondary side. This reflected voltage $V_{INR} = V_{IN}/n$ (where $n = n_P/n_S$) gets impressed at the secondary-side switching node. ***From there on, we have in effect a simple dc-dc buck stage, with an input voltage of V_{INR} and an output of V_O*** (see *Figure 3-6*). Therefore the design of the forward converter's choke is *not* going to be covered here, as it is designed using the same procedure as that of any buck inductor. However, the forward converter's transformer is another story altogether!

Note: Regarding choke design, we should keep in mind that for *high-current* inductors, as would be found in a typical forward converter, the calculated wire gauge may be too thick (and stiff) for winding easily over the core/bobbin. In that case, several thinner wire gauges may be twisted together to make the winding more flexible and easier to handle in production. Further, since choke and inductor design has usually little to do

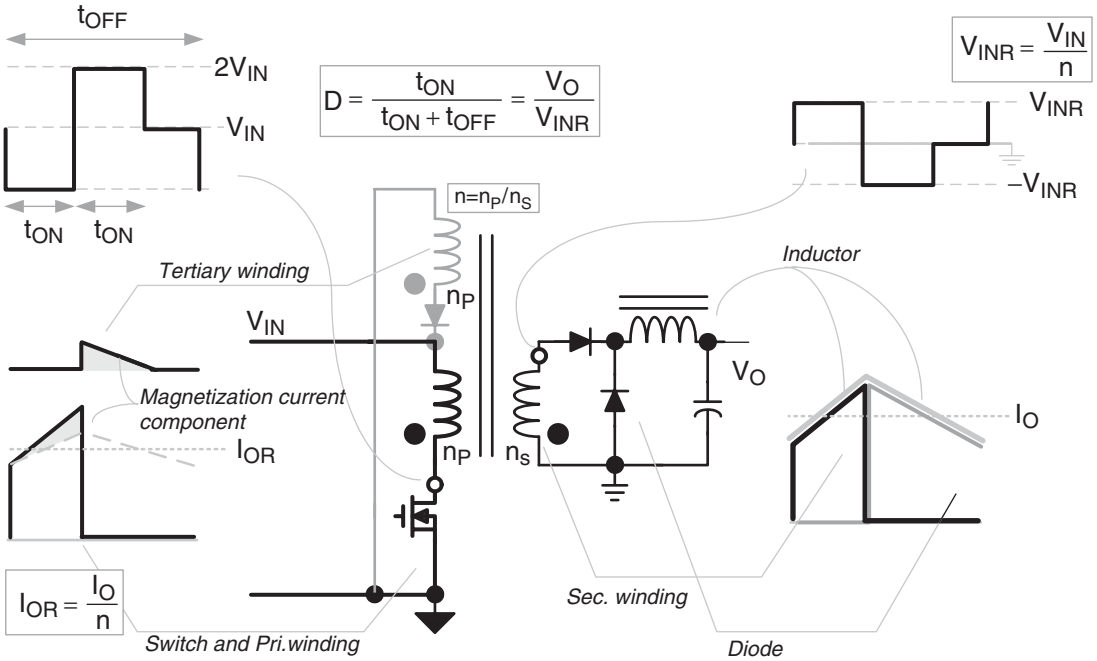


Figure 3-6: The single-ended Forward Converter

with high-frequency skin depth considerations, we can choose strands of almost any practical diameter, so long as we have enough net copper cross-sectional area to keep the temperature rise to within about 40 to 50°C.

Unlike a flyback transformer, the forward converter's secondary winding conducts at the same time as the primary winding. This leads to an *almost* complete flux cancellation inside the core. But there is one component of the primary current waveform which remains the same, irrespective of the load. This is the *magnetization current* component — shown in gray on the left side of *Figure 3-6*. At zero load, this is the entire current through the primary winding and switch (assuming duty cycle remains fixed). As soon as we try to draw some load current, the secondary winding current increases, and so does the primary winding current. Each current increases proportionally to the load current, and so their increments too are mutually proportional — the proportionality constant being the turns ratio. But more significantly, they are of *opposite sign* — that is, looking at *Figure 3-6*, we see that the current enters the dotted end of the transformer on the primary side, and on the secondary side, it leaves by the dotted end at the same time. Therefore, the net flux in the core of the transformer remains unchanged from the zero load condition (assuming D is fixed) — because the core just never "sees" any change in the net ampere-turns flowing through its windings. All conditions inside the core, i.e. the flux, the magnetic fields, the energy stored, and even the core loss, are only dependent on the magnetization current. Of course the windings themselves have a different story to tell — they bear the entire brunt, not only of the actual load current, but the *sharp edges* and consequent high-frequency content of the pulsed current waveforms.

The magnetization current component is *not* coupled by transformer action to the secondary. In that sense, it is like a "parallel leakage inductance." We need to subtract this component from the total switch current, and only then will we find that the primary and secondary currents scale according to the turns ratio. In other words, *the magnetization current does not scal*e — it stays confined to the primary side.

But in fact, the magnetization current is the only current component that is storing any energy in the transformer. So in that sense, it is like the flyback transformer! But, if we are to achieve a steady state, even the transformer needs to be "reset" every cycle (along with the output choke). But unfortunately, the magnetization energy is effectively "uncoupled," simply because of the output diode direction, and so we can't transfer it over to the secondary side. If we don't do anything about this energy, it will certainly destroy the switch by a spike similar to the leakage in a flyback. We don't want to burn it either, for efficiency reasons. Therefore, the usual solution is to use a 'tertiary winding' (or "energy recovery winding"), connected as shown in *Figure 3-6*. Note that this winding is in flyback configuration *with respect to the primary winding*. It conducts only when the switch turns OFF, and thereby freewheels the magnetization energy back into the input capacitor. There is some loss associated with this "circulating" energy term, because of the diode drop and

resistance of the tertiary winding. Note however, that any bona-fide leakage inductance energy also gets recycled back into the input by the tertiary winding. So we don't need an additional clamp for it.

For various subtle reasons, like being able to ensure the transformer resets *predictably under all conditions*, and also for various production-related reasons, the number of turns of the tertiary winding is usually kept exactly the same as the primary winding. Therefore by transformer action, the voltage at the primary-side switching node (drain of the mosfet) must rise to $2 \times V_{IN}$ when the switch turns OFF. Therefore, in a universal-input off-line *single-ended* (i.e. single-switch) forward converter, we need a switch rated for at least 800 V.

As soon as the transformer is reset (i.e. the current in the tertiary winding returns to zero), the drain voltage suddenly drops to V_{IN} — that is, *no voltage* is then present across the primary winding — and therefore there is no voltage across the secondary winding either. The catch diode of the output stage (i.e. the diode connected to the secondary ground in *Figure 3-6*) then freewheels the energy contained in the choke. Note that there is actually some ringing on the drain of the mosfet for a while, around an average level of V_{IN}, just after transformer reset occurs. This is attributable to various undocumented parasitics (not displayed in the figure). The ringing however does contribute significantly to the radiated EMI.

Note that even prior to transformer reset, the secondary winding has not been conducting for a while — simply because the output diode (i.e. the one connected to swinging end of the secondary winding) has been reverse-biased during the time the tertiary winding was conducting.

Note also that the duty cycle of such a forward converter can under no circumstances ever be allowed to exceed 50%. The reason for that is we have to unconditionally ensure that transformer reset will always occur, every cycle. Since we have no direct control on the transformer current waveforms, we have to just *leave enough time* for the current in the tertiary winding to ramp down to zero on its own. In other words, we have to allow *voltseconds balance* to occur naturally in the transformer. However, because the number of turns in the tertiary winding is equal to the primary turns, the voltage across the tertiary winding is equal to V_{IN} when the switch is ON, and is also equal to V_{IN} (opposite direction) when the switch is OFF. Reset will therefore occur when t_{OFF} becomes equal to t_{ON}. So, if the duty cycle exceeds 50%, t_{ON} would certainly always exceed t_{OFF}, and therefore transformer reset would never be able to occur. That would eventually destroy the switch. Therefore, just to allow t_{OFF} to be large enough, the duty cycle must always be kept to less than 50%.

We realize that the forward converter transformer is always in DCM (its choke is usually in CCM, with an *r* of *0.4*). Further, since the flux in the transformer remains unchanged for all loads, we can logically deduce that *no part* of the energy flowing through it into the output must be being stored in the transformer. So the question really is — what does the

power-handling capability of a forward converter transformer depend on? We intuitively realize that we can't use any size transformer for any output wattage! So what governs the size? We will soon see that it is determined simply by *how much copper we can squeeze into the available 'window area' of the core* (and more importantly, *how well we can utilize this available area*), without getting the transformer *too hot*.

Worst-case Input Voltage End

The most basic question in design invariably is — what input voltage represents the worst-case point at which we need to start the design of the magnetics (from the viewpoint of core saturation)? For the forward converter choke, this should be obvious — as for any buck converter, we need to set its current ripple ratio at around 0.4 at V_{INMAX}. But coming to the transformer, we need some analysis before we can make a proper conclusion.

Note that the transformer of a forward converter is in discontinuous mode (DCM), but the duty cycle is determined by the choke, which is in CCM. Therefore, the duty cycle of the transformer also gets "slaved" at the CCM duty cycle of $D = V_O/V_{INR}$, despite the fact that it is in DCM. This rather coincidental CCM + DCM interplay leads to an interesting observation — the *voltseconds across the forward converter transformer is a constant, irrespective of the input voltage*. The following calculation makes that clear, by the fact that V_{IN} cancels out completely:

$$Et = V_{IN} \times \frac{D}{f} = V_{IN} \times \frac{V_O}{V_{INR} \times f} = V_{IN} \times \frac{V_O \times n}{V_{IN} \times f} = \frac{V_O \times n}{f}$$

So in fact, the swing 'Δ' of the current or the field is the same at high input or at low input, or in fact at any input (as long as the choke is in CCM). Since the transformer is in DCM, its peak is equal to its swing, and so the peak too does not depend on V_{IN}. Of course, the peak *switch* current I_{SW_PK} is the sum of the peak of the magnetization current I_{M_PK}, and the peak of the secondary-side current waveform reflected onto the primary side that is,

$$I_{SW\_PK} = I_{M\_PK} + \frac{1}{n}\left[I_O\left(1 + \frac{r}{2}\right)\right]$$

So although the ***current limit of the switch must be set high enough to accommodate I_{SW_PK} at V_{INMAX}*** (since that is where the maximum peak of the reflected output current component occurs), ***as far as the transformer core is concerned, the peak current (and corresponding field) is just I_{M_PK}, which does not depend on V_{IN}***! This is indeed an interesting situation. Note also, that as far as the *choke* is concerned, the peak inductor current is no longer equal to the (reflected) peak switch current (as in a dc-dc buck topology), though the peak diode current still is. Yes, if we subtract the magnetization current from the switch current, and then scale (reflect) it to the secondary side according to the turns ratio, then the peak of that waveform will be equal to the peak inductor current.

So effectively I_M has the property of *input voltage rejection*. We can intuitively understand this in the following way — as the input increases, it tries to increase the slope of the transformer current ramp and thereby get ΔI to increase. However, the output choke, sensing a higher V_{INR}, decreases its duty cycle, and therefore also that of the transformer, and in effect tries to thereby reduce the current swing in the transformer. Coincidentally, these two opposing forces virtually counterbalance each other perfectly, and so there is no net change in the resultant current swing in the transformer.

As a corollary, the *core loss* in the transformer is independent of the input voltage. ***The copper loss, on the other hand, is always worse at low inputs*** (except for the dc-dc buck) — simply because the average input current has to increase so as to continue to satisfy the basic power requirement $P_{IN} = V_{IN} \times I_{IN} = P_O$.

Though we can pick any specific input voltage point for assuring ourselves that the core does not saturate *anywhere within its input range*, since the copper loss is at its worst at V_{INMIN}, we conclude that ***the worst-case for a forward converter transformer is at V_{INMIN}***. For the choke, it is still V_{INMAX}.

Window Utilization

Looking at a typical winding arrangement on an 'ETD-34' core and bobbin in *Figure 3-7*, we see that the plastic bobbin occupies a certain part of the space provided by the core — thus reducing the available window 'Wa' from 171 mm^2 to 127.5 mm^2 — that is, by 74.5%. Further, if we include the 4 mm "*margin tape*" that needs to be typically provided on either side (to satisfy international safety norms regarding clearance and 'creepage' requirements between primary and secondary sides), we are left with an available window of only 78.7 mm^2 — that's a total reduction of 78.7/171 = 46%. In addition to this, looking at the left side of *Figure 3-8*, we see that for any given wire, only 78.5% of the square area it "physically occupies" (or will occupy in the transformer) is actually conducting (copper). So in all, this leads to a total reduction of the available window space by $0.46 \times 0.785 = 36\%$.

We realize some more space will also be lost to interlayer insulation (and any EMI screens if present), and so on. Therefore, finally, we estimate that perhaps only 30 to 35% of the available *core window area* will actually be occupied by copper. That is the reason why we need to introduce a 'window utilization factor' *K* (later we will set it to an estimated value of 0.3). So

$$K = \frac{N \times A_{CU}}{Wa}$$

and

$$N = \frac{K \times Wa}{A_{CU}}$$

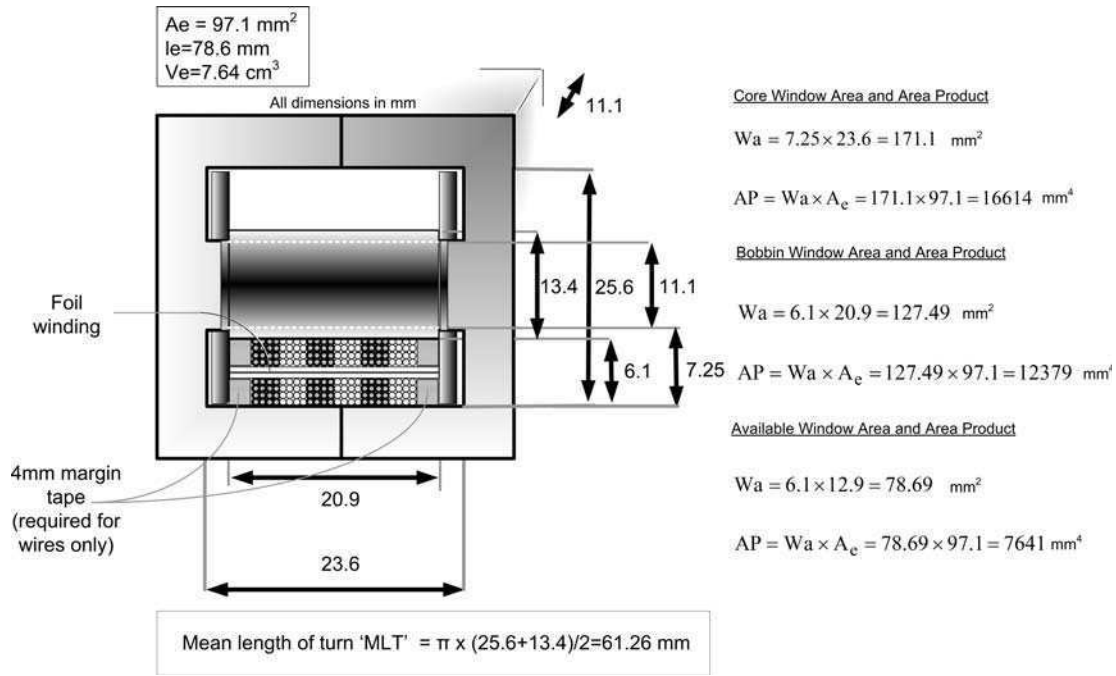


Figure 3-7: An ETD-34 Bobbin Analyzed

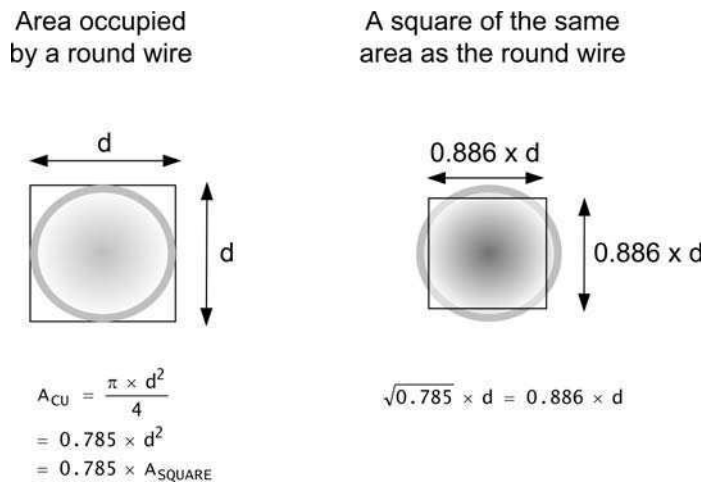


Figure 3-8: The Area Physically Occupied by a Round Wire, and a "Square Wire" of the Same Conducting Cross-sectional Area as a Round Wire

Here A_{CU} is the cross-sectional area of *one* copper wire, and Wa is the entire window area of the core (note that for EE, EI types of cores this is only the area of *one* of its two windows!).

Relating Core Size to Its Power Throughput

We remember that the original form of the voltage dependent equation is

$$\Delta B = \frac{V_{IN} \times t_{ON}}{N \times A} \text{ teslas}$$

Substituting for N, the number of primary turns, we get

$$\Delta B = \frac{V_{IN} \times t_{ON} \times A_{CU}}{K \times Wa \times A} \text{ teslas}$$

Performing some manipulations,

$$\Delta B = \frac{V_{IN} \times I_{IN} \times t_{ON} \times A_{CU}}{I_{IN} \times K \times Wa \times A} = \frac{P_{IN} \times (D/f) \times A_{CU}}{I_{IN} \times K \times Wa \times A} = \frac{P_{IN} \times (D/f) \times A_{CU}}{I_{SW} \times D \times K \times Wa \times A}$$

$$\Delta B = \frac{P_{IN}}{(I_{SW}/A_{CU}) \times K \times f \times Wa \times A} = \frac{P_{IN}}{\left(J_{A/m^2}\right) \times K \times f \times AP}$$

where J_{A/m^2} is the current density in A/m^2, and 'AP' is called the '*area product*' ($AP = A_e \times Wa$). Let us now convert into CGS units for greater convenience. We get

$$\Delta B = \frac{P_{IN}}{\left(J_{A/cm^2}\right) \times K \times f \times AP} \times 10^8 \text{ gauss}$$

where AP is also in cm^2 now. Finally, converting the current density into cmils/A by using

$$J_{cmils/A} = \frac{197{,}353}{J_{A/cm^2}}$$

we get

$$\Delta B = \frac{P_{IN} \times J_{cmils/A}}{197{,}353 \times K \times f \times AP} \times 10^8 \text{ gauss}$$

Solving for the area product

$$AP = \frac{506.7 \times P_{IN} \times J_{cmils/A}}{K \times f \times \Delta B} \text{ cm}^4$$

Let us do some substitutions here. Assuming a typical current density of 600 cmil/A, utilization factor K of 0.3, ΔB equal to 1500 gauss, we get the following fundamental core-selection criterion:

$$AP = 675.6 \times \frac{P_{IN}}{f} \text{ cm}^4$$

Note: In a typical forward converter, it is customary to set the swing in the B-field of the transformer at $\Delta B \approx 0.15$ teslas. This helps reduce core loss, and usually also leaves enough safety margin for avoiding hitting B_{SAT} under say power-up condition at high line. Note that in a flyback, the core loss tends to be much less, because ΔI is a fraction of the total current (40% typically). But since the transformer of a forward converter is always in DCM, therefore the swing in B is now more significant — equal to its peak value, i.e. $B_{PK} = \Delta B$. So, if we set the peak field at 3000 gauss, ΔB would be 3000 gauss too, roughly twice that of a flyback set to the same peak. That is why we must reduce the peak field in a forward converter to about 1500 gauss.

Worked Example (8) — Designing the Forward Transformer

We are building a 200 kHz forward converter for an ac input range of 90 to 270 Volts. The output is 5 V at 50 A, and the estimated efficiency is 83%. Design its transformer.

Input Power

We have

$$P_{IN} = \frac{P_O}{\text{Efficiency}} = \frac{5 \times 50}{0.83} \approx 300 \text{ Watts}$$

Selection of Core

We use the criterion calculated previously:

$$AP = 675.6 \times \frac{P_{IN}}{f} = 675.6 \times \frac{300}{2 \times 10^5} = 1.0134 \text{ cm}^4$$

The area product of the ETD –34 shown in *Figure 3-7* is

$$AP = W\frac{\left[\frac{25.6-11.1}{2}\right] \times 23.6 \times 97.1}{10^4} = 1.66 \text{ cm}^4$$

This is in theory, probably a little larger than required. Bu it is the closest standard size in this range. Later we will see it is in fact just about adequate.

Skin Depth

The skin depth is

$$\delta = \frac{66.1 \times \left[1 + 0.0042\left(T - 20\right)\right]}{\sqrt{f}} \text{ mm}$$

where f is in Hz and T is the temperature of the windings in °C. Therefore assuming a final temperature of T = 80°C (40°C rise over a maximum ambient of 40°C), we get at 200 kHz

$$\delta = \frac{66.1 \times \left[1 + 0.0042 \times \left(60\right)\right]}{\sqrt{2 \times 10^5}} = 0.185 \text{ mm}$$

Thermal Resistance

An empirical formula for EE-EI-ETD-EC types of cores is

$$Rth = 53 \times V_e^{-0.54} °C/W$$

where V_e is in cm^3. Therefore since $V_e = 7.64$ cm^3, for the ETD-34

$$Rth = 53 \times 7.64^{-0.54} = 17.67°C/W$$

Maximum B-field

For a 40°C estimated rise in temperature, the maximum allowed dissipation is

$$P \equiv P_{CU} + P_{CORE} = \frac{\text{deg C}}{Rth} = \frac{40}{17.67} = 2.26 \text{ Watts}$$

Let's *divide this loss equally into copper and core losses* (typical first-cut assumption). So

$$\begin{matrix} P_{CU} = 1.13 \\ P_{CORE} = 1.13 \end{matrix} \text{ Watts}$$

Therefore, the allowed *core loss per unit volume* is

$$\frac{\text{core loss}}{\text{volume}} = \frac{1.13}{7.64} \Rightarrow 148 \text{ mW/cm}^3$$

Using "System B" of *Table 2-5* we get

$$\frac{\text{core loss}}{\text{volume}} = C \times B^p \times f^d$$

where B is in gauss and *f* in Hz. Therefore solving for B

$$B = \left[\frac{\text{core loss}}{\text{volume}} \times \frac{1}{C \times f^{d}}\right]^{1/P}$$

If we are using the ferrite grade "3C85" (from Ferroxcube), we see from *Table 2-6* that $p = 2.2$ and $d = 1.8$ and $C = 2.2 \times 10^{-14}$. Therefore

$$B = \left[148 \times \frac{1}{2.2 \times 10^{-14} \times 2^{1.8} \times 10^{5 \times 1.8}}\right]^{1/2.2} = 720 \text{ gauss}$$

We note that the "B" referred to here is actually, by convention, B_{AC}. So, we get the total allowed swing as

$$\Delta B = 2 \times B = 2 \times 720 = 1440 \text{ gauss}$$

Voltμseconds

Earlier, we had presented the following form of the voltage dependent equation

$$\Delta B = \frac{100 \times Et}{Z \times A} \text{ gauss}$$

where A is the effective area in cm^3. The ***duty cycle of a typical forward converter is set to about 0.35 at low line*** so as to meet the typical 20 ms holdup time requirement, without requiring an inordinately-sized input capacitor. The rectified input at low line is $90 \times 2 = 127$ V. The applied voltseconds is therefore (at any line voltage)

$$Et = V_{IN} \times \frac{D}{f} = 127 \times \frac{0.35}{2 \times 10^{5}} = 222.25 \text{ V}\mu\text{s}$$

Number of Turns

Since $\Delta B = 1440$ gauss, we solve the following equation for N

$$\Delta B = \frac{100 \times Et}{Z \times A} \text{ gauss}$$

$$n_P = \frac{100 \times Et}{\Delta B \times A} = \frac{100 \times 222.25}{1440 \times 0.97} = 15.9 \text{ turns}$$

Note that this says nothing about the required *inductance*. We need these many number of turns, irrespective of the (primary) inductance. Yes, changing the inductance will affect the

peak magnetization and the switch current, because *it changes the proportionality constant connecting B and I.* However, B still remains fixed, independent of the inductance!

Assuming a 0.6 V forward drop across the diode, the required turns ratio is

$$n = \frac{n_P}{n_S} = \frac{V_{IN}}{V_{INR}} = \frac{V_{IN}}{\left(\frac{V_O + V_D}{D}\right)} = \frac{127 \times 0.35}{5 + 0.6} = 7.935$$

Therefore, the number of secondary turns is

$$n_S = \frac{15.9}{7.935} = 2.003 \text{ turns}$$

Note that this could have turned out to be significantly different from an integer. In that case, we would round it off to the nearest (higher) integer, and then recalculate the primary turns, the new flux density swing, and the core loss — similar to what we did for the flyback. But at the moment, we can simply use

$$\begin{aligned} n &= 8 \textit{ (turns ratio)} \\ n_P &= 16 \\ n_S &= 2 \end{aligned} \text{ turns}$$

Secondary Foil Thickness and Losses

The concept of skin depth presented earlier actually represents a single wire standing freely in space. For simplicity, we just ignored the fact that the field from the *nearby windings* may be affecting the current distribution significantly. In reality ***even the annular area we were hoping was fully available for the high frequency current, is not.*** Every winding has an associated field, and when this impinges on nearby windings, the charge distribution changes, and eddy currents are created (with their own fields). This is called the 'proximity effect.' It can greatly increase the *ac resistance* and thus the copper losses in the transformer.

The first thing we need to do to improve the situation is have *opposing* flux lines cancel each other. In a forward converter, that is in fact something that tends to happen automatically, because the secondary windings pass current at the same time as the primary, and in the opposite direction. However, even that can prove totally inadequate, especially at the higher power levels that a forward converter is more commonly associated with. So a further reduction in these proximity losses is achieved by *interleaving* as shown in *Figure 3-9*.

Basically, by splitting the sections, and trying to get primary and secondary layers *adjacent to each other as much as possible,* we can increase cancellation of local adjoining fields. In effect, we are trying to prevent the ampere-turns from *cumulating* as we go from one layer to the next. Note that the ampere-turns are proportional to the local fields that are causing the

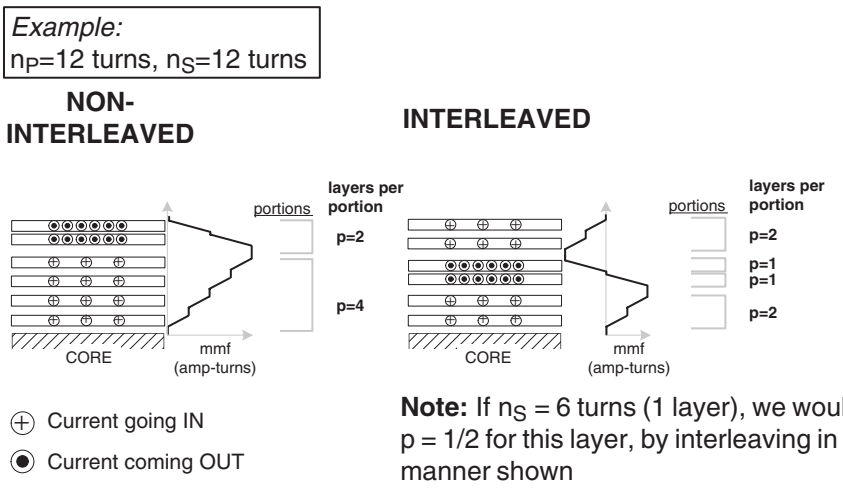


Figure 3-9: How Proximity Losses Are Reduced by Interleaving

proximity losses. However, it is impractical to interleave too much — because we will need several more layers of primary-to-secondary insulation, more terminations, and also more EMI screens at every interface (if required) — all of which will add up to higher cost and eventually lead to possibly higher, rather than lower, leakage. Therefore, ***most medium-power off-line supplies just split the primary into two sections, one on either side of a single-section secondary***.

The other way to reduce losses is to decrease the thickness of the conductor. But there are several ways we can do this. If, for example, we take a winding made up of single-strand wire, and split the wire into several paralleled finer strands in such a way that *the overall dc resistance does not* change in the process, we will find that the ac resistance goes up first before it reduces. On the other hand, if we take a foil winding, and decrease its thickness, the ac resistance falls before it rises again.

In *Figure 3-9*, we have also defined 'p,' the *layers per portion*. Note how *p* gets reassigned when we interleave.

But how do we go about actually estimating the losses? Dowell reduced a very complex multidimensional problem into a simpler, one-dimensional one. Based on his analysis, we can show that there is an *optimum thickness for each layer*. Expectedly, this turns out to be much less than $2 \times \delta$, where δ is the skin depth defined earlier.

Note: In the flyback, we had ignored the proximity effect for the sake of simplicity. But in any case, since the primary and secondary windings do *not* conduct at the same time, interleaving won't help. But interleaving is still carried out in the flyback, in a manner similar to the forward converter. However, the purpose then is to increase coupling between primary and secondary, and thereby reduce the leakage

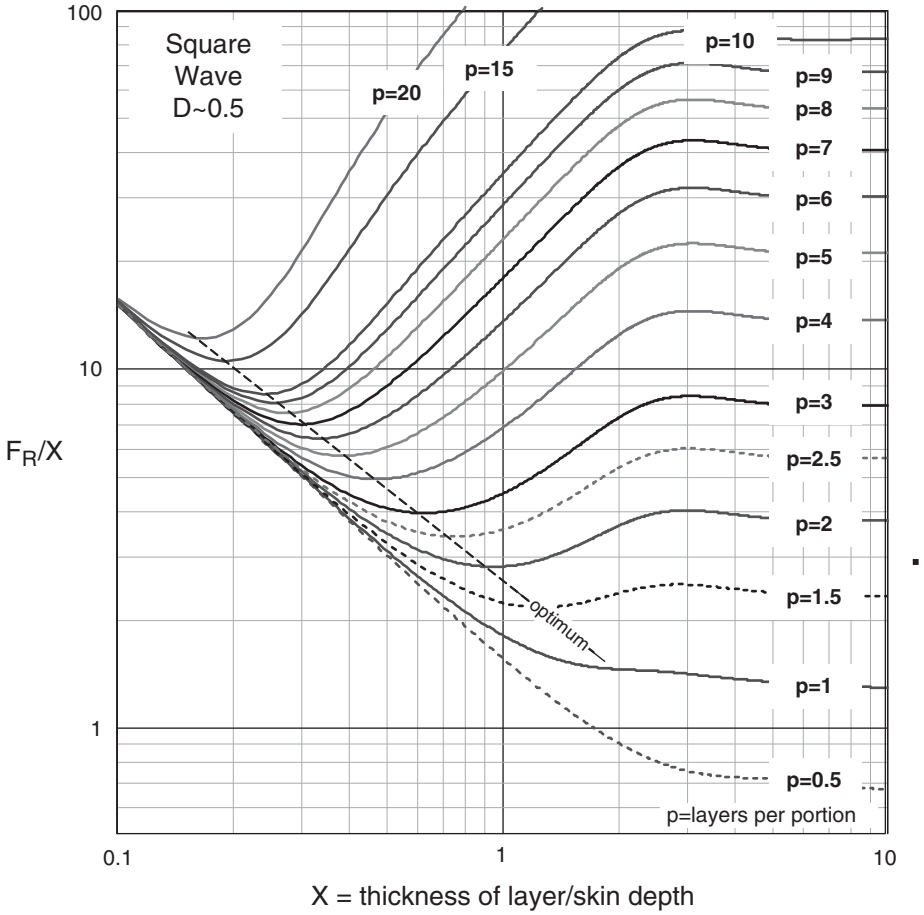


Figure 3-10: Finding the Lowest AC Resistance, as the Thickness of a Foil Is Varied

inductance. However, this also increases the capacitive coupling — unless grounded screens are placed at the primary-secondary interface. Screens are in general helpful in reducing high-frequency noise from coupling into the output, and suppressing common-mode conducted EMI. But they also increase the leakage inductance, which is of great concern particularly in the flyback. Note also that screens must be very thin, or they will develop very high eddy current losses of their own. Further, the ends of a primary-secondary screen should not be connected together, or they will constitute a shorted turn for the transformer.

In *Figure 3-10*, we have plotted out Dowell's equations in a form applicable to a *square* current waveform (unidirectional) in a transformer with foil windings. Note that the original Dowell curves actually plot F_R versus X. But we have plotted F_R/X versus X, where

$$F_R = \frac{R_{AC}}{R_{DC}}$$

and

$$X = \frac{h}{\delta}$$

h being the thickness of the foil. The reason why we have not plotted F_R versus X is that F_R is only the *ratio* of the ac to dc resistance. ***It is not F_R, but R_{AC} that we are really interested in minimizing***. So the "optimum R_{AC}" point need not necessarily be the point of the lowest F_R.

Let us try to understand this for a stand-alone foil (similar to what we did in *Figure 3-3*). If we slowly increase the thickness of the foil, once the foil thickness exceeds 2δ, the ac resistance won't change any further, since the cross-sectional area available for the high-frequency current remains confined to δ on each side of the foil. But the dc resistance continues to decrease as per 1/h — and as a result F_R will increase. So the relationship between R_{AC} and F_R is not necessarily obvious. Therefore, since $F_R = R_{AC}/R_{DC}$, with $R_{DC} \propto 1/h$, we get $R_{AC} \propto F_R/h$. And this is what we really need to minimize (for a foil). Further, since we always like to write any frequency-dependent dimension in reference to the skin depth, we have plotted not F_R/h, but F_R/X, versus X, in *Figure 3-10*.

Note that in *Figure 3-10*, the $p = 1$ and $p = 0.5$ curves do not really have an "optimum." For these, the F_R/X (ac resistance) can be made even smaller as we increase X (thickness). F_R will in fact become much greater than 1. However, we see that *for p = 1 for example, no significant reduction in ac resistance occurs if X exceeds about 2*, that is, thickness of foil equal to twice the skin depth. We can make it thicker if we want, but only for marginal improvement in the secondary winding losses. Further, in the process, we may also take away available area for the primary windings (and any other secondary windings), and that can lead to higher overall losses. *Though we are also cautioned not to fill up all "available space" with copper, especially when we come to (round) wire windings.* That can be shown, not only to increase F_R, but R_{AC} too.

Now let us apply what we have learned to our ongoing numerical example. We start by taking a copper foil wound twice on the ETD-34 bobbin — to form the 5 V secondary winding. Since this is interleaved with respect to the primary, only one turn "belongs" to each split section. So the *layers per portion* for the secondary is $p = 1$. We will calculate the losses, and if acceptable, we will stay with the resulting arrangement.

We can start with a reasonable current density (about 400 cmils/A should suffice here). We use

$$h = \frac{I_O \times J_{cmils/A} \times 10^2}{\text{width} \times 197{,}353} \text{ mm}$$

where h is the foil thickness in mm, I_O is the load current (50 A in our example), and 'width' is the width available for the copper strip (20.9 mm for the ETD-34).

Alternatively, we can directly consult *Figure 3-10* and pick an X of 2.5 for an estimated F_R/X of 1.4. Thus

$$h = X \times \delta = 2.5 \times 0.185 = 0.4625 \text{ mm}$$

The *mean length per turn* ('MLT') of the ETD-34 is 61.26 mm (see *Figure 3-7*), the ("hot") resistivity of copper ('ρ') is 2.3×10^{-5} ohms-mm, so we get the resistance of the secondary winding in ohms as

$$R_{AC\_S} = \left(\frac{F_R}{X}\right) \times \frac{\rho \times MLT \times n_S}{width \times \delta} = (1.4) \times \frac{2.3 \times 10^{-5} \times 61.26 \times 2}{20.9 \times 0.185} = 1.02 \times 10^{-3}$$

Note that since F_R/X is set to 1.4, the corresponding F_R is

$$F_R = 1.4 \times \frac{h}{\delta} = 1.4 \times \frac{0.4625}{0.185} = 3.5$$

This is fairly high, but as explained, it is actually helpful here, because R_{AC} goes down. Now, the current in the secondary looks like a typical switch waveform, with its center equal to the load current (50 A), and a certain current ripple ratio set by the output choke. Its RMS value is

$$I_{RMS\_S} = I_O \times \sqrt{D \times \left(1 + \frac{r^2}{12}\right)} \text{ A}$$

However, we do not yet know what the current ripple ratio of the choke, *r*, is *at 90 VAC*. The *r* has probably been set to 0.4 *at V_{INMAX}, not at V_{INMIN}.* Nevertheless, it is easy to work out the new *r* as follows. The duty cycle is inversely proportional to input voltage. Therefore, if D is 0.35 at 270 VAC, then at 90 VAC it is 0.35/3 = 0.117. *Further, r varies as per (1−D) for a buck stage*. Therefore the value of *r* at 90 VAC is

$$r = \frac{1 - 0.35}{1 - 0.117} \times 0.4 = 0.294$$

So the RMS current in the secondary winding is

$$I_{RMS\_S} = I_O \times \sqrt{D \times \left(1 + \frac{r^2}{12}\right)} = 50 \times \sqrt{0.35 \times \left(1 + \frac{0.294^2}{12}\right)} = 29.69 \text{ A}$$

The heat dissipated in the secondary windings is finally

$$P_S = {I_{RMS\_S}}^2 \times R_{AC\_S} = 29.69^2 \times 1.02 \times 10^{-3} = 0.899 \text{ W}$$

If the losses are not acceptable, we may need to look for a bobbin that will allow a *wider width* of foil. Or we can consider paralleling several thinner foils to increase *p*. For example, if we take four paralleled (thinner) foils in parallel (each insulated from the others), we will get four effective layers for the secondary, and the layers per portion will then become 2.

Primary Winding and Losses

For the secondary, we have finally chosen copper foil of thickness 0.4625 mm (i.e. $0.4625 \times 39.37 = 18$ mil). Let us assume each foil is covered on both sides by a 2 mil thick mylar tape. Since 1 mil is 0.0254 mm, we have effectively added 4×0.0254 mm to the foil thickness. In addition there will be three layers of tape between each of the two primary-secondary boundaries. So in all, the *thickness occupied by the secondary and the insulation*, h_S, is

$$h_S = (n_s \times h) + (n_S \times 4 \times 0.0254) + (12 \times 0.0254) \text{ mm}$$

or

$$h_S = n_S \times \left(h + 0.102\right) + 0.305 \text{ mm}$$

So in our case

$$h_S = 2 \times \left(0.4625 + 0.102\right) + 0.305 = 1.434 \text{ mm}$$

The ETD-34 has an available height inside the bobbin of 6.1 mm. That now leaves $6.1 - 1.434 = 4.67$ mm. Therefore *each section* of the split primary has an available winding height of 2.3 mm only. We should ultimately check that we can accommodate the primary winding we decide on, within this space.

Note that for the primary, the available width is only 12.9 mm (since there is 4 mm margin tape on each side — for the secondary, since we have a foil with tape wrapped over it, we do not need the margin tape). We need to find how best to accommodate eight turns into this available area, with minimum losses.

> **Note:** It is not mandatory to use a particular thickness of insulating tape, provided it is safety-approved to withstand a specified voltage. We can, for example, use 1 mil approved tape or even 1/2 mil, if it suits our production, helps lower the cost, and/or improves performance in some way.

Let us first understand the basic concept for winding wires here. For a *stand-alone* wire, as in *Figure 3-3*, as we increase the diameter of the wire, the cross-sectional area available for the high-frequency current is $(\pi \times d) \times \delta$. And since resistance is inversely proportional to cross-sectional area, we get $R_{AC} \propto 1/d$. Similarly, $R_{DC} \propto 1/d^2$. So $F_R \propto d$. Therefore, $R_{AC} \propto 1/F_R$. This actually means that a higher F_R (bigger diameter) will decrease the ac resistance! That is not surprising, because the annulus available for the high-frequency

current does increase if the diameter increases. However, this is not the way to go when dealing with "non-stand-alone" wire. Because, by increasing the diameter, we will inevitably move on to higher number of layers, and Dowell's equations then tell us that the losses will increase, not decrease.

On the top left side of *Figure 3-11*, we have Dowell's *original* curves, which show how F_R varies with respect to X (i.e. h/δ). The parameter for each curve is *layers per portion* (i.e. *p*). Note that Dowell's curves talk in terms of *foils* only. They don't care about the actual number of turns in the primary or secondary (i.e. from the electrical point of view), but only the *effective layers per portion* (from the *field point of view*). So, when we consider a layer of round wires of diameter 'd,' we need to convert this into an *equivalent foil*. Looking back at the right side of *Figure 3-8*, we see that that this amounts to replacing a wire of diameter *d* with a foil slightly thinner (i.e. with the same amount of copper, but in a square shape). Alternatively, if we want to get a foil of X = 4 for example, we need to start with a wire of

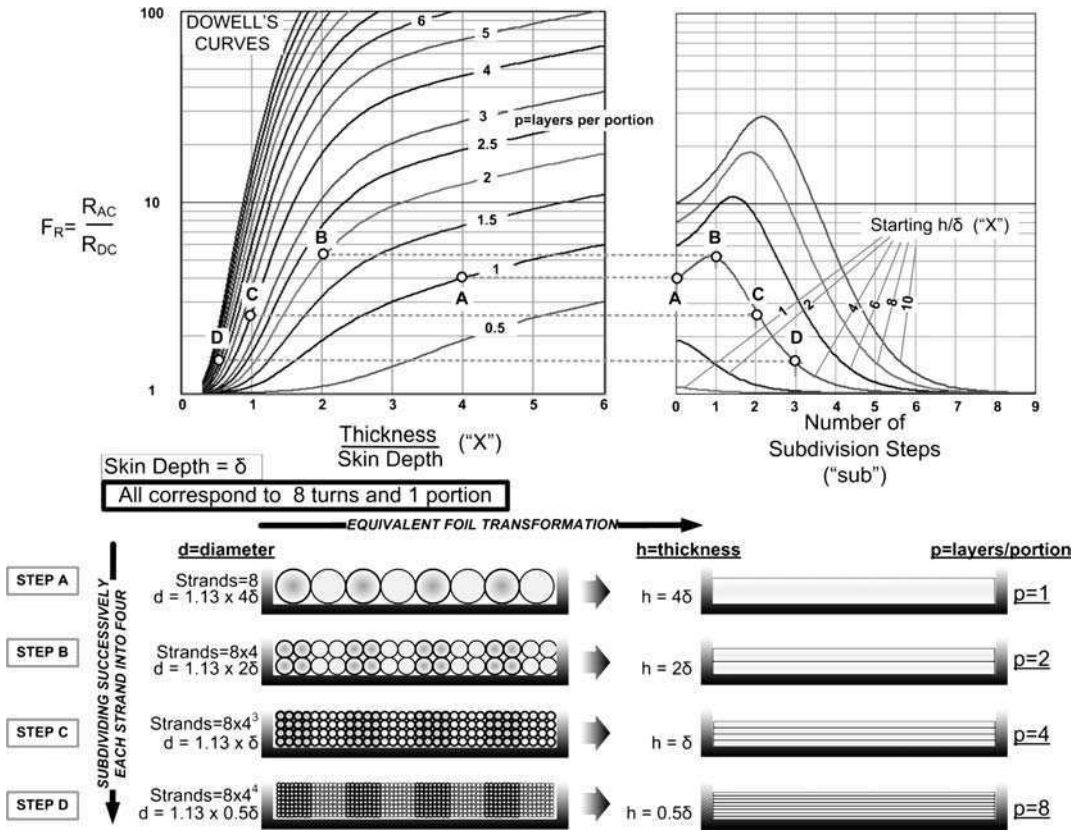


Figure 3-11: Understanding the process of "Subdivision" — Keeping the DC Resistance Unchanged, and How the Equivalent Foil Transformation Process Takes Place

diameter 1/0.886 = *1.13timesX*. Finally, as indicated, all these copper squares then merge (from the field point of view), to give an equivalent layer of foil.

In *Figure 3-11*, we are also conducting a certain "experiment" — as an alternative way of laying out wires optimally. Suppose we have several round wires laid out side by side with a diameter 1.13 × 4δ. Suppose also, that this constitutes *one layer per portion* in a given winding arrangement. This is therefore equivalent to a single-layer foil of thickness 4δ, that is, X = 4. Now using Dowell's curves, the F_R is about ***4*** (points marked "A" in *Figure 3-11*). Suppose we then divide *each strand into four strands*, where each strand has a diameter half the original. Therefore, *the cross-sectional area occupied by copper remains the same* because

$$A = 4 \times \frac{\pi \times \left(\frac{d}{2}\right)^2}{4} = \frac{\pi \times d^2}{4}$$

However, the *equivalent foil thickness is now half of what it was* — 2δ (i.e. X = 2). And we also now have *two layers per portion* from Dowell's standpoint. Consulting Dowell's curves, we get an F_R of about ***5*** now (marked "B"). Since we are keeping R_{DC} fixed in the process, $R_{AC} \propto F_R$. Therefore now, decreasing F_R is a sure way to go, to decrease R_{AC}. So an F_R of ***5*** is decidedly worse than an F_R of ***4***. We now go ahead and subdivide once more, in a similar manner. So we then get four layers per portion, each with X = 1, and F_R has gone down to about ***2.6*** (points marked "C"). We subdivide once more, and we get eight layers per portion, with X = 0.5. This gives us an F_R of about ***1.5*** (marked "D"). This is an acceptable value for F_R.

Note that all these steps have been collected and plotted out in *Figure 3-11* on the right side, ***with the horizontal axis being the number of successive subdivision steps*** (in each step we subdivided each wire into four of the same dc resistance). These steps are being called "sub" (for subdivision step), where *sub* goes from 0 (no subdivision) to 1 (1 subdivision), 2 (2 subdivisions), and so on. We then also realize that with each step, X and p change as per

$$X \rightarrow \frac{X}{2^{sub}}$$
$$p \rightarrow p \times 2^{sub}$$

For example, after four subdivision steps, the foil thickness will drop by a factor of 16, and the number of layers will increase by the same factor. We can then look at Dowell's curves to find out the new F_R.

However, there are a few problems with directly applying Dowell's curves to switching power regulators. For one, the original curves only talked about the ratio of the thickness to the skin depth — and we know skin depth depends on frequency. So implicitly, Dowell's curves provide the F_R for a *sine wave*. Further, Dowell's curves do not assume the current

has any dc value. So engineers, who adapted Dowell's curves to power conversion, would usually first break up the current waveform into its *ac and dc components*, apply the F_R obtained from the curves to the ac component only, compute the dc loss separately (with $F_R = 1$), and then sum as follows:

$$P = I_{DC}^2 \times R_{DC} + I_{AC}^2 \times R_{DC} \times F_R$$

However, in our case, we have preferred to follow the more recent approach of using the actual (unidirectional) current waveform, splitting it into Fourier components, and summing to get the *effective* F_R. The losses are expressed in terms of the thickness of the foil *as compared to the δ at the fundamental frequency* (first harmonic). We also include the dc component in computing this *effective* F_R. That is the reason, when calculating the secondary winding losses, we were able to use the simple equation

$$P = I_{RMS}^2 \times R_{AC} \equiv I_{RMS}^2 \times (F_R \times R_{DC})$$

In that case, the F_R was actually the effective F_R (computed for a square wave with dc level included), though not explicitly stated. However, note that the graphs in *Figure 3-11* are still based on the original sine-wave approach, and the purpose here was only to demonstrate the *subdivision technique* through the original curves.

But in *Figure 3-12*, we have finally modified Dowell's original sine-wave curves. Fourier analysis has been carried out while constructing these curves, and so the designer can apply them *directly* to the typical (unidirectional) current waveforms of power conversion. We will now use these curves to do the calculations for the primary winding of our ongoing numerical example.

But one question may still be puzzling the reader — why are we not using the previous F_R/X curves (see *Figure 3-10*) that we used for the secondary? The reason is the situation is different now. The curves in *Figure 3-10* are Dowell's curves for a square wave, except that on the vertical axis we have used F_R/X, not F_R. That is useful only when we are varying *h* and seeing when we got the lowest R_{AC}. But for the primary windings, we are going to fix the height of the windings in each step of the iterations that follow. We will be using the subdivision technique in each iteration, and therefore *keep the dc resistance constant*. So now the minimum R_{AC} (for a given iteration step) *will be achieved at the minimum F_R, not at the minimum F_R/X*.

The subdivision method was originally presented in *Figure 3-11*, except that now we will use the modified curves in *Figure 3-12*.

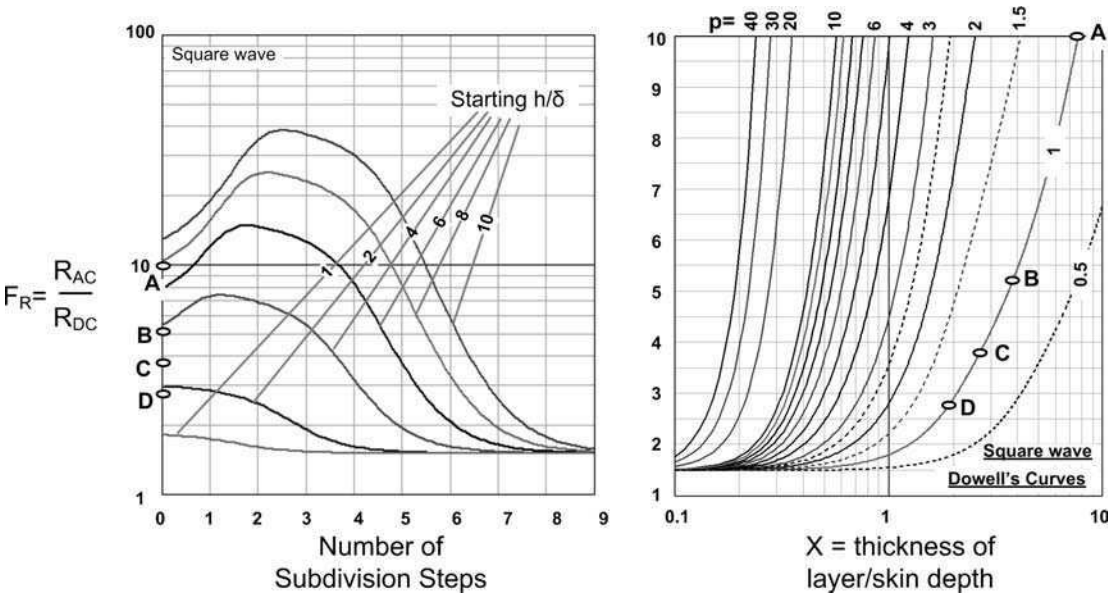


Figure 3-12: Dowell's Curves Modified for Square Current Waveforms, and the Corresponding F_R Curves for the Subdivision Method

First Iteration:

Let us plan to try and fit eight turns on one layer. *Lesser* number of layers will usually be better. We remember that we have 12.9 mm available width on the bobbin. So if we stack eight turns side by side (no gap between them) we will require each of these eight round wires to have a diameter of

$$d = \frac{\text{width}}{\text{turns per layer}} = \frac{12.9}{8} = 1.6125 \text{ mm}$$

We can check that the available height of 2.3 mm is big enough to accommodate this diameter of wire. The *penetration ratio* X is (using the equivalent foil transformation)

$$X = \frac{0.886 \times d}{\delta} = \frac{0.886 \times 1.6125}{0.185} = 7.723$$

The *p* is equal to 1. From either of the graphs in *Figure 3-12*, we can see that the F_R will be about 10 in this case (marked "A"). Further, from the graph on the left side, we can see that we need to subdivide the "X = 7.7" curve (imagine it close to the X = 8 curve) seven times to get the F_R below 2. That would give strands of diameter

$$d \rightarrow \frac{d}{2^{sub}} = \frac{1.6125}{2^7} = 0.0125 \text{ mm}$$

The corresponding AWG can be calculated by rounding off

$$\text{AWG} = 18.154 - 20\log(\text{d})$$

So we get

$$\text{AWG} = 18.154 - 20\log(0.0125) \Rightarrow 56 \text{ AWG}$$

But this is an extremely thin wire, and may not even be available! Generally, from a production standpoint, *we should not use anything thinner than 45 AWG (0.046 mm).*

Second Iteration:

The problem with the first iteration is that we started with a very thick wire, with a very high F_R. So this demanded several subdivisions to get the F_R to fall below 2. But what if we start off with a wire of lesser diameter than 1.6125 mm? We would then need to introduce some wire-to-wire spacing so we can spread the eight turns evenly across the bobbin. However, that would be wasteful! We should remember that if a layer is already assigned and present, we might as well use it to our full advantage to lower the dc resistance — the problem only starts when we indiscriminately increase the number of layers. Therefore in our case, let us try paralleling *two* thinner wires to make up the primary. We still want to keep to one layer (without spacing). That means we will now have 16 wires placed side by side in one layer. We then define a '*bundle*' as the number of wires paralleled to make the primary winding (we will be subdividing each of these further). So in our case

$$\text{bundle} = 2$$

The diameter we are starting off with is

$$\text{d} = \frac{\text{width}}{\text{turns per layer}} = \frac{12.9}{16} = 0.806 \text{ mm}$$

The penetration ratio X is

$$\text{X} = \frac{0.886 \times \text{d}}{\delta} = \frac{0.886 \times 0.806}{0.185} = 3.86$$

The *p* is still equal to 1. From both the graphs in *Figure 3-12*, we can see that the F_R will be about 5.3 in this case (marked "B"). Further, from the graph on the left side, we can see that we need to subdivide five times to get the F_R below 2. That would give strands of diameter

$$\text{d} \rightarrow \frac{\text{d}}{2^{\text{sub}}} = \frac{0.806}{2^5} = 0.025 \text{ mm}$$

This is still thinner than the practical AWG limit of 0.046 mm.

Third Iteration:

So we now parallel *three* wires to make up the primary. That means we will have 24 wires side by side in one layer.

$$\text{bundle} = 3$$

The diameter we are starting off with is

$$\text{d} = \frac{\text{width}}{\text{turns per layer}} = \frac{12.9}{24} = 0.538 \text{ mm}$$

The penetration ratio X is

$$\text{X} = \frac{0.886 \times \text{d}}{\delta} = \frac{0.886 \times 0.538}{0.185} = 2.58$$

The *p* is still equal to 1. From both the graphs in *Figure 3-12*, we can see that the F_R will be about 3.7 in this case (marked "C"). Further, from the graph on the left side, we can see that we need to subdivide four times to get the F_R below 2. That would give strands of diameter

$$\text{d} \rightarrow \frac{\text{d}}{2^{\text{sub}}} = \frac{0.538}{2^4} = 0.034 \text{ mm}$$

But this is still too thin!

Fourth Iteration:

Let us now parallel four wires to start with. We will have 32 wires in one layer.

$$\text{bundle} = 4$$

The diameter we are starting off with is

$$\text{d} = \frac{\text{width}}{\text{turns per layer}} = \frac{12.9}{32} = 0.403 \text{ mm}$$

The penetration ratio X is

$$\text{X} = \frac{0.886 \times \text{d}}{\delta} = \frac{0.886 \times 0.403}{0.185} = 1.93$$

The *p* is still equal to 1. From both the graphs in *Figure 3-12*, we can see that the F_R will be about 2.8 in this case (marked "D"). Further, from the graph on the left side, we can see that

we need to subdivide three times to get the F_R below 2. That would give strands of diameter

$$d \rightarrow \frac{d}{2^{sub}} = \frac{0.403}{2^3} = 0.05 \text{ mm}$$

This corresponds to AWG 44, and would be of acceptable thickness.

Note that by the process of subdivision, the number of layers per portion has gone up as

$$p \rightarrow p \times 2^{sub}$$

So with three subdivisions we have

$$p \rightarrow p \times 2^{sub} = 1 \times 2^3 = 8 \quad \textit{(layers per portion)}$$

that is, eight layers. The penetration ratio has similarly now become

$$X \rightarrow \frac{X}{2^{sub}} = \frac{1.93}{2^3} = 0.241$$

The F_R is now about 1.8 as can also be confirmed from the graph on the right side of *Figure 3-12* (for $X = 0.241$, $p = 8$).

The number of strands each original "bundle" has been divided into is

$$\text{strands} = 4^{sub} = 4^3 = 64$$

So finally, the primary winding consists of four bundles in parallel, each bundle consisting of 64 strands, side by side in one layer, with an F_R of about 1.8.

We can continue the process if we want to get a slightly lower F_R. But at some point we will find the F_R *will start to go up again*. For our purpose, we will take an F_R of less than 2 as acceptable to proceed with the loss estimates.

Note that further tweaking will always be required since when we bunch wires together to form a bundle, they will "stack" in a certain manner that will affect the dimensions from what we have assumed. Further, the diameter of the wire we used was for bare wire, and is slightly less than the coated diameter. Note that in general, *if after winding several layers evenly, we are left with a few turns that seem to need another layer to complete, we are better off reducing the primary number of turns and sticking to the existing completed layers, because even a few turns extra will count as a new layer from the field point of view, and increase proximity losses.*

We can now calculate the losses for the two primary sections combined, since they can be considered to be identical and with the same F_R. The ac resistance in ohms of the entire

primary winding is

$$R_{AC\_P} = (F_R) \times \frac{\rho \times MLT \times n_P}{\pi \times \frac{d^2}{4} \times \text{bundles} \times \text{strands}} = (1.8) \times \frac{2.3 \times 10^{-5} \times 61.26 \times 16}{\pi \times \frac{(0.05)^2}{4} \times 4 \times 64} = 0.08 \text{ ohms}$$

So the loss is

$$P_P = {I_{RMS\_P}}^2 \times R_{AC\_P} = \left(\frac{I_{RMS\_S}}{n}\right)^2 \times R_{AC\_P} = \left(\frac{29.69}{8}\right)^2 \times 0.08 = 1.102 \text{ Watts}$$

Had we gone further and divided the primary into five bundles and then subdivided three times, we would get eight layers with 64 strands of 0.04 mm diameter wire per bundle, and an F_R of 1.65 — which seems better than the 1.8 we got in the last step. But since the wires are so thin to start with, the dc resistance now goes up, and the dissipation will rise to 1.26 W.

Total Transformer Losses

The total dissipation in the transformer is therefore

$$P = P_{CORE} + P_{CU} = P_{CORE} + P_P + P_S = 1.13 + 1.102 + 0.899 = 3.131 \text{ W}$$

The estimated temperature rise

$$\text{deg C} = \text{Rth} \times P = 17.67 \times 3.145 = 55.3^\circ\text{C}$$

What we are seeing is a typical practical situation! The temperature rise is 15°C higher than we were expecting! However, 55°C is perhaps still acceptable (even from the standpoint of getting safety approvals without special transformer materials). Admittedly, there is room for more optimization. However, the next time we do the process, we must note that the core loss is only a third the total loss, not half, as we had initially assumed.

Note also that methods in related literature may predict a smaller temperature rise. But the fact is that these are usually based on the sine-wave versions of Dowell's equations, and we know that will typically underestimate the losses significantly.

CHAPTER

4

The Topology FAQ

This Page Intentionally Left Blank

CHAPTER 4

The Topology FAQ

This section serves to highlight and summarize the gamut of key topology-related design issues that should be kept in mind when actually designing converters (or when appearing for a job interview!).

Questions and Answers

Question 1: For a given input voltage, what output voltages can we get in principle, using only basic inductor-based topologies (buck, boost, and buck-boost)?

Answer: The buck is a *step-down* topology ($V_O < V_{IN}$), the boost only *steps-up* ($V_O > V_{IN}$), and the buck-boost can be used to either step-up or step-down ($V_O < V_{IN}$, $V_O > V_{IN}$). Note that here we are referring only to the *magnitudes* of the input and output voltages involved. So we should keep in mind that the buck-boost also inverts the polarity of the input voltage.

Question 2: What is the difference between a topology and a configuration?

Answer: We know that, for example, a 'down-conversion' of 15 V input to a 5 V output is possible using a buck *topology*. But what we are referring to here is actually a "positive-to-positive" buck *configuration*, or simply, a "positive buck." If we want to convert −15 V to −5 V, we need a "negative-to-negative" buck configuration, or simply, a "negative buck." We see that a topology is itself fundamental (e.g. the buck) — but it can be *implemented* in more than one way, and these constitute its *configurations*.

Note that in the down-conversion of −15 V to −5 V, we use a buck (step-down) topology, even though mathematically speaking, −5 V is actually a *higher* voltage than −15 V! Therefore, only *magnitudes* are taken into account in deciding what the nature of a power conversion topology is.

Similarly, a conversion of say 15 V to 30 V would require a "positive boost," whereas, −15 V to −30 V would need a "negative boost." These are the two configurations of a boost topology.

For a buck-boost, we need to always *mentally* keep track of the fact that it inverts the polarity (see next question).

Question 3: What is an "inverting" configuration?

Answer: The buck-boost is a little different. Although it has the great advantage of being able to up-convert or down-convert on demand, it also always ends up *inverting* the sign of the output with respect to the input. That is why it is often simply referred to as an "inverting topology." So for example, a "positive-to-negative" buck-boost would be required if we want to convert 15 V to −5 V or say to −30 V. Similarly, a "negative-to-positive" buck-boost would be able to handle −15 V to 5 V or to 30 V. Note that a buck-boost *cannot* do 15 V to 5 V for example, nor can it do −15 V to −5 V. The convenience of up- or down-conversion (on demand) is thus achieved only at the expense of a polarity inversion — the conventional (inductor-based) buck-boost topology is useful only if we either desire, or are willing to accept, this inversion.

Question 4: Why is it that *only* the buck-boost gives an inverted output? Or conversely, why can't the buck-boost ever *not*-invert?

Answer: In all topologies, there is a *voltage reversal* across the inductor when the switch turns OFF. So the voltage at one end of the inductor "flips" *with respect to its other end*. Further, when the switch turns OFF, the voltage present at the swinging end of the inductor (i.e. the switching node) always gets "passed on" to the output, because the diode is then conducting. But in the case of the buck-boost, the "quiet end" of its inductor is connected to the *ground reference* (no other topology has this). Therefore, the voltage reversal that takes place at its other end (swinging end) is also a voltage reversal *with respect to ground*. And since this is the voltage that ultimately gets transmitted to the output (which is also referenced to ground), it is virtually "seen" at the output.

Of course the output rail continues to stay inverted even when the switch turns ON, because the diode then stops conducting, and there is an output capacitor present, that holds the output voltage steady at the level it acquired during the switch off-time.

Question 5: Why do we always get only up-conversion from a boost converter?

Answer: *Inductor* voltage reversal during a switch transition occurs in all dc-dc switching topologies — it just does not necessarily lead to an *output* reversal. But in fact, inductor voltage reversal is responsible for the fact that in a buck, the input voltage is always stepped-down whereas in a boost, it is stepped-up. It all depends on where the "quiet" end of the inductor connects to. In the boost, the "quiet" end connects to the *input* rail (in the buck, to the output rail). Therefore, since the swinging end of the boost inductor is connected to ground during the switch on-time, it then flips *with respect to the input rail* during the switch off-time, gets connected to the output through the conducting diode, and thereby we get a boosted output voltage.

Question 6: What is really "ground" for a dc-dc converter?

Answer: In a dc-dc converter there are two input rails and two output rails. But one of these rails is common to both the input and output. This rail is the (power) "ground". The input and output voltages are measured with respect to this reference rail, and that gives them their respective magnitudes and polarity.

Question 7: What is "ground" for the control IC?

Answer: The reference rail, around which most of the internal circuitry of the IC is built, is its local (IC) ground. This rail comes out of the package as the ground pin(s) of the IC. Usually, this is connected on the PCB directly to the power ground. However, there are exceptions, particularly when an IC meant primarily for a certain topology (or configuration) is rather unconventionally configured to behave as another topology altogether (or just a different configuration). Then the IC ground may in fact differ from the power ground.

Question 8: What is "system" ground?

Answer: This is the reference rail for the entire system. So in fact, all on-board dc-dc converters present in the system usually need to have their respective (power) grounds tied firmly to this system ground. The system ground in turn usually connects to the metal enclosure, and from there on to the "earth (safety) ground" (i.e. into the mains wiring).

Question 9: Why are negative-to-negative dc-dc configurations rarely used?

Answer: The voltages applied to and/or received from on-board dc-dc converters are referenced by the rest of the system to the common shared system ground. By modern convention, all voltages are usually expected to be *positive* with respect to the system ground. Therefore, all on-board dc-dc converters also need to comply with the same convention. And that makes them necessarily positive-to-positive converters.

Question 10: Why are *inverting* dc-dc converters rarely used?

Answer: We usually cannot afford to let any given on-board converter attempt to "redefine" the ground in the middle of a system. However, inverting regulators can on occasion be used, especially if the converter happens to be a "front-end" converter. In this case, since the system effectively *starts* at the output terminals of this converter, we maybe able to "define" the ground at this point. In that case, the relative polarity between the input and output of the converter may become a "don't care" situation.

Question 11: Can a buck regulator be used to convert a 15 V input to 14.5 V output?

Answer: Maybe, maybe not! Technically, this is a step-down conversion, since $V_O < V_{IN}$. Therefore, in principle, a buck regulator should have worked. However in practice there are some limitations regarding how *close* we can set the output of a converter in relation to the input.

Even if the switch of a buck regulator is turned *fully* ON (say in an all-out effort to produce the required output), there will still be some remaining forward-drop across the switch, V_{SW}, and this would effectively subtract from the applied input V_{IN}. Note that in this fully ON state, the switcher is basically functioning just like an LDO, and so the concerns expressed in *Chapter 1* regarding the minimum achievable *headroom* of an LDO apply to the switcher too, in this state. As an example — if the switch drop V_{SW} is 1 V, then we certainly can't get anything higher than 14 V output from an input of 15 V.

The second consideration is that even if, for simplicity, we assume *zero* forward voltage drops across both the switch and the diode, we still may not be able to deliver the required output voltage — because of *maximum duty cycle* limitations. So for example, in our case, what we need is a (theoretical) duty cycle of $V_O/V_{IN} = 14.5\ V/15\ V = 0.97$, that is, 97%. However, most buck ICs in the market are not designed to guarantee such a high duty cycle. They usually come with an internally set maximum duty cycle limit ('D_{MAX}'), typically around 90 to 95%. And if that is so, D = 97% would be clearly out of their capability.

A good power supply designer also always pays heed to the tolerance or *spread* of the published characteristics of a device. This spread is usually expressed as a *range* with a specified "min" (minimum), a "max" (maximum), and a "typ" (typical, or nominal). For example, suppose a particular IC has a published maximum duty cycle range of 94 to 98%, we cannot *guarantee* that *all* production devices would be able to deliver a regulated 14.5 V — simply because not *all of them* are *guaranteed* to be able to provide a duty cycle of 97%. Some parts may manage 97%, but a few others won't go much beyond a duty cycle maximum of 94%. So what we need to do is to select an IC with the published 'min' of its tolerance range greater than the desired duty cycle. For example, a buck IC with a published D_{MAX} range of 97.5 to 99% *may* work in our current application.

Why did we say "may" above? If we include the forward drops of the switch and diode in our calculation, we actually get a *higher* duty cycle than the 97% we got using the "ideal" equation $D = V_O/V_{IN}$. The latter equation implicitly assumes $V_{SW} = V_D = 0$ (besides ignoring other key parasitics like the inductor's DCR). So the actual measured duty cycle in any application may well be a couple of percentage points higher than the ideal value.

In general we should remember that whenever we get *too close to the operating limits* of a control IC, we can't afford to ignore key parasitics. We must also account for temperature variations, because temperature may affect efficiency, and thereby the required duty cycle.

Question 12: What role does temperature play in determining the duty cycle?

Answer: As mentioned in *Chapter 1*, it is generally hard to predict the overall effect of temperature on a power-supply's efficiency, and thereby on its duty cycle variation with respect to temperature. Some loss terms increase with temperature and some decrease. However to be conservative, we should at a minimum account for the increase in the forward

drop of the mosfet switch. For low voltage mosfets (rated ~30 V), the increase in R_{DS} (on-resistance) in going from room temperature to "hot" is typically 30 to 50%. So we typically multiply the published room temperature on-resistance by 1.4. For high-voltage mosfets, as used in off-line power supplies, the increase is about 80 to 100%. So we typically multiply the room temperature on-resistance by 1.8.

Question 13: How can we convert an *unregulated* input of 15 V to a regulated output of 15 V?

Answer: The term "unregulated" implies that the stated value just happens to be the 'typical' (usually center) of a certain *range*, which may or may not yet have been defined. So an "unregulated input of 15 V" could well mean say 10 V to 20 V, or 5 V to 25 V, or 12 V to 18 V, and so on.

Of course, ultimately, we do need to know what this input range really is. But it should already be apparent that for a 15 V to 15 V conversion, if the input falls at the lower end of its range, we would need to up-convert, and if the input is at its upper end, we would need to down-convert. Therefore, we must choose a topology capable of performing *both* step-up and step-down conversions *on demand*.

How about the buck-boost? Unfortunately, the standard inductor-based buck-boost also gives us an *inverted* output, which we really don't want here. What we need is a *non-inverting* step-up/step-down topology. Looking around, a suitable candidate for this is the 'SEPIC' (single ended primary inductance converter) topology. See *Figure 4-1*. It is best visualized as *composite* topology — a boost stage followed by a buck cell. Though this "boost-buck"

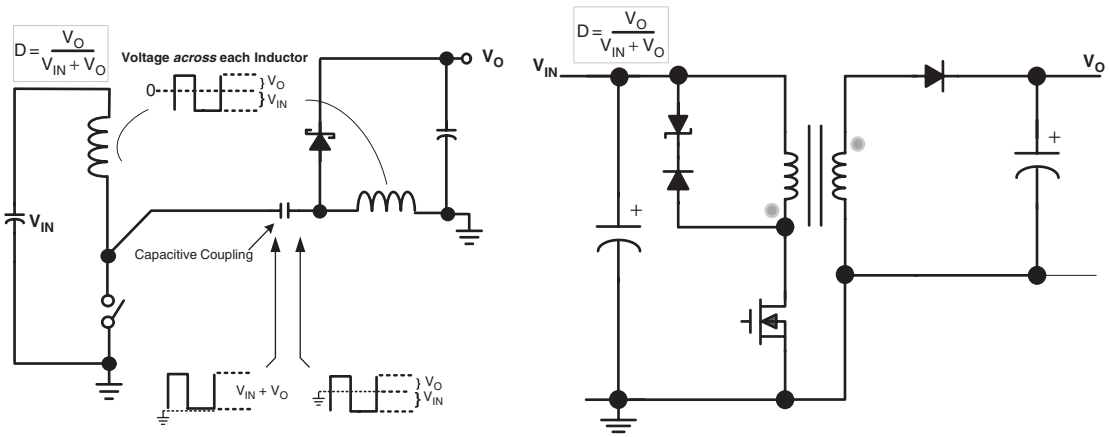


Figure 4-1: Positive-to-Positive Step-up/Step-down Converters

combination needs only one switch, it requires an additional inductor, and also entails significantly more design complexity. We may therefore wish to consider a derivative (or variant) of the conventional buck-boost topology, but with the inductor replaced by a *transformer*. In effect, what we are doing is — we are first separating (isolating) the input from the output, and then reconnecting the windings of the transformer in an appropriate manner so as to correct for the inversion. Thus we get a *non-inverting* or 'non-isolated transformer-based buck-boost' — sometimes simply called a 'flyback' topology.

Question 14: It is much easier to find "off-the-shelf" inductors. So why is a transformer-based buck-boost even worth considering?

Answer: It is true that most designers prefer the convenience of off-the-shelf components, rather than custom-designed components (like transformers). However, *high-power* off-the shelf inductors often come with *two identical windings* wound in parallel (on the same core), (though that may not be immediately apparent just by looking at the datasheet). Further, the ends of these two windings are sometimes completely separated from each other (no galvanic connection between the windings). The reason for this may be that from a production standpoint, it doesn't make sense to try and solder too many copper strands on to a single pin/termination. So the intention here is that the two windings will be eventually connected to each other on the PCB itself. But sometimes, the intention of leaving separate windings on an inductor is to allow *flexibility* for the two windings to be connected to each other either in *series*, or in *parallel*, as desired. So for example, if we place the winding in series, that would reduce the current rating of the inductor, but we would get a much higher inductance. If in parallel, the inductance would come down, but the current rating would increase. However, in low-voltage applications, where safety isolation is not a concern, we can also exploit this inductor structure and use it as a 1:1 transformer. And that would be very helpful in correcting the polarity inversion of the buck-boost. In other words, an off-the-shelf inductor could now serve as a 1:1 transformer!

Question 15: In an inductor with split windings (1:1), how exactly does its current rating and its inductance change as we go from a parallel configuration to a series configuration?

Answer: Suppose each winding has 10 turns and a DCR (dc resistance) of 1 Ω. So if it is used in parallel configuration, we still have 10 turns, but the effective DCR is 1 Ω in parallel with 1 Ω, i.e. 0.5 Ω. When a series configuration is used, we get 2 Ω and 20 turns. We also know that inductance depends on the *square* of the number of turns. So that goes up four times.

What about the current rating? This is largely determined by the amount of heat dissipation the inductor can tolerate. But its thermal resistance (in degC/W) is not determined by the winding configuration, rather by the exposed area of the inductor, and other physical characteristics. Therefore, whether in series or in parallel configuration, we have to maintain the same total I^2R loss. For example, suppose we call the current rating in parallel as "I_P",

and in series "I_S," then as per the DCR in our above numerical example, we get

$$I_P^2 \times 0.5 = I_S^2 \times 2$$

so

$$I_P = 2 \times I_S$$

Therefore, in going from a parallel to a series configuration, the inductance will *quadruple* and the current rating will *halve*.

What happens to the B-field? Don't we have to consider the possibility of saturation here? Well, B is proportional to LI/N (see *Chapter 2*, "DC-DC Converter Design and Magnetics"). So if inductance quadruples, I halves, and N doubles, the B-field is unchanged!

Question 16: Is there any difference between the terms "buck-boost" and "flyback"?

Answer: The answer to that may well depend on whom you ask! These terms are often used interchangeably in the industry. However, generally, most people prefer to call the conventional inductor-based version a (true) "buck-boost," whereas its transformer-based version, isolated or non-isolated, is called a "flyback."

Question 17: When and why do we need isolation? And how do we go about achieving it?

Answer: We must recognize that a (transformer-based) flyback topology may or may not provide us with isolation. Isolation is certainly a natural advantage accruing from the use of a transformer. But to preserve isolation, we must ensure that *all* the circuitry connected to the switch side of the transformer ('primary side') is kept completely independent from *all* the circuitry sitting on the output side ('secondary side'). See *Figure 1-1* in *Chapter 1*.

So for example, if in our attempt to correct the polarity inversion of a buck-boost we *make a connection* between the primary and secondary windings of the transformer, we lose isolation.

Further, to maintain isolation, besides making no galvanic connection between the power stages on either side of the transformer, we must not make any *signal-level* inter-connections either. So we must carry the feedback signal (or any fault information) from the output side to the IC, via one or more 'optocouplers.' The optocoupler manages to preserve primary-to-secondary voltage isolation, but allows signal-level information to pass through. It works by first converting the secondary-side signal into radiation by means of an "led" (light emitting diode), beaming it over to the primary side onto a photo-transistor, and thereby converting the signal back into electrical impulses (all this happening within the package of the device itself).

In high-voltage applications (e.g. off-line power supplies), it may in fact be required by law, to provide electrical isolation between a *hazardous* input voltage level and *user-accessible* ("safe") output terminals of the power supply. Therefore there is a "primary ground" at the input side of the transformer, and a separate "secondary ground" on the output side. The latter is then to the ground of the system, and usually also to the earthed metal enclosure.

Question 18: In an off-line power supply, are the primary and secondary sides really *completely* isolated?

Answer: It is interesting to note that safety regulations specify a certain physical spacing that must be maintained between the primary and secondary sides — in terms of the RMS of the *voltage differential* between them. The question arises — how do we define a voltage *difference* between the two sides of a transformer that are supposedly separate? What is the reference level to compare their voltages?

In fact they do share a connection! As mentioned, the secondary side ground is usually the system ground, and it connects to the metal enclosure and/or to the ground wire of the mains supply ("earth" or "safety ground"). But further down the mains distribution network, the safety ground wire is connected to the "neutral" wire of the supply. And we know that this neutral wire comes back into the primary side. So in effect, we have established a connection between the primary and secondary sides. It does not cause the user any problem, because he or she is also connected to earth. Therefore, the earth potential forms the reference level to establish the voltage difference across the safety transformer, and to thereby fix the primary-to-secondary spacing, and also the breakdown rating of any primary-to-secondary insulation. See Mathcad file in accompanying CD-ROM.

Note that in some portable equipment, only a two-wire ac cord is used to connect it to the mains supply. But the spacing requirement is still virtually unchanged, since a user can touch accessible parts on the secondary side and complete the connection through the earth ground.

Question 19: From the standpoint of an actual power supply design procedure, what is the most fundamental difference between the three topologies that must be kept in our minds?

Answer: In a buck, the average inductor current ("I_L") is equal to the load current ("I_O"), that is, $I_L = I_O$. But in a boost and a buck-boost, this average current is equal to $I_O/(1 - D)$. Therefore, in the latter two topologies, the inductor current is *a function of D* (duty cycle) — and therefore indirectly a *function of the input voltage* too (for a given output).

Question 20: In the three basic topologies, how does the duty cycle change with respect to input voltage?

Answer: For *all* topologies, a high D corresponds to a low input voltage, and a low D to a high input.

Question 21: What do we mean by the "peak current" of a dc-dc converter?

Answer: In any dc-dc converter, the terms "peak *inductor* current," "peak *switch* current," and "peak *diode* current" are all the same — referred to simply as the peak current 'I_{PK}' (of the converter).

Question 22: What are the key parameters of an off-the-shelf inductor that we need to consider?

Answer: The inductance of an inductor (along with the switching frequency and duty cycle) determines the *peak* current, whereas the *average* inductor current is determined by the topology itself (and the specific application conditions — the duty cycle and load current). For a given application, if we decrease the inductance, the inductor current waveform becomes more "peaky," increasing the peak currents in the switch and diode too (also in the capacitors). Therefore, a typical converter design should start by first estimating the optimum *inductance* so as to *avoid saturating the inductor*. That is the most basic concern in designing/picking an inductor.

However, *inductance* by itself doesn't fully describe an inductor. In theory, by choosing a very thin wire gauge for example, we may be able to achieve almost any inductance on a given core, just by winding the appropriate number of turns. But the current that the inductor will be able to handle without saturating is still in question, because it is not just the current, but *the product of the current and the number of turns* ('ampere-turns') that determine the magnetic field present in the inductor core — which in turn determines whether the inductor is saturating or not. Therefore, we need to look out for an inductor with the right inductance and also the required *energy handling capability*, usually expressed in μJ (microJoules). This must be greater than or equal to the energy it needs to store in the application, $\frac{1}{2} \times LI_{PK}{}^2$. Note that the "L" used in this equation carries with it information about the number of turns too, since $L \propto N^2$, where N is the number of turns.

Question 23: What really determines the *current rating* of an inductor?

Answer: There are two limiting factors here. One is the heat developed (I^2R losses), which we should ensure is not excessive (usually 50°C or less). The second is the magnetic field it can withstand without saturating. So most ferrites allow a maximum B-field of about 3000 gauss before saturation starts.

Question 24: Does the maximum allowable B-field depend on the air gap used?

Answer: When designing (gapped) transformers, we need to remember that first, the B-field present within the core material (e.g. ferrite) is the same as the B-field in the air gap. It does not change. Second, though by changing the air gap we can end up decreasing the existing B-field, the *maximum allowable* B-field depends *only on the core material* used — it remains fixed, for example, at about 3000 gauss for ferrites. Note that the H-field is defined as

$H = B/\mu$, where μ is the permeability of the material. So since the permeability of ferrite is much higher than that of air, and since the B-field is the same in both, therefore the H-field is much lower in the ferrite than in the air gap.

Question 25: Why is it commonly stated that in a flyback transformer, the "air gap carries most of the stored magnetic energy"?

Answer: We can intuitively accept the fact that the energy stored is proportional to the volume of the magnetic material. And because of that, we also tend to think the ferrite must be carrying most of the energy, since it occupies the maximum volume — the amount of air enclosed between the ends of the ferrite being very small. However, the stored energy is also proportional to $B \times H$, and since the H-field in the gap is so much larger, it ends up storing typically two-thirds of the total energy, despite its much smaller volume.

Question 26: If air carries most of the stored energy, why do we even need the ferrite?

Answer: An air-cored coil would seem perfect as an inductor, especially since it would never saturate. However, the number of turns required to produce a given inductance would be impractically large, and so we would get unacceptable copper losses. Further, since there is nothing to "channel" (constrain) the flux lines, the air-cored inductor would spew electromagnetic interference (EMI) everywhere.

The ferrite is useful, because it is the very *means* by which we can create such high magnetic fields in the first place — without an excessive number of turns. It also provides us the "channel" for flux lines that we had been looking for.

Question 27: What is the basic design rule for calculating inductance for all the topologies?

Answer: To reduce stresses at various points inside a power supply, and also to generally reduce the overall size of its components, a 'current ripple ratio' ('*r*') of about 0.4 is considered to be a good compromise for any topology, at any switching frequency.

"*r*" is the ratio $\Delta I/I_L$, where ΔI is the swing in the current, and I_L is the average inductor current (center of the swing ΔI). An *r* of 0.4 is the same as $r = 40\%$, or $r = \pm 20\%$. This means that the peak inductor current is 20% above its average value (its trough being 20% below).

To determine the corresponding inductance we use the definition $r = \Delta I/I_L$, along with the inductor equation, to get

$$V_{ON} = L\frac{\Delta I}{\Delta t} = L\frac{I_L \times r}{D/f}$$

solving

$$L = \frac{V_{ON} \times D}{I_L \times r \times f}$$

This gives us the inductance in Henries, when f is in Hz. Note that V_{ON} is the voltage across the inductor when the switch is ON. It is therefore equal to $V_{IN} - V_O$ for a buck, and V_{IN} for a boost and a buck-boost. Also, I_L is the average inductor current, equal to I_O for a buck, and $I_O/(1 - D)$ for a boost and a buck-boost.

Question 28: What is a 'forward converter'?

Answer: Just as the isolated flyback is a derivative of the buck-boost topology, the forward converter is the isolated version (or derivative) of the buck topology. It too uses a transformer (and optocoupler) for providing the required isolation in high-voltage applications. Whereas the flyback is typically suited for output powers of about 75 W or less, the forward converter can go much higher.

The simplest version of the forward converter uses only one transistor (switch), and is thus often called "single-ended." But there are variants of the single-ended forward converter with either two or four switches. So whereas the simple forward converter is suited only up to about 300 W of power, we can use the 'double-switch forward' to get up to about 500 W. Thereafter, the half-bridge, push-pull, and full-bridge topologies can be exploited for even higher powers (see *Figure 4-2*). But note that all of the above topologies are essentially 'buck-derived' topologies.

Question 29: How can we tell whether a given topology is "buck-derived" or not?

Answer: The simplest way to do that is to remember that only the buck has a true LC filter at its output.

Question 30: Which end of a given input voltage range V_{INMIN} to V_{INMAX} should we pick for starting a design of a buck, a boost, or a buck-boost converter?

Answer: Since the average inductor current for both the boost and buck-boost increases as D increases ($I_L = I_O/(1 - D)$) — the design of boost and buck-boost inductors must be validated at the *lower* end of the given input range, that is, at V_{INMIN} — since that is where we get the highest (average and peak) inductor current. We always need to ensure that any inductor can handle the maximum peak current of the application without saturating. For a buck, the *average* inductor current is independent of the input or output voltage. However, observing that its *peak* current increases at higher input voltages, it is preferable to design or select a buck inductor at the *upper* end of the given input range, that is, at V_{INMAX}.

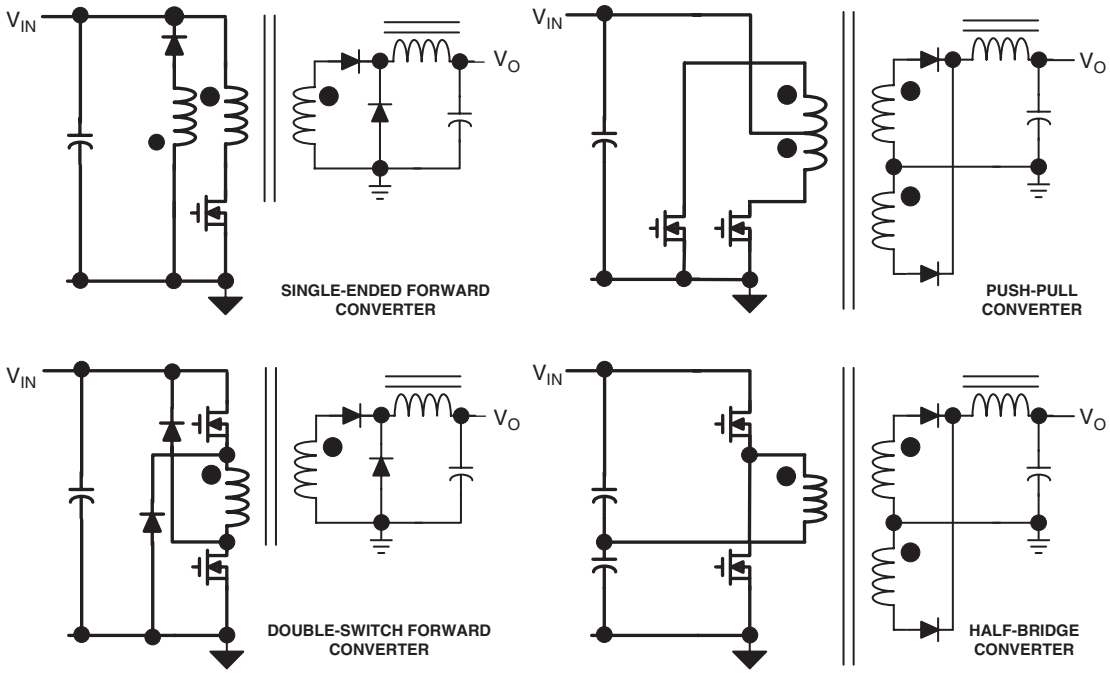


Figure 4-2: Various Buck-derived Topologies

Question 31: Why are the equations for the *average inductor current* of a boost and a buck-boost exactly the same, and why is that equation so different from that of a buck?

Answer: In a buck, energy continues to flow into the load (*via the inductor*) during the entire switching cycle (during the switch on-time *and* off-time). Therefore, the average inductor current must be equal to the load current, that is, $I_L = I_O$.

Note that capacitors contribute nothing to *average* current flow, because, in steady state, just as the voltseconds across an inductor averages out to zero at the end of each cycle, the charge in a capacitor does likewise (charge is the integral of current over time, and has the units Amperes-seconds). If that did not happen, the capacitor would keep charging up (or discharging) on an average, until it reaches a steady state.

However, in a boost or buck-boost, energy flows into the output *only* during the off-time. And it can only be coming *via the diode*. So the average diode current must be equal to the load current. By simple arithmetic, since the average diode current calculated over the full cycle is equal to $I_L \times (1 - D)$, equating this to the load current I_O gives us $I_L = I_O/(1 - D)$ for *both* the boost and the buck-boost.

Question 32: What is the average output current (i.e. the load current) equal to for the three topologies?

Answer: This is simply the converse of the previous question. For the buck, the average output current equals the average inductor current. For the boost and buck-boost, it is equal to the average diode current.

Question 33: What is the average input current equal to for the three topologies?

Answer: In a ***buck***, the input current flows *only* through the switch. It stops when the switch turns OFF. Therefore, *the average input current must be equal to the average switch current*. To calculate the average of the switch current, we know it is ON for a fraction D (duty cycle) of the switching cycle, during which time it has an average value (center of ramp) equal to the average inductor current, which in turn is equal to the load current for a buck. Therefore the arithmetic average of the switch current must be $D \times I_O$, and this must be equal to the input current I_{IN}. We can also do a check in terms of the input and output power

$$P_{IN} = V_{IN} \times I_{IN} = V_{IN} \times D \times I_O = V_{IN} \times \frac{V_O}{V_{IN}} \times I_O = V_O \times I_O = P_O$$

We therefore get input power equal to the output power — as expected, since the simple duty cycle equation used above ignored the switch and diode drops, and thus implicitly assumed no wastage of energy, that is, an efficiency of 100%.

Similarly, the input current of a ***boost*** converter flows through the inductor at all times. So the *average input current is equal to the average inductor current* — which we know is $I_O/(1 - D)$ for the boost. Let us again do a check in terms of power

$$P_{IN} = V_{IN} \times I_{IN} = V_{IN} \times \frac{I_O}{1 - D} = V_{IN} \times \frac{I_O}{1 - \dfrac{V_O - V_{IN}}{V_O}} = V_O \times I_O = P_O$$

Coming to the ***buck-boost***, the situation is not so clear at first sight. The input current flows into the inductor when the switch is ON, but when the switch turns OFF, though the inductor current continues to flow, its path does *not* include the input. So the only conclusion we can make here is that the *average input current is equal to the average switch current*. Since the center of the switch current ramp is $I_O/(1 - D)$, its arithmetic average is $D \times I_O/(1 - D)$. And this is the average input current. Let us check this out:

$$P_{IN} = V_{IN} \times I_{IN} = V_{IN} \times \frac{D \times I_O}{1 - D} = V_{IN} \times \frac{\dfrac{V_O}{V_{IN} + V_O} \times I_O}{1 - \dfrac{V_O}{V_{IN} + V_O}} = V_O \times I_O = P_O$$

We get $P_{IN} = P_O$ as expected.

Question 34: How is the average inductor current related to the input and/or output currents for the three topologies?

Answer: For the ***buck***, we know that average inductor current is equal to the output current, that is, $I_L = I_O$. For the ***boost*** we know it is equal to the input current, that is, $I_L = I_{IN}$. But for the ***buck-boost*** it is equal to the *sum* of the (average) input current and the output current. Let us check this assertion out

$$I_{IN} + I_O = \frac{D \times I_O}{1 - D} + I_O = I_O \times \left(\frac{D}{1 - D} + 1 \right) = \frac{I_O}{1 - D} = I_L$$

It is thus proved. See *Table 4-1* for a summary of similar relationships.

Table 4-1: Summary of relationships of currents for the three topologies

Average Values	Buck	Boost	Buck-Boost
I_L	I_O	$I_O/(1 - D)$	$I_O/(1 - D)$
I_L	I_{IN}/D	I_{IN}	I_{IN}/D
I_L	I_O	I_{IN}	$I_{IN} + I_O$
I_D	$I_O - I_{IN}$	I_O	I_O
I_D	$I_O(1 - D)$	I_O	I_O
I_D	$I_{IN}(1 - D)/D$	$I_{IN}(1 - D)$	$I_{IN}D/(1 - D)$
I_{SW}	I_{IN}	$I_{IN} - I_O$	I_{IN}
I_{SW}	$I_O D$	$I_O D/(1 - D)$	$I_O D/(1 - D)$
I_{SW}	I_{IN}	$I_{IN} D$	I_{IN}
I_O	I_L	I_D	I_D
I_{IN}	I_{SW}	I_L	I_{SW}

Question 35: Why are most buck ICs *not* designed to have a duty cycle of 100%?

Answer: One of the reasons for limiting D_{MAX} to less than 100% is specific to *synchronous* buck regulators (*Figure 4-3*) — when it utilizes a technique called 'low-side current sensing.'

In "low-side current sensing," to save the expense of a separate low-resistance sense resistor, the R_{DS} of the "low-side mosfet" (the one across the "optional" diode in *Figure 4-3*) is often used for sensing the current. The voltage drop across this mosfet is measured, and so if we know its R_{DS}, the current through it is also known by Ohm's law. It becomes obvious that in fact for any low-side current sense technique, we need to turn the high-side mosfet OFF, and thereby force the inductor current into the freewheeling path, so we can measure the current therein. That means we need to set the maximum duty cycle to less than 100%.

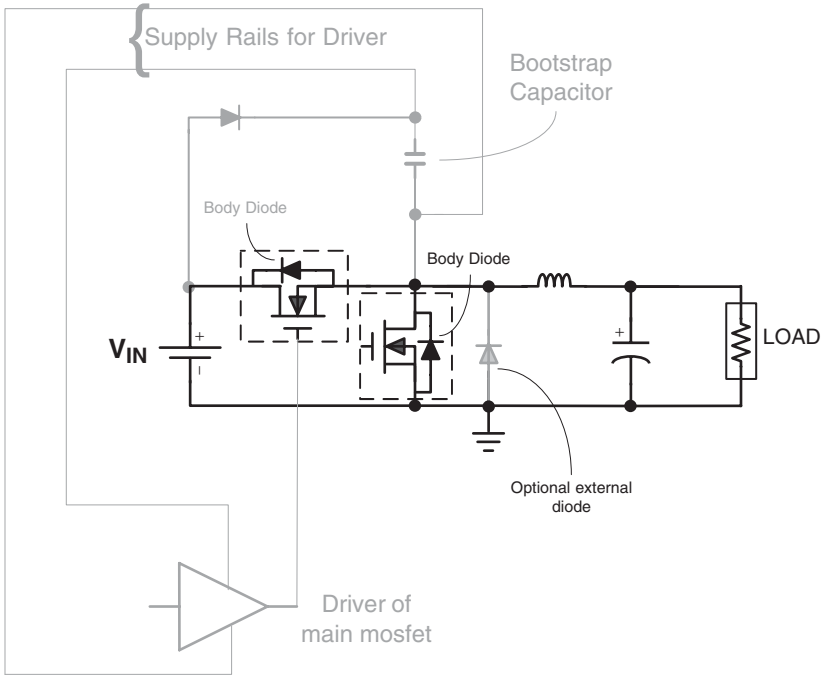


Figure 4-3: Synchronous Buck Regulator with Bootstrap Circuit

Another reason for choosing $D_{MAX} < 100\%$ comes from the use of *n-channel mosfets* in any (positive-to-positive) buck regulators. Unlike an npn transistor, an n-channel mosfet's gate terminal has to be taken several volts *above* its source terminal to turn it ON fully. So to keep the switch ON, when the mosfet conducts, we need to drive its gate a few Volts higher than the input rail. But such a rail is not available! The only way out is to *create* such a rail — by means of a circuit that can *pump* the input rail higher as required. This circuit is called the 'bootstrap circuit,' as shown in *Figure 4-3*.

But to work, the bootstrap circuit demands we turn the switch OFF momentarily, because that is when the switching node goes low and the 'bootstrap capacitor' gets charged up to V_{IN}. Later, when the switch turns ON, the switching node (lower terminal of the bootstrap capacitor) rises up to V_{IN}, and in the process, literally "drags" the upper terminal of the bootstrap capacitor to a voltage higher than V_{IN} (by an amount equal to V_{IN}!) — that happens because no capacitor loses its charge spontaneously! Therefore, the reason for setting the maximum duty cycle to less than 100% is simply to allow a bootstrap circuit (if present) to work!

We will find that a bootstrap circuit is almost always present if an *n-channel* mosfet switch is used in a positive to positive (or just "positive") buck converter, or in a positive to negative

buck-boost, or in a negative to negative (or just "negative") boost. Further, by circuit symmetry we can show that it will also be required (though this time to create a drive rail *below ground*) when using a *p-channel* mosfet in a negative buck, or in a negative to positive buck-boost, or in a positive boost.

Here, we should also keep in mind that the n-channel mosfet is probably the most popular choice for switches, since it is more cost-effective as compared to p-channel mosfets with comparable drain-to-source on-resistance 'R_{DS}.' That is because n-channel devices require smaller die sizes (and packages). Since we also know that the ubiquitous positive buck topology requires a bootstrap circuit when using an n-channel mosfet switch, it becomes apparent why a good majority of buck ICs out there have maximum duty cycles of less than 100%.

Question 36: Why are boost and buck-boost ICs almost invariably designed *not* to have 100% duty cycle?

Answer: We should first be clear that the boost and buck-boost topologies are so similar in nature, that any IC meant for a boost topology can also be used for a buck-boost application, and vice versa. Therefore, such control ICs are generally marketed as being for *both* boost and buck-boost applications.

One of the common aspects of these two topologies is that in both of these, energy is built up in the inductor during the switch on-time, during which *none* passes to the output. It is *delivered to the load only when the switch turns OFF*. In other words, we have to turn the switch OFF to get any energy at all delivered to the output. Contrast this with a buck, in which the inductor, being in series with the load, delivers energy to the load even as it is being built up in the inductor itself (during the switch on-time). So in a buck, even if we have 100% duty cycle (i.e. switch is ON for a long time), we *will* get the output voltage to rise (smoothly). Subsequently, the feedback loop will command the duty cycle to decrease when the required output voltage is reached.

However, in the boost and buck-boost topologies, if we keep the switch ON permanently, we can *never* get the output to rise, because in these topologies, energy is delivered to the output *only* when the switch turns OFF. We can thus easily get into a "Catch 22" situation, where the controller "thinks" it is not doing enough to get the output to rise — and therefore continues to command maximum duty cycle. But with a maximum 100% duty cycle, that means *zero* off-time — so how can the output *ever* rise?! We can get trapped in this illogical mode for a long time, and the switch can be destroyed. Of course, we hope that the current limit circuit is designed well enough to eventually intervene, and turn the switch OFF before the switch destructs! But generally, it is considered inadvisable to run these two topologies at 100% duty cycle. The only known D = 100% buck-boost IC is the LM3478 from National.

Question 37: What are the 'primary' and 'secondary' sides of an off-line power supply?

Answer: Usually, the control IC drives the switch *directly*. Therefore the IC must be located at the input side of the isolation transformer — that is called the 'primary side'. The transformer windings that go to the output are therefore said to all lie on the 'secondary side.' Between these primary and secondary sides lies "no-man's land" — the '*isolation boundary*.' Safety norms regulate how strong or effective this boundary must be.

Question 38: In many off-line power supplies, we can see not one, but *two* optocouplers, usually sitting next to each other. Why?

Answer: The first optocoupler transmits *error* information from the output (secondary side) to the control IC (primary side). This closes the feedback loop, and tells the IC how much correction is required to regulate the output. This optocoupler is therefore often nicknamed the "regulation opto" or the "error opto." However, safety regulations for off-line power supplies also demand that no 'single-point failure' anywhere in the power supply produces a hazardous voltage on the output terminals. So if, for example, a critical component (or even a solder connection) within the normal feedback path fails, there would be no control left on the output, which could then rise to dangerous levels. To prevent this from happening, an *independent* 'overvoltage protection' (OVP) circuit is almost invariably required. This is usually tied to the output rail in parallel to the components of the regulation circuitry. This fault detector circuit also needs to send its sensed 'fault signal' to the IC through a *separate* path altogether, so that its functioning is not compromised in the event of failure of the feedback loop. So logically, we require an *independent* optocoupler — the "fault opto." Note that by the same logic, this optocoupler must eventually connect to the IC (and cause it to shut down) using a pin *other* than the one being used for feedback.

The reason why the two optocouplers are "sitting *next* to each other" is usually only for convenience in the PCB layout — because the isolation boundary needs to pass through these devices, and also through the transformer (see *Figure 1-1* in *Chapter 1*).

Question 39: To get safety approvals in multi-output off-line converters, do we need separate current limiting on each output?

Answer: Safety agencies not only regulate the *voltage* at the user-accessible outputs, but also the maximum *energy* that can be drawn from them under a fault condition. Primary-side current sensing can certainly limit the *total* energy delivered by the supply, but cannot limit the energy (or power) from each output individually. So for example, a 300 W converter (with appropriate primary-side current limiting) may have been originally designed for 5 V at the rate of 36 A and 12 V at the rate of 10 A. But what prevents us from trying to draw 25 A from the 12 V output alone (none from the 5 V)? To avoid running into problems like this during approvals, it is wise to design separate secondary-side current limiting circuits for

each output. We are allowed to make an exception if we are using an *integrated post-regulator* (like the 7805) on a given output, because such regulators have built-in current limiting. Note that any overcurrent fault signal can be "OR-ed" with the OVP signal, and communicated to the IC via the fault optocoupler.

Question 40: How do safety agencies typically test for single-point failures in off-line power supplies?

Answer: Any component can be shorted or opened by the safety agency during their testing. Even the possibility of a solder connection coming undone anywhere, or a bad 'via' between layers of a PCB would be taken into account. Any such single-point failure is expected to usually cause the power supply to simply shut down gracefully, or even fail catastrophically. That is fine, but in the process, no hazardous voltage is permitted to appear on the outputs, even for a moment.

Question 41: What is a synchronous buck topology?

Answer: In synchronous topologies, the freewheeling diode of the conventional buck topology is either replaced, or supplanted (in parallel) with an additional mosfet switch. See *Figure 4-3*. This new mosfet is called the "low-side mosfet" or the "synchronous mosfet," and the upper mosfet is now identified as being the "high-side mosfet" or the "control mosfet."

In steady state, the low-side mosfet is driven such that it is "inverted" or "complementary" with respect to the high-side mosfet. This means that whenever one of these switches is ON, the other is OFF, and vice versa — that is why this is called "synchronous" as opposed to "synchronized" which would imply both are running *in phase* (which is clearly unacceptable because that would constitute a dead short across the input). However, through all of this, the effective *switch* of the *switching* topology still remains the high-side mosfet. It is the one that effectively "leads" — dictating when to build up energy in the inductor, and when to force the inductor current to start freewheeling. The low-side mosfet basically just follows suit.

The essential difference from a conventional buck regulator is that the low-side mosfet in a synchronous regulator is designed to present a typical forward drop of only around 0.1 V or less to the freewheeling current, as compared to a Schottky catch diode which has a typical drop of around 0.5 V. This therefore reduces the conduction loss (in the freewheeling path) and enhances efficiency.

In principle, the low-side mosfet does not have any significant crossover loss because there is virtually no overlap between its V and I waveforms — it switches (changes state) only when the voltage across it is almost zero. Therefore, typically, the high-side mosfet is selected primarily on the basis of its high *switching speed* (low crossover loss), whereas the low-side mosfet is chosen primarily on the basis of its low *drain-to-source on-resistance*, 'R_{DS}' (low conduction loss).

One of the most notable features of the synchronous buck topology is that on decreasing the load, it does *not* enter discontinuous conduction mode as a diode-based (conventional) regulator would. That is because, unlike a bjt, the current can reverse its direction in a mosfet (i.e. it can flow from drain to source or from source to drain). So the inductor current at any given moment can become negative (flowing *away* from the load) — and therefore "continuous conduction mode" is maintained — even if the load current drops to zero (nothing connected across the output terminals of the converter) (see *Chapter 1).*

Question 42: In synchronous buck regulators, why do we sometimes use a Schottky diode in parallel to the low-side mosfet, and sometimes don't?

Answer: We indicated above that the low-side switch is deliberately driven in such a manner that it changes its state only when the voltage across it is very small. That simply implies that during turn-off (of the high-side mosfet), the low-side mosfet turns ON a few nanoseconds *later*. And during turn-on, the low-side mosfet turns OFF just a little before the high-side mosfet starts to conduct. By doing this, we are trying to achieve 'zero voltage (lossless) switching' (ZVS) in the low-side mosfet. We are also trying to prevent "cross-conduction" — in which both mosfets may conduct simultaneously for a short interval during the transition (which can cause a loss of efficiency at best, and possible switch destruction too). However, during this brief interval when both mosfets are simultaneously OFF (the "deadtime"), the inductor current still needs a path to follow. However, every mosfet contains an intrinsic 'body diode' within its structure that allows reverse current to pass through it even if we haven't turned it ON (see *Figure 4-3*). So this provides the necessary path for the inductor current. However, the body diode has a basic problem — it is a "bad diode." It does not switch fast, nor does it have a low forward drop. So often, for the sake of a couple of percentage points in improved efficiency, we may prefer not to depend on it, and use a "proper" diode (usually Schottky), strapped across the low-side mosfet in particular.

Question 43: Why do most synchronous buck regulators use a low-side mosfet with an *integrated* Schottky diode?

Answer: In theory, we could just select a Schottky diode and solder it directly across the low-side mosfet. But despite being physically present on the board, this diode may be serving no purpose at all! For example, to get the diode to take over the freewheeling current quickly from the low-side mosfet when the latter turns OFF requires a *good low-inductance* connection between the two. Otherwise, the current may still prefer the body diode — for the few nanoseconds it takes before the high-side mosfet turns ON. So this requires we pay great attention to the PCB layout. But unfortunately, even our best efforts in that direction may not be enough — because of the significant inductive impedance that even small PCB trace lengths and internal bond wires of the devices can present when we are talking about nanoseconds. The way out of this is to use a low-side mosfet with an *integrated* Schottky

diode; that is, within the same package as the mosfet. This greatly reduces the parasitic inductances between the low-side mosfet and the diode, and allows the current to quickly steer away from the low-side mosfet and into the parallel diode during the deadtime preceding the high-side turn-on.

Question 44: What limits our ability to switch a mosfet fast?

Answer: When talking about a switching device (transistor), as opposed to a converter, the time it spends in transit *between* states is referred to as its "switching speed." The ability to switch fast has several implications, including the obvious minimization of the V-I crossover losses. Modern mosfets, though considered very "fast" in comparison to bjts, nevertheless do *not* respond *instantly* when their drivers change state. That is because, first, the driver itself has a certain non-zero "pull-up" or "pull-down" resistance *through which* the drive current must flow and charge/discharge the internal *parasitic capacitances* of the mosfet, so as to cause it to change state. In the process, there is a certain delay involved. Second, even if our *external* resistances were zero, there still remain parasitic inductances associated with the PCB traces leading up from the gate drivers to the gates, that will also limit our ability to force a large gate current to turn the device ON or OFF quickly. And further, hypothetically, even if we do achieve zero *external* impedance in the gate section, there remain *internal impedances* within the package of the mosfet itself — *before* we can get to its parasitic capacitances (to charge or discharge them as desired). Part of this internal impedance is *inductive*, consisting of the bond wires leading from the pin to the die, and part of it is *resistive*. The latter could be of the order of several ohms in fact. All these factors come into play in determining the switching speed of the device.

Question 45: What is 'cross-conduction' in a synchronous stage?

Answer: Since a mosfet has a slight delay before it responds to its driver stage, though the square-wave driving signals to the high- and low-side mosfets might have no intended "overlap," in reality the mosfets might actually be conducting simultaneously for a short duration. That is called 'cross-conduction' or 'shoot-through.' Even if minimized, it is enough to impair overall efficiency by several percentage points since it creates a short across the input terminals (limited only by various intervening parasitics).

This situation is aggravated if the two mosfets have significant "mismatch" in their switching speeds. In fact, usually, the low-side mosfet is far more "sluggish" than the high-side mosfet. That is because the low-side mosfet is chosen primarily for its low forward resistance, 'R_{DS}.' But to achieve a low R_{DS}, a larger die-size is required, and this usually leads to higher internal parasitic capacitances, which end up limiting the switching speed.

Question 46: How can we try and avoid cross-conduction in a synchronous stage?

Answer: To avoid cross-conduction, a deliberate delay needs to be introduced between one mosfet turning ON and the other turning OFF. This is called the converter's or controller's

'deadtime.' Note that during this time, freewheeling current is maintained via the diode present across the low-side mosfet.

Question 47: What is 'adaptive dead-time'?

Answer: Techniques for implementing dead-time have evolved quite rapidly as outlined below.

- **First Generation (Fixed Delay)** — The first synchronous IC controllers had a *fixed* delay between the two gate drivers. This had the advantage of simplicity, but the set delay time had to be made long enough to cover the many possible applications of the part, and also to accommodate a wide range of possible mosfet choices by customers. The set delay had often to be further offset (made bigger) because of the rather wide manufacturing variations in its own value. However, whenever current is made to flow through the diode rather than the low-side mosfet, we incur higher conduction losses. These are clearly proportional to the amount of dead-time, so we don't want to set too large a fixed dead-time for all applications.
- **Second Generation (Adaptive Delay)** — Usually this is implemented as follows. The *gate voltage* of the low-side mosfet is monitored, to decide when to turn the high-side mosfet ON. When this voltage goes below a certain threshold, it is assumed that the low-side mosfet is OFF (a few nanoseconds of additional fixed delay may be included at this point), and then the high-side gate is driven high. To decide when to turn the low-side mosfet ON, we usually monitor the *switching node* in "real-time" and adapt to it. The reason for that is that after the high-side mosfet turns OFF, the switching node starts falling (in an effort to allow the low-side to take over the inductor current). Unfortunately, the rate at which it falls is not very predictable, as it depends on various undefined parasitics, and also the application conditions. Further, we also want to implement something close to zero-voltage switching, to minimize crossover losses in the low-side mosfet. Therefore, we need to wait a *varying* amount of time, until we have ascertained that the switching node has fallen below the threshold (before turning the low-side mosfet ON). So the adaptive technique allows "on-the-fly" delay adjustment for different mosfets and applications.
- **Third Generation (Predictive Gate Drive™ Technique)** — The whole purpose of adaptive switching is to intelligently switch with a delay just large enough to avoid significant cross-conduction and small enough so that the body-diode conduction time is minimized — and to be able to do that consistently, with a wide variety of mosfets. However, the "predictive" technique, introduced by Texas Instruments, is often seen by their competitors as "overkill." But for the sake of completeness it is mentioned here. Predictive Gate Drive™ technology samples and holds information from the *previous* switching cycle to "predict" the minimum delay time for the

next cycle. It works on the premise that the delay time required for the next switching cycle will be close to the requirements of the previous cycle. By using a digital control feedback system to detect body-diode conduction, this technology produces the precise timing signals necessary to operate very *near the threshold* of cross-conduction.

Question 48: What is low-side current sensing?

Answer: Historically, current sensing was most often done during the *on-time* of the switch. But nowadays, especially for synchronous buck regulators in low output voltage applications, the current is being sensed during the *off-time*.

One reason for that is that in certain mobile computing applications for example, a rather *extreme* down-conversion ratio is being required nowadays — say 28 V to 1 V at a minimum switching frequency of 300 kHz. We can calculate that this requires a duty cycle of $1/28 = 3.6\%$. At 300 kHz, the time period is 3.3 μs, and so the required (high-side) switch on-time is about $3.6 \times 3.3/100 = 0.12$ μs (i.e. 120 ns). At 600 kHz, this on-time falls to 60 ns, and at 1.2 MHz it is 30 ns. Ultimately, that just may not give enough time to turn ON the high-side mosfet fully, "de-glitch" the noise associated with its turn-on transition ('leading edge blanking'), and get the current limit circuit to sense the current fast enough.

Further, at very light loads we may want to be able to skip pulses *altogether*, so as to maximize efficiency (since switching losses go down whenever we skip pulses). But with high-side current sensing we are almost forced into turning the high-side mosfet ON every cycle — just to sense the current!

For such reasons, low-side current sensing is becoming increasingly popular. Sometimes, a current sense resistor may be placed in the freewheeling path for the purpose. However, since low-resistance resistors are expensive, the forward drop across the low-side mosfet is often used for the purpose.

Question 49: Why do some non-synchronous regulators go into an almost chaotic switching mode at very light loads?

Answer: As we decrease the load, conventional regulators operating in CCM (continuous conduction mode — see *Chapter 1*) enter discontinuous conduction mode (DCM). The onset of this is indicated by the fact that the duty cycle suddenly becomes a function of load — unlike a regulator operating in CCM, in which the duty cycle depends *only* on the input and output voltages (to a first order). As the load current is decreased further, the DCM duty cycle keeps decreasing, and eventually, many regulators will automatically enter a random pulse-skipping mode. That happens simply because at some point, the regulator just *cannot* decrease its on-time further, as is being demanded. So the energy it thereby puts out into the inductor every on-pulse starts exceeding the average energy (per pulse) requirement of the load. So its control section literally "gets confused," but nevertheless tries valiantly to

regulate by stating something like — "oops . . . that pulse was too wide (sorry, just couldn't help it), but let me cut back on delivering any pulses altogether for some time — hope to compensate for my actions."

But this chaotic control can pose a practical problem, especially when dealing with current-mode control (CMC). In CMC, usually the switch current is constantly monitored, and that information is used to produce the internal ramp for the pulse-width modulator (PWM) stage to work. So if the switch does not even turn ON for several cycles, there is no ramp either for the PWM to work off.

This chaotic mode is also a variable frequency mode of virtually unpredictable frequency spectrum and therefore unpredictable EMI and noise characteristics too. That is why *fixed-frequency* operation is usually preferred in commercial applications. And fixed frequency basically means no pulse-skipping!

The popular way to avoid this chaotic mode is to "pre-load" the converter, that is, place some resistors across its output terminals (on the PCB itself), so that the converter "thinks" there is some minimum load always present. In other words, we demand a little more energy than the minimum energy that the converter can deliver (before going chaotic).

Question 50: Why do we sometimes *want* to skip pulses at light loads?

Answer: In some applications, especially battery-powered application, the 'light-load efficiency' of a converter is of great concern. Conduction losses can always be decreased by using switches with low forward drops. Unfortunately, switching losses occur every time we actually switch. So the only way to reduce them is by *not* switching, if that is possible. A pulse-skipping mode, if properly implemented, will clearly improve the light-load efficiency.

Question 51: How can we implement *controlled* pulse-skipping in a synchronous buck topology, to further improve the efficiency at light loads?

Answer: In DCM, the duty cycle is a function of the load current. So on decreasing the load sufficiently, the duty cycle starts to "pinch off" (from its CCM value). And this eventually leads to pulse-skipping when the control runs into its minimum on-time limit. But as mentioned, this skip mode can be fairly chaotic, and also occurs only at extremely light loads. So one of the ways this is being handled nowadays is to *not* "allow" the DCM duty cycle to pinch off below 85% of the CCM pulse width. Therefore now more energy is pushed out into a single on-pulse than under normal DCM — and without waiting to run into the minimum on-time limits of the controller. However, now because of the much-bigger-than-required on-pulse, the control will skip even more cycles (for every on-pulse). Thereafter, at some point, the control will detect that the output voltage has fallen too much, and will command another big on-pulse. So this forces pulse-skipping in DCM, and thereby enhances the light-load efficiency by reducing the switching losses.

Question 52: How can we quickly damage a boost regulator?

Answer: The problem with a boost regulator is that as soon as we apply input power, a huge inrush current flows to charge up the output capacitor. Since the switch is not in series with it, we have no control over it either. So ideally, we should *delay* turning ON our switch until the output capacitor has reached the level of the input voltage (inrush stops). And for this, a soft-start function is highly desirable in a boost. However, if while the inrush is still in progress, we turn the switch ON, it will start diverting this inrush into the switch. The problem with that is in most controllers, the current limit may not even be working for the first 100 to 200 ns after turn-on — that being deliberately done to avoid falsely triggering ON the noise generated during the switch transition ("leading edge blanking"). So now the huge inrush current gets fully diverted into the switch, with virtually no control, possibly causing failure. One way out of that is to use a diode directly connected between the input supply rail and the output capacitor (cathode of this diode being at the positive terminal of the output capacitor). So the inrush current bypasses the inductor and boost diode altogether. However, we have to be careful about the surge current rating of this extra diode. It need not be a fast diode, since it "goes out of the picture" as soon as we start switching (gets reverse-biased permanently).

Note also, that a proper ON/OFF function cannot be implemented on a boost topology (as is). For that, an additional series transistor is required, to completely and effectively disconnect the output from the input.

CHAPTER

5

Conduction and Switching Losses

This Page Intentionally Left Blank

CHAPTER 5

Conduction and Switching Losses

As switching frequencies increase, it becomes of paramount importance to reduce the *switching losses* in the converter. These are the losses *associated* with the *transition* of the switch from its on-state to off-state, and back. The higher the switching frequency, the greater the number of times the switch changes state per second. Therefore, these losses are proportional to the switching frequency. Further, of these frequency-dependent loss terms, the most significant are usually those that take place within the switch itself. Therefore, understanding the underlying sequence of events in the switch during each transition, and thereby quantifying the losses associated with each of these events, has become a key expectation of any power supply designer.

In this chapter, we are going to focus mainly on the mosfet, since that is the most widely accepted "switch" in most high-frequency designs today. We will split its turn-on and turn-off transitions into small well-defined *subintervals*, and explain what happens in each of these. The associated design equations will also be presented. Note however, that as in most related literature, we too will be resorting to certain *simplifications*, since modeling the mosfet (and its interplay with the board that it is mounted on) is certainly not a trivial task, to say the least. As a result, it is possible that theoretical estimates can end up underestimating the actual switching losses by a large margin (typically 20 to 50%). The designer should keep that in mind, and may need to eventually incorporate some sort of a "fudge factor," to correspond with reality. However, in our analysis, we have included a "scaling factor" to try and minimize this error.

We will also show how to estimate *driver* requirements, and demonstrate the importance of correctly matching driver capability to the mosfet in a given application. That should ultimately help not only applications engineers to pick better mosfets for their applications, but also IC designers involved in the process of designing driver stages for target applications.

A cautionary note with regard to the terminology — in most of our switching analysis, what we are calling the "load" is the load as seen by the *transistor* — it is not the load of the *dc-dc converter* stage. Similarly the "input voltage" is only the voltage *across the mosfet* when it is OFF — it is *not* the input to the dc-dc converter stage. We will eventually make the required connections into the area of power conversion, but it should be clear that

initially at least, the discussion is more from *the standpoint of the mosfet*, not the topology that it may be a part of.

Switching a Resistive Load

Before we take up inductors, it is instructive to first understand what happens when we switch a *resistive* load.

For simplicity, we are considering an *ideal* situation. So we start with a "perfect" n-channel mosfet in *Figure 5-1*. It behaves in the following manner

- It has zero on-resistance.
- With zero gate-to-source voltage "Vgs" applied at its gate, it is completely non-conducting.
- As we raise the gate-to-source voltage *Vgs* slightly above ground, it starts conducting, and so a drain current 'Id' flows from the drain to the source terminal.
- The ratio of the drain current to the gate voltage is defined as the transconductance 'g' of the mosfet. It is expressed in 'mhos,' that is, *ohm* spelled backward. Nowadays however, mhos is being increasingly called *Siemens*, or 'S.'
- We are assuming that *g* is a *constant* — equal to 1 S for this particular mosfet. So for example, if we apply 1 V at the gate, the mosfet will pass 1 A. If we apply 2 V, it will pass 2 A, and so on.

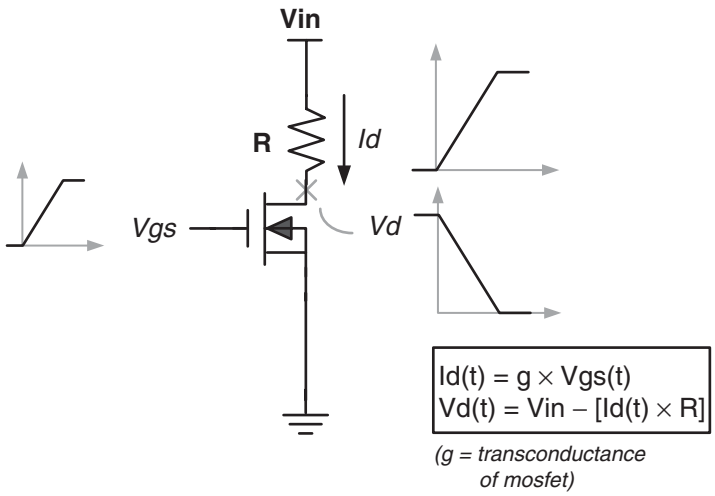


Figure 5-1: Switching a Resistive Load

The application circuit shown in *Figure 5-1* works as follows:

- The applied input voltage is 10 V.
- The external resistance (in series with the drain) is 1 Ω.
- The gate voltage is ramped up *linearly* with respect to time. So at t = 1 s it is 1 V, at t = 2 s it is at 2 V, at t = 3 s it is at 3 V, and so on.

The analysis proceeds as follows ("Vds" is the drain-to-source voltage at any given moment, "Vgs" is the gate-to-source voltage, and "Id" is the drain-to-source current):

- At t = 0, Vgs equals 0 V. Therefore, from the transconductance equation, Id is 0 A. So the drop across the 1 Ω resistor is 0 V (using Ohm's law). Therefore the voltage at the drain of the mosfet, 'Vds,' equals 10 V.
- At t = 1 s, Vgs equals 1. Therefore from the transconductance equation, Id is 1 A. So the drop across the 1 Ω resistor is 1 V (using Ohm's law). Therefore Vds equals $10 - 1 = 9$ V.
- At t = 2 s, Vgs equals 2. Therefore from the transconductance equation, Id is 2 A. So the drop across the 1 Ω resistor is 2 V (using Ohm's law). Therefore Vds equals $10 - 2 = 8$ V.

We proceed ramping up the gate voltage progressively in this manner. When 10 s have elapsed, Vgs is 10 V, Id is 10 A, and Vds is 0 V. *After* 10 s, no further change in Vds or Id can occur, even if Vgs is increased further.

> **Note:** In general, if the gate voltage is increased beyond what it takes to deliver a specified maximum load current, we say that in effect, we are applying "overdrive." This is usually considered wasteful in that sense, but in practice, overdrive helps reduce the on-resistance of the mosfet, and thereby decrease its conduction losses.

The maximum load current in our example is therefore 10 A, and is "Idmax" in *Figure 5-2*. If we plot the drain current and drain voltage with respect to time, we see that the *crossover time*, 'tcross,' is 10 s here. Note that this time is by definition the time for *both* the voltage and the current to complete their transitions.

The energy lost in the mosfet during the transition is

$$E = \int_0^{tcross} Vd(t)Id(t)dt \quad \text{Joules}$$

A conceptual point to keep in mind here is that in related literature, it is often stated (rather inaccurately as we will see) that the "area (jointly) enclosed by the voltage, current, and the time axis is the energy lost in the switch" (during the transition). This is the gray isosceles triangle in *Figure 5-2*. Half of this gray area has been *hatched*. We thus see that within the

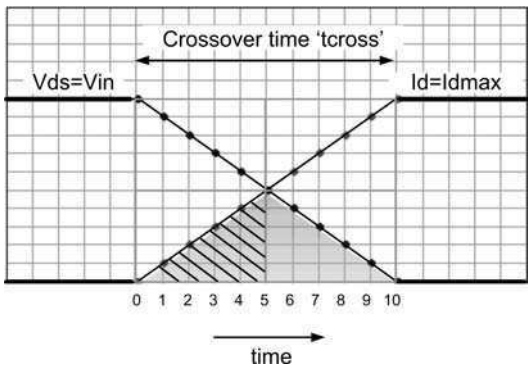


Figure 5-2: The Voltage and Current Waveforms when Switching a Resistive Load

"crossover interval rectangle," there are eight triangles (in all) with the same area as the hatched triangle. Therefore the total gray area is one-fourth the area of the crossover interval rectangle. So *if* the statement about energy being equal to the enclosed area is true, we would have gotten

$$\mathrm{E} = \frac{1}{4} \cdot \mathrm{Vin} \cdot \mathrm{Idmax} \cdot \mathrm{tcross\ Joules}$$

This is *not* correct. In fact, we would have reached the same unfortunate conclusion had we argued on the grounds that during the crossover duration, the *average* voltage is Vin/2 and the *average* current is Idmax/2, and therefore the average cross-product is equal to (Vin × Idmax)/4. This is fallacious too. In general,

$$A_{AVG} \times B_{AVG} \neq \left(A \times B\right)_{AVG}$$

So yes, this *could* in fact have turned out to be true, *if*, *while the voltage was falling, the current had remained fixed*, and vice versa. That is what happens with an *inductive* load, as we will soon see. However, in the case of a resistive load, both the voltage and the current change *simultaneously* during the crossover interval. We clearly need another (better) way to calculate the switching loss for the resistive case.

Let us compute the instantaneous cross-product Vds(t) × Id(t) at $\mathrm{t} = 1, 2, 3, 4 \ldots$ seconds. If we plot these points out, we get the bell-shaped curve shown in *Figure 5-3*. So, to get the energy lost during the crossover, we need to find the net area under *this* curve. But we can see that is not going to be easy, because this curve is rather oddly shaped. In fact, there is no other way than to carry out a formal integration/summation procedure. And for that we have

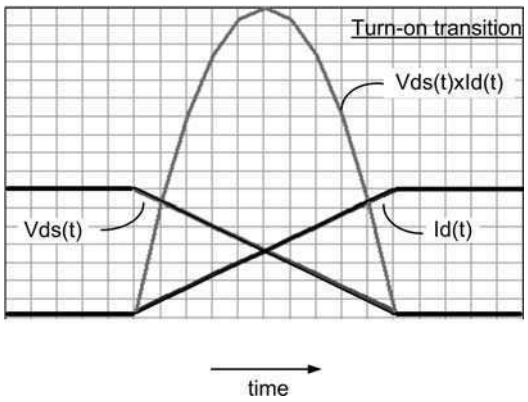


Figure 5-3: The Instantaneous Energy Dissipation Curve for Resistive Switching

to revert to the basic equations for voltage and the current (as presented in *Figure 5-1*). We then integrate their product over time, and we get

$$E = \frac{1}{6} \cdot Vin \cdot Idmax \cdot tcross \text{ Joules}$$

This is the correct result for the energy lost in the switch, during a *resistive* turn-on transition.

If we now turn the mosfet OFF in the same way (with the crossover time kept fixed), we will get exactly the same energy loss term again, though this time with the voltage *rising* and the current *falling*.

We can thus also conclude that if we switch *repetitively* at the rate of *fsw* Hz, the net dissipation, that is, total energy lost per unit time as heat, is equal to

$$Psw = \frac{1}{3} \cdot Vin \cdot Idmax \cdot \ tcross \cdot fsw \text{ Watts}$$

This is therefore the *switching loss* (in the switch) for the case of a resistive load.

> **Note:** Note that to be precise, this particular term more correctly should be called the 'crossover loss,' as was first pointed out in *Chapter 1*. The crossover loss (i.e. specifically attributable to the *V-I overlap*) is not necessarily the entire switching loss taking place in the switch, as we will see.

Now, suppose we had ramped *up* the gate voltage at a rate of 1 V per second as before, but ramped *down faster*, say, at the rate of 2 V per second. Then the turn-on time and the turn-off transition times would be different. So in that case we need to *split* up the crossover

loss 'Psw' as follows:

$$\begin{aligned}\text{Psw} &= \text{Pturnon} + \text{Pturnoff} \\ &= \frac{1}{6} \cdot \text{Vin} \cdot \text{Idmax} \cdot \ \text{tcross}_{\text{on}} \cdot \text{fsw} + \frac{1}{6} \cdot \text{Vin} \cdot \text{Idmax} \cdot \ \text{tcross}_{\text{off}} \cdot \text{fsw}\end{aligned}$$

where '$\text{tcross}_{\text{on}}$' and '$\text{tcross}_{\text{off}}$' are the crossover times during turn-on and turn-off respectively.

Now suppose the value of the external resistor was made larger, say 2 ohms instead of 1 ohm. Then the voltage at the drain would have swung from 10 V to 0 V in only 5 s. And by that time, the drain current would have reached only 5 A. The gate voltage would at that moment be only at 5 V. However, *no further change in Id is possible* (even if we increase Vgs further). Therefore, though the crossover interval has become half of what it was, *the rise time of the current is still equal to the fall time of the voltage* (i.e. 5 s). This is a characteristic *only* of *resistive* loads (since V = IR applies to them).

The rules of the game change considerably, when we have an inductive load. In fact the calculation becomes simpler — ironically because the simplicity (and predictability) of Ohm's law is lost.

Switching an Inductive Load

When we switch an inductive load (with a freewheeling path present of course!), we will get the waveforms shown in *Figure 5-4* (idealized). At first sight they may seem similar to the resistive load waveforms of *Figure 5-2*. But on closer examination, they are very different. In particular, we see that *when the current is swinging, the voltage remains fixed*, and *when the voltage is swinging, the current remains fixed.*

Let us calculate the crossover loss under these conditions. We can do a formal integration as before. But this time, we realize there is in fact an easy way out! Since *one* of the parameters (V or I) is fixed when the *other* is varying, we can now justifiably take the *average* value of the current, Idmax/2, and the *average* value of the voltage, Vin/2, to find the average cross-product. In this manner, we arrive at the energy lost (in Joules) during the turn-on transition

$$\begin{aligned}\text{E} &= \left[\frac{\text{Vin}}{2} \cdot \text{Idmax} \cdot \ \frac{\text{tcross}}{2}\right] + \left[\text{Vin} \cdot \frac{\text{Idmax}}{2} \cdot \frac{\text{tcross}}{2}\right] \\ &= \frac{1}{2} \cdot \text{Vin} \cdot \text{Idmax} \cdot \ \text{tcross}\end{aligned}$$

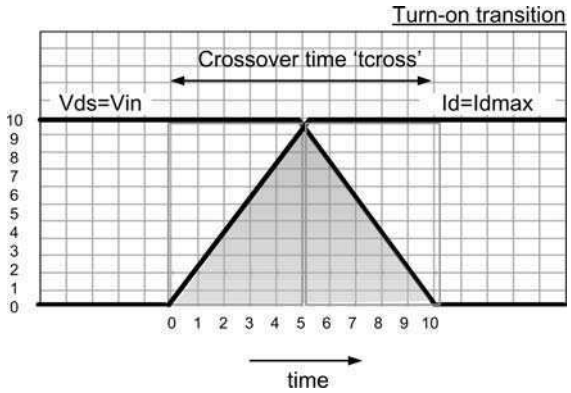


Figure 5-4: The Voltage and Current Waveforms when Switching an Inductive Load

Note that for the same reason as indicated above, we can now justifiably think in terms of the *area* enclosed. By simple geometry, the gray area in *Figure 5-4* is half the rectangular area, and so we get the same result as above.

We realize that our ability to avoid integration (and use simpler arguments to calculate the crossover loss) is just a piece of "good luck" here — specific to the case of an *inductive* load.

Finally, when we switch *repetitively*, the inductive switching loss is

$$\text{Psw} = \text{Vin} \cdot \text{Idmax} \cdot \text{tcross} \cdot \text{fsw Watts}$$

Note: We may superficially conclude that switching an inductive load leads to a dissipation three times greater than a resistive load. That is indeed true, but only *under the exact same conditions*. In reality, the value of Idmax is *fixed* for the case of a resistive load (depending on the value of the resistance used). But for an inductive load, the current can be virtually anything — there is no set "Idmax" as such anymore — it is whatever current that happens to be flowing through the inductor at the instant of switching (either just before or after).

A basic question still remains — *why* are the inductive waveforms so different from the resistive case? To answer that, we have go back to our previous analysis of the resistive load case. There we will see that we had invoked Ohm's law to find the voltage across the switch. But with an inductor, Ohm's law clearly does not apply. So to get the waveforms shown in *Figure 5-4*, we have to recollect something we learned in *Chapter 1* — when we turn the switch OFF, the inductor will create *whatever voltage is necessary to maintain the continuity of current through it.* Let us now show this principle at work in an actual buck converter, for example (see *Figure 5-5*).

In *Figure 5-5*, we first consider the *turn-on* transition (on the left). Just prior to this, the diode is obviously carrying the full inductor current (circled "1"). Then the switch starts to turn ON, trying to share some of this inductor current (circled "2"). The diode current therefore must fall correspondingly (circled "3"). However, the important point is that while the switch

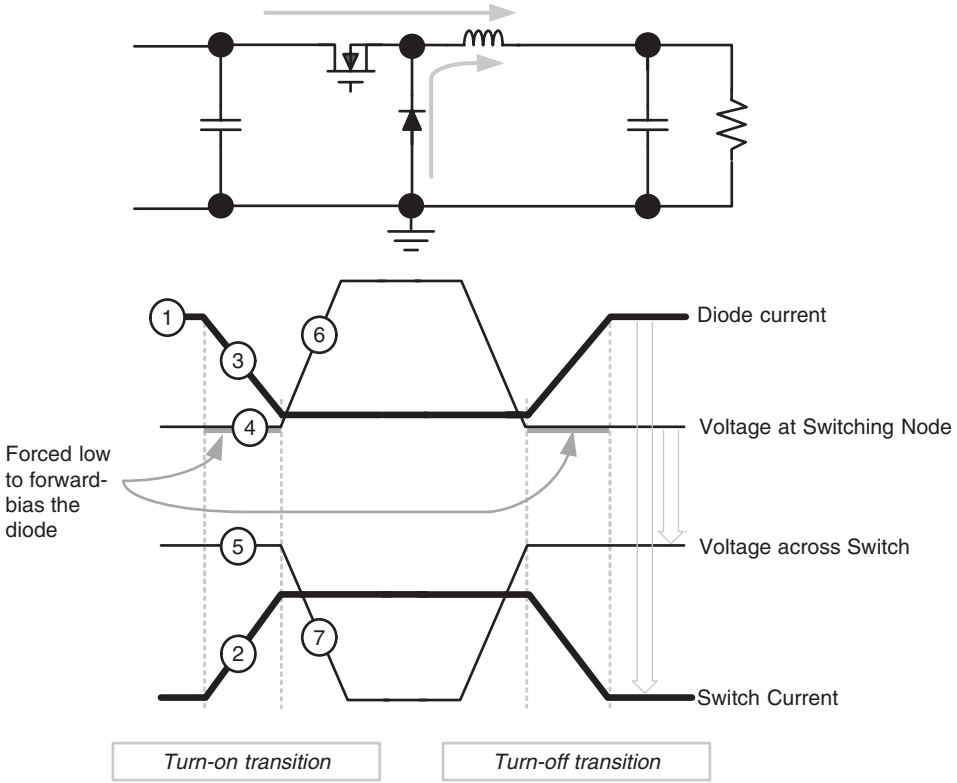


Figure 5-5: Analyzing the Transitions in a Buck Converter

current is still in transit, the diode has to be able to pass *some* current (the remainder, or leftover amount of the inductor current). But, to provide even *some* of the inductor current, the diode must remain *fully* forward-biased. Therefore, nature (i.e. induced voltage in this case) *forces* the voltage at the switching node to remain slightly *below ground* — so as to keep the anode of the diode about 0.5 V higher than the cathode (circled "4"). Then, by Kirchhoff's voltage law, the voltage across the switch *stays high* (circled "5"). Only finally, when the *entire* inductor current has shifted to the switch, does the diode "let go." With that, the switching node is released, and it flies up close to the input voltage (circled "6") — and so now, the voltage across the switch is allowed to fall (circled "7").

- We therefore see that *at turn-on, the voltage across the switch does not change until the current waveform has **completed** its transition*. We thus get a significant V-I overlap.

If we do a similar analysis for the turn-off transition (right side of *Figure 5-5*), we will see that for the switch current to start decreasing by even a small amount, the diode must first be "positioned" to take up *any* current coming its way. So the voltage at the switching node

must *first* fall close to zero, so as to forward-bias the diode. That also means the voltage across the switch must first transit fully, *before* the switch current is even allowed to decrease slightly (see *Figure 5-5*).

- We therefore see that *at turn-off, the current through the switch does not change until the voltage waveform has* ***completed*** *its transition*. We thus get a significant V-I overlap.

We see that the fundamental properties and behavior of an inductor, as described in *Chapter 1*, are ultimately responsible for the significant V-I overlap during crossover.

The same situation is present in the case of any switching topology. Therefore, *the switching loss equation presented earlier also applies to all topologies*. What we have to remember is, that in our equations, we are referring to the voltage *across* the switch (when it is OFF), and the current *through* it (when it is ON). In an actual converter, we will need to ultimately relate these V and I to the actual input/output rails and load current of the application. The procedure for that is described later.

Switching Losses and Conduction Loss

The underlying motivation for initiating *switching* in modern power conversion is often simplistically stated as follows — by switching the transistor, either the voltage across the transistor is close to zero, or the current through it is close to zero, and therefore the dissipation cross-product "V × I" is also almost zero. We have seen that during the transition, that doesn't really hold true anymore (the V-I *overlap*). Similarly, we should keep in mind that though the V × I losses are much closer to the ideal or "expected" value of zero when the switch is OFF, there are considerable losses when the switch is ON. That is because when the switch is OFF, it is *really* so — the *leakage current* through a modern semiconductor switch is almost negligible. However, when the switch is ON, the voltage across it is *not even close to zero* in many cases. One of the highest reported forward drops is in the "Topswitch®" (an integrated switcher IC meant for medium off-line flyback applications) — *over 15 V* (over rated current and temperature)! In general, there will remain a significant V × I loss term even after the inductor current has shifted entirely from the diode to the switch. This particular loss term is clearly the *conduction loss*, P_{COND} (of the switch). It can in fact be comparable to, or even greater than, the crossover loss.

However, unlike the crossover loss, the conduction loss is *not* frequency-dependent. It does depend on *duty-cycle,* but not on frequency. For example, suppose the duty cycle is 0.6; then in a measurement interval of say, one second, the *net* time spent by the switch in the ON-state is equal to 0.6 seconds. But we know that conduction loss is incurred *only* when the switch is ON. So in this case, it is equal to a × 0.6, where "a" is an arbitrary proportionality constant. Now suppose the frequency is doubled. Then the net time spent in

the on-state (in 1 second) *is still 0.6 seconds*. So the conduction loss remains a × 0.6. But now, suppose the duty cycle changes from 0.6 to 0.4 (the frequency can be even doubled in the process), the conduction loss is reduced to a × 0.4. So we realize that conduction loss can't possibly depend on frequency, only on duty cycle.

We can pose a rather philosophical question — why is it that the switching loss is frequency-dependent, but *not* the conduction loss? That is simply because the conduction loss *coincides with the interval in which power is being processed in the converter*. Therefore, as long as the *application* conditions do not change (duty cycle fixed, input and output power fixed), neither can the conduction loss.

The equation to calculate the conduction loss of a mosfet is simply

$$P_{COND} = I_{RMS}{}^2 \times Rds \text{ Watts}$$

where 'Rds' is the on-resistance of the mosfet. I_{RMS} is the RMS of the switch current waveform. It is equal to

$$I_{RMS} = I_O \times \sqrt{D \times \left(1 + \frac{r^2}{12}\right)} \quad \textit{(buck)}$$

$$I_{RMS} = \frac{I_O}{1 - D} \times \sqrt{D \times \left(1 + \frac{r^2}{12}\right)} \quad \textit{(boost and buck-boost)}$$

where I_O is now the load current of the *dc-dc converter* stage, and D is its duty cycle. Note that to a first approximation (current ripple ratio assumed very small), this is equal to

$$I_{RMS} \approx I_{DC} \times \sqrt{D} \quad \textit{(buck, boost, and buck-boost)}$$

where I_{DC} is the average inductor current and 'I_{RMS}' is the RMS of the switch current waveform.

The *diode conduction loss* is the other major conduction loss term in a power supply. It is equal to $V_D \times I_{D\_AVG}$, where V_D is the diode forward-drop. I_{DAVG} is the average current through the diode — equal to I_O for the boost and the buck-boost, and $I_O \times (1 - D)$ for the buck. It too is frequency-independent.

We realize that the way to reduce conduction losses is by lowering the forward-drops across the diode and switch. So we look for diodes with a low drop — like the Schottky diode. Similarly, we look for mosfets with a low on-resistance "Rds." However, there are compromises involved here. The leakage current in a Schottky diode can become significant as we try to choose diodes with very low drops. We can also run into significant body

capacitance, which will end up being more dissipative. Similarly, the speed at which the mosfet switches can be adversely affected as we try to reduce its Rds.

A Simplified Model of the Mosfet for Studying Inductive Switching Losses

In *Figure 5-6*, on the left, we have the basic (simplified) model of the mosfet. In particular, we observe that it has three parasitic capacitances — between its drain, source, and gate. These "small" *interelectrode capacitances* are the key to maximizing switcher efficiency, especially at higher switching frequencies. Their role in the switching transition needs to be understood clearly.

We have seen that the basic reason why we get *any* crossover loss in the first place is because there is an unavoidable V-I overlap during every switching transition. That overlap occurs because the inductor keeps trying to force current, and tries to create suitable conditions for that to happen seamlessly, as we switch. But the reason why this overlap lasts *as long* as it does is mainly because these three interelectrode capacitors are *demanding to be charged or discharged* (as the case may be) at every switching event — so that they can reach their new dc levels, commensurate with the altered state of the switch. So crudely stated, if these capacitances are "big," they take a longer time to charge or discharge, thus increasing the crossover (overlap) time. And that in turn increases the crossover loss. Further, since the charging and discharging paths of these capacitors often include the *gate resistor*, the value of the gate resistance also considerably impacts the transition time, and thereby the switching loss.

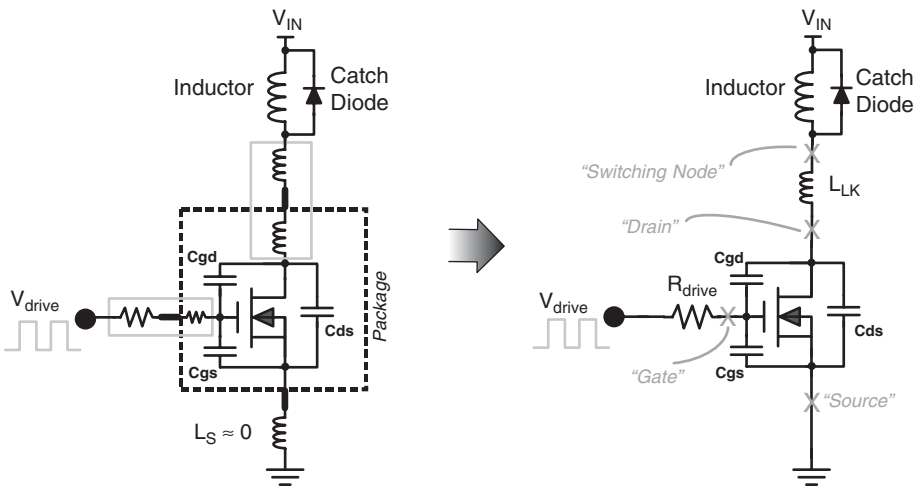


Figure 5-6: Simplified Model of Mosfet

On the right side of *Figure 5-6*, we have further simplified our simple model. So we have lumped the internal and external inductances present at the drain into a single leakage inductance "Llk." Note that we are ignoring any gate-to-source inductance, thus implicitly assuming the PCB layout is very good in this regard. We also lump the small resistor present internally inside the mosfet, along with the *external* gate resistor (if present), and the *driver resistance* (its internal pull-up or pull-down) — to give a single effective 'Rdrive,' or *drive resistance*.

Note that in *Figure 5-6*, the main inductor is "coupled" — because it has a freewheeling path available. But the leakage (or parasitic) inductance is "uncoupled," because it has no path to send forth its energy. It therefore expectedly "complains" — in the form of a voltage spike (whenever we try to change the current through it). However, in our analysis, we will be assuming this *leakage inductance is very small* (though not necessarily negligible either). We will find that this results in certain *artifacts* in the switching waveforms, which makes them appear slightly different, as compared to the idealized inductive switching waveforms shown in *Figure 5-4* and *Figure 5-5*. However, it turns out that these artifacts are mainly of academic interest (provided of course that Rdrive is "small"). In addition, the artifacts in question typically help *decrease* the crossover losses slightly. Therefore, the idealized waveforms are more "conservative" in that sense, and we would do well just sticking to them.

Turning our attention to the "circuit" shown in *Figure 5-6*, we should be clear that this circuit doesn't really *work!* We know from our discussions in *Chapter 1* that we can never hope to achieve a *steady state* without at least an *output capacitor* present — to charge up and thereby help stabilize the voltseconds across the inductor. So this circuit is clearly an *idealization* — it only helps us to perform a *paper-analysis* of a *particular switching transition.*

Note that ultimately, the switch cares only about the *voltage that appears across it* when it turns OFF, and the *current passing through it* when it is ON. That is why this simple circuit can be safely accepted as representative of what happens in *any* topology, *at the moment of transition*. For instance, we could take both the leakage and the main inductor in *Figure 5-6*, and place them on the *source*-side of the mosfet instead. As long as the gate drive is still well coupled to the source (i.e. no inductance *between* gate and source), nothing really changes. That is no surprise, because we know that if a certain component (or circuit block) "A" is in series with "B," we can always interchange their positions and make B in series with A, without changing a thing.

Finally, we should keep in mind that what we are calling the "drain" in our analysis is not necessarily the *pin* of the *package* (of the same name). Nor the switching node! The inductance Llk separates these points as indicated in *Figure 5-6*. Therefore, for example, though the switching node is necessarily clamped close to the "Vin" rail when the diode is freewheeling, the drain of the device may momentarily show a slightly different voltage (clearly equal to the voltage appearing across Llk).

The Parasitic Capacitances Expressed in an Alternate System

We will now progress to a detailed study of the inductive switching transitions of a mosfet. For that, we will be splitting up the turn-on and turn-off into several subintervals of interest. We will learn that *for most of these subintervals, the gate behaves as a simple input capacitance — that is being charged (or discharged)* through the resistor 'Rdrive.' The situation is identical to the simple RC circuit we discussed in *Chapter 1*. In effect, the gate is "blind" to what all may be happening between drain and source (on account of the transconductance of the mosfet).

If we look *into* the gate, from the viewpoint of the ac drive signal, the effective input charging capacitance is the parallel combination (arithmetic sum) of Cgs and Cgd. We are going to call this simply the *gate* or *input capacitance* 'Cg' in our discussion. So

$$\mathrm{Cg} = \mathrm{Cgs} + \mathrm{Cgd}$$

The *time constant* of the charging/discharge cycles of the gate is therefore

$$\mathrm{Tg} = \mathrm{Rdrive} \times \mathrm{Cg}$$

> **Note:** Here we seem to be indirectly suggesting that the drive resistance is the same for turn-on and turn-off. That need not be so. All the equations we will present can easily take any existing difference in the turn-on and turn-off drive resistances into account. So in general, we will have *different* crossover times for the turn-on and turn-off transitions. Also note, that in general, *within* a certain crossover interval (turn-on or turn-off), the actual time it takes for the *voltage* to transit need not be the same as the time the *current* takes (unlike the case of a resistive load).

An *alternative* system of writing the capacitances is in terms of the *effective* input, output, and reverse transfer capacitances — that is, *Ciss*, *Crss*, and *Coss* respectively. These are related to the interelectrode capacitances as follows

$$\mathrm{Ciss} = \mathrm{Cgs} + \mathrm{Cgd} \equiv \mathrm{Cg}$$
$$\mathrm{Coss} = \mathrm{Cds} + \mathrm{Cgd}$$
$$\mathrm{Crss} = \mathrm{Cgd}$$

So we can also write

$$\mathrm{Cgd} = \mathrm{Crss}$$
$$\mathrm{Cgs} = \mathrm{Ciss} - \mathrm{Crss}$$
$$\mathrm{Cds} = \mathrm{Coss} - \mathrm{Crss}$$

In most vendors' datasheets, we can usually find Ciss, Coss, and Crss under the section "typical performance curves." We will then see that these parasitic capacitances are a *function of voltage*. Clearly, that can significantly complicate any analysis. So as an approximation, *we are going to assume that the interelectrode capacitances are all constants*. We will consult the typical performance curves of the mosfet, and then pick the value of the capacitance c*orresponding to the voltage that appears across the mosfet* when it is OFF (in our given application). Later, we will show how to minimize this error, by the use of a certain "scaling factor."

Gate Threshold Voltage

The "perfect mosfet" we talked about earlier (*Figure 5-1*) started conducting the moment we raised the gate voltage above ground (i.e. source). But an actual mosfet has a certain *gate threshold voltage* 'Vt.' This is typically 1 to 3 Volts for 'logic-level' mosfets, and about 3 to 5 Volts for high-voltage mosfets. So basically, we have to exceed the stated threshold voltage to get the mosfet to conduct at all ("conduction" defined typically as a current in excess of 1 mA).

Because Vt is not zero, the definition of transconductance also needs to be modified slightly from

$$g = \frac{Id}{Vgs} \Rightarrow g = \frac{Id}{Vgs - Vt}$$

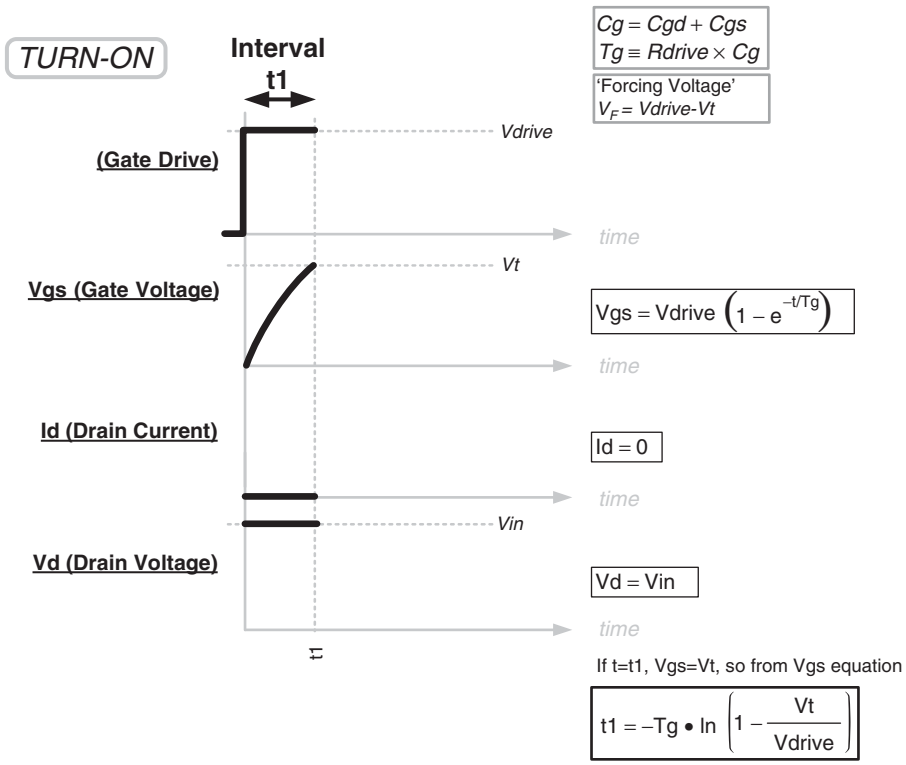
Note that in our analysis, we are making another simplifying assumption — that the transconductance too is a constant.

Finally, with all this background information, we can start looking closely at what actually happens during the turn-on and turn-off transitions.

The Turn-on Transition

We have divided this interval into *four* subintervals as detailed individually in *Figures 5-7* through *5-10*. For quick reference and ease of understanding, the relevant explanations and comments for each sub-interval are also provided within their respective figures.

Briefly, the interval t1 is just the time to get to the threshold Vt. During this time, we just have a simple RC charging circuit. In t2 also, the exponential rise continues, but this time, the drain current starts ramping up. But for all practical purposes, the gate doesn't "know" anything has changed, because the transconductance is fully responsible for the drain current (and further, there is no change in the drain voltage). But in t3, the diode is allowed to stop



Since Vgs is less than Vt, Drain Current is zero, and Drain Voltage is fixed at Vin.

Since Drain Voltage is fixed, there is almost no current injected into Gate (through Cgd) attributable to any variation of Vd.

There is a small Cgd-current due to the increasing Vgs. But this is factored in by the use of Cg=Cgd+Cgs instead of just Cgs in the time constant Tg.

During turn-on, in a switching power supply, the voltage at the switching node (Note: this is physically separated from Vd by leakage Llk) cannot change unless the inductor current completely changes-over from the free-wheeling diode back to the switch — because the diode must stay forward-biased if it is to carry any current at all. Note: diode drop is being ignored.

Figure 5-7: First Interval of Turn-on

conducting (since all the inductor current has by now shifted over into the switch). So now the drain voltage swings. But in doing so, *it injects a current through Cgd*. Note that this capacitance, despite being usually rather small, has probably the greatest effect on the crossover time — because of the fact that it *directly* injects current from a high switching voltage node (drain) on to the gate. Just prior to the interval t3, Cgd has a relatively high voltage across it. But when the switch is fully ON, the voltage across Cgd must decrease to its new final low value. Therefore, *during t3*, Cgd is essentially discharging. So the question is — *what is the path the Cgd discharge current takes*? We can analyze that as follows — having reached the gate, this discharge current has *two* choices — either to go through Cgs and/or through Rdrive. But the gate is already at the constant level of Vt + Io/g — that being

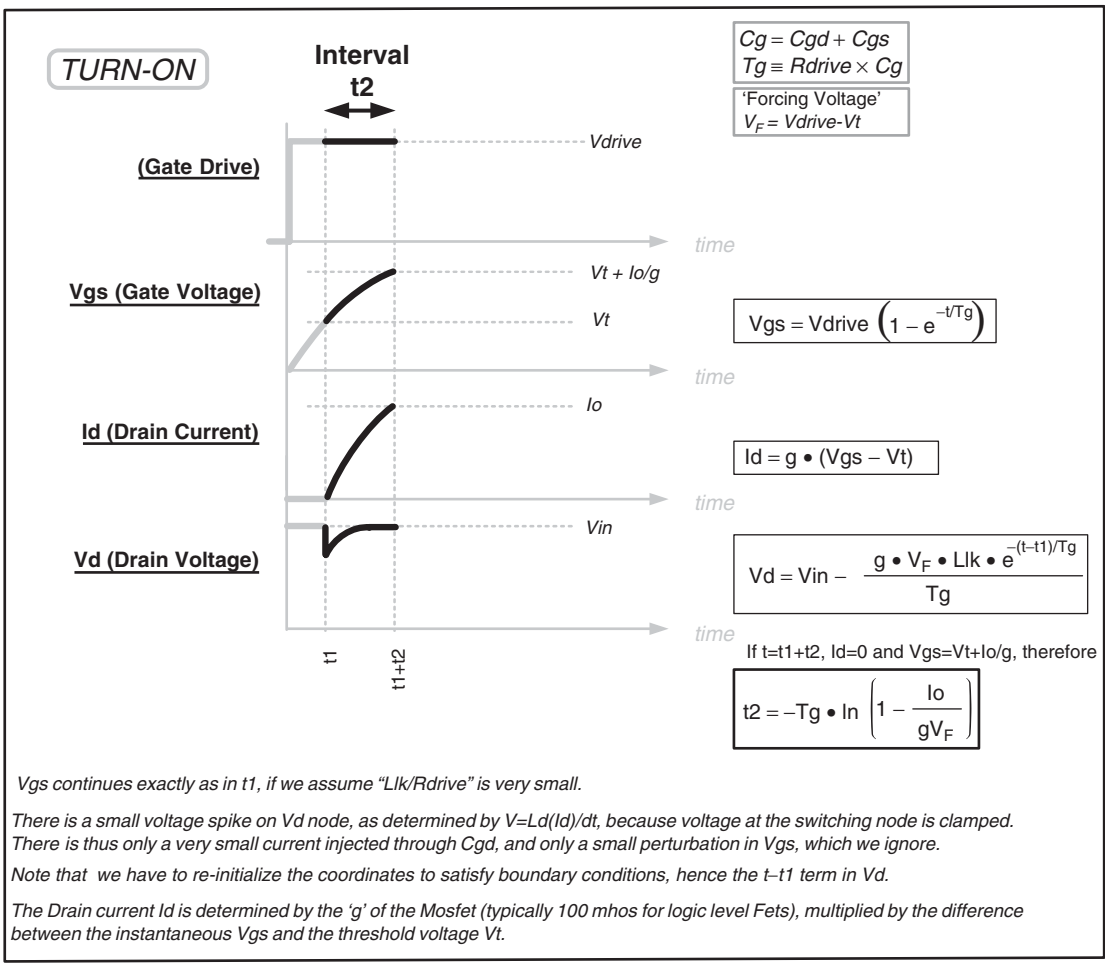
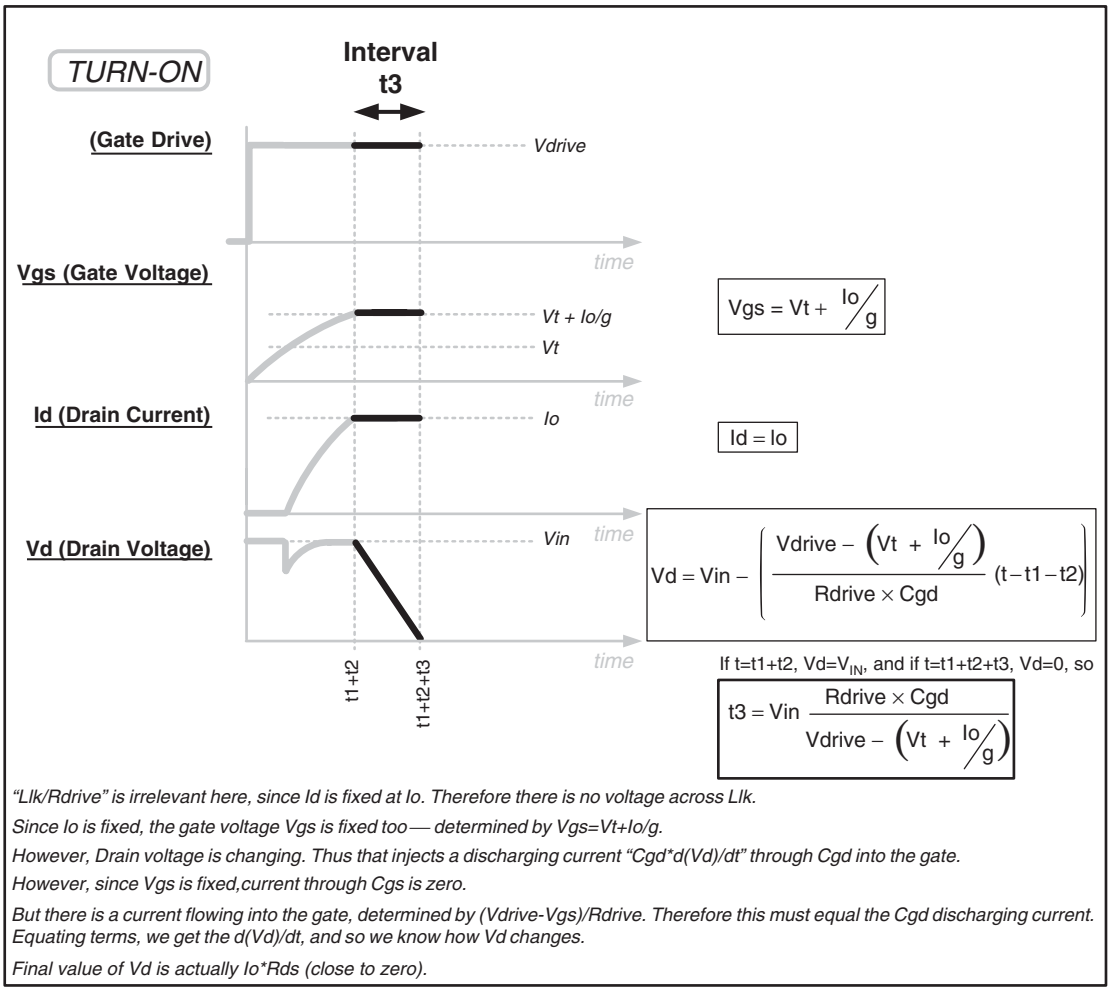


Figure 5-8: Second Interval of Turn-on

the gate voltage level required by the mosfet to support the full inductor current Io. So, to a first approximation, the voltage across Cgs (gate voltage) *need not*, and *does not* change. And further, since the general equation for the current through any capacitor is I = CdV/dt, the current through Cgs must be zero, because there is no *change* in the voltage across it during this sub-interval. Therefore we conclude that *all* the current coming through Cgd into the gate node gets diverted through Rdrive! But the voltage across Rdrive is fixed — one end of it is at Vdrive, the other at Vt + Io/g. Therefore the current through it is predetermined by Ohm's law. Which means that Rdrive is actually in full control of the current through Cgd during the interval t3. However, the current through Cgd also obeys the equation I = C × dV/dt. So if I is fixed at a certain value (by Rdrive), we can calculate the corresponding dV/dt across Cgd, *and thereby calculate Vd*. In effect, this means that Cgd and



"Llk/Rdrive" is irrelevant here, since Id is fixed at Io. Therefore there is no voltage across Llk.

Since Io is fixed, the gate voltage Vgs is fixed too— determined by Vgs=Vt+Io/g.

*However, Drain voltage is changing. Thus that injects a discharging current "Cgd*d(Vd)/dt" through Cgd into the gate.*

However, since Vgs is fixed,current through Cgs is zero.

But there is a current flowing into the gate, determined by (Vdrive-Vgs)/Rdrive. Therefore this must equal the Cgd discharging current. Equating terms, we get the d(Vd)/dt, and so we know how Vd changes.

*Final value of Vd is actually Io*Rds (close to zero).*

Figure 5-9: Third Interval of Turn-on

Rdrive are together determining the rate of fall of drain voltage during t3 (and thus the transition time of the voltage). The *plateau* in the gate voltage waveform during t3 is called the 'Miller plateau' — referring to the effect of the reverse transfer capacitance Cgd. Finally, after the voltage too has completed its swing, the current through Cgd stops completely, and so once again, the gate behaves as a simple RC charging circuit. Note that during t4, the gate is in effect being *overdriven* — there is no change in the drain current anymore (which is already at its maximum possible value). However, *driver dissipation continues during t4*.

The 'crossover time,' being the time during which both the current *and* voltage are transiting, is t2 + t3. As indicated, to know the driver dissipation, we need to consider the *entire* duration t1 + t2 + t3 + t4. Note that by definition, at the end of t4, the gate voltage is at

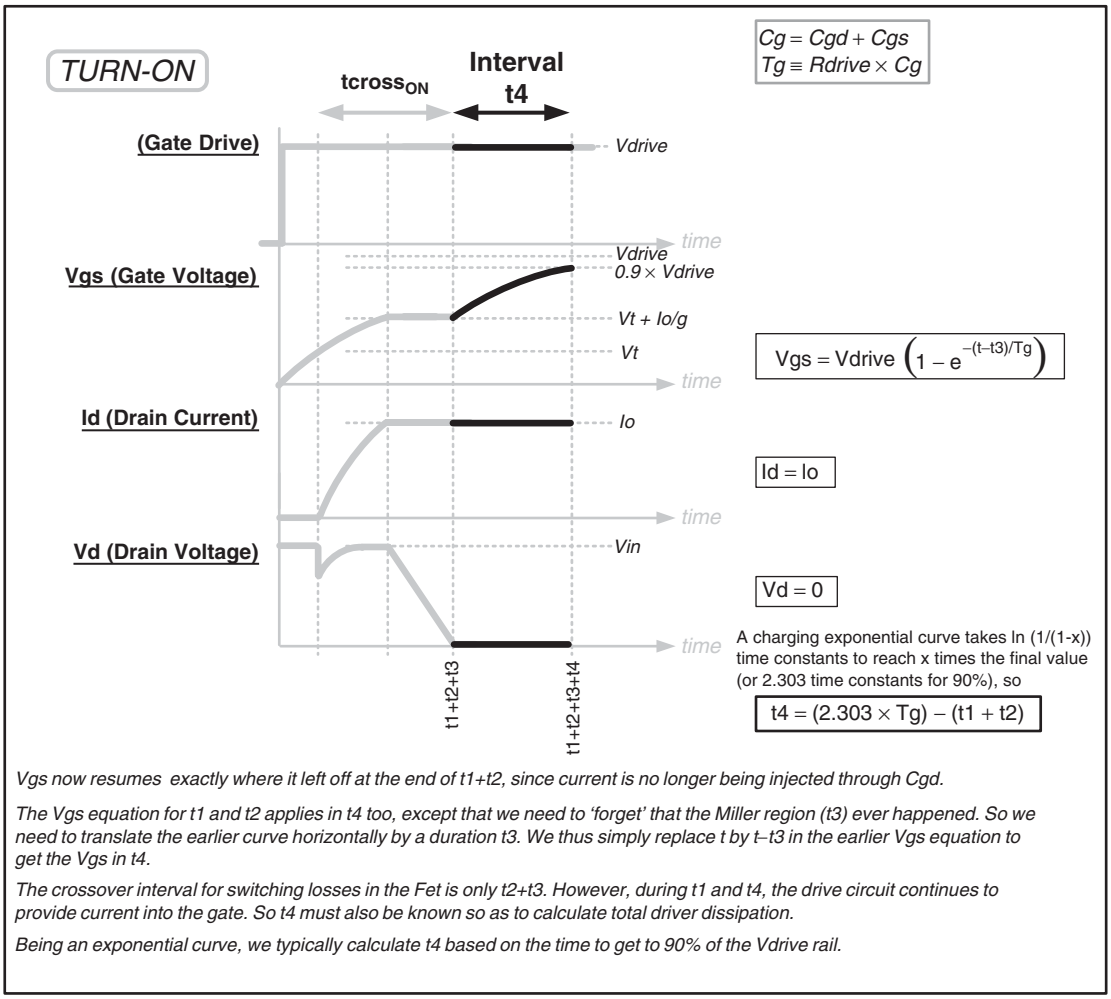


Figure 5-10: Fourth Interval of Turn-on

90% of its asymptotic level (Vdrive). So we can safely assume that for all practical purposes, the driver does very little after this point. Therefore, at the end of t4, the transition is considered complete — from the viewpoint of the switch, and also the driver.

The Turn-off Transition

In a similar manner as for turn-on, we have divided the turn-off interval into *four* subintervals, as shown in *Figures 5-11* through *5-14*.

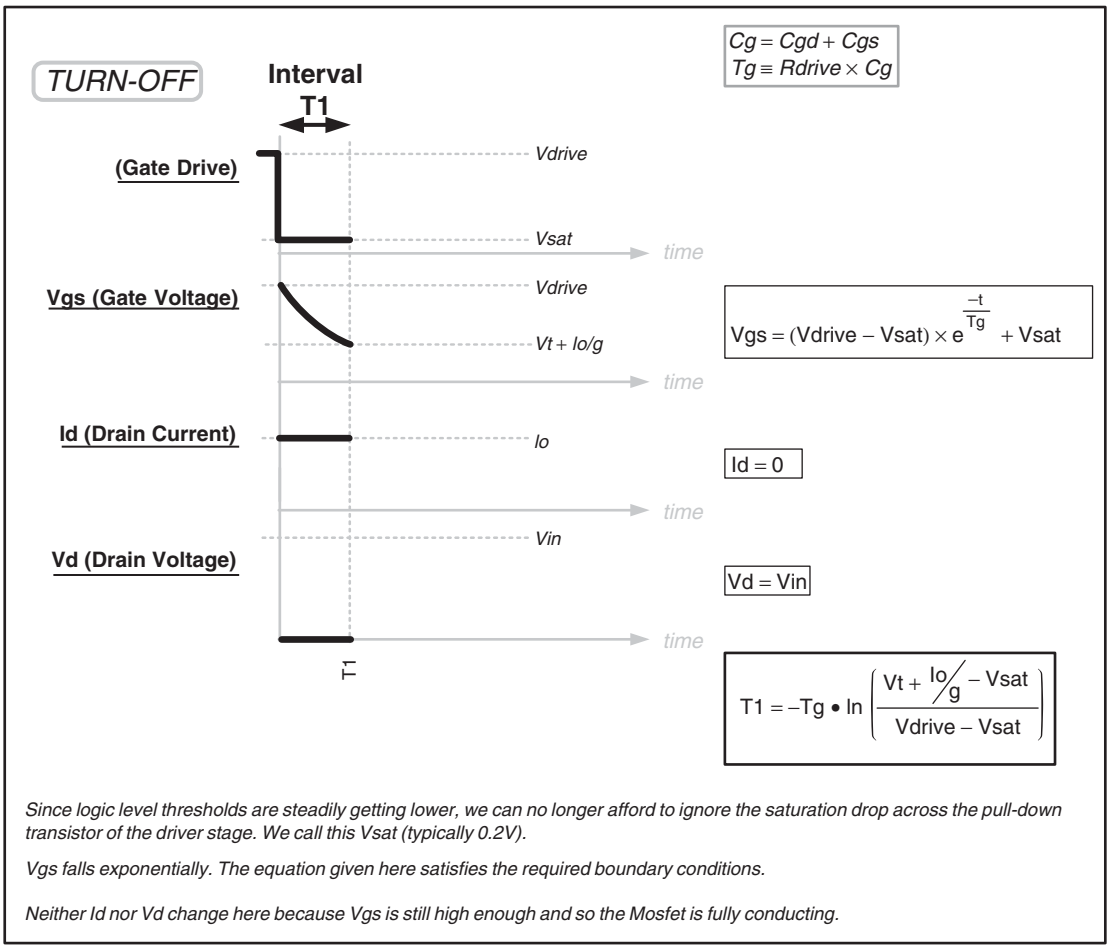


Figure 5-11: First Interval of Turn-off

Briefly, the interval T1 is the time for the "overdrive" to cease; that is, the gate returns to the *sustaining* level Vt + Io/g (the minimum gate voltage required to support the full drain current Io). During this time, there is no change in the drain current, nor in the drain voltage, and so in effect we once again have a simple RC discharging circuit. In T2, the gate voltage again *plateaus*. The reason for that is that the drain voltage must *first* swing close to Vin, and thereby "position" the diode to get forward-biased and be ready to start taking up the current that the switch will progressively shed (see *Figure 5-5*). So T2 is the time for the voltage transition to complete. During T1 and T2 therefore, no change in the drain current occurs. And with logic similar to what we presented for the turn-on sub-interval t3, during T2 the rate of *rise* of the voltage Vds is once again determined (only) by Rdrive and Cgd. Finally, in T3, the current starts falling toward zero. The gate voltage falls *exponentially* (as an

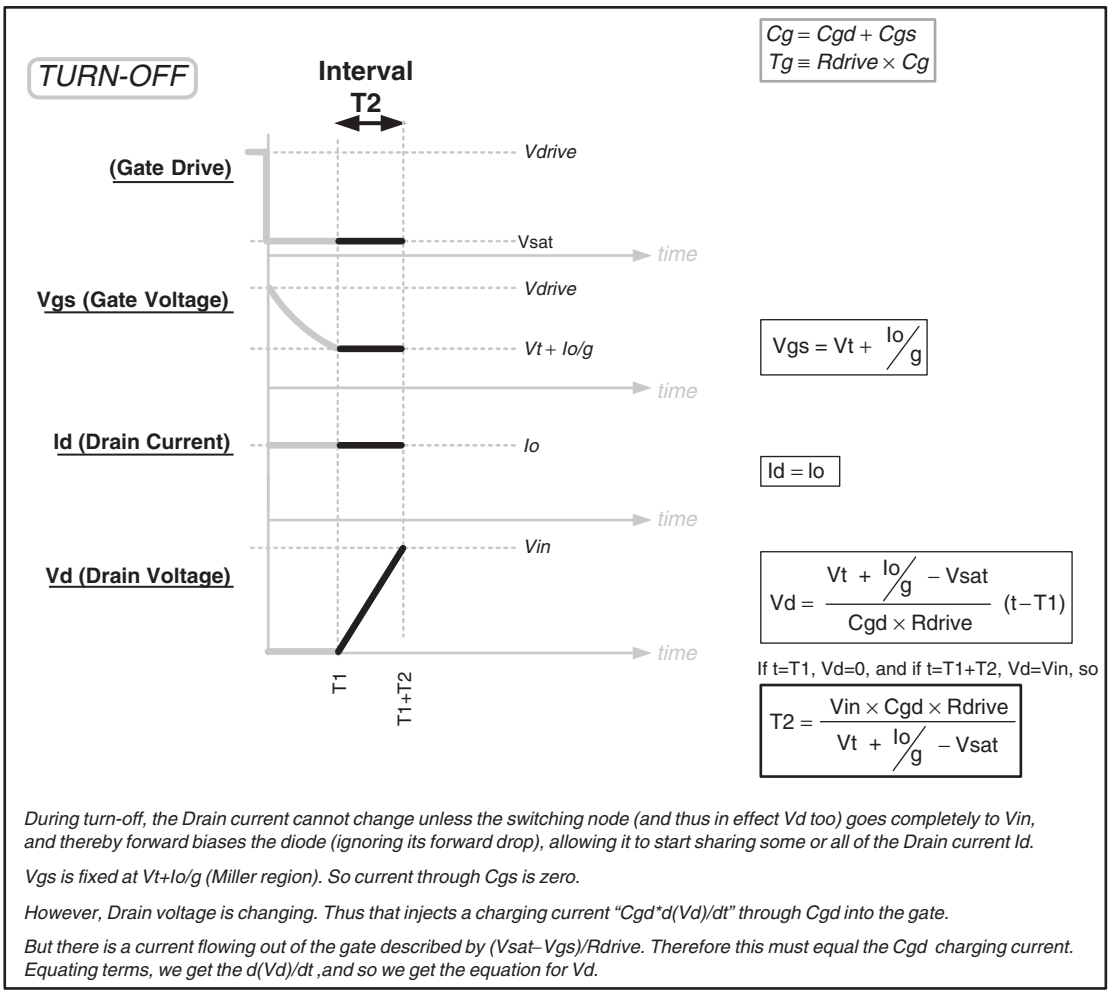
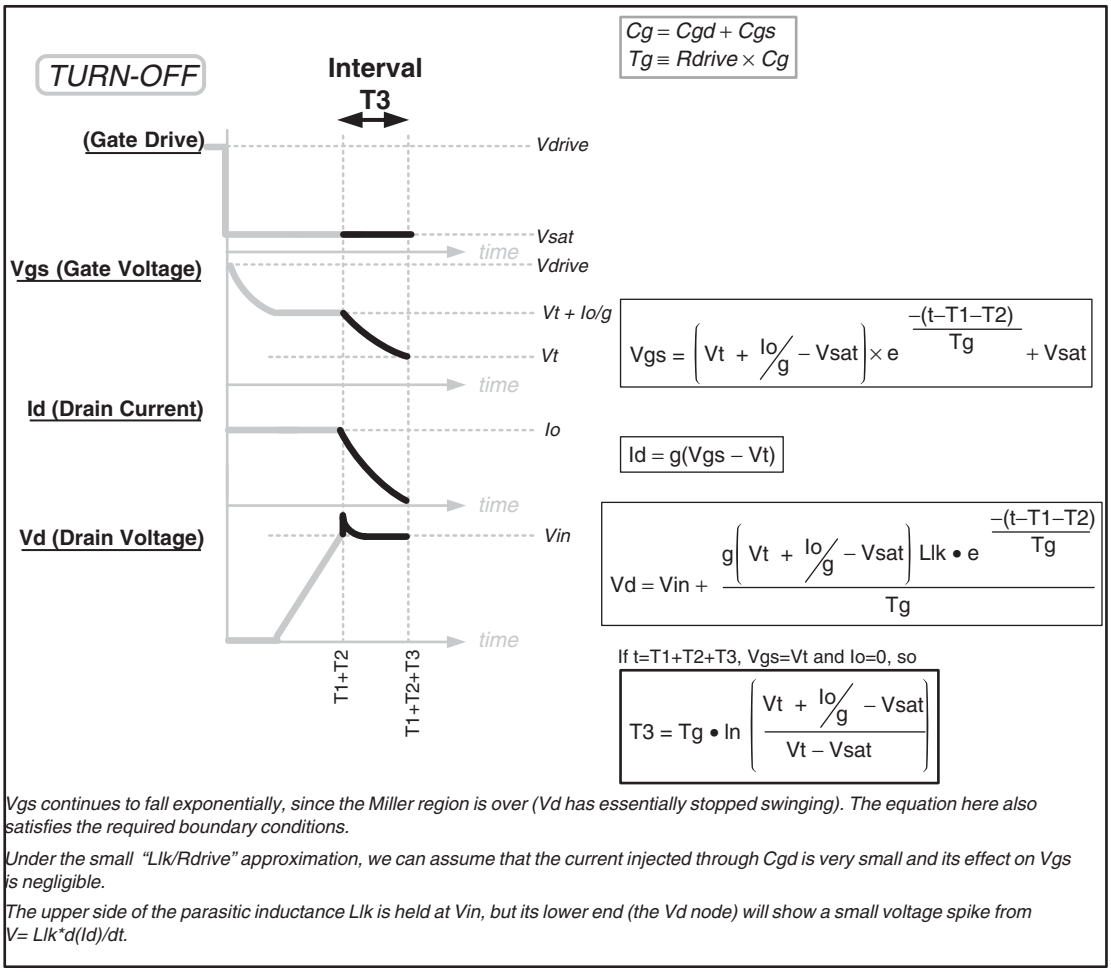


Figure 5-12: Second Interval of Turn-off

RC circuit) — down to Vt, at which moment, the end of subinterval T3 is declared. The transition is now complete *as far as the switch is concerned*. But after that, during T4, the RC exponential discharge continues down to 10% of the initial gate drive amplitude. As before, driver dissipation occurs over T1 + T2 + T3 + T4, whereas crossover occurs during T2 + T3.

Gate Charge Factors

A more recent way of describing the parasitic capacitor-based effects in a mosfet is in terms of *gate charge factors*. In *Figure 5-15*, we show how these charge factors, Qgs, Qgd, and Qg,



Vgs continues to fall exponentially, since the Miller region is over (Vd has essentially stopped swinging). The equation here also satisfies the required boundary conditions.

Under the small "Llk/Rdrive" approximation, we can assume that the current injected through Cgd is very small and its effect on Vgs is negligible.

*The upper side of the parasitic inductance Llk is held at Vin, but its lower end (the Vd node) will show a small voltage spike from V= Llk*d(Id)/dt.*

Figure 5-13: Third Interval of Turn-off

are defined. On the right column of the table in the figure, we have given the relationships between the gate charge factors and the capacitances, *assuming the latter are constants*. Gate charge factors represent a more accurate way of proceeding, since the interelectrode capacitances are such strong functions of the applied voltage. However, our entire analysis of the turn-on and turn-off intervals so far has been implicitly based on the assumption that the interelectrode capacitances are constants. A possible way out of this, one that also helps reduce the error in our switching loss estimates, is detailed in *Figure 5-16*, using the Si4442DY (from Vishay) as an example.

Basically, we are using the gate charge factors to tell us what the *effective* capacitances are (and the voltage swings from 0 to Vin). We see that the *effective* input capacitance (Ciss), for

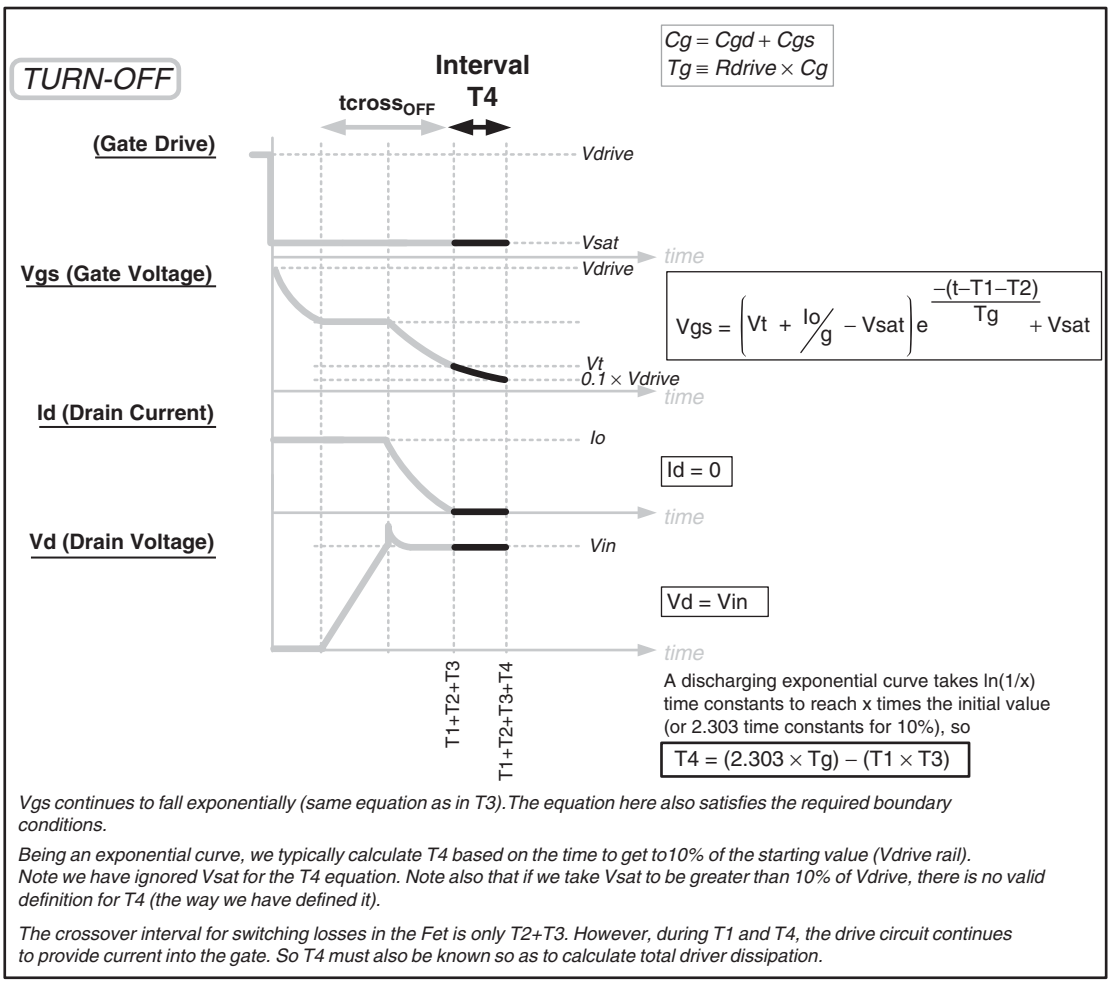


Figure 5-14: Fourth Interval of Turn-off

example, is about 50% greater than the *single-point* Ciss value that we would have read off from the typical performance curves (i.e. 6300 pF instead of 4200 pF). That factor accounts for the fact that as the voltage falls, the capacitance increases. Note that we could have calculated a *scaling factor* individually, for each capacitance. But it is simpler to use, say Ciss, to first find a "universal" scaling factor — and then apply it across the board to *all* the capacitances. In this manner, we arrive at the effective interelectrode capacitances quoted in *Figure 5-16*. These are the values we should use for our switching loss calculations

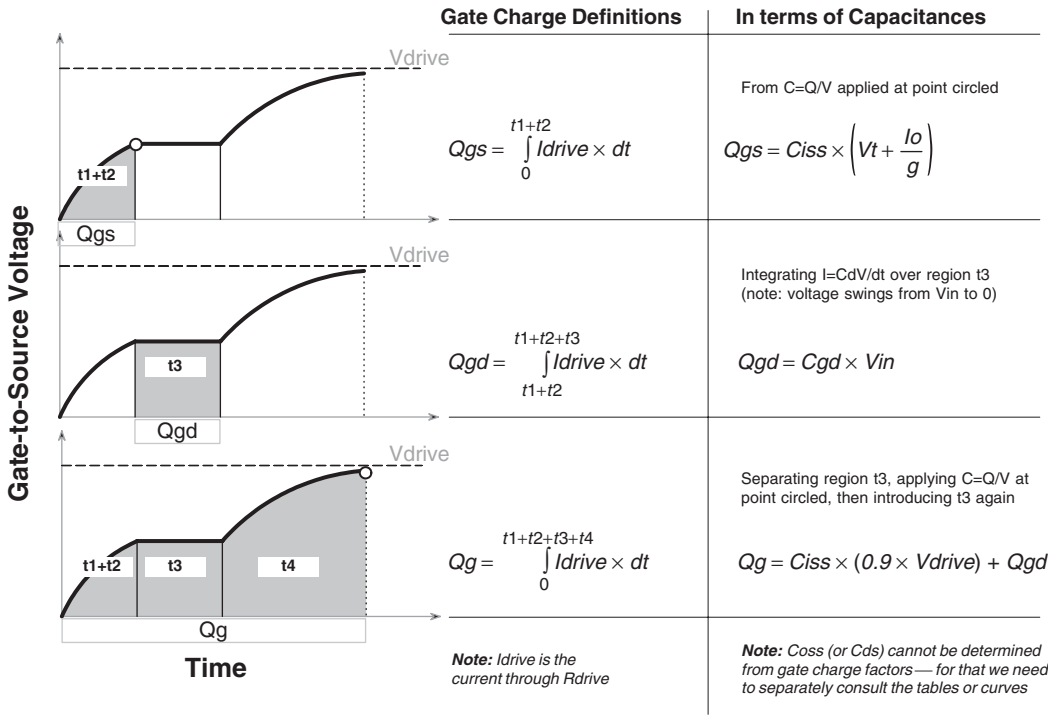


Figure 5-15: Gate Charge Factors of a Mosfet

(in preference to those provided by directly reading off Ciss, Coss, and Crss from their curves). Note that for finding the scaling factor, if we had looked at Crss (Cgd) instead of Ciss, then we would find that the calculated effective capacitance is only 40% higher (than what we would read directly from the curves). So the scaling factor can, in general, be fixed at around 1.4 to 1.5 typically.

Worked Example

We are switching 22 A at 15 V through a Si4442DY mosfet, at 500 kHz. The total pull-up drive resistance, by which the gate is driven by a pulse of amplitude 4.5 V, is 2 ohms. At turn-off, it is pulled-down (to source) by a total drive resistance of 1 ohm. Estimate the switching losses and the dissipation in the drive.

From *Figure 5-16*, we have Cg = Cgs + Cgd = 6300 pF.

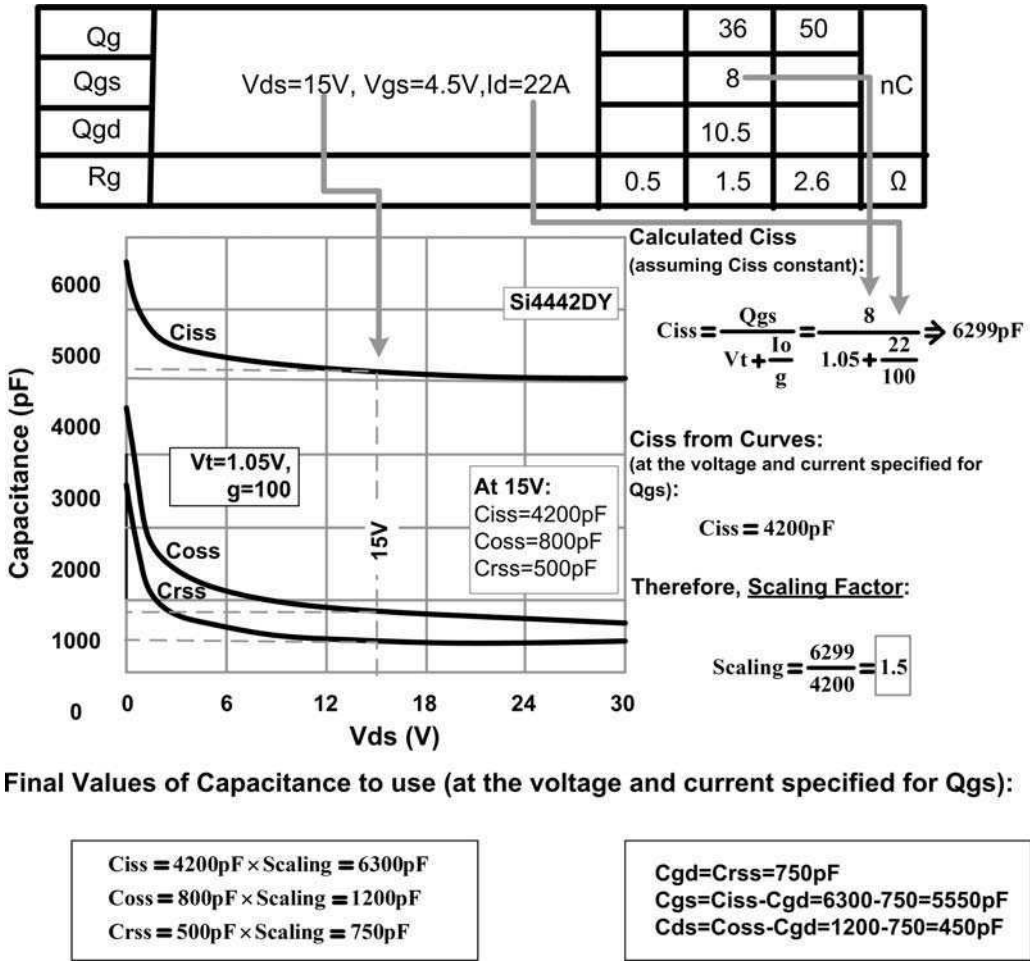


Figure 5-16: Estimating the *Effective* Interelectrode Capacitances from the Gate Charge Factors (Si4442DY as an example)

Turn-on

The time constant is

$$\mathrm{Tg} = \mathrm{Rdrive} \times \mathrm{Cg} = 2 \times 6300\ \mathrm{pF} = 12.6\ \mathrm{ns}$$

The time for the current to transit is

$$t2 = -\mathrm{Tg} \times \ln\left(1 - \frac{\mathrm{Io}}{g \times (\mathrm{Vdrive} - \mathrm{Vt})}\right) = -12.6 \times \ln\left(1 - \frac{22}{100 \times (4.5 - 1.05)}\right)$$

$$t2 = 0.83\ \mathrm{ns}$$

The time for the voltage to transit is

$$t3 = Vin \times \frac{Rdrive \times Cgd}{Vdrive - \left(Vt + \frac{Io}{g}\right)} = 15 \times \frac{2 \times 0.75}{4.5 - \left(1.05 + \frac{22}{100}\right)}$$

$$t3 = 6.966\,ns$$

So the crossover time during turn-on is

$$tcross\_turnon = t2 + t3 = 0.83 + 6.966 = 7.8\ ns$$

The turn-on crossover loss therefore is

$$Pcross\_turnon = \frac{1}{2} \times Vin \times Io \times tcross\_turnon \times fsw$$

$$= \frac{1}{2} \times 15 \times 22 \times 7.8 \times 10^{-9} \times 5 \times 10^{5}$$

$$Pcross\_turnon = 0.64\ Watts$$

Turn-off

The time constant is now

$$Tg = Rdrive \times Cg = 1 \times 6300\ pF = 6.3\ ns$$

The time for the voltage to transit is

$$T2 = \frac{Vin \times Cgd \times Rdrive}{Vt + \frac{Io}{g}} = \frac{15 \times 0.75 \times 1}{1.05 + \frac{22}{100}}$$

$$T2 = 8.858\ ns$$

The time for the current to transit is

$$T3 = Tg \times \ln\left(\frac{\frac{Io}{g} + Vt}{Vt}\right) = 6.3 \times \ln\left(\frac{\frac{22}{100} + 1.05}{1.05}\right)$$

$$T3 = 1.198\ ns$$

So the crossover time during turn-off is

$$\text{tcross\_turnoff} = \text{T2} + \text{T3} = 8.858 + 1.198 = 10 \text{ ns}$$

The turn-off crossover loss therefore is

$$\begin{aligned}\text{pcross\_turnoff} &= \frac{1}{2} \times \text{Vin} \times \text{Io} \times \text{tcross\_turnoff} \times \text{fsw} \\ &= \frac{1}{2} \times 15 \times 22 \times 10 \times 10^{-9} \times 5 \times 10^{5} \\ \text{pcross\_turnon} &= 0.83 \text{ Watts}\end{aligned}$$

So finally, the ***total crossover loss*** is

$$\text{Pcross} = \text{Pcross\_turnon} + \text{Pcross\_turnoff} = 0.64 + 0.83 = 1.47 \text{ Watts}$$

Notice that *we have not even used Cds* so far! This particular capacitance does not affect the V-I overlap (since it is not connected to the gate). But it still needs to be considered! Every cycle, it charges up during turn-off, and then during turn-on it dumps its stored energy inside the mosfet. This is, in fact, *the additional loss term that needs to be added to the crossover loss term, so as to get the total switching loss in a mosfet*. Note that in low-voltage applications, this additional term may seem insignificant, but in high-voltage/off-line applications, it does affect the efficiency noticeably. Let us calculate what it is in our case:

$$\text{P\_Cds} = \frac{1}{2} \times \text{Cds} \times \text{Vin}^2 \times \text{fsw} = \frac{1}{2} \times 450 \times 10^{-12} \times 15^2 \times 5 \times 10^{5} = 0.025 \text{ Watts}$$

So the ***total switching loss*** (in the switch) is

$$\text{Psw} = \text{Pcross} + \text{P\_Cds} = 1.47 + 0.025 = 1.5 \text{ Watts}$$

The ***driver dissipation*** is

$$\text{Pdrive} = \text{Vdrive} \times \text{Qg} \times \text{fsw} = 4.5 \times 36 \times 10^{-9} \times 5 \times 10^{5} = 0.081 \text{ Watts}$$

Note that typically, *the above driver dissipation equation underestimates the actual driver dissipation by almost 20%* — as can be confirmed by integrating the product of the drive current and the voltage across it, over each sub-interval. The reason for the error is simply the Miller plateau — because during this interval, some additional current (other than from the stored charge Qg), gets injected into the drive resistor. So our *corrected* driver dissipation estimate is $1.2 \times 0.081 = 0.097$ W. The driver supply rail current is $0.081/4.5 = 18$ mA.

Applying the Switching Loss Analysis to Switching Topologies

Now we try to understand how our preceding analysis pertains to an actual switching regulator application — in particular, what "Vin" and "Io" are, with respect to the topology.

For a ***buck***, we know that at *turn-on*, the instantaneous switch (and inductor) current is $I_O \times (1 - r/2)$, where *r* is the current ripple ratio, and "I_O" is the load current *of the dc-dc converter*. At turn-off, the current is $I_O \times (1 + r/2)$. Usually, we can ignore the current ripple ratio and take the current as I_O for both the turn-on and the turn-off analysis. So the load current of the dc-dc converter, I_O, becomes the same as the "Io" used so far in the switching loss analysis. Similarly, in a ***boost and buck-boost***, the current "Io" in our switching loss analysis, is actually the average inductor current $I_O/(1 - D)$.

Coming to the voltage across the mosfet when it turns OFF (i.e. "Vin" in the switching loss analysis) — for the ***buck***, this is almost equal to the input rail of the dc-dc converter V_{IN} (a diode drop more in reality). Similarly, for a ***buck-boost***, the voltage "Vin" is almost exactly equal to $V_{IN} + V_O$, where V_O is the output rail of the dc-dc converter. For a ***boost***, the voltage "Vin" is equal to V_O, that is, the output rail of the converter. Note that if we are dealing with an isolated ***flyback***, the voltage at turn-off really is $V_{IN} + V_Z$, where V_Z is the voltage of the zener clamp (placed across the primary winding). However, at turn-on, the voltage across the mosfet is only $V_{IN} + V_{OR}$ (V_{OR} being the *reflected output voltage*, i.e. $V_O \times n_P/n_S$). In a single-ended ***forward converter***, we have $2 \times V_{IN}$ at turn-off, and only V_{IN} at turn-on. Note that in all cases discussed above, we are assuming CCM.

We have tabulated these results in *Table 5-1* for convenience.

Note that if we were in DCM, there is in principle *no* switching loss at turn-on — because there is no current flowing in the inductor by that time. At turn-off, the current at transition is $I_{PK} = \Delta I$, which can be found using $V = L \times \Delta I/\Delta t$.

Table 5-1: Connecting the switching loss analysis with actual topologies

	"Vin"		"Io"	
	Turn-on	**Turn-off**	**Turn-on**	**Turn-off**
Buck	V_{IN}		I_O	
Boost	V_O		$I_O/(1 - D)$	
Buck-Boost	$V_{IN} + V_O$		$I_O/(1 - D)$	
Flyback	$V_{IN} + V_{OR}$	$V_{IN} + V_Z$	$I_{OR}/(1 - D)$	
Forward	V_{IN}	$2 \times V_{IN}$	I_{OR}	
$V_{OR} = V_O \times n$, $I_{OR} = I_O/n$ where $n = n_P/n_S$				

Worst-case Input Voltage for Switching Losses

We must return now to the all-important question — *when we have a wide-input voltage range, what specific input voltage point represents the worst case for calculating switching losses?*

The switching loss equation is generically

$$\text{Psw} = \text{Vin} \cdot \text{Io} \cdot \text{tcross} \cdot \text{fsw Watts}$$

We note that see that in all cases, this loss depends on the *product of Vin and Io*. But by now, we know what Vin and Io are — from *Table 5-1*. So we can analyze the situation for each topology as follows

- For a ***buck***, "Vin × Io" = $V_{IN} \times I_O$. So the maximum loss will obviously occur at V_{INMAX}.
- For a ***boost***, "Vin × Io" = $V_O \times I_O/(1 - D)$. So the maximum loss will occur at D_{MAX}, that is, at V_{INMIN}.
- For a ***buck-boost***, "Vin × Io" = $(V_{IN} + V_O) \times I_O/(1 - D)$. We also know that $D = V_O/(V_{IN} + V_O)$. So plotting "Vin × Io," we get *Figure 5-17* (a typical case). Note that the curve is *symmetrical* around D = 0.5 — and that is the point of *minimum* switching losses. Below that point, the *voltage* increases significantly, and above that, the *current* increases significantly. Either way, the switching losses *increase* as we move away from D = 0.5. Therefore, in general, we must first examine the input range of our application, and see which of its ends is *furthest* from D = 0.5. For example, if in our application, the input range corresponds to a duty

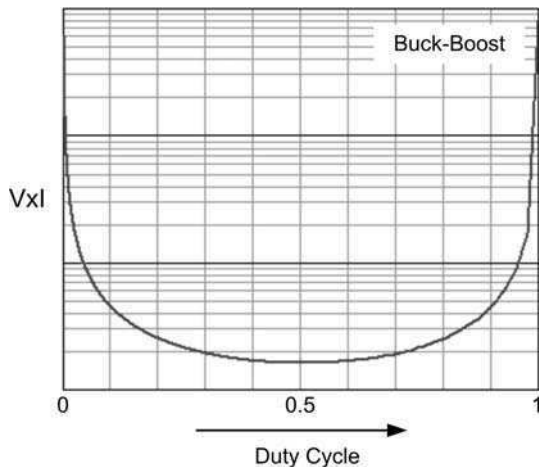


Figure 5-17: Switching Loss Variation with Respect to Duty Cycle, for the Buck-boost

cycle range of 0.6 to 0.8, we need to do the switching loss calculation at D = 0.8, that is, at V_{INMIN}. However, if the duty cycle range is say, 0.2 to 0.7, we need to do the calculation at D = 0.2, that is, at V_{INMAX}.

How Switching Losses Vary with the Parasitic Capacitances

In *Figure 5-18*, we have taken the Si4442DY, and "varied" its Ciss — just to see what can happen as a result of that. On the right vertical axis, we have the corresponding (estimated) switching loss. Note that in computing the loss curve, a "scaling factor" of 1.5 has been applied to the Ciss values given on the left vertical axis (though this is not obvious).

The gray vertical dashed line (annotated "35 nC") represents the Si4442DY *as it is*. So under the stated conditions, we have an estimated switching loss of 2.6 W. If we increase Ciss by 50%, that is, 4200 pF to 6300 pF, we see that Qg will go up to about 47 nC, and the loss to 2.8 W only.

Note: In the actual calculations, using the scaling factor of 1.5, "4200 pF" is actually 6300 pF, and "6300 pF" is actually 9450 pF.

In *Figure 5-19*, we take the Si4442DY, and "vary" its Crss — just to see what can happen as a result of that. The gray vertical dashed line (annotated "35 nC") represents the Si4442DY *as it is*. So under the stated conditions, we have an estimated switching loss of 2.6 W. If we increase Crss by 50%, that is, 500 pF to 750 pF, we see that Qg will go up to about 39 nC only, but the loss goes up to 3.1 W.

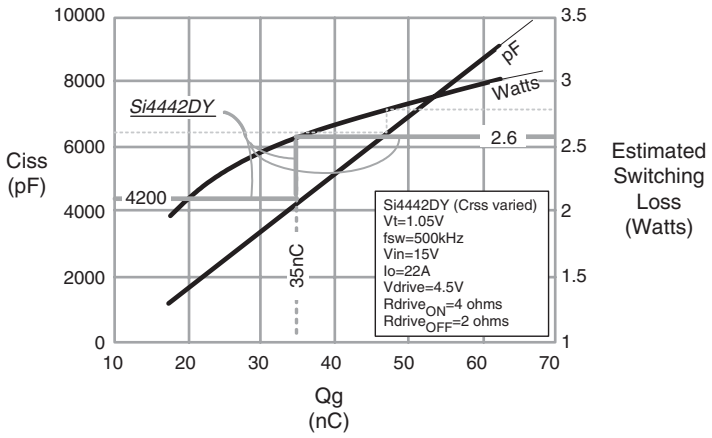


Figure 5-18: Varying the Ciss of the Si4442DY

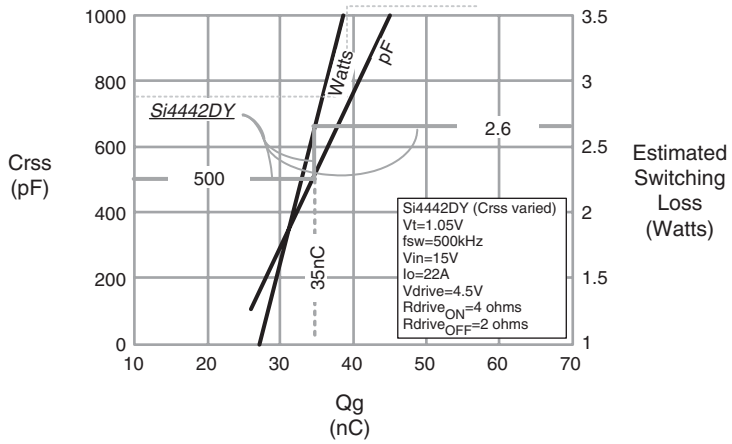


Figure 5-19: Varying the Crss of the Si4442DY

In other words, Qg will certainly affect driver dissipation, but it is not necessarily a good indicator of the *switching losses* — it is more *helpful to try and minimize Qgd (or Crss) when selecting mosfets*, rather than just looking for a "low-Qg" mosfet.

Note: In the worked example, we had estimated the losses to be 1.5 W. There we had a pull-up of 2 ohms, and a pull-down of 1 ohm. Whereas in *Figure 5-18*, we have basically doubled the pull-up and pull-down resistors. However, the switching loss has not doubled — it is only 73% more.

Optimizing Driver Capability vis-à-vis Mosfet Characteristics

In *Figure 5-20*, we have two separate graphs. The one on the left has a fixed pull-up of 4 ohms. On the x-axis, we are therefore, in effect, varying only the pull-down. So if for example, the x-axis is at 2, the pull-down resistor is 4 ohms/2 = 2 ohms. If the x-axis is at 4, the pull-down resistor is at 4 ohms/4 = 1 ohm. We see that as expected, the losses decrease as the pull-down is improved. We also see the effect of "varying" the threshold voltage. So, lower threshold voltages also help lower the switching losses — *provided the pull-down is not too "weak*." On the right graph similarly, we have the results for a fixed pull-up of 10 ohms. We can thereby estimate the effect of varying the pull-up too, on the overall losses.

Finally, in *Figure 5-21*, we are keeping the *pull-up + pull-down constant*, as we vary the *ratio* of the pull-up and pull-down resistors. This is from the IC designer's viewpoint — suppose he or she has roughly allocated a certain die area for the driver stage, say simplistically fixed the pull-up + pull-down. Then the question is — how should the available drive capability be *distributed* between the pull-up and the pull-down sections. For example, if pull-up + pull-down = 6 ohms, is it better to split this as — pull-up = 4 ohms

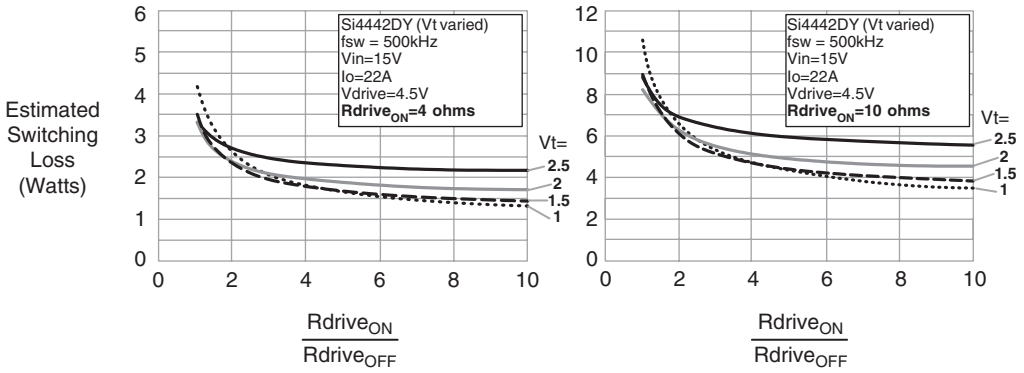


Figure 5-20: Varying the Threshold Voltage of the Si4442DY, and the Drive Resistances (keeping pull-up resistance fixed)

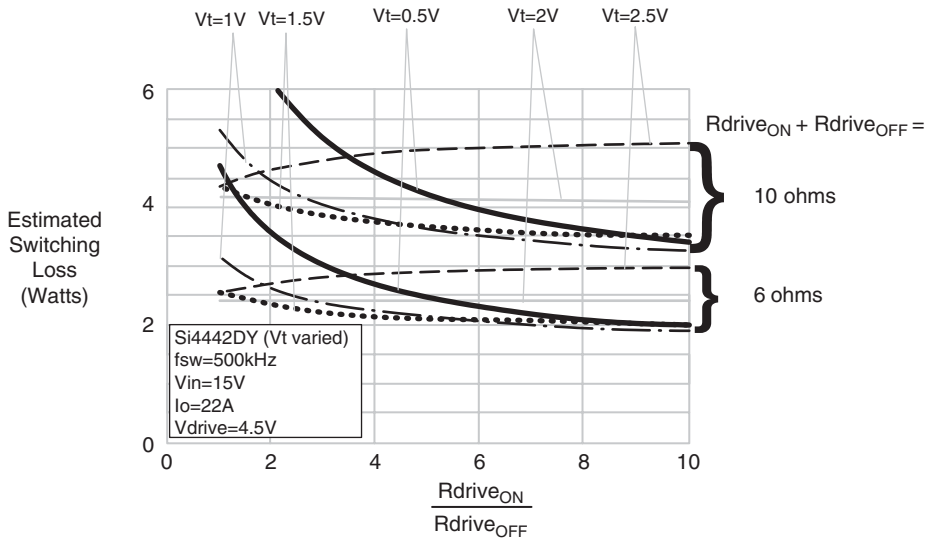


Figure 5-21: Varying the Threshold Voltage of the Si4442DY, and the Drive Resistances (keeping total drive resistance i.e. pull-up + pull-down, fixed)

and pull-down = 2 ohms, or say, pull-up = 3 ohms and pull-down = 3 ohms, or pull-up = 2 ohms and pull-down = 4 ohms, and so on? We see that the answer to that *depends on the threshold voltage*. So we need to have an idea of the mosfets we are planning to use, *before* we decide on the optimum ratio. From *Figure 5-21* we see that *if the threshold is greater than 2 V, improving the pull-up (at the expense of the pull-down) will help*, and so for example — pull-up = 4 ohms and pull-down = 2 ohms will be preferable to pull-up = 5 ohms and

pull-down = 1 ohm. However, *if the threshold voltage is below 2 V*, we see that the reverse is true — so now, *improving the pull-down (at the expense of the pull-up) will help*.

Note: Some vendors provide a rather wide range ("MIN" to "MAX") for threshold voltage. Often, they do not even provide a "TYP" value. But surprisingly, some do not even provide the threshold voltage at all! They simply state that their mosfet is "capable of 4.5 V drive" (as for example most of the mosfets from www.renesas.com).

CHAPTER

6

Printed Circuit Board Layout

This Page Intentionally Left Blank

CHAPTER 6

Printed Circuit Board Layout

Introduction

A great many customer "complaints" regarding switcher ICs are ultimately traced to poor PCB (printed circuit board) layout practices. When designing a PCB for a switching regulator, we need to be aware that the final product is going to be only as good as its layout. Certainly, some ICs are more noise sensitive than others. Sometimes, the "same" part from several vendors can also have starkly varying noise sensitivities (see *Appendix 1* for a case involving the popular 384x series). Further, some ICs are architecturally more noise sensitive than others (for example current mode controllers are far more "layout-sensitive" than voltage mode controllers). We also have to face the fact that virtually no semiconductor manufacturers characterize the noise sensitivity of their products (often letting the customers discover it for themselves!). However, as designers, we can certainly, with poor attention to layout, pull off the near-impossible — turn a comparatively stable IC into a jittery and nervous part — one that can malfunction and even cause catastrophic consequences (switch failure). Further, since very few of these problems can be easily corrected, or "band-aided," at a later stage, it is very important to get the layout right at the very beginning.

Most of the layout recommendations in this chapter revolve around simply assuring basic functionality and performance. Though luckily, the beleaguered switcher designer will be happy to know, in general, the electrical aspects are all related — pointing in the same general direction. So for example, a good layout, that is, one that helps the IC function properly, also leads to reduced electromagnetic emissions, and vice-versa. There are some exceptions to this trend however, particularly when it comes to the practice of indiscriminate "copper-filling" (or copper "flooding") on PCBs, which we will touch upon later. Subsequently, the reader can try to gain more insight into the practical aspects of making switching regulators, by reading the chapters dedicated to the topic of EMI, later in this book.

Trace Section Analysis

A switch transition (crossover) occurs when the switch changes from an on-state (switch closed) to an off-state (switch open), or back. It lasts typically less than 100 ns. But most of the trouble starts right here! In fact, the noise has little to do with the basic *switching frequency* of the converter itself — it is the *transition* that is responsible for most of the

noise, and all its attendant problems. The smaller the switch transition time, the more are the possible consequences, as we see.

The first requirement for the designer is to understand the flow of power-related currents in the converter. This leads to an identification of the troublesome or "critical" traces of the PCB; we must pay the closest attention to these traces. We will also see that this identification process is very "topology-dependent." So we can't, for example, design the PCB for a buck-boost, the same way we would do it for a buck. The rules change significantly! We may thus also realize that very few PCB layout persons out there would understand this too well! Therefore it really is a good idea for the power supply designer to do the layout personally, or at the very least, closely supervise the PCB person in the act.

Some Points to Keep in Mind During Layout

Let's summarize these for quick reference purposes:

- During a crossover transition the current flow in certain trace sections has to suddenly come to a *stop*, and in certain others it has to *start* equally suddenly (within 100 ns or less typically, which is the switch transition time). These trace sections are identified as the "critical traces" in any switcher PCB layout. A very *high dI/dt* is created in them, during every switch transition. See *Figure 6-1*. Expectedly, these traces end up "complaining" vociferously in the form of small, but potent, voltage spikes across them. If *Chapter 1* has been fully understood by now, we realize that this is just the equation $V = L \times dI/dt$ playing its part — with the "L" being the parasitic inductance of the PCB trace. The rule-of-thumb for the inductance presented by a trace is *20 nH per inch of trace length*.
- Once generated, these noise spikes can not only appear at the input/output (causing performance issues), but also infiltrate the IC control section, causing it to behave anomalously, and unpredictably. We could even end up briefly losing the usual current limiting function too, leading to disastrous consequences.
- Mosfets switch faster than 'bjts' (bipolar junction transistors). The transition times of a mosfet can be of the order of 10 to 50 ns, as compared to a bjt's 100 to 150 ns. But that also makes the "spikes" far more severe in the case of the converters that use mosfet switches — because of the much higher dI/dt's they can generate in the critical trace sections of the PCB.

  Note: One inch of trace switching, say 1 A of instantaneous current in a transition time of 30 ns, gives a spike of 0.7 V. For 3 A, and two inches of trace, the induced voltage tries to be 4 V!

  Note: It is almost impossible to "see" the noise spikes. First of all, various parasitics help limit/absorb them somewhat (though they can still retain the capability to cause "controller upset"). Further, the moment we put in an oscilloscope probe, the 10 to 20 pF of probe capacitance can also absorb the spikes, and we would probably see nothing significant. In addition, probes pick up

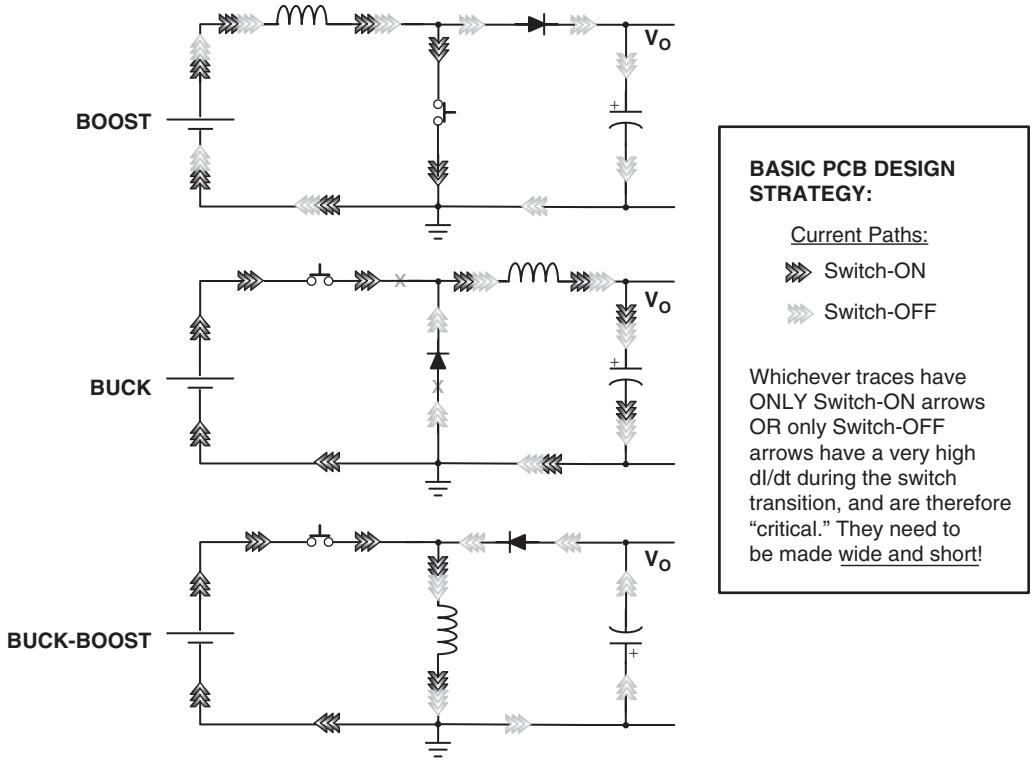


Figure 6-1: Identifying the Critical Trace Sections for the Three Topologies

so much normal switching noise through the air anyway, that we are never even sure of what we may be seeing!

- Integrated switchers ICs (or simply "switchers") have the switch in the same package as the control. Though that makes for convenience and low parts count, such ICs are usually more sensitive to the noise spikes generated by the parasitic trace inductances. That is because the 'switching node' of the power stage (its "swinging node," i.e. the one connecting the diode, switch, and inductor), is a pin on the IC itself, so that the pin conducts any unusual high-frequency noise at the switching node straight into the control sections, causing "controller upset."
- Note that while prototyping, it is a bad idea to insert a current probe (through a loop of wire) anywhere in a critical path (learn to recognize these in *Figure 6-1*). The current loop becomes an additional inductance that can increase the amplitude of the noise spikes dramatically. Therefore practically speaking, it often becomes virtually impossible to measure the switch current or the diode current individually (especially in the case of switcher ICs). In such cases, only the inductor current waveform can really be measured properly.

- Note that in the buck and the buck-boost, the input capacitor is also included in a critical path. That implies we need very good *input decoupling* in these topologies (for the power section). So, besides the necessary bulk capacitor for the power stage (typically a tantalum or aluminum electrolytic of large capacitance), we should also place a small ceramic capacitor (about 0.1 to 1 μF) directly between the *quiet* end of the switch (i.e. at the supply side) and the ground — and also *as close as possible to the switch*.
- In *Figure 6-1*, the control section (IC) has not been shown. However, we should remember that the control circuitry usually needs good *local* decoupling of its own. And for that we need to provide a small ceramic capacitor *very close to the IC*. Clearly, especially when dealing with switchers, the decoupling ceramic for the power stage can often do "double-duty" as the decoupling capacitor of the control too (note that this applies to the buck-boost and the buck only, since the input power-decoupling capacitor is required only for them).
- Sometimes, more effective control IC decoupling may be required — in which case we can use a small resistor (typically 10 to 22 Ω) from the input (supply) rail, going to a (separate) ceramic capacitor placed directly across the input and ground pins of the IC. This constitutes a small 'RC filter' for the IC supply.
- Note that in all topologies, the inductor is not in the critical path. So we need not worry much about its layout, at least not from the point of view of noise. However, we have to be wary of the electromagnetic field the inductor creates, because that can impinge on nearby circuitry and sensitive traces, and cause similar (though usually not so acute) problems. So generally, it is a good idea to try and use "shielded inductors" for that reason, if cost permits. If not, it should be positioned a little further from the IC, in particular keeping clear of the feedback trace.
- In the boost and the buck-boost, we see that the output capacitor is in the critical path. So this capacitor should be close to the control IC, along with the diode. A paralleled ceramic capacitor can also help, provided it does not cause loop instability issues (especially in voltage mode control – see *Chapter 7*).

  In the buck however, note that though the output diode needs to be positioned close to the IC/switch, the output capacitor is not critical (its current is smoothened by the inductor). If we place a ceramic capacitor in parallel to the output capacitor, it is only for the purpose of decreasing high-frequency noise and ripple at the output even further. But it is really not mandatory, and can cause severe loop instability, particularly with voltage mode control, especially if the effective series resistance (ESR) of the output capacitor section becomes too low (less than 100 mΩ typically).
- The position of the diode is critical in all topologies. It leads to the switching node and from there on, straight into the IC when using switcher ICs. However, in buck

converter layouts in which the diode has unfortunately been placed a little too far away from the IC, the situation can usually be rectified even at a later stage, by means of a small series RC snubber connected between the switching node and ground (across the catch diode, close to the IC). This RC typically consists of a resistor (low-inductive type preferred), of value 10-100 ohms, and a capacitor (preferably ceramic), of value about 470 pF to 2.2 nF. Note that the dissipation in the resistor is $C \times V_{IN}^2 \times f$. So not only should the wattage of the resistor be appropriate for the job, but the capacitance should not be increased indiscriminately, to avoid compromising the efficiency significantly.

- A first approximation for the inductance of a conductor (wire) having length 'l' and diameter 'd' is

$$L = 2l \times \left(\ln \frac{4l}{d} - 0.75 \right) \text{ nH}$$

where l and d are in centimeters. Note that the equation for a PCB trace is not much different from that of a wire.

$$L = 2l \times \left(\ln \frac{2l}{w} + 0.5 + 0.2235 \frac{w}{l} \right) \text{ nH}$$

where 'w' is the width of the trace. Note that for PCB traces, the inductance hardly depends on the thickness of the copper on the board.

The logarithmic relationship above indicates that if we halve the length of a PCB trace, we can make its inductance halve too. But we have to increase its width almost 10 times, to get its inductance to halve. In other words, simply making traces "wide" may not do much — we need to keep trace lengths *short*.

- The inductance of a 'via' (through-hole) is given by

$$L = \frac{h}{5} \left(1 + \ln \frac{4h}{d} \right) \text{ nH}$$

where 'h' is the height of the via in mm (equal to the thickness of the board, commonly 1.4 to 1.6 mm), and 'd' is the diameter of the via in mm. Therefore a via of diameter 0.4 mm on a 1.6 mm thick board gives an inductance of 1.2 nH. That may not sound like too much, but has been known to cause problems in switcher ICs, especially those using mosfets, for which an input ceramic decoupling capacitor for the IC becomes almost mandatory. Therefore, it is strongly advised that this capacitor be placed extremely close to where the pins of the IC actually contact the board, and further, there should be no intervening vias either, between this capacitor and the solder pads of the pins.

- Increasing the width of certain traces can in fact become counterproductive. For example, for the (positive) buck regulator, the trace from the switching node to the diode is "hot" (swinging). Any conductor with a varying voltage on it, irrespective of the current it may be carrying, becomes an E-field antenna if its dimensions are large enough. Therefore the area of the copper around the switching node needs to be reduced, not increased. That is why we need to avoid the tendency of indiscriminate "copper filling" — the only voltage node that really qualifies for copper filling is the ground node (or plane). All others, including the input supply rail, can start radiating significantly because of the high-frequency noise riding on them. By making large planes, we also increase the probability of that plane picking up noise from nearby traces and components, by means of inductive and capacitive coupling.
- The so-called "1-oz" board in the United States is actually equivalent to 1.4 mils copper thickness (or 35 μm) on the board. Similarly "2-oz" is twice of that. For a moderate temperature rise (less than 30°C) and currents less than 5 A, we can use a minimum 12 mils width of copper per amp for 1-oz board, and at least 7 mils width of copper per amp for a 2-oz board. This rule-of-thumb is based on the dc resistance of the trace only. So to decrease its inductive impedance and ac resistance, higher trace widths may be required.
- We have seen that the preferred method to reduce trace inductance is to reduce length, not increase width. Beyond a certain point, widening of traces does not reduce inductance significantly. Nor does it depend much on whether we use 1-oz or 2-oz boards. Nor if the trace is "unmasked" (to allow solder/copper to deposit and thereby increase effective conductor thickness). So, if for any reason, the trace length cannot be reduced further, another way to reduce inductance is by *paralleling the forward and return current traces*. Inductances exist because they represent stored magnetic energy. The energy resides in the magnetic field. Therefore conversely, if the magnetic field could be cancelled, the inductance vanishes. By paralleling two current traces, each carrying currents of the same magnitude but in opposite direction, the magnetic field is greatly reduced. These two traces should be parallel and very close to each other on the same side of the PCB. If a double-sided PCB is being used, the best solution is to run the traces parallel (over each other) on *opposite* sides (or adjacent layers) of the PCB. These traces can, and should be, fairly wide to improve mutual coupling and thereby the field cancellation. Note that if a ground plane is used on one side, the return path automatically 'images' the forward current trace, and produces the sought after field cancellation.
- In high-power off-line flybacks, the trace inductances on the secondary side reflect on to the primary side, and can greatly increase the effective primary-side leakage inductance and impair the efficiency (see *Chapter 3*). The situation gets worse when

we have to stack several output capacitors in parallel, just to handle the higher RMS currents — long traces seem inevitable here. However, one way to decrease the inductance is by the field cancellation principle discussed above. This is shown implemented in *Figure 6-2*. Two copper planes (or big copper islands) are allocated, starting from the output diode. One of these planes is the ground plane, the other being the output voltage rail. By using two large parallel planes carrying forward and return currents, the inductance almost completely cancels out, and leads to a very good high-frequency freewheeling path as desired. Note that in the bargain, we also get excellent current sharing between the output capacitors.

- In single-sided boards, a popular way to ensure current sharing between several paralleled output capacitors is shown in *Figure 6-3*. It doesn't minimize inductance, but it does ensure that the life of the first downstream capacitor does not come to a premature end (simply because of "current hogging"). Note that in the "improved" layout on the right side of the figure, the total distance from the diode through each capacitor is roughly equal in all three cases shown — thus leading to more precise sharing.
- With multi-layer boards, it is a common practice to almost completely fill one layer with ground (if so, it should preferably be the layer immediately below the power components/traces). There are people who, usually rightly so, consider this a panacea for most problems. As we have seen, every signal has a return, and as its harmonics get higher, the return current, rather than trying to find the path of least dc resistance

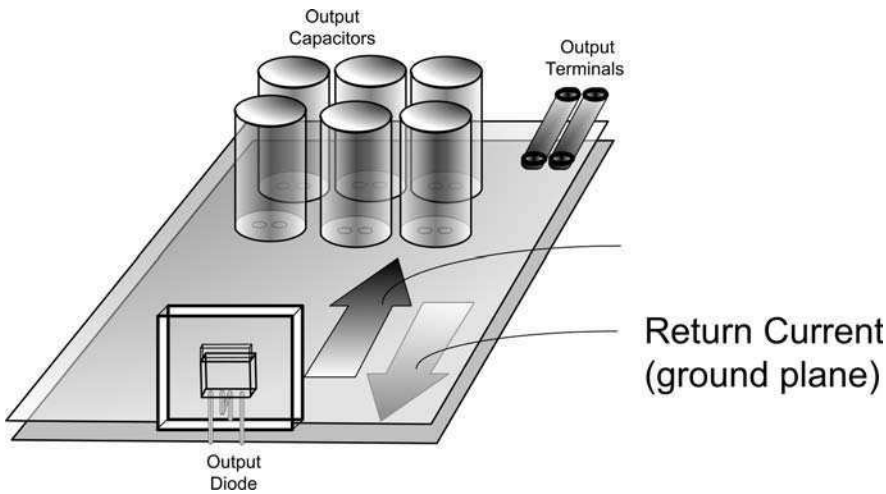


Figure 6-2: How to Achieve Low-inductance Connections to Output Capacitors of a Flyback.

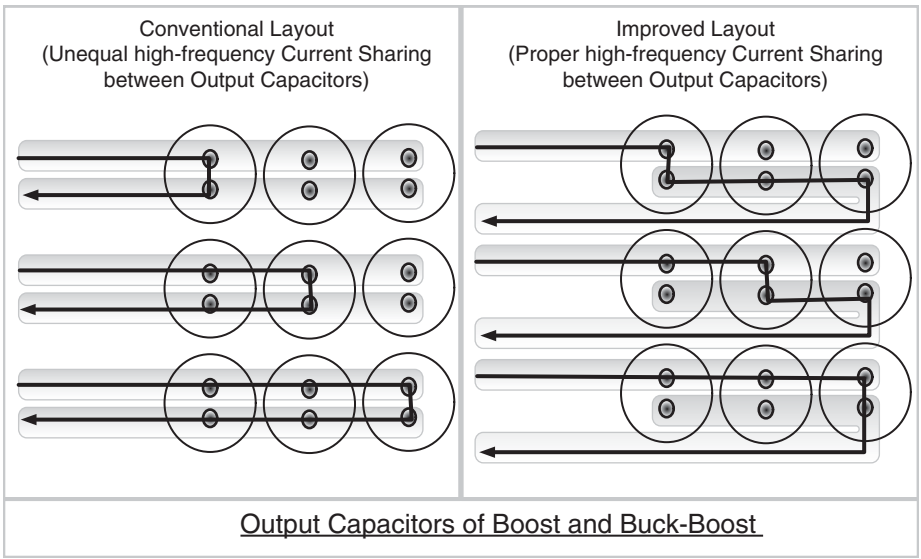


Figure 6-3: How to Get Output Capacitors of a Flyback to Share Current

(straight line), tries to reduce the inductance by imaging itself directly under the signal path even though that may be "zigzagging" away on the board. So by leaving a large ground plane, we basically "allow" nature to do its thing — searching and finding the path of least impedance (lowest dc resistance or lowest inductive impedance, depending upon the frequency of the harmonic). The ground plane also helps thermal management as it couples some of the heat to the other side. The ground plane can also capacitively link to noisy traces above it, causing general reduction in noise/EMI. However, it can also end up radiating if caution is not exercised. One way this can happen is to have too much capacitive coupling from noisy traces. No ground plane is perfect, and when we inject noise into it, it may get affected, especially if the copper is too thin. Also, if the ground plane is partitioned in odd ways, either to create thermal islands, or to route other traces, the current flow patterns can become irregular. No longer can return paths in the ground plane pass directly under their forward traces. The ground plane can then end up behaving as a slot antenna too, in terms of EMI.

- The only important *signal* trace to consider is usually the feedback trace. If this trace picks up noise (capacitively or inductively), it can lead to slightly offset output voltages — and in extreme cases (though rare), even instability or device failure. We need to keep the feedback trace short *if possible* so as to minimize pickup and keep it away from noise or field sources (the switch, diode, and inductor). We should never pass this trace *under* the inductor, or *under* the switch or diode (even if on opposite

sides of the PCB). We should also not let it run close to and parallel, for more than a few millimeters at most, to a noisy (critical) trace, even on adjoining layers of the board. Though if there is an intervening ground plane, that should provide enough shielding between layers.

Keeping the feedback trace short may not always be physically feasible. We should realize that keeping it short certainly is not of the highest priority. In fact, we can often deliberately make it long, just so that we can assuredly route it away from potential noise sources. We can also judiciously cut into the "quiet" ground plane to pass this particular trace through, so that in effect, it is surrounded by a "sea of tranquility."

Thermal Management Concerns

Larger and larger areas of copper do not help, especially with thinner copper. A point of diminishing returns is reached for a square copper area of size 1 inch × 1 inch. Some improvement continues until about 3 inches (on either side), especially for 2-oz boards and better. But beyond that, external heatsinks are required. A reasonable practical value attainable for the thermal resistance (from the case of the power device to the ambient) is about 30°C/W. That means 30°C rise for every watt of dissipation inside the IC.

To calculate the required copper area, we can use as a good approximation the following empirical equation for the required copper area:

$$A = 985 \times Rth^{-1.43} \times P^{-0.28} \text{ sq.inches}$$

Here P is in watts and Rth is the desired thermal resistance in °C/W (degrees Centigrade per Watt).

For example, suppose the estimated dissipation is 1.5 W. We want to ensure that, at a worst-case ambient of 55°C, the case of the part does not rise above 100°C (safe temperature for the PCB material — do not exceed!). Therefore the Rth we are looking for here is

$$Rth = \frac{\Delta T}{P} = \frac{100 - 55}{1.5} = 30°C/W$$

Therefore, the required copper area is

$$A = 985 \times 30^{-1.43} \times 1.5^{-0.28} \text{ sq. inches}$$

$$A = 6.79 \text{ sq. inches}$$

If this area is square in shape, the length of each side needs to be $6.79^{0.5} = 2.6$ inches. We can usually make this somewhat rectangular or odd-shaped too, as long as we preserve

the total area. Note that if the area required exceeds 1 square inch, a 2-oz board should be used (as in this case). A 2-oz board reduces the thermal 'constriction' around the power device and allows the large copper area to be more effectively used for natural convection.

We should not think that heat is lost only from the copper side. The usual laminate (board material) used for SMT (surface mount technology) applications is epoxy-glass 'FR4,' which is a fairly good conductor of heat. So some of the heat from the side on which the device is mounted does get across to the other side, where it contacts the air and helps lower the thermal resistance. Therefore, just putting a copper plane on the other side also helps — but only by about 10 to 20%. Note that this "opposite" copper plane need not even be electrically the same point — it could for example just be the usual ground plane. A much greater reduction of thermal resistance (by about 50 to 70%) can be produced if a cluster of small vias ("thermal vias") are employed to conduct the heat from the component side to the opposite side of the PCB.

Thermal vias, if used, should be small (0.3 to 0.33 mm barrel diameter), so that the hole is essentially filled up during the *plating* process. Too large a hole can cause 'solder wicking' during the reflow soldering process, which leads to a lot of solder getting sucked into the holes, thereby creating bad solder joints for components in the vicinity. The 'pitch' (i.e. the distance between the centers) of several such thermal vias in a given area is typically 1 to 1.2 mm. A grid of several such vias can be placed very close to, and alongside, a power device, and even under its tab (if present).

CHAPTER

7

Feedback Loop Analysis and Stability

This Page Intentionally Left Blank

CHAPTER 7

Feedback Loop Analysis and Stability

Transfer Functions, Time Constant and the Forcing Function

In *Chapter 1* we had discussed a simple series resistor-capacitor (RC) charging circuit. What we were effectively doing there was that by closing the switch we were applying a *step voltage* (stimulus) to the RC network. And we studied its "response" — which we defined as the *voltage appearing across the terminals of the capacitor*.

Circuits like these can be looked upon as a "black box", with two terminals coming in (the input, or the *excitation*) and two leaving (the output, or the *response*). One of the rails may of course be common to both the input and output, as in the case of the ground rail. This forms a "two-port network". Such an approach is useful, because power supplies too, can be thought of in much the same way — with two terminals coming in and two leaving, exposed to various disturbances/stimuli/excitations.

But let us examine the original RC-network in more detail first, to clarify the approach further. Let us say that the input to this RC network is a voltage *step* of height 'v_i'. The output of this network is taken to be the voltage across the capacitor, which we now call 'v_o' here. Note that v_o is a function of *time*. We define the *ratio* of the output to the input of any such two-port network, i.e. v_o/v_i in this case, as the 'transfer function'. Knowing how the RC network behaves, we also know the transfer function of this two-port network, which is

$$\frac{v_o(t)}{v_i} = 1 - e^{-t/RC}$$

Note that in general, a transfer function need not be "Volts/Volts" (dimensionless). In fact, neither the input nor the output of a two-port network need necessarily be voltage, or even similar quantities. For example, a two-port network can be as simple as a *current sense resistor*. Its input is the current flowing into it, and its output is the sensed voltage across it. So, its transfer function has the units of *voltage divided by current, that is, resistance*. Later, when we analyze a power supply in more detail, we will see that its *pulse width modulator* ("PWM") section for example, has an input that is called the 'control voltage', but its output is the dimensionless quantity — *duty cycle* (of the converter). So the transfer function in this case has the units of Volts^{-1}.

Returning to the RC network, we can ask — how did we actually arrive at the transfer function stated above? For that, we first use Kirchhoff's voltage law to generate the following differential equation:

$$v_i = v_{res}(t) + v_{cap}(t) = i(t)R + \frac{q(t)}{C}$$

where i(t) is the charging current, q(t) is the charge on the capacitor, $v_{res}(t)$ is the voltage across the resistor, and $v_{cap}(t)$ is the voltage across the capacitor (i.e. $v_o(t)$, the output). Further, since charge is related to current by dq(t)/dt = i(t), we can write the preceding equation as

$$v_i = R \times \frac{dq(t)}{dt} + \frac{q(t)}{C}$$

or

$$\frac{dq(t)}{dt} + \frac{1}{RC}q(t) = \frac{v_i}{R}$$

At this point we actually "cheat" a little. Knowing the properties of the *exponential function* $y(x) = e^x$, we do some educated reverse-guessing. And that is how we get the following solution:

$$q(t) = Cv_i\left(1 - e^{-t/RC}\right)$$

Substituting $q = C \times v_{cap}$, we then arrive at the required transfer function of the RC network given earlier.

Note that the preceding differential equation for q(t) above is in general a "*first-order*" differential equation — because it involves only the *first* derivative of time.

Later, we will see that there is in fact a *better way* to solve such equations — it invokes a mathematical technique called the 'Laplace transform.' But to understand and use that, we have to first learn to work in the '*frequency domain*' rather than in the 'time domain,' as we have been doing so far. We will explain all this soon.

We note in passing, that in a first-order differential equation of the previous type, the term that *divides* q(t) ('RC' in our case) is called the *'time constant*.' Whereas, the *constant term* in the equation ('v_i/R' in our case) is called the *'forcing function*.'

Understanding 'e' and Plotting Curves on Log Scales

We can see that the solution to the previous differential equation brought up the *exponential constant 'e'*, where **e ≈ 2.718**. We can ask — why do circuits like this always seem to lead

to *exponential-types* of responses? Part of the reason for that is that the *exponential function* e^x does have some well-known and useful properties that contribute to its ubiquity. For example

$$\frac{d(e^x)}{dx} = e^x \quad \text{and} \quad \int (e^x)\,dx = e^x + c \quad (c \text{ is a constant})$$

But this in turn can be traced back to the observation that the exponential constant *e* itself happens to be one of the most *natural* parameters of our world. The following example illustrates this.

Example: *Consider 10,000 power supplies in the field with a failure rate of 10% every year. That means in 2005 if we had 10,000 working units, in 2006 we would have 10,000 × 0.9 = 9000 units. In 2007 we would have* $9000 \times 0.9 = 8100$ *units left. In 2008 we would have 7290 units left, in 2009, 6561 units, and so on. If we plot these points — 10,000, 9000, 8100, 7290, 6561, and so on, versus time, we will get the well-known decaying exponential function (see Figure 7-1).*

Note that the simplest and most obvious initial assumption of a *constant* failure rate has actually led to an *exponential* curve. That is because the exponential curve is simply a succession of evenly spaced data points (very close to each other), that are in simple *geometric progression,* that is, the ratio of any point to its preceding point is a constant. Most natural processes behave similarly, and so *e* is encountered very frequently.

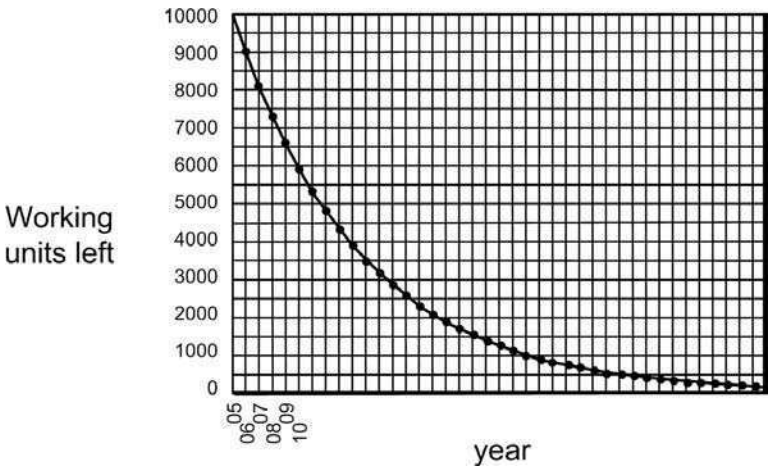


Figure 7-1: How a Decaying Exponential Curve Is Naturally Generated

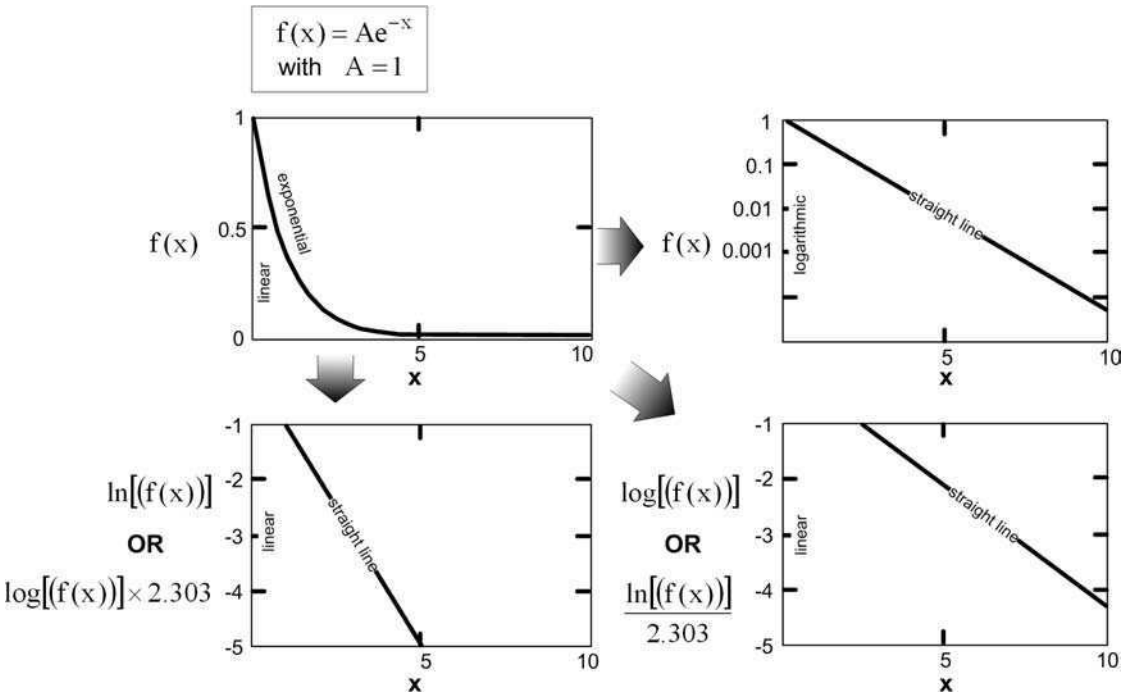


Figure 7-2: Plotting the Exponential Function with a Logarithmic y-scale or Plotting the Log of the Function with Linear y-scale — Both give a Straight Line

In *Figure 7-2* we have plotted out a more general exponentially decaying function, of the form $f(x) = A \times e^{-x}$ (though for simplicity, we have assumed $A = 1$ here). Let us now do some experiments *with the way we set up the horizontal and vertical axes*.

If we make the vertical scale (*only*) *logarithmic* rather than linear, we will find this gives us a straight line. Why so? That actually comes about due to a useful property of the logarithm as described next.

The definition of a logarithm is as follows — if $A = B^C$, then $\log_B(A) = \log(C)$, where $\log_B(A)$ is the "base-B logarithm of A." The commonly referred to "logarithm," or just "log," has an *implied* base of 10, whereas the natural logarithm "ln" is an abbreviation for a base-e logarithm.

So, we get the following sample relationships:

$$\log(10) = 1 \quad \log(100) = \log(10^2) = 2 \quad \log(10^x) = x\log(10) = x$$
$$\ln(e) = 1 \quad \ln\left(e^x\right) = x\ln(e) = x$$

Now, if we take the natural log of both sides of the equation $f(x) = A \times e^{-x}$, and use the last of the equations in the previous equation set, we get

$$\ln[f(x)] = \ln(e^{-x}) + \ln(A) = -x + \ln(A)$$

Further, if we compare this equation with the well-known standard *equation of a straight line* $y(x) = mx + c$ (where *m* is the slope and *c* is its intersection on the y-axis), we realize that, *if we plot ln f(x) on the vertical ("y") axis instead of f(x)* (*x* being the horizontal or "x" axis), we will get a straight line.

In general, plotting *a function on a log scale or plotting the log of a function on a linear scale is one and the same thing*.

But what if we had plotted log f(x) on the vertical axis instead of ln f(x)? If we think about it, we realize that is the same as asking — *what is the log_{10} of e, or equivalently, what is the log_e of 10?* As a matter of fact, there is not much difference really — because the base-10 and base-e logarithms are *proportional* to each other. This is probably more easily remembered when expressed in words — *if the log of any number is multiplied by 2.303, we get its natural log. Conversely, if we divide the natural log by 2.303 we get its log*. This follows from

$$\ln(10) = 2.303 \quad \text{and} \quad \frac{1}{\log(e)} = 2.303$$

Therefore, plotting any arbitrary function using a log scale (base 10) will always give us the same basic "shape" as plotting the natural log of the function. And if the function is an exponential one to start with, we will get a straight line in either case (*provided of course the horizontal axis is kept linear*). (See *Figure 7-2*.)

Time Domain and Frequency Domain Analysis

If we have a circuit (or network) constituted only of resistors, the voltage at any point in it is *uniquely defined* by the applied voltage. If the input varies, so does this voltage — instantly, and *proportionally* so. In other words, there is no 'lag' (delay) or 'lead' (advance) between the two. *Time* is not a consideration. However, when we include reactive components (capacitors and/or inductors) in any network, it becomes necessary to start looking at how the situation *changes over* time in response to an applied stimulus. This is called '*time domain analysis*.'

But we also know that any *repetitive* waveform, of almost arbitrary shape, can be decomposed into a sum of several *sine (and cosine)* waveforms of frequencies that are *multiples* of the basic repetition frequency 'f' ("the fundamental frequency"). That is what 'Fourier analysis' is all about. Note that though we do get an infinite series of terms, it is

composed of frequencies spaced *apart* from each other by an amount equal to f; that is, we do *not* get a *continuum* of frequencies when dealing with *repetitive* waveforms. But later, as we will see, when we come to more arbitrary waveshapes (*nonrepetitive*), we need a continuum of frequencies to decompose it.

The process of decomposition into frequency components implies that the components are mutually "independent." That is analogous to what we learn in our high-school physics class — we split a vector (applied force for example) into "orthogonal" x and y components. We perform the math on each component considered separately, and finally sum up again to get the final vector.

In general, understanding how a system behaves *with respect to the frequency components of an applied stimulus* is called '*frequency domain analysis*.'

Complex Representation

A math refresher is helpful here.

We remember that the impedance of an inductor is $L\omega$ and that of a capacitor is $1/C\omega$. Here $\omega = 2\pi f$ is the *angular* frequency in radians/s, f being the repetition frequency (say of the Fourier component) under consideration. Since both these types of reactive components introduce a *phase shift* (lag or lead) between their respective voltages and currents, we can no longer just add the voltages and currents arithmetically in any circuit that contains them. The solution once again is to use a form of vector analysis, except that now, a given voltage or current vector has *two* components — magnitude and phase. Further, unlike a conventional vector, these components are *dissimilar* quantities. Therefore, we cannot use conventional vector analysis. Rather, we invoke the use of the imaginary number $j = \sqrt{(-1)}$ to keep the phase and magnitude information distinct from each other, as we perform the math.

Any electrical parameter is thus written as a sum of real and imaginary parts:

$$A = Re + jIm$$

where we have used 'Re' to denote the *real* part of the number A, and 'Im' its *imaginary* part. From these components, the actual magnitude and phase of A can be reconstructed as follows:

$$\|A\| = \sqrt{Re^2 + Im^2} \quad \textit{(magnitude of complex number)}$$

$$\phi = \tan^{-1}\left(\frac{Im}{Re}\right) \text{ radians} \quad \textit{(argument of complex number)}$$

Impedance also is broken up into a vector in this complex representation — except that though it is frequency-dependent, it is (usually) not a function of time.

The 'complex impedances' of reactive components are

$$Z_L = j \times L\omega$$

$$Z_C = \frac{1}{j \times C\omega}$$

Note that $1/j = -j$. We see that the imaginary number 'j' is useful, because in effect, it also carries with it information about the 90° phase shift existing between the voltage and current in reactive components. So, the previous equations for impedance indicate that in an inductor, the current *lags* behind the voltage by 90°, whereas in a capacitor the current *leads* by the same amount. Note that *resistance* only has a real part, and it will therefore always be aligned with the x-axis of the complex plane (i.e. zero phase angle).

To find out what happens when a complex voltage is applied to a complex impedance, we need to apply the complex versions of our basic electrical laws. So Ohm's law, for example, now becomes

$$V(\omega t) = I(\omega t) \times Z(\omega)$$

We mentioned that the exponential function has some interesting properties. But a sine wave too has rather similar properties. For example, the rate of change of a sine wave is a cosine wave — which is just a sine wave phase shifted by 90°. It therefore comes as no surprise that we have the following relationships:

$$e^{j\theta} = \cos(\theta) + j\sin(\theta) \qquad \sin(\theta) = \frac{e^{j\theta} - e^{-j\theta}}{2j}$$

$$e^{-j\theta} = \cos(\theta) - j\sin(\theta) \qquad \cos(\theta) = \frac{e^{j\theta} + e^{-j\theta}}{2}$$

Note that in electrical analysis, we set $\theta = \omega t$. Here θ is the angle in *radians* (180° is π radians). Also, $\omega = 2\pi f$, where ω is the angular frequency in radians/s and f the (conventional) frequency in Hz.

As an example, using the preceding equations, we can derive the magnitude and phase of the exponential function $f(\theta) = e^{j\theta}$:

$$\text{Magnitude}(e^{j\theta}) = \sqrt{\cos(\theta)^2 + \sin(\theta)^2} = 1$$

$$\text{Argument}(e^{j\theta}) = \tan^{-1}\left(\frac{\sin(\theta)}{\cos(\theta)}\right) = \tan^{-1}\tan(\theta) = \theta$$

Note: Strictly speaking, a (pure) sine function is one with a phase angle of 0 and an amplitude of 1. However, when analyzing the behavior of circuits in general, a "sine wave" is actually a sine-*shaped* waveform of *arbitrary* phase angle and magnitude. So it is represented as $A_O \times e^{j\omega t}$ — that is, with magnitude A_O, and phase angle ωt (radians). For example, an applied "sine-wave" input voltage is $V(t) = V_O \times e^{j\omega t}$ in complex representation.

Nonrepetitive Stimuli

No stimulus is completely "repetitive" in the true sense of the word. "Repetitive" implies that the waveform has been exactly that way, since "time immemorial," and remains so forever. But in the real world, there is a definite moment when we actually *apply* a given waveform (and another when we remove it). Even an applied sine wave, for example, is not repetitive at the *moment* it gets applied at the inputs of a network. Much later, the stimulus may be considered repetitive, provided sufficient time has elapsed from the moment of application that the initial transients have died out completely. This is, in fact, the implicit assumption we always make when we carry out "steady state analysis" of a circuit.

But sometimes, we *do* want to know what happens at the *moment of application* of a stimulus — whether subsequently repetitive, steady, or otherwise. Like the case of the step voltage applied to our RC-network. If this were a power supply, for example, we would want to ensure that the output doesn't 'overshoot' (or 'undershoot') too much.

To study any such nonrepetitive waveform, we can no longer decompose it into components with *discrete* frequencies as we do with repetitive waveforms. Now we require a spread (continuum) of frequencies.

Further, to allow for waveforms (or frequency components) that can *increase or decrease* over time (disturbance changing), we need to introduce an additional (real) exponential term $e^{\sigma t}$. So whereas, when doing steady state analysis, we represent a sine wave in the form $e^{j\omega t}$, now it becomes $e^{\sigma t} \times e^{j\omega t} = e^{(\sigma + j\omega)t}$. This is therefore a "sine wave," but with an *exponentially increasing* (σ positive) *or decreasing* (σ negative) *amplitude*. Note that *if we are only interested in performing steady state analysis, we can go back and set* $\sigma = 0$.

The s-plane

In traditional ac analysis in the complex plane, the voltages and currents were complex numbers. But the frequencies were always real. However now, in an effort to include virtually arbitrary waveforms into our analysis, we have in effect created a complex frequency plane too, $(\sigma + j\omega)$. This is called *s-plane*, where $s = \sigma + j\omega$. Analysis in this plane is just a more *generalized* form of frequency domain analysis.

In this representation, the reactive impedances become

$$Z_L = Ls$$

$$Z_C = \frac{1}{Cs}$$

Resistance still remains just a resistance (no dependency on frequency or on *s*).

Let us also see how impedances add up in the s-plane — in particular when we parallel or series-combine reactances.

The effective impedance of a series combination is

$$Z(s) = Z_1(s) + Z_2(s) \quad \textit{(series impedances)}$$

For a parallel combination, we know that the *reciprocals* of the impedances add up to give us the *reciprocal* of the effective impedance. So

$$\frac{1}{Z(s)} = \frac{1}{Z_1} + \frac{1}{Z_2} \quad \textit{(parallel impedances)}$$

or

$$Z(s) = \frac{Z_1 \times Z_2}{Z_1 + Z_2}$$

Therefore, two inductors in *series* have an effective impedance equal to their sum

$$Ls = L_1 s + L_2 s$$

So the effective inductance is

$$L = L_1 + L_2 \quad \textit{(series inductance)}$$

Similarly two inductors in parallel behave as an effective inductance of

$$L = \frac{L_1 \times L_2}{L_1 + L_2} \quad \textit{(parallel inductance)}$$

Note however, that *capacitors **seem** to behave in an "opposite" manner to resistors and inductors*. Two capacitors in *parallel* give us an effective capacitance equal to their sum:

$$\frac{1}{Z(s)} = \frac{1}{Z_1} + \frac{1}{Z_2} \Rightarrow Cs = C_1 s + C_2 s \Rightarrow C = C_1 + C_2$$

$$C = C_1 + C_2 \quad \textit{(parallel inductance)}$$

This *looks* like a *series* combination (for resistors and inductors), but in reality, it is a *parallel* combination (of capacitors). Similarly, two capacitors in series give us

$$Z(s) = Z_1(s) + Z_2(s) \Rightarrow \frac{1}{Cs} = \frac{1}{C_1 s} + \frac{1}{C_2 s}$$

$$C = \frac{C_1 \times C_2}{C_1 + C_2} \quad \textit{(series capacitance)}$$

This looks like a parallel combination (for resistors and inductors), but in reality, it is a series combination (of capacitors).

To calculate the response of complex circuits and stimuli in the s-plane, we will need to use the above impedance summing rules, along with the rather obvious s-plane versions of the electrical laws. For example, Ohm's law is now

$$V(s) = I(s) \times Z(s)$$

Finally, the use of *s* gives us the ability to solve the differential equations arising from an almost arbitrary stimulus, in an *elegant* way, as opposed to the brute-force method in the time domain. The technique used to do this is the 'Laplace transform.'

> **Note:** Any such decomposition method can be practical only when we are dealing with "mathematical" waveforms. Real waveforms may need to be approximated by known mathematical functions for further analysis. And very arbitrary waveforms will probably prove intractable.

Laplace Transform

The Laplace transform is used to map a differential equation in the 'time domain' (i.e. involving 't') to the 'frequency domain' (involving 's'). The procedure unfolds as explained next.

First, the applied time-dependent stimulus (one-shot or repetitive — voltage or current) is mapped into the complex-frequency domain, that is, the s-plane. Then, by using the s-plane versions of the impedances, we can transform the entire circuit into the s-plane. To this transformed *circuit* we apply the s-plane versions of the basic electrical laws and thereby analyze the circuit. We will then need to solve the resultant (transformed) differential equation (now in terms of *s rather than t*). But as mentioned, we will be happy to discover that the manipulation and solution of such differential equations is much easier to do in the s-plane than in the time domain. In addition, there are also several lookup tables for the Laplace transforms of common functions available, to help along the way. We will thus get the response of the circuit in the frequency domain. Thereafter, if so desired, we can use the

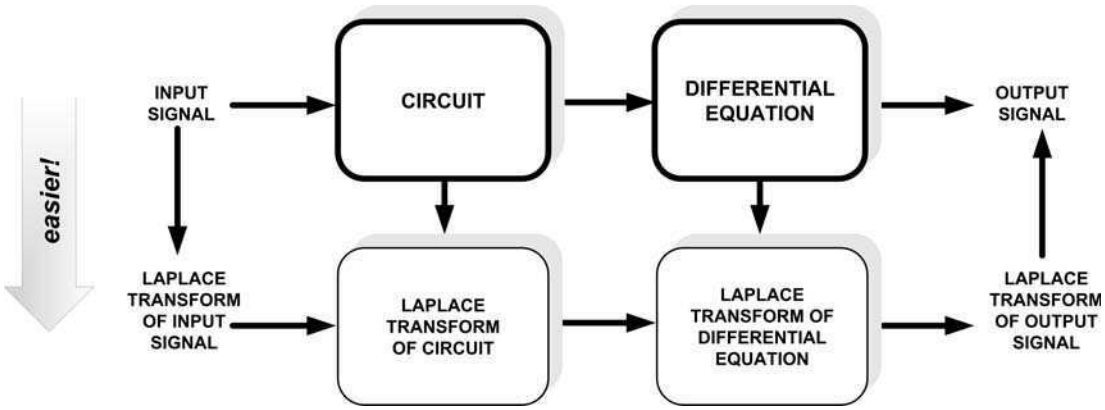


Figure 7-3: Symbolic Representation of the Procedure for Working in the S-plane

'inverse Laplace transform' to recover the result in the time domain. The entire procedure is shown symbolically in *Figure 7-3*.

A little more math is useful at this point, as it will aid our understanding of the principles of feedback loop stability later.

Suppose the input signal (in the time domain) is u(t), and the output is v(t), and they are connected by a general second-order differential equation of the type

$$c_2\frac{d^2u(t)}{dt^2} + c_1\frac{du(t)}{dt} + c_0u(t) = d_2\frac{d^2v(t)}{dt^2} + d_1\frac{dv(t)}{dt} + d_0v(t)$$

It can be shown that if U(s) is the Laplace transform of u(t), and V(s) the transform of v(t), then this equation (in the frequency domain) becomes simply

$$c_2s^2U(s) + c_1sU(s) + c_0U(s) = d_2s^2V(s) + d_1sV(s) + d_0V(s)$$

So

$$V(s) = \frac{c_2s^2 + c_1s + c_0}{d_2s^2 + d_1s + d_0}U(s)$$

We can therefore define G(s), the *transfer function* (i.e. output divided by input in the s-plane now), as

$$G(s) = \frac{c_2s^2 + c_1s + c_0}{d_2s^2 + d_1s + d_0}$$

Therefore

$$V(s) = G(s) \cdot U(s)$$

Note that this is analogous to the time-domain version of a general transfer function f(t)

$$v(t) = f(t) \cdot u(t)$$

Since the solutions for the general equation G(s) above are well researched and documented, we can easily compute the response (V) to the stimulus (U).

A power supply designer is usually interested in ensuring that his or her power supply operates in a stable manner over its operating range. To that end, a sine wave is injected at a *suitable* point in the power supply, and the frequency swept, to study the response. This could be done in the lab, and/or "on paper" as we will soon see. In effect, what we are looking at closely is the response of the power supply to any frequency *component* of a repetitive or nonrepetitive impulse. But in doing so, we are in effect only dealing with a steady *sine wave* stimulus (swept). So we can put $s = j\omega$ (i.e. $\sigma = 0$).

We can ask — why do we need the complex s-plane at all if we are going to use $s = j\omega$ anyway at the end? The answer to that is — we *don't* always. For example, at some later stage we may want to compute the exact response of the power supply to a specific disturbance (like a step change in line or load). Then we would need the s-plane *and* the Laplace transform. So, even though more often we end up just doing steady state analysis, by having already characterized the system in the framework of *s*, we retain the option to be able to conduct a more elaborate analysis of the system response to a more general stimulus if required.

A silver lining for the beleaguered power supply designer is that he or she usually doesn't even need to know how to actually compute the Laplace transform of a function — unless, for example, the exact *step* response is required to be computed exactly. Just for ensuring stability margins, it turns out that a steady state analysis serves the purpose completely. So typically, we do the initial math in the complex s-plane, but at the end, to generate the results of the margin analysis, we again revert to $s = j\omega$.

Disturbances and the Role of Feedback

In power supplies, we can either change the applied input voltage or increase the load (this may or may not be done *suddenly*). Either way, we always want the output to remain well regulated, and therefore, in effect, to "reject" the disturbance.

But in practice, that clearly does not happen in a *perfect* manner, as we may have desired. For example, if we suddenly increase the input to a buck regulator, the output initially just

tends to follow suit — since $D = V_O/V_{IN}$, and D has not immediately changed. To maintain output regulation, the control section of the IC needs to sense the change in the output (that may take some time), correct the duty cycle (that also may take some time), and then wait (a comparatively longer time) for the inductor and output capacitor either to give up some of their stored energy or to gather some more (whatever is consistent with the conditions required for the new *steady state*). Eventually, the output will hopefully settle down again.

We see that there are several such *delays* in the circuit before we can get the output to stabilize again. Minimizing these delays is clearly of great interest. Therefore, for example, using smaller filter components (L and C) will usually help the circuit respond faster.

However, one philosophical question still remains — how can the control circuit ever know *beforehand how much correction* (in duty cycle) to precisely apply (when it senses that the output has shifted from its set value on account of the disturbance)? In fact, it usually doesn't! It can only be designed to "know" what *general direction* to move in, but not by *how much*. Hypothetically speaking, we can do several things. For example, we can command the duty cycle to change *slowly and progressively*, with the output being monitored continuously, and stop correcting the duty cycle at the exact moment when the output returns to its regulation level. However, clearly this is a slow process, and so though the duty cycle itself won't "overshoot," the output will certainly overshoot or undershoot as the case may be, for a rather long time. Another way is to command the duty cycle to change *suddenly* by a *large arbitrary* amount (though of course in the right direction). However, now the possibility of *over-correction* arises, as the output could well "go the other way," before the control realizes it. And further, when it does, it may tend to "overreact" once more… and so on. In effect, we now get "ringing" at the output. This ringing reflects a basic *cause-effect uncertainty* that is present in any feedback loop — the control never fully knows whether the error it is seeing, is truly a *response* to an external disturbance, or its own attempted correction coming back to haunt it. So, if after a lot of such ringing, the output does indeed stabilize, the converter is considered only 'marginally stable.' In the worst case, this ringing may go on forever, even escalating, before it stabilizes. In effect, the control is then "fully confused," and so the feedback loop is deemed "unstable."

We see that an "optimum" feedback loop is *neither too slow, nor too fast*. If it is too slow, the output will exhibit severe overshoot (or undershoot). And if it is too fast (over-aggressive), the output may ring severely, and even break into full instability (oscillations).

The study of how any disturbance *propagates* inside the converter, either getting attenuated or exacerbated in the process, is called 'feedback loop analysis.' As mentioned, in practice, we test a feedback loop by deliberately injecting a small disturbance at an appropriate point inside it (*cause*), and then seeing at what magnitude and phase it returns to the same point (*effect*). If, for example, we find that the disturbance reinforces itself (at the right *phase*), cause-effect separation will be completely lost, and instability will result.

The very use of the word "phase" in the previous paragraph implies we are talking of *sine waves* once again. However, this turns out to be a valid assumption, because as we know, arbitrary disturbances can be decomposed into a series of sine wave components of varying frequencies. So the signal we "inject" (either on the bench, or on paper) can be a sine wave of constant, but arbitrary amplitude. Then, by sweeping the frequency over a wide range, we can look for *frequency components* (sine waves) that have the *potential* to lead to instability — assuming we can have a disturbance that happens to contain that particular frequency component. If the system is stable for a wide range of sine wave frequencies, it would in effect be stable when subjected to an arbitrarily-shaped disturbance too.

A word on the *amplitude* of the disturbance. Note that we are studying only *linear systems*. That means, if the input to a two-port network doubles, so does the output. Their *ratio* is therefore unchanged. In fact, that is why the transfer function was never thought of as say, being a function of the amplitude of the incoming signal. But we do know that in reality, if the disturbance is too severe, parts of the control circuit may "rail" — that means for example an internal op-amp's output may momentarily reach very close to its supply rails, thus affording no further correction for some time. We also realize that there is no perfectly "linear system." However, any system can be approximated by a linear system if the stimulus (and response) is "small" enough. That is why, when we conduct feedback loop analysis of power converters, we talk in terms of 'small-signal analysis' and 'small-signal models.'

Applying these facts to our injected (swept) sine wave, we realize that its amplitude should not be made too large, or else internal "railing" or "clipping" can affect the validity of our data and the conclusions. But it must not be made too small either, otherwise switching noise is bound to overwhelm the readings (poor signal to noise ratio). A power supply designer may have to struggle a bit on the bench to get the right amplitude for taking such measurements. And that may depend on the frequency. Therefore, more advanced instruments currently available allow the user to *tailor* the amplitude of the injected signal with respect to the (swept) frequency. So for example, we can demand that at higher sweep frequencies, the amplitude is set lower than the amplitude at lower frequencies. If we are looking at the switching waveform on an oscilloscope, we should see a small jitter — typically about 5 to 10% around the "edge." Too small a jitter indicates the amplitude is too small, and too large a jitter can cause strange behavior, especially if we are operating very close to the "stops" — the minimum or maximum duty cycle limits of the controller, and/or the set current limit.

Transfer Function of the RC Filter

Let us now take our simple series RC network and transform it into the frequency domain, as shown in *Figure 7-4*. As we can see, the procedure for doing this is based on the well-known equation for a dc voltage divider — now extended to the s-plane.

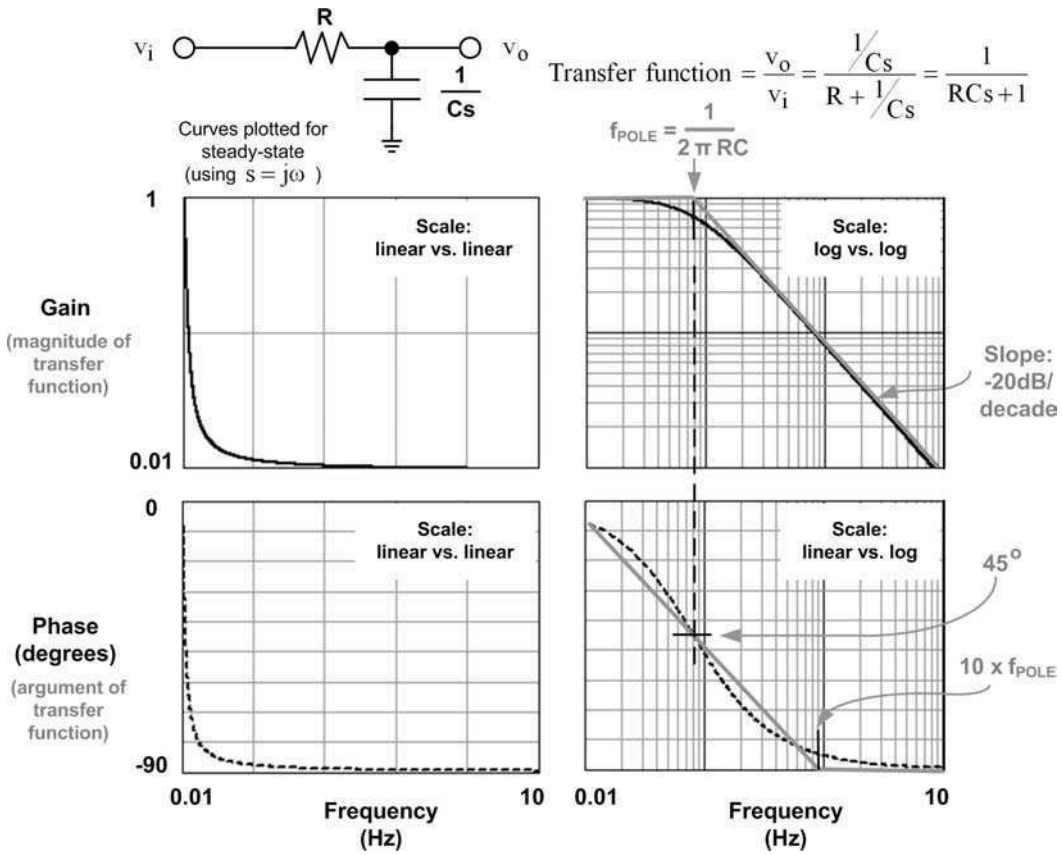


Figure 7-4: Analyzing the First-order Low-pass RC Filter in the Frequency Domain

Thereafter, since we are looking at only *steady state* excitations (not transient impulses), we can set $s = j\omega$, and plot out a) the magnitude of the transfer function (i.e. its '*gain*'), and b) the argument of the transfer function (i.e. its *phase*) — both in the frequency domain of course. This *combined gain-phase plot is called a* ***'Bode plot'***.

Note that *gain and phase are defined only in steady state* as they implicitly refer to a *sine wave* ("phase" has no meaning otherwise!).

Here are a few observations:

- In the curves of *Figure 7-4*, we have preferred to convert the phase angle (which was originally in *radians*, $\theta = \omega t$), into degrees. That is because most of us feel more comfortable visualizing degrees, not radians. To do this, we have used the following conversion — degrees $= (180/\pi) \times$ radians.
- We have also similarly converted from the 'angular frequency' (ω) to the usual frequency (in Hz). Here we have used the equation — Hz = (radians/second)/(2π).

- By varying the type of *scaling* on the gain and phase plots, we can see that the gain becomes a *straight line* if we use *log vs. log* scaling (remember that the exponential function needed log vs. linear scaling to appear as a straight line). We thus confirm by looking at these curves, that the gain at high frequencies starts decreasing *by a factor of 10 for every 10-fold increase in frequency*. Note that by definition, a 'decibel' or 'dB' is **dB = 20 × log (ratio)** — when used to express voltage or current ratios. So, a 10:1 voltage ratio is 20 dB. Therefore we can say that the gain falls at the rate of *−20 decibels per decade* at higher frequencies. Any circuit with a slope of this magnitude is called a 'first-order filter' (in this case a low-pass one).
- Further, since this slope is *constant*, the signal must also decrease by a factor of 2 for every doubling of frequency. Or a factor of 4 for every quadrupling of frequency, and so on. But a 2:1 ratio is 6 dB, and an "octave" is a doubling (or halving) of frequency. Therefore we can also say that the gain of a low-pass first-order filter falls at the rate of *−6 dB per octave* (at high frequencies).
- If the x and y scales are scaled and *proportioned identically*, the angle the gain plot will make with the x-axis is $-45°$. The slope, that is, tangent of this angle is then $\tan(-45°) = -1$. Therefore, a slope of −20 dB/decade (or −6 dB/octave) is often simply called a "−1" slope.
- Similarly, when we have filters with *two* reactive components (an inductor *and* a capacitor), we will find the slope is −40 dB/decade (i.e. −12 dB/octave). This is usually called a "−2" slope (the angle being about $-63°$).
- We will get a straight-line gain plot in either of the two following cases — *a)* if the gain is expressed as a simple ratio (i.e. Vout/Vin), and plotted on a log scale (on the y-axis), or *b)* if the gain is expressed in decibels (i.e. 20 × log Vout/Vin), and we use a linear scale to plot it. Note that in both cases, on the x-axis, we can either use "f" (frequency) on a log scale, or 20 × log (f) on a linear scale.
- We must remember that the log of 0 is indeterminate ($\log 0 \rightarrow -\infty$), so we must not let the *origin* of a log scale ever be 0. We can set it *close* to zero, say 0.0001, or 0.001, or 0.01, and so on, but certainly not 0.
- The bold gray straight lines in both the right-hand side graphs of *Figure 7-4* form the 'asymptotic approximation.' We see that the gain asymptotes have a *break frequency* or 'corner frequency' at $f = 1/(2\pi RC)$. This point can also be referred to as the 'resonant frequency' of the RC filter.
- The error/deviation from the actual curve is usually very small if we replace it with its asymptotes (for first-order filters). For example, the worst-case error for the gain of the simple RC network is only −3dB, and occurs at the break frequency.

Therefore, the asymptotic approximation is a valid "short cut" that we will often use from now on to simplify the plots and their analysis.

- With regard to the asymptotes of the phase plot, we see that we get *two* break frequencies for it — one at $1/10^{th}$, and the other at 10 times the break frequency of the gain plot. The change in the phase angle at each of these break points is 45° — giving a total phase shift of 90° spanning two decades (symmetrically around the break frequency of the gain plot).
- Note that at the frequency where the single-pole lies, the phase shift (measured from the origin) is always 45° — that is, half the overall shift — *whether we are using the asymptotic approximation or not*.
- Since *both* the gain and the phase *fall* as frequency *increases*, we say we have a 'pole' present — in our case, at the break frequency of $1/(2\pi RC)$. It is also a "single-pole," since it is associated with only a -1 slope.
- Later, we will see that a 'zero' is identifiable by the fact that *both* the gain and phase *rise* with frequency.
- The output voltage is clearly always *less* than the input voltage — at least for a (*passive*) RC network. In other words, the gain is less than 1, at any frequency. Intuitively, that seems right, because there seems to be no way to "amplify" a signal, without using an *active* device — like an op-amp for example. However, as we will soon see, if we use filters that use *both* types of reactive components (L and C), we *can* in fact get the output voltage to exceed the input (but only at certain frequencies). And that incidentally is what we more commonly regard as "resonance."

The Integrator Op-amp ("pole-at-zero" filter)

Before we go on to passive networks involving *two* reactive components, let us look at an interesting *active* RC based filter. The one chosen for discussion here is ***the 'integrator,' because it happens to be the fundamental building block of any 'compensation network.'***

The inverting op-amp presented in *Figure 7-5* has only a capacitor present in its feedback path. We know that under steady dc conditions, all capacitors essentially "go out of the picture." In our case we are therefore left with *no* negative feedback at all *at dc* — and therefore infinite dc gain (though in practice, real op-amps will limit this to a very high, but finite value). But more surprisingly perhaps, that does not stop us from knowing the precise gain at *higher* frequencies. If we calculate the transfer function of this circuit, we will see that something "special" again happens at the point $f = 1/(2\pi \times RC)$. However unlike the passive RC filter, this point is not a break-point (or a pole, or zero location). It happens

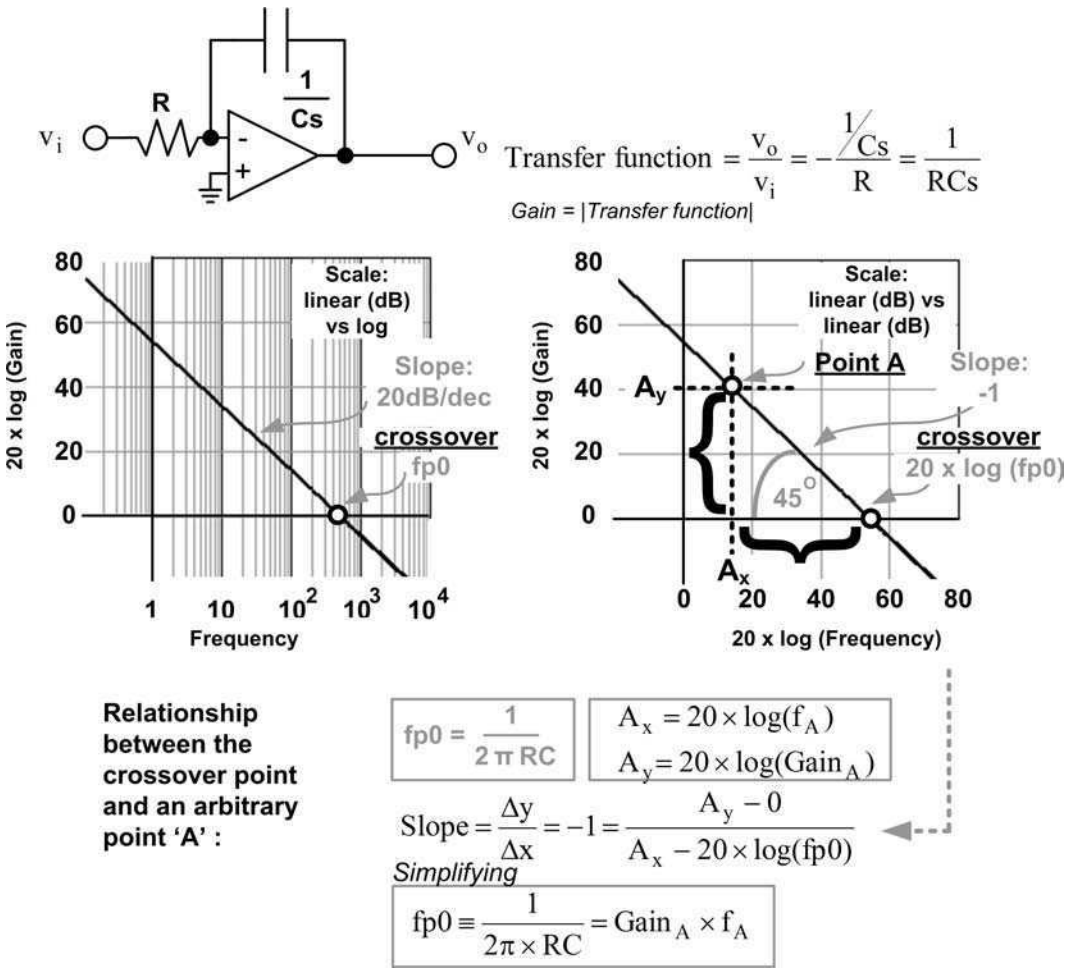


Figure 7-5: The Integrator (pole-at-zero) Operational Amplifier

to be the point where the *gain is unity* (0 dB). We will denote this frequency as "fp0." This is therefore the *crossover frequency of the integrator*. "Crossover" implies that the gain plot intersects the 0 dB (gain = 1) axis.

Note that the integrator has a single-pole — at "zero frequency". Therefore, we will often refer to it as the "pole-at-zero" stage or section of the compensation network. This pole is more commonly called the pole at the origin or the dominant pole.

The basic reason why we will *always* strive to introduce this pole-at-zero is that without it we would have very limited dc gain. The integrator is the simplest way to try and get as high a dc gain as possible.

On the right side of *Figure 7-5*, we have deliberately made the graph perfectly *square* in shape. We have also assigned an *equal* number of grid divisions on the two axes. In addition, to keep the x and y scaling identical, we have plotted *20 × log(f)* on the x-axis (instead of just log(f)). Having thus made the x- and y-axes identical in all respects, we realize why the slope is called "−1" — *it really does fall at exactly 45° now* (visually too).

Therefore, by plotting 20 × log(gain) versus 20 × log(f), we have obtained a straight line with a −1 slope. This allows us to do some simple math as shown in *Figure 7-5*. We have thus derived a useful relationship between an arbitrary point "A" and the crossover frequency 'fp0'

$$\text{fp0} \equiv \frac{1}{2\pi \times \text{RC}} = \text{Gain}_\text{A} \times \text{f}_\text{A}$$

Note that in general, the transfer function of a "pole-at-zero" function such as this will always have the following form

$$\frac{1}{\text{Xs}} \quad \textit{(pole-at-zero transfer function)}$$

The crossover frequency is

$$\text{fcross} = \frac{1}{2\pi\text{X}} \quad \textit{(crossover frequency)}$$

In our case, X is the time constant RC.

Mathematics in the Log Plane

As we proceed toward our ultimate objective of control loop analysis and compensation network design, we will be *multiplying* transfer functions of *cascaded* blocks to get the *overall* transfer function. That is because the output of one block forms the input for the next block, and so on.

It turns out that the mathematics of gain and phase is actually easier to perform in the log plane, rather than in a linear plane. Some simple rules that will help us later are as follows:

- When we combine transfer functions, *decibels add up*. So for example, if we take the product of two transfer functions A and B (cascaded stages), we get C = AB. This follows from the property log(AB) = log(A) + log(B). In words, the gain of A in decibels plus the gain of B in decibels, gives us the gain of C in decibels.
- The overall phase shift is also the *sum of the phase shifts* produced by each of the cascaded stages. So phase angles also add up.

- From the upper half of *Figure 7-6*, we see that ***if we know the crossover frequency (and the slope of the line), we can find the gain at any frequency.***
- Suppose we now shift the line *vertically* (keeping the slope constant) as shown in the lower half of *Figure 7-6*. Then, by the equation provided therein, we can calculate by what amount the crossover frequency shifts in the process.

Transfer Function of the LC Filter

In a buck, there is a *post*-LC filter present. Therefore this filter stage can easily be treated as a cascaded stage following the switch. The overall transfer function is then very easy to compute as per the rules mentioned in the previous section. However, when we come to the boost and buck-boost, we *don't* have a *post*-LC filter — there is a switch/diode connected between the two reactive components that alters the dynamics. However, it can be shown, that even the boost and buck-boost can be manipulated into a 'canonical model' in which an *effective post-LC filter* appears at the output — thus making them as easy to treat as a buck. The only difference is that the original inductance L (of the boost and buck-boost) gets replaced by an equivalent (or *effective*) inductance equal to $\boldsymbol{L/(1-D)^2}$. The "C" remains the same in the canonical model.

Since the LC filter thus becomes representative of the output section of *any* typical switching topology, we need to understand it better, as we now do using *Figure 7-7*:

- For most purposes, we can assume that the break frequency of the gain plot does *not depend on the load* or on the associated parasitic resistive elements of the components. So the resonant frequency of the filter-plus-load combination can be taken to be simply $1/(2\pi\sqrt{(LC)})$, that is, no resistance term is included.
- The LC filter gain *decreases* at the rate of "−2" at high frequencies. The phase also *decreases* providing a total phase shift of 180°. So we say we have a "double-pole" at the break frequency $2\pi\sqrt{(LC)}$.
- Q is the 'quality factor' (as defined in the figure). In effect, it quantifies the amount of "peaking" in the response at the break frequency. Very simply put, if for example $Q = 20$, then the output voltage at the resonant frequency is 20 times the input voltage. On a log scale, this is written as $20 \times \log Q$, as shown in the figure. If Q is very high, the filter is considered "under-damped." If Q is very small, the filter is "over-damped." And if $Q = 0.707$, we have 'critical damping.' In critical damping, the gain at the resonant frequency is 3 dB below its dc value, that is, the output is 3 dB below the input (similar to an RC filter). Note that −3 dB is a factor of $1/\sqrt{2} = 0.707$ — that is, roughly 30% lower. Similarly, +3 dB is $2 = 1.414$ (i.e. roughly 40% higher).

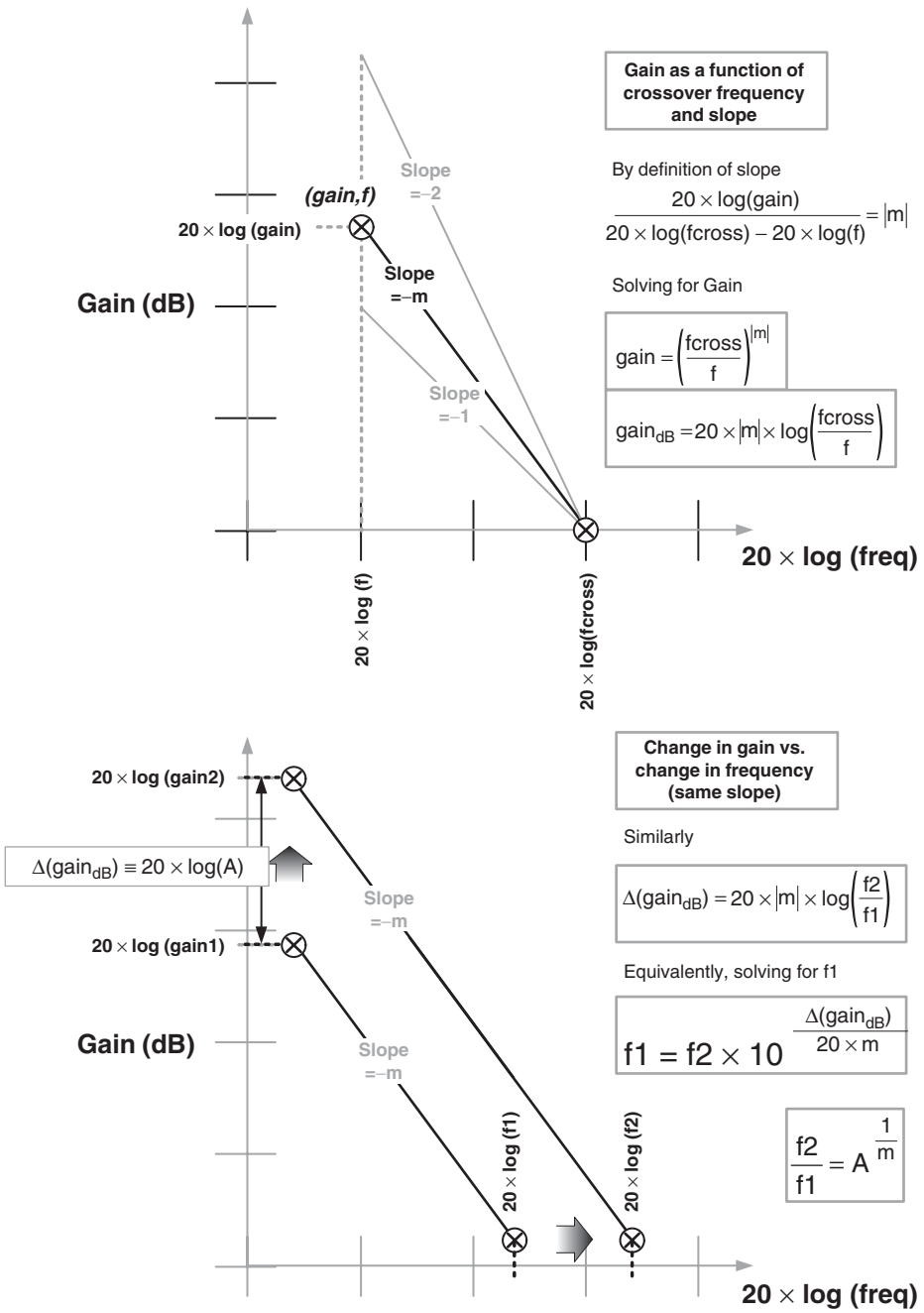


Figure 7-6: Math in the Log Plane

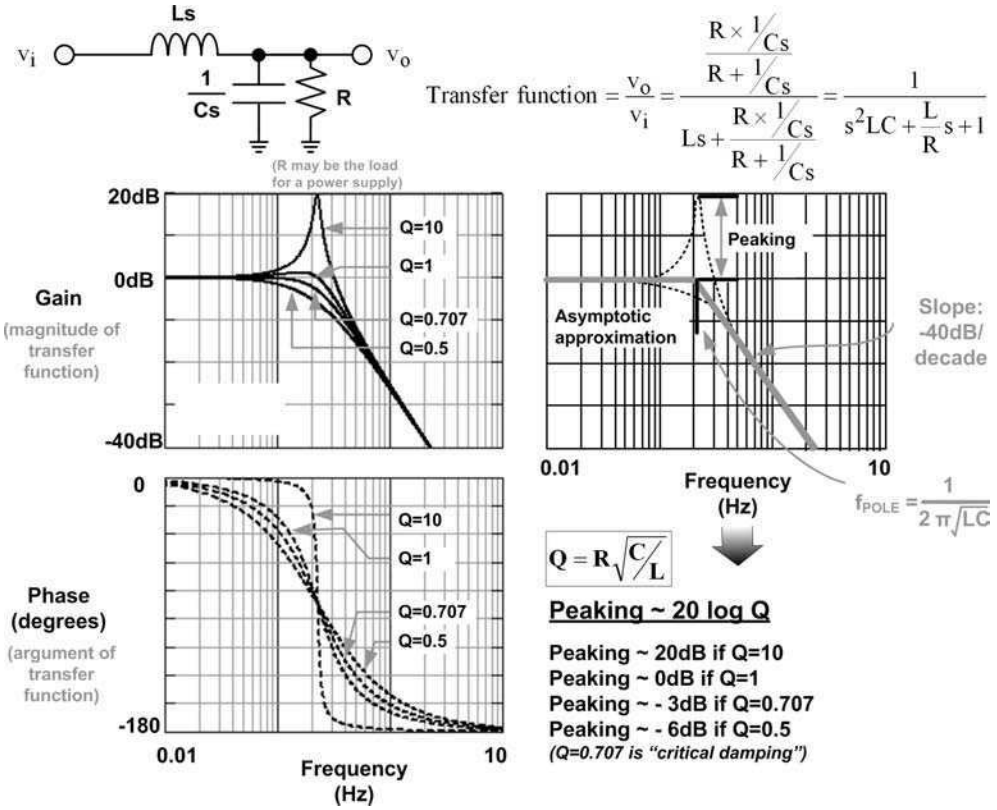


Figure 7-7: The LC Filter Analyzed in the Frequency Domain

- The effect of resistance on the break frequency is usually minor, and therefore ignored. But the effect of resistance on the Q (i.e. the peaking) is significant (though eventually, that is also usually ignored too). However, we should keep in mind that the higher the associated *series* parasitic resistances of L and C, the *lower* the Q. On the other hand, at lower output powers, the resistor *across* the C (i.e. load resistor) is high, and this actually *increases* the Q. Remember that a high *parallel* resistance is in effect a small *series* resistance, and vice versa.
- We can use the asymptotic approximation for the LC *gain* plot as we did for the RC filter. However, the problem with trying to do the same with the *phase* of the LC is that there will be a very large error, more so if the Q (defined in the figure) becomes very large. If so, we can get a very *abrupt* phase shift of 180° close to the resonant frequency. This sudden phase shift in fact can become a real problem in a power supply, since it can induce "conditional stability" (discussed later). Therefore, a certain amount of damping helps from the standpoint of phase and possible conditional stability.

- Unlike an RC filter, the output voltage in this case can be *greater* than the input voltage (at around the break frequency). But for that to happen, Q must be greater than 1.
- Instead of using Q, engineers often prefer to talk in terms of the damping factor defined as

$$\text{damping factor} = \zeta = \frac{1}{2Q}$$

- So a high Q corresponds to a low ζ.

From the equations for Q and resonant frequency, we can conclude that if *L is increased, Q tends to decrease, and if C is increased, Q increases*.

Note: One of the possible pitfalls of putting too much output capacitance in a power supply is that we may be creating significant *peaking* (high Q) in its output filter's response. And we know that when that happens, the phase shift is also more abrupt, and that can induce conditional instability. So generally, if we increase C but simultaneously increase L, we can keep the Q (and the peaking) unchanged.

Summary of Transfer Functions of Passive Filters

The first-order (RC) low-pass filter transfer function can be written in different ways as

$$G(s) = \frac{1/RC}{s + 1/RC} \quad \textit{(RC low-pass)}$$

$$G(s) = \frac{1}{1 + \frac{s}{\omega_0}} \quad \textit{(RC low-pass)}$$

$$G(s) = K\frac{1}{s + \omega_0} \quad \textit{(RC low-pass)}$$

where $\omega_O = 1/(RC)$. Note that the "K" in the last equation is a constant multiplier often used by engineers who are more actively involved in the design of filters. In this case, $K = \omega_O$.

For the second-order filter, various equivalent forms seen in literature are

$$G(s) = \frac{1/LC}{s^2 + s\left(1/RC\right) + 1/LC} \quad \textit{(LC low-pass)}$$

$$G(s) = K\frac{1}{s^2 + \frac{\omega_0}{Q}s + \omega_0^2} \quad \textit{(LC low-pass)}$$

$$G(s) = \frac{1}{\left(\frac{s}{\omega_0}\right)^2 + \frac{1}{Q}\left(\frac{s}{\omega_0}\right) + 1} \quad \textit{(LC low-pass)}$$

$$G(s) = \frac{1}{1 + 2\zeta\left(\frac{s}{\omega_0}\right) + \left(\frac{s}{\omega_0}\right)^2} \quad \textit{(LC low-pass)}$$

where $\omega_O = 1/(LC)^{1/2}$. Note that here, $K = \omega_O^2$. Also, Q is the quality factor, and ζ is the damping factor defined earlier.

Finally, note also that ***the following relations are very useful when trying to manipulate the transfer function of the LC filter into different forms***

$$L/R = 1/\omega_0 Q \quad \text{and} \quad 1/RC = \omega_0/Q \quad \textit{(LC filter)}$$

Poles and Zeros

Let us try to "connect the dots" now. Both the first- and second-order filters we have discussed gave us *poles*. That is because *they both had 's' in the denominators of their transfer functions* — if *s* takes on specific values, it can force the denominator to become zero, and the transfer function then becomes infinite, and we get a *pole* by definition. The values of *s* at which the denominator becomes zero are the resonant (or break) frequencies, that is, the locations of the poles. For example, a hypothetical transfer function "1/s" will give us a pole at *zero frequency* (the "pole-at-zero" we talked about earlier).

Note that the gain, which is the magnitude of the transfer function (calculated by putting $s = j\omega$), won't necessarily be infinite at the pole location. For example, in the case of the RC filter, we know that the gain is in fact *always* less than or equal to unity, despite a pole being present at the break frequency.

Note that if we *interchange the positions* of the two primary components of each of the passive low-pass filters we discussed earlier, we will get the corresponding 'high-pass' RC and LC filters respectively. If we calculate their transfer functions in the usual manner, we will see that besides giving us poles, we also now get single- and double-*zeros* respectively (both at *zero* frequency) as indicated in *Figure 7-8*. So, zeros occur whenever (and wherever) the *numerator* of the transfer function becomes zero.

Zeros are "anti-poles" in many senses. For one, their presence is indicated by both the gain and the phase *increasing with frequency* — opposite to a pole. Further, zeros also "cancel" poles if they happen to fall at the *same* frequency location.

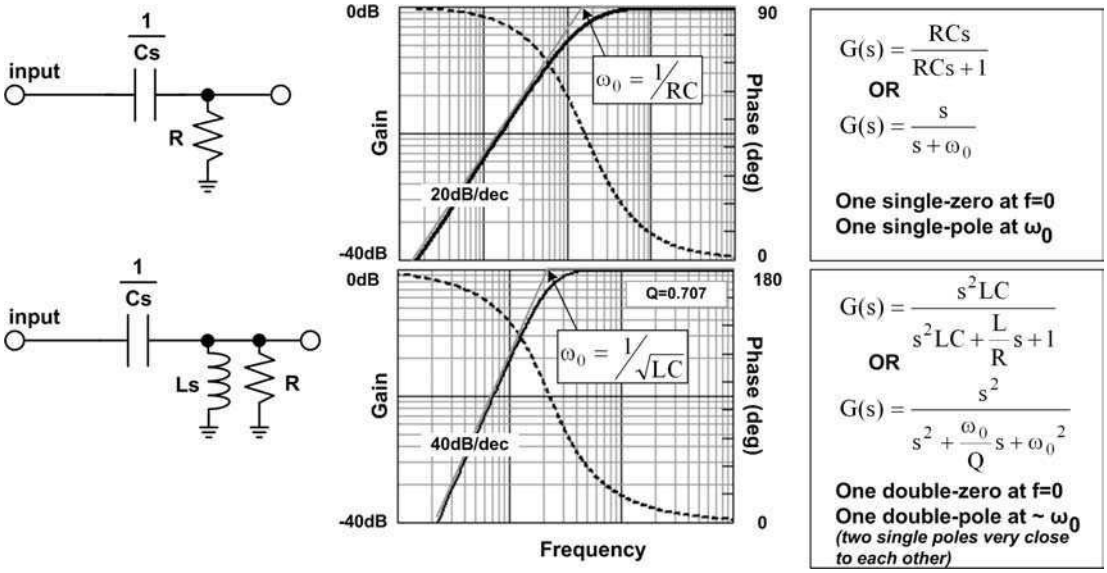


Figure 7-8: High-pass RC and LC (first-order and second-order) filters

We had mentioned that gain-phase plots are called Bode plots. In the case of *Figure 7-8*, we have drawn these on the same graph, just for convenience. Here the solid line is the gain, and to read its value, we need to look at the y-axis on the *left* side of the graph. Similarly, the dashed line is the phase, and for it, we need to look at the y-axis on the *right* side. Note that just for practice, we have once again reverted to plotting the gain (expressed as a simple ratio) on a log scale. The reader should hopefully by now have learnt to correlate the *major* grid divisions of this type of plot with the corresponding dB. So a 10-fold increase is equivalent to +20 dB, a 100-fold increase is +40 dB, and so on.

Now we can generalize our approach. A network transfer function can be described as a ratio of two polynomials:

$$G(s) = \frac{V(s)}{U(s)} = k\frac{a_0 + a_1 s + a_2 s^2 + a_3 s^3 + \ldots}{b_0 + b_1 s + b_2 s^2 + b_3 s^3 + \ldots}$$

This can be factored out as

$$G(s) = K\frac{(s - z_0)(s - z_1)(s - z_2)\ldots}{(s - p_0)(s - p_1)(s - p_2)\ldots}$$

So the zeros (i.e. leading to the numerator being zero) occur at the complex frequencies $s = z_1, z_2, z_3 \ldots$ and so on. The poles (denominator zero) occur at $s = p_1, p_2, p_3 \ldots$ and so on.

In power supplies, we usually deal with transfer functions of the form

$$G(s) = K\frac{(s + z_0)(s + z_1)(s + z_2)\ldots}{(s + p_0)(s + p_1)(s + p_2)\ldots}$$

So the "well-behaved" poles and zeros that we have been talking about are actually in the *left-half* of the complex frequency plane ("LHP" poles and zeros). Their locations are at $s = -z_1, -z_2, -z_3, -p_1, -p_2, -p_3 \ldots$, and so on.

Interaction of Poles and Zeros

We can break up this analysis in two parts:

1. *For poles and zeros lying along the same gain plot* (i.e. belonging to the same stage) — the effect is *cumulative in going from left to right*. So suppose we are starting from zero frequency and move right, toward a higher frequency, and we first encounter a double pole. We know that the gain will start falling with a slope of -2 beyond the resonant frequency point. But as we go further to the right, suppose we now encounter a single-zero. This will impart a *change* in slope of $+1$. So the net slope of the gain plot will become $-2 + 1 = -1$ (after the zero location). Note that despite a zero being present, the gain is still *falling*, though at a lesser rate. In effect, the single-zero canceled *half* the double pole, so we are left with the response of a single pole (to the right of the zero).

   The phase angle also cumulates in a similar manner, except that in practice a phase angle plot is harder to analyze. That is because phase shift can take place slowly over *two decades around* the resonant frequency. We also know that for a double pole (or double-zero), the change in phase may in fact be very abrupt at the resonant frequency. However, *eventually*, the net effect is still *predictable*. So for example, a double pole followed by a single-zero will start with a phase angle of 0° (at dc) and then tend toward –180°. But about a decade below the location of the single-zero, the phase angle will gradually start increasing (though still remaining negative). It will eventually settle down to $-180° + 90° = -90°$ at high frequencies.

2. *For poles and zeros lying along different gain plots* (all coming from cascaded stages) — we know that the overall gain in decibels is the sum of the gain of each (also in decibels). The effect of this math on the pole-zero interactions is therefore simple to describe. If for example, *at a specific frequency*, we have a double pole in one plot and a single-zero on the other plot, then the overall gain plot will have a single pole at this break frequency. So we see that poles and zeros tend to "destroy" each other, as we would expect since zeros are "anti-poles" as mentioned previously.

But poles and zeros also add up *with their own type*. For example, if we have a double pole on one plot, and a single-pole on the other plot (at the same frequency), the net gain will fall with a slope of −3 after the break frequency. Phase angles also add up similarly.

These rules will become even clearer, a little later, when we actually try to work out the open-loop gain of a converter.

Closed and Open Loop Gain

Figure 7-9 represents a general feedback controlled system. The 'plant' (also sometimes called the "modulator") has a 'forward transfer function' G(s). A part of the output gets fed back through the feedback block to the control input, so as to produce regulation at the output. Along the way, the feedback signal is compared with a reference level, which tells it what the desired level is for it to regulate to.

H(s) is the "feedback transfer function," and we can see this goes to a summing block (or node) — represented by the circle with an enclosed summation sign.

Note: The summing block is sometimes shown in literature as just a simple circle (nothing enclosed), but sometimes rather confusingly as a circle with a multiplication sign (or x) inside it. Nevertheless, *it still is a summation block*.

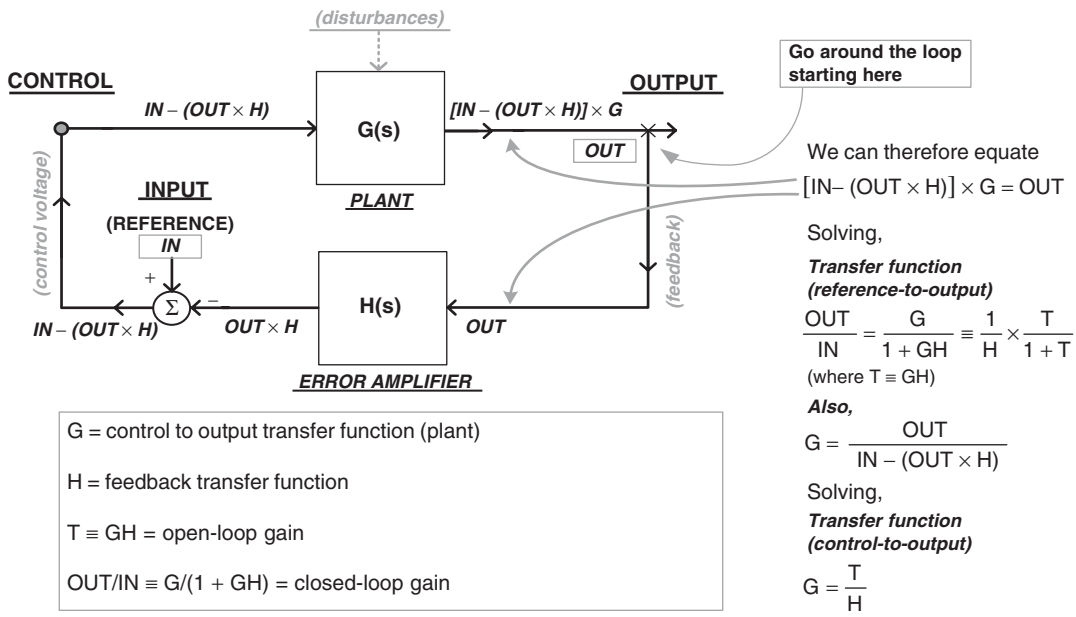


Figure 7-9: General Feedback Loop Analysis

One of the inputs to this summation block is the reference level (the 'input' from the viewpoint of the control system), and the other is the output of the feedback block (i.e. the part of the output being fed back). The output of the summation node is therefore the 'error' signal.

Comparing *Figure 7-9* with *Figure 7-10*, we see that in a power supply, the *plant* itself can be split into several cascaded blocks. These blocks are — the *pulse width modulator* (not to be confused with the term 'modulator' used in general control loop theory for the *entire plant* itself), the *power stage* consisting of the driver-plus-switch, and the *LC filter*.
The feedback block, on the other hand, consists of the voltage divider (if present) and the compensated error amplifier. Note that we may prefer to visualize the error amplifier block as two cascaded stages — one that just computes the error (summation node), and another that accounts for the gain (and its associated compensation network). Note that the basic principle behind the pulse width modulator stage (which determines the shape of the pulses driving the switch), is explained in the next section, and in *Figure 7-11*.

In general, the plant can receive various 'disturbances' that can affect its output. In a power supply these are essentially the *line and load variations*. The basic purpose of feedback is to reduce the effect of these disturbances on the output voltage.

In *Figure 7-9* we have derived the *open-loop gain*, which is simply the magnitude of the product of the forward and feedback transfer functions — that is, obtained by going *around* the loop. On the other hand, the magnitude of the *reference-to-output* transfer function is called the *closed-loop gain*. Note that the word "closed" has really nothing to do with the feedback loop being *literally* "open" or "closed." Further, "GH" is called the 'open-loop transfer function' — again, irrespective of whether the loop is literally "open," say for the purpose of measurement, or "closed" as in normal operation. In fact, in a typical power supply, we can't ever even hope to break the feedback path. Because the gain is typically so high, even a minute change in the feedback voltage will cause the output to swing wildly. So in fact, we always need to "close" the loop (and thereby dc-bias the converter into regulation), before we can measure the so-called "open-loop" gain.

As a further proof of this, note that in *Figure 7-9*, if we go around the cascaded stage consisting of G and H, and calculate the ratio of the input signal to the output, we get

$$\frac{\mathrm{IN}-\left(\mathrm{OUT}\times\mathrm{H}\right)}{\mathrm{OUT}\times\mathrm{H}}=\frac{\mathrm{IN}}{\mathrm{OUT}\times\mathrm{H}}-1=\left(\frac{1+\mathrm{GH}}{\mathrm{G}}\right)\frac{1}{\mathrm{H}}-1=\frac{1}{\mathrm{GH}}+1-1=\frac{1}{\mathrm{GH}}$$

Therefore, the ratio of the *output to the input*, that is, the transfer function of the cascaded G and H blocks, is equal to GH — which is simply the open-loop gain. Therefore, even with the loop "closed," as we go around, we are always going to get the *open-loop gain* GH. Note that the phrase "closed-loop gain" actually refers to the change in the output, if we change the *reference voltage* slightly.

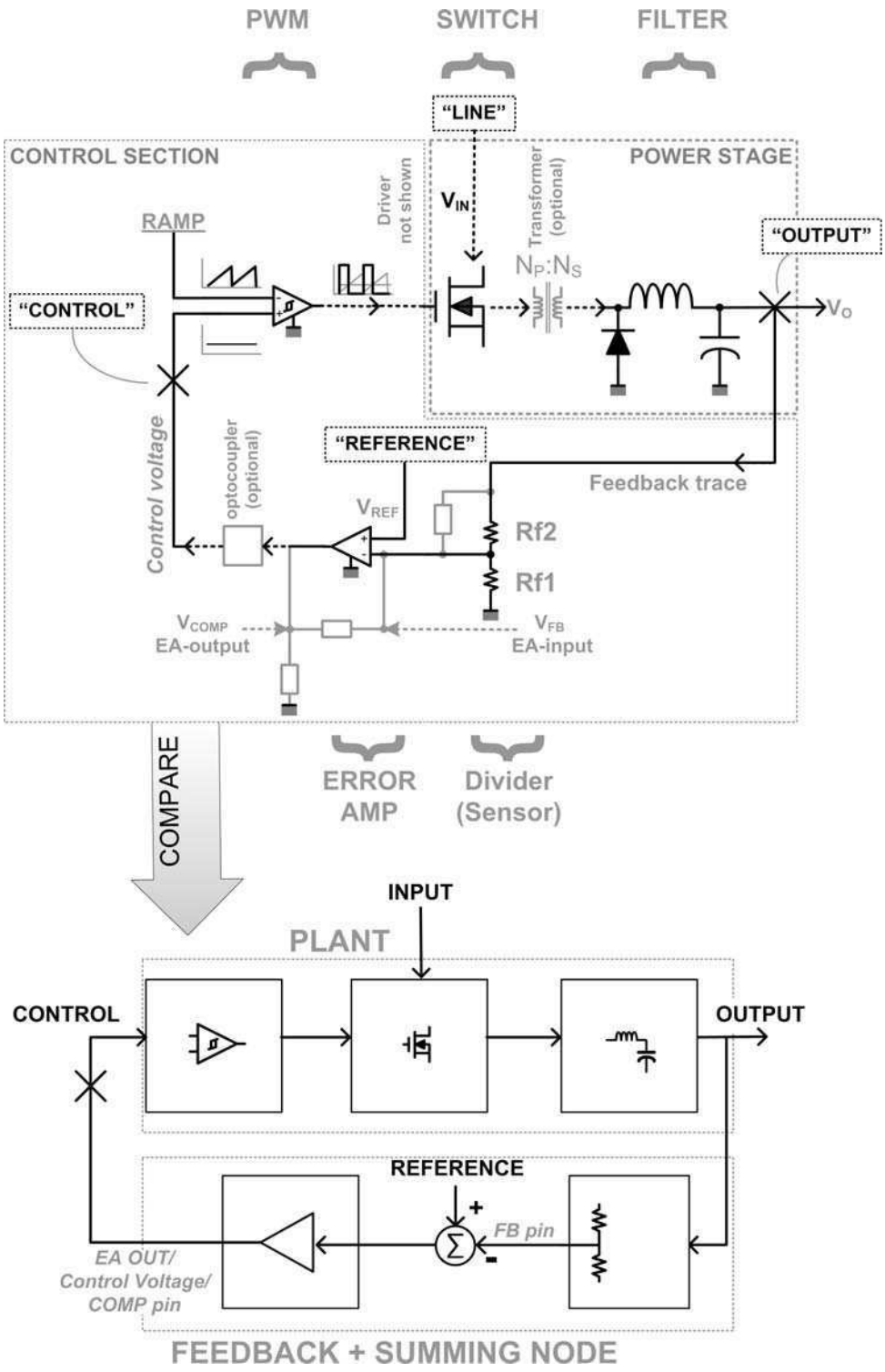


Figure 7-10: A Power Converter and Its Plant and Feedback Blocks

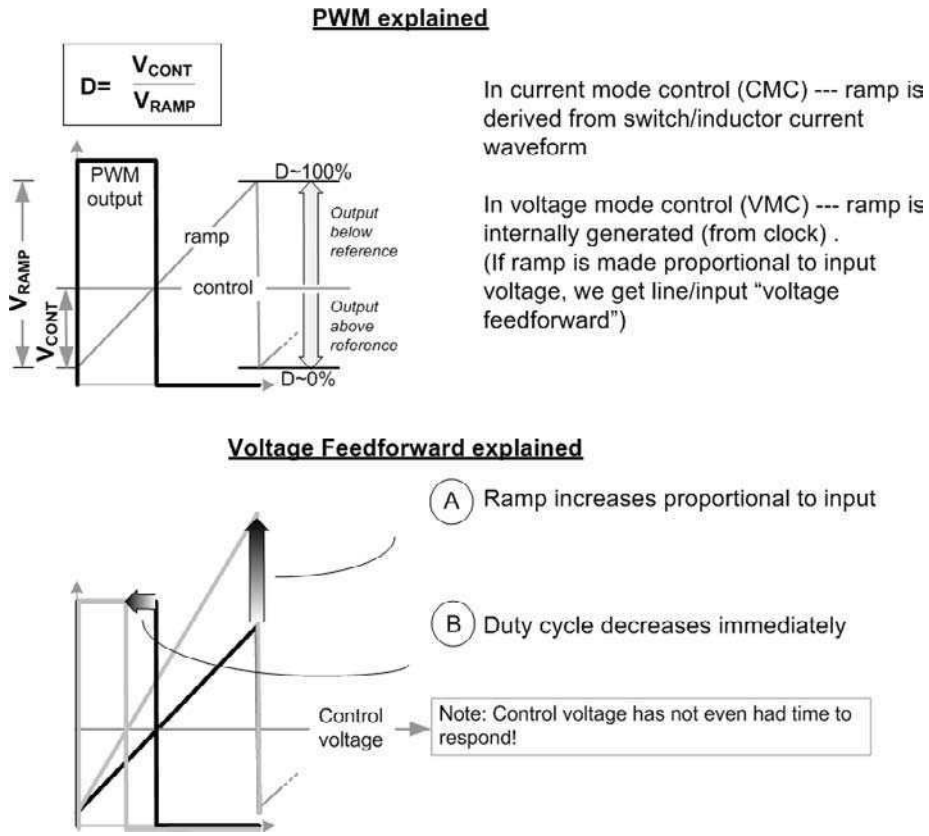


Figure 7-11: Combining Blocks and Thus Showing That "Open-loop" Gain Is Actually in a Closed Loop

The Voltage Divider

The output V_O of the power supply first goes to a *voltage divider*. Here it is in effect, just stepped-down, for subsequent comparison with the reference voltage 'V_{REF}.' The comparison takes place at the input of the *error-amplifier,* which is usually just a conventional op-amp (voltage amplifier).

We can visualize an ideal op-amp as a device that varies its output so as to virtually equalize the voltages at its input pins. Therefore in steady state, the voltage at the node connecting Rf2 and Rf1 (see "divider" block in *Figure 7-10*) can be assumed to be (almost) equal to V_{REF}. Assuming that no current flows out of (or into) the divider at this node, using Ohm's law

$$\frac{Rf1}{Rf1 + Rf2} = \frac{V_{REF}}{V_O}$$

Simplifying,

$$\frac{Rf2}{Rf1} = \frac{V_O}{V_{REF}} - 1$$

So this tells us what *ratio* of the voltage divider resistors we must have, to produce the desired output rail.

Note however, that in applying control loop theory to power supplies, we are actually looking only at *changes (*or perturbations), ***not the dc values*** (though this was not made obvious in *Figure 7-9*). ***It can also be shown that when the error amplifier is a conventional op-amp, the lower resistor of the divider, Rf1, behaves only as a dc biasing resistor and does not play any (direct) part in the ac loop analysis.***

Note: The lower resistor of the divider, Rf1, does not enter the ac analysis, provided we are considering *ideal* op-amps. In practice, it *does* affect the bandwidth of a real op-amp, and therefore may on occasion need to be considered.

Note: If we are using a spreadsheet, we will find that changing Rf1 does in fact affect the overall loop (even when using conventional op-amp-based error amplifiers). But we should be clear that that is only because by changing Rf1, we have changed the duty cycle of the converter (its output voltage), which thus affects the plant transfer function. Therefore, in that sense, the effect of Rf1 is only *indirect*. We will see that Rf1 does not actually enter into any of the equations that tell us the locations of the poles and zeros of the system.

Pulse Width Modulator Transfer Function (gain)

The *output of the error amplifier* (sometimes called "COMP," sometimes "EA-out," sometimes "control voltage") is applied to one of the inputs of the pulse width modulator ('PWM') comparator. On the other input of this comparator, we have an applied sawtooth voltage ramp — either internally generated from the clock when using "voltage mode control," or derived from the current ramp when using "current mode control." Thereafter, by normal comparator action, we get pulses of desired width, with which to drive the switch.

Since the feedback signal coming from the output rail of the power supply goes to the *inverting* input of the error-amplifier, if the output is below the set regulation level, the output of the error amplifier goes high. This causes the pulse width modulator to increase pulse width (duty cycle) and thus try and make the output voltage rise. Similarly, if the output of the power supply goes above its set value, the error amplifier output goes low, causing the duty cycle to decrease. See the upper half of *Figure 7-11*.

As mentioned previously, the output of the pulse width modulator stage is *duty cycle*, and its input is the 'control voltage' or the 'EA-out.' So, as we said, the gain of this stage is not a

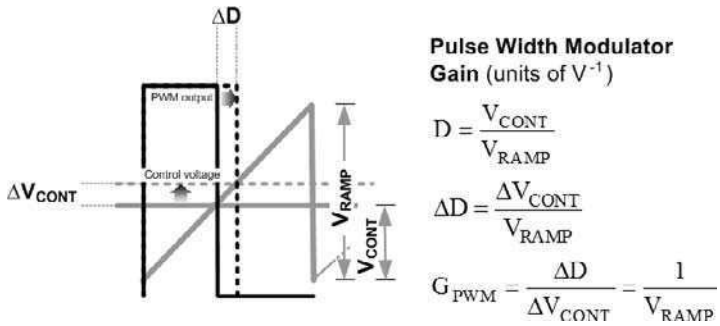


Figure 7-12: Gain of Pulse Width Modulator

dimensionless quantity, but has units of 1/V. From *Figure 7-12* we can see that this gain is equal to $1/V_{RAMP}$, where V_{RAMP} is the peak-to-peak amplitude of the ramp sawtooth.

Voltage Feedforward

We had also mentioned previously, that when there is a disturbance, the control does not *usually* know beforehand how much duty cycle correction to apply. In the lower half of *Figure 7-11*, we have described an increasingly popular technique being used to make that really happen (when faced with *line* disturbances). This is *called input-voltage/line feedforward*, or simply "feedforward."

This technique requires that the input voltage be sensed, and the slope of the comparator sawtooth ramp increased, if the input goes up. In the simplest implementation, a doubling of the input causes the slope of the ramp to double. Then, from *Figure 7-11*, we see that if the slope doubles, the duty cycle is immediately halved — as would be required anyway if the input to a *buck* converter is doubled. So the duty cycle correction afforded by this "automatic" ramp correction is exactly what is required (for a buck, since its duty cycle $D = V_O/V_{IN}$). But more importantly, this correction is virtually *instantaneous* — we didn't have to wait for the *error amplifier to detect the error on the output (through the inherent delays of its RC-based compensation network scheme), and respond by altering the control voltage.* So in effect, by input feedforward, we have *bypassed all major delays*, and so line correction is virtually instantaneous (i.e. "perfect" rejection of disturbance).

We just stated that in its *simplest* form, feedforward causes the duty cycle to halve if the input doubles. Let us double-check that that is really what is required here. From the dc input-to-output transfer function of a *buck* topology,

$$D = \frac{V_O}{V_{IN}}$$

Therefore, on doubling the input,

$$\frac{V_O}{(2 \times V_{IN})} = \frac{D}{2}$$

which is what we are doing anyway, through feedforward. So, we can see that this simple feedforward technique will work very well for a buck. However, for a boost or a buck-boost, the duty-cycle-to-input proportionality is clearly not going to be the best answer.

Using voltage feedforward to produce automatic line rejection is clearly applicable only to *voltage mode control*. However, the original inspiration for this idea came from *current mode control*. But in current mode control, the ramp to the PWM comparator is being generated from the inductor current waveform. In a buck topology for example, the *slope* of the inductor current up-ramp is equal to $(V_{IN} - V_O)/L$. So if we double the input voltage, we do *not* end up doubling the slope of the inductor current. Therefore, neither do we end up halving the duty cycle. But in fact, we need to do exactly that, if we are looking for complete line rejection (from $D = V_O/V_{IN}$). In other words, voltage mode control with feedforward in fact provides *better* line rejection than current mode control (for a buck).

Power Stage Transfer Function

The power stage formally consists of the switch plus the (equivalent) LC filter. Note that this is just the plant *minus* the pulse width modulator. (see *Figure 7-10*).

We had indicated previously that whereas in a buck, the L and C are really connected to each other at the output (as drawn in *Figure 7-10*), in the remaining two topologies they are *not*. However, the small-signal (canonical) model technique can be used to transform these latter topologies into equivalent ac models — in which, for all practical purposes, a regular LC-filter does appear after the switch, just as for a buck. With this technique, we can then justifiably separate the *power stage* into a *cascade* of two separate stages (as for a buck):

- A stage that effectively converts the duty cycle input (coming from the output of the PWM stage), into an output voltage
- An equivalent post-LC filter stage, that takes in this output and converts it into the output rail of the converter

With this understanding, we can build the final transfer functions presented in the next section.

Plant Transfer Functions of All the Topologies

Let us discuss the three major topologies separately here. Note that we are assuming *voltage mode control* and *continuous conduction mode*. Further, the "*ESR zero" is also not included here* (introduced later).

Buck Converter

Control-to-Output Transfer Function

The transfer function of the *plant* is also called the 'control-to-output transfer function' (see *Figure 7-10*). It is therefore the output voltage of the converter, divided by the 'control voltage' (EA-out). We are of course talking only from an *ac* point of view, and are therefore interested only in the *changes* from the dc-bias levels.

The control-to-output transfer function is a product of the transfer functions of the PWM, the switch and the LC filter (since these are *cascaded* stages). Alternatively, the control-to-output transfer function is a product of the transfer functions of the PWM stage and the transfer function of the 'power stage'.

We already know from *Figure 7-12* that *the transfer function of the PWM stage is equal to the reciprocal of the amplitude of the ramp*. And as discussed in the previous section, the power stage itself is a cascade of an equivalent post-LC stage (whose transfer function is the same as the passive low-pass second-order LC filter we discussed previously), *and a stage that converts the duty cycle into an output voltage*. We are interested in finding the transfer function of this *latter stage*.

The question really is — what happens to the output when we perturb the duty cycle slightly (keeping the input to the converter, V_{IN}, constant)? Here are the steps for a buck:

$$V_O = D \times V_{IN} \quad \textit{(buck)}$$

Therefore, differentiating

$$\frac{dV_0}{dD} = V_{IN}$$

And this is the required transfer function of the intermediate *duty-cycle-to-output* stage!

Finally, the control-to-output transfer function is the product of three (cascaded) transfer functions, that is, it becomes

$$\frac{1}{V_{RAMP}} \times V_{IN} \times \frac{1/LC}{s^2 + s\left(1/RC\right) + 1/LC} \quad \textit{(buck: plant transfer function)}$$

Alternatively, this can be written as

$$\frac{1}{V_{RAMP}} \times V_{IN} \times \frac{1}{\left(\frac{s}{\omega_0}\right)^2 + \left(\frac{s}{\omega_0 Q}\right) + 1} \quad \textit{(buck: plant transfer function)}$$

where $\omega_0 = 1/\sqrt{(LC)}$, and $\omega_0 Q = R/L$.

Line-to-output Transfer Function

Of primary importance in any converter design is not what happens to the output when we perturb the *reference* (which is what the *closed loop* transfer function really is), but what happens at the output when there is *a line disturbance*. This is often referred to as 'audio susceptibility' (probably because early converters switching at around 20 kHz would emit audible noise under this condition).

The equation connecting the input and output voltages is simply the dc input-to-output transfer function, that is,

$$\frac{V_O}{V_{IN}} = D \quad \textit{(buck)}$$

So this is also the factor by which the input disturbance first gets scaled, and thereafter applied at the input of the LC filter. But we already know the transfer function of the LC low-pass filter. Therefore, the line-to-output transfer function is the product of the two, that is,

$$D \times \frac{1/LC}{s^2 + s\left(1/RC\right) + 1/LC} \quad \textit{(buck: line transfer function)}$$

where R is the load resistor (at the output of the converter).

Alternatively, this can be written as

$$D \times \frac{1}{\left(\frac{s}{\omega_0}\right)^2 + \frac{s}{\omega_0 Q} + 1} \quad \textit{(buck: line transfer function)}$$

where $\omega_0 = 1/\sqrt{(LC)}$, and $\omega_0 Q = R/L$.

Boost Converter

Control-to-Output Transfer Function

Proceeding similar to the buck, the steps for this topology are

$$V_O = \frac{V_{IN}}{1 - D}$$
$$\frac{dV_O}{dD} = \frac{V_{IN}}{(1 - D)^2}$$

So the control-to-output transfer function is

$$\frac{1}{V_{RAMP}} \times \frac{V_{IN}}{(1 - D)^2} \times \frac{{}^{1}\!/\!_{\underline{L}C} \times \left(1 - s\left({}^{\underline{L}}\!/\!_{R}\right)\right)}{s^2 + s\left({}^{1}\!/\!_{RC}\right) + {}^{1}\!/\!_{\underline{L}C}} \quad \textit{(boost: plant transfer function)}$$

where $\underline{L} = L/(1 - D)^2$. Note that *this is the inductor in the "equivalent post-LC filter" of the canonical model. Note that C remains unchanged.*

This can also be written as

$$\frac{1}{V_{RAMP}} \times \frac{V_{IN}}{(1 - D)^2} \times \frac{\left(1 - \frac{s}{\omega_{RHP}}\right)}{\left(\frac{s}{\omega_0}\right)^2 + \frac{s}{\omega_0 Q} + 1} \quad \textit{(boost: plant transfer function)}$$

where $\omega_0 = 1/\sqrt{(\underline{L}C)}$, and $\omega_0 Q = R/\underline{L}$.

Note that we have included a *surprise term* in the numerator above. By detailed modeling it can be shown that both the boost and the buck-boost have such a term. This term represents a zero, but a different type to the "well-behaved" zero discussed so far (note the sign in front of the *s*-term). If we consider its contribution separately, we will find that as we raise the frequency, the gain will *increase* (as for a normal zero), but simultaneously, the phase angle will *decrease* (opposite to a "normal" zero, more like a "well-behaved" pole). We will see later that if the overall open-loop phase angle drops sufficiently low, the converter can become unstable because of this zero. That is why this zero is considered undesirable. Unfortunately, it is virtually impossible to compensate for (or "kill") by normal techniques. The only easy route is literally to "push it out" — to higher frequencies where it can't affect the overall loop significantly. Equivalently, we need to reduce the bandwidth of the open-loop gain plot to a frequency low enough that it just doesn't "see" this zero. In other words, the crossover frequency must be set much lower than the location of this zero.

The name given this zero is the *'RHP zero'*, as indicated earlier — to distinguish it from the "well-behaved" (conventional) left-half-plane zero. For the boost topology, its location can be found by setting the numerator of the transfer function above to zero, that is, $s \times (\underline{L}/R) = 1$. So the frequency location of the boost RHP zero is

$$f_{RHP} = \frac{R \times (1 - D)^2}{2\pi \underline{L}} \quad (boost)$$

Note that the very existence of the RHP zero in the boost and buck-boost can be traced back to the fact that these are the *only* topologies where an actual LC post-filter doesn't exist on the output. Though, by using the canonical modeling technique, we have managed to create an *effective* LC post filter, the fact that in reality there is a switch/diode connected between the actual L and C of the topology, is what is ultimately responsible for creating the RHP zero.

Note: Intuitively, the RHP zero is often explained as follows — if we suddenly increase the load, the output dips slightly. This causes the converter to increase its duty cycle in an effort to restore the output. Unfortunately, for both the boost and the buck-boost, energy is delivered to the load *only during the switch off-time*. So, an increase in the duty cycle decreases the off-time, and there is now, unfortunately, a *smaller* interval available for the stored inductor energy to get transferred to the output. Therefore, the output voltage, instead of increasing as we were hoping, dips even further for a few cycles. This is the RHP zero in action. *Eventually*, the current in the inductor does manage to ramp up over several successive switching cycles to the new level consistent with the increased energy demand, and so this strange situation gets corrected — provided full instability has not already occurred!

Line-to-output Transfer Function

We know that

$$\frac{V_O}{V_{IN}} = \frac{1}{1 - D} \quad (boost)$$

Therefore we get

$$\frac{1}{1 - D} \times \frac{{}^{1}\!/_{\underline{L}C}}{s^2 + s\left({}^{1}\!/_{RC}\right) + {}^{1}\!/_{\underline{L}C}} \quad (boost: line\ transfer\ function)$$

Alternatively, this can be written as

$$\frac{1}{1 - D} \times \frac{1}{\left(\frac{s}{\omega_0}\right)^2 + \frac{s}{\omega_0 Q} + 1} \quad (boost: line\ transfer\ function)$$

where $\omega_0 = 1/\sqrt{(\underline{L}C)}$ and $\omega_0 Q = R/\underline{L}$.

Buck-boost Converter

Control-to-output Transfer Function

Here are the steps for this topology:

$$V_O = \frac{V_{IN} \times D}{1 - D}$$
$$\frac{dV_O}{dD} = \frac{V_{IN}}{(1 - D)^2}$$

(Yes, it is an interesting coincidence — the slope of $1/(1 - D)$ calculated for the boost is the same as the slope of $D/(1 - D)$ calculated for the buck-boost!)

So the control-to-output transfer function is

$$\frac{1}{V_{RAMP}} \times \frac{V_{IN}}{(1 - D)^2} \times \frac{{}^{1}\!/_{\underline{L}C} \times \left(1 - s\left({}^{\underline{L}D}\!/_{R}\right)\right)}{s^2 + s\left({}^{1}\!/_{RC}\right) + {}^{1}\!/_{\underline{L}C}} \quad \textit{(buck-boost: plant transfer function)}$$

where $\underline{L} = L/(1 - D)^2$ is the inductor in the *equivalent* post-LC filter.

Alternatively, this can be written as

$$\frac{1}{V_{RAMP}} \times \frac{V_{IN}}{(1 - D)^2} \times \frac{\left(1 - \frac{s}{\omega_{RHP}}\right)}{\left(\frac{s}{\omega_0}\right)^2 + \frac{s}{\omega_0 Q} + 1} \quad \textit{(buck-boost: plant transfer function)}$$

where $\omega_0 = 1/\sqrt{(\underline{L}C)}$, and $\omega_0 Q = R/\underline{L}$.

Note that, as for the boost, we have included the RHP zero term in the numerator. Its location is given by

$$f_{RHP} = \frac{R \times (1 - D)^2}{2\pi L \times D} \quad \textit{(buck-boost)}$$

Line-to-output Transfer Function

We know that

$$\frac{V_O}{V_{IN}} = \frac{D}{1 - D} \quad \textit{(buck-boost)}$$

Therefore,

$$\frac{D}{1-D} \times \frac{{}^{1}\!/_{LC}}{s^2 + s\left({}^{1}\!/_{RC}\right) + {}^{1}\!/_{LC}} \quad \textit{(buck-boost: line transfer function)}$$

This is alternatively written as

$$\frac{D}{1-D} \times \frac{1}{\left(\frac{s}{\omega_0}\right)^2 + \frac{s}{\omega_0 Q} + 1} \quad \textit{(buck-boost: line transfer function)}$$

where $\omega_0 = 1/\sqrt{(LC)}$ and $\omega_0 Q = R/L$.

Note that the plant and line transfer functions of all the topologies do not depend on the load current. That is why gain-phase plots (Bode plots) do not change much if we vary the load current (provided we stay in CCM).

Note also that so far we have ignored a key element of the transfer functions — the ESR of the output capacitor. Whereas the DCR usually just ends up decreasing the overall Q (less "peaky" at the second-order (LC) resonance), the *ESR actually contributes a zero to the open-loop transfer function.* Because it affects the gain and the phase significantly, it usually can't be ignored — certainly not if it lies *below* the crossover frequency (at a lower frequency).

Feedback Stage Transfer Functions

Let us now lump the entire feedback section, including the voltage divider, error amplifier, and the compensation network. However, depending on the *type of error amplifier used*, this must be evaluated rather differently. In *Figure 7-13* we have shown two possible error amplifiers often used in power converters.

The analysis for these two cases is as follows

- The error amplifier can be a simple voltage-to-voltage amplification device, that is, the traditional "op-amp" (operational amplifier). This type of op-amp requires *local* feedback (between its output and inputs) to make it stable. Under steady dc conditions, both the input terminals are virtually at the same voltage level. This determines the output voltage setting. But, as discussed previously, though both resistors of the voltage divider affect the dc level of the converter's output, from the ac point of view, *only the upper resistor enters the picture*. So the lower resistor is considered purely a dc *biasing resistor*, and therefore we usually ignore it in control loop (ac) analysis.

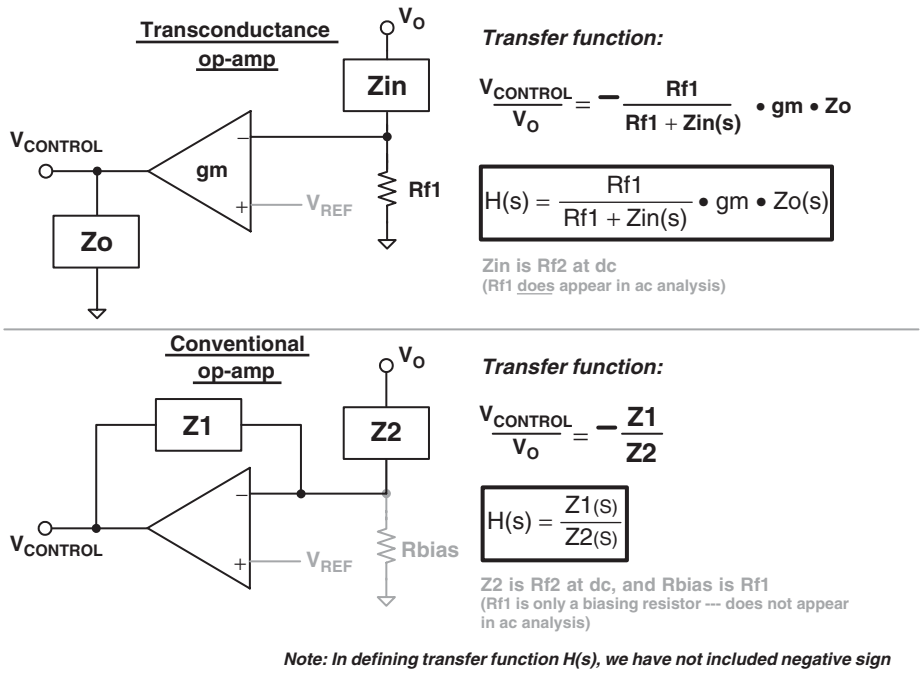


Figure 7-13: Generic Representation of Feedback Stages

- The error amplifier can also be a voltage-to-current amplification device, that is, the "gm op-amp" (transconductance operational amplifier). This is an open-loop amplifier stage with no local feedback — the loop is in effect completed externally, and that again causes the voltages at its input terminals to get equalized (like a regular op-amp). If there is any difference in voltage between its two pins, ΔV, it converts that into a current, ΔI, flowing out of its output pin (as determined by its transconductance $gm = \Delta I/\Delta V$). Thereafter, since there is an impedance Z connected from the output of this op-amp to ground, the *voltage* at the output pin of this error amplifier (i.e. the voltage across Z — also the control voltage) changes by an amount equal to $\Delta I \times Z$. For example, in our converter, if V_{FB} (the voltage from the divider, applied to the inverting pin) *increases* slightly above V_{REF}, this will cause the op-amp to source *less* current. That will *decrease* the control voltage (across Z) and so the duty cycle will *decrease*. Ultimately, because of the high gain, the system again settles down only when the voltages on the input pins of the op-amp become virtually the same. ***For the gm op-amp, both Rf2 and Rf1 enter into the ac analysis***, because together they determine the error voltage at the pins, and therefore the current at the output of the op-amp. Note that the divider can in this case be treated as a simple (step-down) *gain block* of Rf1/(Rf1 + Rf2), cascaded with the op-amp stage that follows.

Note: We may have wondered — why do we always use the *inverting* terminal of the error amplifier for applying the feedback voltage? The intuitive reason for that is that an *inverting* op-amp has a dc gain of Rf/Rin, where Rf is the feedback resistor (from the output of the op-amp to its negative input terminal), and Rin is the resistor between its inverting terminal and the input voltage source. So the output of an inverting op-amp can be made smaller than its input, if so desired (i.e. gain <1). Whereas, a *noninverting* op-amp has a dc gain of 1+ (Rf/Rin), where R_{in} in this case is the resistor between its inverting terminal and ground. So its output will always be greater than its input (gain >1). This restriction on the dc gain has been known to cause some strange and embarrassing situations in the field, especially under certain abnormal conditions. Therefore, a non-inverting error op-amp is generally not favored.

Lastly, note that by just using an *inverting* error amplifier, we have in effect also applied a $-180°$ phase shift "right off the bat"! We will see in the following section, that this increases the possibility of oscillations by itself.

Closing the Loop

We are now in a position to start tying all the loose ends together! For each of the three topologies, we now know both the forward transfer function G(s) (control-to-output) and also the feedback transfer function H(s). Going back to the basic equation for the closed-loop transfer function,

$$\frac{G(s)}{1 + G(s)H(s)} \quad \textit{(closed-loop transfer function)}$$

we see that it will "explode" if

$$G(s)H(s) = -1$$

But G(s)H(s) is simply the transfer function for a signal going through the G(s) block, and then through the H(s) block, that is, the *open-loop* transfer function. We know that the gain is the magnitude of the transfer function (using $s = j\omega$), and its phase angle is its argument. Let us calculate what these are for the transfer function of -1 above.

$$\text{Gain} = \|-1\| = \sqrt{\text{Re}^2 + \text{Im}^2} = \sqrt{(-1)^2 + 0^2} = 1 \quad \textit{(magnitude)}$$

$$\text{Phase} = \phi = \tan^{-1}\left(\frac{\text{Im}}{\text{Re}}\right) = \tan^{-1}\left(\frac{0}{-1}\right) = \tan^{-1}(0) = 180° \quad \textit{(argument)}$$

Note: When doing the $\tan^{-1}$ operation, we may often need to *visualize* where the number is actually located in the complex plane. For example, in this case, tan of 0° and tan of 180° are both zero, and we wouldn't have known which of these angles is the right answer — unless we *visualized* the number in the complex plane. In our case, since the number was *minus* 1, we correctly placed it at 180° instead of 0°.

So we see that ***the system is unstable if a disturbance (of certain frequency) goes through the plant and feedback blocks, and returns at 180° angle, with the same magnitude***.
The summing block that follows has one negative and one positive input, since it represents a *negative feedback* system. This implies that another 180° shift will occur after the signal leaves the H block. *So we conclude that in general, a system would be unstable if a disturbance goes around the loop and comes back to the same point, with the same magnitude and the same phase.*

In a typical gain vs. frequency plot, we will see that a gain of 1 usually occurs at only one specific frequency, and this is called the 'crossover' frequency. Beyond this point the gain becomes less than 1 (i.e. below the 0 dB axis).

The stability criterion is therefore equivalent to saying that ***the phase shift (of the open-loop transfer function) should not be equal to 180° (or −180°) at the crossover frequency***. But we need to ensure a certain *margin of safety* — in terms of the degrees of phase angle at the crossover frequency required to *prevent* this from happening. This safety margin is called the *'phase margin.'*

Note that the possibility of instability applies *only at the crossover frequency*. For example, it surprisingly doesn't matter even if the response somehow manages to come back *larger than the stimulus itself, even at the same phase angle* — that does *not* cause instant instability, simply because it can be shown this particular vector condition can't support any *further increase* in the amount of disturbance.

How much phase margin is enough? In theory, even an overall phase shift of −179° (i.e. a phase margin of 1°) would not produce full instability — though there would certainly be a lot of ringing at every transient, and at best it would be very, very, marginally stable. But component tolerances, temperature variations, and even small changes in the application conditions can change the loop characteristics significantly, ushering in full-blown instability.

It is generally recommended that the phase lag introduced by the successive G and H blocks be about *45° short of −180°* — that is, an overall phase lag of −135°. That gives us a phase margin of 45°. On the other hand, a phase margin of, say, 80° is certainly very stable, but is also usually not very desirable. Under transients, though there is no ringing (after the first overshoot or undershoot), the correction is too slow, and the thus the amount of overshoot/undershoot can become quite significant. A phase margin of 45° would generally be seen to cause just one or two cycles of ringing, and the overshoot/undershoot would also be minimal.

> **Note:** Under very *large* line or load steps, we will actually no longer be operating in the domain of the "small-signal" analysis which we have been performing so far. In that case, the *initial* overshoot/undershoot at the output is almost completely determined simply by how large a bulk capacitance we have placed at the output. That capacitance is needed to "hold" the output steady, until the control loop can enter the picture and help stabilize the output.

Criteria for Loop Stability

We should remember that phase angle can start changing gradually — starting at a frequency *even 10 times lower* than where the pole or zero may actually reside. We have also seen that a second-order double-pole (−2 slope with *two* reactive components) can cause a very sudden phase shift of about 180° at the resonant frequency if the Q is very high. Therefore, in practice, it is almost impossible to estimate the phase at a certain frequency, with certainty — nor therefore the phase margin — unless a certain strategy is followed!

Therefore, one of the most popular (and simple) approaches to ensuring loop stability is as follows:

- Ensure that the *open-loop gain crosses the 0 dB axis with a −1 slope*.
- We also want to maximize the bandwidth to achieve quick response to extremely sudden load or line transients. By sampling theory, we know that we certainly need to set the crossover frequency to less than half the switching frequency. So, in practice, most designers *set the crossover frequency at about one-sixth the switching frequency* (for voltage mode control).
- Ensure that the crossover frequency is well below any *troublesome* poles or zeros — like the RHP zero in continuous conduction mode (boost and buck-boost — with voltage mode or current mode control), and the "subharmonic instability pole" in continuous conduction mode (buck, boost, and buck-boost — with current mode control). The latter pole is discussed later.

Plotting the Open-loop Gain and Phase with an Integrator

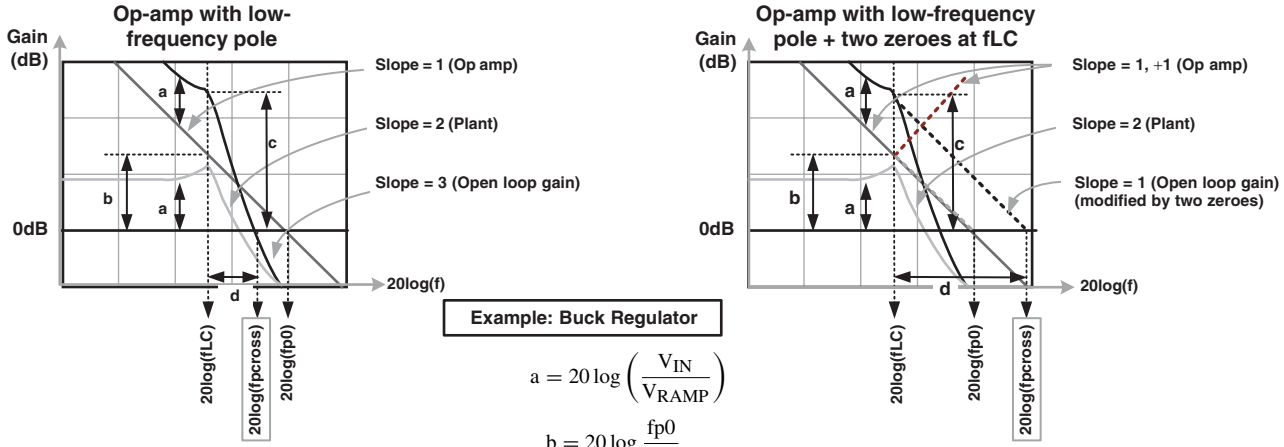
We are interested in plotting the gain and phase of the open-loop transfer function T(s). This is the product of the transfer functions of the G (plant) and H (feedback) blocks in cascade. Review the rules presented earlier in the section "mathematics in the log plane."

Let us start by taking a typical plant and then following it up with the simple integrator (first described in *Figure 7-5*). On the left-hand side of *Figure 7-14* we have plotted the plant gain, the gain of the op-amp integrator, and the overall gain. We can see that the latter is simply the sum of the previous gains (when expressed in dB). We also see that the plant gain falls off after its resonant frequency at a slope of −2. But the integrator slope is falling at −1. Therefore the overall (cascaded) slope is −1 before the double pole, and *−3 thereafter*. That is why we need a "compensation network" around the error amplifier, to meet our simple loop design strategy mentioned in the previous section.

In particular, at lower frequencies we can clearly see that the open-loop gain is *offset vertically* by an amount equal to the plant gain. If the plant is a buck, then we know its

$$fLC = \frac{1}{2\pi\sqrt{LC}} \quad A = 20\log\left(\frac{V_{IN}}{V_{RAMP}}\right) \quad (\textit{Buck})$$

$$fLC = \frac{1-D}{2\pi\sqrt{LC}} \quad A = 20\log\left(\frac{V_{IN}}{V_{RAMP}} \times (1-D)^2\right) \quad (\textit{Boost and Buck-boost})$$



Example: Buck Regulator

$$a = 20\log\left(\frac{V_{IN}}{V_{RAMP}}\right)$$

$$b = 20\log\frac{fp0}{fLC}$$

$$c = a + b$$

$$20\log fcross = 20\log fLC + d$$

$$d = \frac{c}{|slope = -3|} \qquad d = \frac{c}{|slope = -1|}$$

$$fb0 = \frac{V_{IN}}{V_{RAMP}} \times \left(\frac{fcross}{fLC}\right)^3 \times fLC \equiv \frac{1}{2\pi \times RC} \qquad fb0 = \frac{V_{IN}}{V_{RAMP}} \times fcross \equiv \frac{1}{2\pi \times RC}$$

Figure 7-14: Calculating the Open-loop Gain and Stabilizing the Loop

transfer function is

$$\frac{V_{IN}}{V_{RAMP}} \times \frac{{}^{1}/_{LC}}{s^2 + s\left({}^{1}/_{RC}\right) + {}^{1}/_{LC}} \quad \textit{(buck)}$$

So, the vertical offset of this gain plot is

$$20 \times \log\left(\frac{V_{IN}}{V_{RAMP}}\right) \text{ dB} \quad \textit{(buck)}$$

The crossover frequency of the integrator, 'fp0,' is related to the crossover frequency of the open-loop gain 'fcross' by the rule indicated in *Figure 7-6*

$$\text{fp0} = \frac{V_{RAMP}}{V_{IN}} \times \left(\frac{\text{fcross}}{\text{fLC}}\right)^3 \times \text{fLC} \equiv \frac{1}{2\pi \times \text{RC}} \quad \textit{(integrator only)}$$

Therefore, *if we have a certain crossover frequency, 'fcross,' in mind, this tells us the 'RC' that the integrator section must have, to make that happen.*

However, we still have a problem. Though we can clearly set the open-loop crossover frequency almost where we want, *we are still not crossing over with a −1 slope as desired.*

Canceling the Double Pole of the LC Filter

In the right-hand side of *Figure 7-14* we have *hypothetically* (so far) modified the compensation of the previous section to now include two single-order zeros — placed exactly where the plant's double pole lies — *thus effectively canceling the latter out completely*. The op-amp still provides the necessary pole-at-zero from its "integrator section," and if extrapolated, this would cross over at *fp0* with a slope of −1. But now, with the two-zeros present, *the open-loop gain also now falls at a slope of −1* (except for the slight LC peaking along the way). Note that there is a *vertical offset* between these two curves by an amount *equal to the plant gain at low frequencies* (which we already know from above). Therefore we can easily apply the equations given in the lower half of *Figure 7-6*. Simplifying, we will get the following relationship between 'fcross' (the crossover frequency of the open-loop gain) and 'fp0'

$$\text{fp0} = \frac{V_{RAMP}}{V_{IN}} \times \text{fcross} \equiv \frac{1}{2\pi \times \text{RC}} \quad \textit{(integrator + 2 zeros at LC resonant freq.)}$$

This equation therefore connects the desired crossover frequency with the required 'RC' product of the *integrator section of the op-amp.* We can thus adjust the RC to get the required crossover.

Note that for a boost or buck-boost, the only change required in the above analysis is

$$L \Rightarrow {L}/{(1-D)^2} \quad \textit{(boost and buck-boost)}$$

$$\frac{V_{IN}}{V_{RAMP}} \Rightarrow \frac{V_{IN}}{V_{RAMP} \times (1-D)^2} \quad \textit{(boost and buck-boost)}$$

Our compensation analysis seems complete. However, there is one last complication still remaining. We may need *at least one pole* from our compensation network. This is for canceling out the 'ESR zero' coming from the output capacitor. We have been ignoring this particular zero so far, but it is time to take a look at it now.

The ESR Zero

We ignored the ESR of the output capacitor in *Figure 7-14*, and also in the list of transfer functions provided earlier. For example, earlier we had provided the following control-to-output transfer function for a buck

$$\frac{V_{IN}}{V_{RAMP}} \times \frac{1}{\left(\frac{s}{\omega_0}\right)^2 + \frac{1}{Q}\left(\frac{s}{\omega_0}\right) + 1} \quad \textit{(buck: control-to-output transfer function)}$$

where $\omega_0 = 1/\sqrt{(LC)}$. *The ESR zero adds a term to the numerator*. A full analysis shows that the control-to-output transfer function now becomes

$$\frac{V_{IN}}{V_{RAMP}} \times \frac{\left(\frac{s}{\omega esr} + 1\right)}{\left(\frac{s}{\omega_0}\right)^2 + \frac{1}{Q}\left(\frac{s}{\omega_0}\right) + 1} \quad \textit{(buck: complete control-to-output transfer function)}$$

where $\omega esr = 1/((ESR) \times C)$ is the frequency (in radians per second) at which the ESR zero is located. Judging by the sign in front of the *s*-term in the numerator, this is a "well-behaved" (left-half-plane) zero. But it does increase the gain by $+1$ after the ESR zero location. So if the plant gain was falling at a slope of -2 (past its double pole, not cancelled), as soon as it encounters the ESR zero, its slope will change to $-2 + 1 = -1$. If we were using only an op-amp integrator (no LC-pole cancellation), the open-loop gain will then fall with a slope of -2 instead of -3. But that is not enough. However, if we introduce just one zero (besides the pole-at-zero and the ESR zero), we can get the -1 intersection at crossover, as we are seeking.

However we should remember that ESR is a *parasitic* and is hard to control and define well. In fact, vendors of capacitors rarely provide "min/max" limits for this parameter (usually

assuming rather that the MAX of the ESR is the only concern to us). In addition, trace lengths will also contribute to the effective ESR, thereby changing the location of the ESR zero significantly from what we may have been expecting. In addition, the ESR zero is also likely to keep changing with respect to temperature, *and also over time* (especially for aluminum electrolytics). *So the preferred strategy is to estimate rather crudely where the ESR zero is, and try to cancel its effect altogether — by means of a pole provided by the compensation network — at around the same location as the ESR zero*.

In general, for making the control loop less sensitive to high-frequency switching noise, *designers often put another pole roughly at about 10 times the crossover frequency (sometimes at half the switching frequency)*. So now the gain will cross the 0 dB axis with a slope of -1, but at higher frequencies it will drop off more rapidly, with a -2 slope. But why did we pick *10 times* the crossover frequency? Because we know that the phase introduced by this high-frequency pole will actually start making itself felt at one-tenth the frequency of the pole, and we didn't want to adversely impact the phase angle in the vicinity of the crossover frequency (i.e. the phase margin). Later on we will realize that we can actually move this high-frequency pole much closer to the crossover frequency. In fact that is often *desirable*, because it "improves" the phase margin (by *reducing* it and bringing it closer to the ideal value of 45°).

Designing a Type 3 Op-amp Compensation Network

Three types of error amplifier compensation schemes are used most often — called the Type 1, Type 2, and Type 3 in order of increasing complexity and flexibility. The former two are just a subset of the latter, so we prefer to do a Type 3 compensation to demonstrate the full scope (though often, a Type 2 compensation should suffice).

The transfer function of a Type 3 error amplifier as shown in *Figure 7-15* can be worked out easily in the manner we did before. It is given by

$$\frac{\omega p0}{s} \times \frac{\left(\frac{s}{\omega z1}+1\right)\left(\frac{s}{\omega z2}+1\right)}{\left(\frac{s}{\omega p1}+1\right)\left(\frac{s}{\omega p2}+1\right)} \quad \textit{(Type 3 feedback transfer function)}$$

where $\omega p0 = 2\pi(fp0)$, $\omega z1 = 2\pi(fz1)$, and so on. Note that we are ignoring the minus sign in front of this transfer function, as we are separating out the 180° phase shift inherent in negative feedback systems.

There are two poles "p1" and "p2" (besides the pole-at-zero "p0"), and two zeros, "z1" and "z2" provided by this compensation. Note that several of the components involved play a

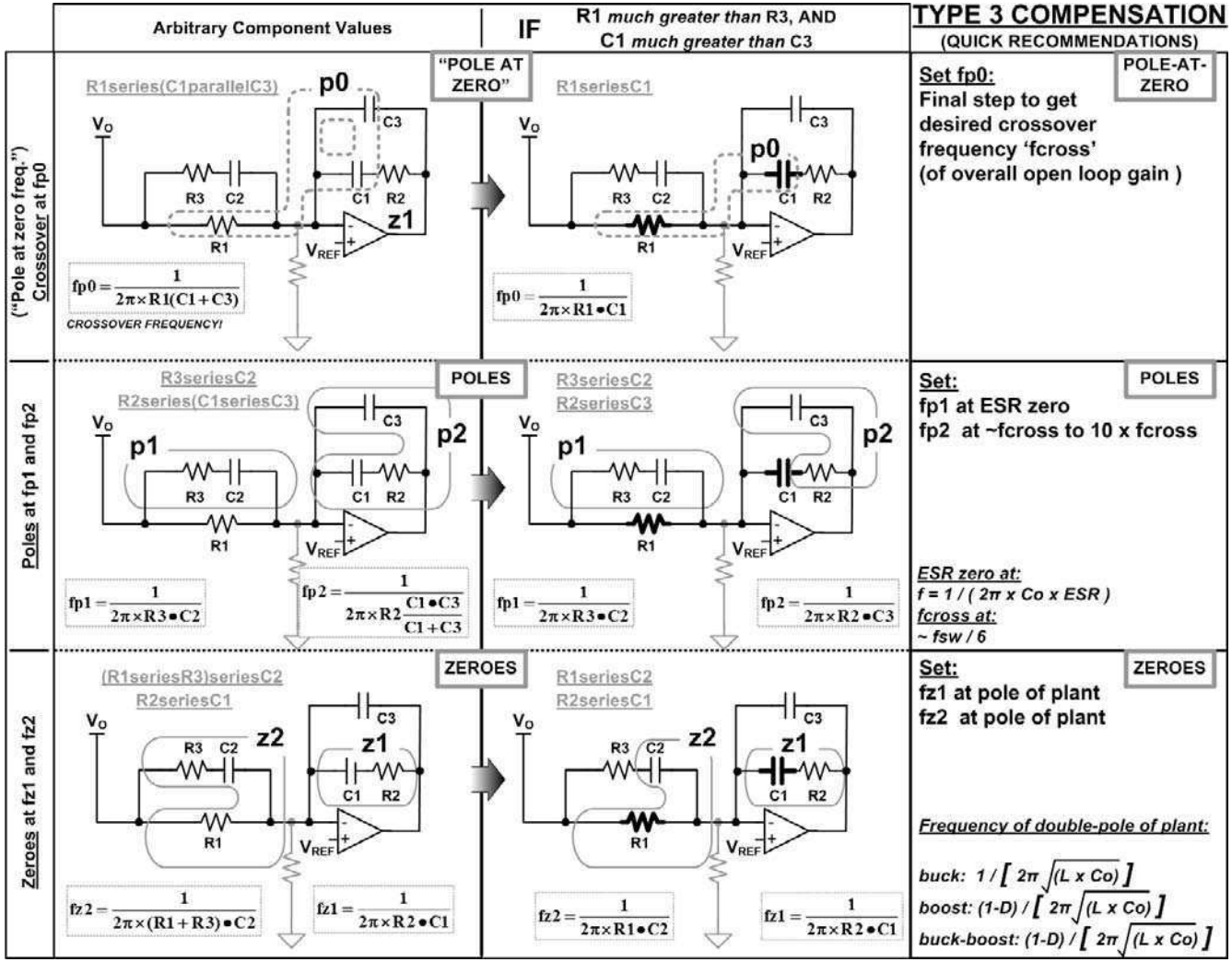


Figure 7-15: Type 3 Conventional Operational Amplifier Compensation

dual role in determining the poles and zeros. So the calculation can become fairly cumbersome and iterative. But a valid simplifying assumption that can be made is that *R1 is much greater than R3*, and *C1 much greater than C3*. So the locations of the poles and zeros are, finally,

$$\text{fp0} = \frac{1}{2\pi \times \text{R}_1(\text{C}_1 + \text{C}_3)} \approx \frac{1}{2\pi \times \text{R}_1\text{C}_1}$$

$$\text{fp1} = \frac{1}{2\pi \times \text{R}_3\text{C}_2}$$

$$\text{fp2} = \frac{1}{2\pi \times \text{R}_2\left(\frac{\text{C}_1\text{C}_3}{\text{C}_1 + \text{C}_3}\right)} \approx \frac{1}{2\pi \times \text{R}_2\text{C}_3}$$

$$\text{fz1} = \frac{1}{2\pi \times \text{R}_2\text{C}_1}$$

$$\text{fz2} = \frac{1}{2\pi \times (\text{R}_1 + \text{R}_3)\text{C}_2} \approx \frac{1}{2\pi \times \text{R}_1\text{C}_2}$$

Note that the reference designators of the components have changed in this section for convenience. What we are now calling "R1" was "Rf2" when we previously discussed the voltage divider. Similarly, the gray unnamed resistor in *Figure 7-15* was previously called "Rf1."

Let us take up a practical example to show how to proceed in designing a feedback loop with this type of compensation.

Example: *Using a 300 kHz synchronous buck controller we wish to step down 15 V to 1 V. The load resistor is 0.2 Ω (5 A). The ramp is 2.14 V from the datasheet of the part. The selected inductor is 5 μH, and the output capacitor is 330 μF, with an ESR of 48 mΩ.*

We know that the plant gain at dc for a buck is $\text{V}_{\text{IN}}/\text{V}_{\text{RAMP}} = 7.009$. Therefore, (20 × log) of this gives us 16.9 dB. The LC double pole for a buck is at

$$\text{fLC} = \frac{1}{2\pi \times \sqrt{\text{LC}}} = \frac{1}{2\pi \times \sqrt{5 \times 10^{-6} \times 330 \times 10^{-6}}} \Rightarrow 3.918 \text{ kHz}$$

We want to set the crossover frequency of the open-loop gain at one-sixth the switching frequency, that is, at 50 kHz. Therefore we can solve for the integrator's "RC" by using our previous (simplified) equation

$$\text{fp0} = \frac{\text{V}_{\text{RAMP}}}{\text{V}_{\text{IN}}} \times \text{fcross} \equiv \frac{1}{2\pi \times \text{RC}}$$

So in our case

$$R_1C_1 = \frac{V_{IN}}{2\pi \times V_{RAMP} \times \text{fcross}} = \frac{15}{2\pi \times 2.14 \times 50 \times 10^3} = 2.231 \times 10^{-5}\ \text{s}^{-1}$$

If we have selected R_1 *as, say, 2* kΩ, C_1 is

$$C_1 = \frac{2.231 \times 10^{-5}}{2 \times 10^3} \Rightarrow 11.16\ \text{nF}$$

The crossover frequency of the integrator section of the op-amp is

$$\text{fp0} = \frac{1}{2\pi \times R_1C_1} = \frac{10^5}{2\pi \times 2.231} \Rightarrow 7.133\ \text{kHz}$$

The ESR zero is at

$$\text{fesr} = \frac{1}{2\pi \times 48 \times 10^{-3} \times 330 \times 10^{-6}} \Rightarrow 10.05\ \text{kHz}$$

The required placement of zeros and poles is

$$\text{fz1} = \text{fz2} = 3.918\ \text{kHz}$$
$$\text{fp1} = \text{fesr} = 10.05\ \text{kHz}$$
$$\text{fp2} = 10 \times \text{fcross} = 500\ \text{kHz} \quad \textit{(set fp2 = fcross for better results — see later)}$$

The remaining components are calculated on the basis of this desired placement of poles and zeros. But now we can use the complete (*non*simplified) versions of the previous equations for fp0, fz1, and so on. So we get the following solutions to these simultaneous equations

$$C_2 = \frac{1}{2\pi \times R_1}\left(\frac{1}{\text{fz2}} - \frac{1}{\text{fp1}}\right) = \frac{1}{2\pi \times 2 \times 10^6}\left(\frac{1}{3.918} - \frac{1}{10.05}\right) \Rightarrow 12.4\ \text{nF}$$

$$R_2 = R_1\frac{\text{fp0}}{\text{fLC}} = 2 \times 10^3 \times \frac{7.133}{3.918} \Rightarrow 3.641\ \text{k}\Omega$$

$$C_3 = \frac{1}{2\pi \times \left(R_2\text{fp2} - R_1\text{fp0}\right)} = \frac{1}{2\pi \times \left(R_2\text{fp2} - R_1\text{fp0}\right)}$$

$$= \frac{10^{-6}}{2\pi \times \left(3.641 \times 500 - 2 \times 7.133\right)} \Rightarrow 88.11\ \text{pF}$$

$$R_3 = \frac{R_1 \times \text{fz2}}{\text{fp1} - \text{fz2}} = \frac{2 \times 10^3 \times 3.918}{10.05 - 3.918} \Rightarrow 1.278\ \text{k}\Omega$$

Note that as usual, for a boost or buck-boost, the only change required in the above analysis is

$$L \Rightarrow {L}/{(1-D)^2} \quad \textit{(boost and buck-boost)}$$

$$\frac{V_{IN}}{V_{RAMP}} \Rightarrow \frac{V_{IN}}{V_{RAMP} \times (1-D)^2} \quad \textit{(boost and buck-boost)}$$

However, in all cases, we must ensure that the selected crossover frequency is at least an order of magnitude below the RHP zero (location provided previously)*!!*

Optimizing the Feedback Loop

In *Figure 7-16* we have plotted the results of the previous example, and we can see that though the crossover frequency is high enough, *the phase margin is rather too generous*.

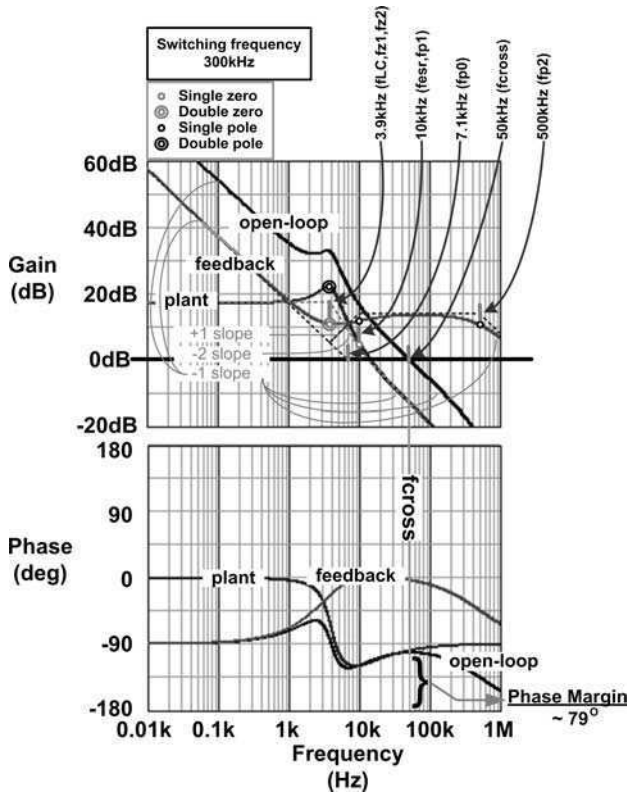


Figure 7-16: Plotting the Results for the Type 3 Compensation Example (nonoptimized)

A very high phase margin may be "very stable," with almost no ringing, but there will be greater undershoot/overshoot.

But what did we expect? Let's analyze this further. For now, let us ignore the high-frequency pole "p2" of the error amplifier. Let us also ignore "p1" since it mutually cancels out with respect to the ESR zero coming from the plant. Therefore the plant just has a double pole at frequency fLC, and this tends to give it the ominous shift of 180° we talked about. But our two single-order zeros from our error amplifier mutually cancel this double pole. So that effectively leaves us with just the pole-at-zero coming from the integrator section of the error amplifier. And like any single pole, this provides a net −90° phase shift to the open-loop characteristics. The phase margin is therefore expected to be 90° short of −180°, that is, 90°. *It is no surprise we got an actual phase margin of 79° as seen in Figure 7-16.* The conclusion is that we are certainly very stable, with no possibility of oscillations, but we are also *over-damped*. We can however accept this as a viable solution too, except perhaps in more critical applications. But if the application is critical, what we would really want to do is to be able to maintain the present crossover frequency of 50 kHz ($1/6^{th}$ of the switching frequency), but to make the phase margin slightly "worse" — for *better* transient response.

We should also have realized by now that intuitively, *poles are generally responsible for making matters "worse"* — since they always introduce a phase lag, leading us closer to the danger level of −180°. On the other hand, *zeros boost the phase angle* (phase lead), and thereby help increase the phase margin away from the danger level.

Therefore, logically, to decrease the existing phase margin of 79° to say 45°, we need *another pole* — that is, to make matters "worse," though in a *calculated manner*. And in fact, that is what *p2* does (we have just not been considering it in our present discussion). Looking back at our worked example and the resultant gain-phase plots, we realize that the reason its effect at the crossover frequency was minimal was that we had *placed it too far out* (10 × fcross). Let us therefore now *change its criterion* and place it *exactly* at the crossover frequency. So we now have

$$\text{fp2} = \text{fcross} = 50 \text{ kHz}$$

We can guess that the phase shift introduced by a single pole at its resonant frequency is 45°, so the new phase margin should be around 79° − 45° = 34°. We plot the gain-phase plots with this new high-frequency pole criterion (and freshly calculated compensation component values) and we get the curve shown in *Figure 7-17*.

In *Figure 7-17*, we see that the phase margin is now almost exactly 45°. The reason it is a little more than our initial estimate of 34° (though we desired 45°), is that the *crossover frequency has decreased slightly* to 40 kHz.

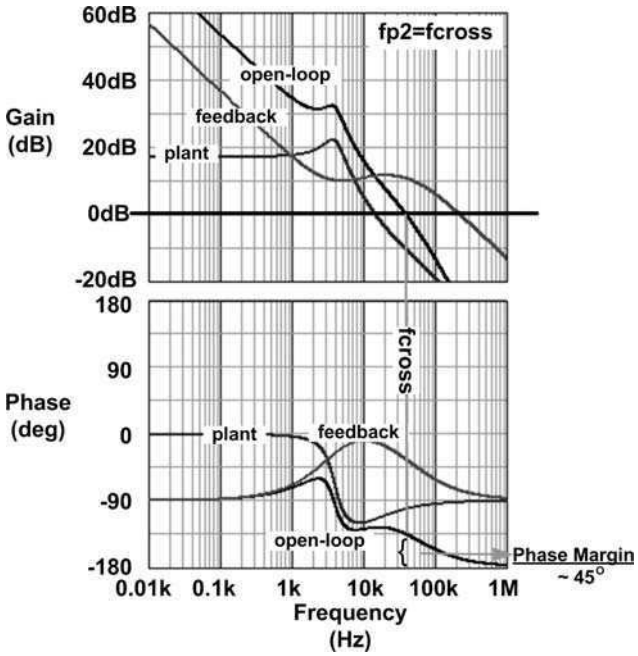


Figure 7-17: Plotting the Results for the Type 3 Compensation Example (optimized)

Note that in terms of the asymptotic approximation, the open-loop gain crosses the 0 dB axis with a slope of −1, but then *immediately thereafter falls off at a slope of −2*. But since the pole is very close to the crossover frequency, the gain in reality falls by 3 dB at this break point (as compared to the asymptotic approximation). So the actual crossover occurs a little earlier. The reason the phase is affected by almost 45° at the crossover frequency is that phase starts changing a decade below where the pole really is.

It can therefore be shown that by trying to place the high-frequency pole exactly at the crossover frequency, the crossover frequency itself shifts downward by almost exactly 20%. So the corollary to that is — if we are starting a compensation network design in which we are going to use the high-frequency pole in this manner, we should initially target a crossover frequency about 20% higher than we actually desire (and get).

Note that engineers use various "tricks" to improve the response further. For example, they may "spread" the two zeros symmetrically around the LC double pole (rather than coinciding with it). One reason to put a zero (or two) slightly before the LC pole location is that the LC pole can produce a very dramatic 180° phase shift, and this can sometimes lead to conditional stability. So the zero absorbs some of that "abruptness" in a sense.

Conditional stability is said to occur if the phase gets rather too close to the −180° danger level at some frequency. Though oscillations do not normally occur at this point, simply because the gain is high (*crossover* is not taking place at this location), under large-signal disturbances, the gain of the converter can suddenly fall momentarily toward 0 dB, thus increasing the chance of instability. For example, if there is a very large change in line and load, the error amplifier output may "rail," that is, reach a value close to its internal supply rails. Its output transistors may then saturate, taking a comparatively long time to recover and respond. So the gain would have effectively decreased suddenly, and it could end up crossing the 0 dB axis at the same location where the phase angle happens to be −180° — and that would meet the criterion for full-blown instability.

Input Ripple Rejection

Why are we interested in maximizing the open-loop gain T = GH *anyway?* Because it can be shown that the effect of line and load variations on the output *is decreased by virtually the same factor* '*T*.'

For example, looking at the earlier equation, we see that the *line-to-output* transfer function for the buck is the same as its control-to-output transfer function, except that the V_{IN}/V_{RAMP} factor is replaced by D. So for example, if $V_{RAMP} = 2.14$ V, and D = 0.067 (as for 1 V output from a 15 V input), then the control-to-output gain at low frequencies is

$$20 \times \log\left(\frac{V_{IN}}{V_{RAMP}}\right) = 20 \times \log\left(\frac{15}{2.14}\right) = 16.9 \text{ dB}$$

and the line-to-output transfer gain at low frequencies is therefore

$$20 \times \log\,(D) = 20 \times \log\,(0.067) = -23.5 \text{ dB}$$

The latter represents *attenuation*, since the response at the output is *less* than the disturbance injected into the input. But both these are *without feedback considered* (or with the error amplifier set to a gain of 1, and with no capacitors used anywhere in the compensation). So when feedback *is* present ("loop closed"), it can be shown by control loop theory that the line-to-output transfer function changes to

$$\text{line-to-output}_{\text{withfeedback}} = \left(\frac{1}{1+T}\right) \times \text{line-to-output}_{\text{withoutfeedback}}$$

where T = GH. Since T (the open-loop transfer function) at low frequencies is very large, so $T + 1 \approx T$. Further, since $20 \times \log\,(1/T) = -20 \times \log\,(T)$, we conclude that *at low frequencies, the additional attenuation provided when the loop is closed is equal to the*

open-loop gain. For example if the open-loop gain at 1 kHz is 20 dB, it attenuates a 1 kHz line disturbance by an *additional* 20 dB — over and above the attenuation already present without feedback considered.

Let's conduct a more detailed analysis. Suppose we are interested in attenuating the 100 Hz (low-frequency) ripple component of the input voltage in an off-line power supply to a very small value. If our crossover frequency is 500 kHz, then by using the simple relationship derived in *Figure 7-6*, we can find the open-loop gain at 100 Hz (here we are assuming we carried out the recommended compensation scheme, which leaves us effectively with only a pole-at-zero response in the open-loop gain). So the gain at 100 Hz is

$$\text{Open-Loop-Gain}_{100\text{Hz}} = \frac{\text{fcross}}{100\,\text{Hz}} = 500$$

Expressed in dB, this is

$$20 \times \log\left(\text{Open-Loop-Gain}_{100\text{Hz}}\right) = 20 \times \log\,(500) = 54\ \text{dB}$$

So the *additional* attenuation is 54 dB here. But we already had $|20 \times \log\,(D)| = 23.5$ dB of attenuation. So by introducing feedback, the total attenuation of the 100 Hz input ripple component has increased to $54 + 23.5 = 77.5$ dB. This is equivalent to a factor of $10^{77.5/20} = 7500$. So if, for example, the low-frequency ripple component at the input terminals was ±15 V, then the output will see only $\pm 15/7500 = \pm 2$ mV of disturbance.

Load Transients

Suppose we suddenly increase the load current of a converter from 4 A to 5 A. This is a "step load" and is essentially a *nonrepetitive* stimulus. But by writing all the transfer functions in terms of *s* rather than just as a function of jω, we have created the framework for analyzing the response to such disturbances too. We will need to map the stimulus into the s-plane with the help of the Laplace transform, multiply it by the appropriate transfer function, and that will give us the response in the s-plane. We then apply the inverse Laplace transform and get the response with respect to time. This was the procedure symbolically indicated in *Figure 7-3*, and that is what we need to follow here too. However, we will not perform the detailed analysis for arbitrary load transients here, but simply provide the key equations required to do so.

The 'output impedance' of a converter is the change in its output voltage due to a (small) change in the load current. With feedback not considered it is simply the parallel

combination of R, L and C. So

$$Z_{out\_withoutfeedback} = R||{}^{1}\!/_{Cs}||\underline{L} = \frac{s\underline{L}}{1 + s\frac{\underline{L}}{R} + s^2LC}$$

where R is the load resistance, and $\underline{L}$ is the actual inductance L for a buck, but is $L/(1 - D)^2$ for a boost and a buck-boost. We have taken the three components in parallel as indicated by the symbol "||."

With feedback considered, the output impedance now decreases as follows:

$$Z_{out\_withfeedback} = \frac{1}{1 + T} \times Z_{out\_withoutfeedback}$$

Even without a detailed analysis (using Laplace transform), this should tell us by how much the output voltage will *eventually* shift (settle down to), if we change the load current.

Type 1 and Type 2 Compensations

In *Figure 7-18*, we have also shown Type 1 and Type 2 compensation schemes (though with no particular strategy in placing the poles and zeros). These are less powerful schemes than Type 3. So whereas Type 3 gives us one pole-at-zero AND 2 poles AND 2 zeros, Type 2 gives us one pole-at-zero AND 1 pole AND 1 zero. Whereas Type 1 gives us ONLY a pole-at-zero (simple integrator).

We know that we always need a pole-at-zero in the compensation, for achieving high dc gain, good dc regulation, and low-frequency line rejection. So the -1 slope coming from the pole-at-zero adds to the -2 slope from the double pole of the LC filter and this gives us a *-3 slope* — that is if we don't put in any more zeros and poles. But still we want to intersect the 0 dB axis with a -1 overall slope. So that means we definitely need two (single-order) zeros to force the slope to become -1.

Type 2 compensation can also be made to work, because though it provides only one zero, *we can use the zero from the ESR of the output capacitor* (despite its relative unpredictability). We remember, in Type 3 we canceled out the ESR zero completely, citing its relative unpredictability. But now we can consider using it to our advantage, if that is indeed possible. Of course, for the Type 2 scheme to work, the ESR zero must be located at a *lower* frequency than the intended crossover frequency.

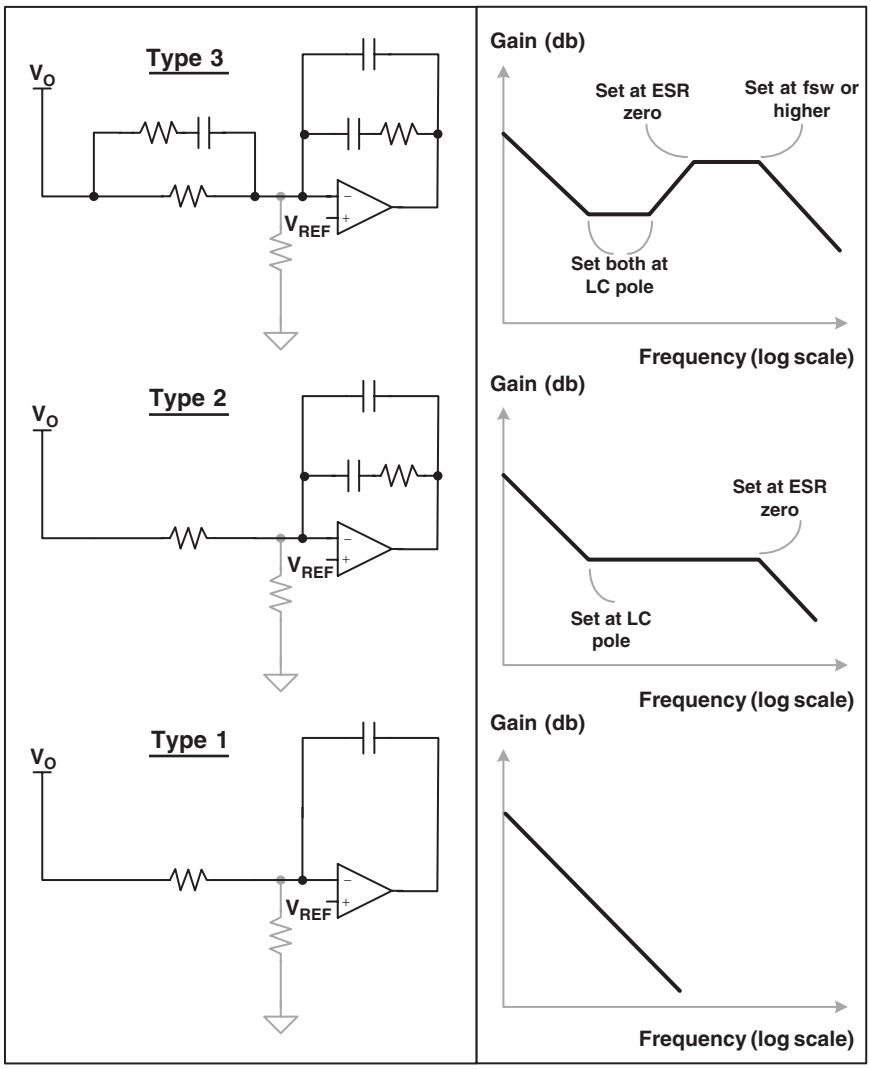


Figure 7-18: Type 1, Type 2, and Type 3 Compensation Schemes (poles and zeros arbitrarily placed and displayed)

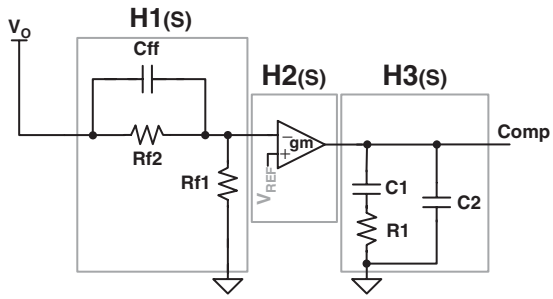
Alternatively, we can easily use Type 2 compensation with current mode control, as explained later.

Type 1 compensation provides only a pole-at-zero, and in fact can *only* work with current mode control (that too with the ESR zero below crossover). Note that it is just a simple integrator.

Transconductance Op-amp Compensation

The final stages of the analysis of voltage mode controlled converters are reserved for the transconductance op-amp. In *Figure 7-13* we had presented its transfer function rather generically. Now let us consider the details.

We can visualize this feedback stage as a product of three transfer functions, H1, H2, and H3 as shown in *Figure 7-19*. When we plot the separate terms out as in *Figure 7-20*, we see that this looks superficially like Type 3 compensation — but it is not! Though it provides two zeros and two poles (besides the inevitable pole-at-zero), we see a big difference — especially in the behavior of H1 (the input side). The problem here is that the dc level of H1 is fixed by the resistive divider (now called Rf2 and Rf1). Its high-frequency value is also fixed at 0 (think of the capacitor Cff conducting fully). So along the way, it *transits* by means of one zero, followed by a pole. But the slope during this transition *has to be* +1 as for any single-order (RC) combination. What all this means is that effectively, *if we fix pole*



$$H(s) = H1(s) \times H2(s) \times H3(s)$$

$$H1(s) = \frac{Rf1}{Rf1 + \dfrac{Rf2 \cdot \frac{1}{Cff \cdot s}}{Rf2 + \frac{1}{Cff \cdot s}}} = \frac{Rf1 \times (Rf2 \cdot Cff \cdot s + 1)}{(Rf1 + Rf2) \times \left(\dfrac{Rf1 \cdot Rf2}{Rf1 + Rf2} \cdot Cff \cdot s + 1\right)}$$

Zero at $1/(2\pi \times Rf2 \times Cff)$

Pole at $1/(2\pi \times Rf2 \parallel Rf1 \times Cff)$

(Rf1 || Rf2 is parallel combination)

Dc level (in dB) = $20 \times \log (Rf1/(Rf1 + Rf2))$

$$H2(s) = gm$$

Dc and high-freq level (in dB) = $20 \times \log (gm)$

$$H3(s) = Zout(s) = C2 \text{ parallel } (C1 \text{ series } R1) = \frac{1 + R1 \cdot C1 \cdot s}{s \cdot (C1)(1 + R1 \cdot C2 \cdot s)} \quad \text{if } C1 >> C2$$

Zero at $1/(2\pi \times R1 \times C1)$

Pole at $1/(2\pi \times R1 \times C2)$

Pole-at-zero crossover at $1/(2\pi \times C1)$

Figure 7-19: "Full-blown" Transconductance Operational Amplifier Compensation

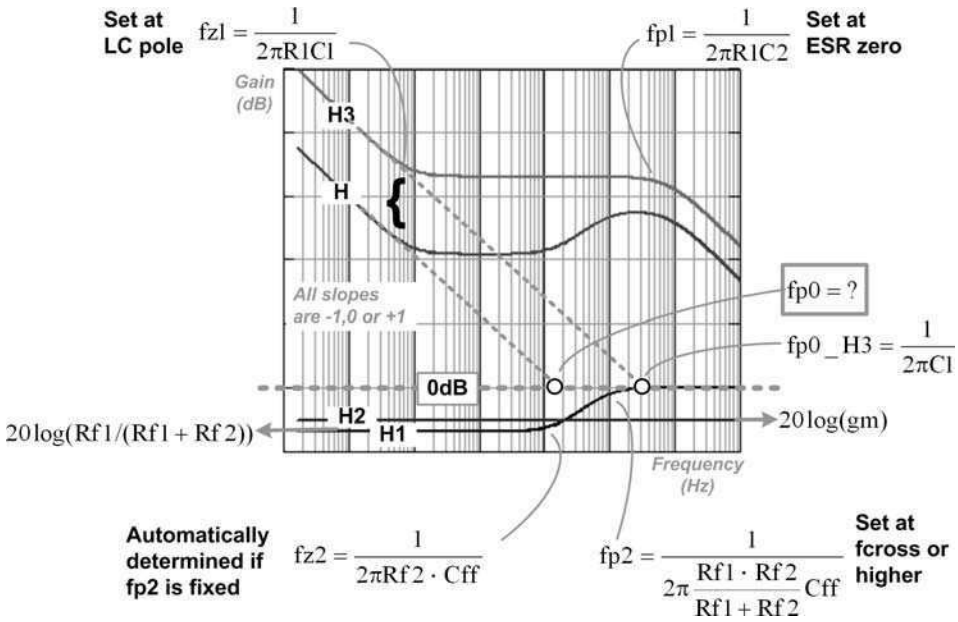


Figure 7-20: The Intermediate Feedback Blocks of the Transconductance Op-amp

fp2 at some frequency, the location of the zero fz2 is automatically defined. There is no flexibility in using this zero and pole pair. For example, if we try to fix *both* zeros of the overall compensation network at the LC double-pole frequency, the pole fp2 will be literally dragged along with fz2, and so the overall open-loop gain would finally fall at −2 slope again, not at −1 as desired. *Therefore, the zero of H1 can only be used if the associated pole fp2 is at or beyond the crossover frequency.* Our final strategy for placing the poles and zeros is indicated in *Figure 7-20.*

It actually requires a great deal of mathematical manipulation to solve the simultaneous equations and to come up with component values for a desired crossover frequency. Therefore, the derivation is not presented here, but the steps are in accordance with the basic math-in-the-log-plane tools presented in *Figure 7-6*. The final equations are presented below, through a numerical example, similar to what we did for Type 3 compensation.

Note: The way we have separated the terms of the transconductance op-amp, the "pole-at-zero" (fp0_H3) seems to be dependent only on C1 (no resistance term). However, we could have also "clubbed" the voltage divider section H1 along with H3 (since these are simply cascaded blocks, in no particular order). Then the "pole-at-zero" would have *appeared* differently (and also included a resistance term). However, whichever way we proceed, the *final* result, that is, H, will remain unchanged. In other words, H1, H2, and H3 are just intermediate mathematical constructs in calculating H (with no obvious physical meaning of their own necessarily). That is why the actual pole-at-zero frequency of the entire feedback block is designated as fp0, not fp0_H3.

Example: *Using a 300 kHz synchronous buck controller we wish to step down 25 V to 5 V. The load resistor is 0.2 Ω (25 A). The ramp is 2.14 V from the datasheet of the part. The selected inductor is 5 μH, and the output capacitor is 330 μF, with an ESR of 48 mΩ. The transconductance of the error amplifier is gm = 0.3 (units for transconductance are "mhos," i.e. ohms spelled backwards). The reference voltage is 1 V.*

The LC double pole occurs at

$$\text{fLC} = \frac{1}{2\pi \times \sqrt{\text{LC}}} = \frac{1}{2\pi \times \sqrt{5 \times 10^{-6} \times 330 \times 10^{-6}}} \Rightarrow 3.918 \text{ kHz}$$

We choose our target crossover frequency "fcross" as 50 kHz. We pick Rf2 = 4 kΩ and Rf1 = 1 kΩ based on the voltage divider equation, the output voltage, and the reference voltage. Then

$$\text{Cff} = \frac{(\text{Rf1} + \text{Rf2})}{2\pi \times (\text{Rf1} \cdot \text{Rf2}) \times \text{fcross}} \Rightarrow 3.98 \text{ nF}$$

The crossover of the overall feedback gain (H) occurs at a frequency "fp0" as indicated in *Figure 7-20*, where

$$\text{fp0} = \frac{V_{\text{RAMP}} \times (\text{Rf1} + \text{Rf2})}{(2\pi)^2 \times \text{fLC} \times \text{Rf2}^2 \times \text{Cff}^2 \times V_{\text{IN}} \times \text{Rf1}} \Rightarrow 10.9 \text{ kHz}$$

So

$$\text{C1} = \frac{1}{2\pi \times \text{fp0}} \times \frac{\text{Rf1}}{\text{Rf1} + \text{Rf2}} \times \text{gm} \Rightarrow 0.87 \ \mu\text{F}$$

$$\text{R1} = \frac{1}{2\pi \times \text{fLC} \times \text{C1}} = 46.5 \ \Omega$$

$$\text{C2} = C_{\text{OUT}} \times \frac{\text{ESR}}{\text{R1}} \Rightarrow 0.34 \ \mu\text{F}$$

We have presented the computed gain-phase plots in *Figure 7-21*. We see that we have an adequate 40° of phase margin, and the crossover frequency is 40 kHz. Note that for visual clarity we used a rather small output capacitance and a large ESR for this particular application and power level. And that is why C1 is not very much larger than C2. But the intention was to shift the ESR zero to less than the crossover frequency, just to demonstrate the principle and be able to plot it out easily. However, the equations and above procedure are valid for any output capacitance and ESR.

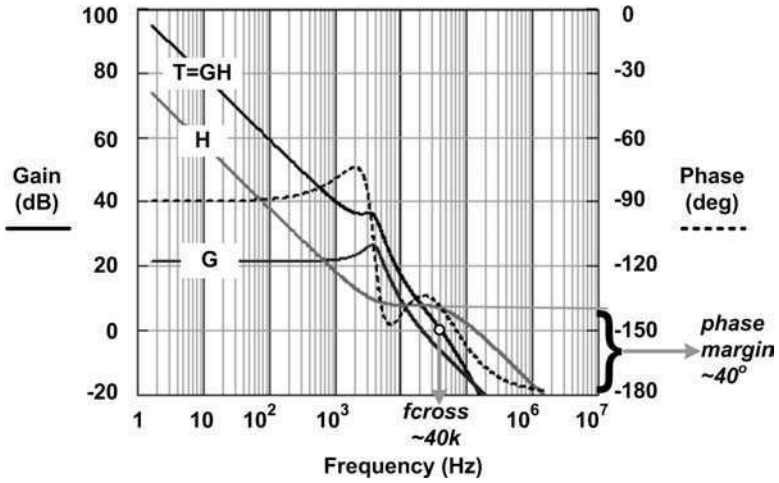


Figure 7-21: Plotting the Results of the "Full-blown" Transconductance Op-amp-based Compensation Example

Simpler Transconductance Op-amp Compensation

There is a practical difficulty involved in using the "full-blown" transconductance op-amp compensation scheme discussed above — because the pole and zero from H1 are *not independent*. They will even tend to coincide if say Rf2 is much smaller than Rf1 (i.e. if the desired output voltage is almost identical to the reference voltage). In that case, the pole and zero coming from H1 will cancel each other out completely. Therefore, we can't proceed anymore, because we were counting on the zero from H1 to change the open-loop gain from −2, to −1, "just in time" before it crossed over.

But there is one other available zero that we can perhaps use — coming from H3. Unfortunately, we have used that up already to cancel part of the LC double pole. (Note that if we hadn't done that, then along with the −1 slope imparted to it by the integrator section (pole-at-zero), the open-loop gain would have had a slope of −3 after its LC double pole location). Therefore to correct the open-loop gain slope from −2 to −1, we now must use the only other available zero — *the ESR zero*, which we had actually canceled out deliberately by a pole in the "full-blown" compensation scheme described previously. *But if we do that, we no longer need fp1, or C2*. So we are left with the simple transconductance stage shown in *Figure 7-22*. The equations for this, based on the *new strategy*, are presented in the following example.

Example: *Using a 300 kHz synchronous buck controller we wish to step down 25 V to 5 V. The load resistor is 0.2 Ω (25 A). The ramp is 2.14 V from the datasheet of the part. The*

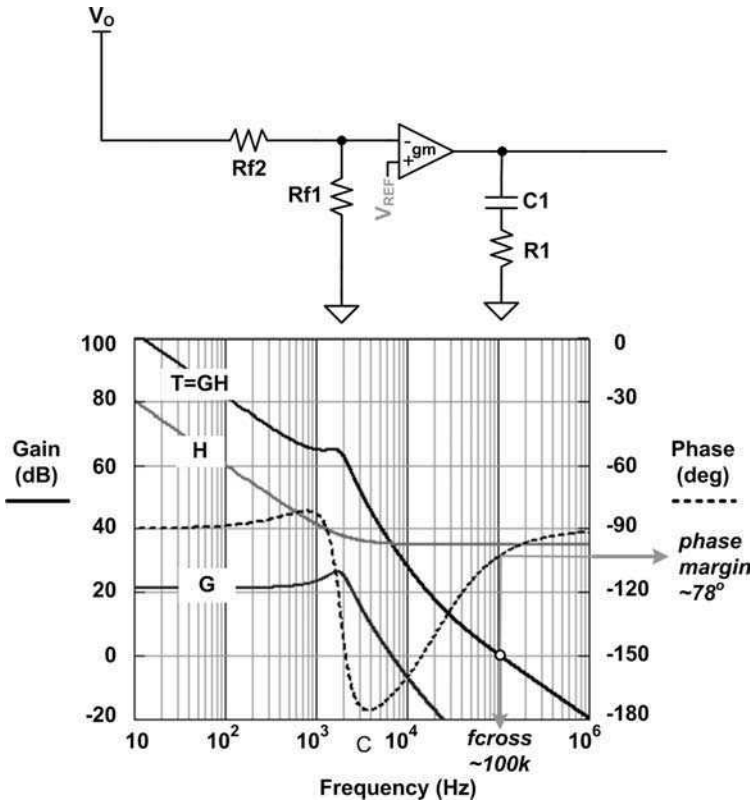


Figure 7-22: Plotting the Results for the Simpler Transconductance Op-amp-based Compensation Example

selected inductor is 5 μH, and the output capacitor is 330 μF, with an ESR of 48 mΩ. The transconductance of the error amplifier is gm = 0.3 (mhos), and the reference voltage is 1 V.

The LC double pole occurs at

$$\text{fLC} = \frac{1}{2\pi \times \sqrt{\text{LC}}} = \frac{1}{2\pi \times \sqrt{50 \times 10^{-6} \times 150 \times 10^{-6}}} \Rightarrow 1.84 \text{ kHz}$$

We choose our target crossover frequency "fcross" as 100 kHz.

The crossover of the feedback gain (H) occurs at a frequency fp0

$$\text{fp0} = \frac{\text{V}_{\text{RAMP}} \times \text{fcross}}{2\pi \times \text{fLC} \times \text{ESR} \times \text{C}_{\text{OUT}} \times \text{V}_{\text{IN}}} \Rightarrow 105 \text{ kHz}$$

So

$$C1 = \frac{1}{2\pi \times fp0} \times \frac{Rf1}{Rf1 + Rf2} \times gm \Rightarrow 2.32\ \mu F$$

Note, if the divider is not present (i.e. $V_O = V_{REF}$), the gain of the stage Rf1/(Rf1 + Rf2) above should be set equal to 1. Further,

$$R1 = \frac{1}{2\pi \times fLC \times C1} = 37.35\ \Omega$$

We have presented the computed gain-phase plots in *Figure 7-22*. We see we have a generous 78° of phase margin and a crossover frequency of 100 kHz. Based on the logic presented for the Type 3 compensation scheme (nonoptimized case, see section "optimizing the feedback loop"), the phase margin in this case is also expected to be around 90°. Here too, one way to reduce the phase lag (if so desired) is to reintroduce the pole fp1 from H3 — *by reintroducing C2* (as in *Figure 7-19*). So we can then set the position of this new pole *at the crossover frequency*, for best effect. The value for C2 required for that is

$$C2 = \frac{1}{2\pi \times R1 \times fcross} \Rightarrow 42.6\ nF$$

Note that by reintroducing C2, the computed crossover again occurs slightly earlier (by about 20%) — at around 80 kHz, instead of 100 kHz. The phase margin is now 36° (closer to the optimal).

Also note that for this simpler compensation scheme to work, *the ESR zero must lie between the LC pole frequency and the selected crossover frequency*.

As before, for a boost or buck-boost, the only change required in the preceding analysis is

$$L \Rightarrow {L}\Big/{(1 - D)^2} \quad \textit{(boost and buck-boost)}$$

$$\frac{V_{IN}}{V_{RAMP}} \Rightarrow \frac{V_{IN}}{V_{RAMP} \times (1 - D)^2} \quad \textit{(boost and buck-boost)}$$

However, we must also always ensure that the selected crossover frequency is at least an order of magnitude below the RHP zero!!

Compensating with Current Mode Control

The plant transfer functions presented earlier were only for voltage mode control. In current mode control, the ramp to the pulse width (duty cycle) modulator is derived from the

inductor current. It can be shown that in the process, the inductor effectively goes "out of the picture" and so there is no double LC pole anymore. So the compensation is *supposedly* simpler, and the loop can be made much faster. But the actual mathematical modeling of current mode control has proven extremely challenging — mainly because there are now *two* feedback loops in action — the normal voltage feedback loop, and a current feedback loop. Various researchers have come up with different approaches, but even as of now, they don't seem to agree with each other completely.

However everyone does seem to agree that *current mode control alters the poles* of the system (as compared to voltage mode control), but the *zeros are unchanged*. So the boost and the buck-boost still have the same RHP zero, as we discussed earlier. *And care is still needed to ensure that the RHP zero is at a much higher frequency than the chosen crossover frequency.*

As mentioned, in current mode control, the ramp to the PWM comparator is derived from the inductor current. Actually, the most common way of producing the ramp is to simply sense the forward drop across the mosfet (or of course by using an external sense resistor in series with it). This small sensed voltage is then amplified by a current sense amplifier to get the voltage ramp, which is then applied to the PWM comparator. At the other pin of the PWM comparator, we have the output of the error amplifier (control voltage). Since the ramp itself gets terminated at the exact moment when it reaches the control voltage level, in effect, we end up *regulating the peak of the inductor current ramp*. That is why it is often said that a converter with current mode control behaves like a current source.

The inductor/switch current ramp is obviously proportional to the voltage ramp at the PWM comparator. So the voltages and currents can be converted into each other through the use of the 'transfer resistance' *V/I*, as defined in *Figure 7-23*. Therefore we realize that we can look at the overall effect *either in terms of currents, or in terms of the voltages* — as shown in *Figure 7-24.* We will see that 'slope compensation' too (discussed below) can be expressed either as a certain *A/s* or as *V/s*. These are *equivalent* ways of talking about the same thing, and they are proportional to each other, through the transfer resistance R_{MAP}.

One of the subtleties of current mode control is that (for all the topologies) we need to add a small ramp to the comparator ramp. This is called *slope compensation*. Its purpose is to prevent an odd artifact of current mode control called subharmonic instability. Subharmonic instability usually shows up as alternate wide and narrow switching pulses. We also discover that the transient response is severely degraded. A bench measurement of a converter that is currently exhibiting this strange pulsing pattern will give us a Bode plot that does not resemble anything we may have been expecting. For one, there is no way to know what the phase margin is.

For subharmonic instability to occur, two conditions have to be met simultaneously — the duty cycle should be close to or exceed 50%, and simultaneously, we should be in CCM

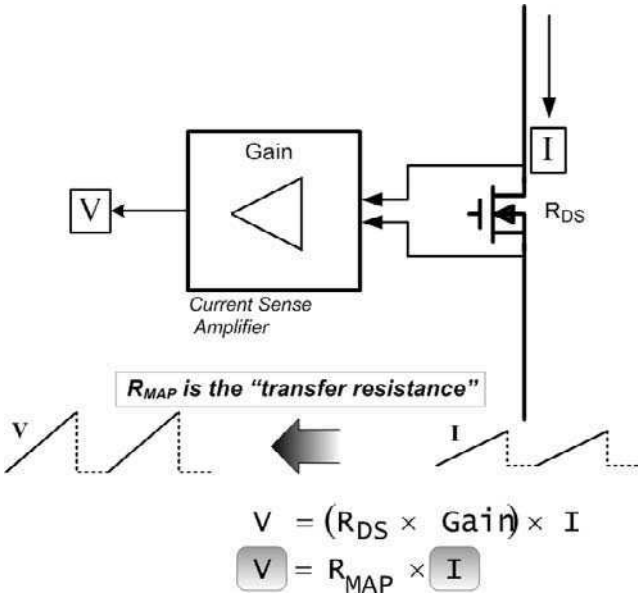


Figure 7-23: How the "Transfer Resistance" Maps the Current in the Switch, into a Voltage Sawtooth at the Comparator Input

(continuous conduction mode). Note that the propensity to enter this state increases as the duty cycle increases (input lowered). So *we should always try to rule out this possibility at* V_{INMIN}. We could certainly avoid this problem altogether, by choosing DCM (discontinuous conduction mode). But otherwise, in CCM, slope compensation is the recognized sure-fix. Though it is interesting to note that by applying slope compensation, we are in effect *blending a little voltage mode control with current mode control*. In fact, it is equivalent to taking the ramp generated from the switch/inductor current, and adding a small fixed voltage ramp to it. Or we can do what is more common — modifying the control voltage as shown in *Figure 7-24*.

If we take the Bode plot of any current mode controlled converter (one that has *not* yet entered this wide-narrow-wide-narrow state), we will discover an unexplained peaking in the gain plot, at exactly *half the switching frequency*. This is the "source" of subharmonic instability. Because, though this point is much past the crossover frequency, it is potentially dangerous because of the fact that if it peaks too much, *it can end up intersecting the 0 dB axis* again — which we know is one of the prescriptions for full instability.

Subharmonic instability is nowadays being modeled as a pole at half the switching frequency. Note that in any case we never consider setting the crossover frequency higher than half the switching frequency. So in effect, this subharmonic pole will always occur at a

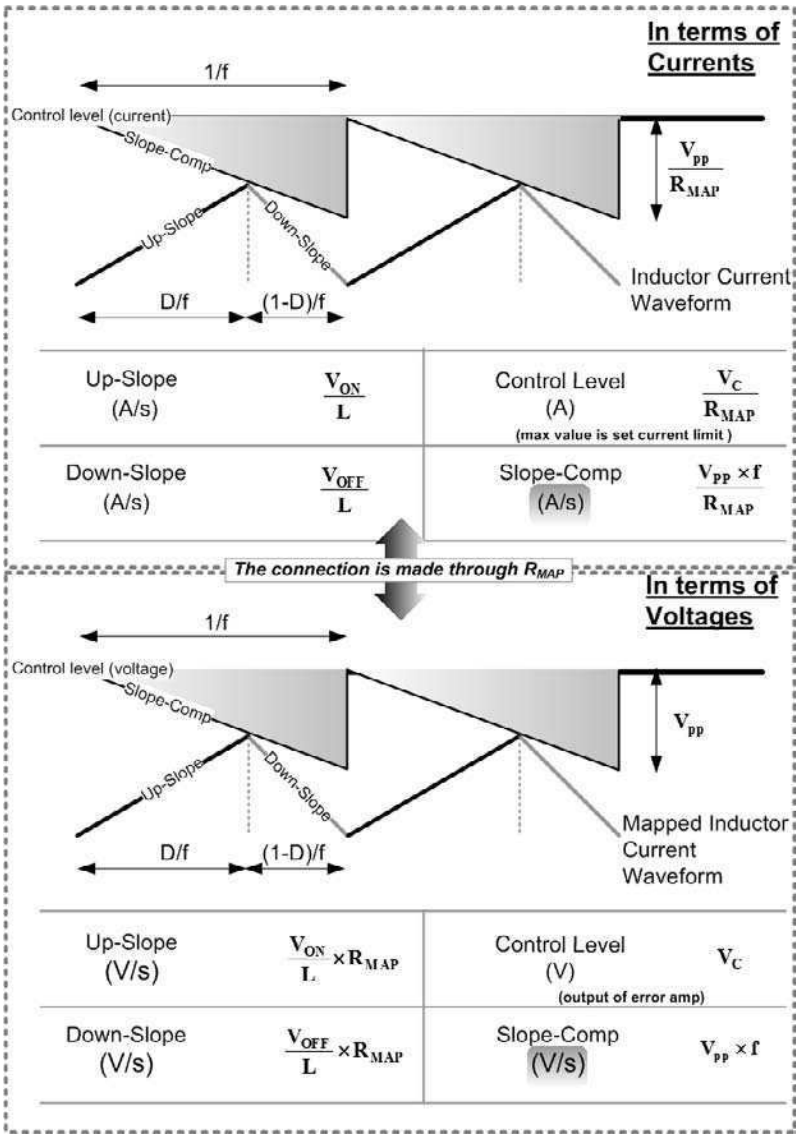


Figure 7-24: Slope Compensation Can Be Expressed Either in Terms of Amperes/Second or as Volts/Second, through the Use of the Transfer Resistance

frequency *greater* than the crossover frequency. However, we also realize that its effect on the phase angle may start at a much lower frequency.

The original models for current mode control did *not* predict this half-switching-frequency peaking (i.e. subharmonic instability). But it has been well known that we need to set a *minimum amount of slope compensation* — the value of which depends on the slopes of the up-ramp and down-ramp of the inductor current. But the criteria used for setting the precise amount of slope compensation, have been slightly differing. Our approach, outlined later, is based on more recent trends.

The subharmonic pole peaking is akin to the double-pole response of an LC filter. So it, too, has a certain "Q", which depends on the slopes of the inductor current, and the applied slope compensation. If the applied slope compensation is too low, the Q will increase, unless we increase the inductance (to decrease the slopes of the inductor current correspondingly). In effect, this means *we need a certain minimum inductance for a given slope compensation*, to remain stable. Alternatively, *we need a minimum slope compensation for a given inductance*, to remain stable. However, slope compensation, the way it is commonly applied, affects the peak current limit too, and that can affect the output power capability of the converter. We have to be careful about that too. In general, we should check that we are being able to meet the required output power, at both ends of the input voltage range.

So how much of slope compensation should we really apply? In the chapter on DC-DC converter design, equations for the minimum inductance were provided. These were actually based on ensuring that *Q never exceeds a value of 2* — that value in turn being based on various bench measurements and data, using typical dc-dc converters. A smaller value of Q will certainly be more "safe", but it will require a larger inductor. A large value of Q will lead to subharmonic instability. Hence $Q = 2$ is normally a good compromise, but it is prudent to confirm the resulting choice of inductor on the bench, on a case-by-case basis.

The design equations presented for the compensation network below, are based on the simpler model from Middlebrook. But they give a good match with far more elaborate models, provided we realize that

- We need to ensure that the RHP zero is much higher than the crossover frequency.
- The half-switching-frequency pole is also higher than the crossover frequency.
- In addition, the half-switching-frequency pole is sufficiently "damped" — by introducing the right amount of minimum inductance vis-à-vis the applied slope compensation (or the appropriate amount of slope compensation, based on the desired or optimum value of inductance — as for example for a current ripple ratio of 0.4).

Note that in our presentation, summarized graphically in *Figure 7-25*, we are even ignoring some other poles from Middlebrook's original model, on the grounds that they usually fall well outside the crossover frequency, and therefore are of little practical interest.

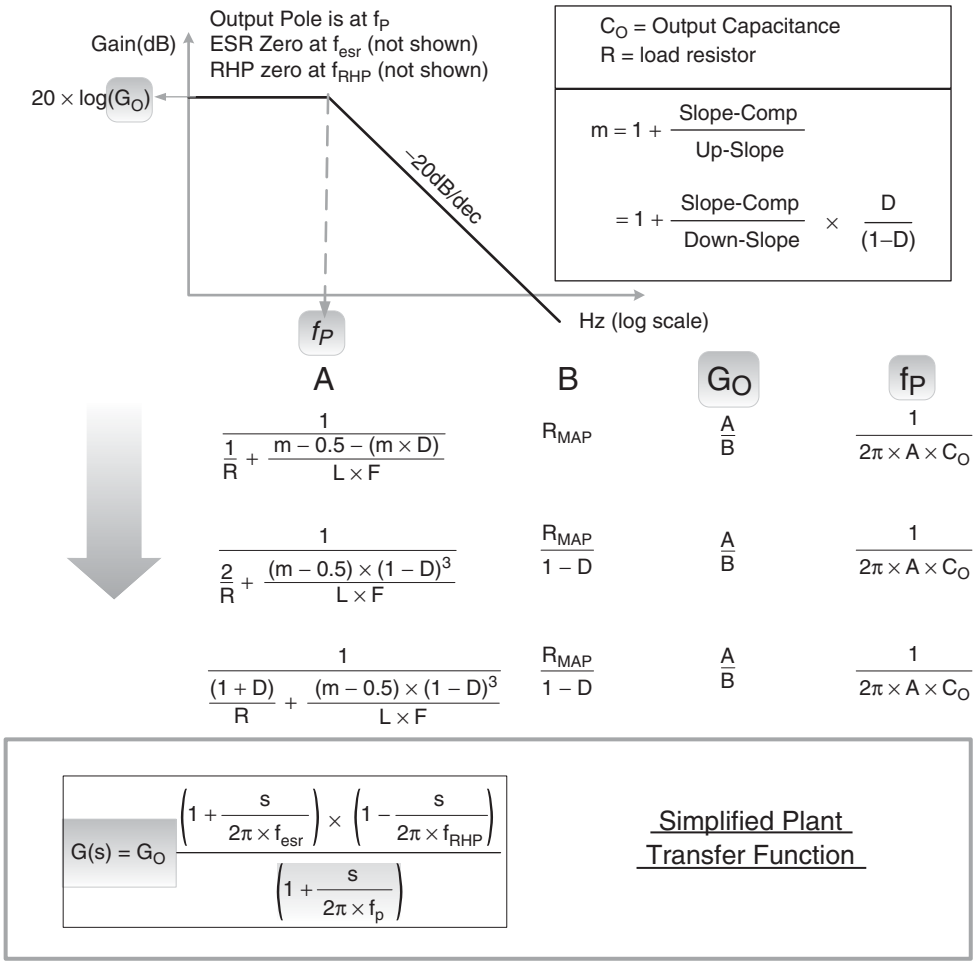


Figure 7-25: Simplified Plant Transfer Function for Current Mode Control

In our simplified model, *we are thus left with only a single pole in the plant transfer function for all the topologies*. This pole comes from the output capacitor and the load resistor (the "output pole"). When we combine it with the inevitable pole-at-zero (from the integrator section of the op-amp), the overall (open-loop) gain will fall with a slope of –2 (after the output pole location). Therefore, we need just one single-zero to cancel part of this slope out, and finally get a −1 slope with which to cross over as desired. Further, this single zero can either be deliberately introduced using Type 2 compensation (in which case we could use its available pole to cancel out the ESR zero) — or we could simply rely on the naturally occurring *ESR zero*. In the latter case, we would need to ensure that the ESR zero is at a frequency *lower* than the crossover frequency. That could indirectly force us to move the crossover frequency out to a higher frequency (but without getting too close to the other

trouble spots mentioned above). And if we can do that, we may be able to use just a Type 1 compensation scheme — because now we don't need even a single pole in our compensation, other than the inevitable pole-at-zero.

The design equations and steps for the *transconductance op-amp* are as follows (see *Figure 7-26*).

1. Choose a crossover frequency 'fcross.' Although we would like to typically target one-third the switching frequency, we must manually confirm that this frequency is significantly below the location of the RHP zero (the equations for the RHP zero were presented earlier, and they still apply here).
2. We realize that, once again, while plotting the open-loop gain, the gain of the integrator will shift vertically by the amount G_O (dc gain of plant). Therefore, using the simple rule in the lower half of *Figure 7-6*, we can find the required *fp0* that will

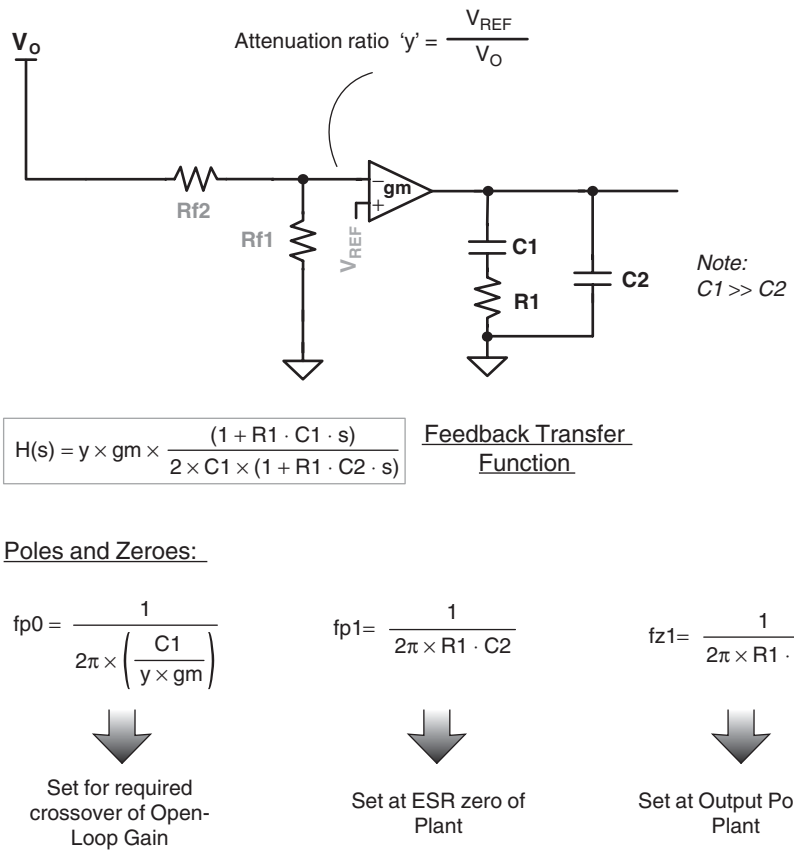


Figure 7-26: Transconductance Op-amp for Current Mode Control

lead to the desired crossover frequency (of the open-loop gain). So

$$\text{fp0} = \frac{\text{fcross}}{A/B}$$

where the values of $G_O = A/B$ are presented in *Figure 7-25*.

3. Calculate C1 using

$$C1 = \frac{y \cdot gm}{2\pi \times \text{fp0}}$$

where y is the 'attenuation ratio' in *Figure 7-26*.

4. Calculate R1 using

$$R1 = \frac{1}{2\pi \times C1 \times f_P}$$

where f_P is the output pole of the plant, as given in *Figure 7-25*.

5. Calculate C2 using

$$C2 = \frac{1}{2\pi \times R1 \times \text{fesr}}$$

where fesr is the location of the ESR zero, that is, $1/(2\pi \times ESR \times C_O)$.

The design equations and steps for the *conventional op-amp* are as follows (see *Figure 7-27*).

1. Choose a crossover frequency 'fcross.' Although we would like to target one-third the switching frequency if possible, we must manually confirm that this frequency is significantly below the location of the RHP zero. The equations for the RHP zero were presented earlier, and they still apply here.

2. We realize that once again, while plotting the open-loop gain, we will find that the gain of the integrator will effectively shift vertically by the amount G_O (dc gain of plant). Therefore, using the simple rule in the lower half of *Figure 7-6*, we can find the required fp0 that will produce the desired crossover frequency (of the open-loop gain). So

$$\text{fp0} = \frac{\text{fcross}}{A/B}$$

where the values of $G_O = A/B$ are presented in *Figure 7-25*.

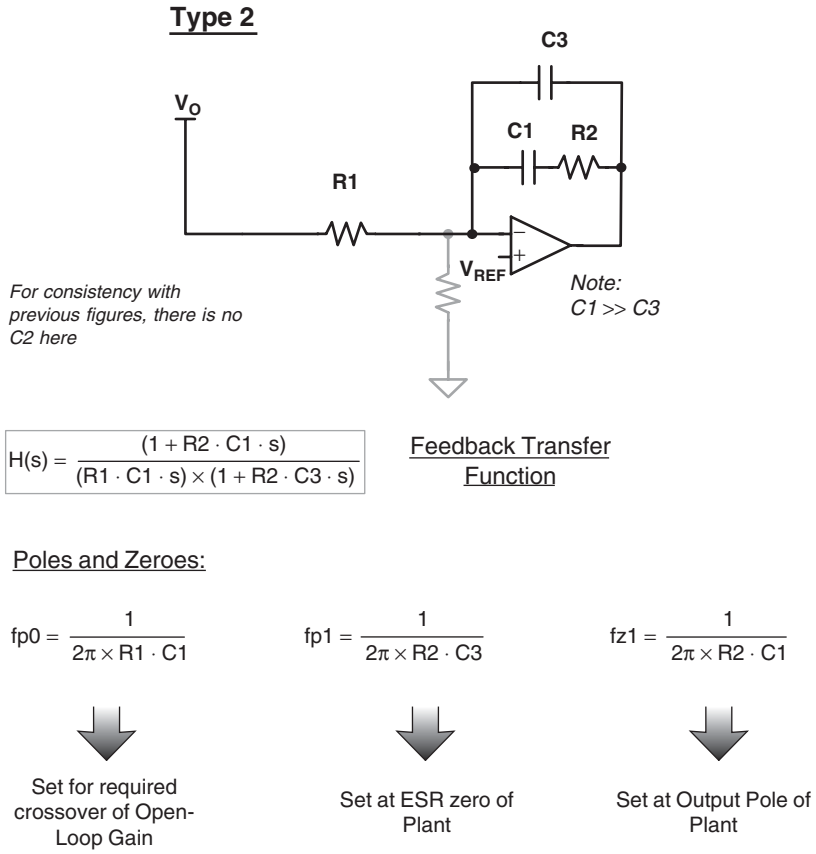


Figure 7-27: Conventional Op-amp for Current Mode Control

3. Calculate C1 using

$$C1 = \frac{1}{2\pi \times R1 \times fp0}$$

where R1 has been chosen while setting the divider.

4. Calculate R2 using

$$R2 = \frac{1}{2\pi \times C1 \times f_P}$$

where f_P is the output pole of the plant, as given in *Figure 7-25*.

5. Calculate C3 using

$$C3 = \frac{1}{2\pi \times R2 \times fesr}$$

where fesr is the location of the ESR zero, that is, $1/(2\pi \times ESR \times C_O)$.

The above design procedure is the same for all the topologies. We just have to use the appropriate row of the table provided inside *Figure 7-25*. Note that for all the topologies, the "L" used is now the actual inductance of the converter (not the "equivalent" inductance of the canonical model).

CHAPTER

8

EMI from the Ground up — Maxwell to CISPR

This Page Intentionally Left Blank

CHAPTER 8

EMI from the Ground up — Maxwell to CISPR

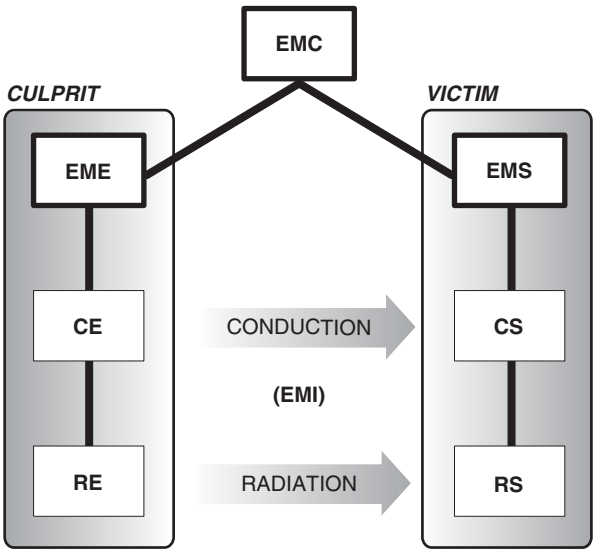
Sooner or later, every power supply designer finds out the hard way that if anything has the potential to cause a return to the drawing board at the very last moment, it is either a *thermal* issue, a *safety* related issue, or a stubborn *EMI* (electromagnetic interference) problem. Of these, the first does get resolved relatively easily — more copper, better heatsinking, and hopefully more air. The safety issues also melt away, with a little prescience during the design phase — and later by some heat shrink tubing, tie-wraps, liberally applied hot-melt glue, RTV (room temperature vulcanizing — i.e. silicone glue), and so on. But we discover that EMI turns out to be a veritable "balloon" — if we try to "push" in the emissions spectrum on one side, it "bulges" out at the other. We manage to achieve compliance with regulatory *conducted* emission limits, only to find it has been at the expense of the *radiated* limits, or vice versa. And sadly, our trusted little bag of tricks (and that dusty Fair-Rite kit of beads) may also sometimes mysteriously let us down. It's then we realize rather acutely — if we can't comply, we can't sell!

The stumbling block to a successful understanding of this very vital but misunderstood area of power conversion is that some of the terminology and descriptions used by *signal integrity engineers* to describe EMI have been bandied about a little too freely in power. Sure there are great similarities, but the devil is in the details. Maybe that's why all the EMI FAQs and 'EMI for Dummies' help-books just didn't seem to help. This general feeling of helplessness has ultimately translated into a popular perception that there is some black magic involved in EMI. Actually, as we will see, all that is really needed is some high school physics, some simple math, and a clear conceptual understanding for all things "power."

That said, EMI is admittedly a challenging area, partly because a lot of *uncharacterized parasitics* enter the stage, each vying for attention. So bench tweaking is not going to be completely avoidable. But with a clear insight into EMI, any *major* redesign should never be required. It is plain however, that for this to happen, the engineer just can't afford to wait until the penultimate moments of the project before he or she takes EMI seriously. It is important to get in early, and get in close — with a clear insight into EMI *as applied to power conversion*. The alternative is the high cost involved in waiting, redesigning, and also days of fruitless and expensive test lab time.

The Standards

The issues of safety and 'electromagnetic compliance' (EMC) are usually clubbed together in most countries. The CE mark (i.e. European Conformity mark) is one such example. Another is the CCC mark (China Compulsory Certification — required by the People's Republic of China, i.e. mainland China). Generally speaking, designated product categories must carry such marks in their respective market regions, and are then assumed to comply with both the applicable safety and EMC standards. In the United States, though, the issues of safety and EMC are taken up separately. The "UL mark" (Underwriters Laboratory Inc.) indicates compliance to product safety standards, whereas "FCC certification" (Federal Communications Commission) reflects compliance with electromagnetic interference (EMI) standards. But EMI is only one note in the gamut of EMC (see *Figure 8-1*). In the United States, unlike Europe, the 'susceptibility' aspect of EMC has been left to the dictate of market forces rather than to the force of law. In Japan, the situation is that though EMC specifications exist, they are stated specifically only for IT (information technology) equipment (or 'ITE'). In addition, the decision to actually affix the relevant mark (called the



EME = Electromagetic Emission
EMS = Electromagnetic Immunity/Susceptibility
CE = Conducted Emission
CS = Susceptibility to Conducted Emissions
RE = Radiated Emission
RS = Susceptibility to Radiated Emissions

Figure 8-1: The EMI/EMC "Tree" (emissions and susceptibility)

VCCI label, for Voluntary Control Council for Interference) is completely voluntary (as the name itself suggests). But we must realize that all such EMI/EMC/Safety marks, whether required by regional/national laws or not, are increasingly perceived by the market as an indication of product quality. Thus, though in theory some of these may be voluntary, in practice, marketing pressures may make them inevitable.

How deep should the power supply engineer really delve into the details of standards and regulations? The road is undoubtedly tricky, because the standards themselves are in a state of continuous evolution. And even though the underlying *intent* is the same, the various international standards can appear quite diverse. Even where supposedly uniform standards have been accepted, there may be national or regional 'deviations.' And even if, hypothetically speaking, the standards were all identical (and there is a definite movement toward that), the certification aspects are never likely to be identical in all countries. The bottom line is that no relevant information that can be provided here, or elsewhere, will turn out to be *durable* enough to withstand the scrutiny of time. So a timely chat with a recognized test lab or consultant should, in any case, always be on a developmental checklist. With that reassurance in mind, the engineer will probably best spend his or her time mastering the *engineering* aspects of EMI/EMC, that being something about which that test house can't really provide detailed guidance.

But the engineer still needs to know *how far* he or she needs to go down the engineering route to achieve necessary compliance — and hopefully, as easily and cost-effectively as possible. The good news here is that, as mentioned, despite the present diversity in regulations, there exists a clear and steady movement toward a set of universally accepted (*harmonized*) standards. The EMI/EMC standards in all the countries and regions mentioned previously, and in fact in most others around the world, are increasingly based on (if not identical to) a well-known set of European standards. Note that, previously accepted norms or standards that were recognizable by familiar prefixes — 'VDE,' 'CISPR,' 'IEC,' and 'ISO,' have all been steadily converging into a single set of pan-European norms *prefixed 'EN.'* A typical power supply engineer is therefore almost always concerned about just achieving compliance with the European EMI standard called "EN 550022" (meant for IT equipment). This standard was originally (and in fact popularly even now) known as CISPR 22. It is the standard we too will be largely focusing on. The corresponding U.S. standard is "FCC Part 15." Despite some differences, the FCC accepts testing to CISPR 22 standards. In fact in 2002, the FCC harmonized Part 15 completely to CISPR 22, laying out a transition period for manufacturers to eventually comply.

As an example of how other countries tend to follow the lead set by Europe and the United States in these matters, take Canada for example. The regulatory Canadian body having jurisdiction over EMI is called 'Industry Canada.' However, its technical requirements are essentially equivalent to those of the FCC. Therefore, if FCC approval

has been obtained (either by meeting Part 15 of the FCC rules or CISPR 22), the equipment need not be retested.

Maxwell to EMI

Light, radio-frequency waves, infrared radiation, microwaves, and so on are all electromagnetic waves (see *Figure 8-2*). For all these, the basic relationship connecting their wavelength λ (in m), their frequency f (in Hz), and the speed of the wave u in the medium of propagation (in m/s), is given by $\lambda = u/f$. For a wave propagating in free space (or air), the speed u is called *'c'* and has the value 3×10^8 m/s. An easy form to remember is

$$\lambda_{\text{meters}} = \frac{300}{f_{\text{MHz}}}$$

The ratio c/u is always greater than 1, and is called the *index of refraction* of the material (through which the wave travels at the speed u). Note that though c is popularly called the velocity of *light*, it is the same for any electromagnetic wave. It can be shown that $c = 1/\sqrt{(\mu_o \varepsilon_o)}$, where μ_o is the *permeability* of free space (vacuum or air) and ε_o is the *permittivity* of free space. μ_o and ε_o are fundamental constants, since they represent the properties of our universe.

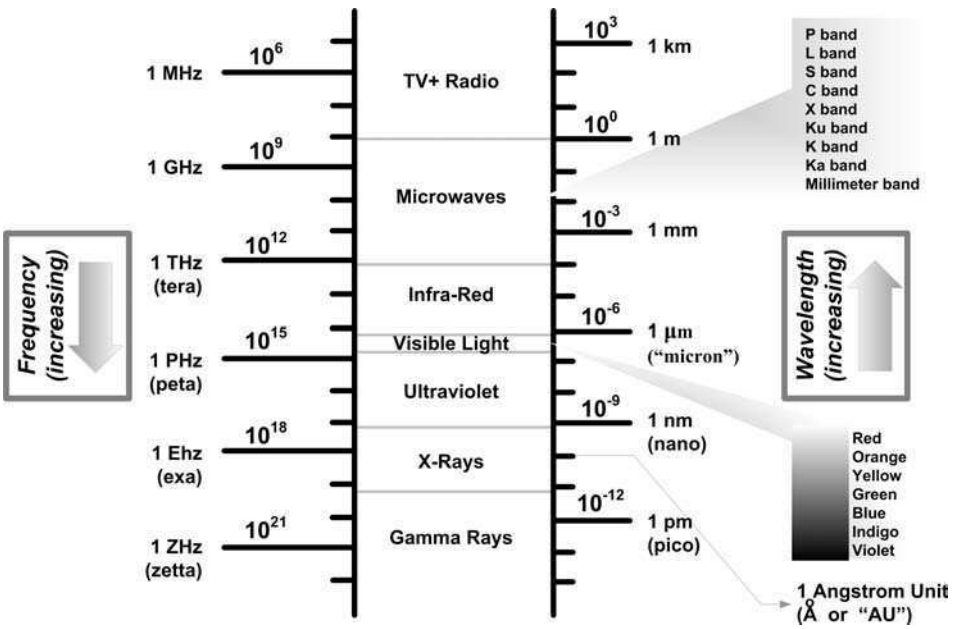


Figure 8-2: The Electromagnetic Environment

It is well-known from our physics class that if a piece of electronic equipment has a dimension close to λ/4, it can end up radiating (or receiving) the corresponding frequency very effectively. This is the principle behind a radio antenna. Note that although a TV antenna is symmetrical around the point where it connects to the cable and has a total physical length of λ/2, it actually has only λ/4 (effective receiving length) on each side. But what if an antenna is much shorter than the 'optimum' of λ/4? Antennas are actually quite effective down to less than λ/10 — which explains why we can pick up almost all the FM stations well enough from a (fixed length) whip antenna on our car. But what if the antenna is much longer than λ/4? In that case, we can consider the antenna as, in effect, being clamped at λ/4 — the remaining length basically superfluous. Therefore, *we should never judge an antenna by its length*!

When we plug a piece of equipment to the ac power lines, its input cable (ac line cord) can combine with the wiring of the building to form an antenna. This can produce strong radiated interference that can affect the operation of other devices in the vicinity. In addition to the radiation process, the emissions can also just conduct through the mains wiring, and thereby directly affect other similarly plugged-in devices. Therefore, there are distinct *radiated emission* limits and *conducted emission* limits specified within all EMI regulatory standards.

We have realized that it would be a mistake to jump to the conclusion that a certain cable length or PCB (printed circuit board) trace is either "too short" or "too long," and therefore *not* contributing to a certain stubborn EMI peak we may be observing. Further, we should keep in mind that any antenna is as good a *receiver*, as it is a *transmitter*. So we could have a situation where radiation is originally generated by the *output* cables, but then picked up by the *input* cables (by radiation), from which point onward it gets *conducted* into the wiring of the building (or/and radiated once again). In fact, we will find that the input and output cables are often responsible for a lot of high-frequency EMI noise, both in the radiated spectrum and the conducted spectrum.

Concerns about cable length take on a whole new meaning when they are coupled with circuits containing modern *high-speed* digital chips. Such chips are themselves powerful EMI emitters, but with the help of inadvertent antennas like the surrounding PCB traces and cables, and also with the inadvertent help of various board, component, and enclosure parasitics, they can put on quite a show! Courtesy Maxwell!

Maxwell showed that whenever an electric field (the 'E-field' — dimensions V/m) *varies with time*, it produces a magnetic field (the 'H-field' — dimensions A/m), *and vice versa.* In fact, the better-known *Faraday's law of induction* (without which no transformer in the world would exist) is actually the first of the set of four Maxwell's unifying equations. So we learn that the E- and H-fields appear simultaneously, the moment the original magnetic or electric source has a time variance. At some distance away, these fields combine to form an electromagnetic wave — that propagates out into space (at the speed of light).

We know that any capacitor that is charged up has an associated E-field residing in the space between the plates, and thus a "resident energy" term of $(1/2) \times CV^2$. We certainly don't expect to have *any* associated H-field, simply because there is *no* current flowing once everything is "steady" (ignoring leakage current). However *during* the capacitor charging process (as the capacitor voltage is rising from zero to its final value), there is a charging current passing through the capacitor. So during that time, both E- and H-fields are generated. Applying the same analogy to a coil passing a steady dc current — it has an associated H-field, with a resident energy of $(1/2)LI^2$ in 'steady state' — but no E-field. Note that in power supplies, there are various subtle connotations to the word 'steady state.' But in the sense used above, a power supply is really not "steady" — except perhaps in its steady *repetitiveness*. In other words, in every switching cycle, the inductor current is actually ramping up and then down, between two extremes. So this *pattern* is certainly steady, but the currents and voltages aren't, since they are constantly changing. So, as the current undulates, there is a constantly changing H-field *and* an accompanying E-field. Together, these two fields constitute an electromagnetic *field* in the vicinity of the coil (forming an electromagnetic *wave* only a certain distance away). That is why, in switching power converters, EMI has become such a major concern. For example, *rod inductors* (often used in post-LC filters at the output) are nicknamed 'EMI cannons' by some engineers. They spew energy all around, and special precautions are needed to prevent this energy from conducting or radiating over to greater distances.

We can ask — what really makes modern digital chips and modern switching converters worse from the standpoint of EMI? That is because of the escalating *frequencies* involved. Smaller and smaller PCB traces and lead lengths can become effective antenna at very high frequencies. So nowadays we are getting painfully aware of the fact that as the frequency increases, so can the intensity of these fields (at a certain distance). Note however, that when talking about *switching power converters*, the "frequency" that we are talking about is not necessarily the basic PWM switching frequency (which is only of the order of say 100–500 kHz — i.e. a time period of a few microseconds). We are referring more to the exceedingly fast *transition* times — which are of the order of 10 to 100 ns only. The Fourier analysis of such a switching waveform will reveal a large amount of very *high-frequency content*, associated with the actual switch transitions. These can cause far more EMI.

Maxwell's equations are usually written out in a way that doesn't fully reveal the following fact very clearly — it turns out that it really does not matter whether we are talking about waveforms of switched voltages (time varying E-fields) or of switched currents (time varying H-fields) — eventually their respective equations are *complementary* and are thus very similar. Further, *these two fields become proportional to each other* at a large distance away, constituting an *electromagnetic wave*, one that can travel great distances on its own. It also follows that if we can 'kill' one component of an electromagnetic wave (either its E-field or H-field), we will manage to kill the entire wave.

If we consider the ratio of the amplitudes of E and H of an electromagnetic wave, we will see that though these fields decrease as the distance from the source increases, their *ratio* remains constant at any point sufficiently far away. The proportionality constant depends on the material of propagation, being equal to $E/H = \sqrt{(\mu/\varepsilon)}$, where μ is the *permeability* of the material (of propagation), and ε its *permittivity*. Note that ε is the electrical analog of the magnetic parameter μ that describes the extent to which a given material allows itself to become magnetized by an external magnetic field. We also note that the units of E are V/m, and H is A/m. Therefore the ratio E/H has the units V/A which is simply a *resistance* (ohms). In air or vacuum (free space), $E/H = \sqrt{(\mu_o/\varepsilon_o)} = 120 \cdot \pi = 377$ ohms. The subscript 'o,' used for the permeability and permittivity above, now refers to free space.

If the fields are very *close* to the source (as compared to the wavelength of their associated EMI wave), they do *not* follow the simple rules described above. See *Figure 8-3* for example, for the possible range of impedances for these so-called 'near-fields.' (Near-fields are defined as those at distances less than $\lambda/6$ from the source.) It also becomes clear why it is often colloquially said that "E-fields have high impedance," whereas "magnetic fields have low impedance." A small circular current loop of trace on a PCB produces magnetic fields, but a strip of copper or metal with a swinging voltage on it (e.g. a heatsink) forms a source of electric fields. Of course, once there is time-variance involved, the H-field leads to an associated E-field, and an E-field produces H-fields. However, only at a great distance

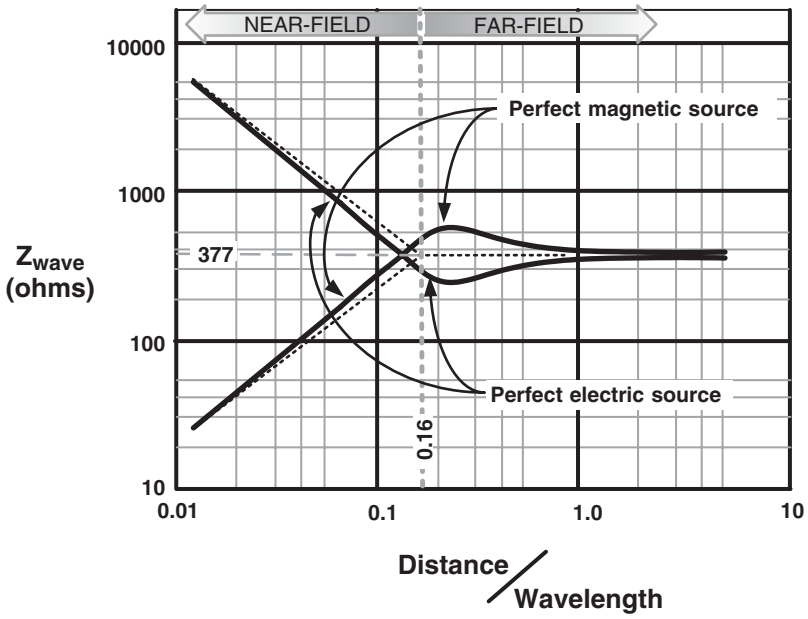


Figure 8-3: Impedances in Free Space

away do the E- and H-fields become proportional to each other and thus form an electromagnetic wave.

As a real-life example of how apparently small-sized circuits can cause major EMI problems, consider a circular current loop (diameter $\ll$ wavelength) enclosing an area "A" (in m^2) carrying an ac current of amplitude "I" (in amperes) and of frequency "f" (in Hz). Its field pattern can be broken up into a *near-field* ($x < \lambda/2\pi$) and a *far-field* ($x > \lambda/2\pi$). The far-field (the electromagnetic wave) at a distance *x* from the center of the coil, when calculated in the plane containing the loop, can be shown to be

$$H = \frac{\pi \cdot I \cdot A}{\lambda^2 \cdot x} \text{ A/m}$$

$$E = 377 \cdot \frac{\pi \cdot I \cdot A}{\lambda^2 \cdot x} \ \mu\text{V/m}$$

where E/H = 377 Ω is called the *wave impedance* in free space, or the *intrinsic impedance* of free space. We can see that as wavelength decreases (frequency increases), the field strength increases. Note that (for far-fields) *both* E and H vary as the *reciprocal of the distance*. But if we get closer and closer to an electric or magnetic source, we could ultimately find field components varying as 1/x, $1/x^2$, and $1/x^3$. But more importantly, we will also see that the dependence on distance of the E-field is no longer the same as that of the H-field. Therefore their *ratio* too now keeps changing as the distance decreases. They are therefore said to constitute an (local) electromagnetic field (not a wave).

In standard radiated EMI tests, the specified measurement range of frequencies is 30 MHz and above (usually up to 1 GHz). An antenna is placed to sense the fields at a specified distance from the device. FCC specifies this distance to be 3 m, whereas CISPR requires 10 m. Let us see why there is actually no conflict here (usually).

The distance from the source that corresponds to the boundary between what is considered a near-field and what is a far-field is given by

$$\frac{\lambda}{2\pi} = \frac{c}{f} \times \frac{1}{2\pi} = \frac{3 \times 10^8}{30 \times 10^6 \times 2\pi} = \frac{10}{2\pi} = 1.6 \text{ m}$$

Note that $\lambda/2\pi = 0.16\lambda$, as was used in *Figure 8-3*. A simpler way to remember the above relationship is as follows — if we call the boundary between the near- and far-field 'X' (expressed in meters), the corresponding frequency (in MHz) is

$$f_{MHz} = \frac{48}{X_{meters}}$$

So a measurement distance of 3 m can be considered a "near-field" distance *only* for frequencies of 48/3 = 16 MHz and *below*. But that is well below the range of frequencies

of interest for a typical radiated EMI test (30 MHz to 1 GHz). Therefore we can conclude that *over the specified test frequency range, we will only see far-fields*. Coming to CISPR, only frequencies lower than 4.8 MHz can present near-fields at the specified distance of 10 m. That is even further below the range of concern in a radiated test. Therefore, knowing that the measured field is *a far-field for either of the two standards*, we can then deduce that the *radiated emissions level must be varying with distance as 1/x*. That makes it finally possible to compare apples to apples. Therefore, the 3 m FCC reading can be easily scaled to a 10 m CISPR reading (and vice versa). The reading at 3 m (FCC) as compared to a reading at 10 m (CISPR) will differ by

$$20 \times \log\left(\frac{E_{3\,m}}{E_{10\,m}}\right) = 20 \times \log\left(\frac{10}{3}\right) = 10.5\ \text{dB}$$

So *if we add 10.5 dB to any radiation spectrum measured as per CISPR, we get the equivalent FCC spectrum*. This variation is referred to in the standards as an "inverse linear distance extrapolation factor of 20 dB per decade." Note that we are talking of a decade of *distance*, not of frequency or amplitude! And 20 dB per decade of distance is simply another way of stating a 1/x dependency ($1/x^2$ would be 40 dB/decade). So for example, a 1/x (far) field at 30 m would be related to the field at 3 m by

$$20 \times \log\left(\frac{E_{3\,m}}{E_{30\,m}}\right) = 20 \times \log\left(\frac{30}{3}\right) = 20\ \text{dB}$$

CISPR 22 also says that if the ambient noise levels are too high at 10 m, we can do the measurement at 3 m, and basically add 10.5 dB uniformly to the 10 m limits.

We also note that radiated tests usually only bother to measure the E-field — since the H-field is then automatically known, being proportional to the E-field. Of course, the near-field and far-field definitions assume a *point* source. Therefore, especially in the case of the 3 m test, some further validation to prove that we really do have only far-fields may become necessary.

Susceptibility/Immunity

Looking at *Figure 8-1* again, we realize that it can also happen that when interference from the *culprit* does affect the operation of a device in the vicinity (the *victim*), it may simply be because the victim device is overly sensitive to EMI. Therefore to assure a prospective buyer that his equipment or device is not going to malfunction whenever someone in the adjacent building just turns on an electric shaver or vacuum cleaner, for example, governments recommend (as in the United States) or mandate (as in the European Union) that a certain

tolerance to incoming EMI be built into the device — or at least be specified clearly for the customer to know beforehand. Now, since all rules must apply equally to all equipment big or small, including the culprit itself in this case, it follows that *any equipment should not emit too much EMI or be too susceptible to it*. These issues are thus two sides of the same coin, and are therefore considered central to the concept of *electromagnetic compatibility* (EMC).

Electromagnetic compatibility (EMC) is therefore defined as the ability of equipment or systems to share the electromagnetic environment (much like we share our freeways). Devices should operate satisfactorily, without either inducing (emitting) or experiencing interference from others. Note that the sensitivity of a device to electromagnetic fields is alternatively described as its *immunity* rather than its *susceptibility*.

Recognizing that the local environment plays a role in defining the amount of received interference and its acceptability, EMI limits are split into two basic application categories:

- *Class A*, corresponding to commercial/industrial equipment/environment
- *Class B*, corresponding to domestic or residential equipment

Clearly, Class B limits will be more restrictive. In fact, these are roughly about 10 dB lower than Class A limits. That is a ratio of about 1:3 in terms of the actual amplitudes of the emission levels.

Like Class A and Class B emission limits, there are also different levels of immunity too. The manufacturer usually specifies the level of tolerance his or her equipment is designed to meet. Based on certain specified test disturbance signals, the generic levels of susceptibility/immunity are:

- Type A: Performance during the test does not fall below a level set by the manufacturer (usually the normal operating level of performance).
- Type B: Though performance degradation is possible during the test, normal performance returns automatically after the test ceases, and with no loss of stored data or any change in the operational characteristics.
- Type C: Restoration of normal performance after the test requires operator intervention (not skilled personnel) and the use of normally accessible controls (like a reset toggle switch on the outside).
- Type D: Nonrecovering failure (possible permanent damage). Includes anything that does not fall within the three types of immunity levels listed above.

Note: There are some situations where the interpretation of the law may not be very clear-cut. We may therefore also get as many opinions as the number of consultants we talk to! For example what if the proposed equipment is meant for an industrial installation, but the same electric utility lines are providing power to a nearby residential area? Should we comply with Class A or Class B? Generally speaking, in such

cases, to avoid last minute delays and a possible redesign, *we should err on the side of caution*. In this case that means we should design to Class B limits at the very outset.

Some Cost-related Rules-of-thumb

A brief look at possible costs:

- The FCC spectrum for digital equipment (currently) begins at 450 kHz, while the equivalent CISPR/EN regulations start at 150 kHz. So FCC compliance can be achieved with a relatively small and inexpensive filter.
- CISPR/EN Class A compliance often requires a filter with at least twice the volume of the FCC-level unit. This filter can therefore be up to 50% more expensive.
- CISPR/EN Class B compliance can require a filter with three to ten times the volume of the FCC unit, and could cost up to 4 times more.

Note: CISPR limits apply to line voltages of 230 VAC, whereas FCC limits are tested at United States line voltage (115 VAC). For a given output power, the input operating current is higher if the input voltage is less. Therefore, in any equipment designed to operate at United States line voltages, thicker copper is required in the filter chokes, and that is somewhat of a cost adder.

EMI for Subassemblies

EMC is generally considered a system-level concern, since from the legal perspective, it applies only to the end-equipment. So a *component power supply* (also called an 'OEM' power supply or a 'subassembly,' e.g. the one inside our desktop computer) does not usually have to meet any EMI/EMC standard per se, unlike a stand-alone power supply. So the ultimate EMC responsibility rests with the system manufacturer. However, take the case of a component off-line power supply (a front-end converter) for example. Here, a major component of the EMI at the input of the system will clearly be coming from the power supply. So it certainly won't help if the power supply itself is producing more EMI than the limits that would apply to the overall system. We should also remember that when the power supply is integrated with the equipment, there are always some hard-to-predict *interactions* between the power supply and the rest of the system — through the connectors, wiring, chassis, grounding, and so on. So *the final EMI is not necessarily just the arithmetic sum (in dB) of the different subassemblies*. Keeping this in mind, the system manufacturer would most likely call out for a front-end converter to maintain its EMI to less than 6 to 10 dB *below* the legal limits. That would usually leave enough headroom for the rest of his circuitry, and also for unexpected interactions. In addition, certification labs themselves may require the submitted prototypes be at least 2 to 3 dB below certification limits — so as to leave a margin for variations in subsequent production. Summing all this up — the practical

resulting situation for front-end OEM converters is simply this — yes, they don't *need* to comply with the legal EMI limits, in fact they need to be much *better*!

What about dc-dc converters that happen to be positioned deep *inside* the equipment? Again, there are no legally applicable EMI/EMC standards for these. Further, since they are likely to be preceded by various circuits and filters, surge suppressors, fuses, capacitors, inrush limiters, and so on (e.g. in the front-end power supply), there is usually an adequate (and fortuitous) EMI barrier already present, that prevents noise from the dc-dc converter from getting on to the ac mains lines via conduction. So, assuming an effective EMI radiation shield is also present (e.g. the grounded metal enclosure), radiated interference may also not be of great concern. Therefore typically, for low-power on-board dc-dc converters, no dedicated input filter stages may be required. However, if such a filter becomes necessary, it can usually just be a simple single-stage LC circuit — possibly even using a small ferrite bead inductor for the L of the filter. And sometimes, just one such filter stage may be good enough to service several paralleled converters.

Nowadays, manufacturers of dc-dc converter modules are often going through the trouble of *profiling* the EMI spectrum present at the (unfiltered) inputs of their products. The purpose is that, whether it is legally required or not, the information will certainly come in handy to the system designer when he or she makes EMI-related decisions. But often, even the *outputs* of modules are being EMI-profiled nowadays. The question arises — why do we even need to test the outputs (for EMI)? In fact, old-school power supply designers are somehow reflexively conditioned into accepting the concept of noise at the input cables, since they visualize the input rails as being rather coarsely regulated, and thus with lots of ripple. They also tend to think that the output voltage rails are relatively 'smooth' since they carry regulated "dc," and can therefore be considered "quiet." But in reality, even the outputs can send forth significant amounts of high-frequency noise, which can often turn out to be even more stubborn than the relatively lower frequency EMI components present at the input. Further, like the input, the output rails from the module or power supply may also be looking into long cable runs (as in telecom applications and distributed power networks). This we know will create a wide-band antenna for EMI. And, as mentioned earlier, the output radiation can be picked up by the input cables, aggravating the problem at the input too.

CISPR 22 for Telecom Ports — Proposed Changes

There is an ongoing debate about extending CISPR 22 to include mandatory full testing on *telecom ports*. Telecommunication ports are defined as those "which are intended to be connected to telecommunications networks (e.g. public-switched telecommunications networks, integrated services digital networks), local-area networks (e.g. Ethernet, Token Ring), and similar networks." There is also a possibility that other product standards will

also start referencing these tests for signal lines in general. The new regulations were due to come into effect in 2001, then again in 2003, but have been successively delayed.
One stumbling block has been that the contemplated test would require a special "ISN" (Impedance Stabilizing Network) — which was apparently in short supply. Note that a more readily available ISN, called the 'LISN' (Line Impedance Stabilizing Network), is commonly used to test the inputs of an off-line power supply. It works by placing an impedance of 50 Ω between each input line and earth ground — so as to closely simulate typical mains wiring impedances. But the proposed telecom ISN would require the impedance to be raised to 150 Ω — that figure being more representative of the actual impedances occurring on typical data networks.

It is increasingly clear that eventually, the law is going to place significant additional EMI requirements on all on-board power converters and any power supply, off-line or otherwise, working inside telecom equipment. Therefore, in the not so unforeseeable future, we could soon all be routinely testing the outputs of every power supply in much the same manner as we test their inputs today. In fact even as of now, many of the big telecom equipment manufacturers, of their own volition, routinely comply with CISPR 22 limits on both the inputs and outputs of their telecom power supplies. Of course, instead of the special proposed ISN, they have just gone ahead and used the more readily available input LISN to test the outputs too.

This Page Intentionally Left Blank

CHAPTER

9

Measurements and Limits of Conducted EMI

This Page Intentionally Left Blank

CHAPTER 9

Measurements and Limits of Conducted EMI

Here we take up the concepts of 'common mode' and 'differential mode' noise. Also, regulatory conducted emission limits and related measuring techniques are discussed.

Differential Mode and Common Mode Noise

Initially, we are going to stick to more conventional descriptions of these parameters. But gradually, we will start discussing certain nuances/differences that can arise in applying the concepts to the area of power conversion.

*Conducted emissi*ons fall into two basic categories:

- Differential mode (DM), also called *symmetric* mode or *normal* mode
- Common mode (CM), also called *asymmetric* mode or *ground leakage* mode.

Looking at *Figure 9-1*, 'L' stands for Live (or "Line" or "Phase"), 'N' for Neutral, and 'E' is the "Safety Ground" or simply, 'earth.' 'EUT' stands for Equipment Under Test. Note that the earth is shown represented by the IEC symbol for *Protective Earth* (ground with a circle around it) and is occasionally labeled 'PE' in literature. The DM noise generator is across the L and N pair. It tries to push/pull a current I_{dm} through these two wires. No current flows through the earth connection on account of this noise source.

Note: There is nothing special about the DM noise current direction indicated in *Figure 9-1*. It can well be the other way around — that is, going *in* through *either* L or N, and coming *out* of the other. In off-line power supplies, we will see that in fact, the direction reverses every ac half-cycle.

Note: The designer may realize that the basic ac input *operating current* of the power supply is also *differential* in that sense — since it flows in through one of the L or N wires and leaves by the other. However, the I_{dm} shown in *Figure 9-1 does not* include this component. That is because the operating current, though differential, is firstly not considered to be "*noise*." Further, its frequency is *low* (twice the line frequency — 100 or 120 Hz, thus being virtually dc). Even its harmonics are well below the range of standard conducted EMI limit curves (150 kHz to 30 MHz). However, *it must not be forgotten that the operating current will dc-bias the noise filter choke*, and can thereby adversely affect the performance of any EMI filter (and also of the current probes being used to gather data). So, though we can ignore the ac (line) component in that sense, it still has an indirect *effect* on the high-frequency input filter.

In *Figure 9-1*, the CM noise source is shown connected at one end to earth. On the other side, it goes through *equal* impedances to each of the L and N lines. It will therefore also

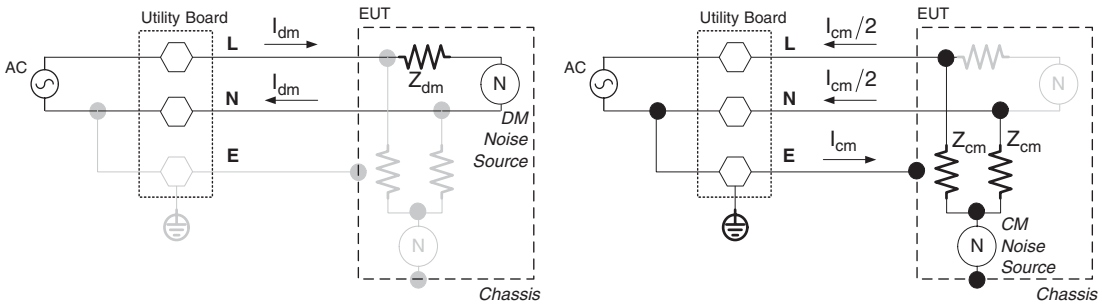


Figure 9-1: Differential and Common Mode Noise

drive *equal* noise currents into the L and N wires — and in the same *direction*. However, note that that *assumes **equal** line impedances too.* We realize that if the impedances are *unbalanced*, we will get an "*asymmetrical* common mode" current distribution (in the L and N wires). And *that is, in fact, a common scenario in actual power supplies*. Also note that this "*asymmetrical* common mode" is equivalent to a mixture of true-CM mixed with some DM (we will demonstrate that a little later).

> **Note:** Since CM noise is itself often called "asymmetric," it is preferable to call this type of operating mode "non-symmetric" rather than "asymmetric."

Engineers often instinctively tend to disregard common mode noise present on the output of their converters. So a typical output noise and ripple measurement is always deliberately *differential* in nature — we spend a long time trying to get the oscilloscope probe positioned correctly on the output terminals (with minimum length of probe ground-wire), simply to avoid picking up common mode noise. But suppose the converter is providing power to an actual subsystem (not a resistive "dummy load"). Looking into the input of this subsystem, we will rarely (if ever) see *equal* (balanced) impedances (i.e. from each of its input terminals to the earth ground). So what happens is that any "common mode" noise existing previously on the output rails of the converter becomes a *differential* input voltage ripple (of high-frequency) for the subsystem. In other words, *common mode noise gets converted into differential mode noise, if the line impedances are unequal*. Therefore, it is no surprise that if an *unbalanced* filter (e.g. a choke present in only one of the two lines) is present at the input of the subsystem, it will only make matters worse. And further, no amount of CMRR (common mode rejection ratio) in the subsystem will help either. True, the subsystem will usually contain a front-end line-to-line input capacitor that will help decouple some of this incoming DM noise. But eventually, the subsystem can start misbehaving. Therefore, reducing common mode noise *at the point of creation* is always the highest priority. Thereafter, *equalizing* the line impedances becomes important. The latter can often be

achieved by placing *balanced* filters at the input of the subsystem (e.g. with two inductors, one on each input line).

Note also that since by the very nature of their creation, common mode currents *usually* have a much higher high-frequency content than differential mode currents, they also have the capacity to cause severe radiation (besides also causing inductive and capacitive coupling to nearby components and circuits). In fact, it is often said that the rule-of-thumb is that *a mere 5 μA of common mode current in a 1 m length of wire can cause FCC Class B radiation limits to be violated.* For FCC Class A limits this number goes up to 15 μA. Note also that *the shortest standard ac power cord is 1 m in length*.

To avoid confusion, we should note that the net common mode current going through the Earth is called "I_{cm}" in our case ($I_{cm}/2$ in each line). However, in related literature, this is often called "$2I_{cm}$" (I_{cm} in each line).

> **Note:** There is nothing special about showing the CM noise current in *Figure 9-1* as coming *out* of the equipment (through both the L and N wires). It could well be in the reverse direction. And like DM noise, it too could be sloshing back and forth, depending on what part of the incoming ac half-cycle we are on, at a given moment.

> **Note:** We will see that in an actual power supply, differential mode noise is initiated by a swinging (pulsating) current — but *the DM noise generator is itself closer to a voltage source*. On the other hand, current mode noise is initiated by a swinging voltage, but the *CM noise generator itself behaves more like a **current** source*. That is actually what makes common mode noise so much more "stubborn" — like any current source, it *demands* a path to flow through. And since its path can include the chassis, *the enclosure can itself become a large high-frequency antenna*.

Let us now do some simple math to split the measured "non-symmetric" currents in the L and N lines into true-CM and DM components. To avoid algebraic errors, we first establish a convention for what is a 'positive direction' for the current flow. Let us assume that in *Figure 9-1*, the direction from right to left is a positive direction, and left to right is negative. We also keep in mind that a current 'I' flowing in one direction on any wire is equivalent to '–I' in the other direction (on the same wire).

Suppose we measure 2 μA going from right to left in one wire (say, the L wire). Then we measure 5 μA going from left to right in the other wire (N). We want to estimate the CM and DM components from these two measurements only.

We have by definition (from *Figure 9-1*)

$$I_L = \frac{I_{cm}}{2} + I_{dm} = 2\ \mu A$$

$$I_N = \frac{I_{cm}}{2} - I_{dm} = -5\ \mu A$$

Solving these simultaneous equations

$$I_{cm} = -3\ \mu A$$
$$I_{dm} = 3.5\ \mu A$$

This means we have a current of 3 μA flowing from right to left in E (common mode component). And we have 3.5 μA (differential mode component) flowing from right to left in L, and left to right in N (differential mode component).

Suppose we measure a current of 2 μA flowing from right to left in the L wire, and ***no*** *current in the N wire. Estimate the CM and DM components from these two measurements.*

We similarly get

$$I_L = \frac{I_{cm}}{2} + I_{dm} = 2\ \mu A$$
$$I_N = \frac{I_{cm}}{2} - I_{dm} = 0\ \mu A$$

Solving,

$$I_{cm} = 2\ \mu A$$
$$I_{dm} = 1\ \mu A$$

So *a non-symmetric mode can be considered part asymmetric (CM) and part symmetric (DM).*

How Conducted EMI Is Measured

For measuring EMI, we need to use an ISN ('Impedance Stabilization Network'). In off-line power supplies, this becomes a LISN (*Line* Impedance Stabilization Network) — also called an AMN (Artificial Mains Network). See *Figure 9-2* for a simplified schematic. Note that the LISN, as recommended for CISPR-22 compliance, is detailed in CISPR 16.

The purpose of the LISN is multifold:

- It is a source of clean ac power to the power supply.
- It provides data to the measurement receiver/spectrum analyzer.
- It provides a stable, *balanced* impedance (as seen by the noise signals emanating from the power supply).
- Most importantly, it makes the measurements repeatable anywhere in the world.

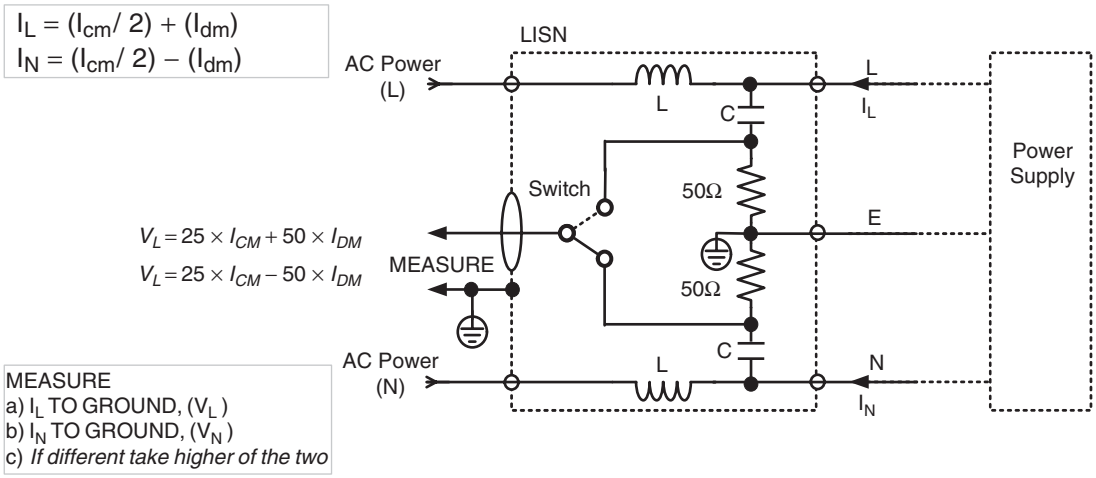


Figure 9-2: Simplified Schematic of LISN

Note that *from the viewpoint of the noise generators in the power supply, it is the LISN that forms their load.*

Let us assume that the values of L and C used in the LISN are chosen so that the following statements hold true, unequivocally:

- The inductance L is low enough not to impede (ac) line current (50/60 Hz) — but high enough to be considered 'open' over the frequency range of interest (150 kHz to 30 MHz).
- The capacitance C is low enough not to pass the ac (line) voltage — but high enough to appear as a 'short' over the frequency range of interest.

Note that *Figure 9-2* doesn't really represent the LISN per se, but is really the *equivalent* schematic required for calculating the level of noise picked up by the receiver. So in fact, the impedance of the cable + receiver has already been factored in — into the values of the components shown (namely the 50 Ω resistors). We know that a typical coaxial cable going to a measuring instrument (analyzer/receiver or oscilloscope etc.) presents a 50 Ω impedance to a high-frequency signal (because of transmission line effects). So when the receiver is measuring the noise, say between L and E, the LISN actually uses a relay/switch to place a real resistor of 50 Ω across the *opposite* pair, that is, between N and E. So in that case, the "50 Ω" shown across L-E is in reality just the impedance of the *cable* going to the receiver. In this way, as we toggle the switch to measure either V_L or V_N, the lines are kept balanced at all times. Note that the choice of 50 Ω also simulates the impedance of typical mains wiring to high-frequency signals. But in any case, the procedure makes the measurement virtually "blind" to the actual impedance of the mains wiring — making it repeatable in any location.

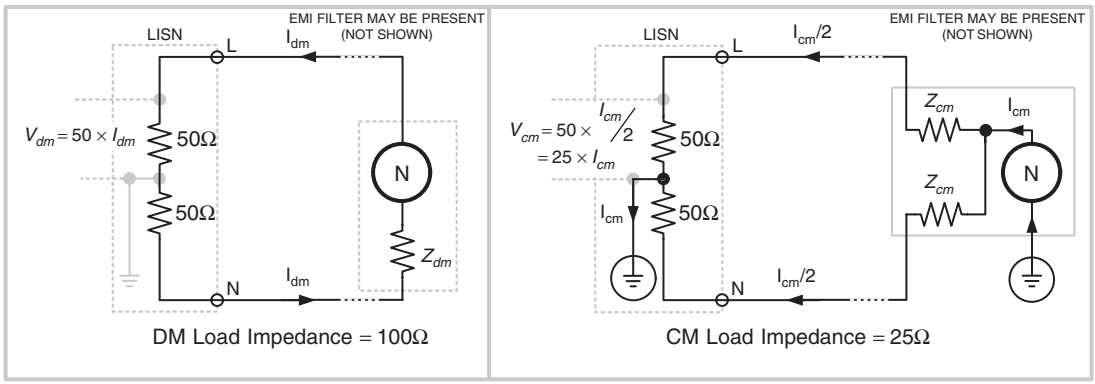


Figure 9-3: Impedance Presented to the CM and DM Noise Generators by the LISN

To know what the measured voltages V_L and V_N are, we now look at *Figure 9-3*. The voltage due to the common mode component is 25 Ω multiplied by the current flowing in the earth connection (i.e. 50 Ω times the current in *each* leg). The voltage due to the differential mode component is 100 Ω times the differential mode current. Therefore the LISN provides the following load impedances to the noise generators (in the absence of any input filter)

- The CM load impedance is 25 Ω.
- The DM load impedance is 100 Ω.

As we flick the switch on the front panel of the LISN, we will measure the following noise voltages:

$$V_L = 25 \times I_{cm} + 50 \times I_{dm}$$

or

$$V_N = 25 \times I_{cm} - 50 \times I_{dm}$$

Both the V_L scan and the V_N scan obviously need to comply individually with the limits.

But how different can the V_L and V_N scans be? In fact, the above two equations have inspired a rather misleading statement often found in related literature — "*if the noise emission is predominantly DM, the V_L and V_N scans will look almost the same. The scans also look identical if the noise is predominantly CM. And if the V_L and V_N scans look very different, that implies that **both** CM and DM emissions are present.*" However, in the case of an off-line power supply, this statement is clearly not true. Because, that would imply that somehow the emissions on the L and N lines are *different*. However, we know that in any typical off-line power supply (with an input bridge rectifier), the L and N lines are

essentially *symmetrical* — both from the viewpoint of the operating current *and* therefore the noise spectrum. So every successive ac half-cycle, the operating current, *and* the noise distribution get transposed from one line to the other. True, at any *given moment*, the noise on L will be quite different from that on N, but when *averaged over several ac cycles* (as any spectrum analyzer would do), equality (symmetry) is restored. Any remnant differences between the V_L and V_N scans can be traced back to some undocumented asymmetries between the two halves of the test circuit, or some severe radiation source impinging asymmetrically on the cables or traces close to the inlet of the power supply.

We observe that the standards *do **not** require us to measure the CM and DM components individually*, but rather a certain sum as described in the preceding equations. However, there are times when engineers do want to see both the CM and DM components separately — for troubleshooting and/or diagnostic purposes. So various people have come up with clever ideas to separate the CM and DM components. Some of these are mentioned below.

- A device called the 'LISN MATE' is rare now. It was invented by an engineer named Nave. It provides about 50 dB attenuation for the DM component, but the CM component comes right through (slightly attenuated — by about 4 dB). The schematic for this is shown in *Figure 9-4*.
- The transformer-based device shown in *Figure 9-5* exploits the fact that common mode voltages cannot cause transformer action — because transformer action requires that a *differential* voltage be applied, so as to produce current in the windings, and thereby causes the flux to swing within the core. Unlike the LISN MATE, in this case both CM and DM noise components are outputted. This device used to be available from AEMC in France, at www.aemc.fr.

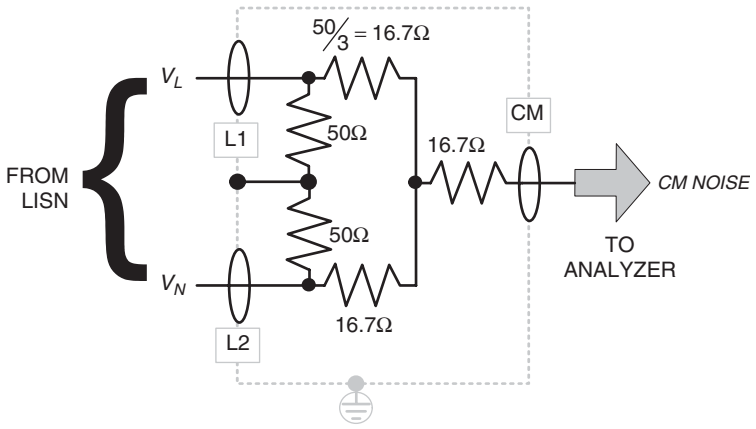


Figure 9-4: The LISN MATE

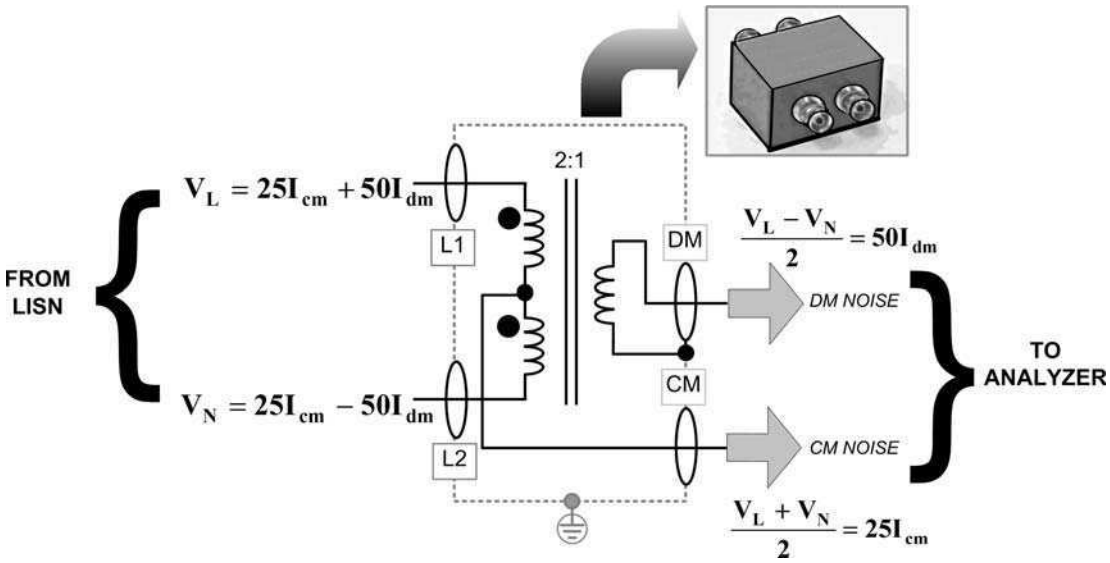


Figure 9-5: A CM and DM Separator

- Both methods above unfortunately require modifications to the standard LISN — because they invoke a certain simultaneous math between the V_L and V_N components. However, a LISN normally provides *either* V_L or V_N at any given moment — not *both* (at the same time, as required here). We can modify the traditional LISN, but that is not only tricky to do, but also hazardous because of the high voltages involved. Therefore, a completely different approach is simply to buy a LISN explicitly designed for the purpose of providing separate CM and DM noise scans (besides providing the necessary "summed up" scan for achieving compliance). An example of such a LISN is the ESA2000 from Laplace Instruments at www.laplaceinstruments.com.

The Conducted Emission Limits

What are we designing our filter to really achieve? In *Table 9-1*, we have provided the CISPR 22 test limits and the FCC limits. We have plotted out the CISPR limits in *Figure 9-6*. Note that in the table, dBμV has also been expressed in terms of mV, in deference to engineers who may have trouble thinking "logarithmically." By definition, a decibel is dB = 20 × log(voltage ratio). So "dBμV" implies a logarithmic comparison to a reference voltage of 1 μV. Note that for *power*, a decibel is 10 × log(ratio).

Note: The FCC does not state average limits, only quasi-peak. However, the FCC does accept certification to CISPR 22. But in that case, "mix-and-match" of the standards is not allowed — it's either one or the other.

Table 9-1: Conducted emission limits

CLASS A (Industrial)								
	FCC Part 15				CISPR 22			
	Quasi-Peak		*Average*		*Quasi-Peak*		*Average*	
Freq (MHz)	dBμV	mV	dBμV	MV	dBμV	mV	dBμV	mV
0.15–0.45	NA	NA	NA	NA	79	9	66	2
0.45–0.5	60	1	NA	NA	79	9	66	2
0.5–1.705	60	1	NA	NA	73	4.5	60	1
1.705–30	69.5	3	NA	NA	73	4.5	60	1
CLASS B (Residential)								
	FCC Part 15				CISPR 22			
	Quasi-Peak		*Average*		*Quasi-Peak*		*Average*	
Freq (MHz)	dBμV	mV	dBμV	mV	dBμV	mV	dBμV	mV
0.15–0.45	NA	NA	NA	NA	66–56.9*	2–0.7*	56–46.9*	0.63–0.22*
0.45–0.5	48	0.25	NA	NA	56.9–56*	0.7–0.63	46.9–46*	0.22–0.2*
0.5–1.705	48	0.25	NA	NA	56	0.63	46	0.2
1.705–30	48	0.25	NA	NA	60	1	50	0.32
* This is a straight line on the dBμV vs. log(f) plot. See worked example								

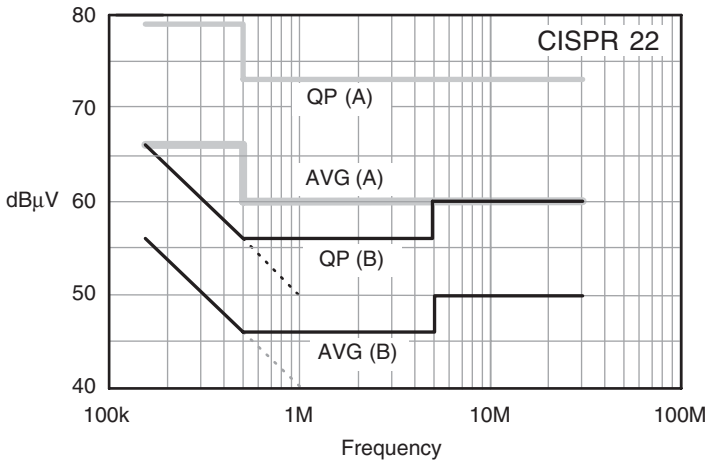


Figure 9-6: CISPR 22 Limits Plotted Out

Example: *What is 1 mV expressed in dBμV?*

$$1\ \text{mV} \rightarrow 20 \times \log \frac{10^{-3}}{10^{-6}}\ \text{dB}\mu\text{V} = 20 \times \log 10^{3}\ \text{dB}\mu\text{V} = 60\ \text{dB}\mu\text{V}$$

In general,

$$(\text{dB}\mu\text{V}) = 20 \times \log\left(\frac{(\text{mV})}{10^{-6}}\right) \quad \textit{(conversion from mV to dB}\mu\textit{V)}$$

Example: *What is 56 dBμV in mV?*

$$56\ \text{dB}\mu\text{V} \rightarrow 10^{56/20} \times 10^{-3}\ \text{mV} = 0.63\ \text{mV}$$

In general,

$$(\text{mV}) = (10^{(\text{dB}\mu\text{V})}) \times 10^{-3} \quad \textit{(conversion from dB}\mu\textit{V to mV)}$$

Returning to *Table 9-1*, we look at the CISPR Class B regions between 150 kHz to 450 kHz and 450 kHz to 500 kHz. Note that this is actually one *continuous* region, with the limit line passing *straight* from 150 kHz through 500 kHz ("straight" on a standard dBμV vs. log(f) plot).

- The equation for a point in the region 150 kHz to 500 kHz (average) for CISPR 22 (conducted) Class B, *average* limits is

  $$(\text{dB}\mu\text{V}_{\text{AVG}}) = -19.07 \times \log(\text{f}_{\text{MHz}}) + 40.28 \quad \textit{(exact)}$$

  This is easier to remember as

  $$(\text{dB}\mu\text{V}_{\text{AVG}}) = -20 \times \log(\text{f}_{\text{MHz}}) + 40 \quad \textit{(almost exact)}$$

- The equation for a point in the region 150 kHz to 500 kHz (average) for CISPR 22 (conducted) Class B, *quasi-peak* limits is

  $$(\text{dB}\mu\text{V}_{\text{QP}}) = -19.07 \times \log(\text{f}_{\text{MHz}}) + 50.28 \quad \textit{(exact)}$$

  This is easier to remember as

  $$(\text{dB}\mu\text{V}_{\text{QP}}) = -20 \times \log(\text{f}_{\text{MHz}}) + 50 \quad \textit{(almost exact)}$$

Note: The published CISPR 22 Class B limits in the range 150 kHz to 500 kHz are themselves actually 'rounded-up' versions of the 'easier to remember' forms provided above. That is why the "exact" equations above look surprisingly stranger than the "almost exact" ones. In fact, what has been done in CISPR 22 is a line was drawn with a slope of −20 dB/dec, passing through the 1 MHz point (extrapolated, see *Figure 9-6*) — at 40 dBμV for the average limits and 50 dBμV for the quasi-peak limits. Thereafter, the calculated y-coordinates at 500 kHz (point of truncation of this portion) and at 150 kHz were rounded off to the nearest decibel.

Example: *What are the CISPR Class B conducted emission limits at a frequency of 300 kHz?*

A frequency of 300 kHz is 0.3 MHz. Therefore, the average limit is

$$-19.07 \times \log(0.3) + 40.28 = 50.25 \text{ dB}\mu\text{V}$$

The quasi-peak limit (which is defined to be 10 dB higher in this region) is automatically 60.25 dBμV.

Looking at the FCC and CISPR 22 quasi-peak limits in *Table 9-1*, we can justifiably ask — do the numbers imply that the FCC standards are actually more stringent than those of CISPR 22 (above 450 kHz)? Not really! The first difference is that FCC measurements are done at the lower U.S. line voltages, whereas the CISPR measurements are done at roughly twice that. So we may be in the position of comparing apples to oranges. Further, *though FCC has no defined average detection limits*, the language *allows for a relaxation of the limits (by 13 dB) if the quasi-peak reading exceeds the average by more than 6 dB*. Therefore, *practically speaking*, equipment compliant to CISPR will always be found compliant to FCC limits.

Quasi-peak, Average, and Peak Measurements

We have not yet explained what the rationale is behind the two types of limits — average and quasi-peak.

Historically, quasi-peak was meant to *simulate human responses* to noise. Humans have a slowly increasing level of aggravation or annoyance to a persistent disturbance. Therefore, to simulate this (subjective) response, there are built-in attack and release rates in quasi-peak detection. The signal level is effectively weighted according to the repetition frequency of the spectral components constituting the signal. So, the result of a quasi-peak measurement will always be dependent on the repetition rate. The higher this frequency, the higher the measured quasi-peak level. Further, because of the finite charge and discharge time constants involved in quasi-peak detection, the spectrum analyzer must sweep considerably slower in this setting. Therefore, *peak detection* can be carried out, and this turns out to be much faster.

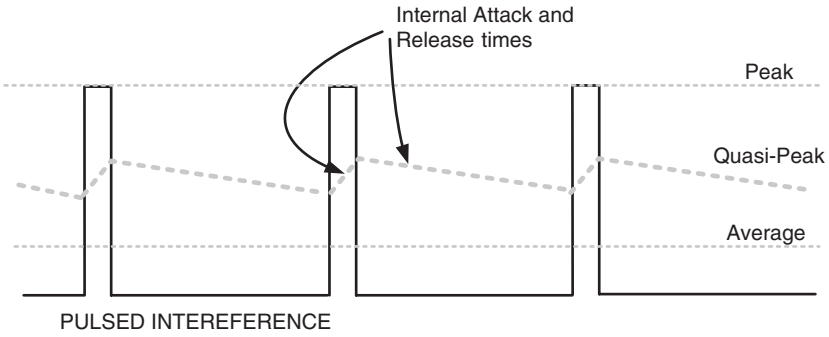


Figure 9-7: Average, Quasi-peak and Peak Readings of a Pulsed Wave

Further, with peak detection, we will always get the *highest reading* (usually followed by quasi-peak and then by the average, see *Figure 9-7*).

So, to perform EMI scans quickly (at least in the initial stages of design), most engineers prefer *peak detection*. The results of this are then *compared against the published quasi-peak limits*, and compliance is sought. If it is achieved, in effect, some additional headroom (safety margin) has been gained, because if a quasi-peak reading had been carried out, it would have certainly given a reading less than the peak reading. The headroom is "nice to have," because it will help account for various *uncontrolled* parasitics that may show up at a later stage. So, quasi-peak detection really needs to be performed only when we are marginally failing the peak detection test (with quasi-peak limits applied).

> **Note:** Usually, if we can meet the quasi-peak limits, we automatically meet the average limits too (using average detection and average limits). But there are stray cases where CISPR 22 quasi-peak limits may be complied with, but the average limit test is failed. That is a bad sign, possibly indicating that a major re-design of the PCB layout and/or transformer is required.

It is often colloquially stated that '*at frequencies below approximately 5 MHz the noise currents tend to be predominantly differential mode, whereas at frequencies above 5 MHz the noise currents tend to be predominantly common mode*.' But this may or may not be true always. Certainly at frequencies above 20 MHz, any conducted noise is most likely attributable to inductive pickup, for example from radiation spilling out from the cables. And we know that that is inherently common mode in nature. However, since radiative pickup is not necessarily the main (or only) source of common mode noise in a switching converter, we should be prepared for surprises. For example, we recall that unequal line impedances can convert a (high-frequency) common mode noise into a (high-frequency) differential mode noise.

Finally, we observe that standard conducted EMI emission limits are typically only up to 30 MHz. We can ask — why weren't the limits set even higher? The reason is that by 30 MHz, any conducted noise is expected to automatically suffer severe attenuation in the mains wiring, and therefore won't really be able to travel far enough to cause interference much down the road. However, since the cables can certainly still radiate (locally), typical EMI *radiation limits* cover the range from 30 MHz to 1 GHz.

This Page Intentionally Left Blank

CHAPTER

10

Practical EMI Line Filters

This Page Intentionally Left Blank

CHAPTER 10

Practical EMI Line Filters

When we start designing EMI filters, we will find that safety issues, thermal issues, and even loop stability concerns are intricately linked to the central issue of EMI. In particular, we must look closely at the aspect of safety, because even though we can attempt to sell equipment that doesn't function satisfactorily, product safety is a legal requirement without which we just cannot sell. So particularly in an off-line application, where the voltages are high enough to cause injury, safety becomes a major concern, even if we are just designing its EMI filter.

In this chapter, we will focus mainly on filters for (single-phase) off-line power supplies. However, tips for dc-dc converters will also be provided along the way.

Safety Issues in EMI Filter Design

The concept of safety and how it impacts the filter section is summed up as follows:

- Any exposed metal (conducting) part (e.g. the chassis or output cables) is capable of causing an electrical shock to the user. To prevent a shock, such parts must be earthed and/or isolated from the high voltage parts of the power supply in some way.
- No *single point failure* anywhere in the equipment should lead the user to be exposed to an electrical shock. There should be *two* levels of protection, so that if one gives way, there is still some protection available.
- Levels of protection that are considered essentially "equivalent" are a) earthing of any exposed metal surface, b) physical separation (typically 4 mm) between any exposed metal and parts of the circuit containing high voltage, and c) a layer of approved insulator between any exposed metal and the high voltage. Note that the insulator must have a minimum dielectric withstand capability of 1500 V ac or 2121 V dc.
- To qualify the preceding slightly — connecting the metal enclosure of the equipment to earth can *sometimes* be considered as an acceptable level of safety protection — there are exceptions to this, as we will soon learn. However, assuming for now that earthing *is* acceptable, we know that to protect the user in case the earth connection fails (maybe due to something as simple as a loose contact), we need to provide one more level of protection. So this could simply be the "4 mm" of separation.

But consider the case of a high-voltage mosfet (switch) mounted on the (earthed) metal enclosure (for better heatsinking). Clearly, we can't provide any level of protection through physical separation. So in this case, we need to place one layer of approved insulator between the mosfet and the enclosure. Note that in this position, the insulator serves as 'basic insulation.'

- What if we have an exposed conductor that is *not* connected to earth (such as for equipment with a two-wire ac cord), or if earthing is itself not an acceptable level of protection for that particular type of equipment as per safety regulations? Then, besides the layer of *basic insulation*, we need another insulating layer (with identical dielectric withstand capability). This is called 'supplementary insulation.' Together these two layers (basic + supplementary) are said to constitute 'double insulation.' We could also use a single layer of insulation, with dielectric withstand properties equivalent to double insulation (i.e. 3000 V ac or 4242 V dc). That would then be called reinforced insulation. So for example, if the equipment is by design, meant only for a two-wire ac cord, we would need two layers of approved insulators (or a single equivalent layer) from primary side to any exposed metal (e.g. output).
- Why do we even bother to connect the enclosure to earth in the first place? In some cases, that is not even considered an acceptable level of protection. And besides, we could achieve two-level protection simply by double (or reinforced) insulation. The main reason for using an earthed metal enclosure is that *we want to prevent radiation from inside the equipment from spilling out*. Without a metal enclosure, whether connected to earth or not, there is very little chance that a typical off-line switching power supply can ever hope to comply with radiated (and possibly conducted) emission limits. That is especially true when switch transition speeds are dropping to a few tens of nanoseconds. Earthing further improves the shielding effect.
- The metal enclosure is rather expectedly eyed by engineers as an excellent and fortuitous heatsink. So in practice, power semiconductors are often going to be mounted on the enclosure (with insulation). However by doing this, we also create leakage paths (resistive/capacitive) from the internal subsystems/circuitry to the metal chassis. And though these leakage currents are small enough not to constitute a safety hazard, they can present a major EMI problem. If these leakage currents are not 'drained out' in some way, the enclosure will charge up to some unpredictable/indeterminate voltage, and will ultimately start radiating (electric fields). That would clearly be contrary to the very purpose of using a metal enclosure. So we then need to connect the enclosure to earth (other than for safety reasons!). We note that even if we didn't have power devices mounted on the enclosure, there could be other leakage paths present to the enclosure. And besides that, an unearthed enclosure would also inductively pick up and reradiate the strong internal electric/magnetic fields.

- Therefore, a) providing a good metal enclosure, and b) properly connecting it to earth is the most effective method of preventing radiated EMI. However, by creating this galvanic connection (to earth), we also now provide a "multi-lane freeway" for the conducted (common mode) noise to flow "merrily" into the wiring of the building. So now, to be able to stay within the applicable *conducted* emission limits, we need to provide a *common mode* filter somewhere.
- Generally speaking, if the equipment is designed *not* to have any earth connection at all (e.g. a two wire ac cord), there will usually be no metal enclosure present either. Ignoring the problem of meeting radiation limits for now, the good news here is that no significant common mode (CM) noise can be created either — simply because CM noise needs an Earth connection by definition. Therefore a CM filter need not be present in this case. However, we must remember that conducted noise limits include not only common mode noise, but differential mode (DM) noise too. So irrespective of the type of enclosure and earthing scheme, *DM filters are always required*.
- One of the simplest ways of suppressing noise is to provide *decoupling* (between the nodes involved). For CM noise this could mean just placing high-frequency *ceramic capacitors* between the L and E wires, and also between the N and E wires, possibly at several points along the PCB. But the problem is that *each of these CM line filter capacitors also unintentionally passes some ac line current into the chassis* (besides the CM noise). The ac component is not considered to be "noise," but it can certainly cause an electrical shock to the user. Therefore, safety regulations restrict the total amount of current that is injected into the earth/enclosure. And this in turn means that in any common mode filter stage *we need to place an upper limit on the net CM filter capacitance*. However, we know that if the "C" of an LC filter is made smaller, then to maintain the desired attenuation level (resonant frequency), we need to correspondingly increase the L. Therefore, it is not surprising that the inductance used for the CM filter stage (in off-line applications) is usually fairly large (several mH).

Practical Line Filters

We now look at a typical power supply line filter, as shown in *Figure 10-1*. Its ultimate purpose is to control *conducted* emissions in general, and therefore it has *two* stages (as highlighted) — one for *differential mode* and one for *common mode*. Let us make some relevant observations.

- Both the CM and DM stages are symmetrical (balanced). From the viewpoint of the noise emerging from the bridge rectifier and flowing toward the LISN, there are in

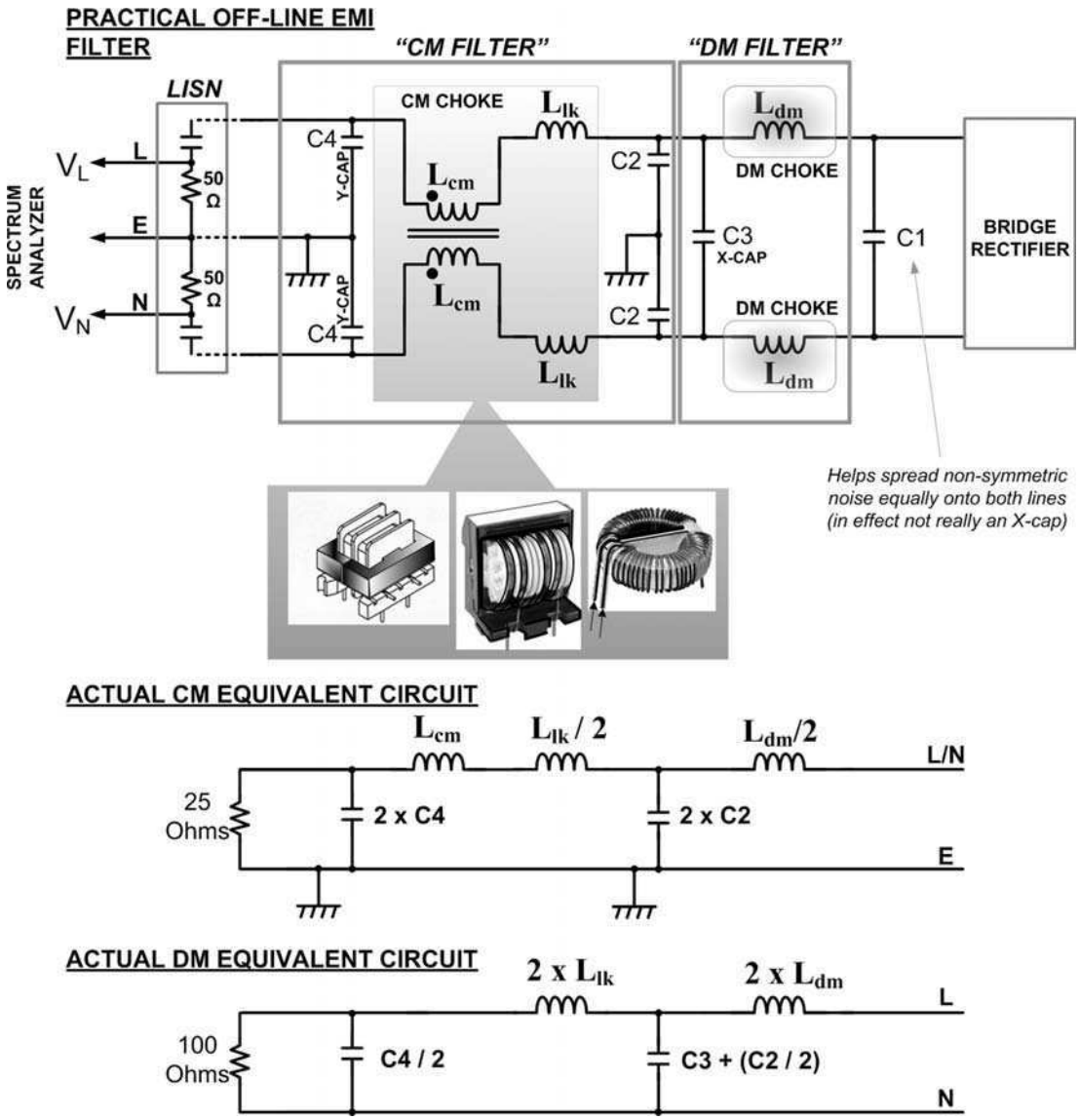


Figure 10-1: Practical Line Filter and the CM and DM Equivalent Circuits

effect two LC filters in cascade (both for DM and CM noise). This filter configuration can provide good high-frequency attenuation (roll-off).

- Occasionally, *unbalanced* filters may be tolerable — for example a single DM choke (i.e. on one line only). Or sometimes, in very low-power applications, just a plain decoupling capacitor (e.g. C1) may suffice. Sometimes tuned filter stages are seen in

commercial off-line power supplies (e.g. from *Weir Lambda*, UK). But there are some anecdotal industry experiences that suggest that under severe line transients or under input surge waveforms, as those typically used for immunity testing, tuned filters can display unexpected oscillations (resonances), ultimately provoking failure of the power supply itself. Therefore, tuned filters are generally avoided in most commercial designs.

- Note that the filter is usually placed *before* the input bridge (i.e. toward the incoming ac line input) — especially because in that position it also suppresses the noise originating from the bridge diodes. Diodes are known to produce a significant amount of medium- to high-frequency noise, especially at the moment they are just turning OFF. Small RC snubbers (or sometimes just a "C") are therefore often placed across each diode of the input bridge. Though sometimes, we can get away simply by choosing diodes with *softer recovery characteristics*.
- Note that input bridge packs using ultrafast diodes are often peddled as offering a significant reduction in EMI. In practice they don't really make much difference — at least not enough to justify their steep cost. In fact typically, the faster a diode, the *greater* are the reverse current and forward voltage spikes that it produces at turn-off and turn-on. So very fast bridges may in fact produce worse EMI scans.
- Typical practical values for the inductance of a CM choke in medium-power converters range from 10 to 50 mH (per leg). The DM choke is always much smaller (in inductance, but not in size as we will see). Typical values for the DM choke are 500 μH to 1 mH.
- In *Figure 10-1*, we have shown both the CM and DM filter stages as being symmetrical (balanced). So for example, we have placed *identical* DM chokes on *each* of the L and N lines. In *Fig. 10-1* we see that in fact *the DM choke is also a part of the CM equivalent circuit* (and vice versa). And since line impedance imbalance can cause CM noise to get converted into DM noise, it is always advisable to keep *both* the CM and DM stages symmetrical (balanced).
- One obvious way to maintain equal CM inductances in both lines is to *wind them on the same core* (e.g. a toroid). That automatically assures a good inductance match (assuming of course that there are an equal number of windings per leg). Note that if we are winding the CM choke ourselves (as during prototyping), we must note the relative direction of the windings, as indicated in *Figure 10-1* (see third sample choke picture). With such a winding arrangement, the *magnetic field inside the core will cancel out completely (in principle) for DM noise. Similarly, the flux due to the operating ac line current will also cancel out* (that too being differential in nature). Therefore the choke will present an impedance only to the CM noise component.

Note: The reader is cautioned that there are several widely used but confusing symbols for the CM choke found in schematics in related literature. But whatever the symbol, as long as it is meant to serve as a common mode choke, the direction of the windings must be as shown for the toroid in *Figure 10-1*.

- If we *reverse* the current direction in one of the windings of a CM choke, then it becomes a DM choke (for both lines). However, now it is also subject to the flux produced by the ac line input current (no cancellation occurs). DM chokes, in general, should always be put through a "saturation check" — because of the impedance they present to the line current.
- We see that DM chokes may need to be quite large, just to avoid core saturation — despite the fact that their inductance is usually much less than that of CM chokes. But in fact, a CM choke can also be very large. That, however, is primarily necessitated *not* by the typically higher inductance required, but more so by the desire to provide the required inductance with the *minimum amount of copper losses*. So, a core *with a high A_L value* is sought, and that usually spells "bigger core." We should also keep in mind that we don't want the core to "topple over" and saturate, on account of small imbalances in symmetry of the windings. So we may ultimately need to oversize the CM choke for various such reasons.
- Theoretically, there is no need for any air gap in a common mode choke, because the flux due to the line current is expected to cancel out completely. In practice, it doesn't fully, mainly due to slight differences in the individual winding arrangement (despite the equal number of turns). At a minimum, this causes the core to get dc-biased in one direction, and thereby cause an imbalance in the inductance it presents to the two lines. This would expectedly degrade the EMI performance, but in extreme cases, the core may even saturate. Note that core saturation in the filter is clearly not a catastrophic event (like the saturation of the main inductor/transformer of the converter can be), but since it is accompanied by severely worsening EMI-suppression efficacy, we need to prevent that too. Therefore, as in a forward converter transformer, a small air gap is usually present, even in a CM choke. This may be an actual air gap (between split core halves), or it may be a *distributed* gap, as in powdered iron cores. Though this lowers the inductance index (A_L) somewhat, the resulting solution is much more immune to production variations, and is also more stable over time. In general, whenever we introduce an air gap, the core starts *partially acquiring the properties of the interposing air* — and since air never saturates, the air-gapped core too has a much softer saturation characteristic.
- We can consider spending some more money and avail ourselves of magnetic materials like "amorphous" cores or 'Kool Mu®' if we want to achieve higher inductance (with higher saturation flux densities), in a smaller size.

- Toroidal CM chokes in particular, when used in off-line applications (i.e. with both windings on the same core) must meet safety requirements relating to the *separation distances* between the windings ('clearance' and 'creepage,' as discussed later). So for example, we cannot simply wind the two windings carelessly overlapping each other — we need to maintain a specified physical separation. Nor can we just use a *bare* toroid core to wind them on — we need an approved coating and/or a suitable bobbin.
- A bare ferrite can be a rather good electrical conductor, especially if it is the more commonly used manganese-zinc ferrite (as opposed to nickel-zinc formulations). This can be confirmed by simply pressing the tips of an ohmmeter at two points on the surface of any bare ferrite lying around in the lab. Further, if we are trying to rely on the enamel coating of a typical copper magnet wire to protect from shorts, we should know that the coating is considered to be just operational/functional insulation, and is not considered to be even basic insulation.
- Note that L_{cm} in *Figure 10-1* is the inductance of *each* leg of the CM choke. Therefore, it is the inductance measured across either winding, with the other winding open. Now, if we repeat this measurement, but instead of keeping the other winding open, we short its ends together, what we measure is the leakage inductance L_k. By definition, the two leakage inductances in each leg are uncoupled, and therefore they cannot be sharing any magnetic path. Therefore the leakage inductance of a CM choke behaves differently from the rest of the choke — differential currents no longer cancel out for this inductance. In effect, L_k presents an inductive impedance to DM noise. This "hidden" inductance of a CM choke has been successfully exploited by filter designers, to serve as an "unintentional" DM choke. Therefore, in low-power converters, we usually won't see any separate DM chokes — just CM chokes. The good news here is that the leakage inductance is effectively an *air-cored* inductor, so it never saturates — even if for some reason, its "parent" CM choke saturates completely. Thus the efficacy of a *leakage-based* DM choke is maintained at any supply current level.

  Note: In any transformer, if we measure its leakage inductance (by shorting the secondary winding), the reading remains virtually unchanged even if we remove the core completely from the bobbin. That is because leakage, by definition, is uncoupled and does not pass through the magnetic core — if it did, it would be "coupled."

- The *inter-winding capacitance* of a choke affects its characteristics significantly at high frequencies. This can be intuitively visualized as providing an easy detour for noise to simply flow past the windings. To minimize the end-to-end capacitance of a toroidal winding, it is recommended that the winding be *single layer*. Also, in *Figure 10-1*, the sample CM choke picture in the middle is better than the one to its left, in terms of minimizing the end-to-end capacitance. That is because of the *split*

introduced in each winding section by the special bobbin used. The split also helps increase the leakage inductance (which helps reduce DM noise). Bobbins with several such splits are also available at a price.

- Line to line capacitors are called 'X-capacitors' ("X-caps"). X-caps when used in off-line applications *before the input bridge* must be safety approved. But *after* the bridge (on the rectified side), it's basically a 'don't care' situation from the safety point of view. Note that since it is essentially a front-end component, approved X-caps are typically impulse-tested up to 2.5 kV peak.
- Line-to-earth capacitors are called 'Y-capacitors.' Since Y-caps are critical in terms of having the potential to cause electrocution if they fail, *approved* Y-caps are typically impulse-tested up to 5 kV peak. Note that Y-caps used *anywhere* on the primary side (in off-line applications) *must always be safety approved*. Depending on the location in the power supply, we may even need two Y-caps in series (basically corresponding to *double* insulation). However, sometimes we can also find Y-caps placed between the *secondary* ground and earth/enclosure (for EMI suppression purposes). In this position, it is usually acceptable to use any ordinary 500 V ac rated capacitor (unapproved).
- Traditionally, off-line X-caps were of special metallized film + paper construction whereas Y-caps were a specially constructed disc ceramic type. However we can also find X-caps that are ceramic, as we can find Y-caps that are film type. It's a choice dictated by cost, performance, and stability concerns. Film capacitors are known to always provide much better stability over temperature, voltage, time, and so on — than most ceramics. In addition, if they are of 'metallized' construction, they also possess *self-healing* properties. Note that ceramic capacitors do not have any inherent self-healing property. However, ceramic Y-caps are specifically designed never to fail *shorted* under any condition, as this would pose a serious safety hazard.
- If for any reason (e.g. filter bandwidth or cost) ceramic is preferred for the Y-cap positions, then we need to carefully account for its basic tolerance, its variation with respect to temperature and applied voltage, and all other long-term variations and drifts. That is because we need a certain filtering efficacy, but at the same time we can't increase the leakage current into the chassis. In this regard, we should keep in mind that the capacitance stated in the datasheet is not just a nominal (or typical) value, but in fact it happens to be a fairly misleading value. For example, the fine print may reveal that the test voltage at which the capacitance is stated is close to, or equal to zero Volts! So the actual capacitance it presents in a working circuit may be very different from its declared value. This is, in general, especially true for ceramic capacitors that use a high dielectric constant ("high-k") material (e.g. Z5U, Y5V, and so on). We should also know that ceramic capacitors *age*, except for COG/NPO

types. A typical X7R capacitor ages 1% for every decade of time (in hours). So its capacitance after 1000 hours will be 1% less than what it was after 100 hours, and so on. Higher dielectric constant ceramics like Z5U can age 4 to 6% for every decade of time. So in effect our filter stage, too, gets less efficacious with time. And we need to account for this in the *initial* design.

- Theoretical filter performance is based on the assumption that we are using "ideal" components. However, real-life inductors are always accompanied by some winding resistance (DCR) and some inter-winding capacitance. Similarly, real capacitors have an equivalent series resistance (ESR) and an equivalent series inductance (ESL). At high frequencies, the inductance will start to dominate, and so a capacitor will basically no longer be functioning as one (from the signal point of view). However, capacitors with *smaller* capacitances generally remain capacitive up to much higher frequencies than do larger capacitances. See *Table 10-1* for some typical self-resonant frequencies (the point above which, capacitors start becoming inductive). Therefore, quite often, a *smaller* Y-cap may help, where a large Y-cap is not yielding results. We can also consider paralleling a larger value Y-cap with a small Y-cap.

Table 10-1: Practical limitations in selecting components and materials for EMI filters

X-Capacitors		**Y-Capacitors**	
Capacitance (pF)	*Resonant Frequency (MHz)*	*Capacitance (μF)*	*Resonant Frequency (MHz)*
1000	53	0.01	13
1500	42	0.022	9
2200	35	0.047	6.5
3300	29	0.1	4.5
4700	21	0.22	2.7
6800	19	0.47	1.9
Magnetic Materials for EMI Chokes			
		Initial Permeability	*Bandwidth (MHz)*
Powdered Iron		60	10
		33	50
		22	100
		10	>100
Ferrite		15000	0.17
		10000	0.3
		5000	1.0
		3000	1.2
		2500	1.5
		1500	3.0

- Surface mount ("SMD") versions of off-line safety capacitors are also now appearing — for example from *Wima*, Germany (http://www.wima.com) and *Syfer*, UK (http://www.syfer.com). But we must realize that it is not enough that the capacitor merely 'complies' with a certain safety standard — the capacitor should actually be *approved* (tested by various safety agencies, and carrying their respective certification marks). From the electrical point of view, one of the great advantages of SMD components is their virtually non-existent ESL. This improves their high-frequency performance in any filter application. On the flipside, some ESR or dc winding resistance ("DCR") is often useful in helping *damp* out oscillations. Without any resistance altogether, oscillations would last forever. That is one of the reasons why engineers sometimes pass one or both of the leads of a standard through-hole Y-cap through a small ferrite bead (preferably of a material with lossy characteristics, like Ni-Zn). This can often help suppress a particular high-frequency resonance involving the Y-cap, which is showing up in the EMI scan. But we must be careful that in doing so, we are not ending up with a radiation problem instead.
- Designers of low-voltage, low-power dc-dc converters may find the "X2Y" patented product range available from *Syfer* (and from the company *X2Y* itself — at http://www.x2y.com) very useful if they need to miniaturize and lower the component count. This is a three-terminal integrated SMD capacitor-based EMI filter that simultaneously provides line-to-line and also line-to-ground decoupling. *Picor* (a subsidiary of *Vicor*) at http://www.picorpower.com is also now selling what is billed as the industry's first active input EMI filter stage for standard 48 V bricks. It may be a viable choice if board space is at a premium, despite its roughly $20 cost.
- We note that a Y-cap is always tested to *higher* safety standards than an X-cap. So *we can always use a Y-cap at an X-cap position*, but *not* vice versa. For example, we can consider placing a ceramic Y-cap in parallel with a film X-cap, so as to improve the DM filter bandwidth.
- Generally, we try to maximize filter performance by increasing its 'LC' product as much as practically possible (thus lowering its resonant frequency). Further, given a choice, we would prefer to harness that improvement by using larger capacitances, instead of impractically-sized inductors. But as we know, the maximum Y-capacitance we can use is limited by safety considerations. X-caps too were limited for many years to a maximum value of 0.22 μF (though occasionally 0.47 μF was also seen). But that was simply availability and component technology limitations. Nowadays, we can get X-caps up to 10 μF. We should be conscious, however, that large input capacitances can cause undesirably high inrush surge currents at power-up. This may also cause eventual failure of the X-cap, especially if it is the very first component after the ac input inlet. Film caps can self-heal from

such an event every time it occurs, but eventually the capacitance gets degraded slowly over time with each successive event. Therefore, despite EMI concerns, we should try and place X-caps *after* any input surge protection element — for example, the NTC (negative temperature coefficient) thermistor, or wirewound resistor, and perhaps even after a front-end choke.

Note: What were traditionally called X and Y capacitors are now more accurately called X2 and Y2 capacitors respectively. From the viewpoint of safety regulations (like impulse voltage rating etc.), the X1 and Y1 are considered virtually equivalent to two X2 and Y2 capacitors in series, respectively. For example Y1 caps are impulse tested to 8 kV. Also, the original terms 'X-caps' and 'Y-caps' have recently started getting defaulted to refer to the more uncommonly used (higher voltage) X1 and Y1 capacitors instead.

Note: In off-line power supplies, for better EMI suppression, we may decide to place Y-caps from the *rectified* dc rails (either one or both) to earth. So sometimes we place Y-caps from primary ground to secondary ground (usually connected to earth ground), or from the HVDC (high voltage dc) rail to secondary ground. In either of these positions, a Y1 capacitor (or two Y2 capacitors in series) may be required.

Note: Safety regulations for Nordic regions (and Switzerland) may require each Y-cap shown in *Figure 10-1* to be actually two Y2 capacitors in series (or a single Y1 capacitor). Historically, this has been necessitated by the fact that earthing is poor in those geographical regions. In fact, it used to be pointed out that even the main conference room of the Norwegian safety agency NEMKO (literally Norwegian Electric Material Control) did not have any earth connection available in the wall outlets. Therefore, practically speaking, *a lack of earth is not considered a fault condition in many parts of the world*, but is just a normal condition (this actually also includes about one-third of homes in the United Sates). Therefore, very often, whether the equipment is supposed to be earthed or not, it is expected to have reinforced insulation anyway. Earthing, if present, is then just for helping out with EMI. We see that Y1 caps will often find use even in single-phase equipment. However, X1 caps are basically meant only for 3-phase equipment, since there is no pressing safety need for such a high voltage rating between the L and N wires in single-phase equipment.

Safety Restrictions on the Total Y-capacitance

Y-caps don't just bypass high-frequency noise, but also conduct some of the *low-frequency line current*. That is what the X-caps do too, the difference being that the Y-caps carry this current into the protective earth/chassis. To prevent a fatal electric shock from occurring, international safety agencies limit the total RMS current introduced into the Earth by the equipment to a maximum of typically 0.25 mA, 0.5 mA, 0.75 mA, or 3.5 mA (depending on the type of equipment and its 'installation category' — i.e. its enclosure, its earthing and its internal isolation scheme). But note that somehow, 0.5 mA seems to have become the industry default design value, even in cases where 0.75 mA or 3.5 mA may have been allowed by safety agencies. *It is important to know how high one can actually go in terms of*

ground leakage current, as this dramatically impacts the size and cost of the line filter, in particular the choke.

Keeping the discussion here at a theoretical level, we can easily calculate that we get *79 μA per nF at 250 VAC/50 Hz*. This gives us a maximum paralleled capacitance of 6.4 nF for 0.5 mA, or 44.6 nF for 3.5 mA, and so on. So, a typical configuration in off-line power supplies consists of four Y-caps, each being 1nF or 1.2 nF or 1.5 nF. Or only two Y-capacitors, each of value 2.2 nF. Note that there may be other parasitic capacitances and/or filter capacitances present, which should be accounted for in computing the *total* ground leakage current, and thereby correctly selecting the Y-caps of the line filter. However we must keep in mind that if for better EMI performance/CM noise rejection, a Y-cap is connected from the rectified dc rails to earth (or from the output rails to earth), there is *no* ground leakage current through these capacitors in principle. Therefore there is no limit on their capacitance either.

Equivalent DM and CM Circuits

The filter in the upper half of *Figure 10-1* reduces to the CM and DM equivalent schematics shown in the lower half of the same figure. The equivalent schematics are from the viewpoint of the noise coming out via the bridge, heading toward the mains wiring (LISN). Some observations are:

- We see that the DM choke acts as a CM filter element too.
- The leakage inductance of the CM choke appears as a DM filter element too.
- Both the Y-caps also appear in the DM equivalent circuit (though arguably they will not add much to the heftier X-capacitance).
- Considering that a very small value for L_{dm} is usually enough (because of the much larger X-capacitance possible), no "intentional" DM choke may be required. The leakage inductance of a common mode choke is roughly 1 to 3% of L_{cm}, depending on its construction. That is usually enough to serve as an unintentional, but effective DM choke.
- Though CM chokes usually have a high inductance (and that is certainly needed — particularly for complying with CISPR 22 limits *below* 500 kHz), a good part of the CM noise is usually found in the frequency range of 10 to 30 MHz. So we must consider the fact that not all ferrites have sufficient *bandwidth* to be able to maintain their inductance (A_L) at such high frequencies. In fact, materials with a *high* permeability tend to have a *lower* bandwidth, and vice versa ("Snoek's law"). Therefore a "high-inductance" CM filter may look good on paper, but may not be as

effective as we had thought, at high frequencies. See *Table 10-1* for typical values of initial permeability vs. bandwidth (bandwidth being defined here as a 6 dB fall in permeability).

- A DM noise generator is more like a *voltage* source. So putting in an LC filter works well for a DM source, as it simply presents a "wall" of impedance that serves to block the DM emissions from entering the mains lines. But this strategy by itself is not going to be very effective for CM noise, because a CM noise source behaves more like a *current* source. And we know that current sources *demand* to keep current flowing, and can therefore surmount any "wall" of impedance we may place in their path (by increasing the corresponding voltage). However, if, besides placing a "wall" of impedance, we *also* present an *alternative route* for the current to keep flowing, we would be successful in preventing the CM noise from entering the mains wiring of the building. Thereafter we could "kill" the noise by dissipating the associated energy. This places a rather unusual-sounding demand on the CM choke — not only do we need high bandwidth, but we should actually *lower its quality factor 'Q', especially at high frequencies.* One way to achieve that is to increase the DCR. But that will impede the line current too, and thereby lower the efficiency of the entire power supply. A better solution is to use a "lossy" ferrite material for the CM choke. The usual ferrite used for power transformers and inductors is predominantly of manganese-zinc composition. But lossy ferrites of *nickel-zinc* composition are actually more helpful in "killing" *high-frequency* CM noise components. Unfortunately, they also have such low initial permeabilities that it is impossible to get the desired high inductance (at lower frequencies). Therefore the lossy CM choke is usually an *add-on* to the normal CM filter stage. It could be just a small bead/toroid/sleeve made of similar lossy material, with both the L and N wires passing through its aperture.

- Engineers are often mystified to find that making the DM choke out of (low permeability) powdered iron or lossy ferrite helps too, when all else has failed — despite all the talk about DM noise being essentially a "low-frequency emission." The reason seems to be as follows — the CM noise in a power supply is actually a *non-symmetric* mode, at its point of creation. Though ultimately, by cross-coupling, it does tend to spread into both the lines equally. It has been shown that non-symmetric noise can be considered as a mix of CM and DM components. Therefore in practice, we do get a fair amount of high-frequency DM noise too — arising out of the non-symmetric CM noise. That is why high bandwidth/low permeability/lossy materials can help in DM noise suppression too.

- The DM and CM filters are *usually* laid out in the order shown in *Figure 10-1*. The basic idea seems to be that the *last* stage the noise encounters (as it travels from the

power supply into the mains) should be a *common mode* filter. Because, if the last stage was a DM stage for example, it may not be very well *balanced* from the viewpoint of the noise emerging from the CM filter. And so the CM noise could get converted into DM noise, as previously explained. However, we do have a DM stage now, to hopefully take care of these additional DM noise components! Therefore, many successful commercial designs have *reversed* the order as drawn in *Figure 10-1*, with the DM stage being placed closer to the power inlet. In brief, there seems to be no hard and fast rule for which stage should come before which other stage.

- A possible location for an additional X-cap is directly on the *prongs* of the inlet socket (at the entrance to the power supply). We remember that in this position any line-to-line capacitor will be exposed to a huge current surge at power-up, and could degrade, if not fail immediately. So if this X-cap position seems to be the last resort, it should at least be made as small as possible (typically 0.047 μF to 0.1 μF). Or we can try *ceramic* capacitors in this position (approved ceramic X-caps or Y-caps should be tried here).
- Similarly, the two front-end Y-caps ("C4" in *Figure 10-1*), or two *additional* Y-caps, can also be connected directly on to the prongs of the ac inlet socket, rather than on the PCB. This can help a great deal if the wires going from the PCB to the mains inlet socket are themselves picking up stray fields (and are therefore beyond assistance from the main filter stage, which unfortunately lies on the PCB — before the point of noise injection).
- Sealed chassis mountable line filters (sometimes with integrated standard *IEC 320* line inlets) are available from several companies like *Corcom* (now part of *Tyco Electronics*) and *Schaffner*, Germany (at www.schaffner.com). Such filters perform excellently but are less flexible to subsequent tweaking, and also far more expensive than board-mounted solutions.

  Note: Incidentally, *Schaffner* also makes some of the most widely used, standard test equipment for immunity surge-testing.

- Note that the performance of most commercially available line filters is specified with *50 Ω at both ends of the filter*. Therefore its actual performance in a real power supply may be quite different from what its datasheet says.
- In general, the traces on the PCB corresponding to the filter section should be *thick and wide for low inductance*. CM noise suppression also usually requires a very good high-frequency connection to the enclosure. So, the relevant traces of the PCB should be connected to the chassis through several *metal standoffs* if possible. However, if standoffs are not feasible, the connection (to the enclosure) should be made via thick braids of fine insulated wire. A "good" connection is usually also

helpful between the enclosure and the earth wire (middle prong of the IEC inlet). In that case braided wire can also be used. In the past, major power supply manufacturers had their own special custom-made metal brackets to connect the earth prong of the IEC inlet socket to the enclosure. But nowadays, standard IEC 320 inlets with *built-in* metal brackets are directly available, such as that from *Methode Electronics Inc.* at www.methode.com.

Some Notable Industry Experiences in EMI

One of the most stubborn cases of conducted EMI failure encountered by the author was ultimately (and rather mysteriously) solved by simply *reversing the orientation of the CM choke* (turning it by 180° on the PCB). It was later deduced that the leakage from the core was being picked up by a nearby trace or component, and so the *phase of the coupling* had somehow become an issue (interference pattern). But since most inductors/chokes are symmetrically built, and also do not carry any marking to distinguish one side from the other, implementing such a fix was not easy in production. However nowadays, with so many similar "orientation-sensitive" cases being reported (even relating to the main inductor of the converter itself), some key inductor manufacturers have taken the step of placing a 'polarity mark' on their inductors/chokes.

In another well-documented EMI problem at a leading power supply manufacturing house, it was discovered that the CM choke had to be rotated by 90° (not 180°) to comply. That clearly spells "bad news" if the unit is already in production, because it means the PCB layout has to be redesigned (and perhaps the power supply needs to be requalified too).

This Page Intentionally Left Blank

CHAPTER

11

DM and CM Noise in Switching Power Supplies

This Page Intentionally Left Blank

CHAPTER 11

DM and CM Noise in Switching Power Supplies

Main Source of DM Noise

Now we turn our attention to a real power supply to see for ourselves where all the buzz is really coming from. First consider what would happen if the input bulk capacitor of the power supply had been a "perfect" capacitor: i.e. with zero effective series resistance (ESR) (ignoring all other capacitor parasitics too). Then any possible differential noise source inside the power supply would be completely bypassed by this capacitor. Clearly, the reason this does not happen is the *non-zero ESR of the bulk capacitor.*

So the ESR of the input capacitor is the major portion of the impedance "Z_{dm}" seen by the DM noise generator. The input capacitor, besides being refreshed by the operating current flowing in through the supply lines, also tries to provide the high frequency pulses of current demanded by the switcher. But whenever current passes through any resistance, such as the ESR in this case, there must be a corresponding voltage drop. So we will see a high frequency voltage ripple across the terminals of the input capacitor. See *Figure 11-1*.

This high-frequency voltage ripple shown in Figure 11-1 is in effect the DM noise generator. It is essentially a *voltage source* (V_{ESR}_hf), but producing noise in the form of a noise current I_{dm}.

However, we should take a closer look at *Figure 11-1*. The input line current flows through the diodes only for a brief moment during the AC cycle. That's when the diodes are forward biased. But during the time the diodes are OFF (highlighted in gray on each waveform), the high frequency switching current still continues to flow through the mosfet. This drives V_{ESR} negative. So the high-frequency ripple continues to be seen on the HVDC rail (marked "V_{IN}"). But surprisingly, noise still appears on the line side of the supposedly reverse biased diodes too. That indicates that the DM noise generator tends to behave as a *current source* when the diodes are OFF (dragging in noise through the reverse-biased diodes). We can look at this from another perspective. The bulk capacitor, because of its non-zero ESR is incapable of providing all the entire high frequency content of the switching current. But the inductor, being essentially a current source, is literally not going to take "no" for an answer. The current must come from somewhere, even if it means dragging the voltage on the anode the bridge rectifier diode momentarily low so as to extract current from that route.

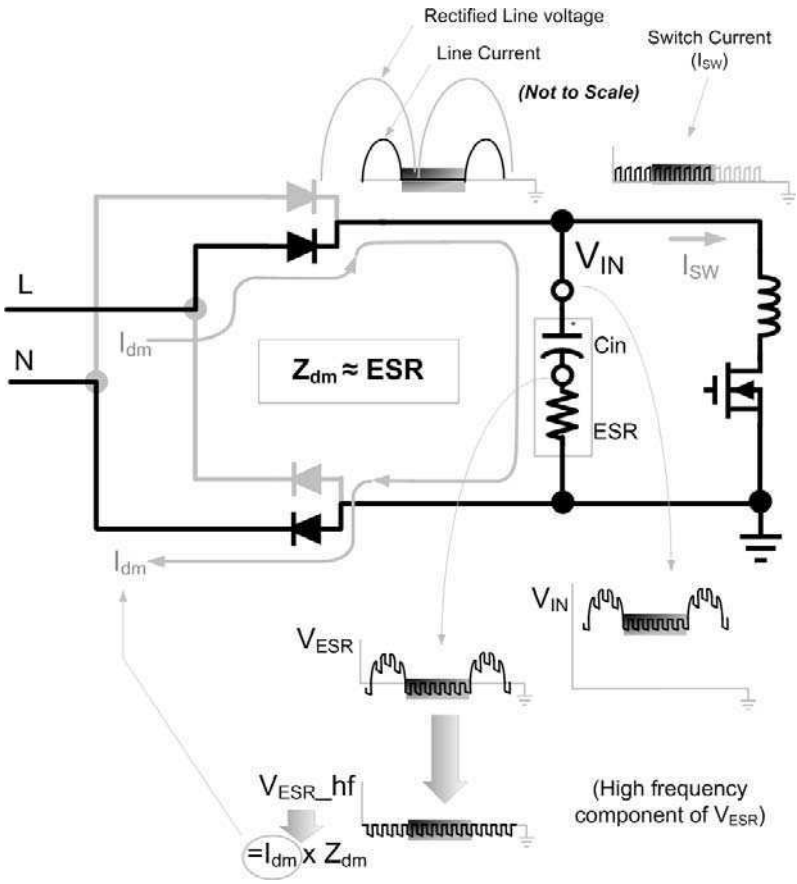


Figure 11-1: How DM noise is created

Therefore, the DM noise generator is modeled as a voltage source during the times when the diodes are ON, but as a current source during the times when the diodes are all OFF. The two models flip-flop back and forth at twice the line frequency. This could make it very hard to analyze. However, it has been seen that *if a small X-cap is placed immediately to the left of the input bridge*, then we can safely assume that the EMI spectrum is dominated by the voltage source, and can thus ignore the current source model. See "C1" in *Figure 10-1*.

Also note that in *Figure 11-1* we have shown I_{dm} as going *into* L and *out of* N. In the opposite half of the ac cycle, *these directions will reverse, along with the ac line current direction*. So the DM current direction "sloshes back and forth" depending on which part of the ac cycle we are on. Of course, from the point of view of the final EMI scan, this makes no difference, as several ac cycles will be looked at by the analyzer for each measurement.

The Main Source of CM Noise

By definition, if there is CM noise, there must be some leakage path to Earth. But in power supplies this path is quite unlike what engineers in other fields may be talking about. For example, in many power supplies, we often use the enclosure to provide us with a fortuitous "infinite heatsink" with which to cool our power devices. We need some electrical insulation, since the tab of the power device is usually the drain of the mosfet, and that point is usually swinging. An insulator for such a purpose needs to be a poor *electrical* conductor, but a good *thermal* conductor — so we can cool the device, while still meeting safety requirements. But we also know that whenever we have two metal plates with an interposing dielectric (the insulator), we create a *capacitance*. From Maxwell's laws we know that if we vary the voltage across these plates, we create a magnetic field, and that is attributable to a current that starts flowing through this parasitic capacitor. In our case, this corresponds to noise current flowing into the earth — in other words "common mode noise." The applicable equation is

$$I = C\frac{dV}{dt}$$

Usually, we don't have much control over the dV/dt, and nor do we really want to reduce it too much in the interest of efficiency. So to reduce this current, we need to reduce C. But a closer look at the root equations reveals a dilemma. The thermal resistance (Rth — in °C/W) is given by

$$R_{th} = \frac{1}{\rho} \times \frac{d}{A}$$

where A is the cross-sectional area of the insulator in m^2 (i.e. the interface area between the device and the heatsink), d is the thickness of the insulator in m, and ρ is the *thermal conductivity* of the insulating material (in W/m-°C). However, the capacitance "C" (in F) is given by

$$C = K \times \varepsilon_O \times \frac{A}{d}$$

where K is the *dielectric constant* of the insulator, and ε_O is the *permittivity of free space* (8.854×10^{-12} F/m). Note that K is dimensionless, being the ratio of the permittivity of the insulating material to the permittivity of air (free space), that is, $K = \varepsilon/\varepsilon_O$. It is also called the *relative* permittivity, ε_r.

Therefore, combining the above two equations, we get Rth as a function of C:

$$R_{th} = \frac{K \times \varepsilon_O}{C \times \rho}$$

We can conclude that

- The relationship between Rth and C does *not* depend on A or d — since only the characteristics of the material remain in the equation above.
- So, if we try to improve (decrease) Rth, the capacitance will definitely increase. And that would clearly increase the CM noise current.
- Because of the inverse proportionality, we can conclude that if we manage to halve the parasitic capacitance, that will give us roughly a 6 dB improvement in EMI—because CM emissions (in dB) would vary according to 20 × log(Ratio of C), and we know that 20 × log(2) ≈ 6 dB. However, we can see from the curve that is also accompanied by a doubling of the thermal resistance of the interface. So if, for example, we previously had a 10°C difference from case to heatsink, we would now have 20°C. And we also know that every 10°C rise of temperature doubles the failure rate of the component (rule-of-thumb). So, we have to weigh the consequences of trying to reduce EMI in this manner against reliability.

Typical values of parasitic capacitance that can be created in a power supply by the insulator are presented in *Table 11-1*. Here we are comparing a traditional insulator material, mica, with a modern choice, silicone rubber.

Note: Mica is a naturally mined mineral (mainly from India). Besides being cheap, it is a very good thermal conductor, and a very poor electrical conductor. Therefore, it was the insulator of choice for many years for mounting power semiconductors on heatsinks. It is still very popular in extra high-voltage applications. However, in power supplies, it fell from favor mainly because of certain production issues — particularly those revolving around the thermal grease that was always required along with it. Besides being messy to apply and hard to control, thermal grease can evaporate slowly (at high temperatures), and this causes a worsening of the thermal resistance over time. Modern materials like silicone rubber have an ability to conform to fairly imperfectly flat surfaces. They therefore require no grease. In fact, the thermal resistance actually falls with time for these insulators.

Table 11-1: Typical mounting capacitances

Package	Area (cm^2)	Material	K	Thickness (mm)	Capacitance (pF)
TO-3	5	Silicone Rubber	5	0.2	111
		Mica	3.5	0.1	155
TO-220	1.644	Silicone Rubber	5	0.2	36
		Mica	3.5	0.1	51
TO-3P	3.25	Silicone Rubber	5	0.2	72
		Mica	3.5	0.1	101
TO-247F	2.8	Silicone Rubber	5	0.2	62
		Mica	3.5	0.1	87

From *Table 11-1* we can see that mica creates higher parasitic capacitances despite a lower K, and that is clearly attributable to the smaller thickness of insulator typically required. The same happens when we use some of the modern, expensive, and yet popular polyimide (not "polyamide"!) insulators which are excellent thermal conductors, but are also very thin. These can be recognized by their typically amber color, and they come in various brand names like Kapton, Kinel, Upilex, Upimol, Vespel, and so on.

So the question is — should we just put in another layer of insulator to solve our EMI problem? In other words, what thickness of insulator do we really need?

The criterion to select a given thickness of insulator is normally based on maximizing thermal performance (as thin as possible) while still complying with any applicable safety requirements, like the required voltage withstand capability. European safety norms require that basic or supplementary insulation be rated at least 1500 V-ac, whereas double or reinforced insulation must be rated over 3000 V-ac. So, for example, a mica sheet of 0.06 mm thickness is typically rated 1000 V-ac, whereas 0.1 mm thick mica is typically rated 1500 V-ac (or 2000 V-ac). Therefore 0.06 mm thick mica usually cannot be used except as *functional* isolation. It can be used in low voltage dc-dc converters, or even in off-line applications where the heatsink is not connected to the chassis/earth. If the line voltage is always less than 130 V-ac (as for equipment intended for use only in the United States), the mandatory dielectric withstand requirement for basic insulation is only 1000 V-ac. Therefore, 0.06 mm mica can be used as basic insulation (with earthing providing the second level of protection). For general acceptability all across Europe, we may always need to place reinforced insulation (rated 3000 V-ac) from primary side to earth — irrespective of earthing (since lack of earthing doesn't count as a fault condition in many regions). In our case, that means two layers of 0.1 mm mica are always required when mounting primary side power devices on to the chassis.

Note: It must be pointed out that some high-end power supply designs (e.g. military grade) use *ceramic* insulators (e.g. beryllium oxide or aluminum oxide, the latter also called alumina). These offer very high thermal conductivities — about 30–50 times better than mica (which has $\rho = 0.7$ W/m-°C) and can therefore be much thicker, so as to reduce the capacitance (they *need* to be thicker too, because they are brittle). We note that beryllium oxide has toxic properties and is therefore not suited for a typical commercial production environment. But use of these ceramic materials can significantly reduce the capacitive noise. There is also an interesting rule called the "45° rule" (degrees of angle not temperature) which has been used successfully by designers of such high-end converters. This rule indicates that you actually decrease the thermal resistance by using larger thicknesses of insulator, basically because more and more of the cross-sectional area of the insulator gets utilized as thickness increases. Note however that like mica, thermal grease is required with these materials too, because of their inherently poor surface finish.

Note: If we want to know how much thermal resistance is typically attributable to thermal grease, we must remember that without this grease we would have air in the spaces between the device and heatsink, and that is a very poor thermal conductor. Thermal grease lowers this interface resistance significantly by filling the

spaces, but it does not establish zero thermal resistance either. We can usually model thermal grease as leaving behind about *0.2°C/W of resistance for each square inch of area of contact*. The thickness of the layer of grease is not significant, only its area of contact. Knowing the total thermal resistance accurately should help in making a better choice of the insulator and trading some thermal resistance off if necessary for lowering the capacitive coupling.

Now we need to understand the *physics* behind common mode noise generation. We will also see why the explanations usually given do not really apply to power supplies. Let us first list two main reasons why this divergence should come as no surprise:

1. In power supplies the main leakage path to earth is not resistive, but capacitive. We also know that in steady state the average current through a capacitor must be zero. So there is no way that a *constant* leakage current can keep flowing into the earth. It must be going *back and forth* so as to keep the average voltage across the parasitic capacitance a constant.
2. In fact, the parasitic capacitance is *not* connected symmetrically to the two input (rectified) dc rails. So why should the ground leakage current end up being shared equally by the two lines?

Now let us look at *Figure 11-2* to see the path the common mode current must actually be taking. Note that we are ignoring the common mode currents that are injected to the secondary side (earthed) through the primary-to-secondary parasitic capacitance present inside the transformer.

We first observe the main path the CM noise current I_{cm} takes (bold arrows) for this particular half of the ac cycle. Note that both schematics (top and bottom) refer to the same ac half-cycle — the top indicates the possible path of current whenever the switch is turning OFF, and the lower schematic, the path when the switch is turning ON. Two diodes are therefore shown as "reverse-biased" all the time, and assuming the diodes are "perfect," only the diodes shown in black tone in the schematics can conduct (even for CM noise). Note that there are also some stray CM paths indicated (dotted arrows), through which a certain amount of noise may be flowing. However, for now let us ignore these extra paths — in particular the component marked "Y-CAP" on the schematic. We can then make the following observations:

- The upper half of *Figure 11-2* shows what happens at the moment the mosfet is turning OFF. The voltage on the drain suddenly goes high. We know that if the voltage across any capacitor changes suddenly, a current is injected through the capacitor, as given by $I = CdV/dt$. This injected current passes into the chassis/earth, and in the process *the capacitor acquires a small amount of charge*.
- The lower schematic shows what happens at the moment the mosfet turns ON. The drain of the mosfet now goes low. So the parasitic capacitance has to give up *all* the

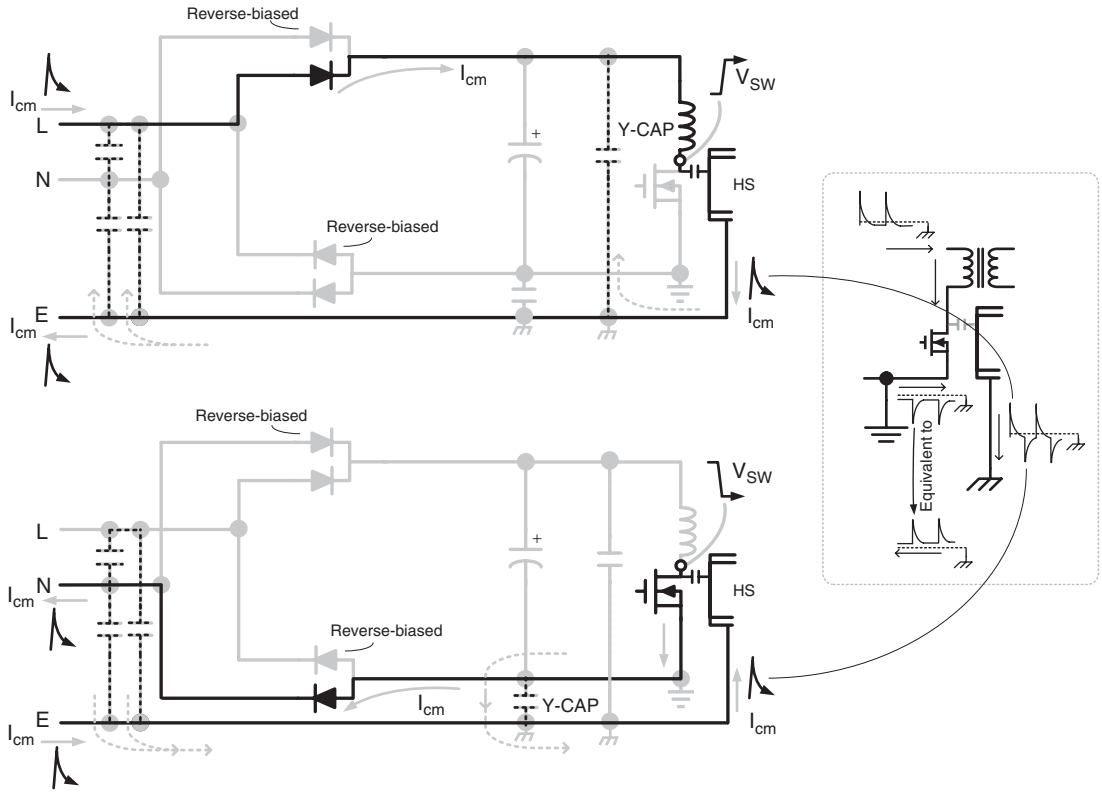


Figure 11-2: How CM Noise Is Created

charge it acquired in the previous step (in steady state). The mosfet therefore turns ON and discharges this parasitic capacitance completely, as indicated.

- We note that when the switch was turning OFF, current was being pulled in through the L wire. And when it turns OFF, the current is pushed out of the N wire. But the latter is equivalent to a current of opposite sign flowing into the N terminal. So eventually, we get the "spiky" CM current shown in the blurb to the right of these schematics. Note that this CM noise is not "dc" as is often suggested in literature.
- We have a *non-symmetrical* CM current flow. That is, we don't have identical currents *at any given instant* in the L and N lines. Further, when the next ac half-cycle comes, the line current *and* the noise current pattern will get *transposed* between the L and N lines. (Calling it CM noise is in that sense really a misnomer — just one of the ways in which various terms seem to have gotten misapplied in this area).
- Whenever we command the mosfet in any power converter to turn OFF, the inductor does *not* allow the current in the mosfet to change — until a freewheeling path is

available. The freewheeling path is provided by the catch diode (not shown in *Figure 11-2*). But for this diode to become "available" (conduct) it must get forward-biased. Which means that the voltage across the mosfet has to rise *fully*, before the current through it even starts to diminish. But for the mosfet voltage to rise up, all the parasitic and non-parasitic capacitances preventing it from doing so must get charged up too. We know that, for example, one of these capacitors is the drain-to-source capacitance (the C_{OSS} of the mosfet). Another such capacitor that we can now identify is the *parasitic mounting capacitance to earth*. Therefore, in its case too, the inductor current is responsible for pushing current through it (thereby charging it up as required). In other words, the parasitic capacitance "brings the whole weight of the inductor to bear" on the situation. And that is the reason why in a switching power supply, the so-called "CM noise generator" is said to behave as a *current source*.

- Now, the drain-to-source capacitance *has* to get charged up for the diode to start freewheeling, because the other end of this capacitance is firmly connected to a fixed voltage rail (the primary side ground). However, in principle, the parasitic capacitance to earth *need not* get charged up at all (for freewheeling to be realized). In fact, we can "enforce" zero current flow through this parasitic capacitance by simply breaking the galvanic connection (continuity) to the earth wire (that is coming in from the mains — assuming no filter stage is present so far). And as expected, this then has no effect on the actual switching process. But what we have done in the process is allowed the enclosure (the other side of this capacitor) to "float." Let us see how that happened. The leakage current through the parasitic capacitor is related to the dV/dt across the parasitic capacitor by the equation $I = CdV/dt$. So if this parasitic charging current is made zero (by breaking its path), the dV/dt must be zero too. However, on one side of this capacitor, we have a fixed dV/dt (with respect to ground) — created by the switching of the mosfet. So the only way the dV/dt *across* this parasitic capacitor can be zero is if both plates of the capacitor have the same dV/dt, that is, no net change in the voltage *across* the capacitor. What all this simply means is that if we don't have a galvanic connection to the earth wire, the enclosure will eventually develop a dV/dt exactly equal to that present on the drain of the mosfet, and it will therefore start *radiating*. So we may have succeeded in improving the conducted emissions spectrum (by virtually *disallowing* CM noise from entering the mains wiring), but we are surely stuck with a radiation problem now.
- Therefore what we really want to do is to *provide* a path for the CM current to flow. By doing this, we can prevent the dV/dt from developing on the chassis. For minimizing noise in general, we must actually ensure that all the grounding (earthing) connections — from the PCB to the enclosure, and on to the earth wire, are *good*. Any intervening PCB traces should also be wide, and of low inductance.

- But having now allowed the I_{cm} to flow, how do we control (or limit) it!? First, we need to prevent it from creating strong electromagnetic fields. So our main goal should be to *minimize the loop area of the CM current path*, so as to prevent it from becoming an effective H-field antenna. We also need to *divert* this current *away* from the mains wiring (by providing an alternate path — and thereby returning the current to its source). We thus realize the important role played by the two *additional Y-caps* marked 'Y-CAP' in *Figure 11-2* (connected between the rectified dc input rails and earth). One or the other, or *both* of these capacitors, are commonly seen in commercial power supplies, and they always help in providing several valuable decibels of additional EMI suppression. They must be placed *very close to the mosfet* — and with low inductance connections (via standoffs in the enclosure for example).
- Since these additional Y-caps hardly pass any ac line frequency leakage current into the earth (being on the rectified side of the bridge), they are not subject to the previously described safety considerations regarding ground leakage currents. Therefore we can make them quite large in capacitance. However, as per safety regulations, we still can't ignore what their *voltage rating* needs to be. So in this position, we usually need two Y2 caps in series (or a single Y1 cap).
- We can see that the CM noise in power supplies tends to be "non-symmetric." However, the X-cap and Y-caps just before the diode bridge *(i.e. toward the incoming supply lines) help in distributing this noise almost equally between the L and N lines*. And that is important if we want the common mode filter that follows to work as envisaged. Otherwise, we will find that it isn't working as well as we expected. And if we didn't know better, we could be needlessly trying to increase the size of the CM choke (but we may need to try increasing the DM choke!).
- Even seasoned engineers are often extremely nervous about chassis-mounting of power devices. Often they can be coaxed into mounting the output diodes in this manner, but *not* the high-voltage mosfet. But actually, if the Y-caps shown in *Figure 11-2* (marked "Y-CAP") are provided for, and *they return the injected noise back very close to the mosfet*, there is really no problem. To help this process, a metal standoff from the enclosure to the PCB should be positioned very close to where the mosfet is mounted, and a Y-cap from the drain can then be connected right there (see *Figure 11-3*).
- Effective CM noise suppression usually requires a very "good" connection to earth. So the earth traces should be very thick and preferably straight — along the length of the PCB — with several metal standoffs if possible, to establish good high-frequency connection from the PCB to the chassis. If this is not done, and supposing the connection is made only toward the ac inlet, and also with wire that is not of very

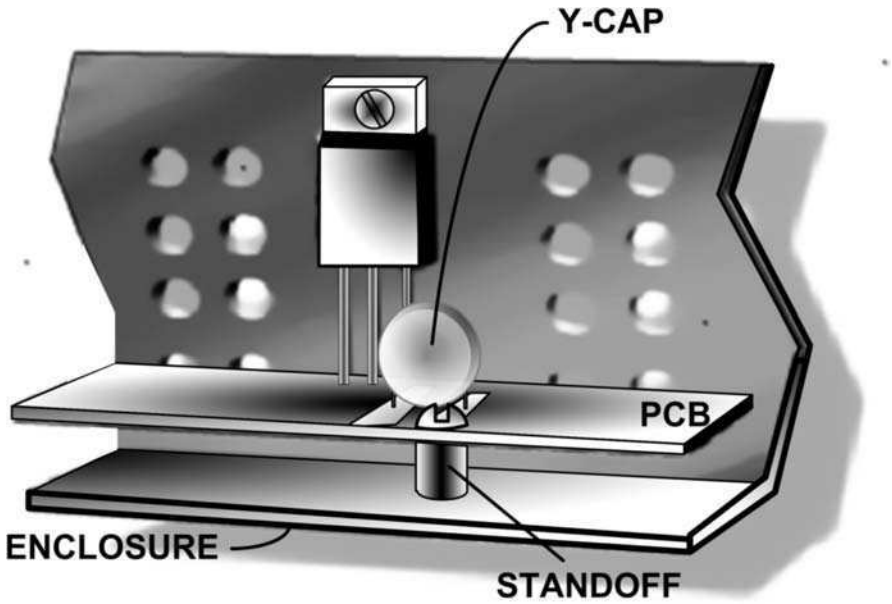


Figure 11-3: How to Mount Power Devices on the Enclosure

low inductance, the enclosure can start radiating as indicated in *Figure 11-4*. We can visualize that *board-mounted IEC inlets* will work much better because of the more direct connection they can provide to help return the CM noise back to its source.

- The entire loop of the PCB traces (up to the input side) as shown in *Figure 11-4* needs to be thick and short. Unfortunately, this often tends to be necessarily long, considering board layout constraints, and all the other components that need to be mounted on it. So in that case we can provide a *high frequency decoupling capacitor from the HVDC to primary ground, very close to the mosfet*.

  Note: Copper traces can't provide a very low inductance if they are *long*, however *wide* they may be. We must remember that though halving the length of any trace does roughly halve its inductance, we have to increase the width of a trace by a factor of 8 to 10 to halve its inductance (see *Chapter 6*).

- Some engineers try to get the "best of both worlds," by mounting the device on the enclosure, but with special insulators — which come with a built-in 'Faraday shield.' This is actually just a thin metal layer sandwiched between layers of insulator. It is supposed to be connected on the PCB to primary ground, and thereby it 'collects' the injected noise and returns it, without letting it pass into the enclosure. However because of safety requirements, such composite insulators are usually very thick, and their thermal resistance is usually unacceptably high — defeating the very purpose of chassis-mounting.

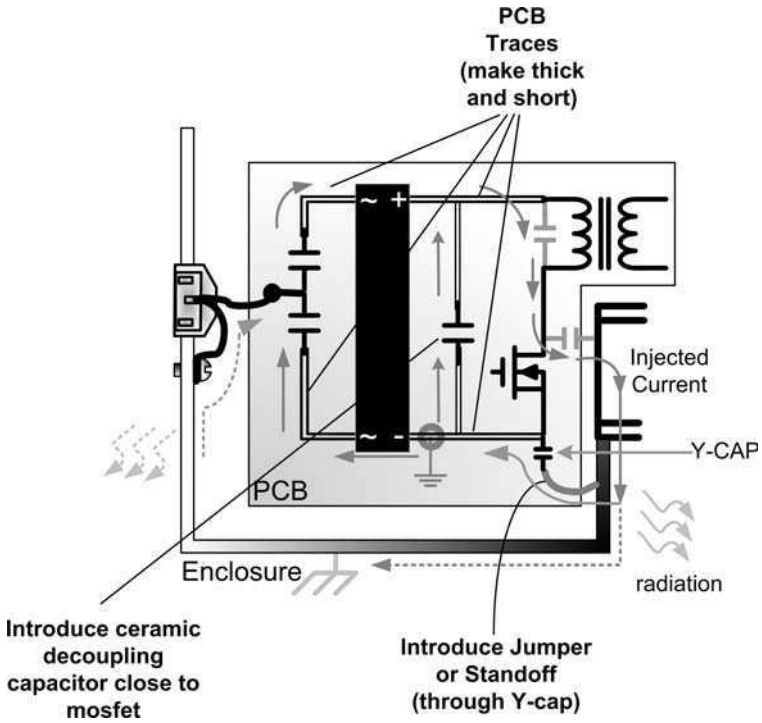


Figure 11-4: Preventing the Enclosure from Radiating

- A "ground choke" should be avoided at all costs. Think of what it can do if we put this in *Figure 11-4*, say on the wire connecting the PCB to the power inlet. See the following discussion.

The Ground Choke

We ask — is it really a good idea to place a small inductor (e.g. a bead or small toroid with a few turns) somewhere in the earth connection? Suppose we place it on the wire connecting the ac inlet to the enclosure (or PCB to inlet). This is then called a 'ground choke' or 'earth choke.' It is commonly found on low-power evaluation boards (from vendors promoting their "clever" IC solutions), but rarely seen on a commercial power supply.

We first note that the idea of such a choke seems to be at odds with our previous suggestion of a *good* high frequency connection to earth. When we place the ground choke, we are basically trying to prevent conducted CM noise from flowing into the mains wiring. But in return, we may have a radiation problem. In addition to that, there are industry-documented cases where the ground choke has caused severe *system* problems. For example if a power

supply is turned on at the peak of the input ac waveform, it produces a very high initial surge of charging current through the Y-caps. If there is a ground choke present, it causes the voltage on the earth traces *and the enclosure* to locally "bump up." Now in most cases, the return of the output rails of the power supply is also connected directly to the enclosure, and forms the ground plane for the entire system. The system would also typically connect to the chassis/enclosure at several points downstream. So this surge-induced bump, around the power supply, causes severe imbalances across the system ground plane — leading to data upsets and even destruction of the subsystems. A similar situation will arise during ESD testing and conducted immunity testing, in which surge voltages are applied from line to line, or from line to earth. So however tempting it may seem to the power supply designer (who is focused only on solving his or her conducted mode EMI problem, and going home!), *a ground choke should be avoided at all costs.* Some high-voltage semiconductor companies, who are only making *open-frame* (enclosure-less/standalone) evaluation boards, seem to have nothing to lose, and everything to gain, by putting in a ground choke. They know that being open-frame anyway, no one expects them to comply with any *radiation* limits. So they quietly push the problem they may have been seeing in their conducted emissions plot — toward a future radiation emissions saga for the systems designer. Beware!

CHAPTER

12

Fixing EMI across the Board

This Page Intentionally Left Blank

CHAPTER 12

Fixing EMI across the Board

1-oz (okay 2-ounces!) of prevention are always better than cure. So here we look at some of the practical design aspects involved in controlling and testing EMI.

The Role of the Transformer in EMI

Very often an engineer resolves a stubborn EMI problem by just 'playing' with the transformer. The transformer comes into the picture in the following ways:

- With its windings carrying high-frequency current, it becomes an effective H-field antenna. These fields can impinge upon nearby traces and cables, and enlist their help in getting transported out of the enclosure, via conduction or radiation.
- Since parts of the windings have a swinging voltage across them, they can also become effective E-field antennas.
- The parasitic capacitance between the primary and secondary windings transfers noise across the isolation boundary. Since the secondary side ground is usually connected to the chassis, this noise returns via the earth plane, in the form of CM noise. The situation is very similar to the tradeoffs required in heatsink mounting issues. In this case, we wish to couple the primary and secondary very close to each other in order to reduce leakage inductance (especially in flyback transformers), but this also increases their mutual capacitance, and thus the CM noise.

Here are some standard techniques that help prevent the preceding:

- In a safety-approved transformer, there are three layers of safety-approved polyester ("Mylar®") tape between the primary and secondary windings — for example the popular #1298 from *3M* (at www.3m.com). In addition to these layers, a copper 'Faraday shield' may be inserted to "collect" the noise currents arriving at the isolation boundary, and diverting them (usually to the primary ground). See *Figure 12-1*. Note that this shield should be a very thin strip of copper foil, so as to avoid eddy current losses and also keep the leakage inductance down. So, it is typically 2 to 4 mils thick, consisting of one turn wound around the center limb.

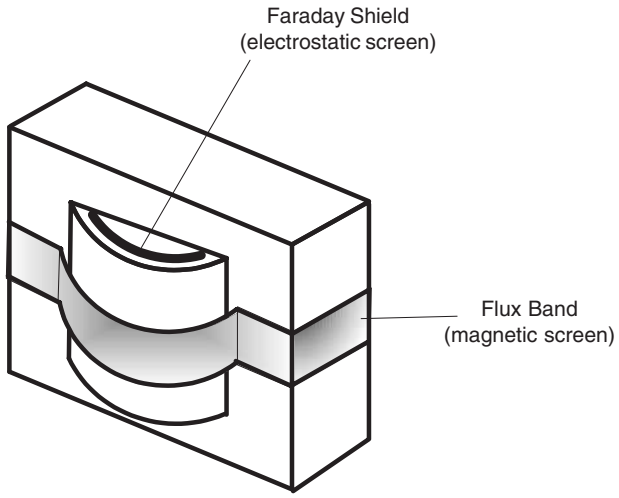


Figure 12-1: Screens used for transformers

A wire is soldered close to its approximate geometric center and goes to the primary ground. Note that the ends of the copper shield should *not* be galvanically connected together, as that would constitute a shorted turn from the viewpoint of the transformer. Some designs also use another Faraday shield, on the secondary side (after the three layers of insulation). This is connected to the secondary ground. However, most commercial ITE (information technology equipment) power supplies don't need either of these shields, provided adequate thought has gone into the winding and construction, as we will soon see.

- There is usually also a circumferential copper shield (or "flux band") around the *entire* transformer. See *Figure 12-1*. The ends of this shield can be, and usually are, shorted (soldered) together. It serves primarily as a radiation shield. This is often left floating in low-cost designs. However, it may be connected to the secondary ground if desired. And if so, safety issues will need to be considered, in regard to the requirement of reinforced insulation between primary and secondary, and also the required primary to secondary 'creepages' (distance along the insulating surface) and 'clearances' (shortest distance through air) as applicable. When the transformer uses an air gap on its outer limbs, the fringing flux emanating from the gap causes severe eddy current losses in the band. So this band is also usually only 2 to 4 mils thick. Like the Faraday shield, this too can often be omitted by good winding techniques.
- To reiterate, from the point of view of EMI, a flyback transformer should be preferably *center-gapped*, that is, no gap on its *outer* limbs. The fringing fields from exposed air gaps become strong sources of radiated EMI besides causing significant eddy current losses in the surrounding copper band.

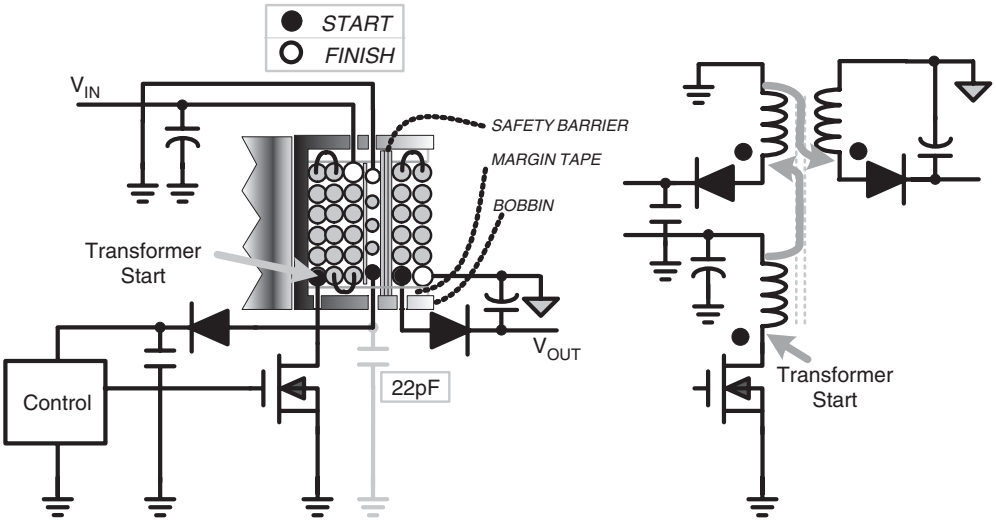


Figure 12-2: Low-noise Transformer Winding Technique

- There is usually an *auxiliary winding* present on the primary side, which provides a low voltage rail for the controller and related circuitry. One end of this is connected to primary ground. Therefore, it can actually double over as a crude Faraday shield if we a) wind it evenly and *spread it out* over the available bobbin width, and b) we help it collect and divert noise by ac-coupling its opposite end (i.e. the diode end) to primary ground, through a *small 22 to 100 pF* ceramic capacitor as shown in *Figure 12-2*.

Figure 12-2 also reveals a *low-noise construction technique*, as applied to a typical flyback transformer. We should compare the right-hand schematic with its equivalent "winding" version on the left. In the following discussion, we note that though transformers with split windings are not being explicitly discussed here, the same principles can be easily extended and applied to them too. Here are some observations on *Figure 12-2*.

- Since the drain of the mosfet is swinging, it is a good idea to keep the corresponding end of the primary winding buried as deep as possible, that is, it should be the *first layer* to be wound on the bobbin. The outer layers tend to shield the fields emanating from the layers below. For sure, the drain end of this winding should *not* be adjacent to the 'safety barrier' (the three layers of polyester tape), because the injected noise current is proportional to the net dV/dt across the two "plates" of the parasitic capacitor (formed by the windings on either side of the interface). Since we really cannot reduce the capacitance much (without adversely impacting the leakage inductance), we should at least try to reduce the net dV/dt across this interface capacitor.

- Comparing the diagram on the left with its schematic on the right, we see that the "start" and "finish" ends of any winding have also been indicated. In particular, all the start ends have been shown with *dots* in the schematic. Note that in a typical production sequence, the coil winding machine always spins the bobbin in the same direction, for every layer and winding placed successively. Therefore, since *all the start ends* (*i.e. dotted ends*) *are magnetically equivalent* — if one dotted end goes high, the other dots also go high at the same moment (as compared to their opposite ends). We can also see that from the point of view of the actual *physical proximities involved*, *every dotted end of a winding automatically falls close to the nondotted end of the next winding* (with the usual fixed winding direction). This means that for the flyback transformer of *Figure 12-2*, the diode end of the secondary winding will *necessarily fall adjacent to the safety barrier*. And because of that we will have a certain amount of dV/dt still present *across* the barrier. But note that this dV/dt is *much smaller than if the drain end of the primary winding was brought adjacent to the safety barrier* (because of the bigger voltage swing on the primary side, due to the large turns ratio). However, the transformer as shown in *Figure 12-2* now has the advantage that the "quiet end" (ground) of the secondary winding is now the *outermost* layer. That is by itself a good shield. So we can safely drop the ubiquitous circumferential shield (flux band). Consider the alternative — suppose we had wound the transformer the "wrong" way, that is, by reversing *all* the start and finish ends shown in *Figure 12-2*. That would have brought the drain end of the primary winding right next to the safety barrier, with the secondary ground end (which is usually connected to the chassis) directly across the isolation boundary. With this winding arrangement, we would have a healthy dose of CM noise injected directly into the chassis/earth — not the best way to achieve compliance for sure!
- When we go through the same reasoning for a forward converter transformer, we will find that with the described winding sequence, we will automatically have the quiet ends of both primary and secondary windings overlooking each other across the safety barrier (isolation boundary). That is because the relative polarities between the primary and secondary windings in a forward converter are opposite to those of a flyback transformer. So now, very little noise will be injected through the parasitic capacitance. That is good. But the outermost layer is not "quiet" anymore, and we could have a radiation problem. So in this case, the circumferential shield may become necessary.
- Another way out of the forward converter "outer surface radiation problem" is to ask production to *reverse* the direction of the secondary winding (only). So for example, if up to the finish of the primary winding, the machine was spinning clockwise, for the secondary we should specify an anticlockwise direction (with expected resistance coming, not from the transformer, but our production staff!). If we do that, the

reasoning given previously for the flyback will now apply unchanged to the forward converter transformer too. So we would now have a "quiet" exterior (without any flux band necessary), though some more common mode noise transferred across the isolation boundary, due to the dV/dt. Note that in general, aiming for a "quiet" exterior seems to be a better option than trying solely to prevent noise injection through the interface capacitance, because the latter can be overcome by various tricks — like having the auxiliary winding double over as a Faraday shield, and so on. But a radiation problem can be hard to manage. We do note however, that a forward converter transformer has no (or very small) air gap, so it is generally considered "quieter" in terms of radiation to start with (as compared to a flyback).

Tip: We don't need to draw *any* current at all from this 'Faraday winding' to make it work. So it need not even be required by our circuitry (for an auxiliary rail). In that case, we could just wrap a few turns of thin wire (spread out evenly), with one end of it connected to primary ground, and the other end via a small 22 pF capacitor to primary ground. This technique certainly saves production costs associated with the making and placing of a formal Faraday shield — not to mention the improvement in efficiency due to the reduced leakage inductance (as compared to what a formal Faraday shield may lead to). In that sense, this informal Faraday shield is a very useful idea, certainly worth trying out.

- When the transistor is mounted on the chassis for thermal reasons, there is a technique that is used to actually try and *cancel the current* injected through the heatsink capacitance. This is done by placing another winding, equivalent to the main winding, and *opposite in phase* (though it can be of much thinner wire). See *Figure 12-3*. The idea is that if the noise current is being *pushed out* from the primary winding, the cancellation winding gets the same current *pulled in*. Therefore in effect, the injected current does a quick "U-turn" back to its noise source. Note that this additional cancellation winding should be very closely coupled to the main winding. Often it is wound *bifilar* with the primary winding (i.e. both wound simultaneously, rather than one on top of the other). However we

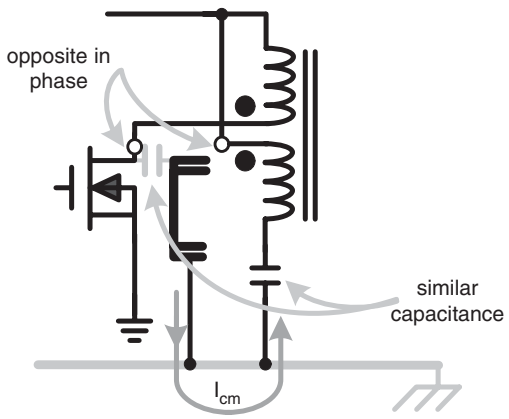


Figure 12-3: Cancellation Winding to Reduce CM Noise

should be aware that in that case, we will have a high voltage *differential* between the two windings at points along their length. So if, for example, there are pinholes in the enamel insulation, there is a danger of flashover and resulting failure of the power supply. The solution is to use wires with "double insulation."

Note: This technique does nothing to cancel the noise injected through the *interface capacitance* (i.e. between the primary to secondary windings). But despite that limitation, a 10 dBμV reduction in conducted EMI is still possible (at various points in the EMI spectrum). So this could certainly be worth trying out, if there is a last-minute problem and a major redesign of the board needs to be avoided. It therefore may be prudent to plan for this winding in advance, including a PCB placeholder for the additional capacitor.

Note: The above idea can clearly be applied to *any* off-line topology (and also all high-power dc-dc converters) — whenever the switch needs to be mounted on the chassis/enclosure (and its tab is swinging). A similar technique may be useful on the secondary side too, if suppose the catch diode is to be mounted on the chassis. However this secondary-side heatsink noise injection is of concern only when the tab of the diode (which is almost invariably the cathode of the diode) *happens to be the switching node* for that particular topology/configuration. So we can work out that the normal boost and flyback topologies don't have this problem, since the cathode end of their diodes are "quiet." However, the (positive-to-positive) buck and the forward converter do have swinging cathodes (tabs), so we should be careful when chassis-mounting their diodes.

- *Rod inductors* are often used in LC post-filtering stages on the output. Because of their open magnetic structure, they have been called "EMI cannons." But they are nevertheless still popular because of their low cost, and also the low "real estate" they need. Some tricks have therefore been developed to control their ill-effects. First, they should be placed vertically (as they normally are). Further, if *two* such rods are being used on a given output, we should wind the two rods identically, but *reverse* the current flow in one of them, as compared to the other (by suitable modification of the PCB). See *Figure 12-4*. So, looking from the top, one rod should be carrying current clockwise and the other anticlockwise. This helps *redirect* the flux from one, back into the other ("U-turn"). In that way, much less "EMI-spilling" occurs.

EMI from Diodes

Here we list some of the things to keep in mind and try out, regarding diodes:

- Diodes are a potent source of low- to high-frequency noise. Slow diodes (like those in a typical input bridge) can also contribute to such noise.
- Input bridges that use ultrafast diodes are available, and their vendors claim significant reduction in EMI. But in practice they don't seem to provide much advantage. They also typically have much lower *input surge current* ratings. In fact

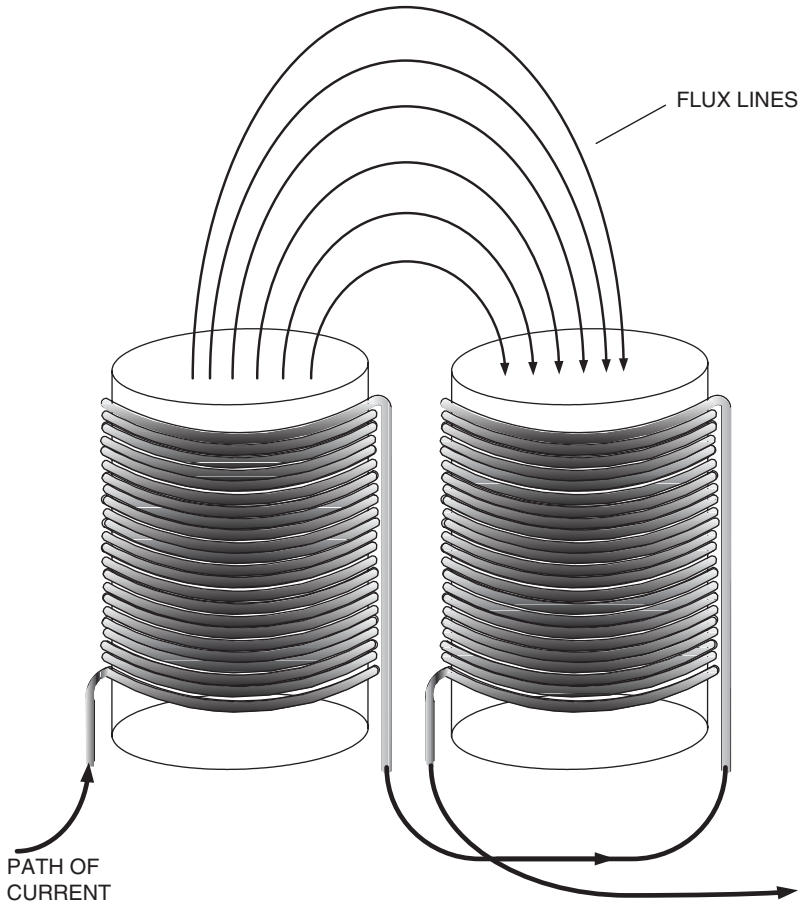


Figure 12-4: Correct Way to Use Rod Inductors

in a front-end position, any component always needs to be able to handle a lot of stress (if not abuse), such as the stresses occurring during power-up at high line.

- To minimize EMI, ultrafast diodes should be selected on the basis of *softer* reverse recovery characteristics. For medium- to high-power converters, RC snubbers are also often placed across these diodes (at the expense of some efficiency). In low-voltage applications, Schottky diodes are often used. Though these diodes have no reverse recovery time in principle, their body capacitance can be relatively high, and can end up resonating with PCB trace inductances. So an RC snubber is also often helpful for Schottkys. Note that if any diode has fully recovered (i.e. zero current) before the voltage across it starts to swing, there is no reverse recovery current. In that case, diodes really don't have to be 'super-super-fast.'

In fact many engineers have reported much lower EMI by choosing *slower* diodes for snubbers/clamps. A popular choice for snubber applications is the soft-recovery fast diode BYV26C (or BYM26C for medium power) from *Philips*.

- It is advisable to have the mosfet switch roughly two to three times *slower* than the reverse recovery time of the catch diode — to avoid shoot-through currents — that will produce strong H-fields (in addition to causing dissipation). Therefore it is not uncommon to intentionally degrade the mosfet switching speed by adding a resistor (typically 10 Ω to 100 Ω in off-line applications) in series with the gate — maybe with a diode across the gate resistor so as to leave the turn-off speed unaffected (for efficiency reasons).
- Small capacitors may often be placed across the mosfet (drain to source). But this can create a lot of dissipation inside the mosfet, since every cycle the capacitor energy is dumped into the mosfet. ($P = 1/2 \times C \times V^2 \times f_{SW}$).
- Ultrafast diodes also exhibit high forward-voltage spikes at turn-on. So momentarily, the diode forward voltage may be 5 to 10 V (rather than the expected 1 V or so). Usually, the snappier the reverse recovery, the worse is the forward spike too. Therefore, at mosfet turn-off, the diodes become strong E-field sources (voltage spike), whereas at mosfet turn-on, the diodes will generate strong H-fields (current spike). A small RC snubber across the diode will help control the forward voltage spike too.
- In integrated switchers, access to the gate of the mosfet may not be available. In that case, the turn-on transition can be slowed by inserting a resistor of about 10 Ω to 50 Ω in series with the bootstrap capacitor. The bootstrap capacitor is in effect the voltage source for the internal floating driver stage. At turn-on, it is asked to provide the high current spike required to charge up the gate capacitance of the mosfet. So a resistor placed in series with this bootstrap capacitor limits the gate charging current somewhat, and thereby slows the turn-on.
- To control EMI, ferrite beads (preferably of lossy nickel-zinc material) are sometimes placed in series with catch diodes (often slipped on to their leads), such as at the output diode of a typical off-line flyback. However, these beads must be very small, as they can have a significant effect on the efficiency of the power supply.

Note: In multioutput off-line flyback converters, we may find larger beads (possibly with more than one turn, and made of the more common manganese-zinc ferrite) in series with the output diodes belonging to some of the *auxiliary outputs* (i.e. those not being directly regulated). But the purpose of these beads is not EMI suppression, but to block some of the voltseconds and thereby improve the "centering" of the outputs.

- A comment about split/sandwich windings. In general, the primary winding may be broken up into two windings, which are then positioned on either side of the secondary winding — so as to reduce leakage inductance in flybacks, and proximity

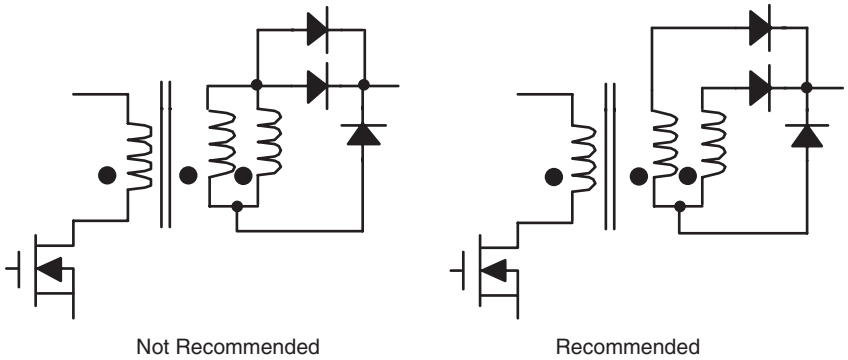


Figure 12-5: Correct Way to Parallel Windings

effect losses in forward converters. This is acceptable for EMI provided the two split winding sections are *in series*. In general, putting windings in parallel is *not* a good idea (especially from the EMI point of view). In high-current power supplies, the secondary winding is also sometimes broken up into two windings (or foils). The intention is usually to increase the current handling capability. See *Figure 12-5*. But these split secondaries are also usually placed *physically apart*, on either side of the primary winding. However, in paralleled windings, the two supposedly "equal" sections are actually always magnetically slightly *different* — because of their different physical positions inside the transformer. Plus, their DCR is also just a little different (different lengths), creating the possibility of an *internal current loop*. The designer may be completely unaware of this, except for the severe tell-tale ringing on the voltage waveform, and a mysteriously bad EMI scan. So if paralleling is *really needed*, it is better to use the scheme in the right-hand side schematic of *Figure 12-5*. Here the forward-drops of the two diodes help "ballast" the windings, and this also helps "iron out" any inequality between the two halves.

Beads, and an Industry Experience — the dV/dt of Schottky Diodes

In a very high-volume power supply design and manufacturing house, the following situation arose. The output Schottky diode of an off-line 70 W flyback had a small but unacceptably mysterious "ppm" (parts per million) failure rate in production testing. Finally, by careful analysis, this was traced to a very slight wiggle (ringing) somewhere in the middle of the turn-off waveform of the diode. Then, by drawing asymptotes, it was seen that the *dV/dt rating* of that particular diode was being momentarily exceeded at the point of the wiggle, thus probably causing its failure (no other overstress could be seen). A small ferrite bead was inserted on the leg of the Schottky diode. This smoothed out the wiggle, reduced the EMI dramatically in the bargain, and as a proof of the hypothesis, no Schottky failures were

seen thereafter. There was a 1 to 2% loss in overall efficiency though. So, not all Schottkys are created equal. We must be conscious of the different dV/dt ratings they can have, depending on their vendors. And also of the leakage current through them, which can not only cause loss in efficiency, but in certain cases anomalous behavior — like premature current limiting at turn-on (especially in integrated switchers where the switch drop is being sensed for implementing current limiting).

Note: On the topic of beads, note that beads also sometimes have been put in series with the mosfet. But we should not put any such bead in the source. If we do so, then during crossover transitions, the source pin (with bead) can develop spikes. And since the gate is referenced to the source, not the drain, this can lead to a spurious turn-on, resulting in reliability issues. Therefore, a bead, if necessary, should be placed only on the drain side of the mosfet. True, this extra uncoupled inductance can also cause a small spike in principle, but in practice, that is rarely an issue. For the same reason, if we want to monitor the current in the mosfet by means of a current probe, we should place the loop of wire (to slip the probe tip on to) on the drain side, never on the source.

Basic Layout Guidelines

For each topology, we need to carefully figure out which PCB trace segments are "critical" in terms of layout and EMI. "Critical" traces are "high-frequency current sections" — in which current is forced to either *start flowing* or *stop flowing* (suddenly) at the instant of turn-on or turn-off. So at each transition we get a very high dI/dt across such traces. Further, from the rule-of-thumb "20 nH per inch of trace," we get a voltage spike according to $V = LdI/dt$. These spikes can not only cause a lot of EMI, but can also infiltrate into the control sections of the IC, causing anomalous behavior (and possibly switch destruction). To minimize the fields, and simultaneously reduce the associated inductance, the *area* enclosed by such high-frequency current loops must be minimized. Therefore, analyzing the topologies from this viewpoint, we get the results summarized in *Table 12-1*. All traces leading to components marked "Critical" must be kept very short (and not too thin!). The corresponding high-frequency loop areas will then be minimized automatically.

We indicated above, that layout concerns and EMI concerns generally overlap. In other words, what is good in terms of layout (ensuring proper performance) is also good for EMI. There is, however, one possible exception to this trend — in particular we need to be careful

Table 12-1: Good component placement

	Input Cap	Output Cap	Catch Diode	Inductor
Buck	Critical	Not Critical	Critical	Not Critical
Buck-Boost	Critical	Critical	Critical	Not Critical
Boost	Not Critical	Critical	Critical	Not Critical

about inadvertently making traces that have a swinging voltage on them too *wide*, as they can become good E-field antennas. The prime example is the trace at the switching node of any topology. We may be wanting to increase its copper area, for the purpose of lowering parasitic inductance and/or helping dissipate heat from the mosfet or catch diode, but *we must do this judiciously*.

The *ground plane* is a very effective method of bringing down the overall level of the EMI emissions. On a multilayer board, if the very next layer to the side containing the power components (and their associated traces) is a ground plane, the EMI can drop by about 10 to 20 dB. This is clearly more cost-effective than opting initially for a "cheap" one- or two-sided board, and then having to use bulky filters later. However the *integrity* of a ground plane should be maintained, as far as possible. For this, we should remember that return currents tend to travel by the shortest *straight line* path at low frequencies. But at high frequencies (or the higher harmonics of the waveform), the return currents tend to image themselves directly under their respective forward traces (on the opposite layer). Therefore, currents, given a chance, automatically try to reduce the area they enclose — as this lowers the self-inductance, and thereby offers the current the *lowest impedance* route possible (at low frequencies, trace impedances are resistive, but at high frequencies they are inductive). So in particular, if we make *ill-considered cuts* in the ground plane (possibly with the intention of "conveniently" routing some other trace), the return currents of the power converter stage (which really need this ground plane) will get diverted along the sides of any intervening cuts. And in doing so, will form effective *slot antennas* on the PCB.

Last-ditch Troubleshooting

It is helpful to separate the CM and DM components to be able to study them and debug a bad EMI scan. But a standard LISN reading only provides a certain weighted *sum* of the *total* conducted noise (CM and DM). Therefore, unless special accessories are available (including a modified LISN), we can only *guess* which part of the EMI scan is mainly DM and which is CM. So we may never know the root cause of the noise either. In *Figure 12-6* we have shown two current probes, wired up in such a way that they are actually performing "simultaneous equation" math on the L and N wires — to separate the CM and DM components (also see *Figure 9-1* in *Chapter 9*). Note that by doing these two measurements at the same time (using two probes rather than one), we have also retained valuable information about the relative phase relationship, existing between the CM and DM components.

> **Note:** The bandwidth and current capability of the current probes used for noise measurements are important. Popular choices for such probes are from *Pearson Electronics* and *Fischer Custom Communications* at www.pearsonelectronics.com and www.fischercc.com, respectively. For very high currents (up to thousands of amperes if necessary), a possible choice are current probes based on the "Rogowski principle." This type of probe is available from several manufacturers, for example *Power*

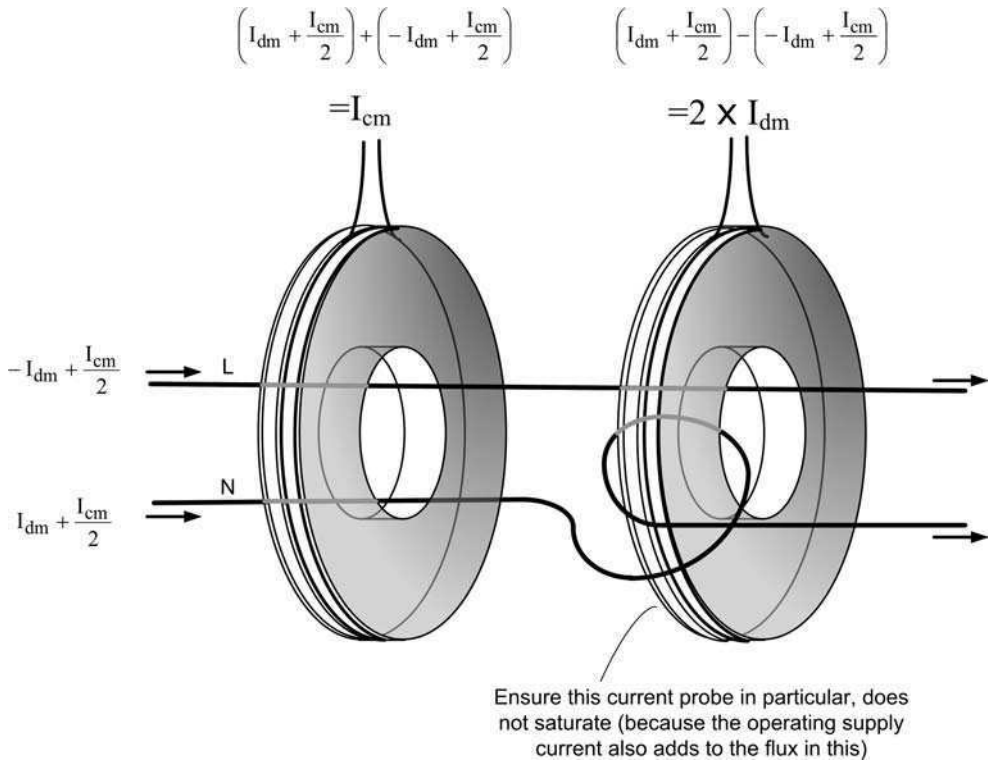


Figure 12-6: Using Two Current Probes to Separate DM and CM Noise Components

Electronic Measurements Ltd. at www.pemuk.com. These probes are not the usual current transformer type. The output from a Rogowski probe depends not on the instantaneous current enclosed, but on the *rate of change of current*. So, instead of just placing several turns around the wire to be sensed, as in a typical current transformer, the Rogowski probe effectively takes an air-cored solenoid and then bends that in a circle around the sensed wire (like a doughnut). Such probes are also considered virtually noninvasive. The usual lab *active* current probes (which also measure dc, and therefore include a Hall sensor) are usually just not suited for these high-bandwidth noise measurements.

Note: When viewing pulse transition times below 100 nanoseconds, or emission noise frequencies above a few megahertz, it is advisable to keep the cable length small. Thereafter, we must terminate the cable at the oscilloscope, or measuring instrument, with a 50 Ω resistor. However, most modern oscilloscopes incorporate a selectable 50 Ω input impedance. Correct termination of cables prevents standing wave effects. Note that, with this 50 Ω termination, *the measured voltage is approximately half of what it really is* because we essentially have a voltage divider formed by the cable and the terminating resistor. Oscilloscopes will usually automatically correct for this, if they 'know' that there is a 50 Ω termination present. Also note that fast-rising pulses can produce spurious ringing, due to high-frequency current crowding on the *surface of the cable shield*. This can be suppressed by threading the measurement cable through one or more ferrite beads (or toroids). For example, Pearson reports that they obtained good results by placing three turns through four ferrite cores of about one inch inside diameter, two inches outside diameter, and 1/2 inch thickness.

Note: A quick diagnostic test for understanding a particular high-frequency conducted EMI problem (measured by a LISN at the *inputs* of the power supply) is to twist the *output* cables of the power supply tightly together (along with their respective return wires). This induces *field cancellation* (also called flux *containment*), thus reducing the radiation from the output cables (if present).

This procedure was actually implemented in full production on a particular high volume commercial design. A few tie-wraps were used to hold the bunch of wires tightly together in the twisted position. This happened to be a last-ditch effort to avoid costly last-minute redesign, just before full production. This "twist-and-tie-wrap" technique is admittedly not very practicable or desirable, in production. But it is cheap. Note that a *ferrite sleeve* slipped over the entire output cable bunch was also working equally well, if not better. But it was disqualified simply because it was far more expensive than three tie-wraps! However, it is interesting to note here, that even though a ferrite sleeve may look like a radiation shield (and smell like one too! — and even produce almost the *same* results as *twisting* the cables produces, it actually works by *reducing the common mode noise currents themselves, not merely by "shielding" the EMI arising due to them*. Twisting, on the other hand, simply tries to cancel the fields of adjacent wires (with their returns). Looking back, in this particular case, the root cause was that there was obviously a significant amount of common mode noise already present *on the output*, which was causing the output cables to radiate. The radiation was thereafter being picked up by the input cables, leading to the failed conducted EMI test.

In *Figure 12-7*, we show a practical technique to separate that part of the conducted emissions spectrum that is due to radiation from within the converter. We see how to learn to

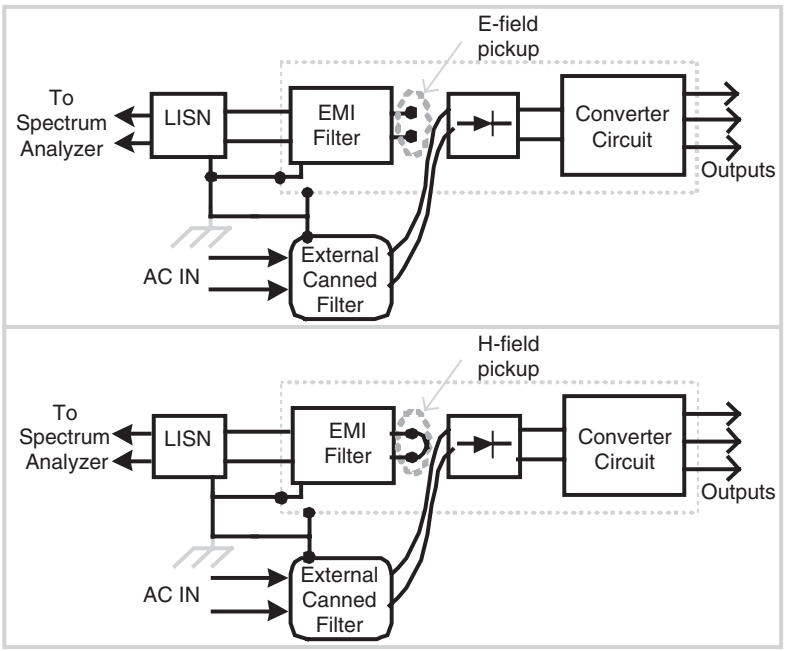


Figure 12-7: Analyzing Magnetic and Electric Field Sources Inside a Power Supply

identify these as E-fields or H-fields. For this, we need to cut the PCB traces just before the input bridge, and then route the ac power from a canned filter outside the enclosure. The ends of the existing filter are kept either open (to receive E-fields) or connected together through a small loop (for seeing the H-fields). The other end of this EMI filter is then routed as usual to the LISN and spectrum analyzer. We can thus see this "extraneous" (radiation-based) noise. It will give us an indication if a heatsink, for example, is causing severe E-fields, or if a certain magnetic component is causing severe H-fields. We can also wave a small plate of thick copper (connected to earth) in suspected areas to see which component may be the actual source of the fields. However, for analyzing the source of magnetic near-fields, a slab of ferrite (from a typical EMI suppression kit) works far better than a copper plate (this can be waved around similarly — but there is no need to earth it).

> **Caution:** AC power is **NOT** to be applied through the LISN. This will cause a serious hazard. Also the plate/slab must be well-wrapped in tape to prevent accidental contact with components.

Are We Going to Fail the Radiation Test?

Most of the smaller companies cannot afford a *precompliance setup* for *radiated* emission tests. However a few of them have a fairly good idea beforehand whether they are going to be successful in that test or not — just by looking hard at the *conducted* EMI scan. What they do is to look carefully at the spectrum in the third region of CISPR 22. This is the flat region from 5 to 30 MHz. They can even scan higher to higher frequencies, if possible. They realize that even though they may have achieved compliance with the conducted limits in this third region, it is *not* good enough! So, they look at the overall *shape of the plot* in this region. If they find that it is gradually rising toward the 30 MHz end, they are quite confident that they have a radiation problem. However, if the plot starts drooping, or remains generally flat as 30 MHz approaches, they are likely to submit the prototype immediately to a lab for the formal radiated limit compliance certification. In other words, we can actually "see" the energy level in the 5–30 MHz region. If there is an undue amount of conducted noise energy in this region, radiation can't be too far off either!

CHAPTER

13

Input Capacitor and Stability Considerations in EMI Filters

This Page Intentionally Left Blank

CHAPTER 13

Input Capacitor and Stability Considerations in EMI Filters

There are certain things we may do unintentionally at the input of the converter that can have a major impact on the performance of the EMI filter, and also the converter itself. If we don't know the rules of the game, we can end up saturating our filter chokes and even inducing instability.

Is the DM Choke Saturating?

Part of the designer's average "EMI troubleshooting day" may involve taking a core from the shelf and placing some turns on it. But the temptation of winding a few more turns (to increase the inductance) may just do it — the core could start saturating, rendering it less and less effective. But how could we have known that?

The simplest equation to check if a core is saturating or not is

$$B = \frac{LI}{NA_e} \text{ teslas}$$

where L is the measured inductance in H, N the number of turns, and A_e is the effective area of the core in m^2. Note that A_e is simply the normal geometric cross-sectional area of the core. If it is an E core, we would take the area of the center limb (or twice the area of each side limb, whichever is smaller, though usually, either way we get the same result). So if we plug in the peak current we can calculate the peak B-field in the choke. If we know the material of the core, we probably already know its *saturation flux density* B_{SAT}. Then we can easily check if the core is saturating or not. Note that B_{SAT} for powdered irons is usually around 10,000 gauss (1000 mT or 1 T) and for ferrites it is 3000 gauss (300 mT, or 0.3 T).

But what is the peak current? This is the term that is usually completely underestimated by most designers, and that is why they often don't realize that their DM choke is ineffective because it is saturating.

The temptation to wind a few more turns on the DM choke (if present of course) in an effort to raise its inductance may just produce enough *ampere-turns* to saturate it. What happens

after that depends a lot on the material itself. Powdered iron may be more forgiving up to a point, but then it also has lower initial permeability to start with. Ferrites can saturate comparatively more sharply (though we know that *air* helps soften that — either in the form of an air gap, or distributed, as in powdered iron). Of course if the DM choke is just the leakage inductance from the CM choke, then in effect it is a completely air-cored coil anyway, and we don't have to worry about it saturating.

So in DM chokes, the major concern is core saturation. But in CM chokes, underestimating the current will make it run excessively hot, due to higher copper losses.

In a typical off-line input stage, with no Power Factor Correction ('PFC') stage present, the input bridge conducts only for part of every ac half-cycle as seen from *Figure 13-1*. The input (bulk) capacitor discharges slowly, during the remaining time, at a rate that depends on the power it is delivering to the converter — which except for the small loss in the filter itself, is essentially equal to the input power, that is, $P_{IN} = P_{OUT}/\eta$, where P_{IN} is the power being drawn from the supply lines, for given output power P_{OUT} at a certain efficiency, η. If we put larger and larger bulk capacitance we can end up with extremely high peak and RMS currents through the input bridge and the filter chokes (along with a lot of low-frequency harmonic content that is regulated separately by line harmonic standards — i.e. PFC-related). The higher currents are attributable to the fact that the bridge *conduction time* becomes shorter as we increase the bulk capacitance. And since a certain amount of average input power is being constantly demanded by the converter, the current "bursts" through the diodes must increase in amplitude, so as to compensate for the shorter gating interval.

Knowledge of the input RMS current is necessary to correctly estimate the copper losses in both the CM and DM chokes, whereas knowledge of the peak current is necessary to correctly estimate the DM core volume (its energy handling capability). In a CM choke, the input ac line current effectively cancels out, so its saturation is usually not considered either a possibility or a concern. That means it is "chopped" out of a vertically offset sine wave of that time period.

The shape of the input current into the power supply is usually described as a 'haversine' — which is simply a sine waveform *offset on its vertical axis* so as to make the minima of the curve coincide with the horizontal axis ($t = 0$). The current waveform shown in *Figure 13-1* is thus a haversine — with a time period equal to the diode conduction time (during which it occurs). That means it is "chopped" out of a vertically offset sine wave of that time period.

The following equations can be derived.

The time for which each diode conducts is

$$t_C = \frac{\cos^{-1}[A]}{2 \times \pi \times f_{LINE}} \text{ seconds}$$

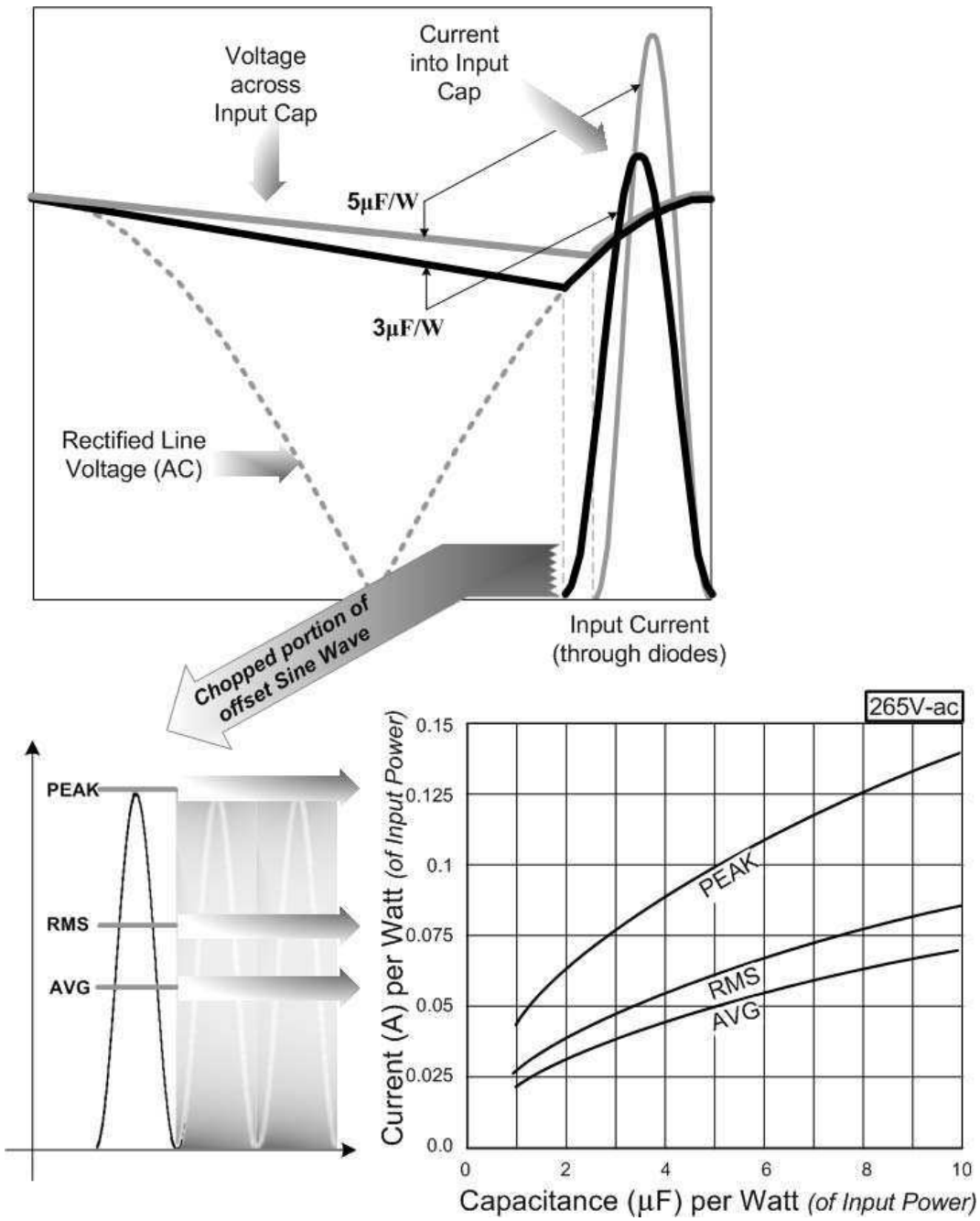


Figure 13-1: The Input Current Waveshape Analyzed

where

$$A = \sqrt{1 - \frac{P_{IN}}{2 \times C \times f_{LINE} \times V_{AC}^2}}$$

C is in Farads above.

The RMS and average values of the current waveform (*calculated only over the diode conduction time*) are, respectively,

$$I_{AVG} = \frac{\sqrt{2} \times \pi \times P_{IN}}{V_{AC} \times [1 + A] \times \cos^{-1}[A]} \text{ Amp}$$

and

$$I_{RMS} = I_{PEAK} \times \frac{\sqrt{1.5}}{2} = 0.612 \times I_{PEAK}$$

where the peak of the current is

$$I_{PEAK} = 2 \times I_{AVG}$$

We have to average the RMS and average values over the full ac cycle, as shown in the following example.

Example: *A power supply delivering 5 A@14 A at 70% efficiency has a 330 μF input capacitor. What are the RMS and peak input currents at 265 VAC/50 Hz?*

$$P_{IN} = \frac{5 \times 14}{0.7} = 100 \text{ W}$$

So A = 0.978, and $t_C = 0.67$ ms. We calculate $I_{AVG} = 4.05$ A, $I_{PEAK} = 8.1$ A, $I_{RMS} = 8.1 \times 0.612 = 5$ A.

We have also provided a graphical method in *Figure 13-1* — for any general case (at a line voltage of 265 V-ac). In our case, the selected input capacitance per (input) Watt is 330/100 = 3.3 μF/W. We locate this value on the horizontal axis, and then we can see that this gives us about 0.05 A for the RMS current on the vertical axis. But the vertical axis is the current *per Watt* of input power. So for our case, the RMS current is 100 × 0.05 = 5 A. This agrees with our numerical calculation above.

Since we have *two conduction intervals per ac cycle time period*, the input average and RMS currents, now calculated *over the whole cycle*, are (with $T = 1/f_{LINE}$)

$$I_{IN\_AVG} = 4.05 \times \frac{2 \times t_C}{T} = 0.27 \text{ A}$$

and

$$I_{IN\_RMS} = \left[5.03^2 \times \frac{2 \times t_C}{T}\right]^{1/2} = 1.3 \text{ A}$$

Therefore, we have to ensure that our DM choke does not saturate with a peak instantaneous current of 8 A. Both the CM and DM chokes must have copper thick enough to handle 1.3 A RMS.

If we are dealing with a wide-input (universal) off-line power supply, *we should check all the currents at low line too*! At low line, despite the fact that the conduction time increases significantly, the required average dc input current is much higher too. So the peak/RMS currents will generally tend to increase, not decrease, as we lower the line voltage. If we assume efficiency is unchanged in going from high line to low line (i.e. P_{IN} is a constant), then a detailed calculation based on the equations presented above reveals that the RMS, peak, and average currents *increase by almost 4% exactly,* in going from 265 V-ac to 90 V-ac (for 3 μF/W). Not much of an increase really! So *Figure 13-1* should suffice. Or we can do a detailed calculation.

Note that if we had PFC, then at 265 V-ac input, we would get an RMS current of only

$$I_{IN\_RMS} = \frac{P_{IN}}{V_{IN\_RMS}} = \frac{100}{265} = 0.38 \text{ A}$$

The peak is also less

$$I_{IN\_PEAK} = I_{IN\_RMS} \times \sqrt{2} = 0.534 \text{ A}$$

But note that, if this is a universal-input power supply, we again need to check the currents at low line too, and rate the chokes according to the worst case. In this case (with PFC), we see that at 90 V-ac, all the currents *increase almost three times* over their values at 265 V-ac (again assuming efficiency is unchanged in going from high line to low line).

However, it is nevertheless clear that *introducing power factor correction in off-line power supplies always leads to much smaller EMI filter chokes.*

Tip: In a power supply without PFC, if we are trying to measure the bridge conduction time with an oscilloscope, we should ensure that the input to the power supply is not from a variac. An electronic programmable ac source is preferable, because the variac tends to appear inductive, and keeps trying to push its residual energy through the bridge for just a little longer. This increases the conduction time, and makes the peak current significantly lower than if the power supply is powered off the wall.

Practical Line Filters in DC-DC Converter Modules

See *Figure 13-2* for an example of how EMI suppression techniques are applied to dc-dc converters. We have shown an industry standard *isolated* "brick" (along with its external EMI filtering). The input to this particular module is a coarsely regulated '–48 V-dc' or '–60 V-dc' bus, forming part of a distributed power architecture for a data/telecom network. Its output is isolated and regulated (e.g. 3.3 V/50 A or 12 V/10 A etc). The –48 V-dc input is usually derived from an off-line telecom power supply (called a "rectifier").

Notice how the traces are laid out in the module's external EMI filter as illustrated in *Figure 13-2*. Note, in particular, the placement of the Y-caps. We should also keep in mind

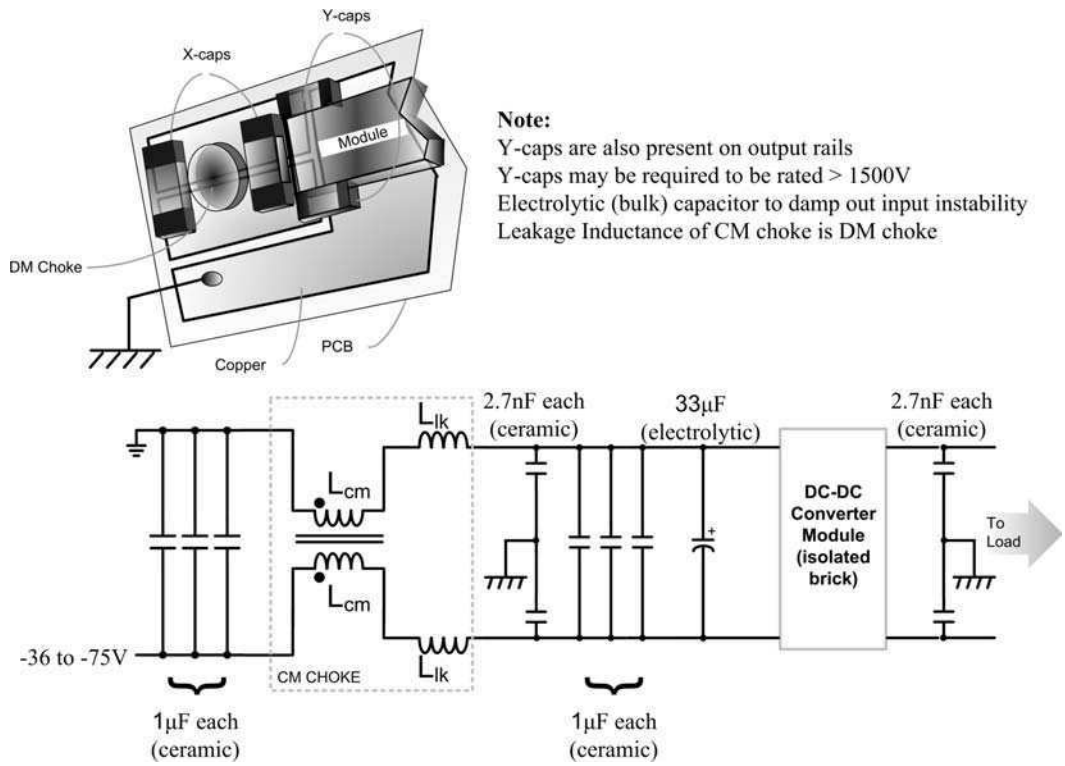


Figure 13-2: Typical EMI Filter for DC-DC Modules (bricks)

that one of the most effective methods of suppressing EMI, especially in board mounted dc-dc converters, is a good ground plane. On a multilayer board, best results are usually obtained by having this plane be the internal layer just below the top (power) component side. Up to 20 dB reduction in noise is possible. The "input instability" issue, as indicated in *Figure 13-2*, is discussed next.

Note: As per the usual safety regulations, voltages below 60 V-dc generally are not considered hazardous and therefore are not subject to the isolation/earthing requirements described earlier. But above 60 V we generally would need reinforced installation. However, in Europe, more recent definitions have evolved concerning 'Telecommunications Network Voltage' (TNV) circuits, which by their very nature are not accessible to humans in general. So for example, above 60 V-dc and below 120 V-dc we have what is called a 'TNV-2 circuit.' Though this requires insulation, it need not be of the reinforced/double insulation category. So in general, the Y-caps shown in *Figure 13-2* can be standard 2 kV rated components. However, if the preceding "rectifier" has reinforced isolation between its own output and the ac mains, then the Y-caps in the module can be just regular 100 V capacitors. Note also, that the Y-caps in a dc-dc converter are between two dc levels, so we can forget about ac ground leakage currents and their related safety issues as previously discussed. This gives us the flexibility of choosing large-capacitance Y-caps.

Note: For protection against ESD (electrostatic discharge) upsets, 0.01 μF caps between the *terminal block contacts* and earth are often also included. These are essentially Y-caps. But note that there have been cases, particularly when these caps were ordinary 50 V multilayer ceramic ("MLCs"), that they got destroyed during the course of an ESD test — simply because they got charged up to excessive voltages! Therefore, these capacitors, and any other Y-caps present, must be evaluated under such abnormal but likely disturbances too. Eventually, we may need to increase the capacitance and/or the voltage rating and/or size of the caps just to protect them from/against over-charging.

Since about 1971 the phenomena of 'input oscillations' or 'input instability' has received quite a lot of attention. It has been shown that instability can occur if the output impedance of the filter is not within a certain "safe" window, as related to the input impedance of the converter (we are talking about the impedances presented to the power flow now — not the CM or DM noise). So, with the modern trend of low-impedance "all-ceramic" solutions in dc-dc converters, the possibility of this particular type of instability is becoming more and more real.

One of the easiest ways to see the impact of the *negative input impedance* of a typical converter is to set it up with *only* ceramic input capacitors (about 10 μF or less) — and then do a "hard power-up." In this type of power-up test, the dV/dt of the applied input is kept intentionally very high. On the bench, this can be done by simply slamming the banana plug from the input of the converter into the output terminals of a (low-impedance/high-current) lab dc power supply. Then, if we monitor the input (supply) pin of the converter with a digital oscilloscope (triggered correctly, and in one-shot acquisition mode), we will see an initial overshoot — that can be as high as 1.5 to 2.5 times the supposed dc voltage level (as set on the lab supply). Note that if the input capacitance is large enough (beyond a certain value), the dV/dt (and overshoot) gets automatically reduced, due to the higher

charging current required for this input capacitor. On the other hand, if the ceramic capacitor is replaced by an aluminum electrolytic (even one with a lower capacitance), the overshoot is dramatically reduced. Tantalum capacitors also produce overshoots under hard power-up, but these are less pronounced than with ceramic.

> **Note:** We should remember that in any case, it's never a very good idea to use tantalums at the *input* of any converter — due to tantalum's inherent surge current limitations. However, if for some reason, tantalums must be used at the input (in any topology), or at the outputs (for a boost or buck-boost), we must ensure that they are 100% surge-tested by their vendors. And even for such surge-tested tantalum capacitors, it is recommended that the maximum voltage applied across them in our application be less than half their voltage rating — that is, a voltage derating of 50%.

We see that it is possible to damage a dc-dc converter, which uses only small-value ceramic capacitors at its input — more so when we already happen to be operating rather close to its maximum input voltage rating.

The designer should note that in *Figure 13-2*, we have placed an *electrolytic capacitor in parallel to the ceramic input capacitors* — for the purpose of damping out "input instability." This needs further explaining. To understand the underlying causes associated with this phenomena, we need to start with the well-known buck converter equations and see what happens if we (hypothetically) "jiggle" the duty cycle, just a little bit, around its steady state value. Note that in a practical situation, this could happen very easily under normal line or load transients. Therefore, expressing the input voltage and the input current as a function of duty cycle

$$V_{IN}(D) = \frac{V_O}{D}$$

$$I_{IN}(D) = I_O \times D$$

So

$$dV_{IN} = -\frac{V_O}{D^2}dD$$

and

$$dI_{IN} = I_O dD$$

Dividing these two equations we get (for a buck converter)

$$\frac{dV_{IN}}{dI_{IN}} = -\frac{V_O}{I_O \times D^2}$$

This is the *incremental resistance* at the input. Let us call this "R_{IN}." So for a buck converter, its incremental resistance in ohms is

$$R_{IN} = -\frac{R_L}{D^2}$$

Here R_L is the load resistance (ohms), and is assumed constant. Note that both the input voltage and the input current always have *positive* values in a (positive-to-positive) buck converter. Therefore the ratio V_{IN}/I_{IN} is also certainly a positive quantity. It's only their *relative change* that is in opposite directions — hence the minus sign in the preceding equation.

Another way to look at it is as follows. For a given output power P_O, if the input voltage *increases* slightly, then the input current must *decrease*. That's because P_O (and therefore roughly P_{IN} too) must remain constant.

$$V_O \times I_O = P_O \approx V_{IN} \times I_{IN}$$

This is clearly true for any topology. We can therefore work out the (negative) incremental input resistance for the other topologies in a similar manner.

Example: *What is the input resistance of a 3.3 V/50 V brick, with an input range of 36–75 V?*

Output power is $3.3 \times 50 = 165$ W. R_L is $3.3/50 = 0.066\ \Omega$. Duty cycle is 0.092 at 36 V input. So the magnitude of R_{IN} is $0.066 / (0.092)^2 = 7.8\ \Omega$. In terms of decibels this is $20\log(7.8) \cong 18$ dBΩ. A similar calculation at 75 V input gives 31 dBΩ. So the (magnitude of the negative incremental) impedance of the buck converter *falls as the input voltage falls*. We will see that this means that this form of *instability is more likely to occur at low input voltages*.

What is it about the interaction of the impedances at the filter-converter interface that causes this instability? Let us see what is really happening as we jiggle the input to the filter (V_{IN}). See *Figure 13-3*.

Here V_{INC} is the voltage that appears at the terminals of the converter. The filter impedance and the converter impedance form a voltage divider.

$$V_{INC} = V_{IN} \times \frac{Z_{INPUT}}{Z_{INPUT} + Z_{SOURCE}}$$

Looking at this equation, we ask — what happens if Z_{INPUT} becomes *negative*? Numbers explain it best. Suppose $Z_{INPUT} = -30\ \Omega$, and $Z_{SOURCE} = 10\ \Omega$. Then V_{INC}/V_{IN} is equal to

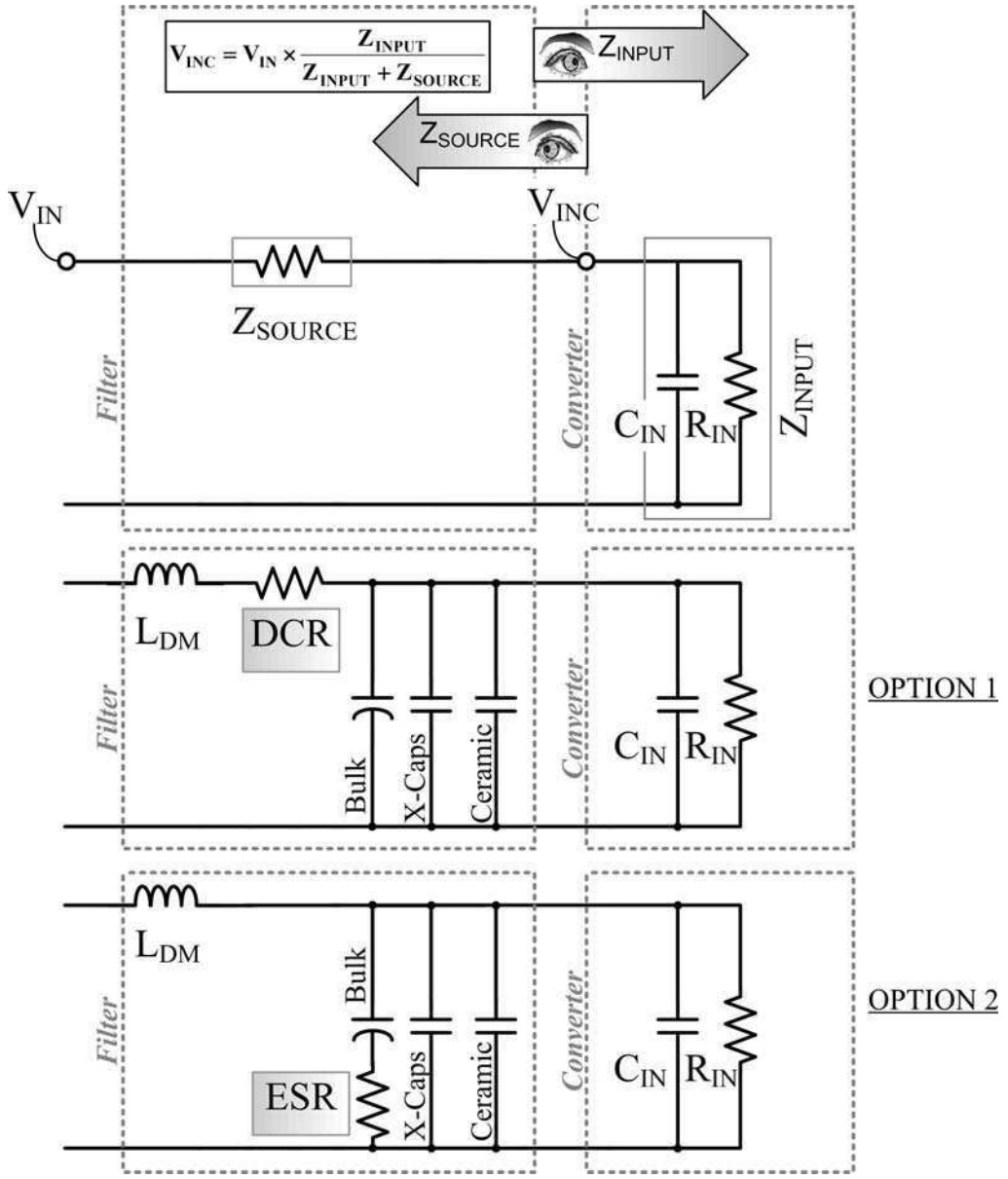


Figure 13-3: Input Interaction and Two Possible Solutions to Increase Damping

$-30/(-30 + 10) = +1.5$. This implies a *50% input overshoot*. However things get uglier. What if $Z_{SOURCE} = 40\ \Omega$ instead of 10 Ω? Now V_{INC}/V_{IN} is equal to $-30/(-30 + 40) = -3$. The sign looks odd! But let's not forget that we are actually talking of *incremental* impedances (the number crunching here is actually very simplified). So the sign of V_{INC} is not really negative, but its *change* is.

What is being indicated is that if we momentarily increase the applied input voltage (at the input of the filter), the input voltage at the converter end falls slightly (though momentarily). The control loop of the converter will however "think" that the input has *fallen* rather than increased, so it will respond incorrectly to the change. And isn't that the usual recipe for output oscillations?!

Therefore, the most basic criterion for avoiding this type of instability is

$$Z_{SOURCE} < |Z_{INPUT}|$$

Now, in reality, the input impedance of the converter is *frequency-dependent* (R_{IN} was just the low-frequency value of Z_{INPUT}). In the more detailed converter model, a parallel capacitance C_{IN} (see *Figure 13-3*) appears across the input of the converter, mainly due to the *output filter components of the converter being reflected into the input*. This causes the (real part of the) input impedance to be *less and less negative* as frequency increases. In *Figure 13-4*, we have shown a typical input impedance plot with respect to frequency. Note that only the magnitude of the converter impedance has been displayed, primarily because the y-axis is in log scale, and we know that log scales cannot be negative.

Z_{SOURCE} (the output impedance of the filter) is also changing with frequency. Looking into the output terminals of the filter (from converter) we see basically a simple parallel LC filter stage. Therefore Z_{SOURCE} has the shape indicated in *Figure 13-4*.

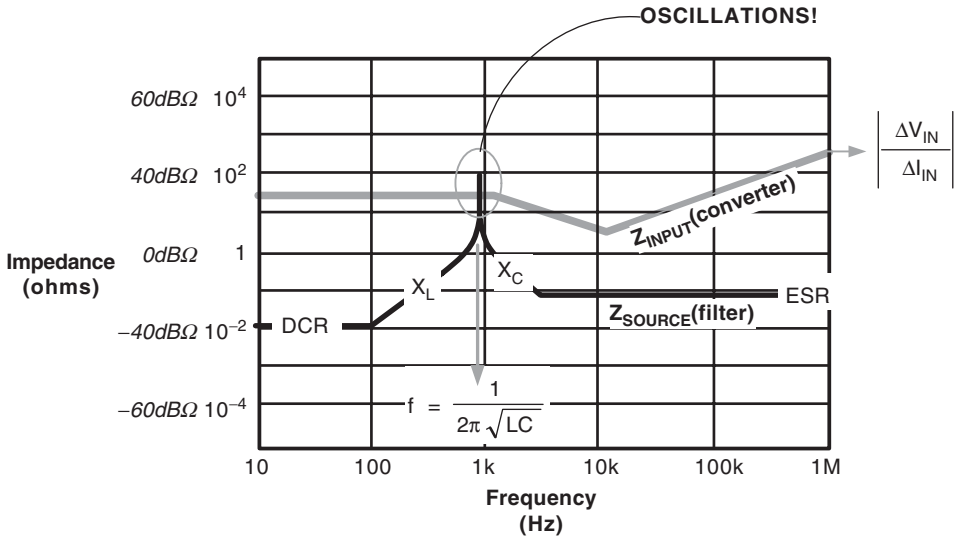


Figure 13-4: Input Filter Interaction and Stability Criteria

The stability criterion therefore means that we are demanding that the output impedance of the filter must be always less than the input impedance of the converter for any frequency. But what happens if the LC filter has *insufficient damping* and therefore has a resonance peak? This is the encircled problem area in *Figure 13-4*, and we can see that in this region we are violating the basic stability criterion. This peak needs to be suppressed. Therefore, in addition to the basic stability criterion, a follow-up criterion must be added to ensure that at resonant frequency, the LC filter peak is properly damped out.

For damping, we could simply add some more resistance to the choke (DCR) as shown in *Figure 13-3*. But that is not a very good idea since the entire operating current also passes through this choke, and the overall efficiency would suffer. Instead, it is preferable to add a slight resistance (ESR) to the capacitor as also shown in *Figure 13-3*. We know that any capacitor in steady state blocks any dc voltage completely. So the input capacitor sees only the ac component of the input current flowing into the switching mosfet. This therefore correspondingly reduces the dissipation required to achieve a given target of damping. However, we also need to maintain good decoupling at the input of the converter (to keep its control sections from getting affected, as also to suppress EMI). Therefore the usual commercially implemented solution for such bricks is to place an additional *high-ESR capacitor in parallel to the existing low-ESR decoupling capacitors*. It can been shown that we need to meet the following conditions to make the system unconditionally stable (the first of these is essentially the basic criterion discussed previously):

$$ESR < |R_{IN}|$$

$$ESR > \frac{L}{C_{BULK} \times |R_{IN}|}$$

$$C_{BULK} \gg C_{INPUT}$$

where C_{INPUT} is *total capacitance at the input terminals of the converter* (including C_{IN}, ceramic capacitors, any X-caps, supply decoupling caps, and so on). C_{IN} is typically a few μF, but without elaborate modeling of the converter, or some type of measurement, its value may be unknown to most designers. But generally speaking, if C_{BULK} is chosen to be much larger, it effectively "swamps" out the effect of C_{IN}, and so the system is stable anyway (the thumb rule is that C_{BULK} should be four to five times the total effective low-ESR input capacitance physically present at the input to the converter, that is, before C_{BULK} was added).

CHAPTER

14

The Math behind the Electromagnetic Puzzle

This Page Intentionally Left Blank

CHAPTER 14

The Math behind the Electromagnetic Puzzle

We now present the main mathematical analysis and tools needed to finally cap a successful filter design. First we should recall our hazy Fourier series class. Fourier analysis is often avoided by power supply engineers, but it can go a long way toward understanding and tackling several key issues like EMI/noise, transformer proximity losses, PFC (power factor correction), and so on.

Math Background — Fourier Series

Collecting some of the basic definitions first:

For a function f(x) with the time period expressed in terms of an angle (2π), we can write

$$f(x) = \frac{1}{2}a_o + \sum_{n=1}^{\infty}\left(a_n \cos nx + b_n \sin nx\right)$$

$$a_n = \frac{1}{\pi}\int_0^{2\pi} (f(x)\cos nx)dx$$

$$b_n = \frac{1}{\pi}\int_0^{2\pi} (f(x)\sin nx)dx$$

Alternatively,

$$f(x) = \frac{1}{2}a_o + \sum_{n=1}^{\infty}\left(c_n \cos(nx - \phi_n)\right)$$

$$c_n{}^2 = a_n{}^2 + b_n{}^2$$

$$\tan \phi_n = \frac{b_n}{a_n}$$

Alternatively,

$$f(x) = \sum_{n=-\infty}^{\infty} d_n e^{jnx}$$

$$d_n = \frac{a_n - jb_n}{2} \quad (n \geq 0)$$

$$d_n = \frac{a_n + jb_n}{2} \quad (n < 0)$$

For a function f(t) with a time period 'T' (units of time):

$$f(t) = \frac{1}{2}a_o + \sum_{n=1}^{\infty}\left(a_n \cos\left[\frac{2\pi nt}{T}\right] + b_n \sin\left[\frac{2\pi nt}{T}\right]\right)$$

$$a_n = \frac{2}{T}\int_0^T f(t)\cos\left[\frac{2\pi nt}{T}\right]dt$$

$$b_n = \frac{2}{T}\int_0^T f(t)\sin\left[\frac{2\pi nt}{T}\right]dt$$

In most math school books, the time period is expressed as 2π. However, in power supplies we know that the period we are interested in is in units of *time* not angle, that is, $T = 1/f$. The normal way to convert angle θ to t is to use the equivalence $\boldsymbol{\theta/2\pi \rightarrow t/T}$**, that is,** $\boldsymbol{\theta \rightarrow 2\pi t/T}$.

Note: The designer should not get confused by the fact that the first term in the expansion is sometimes called "$a_o/2$," sometimes "a_o," or sometimes something else altogether. Either way, in any Fourier expansion of any arbitrary periodic function, the first term is always the area under the waveform calculated over one time period (i.e. its arithmetic average).

Note: If the waveform is "moved" up or down, or side to side, on the graph, *while keeping its basic shape unchanged*, the appearance of the Fourier series may seem to change drastically. However, the ***magnitude*** *of the amplitude of any harmonic* (i.e. "$|c_n|$") remains the same (except for the average term which we don't really care about anyway). We should remember that *only* the magnitudes of the harmonics (the $|c_n|$'s) relate to *measurable* physical effects. And these can change only if the peak-to-peak value of the waveform changes (or of course its basic shape). Otherwise not!

Note: We can first calculate the c_n for the simple case of a waveform of *unity height* (alternatively its peak-to-peak value). Then the c_n for a waveform with a height of "A" (alternatively of peak-to-peak value "A") will simply be A times the previous c_n.

The Rectangular Wave

Let us find the series for the unity height rectangular wave shown in *Figure 14-1*. Textbooks provide the following expansion:

$$f(x) = \frac{\alpha}{\pi} - \frac{2}{\pi}\left(\frac{\sin\alpha \times \cos x}{1} - \frac{\sin 2\alpha \times \cos 2x}{2} + \frac{\sin 3\alpha \times \cos 3x}{3} - \ldots\right)$$

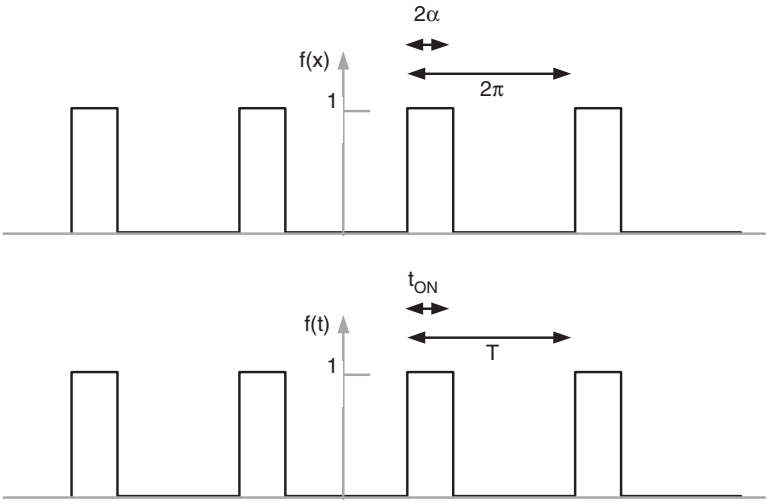


Figure 14-1: A Rectangular Waveform in Radians or in Time

Note that this is in terms of *angles*. Firstly, to avoid confusing the symbol "x" with distance (distance being also often called "x"), it helps to first change it to a more familiar symbol for angle, like "θ." We can then also use the summation sign to write the above expansion in a more compact form:

$$f(\theta) = \text{const} + \sum_{n=1}^{\infty} c_n \cos(n\theta)$$

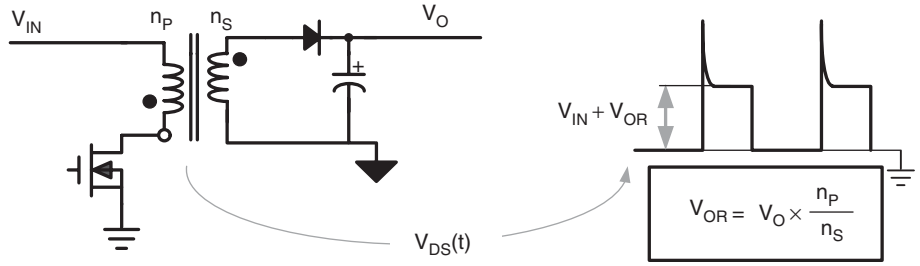
where

$$|c_n| = \left| \frac{2\sin(n\alpha)}{n\pi} \right|$$

Then, applying our conversion rule to go from angles to time (time being called the "t-plane" here), we get

$$f(t) = \text{const} + \sum_{n=1}^{\infty} c_n \times \cos\left[n \times \frac{2\pi \times t}{T} \right]$$

The c_n in this equation can also be explicitly written out, using the same conversion rule (from angle to time). We must remember though, that 2α in the θ-plane corresponds



$$V_{DS}(t) = (V_{IN} + V_{OR}) \times \left[D + \sum_{n=1}^{\infty} \frac{2 \cdot \sin(n\pi D)}{n\pi} \cos\left(\frac{2\pi nt}{T} \right) \right]$$

OR (changing the phase)

$$V_{DS}(t) = (V_{IN} + V_{OR}) \times \left[D + \sum_{n=1}^{\infty} (-1)^n \frac{2 \cdot \sin(n\pi D)}{n\pi} \cos\left(\frac{2\pi nt}{T} \right) \right]$$

Figure 14-2: Fourier Series of Drain Voltage Waveform of a Flyback

to $t_{ON} = D \times T$ in the t-plane (i.e. the switch on-time in a power supply). So we get

$$|c_n| = \left| \frac{2 \times \sin(n\alpha)}{n \times \pi} \right| = \left| \frac{2 \times \sin\left[\frac{n}{2} \times 2\alpha \right]}{n \times \pi} \right| \rightarrow \left| \frac{2 \times \sin\left[\frac{n}{2} \times \left(2\pi \times \frac{t_{ON}}{T} \right) \right]}{n \times \pi} \right|$$

Finally in the t-plane,

$$|c_n| = \left| \frac{2 \times \sin(n \times \pi \times D)}{n \times \pi} \right| \quad \textit{(unity height, rectangular wave, Fourier coefficients)}$$

As an example, let us apply this to the flyback in *Figure 14-2*.

We have provided the full expansion of the voltage on the drain of the mosfet (ignoring the leakage spike and possible ringing).

Note that we now have multiplied all the unity height c_n's by the actual wave amplitude (since we know that all the coefficients scale accordingly). Also, the first term is the average value of the waveform. Note that we could also have written this expansion with alternating signs as follows:

$$V_{DS}(t) = [(V_{IN} + V_{OR}) \times D] + (V_{IN} + V_{OR}) \times \sum_{n=1}^{\infty} \left[\frac{2 \times \sin(n \times \pi \times D)}{n \times \pi} \right] \times \cos\left[\frac{2\pi \times n \times t}{T} \right] \times (-1)^n$$

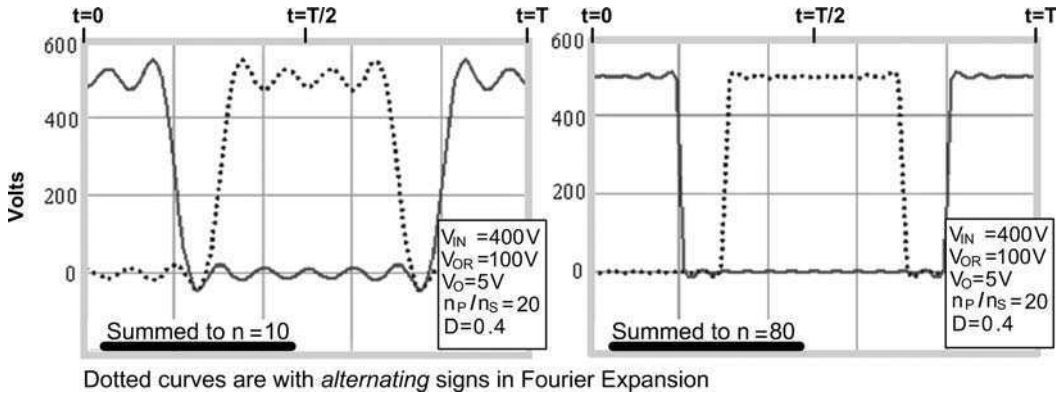


Figure 14-3: Reconstructing the V_{DS} of the Flyback from Its Harmonics

In *Figure 14-3* we have plotted both these expansions out (i.e. with same signs or alternating signs), so that the designer can see a) how including more harmonics helps to reconstruct the original waveform better, and b) the effect of the alternating signs (dotted curve) is just to shift the phase of the entire waveform.

Note: When using a math "crib-sheet" to help in an expansion for a certain "power supply waveform," we can avoid mistakes if we first scale the math function given in the book to unity peak-to-peak value (or unity height as desired). Then we can change its form of expression, from angles to time as discussed. Thereafter, we can multiply all the c_n's by the actual peak-to-peak value (or height) of our power supply waveform. In doing so, the dc value (first term) of the Fourier series (in the crib-sheet) can be completely ignored, as it is usually irrelevant. But if required, it can always be added on later, by examining our particular "power supply waveform" and calculating its arithmetic average, as we normally do (and adding it to the Fourier series thus available).

Analysis of the Rectangular Wave

We are actually interested only in the coefficients c_n, not in the Fourier series per se. We see that the coefficients have the form *sin(x)/x.* Plotting them out in *Figure 14-4*, first with linear scaling, and then with log scaling, we make the following key observations

- For the rectangular wave, the c_n's have the form sin x/x.
- For EMI suppression it doesn't matter if, say, the odd harmonics are present, or the even, or both. We are concerned only with the *envelope* of the emissions, since that is what we need to design the filter for (and to keep below the EMI limit lines).
- Therefore, on the log plot we can see that until x = 1 the function sin(x)/x is flat. After that it rolls-off as x increases. The envelope (with log scaling) falls at the rate of *20 dB/decade* — we know that −20 dB is one-tenth, and a decade is 10 times.

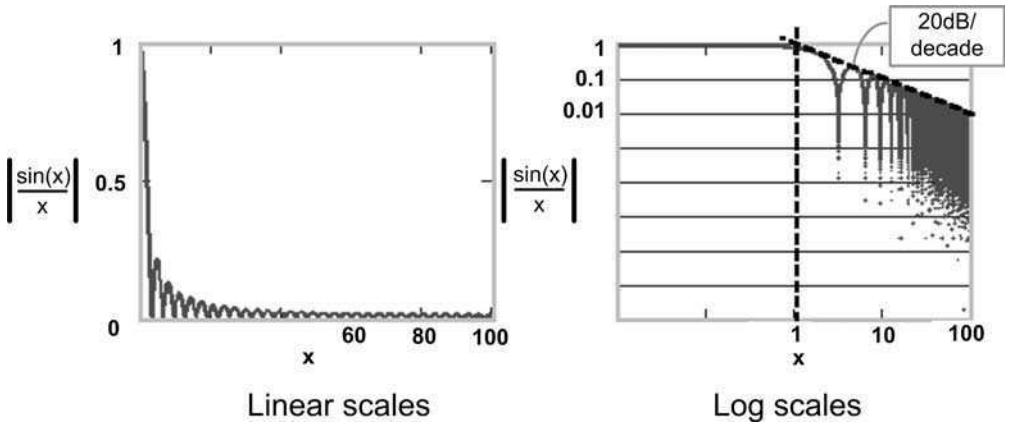


Figure 14-4: $|c_n|$ Plotted out for a Rectangular Waveform

The Trapezoid

Now we take the rectangular wave discussed above and make it a little more realistic by introducing nonzero rise and fall times. By a similar procedure as for a rectangular wave, we can get the following equation (for the case of equal rise and fall times):

$$c_n = A \times \frac{2 \times (t_{ON})}{T} \times \left[\frac{\sin\left\{ \frac{n \times \pi \times t_R}{T} \right\}}{\frac{n \times \pi \times t_R}{T}} \right] \times \left[\frac{\sin\left\{ \frac{n \times \pi \times (t_{ON})}{T} \right\}}{\frac{n \times \pi \times (t_{ON})}{T}} \right]$$

where $t_{RISE} = t_{FALL} = t_R$, and A is the amplitude (peak-to-peak). We are again ignoring any signs as being essentially irrelevant.

Clearly we will get *two* break points now. The first break point occurs at

$$\frac{n \times \pi \times t_{ON}}{T} = 1$$

that is,

$$n_1 = \frac{T}{\pi \times t_{ON}}$$

Since n = frequency of harmonic/fundamental frequency, that is, $n = f \times T$, we get the corresponding break frequency to be

$$f_{BREAK\_1} = \frac{1}{\pi \times t_{ON}} = \frac{0.32}{t_{ON}} \quad \textit{(first break frequency)}$$

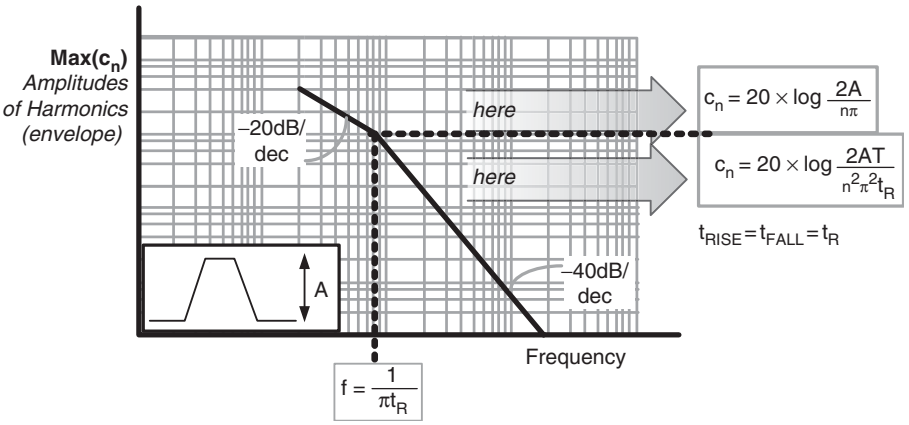


Figure 14-5: Envelope of Amplitudes of Harmonic of a Trapezoid

The second break point is at

$$n_2 = \frac{T}{\pi \times t_R}$$

that is,

$$f_{BREAK\_2} = \frac{0.32}{t_R} \quad \textit{(second break frequency)}$$

We know what to expect too — that *after the second break point, the net roll-off will be at 20 + 20 = 40 dB/decade*. See *Figure 14-5*.

Note that *n* must be an integer to have any physical meaning. The first breakpoint, therefore, may not be even visibly apparent. What we will perceive is that the envelope ramps down almost from the lowest frequency, at the rate of 20 dB/decade.

We can ask — when does n_1 become higher than n = 2? We can solve to get the condition

$$\frac{T}{\pi \times t_{ON}} > 2$$

that is,

$$D = \frac{t_{ON}}{T} < \frac{1}{2\pi} = 16\%$$

We can ask — when does n_1 get to be between n = 1 and n = 2? Solving

$$\frac{T}{\pi \times t_{ON}} < 1$$

that is,

$$D = \frac{t_{ON}}{T} > \frac{1}{\pi} = 32\%$$

In other words, only for very narrow duty cycles we might get to "see" the first break point. Further, below the first break frequency, the envelope of the harmonics becomes "flat." Theoretically, the first break point should be calculated and the envelope should be made flat below this frequency. But other than that, we can use the following equations to describe all the c_n (note that in these equations, the c_n are no longer the actual coefficients of the Fourier expansion, rather they represent the *envelope*):

$$c_n = 20\log\left(\frac{2A}{n\pi}\right)$$

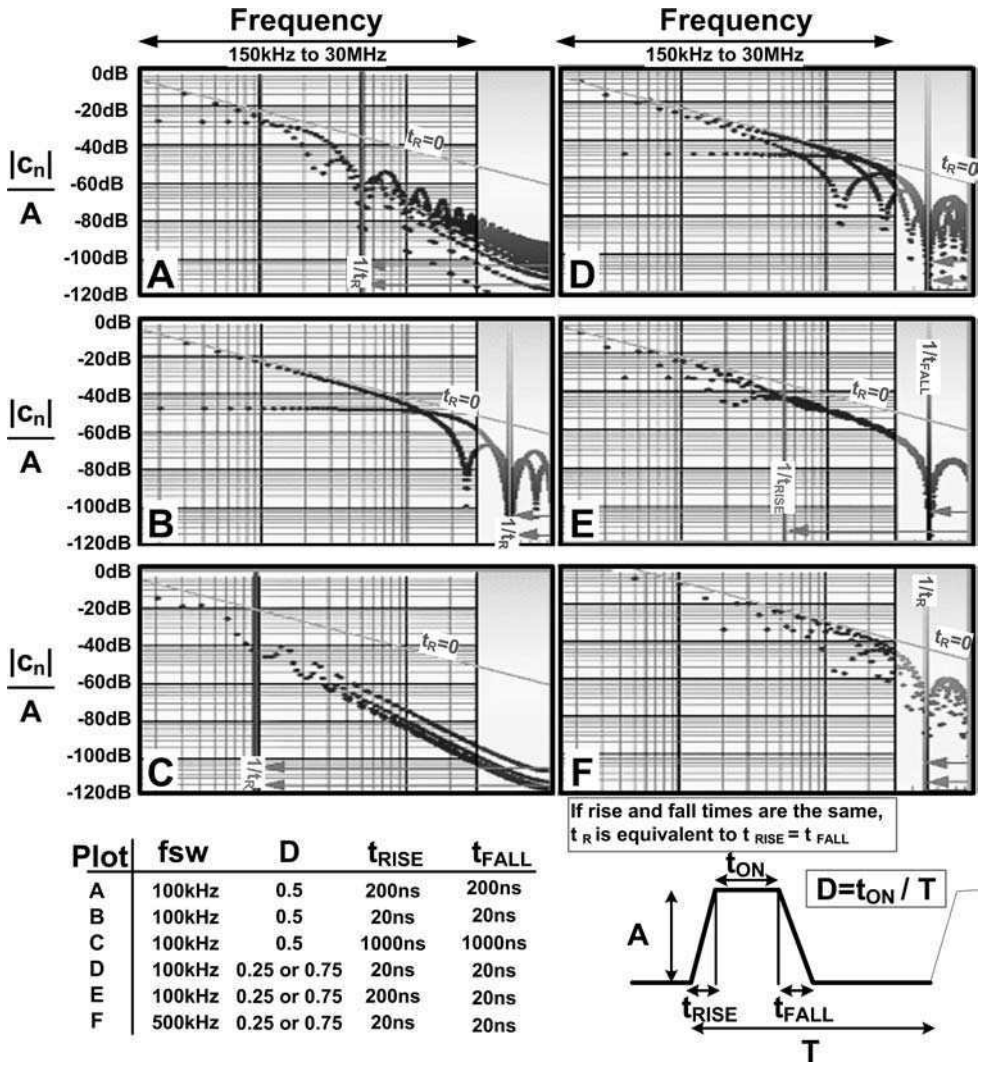
$$c_n = 20\log\left(\frac{2A}{n^2\pi^2 t_R f_{SW}}\right)$$

The first of the two equations above is valid between the *first and second* break points, and the second equation is valid for all frequencies *higher than the second* break point. Note that the switching frequency is $f_{SW} = 1/T$. The trapezoid rise and fall times are shown in *Figure 14-6*. Note how t_{ON} has been defined here.

The EMI from a Trapezoid

Note that the trapezoid is similar to the current in the mosfet (assuming a very high inductance, i.e. the 'flat-top approximation'). This current, we know, determines the DM noise. It is therefore helpful to see from *Figure 14-6* what a trapezoidal spectrum looks like for different duty cycles, frequencies, rise and fall times, and so on. In all cases, we have also provided in the form of a gray line, the extreme case, that is, with zero rise and fall times (the rectangular wave). We know that this gray line ($t_R = 0$) always falls at 20 dB/decade. Note that each curve is displayed in the window of concern from the point of view of *conducted* EMI limits, that is, from 150 kHz to 30 MHz. Note that each point in a spectrum is an actual mathematically generated harmonic amplitude. But as mentioned, ultimately we are only concerned with the *envelope* of these clusters of points. One observation to make is that if t_{RISE} is very small but t_{FALL} is large, and vice versa, the envelope falls with a slope somewhere between 20 dB/decade and 40 dB/decade.

Note: Some of the harmonic envelopes in *Figure 14-6* show a slight rise at the high-frequency end. But this is only an artifact of the computational resolution used in the program (for improving the speed of calculations). Therefore, no meaning should be ascribed to it.



Plot	fsw	D	t_{RISE}	t_{FALL}
A	100kHz	0.5	200ns	200ns
B	100kHz	0.5	20ns	20ns
C	100kHz	0.5	1000ns	1000ns
D	100kHz	0.25 or 0.75	20ns	20ns
E	100kHz	0.25 or 0.75	200ns	20ns
F	500kHz	0.25 or 0.75	20ns	20ns

Figure 14-6: Amplitudes of Fourier Harmonics for Various Switching Voltage Waveforms

The Road to Cost-effective Filter Design

There is no use designing a filter if we don't know how far we really need to go to achieve compliance. So it is a good idea to take stock of where we are at this point.

- If the converter has a switching frequency of f_{SW}, the harmonics are f_{SW}, $2f_{SW}$, $3f_{SW}$, and so on. The harmonics tend to have lower and lower amplitudes than the

fundamental frequency (first harmonic). But for simplicity, let us for now ***assume that the amplitudes of these harmonics are flat with respect to frequency***.

- We normally always start by trying to achieve compliance at the lowest frequency, that is, f_{SW}. But if the switching frequency drops below 500 kHz, certain factors start working starkly in our favor, and some against. We will discover that soon.
- A question we can ask is — should we try to use a complicated multistage filter — to achieve, say, an extreme attenuation of 60 or 80 dB/decade — or should we stick to the usual filters, which provide only about 40 dB/decade?
- We can also ask — if the switching frequency is less than 150 kHz, how low should we try to keep the EMI level at the fundamental frequency (since it is out of the range of the EMI limits anyway)?
- These questions can be answered only if we are made aware of what factors work in our favor and which don't, and what is the overall effect of all that on the techniques we need to apply to ensure compliance. In fact, if we design the filter with only the information presented so far, we will probably end up with an over-designed filter. We therefore need to see the interplay of the "forces arraigned against us," especially in the low-frequency region (150 kHz to 500 kHz).
- We observe that the conducted emissions limit lines (CISPR 22) actually allow for progressively *higher emissions* on our part — below 500 kHz.
- But in addition, the sensitivity of the standard LISN also decreases as the frequency falls off. This effectively allows us *more noise* too. Roughly, the LISN impedance falls from 50 Ω at about 500 kHz to about 5 Ω at very low frequencies — at an approximate rate of *10 db/decade below 500 kHz*. See *Figure 14-7*.
- However, we note that unfortunately, our EMI filter becomes less effective at low frequencies — being a low-pass stage (typical attenuation of 40 dB/decade above its resonant frequency).

Let us see what all this nets us. Suppose by suitable design we have achieved compliance at the lowest frequency. So, if the switching frequency is *less than 150 kHz*, that would mean that we have about 2 mV (66 dBμV) of noise emissions at 150 kHz (see *Table 9-1*). Now let us go in the *reverse direction*; that is, from low frequency to high frequency. This is what happens (see *Figure 14-8*):

1. The LISN sensitivity increases (at the rate of ~10 dB/decade). So we would start getting higher and higher noise readings.
2. But the EMI filter starts becoming more and more effective, attenuating the signal at a typical rate of 40 dB/decade.

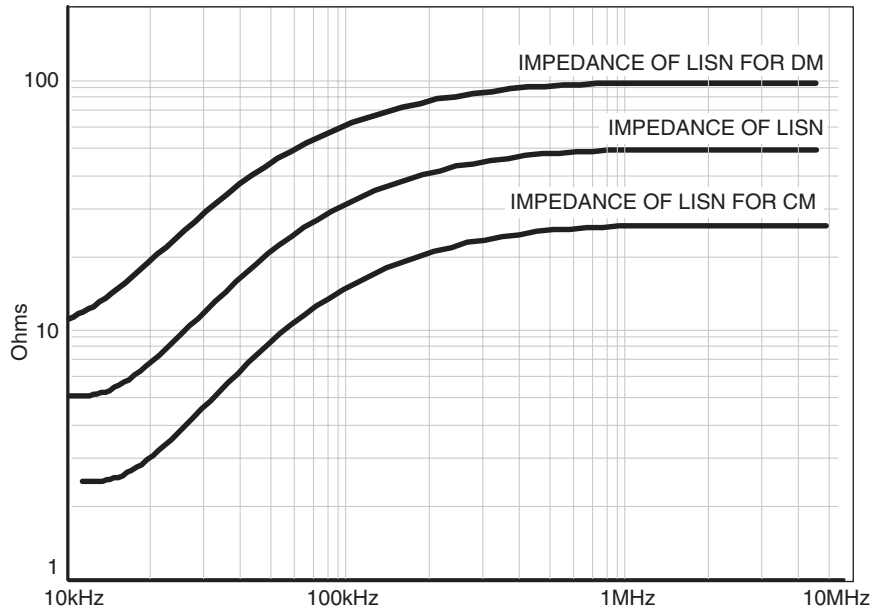


Figure 14-7: LISN Impedance at Low Frequencies

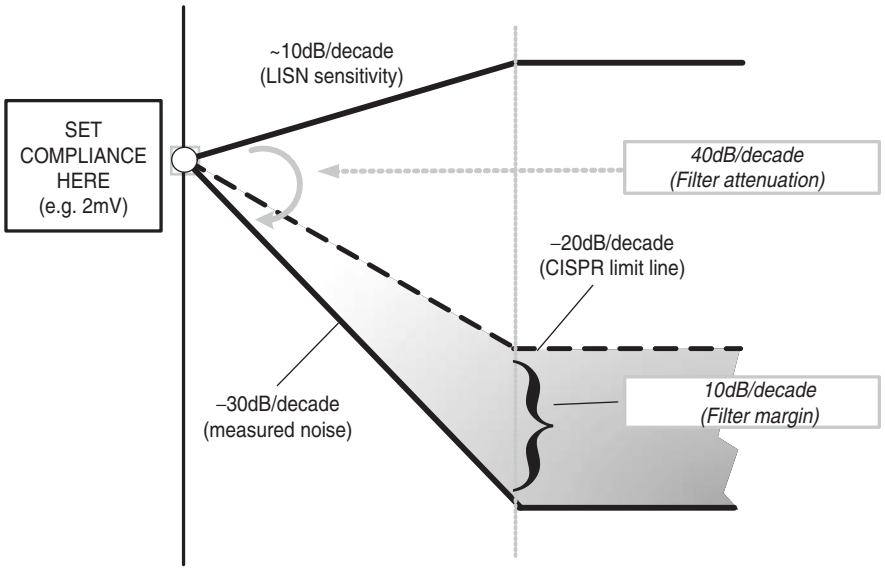


Figure 14-8: How Decreasing Low-frequency LISN Sensitivity and Increasing CISPR 22 Limit Lines "Help" Achieve Automatic "Headroom" in Noise Measurements

3. This swamps out the increasing LISN sensitivity, and so our measured noise actually *falls* at the rate of $40 - 10 = 30$ dB/decade. (assuming a flat emission spectrum).
4. But the limit lines are asking for us to decrease the noise level at the rate of only 20 dB/decade.
5. Therefore the measured noise level continues to fall below the limit lines with an increasing headroom of $30 - 20 = 10$ dB/decade.

That is why we need to achieve compliance at the lowest frequency first.

In reality, there will be some additional spikes in the EMI scan, due to parasitics we didn't model. But we should deal with them individually, at the *board level*, rather than try to bring the entire EMI spectrum down by a brute-force over-designed filter. It is therefore important to be aware of the "trends" as described above. However, in the solved examples that will follow, we will actually be ignoring the change in the LISN sensitivity, mainly for simplicity. In effect we are just assuming that the fall in LISN sensitivity provides us some additional headroom for unexpected spikes.

Practical DM Filter Design

In *Table 14-1* we have provided the height of the switch current trapezoid for all the main topologies (with the flat-top approximation).

The voltage across the ESR of the input bulk capacitor is therefore

$$v = ESR \times I_{SW} \text{ Volts}$$

If there was no filter present, the switching noise current received by the LISN would be

$$I_{LISN} = \frac{v}{100} = \frac{ESR \times I_{SW}}{100} \text{ Amps}$$

(since the LISN has an impedance of 100 Ω for DM noise).

Table 14-1: Switch currents (center of ramp) for different topologies

Topology	I_{SW} *(switch current)*
Buck	I_O
Forward	$I_O \times (n_S/n_P)$
Buck-Boost	$I_O/(1 - D)$
Flyback	$I_O \times (n_S/n_P)/(1 - D)$
Boost	$I_O/(1 - D)$

However the analyzer measures the noise across *one* of the two effective series 50 Ω resistors in the LISN. So the measured level of noise is

$$V_{LISN\_DM\_NOFILTER} = I_{LISN} \times 50 = \frac{ESR \times I_{SW}}{2} \text{ Volts}$$

We have assumed that C_{BULK} is very large, and that it has no ESL, and also that its ESR is much less than 100 Ω.

Example: *What is the DM noise spectrum measured at the LISN for a 5 V@15 A flyback at an input of 265 VAC, with a transformer turns ratio of 20? We are using an aluminum electrolytic bulk capacitor whose datasheet states that it has a capacitance of 270 μF, a dissipation factor (tangent of loss angle) of tan δ = 0.15 as measured at 120 Hz, and a frequency multiplier factor of 1.5 at 100 kHz.*

First the ESR is to be computed at the 120 Hz test frequency. By definition

$$ESR_{120} = \frac{\tan\delta}{2\pi f \times C} = \frac{0.15 \times 10^6}{2 \times 3.142 \times 120 \times 270} = 0.74 \text{ ohms}$$

At a high frequency, the ripple current is allowed to increase by the frequency multiplicative factor of 1.5. Therefore, since the heating ($I^2 \times ESR$) must still be the same, it means that the ESR at a high frequency must be $1/(1.5)^2$ times the ESR at low frequency. Therefore for our purpose,

$$ESR = \frac{1}{1.5^2} \times 0.74 = 0.33 \text{ ohms}$$

And so

$$V_{LISN\_DM\_NOFILTER} = \frac{ESR \times I_{SW}}{2} = 0.17 \times {}^{n_s}/_{n_p} \times I_O = 0.13 \text{ Volts}$$

This is the *amplitude* of the measured signal. It is the "A" in the corresponding Fourier series. In terms of dBμV its value is $20 \times \log(0.13/10^{-6}) = 102$ dBμV. We can thus predict that the spectrum is as shown in *Figure 14-9* (curve marked "DM Noise").

In *Figure 14-9*, we have also shown how the measured spectrum relates to the CISPR 22 Class B quasi-peak emission limits (bold line). We are also showing the case when the *switching frequency is just below the point where the CISPR 22 limits start*. For example, let us assume that this switching frequency is 100 kHz. For simplicity, *we are also assuming that the common mode noise is* ***not*** *a major contributor at these low frequencies*. In practice we should allocate some dB margin for the CM noise (as calculated a little later). We are

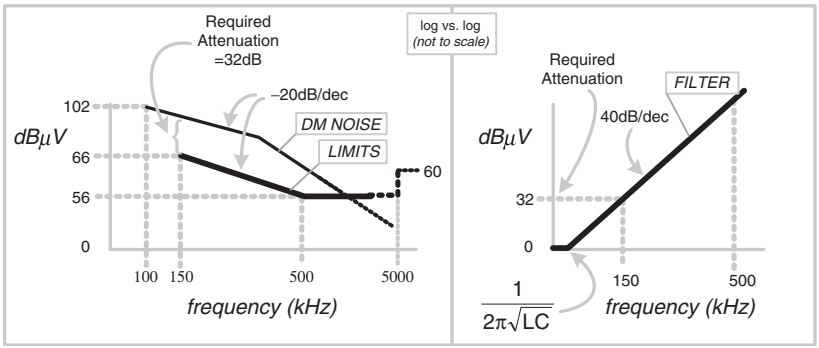


Figure 14-9: DM filter Calculation (see solved example)

also not accounting for any *other* required production margins or headroom required here. Also, as mentioned previously, since we are comparing peak values with *quasi*-peak limits, we automatically get some valuable headroom/margin.

Since we have 102 dBμV at the switching frequency of 100 kHz, the equation of the envelope is (straight line on a log vs. log plot)

$$\mathrm{dB\mu V(f)} = -\mathrm{slope}\,[\{\log(\mathrm{f}) - \log(100)\}] + 102$$

Since the (magnitude of the) slope is 20 dB/decade, at 150 kHz we get

$$\mathrm{dB\mu V(f)} = -20\,[\{\log(150) - \log(100)\}] + 102 = 98\ \mathrm{dB}$$

This is $98 - 66 = 32$ dB higher than allowed. And that means that we need to use a filter to attenuate the noise. We *need to pick a low-pass LC filter which provides an attenuation of 32 dB at 150 kHz*. Knowing this, we can calculate its break frequency. For example if we are using an LC low-pass filter, it has an attenuation characteristic of about 40 dB/decade *above* its break frequency (i.e. $1/2\pi\sqrt{(\mathrm{LC})}$). So from *Figure 14-9* we can see that its equation is

$$32 = \mathrm{slope} \times [\{\log(\mathrm{f}) - \log(\mathrm{f_{BREAK}})\}] = 40 \times [\{\log(150) - \log(\mathrm{f_{BREAK}})\}]$$

Therefore,

$$\log(\mathrm{f}) = \log(150) - \frac{32}{40} = 1.38$$

$$\mathrm{f} = 10^{1.38} = 24\ \mathrm{kHz}$$

We therefore need a filter that has an LC of

$$LC = \left(\frac{1}{2\pi \times 24000}\right)^2 = 4.4 \times 10^{-11}\ \text{sec}^2$$

Therefore if C is, say, 0.22 μF, we get L = 200 μH.

Note that if this is a standard two-stage off-line filter, we need to look at the DM equivalent circuit presented in *Figure 10-1* in *Chapter 10* to realize what exact values of the various X-capacitors in the figure this "C" really corresponds to in that schematic.

> **Note:** We can ask — since the break point associated with the rise and fall times didn't enter the picture here, does that mean that it doesn't matter how fast we turn-on and turn-off the mosfet? Yes from the DM noise viewpoint it *really doesn't matter much*. However there are parasitics that we have ignored (chiefly the ESL and trace inductances). And since, unlike the ESR, these will produce *frequency-dependent voltage spikes*, it is in our interest not to keep the mosfet crossover (transition) times too small.

Practical CM Filter Design

There are two ways of going about this. Either way we are assuming that the mosfet heatsink is tied to the chassis. At one end of the parasitic capacitor C_p we are applying the trapezoid that best describes a typical drain waveform. This causes a CM noise current I_{cm} to flow through the earth wire. We assume that the parasitic capacitances and/or X-caps cause this current to split up equally between the L and N wires. So we have $I_{cm}/2$ in each of these two wires.

First Method: (Quick)

We already know the Fourier components of the trapezoid, and so we can consider them individually and determine the injected current *due to each harmonic*. As an example, we take the forward converter. Here the peak to peak amplitude of the drain-to-source waveform ("Vmax" or "A") is twice the supply rail (V_{IN}). From Page 424, we have

$$c_n = A \times \frac{2 \times (t_{ON})}{T} \times \left[\frac{\sin\left\{\frac{n \times \pi \times t_R}{T}\right\}}{\frac{n \times \pi \times t_R}{T}}\right] \times \left[\frac{\sin\left\{\frac{n \times \pi \times (t_{ON})}{T}\right\}}{\frac{n \times \pi \times (t_{ON})}{T}}\right]$$

Since $\sin x/x \sim 1$ if x is very small, we get

$$c_n \approx 2A \times \left[\frac{\sin\left\{\frac{n \times \pi \times (t_{ON})}{T}\right\}}{n \times \pi}\right]$$

So assuming duty cycle is about 50%, the amplitude of the fundamental (first harmonic) is

$$c_1 \cong \frac{2A}{\pi} = \frac{4 \times V_{IN}}{\pi} \text{ Volts}$$

We realize that, as for the DM noise calculation, from the viewpoint of the noise envelope and its required attenuation, only the fundamental harmonic really counts. The current caused by this is

$$I_{cm} = \frac{V_{DS}}{25 - j\frac{T}{2\pi \times C_p}}$$

(since the LISN presents an impedance of 25 Ω to I_{cm}). But the voltage measured across the 50 Ω resistor of the LISN is due to $I_{cm}/2$ flowing through it. So

$$V_{cm} = \frac{I_{cm}}{2} \times 50 = I_{cm} \times 25 \text{ Volts}$$

Simplifying,

$$V_{cm} = \frac{4 \times V_{IN}}{\pi - \frac{j}{50 \times C_p \times f_{SW}}} \text{ Volts}$$

$$|V_{cm}| = \frac{200 \times V_{IN} \times C_p \times f_{SW}}{\sqrt{(50\pi \times C_p \times f_{SW})^2 + 1}} \text{ Volts}$$

In terms of dBμV this is

$$V_{LISN\_CM\_NOFILTER} = 20\log(\frac{|V_{cm}|}{10^{-6}}) = 120 + 20\log(|V_{cm}|) \text{ dB}\mu\text{V}$$

So for example, if $V_{IN} = 100$ V(A = 200 V), $C_p = 200$ pF, $f_{SW} = 100$ kHz, we get $V_{cm} = 0.4$ V or 112 dBμV (for the first harmonic). We can follow a similar procedure as for the DM filter to calculate the LC of the common mode filter, and thereby the L_{cm} and Y-caps corresponding to *Figure 10-1*.

Second Method: (Detailed)

This method will provide the entire *harmonic content* of the injected current. Looking at *Figure 14-10* we see the exponential edges of the current wave shape — this can be proven by a Laplace transform analysis. The "25" in the figure appears because 25 Ω is the

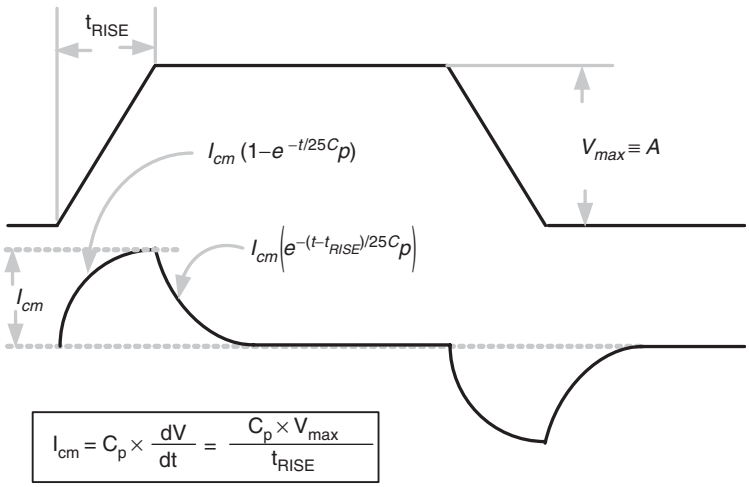


Figure 14-10: CM noise current Injected through Heatsink Mounting Capacitance

impedance the LISN presents to the CM current. Therefore the measured noise at the LISN when $t < t_{RISE}$ is

$$V_{cm\_1} = I_{cm} \times 25 = \frac{25 \times V_{max} \times C_p}{t_{RISE}} \times (1 - e^{-t/25C_p})$$

and for $t > t_{RISE}$ it is

$$V_{cm\_2} = I_{cm} \times 25 = \frac{25 \times V_{max} \times C_p}{t_{RISE}} \times \left(e^{-(t - t_{RISE})/25C_p} \right)$$

The same situation occurs when the mosfet turns ON — only the directions are reversed. We can do a Fourier analysis of the voltage waveform and as per Mark Nave's paper, we get

$$V_{cm} = \frac{50 \times V_{max} \times C_p}{T} \times \left[\frac{\sin\left\{ \frac{n \times \pi \times t_R}{T} \right\}}{\frac{n \times \pi \times t_R}{T}} \right] \times \left[e^{-jn\pi\left(\frac{t_{RISE}}{T}\right)} - e^{-jn\pi\left(\frac{t_{RISE}}{T} + 2D\right)} \right]$$

where D is the duty cycle, and we have assumed that t_{RISE} is the same as t_{FALL}. The sinx/x term again contributes an additional 20 dB/decade roll-off after the break frequency (described by x = 1). The term in the rightmost bracket has no roll-off. Its magnitude changes between the limits 0 and 2 as the harmonic number changes or/and duty cycle changes. So for all practical purposes, if we are just interested in the envelope, we can take its maximum value of 2 to get

$$V_{cm} = \frac{100 \times V_{max} \times C_p}{T} \times \left[\frac{\sin\left\{ \frac{n \times \pi \times t_R}{T} \right\}}{\frac{n \times \pi \times t_R}{T}} \right]$$

This equation is therefore flat untill the break frequency, after which it rolls off at 20 dB/decade. The flat part (the 'pedestal') can be found using the approximation $\sin x/x \cong 1$. It is

$$V_{cm} = \frac{100 \times V_{max} \times C_p}{T}$$

For our example of $V_{IN} = 100$ V (A = 200 V), $C_p = 200$ pF, $f_{SW} = 100$ kHz, we get

$$V_{cm} = \frac{100 \times 200 \times 200 \times 10^{-12}}{10^{-5}} \text{ Volts}$$

So $V_{cm} = 0.4$ V or 112 dBμV (for the first harmonic). This is the same result we got by the first method.

We make the following key observations with regard to common mode noise:

- The envelope is initially always flat and fixed at $\mathbf{100\ V_{max} C_p f_{SW}}$. At the break point described by $f = 1/(\pi t_{RISE})$, the envelope rolls-off at 20 db/decade.
- The pedestal (flat part) does *not* depend on the rise and fall times contrary to popular perception. So the envelope does change, but *not* at the low-frequency end. And for EMI purposes it is that end that is our starting point for a filter design — any subsequent roll-off is not going to affect the filter design.

Note: Since the pedestal of the common mode noise envelope is independent of the rise and fall times, does that mean that it doesn't matter how fast we turn-on and turn-off the mosfet? Yes it doesn't. But read the similar note previously given for the DM filter design section. (see page 430).

APPENDIX

1

Focusing on Some Real-world Issues

This Page Intentionally Left Blank

APPENDIX 1

Focusing on Some Real-world Issues

This is a compilation of articles written by the author. Barring some minor edits and graphical improvements, these are presented here exactly as they first appeared on the EE Times website www.planetanalog.com. Each article carries an apt (often irrepressible) introduction, and sometimes also a closing comment — from the Editor of Planet Analog, Mr. Stephan Ohr himself (who incidentally, also had a major hand in tweaking the language herein for, you guessed it — greater acceptability and lesser controversy!).

The purpose of these articles is to focus on specific issues that have come to the attention of the author over his many years in the industry, working across several continents. That is why they are written in a very personal style. But the content is hopefully not missing altogether!

Sounds Like Worst-case, But There's Danger Lurking in the Middle

(February 2004)

Sanjaya Maniktala describes a PC power supply failure that seemed widespread in Japan, but seldom occurred in the U.S. Was it something in the sushi? he asks. It turns out, the failures are related to the way components are specified and tested. This amusing and instructive piece kicks off a new column on power supply design. Sanjaya's comments will appear monthly at Planet Analog online.

In hindsight, this probably seemed routine enough to inspire some complacency. An OEM PC power supply was all set to go into full production. The Design Integrity Team put it through its paces. Maximum loads were applied at extreme ambients. Stress levels were verified, life predictions matched up, vibration testing, EMI, safety etc., were a cinch. The product was released and no problem was noticed in the several thousand units first shipped to the American market. Then the failures started to show up in droves. All came mysteriously from Japan. Must've been something in the sushi!

What happened was actually so simple that several people must have kicked themselves (and each other). Turns out this power supply had not been tested in Standby mode! Why test in Standby mode?? That's only a handful of watts compared to the 550 W max load. However, when fully operational, the power supply had a fan running off its main 12 V output. In Standby mode the fan stopped as all the outputs collapsed. Well, all outputs but one! A small standby integrated converter was also present on-board delivering a low power 5 V

housekeeping rail. Unfortunately, it was left freestanding by an otherwise experienced engineer who thought its low power wasn't worth his while. Its temperature in Standby mode thus made history. But we also learned that the Japanese actually initiate the Standby function of their PCs rather than leaving the computer idle, as Americans would.

Also remember that a buck converter's input capacitor sees the maximum RMS current at D = 0.5. If your input voltage range is, say, 15 V to 48 V, and you test it "diligently" at both the input corners, you still may never know how long your capacitor will really last. Or take an interleaved buck converter with two independent channels running out of phase. Here we are actually relying on both channels being fully loaded to reduce the input capacitor's RMS current. But in fact this current can be even higher if one channel is unloaded. Coming to magnetics, we also now know that the surface temperature of a core-loss dominated choke means nothing if there is a variable-speed fan present. Remember a few years ago when major manufacturers had field returns, in which the powdered iron chokes had virtually returned to their original powdered form because of prolonged core temperatures. This had then prompted vendors to scramble to characterize life expectancy. Try www.micrometals.com for more information.

Editor's Note:

Sanjaya Maniktala, author of the popular Planet Analog series on EMI in power supplies, has over 15 years of design experience with companies like Artesyn Technologies in the Bay Area and Siemens AG in Leipzig. He holds graduate degrees from the Indian Institute of Technology, Bombay and Northwestern University, Illinois, as well as patents on the Floating Buck Regulator topology. This article kicks off a regular column at Planet Analog online. "Write me at sanjayamaniktala@yahoo.com with your favorite bedtime story!" he bids.

Loop Design Sometimes Compensates for Lower-quality Switchers

(March 2004)

Sanjaya Maniktala is back with a new installment of his popular power supply design column. The question he asks is how to make up for the erratic quality of the 3842 and 3844 switching power supply controllers now on the market. An adjustment of the RC components in the hysteresis loop allowed a power supply company to safely utilize the batch lot of jittery components they were stuck with — some 50,000 of them.

The 3842/43/44/45 series of pulse width modulators are possibly the most popular controllers for off-line applications for several years now. Originally from Unitrode (now Texas Instruments, "TI"), I can find 'clones' from at least 12 more semiconductor companies on the web (see if you can better the count). The list may in fact be quite endless. Their quality is however often questionable, though I do admit that their Application Information seems all rather well written (hmmm, but did I read that somewhere else?).

Practically speaking these 'equivalents' can differ quite a bit. At one time we even had basic functionality problems due to an insurmountable jitter from an 'alternate source' part.

No amount of decoupling on any pin was helping, and it was thus obvious that it was the result of an internal noise feedthrough from the driver stage to the clock, leading to premature pulse termination and unpredictable frequency. The 3844 from the slightly more expensive vendor worked perfectly on the same board and was in fact being shipped out in millions (this was a flyback for a very well-known computer manufacturer). Now trying to return the 50,000 jittery devices to the equally well-known (though now clearly jittery) vendor was met with the familiar 'show me where in the datasheet' attitude.

I did manage to band-aid the problem shortly after my arrival at the Singapore-based company I worked for, and they did manage to slip the bad parts into their high volume production, with no future 'ppm' issues either. Though understandably they didn't return to the previous vendor to renew their learnings on what constitutes a 'guaranteed spec' and what is not.

This was the simple logic I applied to solve the problem: in the 3844, the clock pulse is generated by a simple RC charge-discharge cycle taking place between two fixed voltage thresholds. At the falling edge of the PWM pulse, noise was getting injected onto the ramp and could fool the internal comparator into 'thinking' that the timing capacitor had reached its upper threshold (see *Figure A-1*), at which point the discharge cycle would start prematurely (not shown in figure for reasons of clarity). Now what if we decreased the discharge time by decreasing the C, but simultaneously increased the R to maintain the same frequency? Now *the ramp is actually slightly lower at the instant where the falling edge of the PWM pulse occurs*. We can see that the noise margin has thus improved.

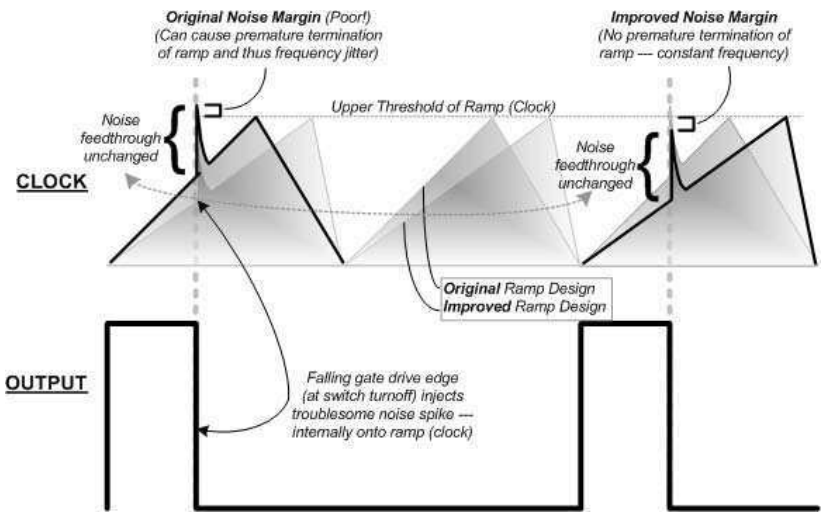


Figure A-1: Noise in the 3844

We must remember that normally, if there is a major change on the primary side of an off-line power supply, we definitely need a fresh approval from safety agencies. But this minor change in the RC combination apparently just merited a notification. End of story.

Please do write me with your personal experiences whenever you can. I will certainly acknowledge you. I can be reached at odd hours of the day and night at sanjayamaniktala@yahoo.com. Steve Ohr will hear about it too!

Re-inventing the Wheel . . . as a Square

(April 2004)

We've all seen renderings of the disconnect between marketing, engineering, and customers in product development – using a child's backyard swing as the example. What marketing requested shows a three-tiered swing, with cushions and lollypop dispensers. What engineering came up with shows a rocket-propelled platform, springing back-and-forth through a steel-reinforced tunnel. What the customer really wanted, of course, was a truck tire hanging by a rope from a tree limb. This month, Sanjaya Maniktala comments on some magnificently-designed power supplies.

A lot has been said about not reinventing the wheel. But how about not repeating errors? 'Errors' are like wheels that we shouldn't even have tried to invent, let alone re-invent, because this wheel was probably square in shape to start with, and there was never any chance of it succeeding.

Yet some may be surprised to know how often this may already have happened. And will happen again. It's just that we don't hear much about it.

If we had inside knowledge, we would often find that such projects usually had impeccable beginnings. Clear design goals, a thoughtful strategy, solid design trade-offs, but then one embarrassing turn to the left. The result was a product that had no known identifiable sire, and in fact no one even remembered ever having worked on it! We all know that marketing, given its nature, always tries to emphasize successes. And no doubt they would be working overtime to gloss this one up. But engineers, with an eye on not repeating known errors in future development, are always keenly interested in what should or could have been done. If only they knew.

Here is a list of some eyebrow raising situations which caught my attention over the years. The examples leave one incredulous.

Example 1

A telecom project required a rack of several 3000 W Power Factor Corrected (PFC) hot-pluggable power supplies. Two brilliant teams went about it, one writing C++ code and the other designing the power sections and the backplane. This was to operate off a 3-phase AC mains supply.

The engineers thought the best way out was to parallel three single phase 1000 W power-trains; that is, each power-train would be running off a different input Phase and the common Neutral. In doing so, the required minimum voltage rating of the FET switches would be the usual 450 V or so, as for any single-phase PFC stage, rather than other techniques which would usually require FETs with roughly twice the voltage rating. The project did get completed.

Then the Marketing guys appeared and informed them that they just couldn't sell it —!! Because in many countries and areas, a 3-phase incoming supply point does not even include the Neutral wire (which in any case is not even designed to carry that much of return current either). This was a complete dead-end. But couldn't the Marketing guys have got, involved a little earlier? Say about 2 years before the project came to a head?

Example 2

Users of a control IC meant for a Flyback topology should recognize that the maximum allowable duty cycle should never be set to 100%. A 100% duty cycle basically means the switch is no longer switching, and could just end up staying ON permanently. So if the output voltage is low, and the IC is trying to get it to rise by increasing the duty cycle 'D', and if D is 100%, there is now unfortunately zero available time for the current to freewheel into the output. So how can the output ever rise up try as hard as the controller insists? Yes such an IC has unfortunately been released into the general market. Look around!

Example 3

Another product is a fairly popular off-line switcher IC family meant for Flyback applications. In an off-line case, the value of Dmax has several more implications. Here we must recall that the earlier generation of this switcher family had a maximum duty cycle of about 67%. When the next generation was conceived, the one-man product-definition team heuristically assigned a Dmax of 78%. His idea was that by 'allowing' a wider input current pulse, we would automatically get a lower current pedestal, and this would enhance the 'power capability' of his device (since this figure was being based purely on current limit, not on dissipation).

This design strategy could actually have succeeded. But there is a catch. Let us consider what happens if we just remove the input power. By allowing the duty cycle to go up to such a high level, the momentary peak currents actually increase much more now, as compared to a case where the Dmax is set lower. This has severe implications on the transformer, since its size is related to the saturation level. There are also some other subtle issues. For example, it can be shown that the dissipation in the zener clamp can

also go up significantly, thus worsening the overall efficiency. So the 'advantage' if any, turned out to be an illusion.

Example 4

The popular 3842/3844 series I wrote about earlier, despite its popularity was apparently hastily conceived. Here we actually have a current mode controller that has no built-in slope compensation!! One would think that everything required for a particular topology should be inherent in the design. After all, we don't buy a bicycle from a store and then go out looking for a pair of tires for it! In this case we do just that.

Example 5

One of the first semiconductor companies to come up with ICs for implementing a boost PFC pre-regulator had got it all wrong, and they admitted that to us privately: You can never hope to do proper sine wave shaping with *peak* current sensing! You need to do *average* current mode control, because it is the average current drawn that forms the input current waveshape. Their competitor understood this, and so despite having broken in later, they quickly became the market leaders in PFC ICs. The former company changed hands, and is virtually unknown as a separate entity today.

That's all for now. Please do write me at sanjayamaniktala@yahoo.com.

The Mighty Zener

(May 2004)

We are by now sadly conditioned to expecting that nothing good is ever going to come our way, at least not easily and cheaply, writes Sanjaya Maniktala in the latest installment of his power supply design column. Surprise, surprise: The gate-source zener on your power FET (a couple of cents) can really save your bacon on safety issues.

Looking at any typical off-line power converter, we will usually find a controller IC driving a high voltage FET. At the gate, besides the usual pull-down resistor to (primary) ground, we may also see a paralleled 500 mW/18 V zener diode. Yes, this diode does cost a few cents, but omitting it can be even more costly! This is just one of those examples of minor 'details' that will ultimately distinguish a bad power supply design from a really good one.

On the face of it, this zener does look as if it is there simply to 'protect' the gate oxide layer under various transients and noise spikes that may be encountered in the field. But while the effects of this one single component can be subtle, they can be dramatically helpful. This is admittedly a rare situation, especially in power conversion, where we are by now sadly conditioned to expecting that nothing good is ever going to come our way, at least not easily

and cheaply enough, without some sort of indirect or unforeseen price to pay somewhere along the way. In fact, here too, we do have our share of some rather reflex-action pessimism abounding. People still seek to question the basic wisdom and validity of this rather crucial zener.

But let us start off by first describing how I personally encountered this issue. The Singapore based design-cum-manufacturing outfit I worked in at that time had a policy of never (knowingly) putting in even one superfluous 1-cent component. They considered that equivalent to 'shipping free parts with every power supply.' They understood that at their current manufacturing volumes, they could probably hire another power supply design engineer for every zener diode they could eliminate from the schematic.

Yet even they couldn't ultimately avoid this rather stubborn 'gate-source' zener. Note that this is actually a 'gate to ground' zener, because there usually is a current sense resistor between source and ground.

The trouble actually started after their first manufacturing sample had been built with no such zener in place. It was submitted for the mandatory UL1950 safety approval. At the test house, various abnormal tests were carried out. In some of these tests the switching FET exploded. And in all such destructive tests, the controller driving the FET failed too, sometimes quietly, and sometimes quite spectacularly. But either way, this was perfectly OK as per UL, since 'safety' was the only concern here, and this should never have been compromised. But in fact, in one case safety did get affected!

In this case, the optocoupler, which was as usual connected to the controller for regulation purposes, cracked open. Now that was something unacceptable to UL, since it meant that the 'sacred' primary to secondary insulation barrier (inside the opto) had somehow gotten breached. That could conceivably lead to a hazardous voltage level from the mains line input (primary side) reaching a user who may have been in physical contact with the system (secondary side) at that very moment. Of course, thereafter, the input fuse would also blow up, disabling the entire system. But a fuse can never be relied upon to blow up fast enough to prevent electrocution, its main use is only in preventing a fire.

This is the rapid chain of events that had apparently occurred:

1. The FET blew up and its drain and source shorted together.
2. The resulting high current ripped through the current sense resistor in the source, causing this MOF (Metal Oxide Film) resistor to fail open.
3. The inductor current coming in through the drain, still needing a path to freewheel through, diverted into the gate, raising its voltage and then entering the controller IC.
4. The controller IC then failed and the high voltage/current damaged several components connected to the pins of the IC, including the optocoupler!

5. The optocoupler cracked, and its safety barrier was breached.
6. Finally, the fuse blew (but too late!).
7. About half an hour later, the 'prime culprit' (one sleepy and hapless power supply design engineer somewhere out there) receives a midnight call from his fulminating boss. Surely the topic this time is not any upcoming promotion.

Now, had we pored over some older power supply designs we may have seen that a transient voltage suppressor ('TVS') was often 'mysteriously' placed across the current sense resistor. In fact, its purpose was to circumvent this very chain of events. It takes us from the end of Step 2 straight to step 6. A TVS is basically just a rugged zener diode, and one that is designed to always fail in a shorted condition. So when the MOF resistor fails in Step 2, the resulting rising voltage would cause the TVS to almost simultaneously fail too, thus maintaining galvanic continuity for the current to keep flowing from drain to source to ground, till the fuse interrupts. So in this case, the current wouldn't need to divert into the gate (and the IC) as happened in Step 3.

But we note that a TVS is a fairly expensive solution. So in Singapore, we decided to try a zener between gate and ground. In this location, the zener also always fails in a shorted condition. It thus protects the controller IC and all its associated components (including the opto), till the fuse interrupts.

During debugging stages, or in initial prototyping, a gate-source zener comes in real handy too. It not only saves a lot of soldering/desoldering, but it dramatically extends the life of the constantly reworked board. Because though the FET fails, the controller IC and all its associated components always survive. So even after an otherwise 'impressive' blow-up, we usually just need to replace the FET, the current sense resistor, the zener, the pull-down gate resistor, and input fuse, to be up and running again in half an hour.

And the skeptics: some are still convinced that the small anode-cathode zener capacitance can combine with the input capacitance of the FET and the lead and trace inductances to form a high-Q pi type of tank circuit (C-L-C). So they recommend a small resistor of about 10 ohms placed between the zener and the gate lead, to damp out any oscillations. Yes, inarguably, the zener must be very close to the FET, but about the oscillations?!! Well, one prominent FET manufacturer earlier was quite sure that this doomsday scenario could really happen, and had even stated as much in a certain Application Note (though this section was later removed). On further inquiries (from me in particular) they backed off, and in fact provided fresh data to actually disprove their own earlier assertion. So ultimately, privately they blamed it on one lone engineer of theirs, who didn't quite 'follow the book' when he reported he saw 'oscillations.' Probably a bad scope probe. We will never know.

As you can see, an engineer's job is not getting any easier. If you have time, do drop me a line, at sanjayamaniktala@yahoo.com . . . just don't blame me if your converter circuit isn't working.

Better Do the Math: Ignore Transfer Functions at Your Own Peril

(June 2004)

It seems that there must be at least two distinct groups of people working in power, writes our power supply guru. One group consists of academicians who invoke integral calculus equations to describe a buck regulator. Another group of power supply professionals may wonder if this is going to really help them design better products. Sanjaya Maniktala says the math CAN translate into a better product, not just some yellowing seminar material fighting for immortality on a dusty shelf.

It seems that there must be at least two distinct groups of people working in power. One consists of academicians who won't hesitate to invoke integral calculus to derive the dc transfer function of a buck topology (and then write textbooks filled with equations). Another group is practicing power supply professionals, many of whom tend to believe that anything even remotely abstract is not going to really help them design better products. I personally feel that there is a valuable middle ground available to all of us here.

By presenting examples of interesting "transfer function interactions," I am hoping to show that it does actually help significantly if engineers try to think in a relatively abstract manner, and it can translate into a better product, not just some yellowing seminar material fighting for immortality on a dusty shelf.

The dc transfer function of a topology is simply the expression connecting the input and output voltages. Most engineers realize that it follows directly from the fact that we have a voltseconds law in existence, which must be diligently upheld by the concerned inductor of any *viable* topology. 'Viable' implies that the topology (discovered or yet-to-be) can exist indefinitely in a steady and stable state. If not, we will certainly hear about it from the switch.

Now consider the equation for the output of a buck-boost in discontinuous conduction mode ('DCM')

$$V_O = \frac{D^2 \cdot V_{IN}^2 \cdot 10^6}{2 \cdot I_O \cdot L \cdot f} \text{ Volts}$$

where V_O is the output voltage of the buck-boost, L is in μH, and f is in Hz.

This therefore has the following dependency

$$V_O \propto D^2 \cdot V_{IN}^2$$

Now consider the dependency of a buck converter in continuous conduction mode ('CCM'):

$$D \propto \frac{1}{V_{IN}}$$

So the point here is as follows: if we have a composite topology in which the duty cycle of a buck in CCM is used to drive a buck-boost in DCM, we can get the dependency on V_{in} above to cancel out completely as follows:

$$V_O \propto \frac{1}{V_{in}^2} \cdot V_{in}^2 = \text{constant}$$

What does this tell us? If the output of the buck-boost is independent of the input voltage, clearly *we must have inherent line regulation.* And it is for free!

In practice, if we have a PWM controlled *master* buck stage, whose switching waveform is used to drive a *slave* buck-boost stage which has no independent PWM (just the switch being toggled), then we know that the output of the buck is certainly well-regulated (because we have a PWM and independent regulation loop), but the output of the buck-boost is also *partially* regulated with respect to line variations. We have no load regulation, but if the load of the buck-boost is fairly constant, we may not need it either. Or we can clamp the output of the buck-boost with a zener to give it constant load characteristics. Do note that this isn't bad considering we have only one PWM control!

We can also use the fact that the output voltage of a discontinuous mode converter at a given duty cycle depends on its inductance. So we can 'tune' the slave buck-boost to have the required output level (at its expected maximum load current) by a careful choice of inductance. Within a valid range, this technique provides completely adjustable auxiliary output voltages, something we cannot normally expect from composite topologies based only on continuous conduction modes.

Note that the zener on the output of this slave converter is almost completely nonconducting when the slave converter is working at its designed (maximum) load. The efficiency is therefore as high as we normally expect from any conventional switching power converter. However, if the load on the slave decreases, the zener comes into play and starts automatically shunting the balance of the current away. It is then behaving as a conventional shunt regulator. Therefore load regulation, which is taken for granted when dealing with single or multi-CCM stages, is not 'automatic' here. It is being 'enforced' by the zener, but luckily, if the inductance has been chosen correctly, this needs to happen only at less than maximum loads.

So how did we manage to achieve *automatic line regulation*? As the input voltage increases, the feedback loop of the regulated buck converter commands its duty cycle to decrease to maintain output regulation. It just so happens that this decrease in duty cycle is exactly what was required by the discontinuous mode buck-boost to 'regulate' its own output almost perfectly.

Have you also ever wondered why the number of turns on a single-ended forward converter transformer (not its output choke) can be calculated either at the high input voltage extreme or the low input voltage, or in fact at any input voltage? Here we have a transformer operating in DCM, but with a duty cycle dictated by a CCM equation (since it is coming from the output choke). Only then do we get a surprising transfer function coincidence: that when the switch is ON, the product of the applied voltage across the transformer and its duration (the voltseconds) turns out to be a constant, irrespective of the input voltage. Therefore any input voltage will give us the same number of required primary turns.

Now I need to transfer some of my functioning into other worldly interactions. But don't forget to write me at sanjayamaniktala@yahoo.com with your comments, insight, and power-horror stories.

Aluminum Cap Multipliers — Why We Can't Have Them and Eat Them Too

(July 2004)

With virtually the highest available CV (capacitance times voltage) capability, accompanied by the lowest cost, aluminum capacitors are still not even close to getting canned into history books, as some would think. Some of our younger engineers get rather charged up thinking about ceramic and modern polymer technologies, writes columnist Sanjaya Maniktala. They should really be paying closer attention to aluminum electrolytics.

With virtually the highest available CV (capacitance times voltage) capability, accompanied by the lowest cost, aluminum capacitors are still not even close to getting canned into history books, as some would think. Some of our younger engineers get rather charged up thinking about ceramic and modern polymer technologies, but they should really be paying closer attention to the viability and finer design aspects of the still undying aluminum electrolytic capacitor (hereafter called an 'elko').

So why not use an elko?? OK, it has a higher ESR. Granted! But let's not forget that 'all-ceramic' solutions can exhibit dangerous input oscillations, and it is now actually being recommended that to damp out these oscillations we should put a high-ESR elko in parallel to the existing input ceramic cap. We may also require a higher ESR just to ensure stability when using voltage-mode control.

To cut to the chase, let us therefore assume that we finally see the need to use an elko in a particular location. Now the main concern with such a component is its life expectancy. Eventually, the electrolyte inside will evaporate causing the capacitance to decrease, and beyond a certain level we would declare the capacitor 'dead' (worn out). We can clearly understand that a few factors will play key roles in this process:

a) The hermeticity of the end seals of the capacitor. However, no joint is one hundred percent perfect, and so some evaporation will take place slowly over time. But we see the need to pick a vendor with a high (and consistent) quality.

b) The surrounding temperature. The heat could come from nearby components or through internal heat dissipation. If we lower the temperature, the evaporation rate will decrease, and extend the life. We will see a little later how this leads to the published 'temperature multipliers.'

c) The core temperature. We expect that there will be hot-spots inside the capacitor since we have less-than-perfect thermal conductivity inside it. As a worst-case, that is the temperature to consider when calculating life. In fact the entire life expectancy calculation is reduced to accurately predicting this core temperature (since we can't measure it).

d) The ESR. This would certainly affect the internal heat dissipation, possibly raising the temperature and aiding the evaporation process.

e) The frequency. Since ESR can be a function of frequency, the frequency will indirectly affect the life of the capacitor. We will see that this leads to the published 'frequency multipliers.'

The most important datasheet parameter is the ripple current rating. This is typically stated in Amperes RMS at 120 Hz and 105°C. It essentially means that if the ambient temperature is at the maximum rating of 105°C, we can pass a (low frequency) current waveform with the stated RMS, and in doing so we will get the stated life. The declared life figure is typically 2000 hours to 10,000 hours under these conditions. Yes there are lower grade 85°C capacitors available, but they are rarely used, as they can hardly meet typical life requirements at high ambients.

Let us now understand what a frequency multiplier tells us. The ESR of an elko is also usually stated at 120 Hz. The vendor may have directly provided a ripple current rating at 100 kHz in addition to the 120 Hz number. If not, he would certainly have provided 'frequency multipliers.' A typical frequency multiplier is 1.43 at 100 kHz. That means that if we are allowed 1 A ripple current at 120 Hz, then at 100 kHz we are allowed 1.43 A. This, by design, will produce the same heating (core temperature rise over ambient) as 1 A causes at 120 Hz. Therefore this is also equivalent to saying that the ESR at 100 kHz is related to the ESR at 120 Hz by the following equation:

$$\left(\frac{I_{100\ kHz}}{I_{120\ Hz}}\right)^2 = \frac{esr_{120\ Hz}}{esr_{100\ kHz}} = (1.43)^2 = 2.045$$

Thus the high-frequency ESR is about half the low-frequency ESR. Frequency multipliers should be used always, or we will overestimate the heating and underestimate the life, possibly forcing us to move to a larger cap size.

Temperature multipliers we have to be more careful about. And we have to clearly understand what they really imply.

The datasheet usually provides certain 'temperature multipliers' for the allowable ripple current. For example for the old but well-known LXF series from Chemicon, the numbers provided are

1. At 65°C the temperature multiplier is 2.23.
2. At 85°C the temperature multiplier is 1.73.
3. At 10°C the temperature multiplier is 1.

This means that if for example the rated ripple current is 1 A (at a maximum rated ambient of 105°C), then we can pass 1.73 A at an ambient of 85°C, and 2.23 A at an ambient of 65°C. *But in doing so, the core temperature will remain the same.*

So what is the actual story the temperature multipliers are telling us? The amount of heating and the core temperature rise are proportional to I_{RMS}^2, so if we assume that in every case the final core temperature was the same, that is, T_{CORE}, then comparing the 105°C ambient case with that at 85°C:

$$\frac{T_{CORE} - 105}{T_{CORE} - 85} = \frac{{I_{105}}^2}{{I_{85}}^2} = \frac{1}{1.73^2} = \frac{1}{3}$$

We can thus solve for T_{CORE} to get

$$T_{CORE} = 115°C$$

This says that if we pass 1.73 A at 85°C, or 1 A at 105°C, the core temperature will be 115°C in either case. In fact, for most 105°C rated capacitors, we will have roughly 5°C differential from ambient to the outer can and then another 5°C from the can to the innards (the core), giving us a total of 10°C from ambient to core.

Let us check our reasoning by confirming the 65°C multiplier

$$\frac{115 - 105}{115 - 65} = \frac{10}{50} = \frac{{I_{105}}^2}{{I_{65}}^2}$$

So the multiplier must be $5^{0.5} = 2.236$, which agrees with the published datasheet value. Therefore we see that from the vendor's published ripple current temperature multipliers, we can easily deduce his designed-in maximum core temperature.

The problem with this is that if the core temperature is at its maximum rated 115°C, the life would always just be the declared 2000 hours or so. But that is hardly enough to get us through even one quarter of a year. We usually need at least about 44,000 hours (5 years)

of life expectancy from all elkos used in a typical commercial power supply. How do we get there? We do that by reducing the core temperature thereby slowing the evaporation rate of the electrolyte. Does this imply we should not be using temperature multipliers to increase the current?

There is actually another complication. It has been determined that not only is the absolute value of the core temperature important, but the differential from can to core is critical too. So if we increase the differential beyond the designed-in 5°C, the life can deteriorate severely, even if the can is held at a lower temperature. But the designed-in differential of 5°C occurs ONLY when we pass the maximum specified ripple current (*no temperature multipliers* applied), and that is irrespective of the ambient. Which means that as a matter of fact we cannot use any temperature multipliers at all. So, if the cap is rated to pass 1 A at 105°C, then even at an ambient of say 65°C, we are allowed to pass only 1 A, NOT 2.23 A.

When the differential is decidedly kept equal to or less than the designed-in value, the life of the elko is then determined by the familiar doubling rule — for every 10°C fall in core temperature (from its maximum rated), the life doubles. That is how we can finally get the required 44 khours. For example if the core is correctly estimated to be at 65°C, then the calculated life of a 2000 hour capacitor is actually $2000 \times 2 \times 2 \times 2 \times 2 \times 2 = 64$ khours.

But we do see that we can't have our cake and eat it too. We can increase the ripple current (but not the life) by applying the temperature multipliers OR we can increase the life (but not the ripple current) by not applying these multipliers. We just can't have it both ways!

Elkos give us lots of benefits as it is, and we just shouldn't be asking for any more.

Do write me at sanjayamaniktala@yahoo.com. Please copy Steve at sohr@cmp.com too, just so he knows I am not babbling.

Limit Your Peak Current, Not Your Reliability

(August 2004)

Pop. Pop. Pop. Pop. It's not the Fourth of July. It could be your own dc-dc design blowing up, worries our power supply expert, Sanjaya Maniktala. You do want to check out your design under a variety of adverse conditions, he suggests. Fault management is the key to designing power supplies that last.

Now that you finally got your dc-dc converter working, is it really time to start popping the champagne? Shouldn't you at least wait to see whether it even survives the very first abnormal condition that comes its way? You don't really want to be the one to have to decipher (in your rather understandably hazy and exulted condition) whether that second pop came from the adjoining bottle or from your very own power supply. Fault management is the key to designing power supplies that last. Unfortunately, all this also has the immediately

sobering potential to completely unmask the full extent of your design capabilities! Therefore, ensure that your current limiting is up to the task before you begin to celebrate.

In a typical high-power off-line supply, fault management requires a clear understanding of the required shutdown and subsequent recovery sequencing of the various constituent stages: the power factor correction stage, the auxiliary (house-keeping) power supply, and the power trains. This is complex for sure. Therefore, in single-chip converters, we tend to think that the task of surviving abnormal conditions is trivial. But not so fast buster! A great deal depends on how quickly the current limiting acts. Also, whether it is set correctly to start with. And whether it can even do its expected job. (For the latter aspect, we will delve into the subtleties of 'frequency foldback' in our next column. Here we first do all the necessary spadework.)

In the high-voltage off-line power industry, the underlying design philosophy is to size the transformer *as per the current limit.* Which means that the transformer should not saturate even if the current hits the limit, as it usually will during any normal power-up or under shorts or overloads. Of course it is always good to have the flexibility to set the current limit 'correctly' — to be *just* able to meet the required holdup time, step load capability, or any specified transient peak load requirement (as for disk drive motors, incandescent lamp loads, and so on). So though the thickness of the copper windings is certainly determined only by the continuous operating current (long-term heating effects), the actual physical size of the magnetic core must be set strictly according to the current limit alone. And this may or may not be related to the continuous operating current, as could easily happen if we use some popular integrated switcher families which provide only a discrete range of built-in fixed current limits.

Implicit in this design strategy is the realistic realization that we just cannot set a given current limit, and expect it to be enforced fast enough to be able to save the switch if the transformer starts saturating, even momentarily. While this was blatantly obvious for bipolar transistors in the past, though the situation has improved with the advent of FETs, it is not enough to have changed this basic design philosophy. At least not for high-voltage applications.

The design philosophy prevalent within the low-voltage dc-dc semiconductor industry is so vastly different from that of the off-line power industry, that transmigratory power engineers (nomads like me in an eternal search for a meaningful home) have a hard time reconciling initially. Here too, we can have families of integrated switchers with fixed current limits, but we almost invariably end up totally ignoring the current limit (just so long as it is high enough to guarantee the desired output power). So we size the inductor according to the continuous operating load current — no more, no less. For example we may use a switcher with a 5 A current limit for a 2 A application, and use only an inductor rated for 2 A. We could also use the same switcher for a 4 A application and we would then use a 4 A

inductor. Here we are assuming that since the switch is obviously OK with a current of 5 A, then even if we use a 2 A inductor, and it saturates momentarily, the current limit circuitry is capable of acting fast enough to immediately turn the transistor OFF the exact moment that it attempts to exceed 5 A. So the switch can never be destroyed, since its legal limit is assiduously enforceable.

Now this scenario is clearly more likely to be true for FETs than bipolars, because the former have virtually no storage/delay times and can react almost immediately. But "almost" is not always good enough. Not in power. For example, if the input voltage is raised, the *saturating inductor's current may ramp up so steeply*, that even during the minimum ON time (minimum pulse width), the current may spike up high enough to damage the switch. In this case, we should ensure that the inductor is not saturating at the moment the current attempts to exceed the set current limit. That gives the circuitry enough time to safely turn the switch OFF.

While testing and evaluating the LM2590HV to LM2593HV family of high-voltage devices we observed some failures if we used very small sized inductors when the input voltage was higher than 40 V. Thus our official recommendation in the datasheets for the inductor selection is somewhat of a cross between the two design philosophies talked about above — we have stated that if the input is less than 40 V, the inductor should be sized as per the continuous rated load current, but if the input exceeds 40 V, we should size the inductor as per the fixed current limit of the device (disregarding the load current totally).

We should also keep in mind that the above mentioned devices are bipolar-based integrated switchers. We can expect the situation to be better if we use integrated devices which use FET switches. On the other hand, if we are using controllers, then even with FET switches we should be cautious, because the switching speeds and corresponding delays are likely to be worse than with integrated switchers. It all boils down to our ability to set the current limit accurately and to enforce it fast enough. It is not a trivial proposition.

The reader may wonder how we ensure rock-solid reliability for low-voltage applications if we disregard the current limit altogether, as we effectively seem to be doing above. Well in reality we don't, but this aspect is cleverly hidden from most users, though not deliberately so, and certainly not to their disadvantage! Most integrated switcher families are actually virtually 'bullet-proof.' And one of the key methods to ensure this is by providing a *second level of current limit protection*, usually not even mentioned in the datasheet, since it is rarely encountered and almost completely transparent to the user. But this second limit is typically about 20 to 30 percent higher than the first (declared) current limit. If this is ever encountered, as with severely saturating inductors and high input voltages, the IC is designed to enter a foldback condition. As we said, more on this subtlety the next time!!

Till then don't forget to drop me a line at sanjayamaniktala@yahoo.com and do copy Steve at sohr@cmp.com.

Reliability Is No Flash in the Pan

(September 2004)

Don't be so quick to pop the champagne when your new power supply prototype seems to work, Sanjaya wags a finger. Lest you want a lawsuit, ensuring reliability requires a disciplined effort around a second current limit. It's a long lonely trudge back to the drawing board, he says.

That was some night of celebration! Now past the ebbing wave of our once-blissful incognizance, we are become increasingly aware that the fireworks display back there did not herald the onset of any new-found freedom – rather a lonely trudge back to the drawing board in the very near future, accompanied only by the growing realization that a) switching power conversion is not half as easy as it sounds, and b) reliability should never be considered to be a flash in the pan.

In the previous month's column we revealed that many integrated switcher ICs are virtually indestructible by design. In that case, why even bother about knowing all about the peak currents? To answer this delicately and with a measure of natural human kindness, the second current limit not only helps make the IC 'bulletproof,' but also 'idiot-proof.' This second current limit, if present, is not even supposed to be encountered under normal operation, or even under "normal abnormal conditions!" That sounds oxymoronic, but in fact most of us practicing engineers would have to struggle extremely hard to even get the peak current to slip past the first current limit unnoticed, so as to trigger the IC into this last-ditch survival mode. But if that happens, unfortunately, now the problem is that the converter output just folds back (low output voltage at higher loads). This in turn is the result of the second level current limit comparator being designed to either cause a smooth, progressive reduction in the switching frequency ('frequency foldback'), or just skipping several on-pulses (the number of pulses skipped being roughly proportional to the amount by which the peak current is exceeding the second level threshold — an averaged sort of frequency foldback).

But before looking further at frequency foldback, let us again beat upon some of the points from the previous month's column, and refine them further. We saw that for a low-voltage integrated switcher, we can for example use a "5 A device" for a 3 A load, and then its OK to use only a 3 A inductor. Note that we have been talking only about buck switchers so far. A "5 A" buck switcher is by definition meant for a maximum load current ('Io') of 5 A, and therefore its current limit ('Iclim') will usually always be internally set a little higher, maybe between 6 to 6.5 A. But this train of logic relies on a property that is specific only to the buck topology: its average inductor current equals its load current. We also recollect that the usual design procedure for selecting the inductance for any dc-dc topology is to set the peak inductor current about 20% higher than the average inductor current. We can show that this '20% inductor criterion' leads to an 'optimum' of sorts from the viewpoint of all the power components of the converter (for more details see application note AN-1197 at

http://power.national.com). So for a buck, with a 5 A load current, we will typically have a peak inductor current of $5 \times (1 + 0.2) = 6$ A. But remember, this holds only for a buck!

For 'non-buck' topologies we have to be careful because there *the load current has no simple relationship to the average inductor current.* And therefore nor to its peak or to the required current limit either. This means that if we come across a device declared to be say a "5 A buck-boost" (or boost) switcher IC, we should remember that this just means that 5 A is its set *current limit, NOT its max load current*. The max load for a non-buck topology depends on the input-output conditions of the particular application. For the boost and the buck-boost, the average inductor current is not the load current 'Io' but is Io divided by $(1 - D)$, where D is the duty cycle. (See AN-1246 at http://power.national.com for more details). This leads to the following basic design rule for the two 'non-buck' topologies: *the worst-case average (and peak) inductor current occurs at the lowest input voltage*. That is thus the input end at which we must design or select our inductor. In contrast, for a buck, we always pick its inductance at the maximum input voltage of the application, because that is the end at which its peak current is the maximum.

We need to refine the above numerical computations slightly and to also see the *importance of tolerances in reliability and cost*. We said that for a low-voltage 5 A buck we can pick a 3 A rated inductor if our max load current is 3 A. That is not completely accurate. Though the chosen inductor is allowed to have a 3 A continuous rating (based on the copper thickness and core loss), its peak or saturation rating (i.e. the current at which the core shouldn't saturate even momentarily) must be such that it can handle the peak current, which by design is typically about 20% higher than the average value, that is, $3 \times 1.2 = 3.6$ A. But we also know that if this happens to be a "high-voltage" application (defined here as an input greater than about 40 V) we may have to size our inductor according to the current limit, not the load current. That number is 5 A in this case. But very roughly so!

Here we should actually check the IC datasheet to see the MAX value of the current limit range. For example, for a 5 A switcher the MIN value may be say 6 A (set high enough simply to guarantee full 5 A load capability with the usual 20% inductor design criterion mentioned above), but the MAX value of the current limit may be say 7 A over temperature and process variations, and depending upon how it may have been trimmed in production. So now, though the inductor needs to have a minimum continuous rating of only 3 A (for a load current of 3 A), the inductor must be able to handle peaks of 7 A! That primarily determines the size and cost of the inductor. Not the amount of copper. In fact, for a high voltage application, the load current is not really important anymore, because the weight of copper used doesn't really affect the inductor cost much, provided of course that the available winding window area is enough to accommodate the required number of turns, and so we are not being forced to pick a larger core size simply to accommodate the required windings, or because the overall core temperature is too high as a result of excessively thin wire gauge.

We also now realize that all "5 A" switchers from different vendors are not necessarily the same! The *spread/tolerance of the current limit* is very important as it determines the size and cost of the inductor when dealing with high voltage ICs. The MIN value of the current limit is important because it determines the minimum guaranteed power output, and the MAX value of the current limit determines the size of the inductor. Therefore, some key manufacturers of high voltage integrated circuits pride themselves on how 'tight' their tolerance is on the set current limit. That is indeed admirable, but only provided we are comparing apples to apples; that is, comparing only high voltage integrated switcher families with fixed current limits from different vendors. This comparison makes no sense if we have a device (e.g., a controller) where we can set the current limit externally. In that case we can match the current limit much more accurately to the load current, and thus get the smallest possible core size. That helps reduce core size. (More on this aspect in the next month's column and how such vendors virtually trick us into 'thinking' that the core size has decreased.)

For non-buck topologies, we now realize that say for a buck-boost application with an input of 15 V to 25 V and an output of 15 V output, the max duty cycle is 50% and this occurs at the minimum input of $V_{in} = 15$ V. So if the load current is 5 A, the average inductor current is $5/(1 - 0.5) = 10$ A. With the 20% inductor criterion, the peak switch current will be $10 \times 1.2 = 12$ A. So the MIN of the current limit must be set higher than 12 A. And then depending on the available accuracy for current limit, the MAX may be as high as say 20 A. Clearly, a 5 A switcher is not going to suffice, nor an inductor rated only for 5 A!

A related issue is the case when a buck IC is used in a so-called 'inverting configuration'. We should realize that in doing so actually the topology has in effect *changed from a buck to a buck-boost*. So now we just cannot get 5 A of load current from a declared '5 A' IC. How much load current is possible depends on the specific input-output conditions. So again, our peak current is not close to 5 A, nor should the inductor rating be "5 A". Or we will certainly be frozen into a July 4th timeframe forever.

From an IC designer's point of view and even our applications level understanding, we must carefully recognize another basic problem with the very concept of current limit. Suppose we now have a very 'fast acting' current limit, and we also use 'blazingly fast' FETs (very low gate charge). Does that mean we are 100% protected? Not necessarily! What does the current limit comparator really do? All it can do is to command the duty cycle to reduce further when we hit the current limit. But it can't make the pulse width narrower than a certain *minimum on-time*. This small minimum pulse width of about 100–150 ns is usually required for the internal circuitry to be able to sense the current every cycle, and we also actually have to turn the switch ON every cycle for the purpose. In fact this minimum pulse width may need to be set even higher, say around 150–250 ns, especially if we are using controllers (as opposed to integrated switchers) since a 'good' controller IC must handle a

relatively wide range of FET characteristics, and various possible PCB layouts and their corresponding trace delays and glitches.

Current mode control may also be worse off in this regard, because of the need to incorporate a certain higher than usual minimum leading edge noise blanking time. Generally, high frequencies will aggravate the situation further, since the same minimum pulse width corresponds to a much higher minimum duty cycle at high frequencies. What this leads to is the very curious situation as indicated in *Figure A-2* (for a buck). Consider a 'hard' power-up, that is, a sudden application of input power, with a high dV/dt, as say with a banana plug slammed into the lab dc power supply. Initially, there is no output voltage rail present, so the current ramps up at the rate of Vin/L, eventually steadying to (Vin − Vo)/L. But during the off-time, at initial powerup, the ramp-down rate is much smaller, as it depends basically only on the diode voltage drop of 0.5 Volts or so. So every cycle, the net current rises just a little higher, irrespective of any current limit. This is a clear case of current staircasing. Over several cycles, depending on the amount of output capacitance and the input dV/dt, the inrush current may go up to very high and almost unpredictable levels. Even 'soft-start,' if available, may serve absolutely no purpose at all, at least not in controlling the inrush current. The situation actually gets worse with a 'good' diode, that is, one where the diode voltage drop is less, because it is this diode drop that stands almost alone in trying to get the current to ramp down. The same thing happens if we short the output!

One solution to this problem is to reduce the frequency (or just omit on-pulses as shown in the lower diagram of *Figure A-2*). This effectively increases the off-time and reduces the minimum duty cycle, and thus gives enough time for the current to ramp down. See the datasheet for LM1572 (at http://power.national.com) for a deeper explanation of such a practical frequency foldback scheme. But note that if too much off-time is available (an excessive amount of frequency foldback), the average current may not be able to rise high enough to meet even the initial load requirement under startup. That could well lead to a scenario that is rather uncomfortably well known in the semiconductor industry — that of an IC which has a declared and mysterious "startup problem" under certain types of loads. In general, we have to be very careful in implementing any type of foldback scheme. Foldback is a two-edged sword. It might help control responses to abnormal conditions, but may also end up encroaching on normal responses.

Another method being used nowadays for synchronous buck ICs is to sense the current during the off-time (i.e. across the lower FET). Note that this allows us to skip on-pulses entirely and is in effect a frequency foldback of sorts. Though its main purpose in powering modern core processors is to be able to carry out a high-to-low conversion at high switching frequency. For example, from 20 V to 1 V at 1 MHz would require 1/20th of 1 μs, that is, an on-pulse no wider than 50 ns. That doesn't leave us much time to be able to sense current in the high-side FET. So the only option is to sense current in the low-side FET. On the face

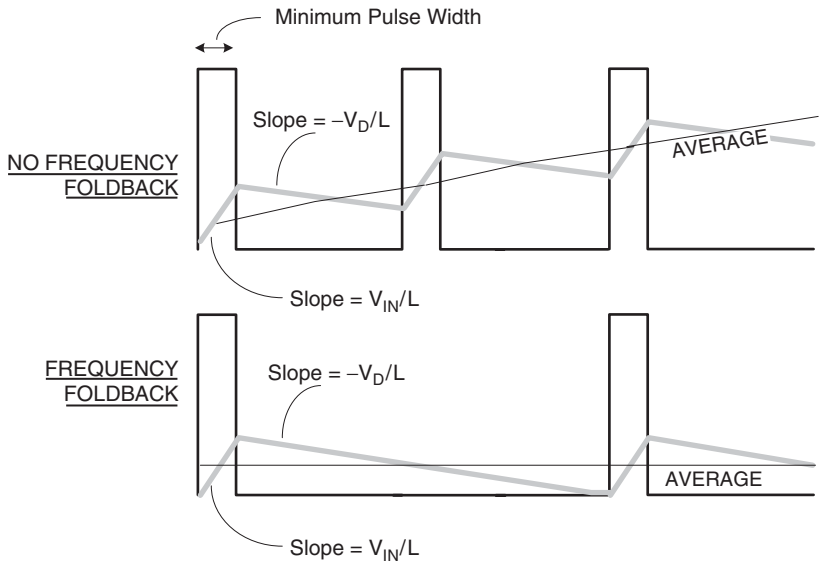


Figure A-2: Frequency Foldback

of it, low-side sensing should also help restrict the current under startup and overloads. But it has its own set of peak current and reliability problems, and these will form the subject of a column in the near future.

Write me at sanjayamaniktala@yahoo.com. Please don't hesitate to ask for pdfs of my older articles, as some of you have already done. And also don't forget to also write Steve (at sohr@cmp.com) and give us the good news that the fireworks display is over at your end.

Editor's Note:

Sanjaya has a new book out, which the publisher, McGraw-Hill Professional, is promoting as "The Bible" for power supply designers. On Amazon.com, the title is Switching Power Supply Design & Optimization.

The Incredible Shrinking Core

(October 2004)

Magnetics is a terrible embarrassment to many engineers, writes Sanjaya Maniktala, in this month's column on power supply design. I suspect they often end up pretending it doesn't really matter ("magnetics-denial"), he says. But dc-dc size, it turns out, is more a function of reliability than of switching frequency. Undersizing the core, he reminds, can have some serious consequences.

Magnetics is a terrible embarrassment to many engineers. I suspect they often end up pretending it doesn't really matter — 'magnetics-denial' — "Oh, I just toddle up to the bin and pick any inductor that works."

Considering that the entire movement in switching power conversion in the last decade to higher and higher frequencies is driven mainly by the burning desire to shrink magnetic components, there must be something wrong with this rather cool, laid-back attitude. It is of course my personal opinion that some engineers often make the problem sound even more complex, by spending all their remaining waking hours twiddling with the Nyquist criterion, double-edge modulation, and so on.

To put it in perspective, none of these issues have really ever been a show-stopper in any practical design scenario, nor have they allowed us to eventually reduce the size of the power supply. We note that size must ultimately dove-tail with reliability, because if we undersize the core for example, we will certainly cause core saturation and a fair amount of resulting silicon shrapnel in the lab! We saw that last month. Now consider the equations:

$$1)\ E = \frac{\mu\mu_0 N^2 I^2 A_e}{2l_e} \times \frac{1}{z}$$

$$2)\ B = \frac{\mu\mu_0 NI}{l_e} \times \frac{1}{z}$$

$$3)\ E = \frac{B^2 V_e}{2\mu\mu_0} \times z$$

$$4)\ L = \frac{\mu\mu_0 N^2 A_e}{l_e} \times \frac{1}{z}$$

where

$$z = \frac{le + \mu l_g}{l_e}$$

μ = relative permeability
μ_e = effective permeability
(MKS units)

Above, I have extracted four key equations from my recent book. I hope to give you a simple insight into the art of reducing the core size. Here we are assuming that core and copper losses are not the limiting factor (as is usually the case with the modern geometries and materials), and that the inductor size is simply related to the energy storage requirement or $\frac{1}{2} \times L \times Ipeak^2$. In my book, I have introduced a useful variable called the 'z-factor,' defined

above, since I found it helps simplify the equations considerably. (Also, see the magnetics on-line seminar at http://www.national.com/onlineseminar/2004/magnetics/magnetics.html.)

Here is the logic: from the fourth equation, we see that to be able to keep the inductance fixed as z goes from 1 to 10 (air gap increased), we only need to increase the number of turns N by $10^{0.5} = 3.2$ times. Therefore, from the second equation we can see that if z went from 1 to 10, but the ampere-turns NI was increased only by a factor of 3.2 (so as to keep L fixed, as we usually want to do), then the operating B-field would be reduced to 1/3rd of its original value.

From the first equation, the energy stored in the core has remained unaltered in the process, though from the third equation we can see that its overload capability (i.e. measured up to a certain saturation flux density B_{SAT}) has increased 10 times. So, any "headroom," as measured from the operating B-field value to the saturation level (B_{SAT}), or from the operating energy storage level to the peak energy handling capability must have increased considerably, even though inductance has been kept a constant in this case.

All this could translate to a much higher field reliability where the converter will likely encounter severe abnormal or transient line/load conditions. However, if all the "bells and whistles" are present in the design of the control circuitry (e.g. feedforward, primary/secondary current limit, duty cycle clamp, and the like), and they serve to protect the converter adequately against any such abnormal conditions, this gives us a great opportunity to select a smaller core for the same power level. In doing so we would be essentially returning to the point of optimum core size, which is defined as that where the peak operating flux density B_{PEAK} is just under B_{SAT} (with current limiting and/or duty cycle clamping present to ensure that B_{SAT} is never exceeded, even for a cycle).

What do we learn here? That by increasing the gap of the core we can move to smaller core sizes. Yes, powdered iron cores for example have a distributed air gap, and come in various "effective permeabilities." So actually, *lower* permeability materials should in principle always lead to smaller core sizes, as they have a larger air gap in effect. All this is rather counter-intuitive I admit. The restricting factor is that to use very low permeability materials, we need more and more turns, and so we will either just run out of enough window space to accommodate these extra turns, or we will have our copper losses mount to the extent that the core size becomes a secondary issue.

Now returning to another issue I promised to touch upon in last month's column. High-voltage off-line integrated flyback switcher ICs are available from several vendors, but they are restricted because they usually come only in a family of fixed current limits. So if for example we have a 5 A part, the next lower part being a 3 A part, the 5 A part is certainly optimum for peak currents slightly below 5 A. But what if the peak current in our particular application is 4 A? For lack of a suitably matching part, we would now be forced to use a

5 A IC for a 4 A application, but we would still need to size the core for 5 A! We may be able to reduce the copper diameter in going from a 5 A application to a 4 A application, but not the physical size of the inductor, as it must still continue to withstand 5 A, which it would see under sudden step-load changes (or even under normal power-up or power-down).

So how do some such vendors manage to showcase a smaller magnetic component for the 4 A application? By increasing the current ripple ratio 'r'! By the previous equations, it can also be shown that if L is allowed to decrease (fewer turns), the energy storage requirement decreases and so we can reduce the core size. Inductors operated with large current ripple ratios are therefore always smaller, though the problem is that they just transfer the burden to the input/output capacitors (more filtering required). But you may not notice that immediately!

Write me at sanjayamaniktala@yahoo.com. Please don't hesitate to ask for pdfs of my older articles, as some of you have already done. And also don't forget to write Steve (at sohr@cmp.com).

Editor's Note:

Sanjaya has a new book out, which the publisher, McGraw-Hill Professional is promoting as "The Bible" for power supply designers. Available through Amazon, the title is Switching Power Supply Design & Optimization.

Plain Lucky We Don't Live in a PSpice World!

(November 2004)

We have a natural ally in nature, writes columnist Sanjaya Maniktala. Design problems occur, he says, when we schedule a confrontation with natural forces. Nature doesn't have "convergence" problems, like PSpice often does. Try simulating the air flow over a heat sink.

We should keep that in mind, that all we need to do sometimes is to just sit back and allow nature to do its 'thing.' We have a natural ally in nature. Design problems actually really start, only when we have somehow managed to adopt what is in effect a schematic confrontation with natural forces. Luckily, nature doesn't have 'convergence' problems, like PSpice often does. All this may sound like some vague debating point in some esoteric man vs. machine philosophy panel discussion but it is actually just plain design common-sense. Appreciating these finer aspects of nature can help us succeed with a host of seemingly mundane or challenging engineering pursuits. And it's almost faster than getting our grand simulator engine to essentially mimic nature itself.

As a seemingly trivial example, we all know that when an object heats up, the air around it moves upward trying to cool it down. Eventually, we achieve thermal equilibrium. But have you noticed that oddly enough, the higher the dissipation, the lower the thermal resistance (expressed in °C per Watt). This is because the rising air turns "turbulent" under higher

dissipations, in an effort to help us even further. This phenomenon, once understood, is actually exploited in creating special pin fin heatsinks that try to provoke turbulent air flow even at lower dissipations. You can try "Pin Fin Heatsinks" (http://powerelectronics.com/mag/power_pin_fin_heatsinks/) and "Equations of Natural Convection" (http://powerelectronics.com/mag/power_simple_equations_simplify/index.html) for more information. Carrying our learning to the extreme, small miniature fans are sometimes mounted on high dissipation ICs blasting air perpendicularly at very close quarters onto the exposed hot surface. This is called 'impingement air flow,' and it leads to a further and really dramatic reduction in thermal resistance. A good read on this technique is available at the ETD Library of Louisiana State University (http://etd02.lnx390.lsu.edu/docs/available/etd-0411103-105200/unrestricted/Chapter1.pdf).

Now let's take a typical switching converter. If we apply a voltage across an inductor during the switch on-time, we get a corresponding increment in current from the basic equation $V = LdI/dt$. But then we turn the switch OFF. Now we will find that a certain voltage automatically appears across the inductor. Its magnitude may be undefined initially (as during initial power-up), but what we can be sure of always is that it is of reverse polarity to the voltage we applied during the on-time. If we think hard, we realize that our contribution as engineers is simply that we managed to create a circuit schematic (or 'topology') where we allowed nature to develop this reverse inductor voltage. But if for example, in a typical converter, we put the diode in the wrong way (or forget the diode altogether!), we may just be preventing this voltage reversal, causing instantaneous combustion.

From $V = LdI/dt$, we also see that to get the increment in current during the on-time to exactly equal the decrement, every cycle, we need the quantity 'V multiplied by time' during the on-time to be equal in magnitude to the same quantity during the switch off-time. In fact, that is the fundamental voltseconds law of power conversion. It leads directly to the expression of duty cycle in terms of input and output voltages.

But what drove nature to strive to do all this? To put it bluntly, we are lucky we don't live in a 'PSpice world.' Luckily, most natural processes tend to converge in our world and without 'user intervention!' We can foresee that if the current doesn't *decrease to exactly the same instantaneous value it had at the start of the cycle*, then every cycle there will be a small net increase in current. After a million switching cycles (nowadays that could take just one second!) this 'small' net increase will not be so 'small' anymore. Ultimately, we would probably never achieve a measurable or stable 'steady state' on the bench. And any 'switch' we may develop will ultimately combust under these escalating currents. Keep in mind that if it is not current, the switch will certainly fail due to excess voltage, because nature goes all out to help us, even increasing the reverse voltage (if that is schematically possible) to force a 'reset' (i.e. convergence in effect). Too bad if we didn't leave any available doors open to allow nature to step in to help us out here. On the other hand, we would be hard pressed to

make a typical PSpice-based converter circuit stabilize without explicitly implementing closed-loop feedback.

Note that the voltage reversal can be considered from the electromagnetic viewpoint as a result of Faraday's law of induced EMF (or Lenz's law). It is interesting to recognize that not only 'transformer action' would not be possible without this law, but no stable dc-dc inductor-based topology could exist either, because Faraday's law is simply the voltseconds law in another form (or vice versa). Without Faraday, we have no voltseconds law either. And without this, there would be no switching power conversion for one!

But what has all this to do with PSpice? Well, on a more subtle level, and in the same spirit of things, nature also tries to lend an additional helping hand by imparting 'parasitics' to every component we use. These actually help many processes converge eventually (or stabilize), even if we have partially overlooked some crucial design aspect or abnormal operating condition. Yes, these parasitics do seem like a nuisance usually, but they have the potential to even temporarily stabilize an inherently flawed new topology. Of course we usually don't want to operate a converter that depends on parasitics for its functioning. Though we do something similar when implementing ZVS ('zero voltage switching'). Parasitics often just 'soften' an abnormal or excessive application condition though we may not realize it (that's until we run PSpice!). As a trivial example, the ESR helps limit the inrush current into the input capacitor. We also know that even the small trace inductances leading to the input capacitor can help dramatically reduce shoot-through (cross-conduction) currents in synchronous buck converters.

Let us also consider what happens if we suddenly overload a normal nonsynchronous buck converter ... say by placing a dead short on the output? The duty cycle is still way up initially, not having had time to respond. So the converter 'thinks' the output voltage is still high and since its duty cycle is unchanged, it actually continues to try to deliver the normally required output voltage. But we know for a fact that the actual voltage on the output terminals has been forced to zero by the short. So to where did the excess Volts disappear? In fact, the full calculated output voltage momentarily appears across the diode and the dc resistance of the inductor. And the current must therefore increase (overload current) such that the following equation is satisfied unequivocally during the initial moments of the short: $\mathrm{Vd} + \mathrm{I} * \mathrm{DCR} = \mathrm{Vo}$, where Vd is the diode forward drop, and DCR the dc resistance of the inductor. So in a fault condition on the output, the DCR of the inductor and the diode drop actually both help in reducing the overload current. 'Good diodes' (with low forward drop) make the overload currents even higher. Note also, that in the latter case, it is not only the fact that we have a diode drop that helps reduce the overload, but the fact that this drop actually increases with increasing current, thus effectively helping out when needed the most. In fact, this effect was belatedly replicated by placing a series resistance (ohmic) term in the PSpice diode model.

Do write me at sanjayamaniktala@yahoo.com and copy Steve at sohr@cmp.com if you want to express convergence of views on this topic. But it's also OK, even if you want to momentarily explode and vent yourself after reading my little viewpoint. Just as long as we all manage to ultimately stabilize the resulting situation! And quite naturally so!

Why Does the Efficiency of My Flyback Nose-dive?

(December 2004)

Sanjaya's back with a confession: "I learned a lot at that little company I worked for way-back-when." What he learned is that flyback switchers have problems with leakage inductance which can come around to bite your tail. Also, a thing-or-two about PSpice, Andrew Lloyd Weber musicals, consultancies in the dot-com era, and the little Precision Advocate that lives inside us. A December smorgasbord, Sanjaya's most personal column yet. Enjoy!

Looking back at the six or seven power conversion companies I have worked in so far, I think it is somewhat intriguing to realize that professionals often learn (and contribute) the most in smaller companies. Just a plain coincidence that these companies are also often the ones engineers are least likely to ever want to admit having worked for, many years later! Call it a 'trial by fire' or 'hardening by heat treatment.' Whatever! I remember Bruce Carsten had written a last-page article in some forgotten publication over a decade ago titled "Why Innovations Seem to Come from Smaller Companies." A very pointed and thought-provoking article, one that I glanced over several times over the years. I just recently threw it away after cleaning up after completing my book.

This month however, I decided I was finally going to come to terms with my past in a way, and try to force myself to remember in vivid (and somewhat painful) detail how I learned to deal specifically with leakage inductance many years ago, while working in a rather small outfit making innovative integrated switcher ICs for off-line flyback applications. In doing so, it will also be clear to you, what really hampers the flyback topology itself, at higher output power levels and low output voltages.

But before I get into that, I will take stock of some of the interesting correspondence I received in response to last month's column. In particular I had an interesting Email exchange with Paul Tuinenga, co-founder of MicroSim (creators of PSpice, now part of Cadence). I will let Paul state it best in his own words: "I have this recollection of a passage from an out of print book (I think it was) 'Paper Money' by Adam Smith, in which the author tells of a conversation with a Southern stock-broker — 'The computer is like a dog. Very useful. Wouldn't think of hunting without one. They spot birds and retrieve. But you don't give the gun to the dog'!" Paul and I ended up being pretty much on the same page as he also agreed with my basic sentiment that "PSpice essentially includes all the equations… like say Kirchhoff's laws. So it does a great job in predicting the final outcome (usually… !). However, in using it, we engineers therefore tend to forget the actual

equations ourselves — which is a curse for any good engineer, since he loses the power of 'optimization' so essential to a good designer . . . he just forgets to think! Not the direct fault of the machine though. In a way, he gets blinded by the luxury of a powerful machine . . . which 'does' but can't 'think.' I feel he can certainly use it to AID design, not as a substitute for design."

Returning to the flyback topology, working in this particular switcher IC company, I had been trying to come up with a more accurate set of 'quick selection curves' for their next-generation switcher IC family. Their older Excel spreadsheet and the corresponding published efficiency curves were really not holding up very well in actual bench verification. Seemed to be finally giving credence to the constant griping by previous customers that the efficiency and max power curves were "unachievable" and "how the hell did you come up with these anyway?" Not that the company was initially really looking to correct these apparent inaccuracies, since when it got assigned to me, the decision was based mainly on some rather astute financial and business sentiments (I didn't say 'Machiavellian').

Actually this story itself is interesting enough to merit a slight detour here, since it keenly epitomizes the touch-and-feel of the entire dot-com era in the Silicon Valley area (while it lasted). The previous applications senior engineer (the one who had created the original Excel spreadsheet and thus indirectly ensured only he understood it fully) had quit suddenly. He had then thoughtfully set up a lucrative private consulting agency — but not before taking with him a whole lot of stock options from this flyback company ("high flying adored did you believe in your wildest moments, all this would be yours" - Evita).

The consulting activity actually was in parallel to a 'full-time' senior position that he also took up in a rather sleepy company (you guessed it, a 'big' one). I just don't know whether this latter company knew/didn't know/didn't care/thought it was perfectly OK, or even admired any human's ability to 'multiplex' so dramatically. I couldn't do it for sure! Oh yes, talented Mr. Ripley was also managing to teach evening EE classes at the local university in his 'spare time'?! Several years later, after this 'big' company got swallowed, and then re-swallowed successively, by bigger and bigger companies, it apparently just became too big even for Mr. Ripley, and so one fine day, along with the whole dot-com era, he too got a pink slip. But, till that transpired, he was more than willing to come back, again and again, to the previous (and precious) flyback IC company, generating all the efficiency and selection curves for every future product family they desired (essentially by magically reconfiguring his previous spreadsheet, as only he could).

Yes, he was certainly counting on making much more than a 40-hours-per-week exempt employee like me. However, company management may have been on to him, opening the door just a little for him (a few hours per week over a few months), perhaps with the unstated intention of transferring his expert knowledge back in-house, and afterward dumping him. All this while, with some help from him, I had successfully developed far

more elaborate Mathcad models which I then used to put out the quick-selection curves for two new product families. In the process I also showed the company what they needed to do to improve their efficiency estimates, especially in regard to leakage inductance. Job done, I too then quit suddenly. I probably left the company with the unenviable choice of either figuring out the older, flawed but simpler Excel program, or running with the more accurate and bench-verified, but horribly complex Mathcad program. I really don't know what they did after that, though I suspect they may have had to play footsie with the other engineer again, at least to make sense of the spreadsheet.

During the bench verification process, I noticed that for 12 V outputs, we got a nice fit with the theoretical efficiency estimates on that spreadsheet, but for lower voltage outputs (5 V, for example) the bench test verification went way off the efficiency estimate curves, especially for high loads. Why so? We went over every loss term, often with a microscope, both on the bench and in the program, refining the models more and more, almost to the extent of making the entire Mathcad file completely iterative. (It would take roughly 24–36 hours to generate the final efficiency curves from the moment I would hit 'calculate' on a 600 MHz PC. The company had to put three PC workstations in my cubical to ensure I would get *some* work done while the simulation was running!) But no luck getting the actual measured efficiency to match the simulation! We were certainly getting much better with each iteration, but still couldn't explain where the remaining couple of efficiency points were going.

I rechecked my program several times, dotting the i's and crossing the t's (the little German living inside of me ever since my Leipzig days) but it looked solid every way I attacked it. In desperation, I then started poring over some old literature looking for clues, and this one from a very old Philips publication "3C85 Handbook" caught my attention: "Leakage will reflect from secondary side to primary side according to square of turns ratio of the transformer." That was it! We all know it is standard design practice for any universal-input off-line flyback to keep the reflected output voltage 'V_{OR}' fixed at an ideal of about 105 V. The V_{OR} is the output voltage multiplied by the turns ratio. So basically, the required turns ratio for a 5 V output is

$$\frac{n_P}{n_S} = \frac{V_{OR}}{V_O} = \frac{105}{5} = 21$$

For a 12 V output the turns ratio is

$$\frac{n_P}{n_S} = \frac{V_{OR}}{V_O} = \frac{105}{12} = 8.75$$

The secondary side leakage (uncoupled) inductance is associated not only with the actual transformer windings, but the lead-out terminations and even the PCB traces leading to and

returning from the diode and output capacitor. Assume two inches of total secondary side trace length for example, and remembering the thumb rule of 20 nH/inch, we get say 40 nH secondary leakage inductance. For a 12 V output, this will reflect to the switch side as an effective leakage of $40 \times (8.75)^2 = 3062$ nH or about 3 μH. This will be added to the existing primary side leakage (typically about 10 μH) to give about 13 μH. The associated energy will be dissipated in the zener clamp. However for a 5 V output, the same 40 nH gets reflected as $40 \times (21)^2 = 17640$ nH or about 18 μH! This will give a total primary side leakage of $10 + 18 = 28$ μH! Triple the first estimate. Enough to inflict trauma on the wimpy zener clamp which probably was just expecting a nice sunny day paddling away on the beach, but got hit by a tsunami instead.

As soon as I understood this, and learned to correctly estimate and perform an 'in-PCB' measurement of effective leakage (not by just shorting the secondary side pins of the transformer, but by placing thick shorts across the diode and output cap), I modeled it into the Mathcad file. The fit to bench results was almost too good to be true - within 1% over the entire load range. After much more testing, by several engineers in fact, the company finally acknowledged this to be the missing piece of the Flyback jigsaw puzzle, and then published the new product family quick-selection curves I had generated and verified, and also the guidelines on leakage inductance measurement. Of course they didn't want to 'alarm' previous customers by going back and correcting the curves of the previous family (which they now knew were super-optimistic). Plain marketing sense!

That is all I could manage this month after a rather hearty and long Thanksgiving break, and that coming in just after a tiring 4-hour course on Magnetics I had to prepare for and present at Power Systems World 2004 (in which, incidentally, I had Keith Billings himself in the adjoining room giving a similar 4-hour course on hold-your-breath: Magnetics! What timing! Luckily (for me) we managed to split the attendees almost evenly. I walked in to say Hi to Keith too.

After suitable reflection, do feel free to write me at sanjayamaniktala@yahoo.com with any free-wheeling comments you may have. Just don't forget to copy Steve at sohr@cmp.com so he can keep an eagle eye on where this column is headed!

It's Not a Straight Line: Computing the Correct Drain to Source Resistance from V-I Curves

(January 2005)

Is the V-I curve of a MOSFET switch really a straight line as we imagined? The RDSON is clearly a function of the current through the MOSFET. But with the device alternating between peaks and valleys, what current value do we use? We can do a 'worst-case analysis', based on the highest RDSON (an instantaneous value) along the V-I curve. But is that value really 'worst-case', or is it even worse than 'worst-case'?! Power supply guru

Sanjaya Maniktala celebrates his anniversary on Planet Analog with some observations on the proper use of RDSON.

Performing an efficiency calculation for a power converter will certainly require knowledge of the drain to source on-resistance (hereby called R_{DSON}) of the switch. In doing so we may refer to the V-I characteristics of the said mosfet. We will probably be thinking that all we need do is to find the slope — "V/I" — to get the R_{DSON}.

But hold on just a minute! Is the V-I curve really a straight line as we imagined? If not, the R_{DSON} is clearly a function of the current through the mosfet. So what is the R_{DSON} we need to take for our calculation? We can do a "worst-case analysis" based on the highest R_{DSON} (slope) along the V-I curve, which would always be found to occur at the highest instantaneous value of current, that is, the peak switching current. But is that value really "worst case," or is it even *worse* than "worst-case"?!

The current through the mosfet is actually varying every cycle between two values, the peak and the trough. It is certainly not fixed at the peak value (at which we may be finding the "worst-case" slope). We are not interested in finding the worst-case *instantaneous* value of the R_{DSON}, what we want is the worst-case value over the *entire switching cycle*. In a switching converter, the R_{DSON} is actually varying smoothly between two values, just as the current is.

By a rather painful analysis, it can be shown that a very close fit to the exact integration-based calculation is obtained simply by (a) finding the R_{DSON} at the extreme current values: the peak and trough, and (b) averaging these two values to get the effective R_{DSON} over the entire cycle. Simple enough!

But hold on a minute longer. Look at the published V-I curve (the black part of *Figure A-3*). This represents a typical integrated switcher device, rated for 1.5A. The device is therefore supposed to function up to 100°C at 1.5 A. But does the curve extend all the way? Not at all! The 100°C is mysteriously truncated! No other information is available in the datasheet. The only way out for us as designers is to try to extrapolate the V-I curve. See gray part.

We now see that the curve intersects at a whopping drop of 17 V at 1.5 A at 100°C! Maybe that is what the vendor didn't want to circle out for us. But at least we can now find the effective R_{DSON}.

However, we could *erroneously* take the average over the *entire* range to get

$$RDS_{AVG} = \frac{17}{1.5} = 11.3\ \Omega$$

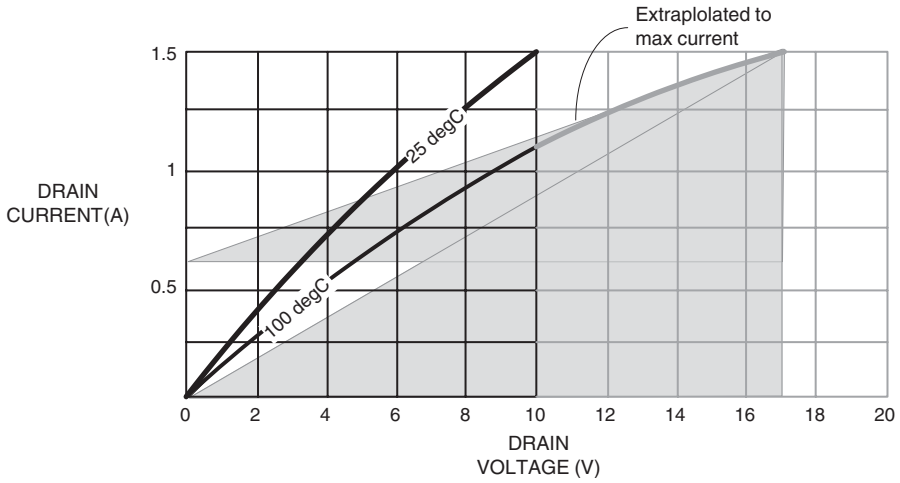


Figure A-3: The Missing Half of the Topswitch Datasheet Curve

This estimate is way too optimistic. As mentioned previously, a more correct estimate is the RDS averaged over the RDS at the extreme values. At peak value

$$\mathrm{RDS_{MAX}} = \frac{17}{1.5 - 0.625} = 19.4\ \Omega$$

We also know that the $R_{DSON-MIN}$ is 10 Ω from the datasheet since the R_{DSON} is rather typically stated at only $1/10^{th}$ the maximum current, that is, 10 Ω at 150 mA in this case.

So the correct effective R_{DSON} is the average of the two

$$\mathrm{RDS} = \frac{\mathrm{RDS_{MIN}} + \mathrm{RDS_{MAX}}}{2} = \frac{10 + 19.4}{2} = 14.7\ \Omega$$

Now that we know the RDS is 50% higher than what was indicated to us, we also have a better idea of the conduction loss in the mosfet.

Please drop me a note at sanjayamaniktala@yahoo.com and copy sohr@cmp.com. One year of this column is now complete . . . many thanks to you all.

Don't Have a Scope? Use a DMM, Dummy!

(February 2005)

Sanjaya's got a gremlin on one of his shoulders; a cherub on the other. The gremlin says, "You call yourself an engineer?! You'll never get this power supply to stabilize. Why don't you just quit?" The cherub says: "Back of

the envelope calculations can work. Look what I brought you from Micro Center. Put this little meter to work." Somewhere between computational theory and the simplest hands-on measurement, Sanjaya has a window on switching regulator efficiency. If only these voices would stop...

Ouch that hurts! Heaping insult on top of a latent injury, and then adding some disgrace for good measure too! And just as I was getting around to finally using the 'Math function' on my $3000 digital storage oscilloscope (DSO) to carefully measure the duty cycle of the switching regulator (which I may add, was strongly requested by you in the first place)... And now you come back to claim that the $9.99 Velleman digital multimeter (DMM) from Micro Center is enough to do the job! And even more accurately?! Hmm. Wise Guy! Now you know why I have recently started to just hate everything about power!! And you know what!? I think I am going to have to cut and run pretty soon!

Now wait a minute (the voice of reason), relax. Let's get some information first. Just the basic facts. I promise to ease your pain, get you on your feet again. But in return you must understand that we really can't afford to overlook the numbers. For that is where the subtle secrets of power conversion usually reside: in those very 'trivialities' that you were always taught to ignore. "Don't get bogged down with comfortably numbing details (if you ever want to amount to anything in life)," that other voice said! "Try to see the big picture." The DSO? You see that! What do you want with a DMM? Let's take the buck regulator for starters (the voice of reason replies). The simplified 'textbook equation' for duty cycle is $D = V_O/V_{in}$. So if we were say stepping down from 20 V to 5 V, the calculated duty cycle is $D = 5/20 = 0.25$. Suppose the load is 2 A. That's an output power of 5 V $\times$ 2 A = 10 W. Now, the center of the ramp portion (average) of the inductor current is always going to be fixed at 2 A for a buck topology, come what may. That follows from Kirchhoff's first law (and believe me, you can't afford to ever get on the wrong side of that guy!).

Then by simple waveform analysis, the average input current (which is the average switch current for this topology) can be calculated from the rectangular waveform of height 2 A with a duty cycle of 0.25. This gives an average input current of 2 A $\times$ 0.25 = 0.5 A. So the input power is 20 V $\times$ 0.5 A = 10 W. We can see that that is exactly the same as the output power, and therefore we are at 100% efficiency.

Well, what did you expect? (That voice again, the other one.) The very act of using the simplified 'textbook equation' for duty cycle was tantamount to assuming 100% efficiency. Where are the loss terms? We have clearly been guilty of ignoring the forward voltage drops across the switch, the diode, the inductor, the capacitor, besides assuming zero switching losses too!! In fact, as soon as we start to consider the non-ideal characteristics of real components, we will see that the inclusion of each non-ideality contributes a little to the increase in the duty cycle over its baseline 'textbook' value.

Which is precisely why we wanted to know the actual duty cycle in the first place. But without an oscilloscope with 'Math function' capability, are we really stuck? Let's think

again. We are expecting the average input current to be 0.5 A. Now let's measure it using the DMM. Suppose we read 0.6 A. Right off the bat, we now know the efficiency is (2 A × 5 V)/(0.6 A × 20 V) = 10/12 = 83.3%. Two extra watts are lost somewhere. That's right. An additional 0.1 A drawn from the input at 20 V is 20 V × .1 = 2 W. So that ties up. But for the buck topology, the center of the current waveform MUST still be the load current. Therefore the only way the average input current, as calculated from the rectangular waveform, can be 0.6 A, is if the duty cycle has increased to 0.6 A/2 A = 0.3. *So now we do know the duty cycle!* Look Ma, no DSO!

In fact, at this point we can declare completion of our initial assignment, that of using a DMM instead of a DSO to find out the actual duty cycle. But we can take this opportunity to delve a little deeper too. Suppose the diode drop is known to be 0.4 V. The actual duty cycle of the current through it is 1 − 0.3 = 0.7. The estimated loss in it is 2 A × 0.4 V × 0.7 = 0.56 W. However, we know that we are dissipating 2 W. But we have computed that only 0.56 W is being lost in the diode. That leaves 2 − 0.56 = 1.44 W still to be explained. We can keep going in this manner and try to account more and more accurately for all the wasted energy. But we can see that the actual duty cycle is the key to a full efficiency analysis.

It is also interesting to note that had we used our initial duty cycle estimate of 0.25 to calculate the diode loss, we would have estimated it to be 2 A × 0.4 V × 0.75 = 0.6 W. That is more than the 0.56 W calculated above. So in reality, losses tend to be lesser than we may have first thought. Why so? This is just another subtle example of how nature moves in mysterious ways to rescue us, as it tries to create conditions for natural processes to converge. Do see my Nov 2004 article on this topic. If we are interested in creating a mathematical model of our power supply in gory detail, we should realize that the entire mathematical process needs to be both iterative and convergent too! For example, as described in last month's column, the voltage drop across the switch is a function of current. So suppose the switch voltage drop is a little higher than first estimated, for any reason, like say a higher ambient temperature. Now the switch would dissipate more of the incoming energy (conduction loss term). Therefore, the converter, in an effort to keep delivering the required output power will have to increase the input power — and therefore the input current. But this will cause the switch voltage drop to increase further too, and so the input current will have to increase a little more. So on and so forth, till mathematical convergence occurs. Nature converges similarly, only *much, much faster!* Like almost at the speed of light!?

The number crunching is actually quite different for the non-buck topologies. Try to work through a numerical analysis for any other topology. And if you don't get it easily enough, do feel free to write me at sanjaya.maniktala@nsc.com and I will take it up in next month's column (but definitely copy sohr@cmp.com if you want real results). In fact, this month's column was itself inspired by a rather persistent but nice caller called 'Jag' from Sunnyvale, California. He was asking me just such a question, based on the boost topology. Actually, nowadays I do get several Emails and calls from engineers from different parts of the world, seeking advice or help to solve a troublesome and usually non part-specific technical problem. And I am always glad to help, provided of course I have the time and acumen to think it through, and 'get it.' Though sometimes I admit, in my case too, my lips move, but you won't hear what I am saying! Good Luck.

Are We Making Light of Electronic Ballasts?

(March 2005)

Way to go, Sanjaya! Rave reviews of Sanjaya Maniktala's book on switching power supply design identify it as the new "bible," the absolute definitive text on this subject. Planet Analog is proud to have "discovered" him, to have published excerpts, and invites you to read his current column. This one is on the reliability of electronic lighting ballasts in India, where the wiring is frayed, the temperature sizzles, and ownership of fluorescent tubes constitutes bragging rights. Afterward, we'll send you to Amazon to see some of those terrific reviews. First, click here . . .

"Power conversion" is a conversion from one form of energy to another. It doesn't preclude the conversion of input energy into light, instead of more conventional load profiles that power supply engineers are generally accustomed to. In fact, I found out the hard way how difficult this area of power conversion can really be. Far more difficult than your average dc-dc converter!

My first exposure to electronic ballasts was, it now seems, light years ago (no pun intended), while I was working in the swank and well-equipped central R&D labs in Bombay of one of India's largest electrical manufacturers. That company is to India, probably what Siemens is to Germany — with a huge and diversified market share in almost everything 'electrical.'

India, of all the hot, confusing, bustling places on earth, may actually form the best place to see what the brouhaha about ballasts is all about. It is the perfect testing ground (and potential graveyard) for all electronic ballasts, big and small, the ultimate leveler if ever there was one.

We are aware that electronic ballasts are universally known to provide more stress-free working environments, attributed to their steady flicker-free light, besides other nice features like instant-start, the more recent auto-dimming capability, and power-factor correction. Though electronic ballasts have been available for a dozen years, it is very surprising that sales have barely started to exceed copper-wound (or 'magnetic') ballasts, which are known to be physically bulkier, heavier, and much less energy efficient. (Touch one and you will surely scream 'yeeeooowwaaaaaasste.')

True, electronic ballasts cost more, but you are supposed to get paid back handsomely in a few years in reduced energy costs. Some governments also continue to contemplate subsidies to help consumers afford such ballasts. But I don't think it has happened yet.

In a typical office environment, the light bills can amount to 40% of the total energy costs. There is great need to conserve energy in lighting, much as we are doing with standby power requirements of appliances. Community-conscious and inherently progressive organizations, like my employer for example, have installed PIR (passive infra-red, i.e. body-heat) sensors almost everywhere to turn the lights on and off as needed. I suspect they also use electronic ballasts. Remember, PIR sensors do help at night, but clearly can't save energy on their own during the day. For that electronic ballasts must be used in conjunction.

(As an aside, we may not conserve THAT much energy if engineers like me continue to leave their computers and monitors on the whole day, night, through weekends, and on long breaks, spicing up their screensavers with graphic-intensive witticisms like "go away," "scat," or simply "attending a meeting" refreshing over and over again. I had heard of endless meetings, but a 10-day blinking message can really take the cake).

California, with its high energy costs is certainly waking up to the potential savings from electronic ballasts this year. See a very useful discussion of ballast sales and impending regulations starting 2005 with the papers at http://www.aboutlightingcontrols.org/education/papers/ballasts.shtml. Another very nice piece on the myths associated with ballasts and tubes can be found at http://www.energyusernews.com/CDA/Article_Information/Fundamentals_Item/0,2637,119815,00.html.

Note that the fluorescent tube also responds rather positively, in terms of its own life and performance, when driven at over 20 kHz (as in an electronic ballast), rather than with 60 Hz (as in magnetic ballasts). Tube-replacement costs are thus significantly reduced with electronic ballasts. But, wait! What about the life of the ballast itself? That is another story altogether!

Electronic ballasts unfortunately have been plagued by failures (see http://lightingdesignlab.com/commercial/articles/Energy_Efficiency.htm). So how do we as traditional power supply engineers provoke incipient failures in power supplies? We increase the stress levels, especially during design phases. Simple burn-ins are actually too kind to the system!
To demonstrate an exaggerated burn-in situation, let's take an 'Amazing Race' detour, and parachute down directly into a remote residential area somewhere in the heartland of India.

Oh I know you feel suddenly out-of place! Here time stands still. Well, almost! After a moment's rest, you can probably observe that the AC utility lines and the frayed household wiring haven't been replaced for several decades now. But worse! Here we happen to be ensconced by several industrial units using heavy electrical machinery, even during the night.

The mains input in India is officially supposed to be 230 VAC or is it 220 V or 240 V? Even I will never know. In fact, that may be a rather redundant question, especially here, because the voltage is known to drop down to a steady level of almost 120 VAC (no typo here: one hundred and twenty!). More so in the summer months (9 of these in India) when all the fans and air-circulators try to come on at the same time. The standard household incandescent bulb is now seen glowing faintly in the distance, too dim to even see where the TV and fridge lie gasping for breath.

So it may confer minor bragging rights with inquisitive neighbors to possess fluorescent tubes in every room of the house! The sage had mentioned these tubes provide a more acceptable illumination level even at such low input voltages. Provided they work!! To get them to work, magnetic ballasts are literally powerless. An electronic ballast can work in

principle, since it is basically the flyback (buck-boost) topology principle. Aha! Out with the sage!... I now see a power conversion engineer firmly entrenched here too! And this poor guy has to ensure his circuit design can get the tube to fire, and continue to run at such a low input voltage.

However, we can't be in the business of designing and selling one ballast design for one area and another circuit design for another area or locality. So now let's take the same ballast and move inside an industrial facility (with better local wiring). We are not surprised to find this is where the maximum usage of fluorescent lighting occurs. Rows upon rows of multi-tube ceiling fixtures, as in all countries. Sounds familiar and encouraging. I too feel almost at home now.

But now we sadly note that the utilities often raise the voltage at the substation end, just to compensate for some arbitrary calculated ohmic drops across miles and miles of lines. But suppose we happen to be the unlucky business unit sitting 'up close and personal' to the actual distributing substation. The intervening ohmic drops are thus virtually negligible. What we get coming into our facility 24/7/365 is a steady overvoltage of over 270 VAC!

But now if we dare to use heavy industrial machinery on our premises, we also know that that can unleash huge inductive spikes back into the mains coinciding with the solenoids and motors turning off. As any typical ballast manufacturer in much of the third-world (and Eastern European, especially the former Soviet bloc), we are faced with the daunting design task of ensuring rock-solid reliability under steady voltage variations of about 100–300 VAC, overlaid with huge spikes.

In fact the relevant qualifying test is usually based on keeping the ballast in operation, and simultaneously applying at the input, the well-known *8/20 μs lightning surge test*. No short-term or long-term damage should ever occur. Note that these line surges are of frequent occurrence in these areas (rather than a rare 'one in a blue-moon' type of thing), so we just cannot rely on MOV's (metal oxide varistors) which have inherent lifetime/wearout issues.

Nor can we rely on TVS's (silicon transient voltage suppressors) because the latter often can't even handle the type of energy a single 8/20 μs spike throws at them, least of all a succession of spikes at a certain constant rate per minute.

But that's not all! Any ballast in the world must be able to survive the 'deactivated tube' test. That is where the gas has leaked out, but the heating filaments are still present, and so the circuit keeps trying to start the tube endlessly (at the elevated voltage and frequency needed to cause it to strike). In fact, this particular test killed every ballast we ever tested in Bombay with no exception (except the one I finally designed! You knew I would say that, but it's true!). We burned out every known name-brand ballast we actually imported at that time

from the USA, Europe, Singapore, Korea, Japan (you-name-it) into India. We personally performed the last rites. And it was a virtual crematorium.

So India is possibly a great place to hone the design skills for any mains input ('off-line') power conversion device. Remember that when you set up a design center in Bangalore! This should help take the sting out of it. Next month will get into the nitty-gritty technicalities, and tell you exactly how the simple electronic ballast actually works, and how we ended up enhancing the reliability besides reducing the manufacturing cost by a factor of almost 2 in the course of what was probably the most successful R&D technology transfer project in that company. Not that they remember me anymore!

Don't forget to write me at sanjayamaniktala@yahoo.com and do copy Steve at sohr@cmp.com, if you don't want your energy to be wasted! And all power to you!

Editors' Note:

Sanjaya's book, Switching Power Supply Design & Optimization*, is published by McGraw-Hill. User reviews at Amazon.com identify it as the new "bible" on switching power supply design."*

More on Designing Reliable Electronic Ballasts

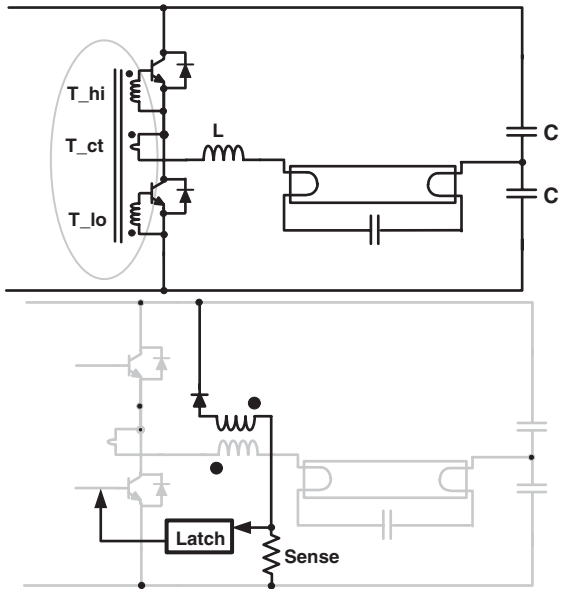
Long, long ago, in a land far away (actually, not so far away with the Internet and 800 numbers), our hero tried to figure out how to keep electronic ballasts from blowing up. The key, writes Sanjaya in this installment of his power design column, is in the ferrite inductor in series with the lighting tube. While its basic purpose is to limit the current, there is a good deal of resonant frequency energy coming off-line. Your choice to use it for good or evil.

This month I need to fulfill the promise I made about explaining what all we did with electronic ballast technology in India, thousands of years ago (or so it seems).

The most common (and commercially viable) fluorescent ballasts still use bipolar transistors (BJTs), not mosfets. They are also self-oscillating, and therefore need no PWM control IC. This is actually an advantage. Engineers who have worked with self-oscillating topologies (like the well-known Royer oscillator) know not to underestimate them. They are inherently self-protecting and tend to be very rugged. Short their outputs and the frequency automatically adjusts itself to maintain critical conduction mode, so there is no staircasing of the current or magnetic flux in the inductor.

In the electronic ballast too, there is a ferrite inductor in series with the tube. Its basic purpose is to limit the current, as in conventional copper ballasts. The difference being that copper ballast needs a large line-frequency choke made of iron/steel, whereas the electronic ballast is much smaller, lighter, and made of ferrite material. The size advantage is exactly what we expect and invariably achieve by creating switching action at a very high frequency. However, one basic difference as compared to a conventional half-bridge switching power supply (which the electronic ballast shown in *Figure A-4* resembles) is that the ballast is

Figure A-4: The Basic Electronic Ballast and the Improvements



actually a resonant topology. The L in series with the tube forms a series-resonant circuit with the two C's of the half-bridge (which are effectively in parallel from the AC point of view). So the current sloshes back and forth, and it makes perfect sense to therefore make the circuit self-oscillate, with the help of base-drive transformers as shown in the oval circle.

Note that when the tube has not fired, that is, when we first apply AC input, the small unnamed capacitor across the tube is the effective series capacitance of the resonant tank circuit. It is the starter cap. The oscillations are in fact at even a higher frequency than in normal operation on startup. But what's more, since there is almost no damping resistance during startup, a high voltage is created across the tube, to get it to fire. We must remember from our high-school EM course that a series LC presents a very low input impedance at its resonant frequency, and at that moment, the voltages across the L and the C can both be very high, though opposite in phase, thereby effectively canceling out as far as the input is concerned.

This is just the opposite as compared to a textbook parallel LC tank circuit, which presents a very high impedance at its resonant frequency, and in which the currents in each component can be very high, though opposite in phase, thus effectively canceling out as far as the input is concerned. We also remember that if we drive such a low-impedance series LC tank circuit with the driving frequency equal to its natural resonant frequency (which is what we really do by using a self-oscillatory scheme), the oscillations build up every cycle, and so, though the input voltage remains the same, the currents and the voltages across each reactive component keep building up every cycle. Finally, the tube 'fires,' thus effectively bypassing the small

starter capacitance. Thereafter, the circuit lapses into a more stable, damped, and lower frequency oscillation based on the resonant frequency created by the C's of the half-bridge.

We can clearly see one major problem already. That is, what if the tube does not fire? This is a real-world possibility, since the seals at the ends of the tube may leak, thus affecting the 'vacuum' inside the tube over a period of time. In this situation, we are expecting to replace the tube, not the ballast! But in a virtually undamped LC circuit, the oscillations will build up every cycle, and eventually the transistors, which see the same current when they turn on, will be destroyed. This is what leads to the 'deactivated tube' test. The tube does not fire and the filaments at the end of the tube are typically of such low resistance, that they really can't damp out the steadily escalating oscillations. Some engineers therefore try to place an additional resistor in series with the small starter capacitance, but this certainly affects the ability to start the tube, especially at lower mains input voltages.

A PTC (a thermistor with a positive temperature coefficient) can be used, but it is an expensive solution and also has response-time limitations. In the case of the existing ballast design (just before we set to work on it), the previous engineers had tried to circumvent the deactivated tube failures by using more expensive and hefty 'horizontal deflection' transistors (the well-known BU508A). But these have low gain, and they run inefficiently and get hot. So now heatsinks had to be added. In addition, there was still a need to turn off the ballast after several such unsuccessful attempts to fire the tube. So in came an expensive mechanical thermal overload relay, fastened to the heatsinks. But then they found out that it just failed to act fast enough to protect the transistors, given the high heatsink thermal capacity involved.

Our contribution was to use the principle of the flyback converter to recover energy back from the inductor. See the second *Figure A-5*. So an additional winding (with turns ratio and polarity carefully worked out) recovered the excess energy and delivered it back to the main input bulk capacitor. But to eventually cease the switching operation, a sense resistor (or diode in our case) was added that would charge up a capacitor and eventually trigger a small NPN-PNP latch sitting on the base of the lower transistor. Now mains resetting would be required to make the ballast try again. Fail-safe really.

Another critical area of improvement was in the base drive. Historically, many vendors use a single-inductor approach (see first schematic in *Figure A-5*). However, we should remember that the key to turning ON a bjt efficiently is a little different from the way we want to turn it OFF. In particular, the turn-on must be a little slower (delayed) and it has been shown that the actual crossover duration is significantly reduced if in fact we don't do a hard turn-on. On the other hand, for turn-off we do want to create a hard turn-off, yanking the base momentarily, several Volts below the emitter voltage.

The problem with the single-inductor drive is that the turn-on waveform of one transistor is the exact inverse of the turn-off waveform of the other. So there is no possibility of driving

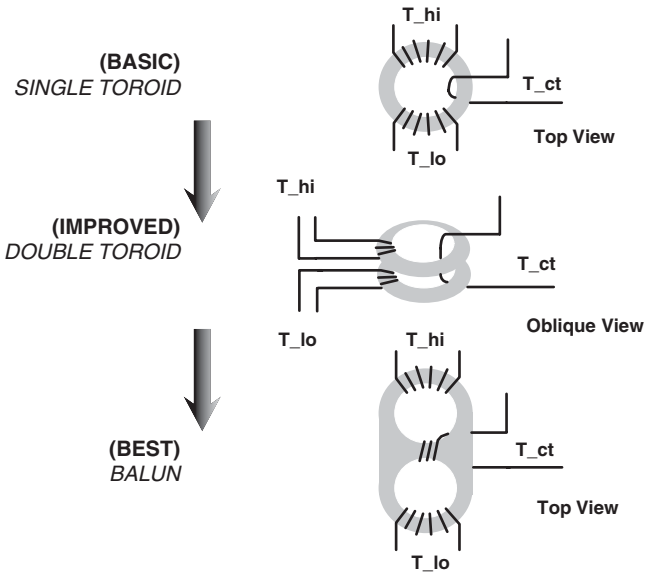
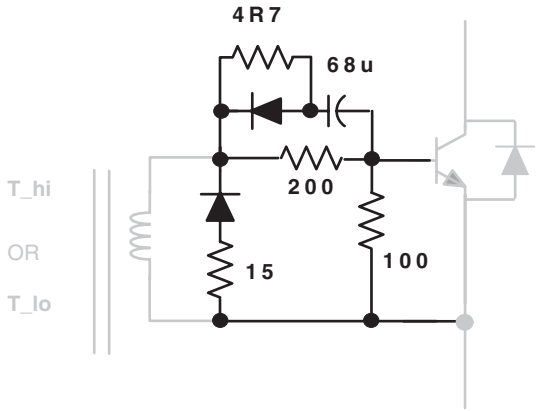


Figure A-5: Different Toroidal-type Drive Transformers

them appropriately and differently, and thereby efficiently. The transistors can thus run very hot with the single-inductor base drive. In *Figure A-5*, we have 'hi' for the high-side transistor, 'lo' for the low-side, and 'sw' stands for the primary winding, that is, the loop of wire passing through the base-drive current transformer from the switching node. To conquer the limitations of the single-inductor drive, engineers often use the double-toroid approach. However, if the permeabilities and dimensions of the two toroids are not well-matched, there are again discrepancies during the crossover, resulting in losses. But realizing that the actual permeabilities of the two toroids are not really important, but their relative permeability is, we started using an innovative 'balun' core to drive the transistors. The advantage is that 'both halves' of the balun are created in the same batch, so though the permeability may have a lot of process tolerance, the two halves are still very well-matched. Besides, since the two halves possess uncoupled inductance, that allows us to create the appropriate turn-on and turn-off waveforms by a little 'wave-shaping' circuit as shown in *Figure A-6*. Baluns, usually made for RF suppression, and on Ni-Zn ferrite, can be made to order with the more preferred Mn-Zn material. Then they require lesser turns and run very cool themselves too.

With these improvements, we could do away with the heatsinks altogether. The transistors would now run cool-to-the-touch, even while free-standing. No thermal trip was required. The transistors could also now be the cheaper and higher gain MJE13005. As for the input surge test, we replaced the filter with a smaller differential mode toroidal filter deliberately made of lossier Ni-Zn ferrite material, and simultaneously increased the input bulk

Figure A-6: Base-drive Enhancement Circuit



capacitance slightly, for that is the only way to really pass the lightening surge test without MOVs, and so on.

Last, I will pass along an interesting e-mail I received from Pat Rossiter in Denver, describing that such reliability problems exist not just in India, but right here in the US too. He writes:

> *Dear Sanjaya / Mr.Maniktala [please circle correct choice]: I read with great interest your article in today's Planet Analog Newsletter. I'm not as technically versed as I'd like to be, but it looks like the ballast that you invented (way to GO!) would be just the ticket for our lights here at Yellow Cab. We are a stone's throw from a sub station and our power fluctuates a bit and in the past month I have replaced 5 ballast packs. The point of this email–and it is certainly about time that I got to it–is where can I get the ballast packs you described? Eagerly awaiting your illumination on the matter.*

With that I put my pen down again for this month. Hope to hear from you at sanjayamaniktala@yahoo.com. But do copy the person who started this column off: Steve Ohr at sohr@cmp.com. Thanks a lot!!

The Organizational Side of Power Management: One Engineer's Perspective

Even with jobs going overseas, the age-old conflict between engineering and marketing has not diminished. In this departure from diodes, mosfets, and loop currents, Sanjaya looks at the organizational side of power management projects. The vocal, technology-grounded engineer may too often be the unwanted child in a "King's Court," he worries.

Talking endlessly about quaint base-drive toroids, ticklish lifetime issues with aluminum capacitors, neat though quirky current-sensing techniques, and what have you, has only one purpose in mind: that of advancing technology. But do we, as engineers, usually solidly devoted to our craft, really manage to achieve that to the full extent we hoped and struggled for? Or is it a case of "one step forward" and then "two steps backward" (for reasons beyond

our control)? A glass that at best remains tantalizingly half-empty and half-full — all because of a darn leak, that we just forgot to factor into our calculations!

But what about the organizational side? Advances in science and technology, especially in power conversion, hinge on a few basic and fairly obvious commonsensical principles. But surprisingly, these are the very ones often overlooked at an organizational level. Let me try to categorize some of my personal observations here — see if you agree:

Communication

A lot can be learned by simply sharing experiences: things that worked *and things that didn't.* Why should engineers always end up repeating mistakes and learning the hard way — potential bugs or possible catastrophes that they could've known about beforehand, just by listening and thinking. Couldn't they also take what in fact was already proven to be the best available engineering solution at that time, and develop it further? Creativity has its place, but let's not re-invent the wheel please!

However, when organizations grow, they usually start subdividing, and fairly quite arbitrarily too. So the Power team becomes separate "AC/DC" and "DC/DC" groups. Whereas we all know that at the heart of any ac/dc switcher is none other than a dc/dc switcher! At a later stage, the DC/DC group may become "Portable Power" and "Power Management" groups. But both still use the same topologies duh! Then sooner or later, Power Management may bifurcate into "high power" and "low power." Does all this imply any radical change in *engineering* principles? Not really! So the end result is that engineers, the ones that are expected to generate products and revenue in the first place, simply don't run into each other anymore, or get to talk about their experiences to each other, even over a coffee machine.

What's worse: if the assignments are on the unimaginative basis of "one project, one engineer," then even within any such finely divided sub-group, there is almost no sharing of engineering information thereafter.

Yes, I agree: marketing or sales or even Field Application engineers may need to be divided to get more "business focus," but for engineering focus, you actually need to get the engineers *together*, not drive them apart. Engineers always thrive when they share. No one really benefits in the long run if the pinky, for example, no longer knows what even the middle finger is up to. And power conversion is just too tricky an area to take that chance.

Integrity

Engineers are trained to respect only facts and data! That's the key to their success as engineers: standing behind every robust and brilliant product they create. Unfortunately, that strength also sometimes isolates them from the rest of the crowd. Of course there are people

whose legitimate job is to slightly blur the boundaries between fact and fiction — to create nebulous and fuzzy perceptions in others minds. Sales, Marketing, PR for example?

But conflicts over integrity can turn out to be quite debilitating to an engineer in the long run. Like what to write in a datasheet. Or what can or can't be designed and how soon. Or "promote this part number please… I don't care what you are writing in your App Note if you don't manage to sell my part first and foremost!" In fact, I know a company where the entire Applications Engineering department reports to the person who doubles over as the Marketing manager too. That may not have been such a bad idea provided that person was *once* an engineer, or at least had a thorough grasp of technical details. But that is usually not so. So one can easily envisage a situation where the Marketing person promises the customer the moon and the stars (all at daybreak tomorrow!), and then goes back to drive his engineers to build his (already committed) palace of dreams overnight. Or else!

Resources

Engineers relish challenges. Changes to their daily routine excite them. Do not hand them the humdrum job of simple repetitive power-up testing of 50 odd boards for a customer. That kills their spirit. Yet how often have we seen that, come layoff time, the first persons to go are the CAD guys, then the technicians, and then the documentation expert! On paper, the concerned manager can show impressive savings to his superiors. Headcount is what it is all about! But a few months later, engineers are still trying to grapple with ORCAD or Protel just to do their simple PCBs.

And in addition, they have to learn a rather complicated language whose only purpose is to transcribe what they already know and have written, into a format suitable for the datasheet standard used by the company. And yes! Those damn 50 boards are still waiting on the bench! Why couldn't these engineers have been doing what they are best trained to do, and also enjoy the most — you guessed it: engineering!? In fact right now, they may have just become the most highly paid technicians around.

Peer Environment

Technology may never gain a foothold in a "king's court," where you are either rewarded with largesse for being vehemently agreeable, or unceremoniously sentenced to the dark dungeons for the rest of your life. Engineers like to speak out — but usually only when they are sure of their facts and have incontrovertible data to back themselves up. They therefore deserve and need a "peer environment," where they are judged (primarily) by the respect received from their peers, the king be damned (on occasion)!

It must be kept in mind that this can really bother the king sometimes! So managers who supervise engineers should be fairly competent at a technical level themselves and respect

data and facts equally. They can't attempt to win a technical argument by throwing rank on their subordinates. Nor should they ever go around, God forbid, trying to subsequently shoot the "emotional and/or disrespectful" engineer down ("that'll teach him"). Surprisingly that does happen more than we dare admit. Not only does the good engineer pay the price, but so does technology in the long run.

Do write me at sanjayamaniktala@yahoo.com. If you think I'm being a busybody, don't hesitate to tell Steve at sohr@cmp.com to ask me to cut these musings out — and get back to mosfets and diodes please!! Alternately, voice your opinion (albeit, multiple choice) on the Planet Analog home page.

Final Note: What better way is there to prove this than by personal experience! This was the only article not sponsored under the Author Encouragement Program of my erstwhile analog chipmaker company. But at least their Power marketing director (my previous Apps Manager cum Boss) finally did read the first article of my series!

This Page Intentionally Left Blank

APPENDIX

2

Reference Design Table

This Page Intentionally Left Blank

APPENDIX 2

Reference Design Table

	Buck	Boost	Buck-Boost
Duty Cycle	$\frac{V_O + V_D}{V_{IN} - V_{SW} + V_D}$	$\frac{V_O - V_{IN} + V_D}{V_O - V_{SW} + V_D}$	$\frac{V_O + V_D}{V_{IN} + V_O - V_{SW} + V_D}$
V_{IN_50} (V)	$(2 \cdot V_O) + V_{SW} + V_D \approx 2 \cdot V_O$	$\frac{1}{2} \cdot [V_O + V_{SW} + V_D] \approx V_O/_2$	$V_O + V_{SW} + V_D \approx V_O$
Output Voltage, V_O (V)	$V_{IN} \cdot D - V_{SW} \cdot D - V_D \cdot (1 - D)$	$\frac{V_{IN} - V_{SW} \cdot D - V_D \cdot (1-D)}{1-D}$	$\frac{V_{IN} \cdot D - V_{SW} \cdot D - V_D \cdot (1-D)}{1-D}$
Voltμseconds (Vμs)	$\frac{V_O + V_D}{f} \cdot (1-D) \cdot 10^6$	$\frac{V_O - V_{SW} + V_D}{f} \cdot D \cdot (1-D) \cdot 10^6$	$\frac{V_O + V_D}{f} \cdot (1-D) \cdot 10^6$
L (μH)	$\frac{V_O + V_D}{I_O \cdot r \cdot f} \cdot (1 - D) \cdot 10^6$	$\frac{V_O - V_{SW} + V_D}{I_O \cdot r \cdot f} \cdot D \cdot (1-D)^2 \cdot 10^6$	$\frac{V_O + V_D}{I_O \cdot r \cdot f} \cdot (1 - D)^2 \cdot 10^6$
Inductor Current Ripple Ratio 'r'	$\frac{V_O + V_D}{I_O \cdot L \cdot f} \cdot (1 - D) \cdot 10^6$	$\frac{V_O - V_{SW} + V_D}{I_O \cdot L \cdot f} \cdot D \cdot (1-D)^2 \cdot 10^6$	$\frac{V_O + V_D}{I_O \cdot L \cdot f} \cdot (1 - D)^2 \cdot 10^6$
ΔI_L (A)	$\frac{V_O + V_D}{L \cdot f} \cdot (1 - D) \cdot 10^6$	$\frac{V_O - V_{SW} + V_D}{L \cdot f} \cdot D \cdot (1-D) \cdot 10^6$	$\frac{V_O + V_D}{L \cdot f} \cdot (1 - D) \cdot 10^6$
RMS Current in Input Cap (A)	$I_O \cdot \sqrt{D \cdot \left[1 - D + \frac{r^2}{12}\right]}$	$\frac{I_O}{1 - D} \cdot \frac{r}{\sqrt{12}}$	$\frac{I_O}{1 - D} \cdot \sqrt{D \cdot \left[1 - D + \frac{r^2}{12}\right]}$
I_{PP} in Input Capacitor (A)	$I_O \cdot \left[1 + \frac{r}{2}\right]$	$\frac{I_O \cdot r}{1 - D}$	$\frac{I_O}{1 - D} \cdot \left[1 + \frac{r}{2}\right]$
RMS Current in Output Cap (A)	$I_O \cdot \frac{r}{\sqrt{12}}$	$I_O \cdot \sqrt{\frac{D + \frac{r^2}{12}}{1 - D}}$	$I_O \cdot \sqrt{\frac{D + \frac{r^2}{12}}{1 - D}}$
I_{PP} in Output Capacitor (A)	$I_O \cdot r$	$\frac{I_O}{1 - D} \cdot \left[1 + \frac{r}{2}\right]$	$\frac{I_O}{1 - D} \cdot \left[1 + \frac{r}{2}\right]$
Energy Handling Capability (μJoules)	$\frac{I_O \cdot V\mu s}{8} \cdot \left[r \cdot \left(\frac{2}{r} + 1\right)^2\right]$	$\frac{I_O \cdot V\mu s}{8 \cdot (1 - D)} \cdot \left[r \cdot \left(\frac{2}{r} + 1\right)^2\right]$	$\frac{I_O \cdot V\mu s}{8 \cdot (1 - D)} \cdot \left[r \cdot \left(\frac{2}{r} + 1\right)^2\right]$
RMS Current in Inductor (A)	$I_O \cdot \sqrt{1 + \frac{r^2}{12}}$	$\frac{I_O}{1 - D} \cdot \sqrt{1 + \frac{r^2}{12}}$	$\frac{I_O}{1 - D} \cdot \sqrt{1 + \frac{r^2}{12}}$
Average Current in Inductor (A)	I_O	$\frac{I_O}{1 - D}$	$\frac{I_O}{1 - D}$

	Buck	**Boost**	**Buck-Boost**
RMS Current in Switch (A)	$I_O \cdot \sqrt{D \cdot \left[1 + \frac{r^2}{12}\right]}$	$\frac{I_O}{1-D} \cdot \sqrt{D \cdot \left[1 + \frac{r^2}{12}\right]}$	$\frac{I_O}{1-D} \cdot \sqrt{D \cdot \left[1 + \frac{r^2}{12}\right]}$
Peak Current Switch/Diode/ Inductor (A)	$I_O \cdot \left[1 + \frac{r}{2}\right]$	$\frac{I_O}{1-D} \cdot \left[1 + \frac{r}{2}\right]$	$\frac{I_O}{1-D} \cdot \left[1 + \frac{r}{2}\right]$
Average Current in Switch (A)	$I_O \cdot D$	$I_O \cdot \frac{D}{1-D}$	$I_O \cdot \frac{D}{1-D}$
Average Current in Diode (A)	$I_O \cdot (1-D)$	I_O	I_O
Average Input Current (A)	$I_O \cdot D$	$\frac{I_O}{1-D}$	$I_O \cdot \frac{D}{1-D}$
Output Voltage Ripple (±mV)*	$1/2 \cdot I_O \cdot r \cdot ESR(m\Omega)$	$1/2 \cdot \frac{I_O}{1-D} \cdot \left[1 + \frac{r}{2}\right] \cdot ESR(m\Omega)$	$1/2 \cdot \frac{I_O}{1-D} \cdot \left[1 + \frac{r}{2}\right] \cdot ESR(m\Omega)$
$r = \Delta I_L / I_L$, L in µH, f in Hz, All voltages and currents are magnitudes * ESL ignored $V_{(IN\_50)}$ is the input voltage at which $D = 0.5$			

References

[1] Nave, M. J.; "The Effect of Duty Cycle on SMPS Common Mode Emissions: Theory and Experiment", IEEE 1989 National Symposium on 23-25 May, 1989

[2] Cochrane, D.; Chen, D. Y.; Boroyevic, D.; "Passive Cancellation of Common-Mode Noise in Power Electronic Circuits", IEEE Transactions on Power Electronics, Volume 18, Issue 3, May 2003

[3] Qu, S.; Chen, D. Y.; "Mixed-Mode EMI Noise and Its Implications to Filter Design in Offline Switching Power Supplies", Applied Power Electronics Conference and Exposition, 2000, Fifteenth Annual IEEE, Volume 2, 6-10 Feb. 2000

[4] "Mounting Considerations For Power Semiconductors", On Semiconductor Application Note AN1040/D, May 2001-Rev. 3

[5] Mardiguian, M.; "Controlling Radiated Emissions by Design", Chapman & Hall, ISBN 0442009496

[6] Mardiguian, M.; "How To Control Electrical Noise", 2nd Edition, 1983, Don White Consultants, Inc., State Route #625, P.O. Box D, Gainesville, Virginia 22065, USA

[7] Hayt, H. W. JR.; "Engineering Electromagnetics", Fourth Edition, McGraw-Hill Book Company, ISBN 0070273952

[8] Collett, P. C. E.; "Investigations into Aspects Affecting the Design of Mains Filters for Frequencies in the Range 10kHz-30MHz", ERA Report No. 82-145R, 1983, ERA Technology Ltd., Cleeve Road, Leatherhead, Surrey KT22 7SA, England

[9] "Capacitors for RFI Suppression of the AC Line: Basic Facts", Fourth Edition, Evox-Rifa Application Notes, Evox-Rifa Inc., 300 Tri-State International, Su. 375, Lincolnshire, IL 60069, USA

[10] "Conducted Emission Performance of Ericsson DC/DC power modules: Characterization and System Design", Ericsson Design Note 009, April 2000, Ericsson Microelectronics AB

[11] Ott, H. W.; "Noise Reduction Techniques in Electronic Systems", Second Edition, 1987, John Wiley & Sons, ISBN 0471850683

[12] Ott, H. W.; "Understanding and Controlling Common-Mode Emissions in High-Power Electronics", Applied Power Electronics Conference and Exposition, 2002

[13] Basso, C.; "Conducted EMI Filter Design for the NCP1200", On Semiconductor Application Note AND8032/D

[14] Armstrong, K.; Williams, T.; "EMC Testing", Parts 1 through 6; Cherry Clough Consultants and Elmac Services, UK
[15] Bergh, K.; "CISPR 22 Telecom Ports", NEMKO Seminar, 2001
[16] "EMC of Monitors", Philips Semiconductors Application Note AN 00038
[17] "EMI Testing Fundamentals", Steward Technical Information
[18] Savino, S. E.; Suranyi, G. G.; "Application Guidelines for On-Board Power Converters", Tyco Electronics Application Note, June 1997
[19] "Input System Instability", Synqor Application Note PQ-00-05-01 Rev.01-5/16/00
[20] Collett, P. C. E.; "Investigations into Aspects Affecting the Design of Mains Filters for Frequencies in the Range 10kHz-30MHz", ERA Report No. 82-145R, 1983, ERA Technology Ltd., Cleeve Road, Leatherhead, Surrey KT22 7SA, England
[21] "Capacitors for RFI Suppression of the AC Line: Basic Facts", Fourth Edition, Evox-Rifa Application Notes, Evox-Rifa Inc., 300 Tri-State International, Su. 375, Lincolnshire, IL 60069, USA
[22] Snelling, E. C.; "Soft Ferrites, Properties and Applications", Second Edition, ISBN 0408027606; Butterworths & Co.
[23] "Power Factor Corrector, Application Manual", 1st Edition, October 1995; SGS-Thomson Microelectronics
[24] "Data Handbook, Aluminum Electrolytic Capacitors", PA01-A, 1993 N.A. Edition; Philips Components
[25] "Understanding Aluminum Electrolytic Capacitors", 2nd Edition, 1995; United Chemi-Con Inc.
[26] Micro Linear Corporation Data Book, 1995
[27] "Fair-Rite Soft Ferrites", Databook, 13th Edition; Fair-Rite Products Corp. NY 12589
[28] "Magnetics Designer", Supplementary Information, 1997; Intusoft
[29] "UC3842/3/4/5 Provides Low-cost Current-mode Control", Application Note, U-100A; Unitrode Integrated Circuits
[30] Billings, K. H.; "Switchmode Power Supply Handbook", 1989, ISBN 0070053308; McGraw-Hill Inc.
[31] Pressman, A. I.; "Switching Power Supply Design", 1991, ISBN 0070508062, McGraw Hill Inc.
[32] McLyman, W. T.; "Transformer and Inductor Design Handbook", 2nd Edition, 1988, ISBN 0824778286; Marcel Dekker, Inc.
[33] Unitrode Power Supply Design Seminar, SEM-500, Unitrode Integrated Circuits
[34] "3C85 Handbook", 1987, Ordering Code 9398 345 90011; Philips Electronic Components and Materials
[35] Sum, K. K.; "Intuitive Magnetic Design", Nov 15-16, 2000, Electronic Design Workshops; Penton Media, Inc.

[36] Bloom, G. E.; "DC-DC Switchmode Power Converters, Circuits and Converters", April 25, 2002, National Semiconductor Corporation Seminar Presentation; Bloom Associates Inc., CA-94903
[37] Mulder, S. A.; "Application Note on the design of low profile high frequency transformers, a new tool in SMPS design", 1990, Ordering Code 9398 074 80011; Philips Components Corporate Innovation Materials
[38] Ahmadi, H.; "Calculating Creepage and Clearance Early Avoids Design Problems Later", March/April 2001; Compliance Engineering Magazine
[39] Redl, R.; "Low-Cost Line-Harmonics Reduction", 1995 Seminar in Bremen, Germany; Power Quality Conference
[40] Carsten, B.; "Calculating Skin and Proximity Effect, Conductor Losses in Switchmode Magnetics", 1995, PCIM Conference
[41] "Magnetics® Ferrites", Databook, 1999; Magnetics Inc., Division of Spang and Company
[42] Lee, S.; "Thermal Management of Electronic Equipment", 1996, PCIM Conference
[43] Middlebrook, R. D.; Cuk, S.; "Advances in Switched-Mode Power Conversion: Volumes I, II, and III", TESLAco, 10 Mauchly, Irvine, CA 92618
[44] Middlebrook, R. D.; "Topics in Multiple-loop Regulators and Current-mode Programming", IEEE 1985
[45] Erickson, R. W.; "Fundamentals of Power Electronics", Springer, Second Edition, ISBN 0792372700
[46] "Control Design Lecture Notes", Center for Power Electronics Systems, June 2-6, 2003, Virginia Polytechnic Institute and State University, Blacksburg, Virginia
[47] Maniktala, S.; "Switching Power Supply Design and Optimization", McGraw-Hill Professional, First Edition, ISBN 0071434836

This Page Intentionally Left Blank

Index

This Page Intentionally Left Blank