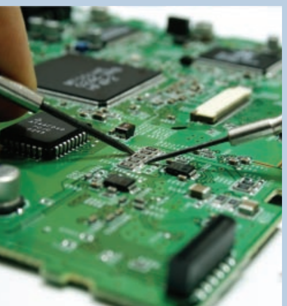
Fundamentals of Power Semiconductors

for Automotive Applications (Second Edition)

Bridging Theory into Practice



www.infineon.com/btip

Bridging Theory into Practice

Fundamentals of Power Semiconductors for Automotive Applications

Edition 2008

Published by
Infineon Technologies AG
81726 Munich, Germany

© 2008 Infineon Technologies AG
All Rights Reserved.

Legal Disclaimer

The information given in this document shall in no event be regarded as a guarantee of conditions or characteristics. With respect to any examples or hints given herein, any typical values stated herein and/or any information regarding the application of the device, Infineon Technologies hereby disclaims any and all warranties and liabilities of any kind, including without limitation, warranties of non-infringement of intellectual property rights of any third party.

Information

For further information on technology, delivery terms and conditions and prices, please contact the nearest Infineon Technologies Office (**www.infineon.com**).

Warnings

Due to technical requirements, components may contain dangerous substances. For information on the types in question, please contact the nearest Infineon Technologies Office.

Infineon Technologies components may be used in life-support devices or systems only with the express written approval of Infineon Technologies, if a failure of such components can reasonably be expected to cause the failure of that life-support device or system or to affect the safety or effectiveness of that device or system. Life support devices or systems are intended to be implanted in the human body or to support and/or maintain and sustain and/or protect human life. If they fail, it is reasonable to assume that the health of the user or other persons may be endangered.

Preface to the second Edition

Dear Reader,

I am proud to introduce the second edition of *Bridging Theory Into Practice (BTiP), the Fundamentals of Power Semiconductors for Automotive Applications*. This book is the second edition of our first book in a series to be developed by the Infineon Technologies Automotive, Industrial and Multimarket (AIM) Group. The series began with the 2006 publication of the first edition of *Bridging Theory Into Practice (BTiP), the Fundamentals of Power Semiconductors for Automotive Applications*. If you've already read the first edition, then you know that the BTiP methodology assists an engineer in bridging the gap between theoretical concepts learned in the classroom and real world electronics design within automotive and industrial applications.

BTiP is a practical methodology that is developed from the knowledge and many years of hands on application experience of Infineon engineers. It continues to be implemented and expanded upon by the Infineon AIM team. Because of the success within Infineon, we couldn't help but recognize the value that the BTiP methodology could bring to a broader audience. We are very proud of the praise and acceptance attained by the first edition of the book. Now, in this second edition, we build upon the basic BTiP concepts and offer our readers in-depth chapters on switching voltage regulators and motor control for automotive applications as well as enhancements to the ESD, EOS and quality/reliability. All fourteen chapters of this edition, when taken together, follow a step-by-step approach to the understanding of automotive power electronics. But because each chapter can stand alone, a more advanced reader can jump directly to a chapter of particular interest. You'll find that we've also enhanced this edition with a detailed glossary, making it an excellent desktop reference.

The BTiP methodology also includes a classroom training element. In fact, this book is the companion to our training course on automotive power semiconductor applications. The trainings are held in fourteen short courses or in a combined four day, high intensity workshop. All trainings are given by experienced Infineon engineers. Automotive OEM's and major transportation electronics suppliers have benefited from our workshops. Please contact your local sales office if you are interested in scheduling a workshop or visit the BTiP web site (www.infineon.com/btip).

I'd like to take this opportunity to thank some individual Infineon AIM team members who have made significant contributions of time and effort to *Bridging Theory Into Practice (BTiP), the Fundamentals of Power Semiconductors for Automotive Applications*. In particular, I would like to recognize the efforts of Bob Beier, Frank Heinrichs, Brian Hemphill, Gunther Krall, Richard Kraus, Keng Ly, Ray Notarantonio, Elvira Palmeda, Chris Spielman, Keith Toby and Shawn Williams. An enormous thank you goes to Mark Budnik, Assistant Professor of Electrical Engineering at Valparaiso University and Tom Pattantyus of Applied Electronic Concepts Co. for their expertise and leadership in writing and editing. This book would not have been possible if it hadn't been for Mark's ability to understand the BTiP initial concept and move our vision to reality. Once our hard work had produced an initial product, it was Tom's knowledge and editing skills that took it to the next level. They both deserve our deepest gratitude. Finally, it is the support of Infineon management and the dedicated efforts of the entire Infineon AIM team that made this book possible.

Your comments on this book and suggestions for future publications are important to us. Please feel free to send us an email at btip@infineon.com.

Sincerely,

Shawn P. Slusser
Vice President
North America Automotive Business

Contents

1.	RLC Load Characteristics and Modeling	11
1.1	Introduction to Capacitors and Resistor Capacitor Networks	12
1.2	Introduction to Inductors and Resistor-Inductor Networks	23
1.3	Example Load Model: Turning on an Incandescent Lamp	34
1.4	Example Load Model: Switching a Relay	36
2.	Introduction to Semiconductors	39
2.1	What is a Semiconductor?	40
2.2	What is a Diode?	43
2.3	What is a Bipolar Junction Transistor?	46
2.4	What is a MOSFET?	48
3.	Transistors and Integrated Circuits	53
3.1	Bipolar Junction Transistors	54
3.2	Metal-Oxide-Semiconductor Field Effect Transistors	61
3.3	Integrated Circuits	66
3.4	Moore’s Law	66
4.	Introduction to Power Dissipation and Thermal Resistance	69
4.1	What is Power?	70
4.2	What is Junction Temperature?	71
4.3	What is Thermal Resistance?	73
4.4	Electrical vs. Thermal Parameters	75
4.5	Thermal Specifications	79
4.6	Heatsinks	81
4.7	Example DC Thermal Calculations	82
5.	Advanced Power Dissipation and Dynamic Thermal Analysis	85
5.1	Electrical and Thermal Parameters	86
5.2	Thermal RC Networks	88
5.3	Understanding the Z_{th} Diagram	92
5.4	Example Transient Thermal Calculations	95
5.5	Complex Waveforms and Superposition	97
6.	MOSFETs, High Side Drivers, and Low Side Drivers	101
6.1	MOSFET Review	102
6.2	Driver Topologies and Loads	102
6.3	Protected Drivers	105
6.4	Selecting the Correct R_{DSon} Static Operation	106
6.5	Selecting the Correct R_{DSon} Dynamic Operation and the Impact of Switching Losses	111
6.6	Capacitive Load In-Rush Current	114
6.7	Switching Off an Inductive Load	116

7. Protected High Side Drivers 125
7.1 What is a Protected MOSFET High Side Driver? 126
7.2 Protection Features 128
7.3 Diagnostic Features 136
7.4 EMI / EMC Considerations 140
7.5 System Implementation 144
7.6 Frequently Asked Questions 146

8. Protected Low Side Drivers 151
8.1 What is a Protected MOSFET Low Side Driver? 152
8.2 Protection Features 152
8.3 Diagnostic Features 155
8.4 EMI / EMC Considerations 156
8.5 System Implementation 158
8.6 Frequently Asked Questions 163

9. Introduction to Linear Regulators 167
9.1 Definition of Power and Basic Concepts Linear Regulators 168
9.2 Linear Regulators 169
9.3 Linear Regulator Functional Blocks 170
9.4 Protection Circuits 175
9.5 Characteristics of Linear Regulators 176
9.6 Auxiliary Functions of Linear Regulators 179
9.7 Stability 182

10. Introduction to Switching Regulators 187
10.1 What is a Switching Regulator? 188
10.2 Why Are Switching Voltage Regulators Needed? 188
10.3 Types of Switching Regulators 189
10.4 General Operation of a Step Down or Buck Switching Regulator 190
10.5 Switching Regulator Components 192
10.6 Step Down (Buck) Switching Regulator Operation 195
10.7 Design Guidelines and Examples 198
10.8 Methods of Control 207
10.9 Special Features 208
10.10 Switching Losses and Efficiency 211

11. Electrostatic Discharge, Electrical Over Stress, and Safe Operating Area 219
11.1 Electrostatic Discharge (ESD) Stress 220
11.2 Electrical Over Stress (EOS) 230
11.3 Safe Operating Area 233

12. Semiconductor Manufacturing 235
12.1 Semiconductor Fabrication 236
12.2 Fabrication Recipes 239
12.3 Semiconductor Packaging 240
12.4 Semiconductor Testing 243

13. Quality and Reliability of Semiconductor Devices 245
13.1 What is Quality? 246
13.2 What is Reliability? 246
13.3 Customer Expectations 248
13.4 Meeting the Demands of ZD to Earn Customer Loyalty 248
13.5 The Power of Preventive and Protecting Process Control 249
13.6 The Making of ZD Semiconductor Devices 249
13.7 Causes of Reliability Issues 250
13.8 Conditions of ZD Technology 250
13.9 What Are ESD and EOS Failures? 251
13.10 Outliers 253
13.11 Latent Defects 254
13.12 Failure Analysis and Problem Solving 254
13.13 Optimization of the Results of F/A 255
13.14 Effective Methods of Problem Solving for Semiconductors 262
13.15 Conclusion 264

14. Introduction to Motor Control 265
14.1 Introduction to Electromagnets and Electric Motors 266
14.2 Electric Motors Used in Automobiles 267
14.3 Interactions between Magnetic Fields 269
14.4 Detailed Description of How the Three Basic Motor Types Operate 271

A. Permanent Magnet DC Motors Models and Application 281
A.1 Motor Loads 282
A.2 Torque Related Units, SI and Traditional Units 282
A.3 Motor Characteristic 283
A.4 Dynamic Braking 285
A.5 Transient Current Demands 286
A.6 Semiconductor Realization of Motor Controllers 289
A.7 Power Calculations 290
A.8 Efficiency 290
A.9 Conclusions 291

Glossary of Terms 293

1. RLC Load Characteristics and Modeling

In this first chapter, we will be examining the two basic passive electrical components: capacitors, and inductors. With these simple components, engineers can develop electrical models of various loads to simplify system design and analysis.

Next, we will examine capacitive electrical components. We will explore the fundamental physical and electrical properties of capacitors. We will then conduct several "experiments" in order to develop a more intuitive feel of how capacitors operate. This will allow us to mathematically describe the operation of capacitors in electrical circuits. Finally, we will examine two simple resistor and capacitor circuits.

This process will be repeated for inductive electrical components.

We will close Chapter 1 with two examples of modeling loads in electrical circuits with resistors, capacitors, and inductors.

1.1 Introduction to Capacitors and Resistor Capacitor Networks

So, what is a capacitor?

A capacitor is an object with the ability to store electrical charge. They usually consist of multiple conductive, electrically isolated parallel plates. The material between the parallel plates is called the dielectric material. This material is an electrical insulator. It does not allow current to flow through it.

- Physical object with the ability to store electric charge (i.e. “electric voltage”)
- Consists of two – electrically isolated – metal electrodes, typically two conductive parallel plates
- Is mostly used to store energy or for filtering purposes
- The isolating material – the dielectric – defines the type of capacitor: e.g. tantalum or ceramic capacitor
- Circuit symbol:

Figure 1.1 Capacitors

In Figure 1.2, we see a drawing of an example parallel plate capacitor. The area of each of the parallel plates is “A”, with a distance “d” separating the two plates. As the area of the parallel plates increases, the capacitor is able to hold more charge. We say that its capacitance increases. If the distance between the plates increases, however, the structure will be able to hold less charge, and its capacitance is said to decrease. Therefore, the larger area and/or smaller the separation the larger the value of capacitors is which enables them to store the more electrical charge between the plates.

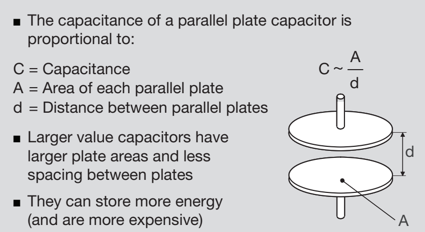


Figure 1.2 Capacitors: Physical Properties

The capacitance of this parallel plate structure is also a function of the dielectric material between the plates. The insulating dielectric material has a characteristic called “permittivity”. This is defined as the ability to store charge (energy) in the presence of an electrical field. If the dielectric material has a relatively high permittivity, it can store more energy than a dielectric material with a low permittivity. Therefore, capacitors, that can store the highest amount of charge, have high permittivity dielectric materials.

Figure 1.3 shows us how the capacitance of a parallel plate structure is calculated. Capacitance is measured in units of Farads (F).

F = Axs / V = C / V

where C in this formula stands for charge (unit is Coulomb or Axs) Therefore, a 1 Farad capacitor can store 1 Coulomb of charge at 1 V.

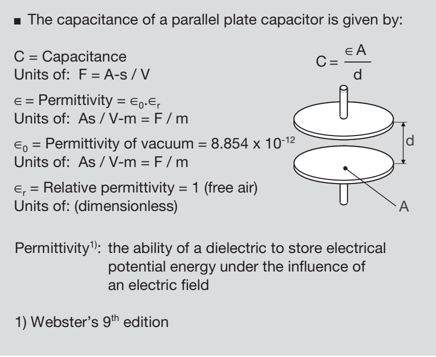


Figure 1.3 Capacitors: Physical Properties

Now, let us look at a hypothetical parallel plate capacitor. The dielectric material between the parallel plates is free air:

$\varepsilon = (1)(8.854 \times 10^{-12}$ F/m)

Our hypothetical area is 1 m^2 and the separation between the plates is 1 mm. This results in a capacitance value of 8.854×10^{-9} F.

Based on this example you can probably see that capacitance values at or above 1 F are rather special. Therefore, capacitors are

usually specified in values of microFarads, nanoFarads, or picoFarads.

- Capacitance of a free air ($\epsilon_r = 1$) parallel plate capacitor with the dimensions of A = 1 m² and d = 1 mm is:

$$C = \frac{\epsilon_r \epsilon_0 A}{d} = \frac{(1)(8.854 \times 10^{-12}\,F/m)(1\,m^2)}{1 \times 10^{-3}\,m} = 8.854 \times 10^{-9}\,F$$

- Typically, capacitance values in the 1 F range are unlikely
- Capacitances typically range from microFarads to picoFarads

  1 microFarad = 1 μF = 10^{-6} F

  1 nanoFarad = 1 nF = 10^{-9} F

  1 picoFarad = 1 pF = 10^{-12} F

Figure 1.4 Relative Size of Capacitance

Now, let us examine in more detail the electrical properties of capacitors. On the last page, we showed that 1 Farad is equivalent to 1 Coulomb per Volt:

1 Farad = 1 Coulomb / Volt

Rearranging the equation, we get:

1 Coulomb = (1 Farad)(1 Volt)

This implies that the amount of charge (in Coulombs, symbol C) that a capacitor can store is equal to its capacitance [C] times the voltage [V] across the capacitor:

Q = C V

Note: It is rather unfortunate that the symbol of both charge and capacitance is "C".

Let us look at another hypothetical example. Imagine a capacitor that has 8 C of charge stored on it. The voltage across the capacitor is found to be 16 V. This allows us to calculate the capacitance:

C = Q / V

C = 8 C / 16 V = 0.5 F

- The stored electrical charge Q in a capacitor is proportional to the voltage V across the capacitor: Q ~ V
- The proportional factor between stored electrical charge and voltage difference is the capacitance value of the capacitor: Q = C V

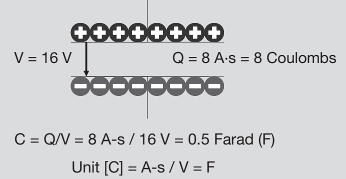


Figure 1.5 Capacitors: Electrical Properties

Now, let us look at what happens when multiple capacitors are used in a circuit. There are two basic configurations: parallel capacitors and serial capacitors (see Figure 1.6).

In a parallel capacitor configuration, we are effectively increasing the area of the top and bottom parallel plates. Therefore, since the capacitance is a linear function of the plate area, capacitors in parallel add their capacitance.

When capacitors are in series, however, things are not quite as simple. If current flows through the two series capacitors, the same amount of charge (Q) will be deposited on each capacitor. Now if V_i is the voltage across capacitor C_i:

$V_1 = Q / C_1$

$V_2 = Q / C_2$

$V = V_1 + V_2$

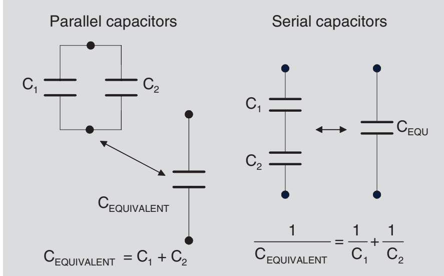


Figure 1.6 Parallel and Serial Capacitance

Substituting for V_1 and V_2:

$V = (Q / C_1) + (Q / C_2)$

Now, let us compare this to a different capacitor, $C_{EQUIVALENT}$.

$V = Q / C_{EQUIVALENT}$

Setting the two voltage equations equal to each other, we find:

$Q / C_{EQUIVALENT} = (Q / C_1) + (Q / C_2)$

Finally, we can divide both sides by Q to find:

$1 / C_{EQUIVALENT} = 1 / C_1 + 1 / C_2$

Let us now look at our first capacitor "experiment." In Figure 1.7, we show an ideal current source is connected directly to a capacitor by a switch. (An ideal current source will supply a defined amount of current into the circuit, regardless of how much voltage is required to do so.)

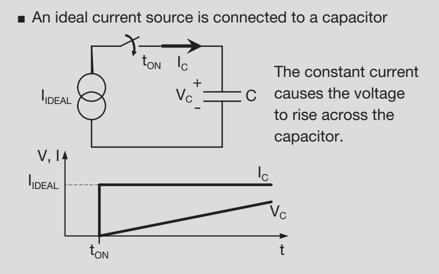


Figure 1.7 Capacitor Experiment #1

The switch is closed at time t_{ON}. We then monitor the current through the capacitor, and the voltage across the capacitor. From the Figure 1.7, we see that the current is a step function. Before t_{ON}, there is an open circuit and no current flows. After t_{ON}, the capacitor current (I_C) is immediately equal to the ideal current source (I_{IDEAL}). There is no impediment to the current flow. At t_{ON}, there is no voltage across the capacitor. As the constant current flows through the capacitor, we note that the voltage across the capacitor (V_C) increases linearly.

In our second experiment, we examine what happens after the ideal current source is disconnected from the capacitor (see Figure 1.8). At t_{OFF}, the switch is opened, and the current through the capacitor falls to 0 A. When this occurs, the capacitor voltage (which was linearly increasing with the current) stops increasing and remains constant at its t_{OFF} value.

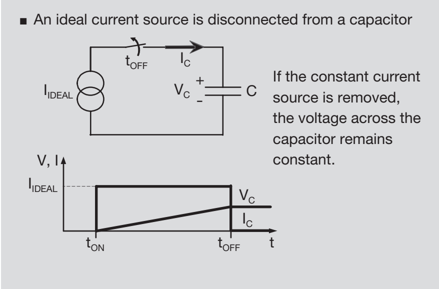


Figure 1.8 Capacitor Experiment #2

Next, we examine what happens when we change the magnitude of the ideal current source. In the Figure 1.9, we show how the voltage across the capacitor (V_{C1}) increases for the first ideal current source value (I_{C1}).

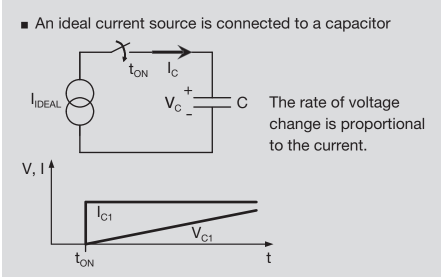


Figure 1.9 Capacitor Experiment #3

Next, we increase the magnitude of the ideal current source to I_{C2} (see Figure 1.10). We notice that the voltage across the capacitor (V_{C2}) is still linearly increasing, but it is increasing at a faster rate. The rate of change of V_C depends on the magnitude of current (I_C).

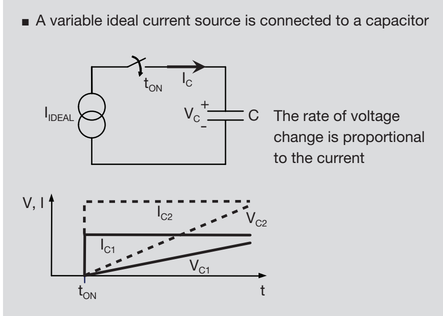


Figure 1.10 Capacitor Experiment #3a

Now an ideal voltage source is connected to a capacitor through a resistor at t_{ON}. (An ideal voltage source will supply a defined voltage bias to a circuit, regardless of how much current is required to do so, similar to a car battery) The results are shown in the Figure 1.11.

Let us consider the voltage across the capacitor (V_C) first. Before t_{ON}, the ideal voltage source is disconnected from the capacitor, so there is no voltage drop across the capacitor. Immediately at t_{ON}, V_C is 0 V. It is only after t_{ON} that the voltage across the capacitor begins to exponentially increase. After a long time period, V_C is nearly equal to V_{IDEAL}.

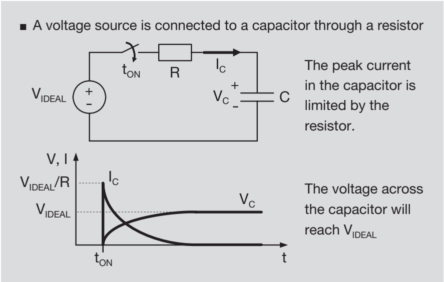


Figure 1.11 Capacitor Experiment #4

Now, let us look at the current flowing through the capacitor (I_C). At t_{ON}, V_C is 0 V, and the ideal voltage source is able to deliver the initial current equal to:

$$I_{C0} = V_{IDEAL} / R$$

As V_C increases, however, the current sourced by the ideal voltage source exponentially decays. Eventually, when V_C is equal to V_{IDEAL}, I_C is 0 A.

Next, let us consider what happens if we look at making the resistor variable.

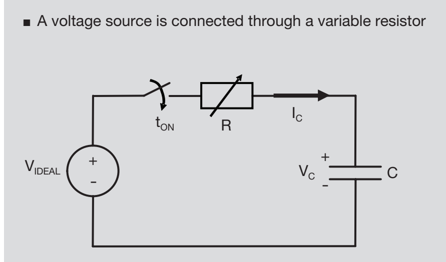


Figure 1.12 Capacitor Experiment #5

First, we let the resistor be equal to R_1. This is the benchmark resistance used for comparison in experiment five. We see in the Figure 1.13 how the capacitor current (I_{C1}) is initially V_{IDEAL} / R_1. As V_{C1} approaches V_{IDEAL}, I_{C1} decays to 0 A (#5a).

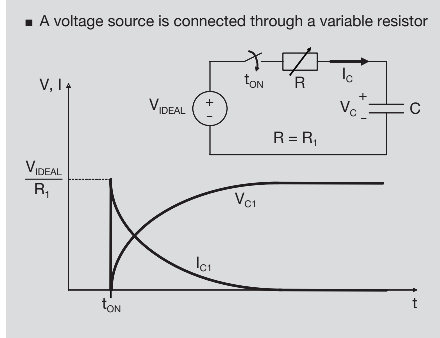


Figure 1.13 Capacitor Experiment #5a

Next, we change the value of the resistor. We let the resistance be R_2, where R_2 is less than R_1 (see Figure 1.14).

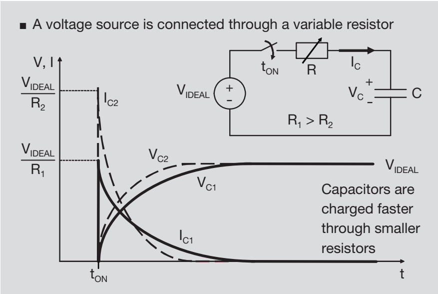


Figure 1.14 Capacitor Experiment #5b

We first note that since $R_2 < R_1$, I_{C2} (V_{IDEAL} / R_2) is greater than I_{C1} at t_{ON}. Next, we see that the capacitor voltage, V_{C2} is increasing to V_{IDEAL} more quickly than in V_{C1}. This results in a more rapid decay of the capacitor current. It can be stated that a given capacitor is charged faster through a smaller resistor.

Now we examine what happens when the resistance is changed a final time (#5c). Let $R_3 > R_1$. As expected, I_{C3} (V_{IDEAL} / R_3) is less than I_{C1}. Also, V_{C3} takes longer to reach V_{IDEAL}, and I_{C3} decays more slowly. A given capacitors is charged more slowly through a larger resistor.

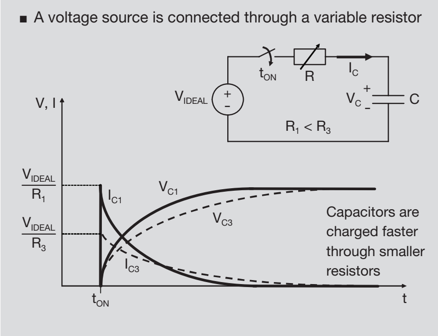


Figure 1.15 Capacitor Experiment #5c

Finally, we conduct our last experiment (#6) on our capacitor circuits, this time with a constant resistance, R. This time, we connect the ideal voltage source to the capacitor through R at time equal to zero seconds, and we examine V_C vs. time.

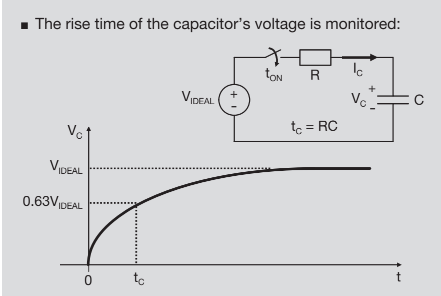


Figure 1.16 Capacitor Experiment #6

As expected, V_C eventually reaches V_{IDEAL}. What we also notice, however, is that at time equal to t_C:

$t_C = RC$

the voltage across the capacitor is approximately 63% of V_{IDEAL}. For example, for $V_{IDEAL} = 1$ V, R = 1 kΩ and C = 1 mF:

$t_C = (1 \times 10^3\ \Omega)(1 \times 10^{-3}\ F) = 1\ s$

One second after the 1 V ideal voltage source is connected, the voltage across the capacitor is approximately 0.63 V. This is true for all values of V_{IDEAL}, R, and C. For example:

$V_{IDEAL} = 2$ V, R = 10 kΩ, C = 1 μF:

$t_C = (1 \times 10^4\ \Omega)(1 \times 10^{-6}\ F) = 0.01\ s$

V_C at 0.01 s = 0.63 (2 V) = 1.26 V

$V_C = 12$ V, R = 37 Ω, C = 22 μF:

$t_C = (37\ \Omega)(22 \times 10^{-6}\ F) = 0.814\ ms$

V_C at 0.814 ms = 0.63 (12 V) = 7.56 V

Therefore, t_C is an important value for resistor and capacitor circuits. It is called the "time constant" of RC circuits. Note, sometimes, t_C is written as τ or τ_C.

If we continue to examine the voltage across the capacitor for time beyond t_C, we notice additional characteristics of RC circuits. For example:

at $t = t_C$, $V_C = 0.63V_{IDEAL}$

at $t = 2t_C$, $V_C = 0.87V_{IDEAL}$

at $t = 3t_C$, $V_C = 0.95V_{IDEAL}$

at $t = 4t_C$, $V_C = 0.98V_{IDEAL}$

at $t = 5t_C$, $V_C = 0.99V_{IDEAL}$

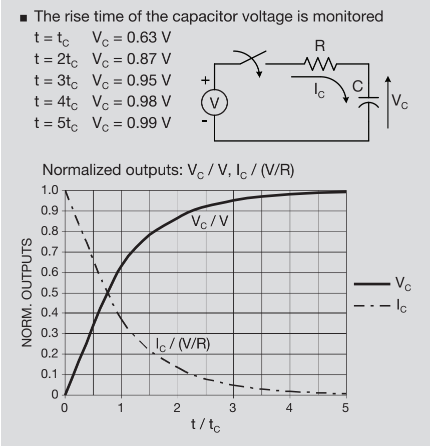


Figure 1.17 Capacitor Experiment #6a

The values of time constant for resistor capacitor circuits are important parameters for design and in analysis.

For example, at $3t_C$ after the voltage source is connected, the designer knows that the capacitor will reach 95% of the source voltage (or final value). Likewise, at $5t_C$, the capacitor will have reached 99% of its final value. The $1t_C$, $3t_C$, and $5t_C$ times are often used to approximate how long it takes to charge a capacitor.

Note, for those interested in a more mathematical description, the voltage across the capacitor and the capacitor current in this circuit is actually given by:

$$V_C = V_{IDEAL}[1 - \exp(-t / t_C)]$$

$$I_C = (V_{IDEAL} / R).[\exp(-t / t_C)]$$

Finally, we can look at developing the mathematical relationship between the current flowing through the capacitor and the voltage across the capacitor. Current (I) is defined as the amount of charge (Q) that flows in a given amount of time (t):

$$I = Q / t$$

Therefore, if we look at the change in charge (ΔQ) in a given amount of time (Δt), we can calculate the average current flowing during time Δt:

$$i = \Delta Q / \Delta t$$

If we let the change in charge and the change in time be very small, we consider this to be a derivative:

$i = dq / dt$ for very small Δt

This can be rearranged:

$$dq = i\,dt$$

This means the change in the charge (dq) is given by the current (charge per second) times the change in time (dt).

- Current is defined as the amount of charge which is transferred in a certain period of time: $I = Q / t$

$$i = \frac{dq}{dt} \quad \text{or} \quad dq = i \cdot dt \qquad (1)$$

Figure 1.18 Development of Mathematical Capacitor Model: I_C vs. V_C

Next, we recall the definition of capacitance:

$$C = Q / V$$

Because C is a constant, we expect that for a small change in the capacitor charge (ΔQ), there will be a small change in the voltage across the capacitor (ΔV):

$$C = \Delta Q / \Delta V$$

Again, letting ΔQ and ΔV be very small:

$$C = dq / dv$$

which can be rearranged:

$$dq = C\,dv$$

- Current is defined as the amount of charge which is transferred in a certain period of time: $I = Q / t$

$$i = \frac{dq}{dt} \quad \text{or} \quad dq = i \cdot dt \qquad (1)$$

- Capacitance is defined as the stored charge on a capacitor vs. the voltage across the capacitor, $C = Q / V$

$$C = \frac{dq}{dv} \quad \text{or} \quad dq = C \cdot dv \qquad (2)$$

Figure 1.19 Development of Mathematical Capacitor Model: I_C vs. V_C

Now we have two equations for dq. If we set them equal, we have:

i dt = dq = dq = C dv

Rearranging yields the fundamental capacitor equation:

i = C dv / dt

- Current is defined as the amount of charge which is transferred in a certain period of time: $I = Q / t$

$$i = \frac{dq}{dt} \quad \text{or} \quad dq = i \cdot dt \qquad (1)$$

- Capacitance is defined as the stored charge on a capacitor vs. the voltage across the capacitor, $C = Q / V$

$$C = \frac{dq}{dv} \quad \text{or} \quad dq = C \cdot dv \qquad (2)$$

- Setting (2) equal to (1) results in:

$$i \cdot dt = C \cdot dv \quad \text{or} \quad i = C\frac{dv}{dt}$$

Figure 1.20 Development of Mathematical Capacitor Model: I_C vs. V_C

Therefore, the current flowing through a capacitor is equal to its capacitance times rate at which the voltage across the capacitor is changing.

Now, let us see the two basic ways that capacitors are used in resistor capacitor circuits. First, the capacitor can be in series with the signal path. The signal of interest must pass through the capacitor (see Figure 1.21 on the left). Second, the capacitor can be from the signal path to ground (see Figure 1.21 on the right).

- In general, there are two basic options for capacitor placement:

C in Series with Signal Path — C from Signal Path to Ground

V_{IN} C R V_{OUT} — V_{IN} R C V_{OUT}

Figure 1.21 Capacitor & Resistor Networks

Let us first look at how these two circuits operate when a constant voltage is applied to the signal path through a switch.

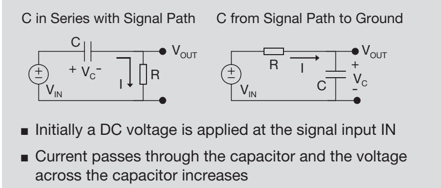


Figure 1.22 Capacitor & Resistor Networks

In both cases, current flows through the capacitor, and the voltage across the capacitor (V_C) starts to increase.

Eventually, the voltage across the capacitor is equal to the input voltage ($V_C = V_{IN}$). At that point, the current flowing through the capacitor has fallen to 0 A.

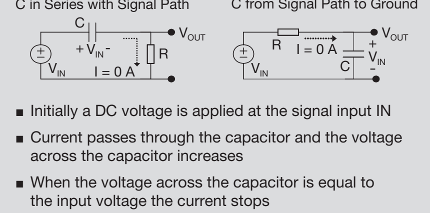
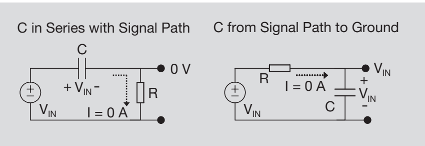


Figure 1.23 Capacitor & Resistor Networks

In the final state (I = 0) if the capacitor is in series with the signal path (see Figure 1.24 on the left), no current is flowing through the resistor. Therefore, V_{OUT} is at 0 V. All of the voltage drop is across the capacitor.

If the capacitor is from the signal path to ground (see Figure 1.24 on the right), V_{OUT} is equal to V_{IN} (the voltage across the capacitor).



- Initially a DC voltage is applied at the signal input IN
- Current passes through the capacitor and the voltage across the capacitor increases
- When the voltage across the capacitor is equal to the input voltage the current stops
- Depending on the capacitor's placement, the V_{OUT} = 0 V or $V_{OUT} = V_{IN}$

Figure 1.24 Capacitor & Resistor Networks

Now, let us look at a circuit with a capacitor in series with the signal path more closely (see Figure 1.25). At time t_1, the input voltage is applied to the series capacitor and resistor. The node at voltage V_X immediately goes to V_{IN}. The voltage across the capacitor (V_C) starts at 0 V, and the initial current is V_{IN} / R. However, as the capacitor is charged, V_C starts to increase, the current and output voltage decay to 0 A and 0 V, respectively.

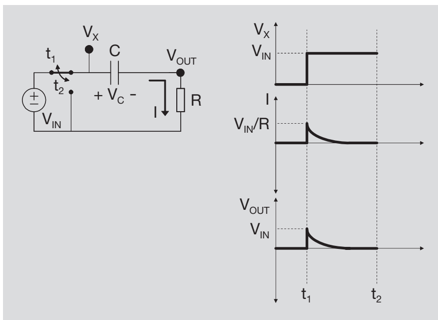


Figure 1.25 Capacitance in Series with Signal Path

Next, we see what happens when the input voltage is switched off at time t_2, and the voltage at V_X is immediately pulled to ground (i.e. $V_X = 0$).

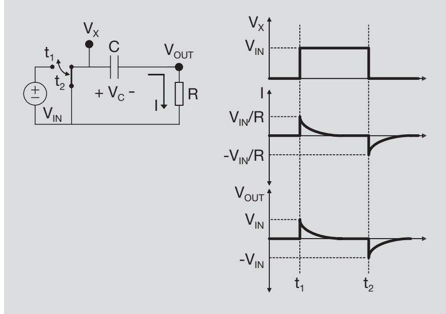


Figure 1.26 Capacitance in Series with Signal Path

The capacitor starts to discharge to ground. Current flows from the capacitor to ground, through node V_{OUT}, through the resistor and back to the capacitor. The current flows in the direction opposite to the charging current, so we consider it to be negative.

In the Figure 1.26, we also see that the output voltage falls below 0 V because the positive capacitor terminal is grounded, the discharge current to ground which then flows through the resistor R, thus the output voltage V_{OUT} is forced below 0 V:

$$V_{OUT} = - I R = -V_{IN} (t = t_2).$$

At time t_2, the output voltage V_{OUT} is equal to V_{IN}.

As the capacitor discharges, voltage across the capacitor starts to fall. Therefore, both the current and V_{OUT} again "decay" to 0 A and 0 V, respectively.

Next, let us look at a circuit with a capacitor from the signal path to ground. The input voltage is again applied to the circuit at time t_1. As before, the voltage at node V_X immediately goes to V_{IN}. Since the voltage across the capacitor (V_C) is initially 0 V, the current again starts at V_{IN} / R. As the capacitor is charged, V_C increases toward V_{IN}, resulting in the current decaying to 0 A.

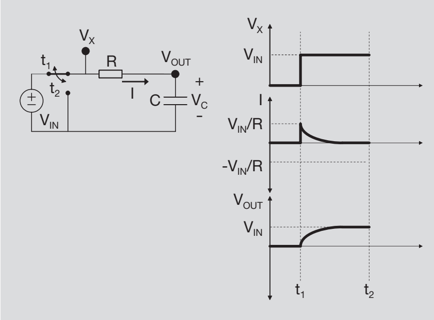


Figure 1.27 Capacitance from Signal Path to Ground

Now, let us see what happens when the V_X voltage is switched to ground at time t_2. V_X is pulled immediately to 0 V.

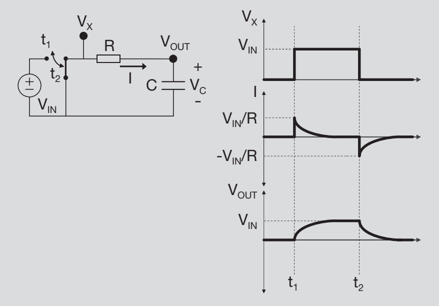


Figure 1.28 Capacitance from Signal Path to Ground

This allows the capacitor to begin discharging. Current flows from the capacitor, to node V_{OUT}, through the resistor, to node V_X, to ground, and back to the capacitor. This current flow is in the opposite direction of the original current, so it is considered to be negative. V_C at time t_2 was V_{IN}, so the current is initially V_{IN} / R. V_{OUT} is equal to V_C, so as the capacitor starts to discharge, the output voltage decays. This results in the current also decaying to 0 A.

Next, let us see what happens when an AC signal is applied to a resistor capacitor circuit.

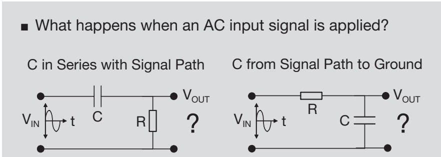


Figure 1.29 RC Networks - AC Signals

First, we need to introduce two new terms: reactance and impedance. Capacitors behave like frequency dependent resistors:

$|X_C| = 1 / (2\pi fC)$

Reactance: the measure of the "AC resistance" of a **capacitor** or an **inductor** (see later) defining both the magnitude of the AC resistance and the phase shift of the circuit element.

- Act like frequency dependent resistor (capacitive reactance, X_C)

$$X_C \propto \frac{1}{fC}$$

- Instead of reactance, impedance (Z) used for circuit elements
- Impedance[1]: The apparent opposition in an electrical circuit to the flow of alternating current that is analogous to the actual electrical resistance to a direct current

1) acc. to Webster's 9th edition

Figure 1.30 Capacitors and AC Signals

Impedance: the measure of the AC resistance of a circuit that contains resistor(s), capacitor(s) and inductor(s) in some combination.

In the "$|X_C|$ vs. frequency" figure, we illustrate how magnitude of the capacitor's reactance, $|X_C|$, varies with the frequency of the AC signal. $|X_C|$ is proportional to the **inverse** of the frequency.

Real life capacitors also have parasitic resistors but in most cases their effect is negligible. In this section only the reactance (X_C) is considered.

- Act like frequency dependent resistor (capacitive reactance, X_C)

$$X_C \propto \frac{1}{fC}$$

- Instead of reactance, impedance (Z) used for circuit elements
- Impedance[1]: The apparent opposition in an electrical circuit to the flow of alternating current that is analogous to the actual electrical resistance to a direct current
- The impedance of a circuit element represents its resistive and/or reactive components
- Besides the magnitude dependency between voltage and current the impedance Z gives also information about the phase shift between the two

Figure 1.31 Capacitors and AC Signals

Therefore, for DC conditions (0 Hz), the impedance is infinite. But, as an AC signal's frequency increases, the capacitor's impedance magnitude falls.

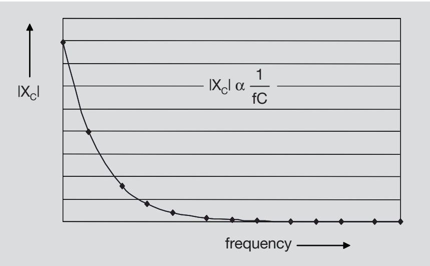


Figure 1.32 Ideal Capacitor's Reactance $|X_C|$ vs. Frequency

Now, let us look at the phase differential (φ) between the voltage across a capacitor (V_C) and the current flowing through a capacitor (I_C). That is, $V_{C,Max}$ and $I_{C,Max}$ do not occur at the same point in time.

In Figure 1.33, we show two curves, the voltage across the capacitor (peak value of $V_{C,peak}$), and the current through the capacitor ($I_{C,peak}$). We note that the peak AC current is equal to the peak AC voltage divided by the magnitude of the capacitor's impedance:

$I_{AC,Max} = V_{AC,peak} / |X_C|$

We see that I_C "leads" V_C by 90°:

When I_C is at its positive peak value, V_C is 0 V.

When I_C decreases to 0 A, V_C is at its positive peak value.

When I_C further decreases to its negative peak value, V_C becomes to 0 V.

When I_C increases to 0 V, V_C attains its most negative value.

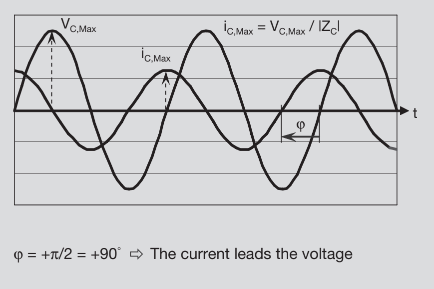


Figure 1.33 Capacitors and AC Signals

For the mathematically inclined, we note this behavior is predicted by our earlier derivation of the mathematical capacitor model:

$i = C\,(dv / dt)$

If we let the voltage across the capacitor to be a sinusoid as in the above figure:

$v(t) = V_{C,peak}\sin(2\pi ft)$,

and noting that the derivative of the sine function is the cosine function, we can see that the current through the capacitor will be:

$i(t) = C.d\,[V_{C,peak}\sin(2\pi ft)] / dt$

$i(t) = 2\pi fC.V_{C,peak}\cos(2\pi ft)$

In summary, capacitors act as frequency dependent resistors. Their reactance determines the relationship between the voltage across the capacitor and the current flowing through the capacitor.

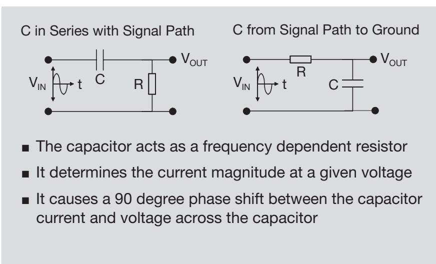


Figure 1.34 RC Networks - AC Signals

The magnitude of the capacitor's reactance is proportional to the reciprocal of applied AC signal's frequency:

$|X_C| = 1 / (2\pi fC) = 1 / \omega C, \omega = 2\pi f$

AC signals pass through the capacitor. Direct current (DC) signals are blocked by the capacitor.

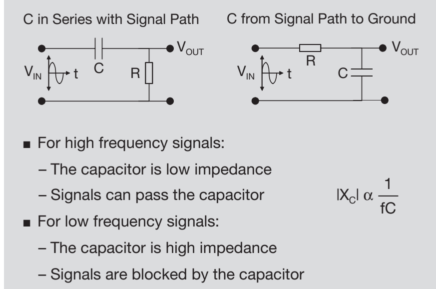


Figure 1.35 RC Networks - AC Signals

This is illustrated in Figure 1.36. First, we see a resistor capacitor circuit in a "high pass" configuration with. We see that for a low frequency input signal, the output is significantly attenuated ($V_{OUT,Max} / V_{IN,Max} = 1.6\ V / 5\ V$). At the highest frequency shown, the output is not significantly attenuated ($V_{OUT,Max} / V_{IN,Max} = 4.5\ V / 5\ V$).

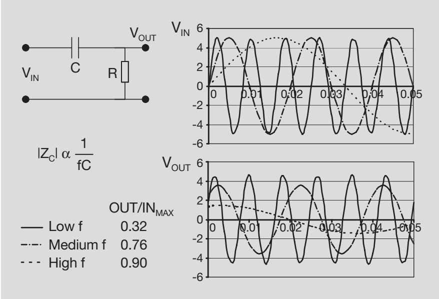


Figure 1.36 C in Series with Signal Path High Pass Configuration

Next, we see a resistor capacitor circuit in a "low pass" configuration with. For a high frequency input signal, the output is significantly attenuated ($V_{OUT,Max} / V_{IN,Max} = 1.95\ V / 5\ V$). At the lowest frequency shown, the output is not significantly attenuated ($V_{OUT,Max} / V_{IN,Max} = 4.8\ V / 5\ V$).

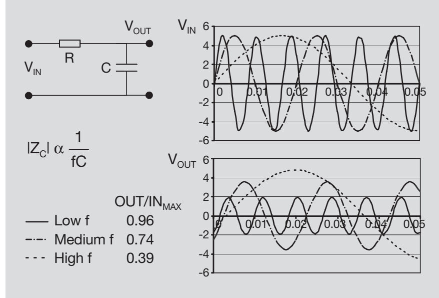
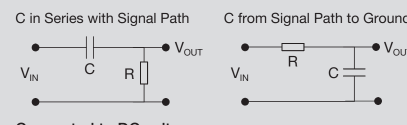


Figure 1.37 C from Signal Path to Ground Low Pass Configuration

So, let us re-view a few points before we close our introduction to capacitors.

They are electrical components that are typically used in one of two different configurations: in series with the signal path or from the signal path to ground.

When a DC voltage is applied to a capacitor, they will allow current to pass only until they are fully charged. Afterwards, they stop all future DC current flow.



Connected to DC voltages:

- Capacitors will allow current to flow only until they are charged
- Once charged, they block future current flow

For AC signals:

- Capacitors act similar to frequency dependent resistors
- Low impedance at high frequencies
- High impedance at low frequencies

Figure 1.38 Capacitor & Resistor Networks Summary

For AC signals, capacitors behave like frequency dependent resistors. At low frequencies, they are very effective in attenuating the current flow. At high frequencies, they are very effective in passing current flow.

1.2 Introduction to Inductors and Resistor-Inductor Networks

So, what is an inductor?

An inductor is an object having the ability to store a magnetic field while current flows through it. An inductor consists of a length of conductive wire, wrapped in a coil around a core material. The core material usually defines the type of inductor coil (for example, an air core inductor or a ferrite core inductor).

- Physical object which can store a magnetic field (electric current)
- Consists of a conductive wire
- Wire is typically a tightly wound coil around a center core (toroid)
- Usually used for energy conversion and for filtering purposes
- The inductor type is usually defined by its core material (for example, air coil or ferrite coil inductors)
- Circuit symbol

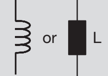


Figure 1.39 Inductors

Figure 1.40 illustrates a toroid wrapped inductor. Let the cross-section area of the toroid be "A", the number of turns of wire wrapped around the toroid "N" and the circumference of the toroid "ℓ". If the number of turns, N, of the wrapped coils increases, or the cross section area of the toroid increases, the inductor can store additional energy (but may be more expensive). If the same number of turns are wrapped more tightly around a toroid of shorter circumference the inductance also increases.

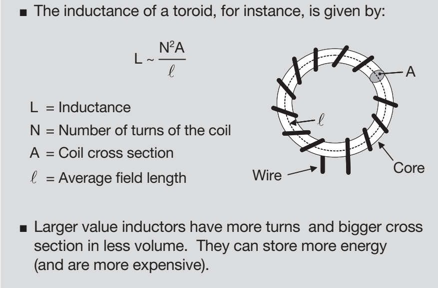


Figure 1.40 Inductors: Physical Properties

The inductance of this toroid structure is also a function of the core material. The core material has a characteristics called permeability (μ). This is defined as the relative ability of a magnetizable material to affect the magnetic flux of a magnetic field. Materials with a high permeability are used to make inductors with higher inductance. Figure 1.41 shows how the inductance of a toroid structure is calculated. Inductance is measured in units of Henrys (H).

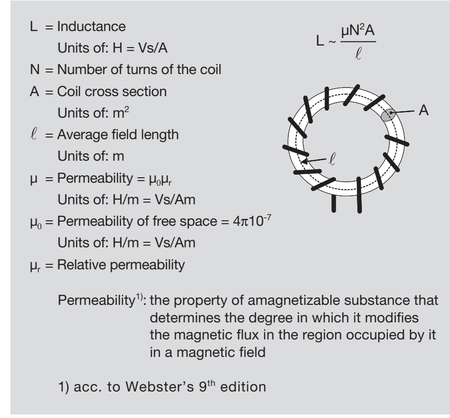


Figure 1.41 Inductance of a Toroid

Now, let us look at a hypothetical toroid inductor. The core material is free air:

$\mu_0 = (1)(4\pi 10^{-7}\ \text{H/m})$

Our hypothetical toroid cross section area is 5 cm^2, the circumference of the toroid is 10 cm, and there are 100 turns wrapped around the toroid. This results in an inductance value of 6.28 x 10^{-5} H.

Based on this example, inductance values at or above 1 H are less common in electronic circuits. Inductor values less than 1 H are usually specified in values of milliHenries (mH), microHenries (µH), or nanoHenries (nH).

- Inductance of a free air toroid ($\mu_r = 1$) with the cross section of A = 5 cm^2, average field length of ℓ = 10 cm, and N = 100 turns is

$$L \sim \frac{(1)(4\pi 10^{-7}\ \text{H/m})(100)^2\ (5 \times 10^{-4}\ \text{m}^2)}{10 \times 10^{-2}\ \text{m}} = 6.28 \times 10^{-5}\ \text{H}$$

- Inductors in the µH range are used in switching regulators
- Small Relays usually have mH values of inductance
- Inductors in general typically range from a few Henries (H) to micro Henries (µH):

1 microHenry = 1 µH = 10^{-6} H
1 milliHenry = 1 mH = 10^{-3} H
1 Henry = 1 H

Figure 1.42 Relative Size of Inductance

Next, let us look at the electrical properties of inductors. When current flows through N tightly wrapped turns a magnetic field is created inside the core. The magnetic field or coil flux (ψ) in the core is proportional to the current (I) flowing through the coils. The proportionality factor is the inductance (L).

$\psi = L\ I$

- The change of magnetic field or coil flux (ψ) in an inductor is proportional to the change of electric current (I) flowing through the inductor's windings: $\psi \sim I$
- The proportional factor between coil flux and current is given by the inductance of the coil: $\psi = L\ I$

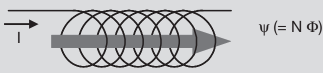


Figure 1.43 Inductors: Electrical Properties

Usually, the magnetic field is described with magnetic flux (Φ), created by the current flowing through a single turn. The magnetic field (magnetic flux) increases linearly with the number of windings (N). Therefore, it is often called the coil flux (ψ), where:

$\psi = N\ \Phi$

Let us look at a simple example. We source 2 A of current through a coil. The observed magnetic flux caused by the current is:

$\psi = 1\ \text{V s}$

This means that the inductor has an inductance of:

$L = \psi / I$

$L = 1\ \text{V s} / 2\ \text{A} = 0.5\ \text{H}$

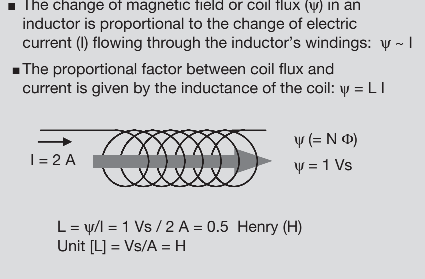


Figure 1.44 Inductors: Electrical Properties

Now, let us look at what happens when multiple inductors are used in a circuit. There are two basic configurations: serially connected inductors and parallel inductors (see Figure 1.45).

In a serial inductor configuration, the net inductance that current must flow through is increased. Therefore, the inductors in series add their inductances.

When the inductors are in parallel, however, they add like parallel resistors:

$1 / L = (1 / L_1) + (1 / L_2)$

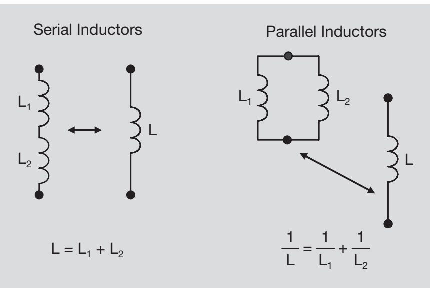


Figure 1.45 Serial and Parallel Inductance

Let us now look at Inductor Experiment #1. It consists of an ideal voltage source being connected directly to an ideal inductor by a switch.

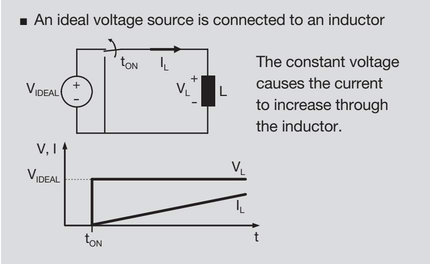


Figure 1.46 Inductor Experiment #1

The switch is closed at time t_{ON}. We then monitor the current through the inductor, and the voltage across the inductor. From the Figure 1.46, we see that the voltage is a step function. Before t_{ON}, there is no voltage across the inductor. After t_{ON}, the inductor voltage (V_L) is equal to the ideal voltage source (V_{IDEAL}). The current (I_L) begins to flow through the inductor, increasing linearly.

In Inductor Experiment #2, we examine what happens after the ideal voltage source is disconnected from the inductor (see Figure 1.47) and the inductor is shorted.

At t_{OFF}, the switch is opened and V_L is shorted to 0 V. When this occurs, the inductor current which was linearly increasing levels off and will theoretically remain constant at its t_{OFF} value.

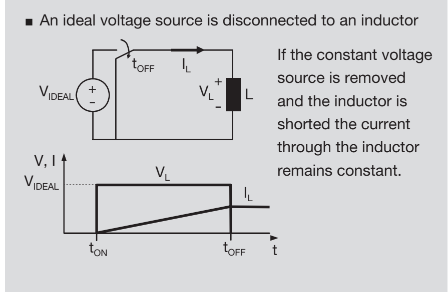


Figure 1.47 Inductor Experiment #2

Next (Inductor Experiment #3), we examine what happens when the magnitude of the ideal voltage source is changed. Figure 1.48 illustrates how the current through the inductor (I_{L1}) increases in response to the first ideal voltage source value (V_{L1}).

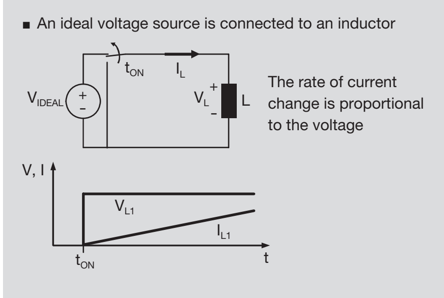


Figure 1.48 Inductor Experiment #3

Next in Inductor Experiment #3a, we increase the magnitude of the ideal voltage source to V_{L2} (see Figure 1.49). We notice that the current through the inductor (I_{L2}) is still linearly increasing, but at a faster rate.

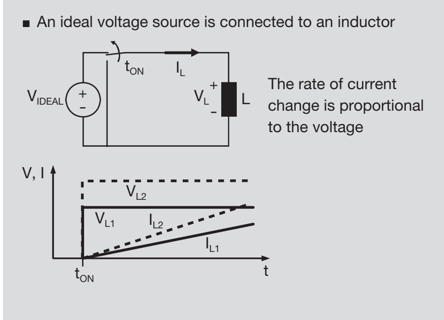


Figure 1.49 Inductor Experiment #3a

Next (Inductor Experiment #4), let us consider what happens when an ideal voltage source is connected to an inductor through a resistor at t_{ON}. The results are shown in Figure 1.50. Let us consider the voltage across the inductor (V_L) first. Before t_{ON}, the ideal voltage source is disconnected from the inductor, so there is no voltage drop across the inductor. Immediately at t_{ON}, $V_L = V_{IDEAL}$. It is only after t_{ON} that the voltage across the inductor begins to slowly decrease. After an extended time interval V_L is almost equal to 0 V.

At t_{ON}, the inductor acts like an open circuit and $I_L = 0$ A. With exponentially increasing current, V_L falls as the voltage across the resistor increases. Eventually, when V_L falls to 0 V and the final value of I_L is:

$$I_L = V_{IDEAL} / R$$

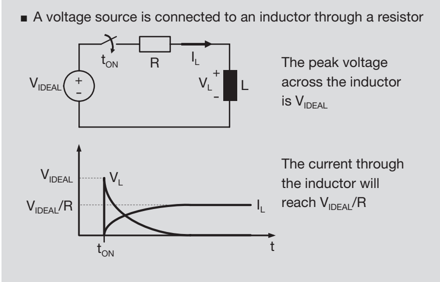


Figure 1.50 Inductor Experiment #4

Next (Inductor Experiment #5), let us consider what happens if we make the resistor variable.

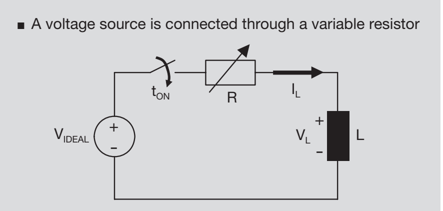


Figure 1.51 Inductor Experiment #5

First, we let the resistor be equal to R_1. It is shown in Figure 1.52 how the inductor voltage (V_{L1}) is initially V_{IDEAL}. R_1 is the benchmark resistance we will use for comparison in Experiments #5a-cAs V_{L1} approaches 0 V, I_{L1} increases to V_{IDEAL} / R_1.

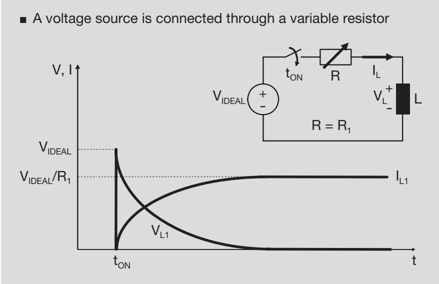


Figure 1.52 Inductor Experiment #5a

Next (Experiment #5b), we change the value of the resistor. We let the resistance be R_2, where R_2 is less than R_1 (see Figure 1.53).

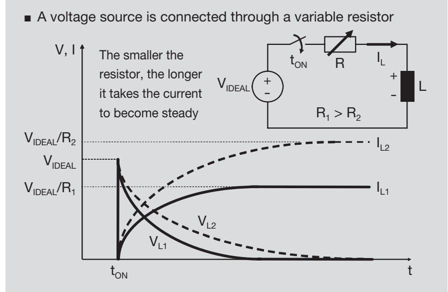


Figure 1.53 Inductor Experiment #5b

We first note that at time t_{ON}:

$$V_{L1} = V_{L2} = V_{IDEAL}$$

However, V_{L2} decays to 0 V more slowly than V_{L1}. This means it takes I_{L2} longer to reach its maximum value, V_{IDEAL} / R_2.

In Experiment #5c it is examined what happens when the resistance is changed once more. Let $R_3 > R_1$. As expected, at t_{ON}:

$V_{L1} = V_{L3} = V_{IDEAL}$

This time, V_{L3} decays to 0 V more quickly than V_{L1}. This means it takes I_{L3} less time to approach its final value, V_{IDEAL} / R_3.

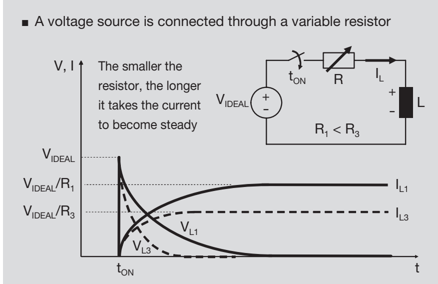


Figure 1.54 Inductor Experiment #5c

The last experiment (Experiments #6, #6a) with inductor circuits a constant resistance, R is used. The ideal voltage source is connected to the inductor through R. Starting at t = 0 as the time changes the inductor voltage V_L is examined.

As expected, V_L eventually decays 0 V. What we also notice, however, is that at time equal to t_C:

$t_C = L / R$

the voltage across the inductor has fallen approximately 63%. For example:

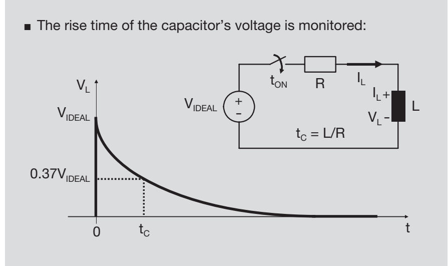


Figure 1.55 Inductor Experiment #6

V_{IDEAL} = 1 V, R = 1 kΩ and L = 1 mH:

$t_C = (1 \times 10^{-3}\ H) / (1 \times 10^{3}\ \Omega) = 1\ \mu s$

One microsecond after the 1 V ideal voltage source is connected, the voltage across the inductor is approximately 0.37 V. This is true for all values of V_{IDEAL}, R, and L. For example:

V_{IDEAL} = 2 V, R = 10 kΩ, L = 1 µH:

$t_C = (1 \times 10^{-6}\ H) / (1 \times 10^{4}\ \Omega) = 100\ ps$

V_{IDEAL} at 100 ps = 0.37 (2 V) = 0.74 V

V_{IDEAL} = 12 V, R = 37 Ω, L = 22 µH:

$t_C = (22 \times 10^{-6}\ H) / (37\ \Omega) = 595\ ns$

V_{IDEAL} at 595 ns = 0.37 (12 V) = 4.44 V

Therefore, t_C is also an important value for resistor and inductor circuits and it is called the "time constant" of RL circuits.
Note, sometimes, t_C is written as τ or τ_C.

If we continue to examine the voltage across the inductor for time beyond t_C, we notice additional characteristics of RL circuits. For example:

at $t = t_C$, $V_L = 0.37V_{IDEAL}$

at $t = 2t_C$, $V_L = 0.13V_{IDEAL}$

at $t = 3t_C$, $V_L = 0.05V_{IDEAL}$

at $t = 4t_C$, $V_L = 0.02V_{IDEAL}$

at $t = 5t_C$, $V_L = 0.01V_{IDEAL}$

Because these values hold constant for all resistor-inductor circuits, designers often have to consider them in their design and analysis of circuits.

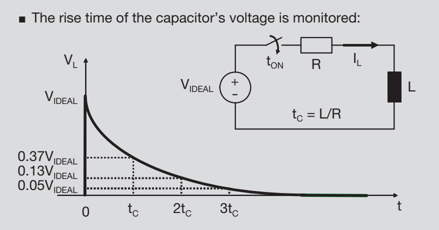


Figure 1.56 Inductor Experiment #6a

For example, at $3t_C$ after the voltage source is connected, the designer knows that the voltage across the inductor will have fallen 95% from its maximum value. Likewise, at $5t_C$, the inductor voltage will have fallen 99% from its maximum value. The $1t_C$, $3t_C$, and $5t_C$ times are often used to approximate how long the decay process will take.

Note, for those interested in a more mathematical description, the value of the voltage across the inductor and the inductor current are given by:

$V_L = V_{IDEAL}[\exp(-t / t_C)]$

$I_L = (V_{IDEAL} / R).[1 - \exp(-t / t_C)]$

Finally, the mathematical relationship between the current flowing through the inductor (I_L) and the voltage across the inductor (V_L) is developed.

The self induced coil voltage in an alternating magnetic field is given by:

$v_{ind}(t) = -N\, d\phi / dt = -\, d\psi / dt$

■ The self induced coil voltage when exposed to an alternating magnetic field is proportional to the change of coil flux vs. time:

$$v_{ind} = -N\frac{d\phi}{dt} = -\frac{d\psi}{dt}$$

Figure 1.57 Development of Mathematical Inductor Model: I_L vs. V_L

The voltage across the inductor, V_L, is always directly opposed to the self-induced voltage, V_{ind}.

$V_L = -V_{ind}(t) = d\psi / dt$

Rearranging yields:

$d\psi = V_L\, dt$

We also recall that the coil flux is proportional to the inductor current (I_L):

$\psi = I_L\, L$

Taking the derivative:

$d\psi / dt = L\, dI_L / dt$

■ The self induced coil voltage when exposed to an alternating magnetic field is proportional to the change of coil flux vs. time:

$$v_{ind} = -N\frac{d\phi}{dt} = -\frac{d\psi}{dt}$$

■ The voltage v applied across an inductor is always directly opposed to the self induced voltage v_{ind}:
$v = -v_{ind} = Nd\phi/dt = d\psi/dt$ (=> $d\psi = vdt$)

$$v = -v_{ind} = N\frac{d\phi}{dt} = \frac{d\psi}{dt} \quad \text{or} \quad d\psi = v\,dt \qquad (1)$$

■ The inductance is defined as coil flux vs. coil current, $L = \psi / I_L$, differentially expressed as:

$$L = \frac{d\psi}{di} \quad \text{or} \quad d\psi = L\,di \qquad (2)$$

Figure 1.58 Development of Mathematical Inductor Model: I_L vs. V_L

If we set the $d\psi$ equations equal to each other, we obtain the relationship between the voltage across the inductor and the current through the inductor:

$v_L\, dt = d\psi = d\psi = L\, di$

$V_L = L\, dI_L / dt$

Therefore, the voltage across the inductor is equal to its inductance times rate of change of inductor current.

$$v = -v_{ind} = N\frac{d\phi}{dt} = \frac{d\psi}{dt} \quad \text{or} \quad d\psi = v\,dt \qquad (1)$$

$$L = \frac{d\psi}{di} \quad \text{or} \quad d\psi = L\,di \qquad (2)$$

- Setting (1) equal to (2), the voltage -current relation for an inductor equals can be found:

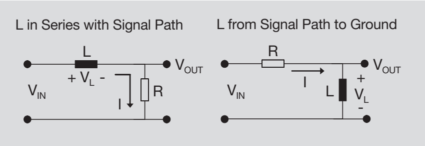
$$v = L\frac{di}{dt}$$

Figure 1.59 Development of Mathematical Inductor Model: I_L vs. V_L

There are the two basic ways that inductors are used in resistor-inductor voltage divider circuits.

First, the inductor can be in series with the signal path (see Figure 1.60, left circuit). Second, the inductor can be from the signal path to ground (see Figure 1.60 right circuit).

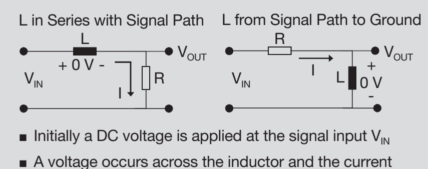
Let us first look at how these two circuits operate when a constant amplitude step voltage is applied to the signal path through a switch. In both cases, the voltage across the inductor is initially V_{IN} and I_L is 0 A. As I_L increases, the voltage across the resistor increases and across V_L decreases.



- Initially a DC voltage is applied at the signal input IN
- A voltage occurs across the inductor and the current through the inductor increases

Figure 1.60 Inductor & Resistor Networks

Eventually, the voltage across the inductor falls to 0 V. The (final) value of the current flowing through the inductor is approaching its maximum of V_{IN} / R.

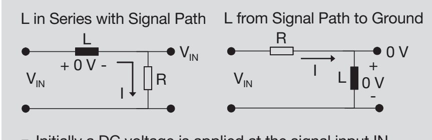


- Initially a DC voltage is applied at the signal input V_{IN}
- A voltage occurs across the inductor and the current through the inductor increases
- When the current through the inductor is at its maximum and remains constant, the voltage across the inductor equals zero

Figure 1.61 Inductor & Resistor Networks

If the inductor is in series with the signal path (see Figure 1.62 on the left), there is no voltage drop across the inductor, therefore V_{OUT} is equal to V_{IN} in the steady state. All the voltage drop is across the resistor.

If the inductor is from the signal path to ground (see Figure 1.62 on the right), V_{OUT} is equal to 0 V, because there is again no voltage drop across the inductor in the steady state.



- Initially a DC voltage is applied at the signal input IN
- A voltage occurs across the inductor and the current through the inductor increases
- When the current through the inductor is at its maximum and remains constant, the voltage across the inductor equals zero
- Depending on the inductor's placement, $V_{OUT} = V_{IN}$ or $V_{OUT} = 0$ V

Figure 1.62 Inductor & Resistor Networks

Now, let us look at a circuit with an inductor in series with the signal path more closely (see Figure 1.63). At time t_1, the input voltage is applied to the series inductor and resistor. The node voltage at V_X immediately becomes V_{IN}. The inductor current starts at 0 A, and begins to increase. Since the $V_{OUT} = I\,R$, V_{OUT} increases until:

$$I = V_{IN} / R$$

$$V_{OUT} = (V_{IN} / R)\,R = V_{IN}$$

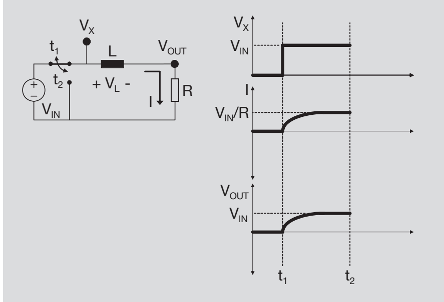


Figure 1.63 Inductance in Series with Signal Path

Next, we see what happens when the inductor is switched to ground ($V_X = 0$) at time t_2.

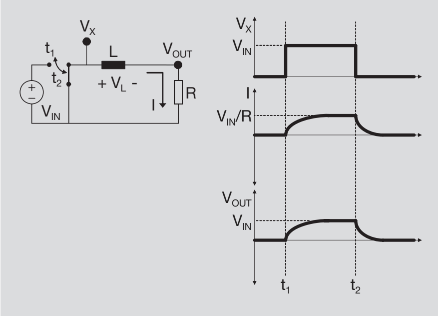


Figure 1.64 Inductance in Series with Signal Path

The self induced negative voltage (V_{OUT}) appears across the inductor which causes in the current to decay to 0 A. As the current decays, V_{OUT} also decays to 0 V.

Next, let us look at the circuit with an inductor from the signal path to ground. The input voltage is again applied to the circuit at time t_1. As before, the voltage at node V_X is immediately equal to V_{IN}. The voltage across the inductor is again initially V_{IN} and $I = 0$ A. The voltage across the inductor decays, and current is increasing to V_{IN} / R.

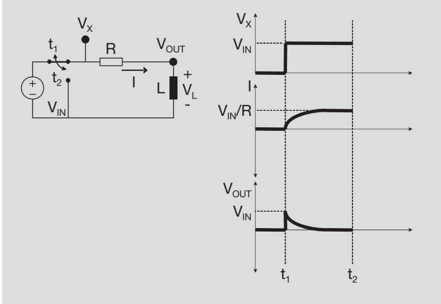


Figure 1.65 Inductance from Signal Path to Ground

Now, let us see what happens when the resistor is switched to ground at time t_2. V_X is immediately equal to 0 V. This allows the inductor to begin discharging.

Current continues in the same direction, from the inductor, to ground, to node V_X, through the resistor, to node V_{OUT}, and back to the inductor. Since the current flows in the same direction, it is positive, but decays to 0 A.

Initially V_{OUT} now falls below ground potential, as the inductor discharges:

$$I R + V_{OUT} = 0 \text{ V}$$

$$V_{OUT} = -I R \text{ (at } t = t_2)$$

Therefore, the output voltage immediately changes to $-V_{IN}$ at $t = t_2$ and decays to 0 V as the current decreases.

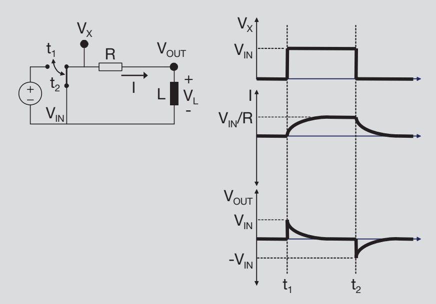


Figure 1.66 Inductance from Signal Path to Ground

What happens when an AC (sine wave) signal is applied to a resistor-inductor circuit?

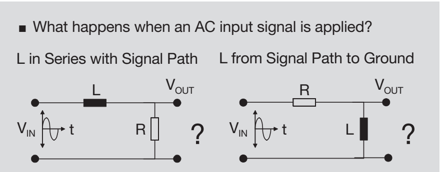


Figure 1.67 RL Networks - AC Signals

First, we need to introduce two new terms: reactance (X_L). Inductors behave like frequency dependent resistors:

$|X_L| = 2\pi fL$

The absolute value of the reactance of an inductor indicates that the inductor behaves like a frequency dependent resistance.

- Act like frequency dependent resistor (inductive reactance, X_L)

  $X_L \alpha\ f\ L$

- Instead of reactance, impedance (Z) used for circuit elements
- Impedance: The apparent opposition in an electrical circuit to the flow of alternating current that is analogous to the actual electrical resistance to a direct current

Figure 1.68 Inductors and AC Signals

The reactance specifically addresses the response an of ideal inductor to an AC signal. A non-ideal inductor always has, at least in theory, a series resistor (i.e. winding resistance) thus the coil **impedance** includes both the resistive and reactive behaviors. The reactance of an inductor actually consists of two types of information. First, the magnitude (absolute value) of the reactance determines the frequency dependent magnitude of the equivalent resistance of the inductor. Second, the reactance also indicates (will be shown how) that there is a 90° leading phase differential between the voltage across the inductor and the current through the inductor.

- Act like frequency dependent resistor (inductive reactance, X_L)

  $X_L \alpha\ f\ L$

- Instead of reactance, impedance (Z) used for circuit elements
- Impedance: The apparent opposition in an electrical circuit to the flow of alternating current that is analogous to the actual electrical resistance to a direct current
- The impedance of a circuit element represents its resistive and/or reactive components
- Besides the magnitude dependency between voltage and current the impedance Z gives also information about the phase shift between the two

Figure 1.69 Inductors and AC Signals

In Figure 1.70, we illustrate how magnitude of the inductor's reactance, $|X_L|$, varies vs. the frequency of the AC signal. $|X_L|$ is directly proportional to the frequency. Therefore, for DC conditions (0 Hz), the ideal inductor looks like a short circuit. But, in an AC circuit as the signal's frequency increases, the magnitude of the inductor's reactance increases.

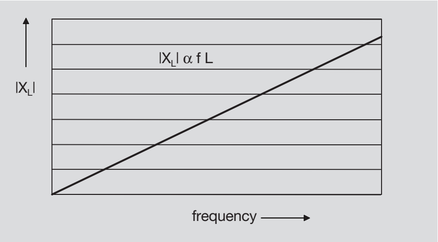


Figure 1.70 Inductor's Reactance Magnitude $|X_L|$ vs. Frequency

Now, let us look at the phase differential (φ) between the voltage across an inductor (V_L) and the current flowing through an inductor (I_L).

In Figure 1.71, we show two curves, the voltage across the inductor (V_L), and the current through the inductor (I_L). We note that the peak value of the AC current is equal to the peak value of the AC voltage divided by the magnitude of the inductor's impedance:

$I_{L,peak} = V_{L,peak} / |X_L|$

Note, $V_{L,peak}$ and $I_{L,peak}$ do not occur at the same point in time. In fact, the current lags behind the voltage by 90°.

When I_L is at its positive peak value, V_L is 0 V.

When I_L becomes 0 A, V_L is at its negative peak value.

When I_L reaches its negative peak value, V_L increases to 0 V.

When I_L increases to 0 A, V_L increases to its positive peak value.

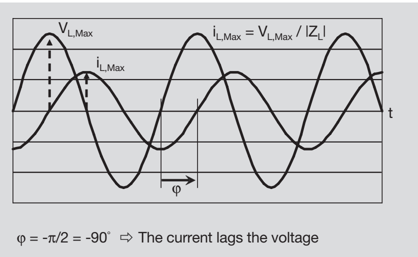


Figure 1.71 Inductors and AC Signals

For the mathematically inclined, we note this behavior is predicted by our earlier derivation of the mathematical inductor model:

$v \sim di / dt$

If the inductor current is a sinusoid:

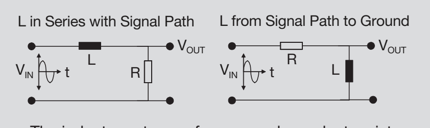
$i_L(t) = I_{L,Max}\sin(2\pi ft)$

We can see that the voltage across the inductor will be:

$v_L(t) = L.d\,[I_{L,Max}\sin(2\pi ft)] / dt$

$v_L(t) = 2\pi fL\, I_{L,Max}\cos(2\pi ft)$

In summary, inductors act as frequency dependent resistors. The reactance provides the relationship between the voltage across the inductor and the current flowing through the inductor.



- The inductor acts as a frequency dependent resistor
- It determines the current magnitude at a given voltage
- It causes a 90 degree phase shift between the inductor current and voltage across the inductor

Figure 1.72 RL Networks – AC Signals

The magnitude of the inductor's reactance is directly proportional to the applied AC signal's frequency:

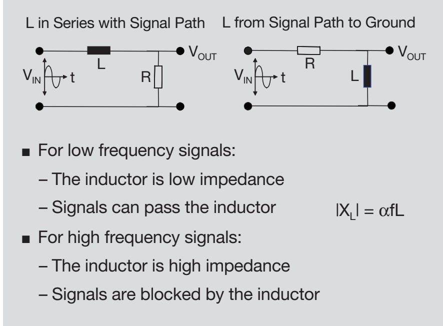


Figure 1.73 RC Networks – AC Signals

This is illustrated in Figure 1.74. First, we see a resistor-inductor circuit in a "low pass" configuration. We see that for a high frequency input signal, the output is significantly attenuated ($V_{OUTpeak} / V_{INpeak}$ = 1.9 V / 5 V). At the lowest frequency shown, the output is not significantly attenuated ($V_{OUTpeak} / V_{INpeak}$ = 4.8 V / 5 V).

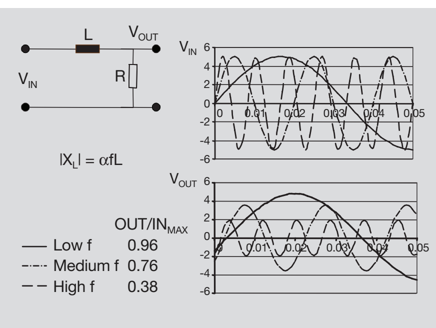


Figure 1.74 L in Series with Signal Path Low Pass Configuration

Next, we see a resistor-inductor circuit in a "high pass" configuration. For a low frequency input signal, the output is significantly attenuated ($V_{OUTpeak} / V_{INpeak}$ = 1.6 V / 5 V). At the highest frequency shown, the output is not significantly attenuated ($V_{OUTpeakx} / V_{INpeak}$) = 4.6 V / 5 V).

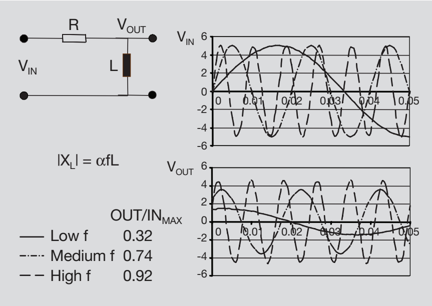


Figure 1.75 L from Signal Path to Ground High Pass Configuration

So, let us review a few points before we close our introduction to inductors.

Inductors-resistors are used, as voltage dividers, in one of two different configurations: the inductor is in series with the signal path or it is in the signal path to ground.

When a DC voltage is applied to an inductor, it initially blocks current. As the inductor current increases, the voltage across the inductor falls, eventually to 0 V, while the inductor current approaches the final value.

For AC signals, inductors behave like frequency dependent resistors. At high frequencies, they are very effective in blocking current flow. At low frequencies they represent low impedance to the current flow.

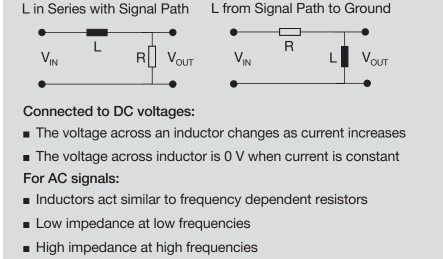


Figure 1.76 Inductor & Resistor Networks Summary

Finally, for reference, we compare the characteristics of electrical and magnetic relations. **Notice the symmetry between the two phenomena:**

Electrical Terms and Symbols		**Units**
Electrical Field Strength	E	V/M
Charge	Q	AS
Current: I = dQ / dt or ΔQ / Δt (rate of change of charge)	I	A
Capacitance: C = Q / V	C	AS/V
Permittivity of Vacuum $\epsilon_0 = 8.854 \times 10^{-12}$	ϵ_0	AS/VM
Energy Stored in a Capacitor: $E = C V^2 / 2$	E	VAS J
Constant Current (I) Charging a Capacitor	V = I t / C	

Magnetic Terms and Symbols		**Units**
Magnetic Field Strength	H	A/M
Coil Flux (= N*Φ)	ψ	VS
Voltage = -N(dΦ / dt) = –NΔΦ / Δt (negative rate of change of flux times the number of turns)	V	V
Inductance: L = ψ / I	L	VS/A
Permeability of Vacuum: $\mu_0 = 4\pi 10^{-7}$	μ_0	VS/AM
Energy Stored in an Inductor: $E = L I^2 / 2$	E	VAS J
Constant Voltage (V) Charging an Inductor	I = V t / L	

1.3 Example Load Model: Turning on an Incandescent Lamp

Let us conduct another experiment. In Figure 1.77, you see a plot from an oscilloscope and an inset picture of an simple electrical circuit. In the circuit, a 14 V ideal voltage source is connected to an incandescent light bulb.

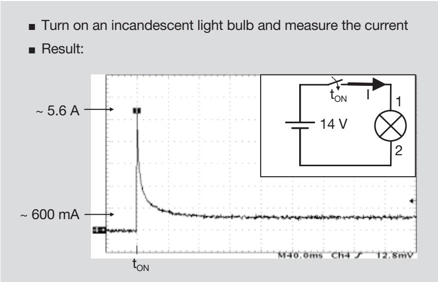


Figure 1.77 Lamp Experiment

At t_{ON}, the voltage source is connected to the light bulb. The current through the light bulb surges to 5.6 A. Then, the current decays slowly, over a period of about 50 ms to a steady-state value of 600 mA.

The question is, can we electrically model this type of behavior with a simple circuit composed of resistors, capacitors, and inductors? We start by assuming the lamp load is a simple black box: we don't know anything of what is inside:

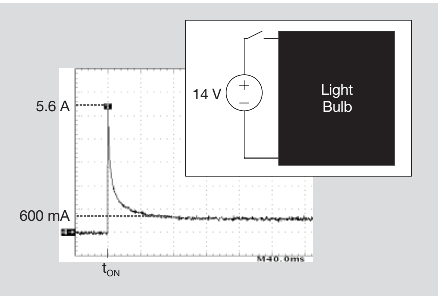


Figure 1.78 Developing a RC Load Model for an Incandescent Light Bulb

First, we consider the steady state case. If the voltage source is connected to the lamp, eventually (~50 ms) the load current settles at 600 mA. Therefore, we model part of the lamp electrically as a resistor:

$R_1 = V / I$

$R_1 = 14\ V / 0.6\ A = 23.3\ \Omega$

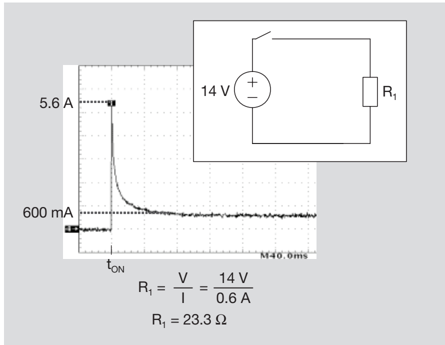


Figure 1.79 Developing a RC Load Model for an Incandescent Light Bulb

Next, we note that the lamp's behavior initially appears capacitive. The load current starts high, but decays similarly to how a series resistor-capacitor is charged. Therefore, we place a capacitor and resistor in parallel to the first 23.3 Ω resistor. (We note that a resistor is in series with the blocking capacitor because the current at t_{ON} is limited by something to 5.6 A.)

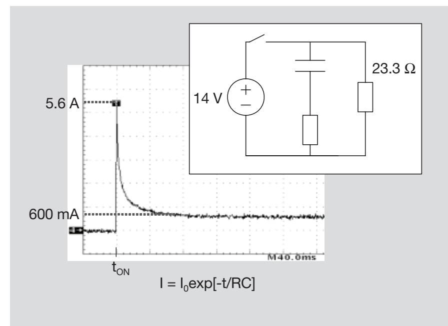


Figure 1.80 Developing a RC Load Model for an Incandescent Light Bulb

How do we calculate the value of the second resistor? At t_{ON}, the voltage across the capacitor is 0 V. (Therefore, we can consider it

a short circuit at t_{ON}). The start of the surge current is totally determined by the parallel equivalent resistance of R_2 and the 23.3 Ω resistor.

$R_{EQUIVALENT} = V / I$

$R_{EQUIVALENT} = 14\ V / 5.6\ A$

$R_{EQUIVALENT} = 2.5\ \Omega$

$1 / R_{EQUIVALENT} = (1 / R_1) + (1 / R_2)$

$1 / R_1 = (1 / 2.5\ \Omega) - (1 / 23.3\ \Omega)$

$1 / R_1 = 0.357\ \Omega^{-1}$

$R_1 = 2.80\ \Omega$

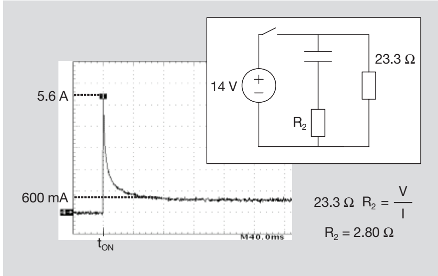


Figure 1.81 Developing a RC Load Model for an Incandescent Light Bulb

Finally, we calculate the value of the series capacitor. The current decay in the series capacitor circuit is given by:

$I = I_0 \exp[-t / t_C]$

where $t_C = R_1C$. We approximate the value of the capacitance by noting that the current value decays approximately 63% in about 10 ms (t_C = 10 ms).

$10\ ms = t_C = R_1C \sim (2.8\ \Omega)(C)$

So, the value of the capacitor is approximately:

$C = 10\ ms / 2.8\ \Omega$

$C = 3.6\ mF$

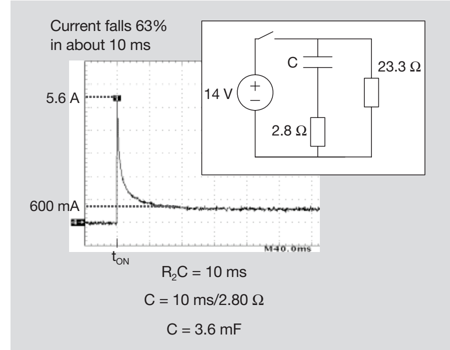


Figure 1.82 Developing a RC Load Model for an Incandescent Light Bulb

To verify our values, we build the lamp model using R_1, R_2 and C and compare the its performance to that of the lamp circuit. Below we show the lamp model (inset) together with the simulated bulb current. You can see that the simulation results are in very good agreement with the incandescent bulb's performance.

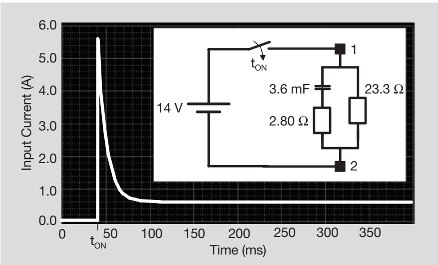


Figure 1.83 Simulation of Lamp RC Model

We will add a few comments to the lamp behavior. When an incandescent bulb is initially turned on, the filament is cold and has a low resistance (about 2.5 Ω in our example). However, as the bulb warms up, the resistance of the filament increases dramatically (to 23.3 Ω in our example). On the other hand, if a bulb is turned-off briefly and then turned back on quickly, the filament does not cool sufficiently to allow the filament resistance to drop back to 2.5 Ω. The resistance in series with the capacitor is a function of the lamp filament temperature. If we wanted to continue to refine our electrical model of the incandescent bulb, we would need to perform additional tests.

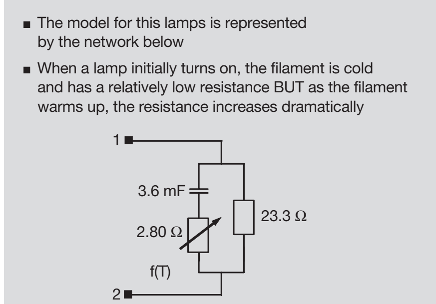


Figure 1.84 A RC Load Model for Incandescent Light Bulbs

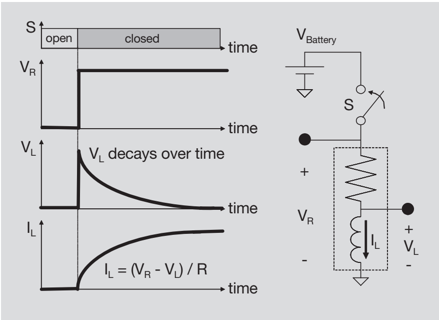


Figure 1.86 Switching On a Relay

The process of turning off a relay, however, is slightly more complicated. When the switch is opened, the relay current has to become 0 A.

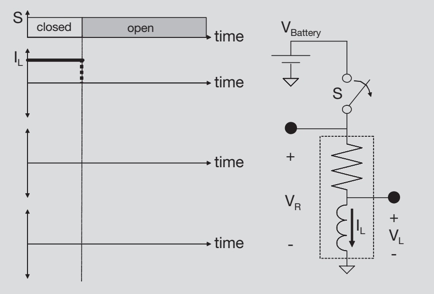


Figure 1.87 Switching Off a Relay

1.4 Example Load Model: Switching a Relay

In Figure 1.85, we show an example of an electrical model of a relay connected to an ideal voltage source by a switch. Relays are modeled as a low value resistor in series with an inductor.

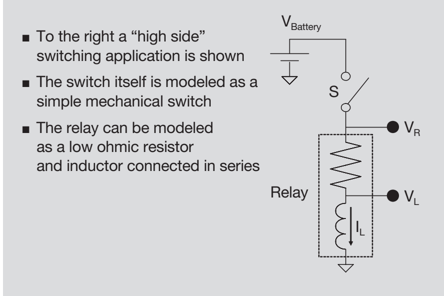


Figure 1.85 Switching a Relay

Let the ideal voltage source $V_{Battery}$ be connected to the relay. When the switch is closed, the common nodes of V_R and V_L immediately will be at $V_{Battery}$, so the current flow is zero. However, as the blocking voltage of the inductor decays, V_L falls to 0 V. The relay current is given by

$I_L = (V_R - V_L) / R \rightarrow 0.$

However, from out mathematical inductor model:

$V_L = L\, di / dt$

So, if the current becomes zero instantaneously that would imply infinite di / dt slope resulting in an inductive voltage surge of infinity (at least in theory).

There will be a voltage drop across the inductor caused by the rapid, decreasing current change Since the change in current is decreasing it causes a large negative V_L pulse. The result of the large voltage pulse is contact arcing and that will provide a path for the inductive and decaying current through the arc.

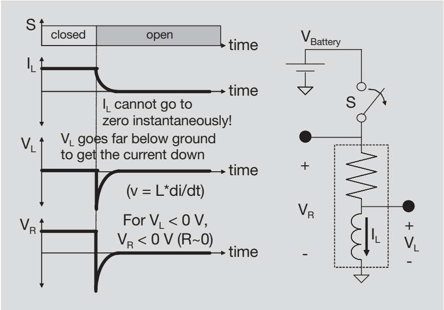


Figure 1.88 Switching Off a Relay (2)

In summary when arcing occurs a current path develops while maintaining a large voltage drop across the contact.

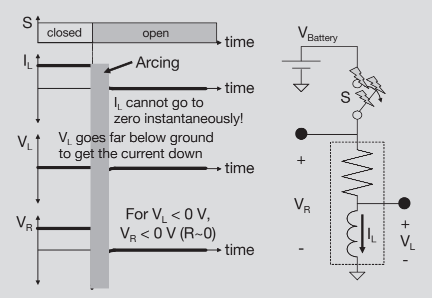


Figure 1.89 Switching Off a Relay (3)

The large negative voltage spike and accompanying arcing can damage sensitive semiconductor components. To minimize this problem, circuit designers will often place a power diode parallel to the relay. The diode will “clamp” node voltage V_R to approximately 0.7 V below ground. When V_R falls to 0.7 V, current will flow through the diode (I_D) and V_R will stay at that level while current is flowing.

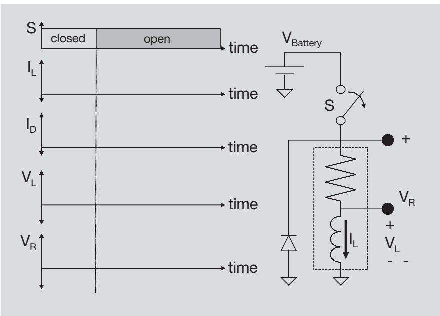


Figure 1.90 Switching Off a Relay no Arcing (1)

So, now when the switch is opened, a recirculation current path exists for the relay. Current flows from the inductor, to ground, through the diode, to node V_R, through the resistor, to node V_L, and back through the inductor.

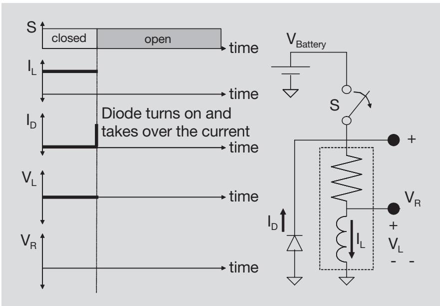


Figure 1.91 Switching Off a Relay no Arcing (2)

If the resistance of the relay is relatively small (approximately 0 Ω), V_R and V_L will be at -0.7 V when the switch is opened. With a constant voltage across inductor, the inductor recirculation current will decay in a linear fashion but much slower:

$$di / dt = -0.7 V / L$$

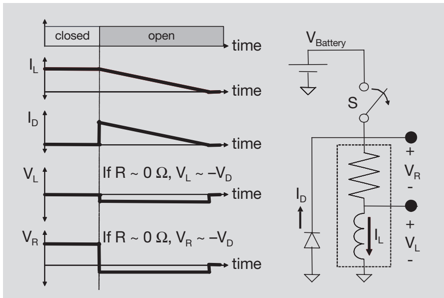


Figure 1.92 Switching Off a Relay no Arcing (#3)

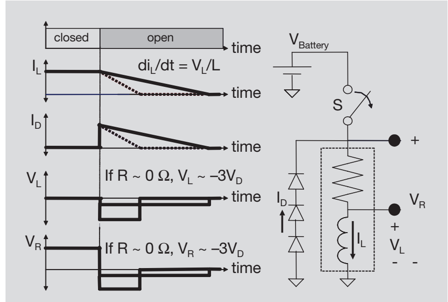


Figure 1.94 Switching Off a Relay no Arcing (#5)

Sometimes, the designer will want to prevent a large negative voltage spike when the relay is turned off, but may also want the recirculation current to decay more quickly. This can be accomplished by adding additional recirculation diodes (three in the Figure 1.93).

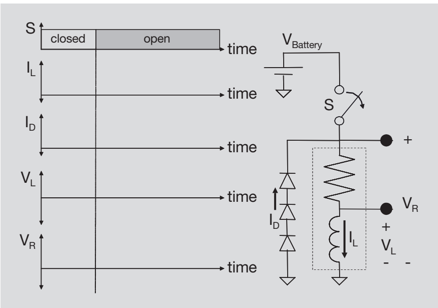


Figure 1.93 Switching Off a Relay no Arcing (#4)

More recirculation diodes will speed up the decay of the current:

di / dt = -2.1 V / L,

see the graphs #4 and #5. The dashed lines represent the shorter decay cycle for the three diode implementation.

2. Introduction to Semiconductors

In this chapter, we will be introducing the subject of semiconductor materials.

We begin the chapter by looking at three broad categories of materials: conductors, insulators, and semiconductors. We will examine similarities and differences in these three broad categories.

Next, we will explain the two fundamental types of semiconductor materials, n-type semiconductors and p-type semiconductors. This will allow us to introduce the simplest of all semiconductor devices, the diode.

Building upon our understanding of diodes, we will examine two types of transistors: bipolar junction transistors (BJTs) and metal-oxide-semiconductor field effect transistors (MOSFETs). The physical structure of the two transistor types will be described. Finally, we will close the chapter with a basic explanation of how the transistors work.

2.1 What is a Semiconductor?

We begin by exploring what makes a semiconductor different from conductors and insulators.

Crystals are solid materials composed of atoms arranged in a uniform repeating pattern. When carbon atoms are arranged in a crystalline structure, a diamond is formed. In a diamond crystal, each carbon atom is bonded to four other carbon atoms. Each of these four carbon atoms are in turn connected to an additional four carbon atoms. This is repeated throughout the entire crystalline diamond.

Silicon atoms can form uniform crystalline structures just like carbon:

- A crystal is a solid composed of atoms in a single, ***UNIFIED*** array
- Notice each ***silicon atom*** is "connected" to its **four nearest neighboring silicon atoms**

– Si – Si – Si – Si – Si – Si – Si – Si –
– Si – Si – *Si* – Si – Si – Si – Si – Si –
– Si – Si – Si – Si – Si – Si – Si – Si –
– Si – Si – Si – Si – Si – Si – Si – Si –

Figure 2.1 Silicon Crystal Structure

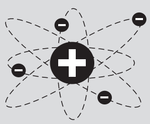
So, why does this occur?

At the center of an atom you will find a (relatively) large nucleus. This nucleus has a positive charge. Around the positively charged nucleus, small negative electrons "orbit". The positive charge of the nucleus is balanced by the identical (but opposite) electron charge. If a nucleus has a positive charge of six, it has six negatively charged electrons. These electrons are arranged in shells, or layers. Many atoms, including carbon and silicon, like to have eight electrons in their outer shell (or outer orbit).

Carbon and silicon atoms, however, only have four electrons in their outer shells Figure 2.2.

Therefore, these atoms like to share electrons with each other.

- Many atoms (including silicon) like to have **EIGHT** electrons in their outer "orbit"
- But, silicon only has **FOUR** outer electrons



- Solution: Silicon atoms share electrons with each other

Figure 2.2 Sharing Electrons

What does this look like?

Figure 2.3 is a two dimensional representation of a silicon crystal.

Let's look first at the silicon atom in the middle (light grey). The four light grey dots represent the four electrons in the medium grey silicon atom's outer shell. As we said before, silicon atoms prefer to have eight electrons in their outer shell. Therefore, they "bond" with other silicon atoms (medium grey). Each medium grey silicon atom shares one of its electrons with a neighboring (i.e. dark) silicon atom. This gives the light grey silicon atom eight electrons in its outer shell (four of its own, and four are shared).

Each of the medium grey silicon atoms also wants eight electrons in their outer shells. They already have four of their own electrons, and they share one of the light grey silicon atoms electrons. Therefore, each of the medium grey silicon atoms needs three more electrons in their outer shell to be happy. Therefore, they bond with three other silicon atoms (dark grey).

Figure 2.3 Silicon Atoms Sharing Electrons in a Crystal

This process is repeated throughout the entire silicon crystal. Each silicon atom in the crystal has eight electrons in their outer shell (four of their own, and four they share with other silicon atoms).

Now, let us turn our attention to what can make a silicon crystal a semiconducting material.

Insulators (such as glass) do not conduct electricity very well. They have a very large electrical resistance. Conductors (like metal) do conduct electricity. They have a very small electrical resistance.

Semiconductors can be thought of as a third type of material. In a pure form, semiconductors act like insulators. They have a very large resistance and are poor conductors of electricity. However, the semiconductor can become very conductive (with very low resistance) when small amounts of differing atoms are added to it.

- A conductor is a material which "conducts" electricity easily (such as metals)
- An insulator is a material which is a very poor conductor of electricity (such as glass)
- A **semiconductor** (silicon) is a material which acts like an insulator, but behaves like a conductor when it is combined with other materials

Figure 2.4 Semiconductor Silicon

The addition of certain atoms (such as phosphorous or boron) to the silicon, however, does not change the overall charge of the crystal. The crystal remains overall charge neutral.

A phosphorous atom can be viewed as a nucleus with five positive charges and five negatively charged electrons in its outer shell. Boron, on the other hand, has a nucleus with a positive charge of three and three negatively charged electrons in its outer shell.

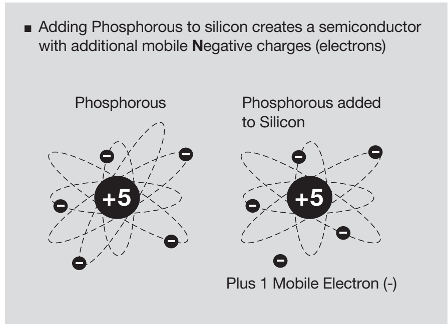


Figure 2.5 N-Type Silicon Charge Neutrality

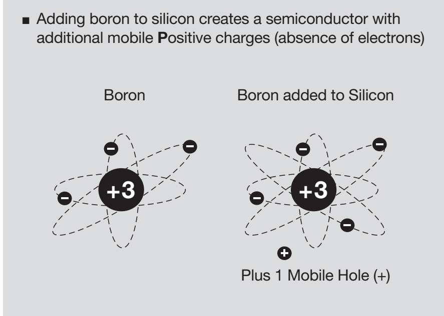


Figure 2.6 P-Type Silicon Charge Neutrality

Figure 2.5 and Figure 2.6 illustrate the effects of phosphorous and boron on silicon.

Figure 2.7 shows two blocks of silicon which are in close proximity to, but not in contact with each other.

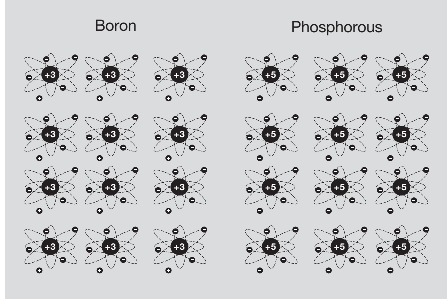


Figure 2.7 Boron and Phosphorous Atoms in a "Sea of Silicon"

The left side block contains boron atoms in a silicon crystal. This results in an absence of electrons (which can be thought of as positive mobile charges called holes) which are free to float throughout the sea of silicon atoms. The left side is a p-type semiconductor. The right side block contains phosphorous atoms in a silicon crystal. This results in negative mobile electrons which are free to float throughout the sea of silicon atoms. The right side is a n-type semiconductor.

Now, we will ignore the boron atoms and phosphorous atoms and look only at the positive mobile holes and negative mobile electrons. Next, we bring the two semiconductor blocks into contact. What do you thing will happen? -------Diffusion! Remember diffusion is a process where materials tend to move from an area of high concentration to areas of lower concentration.

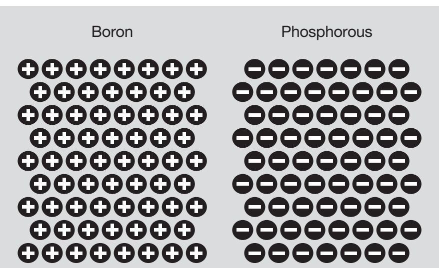


Figure 2.8 Boron and Phosphorous Mobile Charge in a Sea of Silicon

Let us look first at the p-type silicon block on the left side of Figure 2.8. There is a high concentration of positively charged mobile holes floating in the sea of silicon. Some of these holes diffuse to the n-type silicon block on the right where there is a low concentration of mobile holes.

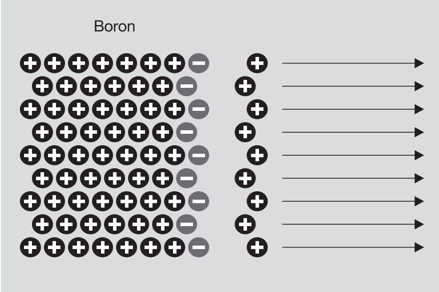


Figure 2.9 Boron Mobile Charge in a Sea of Silicon

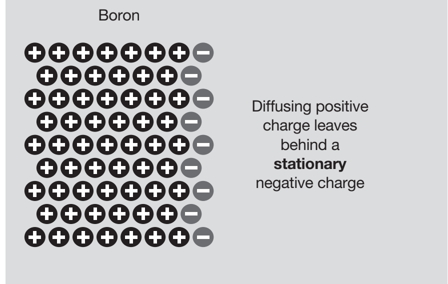


Figure 2.10 Boron and Phosphorous Mobile Charge in a Sea of Silicon

As we said before, n-type and p-type semiconductors are neutrally charged. Therefore, because some of the positive holes have moved to the right, they leave behind a net negative charge.

Next, we look at the n-type silicon block on the right side of Figure 2.8. There is a large concentration of negatively charged mobile electrons floating in the sea of silicon. Some of these electrons diffuse to the p-type silicon block on the left where there is a low concentration of mobile electrons.

Since some of the negative mobile electrons have moved to the left, they leave behind a net positive charge.

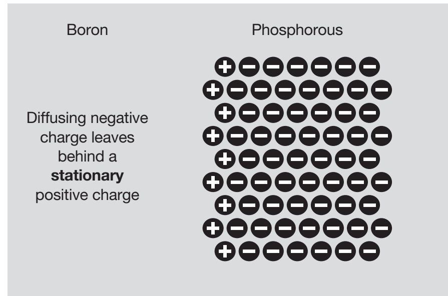


Figure 2.11 Boron and Phosphorous Mobile Charge in a Sea of Silicon

Let us look at the state of charges when we look at both sides together.

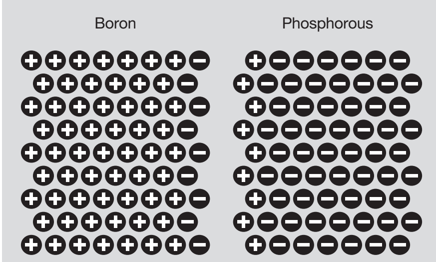


Figure 2.12 Boron and Phosphorous Mobile Charge in a Sea of Silicon

After the diffusion has started, the left hand side remains primarily a p-type semiconductor, but there is a thin strip with a negative charge near the junction between the two semiconductor types.

The right hand side remains primarily a n-type semiconductor, but there is a thin strip with a positive charge near the junction between the two semiconductor types.

The positive mobile holes on the left continue to try to diffuse to the right. As they cross the junction, however, they are repelled by the positive stationary charge that remains after the mobile electrons diffused to the left. Therefore, no more positive mobile holes diffuse to the n-type silicon.

The negative mobile electrons on the right continue to try to diffuse to the left. As they cross the junction, however, they are repelled by the negative stationary charge that remains after the mobile holes diffused to the right. Therefore, no more negative mobile electrons diffuse to the p-type silicon. An equilibrium is achieved between the mobile charges trying to diffuse in one direction and being repelled in the opposite direction by the net opposite charge. The thin strip between the n-type and p-type regions has very few mobile carriers and is often referred to as the “space charge” region or “depletion” region.

2.2 What is a Diode?

When p-type silicon and n-type silicon are in contact, they create a diode. Let us look at this in a simple circuit.

Here the p-type semiconductor and n-type semiconductors are again shown. The negative region in the p-type semiconductor and the positive region in the n-type semiconductor have been created by diffusion and the resulting equilibrium reached.

The remaining positive mobile holes in the p-type silicon try to diffuse to the right, but they are repulsed by the positive layer at the silicon junction (often called the pn junction).

The remaining negative mobile electrons in the n-type silicon try to diffuse to the left, but they are repulsed by the negative layer at the silicon junction (often called the pn junction).

The positive and negative areas at the pn junction create an intrinsic (or “built-in”) potential difference across the pn junction. For silicon this voltage is approximately 700 mV at room temperature.

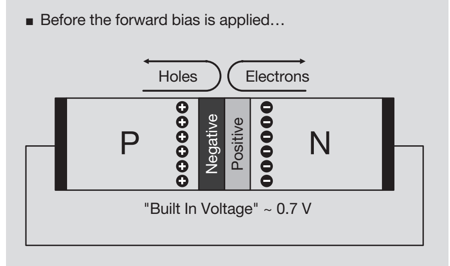


Figure 2.13 Diode under Forward Bias

Now, we examine what happens when a forward bias is applied to the semiconductor. (We call it forward bias when the positive voltage is applied to the p-type semiconductor block.)

Notice the applied bias is in the opposite direction of the built-in voltage at the pn junction. When voltage sources are in series they add. Consequently, applying a forward bias to the semiconductor serves to reduce the magnitude of the built-in voltage and stationary charge at the junction. Thus as the forward bias is increased, more mobile holes and mobile electrons are able to cross the pn junction in larger number without being repelled.

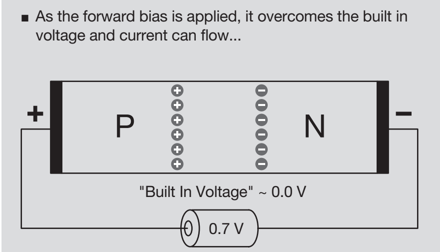


Figure 2.14 Diode under Forward Bias

These carriers cause a current to flow through the semiconductor. As the forward bias is increased, the current through the pn junction will also increase.

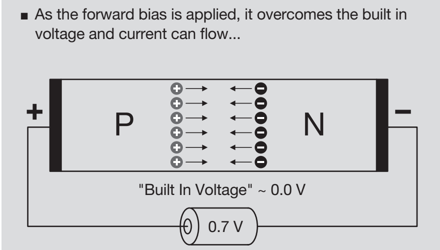


Figure 2.15 Diode under Forward Bias

In an idealized pn junction diode, the current flow vs. applied voltage would look like this:

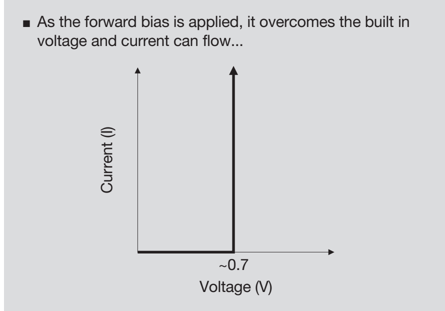


Figure 2.16 Diode under Forward Bias

The current, as a function of the V forward voltage, is exponential in reality instead of piece-wise linear (as illustrated in Figure 2.16).

$$I = I_0 e^{(Vq/kT - 1)}$$

I is the current through the pn junction

V is the voltage across the pn junction

I_0 is a reference current for the pn junction. It varies based upon the size of the pn junction, the amount and type of n-type and p-type dopants and other factors.

q is the electron charge (1.6×10^{-19} C)

k is the Boltzman's constant (1.38×10^{-23} J/K)

T is the temperature of the pn junction in Kelvin (or absolute) degrees

Let us now look at what happens when the applied bias is reversed (reverse bias).

Before the reverse bias is applied, the pn junction is electrically neutral:

- the remaining positive mobile holes in the p-type silicon try to diffuse to the right, but they are repulsed by the positive layer at the silicon junction (often called the pn junction);

- the remaining negative mobile electrons in the n-type silicon try to diffuse to the left, but they are repulsed by the negative layer at the silicon junction;
- the positive and negative areas at the pn junction create an intrinsic (or built-in) voltage across the pn junction.

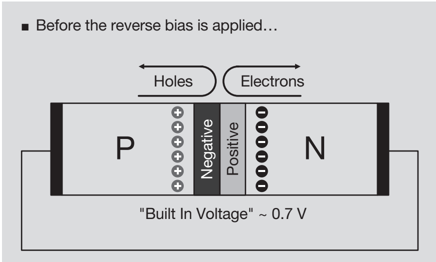


Figure 2.17 Diode under Reverse Bias

Now, let us see what happens when a reverse bias is applied to the semiconductor. (We call it reverse bias when the positive voltage is applied to the n-type semiconductor block.)

Notice the applied bias is in the same direction as the built-in voltage at the pn junction. When voltage sources are in series they add. Therefore, applying a reverse bias to the semiconductor serves to increase the magnitude of the built-in voltage and the stationary charge at the junction.

Positive mobile holes in the p-type semiconductor are drawn to the negative charge on the left side of the semiconductor block. Negative mobile electrons in the n-type semiconductor are drawn to the positive charge on the right side of the semiconductor block. This serves to increase the amount of charge at the pn junction.

As the reverse bias is increased, fewer mobile holes and mobile electrons are able to cross the pn junction.

This results in a very, very small current across the pn junction for reverse bias conditions.

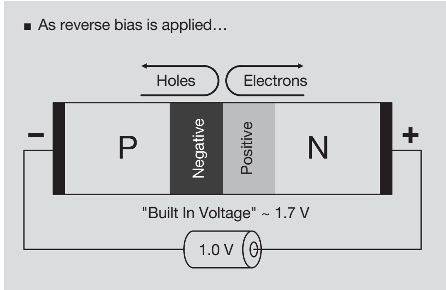


Figure 2.18 Diode under Reverse Bias

In summary, a diode consists of a p-type semiconductor in contact with an n-type semiconductor. At the junction of the two types is a layer of positive and negative charges created by the diffusion of mobile holes and electrons across the junction.

When a forward bias is applied, current through the diode grows exponentially with the bias. The electrical symbol for the diode shows the direction of the **forward** current flow by the solid arrow.

When a reverse bias is applied, very, very little current flows through the diode. The electrical symbol for the diode shows the blocking direction by a black line.

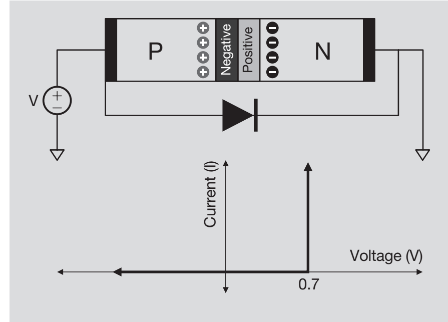


Figure 2.19 Ideal Diode Current-Voltage (I-V) Characteristics

Figure 2.20 shows the ideal (piece-wise linear) current-voltage characteristic compared to the actual exponential characteristic.

Although we have purposely not dealt with the mathematics behind the exponential nature of

current flow, the reader should have some intuitive feeling why current flows in one direction across a diode and not in the other direction due to the basic charge principles and built in voltage at the pn junction that we introduced in this section.

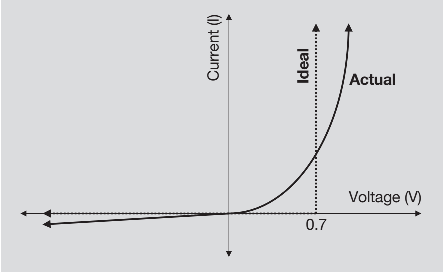


Figure 2.20 Actual Diode Current-Voltage (I-V) Characteristics

Next, we briefly mention that temperature does affect the actual exponential nature of the current flow.

Generally with semiconductors, increasing temperature allows charge to flow more freely through the semiconductor material.

Thus, for the forward voltage region one can see, for a given current level, the voltage decreases with increasing temperature (easier flow of charge).

Also, for the reverse voltage region, at a given reverse voltage, the leakage current flow is greater with increasing temperature (again, easier flow of charge at higher temperature).

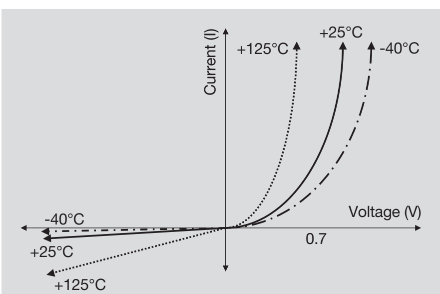


Figure 2.21 Diode I-V Temperature Characteristics

2.3 What is a Bipolar Junction Transistor?

In Figure 2.22 we show three semiconductor types in contact with each other. The middle section is a p-type with mobile holes. The left and right sections are n-type with mobile electrons.

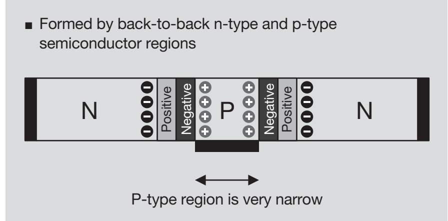


Figure 2.22 Bipolar Junction Transistor (BJT)

Because of the structure of this bipolar junction transistor (BJT), it is often referred to as an NPN BJT.

PNP BJT's are also used. They use a n-type semiconductor for the middle section and p-type semiconductor in the left and right sections.

Notice the two layers of positive and negative charge on either side of the middle section (basically two diode junctions back to back). The center region between these charge layers is very, very, very narrow. We will see why this is important shortly.

Finally, we note that there is an electrical connection to each of the three sections.

In Figure 2.23 an NPN BJT is shown, with each of its sections labeled. The thin center section is called the base which is usually labeled "B" on a transistor diagram. Think of it as the base of operations or the control center of the transistor.

The bottom n-type section is the emitter. This is the source of electrons flowing through the transistor. This terminal is labeled "E" on a transistor diagram.

Finally, the top n-type section is the collector. Electrons in the transistor current will flow from the emitter, through the transistor, and end up

at the collector. The label of this terminal is “C” on a transistor diagram. Note that the direction of current is **opposite** to the electron flow.

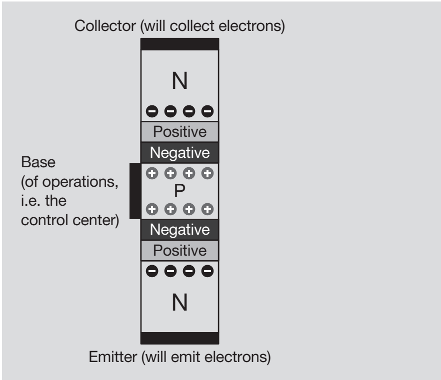


Figure 2.23 NPN Bipolar Junction Transistor (BJT)

Let us now look at the operation of an NPN BJT. We begin by showing the back-to-back diode structure inherent in the NPN BJT.

Let the base and emitter be shorted together. We will see that this effectively “turns off” the NPN BJT. A battery is connected between the collector and the emitter the collector is at a “high“ potential (positive), while the emitter (and the base) are grounded (negative).

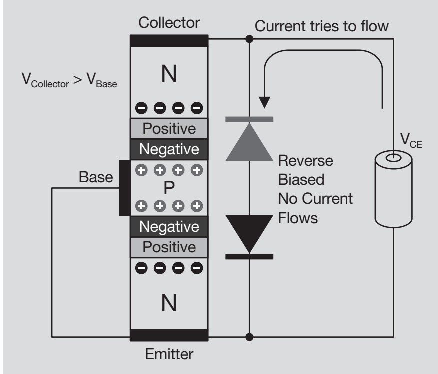


Figure 2.24 NPN Bipolar Junction Transistor (BJT)

We notice that the upper diode, formed by the pn junction between the collector and base, is reverse biased. The voltage applied to the n-type section is greater than the voltage applied to the p-type section. Therefore, when current tries to flow from the battery, it runs into the reverse biased diode, and no current can flow through the transistor.

Now we will take a look at what happens when we also apply a voltage to the base of the NPN BJT.

By applying a positive voltage at the base with respect to the emitter, the bottom diode is forward biased. Mobile electrons can move from the emitter into the base and mobile holes can move from the base into the emitter.

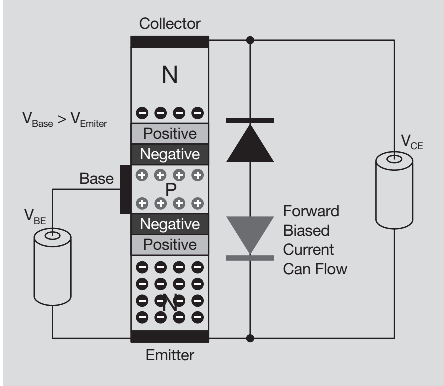


Figure 2.25 NPN Bipolar Junction Transistor (BJT)

Once current begins to flow between base and emitter, an excessive number of electrons congregate in the base region. This is achieved by design since the emitter is doped with a very high concentration of n-type material (much higher than the p-type dopant in the base region).

Since there is a very high concentration of electrons in the base region (and remember we said the base region is physically very thin by design), a large number of electrons will now be able to overcome the built-in base-collector voltage and flow to the collector. The positive voltage at the collector is attracting the electrons and causing them to flow across the built in voltage at the base-collector region.

The current flow from collector to emitter can be quite large compared to the base to emitter

current flow. Thus, the base current can be viewed as an input signal that causes a current path to open or close from emitter to collector.

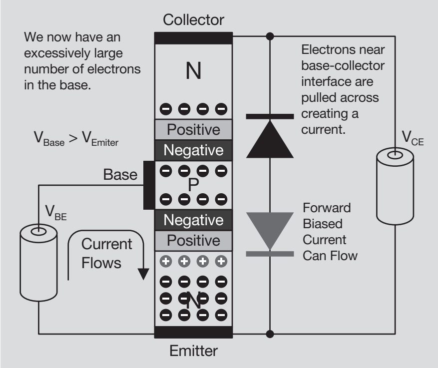


Figure 2.26 NPN Bipolar Junction Transistor (BJT)

The electrical symbol for the NPN BJT is shown in Figure 2.27:

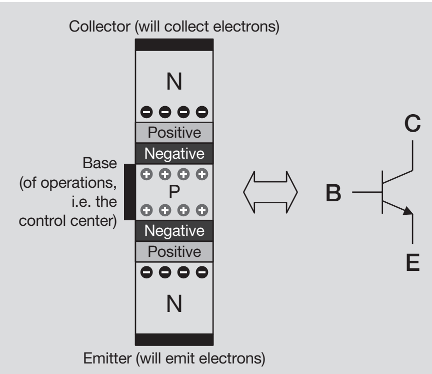


Figure 2.27 NPN Bipolar Junction Transistor (BJT)

The symbol represents a three terminal device (collector, base, and emitter). The arrow is always between the base and emitter, showing the direction of the controlling current flow (standard conventions define the direction for positive current flow - not electron flow). Again, when the transistor is on, current will flow from the collector to the emitter.

2.4 What is a MOSFET?

To understand how a MOSFET transistor works, we will go through the fabrication steps used in manufacturing to build one. The building of an n-channel MOSFET (also called an nMOSFET or an NMOS transistor) is illustrated. To build a p-channel transistor, simply change all the references from p-type to n-type and vice versa.

We begin with a block of p-type semiconductor (called the substrate). Remember, this is silicon with a p-type dopant (boron for example) added. The result is a large number of positive mobile holes in the substrate.

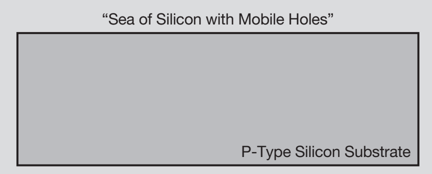


Figure 2.28 N-Channel MOSFET Cross Section (NMOS)

Next, a layer of glass (SiO_2 or silicon dioxide) is grown on top of the substrate by subjecting it to very high temperatures. Early in the presentation we noted that glass (SiO_2) was an excellent insulator. This is an important aspect of the operation of the MOSFET.

Two holes are then etched in the glass with acid, exposing the p-type substrate.

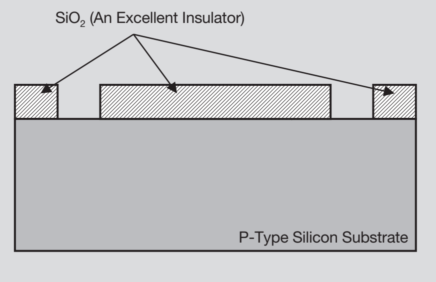


Figure 2.29 N-Channel MOSFET Cross Section (NMOS)

N-type dopants (phosphorous for example) are then injected into the substrate through the holes in the glass.

A sufficient amount of n-type material is added to change the area below the holes from p-type silicon to n-type silicon.

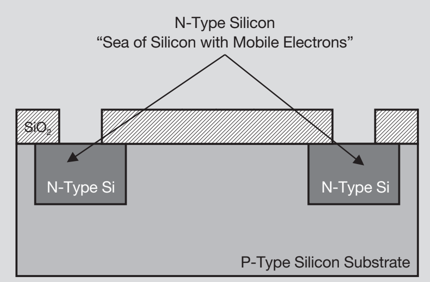


Figure 2.30 N-Channel MOSFET Cross Section (NMOS)

Metal (usually Aluminum) is then deposited across the top of the glass and semiconductor block.

Then some of the metal is etched away with acid to isolate the contacts.

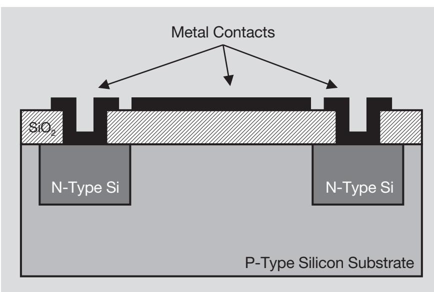


Figure 2.31 N-Channel MOSFET Cross Section (NMOS)

Two of the metal contacts are directly connected to the n-type silicon areas in the semiconductor block.

Referring to Figure 2.32: the contact on the left we have labeled "Source". This is where the electrons in the transistor current will originate from. Often, this is marked as "S" in electrical diagrams. The contact on the right we have labeled "Drain". This is where the electrons in the transistor current will congregate to. Often, this is shown as "D" in electrical diagrams. The electrode "G" is the gate or control electrode. The illustrated MOSFET is source-, and drain-wise symmetric.

There is a third contact labeled "Gate" (symbol is G). The gate is not electrically connected to the silicon substrate but it is isolated by the insulating glass layer. One can think of this contact as the door between the source and the drain. Like the BJT base, the gate is the contact which controls the state of conduction of the MOSFET.

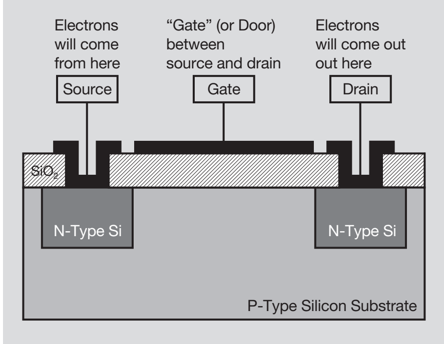


Figure 2.32 N-Channel MOSFET Cross Section (NMOS)

Finally, we show the n-channel MOSFET structure and the circuit symbol used to represent the n-channel MOSFET in electrical schematics.

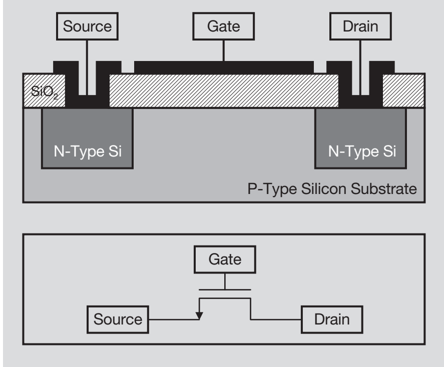


Figure 2.33 N-Channel MOSFET Cross Section (NMOS)

Unlike the BJT, there are several different versions of the MOSFET figure used in schematics that vary considerably, but all of them have two lines representing the gate (where the BJT has only one). The space between the lines indicate the insulating layer found in all MOSFETs between the gate and the silicon substrate.

To understand the operation of the n-channel MOSFET, we first show the negative mobile electrons in the n-type silicon and the positive mobile holes in the p-type silicon.

Next, we point out that the two n-type regions in the p-type substrate again form two back-to-back diodes similar to the BJT.

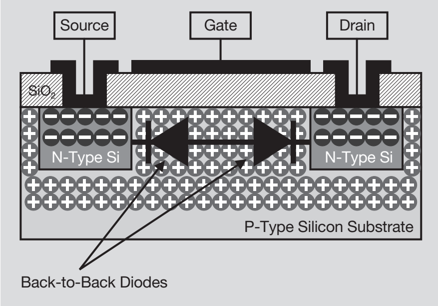


Figure 2.34 N-Channel MOSFET Biasing

Let us see how these back-to-back diodes prevent current from flowing when the MOSFET is off. First, we apply a positive voltage between the drain and source of the MOSFET (drain is positive). The right diode, however, is reverse biased because the n-type silicon is at a higher voltage than the p-type and current does not flow.

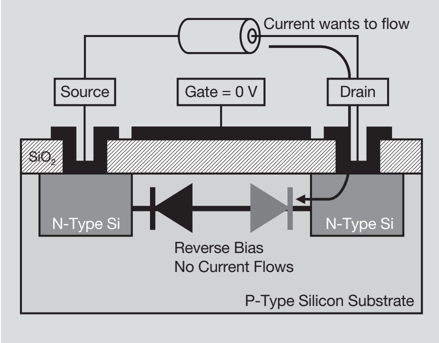


Figure 2.35 N-Channel MOSFET Biasing

Now, let us consider what happens in a real MOSFET. Above 0 degrees Kelvin (absolute zero), there are actually a very small number of negative mobile electrons in the p-type silicon. They may be outnumbered 1,000,000 to 1 by the positive mobile holes, but some small amount of electrons are there.

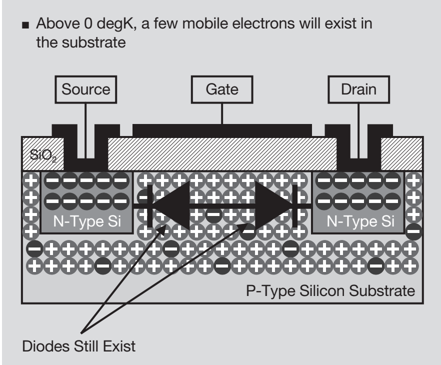


Figure 2.36 N-Channel MOSFET Mobile Carriers

The electrons, however are distributed randomly throughout the p-type substrate. Therefore, the diodes still exist and block the current.

Now we look at a MOSFET when it is turned on. The voltage is still applied between the drain and source. If, in addition, a positive voltage is applied to the gate it charges the top of the insulating glass positively. Since the glass is an excellent insulator, the gate stays positively charged.

The positive charge on the gate serves two purposes. First, it repels the positive mobile holes in the p-type substrate away from the gate. Secondly, it attracts the negative mobile electrons in the p-type substrate toward the gate.

Figure 2.37 N-Channel MOSFET Biasing

After this happens, the silicon directly below the gate is filled with negative mobile electrons and very few positive mobile holes. This cause an n-type conducting channel to form in the MOSFET between the source and the drain directly below the gate.

Electrons now have a path to flow freely from the source to the drain by-passing the back-to-back diodes. As a result, the MOSFET is on and current is allowed to flow from drain to source.

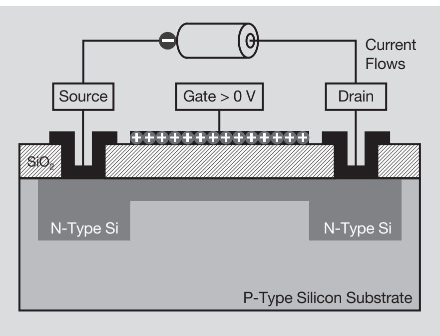


Figure 2.38 N-Channel MOSFET Biasing

However, if the charge on the gate is removed (by grounding the gate), the mobile electrons and holes will diffuse away from their high concentrations to areas of low concentration. When this occurs, the conducting n-channel underneath the gate disappears and the MOSFET will turn off.

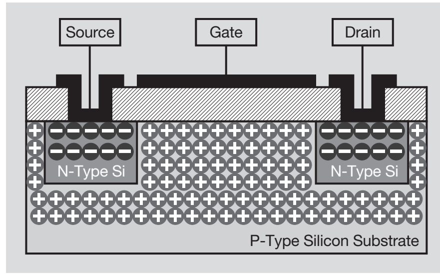


Figure 2.39 N-Channel MOSFET after Turn Off

3. Transistors and Integrated Circuits

In this chapter, we will be continuing our investigation into semiconductor devices from the previous chapter.

We will begin with a more in-depth analysis into the operation of bipolar junction transistors (BJTs) and metal-oxide-semiconductor field effect transistors (MOSFETs). While we will examine each of the two transistors separately, we will cover the same four topics for each:

1) Regions of Operation

2) Control Mechanism (Current or Voltage)

3) Equations and Models

4) Simple Circuits Using the Transistors

Next, we will look briefly into how many transistors can be manufactured in an integrated circuit. From here, we will introduce Moore’s Law, which predicts the rate (and limits) of continued integration in semiconductor components.

3.1 Bipolar Junction Transistors

We begin by looking at bipolar junction transistors (BJTs). First, we will examine the regions (modes) of operation in BJTs. Next, we will see what is meant when a BJT is called a "current controlled" device. We will briefly touch upon the models and equations designers use when implementing a BJT circuit. This will be followed by a detailed examination of a simple circuit using a BJT.

The regions of operation in a BJT refer to the inherent voltage and current waveforms of the transistor at different voltage and current bias points.

In Figure 3.1, the left side shows the basic structure of a NPN BJT. Two diodes are connected back-to-back with the base in between. Additionally, we could examine a PNP BJT, but that is for another day.

The BJT is typically said to be in one of four regions of operation based upon the base-emitter voltage difference (V_{BE}) and the base-collector voltage difference (V_{BC}).

If $V_{BE} < 0$ and $V_{BC} < 0$, then the BJT is said to be in the cut-off region. The transistor is turned off.

If $V_{BE} > 0$ and $V_{BC} < 0$, then the BJT is said to be in the active region. The transistor is turned partially on, but it is still possible to increase the current flow through the transistor.

If $V_{BE} > 0$ and $V_{BC} > 0$, then the BJT is said to be in the saturation region. The transistor is fully turned on.

If $V_{BE} < 0$ and $V_{BC} > 0$, then the BJT is said to be in the inverted region. The transistor is turned on like in the active region, but the collector and emitter have been switched. (Note that based upon non-symmetrical fabrication techniques, some BJTs will not perform as well in the inverted region as they would in the active region.)

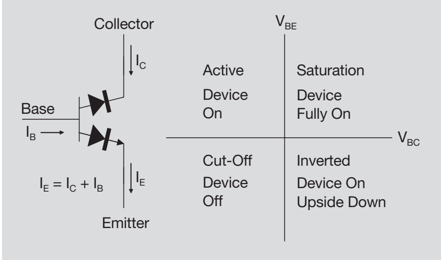


Figure 3.1 Bipolar Regions of Operation

Now, let us look at how the BJT's inputs control the transistor.

First, the NPN BJT needs to sink a base current to turn on. For $I_B = 0$ and for very small base currents, the BJT can be considered to be off. Several curves are shown for different base currents ($I_B = i$ to $5 * i$ in Figure 3.2). Notice that as the NPN BJT sinks additional base current, its collector current increases almost linearly.

The second input is the collector-emitter voltage difference (V_{CE}). For low values of V_{CE}, the NPN BJT is in the saturation region. The current flowing through the BJT (IC) increases almost linearly with voltage across the BJT (V_{CE}). Hence, in the saturation region, the BJT is acting like a resistor.

However, as the voltage across the transistor continues to increase, the BJT loses its ability to source additional collector current. At that point, the increase in the collector current halts and it remains constant for further increases in V_{CE}. This is the active region. In the active region, the BJT is acting like a current source.

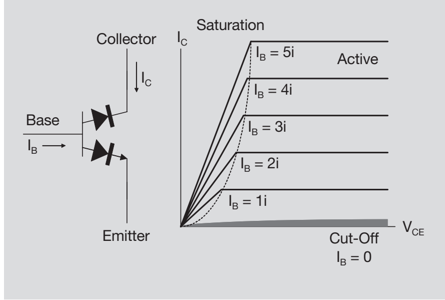


Figure 3.2 Bipolar Regions of Operation

Next, we will look at how the base current controls the operation of BJTs. As was mentioned in the previous chapter, the base can be thought of as "base of operations."

As we saw in Figure 3.2, the collector current (I_C) is dependent upon the current the BJT base sinks (I_B).

For the active region this is true regardless of the collector-emitter voltage difference (V_{CE}). Even if V_{CE} is very high, I_C will be very small for small values of I_B.

BJT bias circuits are designed and implemented to provide the desired amount of current to the base of the BJT.

The circuit bias and expected operating point will be negatively affected if the base current is outside of its expected value.

Therefore, BJTs are called current controlled devices.

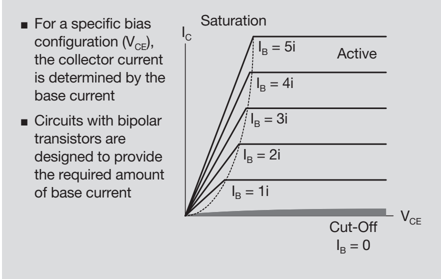


Figure 3.3 Bipolar Transistors are "Current Controlled" Devices

When the BJT is acting as a current source in the active mode, the ideal collector current will be constant regardless of collector-emitter voltage (V_{CE}).

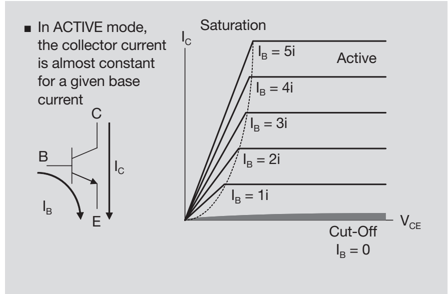


Figure 3.4 Bipolar Transistor Gain (β)

The ratio of the BJT's collector current to base current is known as the BJT's gain. It is denoted as β, but is sometimes spelled out (BETA).

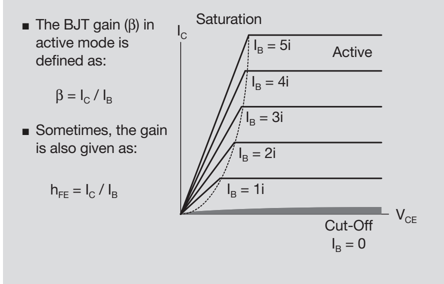


Figure 3.5 Bipolar Transistor Gain (β)

Note, that the value of β appears to be constant across all input values of I_B and V_{CE}.

In practice, there is some variation in β. Figure 3.6 shows how the gain changes for collector currents ranging from 10 μA to 100 mA. At 25 °C, β varies from about 240 - 290 across three orders of magnitude (up to 10 mA). For high current levels, however, non-idealities can cause the gain to drop significantly (approximately 180 for I_C = 100 mA).

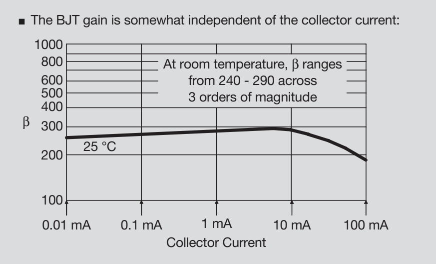


Figure 3.6 Bipolar Transistor Gain (β)

In addition, you will also shortly see that β is a strong function of temperature. At 1 mA, the gain can double over a 150 °C temperature range. Therefore, a bipolar circuit optimized to operate at 125 °C may not function as well at -40 °C.

As long as the BJT remains biased in the active region, the ideal collector current will linearly change with the base current.

Figure 3.7 shows how the collector current of a BJT biased in the active region changes when the base current is changed.

However, the ratio between I_C and I_B will ideally be β under all conditions.

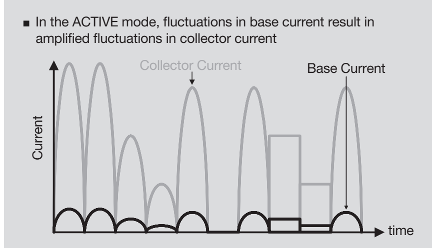


Figure 3.7 Bipolar Transistor Gain (β)

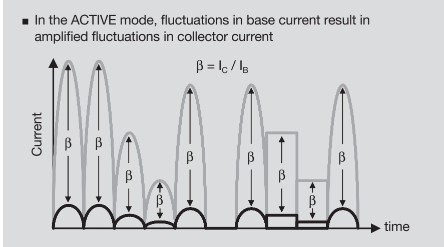


Figure 3.8 Bipolar Transistor Gain (β)

Next we look at a commonly used BJT model and its defining equations.

In 1954, the Ebers-Moll BJT model was proposed. It is still the most commonly used BJT model today.

As we noted before, NPN BJTs can be represented as two diodes. Ebers and Moll modeled the BJT as two back-to-back diodes each in parallel with a current source.

We won't go into any more detail here other than to say that these equations (or some derivative form) are still used today in many circuit simulation software programs.

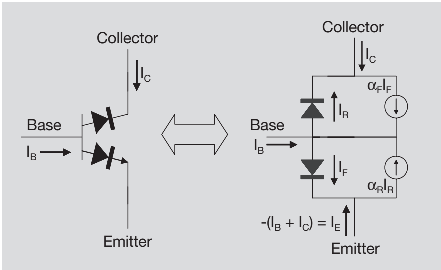


Figure 3.9 Bipolar Junction Transistor Ebers-Moll Model (1954)

Previously, we mentioned that β varies with temperature. We need to look at this more closely to understand one of the problems inherent in circuits using BJTs.

Note, the β of a BJT increases as the temperature increases. This can lead to a problem called "thermal run away".

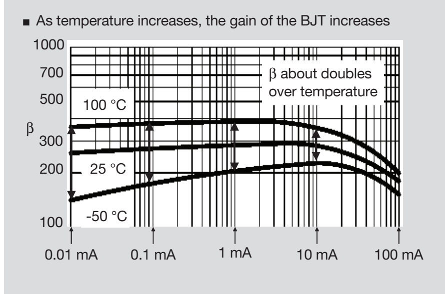


Figure 3.10 Bipolar Junction Transistor Performance vs. Temperature

Thermal run away can occur when the BJT's is affected by excessive power dissipation or ambient temperature.

Care must be taken to ensure that the ambient and operating conditions of the BJT do not lead to a destructive case of thermal run away.

- Since the gain of the transistor increases with temperature, THERMAL RUN AWAY can occur
  1. As the temperature increases, the gain increases
  2. As the gain increases, the collector current increases
  3. As the collector current increases, more power is dissipated
  4. As more power is dissipated, the temperature increases
  5. Go back to step 1
- As thermal run away begins, it can move the BJT away from the expected operating bias point
- Eventually, if the temperature of the device increases above the maximum rated junction temperature ($T_{JUNCTION,MAX}$), the bipolar transistor can be damaged or destroyed

Figure 3.11 Bipolar Junction Transistor Performance vs. Temperature

Alternatively, a small resistor can be placed in series with BJT's emitter. As the current increases through the transistor, this will cause the emitter voltage to increase. Therefore, the base to emitter voltage ($V_{BE} = V_{BASE} - V_{EMITTER}$) will decrease. Thus, there will be a negative feedback effect to reduce the current flowing through the transistor and prevent thermal runaway.

Earlier, we showed this basic graph which shows the BJT's collector current to be flat in the active region.

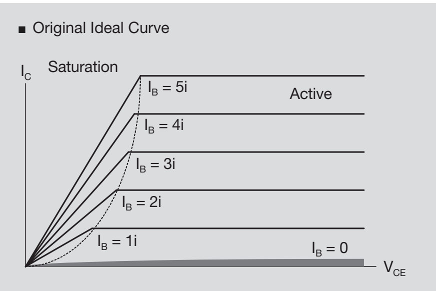


Figure 3.12 Bipolar Junction Transistor Deviations from Ideal Curves

Ideally, the BJT acts as a current source and its collector current is independent of its collector-emitter voltage.

In reality, however, the collector current does increase slightly in the active region as V_{CE} increases.

This is known as the Early effect. It is due to changes in the charge layer widths at the pn junctions within the BJT.

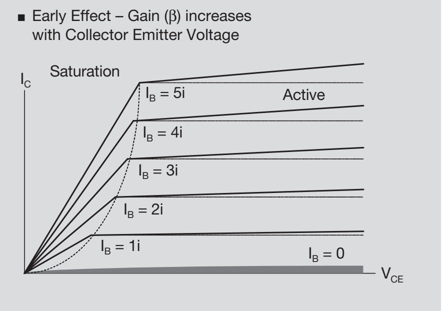


Figure 3.13 Bipolar Junction Transistor Deviations from Ideal Curves

If V_{CE} continues to increase, however, the BJT will enter collector-emitter breakdown.

When the collector-emitter voltage approaches V_{CEO}, the collector current will climb dramatically.

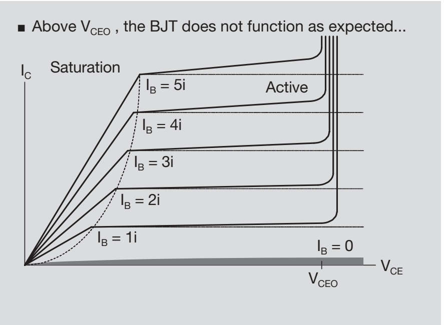


Figure 3.14 Bipolar Junction Transistor Deviations from Ideal Curves

The collector emitter breakdown itself does not damage the BJT. However, if the transistor is heated by the excess power dissipation

beyond the maximum specified junction temperature during the break-down, damage may occur.

Finally, we will examine a simple circuit which utilizes a bipolar junction transistor.

Figure 3.15 shows a simple circuit which uses a NPN bipolar transistor as a low side driver. The 5 V power supply is on the high side of the 4 Ω resistive load, and the transistor is on the low side of the 4 Ω resistive load. Since the 4 Ω resistor is between the collector and a power supply, it is often called the collector resistor, or R_C.

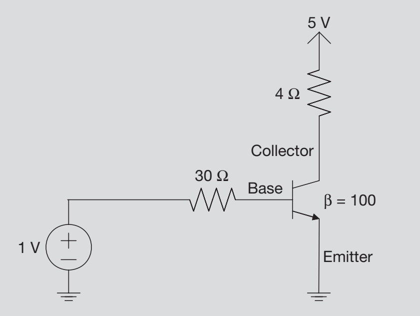


Figure 3.15 Bipolar Transistor Biasing

We begin our analysis of the circuit by remembering that the voltage drop across the NPN base-emitter junction is about 0.7 V. Therefore, about 0.3 V will be dropped across the 30 Ω resistor, resulting in a base current of 10 mA. Since the 30 Ω resistor is between the base and a power supply, it is often called the base resistor, or R_B.

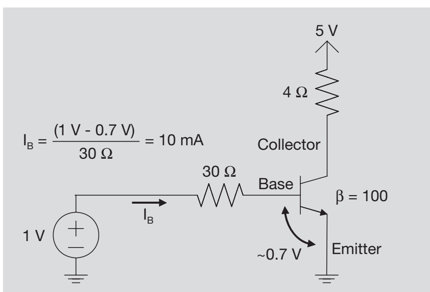


Figure 3.16 Bipolar Transistor Biasing

If we assume that the NPN device operates in the active region and β = 100, the resulting collector current is:

$I_C = \beta\, I_B = (100)(10\text{ mA}) = 1000\text{ mA} = 1\text{ A}$

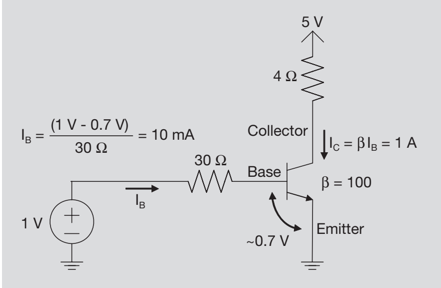


Figure 3.17 Bipolar Transistor Biasing

Therefore, we can calculate the voltage at the collector of the NPN bipolar transistor:

$$
\begin{aligned}
V_C &= 5\text{ V} - R_C\, I_C \\
&= 5\text{ V} - (4\ \Omega)(1\text{ A}) = 5\text{ V} - 4\text{ V} \\
&= 1\text{ V}
\end{aligned}
$$

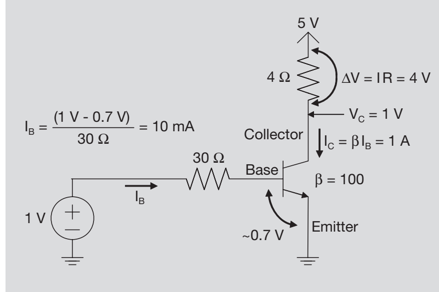


Figure 3.18 Bipolar Transistor Biasing

Next, we examine how this NPN bipolar transistor circuit can operate as an amplifier.

Let us look at what happens when a very small signal (1 mVpp) is added to the base signal. First, we remember what the circuit values were prior to the additional 1 mVpp base signal from Figure 3.18.

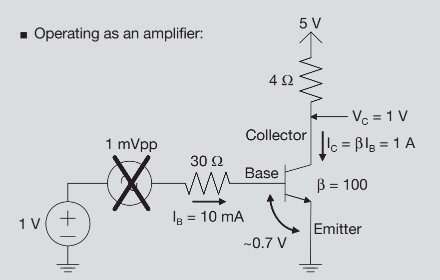


Figure 3.19 Bipolar Transistor Biasing

Now, we insert the 1 mVpp power supply into our original circuit.

The 1 mVpp signal is superimposed upon the 1 V power supply originally connected to the NPN bipolar transistor's base resistor. Since we have already determined that 0.7 V would be dropped across the transistor's base-emitter junction, the additional 1 mVpp signal is dropped across the 30 Ω base resistor.

This results in an additional 33 µApp of base current injected into the NPN bipolar transistor's base.

i_B = 1 mVpp / 30 Ω = 33 µApp

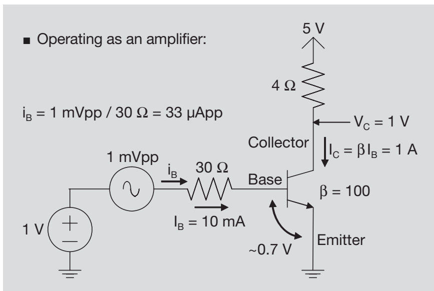


Figure 3.20 Bipolar Transistor Biasing

This results in an additional 3.3 mApp of collector current:

$i_C = \beta\, i_B$ = (100)(33 µApp) = 3.3 mApp

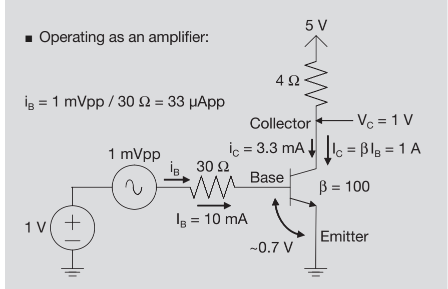


Figure 3.21 Bipolar Transistor Biasing

The small signal collector current then results in a small signal voltage signal on the transistor's collector as the 3.3 mApp of current is dropped across the 4 Ω collector resistor:

$v_C = R_C\, i_C$ = (4 Ω)(3.3 mA) = 13.3 mVpp

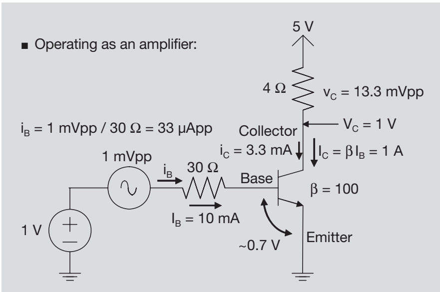


Figure 3.22 Bipolar Transistor Biasing

So, we started with a small 1 mVpp ripple. The output of the circuit is an amplified 13.3 mVpp ripple. We say the circuit gain is:

$A_V = V_{GAIN} = V_{OUT} / V_{IN}$
= 13.3 mVpp / 1 mVpp
= 13.3

Where the gain is often denoted as "A", and the subscript "V" refers to a gain in voltage.

It may not be intuitively obvious, but the gain of this circuit is actually -13.3 (when the input 1 mVpp signal is at its maximum, the output 13.3 mVpp signal is at its minimum).

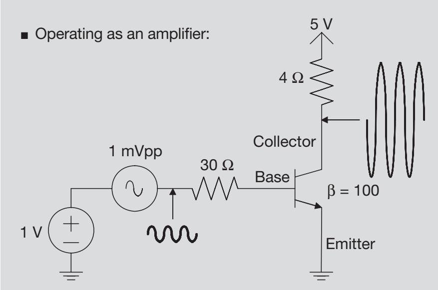


Figure 3.23 Bipolar Transistor Biasing

When the 1 mVpp signal is at +0.5 mV, the base current increases by:

$i_B = v_B / R_B = 0.5\text{ mV} / 30\ \Omega = 16.5\ \mu\text{A}$

The additional 16.5 µA of base current causes the collector current to increase:

$i_C = \beta\ i_B = (100)(16.5\ \mu\text{A}) = 1.65\text{ mA}$

The increased collector current results in a larger voltage being dropped across the 4 Ω collector resistor, and the collector voltage goes down:

$v_C = -R_C\ i_C = (4\ \Omega)(1.65\text{ mA}) = -6.6\text{ mV}$

Therefore, when v_B is at its maximum, v_C is at its minimum. Likewise, when v_B is at its minimum, v_C is at its maximum. Therefore, this circuit is an inverting amplifier, and we denote its gain as:

$A_V = -13.3$

(Non-inverting gain would mean the maximum input voltage results in the maximum output voltage.)

As we noted before, the characteristics of transistors can and will vary from the typical value through manufacturing, temperature, or other factors.

Let us look at the same bipolar circuit, but consider the implications of varying just two of the circuits parameters: the transistor current gain (β) and the transistor base-emitter turn-on voltage (V_{BE}). For this simple example, we will consider the other parameters to be ideal.

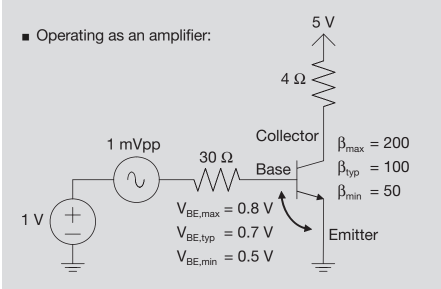


Figure 3.24 Bipolar Transistor Biasing Worst Case Analysis

Figure 3.25 shows a summary of the worst case analysis of the bipolar circuit for variations in the transistor's base-emitter voltage and dc current gain.

From before, for the typical values (β = 100 and V_{BE} = 0.7 V), we have a collector voltage of 1.00 V superimposed with a 13.3 mV sinusoidal signal.

If the dc current gain is 50, we see on the top row of the table that the collector voltage now varies from 1.67 V to 3.67 V, each time with a 6.67 mV superimposed sinusoidal signal. Reducing the dc current gain from 100 to 50 reduces the magnitude of the superimposed sinusoidal signal by 50%. In addition, varying the base-emitter voltage of the transistor causes the dc collector voltage to now vary from 1.67 V to 3.67 V.

If the dc current gain, however, were to remain at 100, and the base-emitter voltage is increased to 0.8 V, the superimposed sinusoidal signal magnitude remains at 13.3 mV. However, it is now superimposed upon a collector dc voltage of 2.33 V.

For the remaining conditions, the table shows that the circuit fails. That is, the circuit will not behave as it was expected to. This failure could be catastrophic and lead to irreparable damage.

Collector Voltage		V_{BE} 0.5 V	0.7 V	0.8 V
	50	1.67 V ±6.67 mV	3.00 V ±6.67 mV	3.67 V ±6.67 mV
β	100	Circuit Fails	1.00 V ±13.3 mV	2.33 V ±13.3 mV
	200	Circuit Fails	Circuit Fails	Circuit Fails

Figure 3.25 Bipolar Transistor Biasing Worst Case Analysis

3.2 Metal-Oxide-Semiconductor Field Effect Transistors

Next, we will look at Metal-Oxide-Semiconductor Field Effect Transistors (MOSFETs). We begin again by looking at their regions of operation.

Figure 3.26 shows a basic n-channel MOSFET (NMOS) device. It has three terminals: Gate, Drain and Source.

In an NMOS, positive current (I_D) flows from the Drain to the Source. The amount of current is controlled by V_{GS}:

$$V_{GS} = V_G - V_S = V_{GATE} - V_{SOURCE}$$

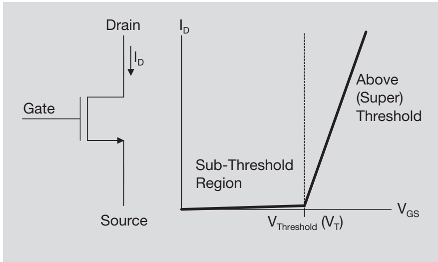


Figure 3.26 MOSFET Two Basic Regions of Operation

When V_{GS} is less than the NMOS threshold voltage ($V_{GS} < V_T$), the NMOS is in the subthreshold region. For most applications, the MOSFET can be considered off.

When V_{GS} is greater than the NMOS threshold voltage ($V_{GS} > V_T$), the NMOS is in the superthreshold region. The NMOS is on and can begin conducting a significant amount of current.

When the NMOS is in the superthreshold region, its operation can be broken into two more regions, linear region and saturation region. To differentiate between the two regions, let us look at how the NMOS's inputs control the transistor.

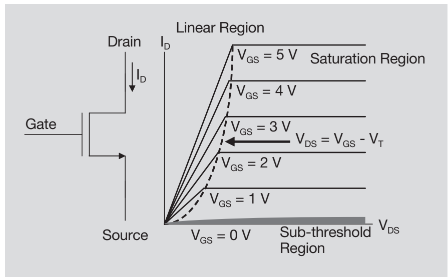


Figure 3.27 MOSFET Super Threshold Regions of Operation

First, we need the NMOS to turn on. Therefore, $V_{GS} > V_T$. Several curves are shown for different levels of V_{GS} (0 V to 5 V in Figure 3.27). Notice, V_{GS} is not the only variable to determine if the transistor is in the linear region or saturation region.

The second input is the drain-source voltage difference (V_{DS}). For low values of V_{DS}, the NMOS is in the linear region. The current flowing through the NMOS (I_D) increases almost linearly with voltage across the NMOS (V_{DS}). Hence, in the linear region, the NMOS is acting like a resistor.

However, as the voltage across the transistor continues to increase, the NMOS loses its ability to source additional drain current. At that point, the increase in I_D halts and it remains constant for further increases in V_{DS}. This is the saturation region. In the saturation region, the NMOS is acting like a current source.

The line that divides the superthreshold region into the linear and saturation regions is:

$V_{DS} = V_{GS} - V_T$

If V_{DS} is less than $V_{GS} - V_T$, the NMOS is in the linear region, and the transistor acts as a resistor.

If V_{DS} is greater than $V_{GS} - V_T$, the NMOS is in the saturation region, and the transistor acts as a current source.

Next, we will look at how the gate voltage controls the operation of a MOSFET. As we saw in the last graph, the drain current (I_D) is dependent upon the gate-source voltage difference (V_{GS}). This is true regardless of the drain-source voltage difference (V_{DS}). Even if V_{DS} is very high, I_D will be very small for small values of V_{GS}.

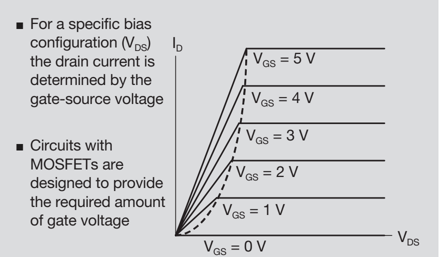


Figure 3.28 MOSFETs are "Voltage Controlled" Devices

MOSFET biasing circuits are designed and implemented to provide the desired amount of gate voltage to the MOSFET. The bias circuit and device can be negatively affected if the gate voltage is outside of its expected value.

Therefore, MOSFETs are called voltage controlled devices.

Recall that bipolar junction transistors were current controlled devices, and we introduced the transistor's gain (β) as the ratio of the collector current to base current:

$\beta = I_C / I_B$

Since MOSFETs are voltage controlled devices, we don't use the same concept of current gain. Rather, we speak of a MOSFET's transconductance (g_m):

$g_m = \Delta I_D / \Delta V_{GS}$

Rearranging the equation, we find that the transconductance is approximately the reciprocal of the transistor's small signal resistance:

$\Delta V_{GS}\, g_m = \Delta I_D$
$\Delta V_{GS} = (1 / g_m)\, \Delta I_D \sim (r)\, \Delta I_D$

$g_m \sim 1 / r$

From the MOSFET's transconductance, you can determine how much the drain current will change based upon a change in the MOSFET's gate-source voltage difference:

$\Delta I_D = (g_m)(\Delta V_{GS})$

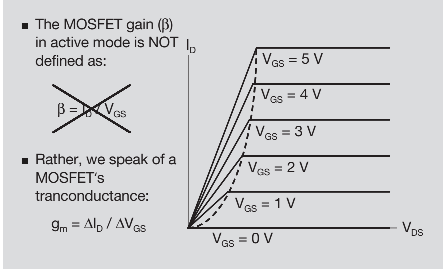


Figure 3.29 MOSFET Transconductance (g_m)

Next, we will briefly look at the equations and models for MOSFETs.

There are many different models that are used for MOSFET circuits. This slide lists four of the most common models in order of increasing complexity and accuracy.

The model most engineering students learn is the square law model. While it no longer accurately represents modern transistors, it still provides a simple introduction to the key device parameters that affect a MOSFET.

The bulk charge theory and more complex models take into account additional device parameters and intricacies that are ignored by the square law model.

- Square Law Model
  - Simple, easy for hand calculations
  - Inaccurate for modern devices
- Bulk Charge Theory
  - Moderately complex for hand calculations
  - Inaccurate for modern devices
- Charge Sheet Model
  - Complex
  - Almost as accurate as the exact charge model
- Exact Charge Model
  - Very complex
  - Very accurate for older and modern devices

Figure 3.30 MOSFET Equations and Models

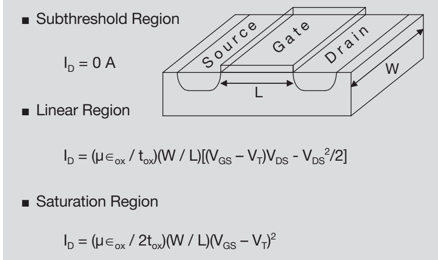


Figure 3.31 MOSFET Square Law Model

The parameters in the equations are:

μ Carrier Mobility in Silicon

ϵ_{ox} Permittivity of Silicon Dioxide

t_{ox} Gate Silicon Dioxide Thickness

W Transistor Channel Width

L Transistor Channel Length

Finally, we examine how MOSFETs can be used to build a simple logic circuit, an inverting gate.

First, in digital logic circuits, MOSFETs always acts as switches.

In an NMOS transistor, the switch is off when $V_{GS} = 0$ V. In an NMOS transistor, the switch is on when $V_{GS} > 1$ V.

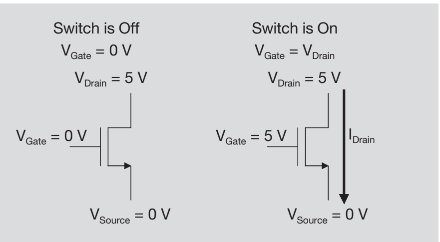


Figure 3.32 n-Channel MOSFET (NMOS) Acting as a Switch

In a PMOS transistor, the switch is off when $V_{SG} = 0$ V. In an PMOS transistor, the switch is on when $V_{SG} > 1$ V.

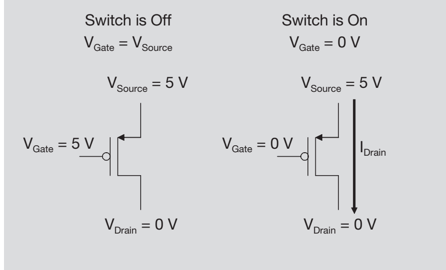


Figure 3.33 p-Channel MOSFET (PMOS) Acting as a Switch

Figure 3.34 is a simple inverter circuit. It uses one PMOS and one NMOS transistor.

It has a single input and a single output.

Since the inverter is a digital logic circuit, the input and output can only take one of two values (HIGH/LOW, YES/NO, 1/0, etc.).

It has a single power connection (+5 V here) and a single ground connection (0 V).

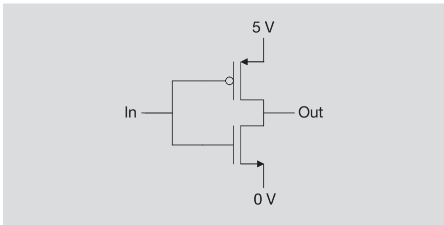


Figure 3.34 Complementary MOSFET “CMOS” Inverter

Let's see what happens when the input is 0 V (input is LOW, or input is "NO").

Recall that this means that there is no positive charge on the NMOS gate to attract electrons into the channel between the source and drain.

Since there is no channel, the NMOS switch does not turn on. It remains off.

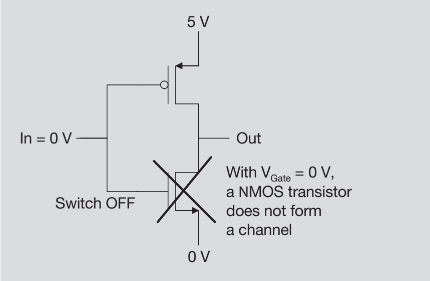


Figure 3.35 Complementary MOSFET "CMOS" Inverter

With a 0 V input, the gate of the PMOS transistor is at 0 V.

Recall that this means that there is a "negative" charge on the PMOS gate. This repels electrons in the PMOS transistor, and a channel is formed between the source and drain.

Since there is a channel, the PMOS switch does turn on.

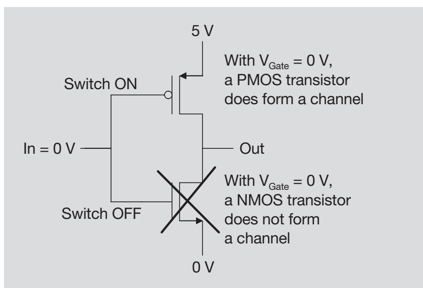


Figure 3.36 Complementary MOSFET "CMOS" Inverter

With the PMOS on, current tries to flow from the 5 V supply. The current flows through the PMOS transistor with very little voltage drop (we can picture the PMOS transistor here as an ideal switch).

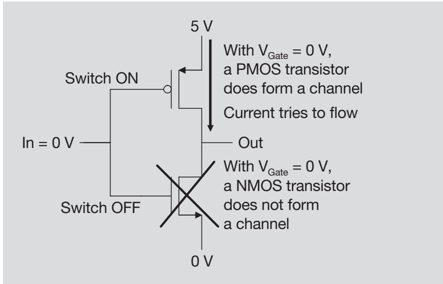


Figure 3.37 Complementary MOSFET "CMOS" Inverter

However, current cannot flow through the switched off NMOS transistor. Therefore, the potential at the output remains 5 V. This value is the inverse of the input value:

Input	Output
0 V	5 V
LOW	HIGH

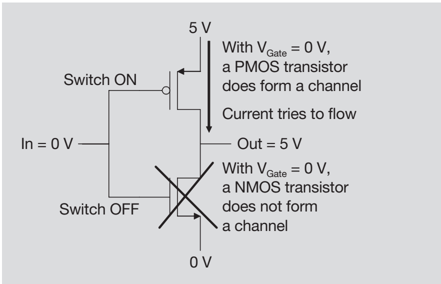


Figure 3.38 Complementary MOSFET "CMOS" Inverter

Now let us consider what happens when the input is raised to 5 V. With a 5 V input, the gate of the NMOS transistor is at 5 V.

Recall that this means that there is a positive charge on the PMOS gate. It does not repel electrons in the channel between the source and drain, and no channel forms.

Since there is no channel, the PMOS switch does not turn on. It remains off.

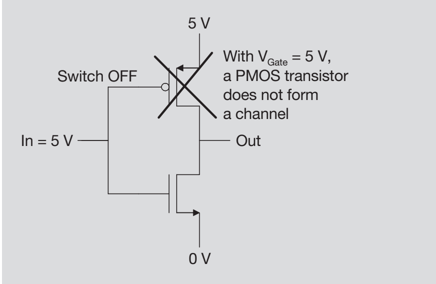


Figure 3.39 Complementary MOSFET "CMOS" Inverter

With a 5 V input, the gate of the NMOS transistor is at 5 V. Recall that this means that there is a positive charge on the NMOS gate. This attracts electrons into the NMOS transistor channel, and a channel is formed between the source and the drain. Since there is a channel, the NMOS switch does turn on.

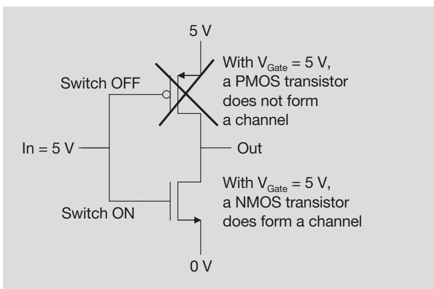


Figure 3.40 Complementary MOSFET "CMOS" Inverter

With the PMOS turned off and the NMOS turned on, current tries to flow from the output to 0 V. If we again picture the NMOS transistor as an ideal switch, the output will go to 0 V.

This value is the inverse of the input value:

Input	Output
5 V	0 V
HIGH	LOW

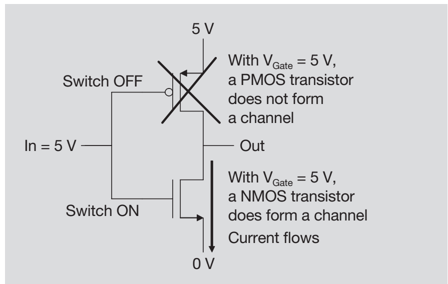
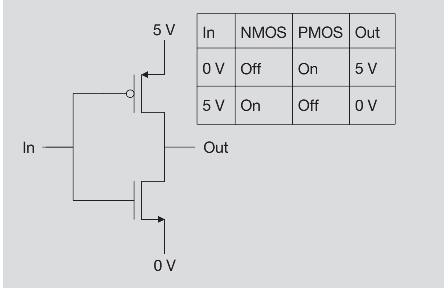


Figure 3.41 Complementary MOSFET "CMOS" Inverter

Therefore, a NMOS and PMOS transistor can be used to implement a simple digital logic gate, an inverter.



In	NMOS	PMOS	Out
0 V	Off	On	5 V
5 V	On	Off	0 V

Figure 3.42 Complementary MOSFET "CMOS" Inverter

- Logic functions are often less susceptible to variations in device characteristics and operating conditions than analog functions
- Both semiconductor component and system variations, however, can impact the CMOS logic performance:
  - – Junction Temperature
  - – System Voltage
  - – Input Voltage Levels
  - – Timing
  - – Transistor Threshold Voltages
  - – Capacitances

Figure 3.43 CMOS Inverter Worst Case Analysis

The performance of advanced logic components (such as microcontrollers, microprocessors, and DSPs) is affected by these and additional parameters.

3.3 Integrated Circuits

Next, we will briefly introduce the topic of integrated circuits.

“Integrated circuit” is the term used for a semiconductor device which integrates more than one active semiconductor component (diode or transistor). Some integrated circuits only contain a few transistors. Others can contain hundreds of millions of transistors.

One very simple integrated circuit is the MOSFET digital inverter we just introduced. Figure 3.44 shows the cross section of an integrated circuit implementing the inverter. You can clearly see the labeled input and output, the power (+5 V), and ground (0 V) connections. The PMOS is on the right, in the n-well. The NMOS is on the left.

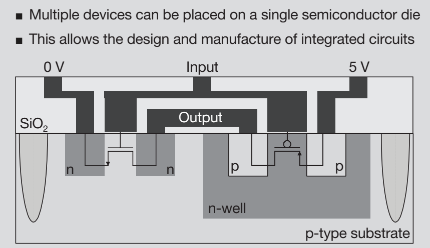


Figure 3.44 Integrated Circuits

As simple as the digital inverter may appear, it still is rather complex.

Figure 3.45 shows us a cross section of the single NMOS transistor. It looks simple, right?

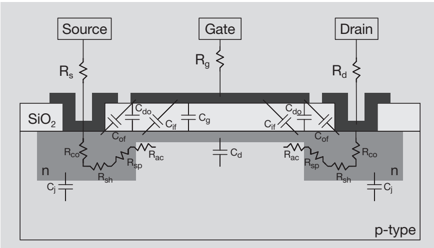


Figure 3.45 Parasitic Resistances and Capacitances

Figure 3.45 shows many of the parasitic resistive and capacitive elements in the NMOS transistor. It does not include any of the parasitic elements existing between the NMOS and PMOS transistors in the digital inverter, or the parasitics in the interconnect between the digital inverter silicon die and the outside world.

Integrated circuits are incredibly complex devices when you consider you might find millions (or billions!) of these transistors in a single semiconductor component.

3.4 Moore’s Law

We will close our introduction to transistors and integrated circuits with a word about Moore’s Law.

In 1965, Gordon Moore wrote an article for “Electronics” magazine. The not so scientific name of the article was:

“Cramming more components onto integrated circuits”

Basically, Moore illustrated that for several years, the number of transistors and diodes in integrated circuits had been doubling every year. He predicted this trend would continue (if not increase) in the short term.

- Gordon E. Moore
  “Cramming more components onto integrated circuits”, Electronics, Volume 38, Number 8, April 19, 1965.
- “The complexity for minimum component costs has increased at a rate of roughly a factor of two per year... Certainly over the short term this rate can be expected to continue, if not increase.”

Figure 3.46 Moore’s Law

Figure 3.47 is based on one of the graphs in Moore’s paper.

It is interesting that the continued integration of functions into increasingly complex integrated circuits was predicted back in 1965 when an integrated circuit with sixty-four transistors was state-of-the-art!

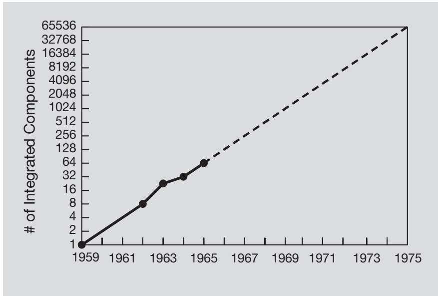


Figure 3.47 Moore's Law

One of the lesser quoted parts of Moore's landmark paper, however, discusses the trend to over-integrate functionality.

- "Clearly, we will be able to build such component-crammed equipment. Next, we ask under what circumstances we should do it. The total cost of making a particular system function must be minimized...."
- "It may prove to be more economical to build large systems out of smaller functions, which are separately packaged and interconnected."

Figure 3.48 Moore's Law Day of Reckoning

Our final figure in this chapter is also from one of the graphs in Moore's paper.

Basically, Figure 3.49 predicts that the ability to integrate transistors and functionality in an integrated circuit will outpace the costeffectiveness of integrating transistors and functionality in the integrated circuit.

This is something all electrical and computer engineers should keep in mind.

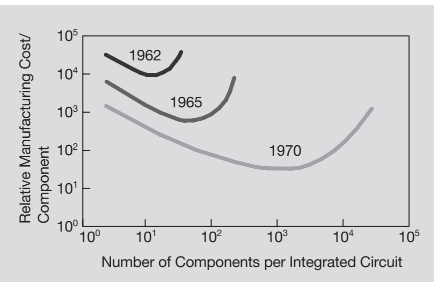


Figure 3.49 Moore's Law Day of Reckoning

4. Introduction to Power Dissipation and Thermal Resistance

In this chapter, we will be examining the factors that influence a semiconductor component's temperature: power dissipation and thermal resistance.

We will begin by examining the concept of electrical power. Afterwards, we will define the term semiconductor junction temperature. We will see why the junction temperature is an important variable in any electronic system using semiconductor components.

Next, we will define thermal resistance and see how different packages vary in cooling effectiveness of a semiconductor component. We will then explicitly draw several analogs between electrical and thermal parameters and electrical and thermal circuits.

After a short introduction to semiconductor thermal specifications and brief description of heatsinks, we will close the chapter with several example thermal calculations.

4.1 What is Power?

In its most general sense, work is defined as the application of power for an amount of time.

So, if we apply power for an amount of time, we can determine the amount of work we are performing.

The units of work are Joules, and the units of power are Watts.

Thus, 1 Joule = 1 Watt * sec

Work is the result of a power applied for a given amount of time

Work = Power * Time

Figure 4.1 What is Power?

In the electrical realm, power is the product of voltage and current:

$P = V * I$

- Electrically, power is a product of a voltage and a current:

Power = Voltage * Current

$P = V * I$

- For example, a battery that can deliver 10 A at 12 V can supply 120 W of power:

$P = 12\ V * 10\ A = 120\ W$

Figure 4.2 What is Power?

In the example, the battery is a source of power. It can be applied for a given amount of time to perform work (for example, to turn on a light).

This equation is almost always used to determine the amount of power that a battery or a power source can deliver. So, if a power supply for an electronics module can deliver 0.8 A at 5 V, it can deliver:

$$P_{DELIVER} = V * I = (5\ V)(0.8\ A) = 4\ V * A = 4\ W$$

If a battery is a source of power, a lamp can be viewed as a consumer of power.

- If a battery can provide 120 W of power, the battery load must consume 120 W of power

120 W Supplied

120 W Consumed

- Some of the power put into the battery load is absorbed and dissipated as heat
- From Ohm's Law (V = IR), the power dissipated as heat in a load is given by:

$P = V * I = (IR) * I = I^2 R$

Figure 4.3 What is Power?

In this example, the battery is capable of supplying 120 W (or 10 A of current at 12 V).

When this battery is connected to the 120 W lamp, the lamp will sink the 10 A of current supplied by the battery, with the entire 12 V dropped across the lamp. Remember that P = V * I. Therefore, the lamp would dissipate (consume) all 120 W delivered by the battery (assuming there was no loss in the wiring from the battery to the lamp).

Instead of always looking at Power as the product of Voltage and Current, it is sometimes easier to use an equivalent equation. By Ohm's Law, V = I * R. Substituting this back into our original equation, we have:

$P = V * I = (I * R) * I = I * I * R = I^2 R$, or

$P = V * I = V * (V / R) = V^2 / R$

These equations are used to calculate the power supplied to, or dissipated in a load. If the resistance of the load (R_{LOAD}) is known, and the current through the load (I_{LOAD}) can be calculated or measured, it is simple to determine the power dissipated in the load:

$$P_{LOAD} = (I_{LOAD})^2 R_{LOAD}$$

So, if the lamp has a load resistance of 1.2 Ω, and the load current is 10 A:

$$P_{LOAD} = (10\ A)^2(1.2\ \Omega) = 120\ A^2{*}\Omega = 120\ W$$

There are only two *electrical* equations you will need to rely upon for the remainder of this chapter as shown in Figure 4.4:

- The important things you must remember here:

$$P = V\,I$$

$$P = I^2 R$$

Figure 4.4 Electrical Power

4.2 What is Junction Temperature?

The term “junction temperature” at first might be a little confusing. It is more appropriate to use the term “semiconductor temperature” instead. We are speaking here of the temperature of the semiconductor die itself. When semiconductor devices were first fabricated, they were very simple. The junction referred to the interface between regions of the semiconductor die, and people started using the term junction temperature. Now, semiconductor devices can have thousands and even millions of different junctions, but we still use the phrase “junction temperature” to refer to the temperature of the semiconductor die.

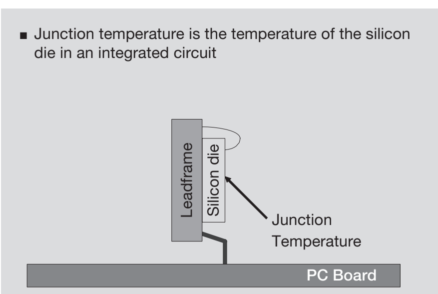


Figure 4.5 Junction Temperature

Just remember that the term “junction temperature” refers specifically to the temperature of the semiconductor device and is often referred to as $T_{JUNCTION}$ or T_J.

The semiconductor die is usually attached to a relatively large (or relatively massive) lead frame, also known as a “case”. The semiconductor die and lead frame are then enclosed inside a package which is mounted on a PC board.

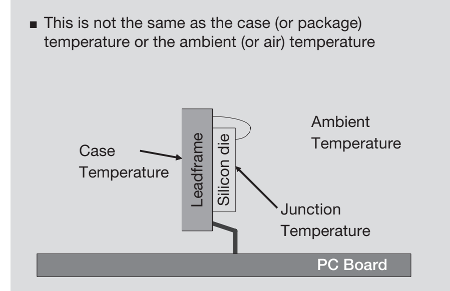


Figure 4.6 Ambient & Case Temperature

In case of power semiconductors some part of the lead frame, however, is usually exposed by the package to help electrically and thermally connect the semiconductor die to the outside world. The temperature of the lead frame (or case) is usually referred to as the case temperature (also known as T_{CASE} or T_C).

The temperature of the air surrounding the package is also important in thermal calculations. It is usually referred to as the ambient air temperature, $T_{AMBIENT}$ or T_A.

In the next several figures, we are going to show the progression of how heat moves within a semiconductor device when it dissipates power.

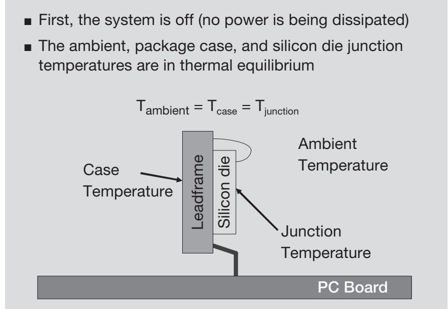


Figure 4.7 Junction, Case, and Ambient Temperatures

Initially, we assume the system is off. The semiconductor device is not connected to a

power source and is not dissipating any power.

Since no power is being dissipated by the device, all three temperatures (ambient, case, and junction) are equal and the system is in thermal equilibrium.

At some point, the system is turned on and the semiconductor die is electrically connected to a power source. From our earlier discussions, we know that power is now being dissipated within the semiconductor device as heat. This causes the semiconductor die temperature ($T_{JUNCTION}$) to increase.

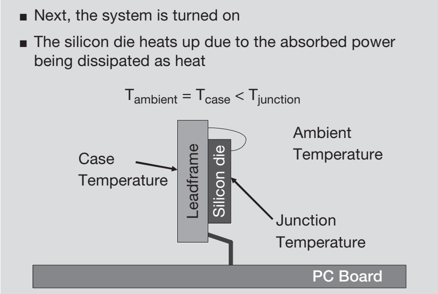


Figure 4.8 Junction, Case, and Ambient Temperatures

1) The semiconductor die generates heat when it dissipates (loses) power.

Since this has "just happened" we can assume that for some length of time, the case or package temperature (T_{CASE}) is still at or near to the ambient air temperature ($T_{AMBIENT}$).

Some time later the hot semiconductor die begins to warm up the lead frame. Since the lead frame is only absorbing heat, its temperature will always be less than the semiconductor device which is generating heat. Therefore, $T_{CASE} < T_{JUNCTION}$ when power is applied.

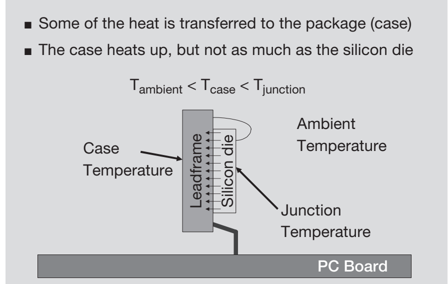


Figure 4.9 Junction, Case, and Ambient Temperatures

2) The lead frame absorbs some of the heat from the semiconductor die and begins to warm up.

After the lead frame begins to absorb heat from the semiconductor die, the ambient air begins to absorb heat from the lead frame (case).

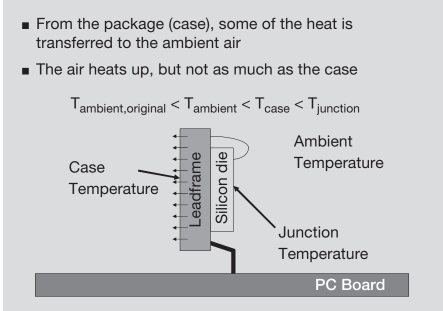


Figure 4.10 Junction, Case, and Ambient Temperatures

3) The ambient air absorbs some of the heat from the lead frame and begins to warm up.

This rise in the ambient air temperature may not be great, but it nevertheless will occur.

So, when a semiconductor device is turned on:

$T_{A,ORIGINAL} < T_A < T_{CASE} < T_{JUNCTION}$

This relationship is important and easy to remember. Since the semiconductor die is the component generating the heat, it is the

warmest. Close to the semiconductor die, the lead frame is also fairly warm. Farther away from the semiconductor die, the ambient air is warmer than when at the start, but cooler than the lead frame.

We now understand the relative temperature levels of a semiconductor device, but why is this all important?

Well, semiconductor suppliers design and manufacture their components to be used within an expected operating temperature range. When the semiconductor component is within this temperature range, they will perform within the specified operating limits.

Outside of this temperature, however, the supplier does not guarantee the semiconductor component will work properly, or even at all.

In Figure 4.11, the specified operating junction temperature (T_J) range is -40 °C to +150 °C. Therefore, for a desired performance, the system designer must ensure the semiconductor component is never subjected to temperatures below -40 °C (also -40 F) or above 150 °C (302 F).

- Semiconductor devices are specified by their manufacturers at a maximum temperature range:

Maximum Ratings

Parameter	Symbol	Values	Unit
Operating temperature range	T_j	-40 ... +150	°C

- Above this temperature (150 °C in the example), the device may not work as well, or it may stop working completely
- Therefore, it is necessary to keep the junction temperature below the maximum rated operating temperature

Figure 4.11 Why is Junction Temperature Important?

Usually, it is not difficult to keep the junction temperature above -40 °C, since most modules themselves are not expected to perform at temperatures below -40 °C.

The upper temperature limit, however, does need to be carefully examined. If the ambient environment of a semiconductor component is already hot (for example, 125 °C), the junction temperature of the semiconductor die is only allowed to increase by 25 °C when power is dissipated.

System designers must ensure that heat can be drawn quickly away from the semiconductor die to the lead frame and ambient environment to minimize the junction temperature increase. This is a function of how well the system conducts heat.

4.3 What is Thermal Resistance?

Thermal resistance is a measure of the ability of an object to block heat-flow. Sometimes, it is easier to think in terms of thermal conductance, the inverse of thermal resistance.

Objects that have high thermal conductance are very good conductors of heat (aluminum is such a material). It has a high thermal conductance and a low thermal resistance.

Objects that have a low thermal conductance are very poor conductors of heat (plastic or quartz, for example) will also have high thermal resistance.

The total thermal resistance of the entire semiconductor system determines how much the junction temperature of a semiconductor will increase by once it is turned on and begins to dissipate power.

Let us begin by looking at an example where the thermal resistance of the semiconductor system is very low. This means that the semiconductor system has high thermal conductance.

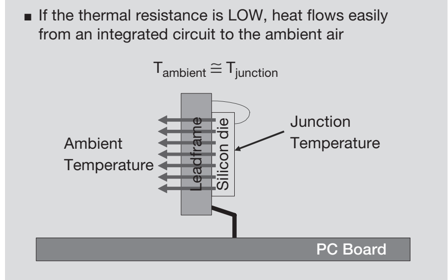


Figure 4.12 Why is Thermal Resistance Important?

Once the semiconductor die begins to heat up, the heat is quickly absorbed by the lead frame. As we said before, the lead frame is relatively large and massive compared to the semiconductor die. The lead frame is said to have a large thermal mass compared to the semiconductor die.

Once the lead frame begins to warm up, the heat flows to the ambient environment. The ambient environment has even a larger thermal mass (i.e. storage capability of heat energy) than the lead frame.

Thermal mass is any mass of matter that absorbs and retains (stores) heat (energy).

The heat generated within the small semiconductor die is shared by the large lead frame and the larger ambient environment resulting in a small increase in $T_{JUNCTION}$.

Imagine if a small piece of metal (0.1” x 0.1”) was subjected to a small amount of heat. The small piece of metal might increase its temperature by 100 °C. However, if the small piece of metal were soldered to a larger piece of metal (for instance a silver dollar), the larger mass of the silver dollar would absorb the heat energy from the 0.1” x 0.1” piece of metal. The net result is that the “system” of the coin and metal fragment would have a much smaller combined increase in temperature (perhaps 10 °C). Finally, imagine that the silver dollar is soldered to an even larger piece of metal (an anvil). The net result of the metal fragment / coin / anvil system might only be a 1 °C temperature rise.

Now, let us begin by looking at an example where the thermal resistance of the semiconductor system is high. This means that the semiconductor system has low thermal conductance.

Once the semiconductor die begins to heat up, the heat is not absorbed well by the lead frame and the heat conductance between lead frame and ambient is even worse. Therefore, the heat generated within the small semiconductor die is NOT shared by the large lead frame and the larger ambient environment and there is a large increase in $T_{JUNCTION}$.

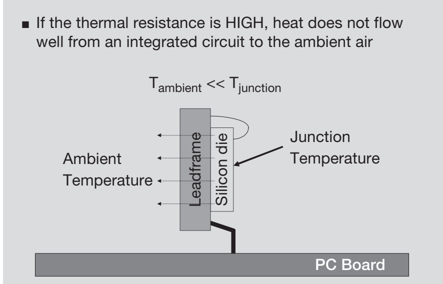


Figure 4.13 Why is Thermal Resistance Important?

Imagine again if a small piece of metal (0.1” x 0.1”) was subjected to a small amount of heat. The small piece of metal might increase in temperature by 100 °C. However, if the small piece of metal were set upon a piece of plastic, very little of the heat from the metal fragment would be absorbed by the plastic. The net result might be that temperature of the small metal piece increases by 98 °C.

If a semiconductor system has a “good” or acceptable thermal resistance to the ambient it will maintain the semiconductor die temperature ($T_{JUNCTION}$) within the specified operating temperature range. Recall that if the semiconductor temperature is outside its operating range, it may not operate as specified.

The longevity of the semiconductor component is also effected by the junction temperature. By keeping the junction temperature below the specified maximum, the long term failure rate of the semiconductor die is minimized and the possibility of damaging the plastic encapsulation package is also eliminated.

In summary, a “good” thermal resistance will:

- Lower the integrated circuit’s junction temperature
- Keep the integrated circuit functioning at a specified (guaranteed) operating temperature
- Minimize the semiconductor long term failure rate
- Minimize problems associated with the glassification of plastic epoxy packages

Figure 4.14 Why is Thermal Resistance Important?

4.4 Electrical vs. Thermal Parameters

At the beginning of this chapter, we have assumed that the reader has a knowledge of simple resistive electrical circuits. Let us briefly review the fundamentals that are needed to understand our simple thermal calculations based on analog DC electrical circuits.

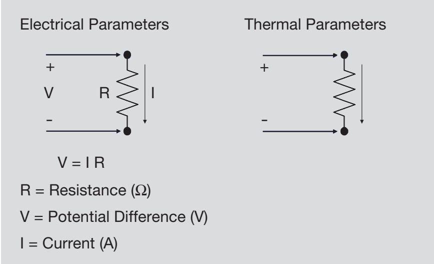


Figure 4.15 Electrical & Thermal Parameters

Ohm's Law states that the voltage drop across an conductor is equal to the product of the current intensity and the electrical resistance of the conductor:

V = I * R

Shifting the frame of reference from the electrical world to the thermal world is all that is needed to develop the "thermal Ohm's Law".

In the thermal world, we also use he thermal resistance (R_{th} or sometimes R_{θ}) i.e. the reciprocal of thermal conductivity.

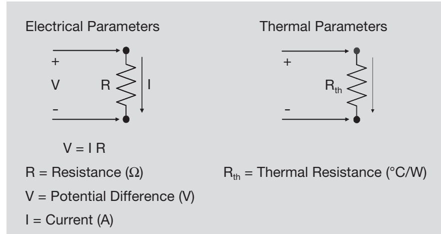


Figure 4.16 Electrical & Thermal Parameters

The thermal resistance is analogous to electrical resistance:

HIGH thermal resistance means the object does not conduct heat well

LOW thermal resistance means the object does conduct heat well

In the electrical world, voltage is viewed as a potential difference. We talk about the "voltage drop" across a resistive load.

In the same manner, in the thermal world, we talk about a temperature difference (or temperature drop), ΔT, across an object.

Think about a house window on a cold winter day. It may be 25 °C (77 F), inside but -10 °C (14 F) outside. There is a large temperature difference (ΔT) across the window.

Temperature difference is analogous to voltage difference.

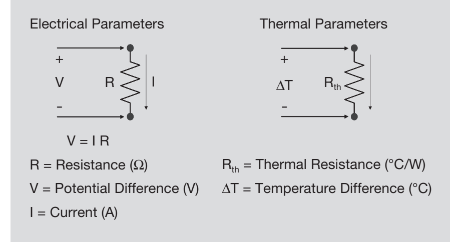


Figure 4.17 Electrical & Thermal Parameters

Finally, in the electrical world, we have current. Current flows through a resistive object from a high voltage to a low voltage. In the thermal world, however, we don't talk about current. Rather, we talk about **power flow**.

Heat flows from a place of high temperature to a place of low temperature. Imagine a pot of water is placed on a hot stove burner. Heat flows from the burner to the pot. Power flows from the burner to the pot. Now the pot can transfer power (heat) to the water and make it boil.

Power flow in thermal calculations is analogous to current flow.

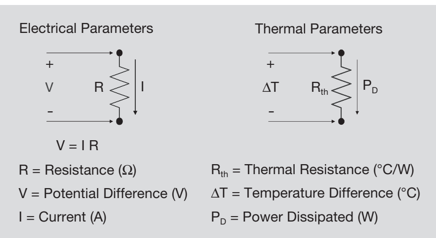


Figure 4.18 Electrical & Thermal Parameters

While the first two analogies were readily apparent, the third (current - power) might be a little more difficult to grasp. Just remember that in an electrical circuit:

> Current flows through a resistive object from a high voltage to a low voltage:
>
> $I = V / R$

And in a thermal circuit:

> Power flows through a resistive object from a high temperature to a low temperature:
>
> $P_D = \Delta T / R_{th}$

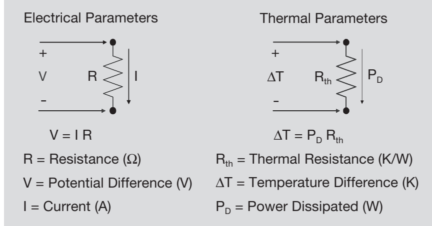


Figure 4.19 Electrical & Thermal Parameters

So, before we perform our first thermal calculation, let us briefly look at the topic of electrical and thermal resistances.

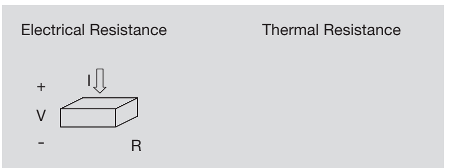


Figure 4.20 Electrical Resistance vs. Thermal Resistance

Let's recall that current flows through a conductor from high potential to low potential (battery + to conductor to battery -).

The electrical resistance of an object is a function of the dimensions of the object and a material parameter, the electrical conductivity.

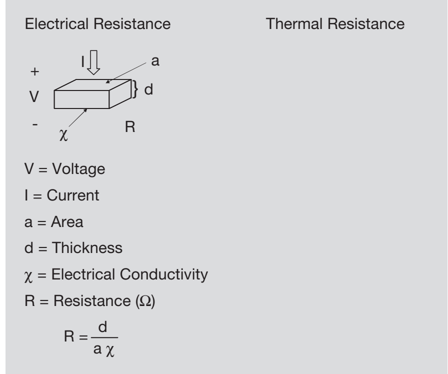


Figure 4.21 Electrical Resistance vs. Thermal Resistance

Electrical conductivity is a parameter of a given material. Regardless of the size or shape of a conductor, its electrical conductivity does not change and high electrical conductivity materials are good conductors of electric current.

The dimensions that determine the resistance of an object are its thickness (or depth) and cross section area. Let's look at thickness first.

The current caused by a constant voltage applied across the resistive object of a given area is directly proportional to its thickness.

The current caused by a constant voltage applied across the resistive object of a given thickness is inversely proportional to its area.

Remember that power flows through a thermally resistive object from a high temperature to a low temperature.

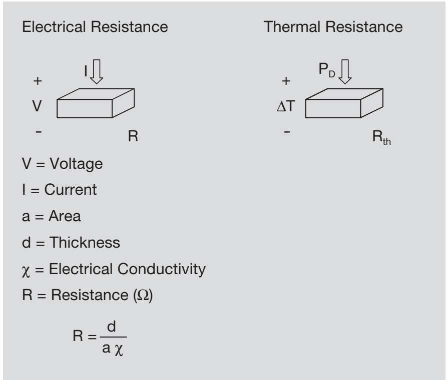


Figure 4.22 Electrical Resistance vs. Thermal Resistance

In our electrical - thermal resistance analogy:

ΔVoltage	- ΔTemperature
Current	- Power
Area	- Area
Thickness	- Thickness
Electrical Conductivity	- Thermal Conductivity

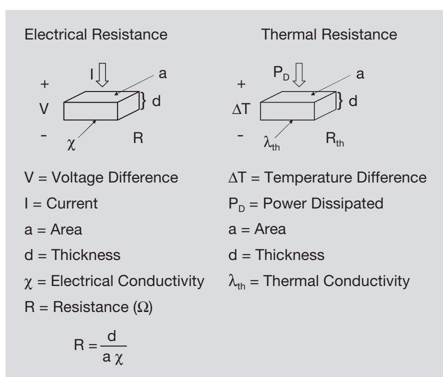


Figure 4.23 Electrical Resistance vs. Thermal Resistance

Like electrical conductivity, the thermal conductivity is a material parameter. Regardless of the size or shape of an object, it's thermal conductivity does not change. A material with a high thermal conductivity is a good conductor of heat and, conversely, a material with a low thermal conductivity is a poor conductor of heat. Note that most materials that are good conductors of heat are also good conductors of electricity.

The thermal resistance of the object is directly proportional to its thickness. This means if the thickness of a thermally resistive object is increased and the same temperature difference is applied to it, the power flowing through the object will decrease. Likewise, if the thickness of a resistive object is decreased and the same temperature difference is applied to it, the power flowing through the object will increase.

Like electrical resistance, the thermal resistance is inversely proportional to the cross section area of the object. This can be illustrated by comparing the heat (power) loss through two identical windows;. twice as much power will be lost through two windows as through one window.

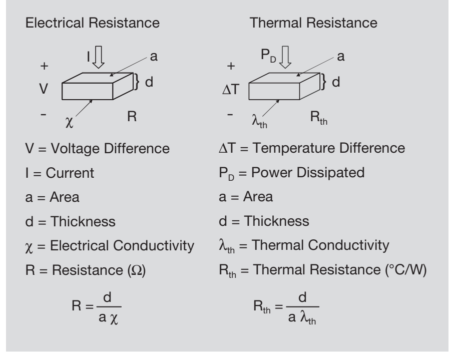


Figure 4.24 Electrical Resistance vs. Thermal Resistance

Let us look at our first thermal calculation. It is a little abstract, but we need to start somewhere. We will solve the same problem in both electrical and thermal realms and will see that the problems are remarkably similar.

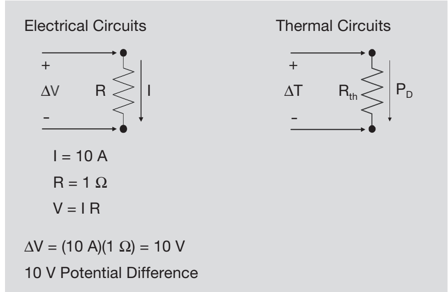


Figure 4.25 Electrical Circuits vs. Thermal Circuits

In the electrical problem, we have a resistive object with an electrical resistance of 1 Ω and 10 A of current is flowing through the resistor.

Now, it is simple to determine the voltage (potential) difference across the resistive object by applying Ohm's Law:

V = I * R
= (10 A) * (1 Ω)
= 10 A*Ω
= 10 V potential difference

In the thermal problem, we begin with the analogous starting information. The thermally resistive object has a thermal resistance of 1 °C/W.

R -> R_{th}

The power flowing through the thermally resistive object is 10 W.

I -> P_D

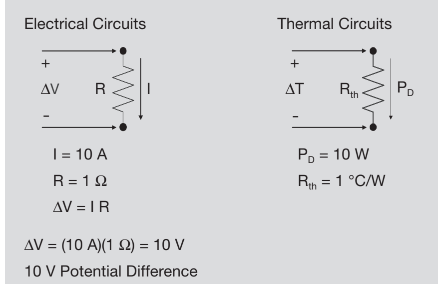


Figure 4.26 Electrical Circuits vs. Thermal Circuits

Using the relationship we established before, we can now determine the temperature difference across the thermally resistive object:

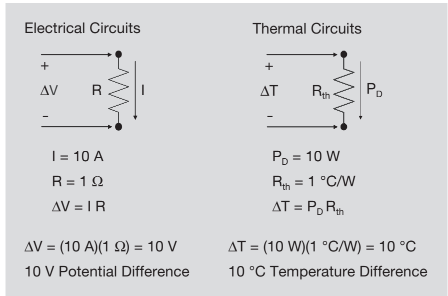


Figure 4.27 Electrical Circuits vs. Thermal Circuits

$\Delta T = P_D * R_{th}$
= (10 W) * (1 °C/W)
= 10 W * 1 °C/W
= 10 °C

Therefore, there is a 10 °C temperature difference across the thermally resistance object, or V -> ΔT.

We might have foreseen this answer. First, the starting values were identical for the analogous parameters. Second, the units of thermal resistance might have given this away. In Figure 4.27, the units of thermal resistance are given as "degrees Centigrade per Watt" (or °C/W).

This means that the thermal resistance will cause a 1 °C temperature difference for every 1 W of power which flows through the thermally resistive object. Therefore, the 10 W of power causes 10 °C temperature difference.

The same analysis can be performed in the electrical realm if the units of resistance are not expressed as Ohms, but rather as "Volts per Amps" (V/A).

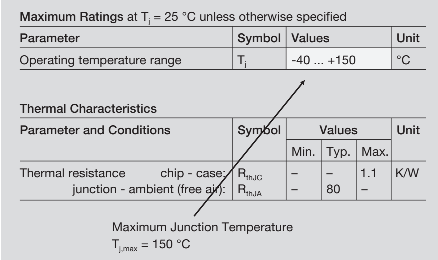
I = 10 A
R = 1 Ω = 1 V/A
ΔV = I * R
= (10 A) * (1 V/A)
= 10 A * V/A
= 10 V

This example may seem trivial and abstract, but the intent is to show how the technique of solving electrical problems can also be applied to solving problems in the thermal domain.

4.5 Thermal Specifications

Consider the specific parameters of semiconductor devices (found in datasheets) which are important in thermal calculations.

First, the maximum junction temperature has already been mentioned. In Figure 4.28 the junction temperature of the semiconductor component must be 150 °C or less to maintain specified operation.



Maximum Ratings at T_j = 25 °C unless otherwise specified

Parameter	Symbol	Values	Unit
Operating temperature range	T_j	-40 ... +150	°C

Thermal Characteristics

Parameter and Conditions	Symbol	Values			Unit
		Min.	Typ.	Max.	
Thermal resistance chip - case:	R_{thJC}	–	–	1.1	K/W
junction - ambient (free air):	R_{thJA}	–	80	–	

Figure 4.28 Thermal Specifications Datasheet Parameters

After the maximum operating temperature is determined, the next important thermal specification is the thermal resistance.

Thermal resistance can be specified in a number of different ways. While this initially may be frustrating, it will allow an experienced designers to optimize their designs for the expected load and power dissipation conditions.

Maximum Ratings at T_j = 25 °C unless otherwise specified

Parameter	Symbol	Values	Unit
Operating temperature range	T_j	-40 ... +150	°C

Thermal Characteristics

Parameter and Conditions	Symbol	Values			Unit
		Min.	Typ.	Max.	
Thermal resistance chip - case:	R_{thJC}	–	–	1.1	K/W
junction - ambient (free air):	R_{thJA}	–	80	–	

Thermal Resistance Junction to Ambient
R_{thJA} = 80 K/W = 80 °C/W

Figure 4.29 Thermal Specifications Datasheet Parameters

K stands for Kelvin (or absolute) degree.
K and °C are interchangeable when applied to *temperature differences*.

In Figure 4.29, the highlighted thermal resistance is R_{thJA} (junction-to-ambient) in free air. The R_{thJA} value represents the entire thermal resistance from the semiconductor die, through the lead frame and package, to the open air. It is the sum of the partial thermal resistances:

$$R_{thjunction\text{-}to\text{-}ambient}$$
$$= R_{thjunction\text{-}to\text{-}case} + R_{thcase\text{-}to\text{-}ambient}$$
$$= R_{thJC} + R_{thCA} = R_{thJA}$$

Next, notice the comment "free air." This is the worst case thermal resistance of the semiconductor component. This number may be significantly improved by using a *heatsink* (we will address this possibility shortly).

The value of the junction-to-ambient free air thermal resistance is 80 K/W. This means that for every 1 W of power that is dissipated by the semiconductor component, the junction temperature will rise 80 K over the ambient.

In Figure 4.30, the highlighted thermal resistance is R_{thJC} (junction-to-case). The value listed is 1.1 K/W. This means that for every 1 W of power dissipated in the semiconductor component, the temperature of the junction will be 1.1 K higher than the case. Note, R_{thJC} is a term in our previous equations:

$$R_{thJA} = R_{thJC} + R_{thCA}$$

So, if we know R_{thJC} and R_{thJA} (worst case), the only thing we have left is R_{thCA} (worst case).

Maximum Ratings at T_j = 25 °C unless otherwise specified

Parameter	Symbol	Values	Unit
Operating temperature range	T_j	-40 ... +150	°C

Termal Characteristics

Parameter and Conditions		Symbol	Values			Unit
			Min.	Typ.	Max.	
Termal resistance	chip - case:	R_{thJC}	–	–	1.1	K/W
	junction - ambient (free air):	R_{thJA}	–	80	–	

Thermal Resistance Junction to Case
R_{thJC} = 1.1 K/W = 1.1 °C/W

Figure 4.30 Thermal Specifications Datasheet Parameters

But, if R_{thJA} is 80 K/W, why is R_{thJC} so much smaller? To answer this question, let us first look at how a semiconductor component is manufactured.

The semiconductor die is first attached to the lead frame with a "die attach" material. The die attach and the lead frame are both manufactured of very low thermal resistance material – meaning they conduct heat very, very well.

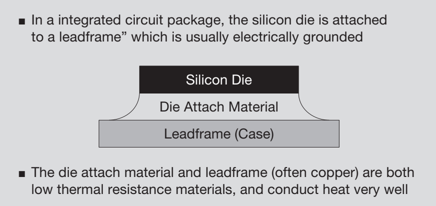


Figure 4.31 R_{thJC} vs. R_{thJA} What is the Package Case?

Before, we said that lead frame and case are often used interchangeably. This is only partially true. The lead frame is the object that the semiconductor die is attached to.

The case is the portion of the lead frame which is exposed to ambient by the encapsulated package and provides the most thermally conductive point between the ambient air and the semiconductor die.

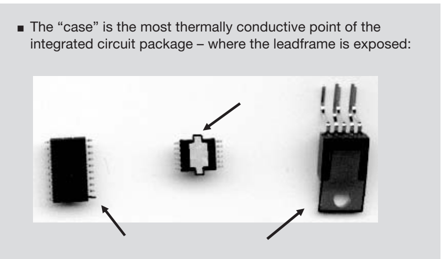


Figure 4.32 R_{thJC} vs. R_{thJA} What is the Package Case?

So, let us look at the temperature difference between the silicon junction and the case:

P_D = 1.5 W
R_{thJC} = 1.1 °C/W

$\Delta T = P_D * R_{thJC}$
$= (1.5 \text{ W}) * (1.1 \text{ °C/W})$
$= 1.65 \text{ °C}$

This is what we expect. For every 1 W of power dissipated within the semiconductor die, there will be a 1.1 °C temperature difference between the junction and the case.

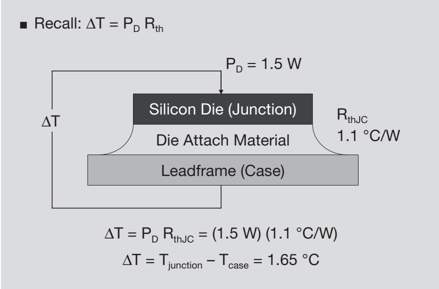


Figure 4.33 R_{thJC} vs. R_{thJA} Case Temperature Difference

Figures 4.33 and 4.35 are good illustrations of why the temperature difference from junction-to-case is so small compared to the temperature difference from case-to-ambient.

- Unlike metal, air is a relatively poor conductor of heat
- Imagine a pot is being heated on the stove
- If you are very close to the pot, you can tell it is hot
- If you touch the pot, you get burned
- There is a large temperature difference from the pot to the air immediately next to the pot
- Therefore, there is a large thermal resistance involved in heat leaving metal and going into the air

Figure 4.34 R_{thJC} vs. R_{thJA}

Every time there is a thermally resistive interface between materials, we expect a larger temperature difference between the materials. This problem is significantly reduced for the junction-to-case interface especially if the semiconductor die is soldered to the lead frame.

Before, it has been mentioned that the datasheet example specifies R_{thJA} and R_{thJC}, so it should be possible to determine R_{thCA} and the temperature difference from case-to-ambient and even junction-to-ambient.

$$R_{thCA} = R_{thJA} - R_{thJ}$$
$$= 80\ °C/W - 1.1\ °C/W$$
$$= 78.9\ °C/W$$

Therefore, the expected temperature difference between the case and the ambient air temperature for every 1 W dissipated within the semiconductor die is 78.9 °C and the temperature difference between the junction and the ambient air temperature for every 1 W dissipated within the semiconductor die is 80 °C.

If 1.5 W is dissipated within the semiconductor die, there will be a 118.35 °C temperature difference between the ambient air and the case. Combining this with the temperature difference from the junction to the case (1.65 °C), 120 °C temperature difference is expected between the semiconductor die and the ambient air.

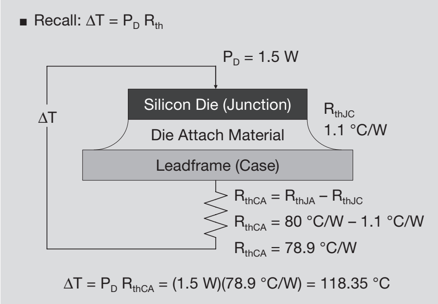


Figure 4.35 R_{thJC} vs. R_{thJA}

From the last calculation the maximum ambient temperature can be determined so, that die temperature is not greater than 150 °C:

$$T_{A,MAX} = T_{J,MAX} - \Delta T_{JUNCTION\text{-}TO\text{-}AMBIENT}$$
$$= 150\ °C - 120\ °C$$
$$= 30\ °C$$

Or, in a slightly more explicit manner:

$$T_{A,MAX} = T_{J,MAX} - R_{thJA} * P_D =$$
$$= 150\ °C - (80\ °C/W)(1.5\ W)$$
$$= 150\ °C - 120\ °C$$
$$= 30\ °C$$

- In Summary:

  $\Delta T_{Junction\text{-}Case} = 1.65\ °C$
  $\Delta T_{Case\text{-}Ambient} = 118.35\ °C$
  $\Delta T_{Junction\text{-}Ambient} = 1.65\ °C + 118.35\ °C = 120\ °C$

- In practice, a 120 °C temperature difference is unrealistic
- A heatsink can be used to reduce the case-to-ambient thermal resistance and the temperature difference

Figure 4.36 R_{thJC} vs. R_{thJA}

This leads us into the topic of heatsinks.

4.6 Heatsinks

Heatsinks can be used to reduce the thermal resistance from the case-to-ambient (R_{thCA}).

Since power (heat) escapes to the ambient air primarily from the surface of the semiconductor case (exposed lead frame), increasing the area of the case will reduce the R_{thCA} value.

Another way to lower the R_{thCA} value, however, would be to use a heatsink. A heatsink is an metallic object which attaches to the semiconductor case (with a very low thermally resistant path) to effectively increase the surface area of the case.

In the previous example the thermal resistance, R_{thCA}, is 80 °C/W. Theoretically, if the case surface area could be doubled or quadrupled, the R_{thCA} value would decrease to 40 °C/W or 20 °C/W, respectively.

In general, the larger the surface area, the lower the R_{thCA} of a heatsink. In practice the thermal resistance is not strictly inversely proportional to the surface area of the heatsink. Still, larger heatsink surface areas will result in lower values of thermal resistance case-to-ambient (R_{thCA}).

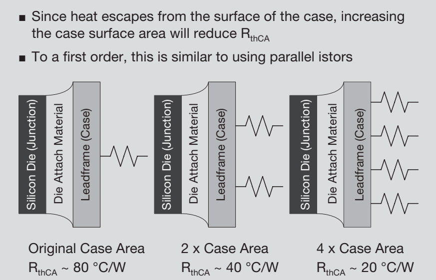


Figure 4.37 Heatsinks

Figure 4.38 shows an example of three different ways to heatsink a TO252 DPAK package to a standard FR-4 board using a 1 oz. copper layer (fire retardant epoxy board with 1 oz. copper evenly spread over 1 square-foot).

In the top right example, the case of the TO252 DPAK package is soldered to the PCB copper directly beneath the package. No additional board space is made available for heatsinking (0 mm^2 of additional board area for heatsinking). From the graph, this results in a R_{thJA} value of about 144 °C/W.

Note how the thermal resistance can be decreased by increasing the amount of heatsinking. In the middle layout, an additional 300 $mm.^2$ (about 0.5 in^2) of board space is also connected thermally and electrically to the TO252 DPAK package. By dedicating the additional 300 mm^2 of board space, the designer is able to reduce thermal resistance almost by 50% (78 °C/W). Finally, in the left example, 600 mm^2 of board space are dedicated to heatsinking, resulting in a thermal resistance of 54.7 °C/W.

Note that the slope of the line decreases as the area of the heatsink increases.

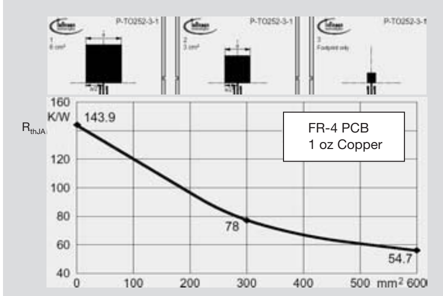


Figure 4.38 Surface Mount Heatsinks (TO252 DPAK)

4.7 Example DC Thermal Calculations

Let us now look at some simple DC thermal calculations.

Here are the relevant thermal characteristics from a semiconductor datasheet.

The maximum operating temperature is 150 °C.

The thermal resistance R_{thJA} is 55 K/W assuming the device is mounted on a FR-4 PCB with an additional copper heatsinking area of 6 cm^2 (using a 70 μm copper thickness or double sided 1 oz. board). Note:

$6\ cm^2 = 600\ mm^2 = 0.93\ in^2 = \text{approx.} = 1\ in^2$

The next semiconductor device that we will examine is an electrical switch. The ELECTRICAL RESISTANCE of the semiconductor switch is R_{DSon} = 24 mΩ (worst case).

Maximum Ratings at T_j = 25 °C unless otherwise specified

Parameter	Symbol	Values	Unit
Operating temperature range	T_j	-40 ... +150	°C

Thermal Characteristics

Parameter and Conditions	Symbol	Values			Unit
		Min.	Typ.	Max.	
Thermal resistance chip - case:	R_{thJC}	–	–	1.1	K/W
junction - ambient (free air):	R_{thJA}	–	80	–	
SMD version, device on PCB [6)]		–	45	55	

6) Device on 50 mm * 50 mm * 1.5 mm epoxy PCB FR4 with 6 cm^2 (one layer, 70 µm thick) copper area for V_{bb} connection. PCB is vertical without blown air.

Load Switching Capabilities and Characteristics

Parameter and Conditions	Symbol	Values			Unit
		Min.	Typ.	Max.	
On-state resistance (pin 3 to pin 1, 5)					
V_{bb} = 5.5 V, I_L = 7.5 A T_j = 25 °C	R_{ON}	–	10	14	mΩ
T_j = 150 °C		–	18	24	
V_{bb} = 12 V, I_L = 7.5 A T_j = 25 °C		–	8	10	
T_j = 150 °C		–	14	18	

Figure 4.39 DC Thermal Calculation MOSFET or Driver

The maximum ambient temperature of the application is 85 °C.

The current flowing through the switch is 5 A.

The power dissipated in the switch can be calculated:

$$
\begin{aligned}
P &= V * I \\
&= (I * R) * I \\
&= I^2 * R \\
&= (5\ A)^2 * (24\ m\Omega) \\
&= 0.6\ W
\end{aligned}
$$

From the datasheet, the thermal resistance with 6 cm^2 of board space dedicated to heatsinking is:

R_{thJA} = 55 °C/W

- Conditions: $T_{ambient}$ = 85 °C, I_{load} = 5 A
- Power Dissipation

  $P_D = I^2 R = (5\ A)^2(24\ m\Omega) = 0.6\ W$

- Thermal Resistance (with 6 cm^2 Copper)

  R_{thJA} = 55 °C/W

- Junction Temperature

  $T_{junction} = T_{ambient} + P_D R_{thJA}$

  $T_{junction}$ = 85 °C + (0.6 W)(55 °C/W) = 118 °C

Figure 4.40 DC Thermal Calculation MOSFET or Driver

Therefore, the junction temperature can be calculated as:

$$
\begin{aligned}
T_{JUNCTION} &= T_{AMBIENT} + \Delta T_{AMBIENT\text{-}JUNCTION} \\
&= T_{AMBIENT} + P_D * R_{thJA} \\
&= 85\ °C + (0.6\ W) * (55\ °C/W) \\
T_{JUNCTION} &= 118\ °C
\end{aligned}
$$

Let’s see if this makes sense. The thermal resistance from junction-to-ambient is 55 °C/W. Therefore, for every 1 W of power dissipated in the semiconductor device, there will be a 55 °C temperature difference between the semiconductor die and the ambient air. For 0.6 W of power, however, there is only a 33 °C (0.6 W * 55 °C/W) temperature difference between the semiconductor die and the ambient air. Based upon an ambient air temperature of 85 °C, this results in a junction temperature of 118 °C.

- Conditions: $T_{ambient}$ = 85 °C, I_{load} = 5 A
- Power Dissipation

  $P_D = I^2 R = (5\ A)^2(24\ m\Omega) = 0.6\ W$

- Thermal Resistance (with 6 cm^2 Copper)

  R_{thJA} = 55 °C/W

- Junction Temperature

  $T_{junction} = T_{ambient} + P_D R_{thJA}$

  $T_{junction}$ = 85 °C + (0.6 W)(55 °C/W) = 118 °C

Figure 4.41 DC Thermal Calculation MOSFET or Driver

Another simple DC thermal calculation is applied to a linear voltage regulator.

Here are the relevant thermal characteristics from a semiconductor datasheet for a voltage regulator.

The maximum operating temperature is 150 °C.

The thermal resistance R_{thJA} is again 55 K/W (°C/W) assuming the device is mounted on a FR-4 PCB with an additional copper heatsinking area of 6 cm^2 (using a 70 µm copper thickness). Recall that 6 cm^2 =approx. 1 in^2.

Maximum Ratings at T_j = 25 °C unless otherwise specified

Parameter	Symbol	Values	Unit
Operating temperature range	T_j	-40 ... +150	°C

Termal Characteristics

Parameter and Conditions	Symbol	Values			Unit
		Min.	Typ.	Max.	
Thermal resistance chip - case:	R_{thJC}	–	–	1.1	K/W
junction - ambient (free air):	R_{thJA}	–	80	–	
SMD version, device on PCB [6)]		–	45	55	

6) Device on 50 mm * 50 mm * 1.5 mm epoxy PCB FR4 with 6 cm^2 (one layer, 70 µm thick) copper area for V_{bb} connection. PCB is vertical without blown air.

Figure 4.42 DC Thermal Calculation Voltage Regulator

The maximum ambient temperature of the application is 85 °C.

The voltage regulator must supply 0.1 A of current at 5 V. The input to the voltage regulator is 14 V.

The power dissipated in the regulator transistor can be approximated as:

$$P = V_{DROP} * I_{OUT}$$

$$= (V_{IN} - V_{OUT}) * (I_{OUT})$$

$$= (14\ V - 5\ V)(0.1\ A)$$
$$= (9\ V)(0.1\ A)$$
$$= 0.9\ V*A$$
$$= 0.9\ W$$

From the datasheet, the thermal resistance with 6 cm^2 of board space dedicated to heatsinking is:

R_{thJA} = 55 °C/W

Therefore, the junction temperature can be calculated as:

$$T_{JUNCTION} = T_{AMBIENT} + \Delta T_{AMBIENT\text{-}JUNCTION}$$

$$= T_{AMBIENT} + P_D * R_{thJA}$$

$$= 85\ °C + (0.9\ W) * (55\ °C/W)$$
$$= 134.5\ °C$$

or there will be a 0.9 x 55 = 49.5 °C temperature rise over the 85 °C ambient temperature.

- Conditions: $T_{ambient}$ = 85 °C, V_{IN} = 14 V, V_{OUT} = 5 V, I_{OUT} = 100 mA
- Power Dissipation

  P_D = V I = (14 V – 5 V)(100 mA) = 0.9 W

- Thermal Resistance (with 6 cm^2 Copper)

  R_{thJA} = 55 °C/W

- Junction Temperature

  $T_{junction} = T_{ambient} + P_D\ R_{thJA}$

  $T_{junction}$ = 85 °C + (0.9 W)(55 °C/W) = 134.5 °C

Figure 4.43 DC Thermal Calculation Voltage Regulator

5. Advanced Power Dissipation and Dynamic Thermal Analysis

In this chapter, we will be examining more of the advanced concepts related to power dissipation and thermal analysis in semiconductor devices.

First, we will examine the corollary between electrical parameters (voltage difference, current, electrical resistance and electrical capacitance) and thermal parameters (temperature difference, dissipated power, thermal resistance, and thermal capacitance).

Those parameters are dependent upon the physical properties of materials. This will lead us into developing the RC models that can be used for thermal analysis. Based on this information, we will introduce the Z_{th} diagram, and how it is used.

After the introduction of the Z_{th} diagram, we will perform a couple transient thermal calculations. We will close Chapter 5 by introducing the concept of analyzing complex power waveforms by applying the superposition principle.

5.1 Electrical and Thermal Parameters

A review is first presented of the analogy between electrical and thermal parameters. It is assumed everyone is very familiar with parameters in the electrical domain.

George Simon Ohm discovered the linear relationship between current flowing through a resistor causing a voltage drop across it. The voltage difference across a resistor is equal to the current through a resistor multiplied by its electrical resistance, or

$V = I \cdot R$

In the thermal domain, Ohm's law can be written as:

$\Delta T = P_D R_{th}$

The temperature difference across an object is equal to the power dissipated in the object multiplied by its thermal resistance.

Electrical resistance in the electrical domain corresponds to thermal resistance in the thermal domain.

Voltage difference (potential) in the electrical domain analogous to temperature difference in the thermal domain, while current in the electrical domain corresponds to dissipated power in the thermal domain.

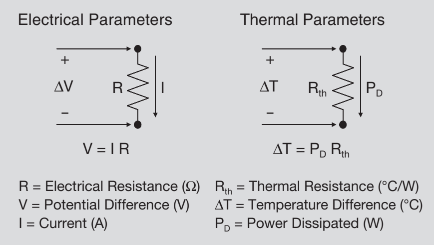


Figure 5.1 Electrical vs. Thermal DC Parameters

The equations defining electrical resistance and thermal resistance are also very similar.

The electrical resistance of any conducting matter is proportional to its thickness and inversely proportional to its area and an electrical property of the matter – its electrical conductivity. The electrical conductivity is a constant for a given type of material.

As the thickness of an object increases, the electrical resistance of the object increases (additional series resistance increases total resistance).

With increasing cross sectional area of the object its electrical resistance decreases (additional parallel resistance decreases total resistance).

There is a complete analogy in the thermal domain.

The thermal resistance of any material is proportional to its thickness and inversely proportional to its cross sectional area and a physical property - its thermal conductivity. The thermal conductivity is a constant for a given type of material.

As the thickness of an object increases, the thermal resistance of the object increases (additional series resistance increases total resistance).

As the cross sectional area of an object increases, the thermal resistance of the object decreases (additional parallel resistance decreases total resistance).

In general, materials that have “good” (high) electrical conductivity also have “good” (high) thermal conductivity.

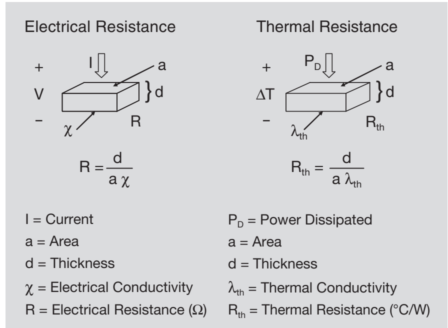


Figure 5.2 Electrical Resistance vs. Thermal Resistance

Usually the units of thermal resistance are given as "°C/W" (centigrade/Watt). Sometimes, the units are also given as "K/W" (degrees Kelvin/Watt). The values, however, remain the same. That is, a thermal resistance of 10 °C/W is the same as 10 K/W. A 1 K change is the same as 1 °C change in temperature and they can be used interchangeably.

For example:

$\Delta T = P_D * R_{th}$ = (10 W) * (1 °C/W) = 10 °C

There is a 10 °C temperature difference across the object.

$\Delta T = P_D * R_{th}$ = (10 W) * (1 K/W) = 10 K

There is a 10 K temperature difference across the object.

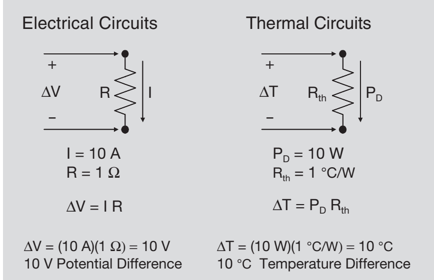


Figure 5.3 Electrical Circuits vs. Thermal Circuits

Just like resistance, capacitance is found in both electrical and thermal domains. An electrically capacitive object has the ability to store electrical energy. A thermally capacitive object has the ability to store thermal energy.

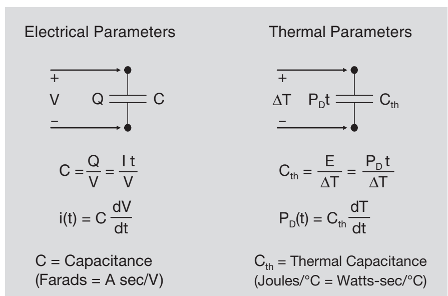


Figure 5.4 Electrical vs. Thermal Parameters

In the electrical domain, capacitance (C) is defined as the ratio of charge on the capacitive object (Q) and the voltage (V) across the capacitive object. In other words, if a given charge Q is held in a capacitor causing a V voltage difference across the capacitor, the object's electrical capacitance is given by Q / V.

Energy must be dissipated in any material to raise its temperature. The thermal capacitance is given by the ratio of energy dissipated in the object and the temperature increase.

Another way of defining energy is that it is the product of power and time. Therefore, if a certain level of power is dissipated in an object for a given amount of time, and the temperature of the object increases by ΔT, then the thermal capacitance of the object is $P_D t / \Delta T$. For example, if 10 W is applied to an object for 1 second, and the temperature rise of the object was 10 °C, then the thermal capacitance of the object would be 1 W-s/°C (note that 1 W-s = 1 J[oule]). If a second object, however, had a thermal capacitance ten times higher (10 J/°C), and the same energy was applied (10 W for 1 second), then the temperature increase would only be 1 °C:

$\Delta T = P_D * t / C_{th}$
= (10 W) * (1 sec) / (10 J/°C)
= 1 °C

An object with a larger thermal capacitance (or thermal mass) is more resistant to temperature changes for a given amount of energy.

5.2 Thermal RC Networks

Based upon our understanding of the similarity of thermal parameters and electrical parameters, we can model thermal systems as a network of thermal resistances and thermal capacitances.

Figure 5.5 shows the cross section of a MOSFET. Starting at the bottom of the MOSFET is the metal (often copper) leadframe. The leadframe is attached to the silicon substrate (n^+ material) by a die attach material. The leadframe also serves as the drain contact for the MOSFET.

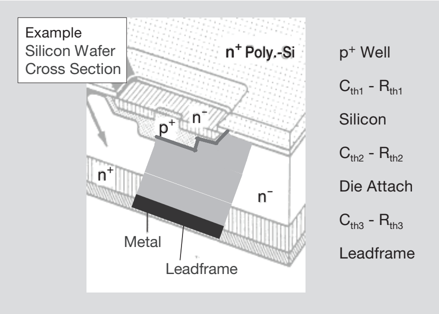


Figure 5.5 Thermal Resistance & Capacitance

On top of the silicon substrate is the silicon epitaxial (also called epi) layer (n^- material). At the top of the epi layer is a p^+ well. This is the source connection for the MOSFET. The interface between the well and the epi layer is called the transistor's "junction". Finally, a n^+ layer is added to the top of the p^+ well to help form an ohmic (non-rectifying) contact for the source. Elsewhere above the epi layer is a thin layer of silicon dioxide and finally a n^+ layer of polycrystalline (or poly) silicon to form the MOSFET's gate.

The first thermal resistive and capacitive elements are found at the MOSFET junction, between the p^+ well and the n^--type silicon. The second thermal resistive and capacitive elements are at the interface of the n^+ type silicon and the solder/die attach. The third thermal resistive and capacitive elements are at the interface of the solder/die attach and the leadframe.

The MOSFET cross section can be represented as a thermal circuit. At the junction, power is being dissipated:

$P_D = I^2 R$

The power has to "flow" out to the ambient environment. This elevates the junction temperature higher than the ambient temperature:

$\Delta T = T_{JUNCTION} - T_{AMBIENT}$

Remember the correlation between the electrical domain and the thermal domain:

Electrical Domain	Thermal Domain
$V = I R$	$\Delta T = P_D R_{th}$

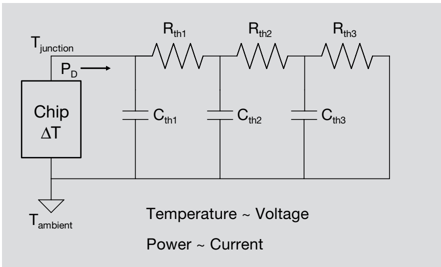


Figure 5.6 Thermal RC Network - Internal

Usually the power semiconductor devices are not just "hanging around" in ambient environment by themselves. Heatsinks are often used to lower the overall impedance of the system. In the past, for leaded devices, "real" mechanical heatsinks would be attached to the semiconductor components with very low thermal resistance and high thermal capacitance to lower the overall system impedance.

Automotive modules with surface mount designs, however, usually are just using copper surface area on their circuit boards. This is not as good a heatsink as the real mechanical devices. The copper on the board will have larger thermal resistance and lower thermal capacitance than its mechanical counterparts. Sometimes, heat pipes are used from one side of the board, down to the other side, and out to a mechanical metal plate to

significantly lower the thermal resistance and increase the thermal capacitance.

The first stage of external thermal resistance and thermal capacitance components is determined by the physical properties of the interface between the semiconductor device and the heatsink (the shaded area). This interface could be solder, glue, or insulating pads. Each interface has unique thermal characteristics. The final stage of thermal resistance and thermal capacitance depends on the physical properties of the heatsink itself.

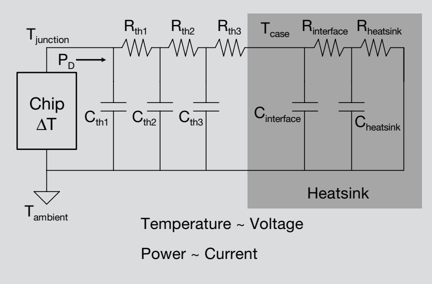


Figure 5.7 Thermal RC Network - Total

Envisioning this network as an electrical circuit (recall that dissipated power is the "current" and the temperature difference is the "voltage"), it is apparent how the current (power) will start to flow.

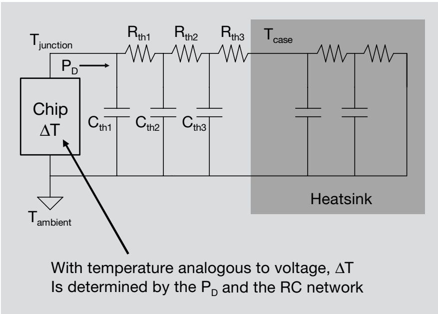


Figure 5.8 Junction Temperature Calculations

You can visualize (in the time domain) how the voltage difference (temperature difference) will increase as the current (power) flows through the network over time.

First, the current (power) will flow through C_{th1}. As it becomes charged up, current (power) will flow through R_{th1} and C_{th2}. This will create a voltage (temperature) difference between $T_{JUNCTION}$ and $T_{AMBIENT}$. When C_{th2} is charged up, current (power) will flow through R_{th1}, R_{th2}, and C_{th3}. This will create a larger voltage (temperature) difference between $T_{JUNCTION}$ and $T_{AMBIENT}$. This continues until all the thermal capacitors are charged up and the current (power) is flowing through the entire series resistance with a maximum voltage (temperature) difference between $T_{JUNCTION}$ and $T_{AMBIENT}$.

The maximum junction temperature that a device can withstand is always specified in the datasheet. Usually, for power devices, the maximum operating junction temperature is 150 °C. Special power devices may have higher maximum junction temperatures.

Many semiconductor components that are not intended for automotive applications have a maximum specified junction temperature of 125 °C or even less.

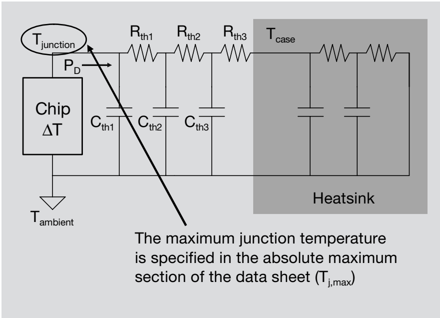


Figure 5.9 Junction Temperature Calculations

In the datasheet, a value for R_{thJC} is typically given for DC conditions. It does not include the thermal capacitor elements and it consists of the sum of the R_{THX} elements between junction and case.

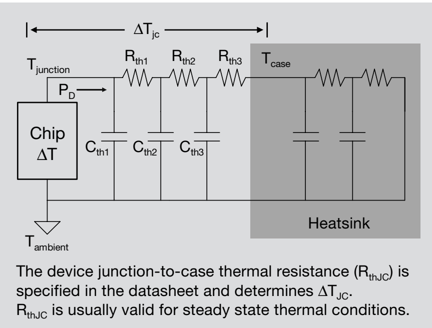


Figure 5.10 Junction Temperature Calculations

The heatsink determines the rise in temperature from the case to the ambient. This temperature difference, ΔT_{CA}, can be very large if the thermal resistance of the heatsink is high. The heatsink can have large effect on the junction temperature when power is dissipated in the semiconductor device.

In addition to selecting a heatsink, the method of attaching the heatsink to the semiconductor device also needs to be specified: grease, pad, etc. Each interface has an associated thermal resistance and capacitance.

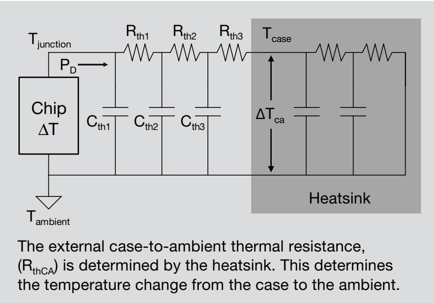


Figure 5.11 Junction Temperature Calculations

Remember, datasheets usually only specify steady state conditions and include only the resistive elements of the above network.

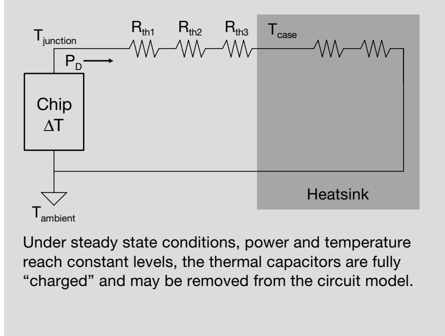


Figure 5.12 Steady State (DC) Junction Temperature Calculations

The next example deals with a DC switch circuit (the switch is always on).

The switch is specified by its resistance when it is on (R_{DSon} = 24 mΩ). The current flowing through the switch is 5 A. Therefore, 0.6 W is dissipated in the switch when it is on.

The thermal resistance, R_{thJA}, specified in this example is 55 °C/W. This means that for every 1 W dissipated in the switch, the junction temperature will increase 55 °C above the ambient temperature. The temperature increase caused by 0.6 W dissipation is 33 °C.

Therefore, for an ambient temperature of 85 °C, the junction is heated to 118 °C temperature when the switch is left on.

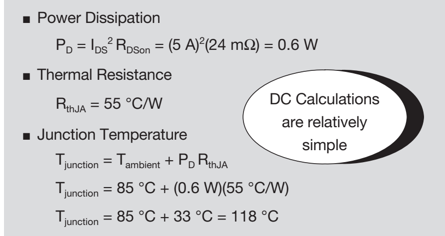


Figure 5.13 Steady State (DC) Junction Temperature Calculation

For transient conditions, the situation looks much more complicated. The network of series thermal resistances and thermal shunt capacitances now form a complex network with

a thermal impedance which can be specified in the frequency domain ($Z_{th}(j\omega)$).

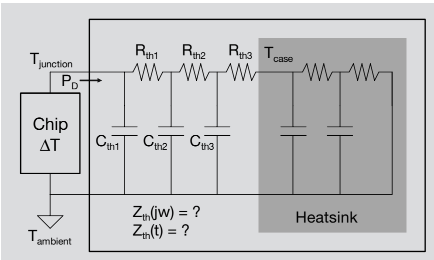


Figure 5.14 Dynamic (AC) Junction Temperature Calculation

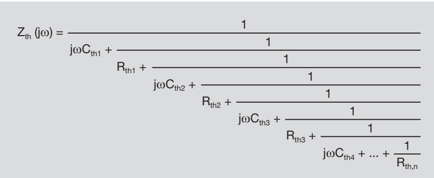


Figure 5.15 Dynamic (AC) Junction Temperature Calculation

Calculating the thermal impedance becomes a very complex problem. This is even more the case as the model becomes more realistic and additional layers of thermal resistance and capacitance are added.

What can we do to simplify this?

One often used method is a parallel to series transformation. The network can be made much more manageable if the shunt capacitors are changed into series capacitors. This will lead us to a simpler thermal impedance calculation.

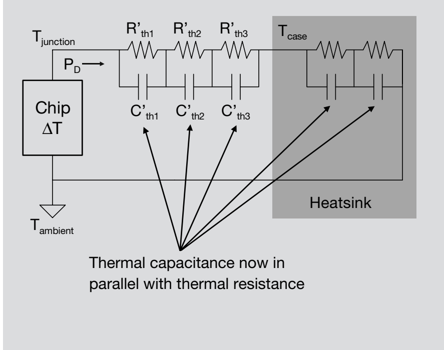


Figure 5.16 Simplified Dynamic (AC) Thermal RC Network

Remember, components R' and C' are mathematical abstractions resulting from the parallel to series transformation.

Previously, the R and C components referred to physical properties of various layers in a semiconductor device. But, all lumped together, these new R' and C' components provide the same thermal impedance as the previous network.

One of the reasons we do this parallel to series transformation is that each parallel RC combination impedance can be calculated simply (and separately) in both the frequency domain and in the time domain.

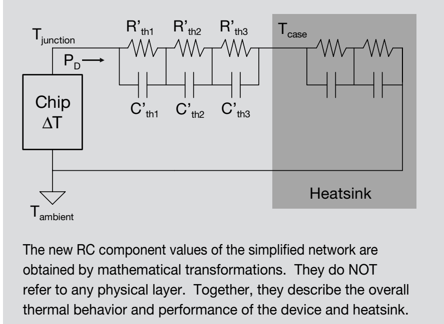


Figure 5.17 Simplified Dynamic (AC) Thermal RC Network

The frequency domain transfer function and the time domain response of the temperature of a single RC circuit to a step function power

input pulse of P_d magnitude is illustrated in Figure 5.18:

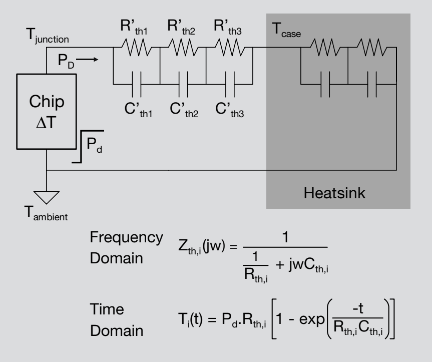


Figure 5.18 Dynamic (AC) Junction Temperature Calculation

To get the dynamic change in temperature from a power dissipation step function input, the Z_{th} transfer characteristics must be established. The $T_i(t)/P_d$ ratio is traditionally known Z_{thJA} or junction to ambient thermal impedance.

To get the total thermal impedance (Z_{th}), one simply sums the impedances of each R'C' pair. This value of Z_{th} multiplied by the power dissipation will result in the temperature difference at any given point in time.

5.3 Understanding the Z_{th} Diagram

Even though this method of calculating Z_{th} is simpler than analyzing the actual thermal RC network, the analysis must still be performed with a electrical circuit modeling program or a complex spreadsheet. For a very quick and simple method, this is probably still too complex - unless you already have all of the model information programmed into your spreadsheet or modeling software.

The tool usually used for quick calculations is the Z_{th} diagram. The Z_{th} diagram is developed by first building a custom test set-up for each device.

Specific power dissipation pulses are created in the semiconductor device for defined lengths of time and changes in the junction temperature are monitored (often with an infrared thermometer). The temperature rises more for stronger pulses (more power) or longer pulses (more time). Remembering that power multiplied by time is energy, more energy dissipated in the device means a larger junction temperature increase.

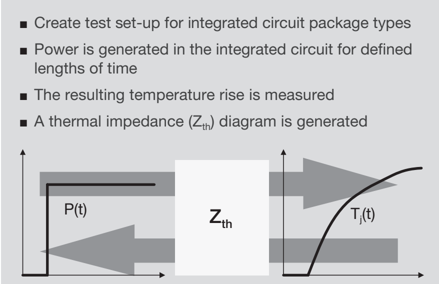


Figure 5.19 Development of the Z_{th} Diagram

After these measurements are made, a graph of the thermal impedance can be created in the time domain based upon pulse time and duty cycle.

This graph is called a Z_{th} diagram.

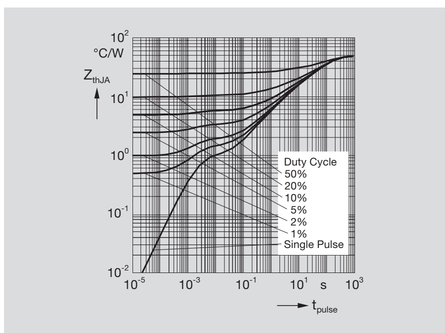


Figure 5.20 Z_{th} Diagram for the TO263 Package

On the x-axis is the time duration of the pulse (from 10 μs to 1000 s). The vertical axis is the thermal impedance (junction to ambient). Several different curves are shown on the graph - each corresponding to a different power pulse duty cycle. For a given pulse length and duty cycle, a Z_{thJA} value can be readily obtained.

A general trend in the graph is apparent. First, as the length of the pulse increases, the thermal impedance increases. This is due to the fact that the thermal "charge" in the thermal capacitors is increasing, and the network becoming more resistive.

Also, as the duty cycle is increased, the thermal impedance is increased. Note, that for each of the curves on the Z_{th} diagram EXCEPT for the single pulse case, each power pulse train of various duty cycles is assumed to continue for infinite length of time.

In case of an ongoing pulse train the junction temperature will rise during each power pulse and it will fall between pulses. If the temperature does not fall to its initial value during the off time, it will rise higher each subsequent time the pulse is applied, and fall not quite as far each time the pulse is turned off, until the temperature reaches an asymptotic limit set by the corresponding Z_{thJA} value from Figure 5.20.

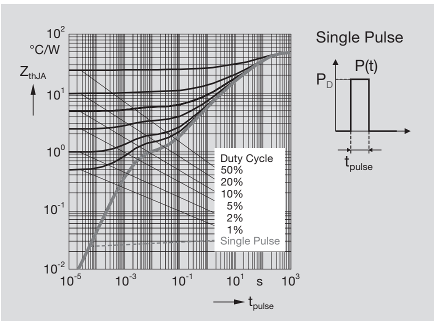


Figure 5.21 Z_{th} Diagram for the TO263 Package

The line corresponding to a single power pulse dissipated in the semiconductor device is shown with a bold line. For very short pulses (for example, 100 µs), the thermal impedance is very low (0.03 °C/W). Z_{thJA} is low because the junction temperature will rise very little if such a short pulse is applied due to the charging of the thermal capacitors. The system is "absorbing" the heat. As the pulse width is increased, the curves start to saturate. At a pulse width of infinite time (the thermal capacitors have been charged with as much thermal energy as they can handle) the Z_{th} diagram lines would saturate at the DC thermal resistance, R_{thJA}. So, for this package, the R_{thJA} value is about 50 °C/W.

As more heatsinking is added, the curve moves to the right – longer pulses are required to generate the same junction temperature increase. The reverse is also true. If the amount of heatsinking is diminished, shorter pulses are required to generate the same junction temperature increase.

Sometimes, Z_{th} diagrams are given for junction to case values (Z_{thJC}). Then, the engineer must determine the thermal characteristics of the application specific heatsink.

The Z_{th} diagram can also be used to find the thermal impedance for periodic events. The highlighted curve in Figure 5.22 is for the 50% duty cycle case. Notice how large the thermal impedance is for even very low duty cycle, repetitive pulses. This is because the pulse train is considered a continuous process and measurements are taken when the thermal system reaches steady state conditions. A higher Z_{thJA} value is observed than when a single pulse of the same width and amplitude heats the thermal system.

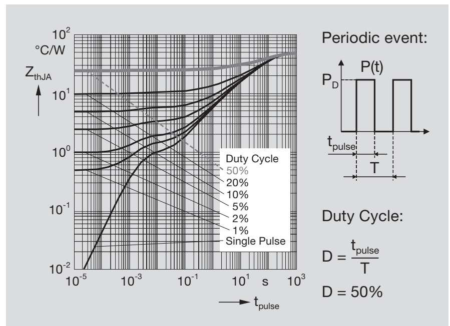


Figure 5.22 Z_{th} Diagram for the TO263 Package

Z_{th} diagrams for four different packages are presented. The differences between the different diagrams are very noticeable. Let us take a look at two of the diagrams (measured with TO252 and SO8 packages).

The TO252 package is rather large and also has a large heatsink within the package. Look at the Z_{thJA} for 600 mm^2 of copper heatsink with a single power pulse of 1 second. It has a thermal impedance of about 4 °C/W. Therefore, if a 1 W pulse is applied to a device in a TO252 package for 1 second, the junction temperature will be 4 °C higher than the ambient temperature. Most of the power has been "absorbed" by the thermal capacitors and only a small junction temperature increase is seen.

Compare this to a rather small SO8 package which has no true integrated heatsink. Look again at the Z_{thJA} for 600 mm^2 of copper heatsink when a single pulse of 1 second is applied. The diagram shows a thermal impedance of about 65 °C/W. Therefore, if a 1 W pulse is applied to a device in a SO8 package for 1 second, the junction temperature will be 65 °C higher than the ambient temperature. Very little of the power has been "absorbed" by the thermal capacitors, and that is why the junction temperature increase is large.

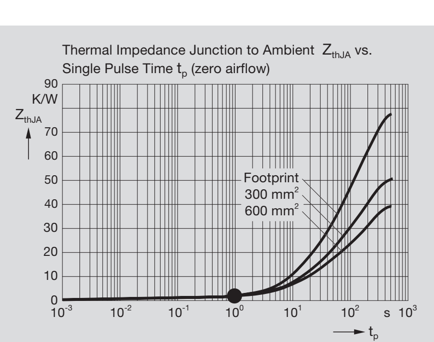


Figure 5.23 Z_{th} Diagram for the P-TO263-5-1 Package (t_p = 1 s, $Z_{thJA} \cong$ 2 °C/W)

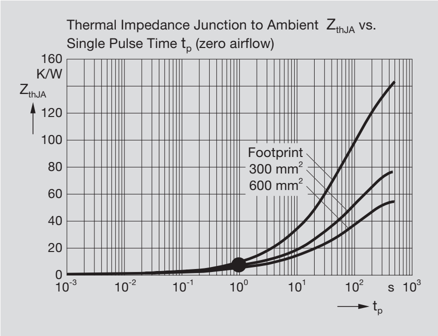


Figure 5.24 Z_{th} Diagram for the P-TO252-3-1 Package (t_p = 1 s, $Z_{thJA} \cong$ 4 °C/W)

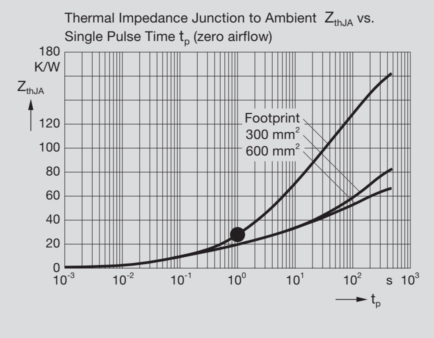


Figure 5.25 Z_{th} Diagram for the SOT223 Package (t_p = 1 s, $Z_{thJA} \cong$ 30 °C/W)

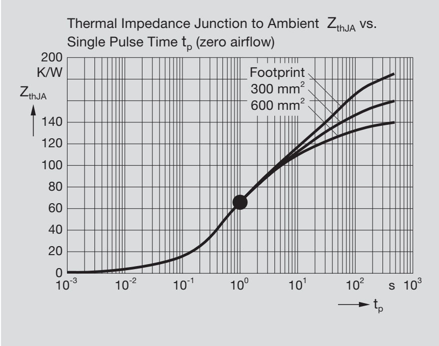


Figure 5.26 Z_{th} Diagram for the SO8 Package (t_p = 1 s, $Z_{thJA} \cong$ 65 °C/W)

5.4 Example Transient Thermal Calculations

Let us say that a single pulse of 400 W is applied to a semiconductor component in a TO263 package for 200 µs at 25 °C starting temperature.

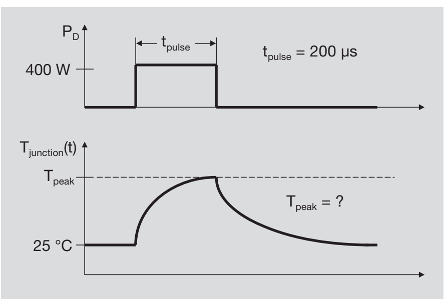


Figure 5.27 Single Pulse in a TO263 Package

While the power is being applied, the junction temperature will continue to rise. It will reach a peak value when the pulse ends. From this point, the junction will begin cooling as the thermal capacitors discharge their thermal energy until the junction temperature again reaches the ambient (25 °C in this example).

The problem is to determine what the peak junction temperature is. We need to verify that the peak temperature remains below the absolute maximum junction temperature specified in the datasheet.

To solve this problem, we examine the single pulse line for 200 µs = 2 x 10^{-4} s in the Z_{th} diagram for a TO263 package which will indicate a thermal impedance of about 0.083 °C/W.

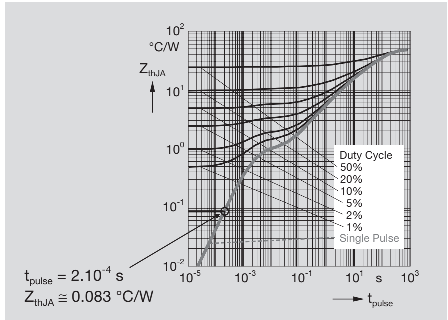


Figure 5.28 Single Pulse in a TO263 Package

The solution now becomes very simple.

With a thermal impedance of 0.083 °C/W, the junction temperature will rise 0.083 °C for every Watt of power dissipated in the device or 33.2 °C with 400 W power dissipation.

Therefore, according to the Z_{th} diagram, the peak junction temperature will be

25 °C + 33.2 °C = 58 °C

Normally, we think of 400 W as being a very large number.

The 400 W, however, is applied for only a very short time (200 µs). Therefore, the junction temperature peak remains below a maximum junction temperature of 150 °C by a large safety margin.

- Power Dissipation

  $P_D = 400$ W

- Thermal Resistance

  $Z_{thja} = 0.083$ °C/W

- Junction Temperature

  $T_{junction,peak} = T_{Ambient} + P_D Z_{thja}$

  $T_{junction,peak} = 25$ °C + (400 W)(0.083 °C/W)

  $T_{junction,peak} = 25$ °C + 33 °C = 58 °C

Figure 5.29 Single Pulse in a TO263 Package

For comparison, Figure 5.30 shows the results of a Saber computer simulation. The simulation shows a 30 °C junction temperature

increase when a 400 W power pulse is dissipated for 200 µs. Comparing this to the 33.2 °C increase from our simple Z_{th} diagram calculation, we can see that sufficiently accurate calculations can be made very quickly using the Z_{th} diagram.

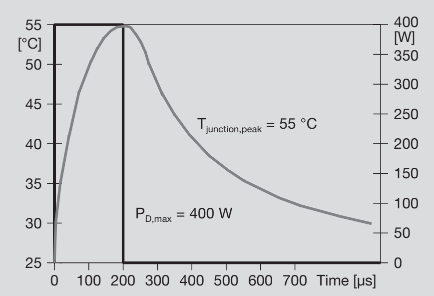


Figure 5.30 Single Pulse - TO263 Package Saber Simulation

For our final example, let us examine the impact of a 50% duty cycle pulse train dissipated in a TO263 package. The pulse width remains 200 µs (so the period is 400 µs).

As the first power pulse is applied, the junction temperature starts to increase. When the thermal pulse is terminated the junction temperature cools slightly. As the second pulse is applied, the junction temperature rises again. When the pulse end the junction cools slightly. This process continues until the junction temperature asymptotically reaches a limit.

The problem is to determine what the peak junction temperature is when the equilibrium temperature is reached. We need to verify that the peak temperature remains below the absolute maximum junction temperature specified in the datasheet.

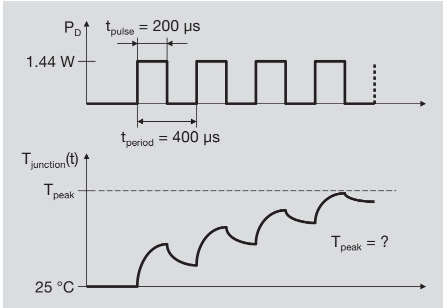


Figure 5.31 50% Duty Cycle in a TO263 Package

To solve this problem reference is made again to the Z_{th} diagram for a TO263 package.

First find the 50% duty cycle line for a pulse width of 200 µs = 2 x 10^{-4} s. We can read from the graph a thermal impedance of about 23 °C/W. Knowing that value the rest of the solution is simple.

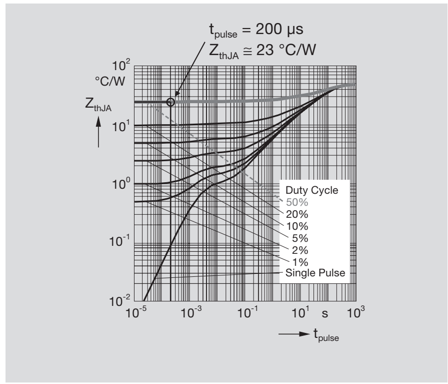


Figure 5.32 50% Duty Cycle in a TO263 Package

With a thermal impedance of 23 °C/W, the junction temperature will rise 23 °C for every Watt of power dissipated in the device. Letting the amplitude of the power pulse be 1.44 W the ΔT is 33.1 °C and the final temperature is 58 °C.

- Power Dissipation

  $P_D = 1.44$ W

- Thermal Resistance

  $Z_{thJA} = 23$ °C/W

- Junction Temperature

  $T_{junction} = T_{ambient} + P_D Z_{thJA}$

  $T_{junction} = 25$ °C + (1.44 W)(23 °C/W)

  $T_{junction} = 25$ °C + 33 °C = 58 °C

Figure 5.33 50% Duty Cycle in a TO263 Package

Note that in the previous example (Figure 5.29), the junction temperature peaked at 58 °C when 400 W was applied once to the device. This time, a relatively small power (1.44 W) is applied, but at a 50% duty cycle for an “infinite” length of time. The peak junction temperature saturates at the same value as for the single 400 W pulse.

5.5 Complex Waveforms and Superposition

Thermal calculations can be extended to more complex power waveforms with the help of the principle of super position. The method will be illustrated with the case of calculating the power dissipation when the MOSFET is switched on.

With reference to the oscillograms, at 10 µs the gate of an NMOSFET is pulled high which initiates the turn-on sequence.

The drain current (I_{DS}) starts to rise and the voltage across the MOSFET, (V_{DS}) begins to fall. The product of the rising current flowing through the MOSFET and the falling voltage across the MOSFET is the power dissipated in the MOSFET. This causes a considerable power loss spike every time a MOSFET is switched on.

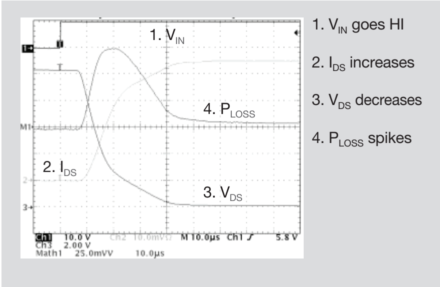


Figure 5.34 Complex Pulse-Superposition MOSFET Turn On

The impulse like trace clearly shows the power dissipation waveform which, in this case, is more complex than those we have previously analyzed. We will show that the junction temperature increase during the power loss spike can be approximated by superimposing several rectangular power pulses (both positive and negative).

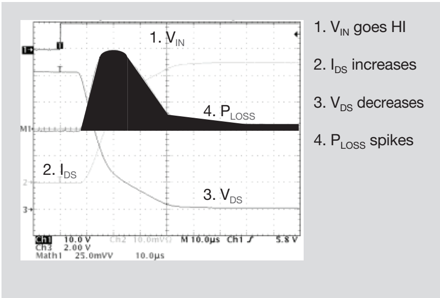


Figure 5.35 Complex Pulse-Superposition MOSFET Turn On

The method of constructing the approximating power pulses is shown next.

The first pulse begins at 20 µs (10 µs after the MOSFET input goes HI) and lasts until 50 µs. The first power pulse will result in an increase in the junction temperature of the MOSFET. This first pulse covers most of the area of the power loss spike, but we can get a more accurate junction temperature calculation by using additional superpositioned power pulses.

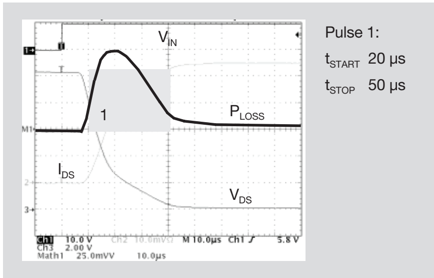


Figure 5.36 Complex Pulse-Superposition MOSFET Turn On

The second pulse begins at 25 µs and also lasts until 50 µs. The second pulse will also result in an increase in the junction temperature of the MOSFET. The addition of the second pulse allows us to cover the peak of the power loss spike, but it also presents us with a new problem. The energy loss represented by the first two pulses will provide an approximation for the power dissipation greater than the actual power spike.

The concept of negative power pulses is introduced to correct for the positive errors caused by the positive pulse rectangles.

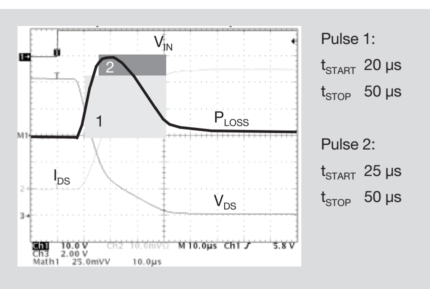


Figure 5.37 Complex Pulse-Superposition MOSFET Turn On

The third pulse starts at 40 µs and also lasts to 50 µs. The third pulse is a negative pulse removing energy from the approximated power dissipation spike comprising of areas 1 and 2. Therefore, the third pulse will result in a superpositioned decrease in the junction temperature of the MOSFET.

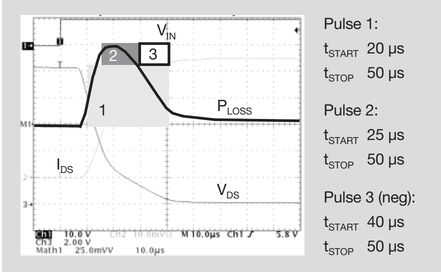


Figure 5.38 Complex Pulse-Superposition MOSFET Turn On

Finally, the fourth power pulse starts at 45 µs and lasts to 50 µs. Like the third pulse, the fourth pulse removes energy from the power dissipation spike. This again results in a superpositioned decrease in the junction temperature of the MOSFET.

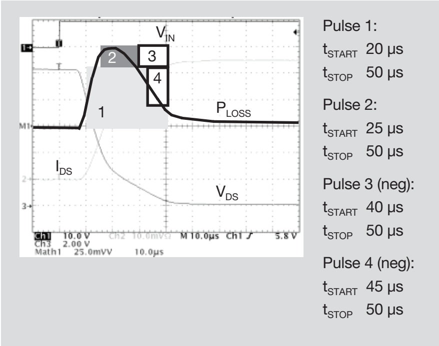


Figure 5.39 Complex Pulse-Superposition MOSFET Turn On

The resulting power loss approximation closely resembles the actual power loss spike.

If additional power pulses were used to more accurately model the actual power loss, the resulting junction temperature increase approximation would be more accurate.

We can now “build up” the junction temperature increase by using the power of superposition.

Figure 5.40 illustrates the superposition of the power pulses from Figure 5.39. The time scale is shifted so that the first power pulse will start at t = 0 µs.

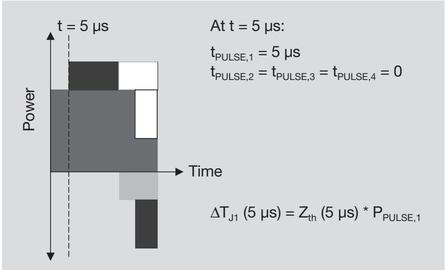


Figure 5.40 Complex Pulse-Superposition MOSFET Turn On

We note that for the first 5 μs, only pulse 1 is active pulses 2-4 have not begun yet. Therefore, the junction temperature increase at 5 μs is only a function of pulse 1. The thermal impedance (Z_{th}) can be read from a Z_{th} diagram single pulse curve with a duration of 5 μs.

We note that from 5 μs to 20 μs, pulse 1 and pulse 2 are occurring. Pulse 3 and pulse 4 have not begun. Therefore, the junction temperature increase at 20 μs is only a function of pulse 1 and pulse 2. The junction temperature increase at 20 μs is the superposition of the junction temperature increase due to pulse 1 and pulse 2.

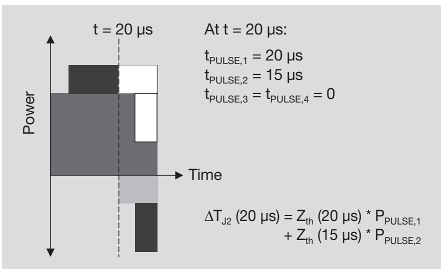


Figure 5.41 Complex Pulse MOSFET Turn On

The thermal impedance of pulse 1 can be read from a Z_{th} diagram single pulse curve with a duration of 20 μs.

The thermal impedance of pulse 2 can also be read from a Z_{th} diagram single pulse curve with a duration of 15 μs.

We note that from 20 μs to 25 μs, pulse 1, pulse 2, and pulse 3 are occurring. Pulse 4 has not begun. Therefore, the junction temperature increase at 25 μs is only a function of pulse 1, pulse 2, and pulse 3. The junction temperature increase at 25 μs is the superposition of the junction temperature increase due to pulse 1, pulse 2, and pulse 3.

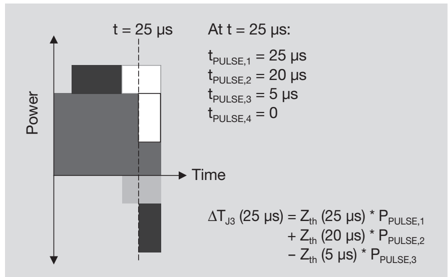


Figure 5.42 Complex Pulse-Superposition MOSFET Turn On

The thermal impedance of pulse 1 can be read from a Z_{th} diagram single pulse curve with a duration of 25 μs.

The thermal impedance of pulse 2 can be read from a Z_{th} diagram single pulse curve with a duration of 20 μs.

The thermal impedance of pulse 3 can be read from a Z_{th} diagram single pulse curve with a duration of 5 μs. Because pulse 3 is a negative power pulse (we are taking power out of the system), pulse 3 results in a junction temperature decrease and the Z_{th}(5 μs) * $P_{PULSE,3}$ value is subtracted from the net junction temperature increase.

We note that from 25 μs to 30 μs, pulse 1, pulse 2, pulse 3, and pulse 4 are acting together.Therefore, the junction temperature increase at 30 μs is caused by all four pulses. The junction temperature increase at 30 μs is the superposition of the junction temperature increase due to pulse 1, pulse 2, pulse 3, and pulse 4.

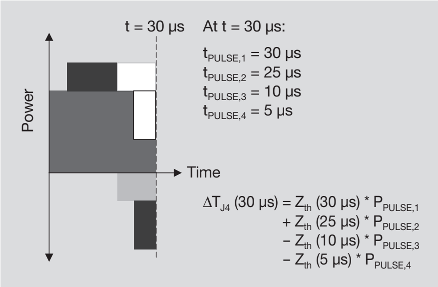


Figure 5.43 Complex Pulse-Superposition MOSFET Turn On

The thermal impedance of pulse 1 can be read from a Z_{th} diagram single pulse curve with a duration of 30 µs.

The thermal impedance of pulse 2 can be read from a Z_{th} diagram single pulse curve with a duration of 25 µs.

The thermal impedance of pulse 3 can be read from a Z_{th} diagram single pulse curve with a duration of 10 µs. Because pulse 3 is a negative power pulse (we are taking power out of the system), pulse 3 results in a junction temperature decrease and the Z_{th} (10 µs) * $P_{PULSE,3}$ value is subtracted from the net junction temperature increase.

The thermal impedance of pulse 4 can be read from a Z_{th} diagram single pulse curve with a duration of 5 µs. Because pulse 4 is a negative power pulse (we are taking power out of the system), pulse 4 results in a junction temperature decrease and the Z_{th} (5 µs) * $P_{PULSE,4}$ value is subtracted from the net junction temperature increase.

Finally, we can go back to the power dissipation curve and illustrate the junction temperature at each of the four points we calculated:

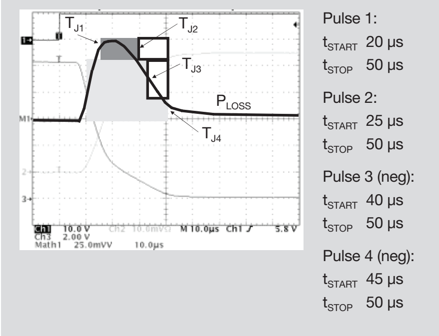


Figure 5.44 Complex Pulse-Superposition MOSFET Turn On

$$\Delta T_{J1}(5\,\mu s) = Z_{th}(5\,\mu s) * P_{PULSE,1}$$

$$\Delta T_{J2}(20\,\mu s) = Z_{th}(20\,\mu s) * P_{PULSE,1} + Z_{th}(15\,\mu s) * P_{PULSE,2}$$

$$\Delta T_{J3}(25\,\mu s) = Z_{th}(25\,\mu s) * P_{PULSE,1} + Z_{th}(20\,\mu s) * P_{PULSE,2} - Z_{th}(5\,\mu s) * P_{PULSE,3}$$

$$\Delta T_{J4}(30\,\mu s) = Z_{th}(30\,\mu s) * P_{PULSE,1} + Z_{th}(25\,\mu s) * P_{PULSE,2} - Z_{th}(10\,\mu s) * P_{PULSE,3} - Z_{th}(5\,\mu s) * P_{PULSE,4}$$

6. MOSFETs, High Side Drivers, and Low Side Drivers

In this chapter, we will examine the Metal Oxide Semiconductor Field Effect Transistor (MOSFET) in two different types of circuits – high and low side drivers.

We begin the chapter with a quick review of MOSFET basics. This will be followed by an introduction to various driver topologies, and how the MOSFET can be used as a solid state electronic switch. We will briefly touch on the topics of protected high side and low side drivers.

The method of selecting MOSFETs for driver applications will then be presented. We will show how the on-state resistance (R_{DSon}) of a MOSFET is determined for both static and switching applications.

Finally, we will examine in detail two special MOSFET driver applications. Technical issues associated with driving both capacitive and inductive loads will be discussed and how using a protected MOSFET driver can help to mitigate some of these challenges.

6.1 MOSFET Review

"MOSFET" is an acronym for "Metal Oxide Semiconductor Field Effect Transistor". The gate of the MOSFET is either "metal" or a polycrystalline silicon material which behaves like a metallic material. (In original MOSFET designs, the gate was metal.) Silicon dioxide ("Oxide") is used to insulate the gate from the underlying semiconductor. It is in the underlying semiconductor that various "n-type" and "p-type" dopants are used to form the "Field Effect Transistor". (It is the "Electric Field" created by the gate-to-source bias voltage which actually turns the MOSFET on and off.)

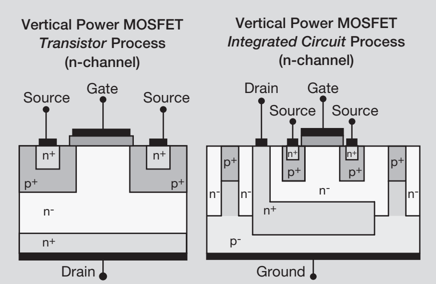


Figure 6.1 Metal Oxide Semiconductor Field Effect Transistor

Here, we show two vertical power MOSFET cross sections. On the left is a power MOSFET from a process optimized for individual power transistors. The MOSFET on the right is from a power integrated circuit process which allows for isolation between various circuit components which allows for integration of other components.

The reader is advised to review Chapters 2 and 3 to refresh his or her knowledge of the basic physics of MOSFETs and their basic operating characteristics.

6.2 Driver Topologies and Loads

In a low side driver configuration, the switch is located between the load and ground.

The n-channel MOSFET is on the "low" side of the load – it is a "low side switch".

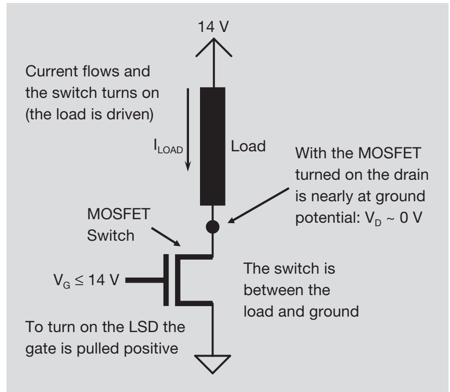


Figure 6.2 Low Side Drive (LSD) Configuration

To turn on the LSD, we apply a positive voltage to the MOSFET gate. When the MOSFET is turned on, the voltage at the transistor drain is usually quite low or nearly at ground potential (because the resistance of the switch is significantly lower than the resistance of the load). The high gate-to-source voltage (V_{GS} = 14 V) and low drain-to-source voltage (V_{DS} ~ 0 V) place the MOSFET in the linear region of operation.

Current can now flow through the load and the MOSFET switch, and the load is "turned on".

Loads can also be driven on the high side. This means that the MOSFET switch is now connected between the supply voltage and the load – the switch is on the "high" side of the load.

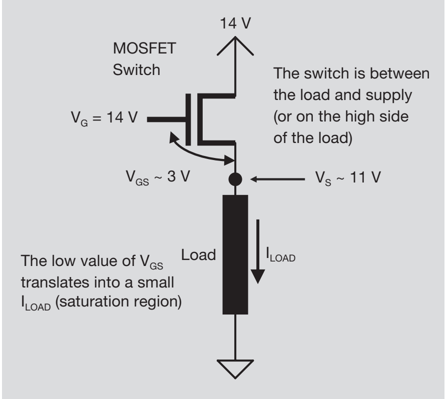


Figure 6.3 High Side Drive (HSD) Configuration

The following description of the operation of the high side driver configuration is based on the Infineon IPD14N06S2-80 (55 V / 80 mΩ) N-channel MOSFET (see the I_D vs. V_{DS} characteristics in Figure 6.4).

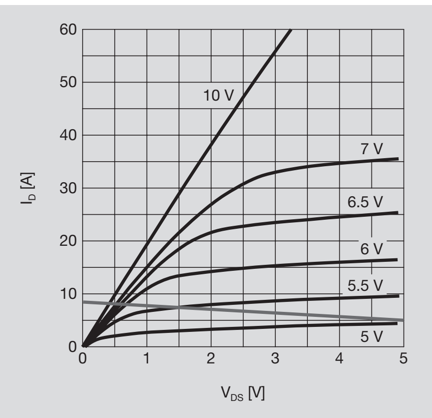


Figure 6.4

Initially, the gate is connected to $\mathbf{V_{BAT} = 14\ V}$ (or the drain and gate voltages are equal: $V_D = V_G$). If the resistive load is 1.8 Ω the voltage drop across the load is **9 V** at $\mathbf{I_D = 5\ A}$. At that current the drain to source voltage, $\mathbf{V_{DS}}$ and the gate to source voltage, $\mathbf{V_{GS}}$ are equal at **5.0 V** (this was approximated using a load line of **1.8** Ω on the above characteristics). The device is operating in the saturation region.

The power dissipation in the MOSFET switch is equal to **P = 5 V x 5 A = 25 W!** This amount of power dissipated in the switch is very wasteful and would require careful and intense cooling.

The problem can be solved by using gate voltages greater than the drain voltage (i.e. greater than V_{BAT}) in order to bias the device in the linear region.

Taking the $\mathbf{V_{BAT} = 14\ V}$ case, assume that $\mathbf{V_G = 24\ V}$. If $\mathbf{V_S \sim V_{BAT}}$, the load current is approximately $\mathbf{I_L = I_D \sim 7.8\ A}$. The gate to source voltage, $\mathbf{V_{GS}}$ is **10 V** which is more than enough to insure operation in the linear region. The MOSFET is **fully enhanced** and the $\mathbf{R_{DSon}}$ is approximately **80 mΩ**.

The power dissipation in this case is given by:

$$\mathbf{P = I_D^2 \times R_{DSon} = 7.8^2 \times 0.08 = 4.8\ W}$$

The dissipation in this case is only 19.2% of the previous one!

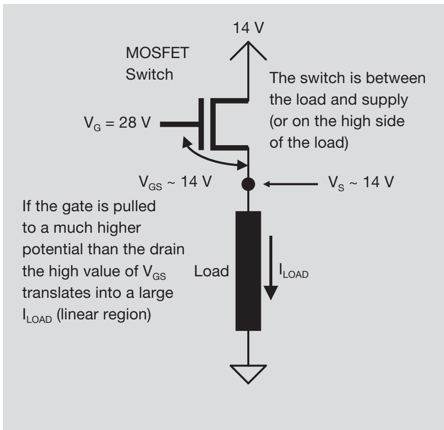


Figure 6.5 High Side Drive (HSD) Configuration

So, the fundamental difference in high side drive applications is that a voltage higher than the drain voltage needs to be applied to the MOSFET gate. The gate voltage is higher than in a low side application.

The decision to use a high side driver or low side driver usually is made by a vehicle OEM during the architectural decision phase. More

than ten years ago, it was common to use low side drive configurations because the switch itself is simpler and less expensive.

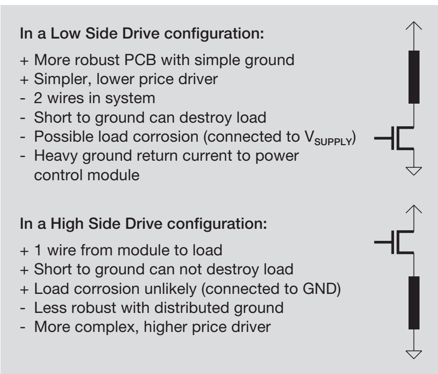


Figure 6.6 Low Side Drivers vs. High Side Drivers

While high side drivers have become more commonplace over the last several years, low side drivers still offer important advantages:

1) Low side drivers are simpler to implement than high side drivers. This may result in a lower price than a comparable high side driver.

2) Low side drivers do not use a distributed ground like high side drivers. Rather, the ground for all the low side drivers is common to the module ground. This can simplify the electrical design of the printed circuit board.

The final drive configuration presented briefly is the H-Bridge drive. The name is descriptive of the circuit topology (see Figure 6.7). Usually, H-Bridge drivers are used to drive motors in two directions (clockwise and counter-clockwise).

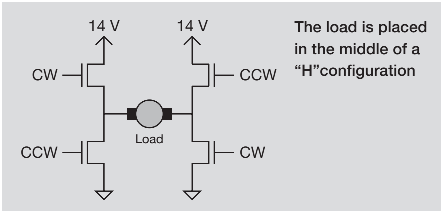


Figure 6.7 H-Bridge Configuration

The motor may have to drive the load in one direction, CW, (for example, to open a window on a car door), then the "CW" legs of the H-Bridge are turned on by raising the voltage at the "CW" inputs. Current will now flow through the motor from the left to the right which will cause the motor to turn in one direction (CW).

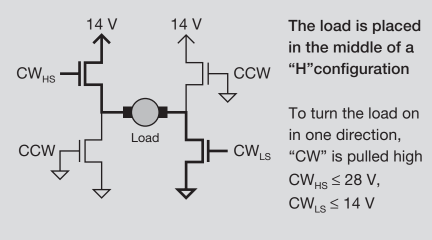


Figure 6.8 H-Bridge Configuration

If we want to drive the motor in the other direction, CCW, the gate voltages will be raised at the "CCW" inputs. Current will now flow through the motor from the right to the left. This causes the motor to work in the CCW direction.

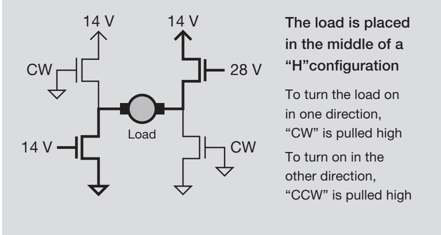


Figure 6.9 H-Bridge Configuration

Each high side switch (right or left) is always used with the opposite low side switch (left or right). The name H-Bride configuration comes from the resemblance of the circuit to the

capital letter “H”. Each high side / low side switch pair is one leg of the “H” and the motor is the crossing between the legs.

In the H-bridge circuit higher gate voltage must be applied to the high side drive MOSFET switches to turn them on in the linear region in order to maximize current flow through the motor and to reduce the power dissipation.

Here are a few words about the types of loads MOSFET high side drivers and low side drivers control.

All real loads have some resistive, inductive, and capacitive elements. The load model will incorporate the various RLC elements for use in driver design and selection.

It is easy to design for resistive loads.

Loads with significant inductive or capacitive components, however, can add design complexities. We will examine these issues in the second half of this chapter.

All real loads are inherently complex:

- Resistive:
  - – Heaters (seats, mirrors, etc.)
- Inductive:
  - – Motors
  - – Relays
- Capacitive and capacitive like:
  - – Lamps (in-rush currents)
  - – Capacitance banks (distributed power supplies such as airbag squib driver power supplies)

Figure 6.10 MOSFET Loads

6.3 Protected Drivers

As a precursor to our future discussions, we will introduce briefly here the concept of protected MOSFET drivers.

For example, Infineon Technologies Corporation manufactures a line of **protected high side drivers entitled PROFETs** (PROtected Field Effect Transistors). A PROFET is a standard MOSFET with additional features that have been integrated to enhance its performance in high side drive applications.

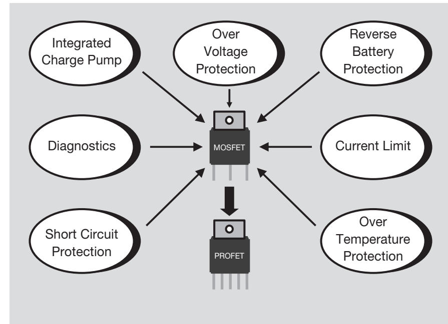


Figure 6.11 PROFETs = PROtected FETs

First, a charge pump is included in the PROFET. As mentioned before, the charge pump circuit boosts the supply voltage. This boosted voltage is then applied to the MOSFET gate when it is turned on to ensure the MOSFET operates in the linear region during hi-side drive operation.

In addition, several protection features are included in the PROFET. For example, the short circuit protection can protect the MOSFET if a short to ground occurs. The over voltage protection ensures the MOSFET will not be damaged during load dump transients or other high voltage events (assuming the transient pulse energy is within specified limits). The reverse battery protection keeps the PROFET safe if the battery polarity is reversed (current may flow, but the device will not be damaged). The current limit feature limits the maximum current which the MOSFET high side switch can source. This again helps to protect the switch from short to ground conditions. The over temperature protection is enabled if the temperature of the MOSFET junction reaches unsafe levels, the switch is turned off stopping the current flow completely.

Finally, diagnostics are included so the PROFET can provide feedback to a microcontroller if a fault condition is detected.

Figure 6.12 is a block diagram of a typical PROFET incorporating the features previously mentioned. Notice the vertical, n-channel

MOSFET and its integrated body diode (the body diode is created by junction of the p^+ wells at the source and the n-type drain). A vertical MOSFET is used because of its low R_{DSon}.

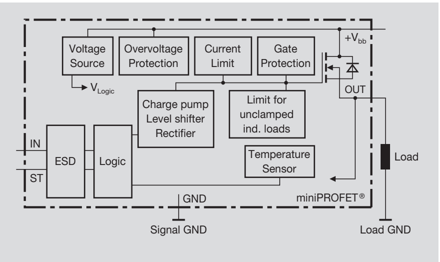


Figure 6.12 PROFET - Block Diagram

Surrounding the vertical power MOSFET (the switch) are the various functional and protection blocks implemented in logic circuitry (illustrated in Figure 6.12). In addition, ESD protection on the logic input (IN) and integrated diagnostic feedback (ST) pins are provided.

Infineon Technologies Corporation also manufactures a line of protected low side drivers called HITFETs (**Hi**ghly **I**ntegrated **T**emperature protected **F**ield **E**ffect **T**ransistors). A HITFET is a standard MOSFET with additional features that have been integrated to enhance its performance in low side drive applications.

Several protection features are included in the HITFET. For example, the short circuit protection can protect the device if a short to $V_{BATTERY}$ occurs. The over voltage protection ensures the HITFET will not be damaged during load dump transients or other high voltage events (again within energy limitations). The current limitation feature limits e.g. the maximum current which the HITFET low side switch can source. This again helps to protect the switch from short to battery conditions. The over temperature protection is enabled if the temperature of the device reaches unsafe levels, turning off the switch and stopping the current flow completely.

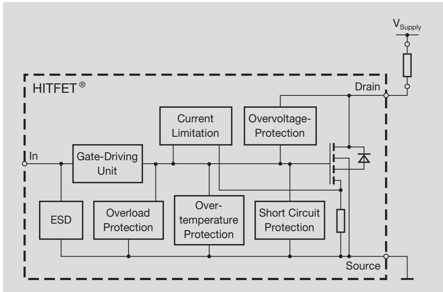


Figure 6.13 HITFET - Block Diagram

6.4 Selecting the Correct R_{DSon} Static Operation

Next, we will look at how to select the correct MOSFET R_{DSon} for power applications.

The determination of R_{DSon} is a function of electrical parameters (such as load current) and also system parameters (including ambient temperature and available heatsinking).

Here is a short review of the power equations used in thermal calculations for static operating conditions. (The reader is reminded that Chapters 4 and 5 provide more detailed description of the thermal calculations.)

- Power Dissipation (switch applications in linear region)
  $P_D = I^2 R_{DSon}$
- Thermal Impedance
  $Z_{thJA} = Z_{thJC} + Z_{thCA}$
- Junction Temperature
  $T_{junction} = T_{ambient} + P_D Z_{thJA}$
- For static operation $Z_{thJA} = R_{thJA}$

Figure 6.14 Basic Power Equations

The power dissipated in a resistive switch is the product of the on-state resistance of the switch and the square of the current flowing through the switch:

$$P_D = I^2 R$$

$$P_{SWITCH,MAX} = (I_{load,max})^2(\max R_{DSon})$$

Next, the complete thermal impedance of a system (Z_{thJA}) is the sum of the thermal

impedance of the device (junction-to-case or Z_{thJC}) and the thermal impedance of the heatsink (case-to-ambient or Z_{thCA}).

$$Z_{thJA} = Z_{thJC} + Z_{thCA}$$

Once the dissipated power in the switch and its thermal impedance are known, the junction temperature of the switch can be determined from Ohm's Law applied to the thermal domain.

The junction temperature must be limited so that it never exceeds the maximum operating temperature specified in the datasheet (usually 150 °C max).

$$T_{junction} = T_{ambient} + P_D Z_{thJA}$$

Finally, for steady state operation, the thermal impedance is equal to the thermal resistance:

$$Z_{thJA} = R_{thJA}$$

If the equations are rearranged, solutions can be obtained for either the maximum load current (for a given R_{DSon}), or the maximum allowed R_{DSon} (for a given load current):

- Rearranging, the equations yield:

$$I_{Load} = \sqrt{\frac{T_{junction} - T_{ambient}}{Z_{thJA} R_{DSon}}}$$

$$R_{DSon} = \frac{T_{junction} - T_{ambient}}{I_{Load}^2 Z_{thJA}}$$

Figure 6.15 R_{DSon} Equations

Many parameters affect the R_{DSon} selection:

- Typically, the following parameters are set by the device:
  - $T_{junction,max}$ - Usually 150 °C
  - R_{DSon} - Function of the silicon die and package
  - Z_{thJC} - Function of the package type (and die size)
- Typically, the following parameters are set by the application:
  - $T_{ambient}$ - Usually 85 °C, 105 °C, or 125 °C
  - I_{Load} - Function of the load resistance
  - Z_{thCA} - Function of the external heatsink

Figure 6.16 Parameters Affecting R_{DSon} Selection

The first group of three entries are set by the choice of the MOSFET switch:

1) Maximum Junction Temperature. For most power devices, this is 150 °C.

2) MOSFET Switch Resistance from Drain to Source (R_{DSon}) when the switch is ON. This is a function of the area of the silicon die, the process used to fabricate the MOSFET, and, to a lesser extent, the type of package.

3) Thermal Impedance from junction to the Package Heatsink. This is primarily a function of the package type. However, there is also a dependence upon the size of the silicon die inside the package. A larger silicon die has a greater surface area to dissipate power resulting in a slightly lower Z_{thJC}.

Maximum Ratings at T_j = 25 °C unless otherwise specified

Parameter	Symbol	Values	Unit
Operating temperature range	T_j	-40 ... +150	°C

Thermal Characteristics

Parameter and Conditions	Symbol	Values			Unit
		Min.	Typ.	Max.	
Thermal resistance chip - case:	R_{thJC}	–	–	1.1	K/W
junction - ambient (free air):	R_{thJA}	–	80	–	
SMD version, device on PCB [6]		–	45	55	

6) Device on 50 mm * 50 mm * 1.5 mm epoxy PCB FR4 with 6 cm^2 (one layer, 70 µm thick) copper area for V_{bb} connection. PCB is vertical without blown air.

Load Switching Capabilities and Characteristics

Parameter and Conditions	Symbol	Values			Unit
		Min.	Typ.	Max.	
On-state resistance (pin 3 to pin 1, 5)					
V_{bb} = 5.5 V, I_L = 7.5 A, T_j = 25 °C	R_{ON}	–	10	14	mΩ
T_j = 150 °C		–	18	24	
V_{bb} = 12 V, I_L = 7.5 A, T_j = 25 °C		–	8	10	
T_j = 150 °C		–	14	18	

Figure 6.17 Datasheet Parameters Affecting R_{DSon} Selection

The first three parameters ($T_{junction,max}$, R_{DSon} and Z_{thJC}) can be found in the datasheet (see Figure 6.17).

First, we look at the maximum operating junction temperature. In this example, the MOSFET driver is specified to operate normally as long as the junction temperature remains at or below 150 °C. This is usually

found in the maximum ratings section of the datasheet.

The second parameter to be found in the datasheet is the thermal resistance (recall that for steady state operations the thermal resistance is equal to thermal impedance). In the example, the junction-to-case (or chip-to-case) thermal resistance is listed as 1.1 K/W (see the thermal characteristics portion of the datasheet). In this example, the units are "degrees Kelvin per Watt" or "K/W". Since 1 K temperature change is equivalent to a 1 °C temperature change, the units "K/W" and "°C/W" are interchangeable when applied to temperature changes.

Note that the datasheet also identifies the thermal resistance under specific system conditions. First, if the MOSFET driver were in free air (a worst case condition), the thermal resistance junction-to-ambient is typically 80 K/W (or 80 °C/W). Next, if the high side driver were mounted on a PCB with 6 cm^2 (about 1 in^2) of copper board area for heatsinking, the typical thermal resistance junction-to-ambient is lowered to 45 K/W (or 45 °C/W).

The final parameter in the datasheet is the on state resistance of the switch between the MOSFET drain and source. In this example, the R_{DSon} is specified for four different conditions. Note that the extreme case conditions (low battery voltage and high temperature) has the worst case on-state resistance (24 mΩ).

Knowing where to find these parameters in a datasheet, let us look at a few sample calculations.

In our first example, it is assumed that the maximum ambient temperature of the electronic module is 85 °C with a battery voltage of 14 V.

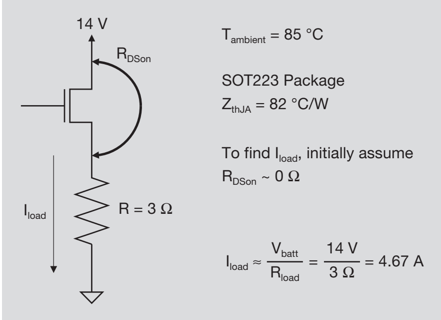


Figure 6.18 R_{DSon} Selection Example Calculation

In addition, a SOT223 package is selected which has a thermal resistance (junction-to-ambient) of 82 °C/W. (We will look at this number more closely later in this chapter.)

For the load current calculation, we assume that the $R_{DSon} << R_{load}$ = 3 Ω and thus R_{DSon} ~ 0 Ω; then the load current calculated by Ohm's Law is 4.67 A.

From the load current, we can determine the maximum allowed R_{DSon}.

We know from the datasheet that the maximum junction temperature is 150 °C. This limits the maximum R_{DSon} to 36 mΩ (significantly smaller than our 3 Ω load).

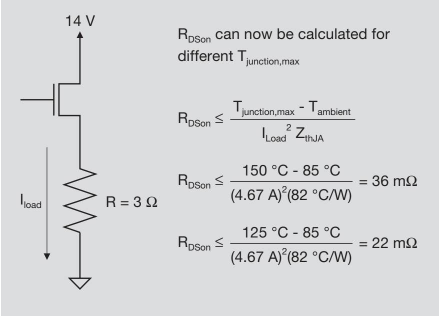


Figure 6.19 R_{DSon} Selection Example Calculation

Let us reflect on what this means. Given an ambient temperature of 85 °C, if a MOSFET driver is used with R_{DSon} = 36 mΩ to switch 4.67 A current in a system with a thermal

resistance of 82 °C/W, the junction temperature will rise to 150 °C. A lower value of R_{DSon} (a more expensive device) will result in a lower junction temperature. This is often done for safety and for providing headroom in a design.

For example, if the maximum desired junction temperature were only 125 °C, the MOSFET would have to have a lower R_{DSon} (22 mΩ). This will require a larger die and a more expensive device.

Where did the 82 °C/W value of thermal resistance come from? In a previous chapter, a fair amount of material was allotted to thermal resistance graphs. The thermal resistance graph for the SOT223 package is shown in Figure 6.20. The graph provides the junction-to-case thermal resistance (R_{thJA}) for a SOT223 package as the copper board space heatsink is varied from 0 mm^2 to 600 mm^2 (6 cm^2 or about 1 in^2).

We arbitrarily chose, in the example, a value of 82 K/W (82 °C/W) corresponding to a SOT223 package mounted on a PCB with about 3 cm^2 (about 0.5 in^2) of board space dedicated to copper heatsinking.

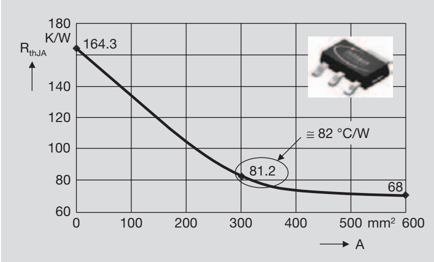


Figure 6.20 SOT223 Heatsink

We can also consider cases with other packages and heatsinks which may have better (or worse) junction-to-ambient thermal resistances. One example is the TO263 package (commonly referred to as D^2-PAK).

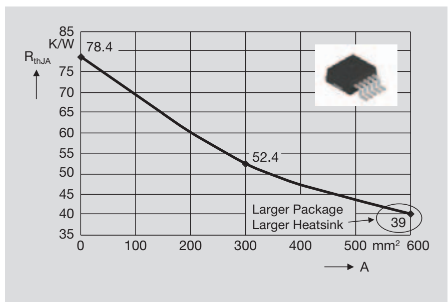


Figure 6.21 TO263 Heatsink

From the graph it is clear that the thermal resistance of the TO263 package is less than that of the SOT223 package for a given copper PCB heatsink area.

For example, with 600 mm^2 of copper board space heatsink, the thermal resistance (junction-to-ambient) of the TO263 package drops to 39 °C/W.

How this can impact the R_{DSon} calculation in the previous example.

Recalling: $T_{ambient}$ = 85 °C
I_{load} = 4.67 A

For a maximum junction temperature of 150 °C, the maximum allowed R_{DSon} = 76 mΩ.

Selecting some safety margin with a 125 °C limit, the new maximum R_{DSon} = 47 mΩ.

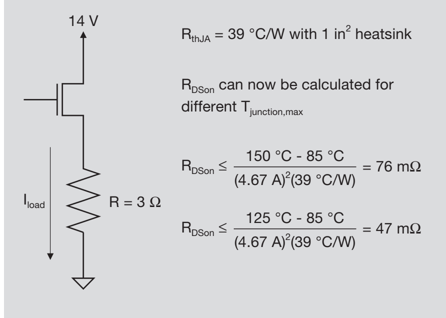


Figure 6.22 R_{DSon} Selection Example Calculation

Note that with the use of larger packages (generally more expensive), the maximum

R_{DSon} becomes larger (less expensive MOSFET silicon die).

As you use larger heatsinks (more expensive board design), the maximum R_{DSon} becomes larger (less expensive MOSFET silicon die).

As you use a narrower safety margin (higher operating temperature), the maximum R_{DSon} becomes larger (less expensive MOSFET silicon die).

Package and Heatsink	R_{DSon} at $T_{junction,max}$ = 125 °C	R_{DSon} at $T_{junction,max}$ = 150 °C
SOT223 (0.5 in²)	22 mΩ	36 mΩ
TO263 (1 in²)	47 mΩ	76 mΩ

Figure 6.23 R_{DSon} vs. Package and Heatsink

In our simple examples, the range of maximum allowable R_{DSon} is

22 mΩ < R_{DSon} < 76 mΩ, or 3.46 ratio of upper limit to lower limit!

This gives an idea of the performance/cost trade-offs that are inherent in the selection of the MOSFET's R_{DSon}.

For reference, the thermal resistance curves for several different packages are included.

As mentioned before, the thermal resistance generally decreases as the package (case) size increases:

The SO-8 package has a higher thermal resistance than the…

SOT223 package which has a higher thermal resistance than the…

TO252 package which has a higher thermal resistance than the…

TO263 package.

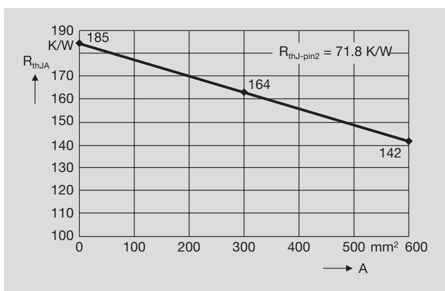


Figure 6.24 R_{thJA} for SO-8 Package

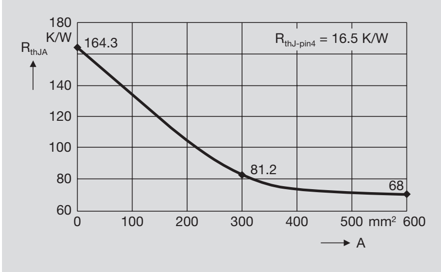


Figure 6.25 R_{thJA} for SOT223 Package

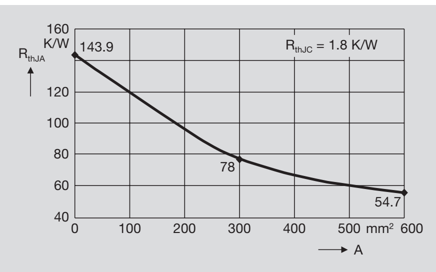


Figure 6.26 R_{thJA} for TO252 Package

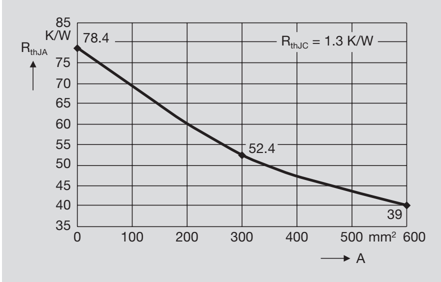


Figure 6.27 R_{thJA} for TO263 Package

6.5 Selecting the Correct R_{DSon} Dynamic Operation and the Impact of Switching Losses

The switching losses that occur in a FET driver switch are examined next when the MOSFET is turned on, turned off, or repeatedly toggled on and off (also known as Pulse Width Modulation or PWM).

Any time the operating state of a semiconductor switch is changed, the turn on and off processes occur within a finite time. The voltage across the MOSFET and the current through the MOSFET do not instantaneously change: it takes time to fully turn on or off the transistor.

During the time of a turn-on transition, switching losses occur as the current through the MOSFET increases causing a voltage drop across the load that decreases the voltage across the MOSFET. During the transition the transient power dissipation in the switch will reach a maximum value.

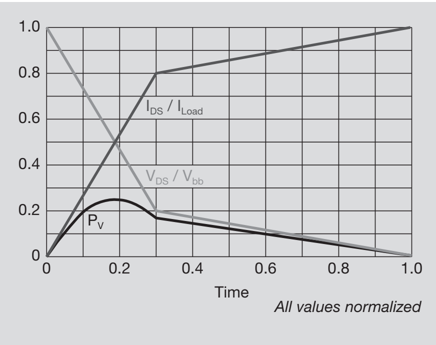


Figure 6.28 Impact of Approximate FET Switching Loss

A real life example of a switching application is a 6 mΩ high side driver used to turn on a 60 W light bulb load (analogous to a head lamp bulb).

The load current through the high side driver and the bulb is shown here. It takes about 200 µs for the high side driver to completely turn on (and for current to reach its maximum value) as shown by Figure 6.29.

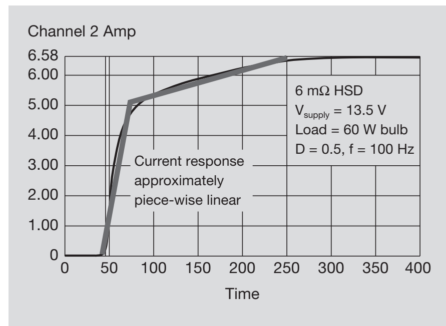


Figure 6.29 PROFET Switching Loss Lamp Turn-On

The current response for this high side driver (an Infineon PROFET device) is approximately piece wise linear. Initially, the current increases rapidly. After about 40 µs, and for the next 160 µs, the rate of current rise decreases. During these two periods the current is approximately linear functions of time. This fact will come into play in the equations to be introduced shortly. The turn-on/off characteristics are determined by the gate drive circuitry contained within the hi-side driver.

A warm light bulb can be considered as a resistive load (more about cold light bulbs momentarily).

If the current through the MOSFET and the voltage across the MOSFET are piece wise linear, then an approximation for switching losses can be made as:

V_{DS} = Voltage transition across the MOSFET

I_{DS} = Current transition through the MOSFET

$P_{SWITCH\ LOSS,AVE} \cong (1/8)(V_{DS})(I_{DS})$

P_{LOSS} = Power Loss During Turn On

t_{SWITCH} = Time to Turn MOSFET On

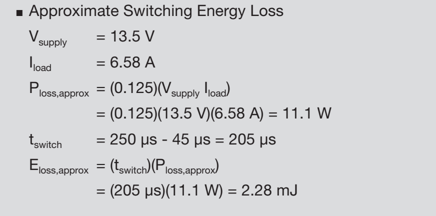
$E_{SWITCH\ LOSS} \cong (P_{SWITCH\ LOSS,AVE})(t_{SWITCH})$

- For a resistive load (with a piecewise linear current and voltage response), the approximate FET switching loss is:

$P_{loss} \sim (0.125)(V_{DS}\, I_{DS})$
$E_{loss} = (P_{loss})(t_{switch})$

Figure 6.30 Approximate FET Switching Loss

These equations are for a single "turn-on" event. Every time the MOSFET is turned on, it will incur these losses. The equations are also useful to calculate the switching losses for our 60 W lamp load example (assuming a warm bulb). The approximate average power lost in the switch is 11.1 W (or 11.1 Joules per second). Since the approximate switching time is 205 µs, the energy loss is 2.28 mJ every time the switch is turned on.



- Approximate Switching Energy Loss

V_{supply} = 13.5 V
I_{load} = 6.58 A
$P_{loss,approx}$ = $(0.125)(V_{supply}\, I_{load})$
= (0.125)(13.5 V)(6.58 A) = 11.1 W
t_{switch} = 250 µs - 45 µs = 205 µs
$E_{loss,approx}$ = $(t_{switch})(P_{loss,approx})$
= (205 µs)(11.1 W) = 2.28 mJ

Figure 6.31 PROFET Switching Loss: Lamp Turn-On

When the current through the MOSFET and bulb is turned off the current as a function of time is shown in Figure 6.32. The current change in this case is approximately linear (in the turn-on case the current change was piece wise linear).

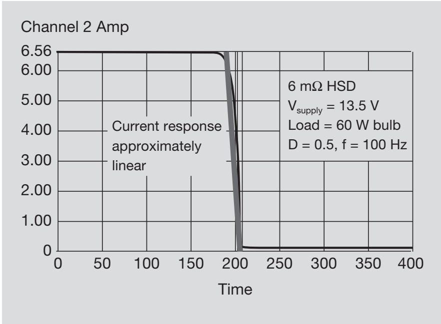
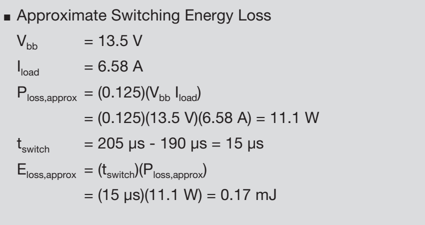


Figure 6.32 PROFET Switching Loss: Lamp Turn-Off

Using the same equations, the switching losses in the MOSFET can be calculated for the turn-off event.



- Approximate Switching Energy Loss

V_{bb} = 13.5 V
I_{load} = 6.58 A
$P_{loss,approx}$ = $(0.125)(V_{bb}\, I_{load})$
= (0.125)(13.5 V)(6.58 A) = 11.1 W
t_{switch} = 205 µs - 190 µs = 15 µs
$E_{loss,approx}$ = $(t_{switch})(P_{loss,approx})$
= (15 µs)(11.1 W) = 0.17 mJ

Figure 6.33 PROFET Switching Loss: Lamp Turn-Off

Notice that the average power loss in the switch during turn off is the same as the power loss during turn on. However, since the switching duration is significantly shorter (15 µs vs. 205 µs), much less energy is dissipated in the MOSFET when the switch is turned off.

The sum of the switching energy loss of the MOSFET during a turn-on/turn-off switching event is approximately 2.45 mJ.

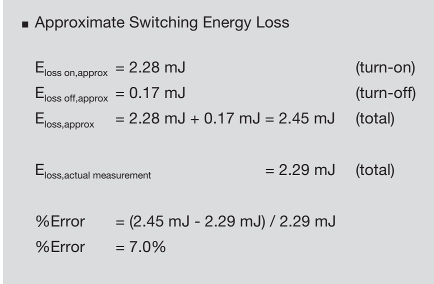


Figure 6.34 PROFET Switching Loss: Lamp Turning On and Off

The result is only 7% higher than the actual energy loss of the system (measured by an oscilloscope). It shows that for quick calculations, the “One Eighth Rule” is a reasonable approximation for piece wise linear V(t) and I(t) functions:

$$P_{LOSS} \cong (1/8)(V_{DS})(I_{DS})$$

$$E_{LOSS} \cong (P_{LOSS})(t_{SWITCH})$$

So, now we can look at the total power dissipation in a MOSFET when it is being used in a PWM application (rapidly being turned on and off).

- The power dissipated in a PWM application is given by:

$P_D = P_{switching} + P_{on}$

$P_{switching} = (F_{switching})(P_{loss\text{-}on}\, t_{turn\text{-}on} + P_{loss\text{-}off}\, t_{turn\text{-}off})$

$P_{loss\text{-}off} \cong (0.125)(V_{supply}\, I_{load})$

$P_{loss\text{-}on} \cong (0.125)(V_{supply}\, I_{load})$

$P_{on} = (I_{load})^2(R_{DSon})(t_{pulse\text{-}on})(F_{switching})$

$T_{junction} = T_{ambient} + P_D\, R_{thJA}$

$D = (t_{pulse\text{-}on})(F_{switching}) = (t_{pulse\text{-}on}) / (T_{Period})$

Figure 6.35 R_{DSon} Calculations for PWM Applications

You do not need to memorize these equations, they are here for your reference. With time and practice, however, you will become more familiar with them.

In a PWM application both switching losses and conductive (on) losses occur:

$$P_D = P_{switching} + P_{on}$$

Here, the total switching power loss is given by the product of the switching frequency and the sum of the energy loss during a turn on and a turn off cycle:

$$P_{switching} = (F_{SW})(P_{loss\text{-}on} * t_{turn\text{-}on} + P_{loss\text{-}off} * t_{turn\text{-}off})$$

The switching power loss equations are the “One Eighth Rule” we previously introduced. The power loss when the MOSFET is on is determined by the product of the power and the Duty Cycle (the product of the time that the MOSFET is on and the pulse repetition rate):

$$P_{on} = (I^2 R)(\text{Duty Cycle})$$

$$P_{on} = (I^2 R)(t_{pulse\text{-}on} * F_{SW})$$

In Figure 6.36 is the equation for the maximum allowed on state resistance of the MOSFET accounting for both switching and conductive losses. There are two portions on the right hand side of the equation. The first term is determined by the static operating losses (I_{load}^2 D), the second term is based upon the switching losses ($1/8VIF_{swit}$).

Each term incorporates the appropriate power and energy loss.

Again, this is not an equation you need to memorize (engineers often just enter this equation into a spreadsheet). In fact, rather than going through all the math, we thought it would be more informative to simply look at how the different parameters affect the R_{DSon} selection.

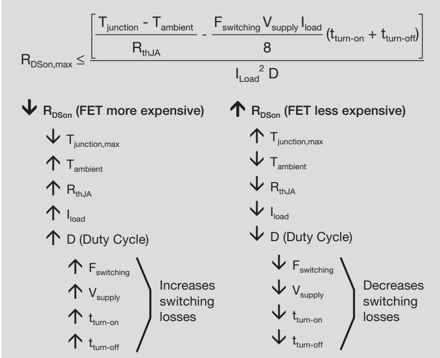


Figure 6.36 R_{DSon} Calculations for PWM Applications

The following is a quick example of how the equation can be used to determine the maximum allowable R_{DSon} of a MOSFET.

The high side switch is being pulse width modulated (PWM operating mode) driving a 2.05 Ω load. The steady state load current is:

$$I_{load} = V_{supply} / R_{load}$$
$$= 13.5\ V / 2.05\ \Omega = 6.57\ A$$

The switching parameters (P_{loss}, $t_{turn\text{-}on}$, $t_{turn\text{-}off}$, $F_{switching}$, Duty Cycle) are from our previous examples.

The package of choice is a TO252 package with 1 in^2 of copper board space for heat sinking.

This allows the use of max. R_{DSon} = 45 mΩ.

Using the same equations, the switching losses in the MOSFET can be calculated.

So, now we can look at the total power dissipation in a MOSFET when it is being used in a PWM application.

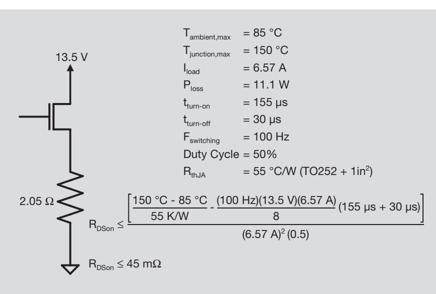


Figure 6.37 R_{DSon} Selection Example Calculation

6.6 Capacitive Load In-Rush Current

The final two sections in this module examine the operation of MOSFET drivers in applications where the load is not purely resistive. These loads generally fall into two categories, capacitive loads and inductive loads. First we will look at capacitive loads and the precautions that need to be taken when switching a capacitive load.

The model shown here is for a light bulb. The RC network will exhibit a large in-rush current when it is initially turned on. This model is also in Chapter 1 (RLC Loads) where the derivation of the component values is shown in detail.

- Lamps and RC networks can experience significant "in-rush" current when they are initially turned on
- When a lamp initially turns on, the filament is cold, and has a relatively low resistance
- As the filament warms up, the resistance increases dramatically (often by an order of magnitude)

23.2 Ω
In
3.6 mF
Out
2.80 Ω

Figure 6.38 Capacitive Load In-Rush Current

The model represents the resistance of the warm filament (23.2 Ω) and the impedance of the cold filament (2.8 Ω in series with the 3.6 mF capacitor). The charging of the capacitor models the warming up of the filament.

The in-rush current vs. time function for a light bulb load is shown in Figure 6.39. The waveform is an actual oscilloscope trace of the current flowing through an initially cold light bulb. The time-base in the waveform is 40 ms/division.

The current peaks very quickly at approximately 5.5 A. As the filament warms (for about 80 ms) its resistance increases, and the load current reaches its steady state value of approximately 0.6 A.

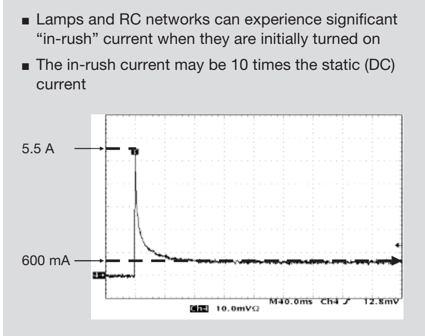


Figure 6.39 Capacitive Load In-Rush Current

When the load resistance is lower than expected and the in-rush current exceeds the maximum current limit, PROFETs and HITFETs may go into a protective current limiting mode. When turning on capacitive loads, like lamps, drivers should be selected such that their current limit value is greater than the peak in-rush current.

The value of the high side driver's current limit can be found in the datasheet. The device shown below limits the maximum current at 70 A (minimum at $T_{junction}$ = 150 °C) to 140 A (maximum at $T_{junction}$ = -40 °C). It is generally NOT recommended to force a protected driver into current limit operation during switching of the "normal/nominal" application load.

Protected drivers are not designed for extended/continuous operation under fault (extreme operating) conditions (such as current limiting). During current limiting, the driver will transition from the linear region of operation to the saturation region and begin to act as a current source, effectively clamping (or limiting) the sourced current to a predetermined value.

In the saturation region, however, the power dissipated (and hence junction temperature) in the protected driver can increase dramatically due to the increasing voltage (V_{DS}) across the MOSFET. Therefore, care must be taken to keep in-rush current levels below the device's current limit threshold.

Parameter and Conditions at T_j = 25 °C, V_{bb} = 12 V unless otherwise specified	Symbol	Values			Unit
		Min.	Typ.	Max.	
Protection Functions [12]					
Short circuit current limit (Tab to pin 1, 5) [13]					
Short circuit current limit at V_{ON} = 6 V [14] T_j = -40 °C	$I_{L6(SC)}$	–	110	140	A
T_j = 25 °C		–	105	–	
T_j = +150 °C		70	90	–	

Figure 6.40 Standard Current Limiting

Here is an example of the how the in-rush current of a load can impact the junction temperature of a high side driver.

In this example, the voltage across the driver (V_{DS}) rapidly decreases. During the time the lamp current rapidly increases to about 27 A and then begins to decrease as the lamp filament heats up. These are actual scope traces of a 50 mΩ hi-side driver turning on a 27 W incandescent bulb (Figure 6.41).

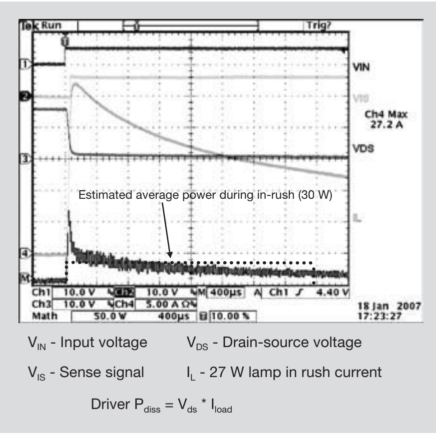


Figure 6.41 Lamp In-Rush Current Example

The actual in-rush power is also shown as a scope trace by multiplying the drain-source voltage and load current. In order to be able to use a Z_{th} diagram to help estimate the transient temperature rise during in-rush, the average power during in-rush must be modeled by a rectangular power pulse. For this example we will estimate the average in-rush power as a rectangular pulse of 30 W lasting for approximately 3 msec. The time duration of 3 msec is somewhat arbitrarily chosen but

here it is used as the point where the in-rush current has decayed to half of its peak value.

Now that the in-rush power is modeled as a rectangular pulse it is valid to use the Z_{th} diagram of the driver in order to estimate the transient temperature rise during current in-rush. Since we are looking at a single in-rush current event, we will look at the single pulse line (Figure 6.42).

For a pulse width of 3 ms, the thermal impedance of our high side driver is ~2.0 °C/W.

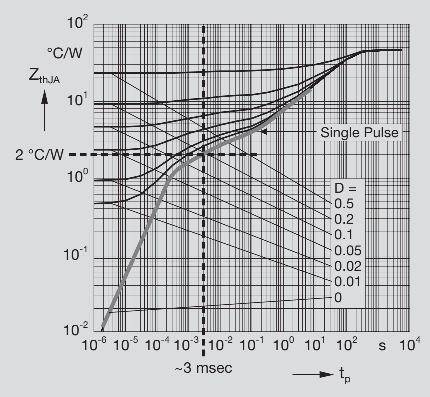


Figure 6.42 Z_{thJA} Chart for Lamp In-Rush Current Example

Now, we can calculate the junction temperature increase of the high side driver during the in-rush current event.

- Approximate junction temperature increase (using Z_{th} diagram and estimated rectangular average in-rush power)

$t_{in\text{-}rush}$ ≅ 3 msec

Z_{thJA} ≅ 2.0 °C/W

$P_{loss,ave}$ ≅ 30 W (estimated from oscilloscope)

$\Delta T_{junction} = Z_{thJA} P_{loss}$
= (2.0 °C/W)(30 W) = 60 °C

Figure 6.43 Lamp In-Rush Current Example Calculations

$T_{junction} = T_{ambient} + Z_{thJA} * P_{loss}$

$T_{junction} - T_{ambient} = Z_{thJA} * P_{loss}$

$\Delta T_{junction} = Z_{thJA} * P_{loss}$
$= 2.0$ °C/W * 30 W = 60 °C

Increasing the transient junction temperature by 60 °C during the in-rush current event is normally considered a somewhat marginal design for long term reliability of the MOSFET driver.

Studies for FET drivers with aluminum metallization have shown that transient temperature increases in the range of approximately 60 °C or higher can cause degradation of the metallization and affect the long term reliability of the device. It is generally recommended to keep the transient junction temperature increase to less than 60 °C during in-rush events from capacitive loads.

6.7 Switching Off an Inductive Load

Finally, we will look at inductive loads and the precautions that need to be taken when turning off an inductive load.

First let's take quick review of Lenz's Law:

- With inductive loads (for example coils and valves), additional switching losses can occur during turn off.
- According to Lenz's Law:
  The electromotive force (voltage) and the induced current in an inductor are in a direction as to tend to oppose the change that produced them.
- Therefore at turn off, the voltage at the output of the high side driver becomes negative to oppose the decreasing inductor current.

Figure 6.44 Switching OFF an Inductive Load

Briefly, it means that the current through an inductor can not be changed instantaneously. To turn off an inductive load, the current must decay over some time to 0 A.

Instantaneous change of current through the inductor would require **very** large voltage transients. The voltage across the inductor will tend to reverse the polarity of voltage that has created the current initially.

This may cause stresses when driving inductors with a high side or low side driver.

Here's an example of what will happen to a protected MOSFET when an inductive load is switched off.

Initially, the high side driver is turned on by increasing the input voltage.

Note that in a protected MOSFET, the input will not actually be tied to the MOSFET gate. Rather, it signals the internal circuitry to apply the gate voltage.

The output voltage quickly reaches its steady state value and the load current increases through the inductor (Figure 6.45).

After some time the load current in the inductor reaches its final value (due to its resistance).

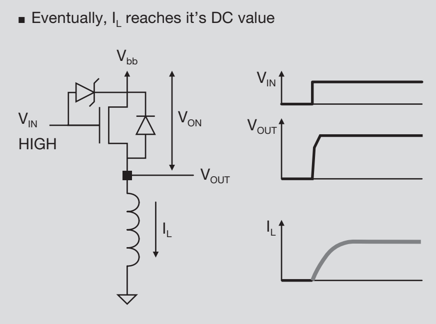


Figure 6.45 Switching OFF an Inductive Load

Next, the switch is turned off by lowering the input voltage.

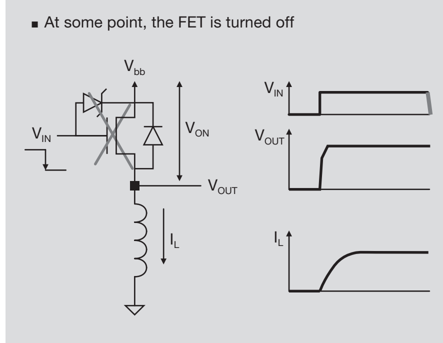


Figure 6.46 Switching OFF an Inductive Load

At this time, the switch turns off, and the MOSFET attempts to stop sourcing current. According to Lenz's Law, however, the current flowing through an inductor cannot be instantaneously changed. The inductor will discharge in some finite time in response to a reversed polarity voltage pulse acting across it.

As the load current is being reduced, (V_{OUT}) will become negative (as the polarity across the inductor flips). For most protected FETs, there are integrated clamping circuits that will only allow the output voltage to go so far negative and then the clamping circuit will turn the FET back-on to essentially clamp the output voltage at a designed level (Figure 6.47).

If the high side driver did not clamp the output voltage, V_{OUT} would continue to fall below ground until it reached the avalanche breakdown of the MOSFET. At this point, an avalanche current would flow from the battery through the MOSFET to the inductor until the inductor current decayed to 0 A. This avalanche process itself is not destructive although it may cause excessive power dissipation, which can permanently damage the high side driver (more on this in a few pages).

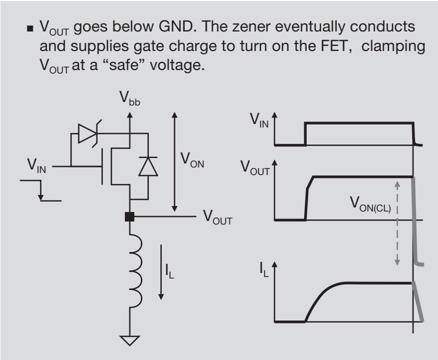


Figure 6.47 Switching OFF an Inductive Load

Once the current discharges to 0 A and all the inductive energy is dissipated, V_{OUT} will go back to 0 V and the turn-off process of the inductive load is complete.

Notice that substantial power will be dissipated in the MOSFET during the turn-off of the inductive load current.

$P_D = V_{ON(CL)} * I_L$

$V_{ON(CL)} = V_{bb} - V_{OUT}$ during clamping

We need to ensure the power dissipated during the turn-off process is limited to a safe value (Figure 6.48).

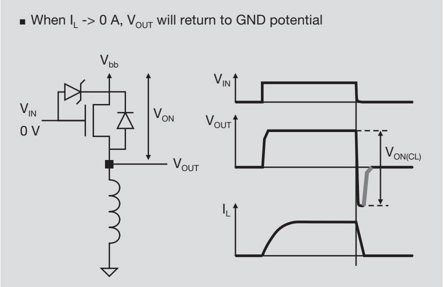


Figure 6.48 Switching OFF an Inductive Load

As stated before, if the diode clamp was not integrated in the protected MOSFET, V_{OUT} would continue to fall below ground until the MOSFET went into avalanche breakdown. This is shown in Figure 6.49.

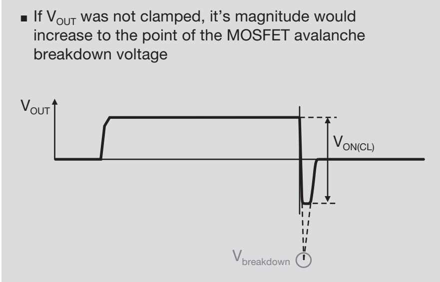


Figure 6.49 Safely Clamping V_{OUT} for Inductive Loads

A MOSFET which turns itself on via a feedback clamping circuit can normally absorb more energy than one which undergoes avalanche breakdown. **Recall that the energy stored in the inductor is $E = 1/2 * L * I_p^2$. This energy must be dissipated when the FET is turned off. Also recall that the power dissipated in the FET is $P_{FET} = V_{DS} * I_D$. By turning the FET on, $V_{DS} = V_{ON(CL)}$, which is designed to be < $V_{DS(AV)}$. The power dissipated in the device during the inductive switch-off event is therefore reduced.**

Thus, a FET operated in the on-state can typically handle more energy than when operated in the avalanche mode, and the protected FET can handle more energy during inductive turn off compared to a standard FET operated in avalanche mode (Figure 6.50).

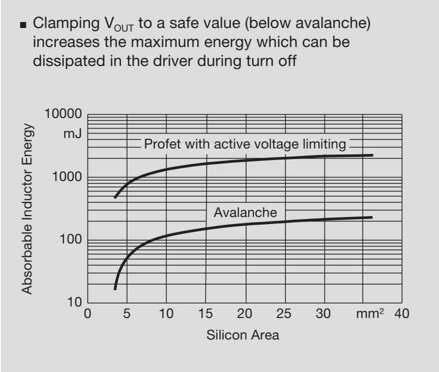


Figure 6.50 Clamping V_{OUT} Increases the Maximum Inductor Energy

In our high side driver datasheets, you will see two important specifications for turning off inductive loads.

At the top of Figure 6.51, in the Maximum Ratings section, an example is given of the maximum amount of energy the high side driver can safely dissipate during a worst case load condition. In this example, a high side driver with a starting junction temperature of 150 °C sourcing 20 A of current can safely absorb an additional 0.3 J of energy during the turn-off of an inductive load.

The bottom half of the figure states the specification of the clamping voltage that was introduced a few pages back. The clamping voltage is equal to the supply voltage minus the negative output inductor voltage at which the MOSFET has turned on and 40 mA of current is flowing:

$V_{ON(CL)} = V_{SUPPLY} - V_{OUT}$
$= V_{DRAIN} - V_{SOURCE}$
$= V_{DS}$ at I_{DS} = 40 mA (test condition)

In this particular device, the clamping voltage is typically 42 V. This means for a 14 V battery voltage, V_{OUT} will typically clamp at -28 V.

- The maximum safe inductive energy which can be dissipated in the FET is found in the maximum ratings section:

Maximum Ratings at T_j = 25 °C unless otherwise specified

Parameter	Symbol	Values	Unit
Inductive load switch-off energy dissipation [3] single pulse I_L = 20 A, V_{bb} = 12 V T_j = 150 °C	E_{AS}	0.3	J

- The clamping voltage is in the electrical characteristics:

Parameter and Conditions at T_j = 25 °C, V_{bb} = 12 V unless otherwise specified	Symbol	Values			Unit
		Min.	Typ.	Max.	
Protection Functions [12]					
Output clamp (inductive load switch off) [15] at $V_{OUT} = V_{bb} - V_{ON(CL)}$ (e.g. overvoltage) I_L = 40 mA	$V_{ON(CL)}$	39	42	–	V

Figure 6.51 Maximum Safe Inductor Energy

In addition to the specified electrical values the datasheets also include a graph illustrating the maximum load inductance for a typical high side driver device (Figure 6.52).

First notice that the graph has been developed for the worst case conditions ($T_{junction}$ = 150 °C and 0 Ω resistance of the inductor). The graph specifies the largest inductor value a typical high side driver can safely turn off based upon the steady state operating current.

Please note that this graph is for a **single turn-off** event of the inductor, and should not be used for repetitive or PWM applications. We will address PWM driving of an inductive load momentarily.

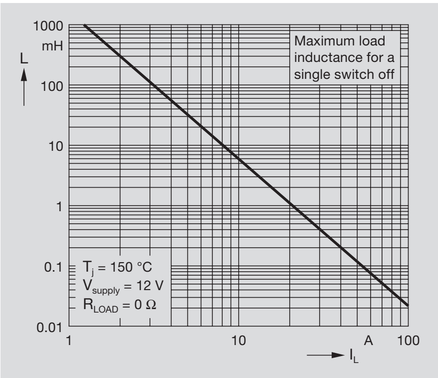


Figure 6.52 Energy Absorbed When Turning Off an Inductive Load

As stated before, when the MOSFET turns off an inductive load, energy will be absorbed by the MOSFET. Some of the energy stored in the inductor is, however, dissipated in form of self heating within the parasitic resistance of the inductor and is not absorbed by the MOSFET.

There is additional energy dissipated in the driver during inductive load turn-off due to the current flowing from the battery to the MOSFET (Figure 6.53).

- The energy absorbed by the high side driver when an inductive load is turned off (E_{loss}) is equal to:

$E_{loss} = E_{SUPPLY} + E_L - E_R$

Where:

E_{supply} is the energy delivered to the MOSFET from the battery

E_L is the energy delivered to the MOSFET from the inductance ($E_L \sim L\, I_L^2 / 2$)

E_R is the energy dissipated by the inductor due to internal self-heating

Figure 6.53 Energy Absorbed When Turning Off an Inductive Load

The total energy absorbed by the protected FET can be found by solving a differential equation. The solution can be approximated as shown in Figure 6.54. It is important to verify with inductive loads whether the E_{loss} value is less than the specified maximum absorbable energy, $E_{AS,MAX}$ (0.3 J in the previous datasheet example).

- This becomes a differential equation:

$$E_{loss} = E_{SUPPLY} + E_{INDUCTANCE} - E_{ESR} = \int V_{ON(CL)} * i_L(t)\, dt$$

- The solution to this equation can be approximated for $R_L > 0\ \Omega$

$$E_{loss} = \frac{L\, I_L}{2\, R_L} \left(V_{SUPPLY} + |V_{OUT(CL)}|\right) \ln\left(1 + \frac{I_L\, R_L}{|V_{OUT(CL)}|}\right)$$

Figure 6.54 Energy Absorbed When Turning Off an Inductive Load

What will happen if the dissipation is too large when inductive load currents are turned off?

It is important to remember that Electrical Over Stress (EOS) conditions can be created when the dissipated energy is in excess of the datasheet ratings.

The large amount of dissipated energy creates "hot spots" on the silicon die melting the interconnect metallization, damaging the silicon and may permanently damage the MOSFET (Figure 6.55).

- Protected FET die after the maximum dissipated energy is exceeded due to switching off an inductive load

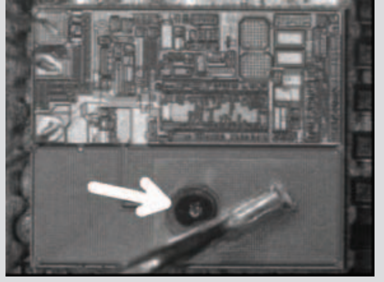


Figure 6.55 What Can Go Wrong?

Special precaution must be taken when repeatedly turning off an inductive load (as in a PWM application).

In a PWM application, every time the current to inductive load is turned off (perhaps 100 times per second or more) the amount of energy previously determined (E_{loss}) is being absorbed by the MOSFET at each cycle. For most applications this may exceed what even a protected FET can handle.

Therefore, one customary practice is to place a fast recirculation diode in parallel to the inductor. This diode clamps the inductor voltage slightly below ground and serves to recirculate the inductor's current during the turn-off phase of a PWM cycle.

Safety does not come without a price: limiting the fly-back to a single diode drop will slow the current decay and thus take longer to turn off the inductive load (Figure 6.56).

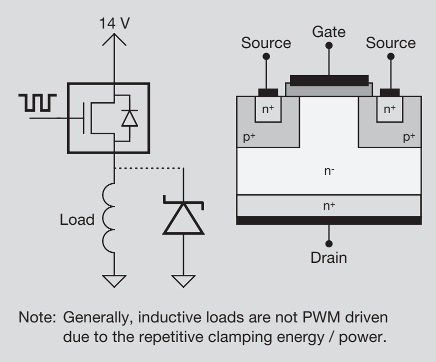


Figure 6.56 Driving a FET with a PWM Input

Finally, we will look at an example of turning off an inductive load with a high side driver and calculate the energy absorbed by the MOSFET.

Initially, the high side driver is switched off by lowering the input voltage. **Again, the input pin is not directly connected to the gate of the high side drive MOSFET** (Figure 6.57)**.**

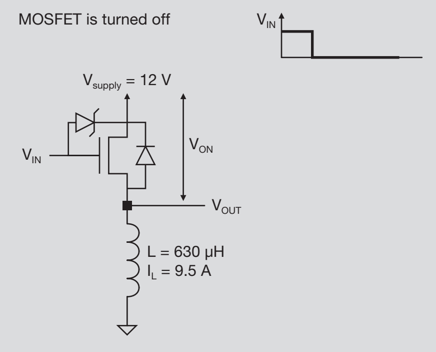


Figure 6.57 Turning Off an Inductive Load

When the MOSFET is initially turned off, V_{OUT} is clamped (Figure 6.57):

$$V_{OUT} = V_{SUPPLY} - V_{AZ} = 12\text{ V} - 42\text{ V} = -30\text{ V}$$

So, in this example we have 42 V across the MOSFET (V_{ON}) as the inductor current decays from 9.5 A to 0 A.

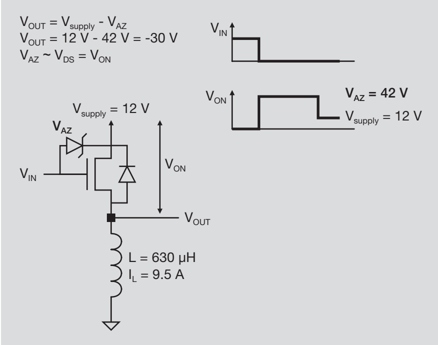


Figure 6.58 Turning Off an Inductive Load

We can calculate the time it takes for the inductor current to decay. Recall the equation for the voltage across a lossless inductor ($V_L = L\ di/dt$). We can rearrange this equation to determine the amount of time required for the inductor current to completely decay.

$$dt = L\ di / V_L$$

$$t_{stop\ decay} - t_{start\ decay} = L * (I_{L,MIN} - I_{L,MAX}) / V_{OUT}$$

$$t_{off} = L * (0\text{ A} - I_{L,MAX}) / V_{OUT}$$

$$t_{off} = - L * (I_{L,MAX}) / V_{OUT}$$
$$= -(630\ \mu\text{H}) * (9.5\text{ A}) / (-30\text{ V}) = 200\ \mu\text{s}$$

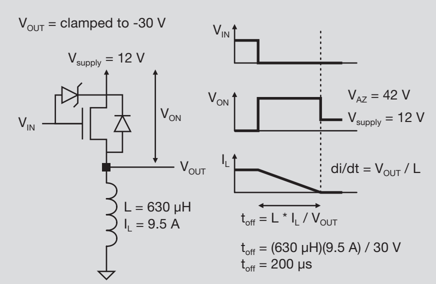


Figure 6.59 Turning Off an Inductive Load

Finally, the power loss in the high side driver is the product of the decaying inductor current (I_L) and the MOSFET drain-to-source voltage (V_{ON}) (Figure 6.60).

The power loss when the load current first begins to decay is equal to:

$$P_{loss,initial} = V_{ON,initial} * I_{L,initial}$$
$$= (42\text{ V})\ (9.5\text{ A}) = 399\text{ W}$$

The power loss when the load current has decayed to 0 A is equal to:

$$P_{loss,final} = V_{ON,final} * I_{L,final} = (42\text{ V})\ (0\text{ A}) = 0\text{ W}$$

With an average power loss of:

$$P_{loss,average} = V_{ON,average} * I_{L,average}$$
$$= (V_{ON,average}) * (I_{L,MAX} / 2)$$
$$= (42\text{ V}) * (9.5\text{ A} / 2)$$
$$= 199.5\text{ W} \sim 200\text{ W}$$

Therefore, energy absorbed by the MOSFET during the turn-off of the inductive load is:

$$E_{loss} = P_{loss} * t_{off} = (200\text{ W}) * (200\ \mu\text{s}) = 40\text{ mJ}$$

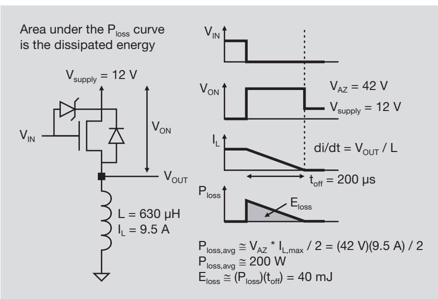


Figure 6.60 Turning Off an Inductive Load

So far, we have examined what happens when using a **high side** driver to turn off an inductive load.

We now briefly turn our attention to the problems that can occur when a low side driver is used to turn off an inductive load.

This time, however, we will see that instead of a negative voltage spike at the output, the use of a low side driver will result in a positive

voltage spike at the output when an inductive load is turned off (Figure 6.61).

- MOSFETs and HITFETs can also be used to drive inductive loads in a low side configuration
- The low side configuration, however, results in a positive voltage spike at the output

V_{SUPPLY} · Negative Voltage Spike · V_{IN} · V_{OUT} · V_{SUPPLY} · Positive Voltage Spike · V_{OUT} · V_{IN}

Figure 6.61 Low Side Drivers and Inductive Loads

Here's an example of what will happen to a protected MOSFET when a low side driver turns off an inductive load (Figure 6.61).

Initially, the low side driver is turned on by increasing the input voltage. Note that in a protected MOSFET, this input voltage will not actually be tied to the MOSFET gate. Rather, it signals the internal circuitry to apply the appropriate gate voltage.

The output voltage quickly reaches its steady state value and the load current increases through the inductor (Figure 6.62).

After some time, the load current in the inductor reaches its final value.

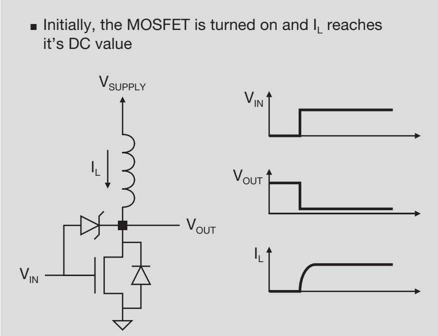


Figure 6.62 Switching an Inductive Load

Next, the switch is turned off by lowering the input voltage.

At this time the switch begins to quickly turn the current off. The reducing current will self-induce a reverse voltage across the inductor (according to Lenz's Law) that drives the output voltage to be more positive than V_{SUPPLY} (Figure 6.63).

When V_{OUT} reaches the Zener clamping voltage, the MOSFET will begin to turn back on creating a path for the reducing inductive current. V_{CLAMP} will last as long as the current is greater than 0 A (Figure 6.64).

Since $V_{OUT} = V_{CLAMP} + V_T$, V_{OUT} is held steady (i.e. clamped) at $V_{CLAMP} + V_T$,
where V_T = FET Threshold Voltage.

If the low side driver did not turn itself on, V_{OUT} would continue to increase until it reached the avalanche breakdown of the MOSFET. At this point, an avalanche current would flow from the battery through the inductor and through the MOSFET until the inductor current decayed to 0 A. This avalanche process itself is not destructive; however, it causes excessive power dissipation which may permanently damage the low side driver.

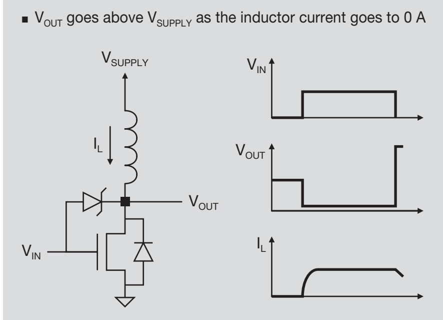


Figure 6.63 Switching an Inductive Load

Once the current discharges to 0 A and all the inductor energy is dissipated, V_{OUT} will return to V_{SUPPLY} and the inductive load turn-off process is complete.

In the case of low-side drivers a large amount of power will be dissipated in the MOSFET

during the current turn-off while driving inductive loads.

During turn-off the dissipated power is equal to:

$P_D = V_{OUT} * I_L$

and it must be ensured that the power dissipated during the turn-off process is limited to a safe value.

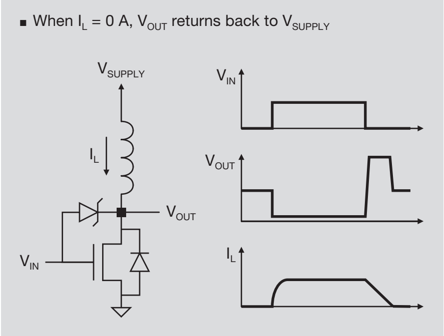


Figure 6.64 Switching an Inductive Load

7. Protected High Side Drivers

In this chapter, we will be introducing the concept of "protected" MOSFET high side drivers (PROFETs).

We will start with a quick review of MOSFETs, and quickly transition to how a protected MOSFET high side driver differs from standard transistors. This will lead us into an introduction to the various types of protection that are available in protected high side drivers. We will also describe the various diagnostic options available in protected MOSFET high side drivers.

We will then discuss how a protected MOSFET high side driver can be easily implemented in an application. We will also examine what EMI issues may arise and how they can be minimized. We will also identify how the functionality of the protected MOSFET high side driver can be improved with the addition of a few external components. Finally, we will end the chapter with a review of the questions a system designer should ask when implementing a protected MOSFET high side driver in their application.

A note to the reader:
the power transistor in all high side drivers is an n-channel MOSFET. The figures do not show the arrows that usually symbolize the polarity of the MOSFET.

7.1 What is a Protected MOSFET High Side Driver?

MOSFET is an acronym for “Metal Oxide Semiconductor Field Effect Transistor”. The gate of the MOSFET is either metallic or polycrystalline silicon material which behaves like a metallic material. (In original MOSFET designs, the gate was metal.) Silicon dioxide (Oxide) is used to separate the gate from the underlying semiconductor. It is in the underlying semiconductor that various n-type and p-type dopants are used to form the “Field Effect Transistor”. (It is the “Field” from the gate down to the semiconductor which actually turns the MOSFET on and off.)

In switching applications, we will operate the MOSFET either in the off state (the switch is opened) or in the linear region (the switch is closed with a finite resistance). In switching applications we usually do not use the MOSFET in the saturation region.

In a high side drive configuration, a MOSFET switch is connected between the supply voltage and the load – the switch is on the “high” side of the load (Figure 7.1).

Figure 7.1 High Side Drive (HSD) Configuration

To turn on the MOSFET, we will again apply a positive voltage to the gate of the MOSFET in an attempt to enhance the channel and allow current to flow through the load. A fundamental property of the MOSFET comes into play here. The maximum voltage a MOSFET can pass to its source is at least one threshold voltage drop (about 1 V) below the applied gate volt. With a load connected, the gate-source voltage will be slightly larger, enough to support the required load current:

$$V_{GS} = V_G – V_S = 14\text{ V} – 11\text{ V} = 3\text{ V}$$

The fact that the drain and gate voltage are equal means that the MOSFET is in the saturation region and will act as a current source (Figure 7.1).

For a given load resistor the load current must be found for equal drain and gate voltage.

$$R_L \times I_L = V_{BAT} - V_S = V_{GS} = V_{DS}$$

From MOSFET datasheets the load current I_L can graphically be found that satisfies the equation. The dissipation is calculated by equation: $P = I_L \times V_{DS}$.

This type of biasing is not desirable for switching applications, because the MOSFET drain-to-source voltage is relatively high and thus the power dissipation in the device is high.

To remedy this situation, we actually pull the MOSFET gate to a higher voltage which will ensure linear MOSFET operation even at $I_L = V_{BAT} / R_L$ load current. This is done with either a charge pump circuit (the circuit will be reviewed later in this chapter).

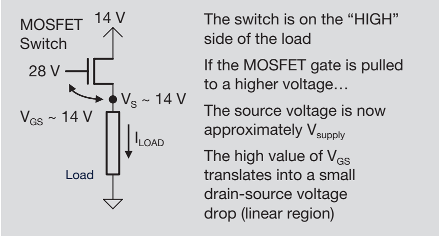


Figure 7.2 High Side Drive (HSD) Configuration

When we apply the higher voltage to the gate, two things change from our previous example. First, the source voltage of the MOSFET rises from a lower voltage to almost V_{BAT} meaning that a very small voltage is dropped across the MOSFET switch. Second, even at the full load current the power dissipation is greatly reduced in the MOSFET (Figure 7.2).

So, the fundamental difference in high side drive applications is that a higher voltage

needs to be applied to the MOSFET gate than in a low side application.

This leads us to a family of devices called in Infineon product listings as **PRO**tected **F**ield **E**ffect **T**ransistors (PROFETs).

What exactly is a PROFET? The power component of a PROFET is a standard n-channel power MOSFET and additional, low power integrated features have been added to improve its performance and protection in high side drive applications (Figure 7.3).

The PROFET performance is improved with the charge pump. As mentioned before, the charge pump circuit boosts the battery voltage. This boosted voltage is then applied to the MOSFET gate when it is turned on to ensure the MOSFET operates in the linear region.

In addition, several protection features are included in the PROFET. These features ensure the transistor will not be damaged if one or more of the operating conditions of the high side switch are improper.

For example, the short circuit protection can protect the MOSFET if a short to ground occurs.

The over voltage protection ensures the MOSFET will not be damaged during load dump transients or other high voltage events. The reverse battery protection keeps the PROFET safe if the battery polarity is reversed (current may flow, but the device or the load will not be damaged). The current limit feature limits the maximum current which the MOSFET high side switch can source. This again helps to protect the switch from short to ground conditions. The over temperature protection is enabled if the temperature of the MOSFET reaches unsafe levels, turning off the switch and stopping the current flow completely.

Finally, diagnostics are included so the PROFET can provide feedback to a microcontroller if any fault condition is present.

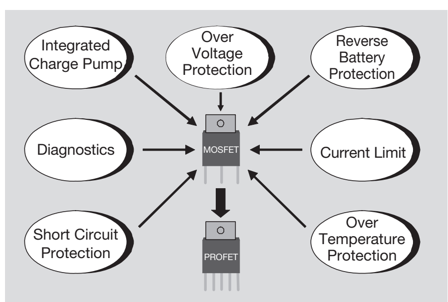


Figure 7.3 PROFETs = PROtected FETs

Figure 7.4 is a die photograph and a block diagram of a typical voltage controlled PROFET incorporating the features we mentioned. Notice the vertical, n-channel MOSFET and its integrated body diode (the body diode is created by junction of the p^+ wells at the source and the n-type drain). A vertical MOSFET is used to minimize the switch's R_{DSon}.

Surrounding the vertical power MOSFET (the switch) are the various functional blocks and protection circuitry implemented in CMOS technology (mentioned earlier). The ESD protection provided on the logic input and the diagnostic feedback pin of the PROFET are also visible.

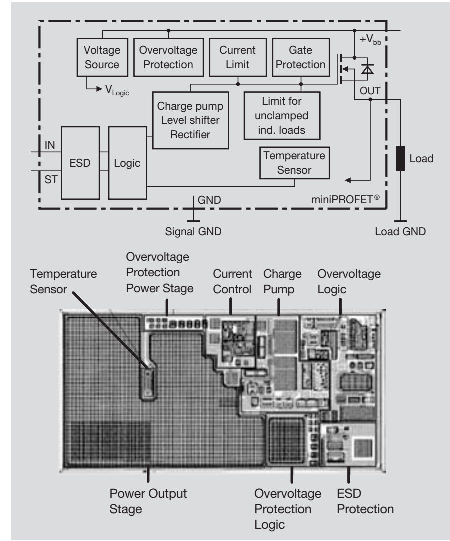


Figure 7.4 Voltage Controlled PROFET Block Diagram and Die View

In addition to voltage controlled PROFETs, Infineon also offers a line of current controlled PROFETs. Note that a current controlled PROFET requires one less pin than a voltage controlled device (no ground pin is required for the current controlled PROFET (Figure 7.5).

The construction and functional blocks of a current controlled PROFET are very similar to those of a voltage controlled PROFET, with the main difference being the input signal.

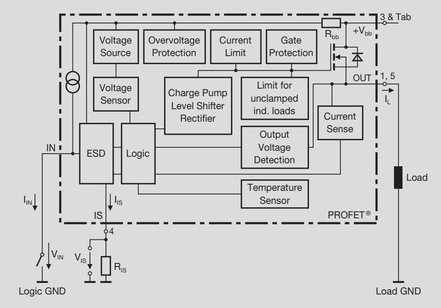


Figure 7.5 Current Controlled PROFET Block Diagram

7.2 Protection Features

Standard MOSFETs are often considered "rugged" devices. They are physically large, single transistor components manufactured in a rather simple fabrication process. However, they are unprotected components - meaning that if the device is subjected to a fault condition, the MOSFET is often not able to protect itself or its load.

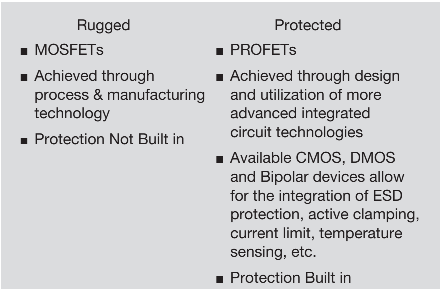


Figure 7.6 Rugged vs. Protected

PROFETs improve upon MOSFET devices by incorporating additional protection circuitry implemented in a fabrication process offering CMOS and DMOS components. Therefore, if the PROFET is subjected to a faulty operating condition, the PROFET can protect itself and its load (Figure 7.6).

The following is a list of the available protection features in Infineon's PROFET portfolio. Many of these may be familiar to you, some are unique to the PROFET family (Figure 7.7).

- Electrostatic Discharge (ESD) Protection
- Overvoltage / Load Dump Protection
- Overvoltage Shut Down Protection and Restart
- Undervoltage Shut Down Protection and Restart
- Reverse Battery Protection
- Reversave™ Battery Protection
- Inductive and Overvoltage Output Clamp Protection
- Thermal Shutdown Protection
- Current Limit Protection
- Short Circuit Shut Down Protection
- Inversave™ Inverse Current Protection
- Loss of Ground Protection
- Loss of Supply Voltage Protection

Figure 7.7 PROtected FET (PROFET) Protection Features

We now will briefly examine each of these protection features. Before describing PROFET protection features, let us briefly review the block diagram of a typical voltage controlled PROFET (Figure 7.8).

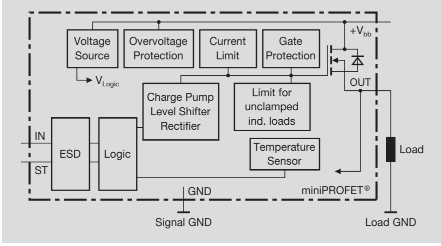


Figure 7.8 Block Diagram Including Protection Features

In the block diagram, we can see that many of the protection features are shown in their own functional blocks (ESD, over voltage protection, current limiting). Additional

protection circuits use one or more of the “sensing” functional blocks (current sense, temperature sensor). Finally, there are those protection circuits that are operating “behind the scenes” and are not readily apparent from such a high level functional diagram.

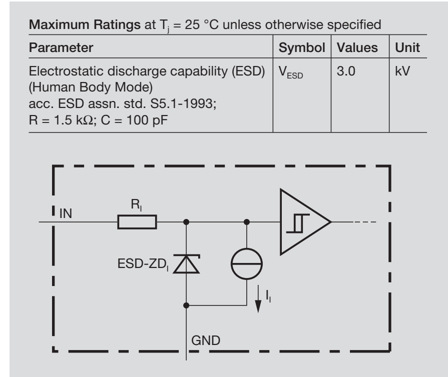
First, we will look at the PROFET input electrostatic discharge (ESD) protection. Figure 7.9 is an example of the ESD protection specification from a PROFET datasheet. Note that the conditions for the ESD test are shown (a human body model test is identified in the example).

Under standard operating voltages, when the IN signal is raised, the zener diode remains off. The signal is then processed by a Schmitt trigger before it actually reaches the control portion of the PROFET. The Schmitt trigger, however, has a high input impedance, so only a small amount of current flows into the device.

When the IN signal exceeds the standard operating voltages, the ESD zener diode turns on pulling current through resistor R_I. This holds the cathode of the zener diode at a safe (clamped) level and protects the Schmitt trigger and the remainder of the PROFET device. Under these conditions, the resistor, R_I, limits the current flowing through the ESD zener diode to a safe level:

$$I_{IN} = (V_{IN} - V_Z) / R_I$$

A pull down current source is included to ensure that the device remains off when the input is floating.



Maximum Ratings at T_j = 25 °C unless otherwise specified

Parameter	Symbol	Values	Unit
Electrostatic discharge capability (ESD) (Human Body Mode) acc. ESD assn. std. S5.1-1993; R = 1.5 kΩ; C = 100 pF	V_{ESD}	3.0	kV

Figure 7.9 ESD Protection

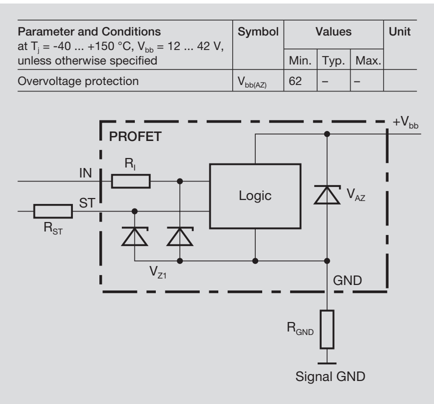
Next, PROFETs are protected if the supply voltage is increased beyond the standard operating conditions (over voltage protection).

Figure 7.10 illustrates the over voltage protection method and a detail from a PROFET datasheet is also shown. In the example, the minimum voltage at which the PROFET’s over voltage protection is activated is 62 V.

The circuit shows how the over voltage protection is functionally implemented. For standard supply voltages:

$$V_{bb} < V_{AZ} ,$$

and the zener diode is not active. However, as the supply voltage increases, it will eventually turn on the zener diode, clamping the supply voltage at the specified value, and protecting the logic portion of the PROFET. Note, that the **external resistor, R_{GND},** is used to limit the current in the zener diode during over voltage events. The voltage drop on R_{GND} is added to V_{AZ} when the over voltage limits are calculated.



Parameter and Conditions at T_j = -40 ... +150 °C, V_{bb} = 12 ... 42 V, unless otherwise specified	Symbol	Values			Unit
		Min.	Typ.	Max.	
Overvoltage protection	$V_{bb(AZ)}$	62	–	–	

Figure 7.10 Over Voltage Protection

Some PROFETs, in addition to the internal clamping protection, turn themselves off if the supply voltage exceeds the over voltage threshold. This also protects the load from seeing too high a voltage. In Figure 7.11, an example of the over voltage shutdown specification from a PROFET datasheet is shown. In this example, the supply voltage range in which the PROFET will turn itself off is 34 V (minimum) to 42 V (maximum).

To prevent a PROFET from oscillating between the normal and shutdown modes, hysteresis is included on the over voltage shutdown. In this example, the typical value of the hysteresis is 0.7 V. For example, if a PROFET's over voltage shut down protection was activated at 36 V, the PROFET would typically turn itself back on when the supply voltage fell to 35.3 V.

The final specification shown in this over voltage shutdown example is the over voltage restart. This specification defines the supply voltage at which the PROFET is guaranteed to turn itself back on after going into over voltage shutdown. So, regardless of whether the PROFET entered over voltage shutdown at 34 V or 42 V, it is guaranteed to turn itself back on when the supply voltage falls to 33 V.

Parameter and Conditions at T_j = 25 °C, V_{bb} = 13.5 V, unless otherwise specified	Symbol	Values			Unit
		Min.	Typ.	Max.	
Operating Parameters					
Overvoltage shutdown T_j = -40 ... +150 °C	$V_{bb(over)}$	34	–	42	V
Overvoltage hysteresis T_j = -40 ... +150 °C	$\Delta V_{bb(over)}$	–	0.7	–	V
Overvoltage restart T_j = -40 ... +150 °C	$\Delta V_{bb(o\ rst)}$	33	–	–	V

Figure 7.11 Over Voltage Shutdown Protection and Restart

Similar to over voltage shutdown, under voltage shutdown will turn off a PROFET if the supply voltage falls below a minimum threshold. Without this protection the operation of logic blocks may become unpredictable and the charge pump may not be able to supply sufficient gate voltage drive as the supply voltage falls.

In Figure 7.12, an example of the under voltage shutdown specification from a PROFET datasheet is shown. The operating supply voltage range in which this particular PROFET will operate properly is 5.5 V (minimum) to 38 V (maximum). As the supply voltage is being turned off, under 5.5 V, the charge pump will turn itself off at some voltage, but before the control logic is shut down. Between 2.5 V and 3.5 V the logic will turn itself off. Note that the turn off voltage of the charge pump will always be higher than that of the logic.

When the supply is first turned on from an off state the logic may come alive between 2.5 V and 3.5 V but the charge pump stays off until the supply voltage reaches the typical value of 4.0 V. The charge pump will definitely turn on (in the extreme case) if the supply is 5.5 V.

This method of under voltage protection will prevent a PROFET from oscillating into and out of under voltage protection operating mode and it will also minimize the hazards of losing control over the power switch.

The method of protection is not the same for **all** PROFET devices. There are differences between newer and older designs as well as voltage controlled and current controlled devices.

The user is strongly advised to pay attention to datasheet details and study the available Application Notes.

Parameter and Conditions at T_j = 25 °C, V_{bb} = 12 V, unless otherwise specified	Symbol	Values			Unit
		Min.	Typ.	Max.	
Operating Parameters					
Operating voltage (V_{IN} = 0) T_j = -40 ... 150 °C	$V_{bb(on)}$	5.5	–	38	V
Undervoltage shutdown	$V_{bIN(u)}$	–	2.5	3.5	V
Undervoltage restart of charge pump	$V_{bb(ucp)}$	–	4	5.5	V

Figure 7.12 Under Voltage Shutdown Protection and Restart

For passenger vehicles the load dump pulse amplitude is 60 V above the battery voltage and the pulse duration between the amplitude levels of 10% - 10% of the peak amplitude is 150 ms. The internal impedance of the load dump source (R_G) is 0.5 Ω (ISO 7637-2).

PROFETs don't necessarily shut down during load dump conditions but if the PROFET has over voltage shut down feature the appropriate shut down limits will apply.

If the PROFET only has over voltage protection the protection levels will depend on the load resistor value (see Figure 7.13). The over voltage protection levels are determined by the Zener-diode voltage and the value chosen for R_{GND} (see Figure 7.10 and Figure 7.14). If the load dump amplitude is insufficient to activate the over voltage Zener-diode or initiate a shut down there won't be any protection at all i.e. the part is robust enough to handle the stress.

Maximum Ratings at T_j = 25 °C unless otherwise specified

Parameter	Symbol	Value	Unit
Load dump protection $V_{LoadDump} = V_A + V_S$ R_I = 2 Ω, t_d = 400 ms, V_{IN} = low or high, V_A = 13.5 V	$V_{Loaddump}$		V
R_L = 13.5 Ω R_L = 27 Ω		73.5 88.5	

- The rated load dump voltage is a function of the generator impedance (R_G) and the load resistance (R_L)
- As R_G and R_L increase, less energy is dissipated in the PROFET, and the maximum allowable load dump voltage increases

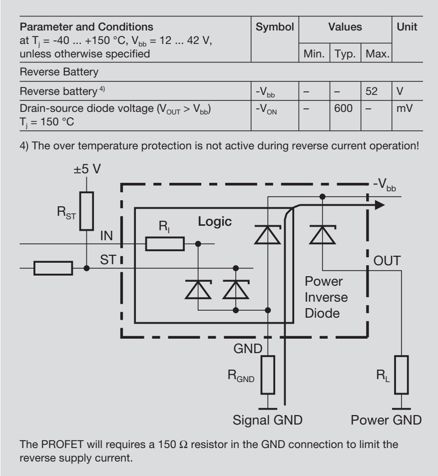
Figure 7.13 Load Dump Protection

PROFETs can also be protected against reverse battery conditions.

In Figure 7.14 there is an example of the reverse battery protection specification from a PROFET datasheet. In this example, the maximum reverse battery voltage the PROFET can withstand is 52 V.

The circuit in Figure 7.14 illustrates how the reverse battery protection is implemented. If the supply voltage and ground connections are reversed, current will flow from the signal ground, through the current limiting resistor (R_{GND}), through a diode, to the supply voltage connection.

Note that the PROFET is not actually functioning during the reverse battery condition, therefore the over temperature protection circuitry is not functional.



Parameter and Conditions at T_j = -40 ... +150 °C, V_{bb} = 12 ... 42 V, unless otherwise specified	Symbol	Values			Unit
		Min.	Typ.	Max.	
Reverse Battery					
Reverse battery [4]	$-V_{bb}$	–	–	52	V
Drain-source diode voltage ($V_{OUT} > V_{bb}$) T_j = 150 °C	$-V_{ON}$	–	600	–	mV

4) The over temperature protection is not active during reverse current operation!

The PROFET will requires a 150 Ω resistor in the GND connection to limit the reverse supply current.

Figure 7.14 Reverse Battery Protection

It is also necessary to limit the reverse battery current through the PROFET's load and MOSFET body diode. It is very likely that the power dissipated in the body diode of the PROFET will be higher than the what the PROFET dissipates during normal operation:

For example, in normal operation:

V_{bb} = 14 V
R_L = 1.99 Ω
R_{DSon} = 0.01 Ω

$$I_{LOAD} = V_{bb} / (R_L + R_{DSon}) = 14\text{ V} / (1.99\ \Omega + 0.01\ \Omega) = 7\text{ A}$$

$$P_{PROFET} = I_{LOAD}^2 R_{DSon} = (7\text{ A})^2(0.01\ \Omega) = 0.49\text{ W}$$

The same calculations can be made for reverse battery conditions:

V_{DIODE} = 0.6 V

I_{LOAD} = $(V_{bb} - V_{DIODE}) / R_L$
= (14 V - 0.6 V) / 1.99 Ω = 6.73 A

P_{PROFET} = $V_{DIODE}\ I_{LOAD}$ = (0.6 V)(6.73 A)
= 4.04 W

As the over temperature protection circuitry is not active during reverse battery conditions it is necessary to provide adequate heat sinking so, that the maximum junction temperature is not exceeded (Figure 7.15).

In this example with ReverSave™ protection:

I_{LOAD} = 7.5 A
R_{DSon} = 13 mΩ (maximum)

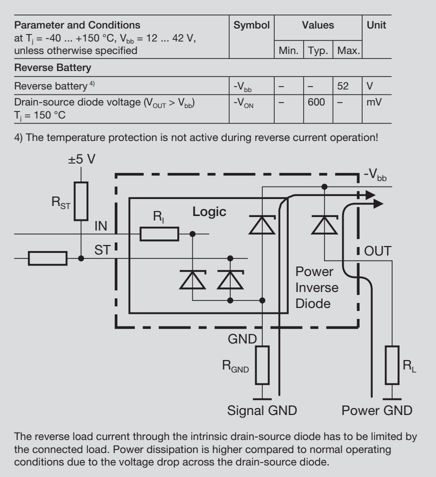
P_{PROFET} = $I_{LOAD}^2\ R_{DSon}$ = (7.5 A)^2(13 mΩ)
= 0.731 W

V_{DS} = I_{LOAD} x R_{DSon} = 97.5 mV << 600 mV

Without ReverSave™ protection:

V_{DIODE} ~ 600 mV

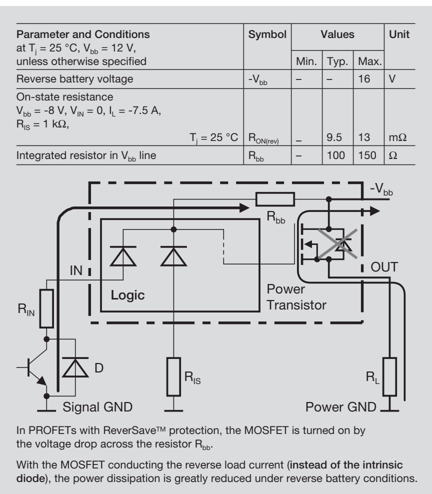
P_{PROFET} = $V_{DIODE}\ I_{LOAD}$ = (0.6 V)(7.5A)
= 4.5 W



Parameter and Conditions at T_j = -40 ... +150 °C, V_{bb} = 12 ... 42 V, unless otherwise specified	Symbol	Values			Unit
		Min.	Typ.	Max.	
Reverse Battery					
Reverse battery [4)]	$-V_{bb}$	–	–	52	V
Drain-source diode voltage ($V_{OUT} > V_{bb}$) T_j = 150 °C	$-V_{ON}$	–	600	–	mV

4) The temperature protection is not active during reverse current operation!

Figure 7.15 Reverse Battery Protection



Parameter and Conditions at T_j = 25 °C, V_{bb} = 12 V, unless otherwise specified	Symbol	Values			Unit
		Min.	Typ.	Max.	
Reverse battery voltage	$-V_{bb}$	–	–	16	V
On-state resistance V_{bb} = -8 V, V_{IN} = 0, I_L = -7.5 A, R_{IS} = 1 kΩ, T_j = 25 °C	$R_{ON(rev)}$	–	9.5	13	mΩ
Integrated resistor in V_{bb} line	R_{bb}	–	100	150	Ω

Figure 7.16 ReverSave™ Reverse Battery Protection

Sometimes, it is not practical to implement a heatsink large enough to protect a PROFET during reverse battery conditions. Therefore, Infineon has developed a line of PROFETs with ReverSave™ protection feature (Figure 7.16).

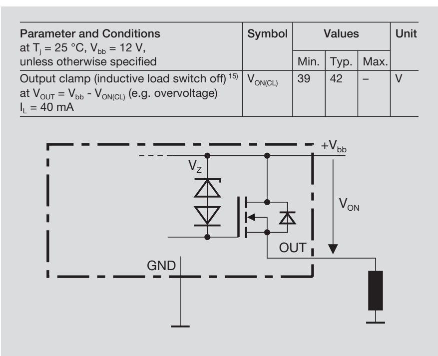
When current flows from the signal ground to the supply voltage connection, voltage is dropped across the resistor, R_{bb}. This voltage is sufficient to partially turn on just the MOSFET within the PROFET and significantly reduce the power dissipation during reverse battery conditions. (The intrinsic MOSFET body diode will not turn on unless the voltage across the MOSFET approaches 600 mV).

PROFETs are also protected when inductive loads are turned off by clamping the on voltage to a safe value ($V_{ON} = V_{bb} - V_{OUT}$).

When the PROFET turns off the current in an inductive load, the MOSFET starts reducing the source current. However, the current flowing through an inductor cannot be changed instantaneously. The reducing inductor current will cause V_{OUT} to go negative as the polarity across the inductor reverses. In a PROFET, the MOSFET gate is tied to the source when V_{IN} is low. Therefore, as the source is pulled below ground, the MOSFET gate is also pulled below ground. At the

specified drain-to-source voltage (V_{ON}), the clamping diode will hold the gate voltage constant at a negative value (Figure 7.17). When that happens V_{OUT} continues to fall a small amount to turn the MOSFET on, thus opening a path for the decaying inductor current to ground. V_{OUT} will be clamped at a negative voltage limit and remains at that level as long as the current > 0. (Refer to Chapter 1 and Chapter 6).

If the diode clamp was not integrated in the protected MOSFET, then at turning off the inductive load V_{OUT} would continue to fall below ground until the MOSFET went into avalanche breakdown. During avalanche breakdown, the inductor current would still decay, but voltage across the MOSFET would be larger than in the clamped case causing significant power dissipation within the MOSFET. The higher voltage drop is due to the fact that not the total drain surface is conducting but only the portions where avalanching is taking place. Therefore, if V_{OUT} is clamped to a safe negative value, no avalanche break down occurs and the power dissipated in the MOSFET is kept below manageable and predictable levels.



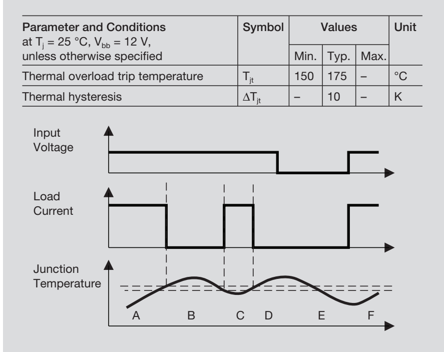
Parameter and Conditions at T_j = 25 °C, V_{bb} = 12 V, unless otherwise specified	Symbol	Values			Unit
		Min.	Typ.	Max.	
Output clamp (inductive load switch off) [15] at $V_{OUT} = V_{bb} - V_{ON(CL)}$ (e.g. overvoltage) I_L = 40 mA	$V_{ON(CL)}$	39	42	–	V

Figure 7.17 Inductive and Over Voltage Output Clamp Protection

PROFETs can be protected from a variety of problems with thermal shutdown. The PROFET is guaranteed to operate within the specification limits up to T_{jt} = 150 °C. The minimum temperature at which the PROFET can enter into thermal shutdown is 150 °C. Above 150 °C (175 °C typical), however, the thermal shutdown will protect the PROFET by turning the device completely off.

To prevent a PROFET from oscillating into and out of the thermal shutdown mode, the thermal shutdown logic operates with hysteresis. In this example, the typical value of the hysteresis is 10 K. Therefore, if a PROFET's thermal shutdown was activated at 155 °C the PROFET would typically turn itself back on when the junction temperature fell to 145 °C. (Recall that a 1 K temperature differential is equivalent to a 1 °C differential.)

Figure 7.18 also shows a series of waveforms showing the operation of a PROFET with thermal shutdown protection:



Parameter and Conditions at T_j = 25 °C, V_{bb} = 12 V, unless otherwise specified	Symbol	Values			Unit
		Min.	Typ.	Max.	
Thermal overload trip temperature	T_{jt}	150	175	–	°C
Thermal hysteresis	ΔT_{jt}	–	10	–	K

Figure 7.18 Thermal Shutdown Protection

Time: A Input Voltage: High
Junction Temperature: < T_{jt}
PROFET Status: On

Time: B Input Voltage: High
Junction Temperature: > T_{jt}
PROFET Status: Off

Time: C Input Voltage: High
Junction Temperature: < T_{jt}
PROFET Status: On

Time: D Input Voltage: High
Junction Temperature: > T_{jt}
PROFET Status: Off

Time: E Input Voltage: Low
Junction Temperature: Falls below T_{jt}
PROFET Status: Off

Time: F Input Voltage: High
Junction Temperature: < T_{jt}
PROFET Status: On

PROFETs can be protected against over current conditions in a number of different ways. First, PROFETs have a protection feature called current limiting.

With current limit protection, a PROFET can limit the maximum current delivered to a load. Therefore, if the load resistance is less than expected (0 Ω with a direct short to ground), the PROFET will modify the MOSFET gate voltage to ensure that its source current does not exceed the specified limited current value.

In this example, two different current limit specifications are shown for a particular PROFET. First, an initial peak current limit value is provided:

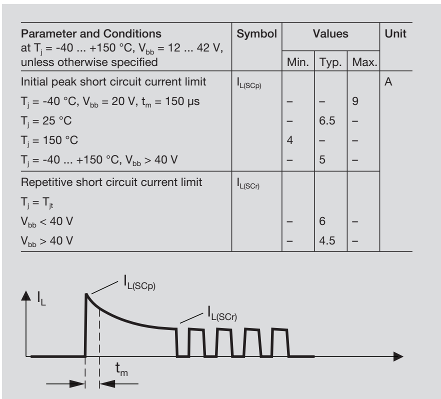
$I_{L(SCp)} \sim I_{LOAD\ (Short\ Circuit,\ peak)}$

Next, the repetitive current limit value is given:

$I_{L(SCr)} \sim I_{LOAD\ (Short\ Circuit,\ repetitive)}$

Note that the initial peak current limit value is higher than the repetitive value.

An example of the load current in a PROFET in current limit is shown at the lower half of the Figure 7.19. When the PROFET is initially turned on, the device immediately goes into current limit ($I_L = I_{L(SCp)}$). The PROFET immediately begins to further limit the current it sources until $I_{L(SCr)}$ is reached while the junction temperature is increasing. When $I_{L(SCr)}$ is reached the waveform shows that the PROFET goes into thermal shutdown as the thermal overload trip temperature is exceeded. Finally, the waveform shows the PROFET toggling between over temperature shutdown and current limit operation.

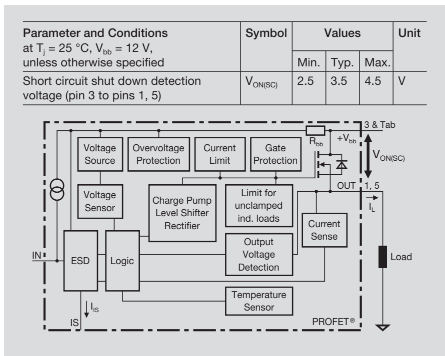


Parameter and Conditions at T_j = -40 ... +150 °C, V_{bb} = 12 ... 42 V, unless otherwise specified	Symbol	Values			Unit
		Min.	Typ.	Max.	
Initial peak short circuit current limit	$I_{L(SCp)}$				A
T_j = -40 °C, V_{bb} = 20 V, t_m = 150 µs		–	–	9	
T_j = 25 °C		–	6.5	–	
T_j = 150 °C		4	–	–	
T_j = -40 ... +150 °C, V_{bb} > 40 V		–	5	–	
Repetitive short circuit current limit	$I_{L(SCr)}$				
$T_j = T_{jt}$					
V_{bb} < 40 V		–	6	–	
V_{bb} > 40 V		–	4.5	–	

Figure 7.19 Current Limit Protection

In addition to current limiting, some PROFETs can be protected against an over current condition by short circuit detection.

An example of the short circuit protection specification from a PROFET datasheet is shown. Instead of specifying the maximum current, the short circuit shutdown detection voltage, $V_{ON(SC)}$, is specified. Figure 7.20 shows the short circuit shutdown detection voltage is measured across the power MOSFET:



Parameter and Conditions at T_j = 25 °C, V_{bb} = 12 V, unless otherwise specified	Symbol	Values			Unit
		Min.	Typ.	Max.	
Short circuit shut down detection voltage (pin 3 to pins 1, 5)	$V_{ON(SC)}$	2.5	3.5	4.5	V

Figure 7.20 Short Circuit Shutdown Protection

$V_{ON(SC)} = V_{bb} - V_{OUT}$

Recall that V_{ON} is the product of the load current and the R_{DSon} of the PROFET:

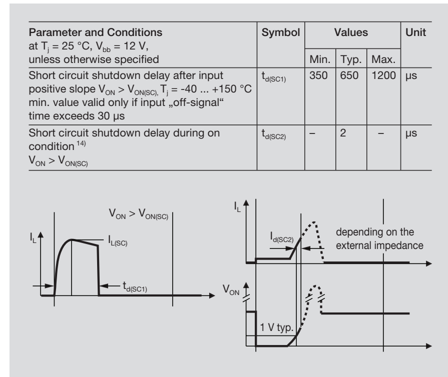
$V_{ON} = I_{LOAD}\ R_{DSon}$

Therefore, when the load current increases, the voltage across the PROFET increases. When it exceeds $V_{ON(SC)}$ (4.5 V maximum), the PROFET turns itself off. The PROFET may be in current limit by that time, meaning the MOSFET is in the saturation mode, not in the linear mode.

PROFETs with short circuit protection have a delay built into the short circuit shutdown circuit.

The first specification is for the initial turn-on of the PROFET (after input current positive slope in case of a current controlled device). For the initial turn on of the PROFET, if $V_{ON} > V_{ON(SC)}$, the short circuit shutdown will typically wait 650 µs following the device turn-on before shutting down the PROFET. This delay allows the PROFET to completely turn on and not mistake a transient in-rush load current behavior with a true short circuit. The figure at the lower left of Figure 7.21 shows a PROFET shutting down after $t_{d(SC1)}$ expires.

The second specification is applicable for any time after the initial turn on of the PROFET. Under such circumstances, the short circuit shutdown delay is typically only 2 µs. The figure at the lower right of Figure 7.21 shows a PROFET shutting down after $t_{d(SC2)}$ expires. Initially, the PROFET is off and V_{ON} is equal to the supply voltage (output is grounded). Next, the PROFET is turned on. Some time later, the load current (and V_{ON}) begin to rise. 2 µs after $V_{ON} > V_{ON(SC)}$, the short circuit shutdown protection turns off the PROFET. Depending upon the external impedance (inductance), the load current and V_{ON} will vary until reaching their respective steady state values ($I_L = 0$ A and $V_{ON} = V_{bb}$).



Parameter and Conditions at $T_j = 25$ °C, $V_{bb} = 12$ V, unless otherwise specified	Symbol	Values			Unit
		Min.	Typ.	Max.	
Short circuit shutdown delay after input positive slope $V_{ON} > V_{ON(SC)}$, $T_j = -40 ... +150$ °C min. value valid only if input „off-signal“ time exceeds 30 µs	$t_{d(SC1)}$	350	650	1200	µs
Short circuit shutdown delay during on condition [14] $V_{ON} > V_{ON(SC)}$	$t_{d(SC2)}$	–	2	–	µs

Figure 7.21 Short Circuit Shutdown Protection

Some PROFETs can be operated in Inversave™ mode to protect against inverse current (current flowing into the PROFET from the output if the output voltage is higher than V_{bb}).

The first specification listed is the on state resistance of the PROFET when the device is operated in inverse current mode (10 mΩ maximum).

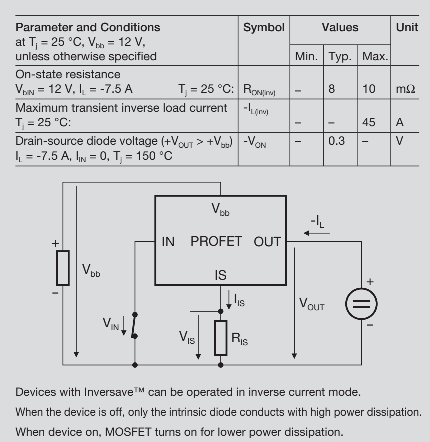
Next, the maximum transient inverse load current specification is given (45 A maximum). Finally, the intrinsic body diode voltage in inverse current mode is provided (600 mV typical at 150 °C).

In inverse current mode, when the PROFET is turned off, the power dissipation of the PROFET is determined by the voltage drop across, and the current through the body diode. For example:

$$
\begin{aligned}
I_{INVERSE} &= 10 \text{ A} \\
P_{PROFET} &= V_{DIODE}\, I_{INVERSE} \\
&= (0.6 \text{ V})(10 \text{ A}) \\
&= 6 \text{ W}
\end{aligned}
$$

When the PROFET is turned on, however, the power dissipation in the device is significantly reduced since the channel is activated and stays on during inverse mode (Figure 7.22):

$$R_{DSon(inverse)} = 10\ m\Omega$$
$$P_{PROFET} = I_{INVERSE}^{2}\ R_{DSon(inverse)} = (10\ A)^2(10\ m\Omega) = 1\ W$$



Parameter and Conditions at T_j = 25 °C, V_{bb} = 12 V, unless otherwise specified	Symbol	Values			Unit
		Min.	Typ.	Max.	
On-state resistance V_{bIN} = 12 V, I_L = -7.5 A T_j = 25 °C:	$R_{ON(inv)}$	–	8	10	mΩ
Maximum transient inverse load current T_j = 25 °C:	$-I_{L(inv)}$	–	–	45	A
Drain-source diode voltage (+V_{OUT} > +V_{bb}) I_L = -7.5 A, I_{IN} = 0, T_j = 150 °C	$-V_{ON}$	–	0.3	–	V

Figure 7.22 Inversave™ Inverse Current Protection

PROFETs can also be protected against a loss of ground.

If the PROFET ground is lost, the input voltages (V_{bb} and V_{IN}) can still be referenced to ground through the load resistance. This will allow the PROFET to protect the system by turning off the load (Figure 7.23).

This situation can occur if the PROFET is part of an electronic control module separate from the load it is driving, for example a fuel pump. If the electronic control module were to lose its ground connection (perhaps through a vehicle crash), this would allow the PROFET to safely turn off the fuel pump.

- With Loss of Ground Protection, V_{bb}, V_{IN}, V_{ST} are still referenced to ground through the output
- This ensures the device will be safely shut off if the ground pin is opened

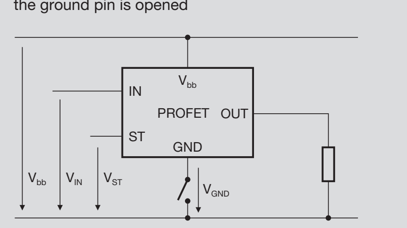


Figure 7.23 Loss of Ground Protection

The final protection feature available in PROFETs is protection from loss of supply voltage (Figure 7.24). Suppose a PROFET is driving an inductive load when the PROFET is suddenly disconnected from the supply voltage. The inductive current must find a path to recirculate and decay to zero. Some PROFETs (mostly monolithic devices) are protected against this situation by allowing the inductive load current to recirculate through the ground pin of the PROFET.

- All PROFETs are protected against a loss of supply voltage for non-inductive loads
- Most PROFETs are also protected against a loss of supply voltage for inductive loads by handling the recirculation current through the GND pin

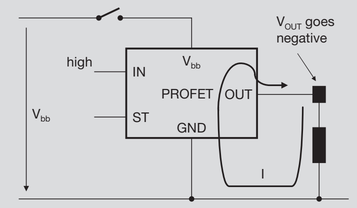


Figure 7.24 Loss of Supply Voltage Protection

7.3 Diagnostic Features

PROFETs have one of two types of diagnostic feedback, digital or analog.

In a PROFET with digital feedback, the status of the device is reported as a voltage on a status line.

In a PROFET with analog feedback, the device status is reported as a variable current flowing out of a status pin (Figure 7.25).

While the analog feedback does require an external resistor to convert the current into a voltage, it can provide additional information regarding the operation of the PROFET.

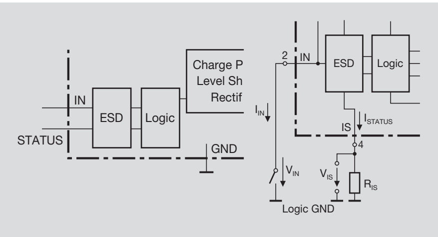


Figure 7.25 PROFET Diagnostic Feedback Digital vs. Analog

Next, we see a simple functional diagram of a voltage controlled PROFET with digital (voltage) diagnostic feedback. The status pin is typically an open drain transistor, requiring an external pull-up. The error state may be high or low, depending on the device (Figure 7.25).

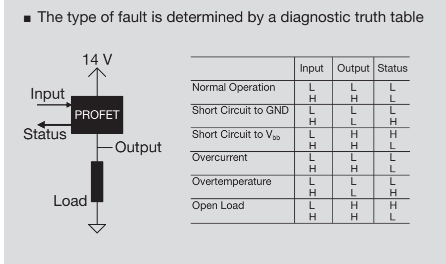
There are various fault conditions that the PROFET might report both in the ON state and the OFF state, again depending on the device.

In the ON state, all PROFETs will report an over temperature condition. There are types that also report an overload (current limit), while other types will report an open load fault based on a current threshold (Figure 7.26).

In the OFF state, some devices will indicate an open load, via a comparator at the output pin (more detail on the detection method follows).

Why are some conditions only reported when the device is on or off? Because some error conditions will look like valid operating conditions when the PROFET is either on or off. This method is also helpful to distinguish between shorted output and open load conditions.

For example, consider what happens when the output is shorted to V_{bb}. When the device is on, the output of the PROFET is expected to be near to the supply voltage under normal operations. The device cannot differentiate between an output shorted to V_{bb} and a normal operation when the PROFET is on. Only when the PROFET is turned off and the output voltage is still at the supply voltage will the device be able to identify that a problem exists (such a problem, by the way, may damage the load).



	Input	Output	Status
Normal Operation	L H	L H	L L
Short Circuit to GND	L H	L L	L H
Short Circuit to V_{bb}	L H	H H	H L
Overcurrent	L H	L H	L L
Overtemperature	L H	L L	L H
Open Load	L H	H H	H L

Figure 7.26 Digital Diagnostic Feedback

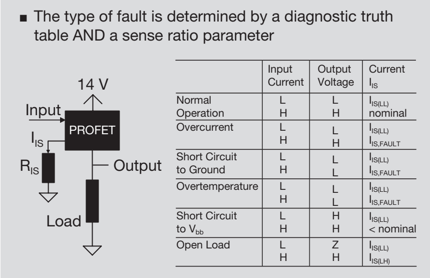
On the left part of Figure 7.27, we see a simple functional block diagram of a current controlled PROFET with analog (current) diagnostic feedback.

$I_{IS(LL)}$ is the status (or sense) leakage current sourced by the PROFET when the PROFET is turned off and the output current is 0 A. This current is about 0.1 µA.

Under normal operation, the magnitude of I_{IS} sourced by the PROFET is proportional to the load current (this is elaborated on the next page). After the device is turned on, if the PROFET determines that there is a fault (over-current, short to ground, or over-temperature), it will change the status current $I_{IS,FAULT}$, to about 5 mA (see the truth table in Figure 7.27).

If the PROFET is turned on and the output is shorted to V_{bb}, the load current (and I_{IS}) will depend on the resistance of the short to V_{bb}.

Finally, if the PROFET is turned on and the output is open, the status current will change from 0.1 µA to $I_{IS(LH)}$ (approximately 1 µA).



	Input Current	Output Voltage	Current I_{IS}
Normal Operation	L H	L H	$I_{IS(LL)}$ nominal
Overcurrent	L H	L H	$I_{IS(LL)}$ $I_{IS,FAULT}$
Short Circuit to Ground	L H	L L	$I_{IS(LL)}$ $I_{IS,FAULT}$
Overtemperature	L H	L L	$I_{IS(LL)}$ $I_{IS,FAULT}$
Short Circuit to V_{bb}	L H	H H	$I_{IS(LL)}$ $<$ nominal
Open Load	L H	Z H	$I_{IS(LL)}$ $I_{IS(LH)}$

Figure 7.27 Analog Diagnostic Feedback

As mentioned previously, under normal operation, when the PROFET is sourcing a load current, the magnitude of I_{IS} is proportional to I_{LOAD} (Figure 7.28).

The proportionality constant is defined as:

$$K_{ILIS} = I_{LOAD} / I_{IS}$$

Therefore, for a given load current and proportionality constant the feedback current is given by:

$$I_{IS} = I_{LOAD} / K_{ILIS}$$

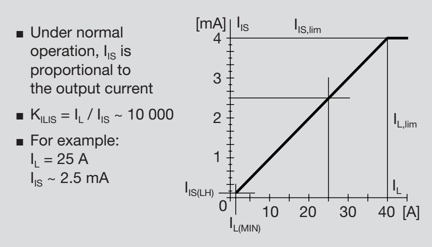


Figure 7.28 Analog Load Current Feedback Via I_{IS} Current

For the load current range greater than $I_{L(MIN)}$ the proportionality constant is quite accurate. Figure 7.28 is typical to a 30 A PROFET.

For very low load currents the PROFET is not expected to normally operate and, as a result, the proportionality constant is not as accurate (see Figure 7.29).

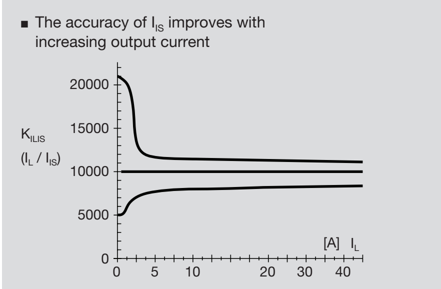


Figure 7.29 I_{IS} Current Sense Ratio

The response at the status pin (STATUS) of a digital (voltage) feedback PROFET is delayed with respect to the input.

The analog (current) feedback status current (I_{IS}) of a PROFET is also delayed with respect to the input.

The shaded areas of the I_L and I_{IS} vs. time waveforms in Figure 7.30 illustrate the times when the status feedback (I_{IS} here) is not valid. Three different delay times are illustrated:

$t_{son(IS)}$ - delay of status current validity when PROFET is turning on

$t_{slc(IS)}$ - delay of status current validity when PROFET is on and load current is changing

$t_{soff(IS)}$ - delay of status current validity when PROFET is turning off

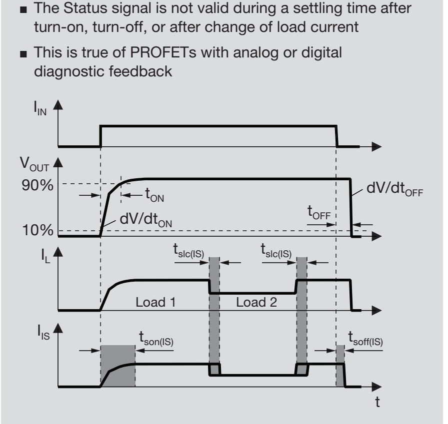


Figure 7.30 Status Signal Setting Time

PROFETs employ one of three methods to detect an open load condition (defined as: the output of the PROFET is disconnected from the load and its potential is floating).

Three Different PROFET strategies

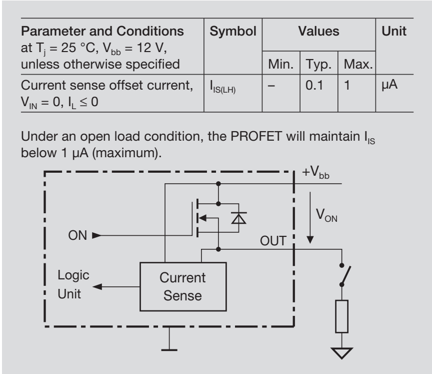
- Open load detection via Sense pin on HiC (High Current) PROFETs and some PROFETs
- Open load detection while PROFET is turned on (for some PROFETs---mostly older types)
- Open load detection while PROFET is turned off (for most PROFETs---mainly newer types)

Figure 7.31 Open Load Detection

Some PROFETs sense an open load condition simply by monitoring the status current, I_{IS}. When the PROFET load is open, the load current will be 0 A. Theoretically, the sense current would also be 0 A (Figure 7.32):

$$I_{IS} = I_{LOAD} / K_{ILIS} = 0 \text{ A} / K_{ILIS} = 0 \text{ A}$$

In practice, the example PROFET guarantees that under such load conditions, the status current will remain below 1 µA (due to internal leakage currents).



Parameter and Conditions at T_j = 25 °C, V_{bb} = 12 V, unless otherwise specified	Symbol	Values			Unit
		Min.	Typ.	Max.	
Current sense offset current, V_{IN} = 0, $I_L \le 0$	$I_{IS(LH)}$	–	0.1	1	µA

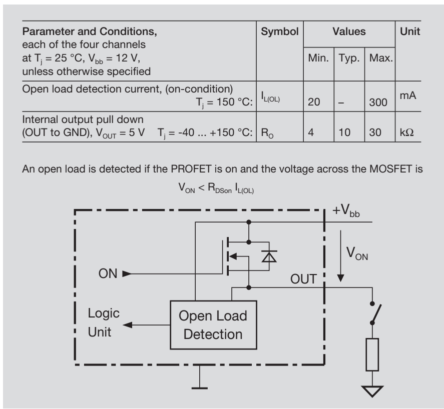
Under an open load condition, the PROFET will maintain I_{IS} below 1 µA (maximum).

Figure 7.32 Open Load Detection-Via Sense Pin

Many of the older PROFETs detect an open load by measuring the voltage across the MOSFET when the device is on (Figure 7.33).

For example, if the load was completely open, V_{OUT} would equal V_{bb} and no voltage is dropped across the MOSFET (V_{ON} = 0 V). Then, an open load would be detected.

This method also allows for the detection of abnormally low output currents (with abnormally low values of V_{ON}).



Parameter and Conditions, each of the four channels at T_j = 25 °C, V_{bb} = 12 V, unless otherwise specified	Symbol	Values			Unit
		Min.	Typ.	Max.	
Open load detection current, (on-condition) T_j = 150 °C:	$I_{L(OL)}$	20	–	300	mA
Internal output pull down (OUT to GND), V_{OUT} = 5 V T_j = -40 ... +150 °C:	R_O	4	10	30	kΩ

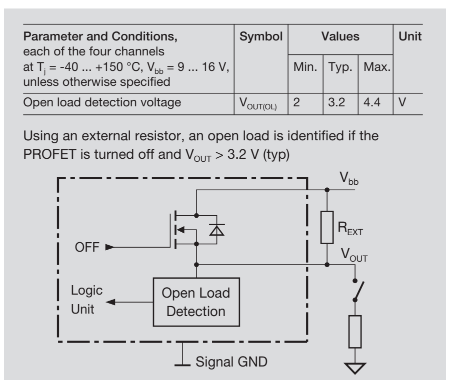
An open load is detected if the PROFET is on and the voltage across the MOSFET is

$V_{ON} < R_{DSon}\, I_{L(OL)}$

Figure 7.33 Open Load Detection-PROFET On

Some PROFETs can also detect an open load condition even when the device is turned off. An external resistor (R_{EXT}) is connected across the MOSFET. This creates a pull up to the supply voltage when the PROFET's MOSFET is turned off and the load is open (Figure 7.34).

In this example, the PROFET will detect and report an open load when the voltage at the output is more than 3.2 V (typical) during off state. This allows the user to select a large value for R_{EXT} and to prevent the load from turning on when the PROFET is turned off.



Parameter and Conditions, each of the four channels at T_j = -40 ... +150 °C, V_{bb} = 9 ... 16 V, unless otherwise specified	Symbol	Values			Unit
		Min.	Typ.	Max.	
Open load detection voltage	$V_{OUT(OL)}$	2	3.2	4.4	V

Using an external resistor, an open load is identified if the PROFET is turned off and V_{OUT} > 3.2 V (typ)

Figure 7.34 Open Load Detection-PROFET Off

7.4 EMI / EMC Considerations

Now we will examine how the use of a PROFET can reduce the electromagnetic emissions from a high side driver application.

The circuit in Figure 7.35 is similar to Figure 7.2. It illustrates how a high side driver can source significant current by using a gate voltage (25 V) higher than the supply voltage (14 V).

How do we get the higher gate voltage?

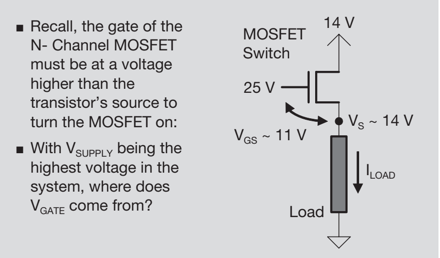


Figure 7.35 MOSFET High Side Drive

Figure 7.36 is a functional schematic diagram of the circuit used in modern charge pumps. It does, however, give adequate insight into the operation of the charge pump for our review here.

When Switch A and Switch B are both open, current flows from V_{SUPPLY} to V_{OUT} through the two diodes, D_A and D_B, to C_B. Therefore:

$$V_{CHP} = V_{SUPPLY} - V_{DA} - V_{DB}$$

For example:

$$V_{CHP} \sim 14\text{ V} - 1\text{ V} - 1\text{ V} = 12\text{ V}$$

(Often, in charge pump circuits, the diode drop is taken to be 1 V for simplicity).

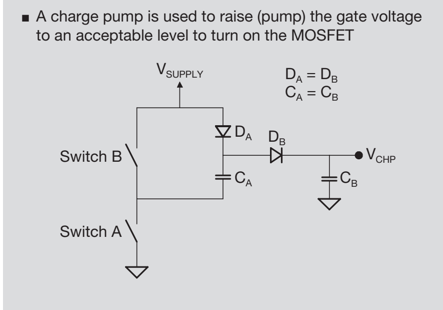


Figure 7.36 Charge Pump Gate Voltage

When Switch A is closed, capacitor C_A begins to charge (Figure 7.37).

Following the closure of Switch A, the capacitor is completely charged, and the voltage across the capacitor is:

$$V_{CA} = V_{SUPPLY} - V_{DA}$$

$$V_{CA} \sim 14\text{ V} - 1\text{ V} = 13\text{ V}$$

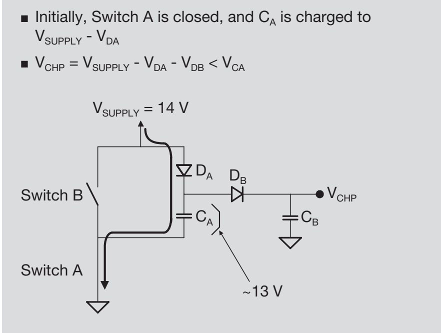


Figure 7.37 Charge Pump Gate Voltage

After C_A is completely charged, Switch A is opened and Switch B is closed (Figure 7.38).

As capacitor C_A is being discharged through diode D_B into capacitor C_B the output voltage (V_{CHP}) starts to increase. Charge then is transferred from capacitor C_A to capacitor C_B.

Remember, initially C_A has 13 V across it. It is acting as a small 13 V battery at the beginning of the charge transfer.

Since the negative terminal of capacitor C_A is step-forced to V_{SUPPLY} its positive terminal will also step to

initial $V_{CA} = V_{SUPPLY} - V_{DA} + V_{SUPPLY} =$
$= 2\ V_{SUPPLY} - V_{DA}$

Current flows from V_{SUPPLY} through capacitor (pseudo battery) C_A and diode D_B into capacitor C_B until $V_{CHP} + V_{DB} = V_{CA}$, because the capacitance of C_A and C_B are equal by assumption.

All the voltage stored on C_A have been added to what has been the initial voltage on C_B. Using the earlier numerical values:

V_{CB} = 12 V initial V_{CA} = 13 V, new V_{CB} = 25 V

The net result is 25 V at the output of the charge pump, our desired gate voltage.

Notice that all through the pumping process (Switch B is closed) diode D_A is reversed biased.

After the completion of charge transfer Switch B is opened and Switch A is closed. This allows capacitor C_A to be recharged to 13 V while capacitor C_B sources the necessary output current. (Diode D_B is reverse biased at this time, isolating capacitor C_B from capacitor C_A.) When C_A is again fully charged, Switch A is opened and Switch B is closed and the process repeats. Note that voltage values in the above description are valid if $C_A = C_B$.

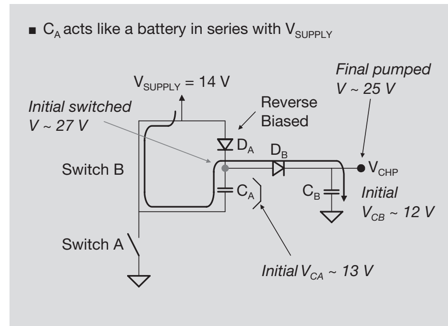


Figure 7.38 Charge Pump Gate Voltage

Switching high voltages and large currents, however, can lead to a new problem: electromagnetic emissions due to the fast switching on and off of the charge pump capacitors (charge/discharge currents).

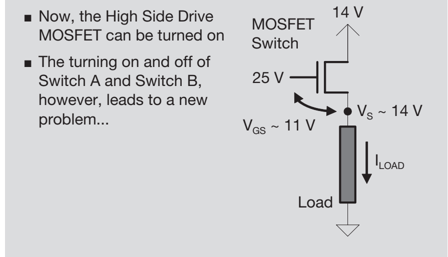


Figure 7.39 MOSFET High Side Drive

Figure 7.40 shows the electromagnetic emissions vs. frequency for an older charge pump design.

Various electromagnetic emission classifications (dashed lines) are shown for reference. In addition, an example specification from an original equipment manufacturer is shown (bold line).

Several points can be seen where the electromagnetic emissions exceed the allowed values.

Note, the discontinuity in the electromagnetic emissions at 30 MHz is due to the automated test system changing the resolution bandwidth on the spectrum analyzer. This in turn causes a change in the noise floor.

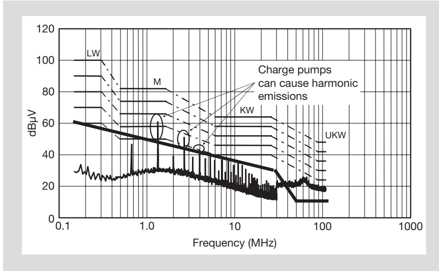


Figure 7.40 Charge Pump Electromagnetic Interference (EMI)

Figure 7.41 shows the electromagnetic emissions vs. frequency for an improved charge pump found in newer PROFETs.

With the new charge pump design, emissions can be reduced significantly.

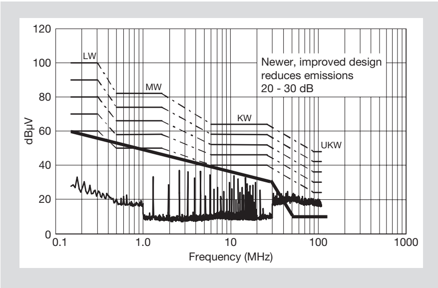


Figure 7.41 PROFET's Improved Charge Pump Reduces (EMI)

If additional suppression of electromagnetic emissions is required, external components can be added. In Figure 7.42, the reduced emission for a protected high side driver are shown with an additional resistor and capacitor.

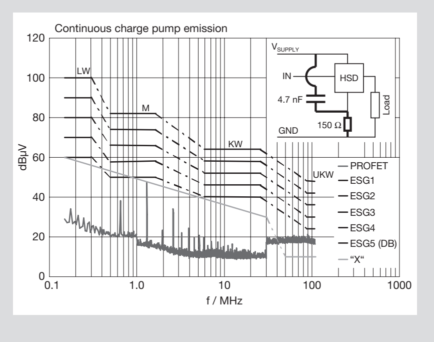


Figure 7.42 Charge Pump Filter Solutions for EMI Reductions

In Figure 7.43, another example of reducing electromagnetic emissions with external components is shown. Here, a larger capacitor is used.

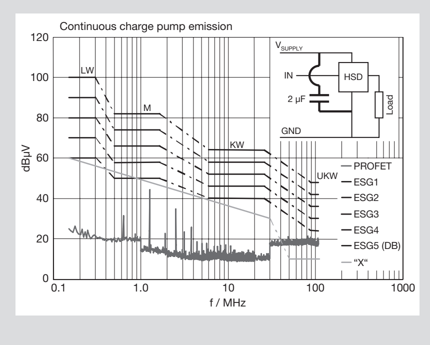


Figure 7.43 Charge Pump Filter Solutions for EMI Reductions

In addition to the charge pump, another source of electromagnetic emissions in protected high side drivers can be just the turning the driver on and off. When this turn-on and turn-off process is repeated, it is often called pulse width modulation (PWM).

PWM operation can be characterized by a number of different parameters. First, is frequency (f). This is the number of times that the turn-on and turn-off is repeated in each second.

Next is period (P). This is the length of time required to complete a cycle:

1) Turn-on

2) Time the driver is on

3) Turn-off

4) Time the driver is off

Note, the period is the reciprocal of the frequency: P = 1 / f.

Finally, duty cycle (D) is the percentage of the period that the driver is on.

Example: let us assume that a driver is being switched on and off 100 times per second with negligible turn-on and turn-off transition times. Each time the driver is turned on, it stays on for 2 ms before it is turned off for 8 ms.

Frequency (f) = 100 Hz

Period (P) = 1 / 100 Hz = 10 ms

Duty Cycle = 2 ms / (8 ms + 2 ms) = 20%

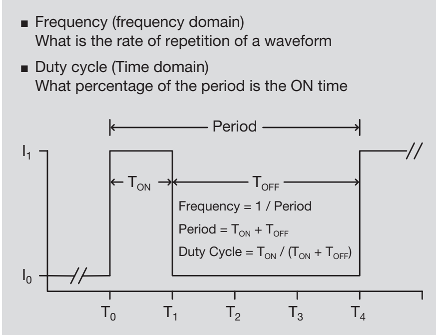


Figure 7.44 PWM Definitions

PWM operation causes increases in the electro-magnetic emission spectra because of the impulsive changes in the driver load current and output voltage.

Many protected high side drivers are used in relatively slow switching applications (below 100 Hz) resulting in an increase of the emission spectra below approximately 1 MHz.

To reduce the emitted energy protected high side drivers "shape" the output current waveforms edges at the beginning and end of the turn-on and turn-off process. The shaping of the transients reduces the electro-magnetic emission levels while still maintaining a slew rate at which the switching power losses are acceptable.

Figure 7.46, Example of Edge Shaping, illustrates the idea. Compromising between switching power loss and electromagnetic emissions, designers will determine an optimum slew rate which is characterized by the speed the output falls from its 90% value to its 10% value.

- Another source of emissions can be PWM operation
- The slew rate and shape of the output voltage and current waveforms cause an increase in the emission spectra
- For slow switching applications (below 100 Hz) this results in an increase of the emission spectra below approximately 1 MHz
- Edge shaping allows reduction of emission levels while maintaining a slew rate which still allows for acceptable switching power loss

Figure 7.45 EMI/EMC Emissions Due to Turn-On and Turn-Off

To provide the maximum theoretical reduction in electromagnetic emissions, the signal prior to the transition and immediately after the transition will exhibit a sinusoidal characteristics (light grey waveform - theoretical ideal). If no edge shaping is performed on the signal, there is a significant change in the slope of the signal turn-on and turn-off (dark gray waveform - slew control only). However, by implementing edge shaping (dashed black line), the instantaneous change in the signal's slope is dramatically reduced, decreasing the electromagnetic emissions.

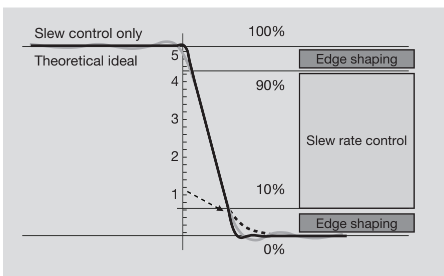


Figure 7.46 Example of Edge Shaping

The next several figures illustrate how the implementation of edge shaping reduces the electromagnetic emissions. In Figure 7.47 we see the emission spectra for a first generation protected high side driver which only implements a certain form of slew rate control.

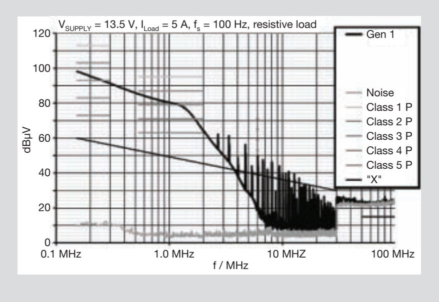


Figure 7.47 EMC Improvements in High Current PROFETs

In Figure 7.48 we see the emission spectra for a second generation protected high side driver which still only implements slew rate control. This time, however, the turn-off is slower, reducing emissions at the cost of higher switching power losses.

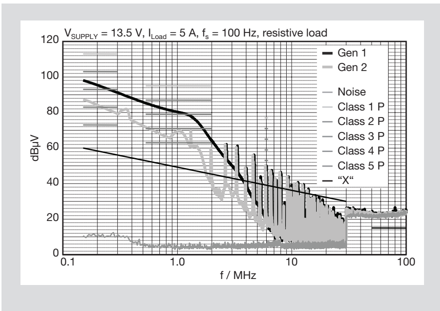


Figure 7.48 EMC Improvements in High Current PROFETs

Figure 7.49, is the emission spectra for a third generation protected high side driver. In addition to slew rate control, this device also features edge shaping circuitry to reduce the electromagnetic emissions at turn-off.

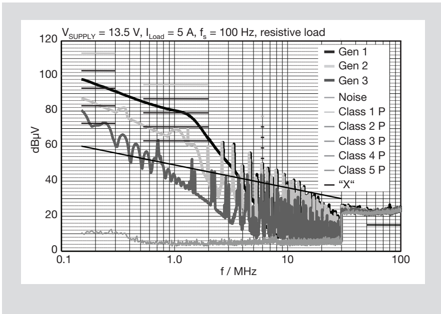


Figure 7.49 EMC Improvements in High Current PROFETs

Figure 7.50, is the final emission spectra for a fourth generation protected high side driver. In addition to slew rate control, it also features edge shaping circuitry to reduce the electromagnetic emissions at turn-off and turn-off.

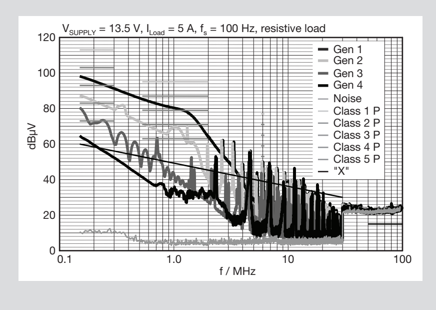


Figure 7.50 EMC Improvements in High Current PROFETs

7.5 System Implementation

We will now identify the external components used in PROFET high side drive circuits.

Figure 7.51 shows a voltage controlled, digital feedback PROFET. To protect the PROFET's logic functional block against over voltage conditions, several external resistors may be required.

R_{GND} limits current through DAZ
R_{ST} protects the microcontroller input pin
R_{IN} may be required to protect the microcontroller output pin

Appropriate values of the resistors can be found in each PROFET datasheet.

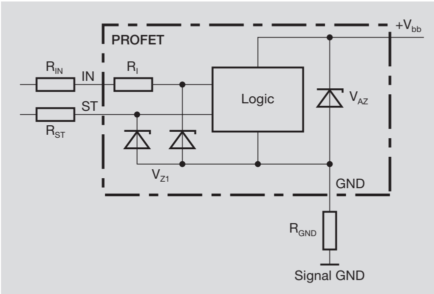


Figure 7.51 Over Voltage Protection of Logic Functional Block

Figure 7.52 shows a voltage controlled, digital feedback PROFET. To protect it from reverse battery conditions, external resistors may be required.

R_{GND} limits current through the logic Zener-diode
R_{ST} protects the microcontroller input pin
R_{IN} may be required to protect microcontroller output pin
R_L limits the current through the power inverse diode

Recommended values of the resistors can be found in each PROFET datasheet.

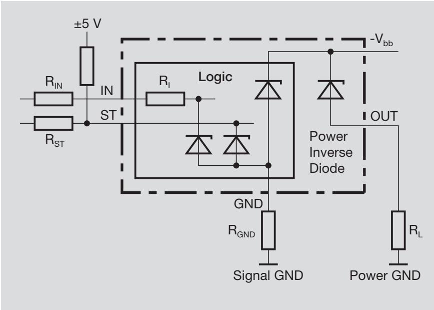


Figure 7.52 Reverse Battery Protection

Warning: the power dissipation during reverse battery conditions can be significantly higher than during normal operation. Let us take a look at an example:

$V_{SUPPLY} = 15$ V
$R_{LI} = R_L + R_{DSon} = 5\ \Omega$
$I_L = V_{SUPPLY} / R_L$
$I_L = 15\ V / 5\ \Omega = 3$ A

For a MOSFET with an R_{DSon} of 100 mΩ, the normal operating mode power dissipation is:

$P_D = I^2 R$
$P_D = (3\ A)^2(100\ m\Omega)$
$P_D = 0.9$ W

Under reverse battery conditions, the reverse current does not flow through a standard MOSFET channel. Rather, the path is through the power MOSFET's intrinsic body diode. Therefore, the power dissipation during reverse battery conditions is approximately given as:

$P_{D,REVERSE} = V\ I = V\ ((15\text{-}07) / (5\text{-}0.1)$
$P_{D,REVERSE} = (0.7\ V)(2.92\ A)$
$P_{D,REVERSE} = 2.04$ W

In order to significantly reduce the power dissipation of a protected high side driver under reverse battery conditions, the MOSFET should be turned-on and divert the reversed load current through its channel.

- Power dissipation during reverse battery can be higher than normal operation due to conduction of load current through the FET body diode
- For example:
  - 3 A load with 100 mΩ R_{DSon} MOSFET in normal mode gives 0.9 W
  - 3 A load thru body diode in reverse battery gives 2.1 W (3 A*0.7 V)
- This leads us to a feature where the MOSFET channel can be turned on during reverse battery operation---ReverSave™

Figure 7.53 Reverse Battery Power Dissipation

Figure 7.54 shows a current controlled, analog feedback PROFET. About 100 mA of current must flow through R_{bb} (from the **IN** or **STATUS** pins) to turn on the MOSFET in inverse mode. Currents above 100 mA in R_{bb} may create excessive power dissipation. R_{IS} will limit the current flowing into the current sense pin and through the R_{bb}.

When using a PROFET with ReverSave™ reverse battery protection, it may be necessary to use a resistor (R_{IN}) to limit the current through R_{bb}, depending on the reverse voltage specification. Appropriate values of the resistors can be found in each PROFET datasheet.

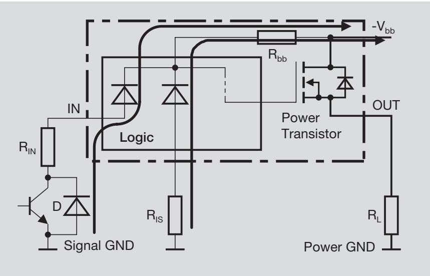


Figure 7.54 ReverSave™ Reverse Battery Protection Circuity

Figure 7.55 shows a current controlled, analog feedback PROFET.

Under large over-voltage (or load dump) conditions (more than 67 V, typical), the voltage dropped across the sense resistor R_{IS} may exceed the safe operating input voltage of a microcontroller I/O pin.

Therefore, the I_S pin can be clamped by an external diode to $V_{Z,VIS}$ if necessary.

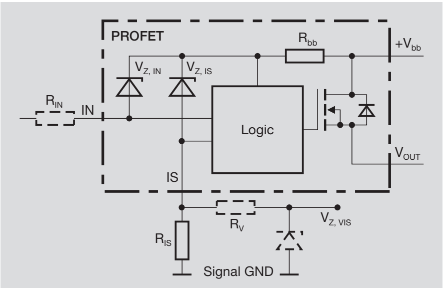


Figure 7.55 IS Pin Over Voltage Protection

7.6 Frequently Asked Questions

Finally, we will look at the questions the designers of any system must ask (and answer) when designing with high side drivers.

- How many channels?
- What is the load current?
- Is the load capacitive and what is the in-rush current?
- Is the load inductive, if yes, what is the inductance and what is the fly-back energy during switch-off?
- Will the load be driven on/off or PWM? What is PWM frequency?
- What is ambient temperature?
- What type of package surface mount or through-hole?
- If surface mount, how much copper area for V_{bb} / tab connection?
- If through-hole, what type of heatsink will be provided for package?
- What diagnostics are needed?
- What application extremes will the device / system be subjected to (reverse battery, load dump, overvoltage,etc.)?

Figure 7.56 PROFET Selection: Customer Questions

Figure 7.56 lists the questions a system designer needs to answer to select an appropriate PROFET high side driver.

Many of these questions are self explanatory, but we will address a few in more detail in the remainder of this chapter.

One of the key parameters of a system when choosing a PROFET high side driver is the expected load current. It is important to know the typical load current, the maximum load current, and how long the PROFET may need to source the maximum load current.

- What is the maximum load current?
- When does the maximum occur?
- What is the typical load current?
- Alternative Question: What is the load resistance?
- Alternative Question: If the load is a lamp, what is it's wattage?
- Recall, the load current is fundamental in determining an appropriate PROFET R_{DSon} value

Figure 7.57 What is the Load Current?

Some times the load current is not directly available. Rather, the characteristics of the load (its resistance or wattage) or the types of the load are given. From these characteristics, the load current can be determined. Based upon these questions, an appropriate R_{DSon} value can be determined (Figure 7.57).

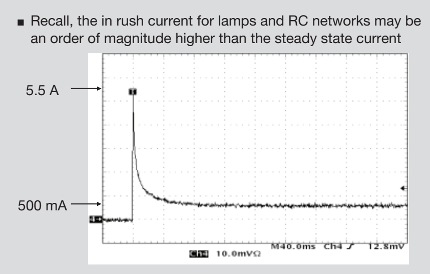


Figure 7.58 Is the Load Capacitive? What is the In-rush Current?

The in-rush current for a light bulb load is shown graphically in Figure 7.58. This is an actual oscilloscope trace of the current flowing through an initially cold filament of a light bulb. The time step in the waveform is 40 ms/division.

The current peaks almost immediately at approximately 5.5 A. As the filament warms (in about 80 ms), its resistance increases and the load current reaches its steady state value. It is important that the maximum current limit value of the chosen PROFET is above the peak load current value (5.5 A) to prevent the PROFET to operate in current limiting mode.

Caution has to be taken when designing any high side driver circuit with an inductive load. While MOSFETs and PROFETs are specified for a maximum load inductance to sustain the turn off energy dissipation, PROFETs have additional circuitry to increase their robustness (e.g. reliability).

- FETs are rated for the max absorbable energy when turning off inductive loads

Maximum Ratings at T_j = 25 °C unless otherwise specified

Parameter	Symbol	Values	Unit
Inductive load switch-off energy dissipation [3] single pulse I_L = 20 A, V_{bb} = 12 V T_j = 150 °C:	E_{AS}	0.3	J

Energy stored in load inductance:

$$E_L = \frac{1}{2} \cdot L \cdot I_L^2$$

While demagnetizing load inductnce, the energy dissipated in PROFET is

$$E_{AS} = E_{bb} + E_L - E_R = V_{ON(CL)} \cdot i_L(t)\ dt,$$

with an approximate solution for $R_L > 0\ \Omega$:

$$E_{AS} = \frac{I_L \cdot L}{2 \cdot R_L} \left(V_{bb} + |V_{OUT(CL)}|\right) \ln\left(1 + \frac{I_L \cdot R_L}{|V_{OUT(CL)}|}\right)$$

Figure 7.59 What is Load Inductance or Energy During Turn-Off?

PROFETs that are simply turned on or off the turn-on, turn-off transient power dissipations are negligible compared to the steady state on dissipation. If the load is driven by a pulse width modulated (PWM) signal, however, switching losses during the switching transitions must be considered. This is particularly true of lighting applications. The duty cycle and PWM frequency are important to know when selecting any high side driver.

Figure 7.60 illustrates the typical automobile loads driven by PROFET-s. Inductive load(s) is (are) relay(s) driven ON and OFF most likely in non-periodic fashion.

The majority of the loads are lamp loads used both in ON-OFF and PWM modes. The in-rush current amplitude in lamps with warmed up filaments and driven in PWM mode is small to insignificant depending on the PWM repetition rate and the average filament temperature. In general PWM lamp drive is used to control the lamp voltage so, that the bulb life is lengthened.

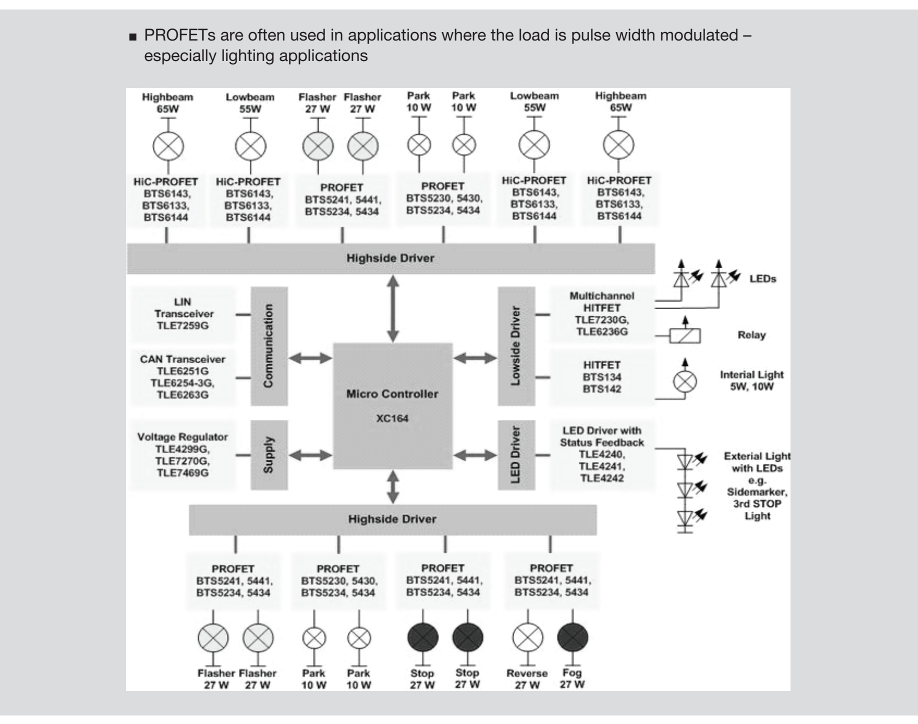


Figure 7.60 Will the Load be On/Off or PWM? What is PWM Frequency?

The maximum ambient temperature is also important when selecting a PROFET R_{DSon} and determining the heat sinking requirements. Higher ambient temperatures will require lower R_{DSon} values and/or additional heatsinking.

- Minimum ambient temperatures is usually -40 °C
- Maximum ambient temperature ranges from 85 °C to 125 °C for most applications:
  85 °C for most non-powertrain applications
  105 °C for some in-dashboard applications
  125 °C for most powertrain applications

Figure 7.61 What is the Ambient Temperature?

Traditional control module manufacturing allowed for the use of through hole components, rather than restricting a system designer to surface mount devices. Heatsinking options are limited with surface mount components. In modern modules surface mount components are preferred which are cooled either by increased Cu-foil surface or heat piping.

- Many applications require all surface mount components
- Surface mount components typically only have excess copper board space heatsinks
- Through-hole components can have large heatsinks for improved power dissipation

Figure 7.62 What Type of Package? Surface Mount or Through-hole?

The thermal resistance of four different surface mount PROFET packages are presented in Figure 7.63:

The graphs are based upon the size of the board space heatsink (copper area).

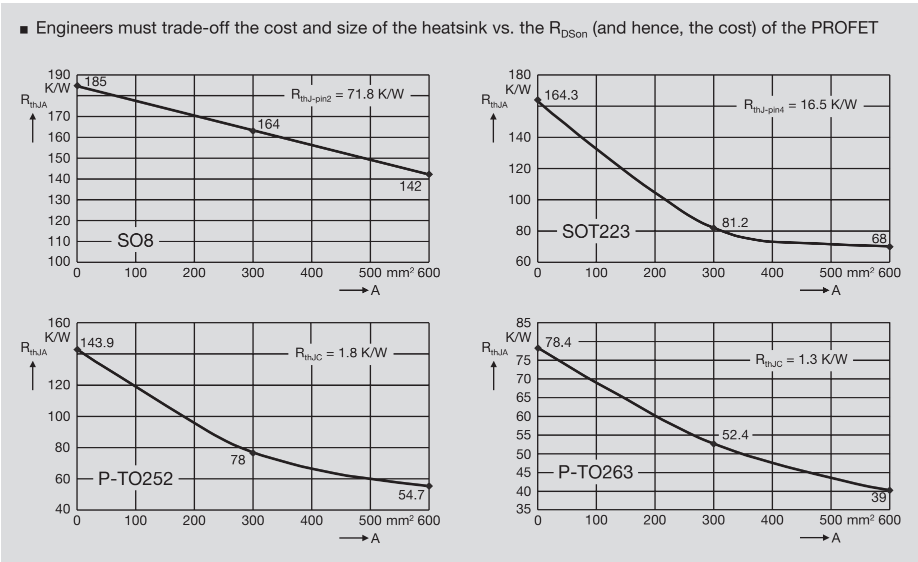


Figure 7.63 If Surface Mount - How Much Board Area is Available for Heatsinks?

Finally, with everything else being equal, there is a cost difference between PROFETs: those with advanced diagnostics will be more expensive than the ones without diagnostics.

- No diagnostics
- General Status (Error Flag)
- General Status and Current Feedback

Figure 7.64 What PROFET Diagnostics are Available?

8. Protected Low Side Drivers

In this chapter, we introduce the concept of "protected" MOSFET low side drivers (HITFETs).

The chapter starts with a quick definition of low-side switches and the description of the differences between protected MOSFET low side drivers and standard transistors. This leads into an introduction to the various types of protection that are available in protected low side drivers. A discussion of the various diagnostic options available in protected MOSFET low side drivers is also included.

The easy application of protected MOSFET low side drivers is demonstrated. We will examine what **E**lectro**m**agnetic **I**nterference (EMI) issues may arise and how they can be minimized. Suggestions are provided how the functionality of the protected MOSFET low side driver can be improved with the addition of a few external components. The chapter ends with a review of the questions a system designer should ask when implementing a protected MOSFET low side driver in an application.

8.1 What is a Protected MOSFET Low Side Driver?

In a low side driver configuration, a MOSFET switch is connected between the load and ground – the switch is on the "low" side of the load (Figure 8.1).

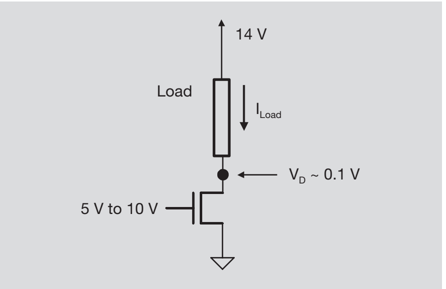


Figure 8.1 Low Side Drive (LSD) Configuration

The reader is advised at this point to review Chapter 2, 3, 6 and 7.

The fundamental difference between low side and high side drivers is that the latter does require the boosted gate voltage.

Like protected high side drivers, protected low side drivers are available to simplify system designs. For example, the **H**igh **I**ntegration, **T**emperature protected **F**ield **E**ffect **T**ransistor (HITFET), is a standard MOSFET with additional features that have been integrated to improve its performance and protection in low side driver applications (Figure 8.2).

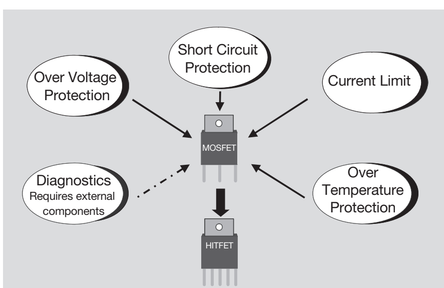


Figure 8.2 HITFET = High Integration Temperature Protected FET

Figure 8.3 and Figure 8.4 show a die layout and a block diagram of a typical HITFET. Notice the vertical, n-channel MOSFET and its integrated body diode (the body diode is created by junction of the p^+ wells at the source and the n-type drain). Vertical MOSFET is used to minimize the R_{DSon} of the MOSFET switch.

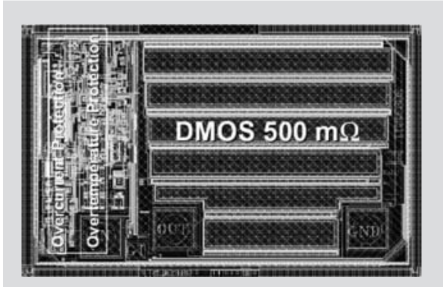


Figure 8.3 HITFET Die Lay-out

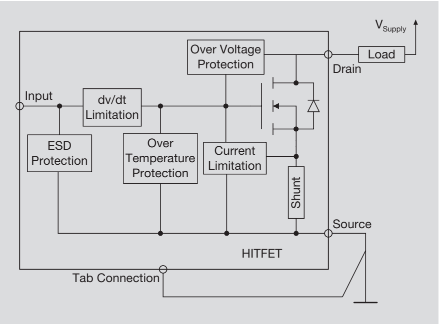


Figure 8.4 High Integration Temperature Protected FET

Left of the vertical power MOSFET (Figure 8.3) are the various functional blocks and protection circuits implemented in CMOS technology, including the ESD protection provided on the logic input (Figure 8.4).

8.2 Protection Features

Standard MOSFET-s are often considered "rugged" devices. They are physically large, single transistor components manufactured in a rather simple fabrication process. However, they are unprotected components - meaning that if the device is subjected to severe fault conditions, the MOSFET is often not able to

protect itself or its load from catastrophic failures.

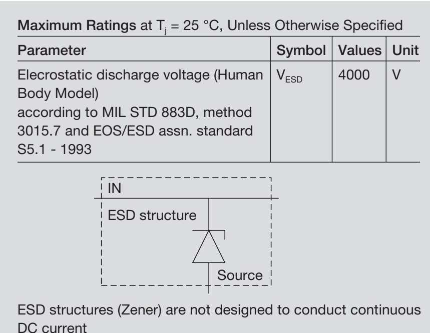
HITFET-s improve upon MOSFET devices by incorporating additional protection circuits implemented in a fabrication process offering CMOS and DMOS (**D**ouble Diffused **MOS**) components. Therefore, if the HITFET is subjected to a fault condition, the device can protect itself and its load.

The following is a list of the available protection features in Infineon's HITFET portfolio (Figure 8.5). Next, we will briefly examine each of these protection features.

- Electrostatic Discharge (ESD) Protection
- Load Dump Tolerant
- Inductive and Over voltage Output Clamp Protection
- Current Limit Protection
- Thermal Shutdown Protection

Figure 8.5 HITFET Protection Features

First the HITFET input electrostatic discharge (ESD) protection is considered. Figure 8.6 is an example of the ESD protection specification from a HITFET datasheet. Note, the conditions for the ESD test are shown (a human body model test is identified in the example).



Maximum Ratings at T_j = 25 °C, Unless Otherwise Specified

Parameter	Symbol	Values	Unit
Elecrostatic discharge voltage (Human Body Model) according to MIL STD 883D, method 3015.7 and EOS/ESD assn. standard S5.1 - 1993	V_{ESD}	4000	V

ESD structures (Zener) are not designed to conduct continuous DC current

Figure 8.6 ESD Protection

Next, HITFET-s are also protected against transient over voltages during load dump events up to a certain energy level. Figure 8.7 is an example of the load dump protection specification from a HITFET datasheet. In this example, values are shown for the load dump protection with a resistive load.

As the load resistance increases, more of the energy in the load dump is dissipated in the load itself and less energy is dissipated in the HITFET. This allows HITFET-s with larger resistive loads to withstand higher load dump voltage transients.

V_{IN}	V_{BAT}	V_{LD} 1)	Load	Pulse Para-meters 1)	Pulse Type 1)
V_{IN} = 0 V and V_{IN} = 10 V	13.5 V	V_{BAT} + V_{PULSE} = 65 V	R_L = 4.5 Ω	t_{PULSE} = 400 ms R_{INT} = 2 Ω 2)	Exponential + DC offset

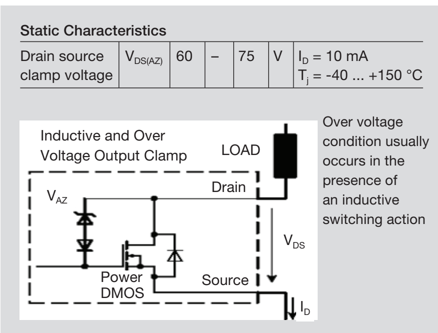
1) Pulse generator set up as per ISO-7637-1
2) Internal impedance of pulse generator

Figure 8.7 Load Dump Protection (BTS3134D)

HITFET-s are also protected when inductive loads are turned off by clamping the drain-source voltage (V_{DS}) to a safe value (see Chapter 1). Without the integrated diode clamp in the protected MOSFET, the inductive turn-off transient, V_{DS} would continue to increase until the MOSFET went into avalanche breakdown. The avalanche process itself is not necessarily destructive.

In avalanche mode excessive power dissipation may occur which can permanently damage the low side driver. During avalanche breakdown, the inductor current would still decay, but the large voltage across the MOSFET will result in significant power being dissipated within the MOSFET. Therefore, if V_{DS} is clamped to a safe value, the power dissipated in the MOSFET can be limited to a predictable level.

When V_{DS} increases as a result of turning the current-off that flows in an inductive load, the inductive switch-off transient causes V_{DS} to increase to the specified voltage, $V_{DS(AZ)}$, then the clamping diode will pull the gate voltage positively so, that the HITFET is turned on again providing a current path for the current stored in the inductive load (Figure 8.8). The HITFET stays turned on until the current decays to 0 A (see Chapter 1).



Static Characteristics

Drain source clamp voltage	$V_{DS(AZ)}$	60	–	75	V	I_D = 10 mA T_j = -40 ... +150 °C

Figure 8.8 Inductive and Over Voltage Clamp

HITFETs can be protected from a variety of problems with thermal shutdown protection. The thermal shutdown of a HITFET may begin at 150 °C. The HITFET is guaranteed to operate normally up to 150 °C. Above 150 °C (175 °C typically), however, the thermal shutdown will protect the HITFET by turning the device completely off (Figure 8.9).

To prevent a HITFET from oscillating into and out of thermal shutdown, hysteresis is included on the thermal shutdown. In this example, the typical value of the hysteresis is 10 K. Therefore, if a HITFET's thermal shutdown was activated at 155 °C, the HITFET would typically turn itself back on when the junction temperature fell to 145 °C (Figure 8.9). Note that 1 K temperature differential is equal to a 1 °C differential).

Figure 8.9 also shows a series of waveforms showing the operation of a HITFET with thermal shutdown protection:

Time: A — Input Voltage: High
Junction Temperature: $< T_{jt}$
HITFET Status: On

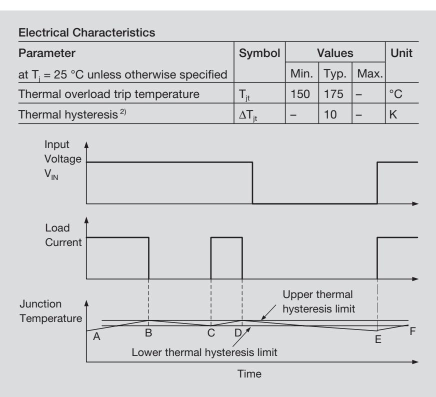
Time: B — Input Voltage: High
Junction Temperature: $> T_{jt}$
HITFET Status: Off

Time: C — Input Voltage: High
Junction Temperature: $< T_{jt}$
HITFET Status: On

Time: D — Input Voltage: High
Junction Temperature: $> T_{jt}$
HITFET Status: Off

Time: E — Input Voltage: High
Junction Temperature: $< T_{jt}$
HITFET Status: On

Time: F — Input Voltage: High
Junction Temperature: $< T_{jt}$
HITFET Status: On



Electrical Characteristics

Parameter at T_j = 25 °C unless otherwise specified	Symbol	Values Min.	Typ.	Max.	Unit
Thermal overload trip temperature	T_{jt}	150	175	–	°C
Thermal hysteresis [2]	ΔT_{jt}	–	10	–	K

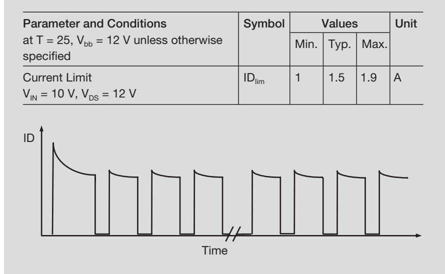
Figure 8.9 Thermal Shutdown Protection

HITFET-s also have a protection feature called current limiting.

With current limit protection, a HITFET can limit the maximum current delivered to a load. Therefore, if the load resistance is less than expected (0 Ω with a direct short to V_{SUPPLY}), the HITFET will ensure that its sourced current does not exceed the specified limited current value (Figure 8.10).

In this example, the minimum current limit specified is 1 A. This means the HITFET is able to supply at least 1 A before it begins to limit the drain current. However, if an application requires 1.1 A load current under worst case conditions, this HITFET would not work. Some devices may limit the current to 1 A before the 1.1 A worst case load current is reached.

The maximum current limit in the example HITFET specification is 1.9 A. This device will not allow load currents of more than 1.9 A. It will limit the load current below or at the 1.9 A maximum threshold.



Parameter and Conditions at T = 25, V_{bb} = 12 V unless otherwise specified	Symbol	Values			Unit
		Min.	Typ.	Max.	
Current Limit V_{IN} = 10 V, V_{DS} = 12 V	ID_{lim}	1	1.5	1.9	A

Figure 8.10 Current Limit Protection

Some HITFET-s have an additional thermal shutdown feature. If they go into thermal shutdown, they will not automatically restart themselves once their junction temperature has cooled. The thermal shut-down protection is said to latch off this type of HITFET.

This is a desirable feature for many applications. Repeated over temperature conditions can potentially damage an integrated circuit (this is why testing for thermal cycling during semiconductor component qualification is needed).

If a short circuit condition were to occur, the part would heat up, go into thermal protection mode and shutdown. Without the latched-off protection the protected low side driver would endlessly turn itself off and on again as determined by the thermal hysteresis band in its internal control circuit. This would continue for as long as the fault condition existed, or until a controlling component (like a microcontroller) turned off the low side driver.

Figure 8.11 is a block diagram of a HITFET with a latching-off thermal shutdown protection circuit.

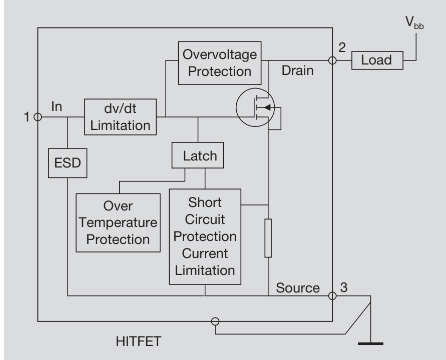


Figure 8.11 Thermal Latch

Example waveforms are shown in Figure 8.12. Note that the input current of the HITFET increases when it is latched off. This can be used to poll the HITFETs operation (more about that in subsequent pages).

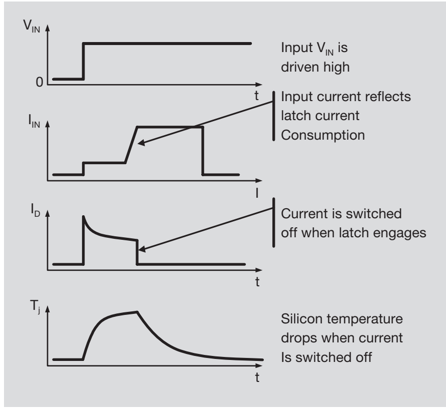


Figure 8.12 Thermal Shutdown Latch Behavior

8.3 Diagnostic Features

HITFET-s do not have a diagnostic feedback pin or STATUS output. Rather, diagnostic feedback can be obtained by monitoring the operation of the HITFET with the addition of a few external components (Figure 8.13).

- HITFETs do not contain internal structures that are specifically intended for diagnostic feedback
- Diagnostic feedback can be obtained by using external Components in conjunction with:
  – Drain - high current output pin - true conductive state of device can be estimated by monitoring the drain voltage
  – I_{IN} - input drive bias can be detected and evaluated against over current or thermal shutdown bias current - this method requires that input current be evaluated.
- NOTE: Care must be taken so as not to significantly reduce the available V_{IN} voltage

Figure 8.13 HITFET Diagnostic Feedback

There are two different ways. Either the voltage on the drain of the HITFET or the input current flowing into the input pin of the HITFET are monitored to determine if the low side driver is functioning as expected.

A circuit for monitoring the drain voltage of the HITFET is shown in Figure 8.14. Attenuation of V_{DS} is needed in most cases because the maximum input voltage of a microcontroller port is usually less than the drain voltage of the HITFET.

This method is simple to implement, but it takes up one analog-to-digital microcontroller input channel.

- Diagnostic feedback is sensed at the drain and applied to a microprocessor Analog to Digital Input.
- Advantage
  Low cost diagnostic and fault correction
- Disadvantages --
  Slower than integrated solution
  Uses one A/D channel per sensed drain

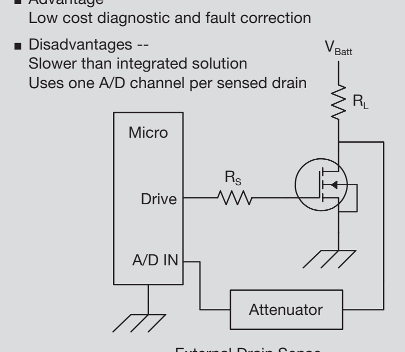


Figure 8.14 HITFET Diagnostic Feedback: Drain Voltage Sensing

The other way to obtain diagnostic feedback on a HITFET is to monitor the input current (see Figure 8.15). Previously, it was shown that the input current increases significantly when a HITFET is in fault mode. If a resistor is placed in series with the input of the HITFET', an analog-to-digital converter input can be used to monitor the voltage drop across the resistor.

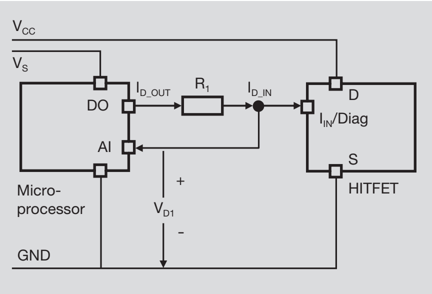


Figure 8.15 HITFET Diagnostic Feedback I_{IN} Current Monitor

8.4 EMI / EMC Considerations

The use of a HITFET can reduce the electromagnetic emissions from a low side driver as discussed next.

When a low side driver is turned on and off, it can result in significant electromagnetic emissions. Before looking at the electromagnetic emissions let us consider some of the parameters of low side drivers.

When the turn-on and turn-off process is repeated, it is often called pulse width modulation (PWM). PWM operation is characterized by different parameters. One such parameter is the repetition frequency (f), or the number of times per second that the driver is turned on and off. The reciprocal of f is the period (P). The period consists of the following parts:

1) Turn-on time (transition from off to on)

2) Time the driver is on

3) Turn-off time (transition from on to off)

4) Time the driver is off

Finally, duty cycle (D) is the percentage of the period that the driver is on.

For example, let us assume that a driver is being switched on and off 100 times per second (Figure 8.16). We will assume that the turn-on and turn-off times are negligible. Each time that the driver is turned on, it stays on for 8 ms before it is turned off for 2 ms.

Frequency (f) = 100 Hz
Period (P) = 1/100 Hz = 10 ms
Duty Cycle = 8 ms / (8 ms + 2 ms) = 80%

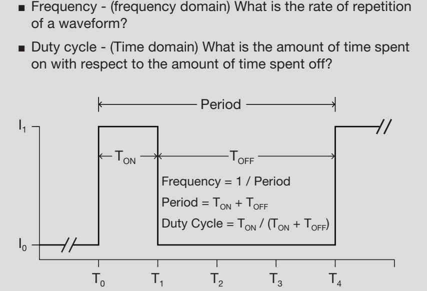


Figure 8.16 PWM Definitions

Every time a low side driver turns on or off, it generates electromagnetic emissions. When a driver is used in PWM applications it may have serious repercussions.

The emissions are dependent on how fast the low side driver turns-on and turns-off. If the turn-on and turn-off are very fast, the emissions are more intense. Therefore, HITFET-s slow down the rate of their turn-on and turn-off. This feature reduces the electromagnetic emissions in HITFET low side driver applications compared to traditional MOSFET drivers.

Figure 8.17, is an example of the slew rate control exhibited by HITFET low side drivers. It lists both the time and slew rate values for turn-on and turn-off and also the maximum allowed drain-to-source voltage (V_{DS}) slew rate.

Dynamic characteristics	Param	Min.	Nom.	Max.	Units
Turn-On time V_{IN} to 90% I_D R_L = 22 Ω, V_{IN} = 0 to 10 V, V_{bb} = 12 V	T_{ON}	–	10	10	µs
Turn-Off time V_{IN} to 10% I_D R_L = 22 Ω, V_{IN} = 10 to 0 V, V_{bb} = 12 V	T_{OFF}	–	10	20	µs
Slew rate On 70% to 50% V_{bb} R_L = 22 Ω, V_{IN} = 0 to 10 V, V_{bb} = 12 V	$-dV_{DS}/dt_{ON}$	–	4	10	µs
Slew rate Off 50% to 70% V_{bb} R_L = 22 Ω, V_{IN} = 10 to 0 V, V_{bb} = 12 V	dV_{DS}/dt_{OFF}	–	4	10	µs

Figure 8.17 HITFET Turn-On / Turn-Off Slew Rate Controlled

For reference, Figure 8.18 illustrates the terms used in the turn-on and turn-off times and slew rate specifications.

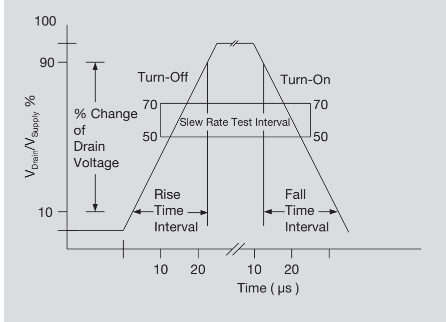


Figure 8.18 HITFET Turn-On / Turn-Off Slew Rate Control

An example of a HITFET's electromagnetic emission spectra (with slew rate control) is shown in Figure 8.19. Note the discontinuity in the graph of electromagnetic emissions at 30 MHz caused by the automated test system changing the resolution bandwidth on the spectrum analyzer. This in turn causes a change in the noise floor.

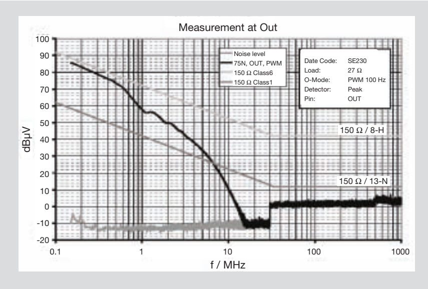


Figure 8.19 Typical HITFET Radiated Emissions Evaluation

8.5 System Implementation

The external components used in HITFET low side driver circuits are identified next.

We begin with the problem of reverse battery faults. When a HITFET is subjected to a reverse battery condition, current flows from the source pin, through the intrinsic body diode, through the drain pin, and into the load. The MOSFET (and its thermal protection) is off during reverse battery conditions and the low resistance channel does not conduct. This can lead to higher levels of power dissipation than in normal (forward) operating modes. The reverse battery current is limited only by the HITFET's load.

- Reverse load current through the intrinsic drain diode in series with the load
- Power dissipation is Higher compared to normal operating conditions due to the voltage drop across the drain to source diode
- Source diode current is limited by the load

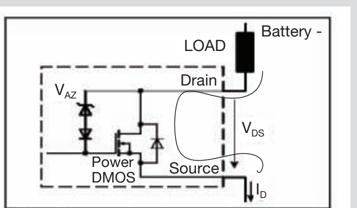


Figure 8.20 Reverse Battery Tolerance

Consider the following example. Recall the basic DC thermal equation:

$$T_{J,Max} - T_{A,Max} = P_{D,Max}\, R_{thJA}$$

For forward operation, the dissipated power is given by:

$$P_D = I_{Load}^2\, R_{DSon}$$

Inserting into the DC thermal equation and solving for $I_{Load,Max}$ results in:

$$I_{Load,Max} = ((T_{J,Max} - T_{A,Max}) / (R_{DSon} * R_{thJA}))^{1/2}$$

For the following values:

$T_{J,Max}$ = 150 °C
$T_{A,Max}$ = 95 °C
R_{DSon} = 0.068 Ω
R_{thJA} = 55 °C/W
$V_{Battery}$ = 14 V

The maximum forward operating current is:

$$I_{Load,Max} = 3.8\text{ A}$$

With a maximum allowed current, the maximum (normal operation) load resistance can be calculated:

$$V = I\, R$$

$$14\text{ V} = (3.8\text{ A})(R_{Load,Min} + 0.068\ \Omega)$$

$$R_{Load,Min} = 3.6\ \Omega$$

For reverse operation, the load current flows through the intrinsic body diode. The power dissipation in the HITFET is given by:

$$P_D = V_{Diode}\, I_L$$

Substituting into the DC thermal equation:

$$I_{L,Max} = (T_{J,Max} - T_{A,Max}) / (R_{thJA}\, V_{Diode})$$

With a diode voltage of 0.7 V, the maximum reverse battery current for this HITFET example is:

$$I_{L,Max} = 1.4\text{ A}$$

With a maximum allowed current, the maximum load resistance for reverse bias conditions can be calculated (Figure 8.20):

$$V = I\, R$$

$V_{Battery} = V_{Diode} + I_{Load} R_{Load}$

$14 V = 0.7 V + (1.4 A) R_{Load}$

$R_{Load,Min} = 9.5\ \Omega$

Therefore, if the HITFET is to survive and protect its load during reverse bias, a larger heatsink (lower R_{thJA}) or a larger load resistance is required. In summary, when designing for reverse battery conditions, the power dissipated in the intrinsic body diode will dominate the HITFET thermal design.

Another advantage of the HITFET protected low side drivers is that they can be used in high side driver applications. Let us look at how this works.

Figure 8.21 is an example of a high side drive application using an ordinary power MOSFET switch that is placed between the power connection (14 V here) and the load.

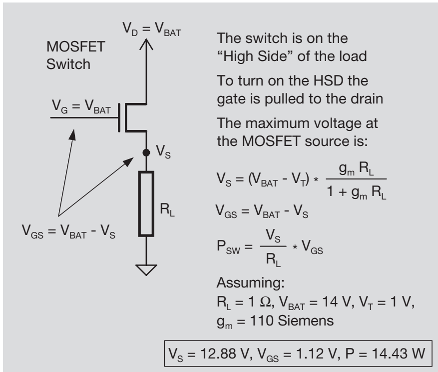


Figure 8.21 High Side Drive (HSD) Configuration

As before, to turn MOSFET on in the high side drive configuration, the gate voltage is pulled high. The source voltage will determine the load (or drain current ID) current:
$I_D = V_S / R_L$.
The load current approximately is given by
$I_D = (V_{GS} - V_T) * gm$.
If the maximum gate voltage is V_{BAT},
$maxV_{GS} = V_{BAT} - V_S$.
Eliminating I_D from the equations yields:

$V_{GS} = V_{DS} = (V_{BAT} - V_T) * gm * R_L / (1 + gm * R_L)$.

In this case V_{GS} is much greater than $I_D * R_{DSon}$ resulting in increased dissipation.

Letting $V_{BAT} = 14$ V, $R_L = 1\ \Omega$, gm = 110 Ω^{-1}

$V_{GS} = V_{DS} = 1.12$ V (based on BTS3160D).

$I_D = V_S / R_L = 12.88 / 1 = 12.88$ A,

$P_D = V_{GS} * I_D = 1.12 * 12.88 = 14.4$ W

The dissipation is unacceptably high. It can be reduced by boosting the gate voltage

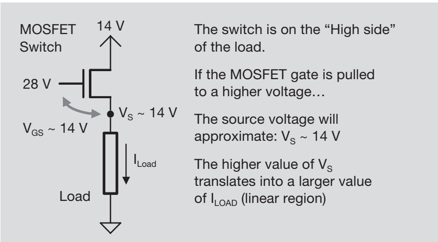


Figure 8.22 High Side Drive (HSD) Configuration

When a higher voltage is applied to the gate, like 28 V, two things change from our previous example. First, the MOSFET is fully enhanced ($V_{GS} >> V_T$), the source voltage of the MOSFET rises to almost 14 V (V_{BAT}) – meaning that the voltage drop across the MOSFET switch is smaller. Second the dissipation in the MOSFET is greatly reduced. The source voltage is given by:
$V_S = V_{BAT} * R_L / (R_L + R_{DSon})$.

Since $V_D = V_{BAT}$

$V_{DS} = V_{BAT} * R_{DSon} / (R_L + R_{DSon})$

Given $R_L = 1\ \Omega$, and $R_{DSon} = 10\ m\Omega$

$V_{DS} = 0.14$ V! << 1.12 V

$V_S = V_{BAT} - V_{DS} = 14 - 1.12 = 13.86$,
$I_D = 13.86$ A

$P_D = 0.14 * 13.86 = 1.94$ W.

So, the fundamental difference in high side drive applications is that a higher voltage needs to be applied to the MOSFET gate than

in a low side application. An example of how this can be accomplished is shown in Figure 8.23.

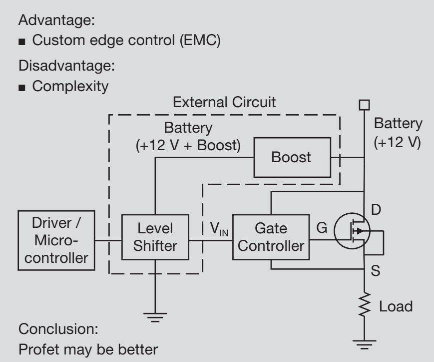


Figure 8.23 Source Follower

A voltage boosting circuit is used to increase the gate voltage above the supply voltage. A logic level signal from the microcontroller is then "shifted" to the new voltage range. A low signal (0 V) from the microcontroller remains 0 V, but a high signal (for example, 5 V) is shifted to the boosted gate voltage. This voltage is then applied to the HITFET gate.

The advantage of this topology is that the designer can implement any external boost circuit, thus customizing the turn-on and turn-off speeds to control the electromagnetic emissions. It does, however, require additional design work. For such applications, an integrated high side drive component may be more attractive.

The final topic in this Chapter is the various edge shaping options available for use in any low side driver application to reduce the electromagnetic emissions. As mentioned in Section 8.4, one source of electromagnetic emissions in low side driver applications can be turning the driver on and off during PWM operations.

Electromagnetic emissions are stimulated by the rate of change of drain current allowed by the switching device. Typically, quicker switching edges result in greater emissions. If the turn-on and turn-off are very fast, the emissions are worse. Therefore, HITFET-s slow down the rate of their turn-on and turn-off. The edge control reduces the electromagnetic emissions in HITFET low side driver applications compared to traditional MOSFET drivers. The electromagnetic emissions can be further reduced with the addition of external components.

First, we begin by noting that the turn-on and turn-off times can only be increased by the addition of the external components. The minimum HITFET slew rate is set internally (reducing harmful emissions), and it cannot further be reduced by adding external components. This option also allows for the implementation of symmetric or asymmetric turn-on and turn-off times. It may be necessary to turn a load off quickly, but the turn-on process can be relatively slow. This can be accomplished with external components while still providing a reduction in electromagnetic emissions. Finally, we must note that increasing the turn-on and turn-off times will, in turn, increase HITFET's switching losses. Therefore, additional care must be taken when performing the thermal analysis.

- Edge rise and fall time can only be increased by the addition of external components
- Slew rate can not be made faster by the addition of external components
- Potentially can modify EMC characteristics (Electro Magnetic emissions)
- Allows for symmetrical or asymmetrical adjustment to rise and fall times as well as slew rate modification
- Additional power is consumed by changing transition times (operation in linear region)

Figure 8.24 HITFET Edge Shaping

There are two ways to externally increase the turn-on and turn-off times:

1) Addition of a low pass filter to the HITFET input

2) Addition of a drain-to-gate feedback (Miller) capacitor

We will begin with the R-C low pass filter.

- Advantages
  – Simple in terms of calculating RC values
  – Is effective at controlling rise and fall time of the device
- Disadvantage
  – Adding a low pass filter to the input
  – Will insert a turn on delay and a turn off delay (dead time) which may modify the intent of PWM applications

Figure 8.25 HITFET Edge Shaping: Simple Low Pass Filter

There are several advantages for using an R-C low pass filter to perform the edge shaping (see Figure 8.26). It very effectively controls the HITFET rise and fall times, and it is simple to implement once the required R and C values are calculated. The associated disadvantages are rather straightforward. The addition of the low pass filter to the input will add a small delay to the turn-on and turn-off of the device in addition to decreasing the slew rate of the HITFET in PWM applications.

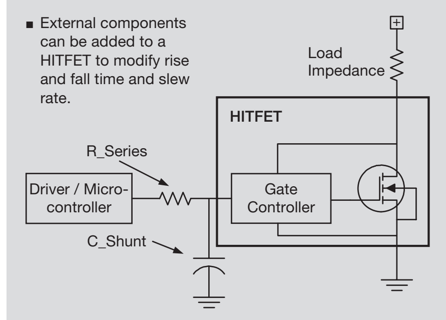


Figure 8.26 HITFET Edge Shaping: Simple Low Pass Filter

Figure 8.27 shows the effect of a simple low pass filter upon the HITFET's input voltage and drain voltage. Three curves are shown (for 1 nF, 100 nF, and 200 nF capacitors). As the capacitance is increased, the slew rate of the HITFET input and drain voltage is decreased, reducing the electromagnetic emissions.

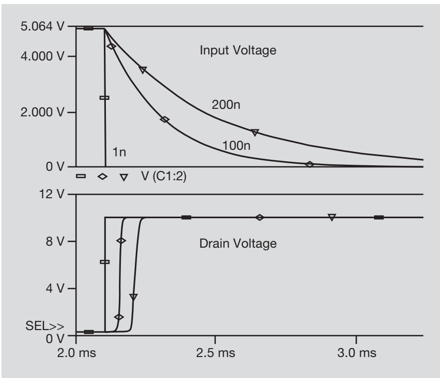


Figure 8.27 HITFET Edge Shaping: Simple Low Pass Filter

The second method of performing edge shaping in HITFET applications is with the addition of an external Miller capacitor. This solution is also effective in controlling the turn-on and turn-off times of the protected low side driver. In addition, it eliminates the turn-on and turn-off delay introduced with the simple low pass filter solution.

However, the calculations involved in selecting the correct Miller capacitance are slightly more complex. The designer must also now consider the load characteristics (resistance) when determining the optimum resistor and capacitor values.

- Advantages
  – Is effective at controlling rise and fall time of the device
  – Does not insert a significant turn-on or turn-off delay
- Disadvantage
  – Calculation of the RC components is more complicated
  – Must now also consider the resistance of the load

Figure 8.28 HITFET Edge Shaping: Miller Capacitor

A HITFET application with Miller capacitor edge shaping is shown in Figure 8.29.

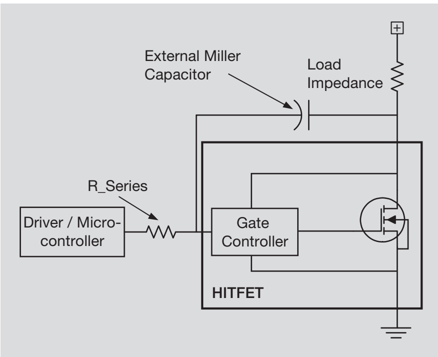


Figure 8.29 HITFET Edge Shaping: Miller Capacitor

In Figure 8.30 and Figure 8.31, we respectively show two examples of the use of Miller capacitors (25nF and 50nF) to perform edge shaping in a HITFET application during turn-on and turn-off. The example illustrates an additional difference when a Miller capacitor is added to perform the edge shaping. Unlike the simple low pass filter solution, using a Miller capacitor introduces an asymmetry in the turn-on and turn-off waveforms.

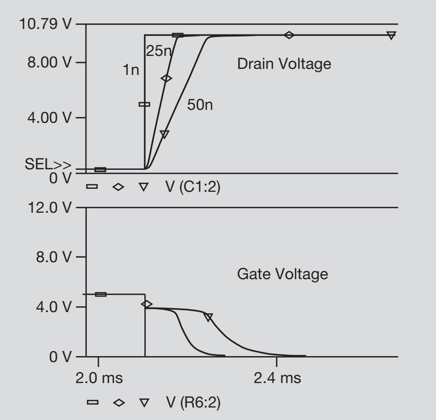


Figure 8.30 HITFET Edge Shaping: Miller Capacitor - Turn-On

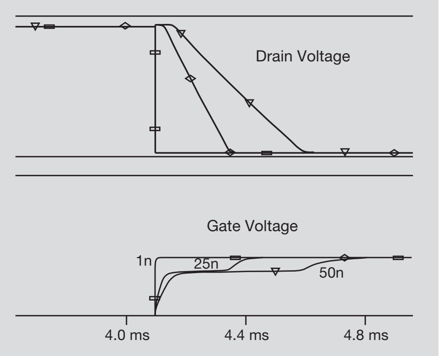


Figure 8.31 HITFET Edge Shaping: Miller Capacitor - Turn-Off

Often, the HITFET input will be driven directly by a microcontroller. In such applications, the HITFET input does not require additional protection. However, if the input is being driven by a non-regulated supply or from an off-board signal, additional precautions may be required.

Input current increases dramatically if the input is taken 0.2 V (or more) below ground. The same is true if the input voltage is pulled 10.0 V (or more) above the source voltage.

- Input protection usually not needed if driven directly from a microprocessor
- Input protection is needed in cases were drive is sourced from a non regulated or out boarded signal source

Test Condition	Parameter	Limit	Unit
Continuous input current	I_{IN}		mA
$-0.2\ V \leq V_{IN} \leq 10\ V$		self limited	
$V_{IN} < -0.2\ V$ or $V_{IN} > 10\ V$		$\lvert I_{IN}\rvert \leq 2$	

Figure 8.32 HITFET Input Protection

HITFET-s are intended to be driven by a 5 V (nominal) signal. In any event, the input voltage must not exceed the specified absolute maximum (typically 10 V) or absolute minimum (usually -0.3 V). In addition, the maximum allowed HITFET input current is 2 mA. This must be limited externally during fault conditions.

- The HITFET input drive circuitry must provide adequate voltage to the gate (4.5 V or more) and must not exceed the maximum allowable input voltage (typically 10 V).
- The maximum specified current allowed to sink or source from the HITFET in pin is 2.0 mA. Current up to 2 mA may be required to operate internal HITFET input protection circuitry.

Figure 8.33 HITFET Input Protection

It is possible, however, to increase the HITFET input voltage above the 5 V nominal value. Doing so will further enhance the drain-source conduction channel. This in turn will decrease the resistance of the HITFET switch and allow for a higher load current.

- HITFETs comply to a specification which uses a nominal 5 V_{IN} drive voltage as a specified operating point.
- Further device enhancement – lower R_{DSon} and higher output current may be achieved by operating the device at a higher V_{IN} voltage (10 V).

Figure 8.34 HITFET 5 V vs. 10 V Operation

This is shown graphically in Figure 8.35.

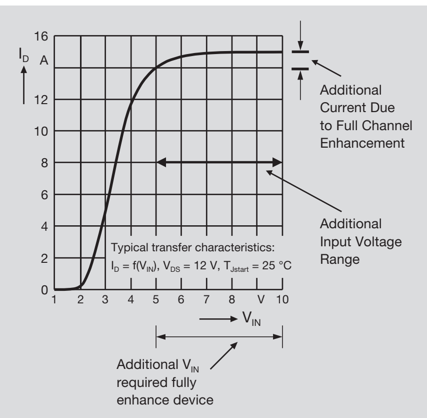


Figure 8.35 HITFET 5 V vs. 10 V Operation

8.6 Frequently Asked Questions

In Figure 8.36 and Figure 8.37 the questions are listed that a system designer needs to answer to select an appropriate protected low side driver.

Many of these questions are self explanatory, but we will address a few in more detail in the remaining pages of this chapter.

- What is the load current?
- Is the load capacitive and what is the inrush current?
- Is the load inductive and the inductance and/or energy during turn-off?
- Will load be on/off or PWM? What is PWM frequency (load states)?
- What is the ambient temperature?
- Can a HITFET be operated as a high side switch

Figure 8.36 Frequently Asked Questions

- What happens if ground (drain leg) opens
- What type of package - surface mount or through-hole?
- If surface mount, how much copper area for V_{bb} / tab connection?
- How is inductive energy evaluated and controlled by the HITFET
- If through-hole, what type of heat sink will be provided for package?
- What diagnostics are needed?
- What application extremes will the device / system be subjected to (reverse battery, load dump, over voltage etc.)?

Figure 8.37 Frequently Asked Questions

One of the key parameters of a system when choosing a HITFET protected low side driver is the expected load current. It is important to know the typical load current, the maximum load current, and how long the HITFET may need to source the maximum load current.

Some times the load current is not immediately available. Rather, the characteristics of the load (its resistance or wattage) are given. From these characteristics, the load current can be determined. Based upon these questions, an appropriate R_{DSon} value can be determined.

- What is the maximum load current?
- When does the maximum occur?
- What is the typical load current?
- Alternative Question: What is the load resistance?
- Alternative Question: If the load is a lamp, what is it's wattage?
- Recall, the load current is fundamental in determining the R_{DSon} value

Figure 8.38 What is the Load Current?

The in-rush current for a light bulb load is shown graphically in (Figure 8.39). An actual oscilloscope plot of the current flowing through an initially cold light bulb is shown in Figure 8.39. The time scale in the waveform is 40 ms/division. The current peaks almost immediately at approximately 5.5 A. As the filament warms (about 80 ms), however, its resistance increases, and the load current reaches a static value. It is important that the peak value (5.5 A) is below the low limit of the maximum current value of the chosen HITFET and does not cause the HITFET to go into current limiting.

- Recall, the in-rush current for lamps and RC networks may be an order of magnitude higher than the steady state current

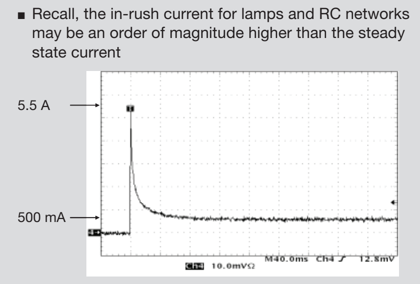


Figure 8.39 Is the Load Capacitive? What is the In-rush Current?

In addition, caution has to be taken when designing any low side driver circuit with inductive loads. While MOSFET-s and HITFET-s are specified for a maximum load inductance (or turn off energy dissipation), HITFET-s have additional circuitry to boost their robustness.

- MOSFETs are rated for the max absorbable energy when turning off inductive loads

Maximum Ratings at T_j = 25 °C unless otherwise specified

Parameter	Symbol	Values	Unit
Inductive load swich-off energy dissipation [3] single pulse I_L = 20 A, V_{bb} = 12 V T_j = 150 °C:	E_{AS}	0.3	J

$$E = \left(\frac{V_{cl}}{V_{cl} + V_{bat}}\right) * \left(\frac{1}{2} L * i_{pk}^2\right)$$

The equations relate the energy absorption capability regarding a *Single* pulse

Where:

L = load inductance

I_{pk} = short circuit load current

V_{cl} = over voltage clip voltage of HITFET

V_{bat} = supply voltage

Figure 8.40 Is the Load Inductive? Inductance and/or Energy During Turn-Off?

The maximum ambient temperature is also important when determining the appropriate HITFET R_{DSon} and heatsinking requirements. Higher ambient temperatures will require lower R_{DSon} values and/or additional heatsinking.

- Minimum ambient temperatures is usually -40 °C
- Maximum ambient temperature ranges from 85 °C to 125 °C for most applications:
  85 °C for most non-powertrain applications
  105 °C for some in-dashboard applications
  125 °C for most powertrain applications

Figure 8.41 What is the Ambient Temperature?

Traditional control module manufacturing allowed for the use of through-hole components, rather than restricting a system designer to surface mount devices. Heatsinking options are limited for system designers working with surface mount components.

- Many applications require all surface mount components
- Surface mount components typically only have excess copper board space heatsinks
- Through-hole components can have large heatsinks for improved power dissipation

Figure 8.42 What Type of Package? Surface Mount or Through-hole?

In Figure 8.43, we see an example of the thermal resistance of four different HITFET packages based upon the size of the board space heatsink.

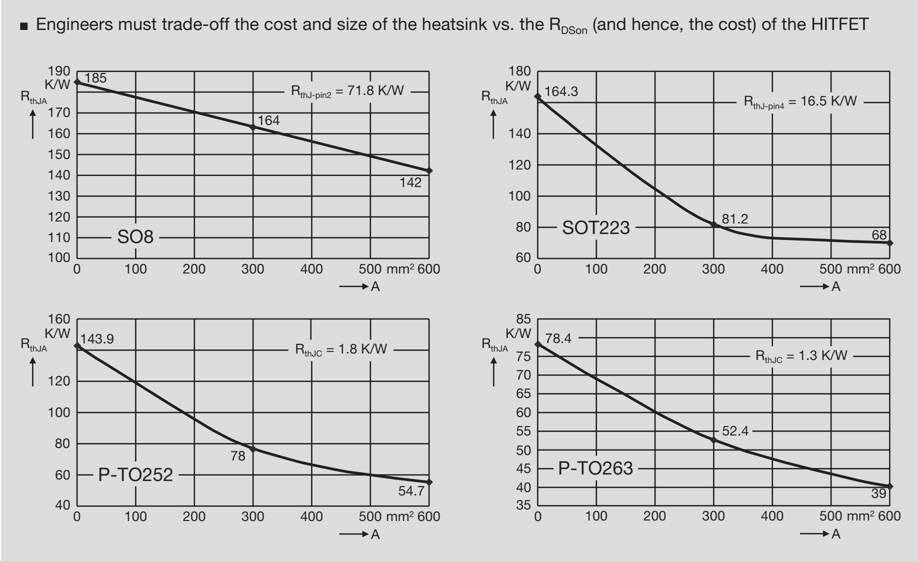


Figure 8.43 If Surface Mount - How Much Board Area is Available for Heatsinks?

9. Introduction to Linear Regulators

In this chapter, a review of electric power and the concept of voltage regulation are presented. First, some fundamental power concepts will be reviewed and answers to the simple question, “What is a linear regulator?” will be sought. Once fundamental concepts are clarified a general explanation will follow of how a linear regulator operates.

The basic concepts and operation of a linear regulator are illustrated in a simple functional diagram. This functional diagram is analyzed in more depth to show what makes up a linear regulator and how a linear regulator operates. After the analysis of the various functional blocks, there are two sections covering some key characteristics and auxiliary functions. The last section of the chapter briefly explains the theory behind control loop stability and applies this theory to linear regulator stability.

9.1 Definition of Power and Basic Concepts Linear Regulators

9.1.1 Power

Let's begin by looking at what power is. Power is the rate of change of energy with respect to time. Since energy can be measured in joules, then power can be expressed as the amount of joules changing per second or watts. Electrical power is the product of an electrical current and voltage. Electrical current is measured in amperes or coulombs per second. Electrical voltage is measured in volts or joules per coulomb. By multiplying voltage and current the remaining units are joules per second or watts (Figure 9.1). For example if a light bulb is connected to a 10 V battery and the current through the lamp load is 1 ampere (A), then the light bulb is consuming 10 watts (10 V x 1 A) of power.

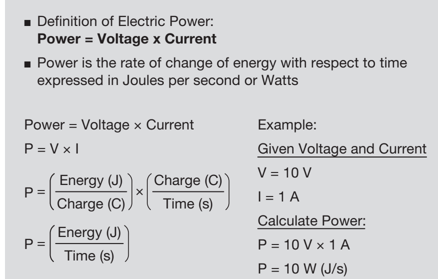


Figure 9.1 Electrical Power Review

9.1.2 Fundamental Concepts of Voltage Regulators

Many applications have loads where the required voltage differs from the source voltage. For example, in an automobile the typical source of energy is supplied by a 12 V battery, while a microcontroller in a car radio only needs 5 V. In applications where the source and load voltages are different, a power supply is often connected between them. A power supply is a device that converts a source voltage into a desired load voltage. The name power supply is a misnomer because it is converting rather than supplying energy. Hence a more accurate name for a power supply is the power converter (see Figure 9.2).

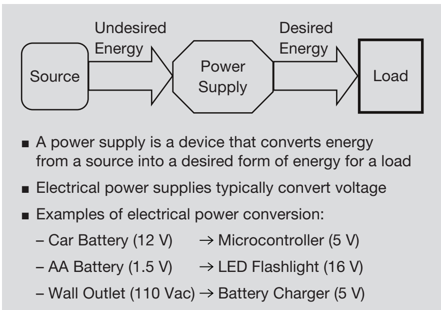


Figure 9.2 Power Supply Definition

Real power supplies require some power to operate and the converted output power is less than the input power. By definition, the efficiency in real power supplies is less than 100%. Efficiency is the ratio of output power to input power. Output power is the power delivered to a load while the input power is the power supplied to the power supply.

Figure 9.3 illustrates the point. If a power supply is drawing 1 A of current from a 12 V source the input power to the power supply is:

$$P_{IN} = V \times I = (12\ V) \times (1\ A)$$

$$P_{IN} = 12\ W$$

Next, the power supply is delivering 2 A of current to a load at 5 V, the power supply's output power is:

$$P_{OUT} = V \times I = (5\ V) \times (2\ A)$$

$$P_{OUT} = 10\ W$$

Although the power supply is delivering more current (2 A) than it consumes (1 A) it is still delivering less power (10 W) than it receives (12 W). The efficiency (η) of a power supply is calculated by taking the ratio of output power to input power.

$$\eta = P_{OUT} / P_{IN} = 10\ W / 12\ W$$

$$\eta = 83\%$$

The remaining 2 W of power in the example is consumed by the power supply and lost as heat (Figure 9.3).

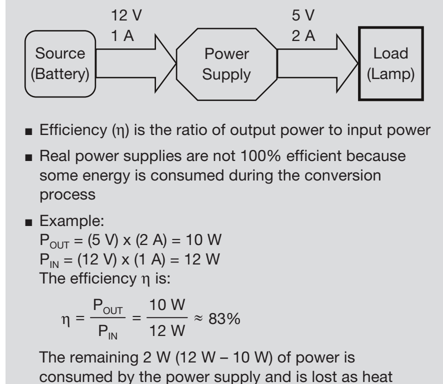


Figure 9.3 Power Supply Efficiency

9.2 Linear Regulators

One of the most common types of electronic power supplies is a linear voltage regulator. A linear voltage regulator, or simply a linear regulator, is a power supply that converts a direct current (DC) input voltage into the desired but smaller DC output voltage. In spite of input voltage variations the output voltage will remain steady as long as the input voltage is greater than the output voltage. Output voltage is kept constant by means of operating a series pass transistor (connected between input and output) in the linear or active region; hence the name “linear” regulator. Figure 9.4 is a cartoon illustration of the principle.

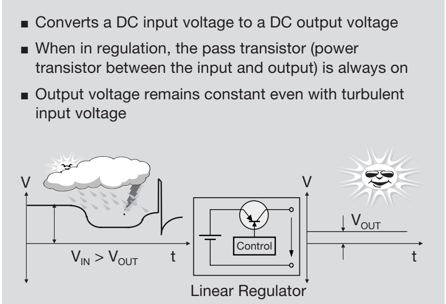


Figure 9.4 What is a Linear Regulator?

Figure 9.4 is a very simple block diagram of a linear regulator that illustrates the basic operation. For simplicity the ground connection is not shown and the focus is on the input and output voltage. The three functional blocks shown are:

Pass Transistor, Output Monitor and Gate Driver (Control).

The Output Monitor senses the output voltage and determines if the output voltage is above, below, or at its expected value. Based on the feedback signal from the Output Monitor, the Gate Driver adjusts the pass transistor to maintain a constant output voltage.

Figure 9.5 is a simple block diagram, and a graph of the input and output voltages illustrates the desired power supply operation. Assume the initial (and desired) output of the linear regulator is 5 V. The Output Monitor senses the output which is at its expected 5 V value and instructs the Gate Driver to maintain its current gate drive voltage at, say 50% of its maximum value.

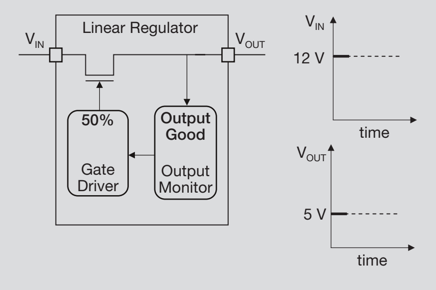


Figure 9.5 How a Linear Regulator Works?

If an unexpected input voltage increase from 12 V to 14 V occurs, as shown in Figure 9.6, the output voltage will also rise until the Output Monitor can react to this change. Typically the reaction time of the Output Monitor is in the order of microseconds.

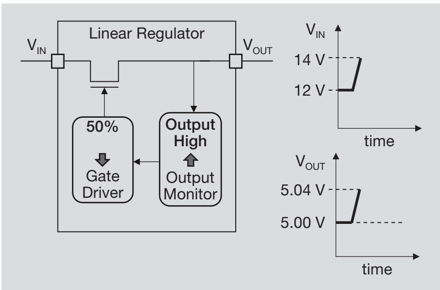


Figure 9.6 How a Linear Regulator Works?

Once the Output Monitor reacts and realizes the output voltage has increased beyond its expected value the Output Monitor commands the Gate Driver to slightly turn off the pass transistor by reducing its gate drive voltage from, say 50% to 40%. The reduction of the gate voltage from 50% to 40% will cause the output voltage to decrease, but, in the illustrated case (Figure 9.7), it is still above 5 V.

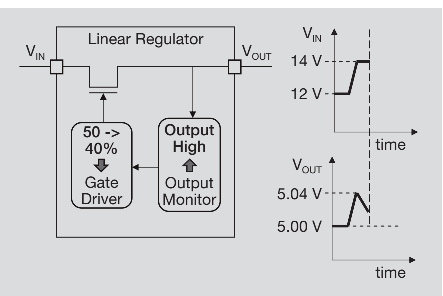


Figure 9.7 How a Linear Regulator Works?

The Output Monitor senses the output voltage is still higher than the expected value and commands the Gate Driver to further reduce the voltage applied to the pass transistor from say 40% to 30% (Figure 9.8).

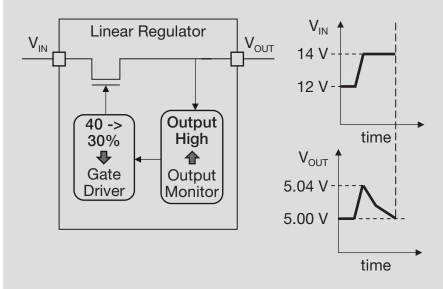


Figure 9.8 How a Linear Regulator Works?

At 30% of gate drive voltage the output voltage returns to its expected value of 5 V. Therefore the Gate Driver maintains the pass transistor at 30% to keep the output voltage at 5 V (Figure 9.9).

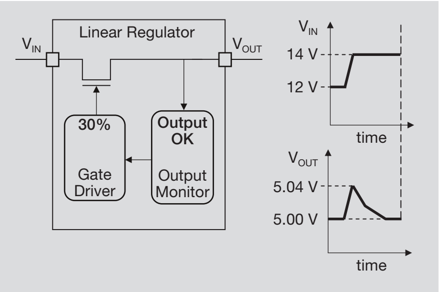


Figure 9.9 How a Linear Regulator Works?

9.3 Linear Regulator Functional Blocks

With reference to the examples in Figure 9.6 - Figure 9.9 the pass transistor is not turned off during the regulation process. By changing the pass transistor's gate voltage, the pass transistor stays on but adjusts the amount of the voltage drop from the input to the output regardless of the current demand (within the specified limits). The on state operation of the pass transistor is analogous to a water pressure regulator.

Figure 9.10 is a functional block diagram of a linear regulator. There are 4 basic blocks shown in this functional diagram and role and operation of each will be explained in this section. The blocks are:

- Bandgap reference
- Series pass transistor
- Error amplifier
- Feedback voltage divider

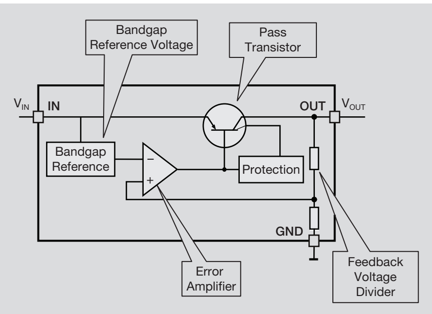


Figure 9.10 Linear Regulator Functional Block Diagram

9.3.1 Bandgap Reference Block

The bandgap reference is an important part of a linear regulator because it provides the reference voltage to accurately maintain the output voltage (regardless of input voltage and output current within their respective specified ranges). The bandgap reference also provides a very stable reference voltage across the entire specified operating temperature range. The voltage reference is stable across a wide temperature range because it uses two offsetting characteristics of semiconductor devices. The base-emitter voltage of a diode-connected NPN bipolar junction transistor approximately decreases at the rate of 2 mV/°C. This decrease in the diode voltage, however, is countered by an increase in the thermal voltage of a similar NPN transistor. The thermal voltage is defined as: $V_t = kT/q$, where k is Boltzman's constant (1.38 * 10^{-23} AsV/K), T is the integrated circuit junction temperature (in K), and q is the electron charge (1.6 x 10^{-19} As). The temperature coefficient of the thermal voltage is positive and equal to 0.086 mV/K.

This voltage is amplified by the factor $A_0 \approx 23.5$ so that the net temperature coefficient of the voltage reference is 0 mV/C as illustrated in Figure 9.11.

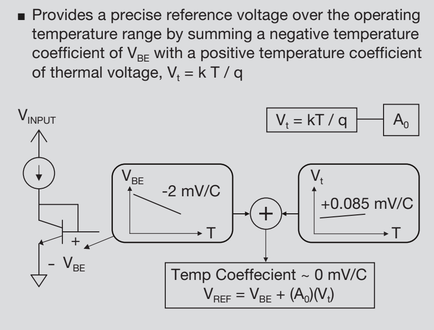


Figure 9.11 Bandgap Reference Voltage (Concept Illustration)

In practice, the bandgap circuit may be similar to the circuit in Figure 9.12. Its reference voltage, V_{REF}, is typically 1.2 V. A "trimming network" of resistors (R_3 - R_5) is added to allow V_{REF} to be permanently adjusted if its output voltage deviates from its expected value due to manufacturing variations.

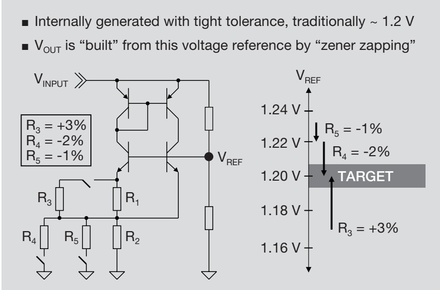


Figure 9.12 Bandgap Reference Voltage Circuit Example

In addition to providing an accurate and stable reference to set and maintain the accuracy of the output voltage, the voltage reference is used for all internal auxiliary circuits. Therefore if the output voltage varies over temperature, then the reset threshold will also vary in the same manner. This behavior is illustrated in Figure 9.13.

- The Bandgap is the internal reference for all circuits, e.g. error amplifier, reset circuitry, output voltage
- Therefore parameter changes due to the temperature will follow the temperature profile of the bandgap reference voltage (e.g. Reset Threshold and Output Voltage vary in the same manner)

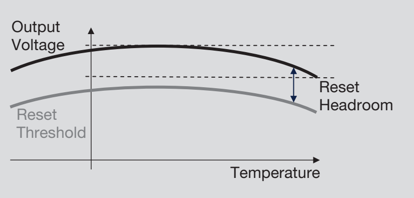


Figure 9.13 Bandgap Reference Voltage and Reset Voltage vs. Temperature

9.3.2 Series Pass Transistor Topologies

Every linear voltage regulator contains a pass transistor in the current path from the power supply input to the output. In the functional diagram example (Figure 9.10) the pass transistor is a PNP bipolar junction transistor. In addition there are three other types of pass transistor topologies used in linear regulators (Figure 9.14).

- “NPN” or Standard Linear Regulator
- “Quasi” Low Drop Out Linear Regulator
- “PNP” or Low Drop Out (LDO) Linear Regulator
- MOS Low Drop Out and Low Quiescent Current Linear Regulator

Figure 9.14 Types of Pass Transistors

A standard or conventional linear regulator uses a NPN bipolar junction transistor as the pass transistor to provide a current from the input to output (Figure 9.15). The minimum voltage difference between the input and output is about 2 V. This means that the input voltage driving the power supply must be at a voltage which is at least 2 V higher than the regulated output voltage. For example:

V_{OUT} = 3.3 V $V_{IN,MIN}$ = 5.3 V
V_{OUT} = 5.0 V $V_{IN,MIN}$ = 7.0 V
V_{OUT} = 12 V $V_{IN,MIN}$ = 14 V

This requirement can be prohibitive for some automotive and battery applications when the output voltage must be within specification even when the input voltage droops. Note that since the input voltage must be higher than the output voltage, linear regulators are not capable of higher (step-up) output voltage conversion.

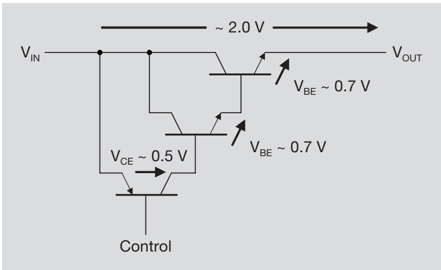


Figure 9.15 “NPN” or Standard Linear Regulator

Some semiconductor suppliers have developed a variant of the standard linear regulator called the “Quasi” Low Drop Out (LDO) regulator. In these regulators one of the NPN transistors has been removed to reduce the required minimum voltage differential between the input and output. The minimum voltage difference between the input and output is about 1.2 V (see Figure 9.16).

V_{OUT} = 3.3 V $V_{IN,MIN}$ = 4.5 V
V_{OUT} = 5.0 V $V_{IN,MIN}$ = 6.2 V
V_{OUT} = 12 V $V_{IN,MIN}$ = 13.2 V

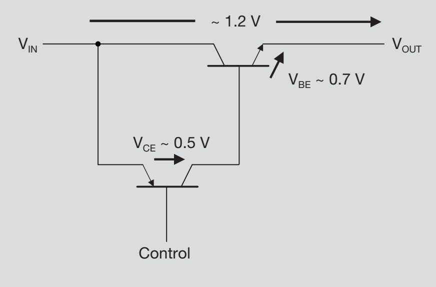


Figure 9.16 “Quasi” Low Drop Out Linear Regulator

In a Low Drop Out (LDO) linear regulator the pass transistor is a PNP bipolar junction transistor. While the PNP transistor requires a larger semiconductor area than a NPN

transistor of comparable current specification it can operate with a smaller voltage differential. This is very important for some automotive and battery applications that need to operate when the battery voltage droops. The minimum voltage difference between the input and output is about 0.5 V (as shown in Figure 9.17).

V_{OUT} = 3.3 V $V_{IN,MIN}$ = 3.8 V
V_{OUT} = 5.0 V $V_{IN,MIN}$ = 5.5 V
V_{OUT} = 12 V $V_{IN,MIN}$ = 12.5 V

Additional design techniques are usually implemented in LDO regulators to minimize the quiescent current. The quiescent current is the amount of current consumed by a linear regulator. A low quiescent current is very important for automotive and battery applications because it conservers battery power.

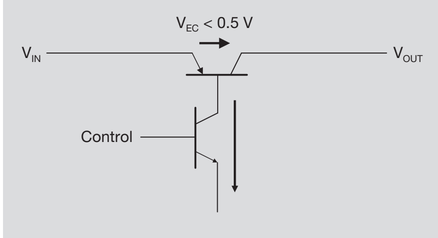


Figure 9.17 "PNP" or Low Drop Out (LDO) Linear Regulator

Finally the MOSFET pass transistor is the last topology we mention. The MOSFET pass transistor significantly reduces the quiescent current demand because it is voltage driven, unlike traditional current driven NPN and PNP LDO pass transistors (Figure 9.18). The price paid for the lower stand-by current in case of a fully integrated regulator is a more complex and noisier control circuit (charge pump). The cost of the greater complexity is mitigated by the inherently lower cost of CMOS technology. In case of a discrete, external pass transistor, a p-channel device can be chosen which will not require the charge-pump gate driver.

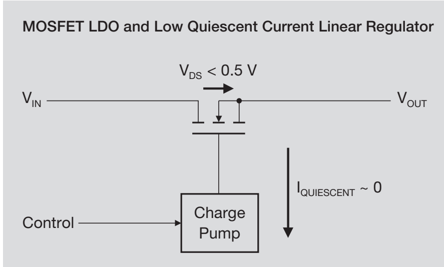


Figure 9.18 N-channel MOSFET Series Pass Transistor

In summary:

NPN or Standard Linear Regulators:

Many regulators with only NPN pass transistors were designed starting in the mid-1960-s. Their functionality and integrated protection is limited. In addition, they are usually not specified to operate under automotive conditions. From a regulator perspective only, standard regulators tend to be very inexpensive. However they will typically require additional protection circuitry adding additional costs to be useful in many automotive applications.

PNP or LDO Linear Regulators:

As stated before, LDO regulators have a low voltage differential from input to output. In addition, LDO regulators can have low quiescent currents mostly due to the design of the linear regulators. They often have lots of additional features (some of which we will cover later) and can be protected against common fault conditions in automotive modules.

MOSFET LDO Linear Regulator with Low Quiescent Current:

MOSFET pass transistors provide a very low drop out voltage and minimal quiescent current. In addition, more complex features may be found in MOSFET based linear regulators because of higher integration capability allowed by the semiconductor process.

The summary of features is captured in Figure 9.19.

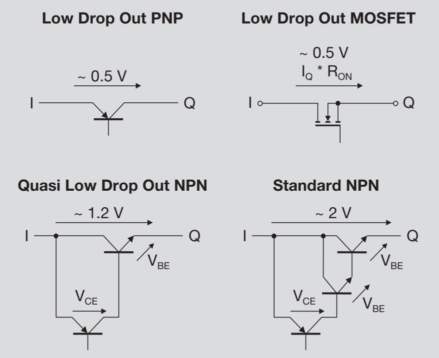


Figure 9.19 Summary of Pass Transistors

9.3.3 Error Amplifier

Recalling that in a linear regulator the pass transistor is always turned on to some extent, but when output current demand is high, the control in the regulator needs to turn on the pass transistor more. When the output current demand is low the regulator needs to turn on the pass transistor less. The extent a pass transistor is turned on is controlled by an error amplifier (Figure 9.20).

The error amplifier senses the output voltage through a voltage divider feedback network which provides a feedback voltage V_{FB} to the error amplifier. If the output voltage is higher than expected the V_{FB} will also be higher than the reference voltage V_{REF}. Assuming a PNP series pass transistor by connecting the feedback voltage V_{FB} to the non-inverting input of the error amplifier will cause its output to increase. Increasing the error amplifier output decreases the V_{EB} voltage of the pass transistor and the transistor is turned down a little bit. This causes the output voltage V_{OUT} to droop back to its expected value.

If the output is lower than expected, V_{FB} will also be lower than V_{REF} causing the output of the error amplifier to decrease. Decreasing the error amplifier output will increase the V_{EB} voltage of the pass transistor. Since the V_{EB} voltage is increased, the transistor is turned on a little bit more. This causes V_{OUT} to move toward its expected value.

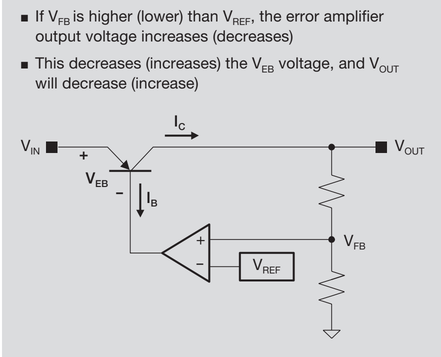


Figure 9.20 Error Amplifier

9.3.4 Voltage Divider

A resistor voltage divider is used to create a feedback signal, V_{FB}. Resistors are sized so the feedback voltage V_{FB} is equal to the bandgap reference voltage V_{REF} at the specified output voltage. The accuracy of the feedback resistors combined with the accuracy of the bandgap reference voltage determine, to a large extent, the total output voltage tolerance. Fixed output linear regulators have internal feedback resistors. Variable or adjustable output voltage linear regulators have external feedback resistors.

The functional position of the feedback circuit is shown in Figure 9.20 and Figure 9.21.

■ Voltage divider feedback from output to ground

■ Resistors are sized that the feedback voltage (V_{FB}) is equal to the bandgap reference voltage (V_{REF}) under typical conditions

■ Resistors are internal for fixed output and external for adjustable output

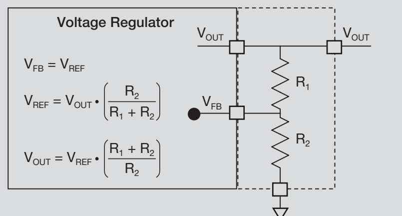


Figure 9.21 Feedback Voltage Divider

9.4 Protection Circuits

Most linear regulators have built in protection features to guard against unintended fault conditions. The typical protection features integrated in linear regulators are current limiting, safe operation monitor and thermal shutdown.

9.4.1 Current Limiter

Current limiting is a protection function that keeps the output current within the specified bounds to protect both the regulator and load. To implement the current limiting protection the linear regulator must be able to monitor the output current. In modern linear regulators the output current is monitored with an alternate pass transistor tap, called current mirror, the output current of which is linearly proportional to that flowing through the series pass transistor. Therefore by monitoring the current through the alternate collector tap a measure of the output current is obtained. The alternate tap is fed into a control block of the linear regulator which controls the flow of current from the input to output.

There are two different modes of current limiting: constant current limiting and a current "foldback" limiting.

Constant current limiting, as the name suggests, limits the output current to a constant maximum allowed value.

Foldback current limiting reduces the output current after the current limit threshold has been exceeded. By reducing or "folding back" the output current the linear regulator power dissipation is reduced. At first glance foldback current limiting may appear to a better technique because it also reduces the power dissipation. However, simpler foldback controls have the potential to "latch up" in the limiting state from which they cannot recover without additional intervention even when the over load condition is removed . The current states of each limiting method are illustrated in Figure 9.22.

- The output current is measured through an alternate collector tap
- If current is too high, the regulator can limit or reduce the output current from increasing

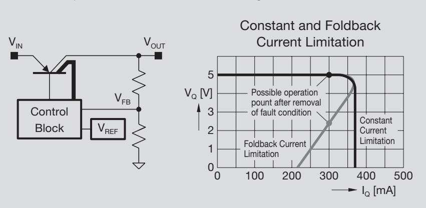


Figure 9.22 Current Limiting Methods

9.4.2 Input Over-voltage Monitoring and SOA Protection

The purpose of the input voltage monitoring is to keep the regulator dissipation within the safe operating area (SOA). Input over-voltage monitoring is part of a protection function that limits the output current when the input voltage is too high and, as a result, the dissipation in the series pass transistor would exceed the limits of the SOA. In many automotive linear regulators the safe operating area monitor begins limiting output current when the input voltage has reached 22 V (see Figure 9.23).

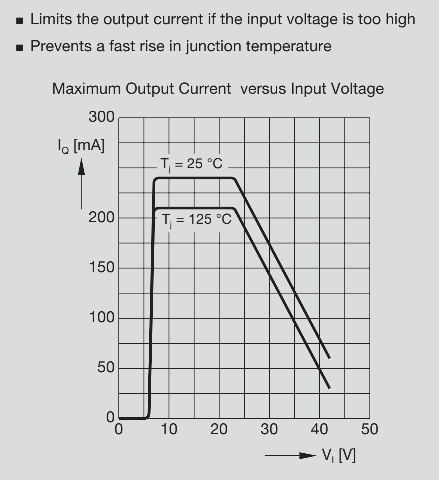


Figure 9.23 Safe Operating Area Monitor

There is a caveat concerning the SOA monitor. The current limiting response to input over-voltage is by a reduction of the output voltage

(refer to 9.4.1). Decrease of output voltage may cause the generation of a reset signal. In order to avoid that to happen the linear regulator must be sized so, that the load current at the maximum input voltage does not shift the operating conditions outside the SOA. Engine control functions must be fully operational at 27 V input voltage (double battery cranking supply). Limiting at 22 V maximum does not cause any problem to most of the body and entertainment functions.

9.4.3 Temperature Monitoring and Thermal Shutdown

Thermal shutdown protects a linear regulator from immediate destruction due to high temperature conditions by turning off the regulator. Most linear regulators have a maximum junction temperature rating of 150 °C. Operating a linear regulator over the maximum junction temperature significantly reduces its life time and pushes the operating variables outside their specified limits.

A simplified implementation of thermal shutdown is shown in Figure 9.24 with a single transistor circuit. During normal operation the base voltage of the NPN transistor in the circuit is not sufficient to turn it on. Therefore no current flows through the NPN device and the OVERTEMP signal is HIGH (no over temperature condition exists). If the temperature increases, then the base-emitter voltage V_{BE}, necessary to turn on an NPN transistor, decreases because of its negative temperature coefficient. Above the maximum junction temperature the required base-emitter voltage is low enough and the NPN transistor turns on causing an OVERTEMP signal to be actively pulled down to ground. The low level OVERTEMP signal informs the control block that the temperature is above the maximum rating and the control block turns off the pass transistor.

- The thermal shutdown prevents the IC from immediate destruction from over temperature
- Exceeding the maximum junction temperature (or running into thermal shutdown) significantly reduces IC lifetime
- As temperature increases the necessary voltage V_{BE} to turn on a NPN decreases
- Above 150 °C the NPN transistor turns on and an OVERTEMP signal goes LOW indicating an over temperature condition and the output turns off

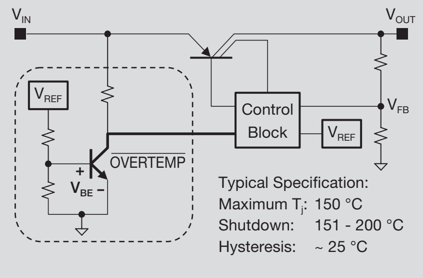


Figure 9.24 Thermal Shutdown

9.5 Characteristics of Linear Regulators

Linear regulators are characterized by many different parametric values. In this section drop out voltage, output current limitation, and thermal resistance will be explained.

9.5.1 Drop Out Voltage

Drop out voltage is the minimum voltage difference between a linear regulator input and output that is required for voltage regulation. Drop out voltage is defined as the difference between the input voltage and the output voltage when the output voltage has dropped 0.1 V from its typical value. For example, consider a 5 V, 2% (±0.1 V) regulator with a drop out voltage of 0.5 V. When the input voltage is reduced to 5.4 V the output voltage would drop from 5.0 V to 4.9 V. Therefore, the minimum input voltage required for a 5 V, 2% regulator is 5.4 V (4.9 V + 0.5 V).

A typical drop-out voltage datasheet entry is shown in Figure 9.25.

Parameter	Symbol	Limit Values			Unit	Condition
		Min.	Typ.	Max.		
Drop Out Voltage $V_{dr} = V_I - V_Q$	V_{dr}	–	250	500	mV	I_Q = 150 mA

- Drop Out Voltage is the minimum voltage differential between a linear regulator input and output that is required for voltage regulation
- Lower drop out voltages enable operation at lower input voltages

Figure 9.25 Drop Out Voltage

9.5.2 Output Current Limits

By knowing the definition of drop out voltage, it is easier to understand how the maximum output current is specified in linear regulators. It may be obvious that if a load tries to pull more current from a power supply than it was designed to deliver, the output voltage will drop. This behavior of power supplies is key to the specification of the maximum output current.

In the first row of the current limit specification (Figure 9.26) the minimum specified current is 300 mA. This minimum is specified where the output voltage has dropped 100 mV below the nominal value. This minimum specification guarantees that the linear regulator will be able to source at least 300 mA of current when the output voltage has dropped 100 mV. If a load attempts to pull more than 300 mA current, then the linear regulator enters current limit mode. In the current limit mode the linear regulator will protect itself by limiting the output current and causing output voltage to sharply drop. In the second row of the current limit specification the maximum specified current is 800 mA. The maximum is specified where the output voltage has dropped to 0 V.

Parameter	Symbol	Limit Values			Unit	Condition
		Min.	Typ.	Max.		
Output Current Limitation	I_Q	300	–	–	mA	1)
Output Current Limitation	I_Q	–	–	800	mA	V_Q = 0 V

1) Measured when the output voltage V_Q has dropped 100 mV from the nominal value obtained at V_I = 13.5 V

Figure 9.26 Output Current Limits

9.5.3 Quiescent or Ground Current

Quiescent current or ground current is the current demand of a linear regulator to perform its functions. Figure 9.27 shows three typical ways the quiescent current is specified. In the first row of the table the quiescent current is specified when the linear regulator has been inhibited or turned off. The current consumed by a regulator when inhibited is extremely low (1 µA). Very low current consumption is especially helpful to conserve battery power in key-off state. In the second row the quiescent current for a very small output current (less than 1 mA) is shown. This indicates that the linear voltage regulator will typically require 65 µA of quiescent current to turn on the output. The third row indicates quiescent current increases at a given output current. This value is important when performing accurate power dissipation calculations at a given V_{IN} to know the power loss heating the linear regulator

Parameter	Symbol	Limit Values			Unit	Condition
		Min.	Typ.	Max.		
Quiescent Current	I_q	–	–	1	µA	Inhibit = L (Device Off) $I_Q = 0$
Quiescent Current	I_q	–	65	105	µA	Inhibit = H (Device On) $I_Q \leq 1$ mA
Quiescent Current	I_q	–	1	2	mA	Inhibit = H (Device On) $I_Q \leq 150$ mA

Figure 9.27 Quiescent (Ground) Current

9.5.4 Thermal Resistance

Thermal resistance of the linear voltage regulator is a measure of how well a regulator conducts heat away from the power semiconductor. Lower thermal resistance values translate into higher currents through the device at the power dissipation limits. The maximum junction temperature allowed for linear regulators is 150 °C.

Thermal resistance is vital to power dissipation calculations and can be expressed in multiple ways. Most datasheets will specify a junction-to-ambient thermal resistance, R_{thJA}. Other thermal resistances commonly specified are

the junction-to-case R_{thJC} and junction-to-pin R_{thJP}. A useful thermal resistance is the junction-to-ambient R_{thJA} because it includes all of the other thermal resistances (from ambient to junction). Figure 9.28 defines junction-to-ambient thermal resistances as a function of the printed circuit board foil area. Given the thermal resistance R_{thJA} and the maximum junction temperature, the maximum allowable power dissipation P_D or the maximum allowable ambient temperature T_A can be calculated. Another practical design method is to calculate the required thermal resistance R_{thJA} since the power dissipation and temperatures are typically known.

Parameter	Symbol	Limit Values			Unit	Condition
		Min.	Typ.	Max.		
Junction-Ambient Thermal Resistance	R_{thJA}	–	104	–	°C/W	Footprint only
		–	74	–	°C/W	300 mm² PCB heatsink area
		–	65	–	°C/W	600 mm² PCB heatsink area

- Thermal resistance indicates how much heat can be conducted by the regulator
- Lower thermal resistance equates to better thermal performance (more power capability or higher ambient temperature operation)

$P_D = (T_J - T_A) / R_{thJA}$

$T_A = T_J - (R_{thJA} * P_D)$

P_D = Power Dissipation
T_J = JunctionTemperature
T_A = Ambient Temperature
R_{thJA} = Junction-to-Ambient Thermal Resistance

Figure 9.28 Thermal Resistance

In the following example, the maximum allowable junction-to-ambient R_{thJA} is calculated based on given operating conditions. Assume a linear regulator is connected to 14 V input source and supplies 5 V output to a load that draws 30 mA and the regulator current demand is 0.5 mA. Therefore the power dissipated is in the device is:

$$(14 - 5) * (0.03 + 0.0005) = 0.275 \text{ W}$$

After calculating the power dissipation, the maximum acceptable thermal resistance R_{thJA} can be calculated for $T_J = 150$ °C and $T_A = 85$ °C:

$$R_{thJA} = (150 - 85) / P_D = 65 / 0.275 = 236.4 \text{ °C/W}$$

In this example, the thermal resistance must be 236 °C/W or less to dissipate 0.275 W of power at 85 °C ambient temperature.

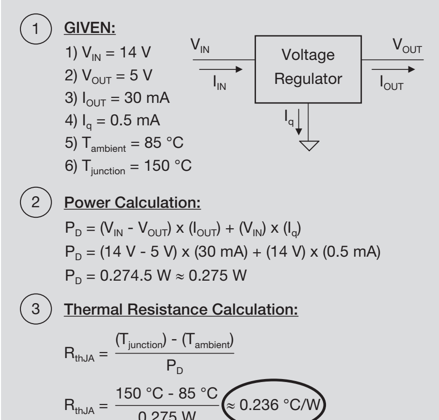


Figure 9.29 Thermal Resistance Example

Larger packages typically have lower thermal resistance. However, the type of package and size of PCB heatsink area can greatly affect the total junction to ambient thermal resistance. For example, a SCT-595 package is much smaller than a DSO-14 package and has a higher (worse) junction to case thermal resistance. The construction of the SCT-595 package differs from the DSO-14 in that it has a large thermal tab that is very effective at conducting heat away when connected to a PCB heatsink area. Therefore a SCT-595 part assembled with 300 mm^2 of PCB heatsink area connected to the thermal tab will result in the same junction to ambient thermal resistance as a significantly larger DSO-14 package.

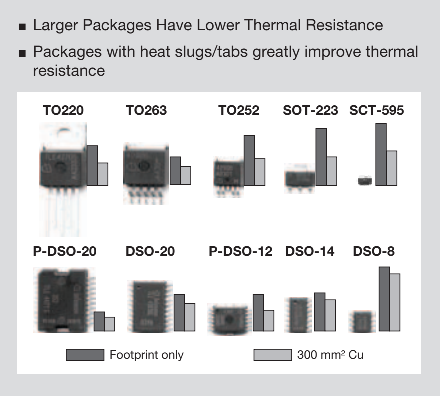


Figure 9.30 Package and Layout Affect Thermal Resistance

9.6 Auxiliary Functions of Linear Regulators

9.6.1 Inhibit

Many linear regulators can be turned off with an inhibit control input. In some automotive and battery applications it is necessary to significantly reduce the quiescent current when the module is off. This can be accomplished by turning off the linear regulator with a logic low signal (0 V) applied to the Inhibit pin. To turn on the regulator a logic high signal (5 V) is applied to the Inhibit pin (Figure 9.31).

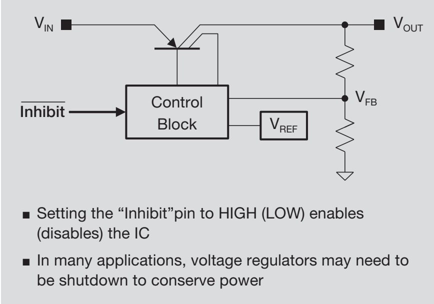
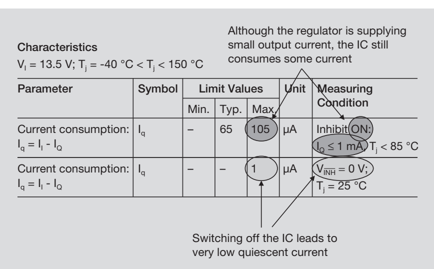


Figure 9.31 Inhibit

In the following example (Figure 9.32) the first row shows the quiescent current when the linear regulator is turned on i.e. not inhibited. The second row shows the quiescent current the device is turned off (inhibited). The quiescent current can be reduced by many orders of magnitude by making use of the INHIBIT input.



Characteristics
$V_I = 13.5$ V; $T_j = -40$ °C < T_j < 150 °C

Parameter	Symbol	Limit Values			Unit	Measuring Condition
		Min.	Typ.	Max.		
Current consumption: $I_q = I_I - I_Q$	I_q	–	65	105	µA	Inhibit ON: $I_Q \leq 1$ mA; $T_j < 85$ °C
Current consumption: $I_q = I_I - I_Q$	I_q	–	–	1	µA	$V_{INH} = 0$ V; $T_j = 25$ °C

Figure 9.32 Inhibit Example

9.6.2 Reset

In addition to Inhibit, Reset is another common auxiliary function found in linear regulators. Reset is a logic low, or logic 0 output signal to the microcontroller occurring in two cases: a) when the output voltage is below its specified low-limit (under voltage reset) and b) when the regulator (and the rest of the system it serves) is being powered up (power-on reset).

Under voltage reset is self explanatory. Power-on reset is a function that delays the initialization of a microcontroller until its clock oscillator has stabilized (Figure 9.33).

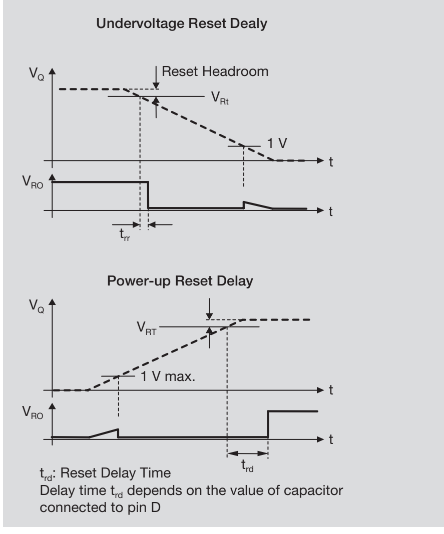


Figure 9.33 Reset

Under voltage reset operates by sensing the output voltage V_{OUT} and comparing it to an internal reset threshold voltage V_{RT}. If the output voltage drops below the reset threshold, then the reset output is active low as long as the low output state exists. The reset output typically connected to a microcontroller reset pin as shown in Figure 9.34.

- The reset circuit compares V_{OUT} to V_{RT}, the reset threshold
- In case $V_{OUT} \leq V_{RT}$ the RO signal is pulled low
  - To guarantee safe operation of a micro controller
  - To indicate low supply voltage

Figure 9.34 Under voltage Reset

Most control modules have a microcontroller and an accompanying clock oscillator. When a microcontroller is first turned on the clock oscillator requires a stabilization period typically in the order of 1 to 10 ms. If a microcontroller begins operating before the oscillator is stable, the microcontroller may not initialize correctly. Linear regulators equipped with the reset function provide a power-on reset delay which prevents a microcontroller from initializing while the oscillator is still stabilizing. Referring to Figure 9.35 and Figure 9.36, a reset output (RO) signal from a linear regulator is asserted low (0 V) and actively holds the microcontroller in the reset state for a certain delay time t_{rd}. The reset delay time t_{rd} is measured from the time where the output voltage has increased above the reset threshold voltage V_{RT} to the time the reset output is released (goes high).

- A microcontroller oscillation stabilization time is typically in the order of ms
- Only when a stable oscillator/clock signal is established, can a microcontroller be correctly initialized
- As V_{OUT} goes above V_{RT}, the reset control block waits t_{rd} (the reset delay time) to allow oscillation stabilization before reset is released and the microcontroller can begin initialization

C = 20 pF ±10 pF for crystal operation

Figure 9.35 Power-On Reset Delay

The length of the reset delay time may be set with a small external capacitor. In addition, the capacitor also sets the reset reaction time t_{RR}. The reset reaction time is the minimum time the output voltage must be below the reset threshold before an under voltage reset output signal is generated (Figure 9.36).

- An external capacitor C_D determines the reset delay time t_{rd}, (the timing until the reset output RO is released) and reset reaction time t_{RR}

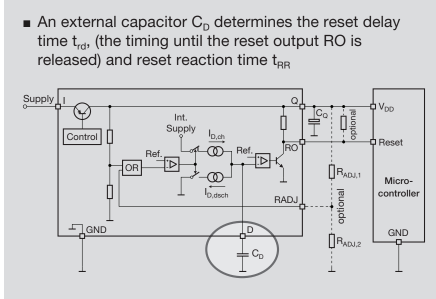


Figure 9.36 Reset Timing

In the following example the delay capacitor value is calculated to achieve a 25 ms reset delay time (see Figure 9.37).

- The reset delay time can be configured by the external C_D

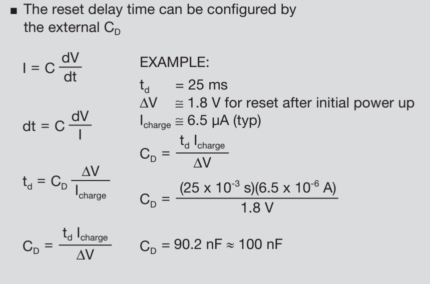


Figure 9.37 How to Calculate Reset Delay time?

After calculating the reset delay capacitance, the reset reaction time can be calculated using this capacitance (C_D) as shown in Figure 9.38.

- The timing capacitor C_D also impacts the reset reaction time
- Reset reaction time, t_{RR}, is defined as the time it takes for the voltage regulator to change the reset output
- V_{OUT} may go dip below V_{RT}, but it must stay lower than V_{RT} for at least t_{RR} before RESET is asserted
- Letting: C_D = 100 nF:

  $t_{RR} = (12\ \text{s/F})(C_D)$

  $t_{RR} = (12\ \text{s/F})(100\ \text{nF}) = 1.2\ \mu\text{s}$

Figure 9.38 Reset Reaction Time

The reset threshold V_{rt} can also be adjusted on some linear regulators with an external voltage divider. This adjustment capability is useful for microcontrollers that can operate at a lower voltage than the standard reset threshold (see Figure 9.39).

- The reset switching threshold, V_{RT}, can be adjusted by using an external voltage divider
- Useful for microcontrollers having lower operating voltages than the standard reset threshold voltage V_{RT}
- The reset adjust switching threshold $V_{RADH,TH}$ is 1.36 V (typ)

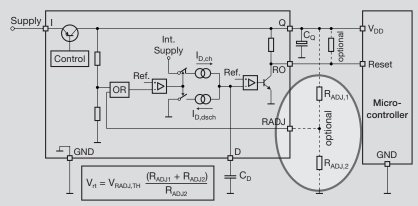


Figure 9.39 Reset Threshold

9.6.3 Watchdog Timer

An advanced form of protection used with microcontroller systems is the watchdog timer. The watchdog monitors the microcontroller to ensure it is operating correctly. The function of the watchdog timer is to monitor the timing of the microcontroller and reset it to a known state of operation in case of an obvious timing error is detected.

The process is illustrated in Figure 9.40 and Figure 9.41. The monitored microcontroller sends a periodic signal to the watchdog circuit to inform its still operating correctly. The periodic signal from the microcontroller to the watchdog is often called a trigger or can be referred to as “petting the watchdog”.

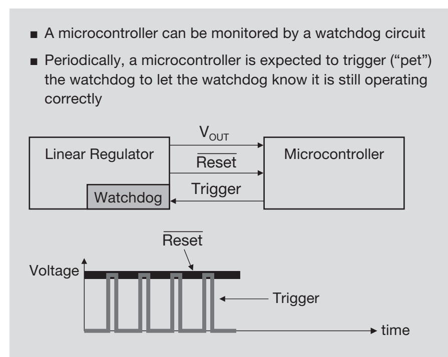


Figure 9.40 Watchdog Normal Operation

For example, a microcontroller may get stuck in a software loop and stops responding to other inputs. If too much time elapses between triggers, then the watchdog senses that something is wrong and resets the microcontroller. Speaking in metaphorical terms, the watchdog barks and wakes up the owner. The reset signal from the watchdog reinitializes the microcontroller to ensure it does not get stuck in an unexpected software loop or erroneous state.

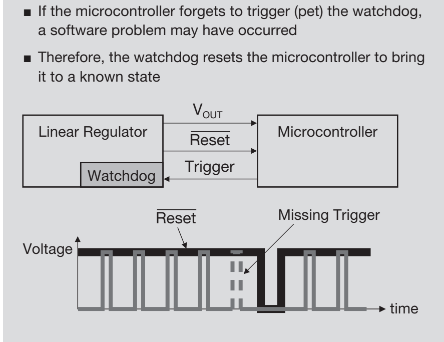


Figure 9.41 Watchdog Fault Operation

For critical applications a more advanced watchdog called a window watchdog is also available. A window watchdog operates in a similar manner as the standard watchdog except a trigger must occur within a certain window or time slot. If a trigger occurs outside of the window or does not occur at all within the designated window, then the window watchdog will reset the microcontroller. If an unintentional trigger occurred, a standard

watchdog would not be able to decipher if this trigger was valid. The timing or window requirement of a window watchdog enables it to detect unintentional triggers.

9.6.4 Early Warning

The final auxiliary function presented in this section is the Early Warning Function. Figure 9.42 shows a simplified Early Warning circuit. It can be used in a voltage regulator by comparing a divided sample of the input voltage (V_{SI}) to a known reference voltage. When the external voltage at the sense input voltage V_{SI} drops below the reference voltage $V_{I,TH}$ an active low warning signal is generated at the sense output (SO) pin.

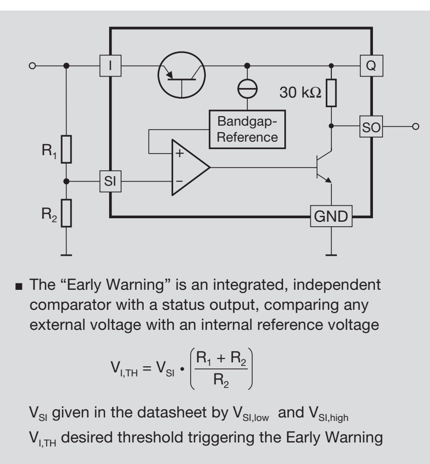


Figure 9.42 Early Warning Circuit

The function of the Early Warning circuit is to make sure that the microcontroller powers down in an orderly fashion. When the input voltage decreases, the output voltage is at risk of falling out of regulation or decreasing below the minimum specified operating voltage. Before the output voltage crosses the reset threshold an early warning signal is transmitted from a linear regulator to a microcontroller. The early warning signal alerts the microcontroller that the supply voltage is dropping and the microcontroller may receive a reset signal soon. This allows a microcontroller to perform any "house cleaning" chores like saving RAM values into EEPROM memory so it can resume operation at its start-up state when it powers up again and the reset is released.

Figure 9.43 illustrates how Early Warning is used.

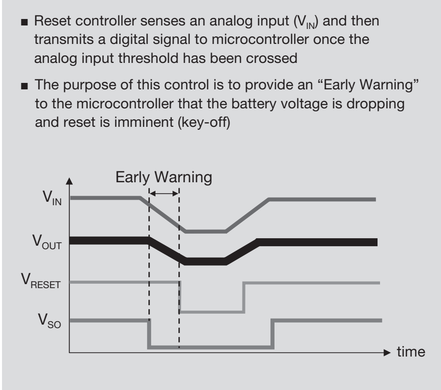


Figure 9.43 Early Warning Example: Low Battery Detection and Warning

9.7 Stability

The stability of linear regulators is a characteristic that describes how well it can maintain the regulated output voltage in spite of changes in the input voltage or load current. If a linear regulator is stable, then the output voltage will be a steady and constant DC voltage. If a linear regulator is unstable, then the output voltage will oscillate (Figure 9.44).

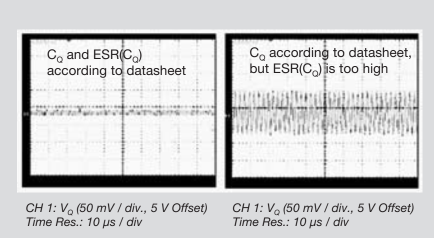


Figure 9.44 Stable and Unstable Output Voltage Displays

A linear regulator is a control system that uses negative feedback to regulate its output voltage to a precise reference. If a control system phase shift reaches 180° where the loop gain is greater than 0 dB (x1), the negative

feedback turns into destabilizing positive feedback and the system either oscillates or latches up. The stability of linear regulators depends on many factors such as load current, temperature, output capacitance and the ESR (equivalent series resistance) of the output capacitor (as shown in Figure 9.45). The most significant factor from an application perspective is the output capacitor because it is external to the regulator IC and the user has control over its selection.

- Linear regulators belong to a class of control systems with negative feedback control loop
- If a control loop phase shift reaches 180° (where the loop gain is > 1 or 0 dB) then negative feedback turns into positive feedback and the system will be unstable (oscillating)
- Control loop stability depends on:
  – Load current
  – Temperature
  – Output capacitance
  – ESR of the output capacitor

Figure 9.45 Review of Control Loop Stability

A control system can be mathematically described by the transfer function of the loop-gain, $G(j\omega)$. The loop gain is the ratio of the output to the input and it can be described graphically with a Bode plot, which shows the loop gain and phase shift as functions of frequency of the input signal.

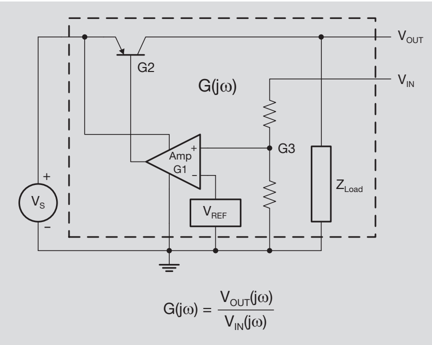


Figure 9.46 Bode Plot Analysis

The Bode plot is constructed by opening the loop at a convenient point (compare Figure 9.46 with Figure 9.20) and injecting a constant amplitude AC sine-signal at different frequencies. The amplitude and phase of the output signal is compared to the input by taking the output/input ratio and approximate phase shift which are plotted at different frequencies, i.e. taking the open-loop frequency response of the regulator. The necessary condition of stability is that at the frequency where the gain is unity the output phase lag with respect to the input be less than -180° (the negative sign symbolizes phase lag). In normal practice the Bode plot is constructed not by measurement but by calculation knowing the AC responses of the different blocks within the system.

A preferred but not necessary rule for control system stability is to have at least 30° of phase margin. Phase margin is the difference between the phase shift and -180° at the frequency where the loop gain has reached 0 dB or unity gain. For example, if the phase shift is -135°, then the phase margin is -135° - (-180°) = 45° (Figure 9.47). Notice that the frequency axis is logarithmic and so is the gain axis. The absolute value of gain, regardless of phase-shift, is given in dB-s defined as:

$$|G_{dB}| = 20 \times \log\left(\frac{|V_{out}|}{|V_{in}|}\right)$$

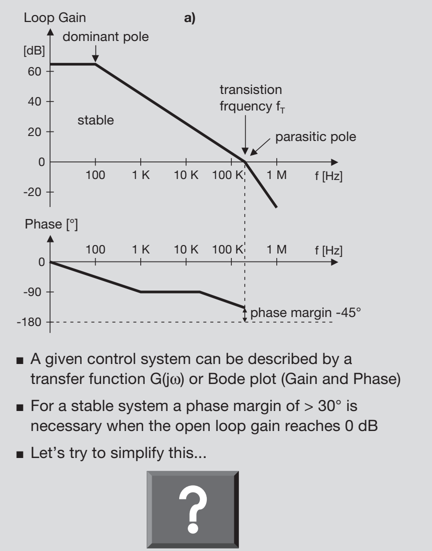


Figure 9.47 Review of Control System Stability

A transfer function $G(j\omega)$ has special frequencies, where the gain and the phase change, which are called poles and zeroes (Figure 9.48). A pole causes the loop gain to change (reduce) at a rate of -20 dB per decade and the phase tends to -90° as the frequency tends to infinity. Conversely, a zero causes the loop gain to change by +20 dB per decade and the phase tends to +90°. A decade is a frequency range between the limits of **f** and **10f** and $\omega=2\pi f$. If a transfer function has 2 poles, the change in phase shift at 0dB loop gain may be close to -180°=2 x (-90°). That, in turn, may result in near 0° phase margin and a very oscillatory control system, which is deemed unstable by our definitions. An ideally stable control system would only have one dominant pole at 0dB loop gain frequency. The maximum phase shift, that a single pole can cause, is only -90° and no oscillatory response is possible.

Transfer Function (Bode Plot) Review of Poles and Zeroes

- Poles and zeroes are the names of frequencies where the gain and phase of a control loop change
  - Poles correspond to the roots of the denominator of $G(j\omega)$
  - Zeroes correspond to roots of the numerator of $G(j\omega)$
- A pole (zero) causes the open loop gain to decrease (increase) at the rate of -20 dB per decade (+20 dB/decade)
- A pole (zero) causes the phase shift to decrease (increase) to -90 degrees (+90 degrees) as $\omega \rightarrow \infty$
  - at $f_{pole}/10$ (decade before), the phase shift starts to decrease (increase) from near 0°
  - at $f_{pole/zero}$ the phase has decreased (increased) to ±45°
  - at 10 $f_{pole/zero}$ (decade after) the phase shift decrease (increase) has reached almost -90° (+90°)
- An ideally stable control system has one dominant pole
  - at the 0 dB loop gain frequency the phase shift is ≈ 90°

Figure 9.48 Review of Poles and Zeroes

The transfer function of a practical linear regulator has multiple poles and zeroes. The first pole is dependent on the output load and output capacitor. The first zero is a function of the output capacitor and its ESR. The remaining poles and zeroes are functions of the internal circuits of the linear regulator IC (Figure 9.49 and Figure 9.50). Therefore a user only has control over the first pole and first zero.

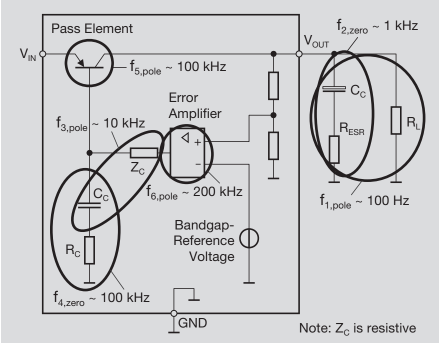
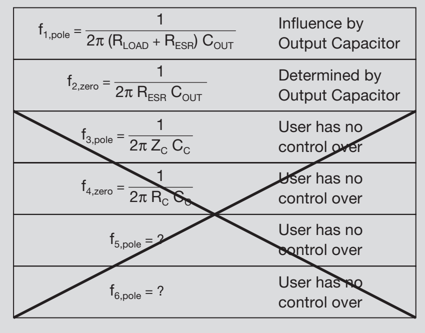


Figure 9.49 Linear Regulator Control Loop Poles and Zeroes



$f_{1,pole} = \frac{1}{2\pi (R_{LOAD} + R_{ESR}) C_{OUT}}$	Influence by Output Capacitor
$f_{2,zero} = \frac{1}{2\pi R_{ESR} C_{OUT}}$	Determined by Output Capacitor
$f_{3,pole} = \frac{1}{2\pi Z_C C_C}$	User has no control over
$f_{4,zero} = \frac{1}{2\pi R_C C_C}$	User has no control over
$f_{5,pole} = ?$	User has no control over
$f_{6,pole} = ?$	User has no control over

Figure 9.50 Linear Regulator Control Loop Poles and Zeroes

Recall, an ideally stable control loop has one pole. Therefore a design technique to stabilize a linear regulator is to cancel the effect of some of the remaining poles with zeroes to make sure that the amplitude response is reducing at the rate of -20dB/decade in the vicinity of the 0 dB frequency. The output capacitor creates a zero and this is the reason it is required for stability (Figure 9.51). The zero from the output capacitor is intended to cancel the effect of the second pole. The third pole and second zero are created internally by the regulator design and cancel each other (note that Z_C is resistive). The fourth and higher order poles typically occur at frequencies above the unity gain (0 dB) frequency where the loop gain is less than 1 and do not have significant impact on the phase margin or stability.

- Make f_1 the dominant pole, by making it rather low, ~ 100 Hz
- Make $f_2 = f_3$ so the first zero (set by the output capacitor) "zeroes out" the second pole

$$f_{2,zero} = \frac{1}{2\pi\, R_{ESR}\, C_{OUT}} = \frac{1}{2\pi\, Z_C\, C_C} = f_{3,pole}$$

- The design of the voltage regulator ensures $f_4 \cong f_5$
- This will result in a stable linear voltage regulator

Figure 9.51 Design Technique for Stability

The following example (Figure 9.52) illustrates the Bode diagram of a stable linear regulator. The dominant or first pole occurs at a 100 Hz. At the frequency of the first pole the gain begins to decrease at -20 dB per decade from 70dB gain(x3162). In addition, the pole causes a phase shift by -90° and this change in phase shift begins approximately one decade before the pole and ends one decade after the pole frequency. The effects of the next two poles on gain and phase shift are negated by zeroes. Finally the fourth pole occurs at the unity gain frequency (200 kHz) causing an ultimate phase shift of -90°. Therefore at the unity gain frequency, where the fourth pole occurs, the over-all phase shift has only changed by an additional -45° and the total of phase shift is only -135° (-90° from first pole and -45° fourth pole). The resulting phase margin is 45° so the regulator is stable.

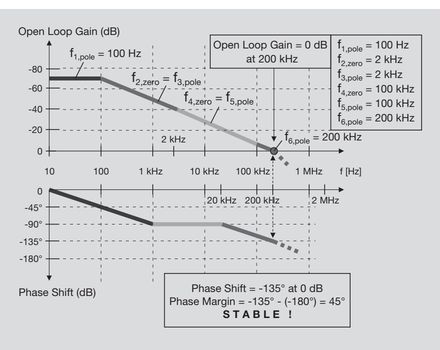
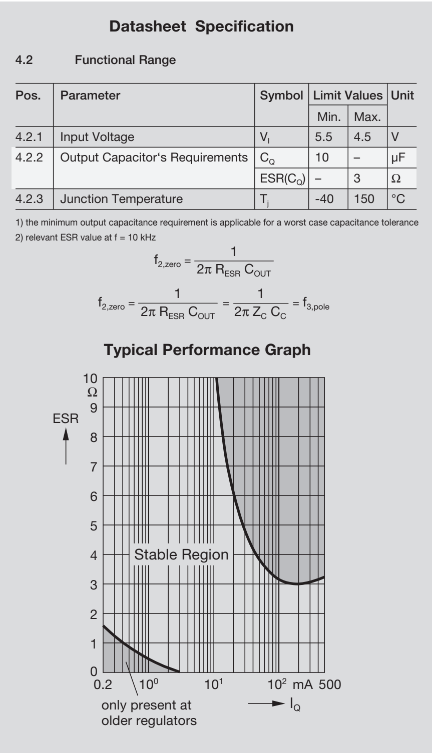


Figure 9.52 Stable Control Loop Example

Linear regulators especially LDO-s with PNP pass transistor, require an output capacitor for stability because it produces a zero which helps to offset the effect of a pole. A stability graph can typically found in a datasheet and shows a region of stability in the plot of ESR versus output current plot (Figure 9.53). In the example a small amount of ESR is required at low loads for stability. The ESR requirement for stability forms a region on the graph which is often referred to as the ESR tunnel. Some linear regulators require a small amount ESR in the output capacitor for stability but newer regulators are designed to be stable with extremely low ESR capacitors. It is very important to study the data sheet (and available application reports) issued by the manufacturer of the linear regulator when the output capacitor is selected.



Datasheet Specification

4.2 Functional Range

Pos.	Parameter	Symbol	Limit Values		Unit
			Min.	Max.	
4.2.1	Input Voltage	V_I	5.5	4.5	V
4.2.2	Output Capacitor's Requirements	C_Q	10	–	µF
		$ESR(C_Q)$	–	3	Ω
4.2.3	Junction Temperature	T_j	-40	150	°C

1) the minimum output capacitance requirement is applicable for a worst case capacitance tolerance
2) relevant ESR value at f = 10 kHz

$$f_{2,zero} = \frac{1}{2\pi\, R_{ESR}\, C_{OUT}}$$

$$f_{2,zero} = \frac{1}{2\pi\, R_{ESR}\, C_{OUT}} = \frac{1}{2\pi\, Z_C\, C_C} = f_{3,pole}$$

Figure 9.53 Stability Tunnel

10. Introduction to Switching Regulators

In this chapter we will introduce in detail the use and operation of switching regulators. We will begin with a quick review of switching regulator fundamentals. We will then examine the selection criteria for a switching regulator and linear regulator.

Next, we will introduce different types of switching regulators. Afterwards, we will present the operation of one of the most common switching regulators, the step down (buck) regulator beginning with a high level overview of the operation, followed by performng an intensive, step-by-step analysis of its workings.

Next we review the variables that go into tailoring a switching power supply for each application. We will examine the performance, size, and cost trade-offs important to switching power supply design. We will also introduce the different types of feedback control methods (voltage, combined voltage and current).

Lastly we will conclude this chapter with a section devoted to power loss. We will show practical examples of calculating power loss and efficiency.

10.1 What is a Switching Regulator?

A switching regulator is a device that converts an input voltage into a desired output voltage. For example, a switching regulator may be used to convert a 12 V automotive battery voltage to a 5 V microcontroller supply typically found in an automotive electronic control unit (ECU). Voltage conversion can also be accomplished by different methods, such as a linear regulator. However, it is the method of voltage conversion that defines the different types of regulators. A switching regulator converts a voltage by periodically switching a transistor on and off, storing energy in low loss elements such as an inductor and a capacitor (Figure 10.1). A more detailed explanation of switching regulator energy conversion will be given later in this chapter.

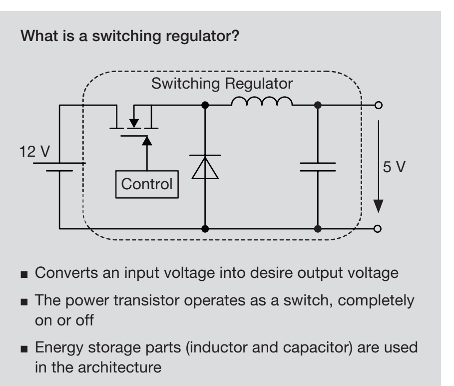


Figure 10.1 Switching Regulator Example

10.2 Why Are Switching Voltage Regulators Needed?

The primary difference between a switching and a linear regulator is in the control of the main (series pass) power transistor. Recall, that in a linear regulator a transistor between the input and output is always turned on in an analog fashion. In addition, an output capacitor is the only required external component to filter the output and stabilize the control loop of a linear regulator. Conversely in a switching regulator a transistor between the input and output is turned on and off periodically in a digital fashion. An external network of components is then used to filter the output and transform the input voltage to the switching regulator to the desired output voltage. The external network of components typically consists of an inductor, a capacitor, and sometimes a diode or MOSFET (Figure 10.1).

Most designers will use a linear regulator whenever it is possible because it is simpler to design and generally cheaper than a switching regulator. So why is a switching regulator used instead of a linear regulator? One simple reason to use a switching regulator is when the desired output voltage is greater than or opposite polarity from the input voltage (Figure 10.2). Recall, a linear regulator can only step down the output voltage or in other words convert an input voltage into a lower output voltage.

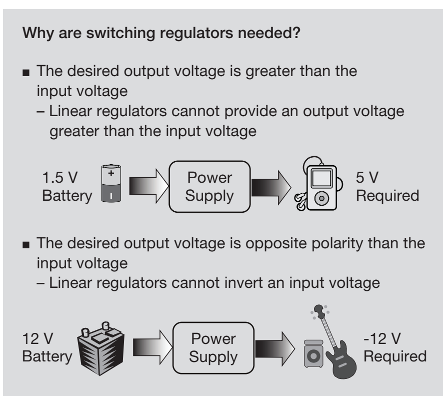


Figure 10.2 Applications only Possible with Switching Power Conversion

Furthermore linear regulators have a much lower efficiency than switching regulators under most conditions. Lower efficiency means more of power is lost as heat. Additionally, linear regulators become less efficient as their input voltages increase and hence power losses increase. This means that linear regulators may need large heat sinks to dissipate the power loss. Under high power conditions the high conversion efficiency of a switching power supply may be the only option (Figure 10.3).

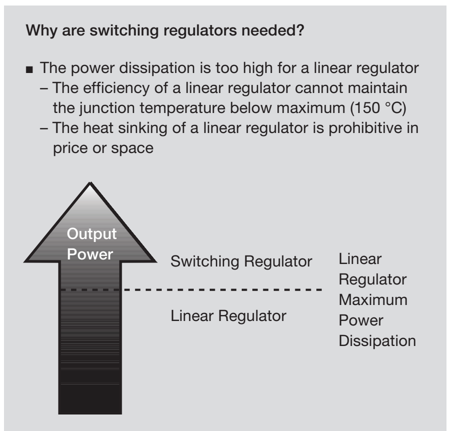


Figure 10.3 Switching Regulators are More Efficient than Linear Ones

10.3 Types of Switching Regulators

Switching regulators can be broadly categorized into four types of converters: AC-DC, AC-AC, DC-AC and DC-DC (Figure 10.4). The most common type of converter used in automotive electronics is a DC-DC converter because most cars have a 12 V_{DC} battery and most electronics operate with a DC voltage lower than 12 V. The discussion of switching regulators in this chapter will be focused on the Down, or Buck DC-DC converters.

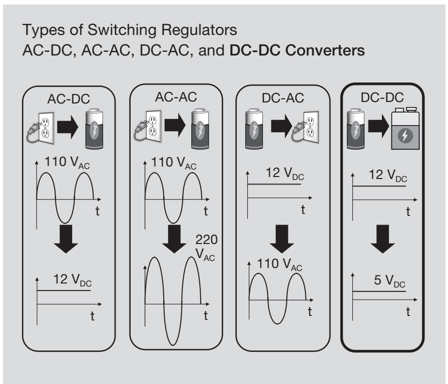


Figure 10.4 Types of Energy Converters

DC-DC converters can be classified into three basic groups; buck, boost, or buck-boost (Figure 10.5). A buck or step down converter delivers a regulated output voltage lower than its input voltage. A boost or step up converter produces a regulated output voltage higher than its input voltage. A buck-boost or inverting regulator generates a higher or lower output voltage that is opposite polarity from the input voltage.

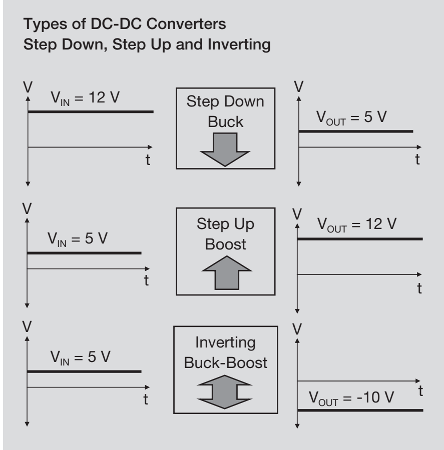


Figure 10.5 Types of DC-DC Converters

The buck, boost, and buck-boost switching regulators consist of the same basic components. The basic components are a series pass or low-side switching transistor, an inductor, a diode and a capacitor. It is the arrangement of these basic components that differentiates these basic dc-dc converters from each other.

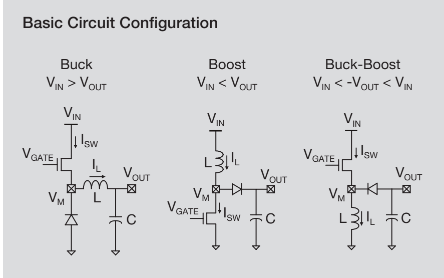


Figure 10.6 Basic Switching Regulator Topologies

If the input voltage is greater than the output voltage ($V_{IN} > V_{OUT}$), then the switching regulator must "buck" or "step down" the voltage and the switching regulator is called a

“Buck Converter” or a “Step Down Converter”. Usually, step down switching regulators can replace linear regulators with a significant efficiency improvement. Notice the input voltage does not have to be a constant DC voltage, but it must always be larger than the desired output voltage. In addition, the output voltage of the switching regulator contains a small amount of AC or ripple voltage.

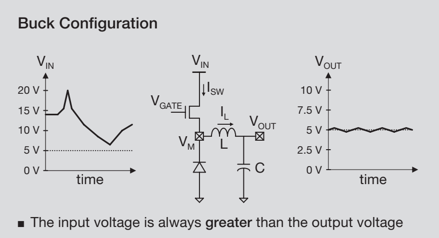


Figure 10.7 Input-Output Voltage Difference in a Buck Regulator

If input voltage is less than the output voltage ($V_{IN} < V_{OUT}$), then the switching regulator must “boost” or “step up” the voltage and the switching regulator is called a “Boost Converter” or a “Step-Up Converter”. A step up switching regulator is not commonly used in automotive applications because most automotive electronics require lower voltages than the available 12 V battery and buck converters or linear regulators are more common for that reason. Notice the input voltage does not have to be a constant DC voltage and must only be smaller than the desired output voltage. In addition, the output voltage of the switching regulator contains a small amount of AC or ripple voltage (Figure 10.8).

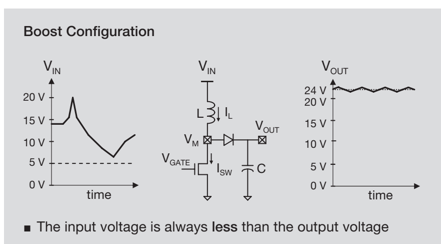


Figure 10.8 Input-Output Voltage Difference in a Boost Converter

In a buck-boost or inverting regulator (Figure 10.9) the input voltage magnitude can fluctuate above, below, or at the expected output voltage. The output voltage is opposite polarity of the input voltage. For example, with reference to the magnitude of the output, a buck-boost may convert 12 V to -5 V (buck mode) or 12 V to -15 V (boost mode). Again, the output of the switching voltage regulator has a small AC voltage (ripple) on the DC output.

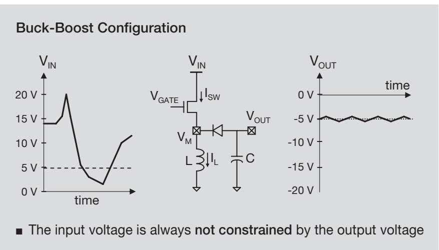


Figure 10.9 Input-Output Voltage Difference in a Buck-Boost Regulator

Other switching power supply topologies exist. However, they do not use the same external filter network components. Some of the more common topologies are listed in Figure 10.10.

Other Switching Voltage Regulator Topologies

- SEPIC
- Push-Pull and Forward Converter
- Flyback Converter

Figure 10.10 Other Switching Regulator Topologies

10.4 General Operation of a Step Down or Buck Switching Regulator

This section will explain the general operation of a switching regulator (buck). Later on in this chapter, a more detailed analysis will be presented on the buck regulator. The focus of the general operation will be on the input and output. For simplicity the ground connection is not included. To begin, a simple functional representation of a switching regulator is shown in Figure 10.11.

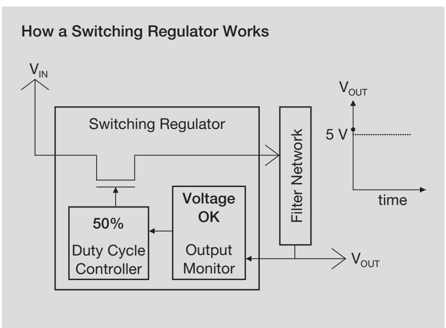


Figure 10.11 Basic Topology of a Buck Converter (grounds are not shown)

There is a transistor between the input voltage and output voltage. The transistor may either by a MOSFET or a bipolar junction transistor. This transistor is called the "switching" transistor or commonly referred to as the "switch". The output of the switching regulator is then fed into a filter network. The output of the filter network is the actual output of the switching regulator supply.

The other two functional blocks shown in the switching regulator diagram are the Output Monitor and Duty Cycle Controller. The Output Monitor circuit senses the output voltage V_{OUT} and provides the input to the Duty Cycle Controller.

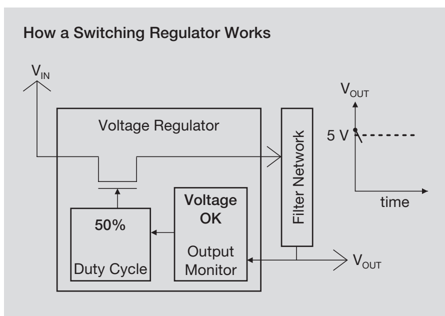


Figure 10.12 Illustration of Buck Regulator Operation

Since the duty cycle is set to 50% in this illustration, the switching transistor is on 50% and off 50% of a switching period. During the transistor off time the output voltage droops slightly because current is no longer supplied from the input to the output. The Output Monitor continues to sense the output voltage and informs the Duty Cycle Controller that the output voltage is still close to its expected value. Therefore, the Duty Cycle Controller maintains a 50% duty cycle.

At the end of the off time, the transistor is turned on again for another 50% of a switching period. Turning on the transistor again adds charge to the filter network and the output voltage rises slightly. The Output Monitor again senses the output and tells the Duty Cycle Controller that the output voltage is still very close to its expected value. Therefore, the Duty Cycle Controller maintains its 50% duty cycle (Figure 10.13).

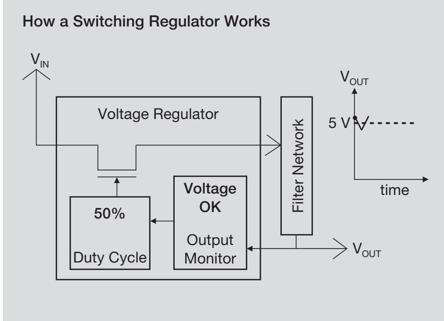


Figure 10.13 Steady State Condition at 50% Duty Cycle

Under steady state conditions the output voltage is slightly decreasing and increasing as the pas transistor is turned off and on. When the switching transistor is turned on, energy is stored in the filter network and the output voltage rises slightly. When the pass transistor is turned off, energy is drawn from the filter network, and the output voltage droops slightly. The switching action creates a steady average output voltage with some small ripple.

Assume the input voltage decreases by 1 V causing the output voltage to fall.

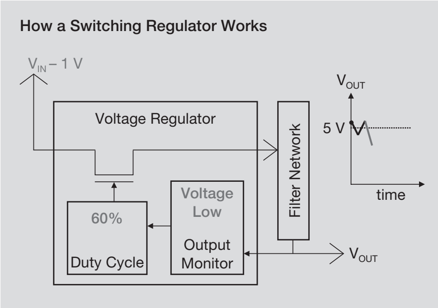


Figure 10.14 Response of Regulator to an Input Voltage Drop

The Output Monitor senses that the output voltage is lower than expected (5 V) and sends a message to the Duty Cycle Controller. In an attempt to increase the output voltage the Duty Cycle Controller increases the duty cycle or the on time of the transistor. Increasing the transistor on time increases the amount of time spent charging the filter network and effectively increases the output voltage over time. After several switching cycles with increased on time the output voltage is restored back to its expected value.

Duty Cycle Controller

Depending on the sensed output voltage the Duty Cycle Controller in a switching regulator turns on and off the transistor and adjusts the percentage of on time unlike in case of linear regulators which always have the transistor turned on.

Also shown is a graph of the output voltage of the switching regulator. Initially the pass transistor is turned on and the output of the switching voltage regulator is slightly above 5 V. The Output Monitor senses the output and tells the Duty Cycle Controller the output voltage is very close to its expected value. Therefore, the Duty Cycle Controller maintains its present duty cycle level at 50% duty cycle.

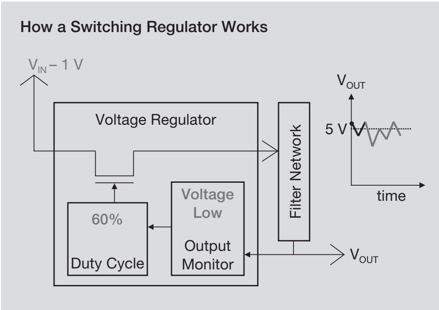


Figure 10.15 Response of Regulator to an Input Voltage Drop

Assume the input voltage returns to its previous value. The Output Monitor identifies that the output voltage is near its expected value and the Duty Cycle Controller returns to the original 50% duty cycle.

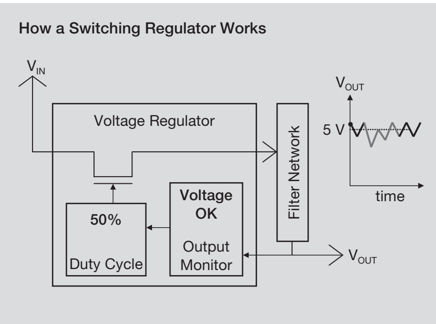


Figure 10.16 Return to the 50% Duty Cycle Mode of Operation

In summary, a switching regulator works under steady state conditions. The pass transistor is turned on, energy is stored in the filter network, and the output voltage rises slightly. When the pass transistor is turned off, energy is drawn from the filter network, and the output voltage droops slightly. The switching process causes the output voltage of a switching regulator to fluctuate about its expected value.

10.5 Switching Regulator Components

Below is a simple functional representation of a switching power supply. For now, we are just concerned with one input (input voltage) and one output (output voltage). For simplicity, the ground connection is not included. In this

section, we examine each of these blocks (Figure 10.17) individually.

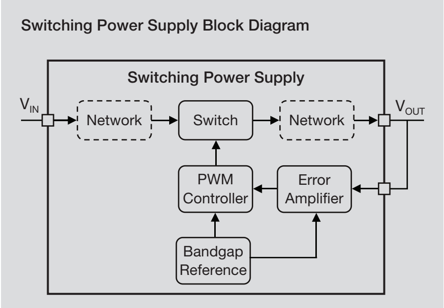


Figure 10.17 Major Switching Regulator Circuit Blocks

10.5.1 Band Gap Reference

We begin with the bandgap voltage reference. It is arguably one of the most important parts of a switching regulator. It provides the reference voltage used by the regulator to develop and maintain an accurate output voltage.

The voltage, V_{REF}, is a created by a circuit called bandgap reference (BGR). It uses several different component features to maintain a very stable reference voltage across the entire specified operating temperature range. The basic concept of the BGR circuit can be shown in Figure 10.18.

The BGR voltage is stable across a wide temperature range because it uses two offsetting characteristics of semiconductor technology. First, the base-emitter voltage of a NPN bipolar junction transistor decreases approximately at -2 mV/°C rate. This decrease in the NPN base-emitter voltage is countered by an increase in the thermal voltage of the semiconductor circuit:

$$V_t = kT / q$$

where k is Boltzman's constant (8.62×10^{-23} Joules/K), T is the integrated circuit junction temperature (in K), and q is the electron charge (1.6×10^{-19} Coulomb or As). The thermal voltage has a positive temperature coefficient of 0.085 mV/K. This is scaled by the factor $A_0 \approx 23.5$ so that the net temperature coefficient of the voltage reference is approximately 0 mV/°C.

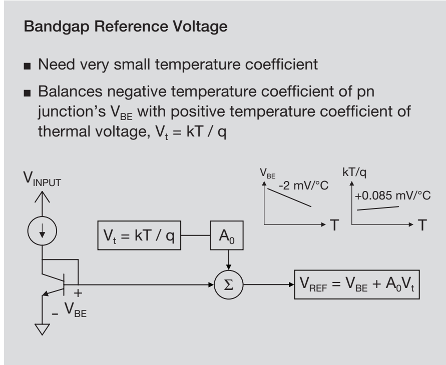


Figure 10.18 Functional Block Diagram of the BGR Circuit

A BGR circuit implementation is shown in Figure 10.19. Its reference voltage, V_{REF}, is typically 1.2 V. A "trimming network" of resistors (R_3 - R_5) is added to allow V_{REF} to be adjusted or "trimmed" during wafer testing if it deviates from its expected value because of manufacturing process variations.

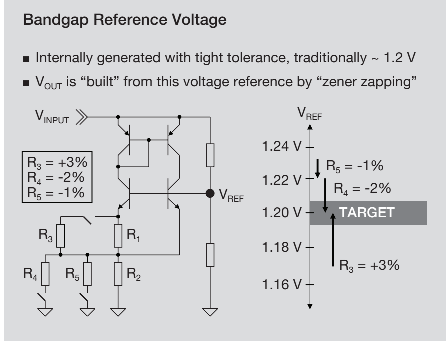


Figure 10.19 An Implementation of the Bandgap Reference Circuit

10.5.2 Error Amplifier

The next major circuit of a switching regulator is an error amplifier. The error amplifier receives two inputs, one is the BGR and the other is from a resistive voltage divider network. The resistor divider R_1 and R_2 creates a signal, V_{DIV}, which is equal to the BGR

voltage for the desired output voltage. The accuracy of the resistor ratio combined with the accuracy of the bandgap reference voltage constitutes a large portion of the overall output voltage tolerance. The error amplifier compares V_{DIV} and V_{REF} to determine how much the output voltage is off its expected value. The error amplifier amplifies the error in the output voltage of the regulator and determines the pulse width of the PWM controller output.

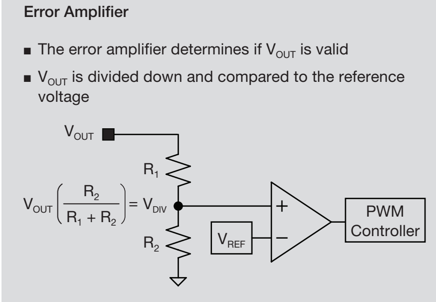


Figure 10.20 Output Voltage Divider and Error Amplifier

10.5.3 PWM Controller

The PWM controller processes the output signal from the error amplifier and adjusts the duty cycle of the switching transistor. If the output voltage is too low, the PWM controller will increase the duty cycle. This allows more energy to be delivered to the load from the input power source, and V_{OUT} will increase. If the output voltage is too high, the PWM controller will decrease the duty cycle. This allows less energy to be delivered to the load from the input power source, and V_{OUT} will decrease (see Figure 10.21).

PWM Controller

- In a switching voltage regulator, the pass transistor is used as a switch - it is either on or off
- The output voltage, however, is an analog value
- PWM controller senses error in V_{OUT} via the error amplifier
- PWM controller updates the duty cycle of the of transistor adjusting the output voltage

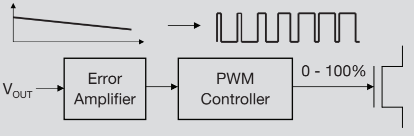


Figure 10.21 PWM Controller

10.5.4 The Series Switch

The switch is most commonly implemented with a bipolar junction transistor (BJT) or an n-channel **m**etal-**o**xide-**s**emiconductor **f**ield **e**ffect **t**ransistor (MOSFET). There are advantages and disadvantages to both types of transistors. Most old switching regulators use BJT transistors. However most new Integrated Circuit (IC) switching regulators use MOSFET transistors. The relative merits are in Figure 10.22.

Switching Transistor Bipolar and MOSFET

	Bipolar	MOSFET
	Base, Collector, Emitter	Gate, Drain, Source
Switch Speed	Slow	Fast
Drive Method	Current	Voltage
Drive Circuit	Simple	Complex
ESD Robustness	High	Low
Voltage Drop	Worse	Better (potentially)

Figure 10.22 Relative Merits of BJT and MOSFET Switching Transistors

In many medium and high power systems, however, the transistor is not physically in a switching regulator integrated circuit (IC). Rather, the switching transistor is an external component which is turned on and off by a switching regulator controller IC. External series pass transistors are generally

MOSFETs. Engineers are able to customize the switching regulator design and take advantage of the latest discrete transistor technologies. It also benefits the control IC because most of the power dissipation is transferred to the external controller IC-s allowing for higher levels of system integration in the switching regulator control IC.

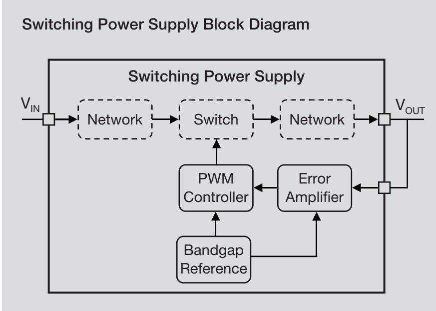


Figure 10.23 Switching Power Supply Block Diagram

10.5.5 External Network (Input and Output Filters)

Regardless of whether the switching transistor is internal or external in a switching regulator, an external network of components is used to transform switched input voltage into the desired output voltage. While the configuration of the external network components varies with the type of switching regulator, the components (inductor, diode, and capacitor) remain the same in most cases.

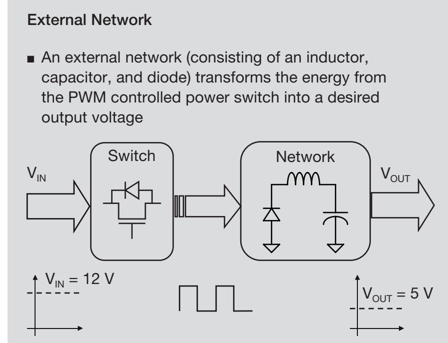


Figure 10.24 Output Network

10.6 Step Down (Buck) Switching Regulator Operation

A detailed examination of a step down or buck switching regulator operation will be explained in this section. In our explanations, we will examine how the following three voltage and two current variables vary:

I_{SW} = Switch Current

I_L = Inductor Current

V_{GATE} = Gate Voltage

V_M = Switching Node Voltage

V_{OUT} = Output Voltage

In this explanation the switching regulator is assumed to be operating in steady state continuous conduction mode (CCM). Steady state means the regulator has already been running for several cycles and CCM means the inductor current is always greater than 0 A. The graphs on the right side of Figure 10.25 reflect how the voltages and currents are varying at various moments in the step down switching regulator cycle of operation.

We begin by considering what happens when the gate voltage is increased and the switching transistor turns on. The voltage across the inductor is given by:

$$V_L = V_M - V_{OUT}$$

The voltage across the inductor is important because it determines the rate of change in current through the inductor at a specific moment in time. Assume, the switching transistor is turned on and there is very little drop across the MOSFET:

$$V_M \approx V_{IN}$$

If V_{IN} and V_{OUT} are relatively constant with only minor variations during a switching period, then V_L will is also be constant:

$$V_L \approx V_{IN} - V_{OUT}$$

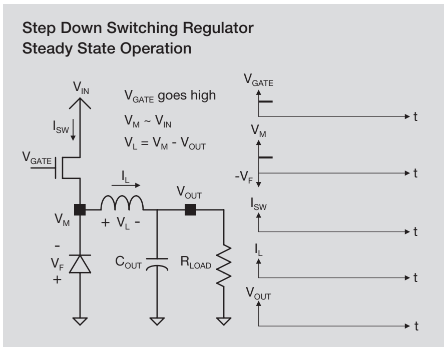


Figure 10.25 Illustration of the Fundamental Waveforms

What will happen to I_{SW} and I_L? The current rate of change through an inductor is given by:

$$\frac{di}{dt} = \frac{V_L}{L}$$

$$\frac{di}{dt} = \frac{V_{IN} - V_{OUT}}{L}$$

$$\frac{di}{dt} = \frac{\text{Constant}}{\text{Constant}} = \text{Constant}$$

Since the rate of change of inductor current is constant it will increase linearly. When the transistor is first turned on the input voltage source (e.g. battery) immediately begins to source the current through the inductor. Therefore, I_{SW} equals I_L when the transistor is on. As the current through the inductor increases, the output capacitor is charged. Recall the definition of capacitance:

$$C = \frac{Q}{V}$$

Rearranging the equation shows that the voltage across a capacitor increases as the charge on the capacitor increases:

$$V = \frac{Q}{C}$$

Therefore, the output voltage increases when the switch is turned on and the inductor current charges the output capacitor.

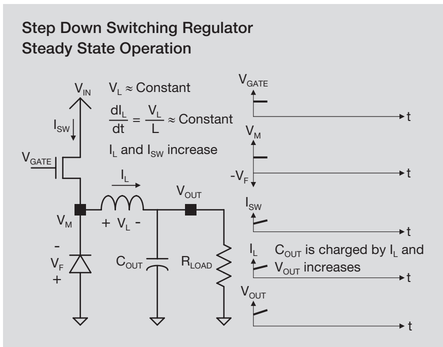


Figure 10.26 Series Pass Transistor is Switched On

After a short time, the pass transistor is turned off. Since the transistor is turned off, the switch current the very quickly falls to 0 A. However, the current through the inductor cannot instantly go to 0 A. Recall the equation for the inductor current rate of change:

$$dI_L / dt = V_L / L$$

and the current will decrease at a rate proportional to the voltage across the inductor. The inductor current cannot instantly go to 0 A because this would require $V_L \approx -\infty$.

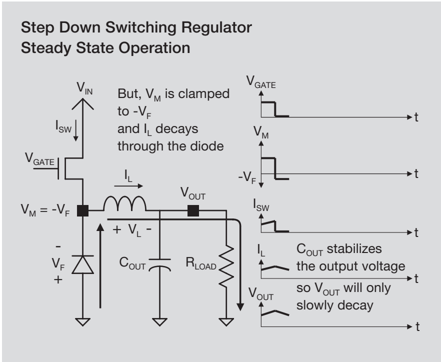


Figure 10.27 Series Pass Transistor is Switched Off

In the real world, the inductor current must decrease over some time and dI_L / dt is negative:

$V_L = V_M - V_{OUT} < 0\ V$

$V_M < V_{OUT}$

Without any limiting the inductor would pull V_M far below ground:

$V_M << 0\ V$

This happens because, it creates a very large negative inductor voltage to quickly reduce the inductor current. The very large voltage transient is very likely damaging the transistor.

To protect the step down switching regulator from a large negative voltage spike at V_M, a diode provides a recirculation path for the inductor current. This diode clamps the voltage at the left side of the inductor to one diode forward voltage drop below ground.

In Figure 10.28 the inductor current recirculation path is shown.

- Current flows out of the inductor at V_{OUT} potential
- Current flows through the load to ground
- Current flows from ground through the recirculation diode and back to V_M
- The recirculating current forward biases the diode and clamps V_M to diode voltage, $-V_F$

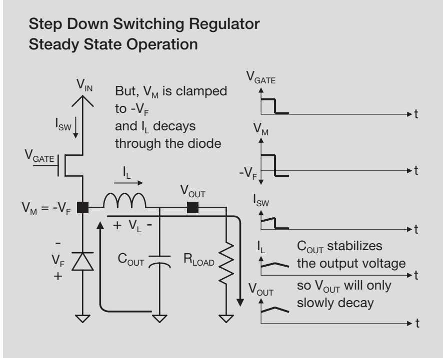


Figure 10.28 Re-circulation During the OFF Cycle

The output capacitor sources some current to the load to help stabilize the output voltage. This results in the output voltage slowly falling when the transistor is turned off. Assuming V_{OUT} is relatively constant with only minor variations during this short switching period then V_L is also constant.

$V_L \approx -V_F - V_{OUT}$

The rate of change of inductor current is equal to:

$$\frac{di}{dt} = \frac{V_L}{L} = \text{Constant}$$

Therefore, the rate of decrease of the inductor current will remain constant over time, and I_L will decrease linearly.

At the end of the off cycle the step down switching regulator will again turn on the transistor. The on cycles and off cycles are repeated over and over again at the rate of the switching frequency. Throughout this process, the output voltage will slightly ramp up and down about its expected value as the inductor and output capacitor are being charged and discharged.

The percentage of time that the transistor is on within a switching cycle is called the duty cycle, D expressed in percents. Note that D is restricted such that:

$0\% < D < 100\%$

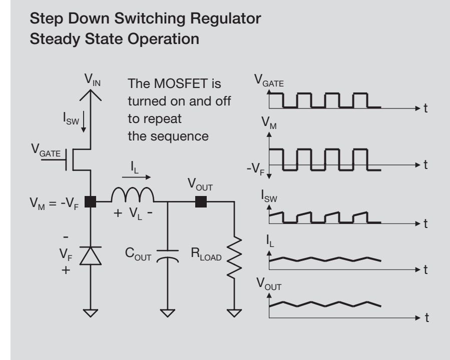


Figure 10.29 Waveforms of a Continuously Operating Buck Converter

If the duty cycle is 50%, the pass transistor is on and off for an equal amount of time. If the

duty cycle is 25%, the pass transistor is off for three times longer than it is on, etc.

To simplify the explanation of switching regulator operation a simple concept, called the "Volts-Second Principle", is introduced. The volts-second principle relates the duty cycle D to the input and output voltage. As it was previously shown, the voltage across the inductor is constant when the pass transistor is on. This will result in a linear increase in the inductor current. We also assume the voltage across the inductor is constant when the transistor is off. Therefore inductor current also decreases linearly when the transistor is off.

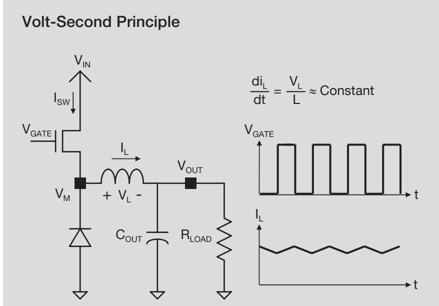


Figure 10.30 Inductor Current Waveform of a Continuously Operating Buck Converter

Figure 10.31 illustrates the waveform of the inductor voltage, V_L in one switching period, T. Initially, the pass transistor is on, and V_L is constant during the on time (DT):

$V_L = V_{IN} - V_{OUT}$ for $0 < t \leq DT$

When the pass transistor is turned off, V_L is again constant, but it is now negative for a time equal to (1 - D)T. Neglecting the voltage drop across the recirculation diode, the inductor voltage is approximately given by:

$V_L \approx - V_{OUT}$ for $DT < t \leq T$

In the steady-state condition, the inductor current must increase and decrease by the same amount. Otherwise, the inductor current would be tending toward positive infinity or zero. Therefore, the total area (or Volt-seconds) under the inductor voltage waveform must be zero within one switching period.

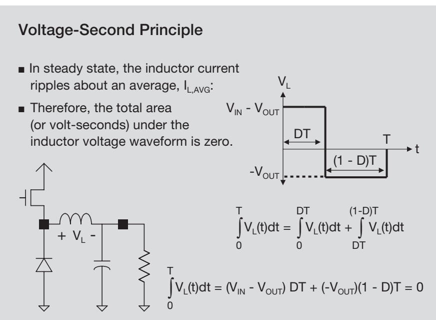


Figure 10.31 Definition of the Volt-Second Principle

If we solve this equation we can express V_{OUT} / V_{IN} transfer ratio by applying the Volt-Second principle. Conversely, the approximate value of the duty cycle of the regulator can also be calculated for given V_{OUT} and V_{IN} values.

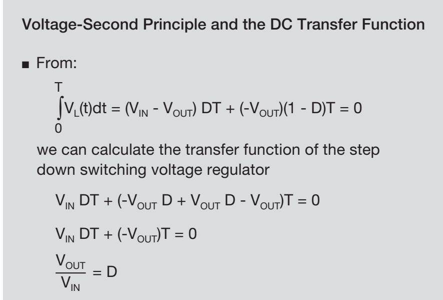


Figure 10.32 Input and Output Transfer Ratio of a Buck Converter

10.7 Design Guidelines and Examples

In this section, we will look at several design guidelines for implementing step down switching power supplies. We will begin by examining in more detail how the V_{OUT} / V_{IN} ratio determines the duty cycle D. Next, we will look at the inductor ripple current, ΔI, and how it is affected by the inductor selection. Third, we will show how ΔI also affects trade-offs between ripple in the output voltage and selection of the output capacitor, C_{OUT}, and its selection will be discussed briefly. Next, we will

present some of the options available to the power supply designer when selecting the recirculation diode and the input capacitor, C_{IN}. Afterwards, we will give a brief introduction to the efficiency of step down switching regulators. Finally, we will close this section by examining a number of different design examples that illustrate many ways to customize a step down regulator.

10.7.1 V_{OUT} / V_{IN} Transfer Ratio and Duty Cycle

First, we consider in more detail how the V_{OUT} / V_{IN} ratio determines the duty cycle of the step down switching power supply. Figure 10.33 reiterates the Volt-second principle we introduced in the previous section.

Solving for the duty cycle, we find:

$$D = \frac{V_{OUT}}{V_{IN}}$$

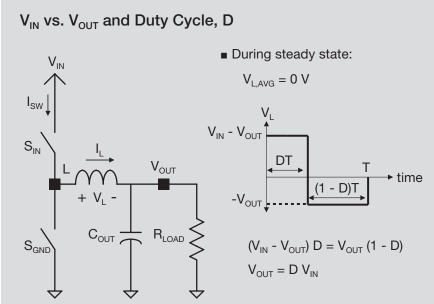


Figure 10.33 Restatement of the Volt-second Principle with an Idealized Buck Converter

Let us consider this for a moment. If V_{IN} is near to the high limit we only need to turn on the pass transistor for a small percentage of time to deliver the output power to the load. The duty cycle will be low. Likewise, if V_{OUT} is nearly the same as V_{IN} the pass transistor will have to be on for a significant portion of the period to deliver the required output power and the duty cycle will be high.

The two statements above are illustrated in Figure 10.34 and Figure 10.35:

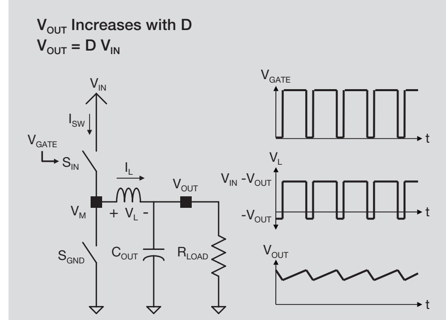


Figure 10.34 Switching Waveforms: V_{IN} - V_{OUT} is Small

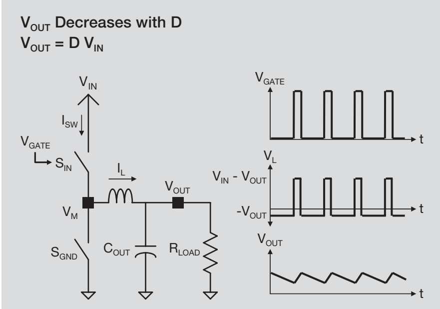


Figure 10.35 Switching Waveforms: V_{IN} - V_{OUT} is Large

The models, we have used, assume that no power is lost in the step down switching regulator. Next we will analyze the effects of the voltage drops across the step down switching transistor and recirculation diode (here modeled as switches, V_{SIN} and V_{SGND}).

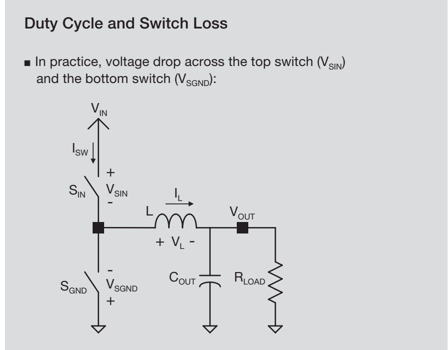


Figure 10.36 Model of a Real Driver Circuit

Figure 10.36 helps to explain how to apply the Volt-second principle to include losses in S_{IN} and S_{GND}. When the pass transistor is on, the inductor voltage is now given by:

$$V_L = V_{IN} - V_{SIN} - V_{OUT}$$

When the pass transistor is off, the inductor voltage is equal to:

$$V_L = -V_{SGND} - V_{OUT}$$

Again, the total area (or Volt-seconds) under the inductor voltage waveform must be zero within any given switching period. Therefore:

$$(V_{IN} - V_{SIN} - V_{OUT})\, D = (V_{SGND} + V_{OUT})(1 - D)$$

Solving for D, we find the duty cycle of a step down switching regulator corrected for loss in the regulator's switches.

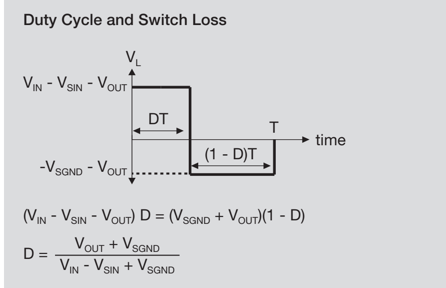


Figure 10.37 Effect of Switch Loss on the Duty Cycle

10.7.2 Inductor Current (steady and ripple)

Next, consider the inductor current and its ripple, ΔI. It is important to remember that I_L is the sum of current I_{SW} flowing from the input when the pass transistor is on and the current I_{GND} recirculating through the ground switch when the transistor is off. I_{SW} (and I_L) increases when the transistor is on. Conversely I_{GND} (and I_L) decays when the pass transistor is off. The inductor current I_L, therefore ramps about an average inductor current, $I_{L,AVG}$ as the switch S_{IN} is turned on and off and S_{GND} turned off and on (Figure 10.38).

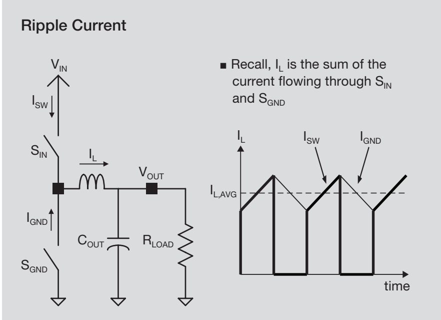


Figure 10.38 Illustration of Ripple Current and Average Current

The inductor current has both a DC and an AC component as illustrated. The DC component is $I_{L,AVG}$. The AC component is the ripple superimposed upon $I_{L,AVG}$. The peak to peak inductor ripple current is designated as ΔI, therefore the maximum inductor current is equal to $I_{L,AVG} + \Delta I/2$, and the minimum $I_{L,AVG} - \Delta I/2$.

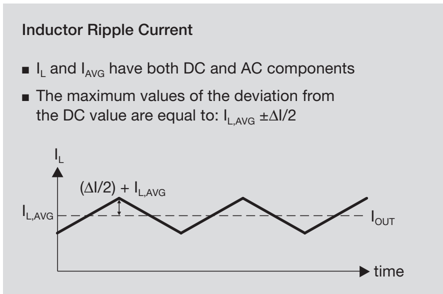


Figure 10.39 Inductor Current vs. Time

The value of the peak to peak inductor ripple current, ΔI, is one of the most important design parameter of a switching regulator. The inductor ripple current will affect almost every other choice a designer has to make for the optimization of a step down switching regulator.

Typical values of ΔI are 30% to 50% of $I_{L,AVG}$. For example, if $I_{L,AVG}$ is 1 A, ΔI could be set between 300 mA and 500 mA. In most applications, lower values of ΔI are desirable. However, lower ripple current may require more complex and expensive components.

10.7.3 Inductor Values

Once a value for ΔI has been selected (in Amperes, not as a percentage), the minimum inductance, L_{MIN}, of a step down regulator can be determined. In addition to ripple current, the minimum inductor L_{MIN} is also a function of the input voltage (V_{IN}), the desired output voltage (V_{OUT}), and the switching frequency (f_{SW}) of the PWM signal.

Selection of ΔI

- The amount of allowed inductor ripple current, ΔI, is one of the key decisions made in designing a power supply
- It is an important factor in correctly sizing the other componen in the power supply
- Typical values of ΔI are 30% to 50% of $I_{L,AVG}$
- Small values of ΔI might be desirable, but can result in more complex and expensive power supplies

Figure 10.40 Summary of Ripple Current Features

Selection of Inductance

- The minimum inductance of a step-down switching voltage regulator is given by

$$L_{MIN} = \frac{(V_{IN} - V_{OUT}) \bullet V_{OUT}}{f_{SW} \bullet V_{IN} \bullet \Delta I}$$

- If ΔI (ripple current) has already been selected and V_{OUT} is a constant:

  L will vary based upon the regulator's switching frequency (f_{SW}) and the input voltage range

Figure 10.41 Selection of Inductor

While V_{OUT} may be relatively constant, V_{IN} can vary considerably over an application's worst case operating conditions. For example, consider a step down switching power supply:

$V_{OUT} = 5\ V$
$f_{SW} = 100\ kHz$
$V_{IN(MIN)} = 9\ V$
$V_{IN(MAX)} = 27\ V$

$$L_{MIN} = \frac{(V_{IN} - V_{OUT}) \bullet V_{OUT}}{f_{SW} \bullet V_{IN} \bullet \Delta I}$$

For $V_{IN} = 9\ V$

$$\Rightarrow L_{MIN} = \frac{(9\ V - 5\ V) \bullet 5\ V}{100\ kHz \bullet 9\ V \bullet 0.4\ A}$$

$$\Rightarrow L_{MIN} = 56\ \mu H$$

For $V_{IN} = 27\ V$

$$\Rightarrow L_{MIN} = \frac{(27\ V - 5\ V) \bullet 5\ V}{100\ kHz \bullet 27\ V \bullet 0.4\ A}$$

$$\Rightarrow L_{MIN} = 102\ \mu H$$

The worst case minimum inductance (~100 µH) occurs at $V_{IN,MAX} = 27\ V$, which is almost twice the required inductance L_{MIN} (55 µH) at $V_{IN,MIN} = 9\ V$.

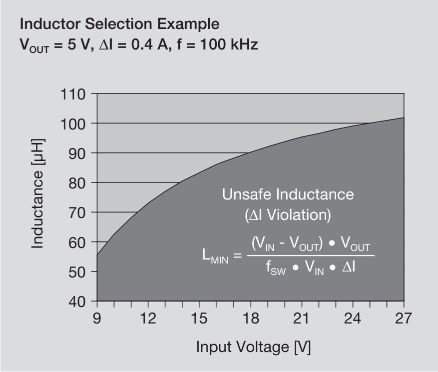


Figure 10.42 Inductance Value vs. Input Voltage

After selecting values for ΔI and L_{MIN}, the power supply designer must next check saturation current, I_{sat} of the inductor. Step down switching regulators operate on the principle of using inductors as energy transfer and storage devices to supply power to the load when the pass transistors are turned off. The energy stored in magnetic field of an inductor is given by:

$$E_L = \frac{1}{2} \bullet L \bullet I_{peak}^2$$

Inductors with a larger value of inductance can store more energy. In a common toroid inductor the inductance value is dependent on its physical size and construction.

Inductor Selection Saturation Current

- Inductors store energy in their magnetic field proportional to their inductance:

$$E_L = (1/2)\, L\, I^2$$

- The inductance of a toroidis determined by several factors:

$$L = \mu N^2 (A / \ell)$$

μ = Permeability of inductor core
N = Total number of turns in the wire coil
A = Area of a single loop in the coil
ℓ = Length of the coil wrapped around the toroid

Figure 10.43 Stored Energy and Inductance of an Inductor

One would expect the maximum energy that can be stored in an inductor for a given current is determined by its size and construction which is true for air-cored inductors. Inductors made with ferromagnetic cores saturate above a certain coil current. The saturation (maximum) current, I_{sat} is specified and if inductor current I_L is above the saturation current rating I_{sat}, the magnetic field in the inductor reaches a limit and the inductor cannot store any additional energy. This phenomenon manifests itself as a reduction of inductance and the rate of reduction may even be higher than the rate of increase of the I^2 term! It is, therefore, necessary to pick an inductor whose saturation current I_{sat} is larger than the peak inductor current in the buck-converter (Figure 10.44).

Inductor Selection Saturation Current

- An inductor, however, can only store a finite amount of energy in its magnetic field
- Above an inductor's saturation current (I_{SAT}), the inductor's permeability decreases significantly, reducing its inductance.
- This results in an increase in the regulator's ripple current:

$$\Delta I = \frac{(V_{IN} - V_{OUT})\, V_{OUT}}{V_{IN}\, f_{SW}\, L}$$

- Therefore, an inductor's (I_{SAT}) must be greater than the switching regulator's peak inductor current:

$$I_{SAT} > I_{L,peak}$$
$$I_{SAT} > I_{L,AVG} + \frac{\Delta I}{2}$$

Figure 10.44 Selection Criterion of Saturation Current

Failure to stay below the saturation current rating of an inductor can result in the step down switching regulator loss of output power or even damage to the switching transistor!

Finally, when selecting an inductor, several other factors need to be considered. These factors include the coil resistance, coil impedance, size, and cost. The designer can choose from a large number of different inductor technologies based upon the switching power supply requirements.

Inductor Technology

- There are a number of inductor technologies to choose from:
  - Drum core
  - Flat coil
  - Toroid
  - Bead
  - Wirewound
  - Planar
- In addition to inductance and saturation current, the inductor technology will also affect:
  - Inductor resistance and impedance
  - Size (length, width, height)
  - Cost

Figure 10.45 List of Inductor Type and List of other Inductor Parameters

10.7.4 Output Capacitor Sizing

After selecting ΔI and the inductor, the trade-offs inherent in selecting an output capacitor must be examined.

The proper selection of capacitor C_{OUT} is important because it, in a large part, determines the output voltage ripple which is a function of its impedance (including its **e**quivalent **s**eries **r**esistance, or ESR).

Larger values of C_{OUT} will stabilize the output voltage and reduce the magnitude of the voltage ripple. Larger values of ESR, however, will result in a greater V_{OUT} ripple voltage.

- The selection of the output capacitor affects the voltage ripple
- The output voltage ripple is a function of the output capacitor's value and its equivalent series resistance (ESR)

$$V_{RIPPLE} = \Delta I \left(ESR_{OUT} + \frac{1}{8\, f_{switching}\, C_{OUT}} \right)$$

- Low ESR capacitors (ceramic and tantalum) are recommended to minimize the output voltage ripple

Figure 10.46 How to Select an Output Capacitor

For some types of capacitors, the ripple due to its ESR can be significant, that's why low ESR capacitors are often used. Figure 10.47 shows how the voltage ripple in a step down switching power supply can significantly increase as the ESR of the output capacitor is increased.

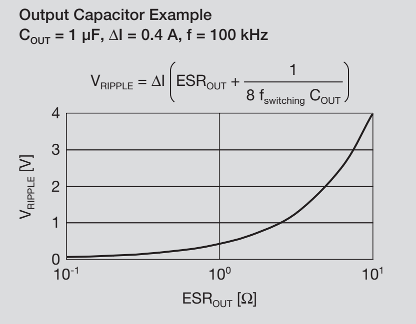


Figure 10.47 V_{RIPPLE} vs. ESR Function

In addition to the output ripple voltage, the output capacitor capacitance affects the amplitude of output voltage overshoot and undershoot caused by large load current changes.

As with the sizing of the inductor, the maximum ripple current through the output capacitor must also be considered. The output capacitor must be able to handle a worst case ripple current equal to I_{ROUT}. The maximum ripple current of the output capacitor will vary with capacitor type and temperature. The designer must ensure that the capacitor's maximum ripple current rating (in A rms) is greater than the RMS value of I_L.

Output Capacitor Current Rating

- The output capacitor must be able to handle a worst case ripple current equal to I_{OUT}
- Capacitors are rated for different maximum ripple currents as a function of their ESR, package (thermal resistance), and ambient temperature

Example Capacitor Parameters

Cap (µF)	Part #	Ripple Current (A rms)		
		25 °C	85 °C	125 °C
100	1	6.0	5.4	2.4
220	2	8.0	7.2	3.2
680	3	10.6	9.6	4.2

Figure 10.48 Dependence of Maximum Ripple Current of a Specific Type of Capacitor on Capacitance and Temperature

The breakdown voltage of the output capacitor is also an important parameter. Below the breakdown voltage limit the capacitor manufacturer guarantees the operation as specified. However, above the specified breakdown voltage the dielectric material between the capacitor's electrodes may be damaged possibly resulting in a permanent capacitor failure.

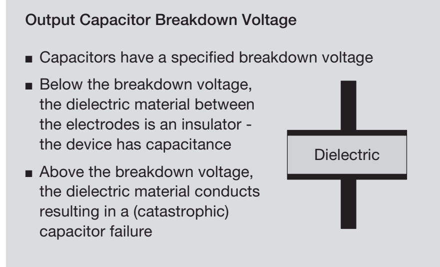


Figure 10.49 Effect of Exceeding the Breakdown Voltage

Like other capacitor parameters, the breakdown voltage is a function of the capacitor dielectric material (type of capacitor). The type of dielectric material will also influence the cost, size and performance vs. temperature of the output capacitor.

The performance of some capacitor types (particularly those with high values of capacitance and ESR) over a wide temperature range may not meet the requirements.

Therefore, in applications that must operate in a wide temperature range tantalum or ceramic output capacitors are often chosen. While not always ideal, tantalum and ceramic capacitors can exhibit significantly better performance in some step down switching power supplies than other capacitors (see Figure 10.50).

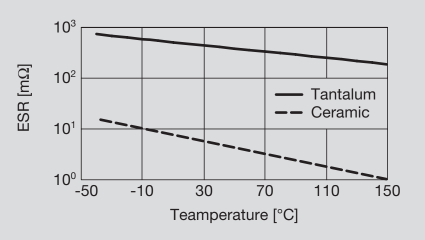


Figure 10.50 ESR vs. Temperature (of a given value)

10.7.5 Input Capacitor Selection

The input capacitor will be subjected to a rapidly changing input current with duty cycle D. To maximize the stability of the input voltage line, a low ESR capacitor should be selected for C_{IN}. The maximum RMS current through the input capacitor, the capacitance and the breakdown voltage are important parameters. The power supply designer must ensure that the maximum ripple current (in A RMS) of the input capacitor is greater than the RMS value of I_{IN}. Figure 10.51 shows how to calculate the rms value of I_{IN}.

Input Capacitor Selection

- The input capacitor is exposed to a square wave current with a duty cycle of V_{OUT} / V_{IN}
- To minimize ripple voltage on the input voltage line, an input capacitor with a low ESR should be used
- The maximum RMS current the input capacitor must withstand:

$$I_{RMS} = I_{LOAD} \sqrt{\frac{V_{OUT}}{V_{IN}}} \sqrt{1 + \frac{1}{3}\left(\frac{\Delta I}{2\, I_{LOAD}}\right)^2}$$

- The input capacitor's breakdown voltage should also be higher than the maximum input voltage

Figure 10.51 Ripple Current Calculation

10.7.6 Selection of the Recirculation Diode

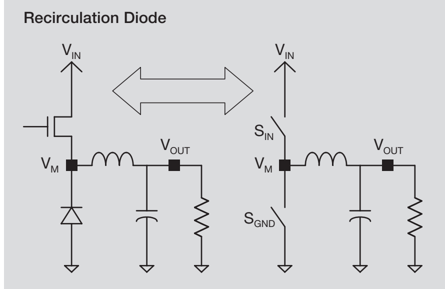


Figure 10.52 Recirculation Diode

Typically, a Schottky diode is chosen for the recirculation diode. The reverse recovery time (the time to switch from forward-bias mode to reverse-bias mode) of Schottky diodes is very short. This minimizes the power loss when the switch transitions from off to on.

In addition, the Schottky recirculation diode must have a breakdown voltage rating greater than the maximum positive voltage seen at the node V_M otherwise the Schottky diode will breakdown and may be permanently damaged.

In some step down switching regulator applications, the recirculation diode is a significant source of power loss. This is especially a problem for applications with a very small duty cycle. As the duty cycle decreases, the recirculation diode is conducting for a larger fraction of each period, increasing its power loss. This negatively impacts the efficiency of the voltage regulator.

Recirculation Diode

- The voltage drop across the recirculation diode affects the switching regulator's duty cycle and efficiency

$$D = \frac{V_{OUT} + V_{SGND}}{V_{IN} - V_{SIN} + V_{SGND}}$$

$$\eta = \frac{P_{OUT}}{P_{OUT} + P_{DISSIPATED}}$$

Figure 10.53 Voltage Drop across the Recirculation Diode Affects the Duty Cycle and the Converter Efficiency

In some applications where efficiency is an issue, the recirculation diode is replaced with a low R_{DSon} MOSFET. This transistor is switched 180° out of phase with the series pass transistor. If the voltage drop across the recirculation MOSFET is smaller than the voltage drop across a Schottky diode (0.3 - 0.4 V), less power will be dissipated when series pass transistor is off.

10.7.7 Efficiency Curves (datasheet)

Calculating the exact power dissipation in a step down switching power supply is a very involved process. One tool that power supply designers use to simplify their efforts is the typical efficiency curves found in most switching power supply datasheets.

Figure 10.54 is an example of a typical efficiency curve from a step down switching power supply datasheet. The graph shows the efficiency of the regulator vs. input voltage for three output currents at a specified output voltage. While such curves cannot be used for detailed analyses, they are very useful for making a first estimate of the power dissipation.

A more detailed estimate of the power dissipation and efficiency of step down switching power supplies will be presented in Section 10.9.

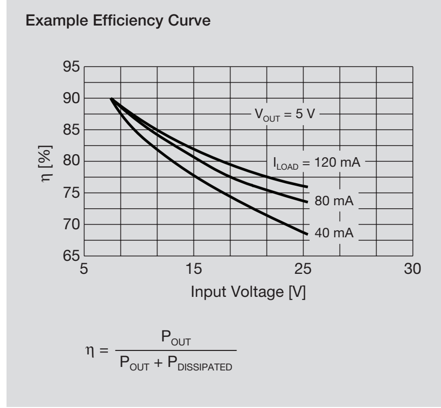


Figure 10.54 Efficiency Curves Examples (found in datasheets)

10.7.8 Examples

Now, that we have examined some of the guidelines for implementing a step down switching power supply let us look at several examples.

10.7.8.1 Example 1

First, we examine the minimum inductance required for a regulator with a 5 V output and a switching frequency of 100 kHz. We vary the input voltage from 6 V to 27 V and the ripple current from 50 mA to 1 A.

In automotive applications components must be selected for the worst case operating conditions. As the graph shows, there is a difference of almost two orders of magnitude in the minimum inductance from the best case value (approximately 10 µH at V_{IN} = 6 V and ∆I = 1 A) to the worst case one (approximately 1 mH at V_{IN} = 27 V and ∆I = 50 mA). This example illustrates why it is so important to define the real, unpadded worst case operating conditions. Overdesign will increase the cost and the size of the regulator!

10.7.8.2 Example 2

Next, we examine the minimum inductance for a step down switching power supply with ∆I = 300 mA and $f_{switching}$ = 100 kHz. This time, we vary V_{IN} from 6 V to 25 V for V_{OUT} = 5 V (see Figure 10.55). A second curve on the graph shows the minimum inductance for V_{IN} from 4 V to 25 V and V_{OUT} = 3.3 V. Again, there is almost an order of magnitude difference in the minimum size of the inductance over the entire input voltage range.

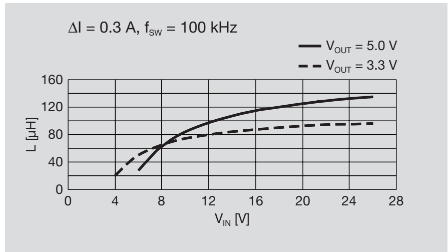


Figure 10.55 Minimum Inductance vs. V_{IN}

10.7.8.3 The Effect of Switching Frequency on the Inductor Value

In the first two design examples, we used 100 kHz switching frequency. Figure 10.56 illustrates the minimum inductance as $f_{switching}$ is varied over two orders of magnitude with ΔI again varying from 50 mA to 1 A.

While it may seem extreme to examine a variation in switching frequencies from 10 kHz to 1 MHz, this range closely mirrors the expected range of most step down switching power supplies available today. Many older (and less expensive) devices use significantly lower switching frequencies, and therefore require inductance values in the milli-Henry range. Some newer step down switching power supplies have switching frequencies in the hundreds of kHz to a few MHz. Such regulators can reduce the inductor size (and cost) significantly.

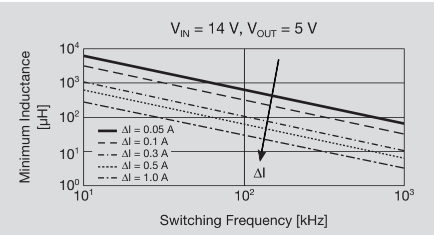


Figure 10.56 Inductance vs. Switching Frequency and Ripple Current

10.7.8.4 Output Ripple

Next, we examine how the output voltage ripple varies depending on the output capacitor and switching frequency. For a ripple current of 400 mA, and a hypothetical set of capacitors with a constant ESR of 0.25 Ω, we vary the value of the output capacitance and the switching frequency. Figure 10.57 clearly shows that for larger output capacitances, the output voltage ripple decreases. The ripple also decreases with increasing frequency. However, with the constant ESR, there will always be a minimum ripple voltage given by the product of the ESR and ripple current. In this example, we have:

$$V_{ripple,min} = (\Delta I)(ESR)$$

$$V_{ripple,min} = (0.4\ A)(0.25\ \Omega)$$

$$V_{ripple,min} = 0.1\ V$$

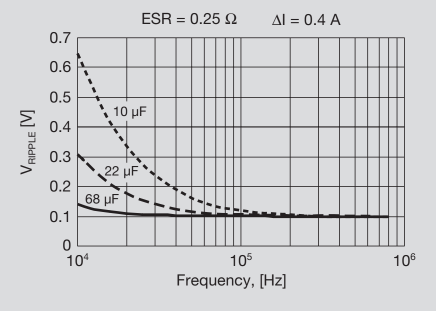


Figure 10.57 V_{RIPPLE} vs. Switching Frequency

In Figure 10.58 we illustrate how the output voltage ripple in a step down switching power supply varies with the ESR and capacitance of the output capacitor. Again, ΔI = 400 mA and $f_{switching}$ = 100 kHz.

As expected, the ripple voltage is highly dependent upon C_{OUT} for low values of ESR. However, as the ESR increases above 0.1 Ω, we see that the ripple voltage can increase appreciably.

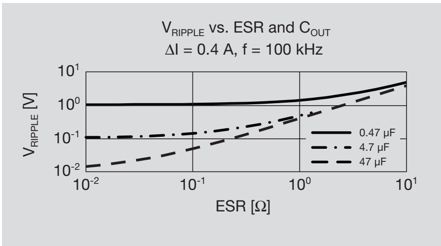


Figure 10.58 V_{RIPPLE} vs. ESR

Finally, we show how the ESR in a hypothetical capacitor can vary with temperature, resulting in variations in the output voltage ripple (Figure 10.59). For ΔI = 400 mA and a constant 22 µF output capacitance value, we vary the temperature from -40 °C to +120 °C. The ESR curve (continuous line) shows that the ESR starts at 300 mΩ at hot temperatures and increases to more than 850 mΩ at -40 °C. As a result, the output ripple voltage (dashed line)

increases from approximately 130 mV at +120 °C to 360 mV at -40 °C.

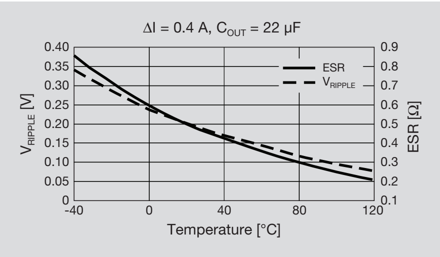


Figure 10.59 V_{RIPPLE} and ESR vs. Temperature

In the following section, we will look inside the step down switching power supplies to see how they are controlled and what can be done internally to improve their performance.

10.8 Methods of Control

In this section, we will examine two different control methods used in step down switching power supplies. We will also introduce several features to improve the operations of the switching regulators.

10.8.1 Pulse Width Modulation

At the heart of a step down switching power supply is the pulse width modulation (PWM) controller. Figure 10.60 shows several of the components of the PWM controller and a graph showing its signals during operation.

V_{OSC} is a saw-tooth waveform created by a current source feeding a capacitor. As the charge builds up on the capacitor, the amplitude of the V_{OSC} signal linearly increases. However, when V_{OSC} is greater than a reference voltage, a comparator turns on a switch which shorts the V_{OSC} node to ground. This quickly discharges V_{OSC}. The comparator then opens the switch and V_{OSC} starts to linearly increase again as it is fed by the current source.

V_{OSC} is fed into the inverting terminal of a second comparator. The feedback signal, V_{ERR} (determined by the step down switching regulator's output voltage), is connected to the non-inverting terminal. The output is the V_{PWM} signal.

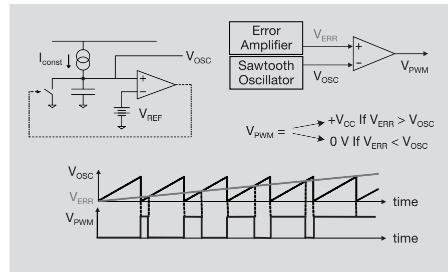


Figure 10.60 Illustration of Pulse Width Modulation

Interface circuits (not shown) control the application of the V_{PWM} signal to the series pass transistor. We will first examine a voltage-mode control loop. Then we describe a combined voltage-mode and current-mode control loop. Individually, each has advantages and disadvantages, but when operating in unison, they provide for a very stable output voltage.

10.8.2 Voltage Mode Control

Figure 10.61 illustrates a typical step down voltage regulator. Its voltage mode control loop is drawn in dashed lines. The output is divided down with a resistive voltage divider, G(s). The output of the voltage divider, GV_{OUT}, will be equal to V_{REF} when V_{OUT} is at the specified value (note, that G < 1). The GV_{OUT} signal and V_{REF} signal are fed into the error amplifier, C(s). The output of the error amplifier, V_{ERR} is converted into the V_{PWM} signal as we saw in Figure 10.60. This is how a voltage-mode control system can increase, decrease, or maintain the duty cycle of the pass transistor. The voltage mode control loop can be considered as the traditional method of regulating the output voltage of a step down switching power supply.

While a voltage mode control loop works well for constant operating conditions, it is limited in its ability to respond quickly to changes in the input voltage or output current. Voltage-mode control monitors the output voltage and adjusts the pass transistors duty cycle. The change in the input energy delivered to the

output to correct the output voltage must first propagate through the inductor. The delay of the input energy through the inductor and the filtering effect of the output capacitor are the main causes of the poor transient response.

Figure 10.61 Block Diagram of Voltage Mode Control

10.8.3 Combination of Voltage Mode and Current Mode Controls

To remedy this problem, most step down switching power supplies use a combination of voltage mode control loop and current mode control loops. In Figure 10.62 the combined loop has been added to a typical step down switching regulator in dashed lines.

In addition to monitoring the output voltage (voltage mode control), this regulator also monitors the input current flowing into the pass transistor (current mode control).

A “sample” of the input current I_{RI} is taken from I_{IN}. I_{RI} is proportional to the input current. For example, if I_{IN} were to increase unexpectedly, I_{RI} will also increase by the same percentage. The current I_{RI} is converted to a voltage, V_{RI}. This voltage can then be compared to an expected value to determine if the duty cycle of the step down voltage regulator needs to be adjusted.

Current mode control is an improved method of regulating the output voltage. Since $I_{IN} = I_L$ when the pass transistor is on, this method allows the control loop to directly sense the inductor current. When the output voltage sags, the current mode control loop will inject more current directly into the inductor to supply more current to the load. By regulating the inductor current directly, the current mode control loop is “closer” to the output voltage. This system provides excellent line and load regulation. For reasons not elaborated here, current mode control also improves the stability.

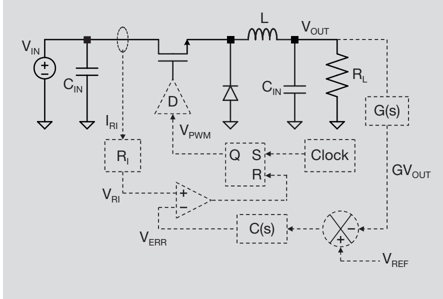


Figure 10.62 Combined Voltage and Current Mode Control (block diagram)

10.9 Special Features

Now let us turn our attention to several features which can be added to the control loop of a step down switching power supply to improve its performance.

10.9.1 Pulse Skipping

One of the primary reasons step down switching power supplies are widely used is their high conversion efficiency. To insure high conversion efficiency over a wide range of output currents of several orders of magnitude, the switching regulator may be put into a “sleep” or standby mode.

Another way to improve the efficiency is by “pulse skipping”. Every time a switching power supply cycles through a switching period its pass transistor must be turned on and off. For low load currents, the power dissipated turning the pass transistor on and off again can significantly reduce the conversion efficiency.

Maximizing Efficiency During Low Output Currents

- Switching regulators are often used to maximize power conversion efficiency
- This must often occur for large variations in operating current
- Switching power losses, however, can dramatically reduce a regulator's efficiency at very low duty cycles

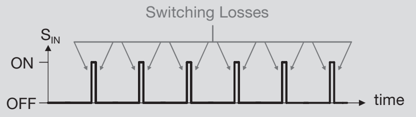


Figure 10.63 Losses Occur Every Time the Pass Transistor is Turned-on

Although various methods exist to implement pulse-skipping, they all fundamentally perform the same function. During times of low output current draw, the step down voltage regulator will simply "skip" some of the turn-on and turn-off cycles of the pass transistor. While this might translate into a longer on-time every other cycle (or every third or fourth cycle, etc.), it can dramatically reduce the number of times the pass transistor is switched on and off.

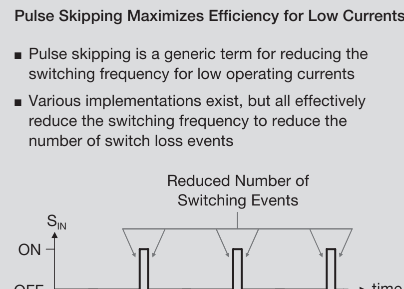


Figure 10.64 Illustration of Pulse Skipping

10.9.2 External Clocking

Next, we consider the benefits of externally clocked switching power supplies. In most step down switching power supplies, the switching frequency is internally set as we saw with the V_{OSC} signal above. The actual frequency of V_{OSC} will vary from its expected value due to processing variations and operating conditions. However, without a definitive switching frequency, the regulator's frequency spectrum will be unknown, which may complicate system design.

To eliminate this problem, some step down switching power supplies are designed to accept a synchronization signal from an external clock source allowing to run the voltage regulator at a fixed frequency.

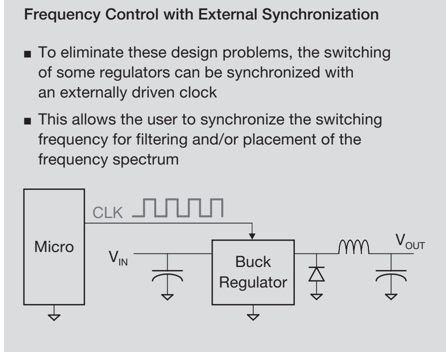


Figure 10.65 External Synchronization

10.9.3 Soft Starting

Another potential design problem is due to the large start-up, or in-rush currents. When a step down switching power supply is turned on for the first time, the output voltage is 0 V causing the switching power supply to draw a very large transient input current as the output capacitor is being charged. This large input current surge may induce voltage transients on the input voltage line which could interfere with the operation of other components powered by the V_{IN} power source.

Current at Start-Up

- When switching voltage regulators are turned on, V_{OUT} is far from its expected value
- The switching power supply will draw a large amount of current from the input power supply

V_{IN}, I_{IN}, S_{IN}, I_L, V_{OUT}, I_C, I_{LOAD}, C_{OUT}, R_{LOAD}, S_{GND}

Figure 10.66 How Inrush Current Occurs at Start-Up

To minimize any problem, many step down switching power supplies are starting softly (soft-start feature). The voltage regulator is designed to slowly increase the duty cycle of the pass transistor as the output voltage reaches its expected value (Figure 10.67). This reduces the possibility of the switching power supply causing unacceptable fluctuations on the V_{IN} power line.

Soft Start Limits Current at Start-Up

- The generic term "soft start" refers to a number of different methods used to limit in-rush current at start-up
- The maximum duty cycle is increased to its nominal value
- This slowly allows the inductor current to increase to its nominal value

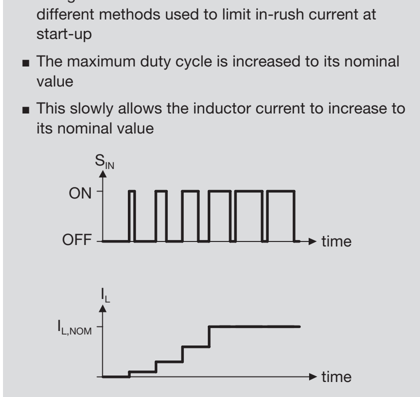


Figure 10.67 Soft Start-Up by Means of Pulse Width Modulation Control

10.9.4 Multiphase Switching

Even with all of the features we discussed above, some conventional step down switching power supplies still cannot provide the expected performance. One inherent limitation is due to the maximum switching frequency. If a large change occurs in the operating conditions of the step down switching power supply, the control loop will not be able to respond quickly enough to keep the power supply in regulation.

Also, even with external (discrete) pass transistors, power dissipation issues can arise in conventional step down switching power supplies in case of high-power systems. Power supply design is further complicated in large current systems (> 20 A) due to correspondingly larger inductor ripple currents and output voltage ripple. For very high power systems, sufficiently large inductors and output capacitors may be prohibitively large and expensive (or even unavailable).

Therefore, power supply designers will often resort to multiphase power supplies. These systems are essentially multiple switching power supplies operating out of phase with each other. Each regulator shares most of the regulation controller function and output load power responsibilities. Multiple switching increases the effective switching frequency of the multiphase regulator. In addition, it shares the power dissipation across a larger number of pass and recirculating transistors, which can simplify the thermal design. Figure 10.68 illustrates the Basic topology of a two-phase switching regulator.

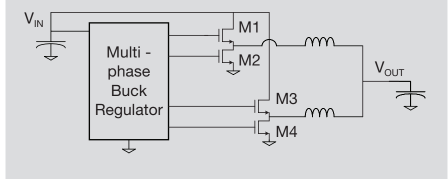


Figure 10.68 Topology of a Two-phase Switching Regulator

Figure 10.69 shows the multiphase step down switching power supply during the first phase of its operation. Transistors M1 and M4 are on,

and M2 and M3 are off. Input current from V_{IN} flows through M1 to V_{OUT}. At the same time, current is recirculating from the lower inductor in the figure through M4. The output current is sourced through both M1 and M4.

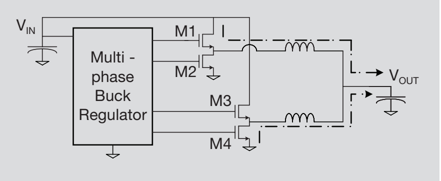


Figure 10.69 Current Flow in the First Phase

In phase two, M1 and M4 are turned off and M2 and M3 are turned on. Input current from V_{IN} flows through M3 to V_{OUT}. At the same time, current is recirculating from the inductor at the top of the figure through M2. The output current is sourced through both M2 and M3 through the lower inductor (Figure 10.70).

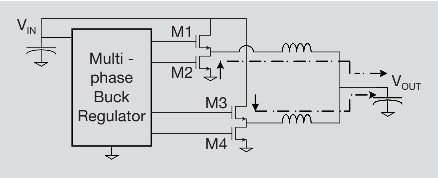


Figure 10.70 Current in the Second Phase

The advantages of multiphase step down voltage regulators in high current applications are:

1) Higher effective frequency improves efficiency (effective operating switching frequency is equal to the product of the controller frequency and the number of phases)

2) Higher effective frequency can allow each phase to work at a lower (potentially optimum) frequency equal to the controller frequency

3) Higher effective frequency also improves the transient response

4) Each external pass and recirculation transistor handles reduced current and dissipates less power thus simplifying the thermal design

5) The net ripple current and output voltage ripple are reduced compared to conventional step down switching power supplies

Multiphase step down switching power supplies are more complex to design and could be more expensive. However, if a conventional single phase power supply cannot provide the required performance, a multiphase power supply may be a good solution.

10.10 Switching Losses and Efficiency

In this final section, we turn our attention to calculating the power dissipation and efficiency of step down switching power supplies. We begin by reviewing the basics of power dissipation in linear voltage regulators. Next, we examine the static and switching losses in the series pass transistor and recirculating diode and power losses in the output capacitor. Finally, we conclude with typical and "worst case" examples of power dissipation and efficiency calculations.

10.10.1 Review of Power Dissipation and Efficiency of Linear Regulators

In linear voltage regulators, the pass transistor is always on. The voltage drop of V_{IN} - V_{OUT} across the pass transistor is present at all times. We can approximate the input power of a linear voltage regulator as:

$$P_{IN} = V_{IN} (I_{OUT} + I_{GND})$$

where I_{GND} is the current needed for the operation.

The output power of a linear voltage regulator is:

$$P_{OUT} = V_{OUT} I_{OUT}$$

Figure 10.71 presents the calculations for its power dissipation and efficiency. In this particular example, the efficiency of the linear voltage regulator is less than 30%. While this

might seem low, it is typical of any linear voltage regulator with a V_{IN} / V_{OUT} ratio of 3-to-1.

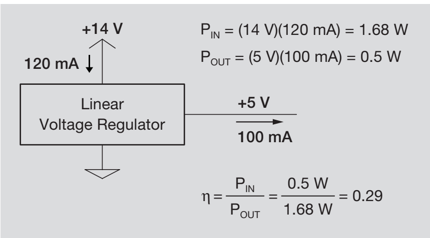


Figure 10.71 Linear Regulator: Power Dissipations and Efficiency

In Figure 10.72 we see the efficiencies of a linear regulator and step down switching power supply vs. load current. At low load currents, the linear regulator efficiency is less than 30% and it only increases to about 43% at high load currents (where the effect of I_{GND} is minimized). For the step down switching power supply, we see that the efficiency is again low for small operating currents but reaches over 80% for larger load currents.

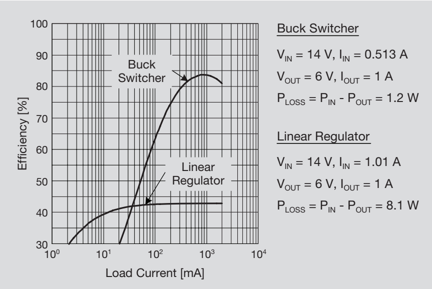


Figure 10.72 Efficiency of Switching and Linear Regulators

10.10.2 Switching Losses

To better understand all the different power loss mechanisms in a step down switching power supply, we begin by considering Figure 10.73. Before time 1, the pass transistor is off. The voltage across the transistor, V_{DS}, is approximately V_{IN}, and I_{IN} is 0 A.

At time 1, the MOSFET starts to turn on. I_{IN} begins to increase, but V_{DS} is relatively unchanged until $I_{IN} \approx I_{RECB}$. The MOSFET is in saturation until time 2 and switching losses occur in the series pass switch.

At time 2, the MOSFET moves out of saturation into the linear region. V_{DS} rapidly falls and I_{IN} begins to increase at a slow rate in response to the rising source voltage. V_{DS} would approach 0 V if the pass transistor were an ideal switch.

At time 3, V_{DS} has fallen to its minimum value, and the MOSFET is fully on. At this point, the switching loss, associated with turning on the pass transistor ends. After time 3, I_{IN} continues to increase as a ramp function because the voltage difference across the inductance is almost constant and is equal to V_{IN} - V_{OUT}. The ramping up current results in V_{DS} increasing slightly as a product of I_{IN} and R_{DSon}. This voltage drop across the non-ideal MOSFET switch will result in a static power loss.

At time 4, the pass transistor begins to turn off. There is an increase in V_{DS} as the MOSFET moves out of the linear region. I_{IN} begins to decrease at a slow rate as long as $V_{OUT} \leq V_{DS}$.

In the interval of 5 to 6, the recirculation diode begins to conduct the inductor current, the pass transistor moves into its saturation region and I_{IN} is falling at a faster rate. Switching power losses will continue to occur until I_{IN} has fallen to 0 A at time 6.

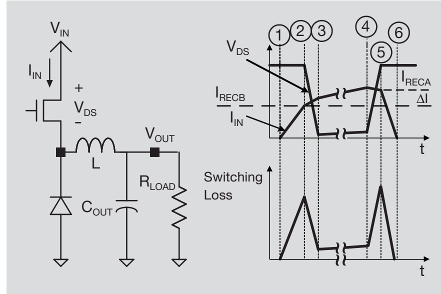


Figure 10.73 Power Losses in Buck Converters

Note: The t_{swon} and t_{swoff} periods are not to scale.

10.10.3 Static Power Losses

First let us consider the static power loss in the pass transistor when it is on (Figure 10.74).

As expected, the power losses when the pass transistor is on are primarily due to the drain-source on resistance, R_{DSon} of the MOSFET.

R_{DSon}, however, is affected by a number of different factors, including the MOSFET gate-source voltage (V_{GS}), the drain-source current (I_{DS}), and the semiconductor junction temperature (T_J).

It is difficult to represent R_{DSon} varying with V_{GS}, I_{DS}, and T_J on a single graph. Therefore, it is common to see R_{DSon} as a function of two of these three variables on a single graph in a datasheet. For example, in Figure 10.74 we see R_{DSon} can be increased several times over a range of gate-source voltages and drain-source currents at constant temperature.

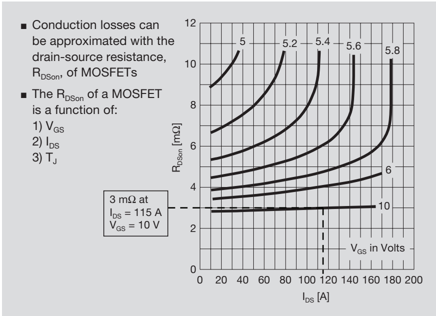


Figure 10.74 Forward Conduction Losses

Figure 10.75 illustrates how R_{DSon} varies with the junction temperature. It is not uncommon for the resistance to double as the junction temperature is increased from room temperature (25 °C) to an absolute maximum temperature (150 °C).

Note that Figure 10.75 only shows a single R_{DSon} curve which would be specified with V_{GS} and I_{DS} parameters. If necessary the R_{DSon} at additional values of V_{GS} and I_{DS} vs. temperature can be estimated. The factor α is calculated from an R_{DSon} vs. T_J curve obtained from characterization data of the MOSFET or by measurements.

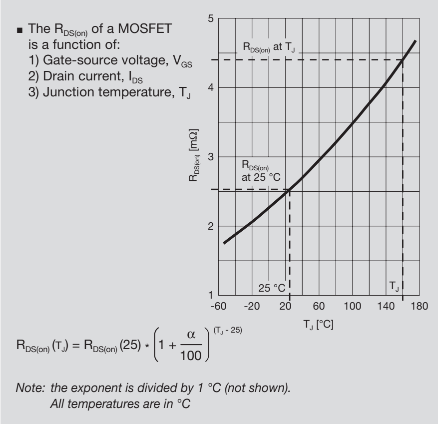


Figure 10.75 Typical Temperature Dependence of R_{DSon}

Next the power loss in the recirculation diode is considered. An example of the forward current through the diode is shown in Figure 10.76. After the pass transistor is turned off, current I_L starts at I_{peak} and slowly decays during the recirculating (i.e. off) period while power, equal to the product of the forward voltage drop and coil current, is dissipated in the diode. When the pass transistor is off, the diode is conducting the entire (yet decaying) inductor current. Since the pass transistor is on for a given duty cycle, D, the diode will be forward conducting for a percentage of time given by (1 - D).

When the pass transistor turns on again, the diode forward current quickly falls toward 0 A. After the diode's forward current falls to 0 A, the current through the diode begins to increase in the opposite direction (reverse recovery) while increasing reverse voltage is appearing at its cathode. After the diode has completed its reverse recovery, the recovery current is equal to 0 A. This transient happens very quickly in Schottky diodes. When slow recovery diodes are used, the recovery takes several times longer, further increasing the power dissipation. Modern, high frequency switchers simply don't work with regular and slow response power diodes, fast recovery diodes may switch in sufficiently short times but their forward voltage drop is higher than that of Schottky diodes, thus further increasing the power losses.

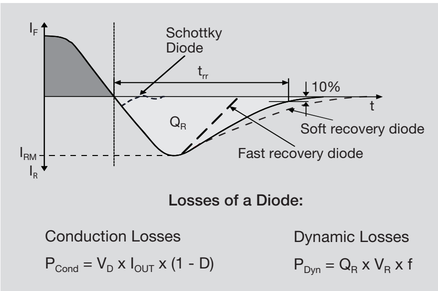


Figure 10.76 Power Losses in the Recirculating Diode

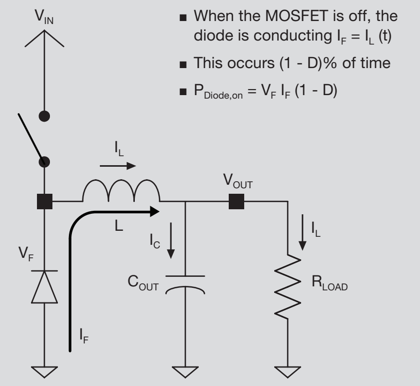


Figure 10.77 Illustration of Recirculating Diode Current

Now, that we have a better understanding of the sources of power dissipation in a step down switching power supply, it is time to examine some of the approximations generally used to estimate power dissipation and efficiency.

Figure 10.78 illustrates the power losses associated with a MOSFET pass transistor. It is comprised of three parts. First, we include the static power loss when the MOSFET is on. Second, we need to consider the power dissipated in the gate driving circuitry charging the MOSFET gate capacitance (C_G). Finally, we include the power dissipated in the MOSFET as it switches on and off.

The current through the high side switch is given by:

$$i_{on}(t) = \left(i_L - \frac{\Delta i_L}{2}\right) + \frac{\Delta i_L}{DT}\, t\text{ , for } 0 < t \le DT$$

Where i_L is the inductor current, Δi_L is the peak-to-peak ripple, D is the duty cycle and T is the period of switching.

The average R_{DSon} power dissipation, when the high side switch is on, is calculated as from the integral:

$$P_{RDSon} = \frac{R_{DSon}}{T}\int_0^{DT}[i_{on}(t)]^2 dt = DR_{DSon}\left(i_L^2 + \frac{\Delta i_L^2}{12}\right)dt$$

During the recirculation period the current is given by:

$$i_{off}(t) = i_L + \frac{\Delta i_L}{2} - \frac{\Delta i_L}{(1 - D)T}t\text{ , for } 0 < T \le (1 - D)T.$$

Notice, that the time is reset to 0 at DT to simplify the integration:

$$P_{Diode} = V_F \int_0^{(1-D)T} i_{off}(t)dt = (1 - D)V_F i_L$$

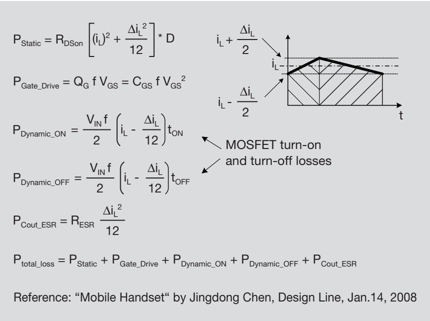


Figure 10.78 Calculation of Dissipation in the Power Transistor and in ESR

Next, we examine the power dissipated in the output capacitor as the inductor current ripple and output voltage ripple charge and discharge C_{OUT}. Figure 10.79 illustrates the output.

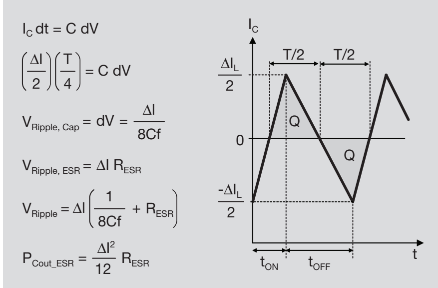


Figure 10.79 Calculations of Losses in the Output Capacitor (caused by ESR)

Finally, we include the switching power loss in the recirculation diode. The switch on losses of the diode are a function of the reverse recovery charge (Q_{rr}) of the diode, the input voltage and switching frequency of the system. Since the turn-off power losses of the diode are significantly smaller than the turn-on power loss, they are typically neglected.

In total, the power losses in a step down switching power supply are the sum of P_{Static}, $P_{GateDrive}$, $P_{Dynamic}$, P_{Cap}, $P_{Diode,on}$, and $P_{Diode,sw}$ as elaborated in Figure 10.80.

- The losses in the buck regulator are approximately equal to the sum of:
  1) MOSFET conduction ("static") losses
  2) Diode conduction losses
  3) MOSFET switching losses
  4) MOSFET gate drive switching losses
  5) Capacitor switching losses
  6) Diode switching losses

$$P_{Loss} = P_{MOSFET_Static} + P_{Diode_on} + P_{MOSFET_sw} + P_{MOSFET_Gate} + P_{Cap_sw} + P_{Diode_sw}$$

Figure 10.80 List of Losses in a Buck-regulator

Once the power loss in the step down switching power supply is calculated it is simple to find the efficiency of the power supply using the equation for η:

$$\eta = P_{OUT} / (P_{OUT} + P_{LOSS})$$

Let us now take a look at the power dissipation and efficiency for a typical case step down switching power supply. The parameters for the system and the power supply are shown in the following list:

System Design Parameters

- V_{IN} = 13.5 V
- V_{OUT} = 5 V
- I_{OUT} = 500 mA

Buck Regulator Parameters

- R_{DSon} = 0.33 Ω
- f_s = 100 kHz
- V_F = 0.4 V
- R_{ESR} = 100 mΩ
- ΔI = 150 mA
- Q_G = 4.6 nC
- Q_{rr} = 140 pC
- V_{GS} = 12 V
- t_{rise} = 100 ns
- t_{fall} = 50 ns

Figure 10.81 Parameters for System and Power Supply

R_{DSon}	The drain-source resistance of the MOSFET pass transistor.
f_s	Switching frequency of the power-supply
V_F	Forward drop across the recirculation diode
R_{ESR}	Equivalent series resistance of the output capacitor
ΔI	Inductor ripple current
Q_G	MOSFET gate charge
Q_{rr}	Diode reverse recovery charge
V_{GS}	Typical gate-source voltage when MOSFET is on
t_{rise}	Turn-on time of MOSFET
t_{fall}	Turn-off time of MOSFET

In the following calculations $I_{OUT} = I_{AVG}$ simplification will be used.

$$I_{OUT} = I_{AVG} = 0.5\text{ A}$$

$$D \approx \frac{V_{OUT} + V_F}{V_{IN} - R_{DSon}\ I_{OUT} + V_F} =$$

$$= \frac{5 + 0.4}{13.5 - 0.33 \times 0.5 + 0.4} = 0.393 = 39.3\%$$

$$P_{Static} \approx R_{DSon}\left[(I_{OUT})^2 + \frac{(\Delta I)^2}{12}\right] \times D =$$

$$= 0.33\left(0.25 + \frac{0.0225}{12}\right) \times 0.393 = 32.7 \text{ mW}$$

$$P_{Diode(on)} \approx V_F\, I_{OUT}(1 - D) =$$

$$= 0.4 \times 0.5 \times (1 - 0.393) = 121.4 \text{ mW}$$

$$P_{Dynamic} \approx \frac{I_{OUT}\ R_{DSon}\ f}{2}\left(I_{OUT} + \frac{\Delta I}{12}\right)\ (t_{d(on)} + t_{d(off)})$$

$$= \frac{0.5 \times 0.33 \times 10^5}{2}\left(0.5 + \frac{0.15}{12}\right)(100 + 50) \times 10^{-9} =$$

$$= 0.63 \text{ mW}$$

$$P_{Gate_Drive} = f \times Q_G \times V_{GS} = 5.52 \text{ mW}$$

$$P_{Cap_sw} = \frac{(\Delta I)^2}{12} R_{ESR} = \frac{0.0225 \times 0.1}{12} = 0.19 \text{ mW}$$

$$P_{Diode_sw} = \frac{1}{4} Q_{rr} V_{IN} f = \frac{1}{4} 140 \times 10^{-12} \times 13.5 \times 10^5 =$$

$$= 0.047 \text{ mW}$$

Summing the loss terms will give:

$$P_{LOSS} = (32.7 + 121.4 + 0.63 + 5.52 + 0.19 + 0.047) =$$

$$= 160.49 \text{ mW}$$

and the efficiency will be equal to:

$$\eta = \frac{V_{OUT}\ I_{OUT}}{V_{OUT}\ I_{OUT} + P_{LOSS}} = \frac{5 \times 0.5}{5 \times 0.5 + 0.16049} =$$

$$= 0.9397 \rightarrow 94\%$$

Next, the same basic step down switching power supply is reexamined while operating under less favorable conditions. Note that “the less favorable conditions” are better than the absolute worst case scenario and only five of the key system and regulator parameters are modified to show how they can affect the power supply efficiency.

Let the maximum input voltage be 26 V corresponding to a double battery condition in an automotive application. The output voltage is assumed to be at the specified minimum of 4.75 V (5% below a 5 V typical value). The maximum load is increased to 550 mA caused by a hypothetical increase in the power demand by the circuits supplied by the switching power supply.

In addition to three system parameters, two other component values are modified: the R_{DSon} value is increased from 0.33 Ω to 1 Ω to account for variations in processing and temperature and the forward drop across the recirculation diode is also increased by 100 mV to 0.5 V.

System Design Parameters	Buck Regulator Parameters
■ V_{IN} = *26 V*	■ R_{DSon} = *1 Ω*
■ V_{OUT} = *4.75 V*	■ f_s = 100 kHz
■ I_{OUT} = *550 mA*	■ V_F = *0.5 V*
	■ R_{ESR} = 100 mΩ
	■ ΔI = 150 mA
	■ Q_G = 4.6 nC
	■ Q_{rr} = 140 pC
	■ V_{GS} = 12 V
	■ t_{rise} = 100 ns
	■ t_{fall} = 50 ns

Figure 10.82 Parameters for System and Power Supply

$$D \approx \frac{V_{OUT} + V_F}{V_{IN} - R_{DSon}\ I_{OUT} + V_F} =$$

$$= \frac{4.75 + 0.5}{26 - 1.0 \times 0.55 + 0.5} = 0.2023 = 20.23\%$$

$$P_{Static} \approx R_{DSon}\left[(I_{OUT})^2 + \frac{(\Delta I)^2}{12}\right] \times D =$$

$$= 1.0 \times \left(0.3025 + \frac{0.0225}{12}\right) \times 0.2023 = 61.6 \text{ mW}$$

$$P_{Diode(on)} \approx V_F\, I_{OUT}\, (1 - D) =$$

$$= 0.5 \times 0.55 \times (1 - 0.2023) = 219.4 \text{ mW}$$

$$P_{Dynamic} \approx \frac{I_{OUT}\ R_{DSon}\ f}{2}\left(I_{OUT} + \frac{\Delta I}{12}\right)(t_{d(on)} + t_{d(off)})$$

$$= \frac{0.55 \times 1.0 \times 10^5}{2}\left(0.55 + \frac{0.15}{12}\right)(100 + 50) \times 10^{-9} =$$

$$= 2.32 \text{ mW}$$

$$P_{Diode_sw} = \frac{1}{4}Q_{rr}V_{IN}f = \frac{1}{4}140 \times 10^{-12} \times 26 \times 10^{5} =$$

$$= 0.091 \text{ mW}$$

$$P_{LOSS} = (61.6 + 219.4 + 0.149 + 5.52 + 0.38 + 0.091) =$$

$$= 287.14 \text{ mW}$$

$$\eta = \frac{V_{OUT}\ I_{OUT}}{V_{OUT}\ I_{OUT} + P_{LOSS}} = \frac{4.75 \times 0.55}{4.75 \times 0.55 + 0.28714} =$$

$$= 0.9009 \rightarrow 90.1\%$$

From the above calculations it is clear that the changes in the system, and voltage regulator parameters did not significantly change the efficiency of the step down voltage regulator. However, it is essential to consider the typical case and worst case(s) when designing a step-down switching power supply. For example the efficiency for several orders of magnitude variation in I_{OUT} (unlike the +10% increase we considered here).

11. Electrostatic Discharge, Electrical Over Stress, and Safe Operating Area

In this chapter, we will be examining some of the most common causes of semiconductor device failure and ways to address or prevent these problems.

To begin, the concept of electrostatic discharge (ESD) is introduced. We will discuss the physics behind ESD and how it can damage semiconductor components. We also review what integrated circuit designers can do to reduce the probability of ESD damage both at device level and system level.

The next section in this chapter is about the electrical over stress (EOS) and the description of how it differs from ESD. We will examine common failure modes from EOS and show how these failures modes are distinctly different from typical ESD damages.

Lastly, the chapter is concluded by introducing the concept of safe operating area (SOA). The SOA is usually specified within a power semiconductor datasheet. Strict adherence to limits of its SOA when operating a semiconductor device will make every designer’s life easier and, more importantly, the design more robust.

11.1 Electrostatic Discharge (ESD) Stress

We are all familiar with electrostatic "shocks".

Charge can build up on a person or an insulating object from a variety of sources. Friction, like walking across a carpet or sliding out of a car seat is a common source and a well known cause of charge build up.

After charging up an object, it is at a higher electrostatic potential than other surrounding objects. Think of this as the object having excess energy and the energy wants to escape.

When the charged up object comes in contact with another object or person at a lower electrostatic potential (less energy e.g. less charge), a transient shock will result. The shock is caused by the sudden transfer of the electrostatic charge called an electrostatic discharge (Figure 11.1).

- We are all familiar with a common form of electrostatic discharge (ESD):

Shaggy Carpet

- ESD is the sudden transfer of electrostatic charge between objects at different electrostatic potentials

Figure 11.1 Electrostatic Discharge (ESD)

It is interesting to note how much voltage potential can result from electrostatic charges. Generally the human body does not feel the electrostatic shock if the static voltage is less than 3 or 4 kV!!! It is only at static voltages above those limits one starts to feel the familiar shock/poke when touching the car door while getting out of the vehicle or when touching another object after walking across a carpeted floor. We will elaborate more on this subject later in this chapter, but as a quick introduction this brings up the first question as to how do ICs survive in the real world that are normally only protected against ESD up to 2 kV but yet many ESD events that we commonly experience are often greater than 3 or 4 kV???? The answer lies in the fact that there are different definitions (and reference points) when talking about electrostatic voltages and charge transfer. When discussing the amount of charge/voltage on an object we are referencing it to "mother earth" (i.e. 0 V potential). Thus the object has 4 kV of electrostatic potential higher than "earth potential". When discussing the robustness of ICs (strictly at the component level) to ESD energy, we are often referring to a static voltage between two different pins on the IC. Thus when you touch an IC between your two fingers, even though you may be charged to 4 kV above earth potential, the amount of electrostatic potential between your two fingers (one finger referenced to another) and thus the potential between the two IC pins, is much less than the 4 kV between your body and earth potential.

- Which levels can occur?
  - Below 3 - 4 kV you see, hear or feel nothing!
  - Just above 4 kV, air-gap-sparks can occur
  - 1 mm → 1 kV (5 mm spark means 5 kV)
- Why does a 2 kV protected device survive the real world?
  - You are charged relatively to earth, not to "pin7"
  - You do not have 4 kV between your thumb and your index finger

Figure 11.2 Where Does ESD Come From?

You may already be aware that these "electrostatic shocks" tend to occur more frequently in dry environments like in the winter time in colder climates. The reason for this is that higher humidity tends to form a slight film of moisture on the surface of objects which helps to distribute the charge evenly around the object, which in turn, reduces the static voltages. The static voltage increases as the charge is more concentrated in one area of the object. This is the main reason why we tend to get more "electrostatic shocks" in the wintertime and that is why electronic manufacturing plants have high quality temperature and humidity controls.

Figure 11.3 illustrates the effect of humidity and type of material on static potential.

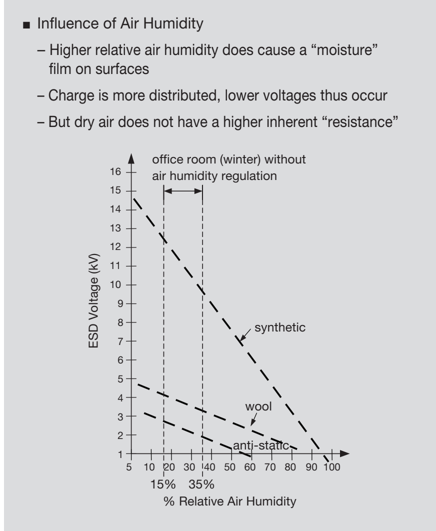


Figure 11.3 Where Does ESD Come From?

11.1.1 Susceptibility of Semiconductor Devices to ESD Stress

We now take a quick look at how electronic devices can be susceptible to damage from electrostatic charge/energy. Figure 11.4 is a side view of a Metal Oxide Semiconductor Field Effect Transistor (MOSFET) that is discussed in more details in the previous chapters.

The transistor can be thought of as an electronic switch. Current can flow between the drain and the source and the gate acts as the control between them. When the switch is closed, current will flow between the drain and source, and the transistor is "on". When the switch is open, current is prevented to flow between the drain and source and the transistor is "off".

When no bias (or zero volts) is applied at the gate, the transistor will be in the off state and to close the switch and allow current to flow between drain and source, a positive voltage (or bias) with respect to the source has to be applied to the gate.

A thin non-conductive layer of glass (silicon dioxide) electrically separates the gate from the body of the transistor.

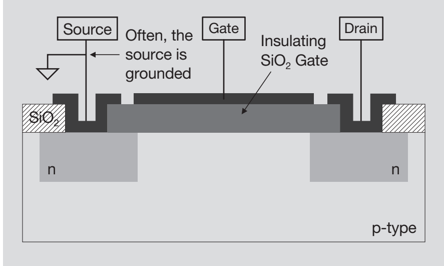


Figure 11.4 MOSFET Gate Susceptibility

Let us see what happens when a voltage is applied to the gate to turn the transistor on. Over time, as the voltage is applied to the gate, charge flows onto the gate.

However, because of the layer of glass (silicon dioxide) the gate is electrically insulated. The charge is stuck and has no where to go. (Figure 11.5)

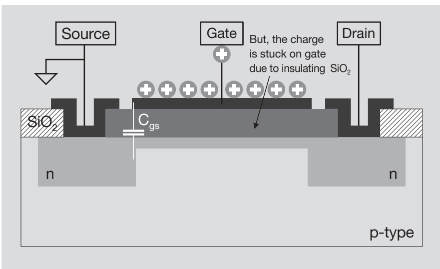


Figure 11.5 Charge is Applied to the MOSFET Gate

We will be using the terms voltage and capacitance a lot during this module, the reader is encouraged to review these concepts in Chapter 1.

The equations relating capacitance, voltage and charge are also applicable to a MOSFET gate-source capacitance. Letting the amount of charge on the gate be equal to Q and the physical dimensions of the gate be **d** (or thickness of SiO_2) and **A** (or area), the gate-to-source voltage (V) is defined as:

$$V = Q * d / (\epsilon_0 \epsilon_{ox} A)$$

Therefore, as more charge is placed on the gate of the transistor, the voltage on the gate increases.

$$V = \frac{Q\,d}{\epsilon_0 \epsilon_{ox} A}$$

- $\epsilon_0 \epsilon_{ox}$ is a constant for a given material (SiO_2)
- As the charge (Q) on the capacitor increases - the voltage across the capacitor increases…

Figure 11.6 Quick Review Summary

Let's see what can possibly happen when this process goes unchecked. If sufficient charge is placed on the gate of the transistor, the voltage can quickly climb to unsafe levels since voltage across a capacitor is given by V = Q / C (charge per capacitance). Sometimes, these voltages can reach 100 V, 1000 V, or even more. The electrostatic energy (and associated electric field) can be large enough that it damages the insulating glass, often creating an electrical short from the edge of the gate to the source where the electric field is strongest.

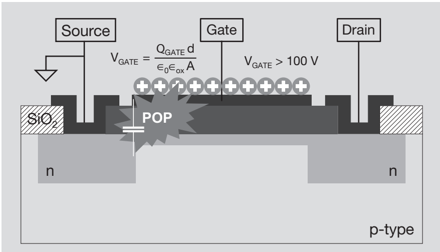


Figure 11.7 Gate Charge Induces a Gate Voltage

The sudden discharge damages the structure and creates a path for gate charge to flow through the silicon dioxide layer and to the source of the transistor which is commonly at ground potential (Figure 11.7).

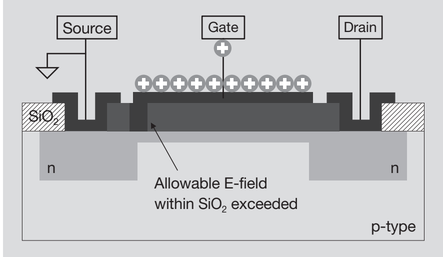


Figure 11.8 Induced Gate Voltage Creates a Hole in the SiO_2

Effectively, this means that the gate and source are shorted together and the gate will be held at or close to ground potential. Due to the short, the transistor cannot be turned on and for all practical purposes the transistor is destroyed (Figure 11.8 and Figure 11.9).

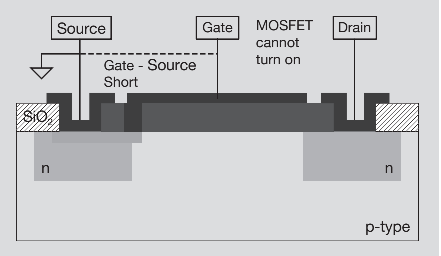


Figure 11.9 Induced Gate Voltage Creates a Hole in the SiO_2

So, let us summarize what we have learned so far by looking at our primary equation:

$$V = Q / C \sim Q * d / A \quad (C = \epsilon_0 \epsilon_{ox} A / d)$$

That is, the gate voltage is proportional to the gate charge times the insulating glass thickness, divided by the gate area (the term $\epsilon_0 \epsilon_{ox}$ is the dielectric constant of the gate capacitance, C).

With technology advances in semiconductor manufacturing transistors are made smaller and smaller. With smaller feature geometry the thickness of the insulating glass layer decreases, but the area (width times length) of the transistor gate decreases at an even faster rate. So, if the same amount of charge is placed on the gate of a large (older) transistor and a small (newer) transistor, the voltage on

the small (newer) transistor will be much higher primarily due to the smaller gate structure.

Therefore, newer, smaller transistors tend to be more susceptible to damage caused by electrostatic discharge if proper precautions are not taken both at the device and system levels.

$$V = \frac{Q\,d}{\epsilon_0 \epsilon_{ox} A}$$

- As the charge (Q) on the capacitor increases - the voltage across the capacitor increases…
- If the transistor decreases in size - the thickness of the SiO_2 gate (d) decreases - but, the area (A) of the gate decreases faster -
- For the same amount of charge, the voltage across the capacitor is higher for a smaller transistor
- More advanced technologies may require additional ESD precautions

Figure 11.10 Gate Susceptibility Summary

Let us look at a quick example (Figure 11.11). We will compare what happens to transistor in a "very" old 3 µm process and a "relatively" old 1.2 µm process during an electrostatic discharge event. In both cases, 1.16×10^{-11} coulomb charge is placed on the gate of a transistor ($\epsilon_{ox} = 3.9$).

- 3 µm Process (Minimum Size Transistor)
  - t_{ox} = 400 A = 4×10^{-8} m
  - L = 3 µm
  - W = 3 µm
  - Q = 1.16×10^{-11} C

$$V = \frac{(1.16 \times 10^{-11}\ C)(4.0 \times 10^{-8}\ m)}{(3.9)(8.85 \times 10^{-12}\ F/m)(3 \times 10^{-6}\ m)(3 \times 10^{-6}\ m)} = 1.5\ kV$$

- 1.2 µm Process (Minimum Size Transistor)
  - t_{ox} = 200 A = 2×10^{-8} m
  - L = 1.2 µm
  - W = 1.2 µm
  - Q = 1.16×10^{-11} C

$$V = \frac{(1.16 \times 10^{-11}\ C)(2 \times 10^{-8}\ m)}{(3.9)(8.85 \times 10^{-12}\ F/m)(1.2 \times 10^{-6}\ m)(1.2 \times 10^{-6}\ m)} = 4.7\ kV$$

Note: $\epsilon_0 \epsilon_{ox} = (3.9)(8.85 \times 10^{-12}\ F/m)$

Figure 11.11 Induced Voltage for 3 µm and 1.2 µm CMOS Processes

For the 3 µm transistor, the gate is 3 µm long and 3 µm wide. The insulating glass layer is 4×10^{-8} m thick. For the 1.2 µm transistor, the gate is 1.2 µm long and 1.2 µm wide. The insulating glass layer is 2×10^{-8} m thick. Using our equation:

$V = Q * t_{ox} / (\epsilon_0 \epsilon_{ox} A)$

the voltage on the gate of the 3 µm transistor is 1,500 V which is approximately the voltage built up on a person walking across carpet on a very humid day. The voltage on the gate of the 1.2 µm transistor, however, is 4,700 V!!! (in spite of the fact that the charge on the gates is exactly the same in both cases!).

Note that the symbol "A" in Figure 11.11 is for Angstrom or 10^{-8} m.

11.1.2 ESD Testing Standards

In order to characterize and quantify the ESD robustness of integrated circuits different ESD standards and tests have been developed over the years. These tests are for trying to "simulate" real world events and applications but unfortunately there is no one, single ESD test that can cover all the different mounting and handling conditions. Different tests have been developed to simulate ESD events resulting from human bodies, from machines (manufacturing equipment), and also from the devices themselves acquiring charge and eventually discharging to ground potential. In addition to all of these different tests, a complete set of different test standards have been developed regarding if the device under test is a component (single integrated circuit) or a complete electronic module (printed circuit board with many ICs). Thus when talking about ESD levels and robustness it is extremely important to understand and know what test standards are being applied.

These test standards (Figure 11.12) all consist of a model, parametric values for the elements used in the model and a test procedure which tells how to apply the standard (amplitude of pulses, how many pulses to apply, use both a positive and negative pulse and what pins to apply the pulses to).

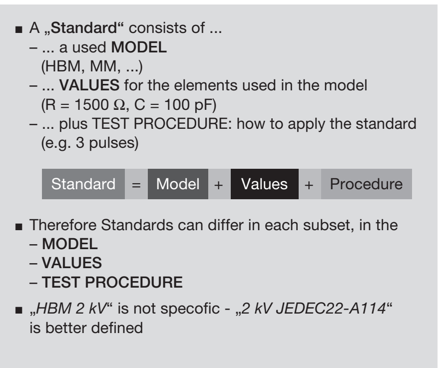


Figure 11.12 ESD Standards & Tests: Overview

One of the most common standards is the so-called “Human Body Model” (HBM) which consists of a capacitor and a series resistor. This standard is meant to replicate if a charged human being comes in contact with an IC. It is important to note again here there is not one HBM standard but several standards which can be seen in Figure 11.13 using different values for the capacitor and resistor.

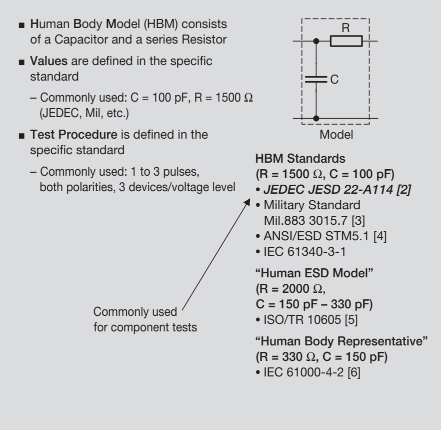


Figure 11.13 ESD Models: Human Body Model

The most common HBM (Figure 11.13) used for component level testing is in the JEDEC standard but it will also be shown later in this chapter that a different standard is used for HBM testing of complete modules.

When testing for ESD robustness the capacitor is charged by an external source to the specified voltage (1 kV, 2 kV, etc.) and then the capacitor is discharged into the device under test through the series resistor.

Typical HBM test current waveforms can be seen in Figure 11.14; note, that the transients are quite fast (nanosecond type rise times) and the current amplitudes are in the range of a couple of amperes.

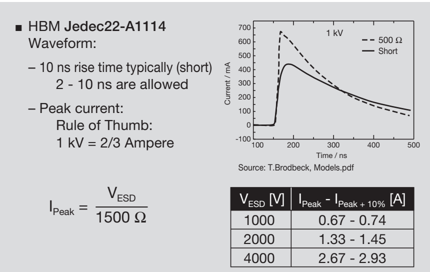


Figure 11.14 ESD Models: Human Body Model -> Waveform

A second common ESD test standard is the machine model (MM, see Figure 11.15) which is aimed at simulating when a charged part of the manufacturing equipment comes into contact with the IC during printed circuit board assembly. This model only uses a capacitor and there is no series resistance. Most standards use 200 pF capacitance value.

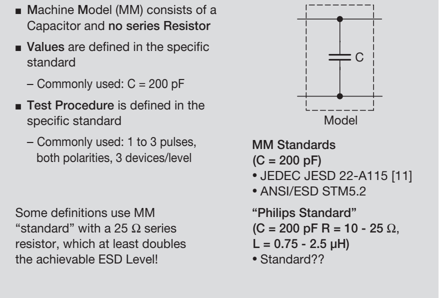
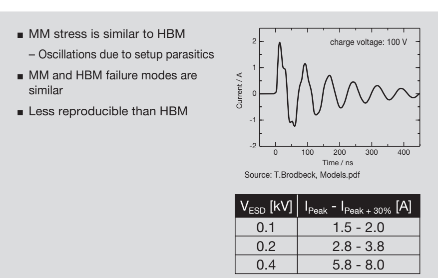


Figure 11.15 ESD Models: Machine Model

The current waveform of the machine model test tends to be oscillatory due to the fact that parasitics in the test set up (stray inductance, etc.) can influence the waveform. Due to the

oscillatory response the machine model test tends to be less reproducible than the human body model test.



V_{ESD} [kV]	I_{Peak} - $I_{Peak + 30\%}$ [A]
0.1	1.5 - 2.0
0.2	2.8 - 3.8
0.4	5.8 - 8.0

Figure 11.16 ESD Models: Machine Model

For these reasons the machine model test seems to be less frequently used and another test called the “Charged Device Model” (CDM) is becoming more popular. In course of both the human body model test and machine model test an external object (the capacitor) is discharged into the device under test. The charged device model test is exactly the opposite in that the test simulates a charged device under test discharging to ground. This is an approximation of the effects when a charged component is touched or handled by a grounded object.

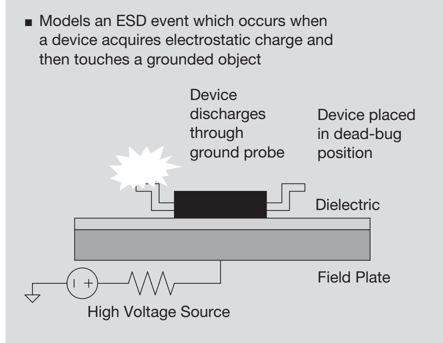


Figure 11.17 Charged Device Model Test

The charged device model current waveform is highly dependent on the package capacitance of the device under test and **the rise time is even faster** (in the range of hundreds of picoseconds).

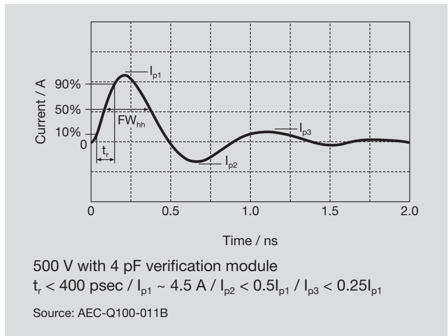


Figure 11.18 CDM Waveform: Highly dependent on die size and package capacitance

Most integrated circuits that are used in automotive applications must be qualified according to the AEC-Q100 qualification test standard originally developed by Ford, GM, and Chrysler. The AEC-Q100 is not itself a test standard but a collection of qualification requirements which references other existing standards. In regards to ESD robustness the IC must pass either 2 KV for the HBM test (JEDEC standard) OR 200 V for the machine model (MM) test (again JEDEC standard). In addition the device must also pass 500 V (750 V for corner pins) using the charged device model (CDM) test according to a JEDEC standard.

11.1.3 Component vs. Module Level Tests

As mentioned previously there is a difference in ESD testing (and test standards) between testing an individual IC (component) and testing an electronic module (system).

The main goal for device level testing is that the **component remains intact and fulfills all specification** limits following the mentioned ESD tests, thus ensuring a relative measure of robustness during handling and manufacturing.

The main objective for system level testing is to ascertain that the **system remains functional** during and after the ESD stress is applied which gives a relative measure of robustness when used by the end customer(s).

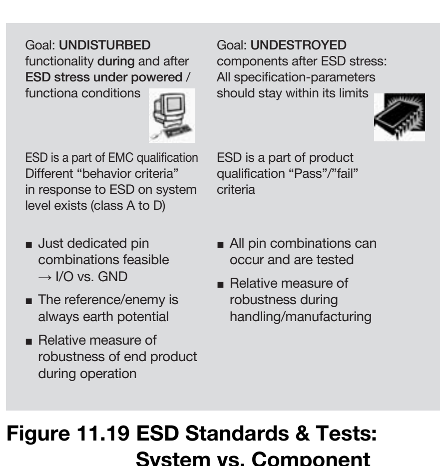


Figure 11.19 ESD Standards & Tests: System vs. Component

As an example we can look at the differences in the human body model tests normally used for component and system level testing. For component level HBM testing one of the most widely used test method is formulated by the JEDEC standard which uses a 100 pF capacitor and 1500 Ω resistor. During this test all possible pin combinations are tested.

For system level HBM testing the so called "Gun Test" is widely applied using a 150 pF capacitor and 330 Ω resistor. Already it can be seen the peak current will be higher for the Gun Test compared to the JEDEC component test (more about that shortly) at comparable voltage levels. Also the system pins or interface are tested with reference to earth ground/potential (Figure 11.20).

Module/System Level	Component Level
Human Body Model (HBM) 150 pF / 330 Ω EN 61000-4-2 (so called "GUN Test")	Human Body Model (HBM) 100 pF / 1500 Ω JEDEC-Norm JESD22-A114-B (MIL-STD883D, method 3015)
Human Body Model (HBM) 150 pF / 2000 Ω ISO 10605	Machine Model (MM) 200 pF / 0 Ω JEDEC-Norm JESD22-A115-A (correlates to HBM)
Human Body Model (HBM) 330 pF / 2000 Ω ISO 10605	Charged Device Model (CDM) Package pF / 0 Ω JEDEC-Norm JESD22-C101-A

Figure 11.20 ESD Test Methods (Models) System vs. Component

As an example we examine the details of the implementation of the JEDEC human body model test for components. This test consists of a high voltage source charging the capacitor and then the capacitor is discharged through a resistor into a number of pins of the integrated circuit selected in a certain combination and shorted together. Some pins are individually tested as described in Figure 11.21.

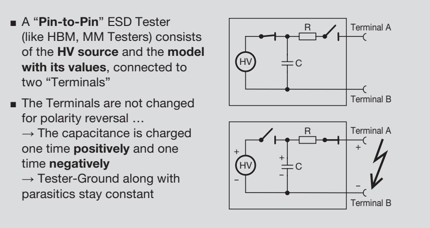


Figure 11.21 ESD Models: Human Body Model -> Component Test

In regards to the pin combinations tested, the pins are normally classified and grouped as supply pins (all supply pins and ground pins) and non-supply pins.

For the supply pin test each individual non-supply pin is tested one at a time (at terminal A for instance) in reference to the first supply pin (at terminal B for instance). The test is then repeated for each supply pin.

For the non-supply pin test each individual non-supply pin is tested one at a time (at terminal A for instance) in reference to all other non-supply pins grouped (connected) together in parallel (at terminal B for instance). This test is repeated for each non-supply pin.

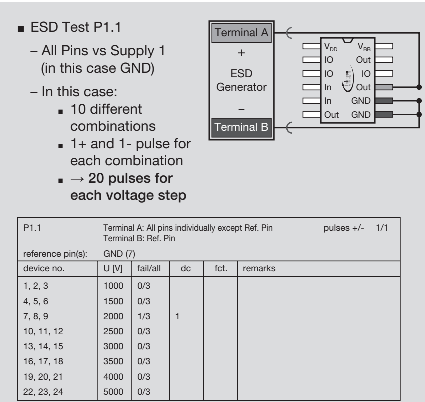
An ESD stress pulse prescribed for the test is applied with one positive pulse and negative pulse for each pin combination. Then the devices are tested with the normal test program to make sure all parameters are, and the functionality is within specification. If all devices are within specification HMB test is then repeated at the next higher pulse level.

The following figures illustrate the different HBM ESD tests: Figure 11.21, Figure 11.22, Figure 11.23 and Figure 11.24.

- 2 different Pin - Combination - Types are tested
  - **Supply - Pin - Tests**
    - All Pins (individually one at a time) at Terminal A vs Supply-X at Terminal B
    - Repeat for Supply-Y, Supply-Z, etc. at Terminal B
  - Non-**Supply - Pin - Tests**
    - All Non-Supply pins (individually one at a time) at Terminal A vs all other non-Suply pins (together) at Terminal B
    - Repeat tests with reversed roles (individual pin to Terminal B grouped pins to Terminal A)
- 1 positive and 1 negative pulse for each pin - combination
- Step - Stress, 500 V, 1 kV, 2 kV and 4 kV are used; different levels and steps can be defined
  - A new set of 3 devices per level is used

ESD Product Qualification Test @ IFX according to
JEDEC EIA / JESD22 - A114 - B[2]
described in IFX Procedure [1]

Figure 11.22 ESD Models: Human Body Model -> Component Test



P1.1	Terminal A: All pins individually except Ref. Pin Terminal B: Ref. Pin				pulses +/- 1/1
reference pin(s):	GND (7)				
device no.	U [V]	fail/all	dc	fct.	remarks
1, 2, 3	1000	0/3			
4, 5, 6	1500	0/3			
7, 8, 9	2000	1/3	1		
10, 11, 12	2500	0/3			
13, 14, 15	3000	0/3			
16, 17, 18	3500	0/3			
19, 20, 21	4000	0/3			
22, 23, 24	5000	0/3			

Figure 11.23 ESD Supply-Pin Test: HBM ESD Each Pin vs. Supply-1 (GND)

- ESD Test P2
  - Each non-supply vs. All other non-supply
  - One non-supply at a time on Terminal A
  - All other non-supplies at Terminal B
  - In this case:
    - 8 different combinations
    - 1+ and 1- pulse for each combination
    - **→ 16 pulses for each voltage step**

Terminal A
+
ESD Generator
–
Terminal B
VDD VBB
IO Out
IO IO
In Out
In GND
Out GND

P2	Terminal A: All non-supply pins individually Terminal B: All other non-supply pins				pulses +/- 1/1
device no.	U [V]	fail/all	dc	fct.	remarks
121, 122, 123	1000	0/3			
124, 125, 126	1500	0/3			
127, 128, 129	2000	0/3			
130, 131, 132	2500	0/3			
133, 134, 135	3000	0/3			
136, 137, 138	3500	0/3			

Figure 11.24 ESD Non-Supply-Pin Test: HBM ESD Each Non-Supply vs. All Other Non-Supply

It can now be seen that this series of tests is quite a lengthy process to fully characterize the component for human body model ESD robustness.

In comparison the system level test will always use earth potential as the "reference pin/point" and applies the test pulses to "normal points of accessibility" during end customer use of the system (Figure 11.25).

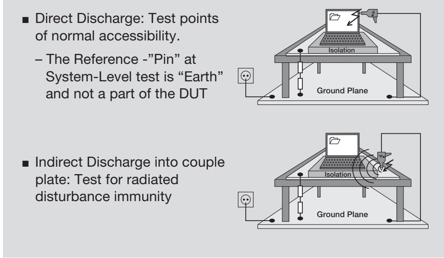


Figure 11.25 ESD Standards & Test: System-Level Test

Occasionally it will be asked if a system level test can be applied to a component in order to try to correlate test results between component and system level test results. This is possible but there are a few differences which cannot be correlated such as the component cannot be powered and the reference pin will be the ground pin and not

earth ground. As can be seen in Figure 11.26 the gun test peak current is approximately 5-times higher than the component level peak current due to the differences in the model capacitor and resistor mentioned previously.

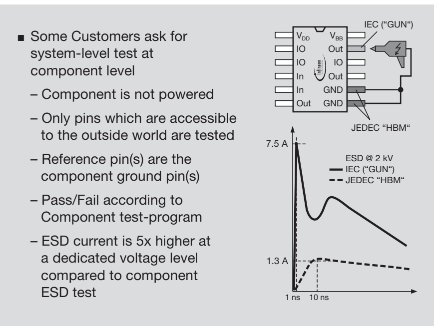


Figure 11.26 HBM: System Level Tests applied to components?

11.1.4 ESD Damage

In general the energy content of the ESD pulses is rather small because of the low current amplitudes and fast pulse times (nsec or psec) even though the voltage amplitudes are in the kV's range. With such low energy level the damage due to ESD pulses will normally show very small physical failure signatures that, in many cases, cannot even be seen with the "naked eye".

As mentioned earlier in the chapter one susceptible area to ESD damage is the MOSFET gate structure. Figure 11.27 and Figure 11.28 are two pictures from a MOSFET damaged in the gate structure very close to the source contact.

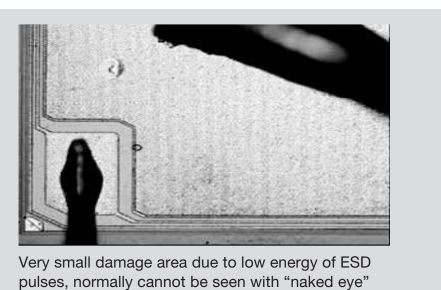


Figure 11.27 HBM ESD Gate Shorted to Source

Without any magnification the damaged area is not visible to the naked eye (marked by the small circle). Upon further magnification the damage is visible: the source contact metal has a slight burn mark right next to the gate structure. This device has the gate shorted to the source (marked with the four arrows) as previously shown earlier in this chapter.

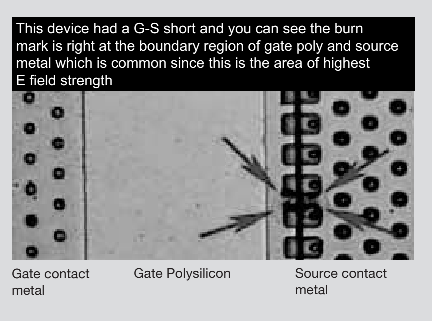


Figure 11.28 HBM ESD Gate Shorted to Source

11.1.5 Pin Protection for ESD Stress

The obvious question is: can ESD sensitive devices be protected in automotive applications? One very common method of protecting sensitive circuits is to place a capacitor between the vulnerable pin and ground. The external protection capacitor can divert the ESD current to ground and protect the sensitive device.

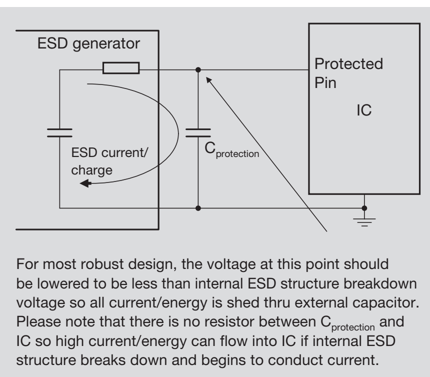


Figure 11.29 Decrease ESD Sensitivity with a Predictable Charge Well Topology

To ensure robustness the voltage level at the sensitive IC pin(s) must be lowered below the breakdown level of the internal ESD structure contained within the IC such that all current will flow through the external capacitor and completely protect the IC. A simple equation can be used to determine the value for the external protection capacitor. It can be seen in Figure 11.29 that a capacitive voltage divider is formed between the ESD generator and the external capacitor. Thus by knowing that charge is equal to capacitance times voltage (Q = C V) and the total charge is transferred from the ESD generator to the external protection capacitor it follows that $C_{ESD} * V_{ESD} = C_{protection} * V_{protection}$. Knowing the ESD capacitance and voltage of the generator as well as the maximum voltage that can be allowed at the IC, the value of the protection capacitor can be calculated from the equation. Figure 11.30 is an example where a gun test at 15 kV is reduced down to 45 V by using an external capacitor of 110 nF.

- System level/gun tests ESD voltages may need to be 15000 V (direct contact)
  - Gun tests uses 330 pF for source capacitor
- For automotive technologies having ESD structures with 40 - 45 V breakdown is common

$C_{protection} = (C_{gun} / V_{br_ESD}) * V_{gun}$
$= (330 \text{ pF} / 45 \text{ V}) * 15 \text{ kV}$
$= 110 \text{ nF}$

Figure 11.30 Decrease ESD Sensitivity with a Predictable Charge Well Topology

It is very common to find integrated protection circuits within the integrated circuit. These circuits can be referenced to, and configured to shed the ESD energy either to ground or to the supply. For instance, in Figure 11.31 a simple zener diode circuit referenced to ground is shown.

When a positive pulse applied to the external pin the zener diode will start to conduct at a certain breakdown voltage and provide a low impedance path to ground. For a negative pulse the diode will conduct in the forward direction and allow the current to flow from ground through the external pin. Also shown is a simple n-channel FET circuit connected between the supply (drain) and the protected pin (source). Notice the diode-resistor circuit between the gate and ground. For a positive pulse at the external pin the ESD current pulse is channeled to the supply via the body diode of the FET. For a negative pulse, since the FET gate is tied to ground, at a certain negative voltage at the source (external pin) the FET will turn on and the pulse current is conducted from supply to the external pin.

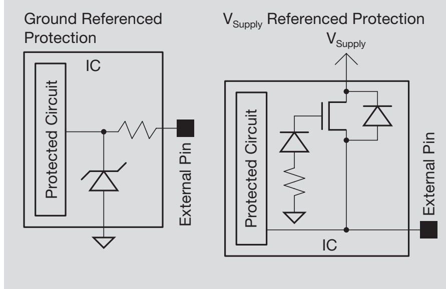


Figure 11.31 Typical Internal IC ESD Protection Circuits

11.1.6 ESD Summary

The summary of the discussions regarding ESD is in Figure 11.32:

- Electrostatic discharge occurs when excessive static charge on an object builds up to a very high voltage (thousands of volts) and causes device damage during contact and subsequent discharge (current flow) with another object
- MOS devices with insulating SiO_2 gates are especially susceptible to ESD damage
- Different test standards have evolved for component level and system level tests and confusion can result if these standards are not understood and clarified in reports and communication
- The very fast (HBM = nsecs / CDM = psecs) ESD pulses have low energy and result in ***VERY*** small physical damage signatures

Figure 11.32 ESD Summary

11.2 Electrical Over Stress (EOS)

Next, we will investigate the causes and results of Electrical Over Stress (EOS). It will be shown how ESD and EOS have different causes and will damage integrated circuits differently. It may be obvious, but electrical over stress is exactly what it says ---- the semiconductor device is subjected to an electrical stress exceeding its specified limits of operation.

- Electrical Over Stress is exactly what it says…

  A device is electrically stressed over it's specified limits in terms of voltage, current, and/or power/energy

- Unlike ESD events, EOS is the result of „long" duration stress events (millisecond duration or longer)
  - Excessive energy from turning off inductive loads
  - Load Dump
  - Extended operation at junction temperatures > 150 °C
  - Repetitive excessive thermal cycling
  - Excessive/extended EMC exposure, etc.
- EOS often results in large scorch marks, discoloration of metal, melted metallization and/or bond wires, and massive destruction of the semiconductor component

Figure 11.33 What is Electrical Over Stress (EOS)?

11.2.1 Description of EOS Damage Signatures

The failure signature from EOS damage can be quite severe resulting in large scorch marks and melted metallization which is quite different from the very minute (size-wise) damage caused by ESD. This is due to the fact that EOS is a much larger energy event typically than the ESD.

The failures from EOS can result in an immediate destruction of the device (hard failure) or the failure can occur later in time after further degradation occurs (latent failure) as a result of smaller and repeated non-destructive but damaging EOS events.

- Failures from EOS can result in the following:
  - Hard failure: failure is immediate and results in a complete non-operational device
  - Soft failure: EOS results in a marginal failure or a shift in parametric performance of the device
  - Latent failure: At first the EOS results in a non-catastrophic damage but after a period of time further degradation occurs resulting in a hard or soft failure

Figure 11.34 What is Electrical Over Stress (EOS)?

11.2.2 Consequences of EOS

The amount of damage the device experiences is related to the duration of, and the peak temperature the silicon reaches in the EOS event. Figure 11.35 is a graph relating lifetime vs. device temperature (time scale is logarithmic). For most automotive components the full lifetime and specification are valid up to 150 °C. For temperatures up to 200 °C the specification and functionality is restricted and the lifetime is reduced *by and order of magnitude*. For temperatures up to approximately 270 °C no functionality or specification is guaranteed and the lifetime is highly reduced although no immediate permanent damage should result. At temperatures above 350 °C the silicon will start to become self-conductive and irreversible damage can occur.

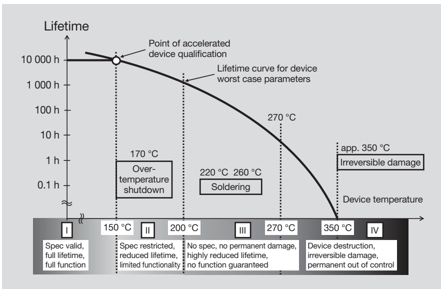


Figure 11.35 EOS: Thermal Lifetime Curve

Let's now take a look at what does the EOS look like? Remember from earlier descriptions that damage from ESD was barely noticeable and sometimes could not even be viewed by the naked eye. Well, damage from EOS is very

different and is usually quite evident even without magnification. Many times this damage consists of either degraded metallization or melted metallization or even melted silicon as illustrated in Figure 11.36.

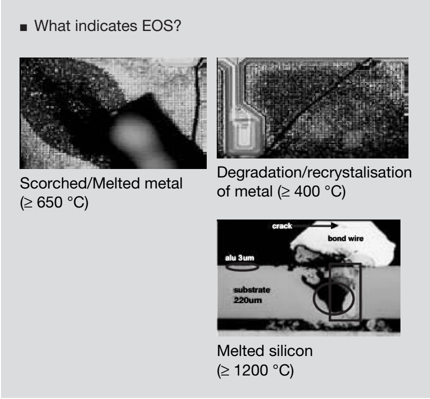


Figure 11.36 What is Electrical Over Stress (EOS)?

Let's look at, in greater details, two different types of EOS failures, one resulting from a one-time stress and another caused by continuous or repetitive stress events.

Figure 11.37 shows the signature of a device failure caused by an excessive inductive turn off energy hit which was discussed in earlier chapters. If the energy stored in an inductive load is too large, the high side or low side driver may be damaged when the inductive load is turned off. Such an event is a singular or one time electrical over stress since the damage and failure occurred immediately just after one pulse was applied. A hot spot or so-called "scorch" can be seen within the FET power stage but also notice the metallization is not degraded (or re-crystallized) near the bond wire and other than the scorch mark the device looks quite normal.

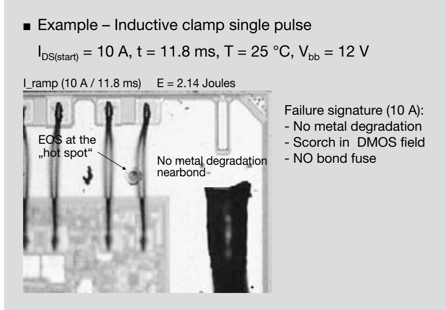


Figure 11.37 EOS: Failure Signature from Excessive Inductive Turn-Off Energy

The signature of an EOS damage caused by many high current pulses applied in a repetitive or continuous manner is illustrated in Figure 11.38. The degradation (seen as discoloring) of the metal around and near the bond wire is very noticeable. This is due to the fact that the highest current density (and thus highest power density and temperature rise) is around the bond wire. Scorch marks are typically near or even under the bond wire.

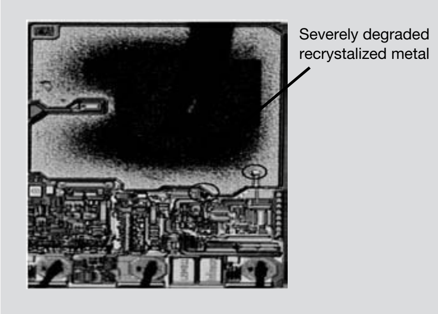


Figure 11.38 EOS: Failure Signature from Repetitive Thermal Cycling Combined with High Current

This brings up the point that EOS can be caused by a single event or limited number of events (10's to < 100) where the single pulse has enough energy to start the scorching or melting process of the metallization (> 600 °C). However, EOS can also result from the continuous and repetitive application of stresses which have lower energy but are applied many times (1000's or more). For this

type of repetitive stress the energy of a single pulse may only increase the die/metal temperature by a 100 °C or so but the continuous heating and cooling of the die causes thermal mechanical stress in the die/package system and eventually a failure will occur mainly due to degraded and cracked metallization. As shown in Figure 11.37 and Figure 11.38 the failure signatures from these two different types of EOS can look significantly different (Figure 11.38).

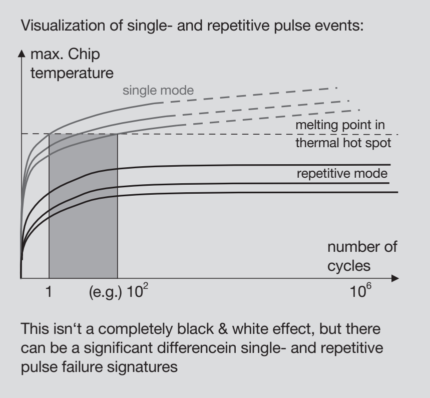


Figure 11.39 EOS Failure Signatures---Generalities

Melted or fused bond wires can also occur from EOS and is usually indicative of an excessive current flow through the bond wires. To give a general feeling of some of the common bond wires used for automotive power semiconductors and their related current capability please refer to Figure 11.40. As can be seen the current ratings of bond wires with different the cross-sectional area, made of different materials can vary considerably.

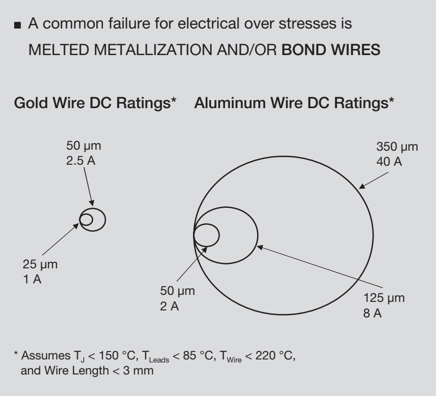


Figure 11.40 Electrical Over Stress (EOS)

11.2.3 Common Causes of EOS

Finally we can ask: how can components be electrically overstressed? Obviously there can be a multitude of possibilities and some of the more common causes are in Figure 11.41.

- Put components in/out of sockets while power already applied (hot plug)
- Applying electrical signals which exceed the ratings of a component
- Applying an input signal to a device before supply voltage and/or ground is connected
- Apply an input signal to a device output
- Use an inexpensive power supply with poor protection features (supply overshoot)
- Failing to provide sufficient noise filtering at the input lines of the board
- Using poor ground connections (high resistance and inductance)

Figure 11.41 How to Electrically Over Stress Components

11.3 Safe Operating Area

Finally we introduce the concept of Safe Operating Area (SOA) and how it is used to avoid electrical over stress.

The safe operating area is specified by a range of various conditions such as operating voltages and currents, junction or ambient temperature, and heatsinking conditions. If a device is used within the SOA, the semiconductor supplier ensures the device will operate as specified with certain reliability and without the threat of electrical over stress.

Because of the wide variety of semiconductor devices and corresponding functions not all safe operating areas can be neatly or easily categorized in tabular form. Rather, the safe operating areas are specified in a variety of different ways with common device/functionality types such as high side drivers, voltage regulators, etc. being defined by a similar set of parameters for safe operating area.

One of the most common ways to specify a safe operating area is in graphical form. Let's look at a couple of examples.

Figure 11.42 is the (SOA) graph for a discrete MOSFET transistor. The test conditions are for single pulse stresses of different power levels and durations when the device case held at 25 °C, soldered to a copper PCB heatsink of 6 cm^2 area and the resulting junction temperature is to stay less than 125 °C.

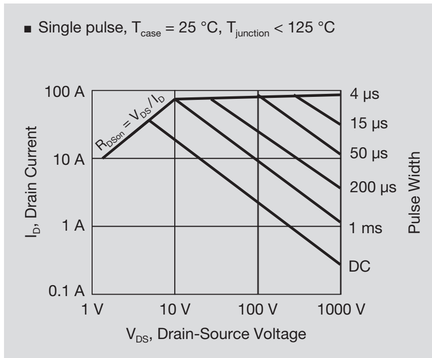


Figure 11.42 SOA Graph for MOS Transistor

The electrical variables used to specify this safe operating area are the drain to source voltage on the horizontal axis and drain current on the vertical axis (notice, that the scale on both axes is logarithmic). The pulse width is the parameter characterizing the different lines on the graph. As an example for interpreting this graph: with a drain to source voltage of 100 V the device can handle continuously approximately 3 A drain current and for a pulse of 1ms in duration the drain current is 10 A under the test conditions specified.

Figure 11.43 shows the SOA graph for a linear voltage regulator. The test conditions are: with a 5 V output voltage and with a 3 cm^2 copper PCB heatsink area the desired junction temperature must be less than 150 °C. The electrical variables are the input voltage to, and output current out of the voltage regulator and the parameters are the different ambient temperature values. Both axes have linear scale.

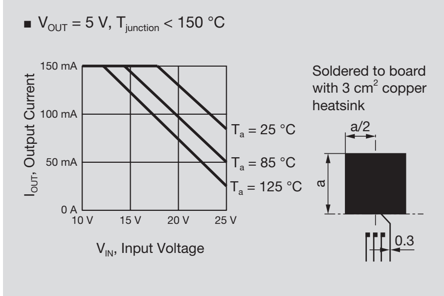


Figure 11.43 SOA Graph for Linear Voltage Regulator

Thus given 15 V input voltage the device can handle approximately 120 mA at 125 °C ambient temperature and about 145 mA at 85 °C ambient. One additional point to notice is that the graph is "cut-off" at 150 mA because this particular device has an internal over-current protection circuit limiting the output current at 150 mA.

12. Semiconductor Manufacturing

In this chapter, the basic steps required to manufacture a semiconductor component are presented, which are fabrication, packaging, and testing.

In course of the presentation the various fabrication processes and the differences between them will be described.

Next the different packaging options available to all semiconductor components are examined.

Finally, we introduce some of the basic requirements for testing semiconductor components.

12.1 Semiconductor Fabrication

Semiconductor fabrication refers to the process where raw, pure semiconductor material (usually silicon) is processed into an integrated circuit or discrete component. It can be thought of as the recipe to make a semiconductor die.

Different types of semiconductor components use different types of fabrication processes. Different processes are used and optimized for: microprocessors, logic components, memory IC-s like DRAM, inear (also called analog) devices and power components. All of these different semiconductor fabrication processes use similar steps and “ingredients”. The number of steps and the quantity of each ingredient, however, is optimized for different types of semiconductor components.

We will be examining in some detail three power semiconductor fabrication processes. Each has its own characteristics which lend itself to specific applications.

- Semiconductor components are fabricated typically from one of several different processes
- While suppliers might use various names, the processes themselves vary little between companies

- Power MOSFET processes
- Bipolar processes
- Combination processes using two or more of the above

Figure 12.1 Semiconductor Fabrication

12.1.1 Power MOSFET Fabrication

First, we examine the Power MOSFET process. It is used to fabricate silicon dice often implementing only one large power transistor. Such silicon devices are often termed “discrete” components. Some semiconductor manufacturers have variations of this process which allow the implementation of multiple large power transistors on a single die with a small amount of control functions.

Such devices are intended for the harshest applications - the highest levels of voltage and current. Because of their robustness, they can be used in applications which would damage more complex and sensitive semiconductor components.

The power MOSFET process is typically a vertical process - the current flows through the transistor in the vertical direction. A planar n-channel power MOSFET is shown in Figure 12.2.

- Power MOSFET processes are for discrete power transistors and integrated logic/power devices
- Power MOSFET transistors are intended for high current and/or high voltage operation. They have a unique construction which allows them to operate under „extreme“ conditions.

Figure 12.2 Power MOSFET Process

When the gate is grounded, no current can flow from the back side of the silicon die (drain) to the top of the die (source). When the gate voltage is raised, the transistor turns on and current flows from the drain to the source.

12.1.2 Bipolar Process

Next, we present a typical bipolar process. Bipolar processes can be used to fabricate silicon die which implement a single large bipolar power transistor or a more complex integrated circuit.

Bipolar deices have several advantages over their MOSFET counterparts. First, they can be used to realize very precise analog functions. In addition, they can operate at a very high frequency. Unfortunately, their high performance also results in significantly higher power consumption than many MOSFET circuits.

Figure 12.3 is an example of a bipolar process which is used to implement a bipolar integrated circuit. The transistors themselves are fabricated in the top of the silicon die (the epitaxial layer). The base, collector, and emitter of the NPN bipolar transistor are

labeled in the figure. Two layers of metal and one pseudo-metal (polysilicon) layer are used to interconnect the individual bipolar transistors together. In addition, the top metal layer is used to form bonding pads (top right corner of the process cross section). These bonding pads will be used to provide the electrical connection between the semiconductor die and the package pins.

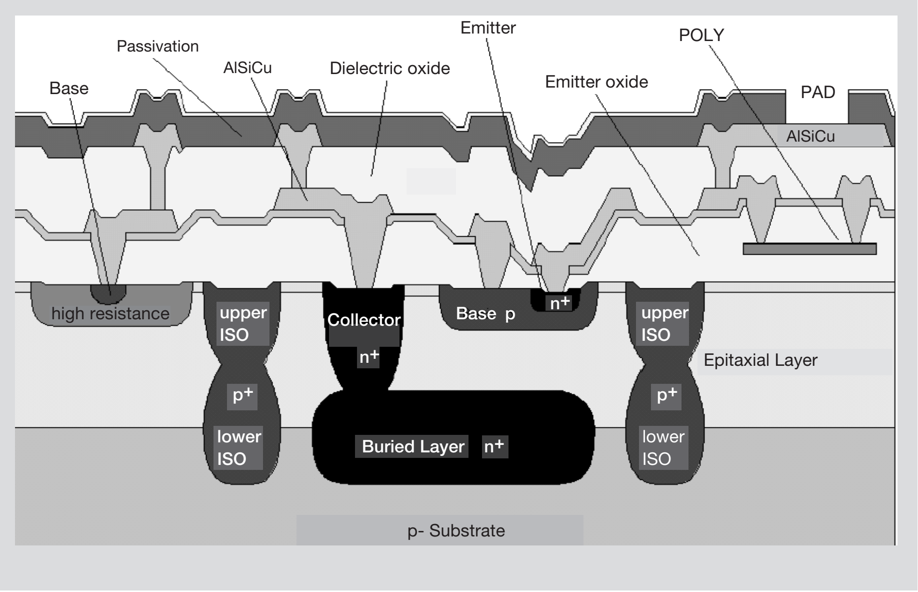


Figure 12.3 Bipolar Process

Finally, processes can be combined to form “super recipes” to make extremely versatile semiconductor components. While these combined processes provide the greatest freedom to integrated circuit designers, they are inherently more complex to manufacture (and therefore, more expensive) than the individual MOSFET and bipolar processes.

12.1.3 BiCMOS and DMOS Processes

BiCMOS (Bipolar + CMOS) processes are used for making precision analog components that do not handle high power loads while require a high level of control.

Smart power or Bipolar-CMOS-PowerMOS processes provide the advantages of the constituent processes, with the added benefit of high power load drivers implemented in a special power MOSFET transistor called “DMOS”. A die cross section in Figure 12.4 and Figure 12.5 show the different silicon structures of a DMOS die.

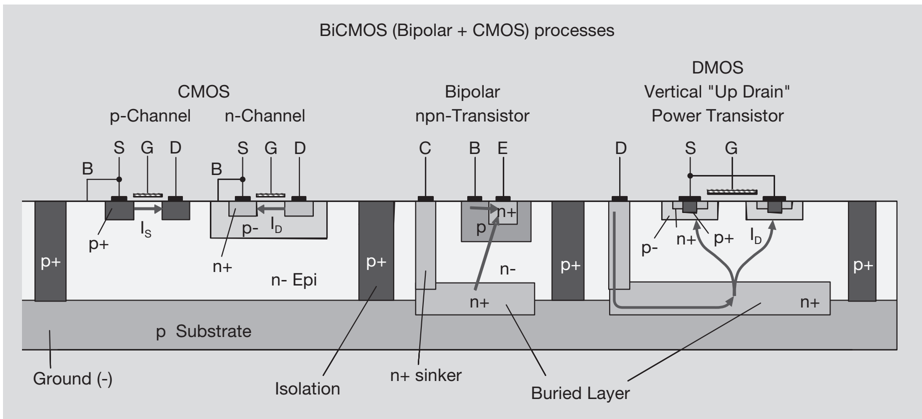


Figure 12.4 BiCMOS-DMOS Combination Process

Another type of combination process combines the ruggedness of the vertical power MOSFET transistor with the control of high density CMOS circuits. While they may not be able to implement the precise analog functionality of bipolar circuits, the reduced complexity of the CMOS-DMOS combination process (lower cost) often makes it an attractive option to integrated circuit designers and users (Figure 12.5.

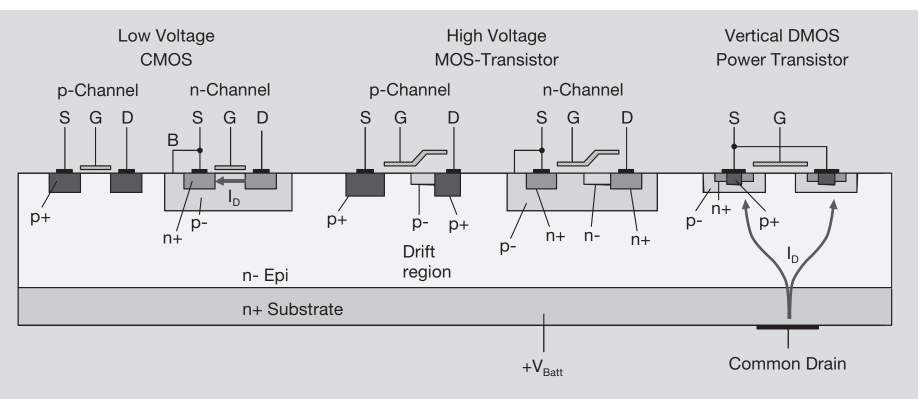


Figure 12.5 CMOS-DMOS Combination Process

12.2 Fabrication Recipes

The method of physical implementation of these silicon die fabrication recipes is presented next. Let's see how a simple CMOS inverter is implemented.

Beginning with a silicon die an additional silicon layer (the epitaxial layer, or epi-layer) is added to the top surface and it is within this epitaxial layer that the transistors will be formed (Figure 12.6).

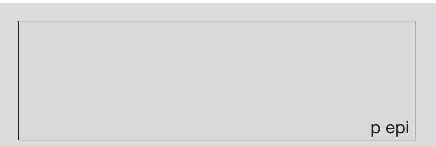


Figure 12.6 Semiconductor Fabrication

Next, the silicon die is literally "baked" in a very hot oven. The high temperature allows the top layer of the silicon die to chemically react with the oxygen in the oven (Figure 12.7).

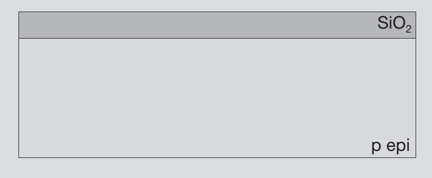


Figure 12.7 Semiconductor Fabrication

The result is silicon dioxide, or glass. Small holes are then etched in the glass with acid. Atoms are injected into the epitaxial layer through the holes in the glass (this step is called doping). The injected atoms change the crystal and electrical properties of the semi-conductor. They form the source and drain of the n-channel MOSFET (Figure 12.8) and create a well in which the p-channel MOSFET is fabricated.

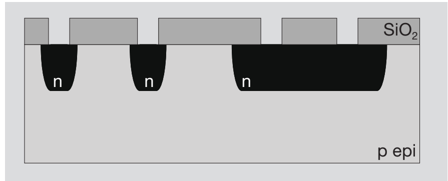


Figure 12.8 Semiconductor Fabrication

Next, the die is returned to the oven and the glass layer is regrown. Small holes are etched above the large n-well and atoms are again injected. This time, the injected atoms form the source and drain of the p-channel MOSFET in the well (Figure 12.9).

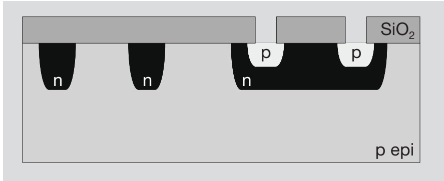


Figure 12.9 Semiconductor Fabrication

The die is again returned to the oven, and the glass layer is regrown. This time, a thin glass layer is formed above the transistor channels. Poly-silicon, also known as poly-crystalline silicon, has nearly metallic conducting properties. It is deposited on the thin glass layer above the channels to form the transistor gates (Figure 12.10).

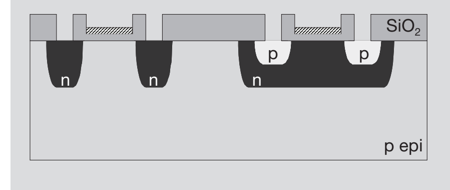


Figure 12.10 Semiconductor Fabrication

The die is returned to the oven again and the glass layer is regrown. Small holes are again etched at the source and drain of each transistor. Metal is then deposited on the die to form the transistor contacts and the interconnect between the transistors.

Then end result is a simple inverter fabricated in a “n-tub” CMOS process (Figure 12.11).

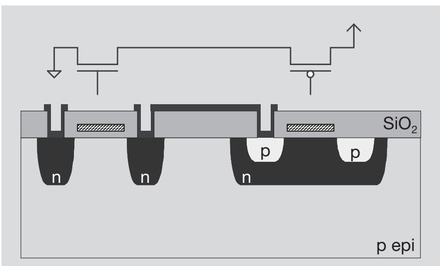


Figure 12.11 Semiconductor Fabrication

12.3 Semiconductor Packaging

The backbone of most power semiconductor packages is the leadframe. It is a piece of metal (usually a copper alloy) supporting the silicon die.

Figure 12.12 shows a leadframe for a 16 pin dual in-line package. There are two small protrusions on each side of the leadframe that are also pins of the package. These four pins will be electrically connected to the back side of the semiconductor die.

The additional pins are formed by additional metal pieces which are not directly connected to the leadframe Figure 12.12).

The sequence of package assembly is described next.

First, a die attach material is placed upon the leadframe. The die attach material is an electrically conductive glue or solder in some cases. It will hold the semiconductor die in place on the leadframe while electrically connecting the back of the semiconductor die to the leadframe.

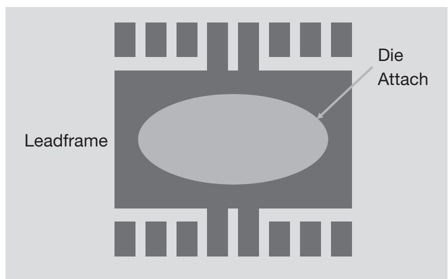


Figure 12.12 Basic Semiconductor Package

The semiconductor die is then secured on the leadframe with the die attach.

As we mentioned before, the semi-conductor die has several bonding pads formed by the top metal layer. These bonding pads will enable the electrical connection between the silicon die and those package pins that are not directly connected to the leadframe. Figure 12.13 illustrates the attached die and bond pads.

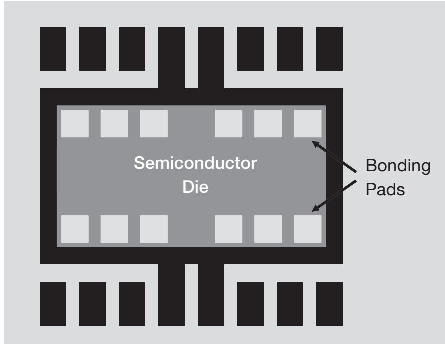


Figure 12.13 Basic Semiconductor Package

The electrical connections between the semiconductor die and the package pins are by means of bond wires (gold or aluminum). These wires are bonded to the semiconductor die and those package pins that are not directly connected to the leadframe (Figure 12.14).

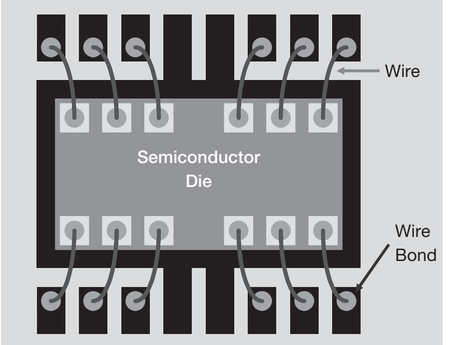


Figure 12.14 Basic Semiconductor Package

Finally, the die, bonding wires, and leadframe are encapsulated in an epoxy compound that is the familiar black plastic package of semiconductor components (Figure 12.15). Notice that all that is left exposed are the pins.

In some power packages, the back of the leadframe is not encapsulated by the epoxy. The exposed leadframe helps to form a better thermal connection with the die improving the cooling of the package.

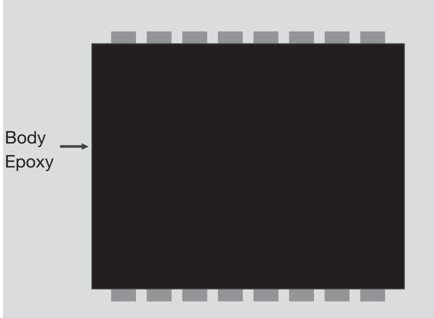


Figure 12.15 Basic Semiconductor Package

12.3.1 Package Classification

There are many different ways to classify semiconductor packages (Figure 12.16).

Access rights to the packages is one of the most convenient one. "Industry standard" packages are most widely used and are generally available to all manufacturers.

Newer package options, however, are often proprietary. They may be owned by one or just a few suppliers. Those packages may either offer significant improvements in cooling or in pin count (Figure 12.16).

- Semiconductor packages can be classified into two groups:

- Industry standard packages
  - Widely used by multiple suppliers
  - Often proven (old) technology
- "Proprietary" packages
  - Package may be used by one or only a few suppliers
  - Often a new technology which has not yet been embraced by the semiconductor industry

Figure 12.16 Semiconductor Packaging

Another way to classify packages is by the type of semiconductor component they house: power components and non-power (or small signal) components (Figure 12.17).

In general, power packages are better at removing heat from the semiconductor die but have a limited number of pins with high current rating. Small signal packages, however, are often optimized to provide maximum number of pins for given printed circuit board footprint areas.

Next, we will take a look at some examples of power and small signal packages.

- Next, packages can be classified into power and small signal packages
- Power packages
  - Designed to extract the maximum amount of heat from the die
  - Wide variation in design styles
  - Typically have "low" pin counts
- Small signal packages
  - Designed for minimum cost and foot-print size
  - Little variation in design styles
  - Typically have "high" pin count options

Figure 12.17 Semiconductor Packaging

12.3.2 Package Types

Figure 12.18 shows six different power packages. The package names provide insight into their construction:

Package Name: SCT595-05-1
Body Type: SCT595
Number of Pins:4

Package Name: P-DSO-20-6
Body Type: P-DSO
Number of Pins: 20

Package Name: SOT223-4-2
Body Type: SOT223
Number of Pins: 4

Package Name: P-DSO-36-10
Body Type: P-DSO
Number of Pins: 36

Package Name: P-TO252-3-1
Body Type: P-TO-252
Number of Pins: 3

Package Name: P-TO263-5-1
Body Type: P-TO263
Number of Pins: 5

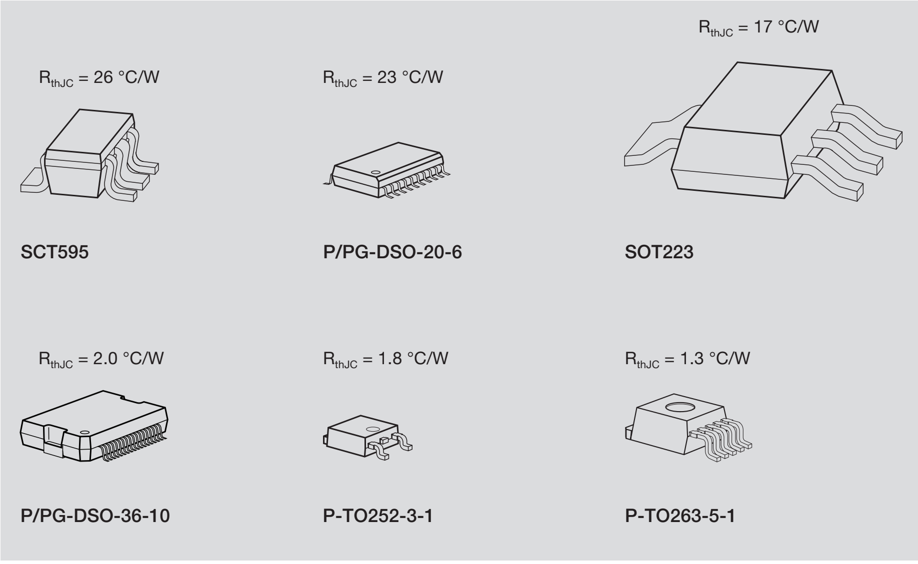


Figure 12.18 Power Packages

Figure 12.18 also shows the thermal resistance (junction to case) of the packages. The lower the number, the better the package is at removing heat from the semiconductor die.

In Figure 12.19 we see three common types of small signal packages.

The top left is the dual in-line package. Its pins are on the two long sides of the epoxy body. When this configuration proved insufficient to the growing pin count requirements, quad style packages were developed. These packages are often square, and have pins on each of their four sides. Finally, grid array packages have been developed for even higher pin counts. These packages use pins or bumps to provide the electrical connection to the package (bumps are shown here). Grid array packages have bumps (or pins) on the bottom of their packages, not on their sides like the dual in-line and quad style packages.

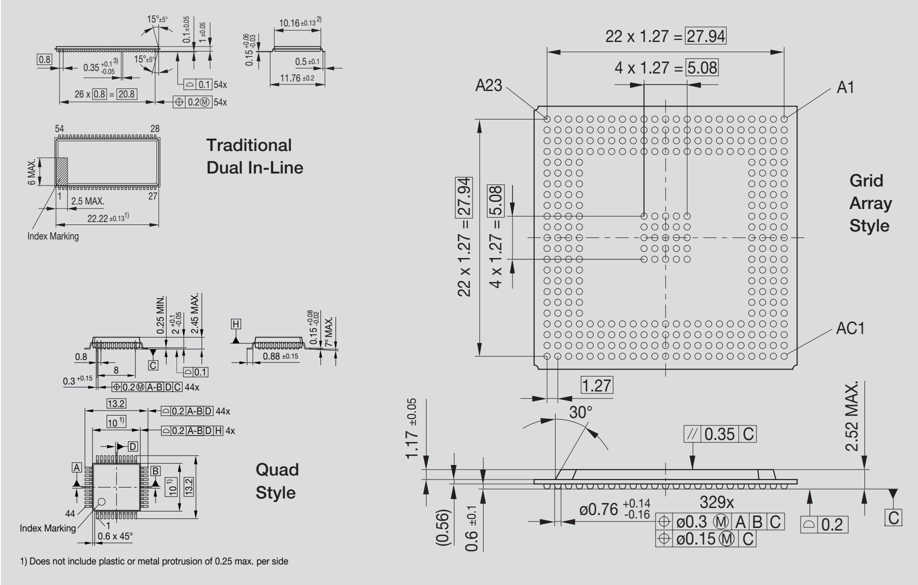


Figure 12.19 Small Signal Packages

12.4 Semiconductor Testing

Concurrently with the semiconductor component design, long before fabrication begins, the testing plan for the semiconductor component is developed. Without thorough tests, semiconductor manufacturers cannot determine if the component will work as expected.

As the semiconductor components themselves are updated or modified, the tests are also modified and improved.

- As the semiconductor is being designed, a thorough test algorithm is being developed to verify the device meets all functional and parametric requirements
- When the first components are manufactured, they are submitted for analysis with the new test algorithm to determine if the component needs to be improved before production begins

Figure 12.20 Testing Development

Developmental testing is usually much more rigorous than production testing.

For example, during development and prototyping, several different tests may be used to verify a semiconductor device meets a certain specification. In Figure 12.21, we illustrate four different tests that may be used to determine if a power MOSFET satisfies its R_{DSon} specification.

In production, however, some of these tests may be eliminated. This reduces the total test time and thus the cost of the silicon component. If the developmental tests were actually used in production, an integrated circuit could be cost prohibitive.

- When a component is released to production, the test algorithm is reviewed
- Some tests may be eliminated and/or modified for production purposes if the supplier can verify the devices functionality through other tests and guard-banding

Prototype Tests for R_{DSon}	Maximum (Ω)
I_{DS} = 1 A, V_{Supply} = 12 V, T_j = 25 °C	0.10
I_{DS} = 1 A, V_{Supply} = 12 V, T_j = 125 °C	0.17
I_{DS} = 4 A, V_{Supply} = 18 V, T_j = 25 °C	0.08
I_{DS} = 4 A, V_{Supply} = 18 V, T_j = 125 °C	0.14
Production Tests for R_{DSon}	**Maximum (Ω)**
I_{DS} = 4 A, V_{Supply} = 12 V, T_j = 25 °C	0.10

Figure 12.21 Production Testing

There are many different reasons why tests may be removed. We list several in Figure 12.22.

- The test itself may be destructive
- The test may be a subset of other tests
- The test may require manual intervention
- The test may require additional hardware
- The test may take too long

- Production testers cost millions of dollars and require an extremely high level of up-keep
- Suppliers want to minimize the level of testing required to ensure a customer's high quality expectation while shipping an affordable component

Figure 12.22 Why are some tests eliminated for production devices?

During production, some quantity of production devices are periodically subjected to a more thorough testing. The results of these monitor tests are used to ensure the production tests remain valid throughout the life of the product.

13. Quality and Reliability of Semiconductor Devices

In this chapter we will cover the importance of having high quality and reliable semiconductor products, common quality and reliability issues and their causes, how they can be prevented and effective problem solving tools that can help to avoid the reoccurrence of known defects.

The competition in automotive, industrial and communication industries is demanding higher acceptable quality levels for semiconductor products. What was once acceptable has become sub-par for the typical end customer. These customers now demand “best-in-class” quality and reliability performance for every semiconductor product that they purchase.

Quality and Reliability are not synonymous. However, these words have been used interchangeably by end customers whether they are buying cell phones or cars. This is a clear indication of what our current end customers are expecting to purchase, namely best-in-class, reliable products.

13.1 What is Quality?

Quality has been defined many different ways in different books by different experts in the field. The International Organization for Standardization (ISO) defined it in Quality ISO 9000: 2000 document as “the degree to which products and services fulfill the requirements and expectations of the customers”. It is simply the degree of how customer expectations are met. While that may sound simple, customer’s expectations can be expressed in a variety of ways. There are objective and explicit customer requirements such as functional performance, which usually include various forms of specifications with defined limits, and there are those expectations that are subjective such promptness of response or quality of service. The real measure of Quality is when the customer comes back to you and not the product.

13.2 What is Reliability?

Reliability is the property of a product which ensures the same good quality performance consistently. It is the measure of the probability that a product will perform a required function without failure under stated conditions for a stated period of time.

13.2.1 How is Reliability Confirmed?

Reliability is assessed by applying industry standard testing or supplier specific testing. Examples of industry standards are listed below:

JEDEC (Joint Electron Device Engineering Council)

EIAJ (Electronic Industries Association of Japan)

MIL-STD-883 (Department of Defense)

Underwriters Laboratories (UL)

IPC (Formerly Institute of Interconnecting and Packaging Electronic Circuits)

AEC (Automotive Electronics Council)

CDF-AEC-Q100/Q101

CE a mandatory conformity mark on many products placed on the single market in the European Economic Area (similar in scope to UL approval)

The testing performed during these examinations is designed to stress both package and silicon in order to precipitate intrinsic and extrinsic reliability failure mechanisms. Intrinsic failure mechanisms are those that are inherent to a particular type of technology, either silicon or package related. Intrinsic failures are typically caused by some form of technology or physical limitation. Extrinsic failure mechanisms are the results of manufacturing limitations. Those are usually based on the capability of a supplier to manufacture devices in a facility without inducing failures within the product. Both types of failures can occur as initial failures or as latent failures. The Table 13.1 lists examples of tests used by IC suppliers to verify reliability:

Table 13.1 Integrated Circuit Reliability Tests

	Stress Test	Failure Mode Verified	Failure Mechanism Verified
Accelerated Lifetime Stresses	Preconditioning (JESD22-A113)	Represents typical industry multiple solder reflow operation	
	Temperature Humidity Bias (JESD22-A101, A110)	Evaluates penetration of moisture while device is biased	Failures include metal corrosion
	Autoclave (JESD22-A102, A118)	Evaluates penetration of moisture under high temperature and pressure	Failures include metal corrosion and mobile contaminant collection
	Temperature Cycling (JESD22-A104)	Evaluates ability to withstand stresses due to temperature changes	Failures include die/package cracks, wire breaks, bond lifts
	Power Temperature Cycling (JA105)	Very rigorous evaluation of temperature changes while device is biased	Failures include die/package cracks, wire breaks, bond lifts
	High Temperature Storage (JA103)	Evaluates effect of long term affects of temperature without bias applied	Failures include oxidation of metal and inter-metallic shorts
Lifetime Stress	High Temperature Operating Life (JESD22-A108)	Evaluates reliability over extended period of time	Failures include dielectric breakdown, electromigration, ion contamination
	Early Life Failure Rate (AEC-Q100-008)	Evaluates infant mortality rate	Failures include dielectric breakdown, electromigration, ion contamination
	NVM Endurance and Data Retention (AEC-Q100-005)	Evaluates memory ability to retain data and sustain data changes	Failure due to data bit(s) retention
Package Assembly Integrity Tests	Wire Bond Shear (AEC-Q100-001)	Evaluates ability to withstand a horizontal push (shear) to the wire bond	
	Wire Bond Pull (MIL-883-STD Method 2011)	Evaluates ability to withstand a vertical pull to the wire bond	
	Solderability (JESD22-B102)	Evaluates ability of package to be soldered to another surface	
	Physical Dimensions (JESD22-B100, B108)	Evaluates ability to meet package drawing	
	Solder Ball Shear (AEC-Q100-010)	Evaluates ability to withstand a horizontal push (shear) to solder ball	
	Lead Integrity (JESD22-B105)	Evaluates ability of leads to withstand bending during assembly or rework	
Electrical Verification Tests	Electrostatic Discharge Human Body Model (AEC-Q100-002)	Evaluates ability to withstand high voltage through a resistance	
	Electrostatic Discharge Machine Model (AEC-Q100-003)	Evaluates ability to withstand high voltage directly	
	Latch-Up (AEC-Q100-004)	Evaluates ability to withstand high current injection	
	Electromagnetic Compatibility	Determines if electromagnetic radiations might be excessive	

The above tests have been designed to reveal intrinsic and extrinsic reliability issues with semiconductor devices as well as verify the lifetime reliability expected by automotive customers.

A reliable semiconductor device is capable of passing the specified stress testing and provides consistent performance over time, temperature, pressure and other acceleration factors defined by the different industry standards or by the customers. There are different reliability test standards which are minimum expectations by automotive customers. Different customers typically specify their additional reliability requirements, if any, before purchasing a device. These requirements are based on their expected product life including any warranty given to end users (the warranty given to car buyers, for example). Also taken into consideration are such factors as: what applications determine the kind of environmental stresses that the product and the device will be exposed to due to being under the hood of the car (elevated temperature, salt spray, etc.), or used in the tire which exposes the product to soap, water and weather conditions. It is important to note that the result of reliability tests can only give an initial data point of the product's capability to pass the various stress tests and provide initial statistics including Mean Time To Failure (MTTF) and Mean Time Between Failures (MTBF). The variations during the manufacturing processes impact the actual product characteristics and will cause the actual characteristics distribution to shift (Figure 13.1).

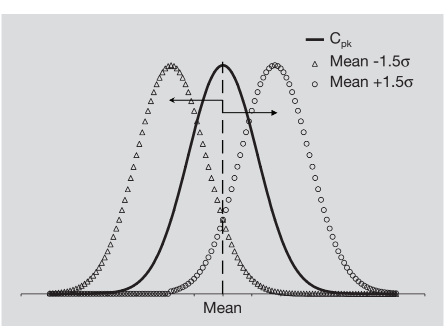


Figure 13.1 Illustration of Process Shifts

Inherent causes of variations that result in process shifts are the changes in the input characteristics which include variations in raw materials, operator's skills and knowledge, machine and process capability, environmental conditions, i.e. temperature, humidity, and measurements used. This is commonly referred to in manufacturing as the 6 Ms being Man, Machine, Material, Methods, Mother Nature, and Measurements. When a problem is detected, these M's are also the starting point for the commonly used problem solving tool called the Fishbone Diagram, used for analyzing what may have changed in any one of the 6 Ms.

13.3 Customer Expectations

The current expectations of customers have redefined quality as meeting their requirements consistently over specified period of time, with the best imaginable service, at the best price. This translates into what the car makers and electronic module manufacturers call the Zero Defect (ZD) requirement. ZD expectation means that absolutely no defects are shipped by suppliers to users during the life of the product.

It may sound like demanding the impossible but this is what will insure return business and loyal customers resulting in a long term business relationship. There are manufacturers who are leading the industry through high quality products and services.

13.4 Meeting the Demands of ZD to Earn Customer Loyalty

The automotive industry is asking for high quality parts because we, as car buyers, do not want to be the one customer that gets, for example, a faulty tire pressure monitoring device out of the one million cars sold. In other words, one part per million (PPM) failure rate is not acceptable, especially if "it's got to be me". If 1 PPM is not acceptable at car level, what quality level is acceptable at components level?

Typically, a car contains 50 electronic control modules and the average electronic control modules have about 300 components. If each component is at a 1 PPM failure rate, the

module failure rate will be at 300 PPM and the potential failure rate of the car will be 15000 PPM. This can translate to 15000 cars serviced or recalled for every 1 million cars sold. The impact of such a failure rate to the reputation of any car company and its customers' loyalty could be devastating which may result in a big loss of business.

To meet the demand for ZD, it is critical to design and manufacture dependable components. Many customers say "garbage in = garbage out". The next sections will discuss important lessons learned and what should be done to prevent and control common semiconductor device quality and reliability issues.

13.5 The Power of Preventive and Protecting Process Control

Since the expectation is zero-defect, the literal meaning of the phrase is to not make any defective product at all. The traditional approach of applying inspection and tests as part of the process control becomes outdated and irrelevant because both practices mean that defect(s) may be created which then might be caught later in the next process steps. Catching them subsequently is not sufficient or even affordable because it is well known that inspection is only 80% effective and testing for all the product characteristics is very costly. It is almost certain that process controls that are still designed this way will pass defective parts to external customers.

ZD can only be achieved by not making defective products in the first place; therefore, controls are to be applied before defective parts are created without relying on inspection and testing. Error-proofing has been widely accepted to achieve this. However, the different levels of error (or mistake) - proofing have not been clearly understood.

Error-proofing means to design a process such, that process errors, that may cause defects are not made either by machine or operator. Clear understanding of the cause and effect is critical, and process control measures are employed in order to eliminate the cause(s) and, therefore, to prevent the creation of a defect.

To illustrate the point, we will examine common quality issues in the manufacture of semiconductor devices. Semiconductor manufacturing is typically divided into the following three major steps: Wafer Fabrication, IC Assembly/Packaging and Final Testing.

13.6 The Making of ZD Semiconductor Devices

Manufacturing of semiconductor devices can consist of 300 or more process steps. The environment in the processing foundries is highly controlled and ultra clean. Each of these hundreds of steps has many process parameters that need to be tightly controlled to guarantee the expected product characteristics consistently. There are three major groups of process steps.

13.6.1 Wafer Fabrication

Wafer Fabrication is the manufacture of semiconductor devices through a multiple repetitive sequence of photographic and chemical process steps so, that integrated circuits are gradually created on a pure, single crystal silicon wafer. The number of process steps is determined by the complexity of the circuit. Most of the integrated circuits have 300 or more process steps. The process steps include deposition, removal of material, patterning, doping, diffusion and wafer testing.

13.6.2 IC Assembly (or packaging)

IC Assembly (or packaging) is the installation of an individual die (or a small cluster of dies) into a single component package to facilitate connection to the application board or module. There can be several hundreds of devices fabricated on a single wafer depending on the size of each chip. There are typically about ten steps in the assembly process: Wafer Sawing to break the wafer into individual die, Die Bonding (Die Attaching) to the lead frame, Wire Bonding, Encapsulation or Molding, Curing, Plating of the leads, Laser Marking, Forming and Singulation.

13.6.3 Device Testing

Device Testing involves checking each packaged device for functionality. While each

integrated circuit was previously tested at wafer level, prior to assembly, the assembled device is retested after packaging. This is to ensure that the die was not damaged during the assembly processes and that the die-to-pin interconnection was completed and conforms to the package design. Tests are done at room temperature but they may also be performed at hot or cold temperatures depending on customer's requirements and demands of the application. Typical lead time from the bare silicon wafer to the packaged and tested devices can take six to eight weeks.

13.7 Causes of Reliability Issues

The major stages of manufacturing semiconductor devices discussed earlier may cause quality and reliability issues categorized as:

13.7.1 Assignable Cause

Assignable Cause - refers to an inadvertent specific incident which is caused either by an operator's mistake or some failure either in the equipment or in the process. It is also referred to as a quality accident because it is not planned for and not intended to happen. Examples include broken saw blades that will cause die chipping and cracks, or wrong test program applied, wrong tools used, or missed process step(s).

13.7.2 Common Cause

Common Cause - means inherent process limitations that cause periodic change in product characteristics. These include typical variation in input process parameters like temperature, pressure, characteristics of tools and raw materials that are limited by the laws of physics and chemistry. The cited factors are called process capability limitations and examples include minute foreign materials in the a clean room, a shift in chemical solutions concentrations caused by variations in incoming raw materials or chemicals and variation in viscosity, ductility and brittleness of raw materials, etc.

13.8 Conditions of ZD Technology

Due to the complexity of both assignable and common causes of quality issues, it is critical that a holistic methodology and a systematic approach is defined and executed across the organization to achieve zero defects in the manufacturing of semiconductor devices. This methodology must address the fundamental belief that the contribution of every employee impacts product quality and that attention to details in every step of the production of the device is a must. This mentality should start from concept development, design, manufacture and continue through shipping.

Advanced Product Quality Planning (APQP) is intended to cover the proactive attention to details in each step of the Product Development Process (PDP). The use of Potential Failure Mode Effect Analysis (PFMEA) for both Design (dFMEA) and Process (pFMEA) and Process Control Plans (PCP) facilitate planning for quality.

Assignable causes that result in a so called "Quality Spill" can be eliminated by the implementation of proactive process controls or predictive maintenance. The key words are "proactive and predictive" which means before the event happens and before an error causes a defect, for example:

- Package Cracks due to worn out cutting tools at the Trim and Form process can be prevented by controlling the number of cycles or shots based on the life of the tool. A simple counter connected to the machine controls its operation and shuts the machine down before the tool is used to cut beyond its known life.

- Over or under etching during the wafer fab process[1] can be prevented by fixing the time or temperature through the use of controllers or sensors which will stop any production beyond set limits.

- Incomplete mold due to wrong mold pellet size or insufficient mold material can be prevented by using an on-line weighing scale that will only start the molding process if the machine detects the right weight.

1) *Wafer fabrication process or fab process can be used interchangeably*

There are many common proactive controls that eliminate potential errors thereby preventing defects. Some of them are available as features or options from equipment manufacturers.

Common causes that result in ongoing quality issues are not easily minimized (error-proofed), like assignable causes. Error-proofing is addressed through a combination of continuously improving process capability and implementing device tests. Process capability index, which is referred to as C_{pk}, is driven by manufacturing process owners by constantly upgrading their equipment as new ways and technology emerge. C_{pk} is defined as:

C_{pk} = the lesser of [(USL-Mean) / σ, (Mean-LSL) / σ]
where
USL upper spec. limit,
LSL lower spec. limit
σ standard deviation

Depending on the component and application $C_{pk} \leq 1.67$ is unacceptable.

13.8.1 Defect Density

The inherent variations in the hundreds of process steps performed under precise control result in a certain defect rate and this rate is expressed in terms of Defect Density. Over time, this rate can be determined for a process step and is used as a measurable metric that needs to be reduced to track improvements at specific process steps.

Defect density issues are common-cause type failure mechanisms that occur due to inherent process limitations during the wafer fabrication process and are not detected at the process source or during device testing. Very often, these defects are not detected prior to shipment to customers, typically due to the need for an acceleration factor like heat or high voltage.

13.8.2 Continuous Improvement (CI) Plans

It is a set of activities that are defined and implemented to reduce the density of the defects in specific process steps based on priority and criticality. These activities target the process and environment variations to reduce or minimize impact to the product. Normally, the variation cannot be totally eliminated, but significant reduction in common cause defects can be achieved through disciplined implementation of process upgrades, material changes and application of more stable tools. This improves the overall process capability, thus, reducing defects.

Some examples of CI activities are:

- Fabrication Clean room Particle reduction activity
- Reduction of manual handling of wafers
- Improved design of Clean room Equipment (carriers, transport carts, wafer alignment tools, etc.)
- Improved equipment control protocols
- Targeted inspections of production wafers after critical process steps
- Change to more stable or durable materials

13.9 What Are ESD and EOS Failures?

We will review and summarize in this chapter how these failure mechanisms impact reliability of semiconductor devices and how they can be avoided.

13.9.1 Electrostatic Discharge or ESD

It is a single event of a fast transient transfer of electrostatic energy which results when two objects with different potentials come into direct contact. ESD can also occur when high intensity electrostatic field exists between two objects in close proximity. ESD damage is one of the primary causes for early life failure of semiconductor devices. An example of an ESD event is the shock one receives sometimes during winter when one touches a doorknob after walking across a carpeted floor.

13.9.2 Prevention of ESD

ESD damage during the manufacture and use of semiconductor devices can be prevented

through the deployment of different approaches which are as follows:

13.9.2.1 Reducing the Build up of Static Charge

This is accomplished by using materials that are less likely to generate static charges in the work place. Assembly line equipment should be free of moving parts such as rollers, stops, etc. made of electrical isolator materials that may generate charge. Carriers used to move products in the production area should be constructed of ESD dissipative materials. Ionizers may be used in critical locations to neutralize charges accumulated during the assembly process. Personnel should observe strict ESD reduction discipline by using wrist and foot straps, protective lab coats, and ensure that the protection they use is in good condition (regular testing of wrist straps, for example).

13.9.2.2 Safe Dissipation of Charge Build-Up

Ensure safe dissipation of any charge build up during the manufacturing process. The key to this approach is that each equipment or object that comes in contact with the semiconductor is connected to a common ground point. Conductive flooring, workstations, storage racks must have direct grounding to a common point in order to implement an effective ESD control program.

13.9.2.3 Designing ESD Protection Structures on the Silicon Chip

Protection circuits exist for different wafer technologies that provide some level of protection to external transients by dampening/reducing the transient energy level of an incoming ESD pulse.

Prevention of charge build up and effective dissipation of accumulated charge are the most effective approaches to reduction of ESD damage to semiconductor products. ESD requirements for storage and handling of ESD sensitive semiconductor devices are spelled out in the industry recognized standards JEDEC 020B/C. These standards are to be implemented as a minimum to prevent damage to sensitive devices.

13.9.3 Electrical Overstress or EOS

EOS is a more general failure mechanism which results from the exposure of a semiconductor to voltage or current conditions beyond the design capability of the device. ESD damage may be considered as a subset of EOS damage. The damage to a semiconductor resulting from either one of these events may be so subtle that determining which event actually happened is not always possible.

Typically, a device damaged by EOS will exhibit physical signatures in the silicon resulting from a “punch-through” of the dielectric material (oxide), a fused or reflowed metal structure on the surface of the semiconductor or junction damage due to thermal runaway.

The dielectric punch-through occurs when a circuit element is exposed to a voltage larger than the dielectric strength of the oxide layer, or to a rapidly changing voltage across the dielectric material causing large currents flowing inside the material. MOS devices are especially vulnerable to both causes. As semiconductor technology evolves, oxide thicknesses are becoming thinner and features smaller making the devices more susceptible to this type of failure.

Fused or reflowed metal structures occur when current density in a circuit element exceeds the rated limit. Metallization lines are heated leading to reconfiguration or even fusing open.

Excessive power dissipation can lead to localized heating at device junctions. This forms a positive feedback mechanism where higher localized temperature leads to lower resistance, which, in turn, leads to more current flow, resulting in more heating. Ultimately, the junction collapses and the silicon melts at the junction.

Elimination of EOS failures requires detailed information about the conditions under which the failure occurred. Most failures are caused by improper test procedures of modules or sub-assemblies. Inadvertent socketing errors, careless power-up techniques and the use of a

component in a circuit beyond its design capabilities are all common causes of EOS damage.

13.10 Outliers

Outliers are those devices whose measured characteristic values are outside the typical variance (normally within ±3σ) of the mean or set target. Note that the value may be well within established specification or operational limits and still be an outlier. Since we have redefined quality as not just meeting requirements but how consistently the product characteristics are from one unit to the next, we can now ask what about those parts that meet the specifications but are not close to the mean? Outliers are the result of the inherent variability in the production process, resulting in a distribution of products around a target value. This value can be a final electrical specification, a mechanical dimension, or a critical process characteristic of wafer fabrication. Statistical process controls form an important control mechanism used during the manufacturing cycle to ensure that outliers are not shipped to customers.

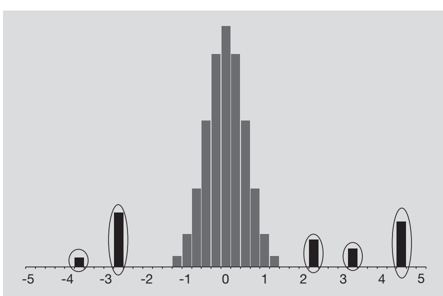


Figure 13.2 Illustration of Outliers

Some statistical techniques have been identified in the semiconductor industry to separate outliers from the other parts that can be shipped. Those common techniques are as follow:

13.10.1 Part Average Testing (PAT)

It is a statistically based method used during parametric testing (wafer probe or final test) of a device to reject components with abnormal characteristics called outliers. Any product characteristic that may drift in time and impact long term consistency of the device is normally a priority cause for PAT. Test data is collected over time, the average for the part is the target and the PAT limit is typically set to Six Sigma.

Mature test programs feature **Dynamic PAT** to continuously update test limits based on accumulated actual performance data. Common characteristics that must have PAT limits according to the Automotive Electronics Council recommendations are:

- Leakage
- Break down voltage
- Stand by current (IDD, ICC)
- Gate Stress
- Voltage Stress

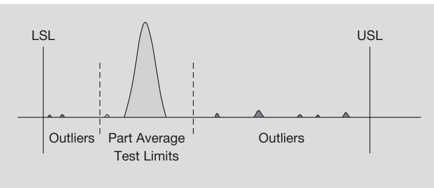


Figure 13.3 Illustration of Part Average Testing (PAT) Limits

13.10.2 Statistical Bin/Yield Analysis (SBA or SYA)

It is another statistical technique used to eliminate outlier lots based on test yields. Similar to PAT limits, tests limits are calculated per test bins by identifying the average yield plus Six Sigma. If test results from any production lot yielding fallouts in excess of the fallout limits in a specific test bin (yield outlier), there could be a specific cause defect in the lot which should be investigated. This lot will be called a maverick lot and is typically not shipped to the customer unless the root-cause and risk are fully understood. Common test bin configurations that have bin limits are:

- Continuity Failures
- PAT failures
- Functional failures
- Kelvin failures

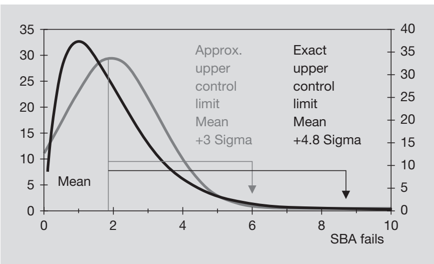


Figure 13.4 Illustration of Statistical Bin/Yield Analysis (SBA)

13.11 Latent Defects

Latent Defects are flaws or other imperfections in a device which surface after the semiconductor component is placed in the application circuit and is exercised or activated. Latent defects are inherent weaknesses in the device which are not detected by examination or routine tests and the conditions of tests are aggravated or the potential causes of failure are accelerated at the point of use. These defects can be caused by either intrinsic or extrinsic factors. Intrinsic factors are bounded by physical limit(s). Extrinsic factors are associated with the manufacturing processes used to produce a particular device. The end result of a latent defect is an overall reduction of the reliability of the device in its intended application.

Preventing and eliminating latent defects can be accomplished by one or more ways:

- Some form of screening such as a burn-in test
- Accelerated reliability life testing
- Qualification testing beyond predicted operating extremes
- Predictive and preventative manufacturing process controls
- Robust product design

13.12 Failure Analysis and Problem Solving

When semiconductor devices fail at the point of use anywhere in the supply chain, the device will have to be analyzed to determine what caused its failure. This process is called Failure Analysis (F/A). The part is returned to the manufacturer where it undergoes physical and/or chemical analysis depending on the nature of the detected or suspected failure.

It is recommended that an initial analysis at board level be performed to isolate the root-cause of the failure at component level. A common practice is a called ABA swapping which is done to see if the problem follows the component. The method is illustrated in Figure 13.5a and Figure 13.5b:

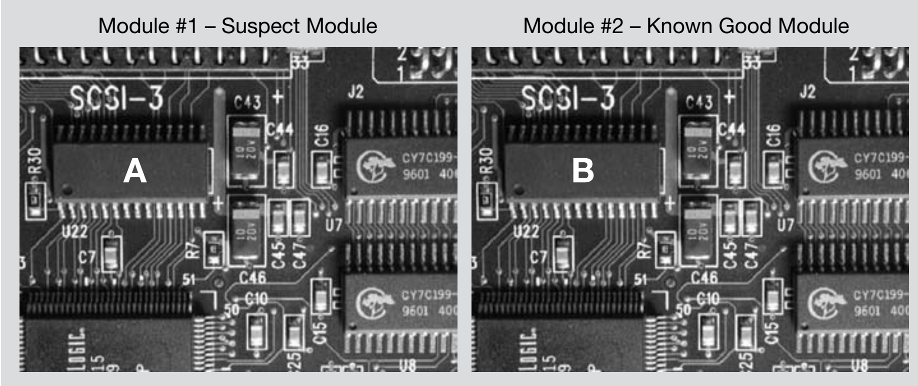


Figure 13.5a Module #1 fails testing due to suspected failure of device "A". Module #2 is a working module for reference.

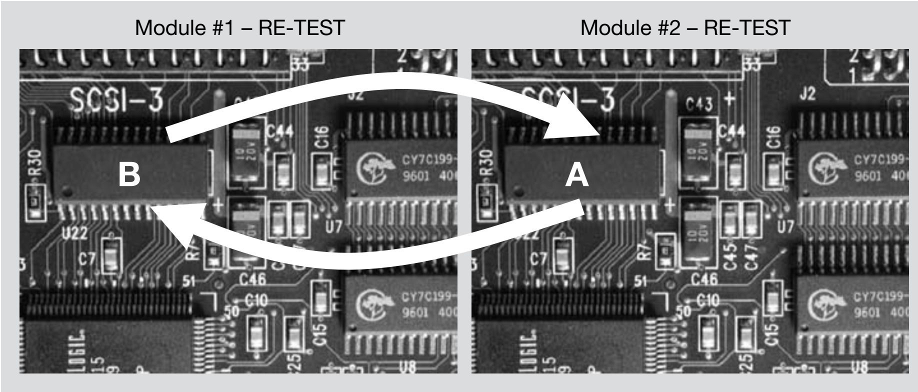


Figure 13.5b Suspect device "A" is replaced with known good device "B" and the suspected "A" part is mounted on Module #2. Both modules are re-tested to determine if the failure follows the part or the module. If Module #2 now fails, device "A" is assumed to be the root cause. If the Module #2 does not fail, the root cause is assumed to be related to component(s) other than part "A" in Module #1.

Figure 13.5a and Figure 13.5b describe the ABA swap technique.

13.13 Optimization of the Results of F/A

Optimizing F/A results comes as a result of combined efforts between the customer and the supplier. The first step in optimizing F/A results is to obtain a very accurate description of the observed problem from the customer. This description includes functional, behavioral, and measured electrical observations of the suspect device. Once the customer and the supplier have described the failure, appropriate electrical and physical analysis techniques and tools can be selected to increase the probability of finding the root cause of the observed failure mode.

13.13.1 Best Practices

Best practices for F/A include accurate failure localization which is the result of obtaining key information regarding the nature of the failure in the device. This key information is collected by several methods. The first method is the use of an automated electrical functional test.

Those tests are, by nature, intended to isolate failures down to functional blocks of the device. Once the functional testing has been performed the test results are then used to further isolate the failure to a particular lower level of the device structures. Behavioral models are used to simulate the flawed circuitry. Once the failure has been identified and compared against the device schematic, more complex physical analysis techniques are used to either stimulate the failure or correct the physical flaw and stimulate the device to function correctly. Once the isolation has been made, the destructive analysis of the device can begin to determine the exact root cause of the observed failure. A typical F/A flow is shown in Figure 13.6.

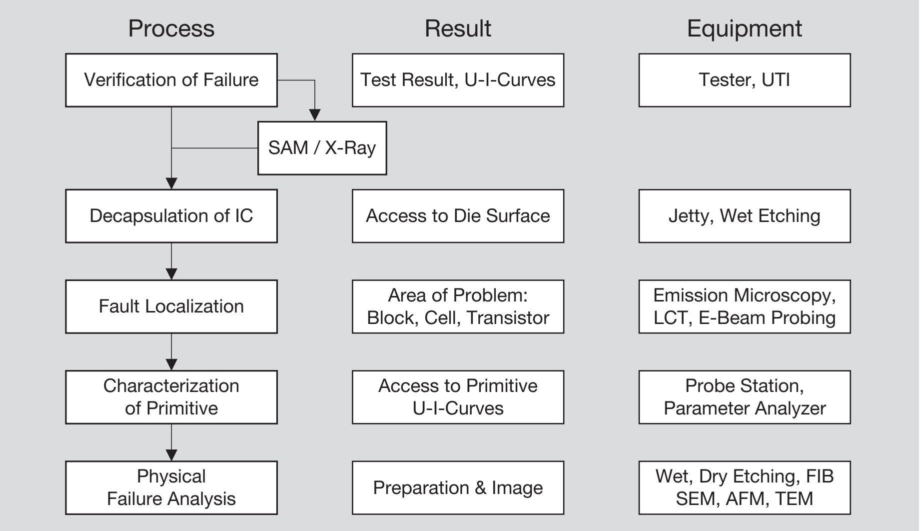


Figure 13.6 Failure Analysis: Processes Flow Chart, Results and Equipment used

Equipment used during F/A: each of the F/A process steps in Figure 13.6 is effectively executed with corresponding equipment or tools to achieve accurate results. Results are interpreted by trained Failure Analysis engineers and the conclusions are utilized for effective problem solving.

NOTES:
U-I-Curves or IV Characteristics: current-voltage characteristics
UTI: universal test instrument
LCT: liquid crystal thermography
E-Beam: electron beam
SEM: scanning electron microscope
FIB: focused ion beam
AFM: atomic force microscopy
TEM: transmission electron microscopy

13.13.2 Verification of Failures (non-destructive)

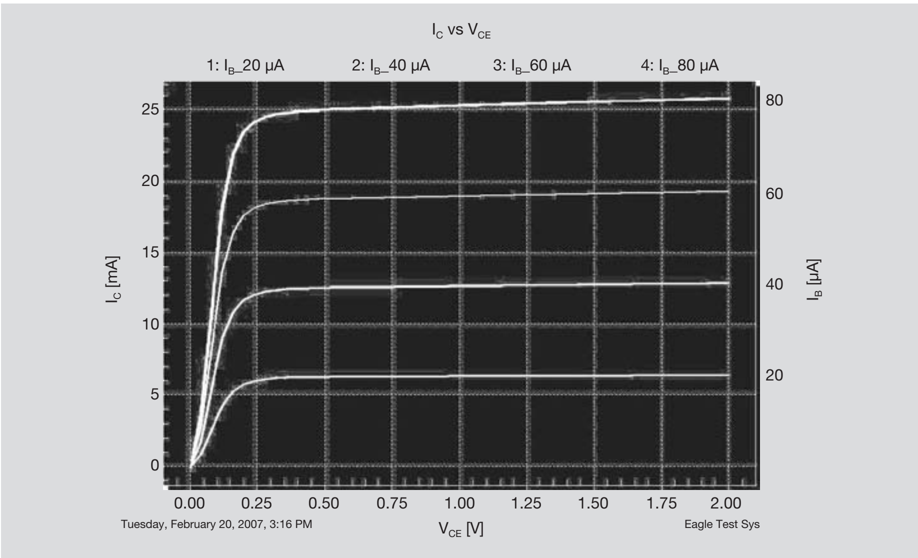
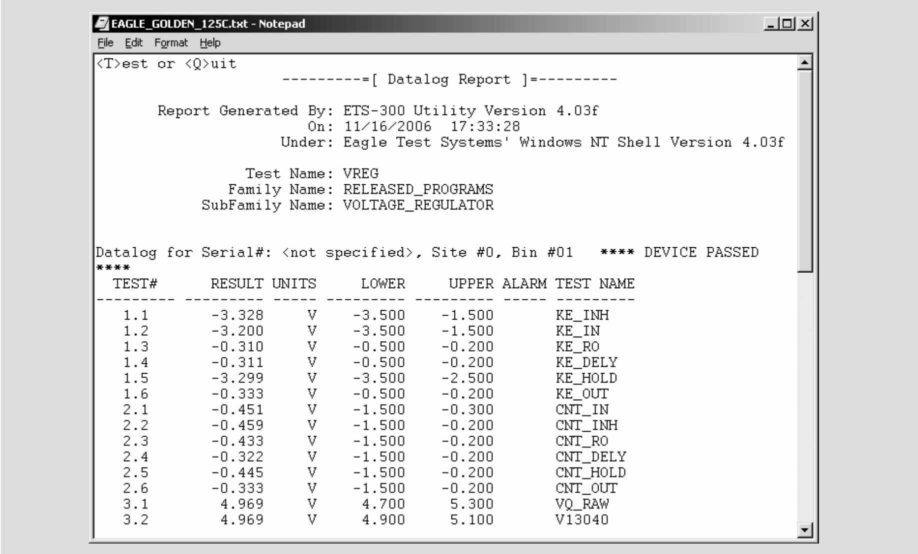


Figure 13.7 **The curve tracer is a widely used tool to verify device IV characteristics**
Horizontal axis: V_{CE} (V) (collector to emitter voltage)
Vertical axis: I_C (mA) (collector current)
I_B: base current (μA)



```
EAGLE_GOLDEN_125C.txt - Notepad
File  Edit  Format  Help
<T>est or <Q>uit
                         ----------=[ Datalog Report ]=----------

       Report Generated By: ETS-300 Utility Version 4.03f
                        On: 11/16/2006  17:33:28
                     Under: Eagle Test Systems' Windows NT Shell Version 4.03f

                 Test Name: VREG
               Family Name: RELEASED_PROGRAMS
            SubFamily Name: VOLTAGE_REGULATOR


Datalog for Serial#: <not specified>, Site #0, Bin #01   **** DEVICE PASSED
****
  TEST#      RESULT UNITS     LOWER     UPPER ALARM TEST NAME
--------- --------- ----- --------- --------- ----- ---------
   1.1       -3.328     V    -3.500    -1.500       KE_INH
   1.2       -3.200     V    -3.500    -1.500       KE_IN
   1.3       -0.310     V    -0.500    -0.200       KE_RO
   1.4       -0.311     V    -0.500    -0.200       KE_DELY
   1.5       -3.299     V    -3.500    -2.500       KE_HOLD
   1.6       -0.333     V    -0.500    -0.200       KE_OUT
   2.1       -0.451     V    -1.500    -0.300       CNT_IN
   2.2       -0.459     V    -1.500    -0.200       CNT_INH
   2.3       -0.433     V    -1.500    -0.200       CNT_RO
   2.4       -0.322     V    -1.500    -0.200       CNT_DELY
   2.5       -0.445     V    -1.500    -0.200       CNT_HOLD
   2.6       -0.333     V    -1.500    -0.200       CNT_OUT
   3.1        4.969     V     4.700     5.300       VQ_RAW
   3.2        4.969     V     4.900     5.100       V13040
```

Figure 13.8 **Automated Test Equipment (ATE) is a computer controlled unit that automates the test process to verify that device parameters and functionality are within the specified limits. The output is typically a text file that shows all the device measurements compared to their specified limits.**

13.13.3 Fault Localization (non-destructive)

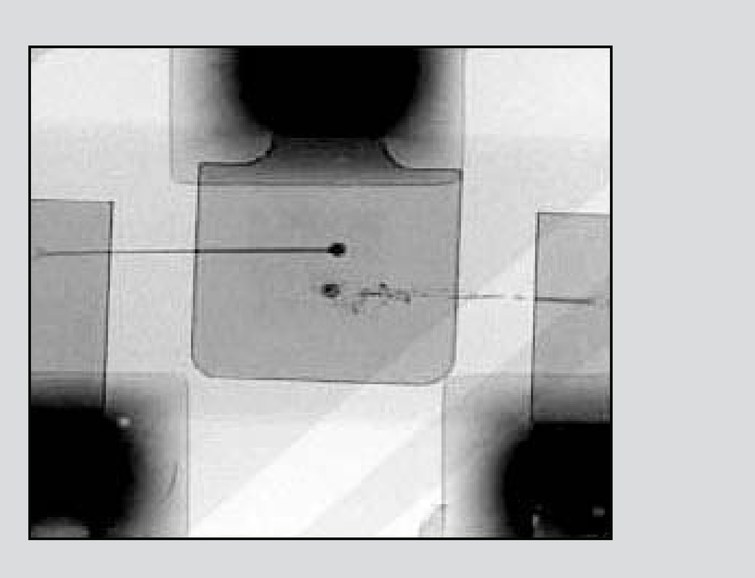


Figure 13.9 X-ray inspection is a common tool to reveal internal structures of the devices.

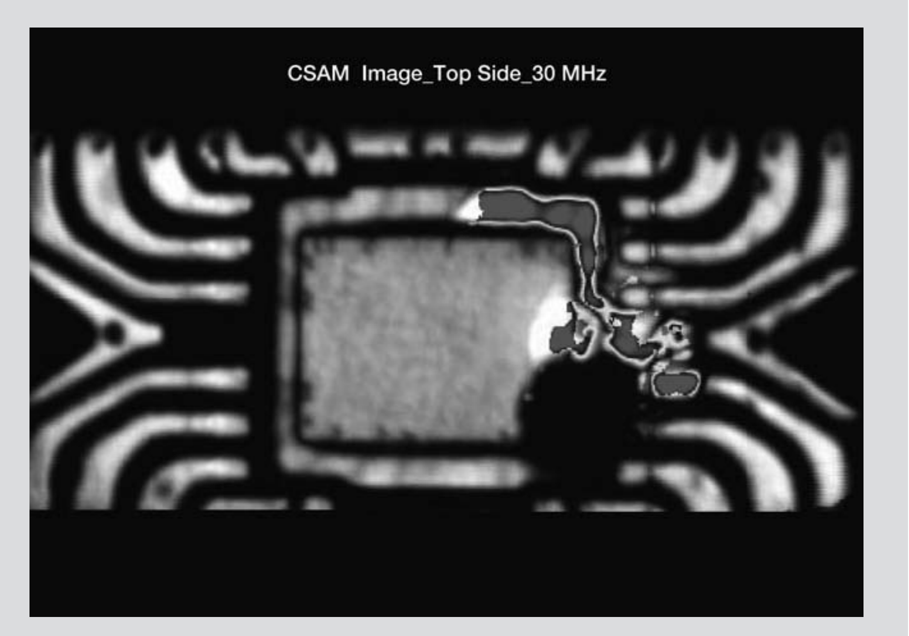


Figure 13.10 Scanning acoustic microscopy (CSAM) uses focused sound waves to detect voids, delamination and cracks within the package of the device.

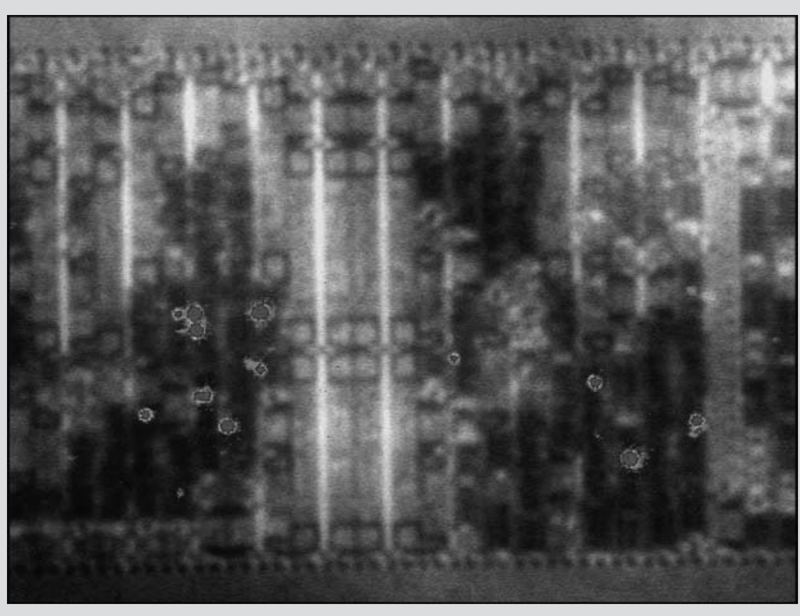


Figure 13.11 Photo Emission Microscope (PEM or EMMI) is an important tool to localize defects in integrated circuits. It detects faintly visible and near-infrared light emitted from silicon under numerous failing conditions.

Figure 13.12 Fluorescent Microthermal Imaging produces static thermal maps of surfaces or/and localizes "hot spots".

Figure 13.13 Liquid Crystal Thermography uses temperature sensitive chemicals and polarized cameras to detect hot spots on the die surface.

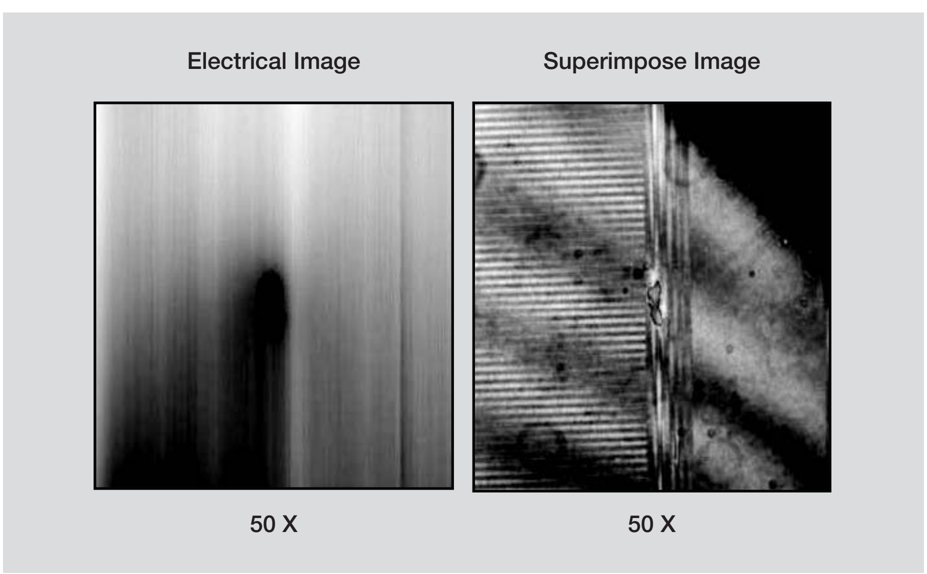


Figure 13.14 Thermally Induced Voltage Alteration (TIVA) is an imaging technique which pinpoints the location of electrical shorts on a device. A laser induces thermal gradients in the device, which changes the amount of power that the device consumes.

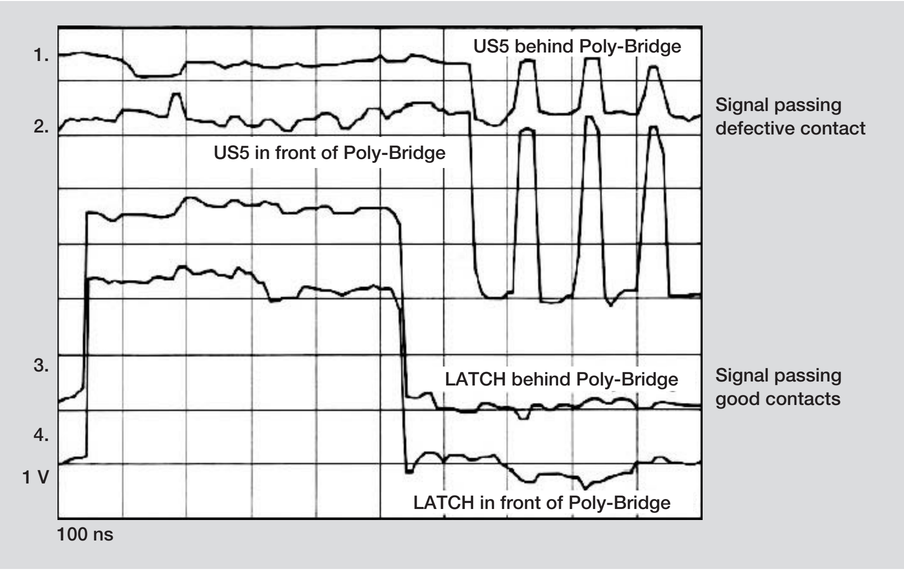


Figure 13.15 In electron beam probing, the voltage distribution at the surface of an IC can be displayed (active voltage contrast) by scanning a focused electron beam.

13.13.4 Decapsulation (destructive) - Dry decapsulation Wet Etching (Fuming Nitric Acid) Physical Analysis -

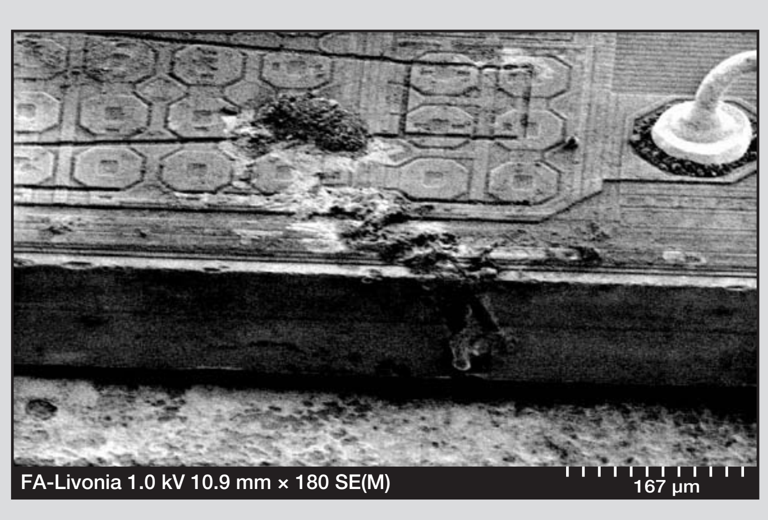


Figure 13.16 Scanning Electron Microscope (SEM) uses an electron beam to induce the emission of secondary electrons from the sample. The secondary electrons are detected to provide high magnification images of the sample.

13.14 Effective Methods of Problem Solving for Semiconductors

When quality issues are not prevented, it is critical that an immediate solution is achieved and lessons are learned from the results of the problem solving. A problem solving event is considered effective when the real root cause, and not the symptom, is identified and eliminated. There are many well publicized problem solving methodologies and they all revolve around the general sequence: Plan, Do, Check and Act (PDCA). The most common practice in the Semiconductor industry to structure the problem solving efforts is the Six Sigma problem solving method.

It is important to note that an effective problem solving process is a pre-requisite for any Zero Defect program. The Six Sigma process is summed up in the DMAIC mnemonic: **D**escribe, **M**easure, **A**nalyze, **I**mprove, and **C**ontrol. The inputs and outputs of the Six Sigma problem solving process are then communicated using the 8D format to both the supplier and customer problem solving teams. This format has evolved from the early use of the 5 phase process developed by the automotive OEMs.

The Infineon 8D process is outlined in Figure 13.17:

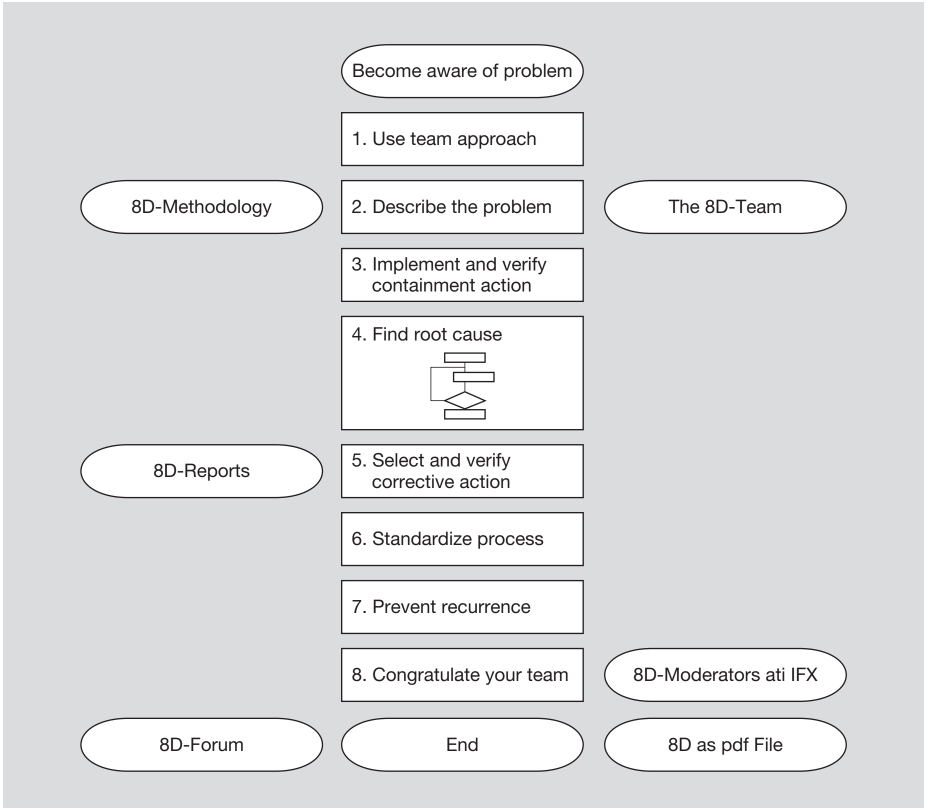


Figure 13.17 8D Process Outline

13.14.1 Problem Solving Tools

Effective problem solving is achieved with the expert use of problem solving tools applied to the process. Different customers require different tools and formats to document and communicate the steps and results of the problem solving procedures. Some common

problem solving tools are shown in Figure 13.18:

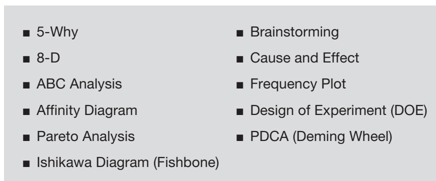
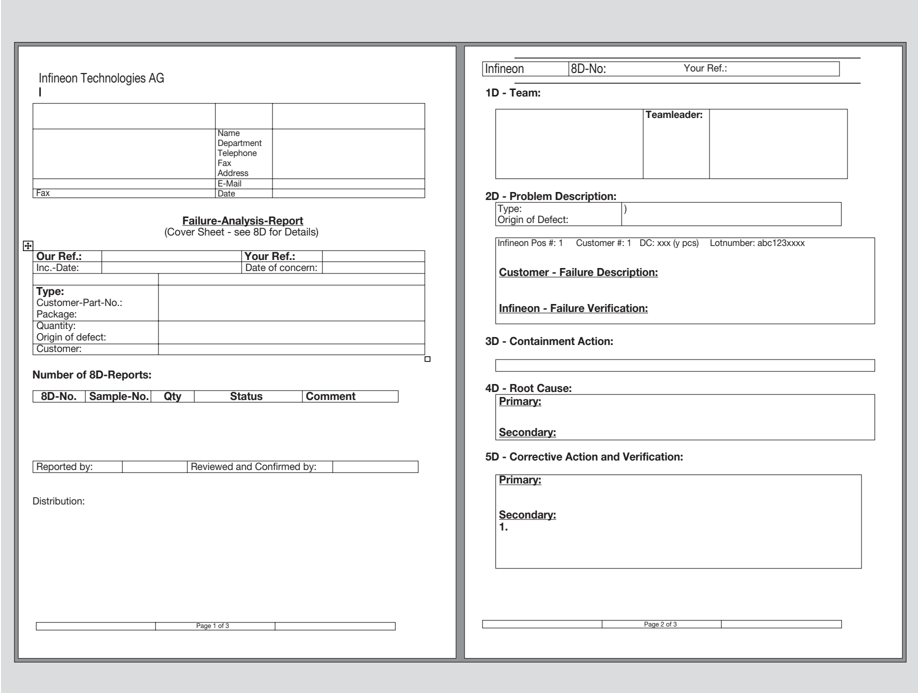


Figure 13.18 Common Problem Solving Tools

13.14.2 Problem Solving Effectiveness

The 8D process has been developed in order to increase the effectiveness of problem solving and facilitate the communication of results. There are many factors that influence the effectiveness of solving a problem. The following factors need to be considered when planning the problem solving activity:

- Is the problem important enough that it requires a team to address?
- Do the potential solutions to the problem exceed the skills of the individual attempting to solve the problem?
- Does the problem have cross functional character?



Infineon Technologies AG

	Name Department Telephone Fax Address	
	E-Mail	
Fax	Date	

Failure-Analysis-Report
(Cover Sheet - see 8D for Details)

Our Ref.:		**Your Ref.:**	
Inc.-Date:		Date of concern:	

Type: Customer-Part-No.: Package:	
Quantity: Origin of defect:	
Customer:	

Number of 8D-Reports:

8D-No.	Sample-No.	Qty	Status	Comment

Reported by:		Reviewed and Confirmed by:	

Distribution:

Page 1 of 3

Infineon	8D-No:	Your Ref.:

1D - Team:

	Teamleader:	

2D - Problem Description:

Type: Origin of Defect:	)

Infineon Pos #: 1 Customer #: 1 DC: xxx (y pcs) Lotnumber: abc123xxxx

Customer - Failure Description:

Infineon - Failure Verification:

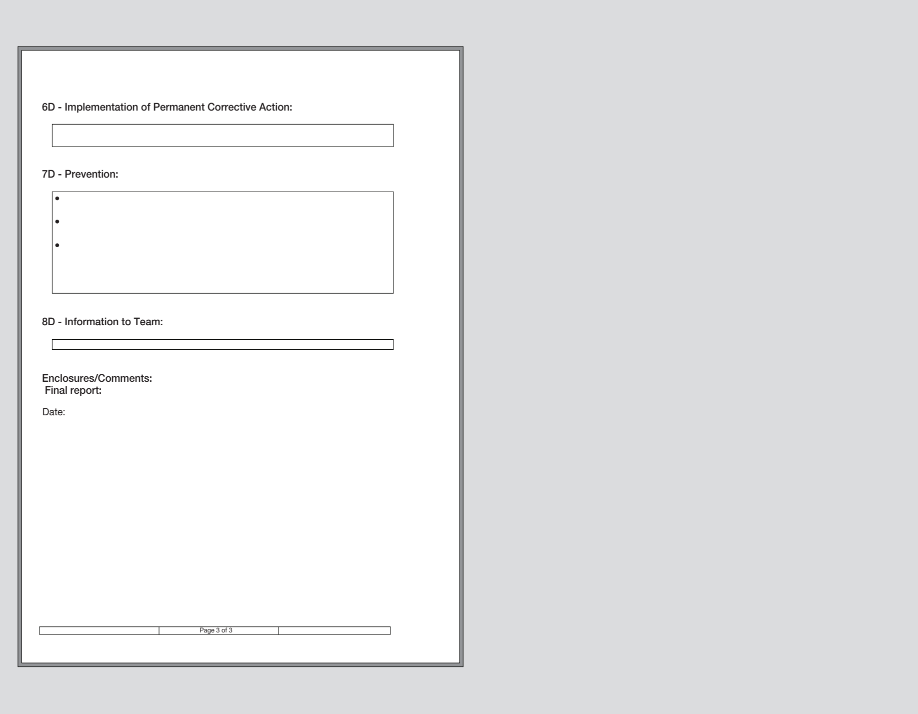
3D - Containment Action:

4D - Root Cause:
Primary:
Secondary:

5D - Corrective Action and Verification:
Primary:
Secondary:
1.

Page 2 of 3

Figure 13.19 Example of Infineon 8D



6D - Implementation of Permanent Corrective Action:

7D - Prevention:

-
-
-

8D - Information to Team:

Enclosures/Comments:
Final report:

Date:

Page 3 of 3

Figure 13.20 Example of Infineon 8D (continued)

13.15 Conclusion

In conclusion we can state that as we design and manufacture complex electronic modules for different control functions found in a modern vehicle, the Quality and Reliability of the semiconductor devices used in those modules are critical determining factors in the manufacture of the high quality cars every driver wants. Quality is when the customer is satisfied to the point where he comes back for new products, not the product is coming back for repair!

Quality and Reliability are not achieved accidentally and they are not free.

14. Introduction to Motor Control

In this chapter, an investigation is presented into motor control. We begin with two simple questions what is an electric motor and what is motor control? These questions form the foundation for this chapter.

There are many different types of electric motors which a designer may choose from. However, the discussion will be concentrated on three types: permanent magnet DC motors, stepper motors, and brushless DC motors. Each is examined for their basic construction, operation, and control circuits.

Before we begin with an introduction to electric motors, it is important to review the basic concepts of magnetism and electromagnets.

Our introduction to motor control concludes with a brief summary of the topics discussed.

14.1 Introduction to Electromagnets and Electric Motors

14.1.1 Electromagnets

A permanent magnet is a piece of iron or steel which produces a magnetic field. The magnetic field causes permanent magnets to attract iron and some other materials (called ferromagnetic). The two ends of the permanent magnet are called poles usually designated as North and South, where the opposite magnet ends attract and like magnet ends repel. Permanent magnets are found in nature for example as magnetite (Fe_3O_4) or lodestones.

Electromagnets behave like permanent magnets, but their magnetic field is not always present. Rather, an electromagnet's magnetic field is temporarily induced by an electric current. To construct a simple electromagnet, we start with a wire wrapped around an iron bar.

How Do You Make an Electromagnet?

- Start with an iron bar
- Wrap a wire around the iron bar
- Connecting a battery causes a current to flow in the wire
- The current induces a magnetic field creating an electromagnet

NORTH
SOUTH
+
-
Current

How Do You Make an Electromagnet?

- Reversing the current direction, reverses the polarity

SOUTH
NORTH
-
+
Current

Figure 14.1 Making of an Electromagnet

When a battery is connected to the wire, current flows. The current flowing in a circular pattern about the iron bar induces a temporary magnetic field. The ends of the iron bar electromagnet are labeled as North and South (Figure 14.1).

If the battery connection is reversed, the current will flow in the opposite direction about the iron bar. This causes the end polarities of the electromagnet to also be reversed (Figure 14.1). When the battery is disconnected, the current around the iron bar stops. With no current, theoretically the induced magnetic field of the electromagnet ceases (Figure 14.2).

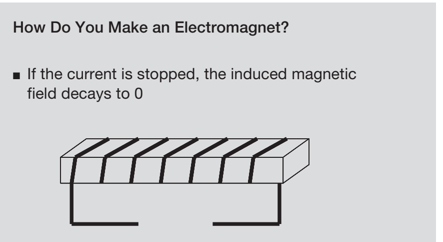


Figure 14.2 Electromagnet is NON MAGNETIC without Current

14.1.2 Electric Motor Basics

The term electric motor can be used in general to indicate any type of mechanism which converts electrical energy into mechanical energy (or motion). Certain electric motors may be used as electric generators that convert mechanical motion (or energy) into electrical energy.

When a voltage potential (i.e. a battery) is applied to an electric motor, current flows, and electrical energy is delivered to the motor. Electric motors convert this electrical energy into mechanical energy. In automotive applications, the mechanical energy is used to perform various tasks, such as opening a sunroof, moving a power seat, or positioning a mirror.

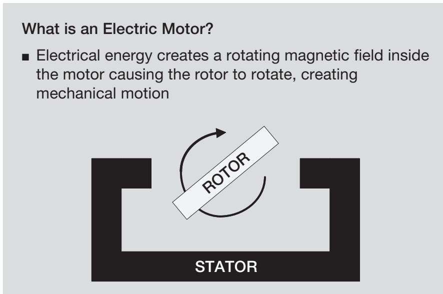


Figure 14.3 Basic Electric Motor

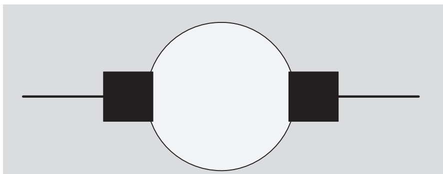


Figure 14.4 Symbol of a PMDC Motor

At the most basic level, the construction of electric motors can be taken down to two components: the stator and the rotor (Figure 14.3). The rotor is the part of the electric motor that rotates when electrical energy is applied. The stator is the part of the electric motor which remains stationary and magnetically interacts with the rotor.

When electrical energy is applied to the electric motor, a rotating magnetic field is created by an interaction of the stator and rotor. This rotating magnetic field causes the rotor to rotate, creating mechanical energy. The circuit symbol of a PMDC motor is shown in Figure 14.4.

14.1.3 Where are Electric Motors Used?

Electric motors are used in many different automotive applications. They are becoming increasingly important components in automotive designs, as they represent an opportunity to provide additional feature content and functionality while, in many cases, improving system performance and/or efficiency. They can be found in engine control systems, vehicle control systems, passenger comfort system, displays, etc. Figure 14.5 lists some of the common applications. Electric motors are also used for vehicle propulsion in hybrid and battery driven cars. While the basic operating principles are similar to the ones presented in this chapter, their control is a special discipline and is not included in this material.

Where Are Electric Motors Used?

- Electric motors are used in many different automotive applications:

Power windows
Power seats
Power mirrors
Fans
Windshield wipers
Windshield washer pumps
Starter motor
Electric radio antenna
Door locks
Information gauges

Sunroof
Brakes
Power steering
Fuel pump
Water pump
Hybrid and electric vehicles
Cruise control
Throttle plate control
Air vents
Others

Figure 14.5 Uses of Electric Motors

- The controlled application of electrical energy to a motor to elicit a desired mechanical response
  – Start / Stop
  – Speed
  – Torque
  – Position
- Significant amount of electronics may be required to control the operation of some electric motors

Figure 14.6 What is Motor Control?

14.2 Electric Motors Used in Automobiles

As one might expect, electrical energy cannot applied indiscriminately to electric motors. The application of electrical energy to some motors is in fact strikingly complex and could fill entire volumes! However, we can generalize and state that motor control is the controlled application of electrical energy to an electric motor to elicit a desired response. As it will be shown shortly, a significant amount of electronics (hardware) and signal processing (software) is often required to properly control some electric motors. Figure 14.6 lists the most common areas of control.

When we refer to motor control, we are actually talking about how we electrically control the torque exerted on the motor shaft and the rotational speed and direction of the shaft.

When a bias is applied to an electro-magnet, it is polarized with North and South poles, just like a permanent magnet. Unlike a permanent magnet, however, once the bias is removed from an electromagnet, it loses it magnetic properties. In addition, an electromagnet's polarity can be reversed (North becomes South and South becomes North) if the applied bias is reversed. Most of the mechanical design and development of a motor is concerned with the relative placement of the permanent and temporary magnets. This in turn dictates how the internal rotating magnetic field is generated, causing the rotor to rotate.

14.2.1 Common Motor Types

There are literally dozens of different types and classifications of motors. Some are listed in Figure 14.7.

- There are many different types and classifications of electric motors:

Permanent magnet DC motor
Stepper motor
Brushless DC motor
Wound field motor
Universal motors
Three phase induction motor
Three-phase AC synchronous motors
Two-phase AC Servo motors
Torque motors
Shaded-pole motor
Split-phase induction motor
Capacitor start motor
Permanent Split-Capacitor (PSC) motor
Repulsion-start induction-run (RS-IR) motor
Repulsion motor
Linear motor
Variable reluctance motor
Unipolar stepper motor
Bipolar stepper
Full step stepper motor
Half step stepper motor
Micro step stepper motor
Switched reluctance motor
Shaded-pole synchronous motor
Induction motor
Coreless DC motor
Others......

Figure 14.7 Types of Electric Motors

- Similar in construction to the introductory example
- Metallic contacts (brushes) are used to deliver electrical energy
- Rotational speed proportional to the applied voltage
- Torque proportional to the current flowing through the motor
- Advantages:
  + Low cost (high volume demand)
  + Simple operation
- Disadvantages:
  – Medium efficiency
  – Poor reliability (brush, commutator wear out)
  – Strong potential source of electromagnetic interference

Figure 14.8 Permanent Magnet DC Motor

However, many of these motor types are rather esoteric; others are only derivations of other more common types of motors. Therefore, in this chapter, we will examine only three of the most common types of electric motors used in automotive applications:

- Permanent magnet DC motors
- Stepper motors
- Brushless DC motors

14.2.2 Permanent Magnet DC Motor

Permanent magnet DC motors are very similar in construction to the example that was outlined in Section 14.1.2 showing a simple stator and rotor. Key features of these electric motors are the spring loaded graphite brushes used to apply the electrical energy and polarize the rotating electromagnet. Another important component is the segmented commutator ring.

Electrically, permanent magnet DC motors spin faster if a higher (DC) voltage is applied to the motor at a given load. They also have more torque for higher operating currents.

The greatest advantages of permanent magnet DC motors are their simple operation and low cost. However, their simplicity (the use of the brush contacts and commutator ring) results in several issues. Permanent magnet DC motors are only moderately efficient in converting electrical energy to mechanical energy. In addition, they suffer from a rather poor reliability record, and they are strong sources of electrical noise (electromagnetic interference, or EMI), which can potentially disrupt neighboring electrical systems (see Figure 14.8).

14.2.3 Stepper Motors

Stepper motors are another type of common electric motor in automotive applications. They differ from other common electric motors in that their construction and operation divides a 360° rotation into a discrete number of steps. For example, a stepper motor with 200 steps will move through 1.8° of rotation with each step.

Conceptually, the operation of stepper motors is relatively simple. A stepper motor controller can move the rotor one step by applying a single voltage pulse to the stator. By applying a different voltage pulse, the controller can move the rotor one step in the opposite direction. The rotational speed is controlled by changing the rate at which the voltage pulses are applied.

There are several advantages to stepper motors. Their discrete step operation makes them ideal for position control applications like traditional analog instrument gauges. In addition, once a controller has moved the rotor to the desired position (desired number of steps), the stepper motor can easily hold its position.

Their step-by-step operation, however, causes these motors to have a very low efficiency converting electrical energy into mechanical energy. In addition, the motors are relatively expensive and require a digital controller and interface for proper control. The basic stepper motor controller is relatively low cost because of the digital nature. The basic advantages and disadvantages of the stepper motors are listed in Figure14.9.

Stepper Motor

- Full rotation of electric motor divided into a number of “steps“
- For example, 200 steps provides a 1.8° step angle
- A stepper motor controller can move the electric motor one step (in either direction) by applying a voltage pulse
- Rotational speed is controlled by changing the frequency of the voltage pulses
- Advantages:
  + Low cost position control (instrument gauges)
  + Easy to hold position
- Disadvantages:
  – Poor efficiency
  – Requires digital control interface
  – High motor cost

Figure 14.9 Relative Merits of Stepper Motors

Brushless DC Motor

- Similar to a permanent magnet DC motor
- Rotor is always the permanent magnet (internal or external)
- Design eliminates the need for brushes by using a more complex drive circuit
- Advantages:
  + High efficiency
  + High reliability
  + Low EMI
  + Good speed control
- Disadvantages:
  – May be more expensive than “brushed“ DC motors
  – More complex and expensive drive circuit than “brushed“ DC motors

Figure 14.10 Relative Merits of Brushless DC Motors

14.2.4 Brushless DC (BLDC) Motors

BLDC motors can be thought of as an improvement upon permanent magnet DC motors. Their design and operation don’t require the spring-like brushes which lead to many of the disadvantages inherent in permanent magnet DC motors. Therefore, BLDC motors can be more efficient and more reliable than permanent magnet DC motors. In addition, BLDC motors generate significantly less electromagnetic interference than permanent magnet DC motors.

The elimination of the brush contacts, however, results in significantly more complex operation and control. In addition, brushless DC motors may be more expensive than comparable permanent magnet DC motors. Extra position sensing components may also be needed (usually Hall-effect sensors). The basic advantages and disadvantages are listed in Figure 14.10.

14.3 Interactions between Magnetic Fields

The magnetic field between two opposing magnetic poles (separated by a gap) is characterized by the flux density vector ***B***.

The positive direction of the magnetic field is from North pole to South pole. Applying the Left Hand Rule, the pointing finger follows the magnetic field into the page.

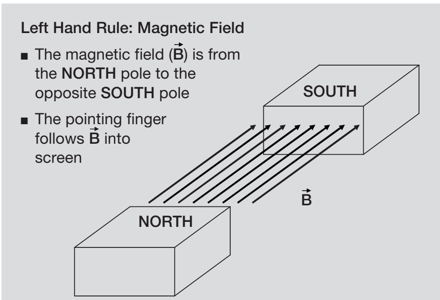


Figure 14.11 Magnetic Field Vectors between Magnetic Poles

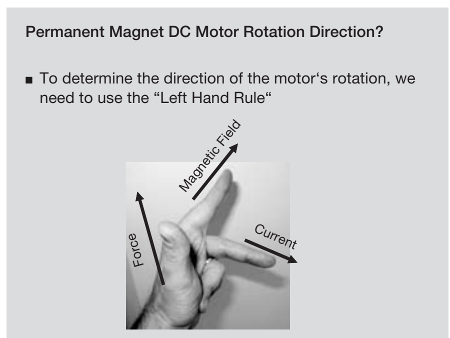


Figure 14.13 Left Hand Rule

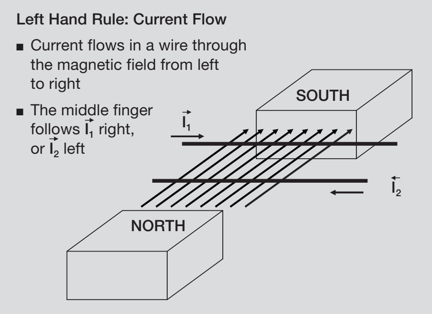


Figure 14.12 Current Carrying Wires in Magnetic Field

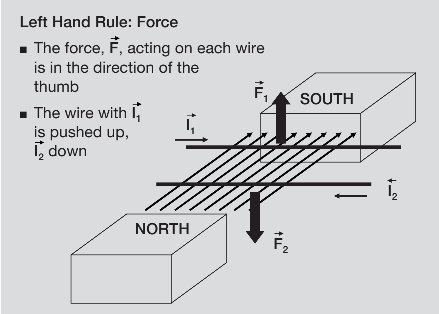


Figure 14.14 Forces Acting on Wires in a Magnetic Field

Assume that currents flow in the wires placed in the magnetic field perpendicular to the orientation of the ***B*** vector. Let the current I_1 flows from the left to the right. The direction of the current flow is followed by the left hand's middle finger. Let the index finger point into the page (B), then the middle finger points to the right edge of the page (I_1). If the current were to flow from right to left (I_2), the middle finger would point to the left side of the page.

Any wire carrying electrical current and surrounded by magnetic field, like in this case, will interact with the field between the magnetic poles. The direction of the force vectors is indicated by the left hand rule: the force acting on the wire conducting current I_1 is upward and the force acting on the other wire (I_2) is downward as shown in Figure 14.14 to Figure 14.16.

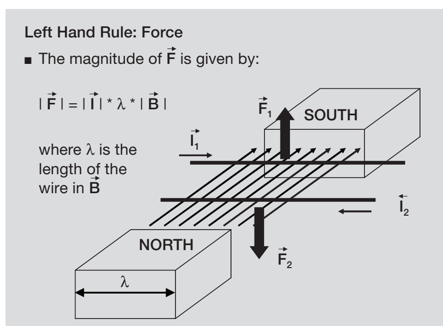


Figure 14.15 Left Hand Rule of Force

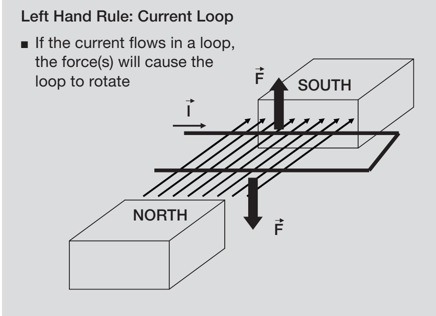


Figure 14.16 A Current Loop in Magnetic Field

The magnitude of the force is the product of the field strength ***B***, the current amplitude I_1 (I_2) and the length of wire in the field λ (see Figure 14.15). If the wire loop in Figure 14.16 is attached to an axle which is placed in the plane defined by the wire pair and parallel with the two wires, the forces would exert a rotating torque on the loop. Depending on the angular position of the loop the forces may attract the wires into the field or repel them.

In every motor type be it DC or AC the magnetic interactions must set up a rotating field pattern that will apply a rotating torque on the rotor.

14.4 Detailed Description of How the Three Basic Motor Types Operate

14.4.1 Permanent Magnet DC (PMDC) Motors Operation

The construction and operation of the PMDC is reviewed first through an example. PMDC motor uses permanent magnets mounted on a stator. While the numbers may vary, the simple example here shows a single North-South magnet on the stator. The stator is mounted “external” to the rotor (Figure 14.17).

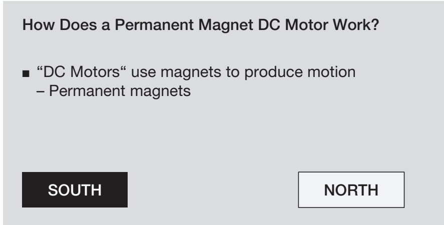


Figure 14.17 Components in a PMDC Motor

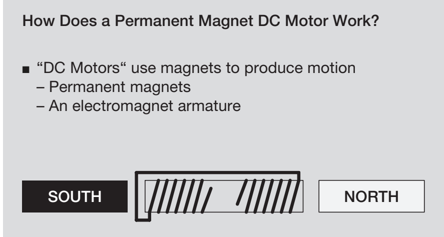


Figure 14.18 Components in a PMDC Motor (continued)

Next, the temporary magnet (electromagnet) rotor of our example PMDC motor is shown. The temporary magnet, with its wire wrappings, is known as an armature. It is mounted “internal” to the external stator. The armature is mounted so that it may spin about its center (Figure 14.18).

The electric motor’s commutator is the connection between the armature’s wire wrapping and the brushes of the permanent magnet DC motor (Figure 14.19).

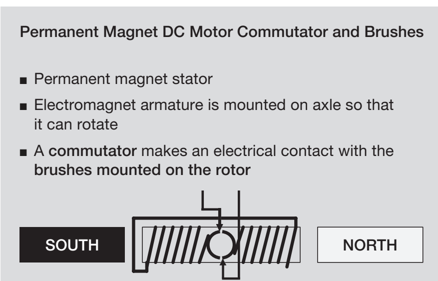


Figure 14.19 Components in a PMDC Motor (continued)

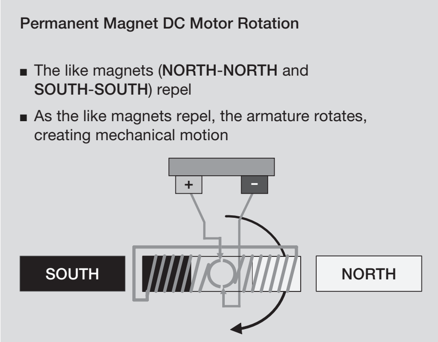


Figure 14.20 Completed PMDC Motor

The commutator of a single coil armature in a permanent magnet DC motor is comprised of two isolated halves of a ring (Figure 14.19). The two halves are mounted on the armature's axle allowing them to rotate with the armature. The permanent magnet DC motor's brushes electrically connect the armature's windings and commutator to the battery (Figure 14.20).

When permanent magnet DC motor is connected to a battery the current flows through the brushes to the commutator and armature's windings (Figure 14.20) and induces a magnetic field polarizing the rotor. The rotating armature just having crossed the gap of the commutator ring changes the polarization of the armature from the previous attraction to repulsion because like poles are now facing each other. The repulsing torque is diminishing as the armature turns but after 90° angular displacement attracting torque is developing because opposing polarities are approaching each other. When the displacement is 180° the split is traversed again and the polarization of the armature magnet is reversed again changing the attracting torque to repulsing one. The moment of inertia of the rotor carries it over the zero torque angular position. This process continues as long as power is supplied to the armature. If the polarity of the connections between the battery terminal and brushes is reversed the armature will rotate in the opposite direction.

14.4.2 Control of PMDC Motors

Uni-directional Control

There are applications where the motor is only required to drive in one direction (for example, fan or blower motors). In this case the drive can be full voltage, continuous DC or PWM controlled DC.

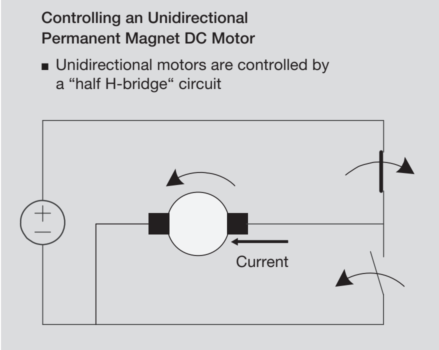


Figure 14.21 Motor is Driven Counter Clock Wise

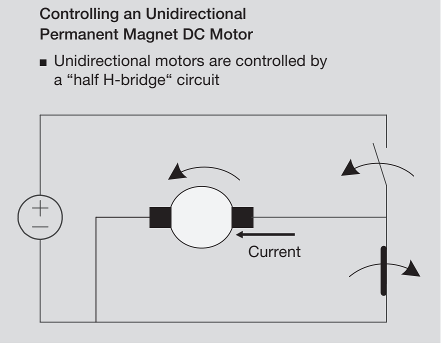


Figure 14.22 Motor is Braked or Stopped

The power control circuit topology is the same in both cases consisting of two switches realized by power transistors in most cases. When the top switch is closed and the lower one is open the motor is driven in the counter clock direction. If the top switch is closed for a sufficiently long time the motor will reach the final angular velocity determined by the motor parameters and the torque demand.

When the power is switched off the top switch opens and the lower switch closes shorting the motor terminals to ground. The motor will be slowed up and eventually stopped not only by the torque load but also by the back-emf generated in the rotating armature (see APPENDIX) which causes a braking torque proportional to the armature angular velocity.

In PWM mode of operation the top and bottom switches open and close in opposite phase to each other presenting a controlled and low impedance DC voltage source to the motor. If a complete switch cycle period is T, the switch is closed for $T_1 < T$ and the supply voltage is V_b the drive voltage across the motor terminal is equal to $V_M = (T_1 / T)\ V_b = D \cdot V_b$, where D is the duty cycle.

Shorting the motor when not driven will also eliminate or greatly reduce the inductive transients caused by the stored magnetic energy in the armature inductance. Transient voltage control is very important when solid state power switches are used to drive PMDC motors.

Bi-directional Control

The bi-directional motor control circuit consists of a DC voltage source and four electrical switches (usually transistors). The circuit is called an H-Bridge circuit because the electric switches and the motor form an “H” shape (Figure 14.23).

To cause the motor to spin in one direction, one switch connected to the positive terminal of the DC voltage source is closed while the other switch connected to the same terminal of the DC voltage source remains open. The states of the two switches connected to the negative terminal of the DC voltage source are the inverse of the ones in contact with the positive terminal. In this example, current can now flow from the DC voltage source, through the top left switch, through the PMDC motor (from the left to the right), through the bottom right switch, and back to the DC voltage source. This causes the motor to spin in the clock wise direction (Figure 14.24) in the illustration. Reversing the switch states to closing the top right and bottom left switches (and opening the others) will reverse the direction of rotation.

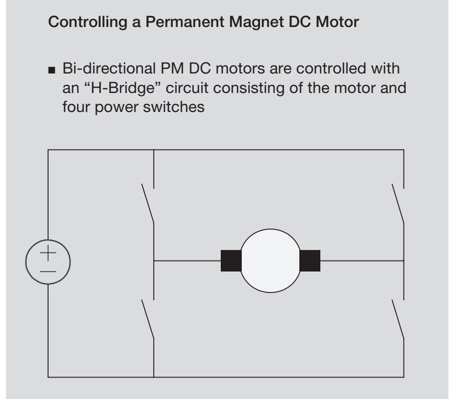


Figure 14.23 Basic H-bridge Circuit Topology

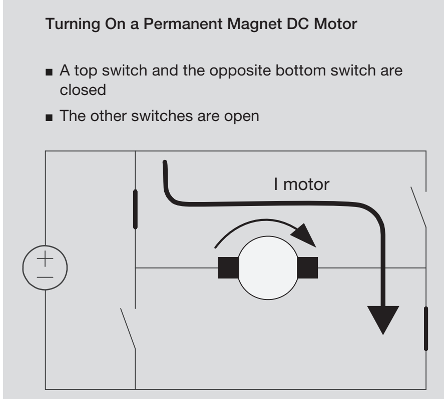


Figure 14.24 Driving the Motor Clock Wise

H-bridge motor controllers may act as DC drives or variable voltage drives using pulse width modulation (PWM). In the former mode the switch pairs are closed and kept closed as long as the motor has to run. In the latter mode

voltage is applied across the motor in a pulsed fashion. In practical motor drives the motor is not allowed to “free-wheel”: if the motor is not driven then its brush terminals are shorted through either by closing both of the top or both of the bottom switches as illustrated in Figure 14.25 and Figure 14.26.

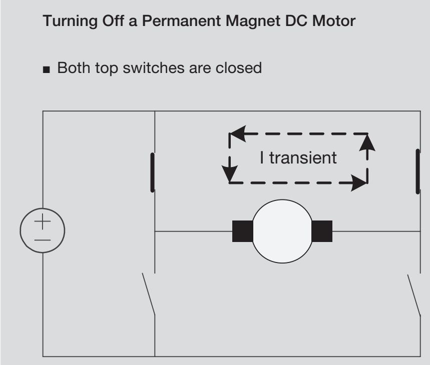


Figure 14.25 Motor Braking with High Side Recirculation

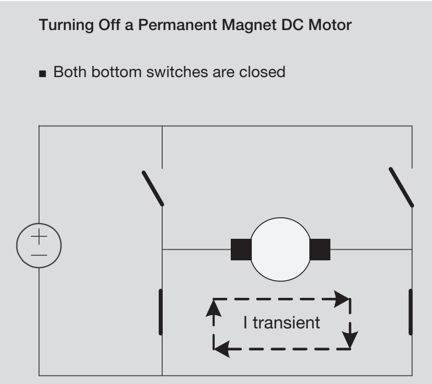


Figure 14.26 Motor Braking with Low Side Recirculation

The illustrated cases represent the braking mode preceded by the drive mode as in Figure 14.24. The dashed loops in both figures symbolize the transient decay of the current stored in the armature inductance. Keeping the armature shorted in the off-state is desirable for another reason: shorted armature develops a braking torque on the armature as soon as it moves. That may prevent the “creeping” of the rotor in response to road vibration.

PWM drive is illustrated in Figure 14.27 and Figure 14.28. In this case the motor current flows left to right and the motor voltage is controlled by pulse width modulation. Notice that in the left leg of the H-bridge, the top left switch is permanently closed and the bottom left switch is permanently open. Voltage control is accomplished by duty cycle control just like in the case of the uni-directional drive circuit.

If the role of the switches is reversed the motor is driven in the opposite direction in a controlled fashion. Care must be taken that the switches in the same leg are never closed together.

In addition, there are several other ways to open and close the four switches beyond those we have discussed. The most common of these is used for dynamic braking (or slowing) the motor. When it is desirable to stop the motor quickly, the polarity of the H-bridge drive may be reversed while the rotor is still running in the previously driven direction. The reverse drive can be applied at full voltage or with PWM control. In either way the torque is reversed on the moving armature and it stops much quicker than if the armature were only shorted or the drive all together were removed by opening all switches. Shorting the armature is a special case of dynamic braking. The various motor control modes are discussed in the APPENDIX.

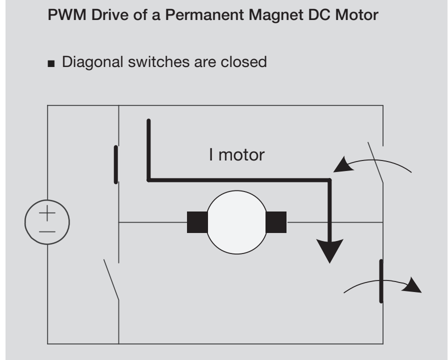


Figure 14.27 PWM Drive Cycle

PWM Recirculation in a Permanent Magnet DC Motor

- Both top switches are closed
- Both bottom switches are open

I motor

Figure 14.28 PWM Recirculation Cycle

Figure 14.27 shows the drive state during the PWM “on” cycle and Figure 14.28 shows the off-cycle. A summary of the different PMDC motor drive methods is in Figure 14.29.

- DC operation
  – Rotational speed of the DC motor is fixed at a given voltage and load
- PWM Operation
  – Average voltage (and rotational speed) can be controlled by opening/closing the switches quickly
- Braking
  – Shorting the terminals or momentarily reversing the drive
- Others

Figure 14.29 Controlling a PMDC Motor Options

- Unlike the permanent magnet DC motor, stepper motors move in discrete steps as commanded by the stepper motor controller
- Because of their discrete step operation, stepper motors can easily be rotated a finite fraction of a rotation
- Another key feature of stepper motors is their ability to hold their load steady once the require position is achieved
- An example application for stepper motors is for implementing traditional “analog“ instrumentation gauges on a dashboard

Figure 14.30 Why a Stepper Motor?

14.4.3 Stepper Motors

In this section stepper motors are discussed. Recall that stepper motors can be rotated in discrete steps and are often used in position control applications. Some of the key features and one potential application are listed in Figure 14.30.

One of the characteristic features of a stepper motor is the large number (~200) of “teeth” on both the rotor and stator. In the example we will present, the stepper motor has an internal rotor with six of permanent magnet poles and an external stator with four electromagnet poles. The electromagnet (or stator) teeth are polarized sequentially. The polarization may also be reversed. Each time the stator poles go through the process, the rotor takes one step.

In a full step stepper motor, angular displacement of a single step is 360°/N (N is the number of teeth).

A micro-stepper type motor can increase the number of discrete steps significantly, sometimes 50000 or more. This requires a large amount of drive signal processing, however, and quickly drives up the complexity of the motor control circuit and algorithms.

- A stepper motor often has an internal rotor with a large number of permanent magnet “teeth”
- A large number of electromagnet “teeth” are mounted on an external stator
- Electromagnets are polarized and depolarized sequentially, causing the rotor to spin one “step”
- Full step motors spin 360°/(# of teeth) in each step
- Half step motors spin 180°/(# of teeth) in each step
- Microstepmotors further decrease the rotation in each step

Figure 14.31 How Does a Stepper Motor Work?

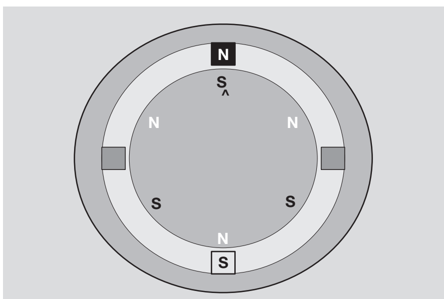


Figure 14.32 Full Step Operation

The series of figures (Figure 14.32 to Figure 14.34) illustrates the operation of the full step stepper motor. The permanent magnets on the internal rotor are marked “N” for North pole and “S” for South pole. The stator is also magnetized as shown. Six rotating poles are assumed in the illustrated case resulting in 30° positional change per step (12 steps per revolution). Each stator electromagnet can be polarized either as shown or in the opposite direction depending on the direction of the current flow in the coils.

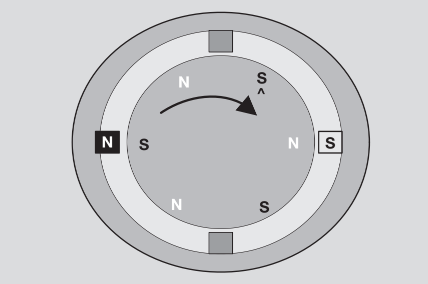


Figure 14.33 Full Step Operation

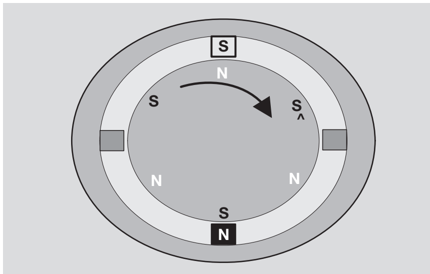


Figure 14.34 Full Step Operation

Now consider the operation of the full step stepper motor as it rotates and holds a position.

First the electromagnet stator is polarized (Figure 14.32). In this particular state of polarization the electromagnet stator causes the rotor to land in the indicated position. One step in the clockwise direction brings the rotor to a stable position and its position is held by the stator electromagnet (Figure 14.33). The step is controlled by de-energizing the stator poles previously energized and energizing the other poles as indicated. The next step energizes the stator in the “vertical direction” with polarity reversed with respect to the starting state (Figure 14.34). The movement of the rotor is indicated with the **caret** character. Notice that as the rotor is stepping clockwise the magnetization of the stator is “turning” counter clock wise.

If the direction of the rotor has to be reversed the polarity of the stator magnetization shown in Figure 14.33 and Figure 14.34 has to be reversed. As it has been stated previously the polarization of the stator is controlled by the current direction. The stator coils can be driven either by an H-bridge driver, or by making the coils center-tapped and using uni-polar low-side (or high-side) drivers. The angular velocity of the rotor is controlled by the pulse rate.

The control of a stepper motor is performed by a stepper motor controller. The controller sends square wave pulses to the stepper motor to advance the motor through the appropriate steps. There are stepper motors with much finer resolution than the illustrated case. The control circuit requirements are essentially the same as in case of the coarser resolution motors but the electromechanical parts are different. Usually both the stator and the rotor are “toothed”.

As previously mentioned, micro-stepping requires significantly more work. The controller must source a quasi sinusoidal current waveform to the stepper motor. This allows the rotor to move in finer steps than allowed even by the half stepping operation we reviewed.

Because the operation of the controller is performed by a digital circuit, stepper motors are also known as digital motors. The short description given in this section is very sketchy; there are a lot of good articles, books and tutorials on the internet for those who want to dig deeper into this topic.

14.4.4 Brushless DC Motors

Another common class of electric motors in automotive applications is the brushless DC motor (also known as BLDC motor). Many of the limitations of the classic permanent magnet “brushed” DC (PMDC) motor are

caused by the brushes pressing against the rotating commutator creating friction:

- as the motor speed is increased, brushes may not remain in contact with the rotating commutator;
- sparks and electric noise may be created as the brushes encounter flaws in the commutator surface or as the commutator breaks contact with the rotor segment that has just been energized;
- brushes eventually wear out and require replacement and the commutator ring itself is subject to wear and maintenance.

BLDC motors avoid these problems with a modified design but require a more complex control system. The name is a misnomer because from the electromagnetic view point they are more akin to synchronous AC motors than to the PMDC motors. In case the BLDC motors the permanent magnet forms the rotor and the electro-magnet is the stator. There is no commutator ring and the rotor position is determined by means of Hall-effect sensors excited by the poles of the rotating permanent magnet. The BLDC motor control reads the signals from the sensors and based upon these inputs the controller knows which electromagnets to turn on or off and when. This control method allows a controller to use an external clock signal to synchronize the switching sequence of the electromagnets and thus the rotation of the BLDC motor allowing precise speed control. Since the position of the rotor is sensed the motor can be driven with a continuous DC supply (given a commutating controller); in this case the rotor speed depends on the source voltage and the torque load on the shaft. The commutation, as mentioned above, is determined by the rotor position sensors.

There are two basic types of BLDC motors:

- BLDC motor with an external (ring) permanent magnet rotor and an internal electromagnet stator
- BLDC motor with an internal permanent magnet cylindrical rotor and an external electromagnet stator

In the following descriptions the motor configuration is internal permanent magnet and external electromagnet as illustrated in Figure 14.35 and Figures 14.36-14.39.

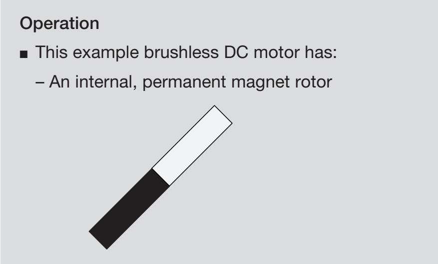


Figure 14.35 Example Brushless DC Motor

The first component is the permanent magnet pivoted at the center of the electromagnets. The electromagnets are arranged evenly around the periphery of the stator area (Figure 14.36).

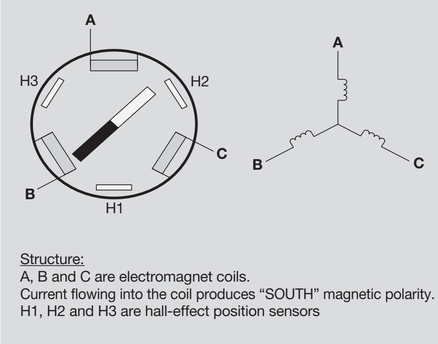


Figure 14.36

The simplest motor construction features only three coils 120 degrees apart and the coils are connected in a three-phase star pattern. The external solid ring corresponds to the common points of the windings ends (star point). At any given time current is flowing through two coils. In addition to the electro-magnets and rotating permanent magnet rotor position sensors are also needed which are typically Hall-effect magnetic flux sensors (Figure 14.36).

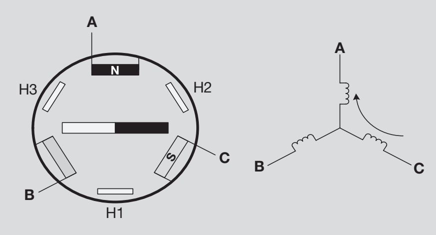


Figure 14.37

Assume that the current drive and the rotor position is as in Figure 14.37. By arbitrary choice current flowing into a coil makes the electromagnet polarity SOUTH (south-pole on the magnets are white, north-poles are black) and un-excited electromagnets are filled with cross hatch.

In Figure 14.37 the north pole of the rotating magnet is pulled by the SOUTH electromagnet end its south-pole by the more distant NORTH electro-magnet.

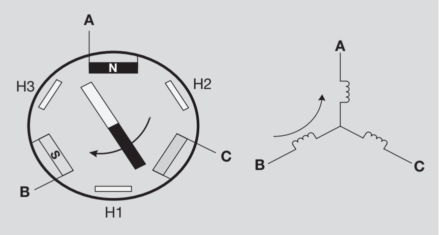


Figure 14.38

Just before the "C" electro-magnet (SOUTH) would grab the north pole of the rotating magnet the field of the south pole of the rotating magnet is picked up by sensor H3 and the current from coil C is transferred to coil B (Figure 14.38).

The rotor is now being pulled by electro-magnets B (SOUTH) and A (NORTH). Just before electromagnet B (SOUTH) would grab the north-pole of the rotor sensor H2 senses the flux of the south pole of the rotating magnet and will switch the current from electro-magnet "A" to "C" (Figure 14.39).

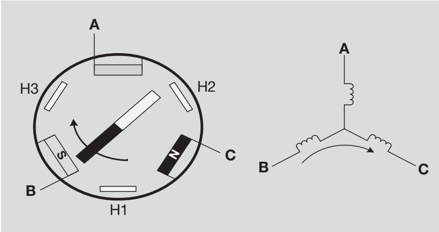


Figure 14.39

In a similar fashion the rotation of the permanent magnet and the excitation sequence of the electromagnets can be traced out.

There are a few things need to be mentioned at this point:

- the rotor normally features more than one pole pair;
- the electro-magnets consists of inter-leaved coils in "ABCABC…" sequence;
- given a proper commutation control BLDC motors are similar to the brushed DC motors described earlier. Increasing voltage increases the speed or torque, and back emf is also present and acts similarly. Speed, torque and back-emf are detailed in the APPENDIX (in terms of permanent magnet DC motors).

An example of the control circuitry for a BLDC motor is shown in Figure 14.40. If the drive current, for example, is from A-to-B switches HS1 and LS2 are closed, similarly from B-to-C switches HS2 and LS3 are closed and C-to-A switches HS3 and LS1 are closed.

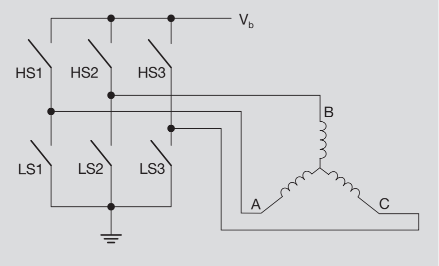


Figure 14.40

The BLDC motor controller is realized with solid state switches; in low power (<1 kW), low voltage (24 V or less) automotive applications N-channel MOSFET half bridge circuits are commonly used. A schematic circuit is in Figure 14.41. In higher voltage, higher current applications (electric propulsion control, for example) isolated gate bipolar transistors or IGBT-s are commonly used.

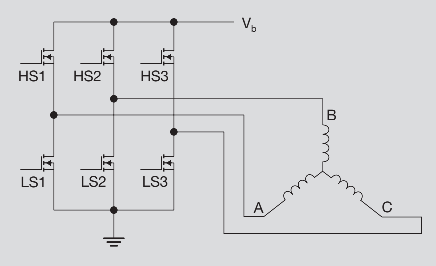


Figure 14.41

The drive control has two different modes:

- torque-speed control by means of supply voltage change;
- control by using pulse width modulation.

In the first case the HSx and LSy switches are operated as controlled by the rotor position sensing Hall-effect switches. The torque delivered by the motor is linearly dependent of the supply voltage (at least in the first approximation).

In the second case the high side driver corresponding to the rotor position is solidly ON and the low side driver is pulse width modulated. Suppose that in the given rotational segment switches HS1 and LS2 are active. Current flows from terminal A to terminal B and HS1 is ON while the rotor is in the appropriate segment but LS2 is pulse width modulated. When LS2 is opened the current stored in windings A-B will re-circulate at first through the parasitic diode of MOSFET HS2 and a few micro seconds later HS2 is closed. Thus the PWM is realized by alternating switches LS2 and HS2. PWM is similar when other coil pairs are driven or when the polarity of the drive is reversed; in our example current would flow from B to A, switch HS2 would be solidly on and switches LS1 and HS1 would be alternatively turned on and off.

14.4.5 Conclusion

While there are many different kinds of motors, most automotive applications use a few different types:

1) Permanent magnet DC motors (PMDC motors)
2) Stepper motors
3) Brushless DC motors (BLDC motors)

The advantages and disadvantages of each of these three motors are shown in the next table:

Types of Electric Motors

	Permanent Magnet DC Motor	Stepper Motor	Brushless DC Motor
Advantages:	+ Low cost (high volume) + Simple operation	+ Position control (low cost control circuits)	+ High efficiency + High reliability + Low EMI + Speed control
Disadvantages:	- Medium efficiency - Poor reliability - Bad EMI	- Poor efficiency - Digital interface - High motor cost	- Maybe higher cost (lower volume) - Complex control

A. Permanent Magnet DC Motors Models and Application

Permanent **m**agnet **DC m**otors (PMDCM) are generally used to drive various types of automotive loads (note they can also be used as generators). Applications, models, and design methods are presented in this chapter. The basic structure and operation of the PMDC is presented in Chapter 14 and relevant terms are in the Glossary which is not repeated in this chapter.

In discussing any electro-mechanical system, like a motor and its load, it is important to use correct and pure units so as to eliminate confusion.

In response to the input power (voltage source, armature current) the motor develops torque on its shaft that is proportional to the current flowing through the motor armature. The units of torque, moment of inertia, angular velocity and angular acceleration are first reviewed to help the reader in the interpretation and application of parameters as published in datasheets. Unfortunately SI units (international system of units form the French - -Système International d'Unités) are often mixed with traditional engineering units that unnecessarily complicate the analysis and simulation of motor drive systems.

For clarity, we will stay with the SI units so that **power** and **work** are respectively given in **watt** and **joule** whether resulting from mechanical or electrical parameters and computations. Another great advantage is that basic motor parameters, like the **k** constant and the mechanical time constant of the loaded and unloaded motor will have simpler forms.

The general block diagram of a motor driving a load is shown in Figure A.1:

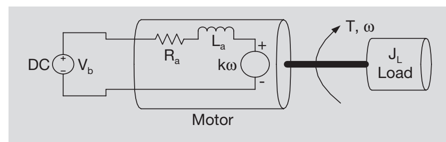


Figure A.1 Block Diagram of PMDCM Driving an Inertial Load

Meanings of symbols in Figure A.1 are: R_a armature resistance, L_a armature inductance, k motor velocity or torque constant, kω back emf, V_b supply voltage, T torque driving the load and ω angular velocity.

A.1 Motor Loads

In the following discussion armature, shaft and rotor are used interchangeably. The load torque may take many different forms as listed below:

A.1.1 Coulomb-Friction Loads

In the first approximation this type of load is independent of the angular position, angular velocity or angular acceleration of the armature and it is called **Coulomb friction**, T_C. Coulomb friction load is always present because of the brushes to commutator contacts and bearing frictions.

A.1.2 Loads Proportional to Angular Velocity

a) Viscous Friction Load: this load is linearly proportional to angular velocity.

$T_v = k_v * \omega$

b) Wind resistance type load is very common when the motor drives a fan. The torque load is proportional to the square of the angular velocity.

$T_\omega = k_\omega * \omega^2$

A.1.3 Loads Proportional to Angular Acceleration

The motor accelerates an inertial load with “J” moment of inertia constant. The torque is directly proportional to the rate of change of angular velocity:

$T_J = J(d\omega/dt)$

(Newton’s second axiom applied to rotary motion). J includes the moment of inertia of the rotor and the load:

$J = J_a + J_L$

A.1.4 Driving Torque

This torque is the motor output and it is directly proportional to the armature current:

$$T_m = k * I_a \quad (1)$$

where: k is the motor constant and I_a is the armature current.

The output torque of the motor, therefore, may have to drive all the listed torques as shown in Figure A.2 and by Equation (2):

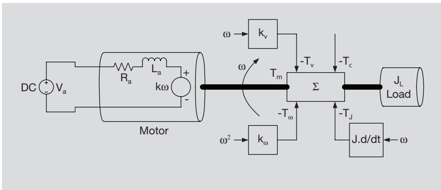


Figure A.2 Block Diagram of PMDCM Driving Different Torque Loads

$$T_m = T_C + T_v + T_\omega + T_J \quad (2)$$

$$J = J_L + J_a \quad (2a)$$

Combining Equation (1) and Equation (2) gives:

$$k * I_a = T_C + k_v * \omega + k_\omega * \omega^2 + J(d\omega/dt) \quad (3)$$

A.2 Torque Related Units, SI and Traditional Units

The definition of the torque is: a force, F perpendicularly acting on an arm of length L. If the force vector is not perpendicular on L the perpendicular component is effective.

A.2.1 Basic Mechanical and Electrical Units

Unit of mass: kg
[M] = kg

Unit of acceleration:
[a] = m/sec^2, where m is meter, sec is second

Unit of force:
[F] = [M].[a] = kg.m/sec^2
[F] = Newton = N (4)

Unit of torque:
[T] = [F].[L] = kg.m/sec^2.m = kg(m/sec)^2
[T] = Newton.meter = N.m (5)

Unit of angle: radian
[ϕ] = rad, note, that rad is dimensionless

Unit of angular velocity:
$[d\phi/dt] = [\omega] = rad/sec$

Unit of angular acceleration:
$[d\omega/dt] = rad/sec^2$

Unit of current:
[I] = A (ampere)

Unit of voltage:
[V] = V (volt)

Unit of resistance:
$[V/I] = \Omega$ (ohm)

Unit of inductance:
$[L] = [V/(dI/dt)] = \Omega.sec$ (ohm.sec)

A.2.2 Torque Constants

Torque constant of motor, k:
[k] = [T] / [I] = N.m/A (6)

If "k" is expressed in SI units the same numerical value gives the torque constant and the back emf constant (related to the voltage induced in the rotating armature by the stator field):

Back emf constant:

[k] = V/(rad/sec) (6a)

Velocity torque constant, k_v:
$[k_v] = [T] / [\omega] = N.m/(rad/sec)$,
$\mathbf{[k_v] = [T] / [\omega] = N.m.sec}$ (7)

$(Velocity)^2$ torque constant, k_ω:
$[k_\omega] = [T] / [d\omega/dt]^2 = N.m/(rad^2/sec^2)$,
$\mathbf{[k_\omega] = N.m.sec^2}$ (8)

Moment of inertia, J:
$[J] = N.m/(rad/sec^2)$
$\mathbf{[J] = N.m.sec^2 = kg.m^2}$ (9)

A.2.3 Traditional Units:

angular displacement: 1 rev = 2π rad,
angular velocity: 1 rev/minute = **1 rpm**
1 rpm = 2π rad/60 sec =
= 0.1047198 rad/sec

weight, force, length, torque:
1 kg force = 1 kg.g = 1_{kgf} = 9.806 N

"g" is the gravitational acceleration
($g = 9.806\ m/sec^2$)

1 pound = 0.45359 kg

1 pound force =1_{lbf} = (0.45359 * 9.806) = 4.448 N

1_{lbf}.foot = (0.45359237 * 9.806) x 0.3048 =
= **1.355728 N.m**

1_{ozf}.in = 1.355728/(16 x 12) = 0.007061 N.m

In modern European and Japanese catalogs one often finds torques expressed in N x m (SI) units while in American/English catalogs oz(force) x in or lb(force) x foot units appear frequently.

A.3 Motor Characteristic

The PMDCM has two kinds of parameters: electrical and electro-mechanical.

A.3.1 Electrical Parameters

Armature resistance: $\mathbf{[R_a] = \Omega}$ (10)

Armature inductance: $\mathbf{[L_a] = \Omega.sec}$ (11)

Note: R_a and L_a are dependent on the angular position of the rotor (armature). The number of commutator segments in contact with the brushes alternates at least between two discrete limits as the armature turns, because the number of armature winding segments connected in parallel by the brushes also changes. Either the average of R_a and L_a values in one revolution are given in parametric tables or the maximum and minimum values.

Electrical Time constant,

$[\tau_e] = L_a / R_a = \Omega.sec/\Omega = sec$ (12)

The armature electrical time constant is an important parameter for designing pulse-width modulated (or PWM) drives. Notice that usually $\tau_e << \tau_a$.

A.3.2 Electro-Mechanical Parameters

a) Motor Constant **k**

This constant has been characterized in Section A.2. Expressed in SI units the torque constant numerically is equal to the back emf constant. This fact is not surprising, for the motor torque is the result of the interaction between the electromagnetic force of the armature (determined by the winding structure and the armature current) and the field of the stator permanent magnet. The back emf voltage is induced in the same armature winding by the same stator magnetic field. The constant **k** is dimensionally the same in both representation as demonstrated below:

[N x m] = [Energy] = [Power x sec] = **[V x A x s]**

[k] = [N x m/A] =
V x A x sec/A = V x sec → **V/(rad/sec)** (13)

In American/English catalogs often two different k values are indicated when traditional engineering units are used for both the “torque constant”, k_T [oz(f) x in/A] or [lbs(f) x ft/A] and for the “back emf constant”, k_ω [V/rpm] (or [V/1000 rpm]). Notice that in this representation the two constants have different values and dimensions.

b) Motor Mechanical Time Constant, τ_a

Assume that the motor is driven from an ideal voltage source, V_b and the only torque load on the shaft is the moment of inertia of the armature, J_a (the Coulomb friction is assumed to be zero). In the first approximation the electrical time constant is neglected ($L_a di_a/dt$ term is not considered). In this case the model is:

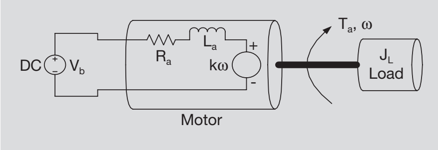


Figure A.3 Model for Deriving the Motor Time Constant

$$T = k * i_a \tag{14}$$

$$i_a = \frac{V_b - k * \omega}{R_a} \tag{14a}$$

$$J_a * \frac{d\omega}{dt} = k * i_a = k * \frac{V_b - k * \omega}{R_a} \tag{14b}$$

$$J_a * \frac{d\omega}{dt} + \frac{k^2 * \omega}{R_a} = k * \frac{V_b}{R_a} \tag{14c}$$

$$\frac{d\omega}{dt} + \frac{k^2}{J_a * R_a} * \omega = k * \frac{V_b}{J_a * R_a} \tag{14d}$$

The solution of equation (14d) with DC voltage source V_b and $\omega(t = 0) = 0$ initial condition is equal to:

$$\omega(t) = \frac{V_b}{k} * \left(1 - \exp\left(-\frac{t * k^2}{J_a * R_a}\right)\right) \tag{14e}$$

J_a in Equation (14a) - Equation (14e) denote the moment of inertia of the armature.

The motor (electro-mechanical) time constant is defined as:

$$\tau_a = \frac{J_a * R_a}{k^2} \tag{15}$$

More comprehensive motor catalogs also contain the motor time constant with the help of which the moment of inertia of the armature may be calculated. Using SI units it is easy to show that the dimension of the right hand side of Equation (15) is, indeed, time.

If J_a is not a catalog entry, its measurement is not a trivial task. First measure the average R_a with a digital ohm meter. Measurement of the k constant is very simple: the motor is disconnected from the voltage source V_b, the rotor is driven to a well defined rpm (rad/sec) and the motor voltage (V_a) is measured across the brushes without any appreciable electric loading. Next, the motor shaft is accelerated with different supply voltages and the acceleration results will yield τ_a. Now J_a can be calculated by using Equation (15). The knowledge of J_a is important in drive systems with low inertial loading or in feedback control systems.

c) Torque vs. Speed Curves

The armature current at 0 rpm (stalled rotor) is given by

$$i = \frac{V}{R_a} \qquad (16)$$

at $\omega = 0$ and the stall torque will be equal to

$$T_{stall} = k * i_a = k * \frac{V}{R_a} \qquad (17)$$

when $\omega \neq 0$,

$$i_a = \frac{V - k * \omega}{R_a} \qquad (18)$$

Substituting Equation (18) into Equation (17) will yield the torque expressed as a function of ω at a given terminal voltage, V:

$$T(\omega) = k * \frac{V - k * \omega}{R_a} = \frac{k}{R_a} * (V - k * \omega) \qquad (19)$$

A general plot of Equation (19) is shown in Figure A.4:

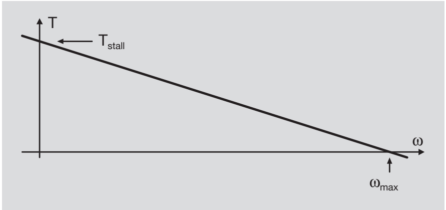


Figure A.4 Torque vs. Speed Function of a PMDCM

The literal interpretation of the torque vs. speed function as given by Equation (19) or its plot (Figure A.4) is as follows:

at zero speed the current demand is V / R_a and the (stall) torque is kx(V / R_a);

at maximum unloaded speed both the torque and the current are zero (assuming no friction losses) and the armature speed is equal to $\omega_{max} = V / k$.

The last sentence is not quite true since the static torque load caused by the brush-to-commutator friction is always present. In general there may also be speed related torque loads caused by the wind resistance acting on the rotating armature. Due to those torque loads the armature current is also greater than zero at the unloaded shaft speed. This operating condition is better described as: at ω_{max} the torque *available* to drive any external load is zero.

The static friction can be determined by measuring the current of the unloaded motor at the prevailing drive voltage.

Given any torque loading T such, that, $0 \leq T \leq T_{stall}$ the corresponding shaft speed is found as shown in Figure A.5. Similarly, given a shaftspeedω_{load}and0 $\leq \omega_{load} \leq \omega_{max}$theavailable torque can also be taken from the speed vs. torque curves.

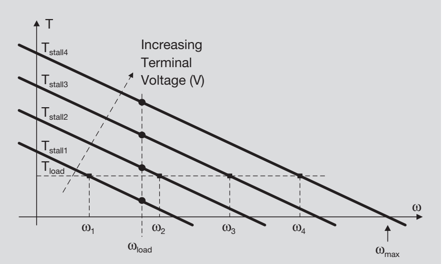


Figure A.5 Torque vs. Speed Functions of a PMDCM at Different Terminal Voltages

A.4 Dynamic Braking

Dynamic braking is accomplished by changing the drive voltage from an initial value to a lower value. In case of a polarity reversal the braking is more aggressive. If the reversed voltage is still present at the terminals when the shaft speed has reached 0 rad/s, the motor shaft begins to rotate in the opposite direction.

Consider the differential equation Equation (14d) i.e. the case of unloaded motor, and seek the solution for the new shaft velocity if the supply voltage is changed as a step function from V to V_{new} ($V_{new} < V$) when the initial shaft speed $\omega_0 = V / k$. Assuming ideal voltage sources and step function switching from V to V_{new}:

$$J_a * \frac{d\omega}{dt} + \frac{k^2}{R_a} * \omega = \frac{k}{R_a} V_{new} ,$$

subject to initial condition

$$\omega(t = 0) = \omega_0 = \frac{V}{k} \qquad (20)$$

The solution of Equation (20) is

$$\omega(t) = \frac{V}{k} * \exp\left(-\frac{t}{\tau_a}\right) + \frac{V_{new}}{k} * \left(1 - \exp\left(-\frac{t}{\tau_a}\right)\right) \quad (21)$$

where $\tau_a = \frac{R_a * J_a}{k^2}$ as in Equation (15).

Three different conditions can be distinguished:

a) $0 < V_{new} < V$

In this case the initial shaft speed is:

$\omega_0 = \frac{V}{k}$ and its final value is:

$$\lim\omega(t \to \infty) = \frac{V_{new}}{k}$$

b) If $V_{new} = 0$, and $\omega_0 = \frac{V}{k}$ the final value is:

$$\lim\omega(t \to \infty) = 0$$

c) Let $V_{new} < 0$, $V > 0$ and $\omega_0 = \frac{V}{k}$,

at $t = t_{rev}$ $\omega(t = t_{rev}) = 0$ and the final speed is:

$$\lim\omega(t \to \infty) = -\frac{V_{new}}{k}.$$

$t = t_{rev}$ (time to reversal) can be calculated from Equation (15) as follows:

$$\omega(t_{rev}) =$$

$$= \frac{V}{k} * \exp\left(-\frac{t_{rev}}{\tau_a}\right) - \frac{V_{new}}{k} * \left(1 - \exp\left(-\frac{t_{rev}}{\tau_a}\right)\right) = 0$$

$$\text{or } \left(\frac{V}{k} + \frac{V_{new}}{k}\right) * \exp\left(-\frac{t_{rev}}{\tau_a}\right) = \frac{V_{new}}{k}, \quad (22)$$

$$t_{rev} = \tau_a * LN\left(1 + \frac{V}{V_{new}}\right) \quad (23)$$

In Equation (22) and Equation (23) V_{new} is the magnitude of the reverse voltage.

WARNING: when the voltage is reversed to decrease the motor speed or to reverse the motor shaft rotation the initial current is GREATER than the stall current because the back emf voltage is ADDED to the reversed terminal voltage as described in Section A.4 and illustrated by Figure A.3.

A.5 Transient Current Demands

When power MOSFET devices are used for driving **PMDCM**s the current demand under different operating modes is one of the main design parameters in determining the stresses on the power control devices. A good practice to analyze the system is to build a computer model and confirm the limit values of control parameters such as:

- maximum voltage experienced by the drivers
- maximum current the drivers must supply
- maximum dissipation (power and energy)
- response time between extreme drive function limits

The dynamics of the armature current are determined by the circuit parameters of the armature (resistance, inductance), the load (moment of inertia, other torque load) and the motor constants. If the only torque loads are the moment of inertia and viscous damping, the system dynamics can be represented by two first-order linear differential equations:

$$J * \frac{d\omega}{dt} + K_v\omega = ki_a, \text{ and } \omega(t = 0) = \omega_0 \quad (24)$$

$$L_a * \frac{di_a}{dt} + R_a i_a = \nu_b(t) - k\omega, \text{ and } i_a(t = 0) = i_0 \quad (25)$$

The simultaneous differential equations will be Laplace transformed and solutions for the Laplace-transform of $\omega(t)$ and $i_a(t)$ will be separated. Assuming that **$\nu_b(t)$ is a step function of V_b magnitude**, and the initial conditions are as shown in Equation (24) and Equation (25), the Laplace transforms are equal to:

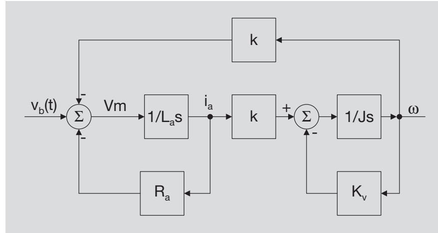


Figure A.6 Functional Block Diagram of PMDC Motor Driving a Viscous Load

$$\Omega(s) = \frac{\omega_0 s^2 + s\left(\omega_0 \frac{R_a}{L_a} + k\frac{i_0}{J}\right) + \frac{kV_b}{JL_a}}{s*\left[\left(s+\frac{1}{\tau_e}\right)*\left(s+\frac{1}{\tau_{Load}}\right)+\frac{k^2}{JL_a}\right]} \qquad (25a)$$

$$I_a(s) = \frac{s^2 i_0 + s\left(\frac{V_b}{L_a} + \frac{K_v i_0}{J} - \frac{k\omega_0}{L_a}\right) + \frac{K_v V_b}{L_a}}{s*\left[\left(s+\frac{1}{\tau_e}\right)*\left(s+\frac{1}{\tau_{Load}}\right)+\frac{k^2}{JL_a}\right]} \qquad (25b)$$

and $\tau_e = \frac{L_a}{R}$, $\tau_{Load} = \frac{JR_a}{(K_v R_a + k^2)}$. (26)

The characteristic polynomial of the system is equal to:

$$s^2 + s\left(\frac{1}{\tau_e} + \frac{1}{\tau_{Load}}\right) + \frac{1}{\tau_e * \tau_{Load}} + \frac{k^2}{JL_a} = 0 \qquad (27)$$

The solution is easier if the numerical parameter values are known. Solution in general terms is more involved especially if the roots of the characteristic polynomial are complex numbers. The initial and final values, however can easily be obtained from the Laplace-transforms by applying the initial value and final value theorems.

$$\omega(t=0) = \lim_{s \to \infty}(s\Omega(s)) = \omega_0$$

and

$$i_a(t=0) = \lim_{s \to \infty}(sI_a(s)) = i_0$$

The final values are equal to:

$$\omega(t=\infty) = \lim_{s \to 0}(s\Omega(s)) = kV_b / (k^2 + R_a K_v)$$

and

$$i_a(t=\infty) = \lim_{s \to 0}(sI_a(s)) = K_v V_b / (k^2 + R_a K_v)$$

In most cases the electrical time constant is orders of magnitude smaller than the mechanical one resulting in substantial simplification of the differential equation (and the functional block diagram). Assuming the same load conditions as in the previous case the system differential equations with $\tau_e << \tau_{Load}$ become:

$$J\frac{d\omega}{dt} = \frac{V_b - k\omega}{R_a} * k - K_V\omega \text{, or}$$

$$J\frac{d\omega}{dt} + \left(K_V + \frac{k^2}{R_a}\right)\cdot\omega = k\frac{V_b}{R_a}, \qquad (28)$$

subject to $\omega(t = 0) = \omega_0$ and

$$i_a(t) = \frac{V_b - k\omega(t)}{R_a}, \qquad (29)$$

where V_b is assumed to be a unit step function.

The system block diagram corresponding to Equation (28) - Equation (29) is in Figure A.7.

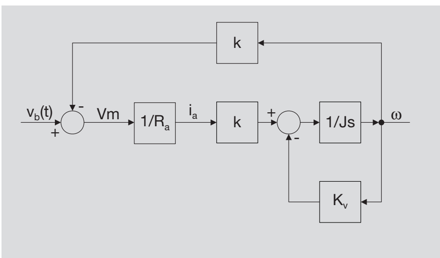


Figure A.7 Simplified Motor Control Block Diagram

The system time constant is equal to:

$$\tau = \frac{JR_a}{K_V R_a + k^2} \qquad (30)$$

and the solution for $\omega(t)$ is worked out similarly to the previous case.

$$\omega(t) = \omega_0 * \exp\left(-\frac{t}{\tau}\right) + \frac{V_b k}{K_V R_a + k^2}\left(1 - \exp\left(-\frac{t}{\tau}\right)\right), \qquad (31)$$

where ω_0 is the initial shaft velocity.

The motor current is given by

$$i_a(t) = \frac{V_b - k\omega(t)}{R_a} =$$

$$= \frac{V_b}{R_a} - \frac{k}{R_a}\left(\omega_0 \exp\left(-\frac{t}{\tau}\right) + \frac{V_b k}{K_V R_a + k^2}\left(1 - \exp\left(-\frac{t}{\tau}\right)\right)\right) \quad (32)$$

The following cases are considered:

- $t = 0$, $\omega(0) = \omega_0$, $v(t) = \text{step}(V_b)$; **then**

$$i_a(0) = \frac{V_b - k\omega_0}{R_a}$$

- $t \to \infty$, $\omega(0) = \omega_0$, $v(t) = \text{step}(V_b)$; **then**

$$\omega(t \to \infty) = \frac{V_b k}{R_a K_V + k^2}, \text{ and}$$

$$i_a(t \to \infty) = \frac{V_b K_V}{R_a K_V + k^2}$$

- $t = 0$, $\omega(0) = \omega_0$, $v(t) = \text{step}(-V_b)$; **then**

$$i_a(0) = \frac{(-V_b) - k\omega_0}{R_a} \quad (33)$$

- $t \to \infty$, $\omega(0) = \omega_0$, $v(t) = \text{step}(-V_b)$; **then**

$$\omega(t \to \infty) = \frac{-V_b k}{R_a K_V + k^2}, \text{ and}$$

$$i_a(t \to \infty) = -V_b \frac{K_V}{R_a K_V + k^2} \quad (34)$$

Equation (31) - Equation (34) provide analytical information of the motor current and motor speed values (both initial and final values) when the electrical time constant is much smaller than the mechanical one and the main load is viscous torque type. An important equation is Equation (33): when the rotor is spinning in one direction and a step voltage of opposite polarity is applied, the motor current will have a large transient with initial amplitude directly proportional to the sum of the back emf caused by the spinning armature and the reversing voltage. This fact MUST always be kept in mind when motor drive systems are designed. Whenever applicable, the simplified model provides a more pessimistic estimation of the maximum current than the detailed model. The inrush current obtained from Equation (32) shows a step response of the current while the evaluation of Equation (25b) defines an exponential current rise/fall to the peak value (and not a step change as in Figure A.8.

The graphs of the speed and current functions that have been calculated from the simplified model are shown in Figure A.9.

EXAMPLE 1 (Figure A.8): $V_b = 8$ V, $R_a = 2\ \Omega$, $K_v = 1.2 \times 10^{-4}$ Nmsec, $k = 267.38 \times 10^{-4}$ Nm/A (or 267.38×10^{-4} Vsec), $J = 10^{-4}$ Nmsec^2, $\omega_0 = 0$ rad/s and $\tau = 0.2094$ sec.

Using Equation (32) - Equation (33) the final current value is: $I_a(t \to \infty)$ = **1.005** A and the final shaft speed value is:
$\omega(t \to \infty)$ = **224** rad/s (= 2138.9 rpm).

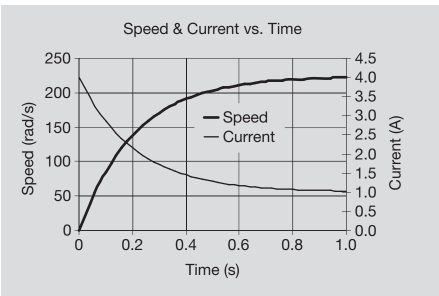


Figure A.8 Motor and Load Accelerated from 0 Initial Shaft Speed

EXAMPLE 2 (Figure A.9): same motor and load parameters, battery voltage is stepped down from a steady state **14 V to 8 V**.

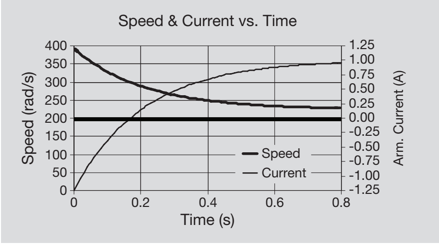


Figure A.9 Supply Voltage is Step-Changed from 14 V to 8 V

EXAMPLE 3: same motor parameters, but supply voltage is changed from an initial steady state **8 V to -8 V**.
Final velocity is -224 rad/s (-2139 rpm), final current value is -1.005 A. The negative signs

indicate sign reversals (i.e. direction of rotation. Notice the large current transient when the drive voltage is reversed!

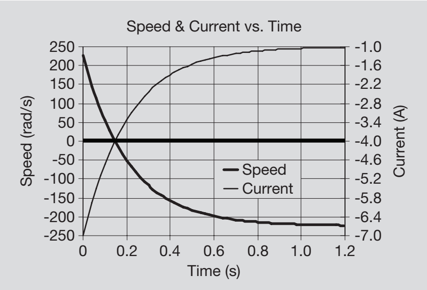


Figure A.10 Supply Voltage Polarity is Reversed from 8 V to -8 V

A.6 Semiconductor Realization of Motor Controllers

The driver topologies depend on the nature of the desired control, namely is it unidirectional or bidirectional? Unidirectional drivers use half-bridge topology and bidirectional drivers have full bridge topology. The driver schematic diagrams are in Figure A.11.

The two transistors in the half bridge are driven 180 degrees out of phase. When Q1 is on Q2 is off and vice versa. Using pulse-width modulated (PWM) drive, Q1 is fully switched on and current flows through Q1 and the load as indicated with the **bold** arrow (Figure A.11). When Q1 is off Q2 is turned on providing a path for the current stored in the motor inductance (Figure A.12). Given period T of the PWM signal the effective DC voltage across the motor is equal to:

$$V_{motor} = V_b\left(\frac{T_{on}}{T}\right) = V_b * D\ , \tag{35}$$

where D is the duty cycle.

During recirculation the current in the motor driven by either of the circuits changes only by a small amount provided the PWM period (T) is shorter than 2-3 times the electrical time constant.

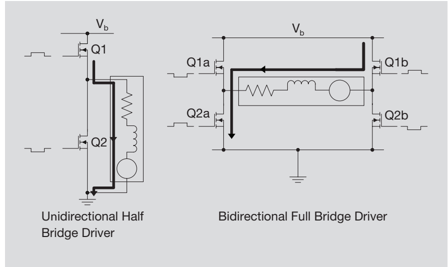


Figure A.11 Bridge Driver Topologies

In the bidirectional bridge driver transistor Q1b is always on transistor Q2b is always off and transistors Q1a and Q2a are alternatively turned on and off with 180° phase difference. The described control causes DC current to flow through the motor as indicated with the **bold** arrow in Figure A.11. Notice, however, that when Q1a and Q1b are both on the recirculation current path is through transistors Q1a and Q1b, and the current in Q1a is from source to drain (i.e. reversed) see Figure A.12. When the roles of the H-bridge transistors are reversed, the drive voltage polarity is reversed.

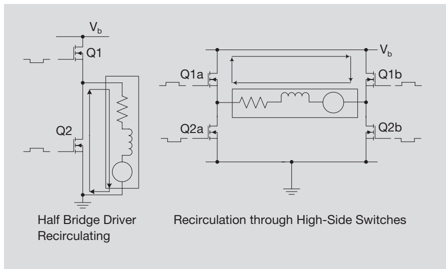


Figure A.12 Half and Full Bridge in Recirculating Mode

The drive and recirculating modes of operation are illustrated in Figure A.11 (drive) and Figure A.12 (recirculation).

The ideal control circuit for a PMDC motor has variable voltage output and low output impedance. The pulse-width modulated topologies approximate the ideal driver:

- during the drive cycle the turned on transistors work in the linear mode and only their R_{DSon} is in series with the low internal impedance battery;

- during recirculation the re-circulating transistor(s) short out the motor terminals with the R_{DSon} resistor(s)..

The output voltage is determined by the battery voltage and the duty cycle (D). During recirculation the voltage is close to zero, therefore averaging the driving and re-circulating voltages yields the output voltage (in the first approximation).

A.7 Power Calculations

A.7.1 Output Power

The mechanical output power of the motor is given by the product of the torque and the shaft angular velocity.

$$P_{mech} = T * \omega \tag{36}$$

Using the torque vs. speed function (Equation (19)) at constant terminal voltage V_T, the mechanical power is given by:

$$P_{mech} = \frac{k}{R_a} * (V_T - k * \omega) * \omega \tag{37}$$

for $0 \le \omega \le \frac{V_T}{k}$

Maximum power output is calculated from Equation (37) by taking the first derivative with respect to ω:

$$\frac{dP_{mech}}{d\omega} = \frac{k}{R_a} * (V_T - 2k\omega)\ ,$$

P_{mech} is maximum at ω_{max} where its derivative is zero with respect to ω, or

$$\omega_{max} = V_T / 2k \tag{38}$$

The mechanical power output at ω_{max} is obtained from Equation (37) and Equation (38):

$$maxP_{mech} = \frac{k}{R_a} * \left(V_T - k * \frac{V_T}{2k}\right) * \frac{V_T}{2k} = \frac{V_T^2}{4R_a} \tag{39}$$

A.7.2 Input Power

The electrical power is the product of the terminal voltage and the armature current:

$$P_{elect} = V_T * i_a$$

At the $maxP_{mech}$ the armature current is given by:

$$i_a = \frac{(V_T - k\omega_{max})}{R_a} = \frac{V_T - k\frac{V_T}{2k}}{R_a} = \frac{V_T}{2R_a} \tag{40}$$

thus the input power is equal to:

$$P_{elect} = V_T * \frac{V_T}{2R_a} = \frac{V_T^2}{2R_a} \tag{41}$$

The ratio of Equation (39) to Equation (41) shows that the efficiency of the motor at the **peak power output** is 50%.

A.8 Efficiency

It must be noted that the idealized model, that assumes zero friction losses, will give false efficiency results if “pushed to the limits”. Just using the equations derived so far would show that at a given terminal voltage, V_T, the peak efficiency is at $\omega = V_T / k$ but that result is a non-sense since at that speed the torque is zero. If, however, the friction losses are considered, at the maximum no-load speed (ω_0) the motor has to deliver T_{fr} torque to overcome the friction. If at $\omega = 0$ the stationary motor torque is T_{st}, the speed of maximum efficiency is given by

$$\omega\ max_eff = \omega_0\left(1 + \frac{T_{fr}}{T_{fr}+T_{st}}\sqrt{\left(\frac{T_{fr}}{T_{st}}\right)^2 + \frac{T_{fr}}{T_{st}}}\right) \tag{42}$$

and the maximum efficiency will be equal to:

$$max_efficiency = \left(1 - \sqrt{\frac{T_{fr}}{T_{fr}+T_{st}}}\right)^2 \tag{43}$$

Equation (42) and Equation (43) have been worked out for small (1 HP<) permanent magnet DC motors.

References:

1) Eric Brasseur: "Data and formules about little elctric motors" http://www.4p8.com/eric.brasseur/emamem.html

A.9 Conclusions

Fundamental concepts and basic applications of armature commutated permanent magnet DC motor drives have been presented in this short paper. Those types of motors are very common, cost effective (because of the huge demand), easily understood and readily analyzable electromagnetic devices. The general principles can be extended to other machines, like brushless DC motors, which also produce back emf, torque dependent on the stator current, recirculation drive, etc.

The differential equations will be helpful in building dynamic computer models whereas the solutions of those equations will ease the hand calculator evaluations. The definition and explanation of the specification entries (and their units) will help in motor selection.

Glossary of Terms

Term	Description/Definition	Source
6Ms	causes resulting in process shifts (Man, Machine, Materials, Methods, Mother Nature and Measurements) or sequence in troubleshooting procedures	
8D	documentation format of SIX SIGMA failure analysis procedures	
AC	alternating current. Electrical current whose magnitude and direction vary cyclically; charge carriers oscillating in conductors or insulators (SI unit: A)	WKP
ACCELERATION	vectorial rate of change of velocity vctor (SI unit: m/sec^2)	
AC-to-DC CONVERTER	a type of circuit (sytem) that converts an AC power source to a DC power source (also known as the RECTIFIER)	
AFM	atomic force microscopy (an F/A tool)	
AMBIENT TEMPERATURE	temperature of the spece surrounding the semiconductor device. Symbol: T_A, Units: °C, K	
ANGSTROM	a unith of length equal to 10^{-10} m (mostly used in optics as a measure of wavelenght of light)	W
ANGULAR ACCELERATION	$d\omega/dt$; rate of change of angular velocity of a rotating object (SI unit: rad/sec^2)	
ANGULAR VELOCITY	ω; rate of change of angular position of a rotating object (SI unit: rad/sec)	
APQP	Advanced Prodact Quality Planning (also see SIX SIGMA)	WKP
ARMATURE	rotating electromagnet in a DC motor exerting TORQUE on the motor shaft by means of interacting with the field of the stator magnet (permanent or electromagnet)	
ARMATURE INDUCTANCE	inductance of the armature winding, L_a measured between the brushes (SI unit: Ωsec = V x sec/A). See also INDUCTACE	
ARMATURE RESISTANCE	resistance of armature winding R_a measured between the brushes (SI unit: Ω = V/A). See also RESISTANCE	
ASSIGNABLE CAUSE	refers to an inadvertent specific incident which is caused either by an operator's mistake or some failure either in the equipment or in the process. It is also referred to as a quality accident because it is not planned for and not intended to happen.	
ATOM	smallest particle of an element that can exist either alone or in combination	W

Term	Description/Definition	Source
AVALANCHE BREAK-DOWN	very high voltage drop in the space charge zone with PN junctions causing the loss of the insulating property of the zone. The high voltage accelerates the charge carriers in the space charge zone so much that new charge carriers are generated in an avalanche-like fashion through impact ionization.	INF
BAND GAP	energy difference between the top of the valence band and the bottom of the conduction band, where electrons are able to jump from one band to another	WKP
BAND GAP REFERENCE	voltage drop across a diode junction compensated for temperature to ensure both precision and stability	
BARRIER LAYER	in semiconductor technology, it is a thin, active or passive area that separates or connects other layers (see also DIFFUSION)	INF
BASE	one of the three electrodes (terminals) of a BJT	
BATTERY	it is a device for storing and supplying DC electrical energy	
BIASING	bias point, also known as operating point, quiescent point or Q-point, is a dc voltage (current) which, when applied to a device, causes it to operate in a certain desired fashion. The application of the bias is BIASING	WKP
BIPOLAR	property of semiconductor components that use both p-doped as well as n-doped semiconductors. The simplest structure is the PN diode	
BJT	bi-polar junction transistor	
BODE-PLOT	combined open loop frequency response of a controller and the load used in a feedback control loop (see FREQUENCY RESPONSE)	
BOLTZMAN-CONSTANT	1.38065 10^{-23} J/K OR 8.617 10^{-5} eV/K	
BOND PAD	electrical terminal on the DIE anchoring one end of the BOND WIRE	
BOND WIRE	metallic connecting element between BOND PAD and LEADFRAME terminal	
BONDING	attaching fine wires to BOND PADS on the DIE and to LEADFRAME contacts	

Term	Description/Definition	Source
BOOST REGULATOR	a SWITCHING REGULATOR that converts a supply voltage source to a higher voltage source	
BOOTSTRAP CIRCUIT	A type of capacitive positive feedback between the gate and source of a switching transistor to raise the gate voltage higher than the drain voltage.see also CHARGE PUMP and SWITCHING REGULATOR	
BREAK DOWN VOLTAGE	in semiconductors, overcoming a voltage or current barrier and the subsequent surge in the injection current	INF
BRUSH (DC MOTOR)	graphite contactors touching the COMMUTATOR, providing contacts between the BATTERY terminals and the ARMATURE winding(s)	
BRUSHLESS DC MOTOR	it is a DC motor with electronically controlled commutation system	
BUCK REGULATOR	a SWITCHING REGULATOR that converts a voltage source to a lower voltage source	
CAPACITANCE	capacitance is a measure of the amount of ELECTRIC CHARGE stored (or separated) for a given electric potential. (SI unit: F = Axsec/V)	WKP
CAPACITOR	device consisting of conductors separated by insulators for storing ELECTRIC CHARGE. Capacitors block DC current and pass AC current	WKP
CASE TEMPERATURE	temperature measured at the portion of the leadframe exposed to the AMBIENT by the encapsulated package	
CHANNEL	current path in a MOSFET between SOURCE and DRAIN. The width of the channel, which determines how well the device conducts, is controlled by an electrode called the GATE, separated from channel by a thin layer of oxide insulation. The insulation keeps current from flowing between the gate and channel.	EEH
CHARGE PUMP	an electronic circuit that uses capacitors as energy storage elements with diodes and/or electronic switches to create either a higher or lower voltage than power source	WKP
CHARGED **D**EVICE **M**ODEL (CDM) TEST	CDM test is used to define how much ESD a device can withstand when the device itself has an electrostatic charge and is discharged to metal contact	WKP
CHIP	see DIE	

Term	Description/Definition	Source
CLAMPING	prevention of the voltage across a load (or current through a load) to exceed a predetermined limit by means of appropriate circuit topology based on the use of non-linear circuit elements or periodic switching of the supply	
CMOS	An MOS technology that works with complementary logic elements consisting of individual p-CHANNEL and n-CHANNEL MOSFETs. This circuit technology considerably reduces the power requirements of the logic elements because power is consumed only during state changes (switching)	INF
CMOS INVERTER	the most basic logic element that consists of a pMOS and an nMOS transistor with connected drains and connected gates. The output is the inverse (or logic negation) of the input.	
COLLECTOR	one of the three electrodes (terminals) of a BJT into which the charge carriers drain	
COLLECTOR CURRENT	current flowing between collector and emitter of a BJT controlled by the current flowing between base and emitter	
COMMON CAUSE	inherent process limitation that causes periodic change in product characteristics	
COMMUTATOR	component in a DC motor that periodically reverses the current in the armature of the motor to alternate the orientation of magnetic field of the armature	
CONDUCTOR	(electrical) conductor, any material that easily permits the flow of electric current	WKP
COULOMB	SI unit of electrical charge; dimension A x sec	WKP
COULOMB FRICTION	force (torque) resisting movement; it is independent of (angular) velocity in the first approximation (SI unit: see TORQUE)	
Cpk	process capability index. Definition: Cpk = the lesser of [(USL-Mean) / σ, (Mean-LSL) / σ]	
CURRENT GAIN (BJT)	ratio of collector current to base current (symbol: β or h_{fe}) of a BJT	
CURRENT LIMITING	reduction of average current supplied by en electronic switch (MOSFET, BJT) by means of repeatedly turning the current off and on or CLAMPING	

Term	Description/Definition	Source
CURRENT MODE CONTROL	a type of FEEDBACK CONTROL LOOP in which the output voltage or the output current of the loop is regulated by comparing a sample of the output quantity to an appropriate REFERENCE	
CURRENT RECIRCULATION	maintaining current built up in an INDUCTOR by means of providing an alternative path for the said current stored in the inductor when means of driving (voltage or current) is switched off. Also see PWM.	
CURRENT SOURCE	source of electrical energy from high internal RESISTANCE (IMPEDANCE) supply	
CYCLE	unit of phase angle equivalent to one oscillation, or 2π radians (SI unit: rad)	
DC	Dirrect current. Movement of charge carriers in the same direction (SI unit: A [ampere])	WKP
DC MOTORS	rotating machines driven from DC electric energy source	EEH
DC-to-DC CONVERTER	an interface circuit (linear or switching) between a DC energy source (input) and a load supplied with DC energy from the converter output	
DECAPSULATION	partial or total removal of the packaging material from a semiconductor die (F/A tool)	
DEFECT DENSITY	inherent variations resulting from the hundreds of process steps performed under precise control	
DEPLETION REGION	also called depletion layer or depletion zone, as well as the junction region or the space charge region is an insulating region within a conductive, doped semiconductor material where the charge carriers have diffused away or have been swept away by an electric field	WKP
DEVICE TESTING	checking the package device for functionality and performance quality	
dFMEA	design FMEA	
DIAGNOSTIC FEEDBACK	some form of indication (flag) from the integrated power device to the controlling device that its operation is either interrupted permanently (latch-off) or temporarily due to exceeding one or more operating limits	
DIE	a finished individual transistor or integrated circuit fabricated on and harvested from a WAFER	

Term	Description/Definition	Source
DIE ATTACH or DIE BONDING	mounting/fixing the die to the LEAD FRAME	
DIELECTRIC CONSTANT	measure of storage capability of insulating materials. $\varepsilon = \varepsilon_0 \times \varepsilon_r$ (SI unit: [V x sec] / [A x m]). See also RELATIVE DIELECTRIC CONSTANT and DIELECTRIC CONSTANT OF VACUUM	
DIELECTRIC CONSTANT OF VACUUM (free space)	$\varepsilon_0 = 8.854 \times 10^{-12}$ (V x sec) / (A x m)	
DIELECTRIC MATERIALS	insulators blocking electric current and storing electric charge in capacitors	
DIFFUSION	the spontaneous spread of molecules and/or charge carriers within the semiconductor material, see also BARRIER LAYER.	
DIODE (semiconductor)	a component that blocks the flow of electrons in one direction (from anode to cathode) and CONDUCTS in the other	INF
DMAIC	**D**escribe, **M**easure, **A**nalyze, **I**mprove and **C**ontrol; sequence of actions used in the six sigma process	
DMOS	**D**ouble Diffused **MOS** (process used in power MOSFET fabrication)	INF
DOPING	The specific admixing of foreign atoms in a semiconductor material. A distinction is made between donor atoms that bring in free electrons and acceptor atoms that deprive the semiconductor atoms of electrons and cause “holes”.	INF
DRAIN	one of the electrodes of a FET designating the power outlet end of a conductive CHANNEL	INF
DRAIN-to-SOURCE BREAK DOWN VOLTAGE	drain-to-source voltage at which the MOSFET begins to conduct in spite of the fact that $\lvert V_{GS} \rvert - \lvert V_T \rvert = 0$	
DUTY CYCLE	portionon of time (angle) of a CYCLE during which a component, device, or system is operated; See PULSE WIDTH MODULATION;	WKP
DYNAMIC BRAKING (relative to PMDCM)	application of retarding torque by reduction or reversal of the motor drive voltage of PMDCM	

Term	Description/Definition	Source
DYNAMIC PAT	continuous updating of test limits based on accumulated actual performance data.	
EARLY EFFECT (BJT)	effective decrease in the base width because of the widening of the base-collector depletion region, resulting in an increase in the collector current with an increase in the collector to emitter voltage.	WKP
E-BEAM	electron beam (F/A tool)	
ELECTRIC CHARGE	Electric charge is a fundamental conserved property of some subatomic particles, definite quantity of electricity (SI unit: A x sec)	WKP
ELECTRIC FIELD	a vector with SI units of newtons per coulomb ($N\ C^{-1}$) or, equivalently, volts per meter ($V\ m^{-1}$). The direction of the field at a point is defined by the direction of the electric force exerted on a positive test charge placed at that point	WKP
ELECTRIC MOTOR	a device that converts electrical energy into mechanical energy	
ELECTRICAL EFFICIENCY	useful power output per electrical power consumed	
ELECTRICAL OVERSTRESS (EOS)	event resulting from the application of voltage or current to a device exceeding its ratings (device is stressed beyond its specified limits of operation)	INF
ELECTROMAGNET	a type of magnet in which the magnetic field is produced by the flow of an electric current. The magnetic field disappears when the current ceases.	EEH, WKP
ELECTRON	electron is a fundamental subatomic particle that carries a negative electric charge. Electron charge = 1.6022×10^{-19} COULOMB, mass = 9.11×10^{-31} kg	WKP
ELECTRON CHARGE	1.6022×10^{-19} COULOMB	
EMC	**e**lectro**m**agnetic **c**ompatibility is the branch of electrical sciences which studies the unintentional generation, propagation and reception of electromagnetic energy with reference to the unwanted effects that such an energy may induce	WKP
EMI	**e**lectro**m**agnetic **i**terference: a (usually unwanted) disturbance caused in a radio receiver or other electrical circuit by electromagnetic radiation emitted from an external source	WKP

Term	Description/Definition	Source
EMITTER	one of the three electrodes (terminals) of a BJT from which the charge carriers originate	
ENCAPSULATION	refers to the design and manufacturing of protective packages for integrated circuits and other kind of semiconductor devices	WKP
ENERGY	ability to do same amount work as its magnitude (SI unit: N x m, joule, watt x sec)	
ENHANCED STATE	conducting MOSFET CHANNEL is established when the gate-to-source voltage (V_{GS}), the drain-to-source voltage (V_{DS}) and the threshold voltage (V_T) satisfy $\lvert V_{DS}\rvert > \lvert V_{GS} - V_T\rvert$	
EOS	see ELECTRICAL OVERSTRESS	
ERROR-PROOFING	design of a process such that errors which may cause a defect are not made either by equipment or operator	
ESD	**e**lectro**s**tatic **d**ischarge: sudden and momentary electric current that flows between two objects at different electrical potentials	WKP
ESR	**E**quivalent **S**eries **R**esistor of a capacitor	
ETCHING	a subtractive process used in microfabrication of semiconductors to remove layers from the surface of a wafer during manufacturing either by disolving or by evaporation	
EXTRINSIC FACTORS	those factors that are associated with the manufacturing processes used to produce a particular device	
FABRICATION or fab. (semiconductor)	the processing sequence of a semiconductor material (silicon most of the time) in the manufacture of transistors or integrated circuits	
FAILURE ANALYSIS (F/A)	investigation to determine the cause of device failures occuring anywhere in th supply chain	
FEEDBACK CONTROL LOOP	a class of control systems which compares outputs to predetermined references and minmizes the error between them by means of a control strategy	
FEEDBACK LOOP STABILITY	behavior of a feedback control loop that ensures the expected results of its control strategy	
FIB	focused ion beam (an F/A tool)	

Term	Description/Definition	Source
FIELD **E**FFECT **T**RANSISTOR (FET)	type of transistor that relies on an electric field to control the shape and hence the conductivity of a “channel” in a semiconductor material	WKP
FISHBONE DIAGRAM	or Ishikawa Diagram. It is one of the seven basic tools of quality management. It is a cause-and-effect diagram that can reveal key relationships among various variables and the possible causes, providing additional insight into process behaviour.	WKP
FLUX DENSITY	**B** vector quantity. |**B**| is given by the FORCE exerted on an ELECTRIC CHARGE moving perpendicular to a magnetic field defined by **B** with VELOCITY **v**. [SI unit: N/(A x sec x msec^{-1}) = V x sec/m^2] Note: bold print letters are vectors	
FMEA	Failure Mode and Effect Analysis	
FORCE	cause of acceleration of a mass (SI unit: Newton, N, N = kg x m/sec^2)	
FORWARD BIAS	voltage applied to a semiconductor junction in direction giving higher current	EEH
FR-4 PCB	double sided and multilayer printed circuit boards made of FR-4 (a fiberglass flame resistant) type material	WKP
FREQUENCY RESPONSE	the combined function of amplitude vs. frequency and the function of phase vs. frequency of a physical system.	
FRERRO-MAGNETIC MATERIALS	materials that concentrate magnetic field lines when placed in magnetic field. Some materials can be permanently magnetized by an external magnetic field	
GATE (FET)	one of the electrodes of a FET controlling the channel current between source and drain.	INF
GATE (logic)	logic gate performs a logical operation on one or more logic inputs and produces a single logic output	WKP
GATE CAPACITANCE	the CAPACITANCE measured between the GATE and the SOURCE electrodes of a MOSFET	
GENERATORS (electric)	rotating machines producing electric energy in response to mechanical drive	
g_m	conductivity of a MOSFET defined as the derivative of drain current with respect to the gate-source voltage (dI_D/dV_{GS}). The unit of g_m is A / V (Siemens).	

Term	Description/Definition	Source
H-BRIDGE	a power control electronic circuit which controls DC electric motors to rotate clockwise or counter-clockwise	
HEATSINK	an object attached to the semiconductor case (with a very low thermally resistant path) to effectively increase the surface (radiating) area of the case	
HITFET	**h**igh **i**ntegration, **t**emperature-protected **f**ield **e**ffect **t**ransistor. It is a MOSFET with added protection and performance enhancing features	
HOLE	positive charge carrier in a semiconductor material. Holes are created by implanting ions in the host semicunductor material that have electron defficiency	
HSD	**h**igh **s**ide **d**river. The term identifies the circuit topology in which the driver is returned (referenced) to the supply voltage and the load to ground; in short, the driver is connected to the **high side** of the load..	
HUMAN BODY MODEL ESD	HBM a 100 pF capacitor charged to 2 kV suddenly discharged through a 1.5 kΩ resistor	
HYSTERISIS	the magnitude of an input signal causing an opposite state change in the output that has been determined by the previous magnitude of the input	
IC ASSEMBLY (PACKAGING)	the installation of an individual die (or a small cluster of dies) into a single component package to facilitate connection to the application board or module.	
IMPEDANCE	measure of opposition to a sinusoidal alternating current of INDUCTORs and CAPACITORs (SI unit |Z| = V / A)	WKP
INCANDESCENT LAMP	electric light source (light bulb) consisting of an electrically heated filament placed in a glass envelope containing low pressure inert gas.	WKP
INDUCTANCE	mesure of the degree of SELF-INDUCTION (SI unit: V x sec / A = H or Henry)	
INDUCTOR	electric component that opposes AC currents and passes DC currents. See SELF-INDUCTION.	
INRUSH CURRENT	initial current demand of an electric load (cold lamp filament or DC motor with stationary shaft) at the first application of supply voltage	
INSULATOR	material that does not conduct electricity	

Term	Description/Definition	Source
INTEGRATED CIRCUIT (IC)	miniaturized electronic circuit (consisting mainly of semiconductor devices, as well as passive components) that has been FABRICATED in the surface of a thin substrate of semiconductor material.	
INTRINSIC FACTORS	manufacturing factors that are bounded by the limits of physical nature	
IV CHARACTERISTICS	current response to a voltage input (or vice versa); also known as U-I Curves.	
JUNCTION SPIKING	under ESD events, large currents flowing through a contact heating up the contact to the point where the aluminum diffuses deep into the silicon shorting out the p-n junction	
JUNCTION TEMPERATURE	in the context of semicoductors: temperature of the semiconductor die. Symbol: T_J. (SI unit: Kelvin degree, K)	
k constant	fundamental parameter of PMDCMs. Definition 1: ratio of rotor torque and armature current. Definition 2: ratio of voltage induced in the armature and the angular velocity of the armature (SI unit: $[kg \times m^2] / [A \times sec^2] = V / [rad / sec]$)	
KELVIN DEGREE	measure of temperature related absolute zero. 1 K temperature **change** is equal to 1 °C temperature **change**. 0 °C = 273.15 K	
LATCH-OFF	terminating the operation of a protected power semiconductor device because of out of range operating condition(s). The latch-off state is removed by either a power-off/power-on sequence or by some other action of the controlling source	
LATENT DEFECTS	flaws or other imperfections in a device which surface after the semiconductor component is placed in the application and is exercised or activated	
LCT	liquid crystal thermography (an F/A tool)	
LDO	**L**ow **D**rop-**O**ut voltage regulator. It is a class of LINEAR REGULATORs with the EMITTER (or SOURCE) of its SERIES PASS TRANSISTOR connected to the power source and the COLLECTOR (or DRAIN) to the load.	
LEADFRAME	or chip carrier: a right-angled metal frame with pin wiring and linking elements in between for package mounting	INF

Term	Description/Definition	Source
LEADFRAME (CASE) TEMEPRATURE	temperature of part of leadframe not covered by the package Symbol: T_C	
LEFT HAND RULE	Fleming's left hand rule (for electric motors) shows the direction of the thrust on a conductor carrying a current in a magnetic field.	WKP
LINEAR POWER SUPPLY	a class of circuits which contiuously transform and rectify AC voltage sources and continuously regulate the rectified voltage with LINEAR REGULATORS	
LINEAR REGION (MOSFET)	the drain-source voltage (V_{DS}) is between V_{GS} and V_T: $\lvert V_T\rvert \leq \lvert V_{DS}\rvert < V_{GS}$	
LINEAR REGULATOR	a class of control systems which compares outputs to predetermined references and minmizes the error between them by means of a control strategy	
LOAD DUMP TRANSIENT	high voltage, high energy impulse in automotive electrical systems caused by the reaction to the alternator disconnect from the battery while remaining connected to the loads. The enrgy stored in the stator of the alternator is **dumped** on the loads before the regulator can reduce the alternator voltage.	
LSD	**l**ow **s**ide **d**river. The term identifies the circuit topology in which the driver is returned (referenced) to ground and the load to the supply voltage or the driver is connected to the **low side** of the load..	
LSL	lower specification limit	
MACHINE MODEL ESD	MM a 100 pF capacitor, charged to 250 V, suddenly and directly discharged	
MAGNETIC FIELD	portion of a space near a magnetic body or a current carrying body in which forces due to the magnetic body or current can be detected.	
MAGNETIC FIELD STRENGTH	measure of MAGNETIC FIELD. (SI unit: A / m)	W
MAGNETIC FLUX	integral of the FLUX DENSITY vector over a directed surface. SI unit: V x sec	
MAGNETIC POLES	interface between the internal magnetic field and the external field of magnetized substances or the interface between the region within a current loop (loops) and regions outside it. Poles are always in pair (dipole) , North and South	

Term	Description/Definition	Source
MAJORITY CHARGE CARRIERS	primarily responsible CHARGE carrier for current transport in a piece of semiconductor. If an intrinsic semiconductor (which does not contain any impurity) is doped with N-type impurity then the majority carriers are electrons and if the semiconductor is doped with P-type impurity then the majority carriers are holes. See DOPING and HOLES	WKP
MAVERICK LOT	any production lot yielding fallouts in excess of the fallout limits in a specific test bin	
MILLER CAPACITOR	**Miller effect** describes the fact that a capacitance between input and output of an inverting amplifier is multiplied by a factor of $(1 – A_v)$, where A_v is the magnitude of the voltage gain of the amplifier.	WKP
MODELING	any deliberate, intelligible cognitive activity aimed at abstracting and reproducing some convenient realm of discourse, features of an object or system of interest of the modeler.	EEH
MODELS of MOSFETs	in terms of increasing complexity: square-law, bulk charge theory, charge sheet model and exact charge model.	
MOLDING COMPOUND	materials used for encapsulating semiconductor devices are known as plastic molding compounds	
MOMENT of INERTIA	J; indisposition to rotating torque, the measure of the ability to maintain rotation in the absence of driving torque (SI unit: kgxm2)	
MOORE's LAW	the empirical observation made in 1965 that the number of transistors on an integrated circuit for minimum component cost doubles every 24 (18?) months.	WKP
MOS TECHNOLOGY	FAB technology based on **m**etal-**o**xide **s**ilicon field effect transistors	
MOSFET	combination of the acronyms **MOS** and **FET**, i.e. a field effect transistor based on MOS technology	INF
MOTOR electrical TIME CONSTANT	the ratio of armature inductance, L_a to the armature resistance, R_a; $\tau_e = L_a / R_a$	INF
MOTOR electro-mechanical TIME CONSTANT	time it takes the armature of a PMDCM to slow to 36.79% of its initial angular velocity with shorted terminals and zero external load; $\tau_a = (J_a \times R_a) / k^2$. J_a: the MOMENT OF INERTIA of armature, R_a: resistance of armature winding.	

Term	Description/Definition	Source
MOTOR OUTPUT POWER	P_m =T x ω; the product of the torque T acting on the motor shaft and shaft angular velocity ω	
MTBF	mean time between failures (a reliability measure)	
MTTF	mean time to failure (a measure of robustness)	
MULTI-PHASE SWITCHING REGULATOR	a class of switching regulators with multiple switches and inductors with apprporiately connected outputs feeding a common load. The phases of the switchings are arranged so as to avoid any failure prone interactions between the switches.	
N-CHANNEL MOSFET	FET that is controlled with a positive gate source voltage and blocked with a positive drain source voltage	INF
NEUTRON	an uncharged elementary particle in nuclei of atoms (except hydrogene). Mass = 1.673 x 10^{-21} kg (same as the proton mass)	
NEWTON (N)	measure of force in SI units (N =kg x m / sec^2)	
NO-LOAD SPEED	angular velocity of a powered motor at zero torque external loading	
NORMAL DISTRIBUTION	distribution of measured parametric values of the same kind (x-axis) and the number of samples having the said value (y-axis) obtained by measuring a multitudes of samples. The shape of the distribution is bell-like, characterized by the mean (μ) of the parametric values where the peak y value is and the spread (σ) or standard deviation of the bell curve.	
NPN TRANSISTOR	**N**egative **P**ositive **N**egative transistor: a layer sequence for bipolar transistors	INF
N-TYPE SEMICONDUCTOR	an impurity of valence-five element implanted in a valence-four semiconductor to increase the number of free negative (electron) charge carriers	WKP
NUCLEUS	the positively charged center of the atom that comprises nearly all the atomic mass consisting of NEUTRONs (except in HYDROGENE) and PROTONS	W
OEM	**o**riginal **e**quipment **m**anufacturer (automobile manufacturing companies)	
ONE-EIGHT RULE	approximate formula for calculating switch-on losses of PROFET-s controlling INCANDESCENT LAMP loads. P_{loss} = approx.V_{bat} x I_{load} / 8	INF

Term	Description/Definition	Source
OPERATING JUNCTION TEMPERATURE RANGE	if the junction temperature of the semiconductor component is within this temperature range, it **shall operate** in a specific manner.	
OPERATIONAL AMPLIFIER	is a DC-coupled high-gain electronic voltage amplifier with differential inputs and, usually, with a single output. In its ordinary usage, the output of the op-amp is controlled by negative feedback which, because of the amplifier's high gain, almost completely determines the output voltage for any given input.	WKP
OPTIMUM SLEW RATE	best compromise between low switch-on (off) losses and response time	
OSCILLATOR	an electronic circuit generating periodic signals	
OUTER SHELL	The VALENCE shell is the outermost shell of an atom in its uncombined state, which contains the electrons most likely to account for the nature of any reactions involving the atom and of the bonding interactions it has with other atoms	WKP
OUTLIER	any device whose measured characteristics value is outside the typical variance normally within $\pm 3\sigma$ of the mean or set target	
OZ-RATING of PCB-s	weight of the copper laminate in ounce per square feet (1 oz/sqft = 0.3051 kg/m^2)	WKP
PART AVERAGE TESTING	a statistically based method used during parametric testing (wafer probe or final test) of devices to reject components with abnormal characteristics called outliers. The average for the part is the target and the PAT limit is typically set to Six Sigma (see DYNAMIC PAT)	
P-CHANNEL MOSFET	FET that is controlled with a negative gate source voltage and blocked with a negative drain source voltage	INF
PCB	**p**rinted **c**ircuit **b**oard; carrier of electronic components providing printed conductor interconnections between them.	
PCP	Process Control Plan	
PDCA	**p**lan **d**o **c**heck **a**ct - general steps in most problem solving methods	
PDP	Product Development Process	

Term	Description/Definition	Source
PERMANENT MAGNET	material that retains its magnetism after removal of the magnetizing force	WKP
PMDCM	**p**ermanent **m**agnet **dc m**otor. A type of electric motors driven by DC power source supplying its electromagnet (via brushes and commutator) which rotates in the field of a stationary permanent magnet;	
PERMEABILITY	Property of a magnetizable substance that determines the degree by which it modifies the MAGNETIC FLUX in the region it occupies in a MAGNETIC FIELD. Symbol: μ	W, WKP
PERMEABILITY of VACUUM (free space)	Symbol: μ_0, value: $\mu_0 = 4\pi \cdot 10^{-7}$ (V x sec) / (A x m)	
pFMEA	process FMEA	
PFMEA	Potential FMEA	
PNP TRANSISTOR	**P**ositive **N**egative **P**ositive transistor: a layer sequence for bipolar transistors	INF
POTENTIAL DIFFERENCE	see VOLTAGE	
POWER	work done in a given time interval (SI unit: joule/sec, W, N x m/sec)	
POWER DISSIPATION	rate of change of electrical energy imparted to a load defined by the product of the momentary voltage drop across the load and the momentary current flowing through the load. (SI unit: W = V x A)	
POWER SUPPLY	the interface circuit between the input power source and the output power delivered to a load	
PPM	parts per million; a measure of failure or reject rate	
PROFET	**pro**tected **f**ield **e**ffect **t**ransistor. It is a MOSFET with added protection and performance enhancing features	
PROTECTION CIRCUIT	special sub-circuits used in power supply or power control systems that interfere with the normal operation so as to prevent system failure in case of out of specification power input, load output and component and/or system temperature values	
PROTON	a charged elementary particle in nuclei of atoms. Charge = 1.6022×10^{-19} C Mass = 1.673×10^{-21} kg	

Term	Description/Definition	Source
P-TYPE SEMICONDUCTOR	an impurity of tri-valent element implanted in a valence-four semiconductor to increase the number of free positive charge carriers (holes).	
PULSE WIDTH MODULATION (or PWM)	signal or power source of rectangular pulses that involves the modulation of its duty cycle, to either convey information over a communications channel or control the amount of power sent to a load	WKP
PULSE-SKIPPING	intermittent reduction of the number of switching cycles of a SWITCHING REGULATOR in order to improve the EFFICIENCY at low load currents	
QUALITY 1	Quality per ISO 9000: 2000 Is the degree to which products and services fulfill the requirements and expectations of the customers.	
QUALITY 2	meeting customers' requirements consistently over specified period of time, with the best imaginable service, at the best price	
QUIESCENT CURRENT	current consumed by an electronic circuit in the stand-by state	
RADIAN	length of arc of a unit radius circle corresponding to its central angle (full circle = 2π rad, no SI dimension)	
R_{DSon} or $R_{DS(on)}$	resistance between the SOURCE and DRAIN terminals of an enhanced MOSFET	
RECTIFIER	see AC-to-DC CONVERTER	
REGIONS OF N-MOSFET OPERATIONS	$V_{GS} < V_T$: sub-threshold region, $I_D = 0$, $V_{GS} > V_T$: super-thershold region, $I_D > 0$	
RELATIVE DIELECTRIC CONSTANT	ε_r, (no dimension). Measure of electric charge storage capability of insulators relative to vacuum. ε_r of vacuum (air very close) is 1.	
RELATIVE PERMEABILITY	μ_r, (no dimension) measure of permeability of a substance relative of vacuum. μ_r of vacuum (air very close) is 1	EEH
RELIABILITY	the ability of a product to give the same good quality result consistently	
RESISTANCE	measure of resistor (SI unit: $\Omega = V / A$). See also RESISTOR	

Term	Description/Definition	Source
RESISTOR	a two terminal device opposing the transfer of charges, i.e. current. Voltage develops across its terminals proportionally to AC or DC CURRENT.	
REVERSE BIAS	opposite of FORWARD BIAS	EEH
RIPPLE CURRENT	periodic small amplitude alternating current superimposed on, or modulating a DC or a lower frequency current	
RIPPLE VOLTAGE	periodic small amplitude voltage superimposed on, or modulating a DC or a lower frequency voltage	
RMS value (electrical engineering)	**RMS** = **R**oot **M**ean **S**quare: the value of the equivalent DC voltage (current) that dissipates the same energy in a unit load resistance as the voltage (current) having AC and DC components	
RMS value (mathematics)	the square root of the sum arithmetic mean of the squares of a set of numbers	W
SAFE OPERATING AREA (SOA)	safe operating area for power semiconductors is defined as the voltage and current conditions over which the device can be expected to operate without self-damage.	WKP
SAM	scanning acoustic microscpoy (an F/A tool)	
SATURATED REGION OF N-CHANNEL MOSFET OPERATION	part of the super thrashold region. The drain-source voltage (V_{DS}) is greater than the difference V_{GS}-V_T: $V_{DS} > (V_{GS} - V_T)$.	
SBA (SYA)	**s**tatistical **b**in **a**nalysis. Test limits are calculated per test bins by identifying the average yield plus Six Sigma. If test results from any production lot yielding fallouts in excess of the fallout limits in a specific test bin (yield outlier), there could be a specific cause defect in the lot which should be investigated (also see MAVERICK LOTS)	
SCHOTTKY DIODE	a diode fabricated with metal-to-semiconductor junction featuring low forward voltage drop, almost zero charge storage and close to unity noise to temperature ratio.	
SELF-INDUCTION	induction of an electromotive force in a circuit by a varying current in the same circuit	W
SEM	scanning electron microscpoy (an F/A tool)	
SEMICONDUCTOR	A crystalline solid material that in its pure form has a conductivity between a conductor and an insulator.	WKP

Term	Description/Definition	Source
SEMICONDUCTOR JUNCTION	see BARRIER LAYER	INF
SERIES PASS TRANSISTOR	the power control element in a LINEAR REGULATOR or a BUCK REGULATOR connected between the power source and the load.	
SI UNITS	Systeme International d'Unités (a.k.a. International Systems of Units or MKS [**M**eter{m} - **K**ilogram{kg} - **S**econd{sec}] system of units)	
SIX SIGMA	Six Sigma is a set of practices originally developed by Motorola to systematically improve processes by eliminating defects.	WKP
SLEW RATE	represents the maximum rate of change of signal at any point in a circuit	
SOFT START (switching regulator, motor driver)	gradual increase of the pulse width of SWITCHING REGULATORS or motor drivers to the required DUTY CYCLE at the time of the first turn-on (i.e. start-up)	
SOURCE	one of the electrodes of a FET; the current-supplying end of the conducting channel in an FET.	
STALL TORQUE	torque exerted by the motor on its forcibly immobilized shaft	
START-UP	bringing an electronic circuit (system) into operating mode from a switched-off state	
STEP DOWN SWITCHING VOLTAGE REGULATORS	See BUCK REGULATOR and SWITCHING REGULATOR	
STEPPER MOTOR	a brushless, synchronous electric motor that can divide a full rotation into a large number of steps	WKP
SUBSTRATE	inactive base (not to confuse with transistor base) material that is used as a carrier in semiconductor manufacturing, see also WAFER.	WKP
SUBSTRATE DIODE	parasitic diode formed between the source and drain of a MOSFET. Diode is reverse biased when MOSFET is forward biased	INF
SWITCHING LOSS	energy dissipated by an electronic swirtch during turning the power on or off to a load	

Term	Description/Definition	Source
SWITCHING REGULATOR	an electronic circuit that uses capacitors and inductors as energy storage elements with periodically switched supply source to create an output voltage of either higher or lower value than the source voltage. The output voltage is regulated by means of feedback and precision reference.	
SWITCHING REGULATOR EFFICIENCY	the ratio of output power to the input power	
SYNCHRONIZATION (of switching regulators)	determining the operating frequency of a SWITCHING REGULATOR or an OSCILLATOR by an external clock signal source	
TEM	transmission electron microscope (F/A tool)	
THERMAL CAPACITY	a measurable physical quantity that characterizes the ability of a body to store heat as it changes in temperature (SI unit: W x sec / K = joule / K). Symbol: C_Θ	WKP
THERMAL CONDUCTIVITY	a measure of how well temperature conducts through a substance (SI unit: W / (m x K)	WKP
THERMAL IMPEDANCE	$Z_{thk} = R_{thk} + 1 / (j\omega C_{thk})$ of the k-th layer (ratio of temperature change vs. step power input)	
THERMAL MASS	in the most general sense, it is any mass that absorbs and holds heat	WKP
THERMAL RESISTANCE	a measure of how poorly temperature conducts through a substance (SI unit: K / W)), symbol: R_{THxy} or $R_{\cup xy}$. "xy" designates the layers like in “ca” = case-to-ambiemt	WKP
THERMAL SHUT-DOWN	protection circuit in protected power devices that interrupt or terminate the operation of the component if its junction temperature exceeds a given limit	
THERMAL TIME CONSTANT	product of THERMAL RESISTANCE and THERMAL CAPACITY: $\tau_\cup = C_\Theta \times R_\cup$	
THRESHOLD VOLTAGE (V_T)	of a MOSFET is usually defined as the gate voltage where a depletion region forms in the substrate (body) of the transistor	WKP
TIME COSTANT	measure of time delay to transient changes. Capacitive-resistive circuit: $\tau = R \times C$, inductive-resistive circuit: $\tau = L / R$	

Term	Description/Definition	Source
TOROIDAL COIL	toroid is a doughnut shaped object. Toroidal coil: wire coiled around a toroid shape magnetic core (made of iron powder, laminated iron or non-permeable coil former) to make an INDUCTOR.	
TORQUE	T; force acting on an arm, cause of ANGULAR ACCELERATION of a MOMENT OF INERTIA; a measure of driving ability of rotating machines (SI unit: kg x m^2 / sec^2)	WKP
TRANSCONDUC-TANCE	reciprocal of resistance (SI unit: S = A / V = 1 / Ω. ["S" stands for Siemens, not to be confused with second].	
TRANSCONDUC-TANCE of MOSFETs	the ratio of a small change in drain current caused by a small change in gate-to-source voltage: $g_m = \Delta I_D / \Delta V_{GS}$ in the superthreshold region.	
USL	upper specification limit	
UTI	universal test instrument; may be bench type or hand-held	
VALENCE	valence, also known as *valency* or *valency number*, is a measure of the number of chemical bonds formed by the atoms of a given element	
VALENCE BOND	intra-melecular forces which hold atoms together in molecules	WKP
VECTOR	a mathematical concept characterized by a magnitude and a direction	WKP
VELOCITY	vectorial rate of change of position; (SI unit: m/sec). Magnitude of velocity vector is speed.	WKP
VISCOUS FRICTION	(torque) force load depending on (angular) velocity (SI unit: see TORQUE, FORCE)	
VOLTAGE	energy required to move a unit charge in an electric field (SI unit: V [volt])	
VOLTAGE GAIN	measure of the ability of a circuit to increase the VOLTAGE amplitude of a signal. VOLTAGE GAIN is a dimensionless number	WKP
VOLTAGE MODE CONTROL	a type of FEEDBACK CONTROL LOOP in which the output voltage of the loop is determined comparing it to a VOLTAGE REFERENCE	
VOLTAGE REFERENCE	the output voltage of a precision and stable voltage regulator	

Term	Description/Definition	Source
VOLTAGE REGULATOR	a voltage converter providing output power to the load at constant voltage regardless of the input voltage or output current (provided they are within specified limits)	
VOLTAGE SOURCE	source of electrical energy with low internal RESISTANCE (IMPEDANCE)	
VOLT-SECOND PRICIPLE	the sum of time x voltage area (or volt-seconds) across an energy storage inductor within one PWM cycle in a DC-DC converter is zero under steady state operating conditions	
WAFER	a disc of semiconductor material on which a number of identical dies (chips) are created	INF
WAFER FABRICATION	the making of semiconductor devices through multiple repetitive sequence of photographic and chemical process steps in which integrated circuit are gradually created on a pure single crystal silicon wafer	
WAFER SAWING	separating individual dies on the same wafer	
WATT	unit of electric power (SI unit: W = V x A)	
WORK DONE	see also ENERGY (SI unit: N x m, joule, W x sec)	
WORST CASE CONDITION	the operating condition of an electronic circuit (sytem) with internal and external parameters taken at values (magnitude AND sign) to produce the largest error from the desired performance	
ZERO DEFFECT (ZD)	The expectation of the user of the device that absolutely no rejects is shipped to them by the supplier during the life of the product.	

SOURCE (ABREVIATION)	**SOURCE**
EEH	ELECTRONICS ENGINEERS' HANDBOOK, D.G. FINK et al. Second edition 1985
INF	Infineon “Semiconductors” Second edition 2004 or training slides
W	Webster's Ninth New Collegiate Dictionary, 1983
WKP	WIKIPEDIA

Bridging Theory into Practice (BTiP) is a practical training methodology that is developed from the knowledge and many years of hands on application experience of Infineon engineers. In this second edition of *the Fundamentals of Power Semiconductors for Automotive Applications*, we build upon the basic BTiP concepts and offer our readers in-depth chapters that follow a step-by-step approach to the understanding of automotive power electronics.

Contents

• RLC Load Characteristics and Modeling • Introduction to Linear Regulators
• Introduction to Semiconductors • Introduction to Switching Regulators
• Transistors and Integrated Circuits • ESD and EOS • Introduction to Power Dissipation
• Semiconductor Manufacturing • Advanced Thermal Analysis
• Quality and Reliability • MOSFETs • Introduction to Motor Control • Protected High Side Drivers
• Permanent Magnet DC Motors • Protected Low Side Drivers • Glossary of terms

"Bridging Theory into Practice" refreshed our fundamental concepts. We were impressed with the simplicity with which the team presented very complex items."

Ford Motor Company - Electrical Electronics Systems Engineering

"A practical, usable guide for the analog electronics design engineer. Whether a neophyte or seasoned veteran in the analog power electronics field, this text identifies and explains real-world constraints and challenges that warrant serious considerations by the designer"

Kraig J. Olejniczak, Ph.D., P.E. - Dean of Engineering, Valparaiso University

Published by Infineon Technologies AG

Second Edition, 2008
Ordering No. B112-H9219-X-X-7600