HBD871/D
Rev. 1, Jul-2012

ON Semiconductor®

Enabling Energy Efficient Solutions

IGBT Applications Handbook

www.onsemi.com

THIS PAGE INTENTIONALLY LEFT BLANK

IGBT Applications

Handbook

HBD871/D
Rev. 1, Jul-2012

ON Semiconductor®

Grafoil is a registered Union Carbide.
Kon–Dux and Rubber–Duc are trademarks of Aavid Thermal Technologies, Inc.
Thermasil is a registered trademark and Thermafilm is a trademark of Thermalloy, Inc.
Kapton is a registered trademark of du Pont de Nemours & Co., Inc.
Sil–Pad is a registered trademark of the Bergquist Company.
CHO–THERM is a registered trademark of Chomerics, Inc.
FULLPAK, ICePAK and Thermopad are trademarks of Semiconductor Componenets Industries, LLC.

PUBLICATION ORDERING INFORMATION

LITERATURE FULFILLMENT:
Literature Distribution Center for ON Semiconductor
P.O. Box 5163, Denver, Colorado 80217 USA
Phone: 303-675-2175 or 800-344-3860 Toll Free USA/Canada
Fax: 303-675-2176 or 800-344-3867 Toll Free USA/Canada
Email: orderlit@onsemi.com

N. American Technical Support: 800-282-9855 Toll Free USA/Canada.
Europe, Middle East and Africa Technical Support:
Phone: 421 33 790 2910
Japan Customer Focus Center
Phone: 81-3-5773-3850

ON Semiconductor Website: **www.onsemi.com**

Order Literature: http://www.onsemi.com/orderlit

For additional information, please contact your local Sales Representative

Table of Contents

Introduction

The IGBT is a four layer, semiconductor device that combines the voltage characteristics of a bipolar transistor (collector – emitter) and the drive characteristics of a MOSFET. The concept was first documented in a Japanese patent by Yamagami, which was filed in 1968. The first devices were of a planar technology, but more recently vertical, trench devices have been made popular.

The popularity of the IGBT has soared in recent years due to an increase in high voltage, high power applications at which they excel. While the switching speeds are slower than a MOSFET, the $V_{CE(SAT)}$ characteristics are a significant improvement over those of a MOSFET at high currents, especially for high voltage devices. They are available in a range of voltage ratings from 300 to over 1200 volts and current ratings of 15 to 100 amps for a single die. IGBT modules have current ratings well into the 100s of amps.

The range of ratings of an IGBT make it well suited for high power applications such as:

- Electric vehicle motor drives
- Appliance motor drives
- Power factor correction converters
- Solar inverters
- Uninterruptable power supplies (UPS)
- Inductive heating cookers
- High frequency welders

This manual contains a collection of application notes that are available to help designers with the thermal, mechanical and electrical challenges of an IGBT based power converter. In addition, ON Semiconductor has a staff of skilled marketing and application engineers to assist customers with specific needs.

AND9070/D

Device Physics of the IGBT

ON Semiconductor®

http://onsemi.com

APPLICATION NOTE

INTRODUCTION

The Insulated Gate Bipolar Transistor (IGBT) evolved from the vertical power MOSFET in the mid 1980's. The simple addition of an extra P−N junction to the drain of the MOSFET changes this unipolar device into a Bipolar Junction Transistor. However, because of the MOS gate structure, this BJT is a voltage controlled device.

This combination of an insulated gate input and bipolar output makes the IGBT an excellent power switch for medium frequency (5 − 50 kHz) and high voltage (200−2000 V) applications. 600 V and 1200 V are the most widely used voltage platforms for white goods and industrial motor drive applications. Improvements in device design and semiconductor processing continue to increase the IGBT's frequency and reduce its conduction voltage drop, while maintaining high blocking voltage capability. The current carrying capability is only limited by the paralleling of large area IGBT's in multiple die modules.

The following sections discuss the design and operation of discrete vertical power IGBT's as large signal switches. This applications note emphasizes the blocking, conduction, and switching of the IGBT for high performance power conversion. Lastly, The Safe Operating Area and temperature effects of the IGBT will be examined.

DEVICE STRUCTURE

Vertical cross−sections of a planar n−channel IGBT are shown in Figure 1. Complementary P−channel IGBT's also exist, but have higher on−state losses and latch−up more easily than N−channel IGBT's. These cross−sections show a single IGBT cell, typically 2 to 10 μm wide, where the products are designed by integrating several million cells in a single silicon chip to provide 10's and 100's of amperes of current depending on the voltage rating.

Many features of IGBTs are adapted from power MOSFETs with high cell densities on the semiconductor die to achieve the desired current rating. But due to the device concept of IGBT with conductivity of modulation it can handle a much higher current density compared to a MOSFET. The high current density enabled by an IGBT allows 3X die size reduction for 600 V, and the advantage for the IGBT increases further as the voltage increases.

Not shown in Figure 1 are the high voltage termination and bonding areas for electrically and thermally connecting the IGBT to a three−terminal package. Termination is typically implemented by a number of P−type rings and field plates which are carefully designed and tailored in order to achieve the desired high voltage.

Figure 1a shows a representative cross section of conventional planar IGBT and its equivalent circuit. In this structure the silicon surface is planar and the channel regions which are created and controlled by the gate bias, form a lateral current path for conduction from the bottom p+ collector to the surface N+ emitter region. The channel is formed by applying positive gate−to−emitter voltage which inverts the surfaces of the P+ tubs from P−type to N−type, thereby creating conduction path for electrons from the n+ emitter, through the channel to the N−type epi.

A drawback for planar IGBTs is that current flow is constricted between the p+ tubs in what is known as the JFET region. Even though techniques are used to increase carrier concentration in the JFET, this region continues to pose a limitation to planar gated devices. The channel operation of an IGBT is the same as the MOSFET conduction, but since an IGBT has a P+ backside collector, the channel current serves as the base current to activate a PNP bipolar transistor. Since a BJT is a conductivity−modulated device the voltage drop (V_{CEsat}) in the IGBT is significantly reduced. Thus the secret of an efficient IGBT is combining the voltage controlled MOS gate with high input resistance, and a low V_{CEsat} bipolar transistor.

As seen in the vertical cross section, the IGBT is composed of a four layer NPNP semiconductor. It is very important to suppress this parasitic NPNP thyristor device by controlling the gains of the interlocked BJTs forming the thyristor. The parasitic NPN transistor is designed to be inactive, as its emitter−base junction is shorted out by the MOSFET source metal. Therefore the basic IGBT is a vertical wide−base PNP transistor, with its base drive provided by the surface MOSFET. The simple four layer device of Figure 1a has several drawbacks affecting switching and SOA.

To overcome these limitations several techniques are devised in modern IGBTs to modify the vertical structures as IGBT design has progressed over the past three decades.

The Punch−Through (PT) IGBT, shown in Figure 1a, implemented by epitaxy, has been predominantly used in lower voltages below 600V for good breakdown voltage and low V_{CEsat}. A PT−IGBT has an extra N+ region (buffer layer) epitaxially grown on the thick P+ substrate prior to the N− epitaxy. During the off−state, the blocking voltage (BV_{ces}) completely depletes the N− region, "punching−thru" to the N+ layer. The additional N+ buffer layer prevents "reach−thru" (voltage collapse from collector to emitter) to the P+ substrate, but allows a relatively thin N− region which minimizes $V_{CE(on)}$ while maximizing $V_{(BR)CES}$. The N+ region also improves the switching speed of the IGBT by reducing excess hole injection from the P+ substrate. The PT−IGBT also has a heavy P+ tub in the emitter region, which improves the V_{BE} shorting of the parasitic NPN, for improved SOA.

PT−IGBT's are typically available in 300 volt to 1200 V $V_{(BR)CES}$ rated devices. Nonetheless, planar IGBTs based on PT epitaxy have disadvantages. It is expensive and difficult to grow defect free epitaxy more than 100 μm thick which is required for IGBTs of 1200 V and above. More seriously epi is grown on heavily doped substrates to reduce V_{CEsat}, but the high doping level injects large concentrations of minority carriers (PNP holes) which take considerable time to be removed during IGBT turn off. As a result, large switching losses occur, limiting the high frequency operation of these IGBTs to only few kHz.

To circumvent this problem, a Non−Punch−Thru (NPT) IGBT as shown in Figure 1b is utilized. NPT−IGBT's have a thicker N− region which is fabricated using N− Float Zone wafers and is wide enough to prevent high voltage punch−thru. A shallow P+ collector is implanted on the backside of the wafer. Precise control of this P+ implant dose and junction depth results in reduced PNP hole injection compared to the PT−IGBT. This precise hole distribution results in a fast switching NPT−IGBT without having to use the lifetime killing processes common in PT−IGBT's. With this improvement in technology, the Float Zone wafer can be thinned to less than 100 μm and back phosphorus implantation is used to form the N+ region to block the electrical field reaching the P+ collector under the high voltage condition resulting in the field stop IGBT, which improves the DC and switching performance. In Figure 1a, the IGBT has a planar−gate structure.

The performance of IGBTs has been improved significantly by using a trench−gate structure similar to that used in trench MOSFETs. Trench gates provide higher channel density (small cell size) and reduce the JFET resistance in the IGBT structure. In conventional planar IGBTs, current crowding and electric field curvature in the small space between p+ tubs does not allow for sufficient conductivity modulation. This effect is alleviated in trench IGBTs due the long distance between p+ tubs. The free carrier concentration in the N−drift region near the emitter is also enhanced, leading to a lower V_{CEsat}. Trench structure also provides a more robust IGBT as it more effectively suppresses the latch−up of the NPNP four layer device. This is due to the fact that the n+ emitter (similar to the MOSFET source) can be made narrower and the p+ body can be more heavily doped. Suppression of NPNP latch−up and control of IGBT saturation current are used in IGBTs to provide products with excellent short circuit withstand capability and robust FBSOA.

In addition to trench gate and engineered cell architecture, advanced IGBTs are dramatically more enhanced by the field stop which is similar to the buffer layer concept presented earlier, but implemented in thin Float Zone substrates, as shown in Fig 1c. Using a thin wafer for a non−punch−through (NPT) IGBT and adding a field stop region enables several improvements. The term field refers to stopping the electric field by the n buffer layer, allowing a thinner wafer to be used for the same high breakdown voltage. Also, controlling the carrier concentration of both the field stop and p+ collector layers, controls the emitter efficiency of the backside junction. Thus, field stop technology provides faster switching and lower V_{CEsat} due to a thinner wafer, and adjustment of these parameters by carrier concentration control.

Freewheeling diodes are co−packaged with IGBTs in inductive load topologies. These high voltage IBGTs are designed with fast switching, low charge and soft recovery characteristics. Recent developments have been successful in the monolithic integration of rectifiers with IGBTs, allowing reverse conduction and much more efficient utilization of silicon real estate. This combination of the IGBT & diode has been enabled by modern field stop techniques. ON Semiconductor products employ the most advanced and proprietary techniques of field stop trench IGBTs providing compelling solutions to a variety of high performance power conversion applications, such as motor drive, renewable energy and white goods applications.

A few final comments comparing these structures follow. The basic IGBT and NPT−IGBT are nearly symmetrical blocking devices with $V_{(BR)CES}$ $V_{(BR)ECS}$. However, the reverse leakage IECS is relatively high (mA range), as the backside P−N junction is sawn through and un−passivated. The PT−IGBT on the other hand is an asymmetric blocking device with $V_{(BR)ECS} << V_{(BR)CES}$ due to the presence of the N+ buffer layer. But like the others, it also has a leaky IECS. Thus, third quadrant operation of most IGBT's is limited to low reverse voltage.

OFF_STATE CHARACTERISTICS

For normal first quadrant operation of a low−side N−channel IGBT, the emitter is grounded, the collector is connected to $+V_{CC}$ through some impedance and the gate V_{GE} controls the collector current. IGBT's are enhancement mode devices, that is, they are normally off for $V_{GE} < V_{th} \approx$ 5 V. For V_{GE} = 0 V, only a small (nA level) leakage current I_{CES} flows through the collector−emitter terminals until V_{CE} reaches the avalanche voltage $V_{(BR)CES}$. The upper P−tub/N− junction supports this forward blocking voltage.

Usually the N- region's thickness and resistivity determine the $V_{(BR)CES}$ breakdown voltage. However, as discussed in the next section, the on-state $V_{CE(off)}$ also increases with N-thickness, so IGBT's are designed for the minimum N-thickness necessary for the $V_{(BR)CES}$ rating. For the 1200 V NPT-IGBT, this results in a wafer thickness of about 0.007 inch.

Besides the N- region, the device designer must consider several other factors that can affect $V_{(BR)CES}$. These include the MOSFET polysilicon gate width (voltage breakdown between the cells), MOSFET P-tub width, depth, and concentration (curvature and punch-thru effects), and the high voltage planar edge termination and passivation. The termination design and passivation process are critical to the long term reliability of any high voltage semiconductor. Likewise for trench IGBTs, trench depth, trench spacing and p+tub concentration are critical for optimized breakdown voltage, along with optimized termination design.

ON-STATE CHARACTERISTICS

When $V_{GE} > V_{th}$, the MOSFET channel is inverted and electrons can flow from the N+ source (IGBT emitter) to the N- drain region, if $V_{CE} >$ 0.8 V. This $V_{CE(offset)}$ 0.8 V must be exceeded to forward bias the backside P-N junction and permit conduction. The MOSFET current $I_D = I_B$, the base current of the vertical PNP. For large V_{CE}, the collector current of the IGBT saturates due to the transconductance of the MOSFET (g_m) and the current gain of the PNP (beta) according to the equation:

$$I_{C(sat)} = g_m(\beta + 1)\left(V_{GE} - V_{th}\right)^2$$

Like any MOSFET, g_m is a function of the gate oxide capacitance, electron mobility, and the channel width/length ratio. Being a wide base BJT, the beta of the PNP is strongly dependant on the N- region minority carrier lifetime, the N+ buffer layer profile (for a PT-IGBT), and the backside P+ doping profile (which internally is the emitter of the PNP). Carrier lifetime is customarily reduced during manufacturing to improve the switching speed of the IGBT. This lowers the PNP beta, and increases the linear region $V_{CE(on)}$.

For hard switching applications, the linear region $V_{CE(on)}$ is a critical parameter to enable fast switching, as it determines the on-state power dissipation, and is a direct trade-off with the switching losses mentioned above. The $V_{CE(on)}$ in the linear region has three components:

$$V_{CE(on)} = V_{CE(offset)} + V_{(N-region)} + V_{DS(surface\ MOSFET)}$$

Provided that a good ohmic contact exists to the collector P+, the slope of the $V_{CE(on)}$ linear region is controlled primarily by the latter two terms. The resistance of the N- region is conductivity modulated by the bipolar action (electron and hole conduction) of the IGBT. This increased conductivity results in significantly lower VCE(on) at a given current density compared to the unipolar MOSFET of equal voltage rating. At I_C = 100 A/cm2, (typically 20 A in a TO-220 package), a fast IGBT's VCE(on) is roughly one third the $V_{DS(on)}$ of a MOSFET at I_D = 100 A/cm2.

Such improvement in on-state voltage drop is not without consequence. The excess hole carriers in the N- region must be injected and removed during turn-on and turn-off, respectively. Thus the IGBT has slower switching capability than the MOSFET, as discussed in the next section. The IGBT switching speed/VCE(on) trade-off can be tailored for the operating frequency to minimize total system dissipation. For PT-IGBT's this is usually done by adjusting the final carrier lifetime. For NPT-IGBT's, the backside P+ profile is adjusted to control the excess hole distribution, so that carrier lifetime killing is usually not required. Using these methods, switching fall times of 100 to 300 nsec can be achieved. Lastly, the third term (V_{DS}), like g_m, is primarily determined by the packing density of the MOSFET cells. A high channel width/length ratio reduces total channel resistance. Also the drain region under the gate is not highly conductivity modulated, especially for planar IGBT, so that the polysilicon gate width must be optimized for $V_{CE(on)}$ as well as $V_{(BR)CES}$.

SWITCHING CHARACTERISTICS

IGBT's are often used in PWM inverters with freewheeling diodes commutating the load current between the power switches. The turn-on switching loss (Eon) is dominated by the voltage fall time (t_{fv}) of the V_{CE} waveform, which exponentially falls to the on-state $V_{CE(on)}$. The t_{fv} exponential time constant is proportional to the minority carrier lifetime and beta of the PNP. The lifetime can be reduced by heavy metal doping or by high energy particle bombardment creating recombination centers in the silicon lattice. The PNP beta can be reduced by increasing the N+ buffer concentration (PT-IGBT) or decreasing the P+ backside effective emitter doping (NPT-IGBT). Use of both lifetime killing and beta reduction, or other novel recombination structures now in development, are expected to push IGBT fall times to the 50 nsec range.

Gate capacitance and gate charge play an important role in turn-on switching characteristics. It is critical to reduce input capacitance and charge to minimize delay times, rise times and drive current for IGBT turn on. In addition, the capacitance and gate charge associated with gate-to-collector (or Miller) capacitance is not only very influential in dynamic performance of MOS-gated devices, but to device robustness and operation stability. The turn-off switching loss (E_{OFF}) is dominated similarly by the IGBT's current fall time (t_{fi}), which is also proportional to the carrier lifetime and PNP beta. As lifetime and beta are reduced to improve switching losses, the on-state $V_{CE(on)}$ increases due to the reduced conductivity of the N region. As mentioned before, the device designer must optimize the IGBT's speed/$V_{CE(on)}$ trade-off for the operating frequency of the application. During switching, or during fault

conditions, high IC and V_{CE} are present simultaneously, which leads to the next topic.

SOA. The Safe Operating Area (SOA) is defined as the loci of points where the load line (V_{CE} and I_C) may safely traverse the output characteristics of a device at a specified junction temperature and pulse width. BJT's specify two types of SOA: Forward Biased (FBSOA) and Reverse Biased (RBSOA). Since the IGBT base region is floating and cannot be reversed biased, and since the PNP beta is normally so low that PNP $BV_{CEO} \approx BV_{CES}$, the IGBT's FBSOA and RBSOA are similar, although RBSOA is slightly less as explained below. IGBT SOA capability is a function of many device design parameters, including lifetime, beta, P-tub doping profile, and MOSFET cell geometry. SOA failures can occur via two distinct mechanisms. The first mechanism is a current induced failure, due to excessive hole carriers in the P-tub region, which internally biases the parasitic NPN, causing latch-up of the parasitic four layer NPNP thyristor. FBSOA is normally a current induced failure. The second mechanism is an electric field induced failure, due to the excessive charge distribution narrowing the depletion region in the N- region, which causes avalanche injection and voltage collapse.

During RBSOA, the electron (MOSFET) current turns off first, leaving an excess of holes in the N- region. These holes add to the donor charge, effectively reducing the N- resistivity and breakdown voltage. RBSOA is normally an electric field induced failure. Both FBSOA and RBSOA are affected by the junction temperature.

A third SOA rating for IGBTs is Short Circuit (SCSOA). During fault conditions such as a shorted load, the IGBT must survive until the protection circuitry detects the fault and shuts down the system. The device must be designed to survive three modes of SCSOA; turning on into the short (mode A), surviving the high I_C and V_{CE} conduction period (mode B), and then turning off safely (mode C). The failure mechanism in each mode is as follows:

- Failure Mode A is FBSOA latchup induced.
- Failure Mode B is simply power dissipation limited.
- Failure Mode C is RBSOA avalanche induced.

The standard specification for SCSOA is surviving for 10 μ sec at a starting T_J = 150°C, non-repetitively (since T_J will exceed 150°C). Special IGBT cell design featuring hole-bypass or ballast structures is necessary to survive these severe SCSOA conditions. The fourth and final SOA condition is Unclamped Inductive Switching (UIS). IGBT's are much weaker in UIS than power MOSFETs, due to the hole current present in the IGBT. Because of this, IGBT's are usually clamped well below $V_{(BR)CES}$, but high dv/dt excursions can cause momentary spikes and UIS failure. One solution to this problem is to add series Rg to the gate drive circuit, limiting the dv/dt. "Smart" IGBT's are also becoming commercially available with current limiting and temperature sensing, providing SOA protection on chip.

TEMPERATURE CHARACTERISTICS

IGBT's with a maximum junction temperature rating of 150°C are typically operated with an average T_J of 75 to 125°C. Therefore the temperature dependence of the on-state and switching characteristics is of primary concern in system design. For this discussion, it is assumed that the device designer has optimized the IGBT for safe operation up to 150°C within the SOA ratings. It remains to describe the effect of temperature on $V_{CE(on)}$, E_{ON}, and E_{OFF}.

As discussed before, the $V_{CE(on)}$ is composed of the $V_{CE(offset)}$, the V(N- region), and the $V_{DS(MOSFET)}$. These three components are affected by several semiconductor parameters that are temperature dependent.

PARAMETER TEMP EFFECT (75°C to 125°C)

- intrinsic carrier concentration increases
- minority carrier lifetime increases
- carrier mobility decreases

$V_{CE(OFFSET)}$ decreases as the intrinsic carrier concentration increases. V(N-region) varies in a complex way, because the lifetime increase lowers V(N-), while the mobility decrease with a raise in V(N-). For PT-IGBT's with lifetime killing, the lifetime effect dominates and V(N-) decreases with T_J. For NPT-IGBT's without lifetime killing, the mobility effect dominates and V(N-) increases with T_J. $V_{DS(MOSFET)}$ also varies in a complex way as V_{th} is also temperature dependent.

For most applications $V_{GE} >> V_{th}$ (high I_C), so then V_{DS} is dominated by the decrease in mobility and V_{DS} increases with T_J. The combined temperature effect of these three components is as follows. The $V_{CE(on)}$ of PT-IGBT's has a slightly negative temperature coefficient. The $V_{CE(on)}$ of NPT-IGBT's has a slightly positive temperature coefficient. This makes the NPT-IGBT the better choice when paralleling IGBT's, as their positive temperature coefficient improves current sharing.

Both E_{on} and E_{off} increase with temperature, as the lifetime increase causes the IGBT fall time to increase, due to both slower carrier recombination and higher PNP beta. The NPT-IGBT fall time has a higher thermal coefficient than the PT-IGBT, as the beta increase causes excess carriers which cannot easily recombine in the higher lifetime NPT-IGBTs.

CONCLUSION

The IGBT has become the power switch of choice in medium and high power applications. The combination of speed, on-voltage, wide SOA, ease of drive, and relatively small temperature coefficients give the IGBT significant advantages over MOSFET's and BJT's. The growth of IGBTs is rapidly expanding due the massive adoption of variable speed (frequency) motor drives, white goods energy efficiency, the proliferation of renewable inverter applications and advent of electric vehicles.

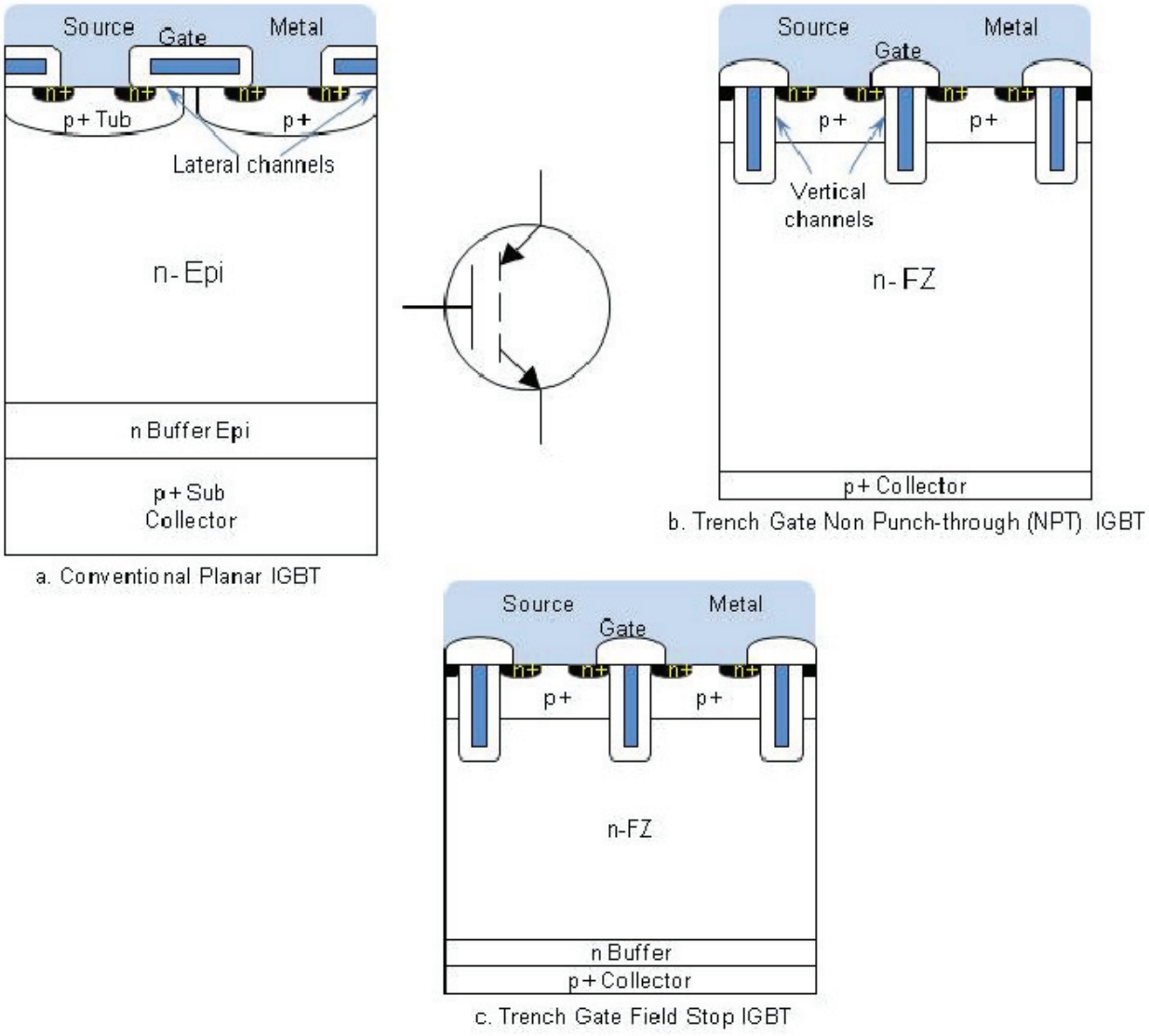


Figure 1. Insulated Gate Bipolar Transistor (IGBT) Structures

Reliability and Quality for IGBTs

ON Semiconductor®

http://onsemi.com

APPLICATION NOTE

INTRODUCTION

In today's semiconductor marketplace two important elements for the success of a company are product quality and reliability. Both are interrelated – reliability is the quality extended over the expected life of the product. For any manufacturer to remain in business, their products must meet and/or exceed the basic quality and reliability standards. ON Semiconductor, as a semiconductor supplier, has successfully achieved these standards by supplying product for the most strenuous applications to perform in the most adverse environments.

It is recognized that the best way to accomplish an assured quality performance is by moving away from the previous methods of "testing in" quality and embracing the newer concept of "designing in" quality. At ON Semiconductor, we use a two−fold approach toward reaching the ultimately achievable level of quality and reliability. First, we develop and implement a process that is inherently reliable. Then we exercise meticulous care in adhering to the specifications of the process every step of the way – from start to finish. This allows the development and application of inspections and procedures that will uncover potentially hidden failure modes. It is this dedication to long−term reliability that will ultimately lead to the manufacture of the "perfect product."

ON Semiconductor approaches the ideal in IGBT product reliability by instigating a four−step program of quality and reliability:

1. Stringent in−process controls and inspections.
2. Thoroughly evaluated designs and materials.
3. Process average testing, including 100% QA redundant testing.
4. Ongoing reliability verifications through audits and reliability studies.

These quality and reliability procedures, coupled with rigorous incoming inspections and outgoing quality control inspections add up to a product with quality built in – from raw silicon to delivered service.

RELIABILITY TESTS

ON Semiconductor IGBT's are subjected to a series of extensive reliability tests to verify conformance. These tests are designed to accelerate the failure mechanisms encountered in practical applications, thereby ensuring satisfactory reliable performance in "real world" applications.

The following describes the reliability tests that are routinely performed on ON Semiconductor's IGBT's.

High Temperature Reverse Bias (HTRB):

The HTRB test is designed to check the stability of the device under "reverse bias" conditions of the main blocking junction at high temperature, as a function of time.

The stability and leakage current over a period of time, for a given temperature and voltage applied across the junction, is indicative of junction surface stability. It is therefore a good indicator of device quality and reliability.

For IGBT's, voltage is applied between the collector and emitter with the gate shorted to the emitter. I_{CES}, $V_{(BR)CES}$, I_{GES}, $V_{GE(th)}$, and $V_{CE(on)}$ are the dc parameters monitored. A failure will occur when the leakage achieves such a high level that the power dissipation causes the devices to go into a thermal runaway. The leakage current of a stable device should remain relatively constant, only increasing slightly over the testing period.

Typical conditions:

V_{CE} = 80% – 100% of maximum rating
V_{GE} = 0 V (shorted)
T_A = 150C or Tj maximum
Duration: 1000 hrs for qualification

High Temperature Gate Bias (HTGB):

The HTGB test is designed to electrically stress the gate oxide at the maximum rated dc bias voltage at high temperature. The test is designed to detect for drift caused by random oxide defects and ionic oxide contamination.

For IGBTs, voltage is applied between the gate and emitter with the collector shorted to the emitter. I_{GES}, $V_{GE(th)}$, and $V_{CE(on)}$ are the dc parameters monitored. Any oxide defects will lead to early device failures.

Typical conditions:

V_{GE} = ±20 V or 100% rated Vge
V_{CE} = 0 (shorted)

T_J = 150°C or T_J maximum
Duration: 1000 hrs for qualification

High Temperature Storage Life (HTSL) Test:

The HTSL test is designed to indicate the stability of the devices, their potential to withstand high temperatures and the internal manufacturing integrity of the package. Although devices are not exposed to such extreme high temperatures in the field, the purpose of this test is to accelerate any failure mechanisms that could occur during long periods at storage temperatures.

The test is performed by placing the devices in a mesh basket, then placed in a high temperature chamber at a controlled ambient temperature, as a function of time.

Typical conditions:

T_A = 150°C on Plastic package
Duration: 1000 hrs for qualification

High Humidity High Temperature Reverse Bias (H^3TRB)

The H^3TRB test is designed to determine the resistance of component parts and constituent materials to the combined deteriorative effects of prolonged operation in a high temperature/high humidity environment. This test only applies to nonhermetic devices.

Humidity has been a traditional enemy of semiconductors, particularly plastic packaged devices. Most moisture related degradations result, directly or indirectly, from penetration of moisture vapor through passivating materials, and from surface corrosion. At ON Semiconductor, this former problem has been effectively addressed and controlled through use of junction “passivation” process, die coating, and proper selection of package materials.

Typical conditions:

V_{CE} = 80% − 100% of maximum rating
V_{GE} = 0 (shorted)
T_A = 85°C
RH = 85%
Duration: 1000 hrs for qualification

Typical conditions:

$V_{GE} \geq 10$ V
ΔT_J = 100°C
$R_{\theta JC}$ = Device dependent
T_{on}, $T_{off} \geq 30$ seconds
Duration: 10 − 15k cycles for qualification

Unbiased Highly Accelerated Stress Test (UHAST)

The UHAST is designed to determine the moisture resistance of devices by subjecting them to high steam pressure levels. This test is only performed on plastic/epoxy encapsulated devices and not on hermetic packages (i.e., metal can devices). Within the chamber a tray is constructed inside to keep the devices approximately two inches above the surface of deionized water and to prevent condensed water from collecting on them. After achieving the proper temperature and atmospheric pressure, these test conditions are maintained for a minimum of 24 hours. The devices are then removed and air dried. Parameters that are usually monitored are leakage currents and voltage.

Typical conditions:

T_A = 131°C
P = 14.7 psi
RH = 100%
Duration: 72 hrs for qualification

Intermittent Operating Life:

The purpose of the IOL test is to determine the integrity of the chip and/or package assembly by cycling on (device thermally heated due to power dissipation) and cycling off (device thermally cooling due to removal of power applied) as is normally experienced in a “real world” environment.

DC power is applied to the device until the desired function temperature is reached. The power is then switched off, and forced air cooling applied until the junction temperature decreases to ambient temperature.

$$\Delta T = \Delta T_C + R_{\theta JC} P_D$$

$$\Delta T_J = 100^\circ C$$

(typically, which is an accelerated condition)

$$\Delta T_C = T_{C\,HIGH} - T_{C\,LOW}$$

The sequence is repeated for the specified number of cycles. The temperature excursion is carefully maintained for repeatability of results.

The Intermittent Operating Life test indicates the degree of thermal fatigue of the die bond interface between the chip and the mounting surface and between the chip and the wire bond interface.

For IGBT’s, parameters used to monitor performance are thermal resistance, threshold voltage, on−resistance, gate−emitter leakage current and collector−emitter leakage current.

A failure occurs when thermal fatigue causes the thermal resistance or the on−resistance to increase beyond the maximum value specified by the manufacturer’s data sheets.

Temperature Cycle (TC):

The purpose of the Temperature Cycle Test is to determine the resistance of the device to high and low temperature excursions in an air medium and the effects of cycling at these extremes.

The test is performed by placing the devices alternatively in separate chambers set for high and low temperatures. The air temperature of each chamber is evenly maintained by means of circulation. The chambers have sufficient thermal capacity so that the specified ambient is reached after the devices have been transferred to the chamber.

Each cycle consists of an exposure to one extreme temperature for 15 minutes minimum, then immediately transferred to the other extreme temperature for 15 minutes minimum; this completes one cycle. Note that it is an immediate transfer between temperature extremes and thereby stressing the device greater than non−immediate transfer.

Typical Extremes

−65/ + 150° C

The number of cycles can be correlated to the severity of the expected environment. It is commonly accepted in the industry that ten cycles is sufficient to determine the quality of the device. Temperature cycling identifies any excessive strains set up between materials within the device due to differences in coefficients of expansion.

Low Temperature Storage Life (LTSL) Test:

The LTSL test is designed to indicate the stability of the devices, their potential to withstand low temperatures and the internal manufacturing integrity of the package. Although devices are not exposed to such extreme low temperatures in the field, the purpose of this test is to accelerate any failure mechanisms that could occur during long periods at storage temperatures.

Typical conditions:

T_A = −65°C on Plastic package

Duration: 1000 hrs for qualification

The test is performed by placing the devices in a mesh basket, then placed in a high temperature chamber at a controlled ambient temperature, as a function of time.

Steady State Operational Life (SSOL) Test

The SSOL test is designed to indicate the integrity of the chip and/or package assembly at steady−state continuous operational life conditions.

For IGBT's, parameters used to monitor performance are thermal resistance, threshold voltage, on−resistance, gate−emitter leakage current and collector−emitter leakage current.

Typical conditions:

$V_{GE} \geq 10$ V

$\Delta T_J = 100°C$

$T_A = 25°C$ Duration: 1000 hours for qualification.

A failure occurs when thermal fatigue causes the thermal resistance or the on−resistance to increase beyond the maximum value specified by the manufacturer's data sheet

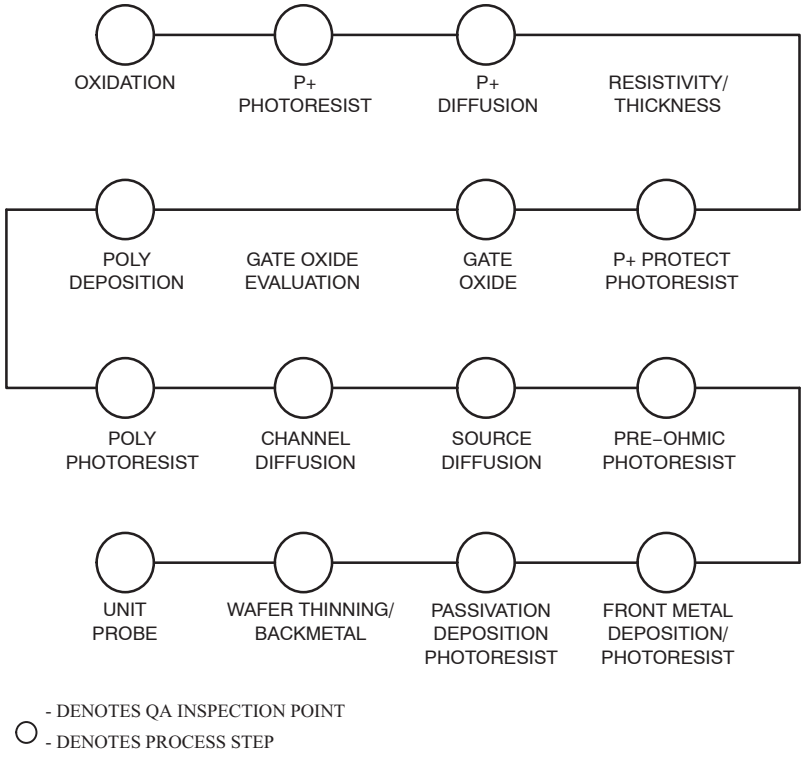
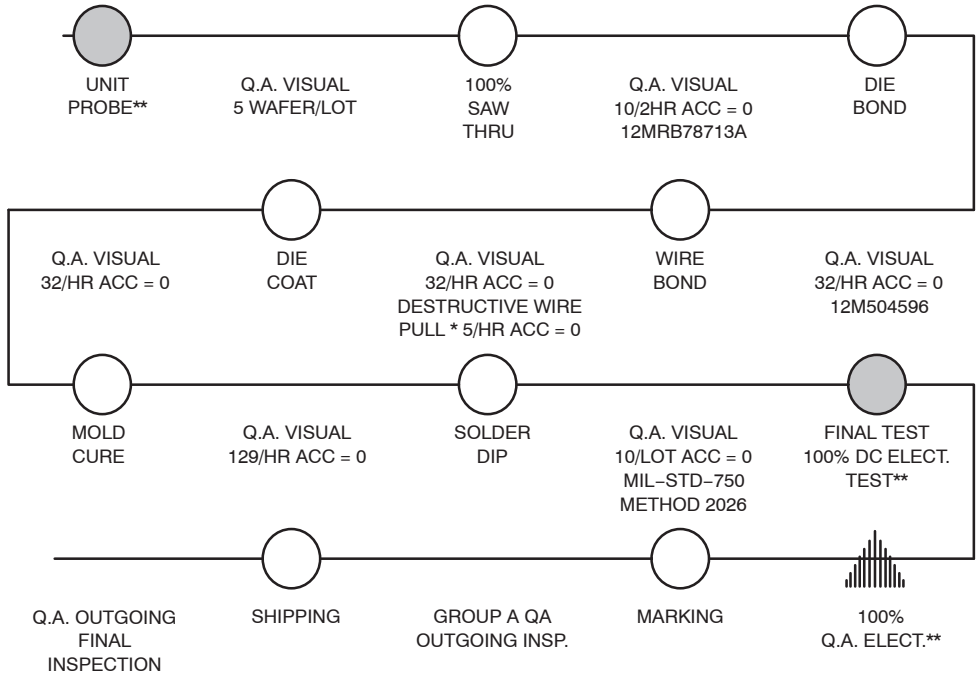


Figure 1. IGBT Wafer Fabrication



*100% NON−DESTRUCTIVE WIRE PULL AFTER WIRE BOND
**100% DC ELECTRICAL TESTING

– DENOTES QA INSPECTION POINT
○ – DENOTES PROCESS STEP

Figure 2. Assembly Process Flow

Environmental Package Related Test Programs:

A.. Physical Dimensions − This test is performed to determine the conformance to device outline drawing specifications.

B.. Visual and mechanical examination − A test to determine the acceptability of product to certain cosmetic and functional criteria such as marking legibility, stains, etc.

C.. Resistance to Solvents − A test to determine the solderability of device terminals.

D.. Terminal Strength − This test is a lead bend test to check for lead strength.

Every manufacturing process exhibits a quality and reliability distribution. This distribution must be controlled to assure a high mean value, a narrow range and a consistent shape. Through proper design and process control this can be accomplished, thereby reducing the task of screening programs which attempt to eliminate the lower tail of the distribution.

Accelerated Stress Testing

The nature of some tests in this report is such that they far exceed that which the devices would see in normal operating conditions. Thus, the test conditions "accelerate" the failure mechanisms in question and allow ON Semiconductor to predict failure rates in a much shorter amount of time than otherwise possible. Failure modes that are temperature dependent are characterized by the Arrhenius model.

$$AF = \frac{EA}{K}\left(\frac{1}{T2} - \frac{1}{T1}\right)$$

AF = Acceleration Factor
EA = Activation Energy (eV)
K = Boltzman's Constant (8.62 x 10E−5 eV/K)
T2 = Operating Temperature, K
T1 = Test Temperature, K

Therefore, the equivalent device hours are equal to the acceleration factor (as determined by the Arrhenius Model) times the actual device hours.

Review of Data

High Temperature Reverse Bias (HTRB) indicates the stability of leakage current, which is related to the field distortion of IGBT's. HTRB enhances the failure mechanism by high temperature reverse bias testing, and therefore is a good indicator of device quality and reliability, along with verification that process controls are effective.

High Temperature Gate Bias (HTGB) checks the stability of the device under "gate bias" forward conditions at accelerated high temperature, as a function of time. This test is performed to electrically stress the gate oxide to detect for drift caused by random oxide defect. This failure mechanism appears in the infant and random zones of the reliability "bath tub curve" at a very low rate of defect.

Intermittent Operating Life (IOL) is an excellent accelerated stress test to determine the integrity of the chip

and/or package assembly to cycling on (device thermally heated due to power dissipation) and cycling off (device thermally cooling due to removal of power applied). This test is perhaps the most important test of all, along with simulating what is normally experienced in a “real world” environment. IOL exercises die bond, wire bonds, turning on the device, turning off the device, relates the device performance, and verifying the thermal expansion of all materials are compatible. ON Semiconductor performs extensive IOL testing as a continual process control monitor that test relates to the “device system**” as a whole. ON Semiconductor also performs extensive analysis and comparison of delta function temperatures. ON Semiconductor has determined that to effectively stress the device a delta T_J of 100°C is necessary which far exceeds many customers’ application and determines the reliability modeling of the device.

Temperature Cycling (TC) is also an excellent stress test to determine the resistance of the device to high and low temperature excursions in an air medium. Where IOL electrically stresses the “device system” from internally, temperature cycle stresses the “device system” thermally from external environment conditions.

High Temperature Storage Life (HTSL), High Humidity Temperature Reverse Bias (H^3TRB), Thermal Shock (TC) and “Pressure Cooker” (Autoclave) are routinely tested, however it is felt by ON Semiconductor Reliability Engineering that HTRB, HTGB, IOL and TC are of primary importance. ON Semiconductor has been in the semiconductor industry for many years and will remain there as a leader with continued reliability, quality and customer relations.

RELIABILITY AUDIT PROGRAM

At ON Semiconductor reliability is assured through the rigid implementation of a reliability audit program. All IGBT products are grouped into generic families according to process technology and package types. These families are sampled quarterly from the raw stock at final test, then submitted for audit testing. The extreme stress testing, in real−time for each product run, may uncover process abnormalities that are detectable by the in−process controls. Typical reliability audit tests include high temperature reverse bias, high temperature gate bias, intermittent operating life, temperature cycling, and autoclave. To uncover any hidden failure modes, the reliability tests are designed to exceed the testing conditions of normal quality and reliability testing.

Audit failures which are detected are sent to the product analysis laboratory for real−time evaluations. This highly specialized area is equipped with a variety of analytical capabilities, including electrical characterizations, wet chemical and plasma techniques, metallurgical cross−sectioning, scanning electron microscope, dispersive x−ray, auger spectroscopy, and micro/macro photography. Together, these capabilities allow the prompt and accurate analysis of failure mechanisms – ensuring that the results of the evaluations can be translated into corrective actions and directed to the appropriate areas of responsibility.

The ON Semiconductor reliability audit program provides a powerful method for uncovering even the slightest hint of potential process anomalies in the IGBT product line. It is this stringent and continuing concern with the reliability audits that gives positive assurance that customer satisfaction will be achieved.

IGBT RELIABILITY AUDIT PROGRAM

Test	Conditions	S/S	Frequency
HTRB	V_{CE} = 80 −100% Max Rating V_{GE} = 0 V T_J = 150°C Duration = 168 Hours (short), 1000 Hours (long)	77pcs	Qty
HTGB	V_{GES} = ± 20 V V_{CE} = 0 V T_J = 150°C Duration = 168 Hours (short), 1000 Hours (long)	77pcs	Qty
IOL	D T_J = 100°C $V_{CE} \geq$ 10 V Duration = 5000 Cycles (short), 15,000 cycles (long)	77pcs	Qty
Solder Heat	1 Cycle @ 260°C for 10 seconds followed by:	77pcs	Qty
Temperature Cycle	100 Cycles (short) 500 Cycles (long) −65 to +150°C Dwell Time ≥ 15 minutes	77pcs	Qty
Pressure Cooker	P = 15 psi, T = 121°C Duration = 48 Hours (short), 96 Hours (long) (Plastic Package Only)	77pcs	Qty
HTSL	T_A = 150°C, 168 hrs	77pc	Qty

Essentials of Reliability:

Paramount in the mind of every semiconductor user is the question of device performance versus time. After the applicability of a particular device has been established, its effectiveness depends on the length of trouble free service it can offer. The reliability of a device is exactly that – an expression of how well it will serve the customer. Reliability can be redefined as the probability of failure free performance, under a given manufacturer's specifications, for a given period of time. The failure rate of semiconductors in general, when plotted versus a long period of time, exhibit what has been called the "bath tub curve".

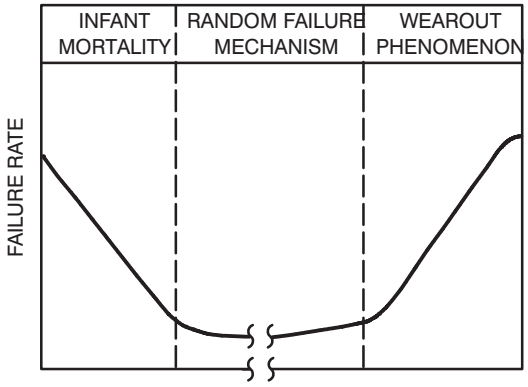


Figure 3. Failure Rate of Semiconductor

Reliability Mechanics

Since reliability evaluations usually involve only samples of an entire population of devices, the concept of the central limit theorem applies and a failure rate is calculated using the λ^2 distribution through the equation:

$$\lambda 3 \frac{\lambda 2(\alpha, 2r + 2)}{2nt}$$

where

$$\alpha = \frac{100 - cl}{100}$$

The confidence limit is the degree of conservatism desired in the calculation. The central limit theorem states that the values of any sample of units out of a large population will produce a normal distribution. A 50% confidence limit is termed the best estimate, and is the mean of this distribution. A 90% confidence limit is a very conservative value and results in a higher λ which represents the point at which 90% of the area of the distribution is to the left of that value.

The term (2r + 2) is called the degrees of freedom and is an expression of the number of rejects in a form suitable to λ^2 tables. The number of rejects is a critical factor since the definition of rejects often differs between manufacturers. Due to the increasing chance of a test not being representative of the entire population as sample size and

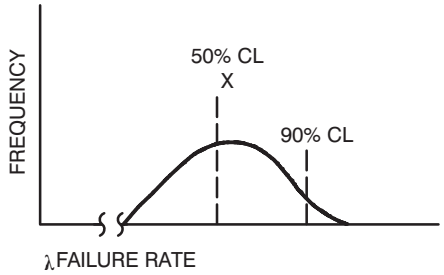


Figure 4. Confidence Limits and the Distribution of Sample Failure Rates

test time are decreased, the λ^2 calculation produces surprisingly high values of λ for short test durations even though the true long term failure rate may be quite low. For this reason relatively large amounts of data must be gathered to demonstrate the real long term failure rate. Since this would require years of testing on thousands of devices, methods of accelerated testing have been developed.

Years of semiconductor device testing has shown that temperature will accelerate failures and that this behavior fits the form of the Arrhenius equation:

$$R(t) = Ro(t)e^{-/KT}$$

Where R(t) = reaction rate as a function of time and temperature

R_o = A constant
t = Time
T = Absolute temperature, °Kelvin (°C + 273°)
= Activation energy in electron volts (ev)
K = Boltzman's constant = 8.62 x 10^{-5} ev/°K

This equation can also be put in the form:
AF = Acceleration factor
T2 = User temperature
T1 = Actual test temperature

The Arrhenius equation states that reaction rate increases exponentially with the temperature. This produces a straight line when plotted on log−linear paper with a slope physically interpreted as the energy threshold of a particular reaction or failure mechanism.

Reliability Qualifications/Evaluations Outline:

Some of the functions of ON Semiconductor Reliability and Quality Assurance Engineering is to evaluate new products for introduction, process changes (whether minor or major), and product line updates to verify the integrity and reliability of conformance, thereby ensuring satisfactory performance in the field. The reliability evaluations may be subjected to a series of extensive reliability testing, such as those outlined in the "Tests Performed" section, or special tests, depending on the nature of the qualification requirement.

Reading ON Semiconductor IGBT Datasheets

ON Semiconductor®

http://onsemi.com

APPLICATION NOTE

Abstract

The Insulated Gate Bipolar Transistor is a power switch well suited for high power applications such as motor control, UPS and solar inverters, and induction heating. If the application requirements are well understood, the correct IGBT can easily be selected from the electrical properties provided in the manufacturers' datasheet. This application note describes the electrical parameters provided in the ON Semiconductor IGBT datasheet.

Part Number

The part numbering convention for ON Semiconductor IGBTs is shown in Figure 1. Many of the device ratings and details are described in the part number and can be understood using this code.

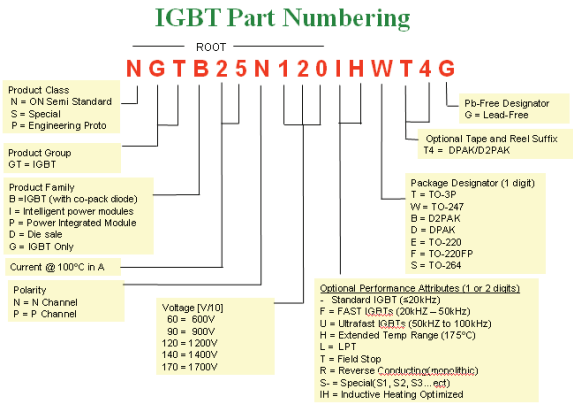


Figure 1. ON Semiconductor IGBT Part Numbering Key

Brief

This section provides a description of the device and lists its key features and typical applications.

Table 1. ABSOLUTE MAXIMUM RATINGS

Rating	Symbol	Value	Unit
Collector−emitter voltage	V_{CES}	600	V
Collector current @ T_C = 25°C @ T_C = 100°C	I_C	 30 15	A
Pulsed collector current, T_{pulse} limited by T_{Jmax}	I_{CM}	60	A
Diode forward current @ T_C = 25°C @ T_C = 100°C	I_F	 30 15	A
Diode pulsed current, T_{pulse} limited by T_{Jmax}	I_{FM}	60	A
Gate−emitter voltage	V_{GE}	±20	V
Power dissipation @ T_C = 25°C @ T_C = 100°C	P_D	 130 55	W
Short circuit withstand time V_{GE} = 15 V, V_{CE} = 400 V, T_J ≤ +150°C	t_{SC}	10	μs
Operating junction temperature range	T_J	−55 to +150	°C
Storage temperature range	T_{stg}	−55 to +150	°C
Lead temperature for soldering, 1/8" from case for 5 seconds	T_{SLD}	260	°C

Stresses exceeding Maximum Ratings may damage the device. Maximum Ratings are stress ratings only. Functional operation above the Recommended Operating Conditions is not implied. Extended exposure to stresses above the Recommended Operating Conditions may affect device reliability.

Absolute Maximum Ratings

The absolute maximum ratings shown in Table 1 are typical for an IGBT. This table sets the limits, both electrical

and thermal, beyond which the functionality is no longer guaranteed and at which physical damage may occur. The absolute maximum rating does not guarantee that the device will meet the data sheet specifications when it is within that range. The specific voltage, temperature, current and other limitations are called out in the Electrical Characteristics table.

Collector−Emitter Voltage, V_{CES}

The maximum rated voltage to be applied between the collector and emitter terminals of the device is specified to prevent the device from entering avalanche breakdown and dissipating excessive energy in the device. The avalanche breakdown voltage varies with temperature and is at its minimum at low temperature. The breakdown voltage of the device is designed to meet the minimum voltage rating at −40°C.

Collector Current, I_C

The maximum collector current is defined as the amount of current that is allowed to flow continuously into the collector for a given case temperature, T_C, in order to reach the maximum allowable junction temperature, T_J (150°C). The collector current can be stated in the following equation form:

$$I_C = \frac{T_J - T_C}{R_{th(j-c)(IGBT)} \cdot V_{CE(sat)}}$$

where $R_{th(j-c)}$ is the thermal resistance of the package and $V_{CE(sat)}$ is the on−state voltage at the specified current, I_C. Since it is the current being sought after, and $V_{CE(sat)}$ is a function of current, the equation must be solved iteratively. An estimate of the $V_{CE(sat)}$ for a given collector current and temperature can found in the typical datasheet curves, discussed later.

It is very important to understand that the absolute maximum collector current is defined based on very specific electrical and thermal conditions. The capability of the IGBT to conduct current without exceeding the absolute maximum junction temperature is highly dependent on the thermal performance of the system, including heatsinks and airflow.

Pulsed Collector Current, I_{CM}

The pulsed collector current describes the peak collector current pulse above the rated collector current specification that can flow while remaining below the maximum junction temperature. The maximum allowable pulsed current in turn depends on the pulse width, duty cycle and thermal conditions of the device.

Diode Forward Current, I_F

The diode forward current is the maximum continuous current that can flow at a fixed case temperature, T_C, while remaining under the maximum junction temperature, T_J. This is determined in similar fashion to the $V_{CE(sat)}$, above.

$$I_F = \frac{T_J - T_C}{R_{th(j-c)(diode)} \cdot V_F}$$

The equation relating I_F and V_F to the temperature rise is the same, although the $R_{th(j-c)}$ for the diode is specified separately.

Diode Pulsed Current, I_{FM}

The pulsed diode current describes the peak diode current pulse above the rated collector current specification that can flow while the junction remains below its maximum temperature. The maximum allowable pulsed current in turn depends on the pulse width, duty cycle and thermal conditions of the device.

Gate−Emitter Voltage, V_{GE}

The gate−emitter voltage, V_{GE} describes maximum voltage to be applied from gate to emitter under fault conditions. The gate−emitter voltage is limited by the gate oxide material properties and thickness. The oxide is typically capable of withstanding greater than 80V before the oxide ruptures, but to ensure reliability over the lifetime of the device, and to allow for transient overvoltage conditions in the application, this voltage is limited to well below the gate rupture voltage.

Power Dissipation, P_D

The maximum power dissipation is determined using the following equation:

$$P_D = \frac{T_J - T_C}{R_{th(j-c)}}$$

where $R_{th(j-c)}$ is the thermal resistance of the package. The maximum power dissipation is given at case temperatures of 25°C and 100°C, where the maximum junction temperature is 150°C.

Short Circuit Withstand Time, t_{sc}

The short circuit withstand time describes the ability of the device to carry high current and sustain high voltage at the same time. The device must withstand at least the rated short circuit withstand time with specified voltages applied from collector to emitter and from gate to emitter. The collector−emitter voltage specified for the test will vary based on the minimum blocking voltage capability of the device. The gate−emitter voltage is usually 15 V. The current flowing through the device under these conditions can far exceed the rated current, and is limited by the IGBT forward transconductance, an electrical parameter described below. The failure mode during this fault condition is usually thermal in nature.

Operating Junction Temperature Range, T_J

This is the junction temperature range in which the device is guaranteed to operate without physical or electrical damage or reduced life expectancy.

Storage Temperature Range, T_{stg}

This is the temperature range in which the device may be stored, without electrical bias, without reducing the life expectancy of the device.

Lead Temperature for Soldering, T_{SLD}

The maximum allowable soldering temperature is limited by the thermal conduction from the leads to the junction and die attach regions of the device. The maximum lead temperature is also dependent on the duration for which the soldering iron is applied to the lead. The maximum time for application of the heat is specified in the conditions of this rating.

THERMAL CHARACTERISTICS

Table 2. TABLE OF IGBT AND DIODE THERMAL CHARACTERISTICS

Rating	Symbol	Value	Unit
Thermal resistance junction to case, for IGBT	$R_{th(j-c)}$	1.1	°C/W
Thermal resistance junction to case, for Diode	$R_{th(j-c)}$	2.4	°C/W
Thermal resistance junction to ambient	$R_{th(j-a)}$	60	°C/W

Thermal Resistance Junction−to−Case, $R_{th(j-c)}$

The value for the thermal resistance given in Table 2 represents the steady−state thermal resistance under dc power conditions, applied to the IGBT. The thermal resistance is derated for a square power pulse for reference in designing pulse width modulated applications and is described in the graph of thermal resistance for varying pulse width and duty ratio, shown in Figure 2, below.

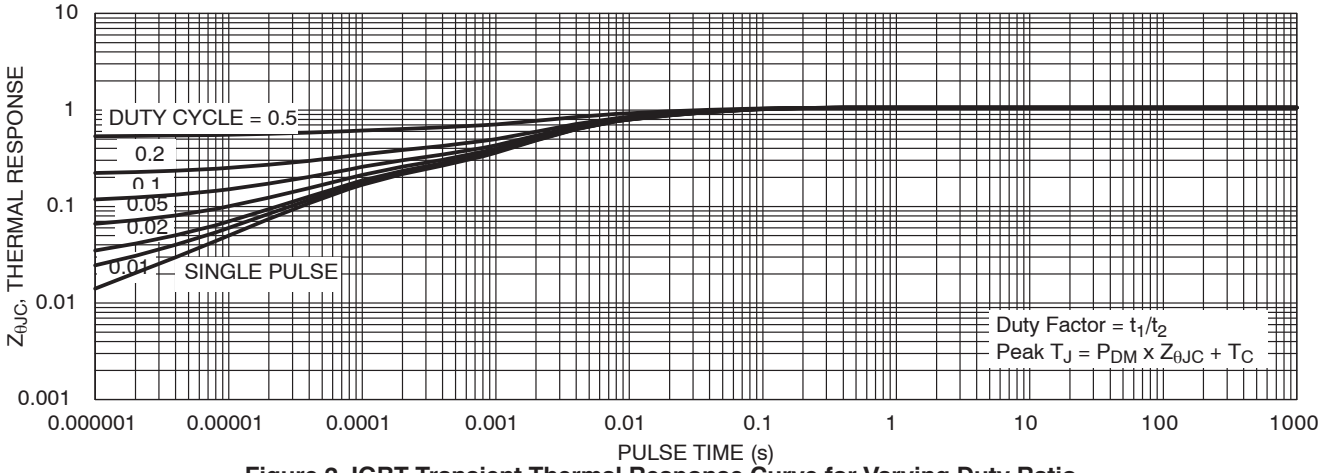


Figure 2. IGBT Transient Thermal Response Curve for Varying Duty Ratio

For a copackaged device such as the NGTB15N60EG the thermal resistance from the junction to case is specified separately for the IGBT and the diode.

Thermal Resistance Junction−to−Ambient, $R_{th(j-a)}$

This is the entire thermal resistance from the silicon junction−to−ambient.

Electrical Characteristics

Static Characteristics

The static, or dc, electrical characteristics are shown in Table 3.

Table 3. IGBT STATIC ELECTRICAL CHARACTERISTICS

Parameter	Test Conditions	Symbol	Min	Typ	Max	Unit
STATIC CHARACTERISTIC						
Collector−emitter breakdown voltage, gate−emitter short−circuited	V_{GE} = 0 V, I_C = 500 μA	$V_{(BR)CES}$	600	−	−	V
Collector−emitter saturation voltage	V_{GE} = 15 V , I_C = 15 A V_{GE} = 15 V , I_C = 15 A, T_J = 150°C	V_{CEsat}	− −	1.7 2.1	1.95 2.4	V
Gate−emitter threshold voltage	V_{GE} = V_{CE} , I_C = 250 μA	$V_{GE(th)}$	4.5		6.5	V

Table 3. IGBT STATIC ELECTRICAL CHARACTERISTICS

Parameter	Test Conditions	Symbol	Min	Typ	Max	Unit
STATIC CHARACTERISTIC						
Collector−emitter cut−off current, gate−emitter short−circuited	V_{GE} = 0 V, V_{CE} = 600 V V_{GE} = 0 V, V_{CE} = 600 V, T_J = 150°C	I_{CES}	− −	10 −	− 200	μA
Gate leakage current, collector−emitter short−circuited	V_{GE} = 20 V, V_{CE} = 0 V	I_{GES}	−	−	100	nA
Forward Transconductance	V_{CE} = 20 V, I_C = 15 A	g_{fs}	−	10.1	−	S

Collector−Emitter Breakdown Voltage, $V_{(BR)CES}$

This is the minimum off−state forward blocking voltage guaranteed over the operating temperature range. It is specified with the gate terminal tied to the emitter with a specified collector current large enough to place the device into avalanche.

Collector−Emitter Saturation Voltage, $V_{CE(sat)}$

$V_{CE(sat)}$ is an important figure of merit, since it is directly related to the conduction losses of the device. This is the voltage drop from collector to emitter for a specified gate voltage and collector current. Both a typical value and a maximum value are specified in the electrical table for both 25°C and 150°C.

In addition to the electrical limits in the table, the datasheet includes a graph describing the dependence of $V_{CE(sat)}$ on temperature, as shown in Figure 3. The graph describes the typical part and does not guarantee performance, but it can be used as a starting point to determine the $V_{CE(sat)}$ for a given temperature. The curves are given for V_{GE} = 15 V and various collector currents.

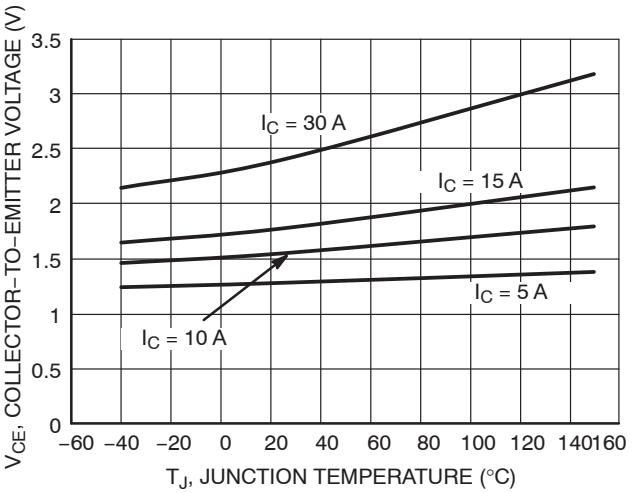


Figure 3. Graph of the Temperature Dependence of $V_{CE(sat)}$

The $V_{CE(sat)}$ values in the electrical parameter table are only given for V_{GE} = 15 V. If the gate of the IGBT is being driven by a different voltage, the output characteristics shown in Figure 4 can also be useful in approximating the $V_{CE(sat)}$. This chart shows the I_C dependence on V_{CE} for various gate−emitter voltages. The datasheet contains output characteristics for T_A = −40, 25, and 150°C.

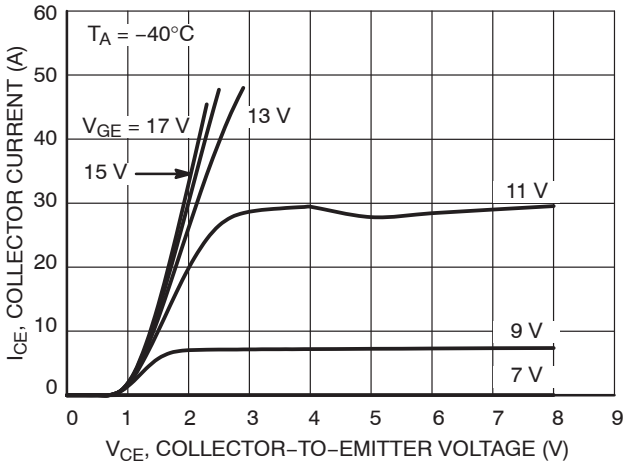


Figure 4. Graph of the Output Characteristics of the IGBT at 25°C

The characteristic curves and typical relationships should never be substituted for worst case design values. Good design practices and board−level design evaluation are critical for a reliable system.

Gate−Emitter Threshold Voltage, $V_{GE(th)}$

This parameter describes the gate to emitter voltage required for a specified amount of collector current to flow. This defines the gate to emitter voltage at which the device enters the on−state. Typically this test is based on a collector current flow proportional to the die size.

Collector−Emitter Cut−off−Current, I_{CES}

This specifies the leakage current one can expect in the off−state forward blocking mode. It is specified at the maximum rated blocking voltage, V_{CES} with the gate−to−emitter voltage equal to zero volts. The maximum allowable value of leakage current occurs at the maximum junction temperature.

Gate Leakage Current, I_{GES}

The absolute maximum value of gate leakage current is typically specified at a gate voltage of 20 V while the collector and emitter are grounded.

Forward Transconductance, g_{fs}

This is the amount of change in collector current for an incremental change in the gate to emitter voltage, measured in Siemens (or Mhos). It is specified at the room temperature rated current of the device, and typically with the device in full saturation, where a further increase in collector−emitter voltage no longer leads to an additional increase in collector current. A typical collector−emitter voltage used for this test is 20 V. Figure 5 illustrates the g_{fs} measurement.

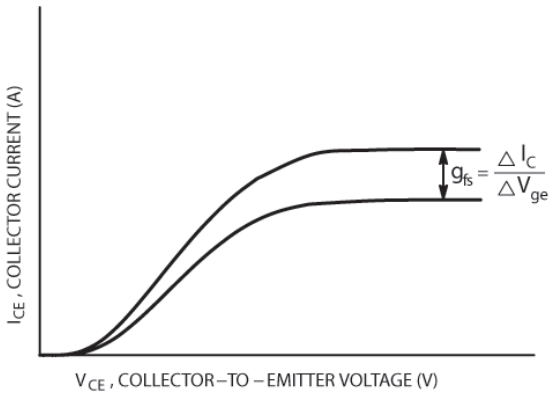


Figure 5. Illustration of the Measurement of IGBT g_{fs}

Dynamic Characteristics

Table 4. IGBT Dynamic Electrical Characteristics

Parameter	Test Conditions	Symbol	Min	Typ	Max	Unit
DYNAMIC CHARACTERISTIC						
Input capacitance	V_{CE} = 20 V, V_{GE} = 0 V, f = 1 MHz	C_{ies}	−	2600	−	pF
Output capacitance		C_{oes}	−	64	−	
Reverse transfer capacitance		C_{res}	−	42	−	
Gate charge total	V_{CE} = 480 V, I_C = 15 A, V_{GE} = 15 V	Q_g		80		nC
Gate to emitter charge		Q_{ge}		24		
Gate to collector charge		Q_{gc}		33		

The dynamic electrical characteristics which include device capacitances and gate charge are given in the electrical table, as shown in Table 4.

IGBT capacitances are similar to those described for power MOSFETs. The datasheet describes the measurable terminal capacitances, C_{ies}, C_{oes}, and C_{res}. They are specified in the electrical table at a fixed collector bias voltage; however, the capacitances are voltage dependant, as can be seen in Figure 6. The capacitances specified on the datasheet are convenient and easily measured. They relate to the pin to pin capacitances shown in Figure 7 and described below.

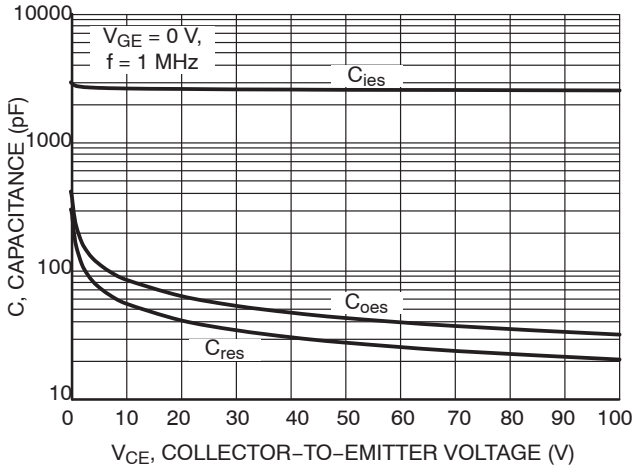


Figure 6. IGBT Capacitance versus Collector−Emitter Voltage Showing Voltage Dependance of C_{oes} and C_{res}

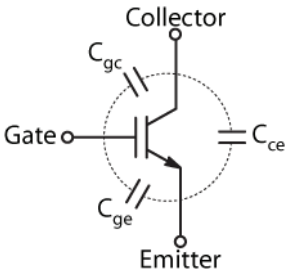


Figure 7. Pin−to−pin Capacitances of the IGBT

$$C_{ies} = C_{ge} + C_{gc} \text{ with } C_{ce} \text{ shorted}$$

$$C_{oes} = C_{gc} + C_{ce}$$

$$C_{res} = C_{gc}$$

Input Capacitance, C_{ies}

The input capacitance is made up of the parallel combination of gate−emitter and gate−collector capacitances, when the collector and emitter are tied together. The gate−emitter capacitance is constant, as it consists mainly of the metal−oxide−semiconductor capacitance. The gate−collector capacitance is a combination of a fixed oxide capacitor and a p−n junction capacitor. This results in a voltage dependence that is slightly more complex than that of a p−n junction.

Output Capacitance, C_{oes}

The output capacitance is formed by the parallel combination of the gate−collector and collector−emitter capacitances. As mentioned above, the gate−collector capacitance is voltage dependant. This is also true for the collector−emitter capacitance. The voltage dependence of the collector−emitter junction is that of a p−n junction.

Transfer Capacitance, C_{res}

The transfer capacitance is composed only of the gate−collector capacitance. Its role in the device operation is critical, as it provides negative feedback between the collector and the gate. This capacitance is responsible for the plateau on the gate charge curve. The change in collector−emitter voltage forces a current through C_{res} which reduces the gate drive current while the collector voltage is changing.

Gate Charge, Total, Q_g

Input capacitance is useful, but in terms of gate drive design, the more important figure of merit is the gate charge. It is used to size the gate drive components and predict switching losses in the driver. To measure gate charge the IGBT gate is driven with a current and the gate voltage change is monitored versus time. The resulting gate voltage versus gate charge curve is shown in Figure 8 for a constant current gate drive signal.

Q_g is the total charge required on the gate to raise V_{GE} to a specified gate voltage. ON Semiconductor devices are specified at V_{GE} = 15 V.

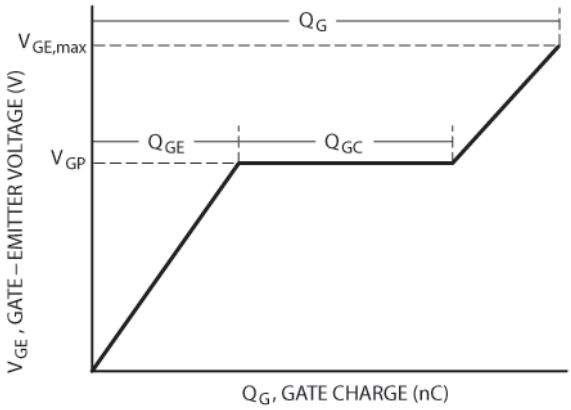


Figure 8. Theoretical Gate Charge Curve showing V_{GP}, Q_G, Q_{GE}, and Q_{GC}

Gate to Emitter Charge, Q_{ge}

Q_{ge} is the amount of charge required to reach the plateau voltage V_{GP}. This charge contributes to turning on the MOS channel, at which time the collector−emitter voltage begins to transition from high to low voltage. The level of V_{GP} is dependent on the load current being switched and can be approximated by determining the V_{GS} that corresponds to the switching current level from the transconductance curves in Figure 5.

Gate to Collector Charge, Q_{gc}

Q_{gc} is the amount of charge required to charge the junction capacitor while the voltage from collector to emitter is decreasing in the transition between the off−state and on−state. This plateau corresponds to the charging of what is also known as the Miller capacitance.

Switching Characteristics

The IGBT switching characteristics are of great importance because they relate directly to the switching energy losses of the device. Switching losses can be substantial, especially at higher frequencies and increasing temperature, where the switching losses increase.

When voltage is applied to the gate, the input capacitance must first be charged to the threshold voltage, $V_{GE(th)}$. This leads to a delay ($t_{d(on)}$) before the IGBT collector current begins to flow. Once the collector current begins to flow, the depletion layer that blocks the voltage during the off−state begins to collapse. The voltage drops to the on−state voltage drop, $V_{CE(sat)}$. This is illustrated in Figure 9.

During turn−off, the gate voltage is reduced to zero and the opposite occurs. The channel for the MOSFET current is closed and the current begins to drop abruptly. The voltage

begins to rise from $V_{CE(sat)}$ as the charge due to current flow is removed. The voltage across the device reaches the supply voltage, and minority carriers that remain in the device after turn−off cause a tail current that continues to flow. This is illustrated in Figure 10.

The switching characteristics are given in the electrical parametric table for T_J = 25 and 150°C. These are shown in Table 5.

Table 5. INDUCTIVE SWITCHING ELECTRICAL CHARACTERISTICS OF THE IGBT

Parameter	Test Conditions	Symbol	Min	Typ	Max	Unit
SWITCHING CHARACTERISTIC , INDUCTIVE LOAD						
Turn−on delay time	T_J = 25°C V_{CC} = 400 V, I_C = 15 A R_g = 22 Ω V_{GE} = 0 V / 15 V	$t_{d(on)}$		78		ns
Rise time		t_r		30		
Turn−off delay time		$t_{d(off)}$		130		
Fall time		t_f		120		
Turn−on switching loss		E_{on}		0.900		mJ
Turn−off switching loss		E_{off}		0.300		
Total switching loss		E_{ts}		1.200		
Turn−on delay time	T_J = 150°C V_{CC} = 400 V, I_C = 15 A R_g = 22 Ω V_{GE} = 0 V / 15 V	$t_{d(on)}$		76		ns
Rise time		t_r		33		
Turn−off delay time		$t_{d(off)}$		133		
Fall time		t_f		223		
Turn−on switching loss		E_{on}		1.10		mJ
Turn−off switching loss		E_{off}		0.510		
Total switching loss		E_{ts}		1.610		

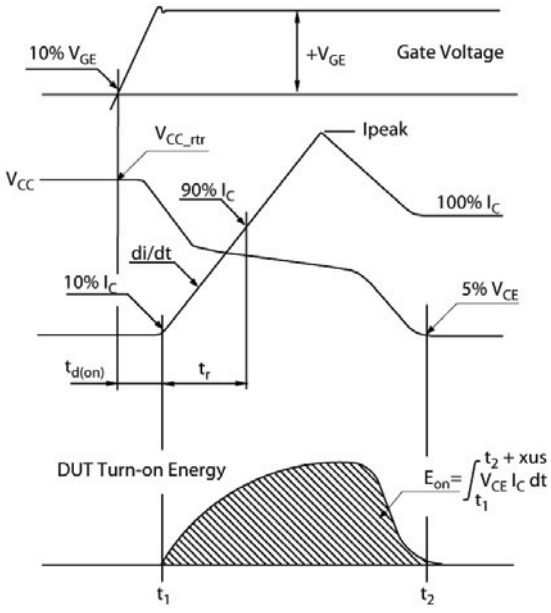


Figure 9. Turn−on Switching Illustration Showing the Definitions of the Turn−on Switching Characteristics

Turn−on Delay Time, $t_{d(on)}$

$t_{d(on)}$ is the time delay between the rising edge of the gate pulse and the rising edge of the IGBT collector current. The measurement considers the point at which both the gate voltage and collector current reach 10% of their final specified value.

Rise Time, t_r

The interval between the time the collector reaches 10% of its specified current value and the time it reaches 90% of its final value is defined as the rise time.

Turn−on Switching Loss, E_{on}

The turn−on switching losses are calculated by integrating the power dissipation (I_C x V_{CE}) over the time interval starting when the collector current reaches 10% of its final value and ending when the collector−emitter voltage reaches 5% of its peak value.

Turn−off Delay Time, $t_{d(off)}$

$t_{d(off)}$ is the time delay between the falling edge of the gate pulse and the falling edge of the collector current. The measurement is the time between the point at which the gate voltage falls to 90% of its maximum value and the collector current reaches 10% of its final specified value.

Fall Time, t_f

The fall time is defined as the time required for the collector current to drop from 90% to 10% of its initial value.

Turn−off Switching Loss, E_{off}

The turn−off switching energy losses are calculated to include the overlap of the rising collector−emitter voltage

and the falling collector current. Because the IGBT is a minority carrier device, the collector current continues to flow after the time where the collector voltage has fully risen. This residual current, called tail current, eventually decays to zero. It is customary to add a fixed length of time to the end of the turn−off time to capture the energy lost during the entire tail current. This added time is denoted as xμs in Figure 10.

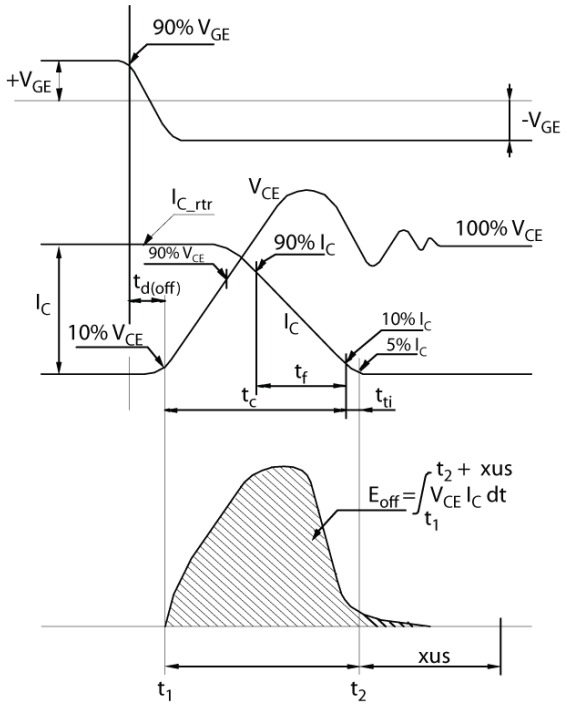


Figure 10. Turn−off Switching Illustration Showing the Definitions of the Turn−off Switching Characteristics

Total Switching Loss, E_{ts}

The total switching losses comprise the sum of the turn−on and turn−off switching losses.

Typical switching time and switching energy loss graphs are given that describe the dependence of the switching characteristics on a variety of system variables. The dependence on junction temperature, collector current, collector−emitter voltage, and gate resistance are all provided to aid in the design process.

Diode Characteristics

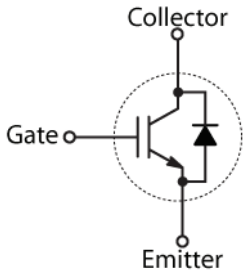


Figure 11. Copackaged IGBT and Freewheeling Diode

IGBTs are frequently used in applications where the load is inductive, such as motor control. These applications are hard switching and require that the IGBT be in parallel with a freewheeling diode. ON Semiconductor offers copackaged IGBT and diode devices. The diode cathode and IGBT collector are connected together and the diode anode and IGBT emitter are also connected, as shown in Figure 11. The freewheeling diode takes the place of the body diode that otherwise exists in a power MOSFET. For IGBTs that are copackaged with a freewheeling rectifier diode, the datasheet will also include electrical specifications for the diode, as shown in Table 6.

Table 6. ELECTRICAL CHARACTERISTICS OF THE DIODE

Parameter	Test Conditions	Symbol	Min	Typ	Max	Unit
DIODE CHARACTERISTIC						
Forward voltage	V_{GE} = 0 V, I_F = 15 A V_{GE} = 0 V, I_F = 15 A, T_J = 150°C	V_F		1.6 1.6	1.85	V
Reverse recovery time	T_J = 25°C I_F = 15 A, V_R = 200 V di_F/dt = 200 A/μs	t_{rr}		270		ns
Reverse recovery charge		Q_{rr}		350		nc
Reverse recovery current		I_{rrm}		5		A
Reverse recovery time	T_J = 125°C I_F = 15 A, V_R = 200 V di_F/dt = 200 A/μs	t_{rr}		350		ns
Reverse recovery charge		Q_{rr}		1000		nc
Reverse recovery current		I_{rrm}		7.5		A

Forward Voltage, V_F

The forward voltage of the rectifier is measured while the IGBT gate and emitter terminals are tied together, ensuring the IGBT is in its off−state. A forcing current enters the emitter terminal and the emitter−collector (anode−cathode) voltage is measured.

Forward voltage is an important parameter in hard switching applications. V_F is specified in the electrical table

for a given current and is specified at T_J = 25 and 150°C. The datasheet also includes a graph showing the I_F−V_F relationship for a typical part at T_J = −40, 25, and 150°C, as shown in Figure 12.

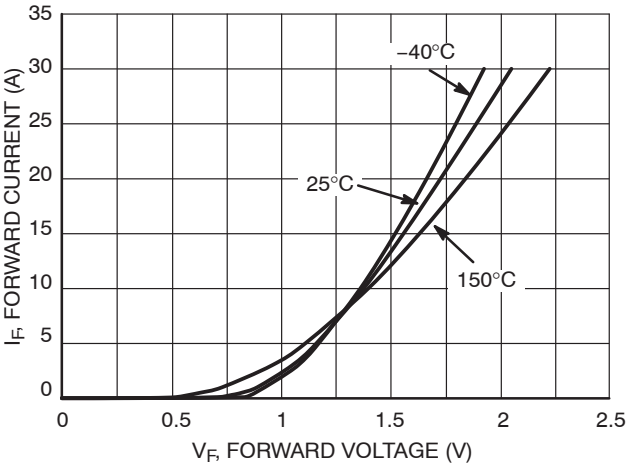


Figure 12. Diode Forward Characteristic Curves for T_J = −40, 25, and 150°C

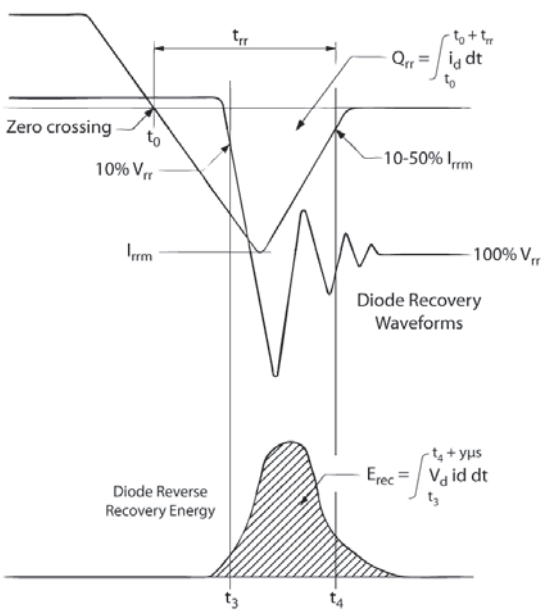


Figure 13. Diode Reverse Recovery Illustration Showing the Definitions of the Reverse Recovery Characteristics

Reverse Recovery Time, t_{rr}

The reverse recovery time, t_{rr}, defines the time the diode takes to enter the reverse blocking state after conducting in the forward direction. It is defined as the length of time required for the reverse current to return to 10% of its peak reverse value (I_{rrm}). It is measured from the point in time where the diode current crosses zero. The time period is labeled in Figure 13.

Reverse Recovery Charge, Q_{rr}

The amount of charge that is recovered from the diode during turn−off is referred to as reverse recovery charge, Q_{rr}. It is calculated by taking the integral of the reverse recovery current over the time period, t_{rr}.

Reverse Recovery Current, I_{rrm}

I_{rrm} is the peak current reached during diode turn off. I_{rrm} depends on the initial forward diode current and the rate of change of the diode current, dI/dt, used to turn the diode off.

AND8220/D

How To Use Thermal Data Found in Data Sheets

Packaging Technology Development

Prepared by: Roger Paul Stout, PE
ON Semiconductor

ON Semiconductor®

http://onsemi.com

APPLICATION NOTE

Table of Contents

Definitions

ambient, T_A = where all the heat ends up, the environment thermally "far" from the device (not to be confused with the case temperature, which may be vastly different from the air temperature just millimeters away).

case temperature, T_C = representative point on the external "case" of the device, location must be well defined along with use of any parameters based on this value.

junction, T_J = the hottest point inside the semiconductor device.

psi–JT, $\Psi_{JT} = \frac{T_J - T_T}{P_d}$ thermal characterization parameter, measured junction to case top (T_T)

psi–Jx, $\Psi_{Jx} = \frac{T_J - T_x}{P_d}$ thermal characterization parameter, measured junction to location defined (T_x)

psi–xA, $\Psi_{xA} = \frac{T_x - T_A}{P_d}$ thermal characterization parameter, measured from location defined (T_x) to ambient

theta–JA, $\theta_{JA} = \frac{T_J - T_A}{P_d}$ overall thermal resistance of device plus external system

theta–JC, $\theta_{JC} = \frac{T_J - T_C}{P_d}$ ideally, thermal resistance of just the device as measured to the case

P_d = total power dissipation of device

min–pad = in reference to a thermal test board, a board having only the minimum amount of metal pads and traces required to mount the device and carry power and signals to and from the device; the traces may actually have significantly more area than the mounting pad itself, and the total size of the board (typically 3″ square), and its thickness (0.062″) may be significantly different than what will be used in an actual application. These variables are only some of those that render min–pad data sheet values of limited use in a real application environment.

1″ pad = in reference to a thermal test board, a board having a nominally 1″ square area of copper plating, at the center of which is mounted the package; typically the additional trace area required to carry power and signals to and from the device will be a small fraction (<10%) of the pad; but for larger devices, e.g. a D2pak, the actual heatsink itself may be a significant fraction of 1–sq–in., so the difference between the min–pad and 1″ pad values will not be as large for large devices as it is for tiny devices. Copper spreader thickness (typically 1–oz, meaning 1–sq.–ft of the material in that thickness would weigh 1–oz), and overall board size (typically 3″ square) and thickness (0.062″) will potentially make the value significantly different than what would be experienced in an actual application. These variables are only some of those that render 1″ pad data sheet values of limited use in a real application environment.

Steady State Data

By "steady state", we mean operating conditions wherein power dissipation in each relevant device, has been constant for a long enough period of time that temperature changes are no longer occurring. Starting from zero power, with all temperatures initially at ambient, the sudden application of constant non–zero power will result in monotonically increasing temperatures. The highest temperatures will eventually be reached, therefore, at steady state. Steady state thermal data is often provided in the form of specific values of thermal resistance or impedance. In addition, other charts may be presented showing how steady–state thermal characteristics typically depend on certain external conditions, such as the amount of heat spreading metal that has been provided on the application board for the specific device in question. In the case of multiple–junction devices, there may be a matrix form of the steady state thermal characteristics as well.

Theta (θ) and Psi (Ψ) Numbers

Theta (θ), sometimes denoted R_θ, values are true "thermal resistances." That is, they tell you that if you know the temperatures at two points (connected by the thermal resistance) then the amount of heat that flows from one point to the other is completely determined by that resistance. Conversely, if you know the heat flow along the path, and you know its resistance, then you can predict the temperature difference that will result due to this heat flow. If there are other heat paths in the system, they have their own characteristics, and they are independent of what happens along the particular path of interest. Typical units would be °C/W.

In the context of semiconductor packages and devices, there are usually going to be, *at most*, two "true" thermal resistances, theta–JA, and theta–JC, and these must be defined carefully. But the single most important fact about these values (what, indeed, makes them "theta"s), is that the total power dissipated by the device flows between the two "points" being described (the junction being one "point," and either ambient or the case temperature being the other "point"). That is, there are no extraneous, parallel thermal paths in the system allowing some of the heat to "leak" away. All heat leaving the junction, the first point, eventually arrives at or passes through the other point – either ambient or case, respectively.

Mathematically, we define these two quantities as follows:

$$\theta_{JA} = \frac{T_J - T_A}{P_d} \quad \text{(eq. 1)}$$

$$\theta_{JC} = \frac{T_J - T_C}{P_d} \quad \text{(eq. 2)}$$

Knowing the appropriate values, actual operating junction temperature therefore could be predicted according to:

$$T_J = \theta_{JA} \cdot P_d + T_A \quad \text{(eq. 3)}$$

or

$$T_J = \theta_{JC} \cdot P_d + T_C \quad \text{(eq. 4)}$$

For theta–JA, obviously, by definition, all the power leaving the junction eventually arrives at ambient, hence the ratio of this temperature difference over the total package power, is a true system thermal resistance. A data sheet may provide one or more theta–JA values for different representative mounting situations. For instance, min–pad board values, and 1″ pad board values; or perhaps a chart of theta–JA vs. copper spreader area. But these values may not apply to a real application, even if the copper area is "correct." Other variables, such as the presence of other power dissipating devices, the nature of the air flow conditions, the thickness and detailed layout of the spreader itself, will all affect the value.

For theta–JC, the two points are the junction (J), and the "case" (C) – it's the definition or selection of the "case" temperature that makes the challenge here. If we are justified in making the assumption that 100% of the dissipated power actually flows past the "C" point we've defined, then again, the ratio of this temperature difference to the total package power, is a true thermal resistance. Typically the only test situation that results in a reasonable approximation of this 100%–of–heat condition is a coldplate test, and that being for power packages clamped directly onto the coldplate. (Even in such a test, from 1–10% of the power may "leak" out through other paths, depending on the particular package design and test fixturing.) One must still carefully designate the location of the "case" measurement. For a good theta–JC measurement, the "case" is usually defined as the center point of the heatsink at the heatsink/coldplate interface, which will be the hottest point on the coldplate, but not necessarily the coldest point on the "case" of the device at all. One practical difficulty in making a good theta–JC measurement, is doing so without disturbing the heat flow. A groove in the surface of the coldplate may reduce the interface area significantly; holes bored through the coldplate disturb the heat flow as well, though if small enough, perhaps not as much. Another issue is the question of how well the case–measuring thermocouple makes contact with the case, when it can't be seen. As an alternative, measuring the temperature at the exposed edge of the heatsink (say on the tab of a Dpak or TO220) may circumvent both these difficulties, but it may yield a significantly different result (perhaps 20–40% higher than the "true" theta–JC value). Clearly, to use a theta–JC value successfully in a real application environment (i.e. to predict the actual junction temperature), the application must assure that virtually 100% of the device power flows through the case. Finally, a theta–JC value as measured on a coldplate, may be vastly different than a corresponding "psi–JC" value, even through the thermocouple location may be identical for both measurements (see following discussion on "psi" values). This arises because the fraction of the heat passing through the "case" point will very likely be substantially less in a non–coldplate test setup, as compared to the coldplate setup. Indeed, if the theta–JC value was derived from a 100%–of–heat condition, it is axiomatic that any corresponding psi–JC value will be lower. For instance, if only 10% of the heat flows past the "case" in a non–coldplate mounting situation, then psi–JC will be one tenth of the theta–JC value!

Clearly, the difference between theta–JA and theta–JC is that theta–JA necessarily includes the entire system, not just the package, whereas theta–JC is idealized as a "package only" property. It is not at all unusual for theta–JC to be a small fraction of theta–JA, which is a way of saying that the thermal design of the external environment is more important in determining the operating junction temperature of a device, than is the thermal design of the device itself.

The problem with these values appearing on data sheets, is that theta–JA will very likely not apply to a particular customer application (because there will be differences in the system external to the package, such as air flow conditions, amount of metal thickness, area, and layout in the board, proximity and power dissipation of neighboring devices, and so forth). So theta–JA may *seem* convenient, because all you need to know is the ambient temperature. In reality, unless the application is exactly like the thermal test situation, a different theta–JA will apply, and the difference may be substantial. Data sheets may present either or both "min pad" values and "1″ pad values", but the real theta–JA in an application may be better than the "1″ pad" value, or it may be worse than the "min–pad" value. In any event, if the system is different, the data sheet's theta–JA is not really a useful value.

Theta–JC *may* be more useful, in the sense that it *may* actually describe the device's characteristics in the real application (including the requirement that essentially 100% of the heat pass through the identified case location). Even so, it is truly useful only if the external system is fully defined as well. The problem here is that one cannot simply assume that the case temperature can be controlled to an arbitrarily chosen value; rather, the design of the external thermal system must ensure that this is true for the amount of power being dissipated by the device. For example, consider a particular TO264 power transistor, with a theta–JC of 0.4°C/W. If max Tj is 150°C and the case could be held at 25°C, then one could *in principle* dissipate 312.5 W [$P_d = (T_J - T_C)/\theta_{JC}$].

However, what sort of external system can "hold" the case at 25°C? How about a water–cooled coldplate capable of sinking 400 W, whose thermal resistance is approximately 0.2°C/W, measured from the mounting point on the coldplate to the "infinite" coolant supply? To sink 312.5 W implies that the coolant itself would have to be held at (0.2°C/W * 312.5 W, or) 62.5°C *below* case temperature, or –37.5°C! In fact, a realistic "real world" application of this

TO264 device might utilize a forced−air−cooled heatsink, with a hundred square inches of surface area, and a net thermal resistance of 0.2°C/W (much similar in capacity to the water−cooled coldplate just illustrated). But in this real−world system, it is *ambient*, not the device case, that is limited to 25°C. Since the total system resistance (theta−JA) is 0.6°C/W (0.4 for the device, plus 0.2 for the heatsink), actual maximum power dissipation is really 208 W, and the case temperature at steady−state equilibrium will be about 66°C.

Psi (Ψ) values, as contrasted with theta values, are not really thermal resistances, though they have the same units. JEDEC defines the term as a "thermal characterization parameter."[1] It is nothing more than the ratio of the temperature difference between two selected points in a system, and the total power dissipation of the device in question. The equation defining it is essentially identical to that for theta, that is:

$$\Psi_{Jx} = \frac{T_J - T_x}{P_d} \qquad \text{(eq. 5)}$$

and

$$\Psi_{xA} = \frac{T_x - T_A}{P_d} \qquad \text{(eq. 6)}$$

observe

$$\Psi_{Jx} + \Psi_{xA} = \frac{T_J - T_x}{P_d} + \frac{T_x - T_A}{P_d} = \theta_{JA} \qquad \text{(eq. 7)}$$

Note that we have defined two variations, one referring the junction to some arbitrary package location x; the other referring the arbitrary package location x to ambient. This misleadingly suggests that the former is mainly a "package" characteristic, and the latter mainly an "environment" characteristic. The reality is that the chosen package location merely *arbitrarily* divides the overall system theta−JA into two pieces that are guaranteed to add up to the correct total (Equation 7). It does *not* follow that location x will have a predictable temperature between the two endpoints as the environment changes around it; it will only be predictable (from *either* endpoint) if the environment does not change. (Contrast this with theta−JC, which, at least ideally, will always yield a predictable temperature relative to the junction, regardless of what happens to the environment beyond the case.) Be that as it may, just as with theta values, knowing the appropriate inputs, operating temperatures might be predicted according to:

$$T_J = \Psi_{Jx} \cdot P_d + T_x \qquad \text{(eq. 8)}$$

or

$$T_x = \Psi_{xA} \cdot P_d + T_A \qquad \text{(eq. 9)}$$

[1] *Guidelines for Reporting and Using Electronic Package Thermal Information, EIA/JESD51−12*, Electronic Industries Association, 2005.

In the context of semiconductor devices and packages, common data sheet psi values include psi−JLn (where a particular lead, n, is designated), psi−JT (where T represents the case top), psi−J−tab (where the tab is the exposed heatsink tab on a suitable power device), and psi−J−board (where perhaps the board directly underneath the center of the package is designated, for instance for a BGA style package).

It is usually possible to know the total power dissipation of the device in question, but it is far more difficult to know what fraction of the heat flows out through the case top, vs. through the leads, vs. through the air gap under the package, and so forth. Though it may be feasible to make temperature measurements at myriad locations all over a package, it is difficult or impossible to make actual heat flow measurements along selected paths. Further, those paths may vary dramatically in their actual thermal resistances as connected to the junction, and in their individual sensitivity to external changes. Therefore, as mounting conditions vary, the relative heat flow along the various possible paths may shift significantly. Psi values reported on data sheets, therefore, must only be used for application temperature estimates, when it is known that the heat flow is similarly distributed. The *minimum* stipulation for valid application of a psi value is that the same fraction of heat flows along the *particular* psi−path as occurred during the lab measurement (regardless of how much difference there might be along other paths); generally this will be difficult or impossible to ascertain.

Consider a particular example of a 2−leaded axial device. It was measured on a thermal test board having a symmetric layout with equal amounts of trace metal to each lead. In this test scenario, theta−JA was 45°C/W, and psi−JL was 15°C/W (same value to each lead, due to symmetry).[2] Now in a particular application, the device is mounted on a board with a 1" square pad allocated to just one of the two leads; the other lead has minimal traces (identical to the original test board). Measurements made on this application board now yield theta−JA of 31°C/W, and extremely non−symmetric values of psi − JL1 = 21°C/W and psi − JL2 = 9°C/W. Depending on which lead temperature were to be used as the reference point, it should be clear that if the data sheet value for psi−JL was used to predict junction temperature, it would either result in extremely high or extremely low values. (Obviously, if psi values were measured for the actual application, there would be no reason to use the data sheet value. The emphasis here is on the extreme difference that the actual psi values have as compared to the data sheet, even though the package is the same, and one of its leads is even mounted exactly as during the data sheet measurements.) If one returns to the concept of true thermal resistance, it may be seen that in the original

[2] It is unfortunate in this particular example that the data sheet was pre−1995, and referred to the value as "Thermal resistance, junction−to−lead," with no explanation of how to apply the value. So is it a *psi*, or a *theta*? Nothing like a small two−to−one discrepancy in resulting calculations!

lab measurements, the true *theta*–JL value would have been 30°C/W, as each lead carried exactly one half the total power.

$$\text{that is, } \theta_{JL} = \frac{T_J - T_L}{\left(\frac{P_d}{2}\right)} = 2\Psi_{JL} \qquad \text{(eq. 10)}$$

In fact, knowing that theta value would allow one to predict junction temperatures precisely, in any application environment, however non–symmetrically the heat spreading metal was arranged, so long as both lead temperatures are measured instead of relying on just one, specifically:

$$T_J = \frac{\theta_{JL}P_d + T_{L1} + T_{L2}}{2} \qquad \text{(eq. 11)}$$

Using Equation 10, it may be of interest to rearrange this expression into:

$$T_J = \Psi_{JL}P_d + \text{avg}(T_{L1}, T_{L2}) \qquad \text{(eq. 12)}$$

Thus in this particular two–leaded device example, it is seen that even psi–JL may be used to advantage if both lead temperatures are measured in the actual application. In more complex package situations, however, there may be multiple leads, some of which have direct connections to the internal "flag" (where the silicon is mounted) and some don't, or there may be other significantly non–symmetric thermal paths. Then, changing the amount of metal trace area associated with individual leads, or adding an external heatsink to the exposed case, may drastically modify the relative heat flows as compared to those present when the data sheet values were determined, and may thus completely invalidate the use of the published psi values. As a rule, when it comes to choosing which lead to characterize for a data sheet, it is preferred to pick the one that carries the largest fraction of the power, if such a lead can in fact be identified. This may well be the one with the largest psi–JL, especially if it has a direct internal connection to the flag and has been given a disproportionately large heat spreader on the outside. Such a value will be the least sensitive to modest changes in the actual application environment. On the other hand, if a data sheet specifies a lead which is known *not* to have the most heat flow, then one cannot be sure in a particular application, whether the actual psi value will be higher or lower than that provided in the data sheet (the two–leaded example above bearing witness).

Multiple Junction Devices and Matrix Formulations

In referring to multiple junction devices, we are generally referring to devices that contain relatively independent electrical components, for which the ratio of power dissipation between the various possible "junctions" may vary widely. This might be an analog device with two different regulated voltage outputs, driving vastly different loads (that from one application to another may differ substantially). It might be a dual–rectifier package, which in one application utilizes both channels fairly equally, but in another application perhaps utilizes one in preference over the other. It might be a single application that from time to time moves between two vastly different operating points (different power distributions); which condition is "worst case" needs to be determined. So long as constant power conditions are considered, a matrix approach utilizing steady state values is a concise method of describing the system's thermal characteristics. It relies on the principle of linear superposition, which states that the temperature rise at any given point in the system is the sum of the independently derived temperature increases attributable to each heat source in the system. Stated in matrix form, we would say it this way:

$$\begin{Bmatrix}\Delta T_{J1}\\ \Delta T_{J2}\end{Bmatrix} = \begin{pmatrix}\theta_1 & \Psi_{12}\\ \Psi_{12} & \theta_2\end{pmatrix}\begin{Bmatrix}P_{d1}\\ P_{d2}\end{Bmatrix} \qquad \text{(eq. 13)}$$

Note that the matrix notation is simply a shorthand version of the following pair of equations:

$$\Delta T_{J1} = \theta_1 \cdot P_{d1} + \Psi_{12} \cdot P_{d2} \qquad \text{(eq. 14)}$$
$$\Delta T_{J2} = \Psi_{12} \cdot P_{d1} + \theta_2 \cdot P_{d2}$$

(Note that if every matrix element is referenced to ambient, for instance, then each temperature rise so computed should be added to ambient, in order to predict the actual temperatures resulting from the two applied power dissipation values). For a two junction device, this means that the temperature rise of the first junction is the sum of its "self heating" characteristic, times its own power dissipation, added to the "interaction" characteristic (how much each junction heats the other) times the power dissipation of the other junction. Another principle applicable to linear systems is known as *reciprocity*, which states that the amount by which one junction *heats* another (in terms of temperature rise per unit power input) is equal to the amount that it will be *heated by* the other, hence the symmetry of the matrix across the main diagonal.

Observe that here we have used psi to represent the interaction terms. This is strictly correct because in general we cannot say that any particular fraction of the heat dissipated at either junction is passing "through" the other junction; *some* of it certainly flows that way, causing the temperature to rise there (indeed, if it did not, the interaction term would be zero and the two junctions would be completely thermally independent). On the other hand, we have used theta for the self heating terms (along the main diagonal). This is *not* strictly correct unless we have the assurance stipulated in the preceding discussion of theta values, that is, the reference point to which theta refers is the ultimate destination of 100% of the power dissipated by (in this case) both junctions. However, even the basic matrix formulation itself, as illustrated above, is an oversimplification of the required mathematical description if multiple reference temperatures are needed.

Suffice it to say, for idealized simple situations, a matrix formulation typically will consist of theta–JA self heating and psi–JA interaction heating values, or possibly theta–JB and psi–JB values, where a board location central to the package is identified and *presumed* to represent the major

heat flow path and a common reference temperature. One way of describing the situations in which this simple matrix approach is valid, would be to say that as long as there is a single temperature *boundary condition* (for instance, ambient, or board), the formulation is applicable. (Then that temperature reference, or boundary condition, becomes the common value to which all temperature increases are related.) But in typical situations where lead, board, or case−top temperatures have been characterized using psi parameters, these additional temperature locations are not true boundary conditions, rather auxiliary reference points where the temperature is measured but not controlled.

Even with this restriction, it should be noted that a matrix description can be readily extended to any number of junctions and auxiliary temperatures, where every heat source of interest yields a self heating theta, and a set of interaction psi's, for every other junction and point of interest. For example, if we have three heat sources, and one lead and one board temperature reference location, we would have:

$$\begin{Bmatrix} \Delta T_{J1} \\ \Delta T_{J2} \\ \Delta T_{J3} \\ \Delta T_{L1} \\ \Delta T_{B} \end{Bmatrix} = \begin{bmatrix} \theta_{J1A} & \Psi_{12} & \Psi_{13} \\ \Psi_{12} & \theta_{J2A} & \Psi_{23} \\ \Psi_{13} & \Psi_{23} & \theta_{J3A} \\ \Psi_{L1-1A} & \theta_{L1-2A} & \Psi_{L1-3A} \\ \Psi_{B1A} & \Psi_{B2A} & \Psi_{B3A} \end{bmatrix} \begin{Bmatrix} P_{d1} \\ P_{d2} \\ P_{d3} \end{Bmatrix} \quad \text{(eq. 15)}$$

Here we do not have a square matrix, because we are using psi characteristics to describe the temperature at some points in the system that are not themselves heat sources. (If they *were* heat sources, we could expand the matrix back to a larger square with two additional heat input variables in the heating vector, but this would necessarily require two additional columns of psi characteristics that need to be measured before computations may be completed. Clearly it is much simpler, if those points really are unheated, to avoid the extra work.) Perhaps obviously, the symmetry consequent of the reciprocity theorem applies only to the square *sub*−matrix representing the temperatures at only the heat sources. Once again, the preceding matrix notation is simply a shorthand version of the following system of equations:

$$\begin{aligned} \Delta T_{J1} &= \theta_{J1A}\cdot P_{d1} + \Psi_{12}\cdot P_{d2} + \Psi_{13}\cdot P_{d3} \\ \Delta T_{J2} &= \Psi_{12}\cdot P_{d1} + \theta_{J2A}\cdot P_{d2} + \Psi_{d3}\cdot P_{d3} \\ \Delta T_{J3} &= \Psi_{13}\cdot P_{d1} + \Psi_{23}\cdot P_{d2} + \theta_{J3A}\cdot P_{d3} \\ \Delta T_{L1} &= \Psi_{L1-1A}\cdot P_{d1} + \Psi_{L1-2A}\cdot P_{d2} + \Psi_{L1-3A}\cdot P_{d3} \\ \Delta T_{B} &= \Psi_{B1A}\cdot P_{d1} + \Psi_{B2A}\cdot P_{d2} + \Psi_{B3A}\cdot P_{d3} \end{aligned} \quad \text{(eq. 16)}$$

to all of which, ambient must be added.

In real life, you would measure three power levels and *three* temperatures (lead, board, and ambient). The matrix method then allows you to calculate *five* temperatures: the three junctions as well as the lead and board temperatures. If there is agreement between the calculated temperatures and the measured values (two opportunities: lead and board), you would then have some confidence that all the characterization inputs (theta's and psi's) were valid for the system under consideration. If there was a significant discrepancy between the calculated and measured values for lead and board, you would have an indication that heat flow distribution in the actual application differs significantly from that during which the psi's originally had been deduced, casting into doubt the calculated junction temperatures as well.

For completeness, we shall demonstrate the additional complexity required for a true multiple heat source, multiple temperature boundary condition model. Suppose we have a six−leaded package with two independent silicon devices (i.e., two junctions, meaning two heat sources) inside. If we can make temperature measurements at all six leads, under the assumption that virtually 100% of the heat generated internally, at either junction, must exit the package along one of the leads, then we could write the following matrix equation, where each lead is treated as a separate boundary condition:

$$\begin{Bmatrix} T_{J1} \\ T_{J2} \end{Bmatrix} = \begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{12} & \Psi_{22} \end{pmatrix} \begin{Bmatrix} P_{d1} + \frac{T_{L1}}{\Psi_{1L1}} + \frac{T_{L2}}{\Psi_{1L2}} + \frac{T_{L3}}{\Psi_{1L3}} + \frac{T_{L4}}{\Psi_{1L4}} + \frac{T_{L5}}{\Psi_{1L5}} + \frac{T_{L6}}{\Psi_{1L6}} \\ P_{d2} + \frac{T_{L1}}{\Psi_{2L1}} + \frac{T_{L2}}{\Psi_{2L2}} + \frac{T_{L3}}{\Psi_{2L3}} + \frac{T_{L4}}{\Psi_{2L4}} + \frac{T_{L5}}{\Psi_{2L5}} + \frac{T_{L6}}{\Psi_{2L6}} \end{Bmatrix} \quad \text{(eq. 17)}$$

Two important differences exist between this complete model and the previous, simpler matrix formulation. First, the left hand side of the equation is the actual junction temperature prediction, *not* a temperature rise over some common reference temperature. Second, the temperature boundary conditions (of which there are six in this "complete" model) show up as quasi−heat−inputs. Each has its own associated weighting, expressed here as yet another psi value. So even though there are only two heat sources, there are actually 16 (15 independent) distinct parameters characterizing this model. Obviously it is one thing to state that such a model may exist; it is quite another thing to experimentally (or otherwise) derive all these coefficients. This is the domain of "compact models," for which a tremendous amount of literature and research has been generated over the past decade. The minimum number of independent external boundaries necessary for a certain desired level of accuracy, whether (and how) temperature nonlinearities must be considered, and even the internal structure of the model, are all significant issues that must be addressed for a thorough development of such a model.

Theta−JA vs. Copper Area

One of the most problematic issues in providing manufacturer's thermal data lies in choosing what data is helpful for purposes of illustration, but which may unfortunately grossly misrepresent how the package will actually perform in a customer's application. An excellent example of this is in the variation of Theta−JA with copper area, as shown in the next figure.

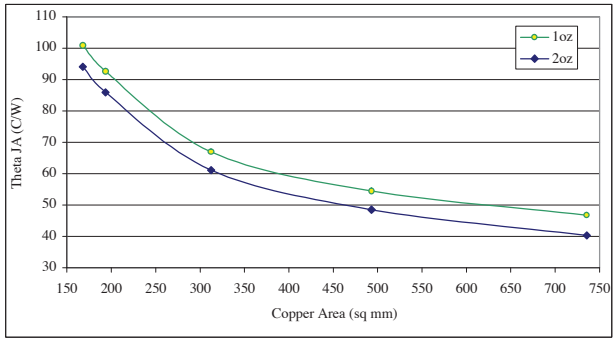


Figure 1. Theta–JA vs. Copper Area for a Dpak

The two parameters explicit on this chart are the thickness of the heat–spreading copper (one curve for each of two representative thicknesses), and the copper area (the x–axis). Even considering just these two items, many questions arise. For instance, what should be done if the copper thickness is outside the range provided? Indeed (though it would probably be considered ridiculous to try to be this precise in the first place), should one interpret linearly even within the range provided? What happens if the copper area falls below the left end or above the right end of the curves provided? Is a linear extrapolation appropriate? Should a curve (of what type?) be fit to the points known, and extended (and if so, how far is appropriate)? (Certainly, any polynomial extrapolation to the right will eventually pass through or end in physically impossible values, such as zero, or negative values. Worse, it may even *increase* as copper area increases.) Does the metal area include the traces directly associated with the device of interest? If so, clearly at some point toward the left ends of these curves, one transitions from large "blocky" areas to long skinny areas. Surely the heat spreading ability will shift markedly at this transition. So if in a particular application, a different amount of purely "trace" metal is utilized than that upon which this chart was derived, significant departures from an "obvious" extrapolation will occur.

However, beyond those explicit parameters, there are many very important factors *not* explicit on this chart affecting the actual theta–JA value that will be experienced. (And even if they were known, perhaps through conscientious provision of footnotes by the manufacturer, and diligent reference to the same by the customer, the same questions would remain regarding how to handle deviations from the stipulated conditions.) For instance, how large is the board beyond the copper area? How much airflow is there? Is the airflow the same on both top and bottom of the board? If the air is "still" (also known as "free convection," being driven by the buoyancy due to the temperature difference in the air as it is heated by the device in question), what is the orientation of the board with respect to gravity, and how does it change with orientation (significantly)?

Finally, of course, how near are the nearest neighboring heat sources? Do they share common metal (i.e. ground planes)? How much separation is there between metal areas not electrically connected?

The moral here is that these charts may be used *only* to gain a very, *very* rough idea of just how much leverage one *might* have, for the device of interest, in adjusting overall system thermal characteristics. Obviously the starting point is to subtract out the "intrinsic" device characteristics (psi–J–lead, for instance), and see if what's left over might possibly give enough margin for a required compensation. If there is, a comprehensive analysis of the external thermal system is mandatory, taking into account at least all the variables that have been highlighted in this discussion.

Transient Data

Transient results may be presented in the form of heating curves, duty cycle curves, thermal–RC–network models, or thermal–RC mathematical models. In conjunction with any of these descriptions may also be found a "sqrt(t)" surface heating model, suitable for particularly short duration heating situations outside the scope of the other approaches.

Heating Curves

A heating curve (also known as a transient response curve, or a single–pulse heating curve), shows how the junction temperature of a device increases with time, given a constant power input at the junction, in some particular environment. In the following chart, it is clear that for times shorter than about 0.2 s, the device in question has the same thermal transient response independent of what sort of board it is mounted on. Somewhere between 0.1 and 1 s the effects of the environment (the board) start to be felt, and clearly by 1000 s the difference *between* those two particular example environments (about 50°C/W) is larger than the original contribution due to the device itself at 0.2 s (5–6°C/W). If two different environments are depicted on the same heating curve chart, therefore, one gets an idea as to when the environment begins to enter into play, and how much is due to the package alone.

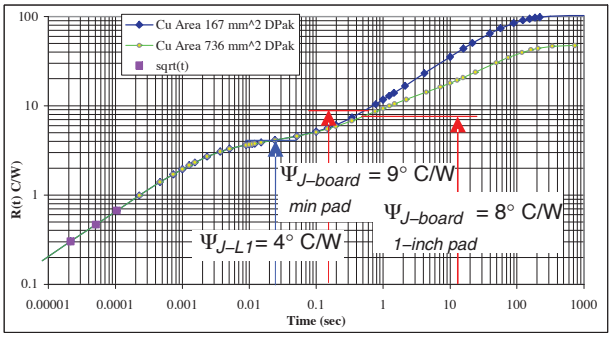


Figure 2. Typical Heating Curves

It may also be useful to note the point(s) on the transient heating curve where some of the other "steady state" theta or psi values occur. For instance, a psi–JL or psi–JB value may be found to occur roughly at the point where two

different environment curves separate. Possibly two different psi–J–tab values are given, one for each of the mounting conditions. In that case, it will be seen that they occur at roughly the same time along the two curves. One may also, therefore, use the psi–JL values, if provided, to help ascertain the time scales at which package effects “end” and environment effects “begin,” even when only one environment’s curve may be presented (for instance, if a 1″ pad curve is given, but no min–pad board curve).

The use of a heating curve is straightforward, in the situation where constant power is applied to a device. If it is desired to know how hot the junction has become at a certain moment, the R(t) value is simply looked up and used just as if it was a theta value:

$$T_J(t) = R(t) \cdot P_d + T_A \qquad \text{(eq. 18)}$$

Sometimes heating curves are referenced to lead temperature or board temperature, rather than ambient, so one must pay attention to this detail in making use of the curve. It should be evident that if the curve ends up at the steady state theta value for the particular application and environment, then this computation yields the steady state junction temperature. The heating curve simply provides a time–variable generalization of the concepts previously discussed in the steady state context.

Heating curves may be used to estimate more complex power situations. For complete generality, a non periodic power input with ramps or smooth curves, rather than square edges, can be modeled using the single–pulse curve. Figure 3, illustrates the basic steps involved in such an analysis.

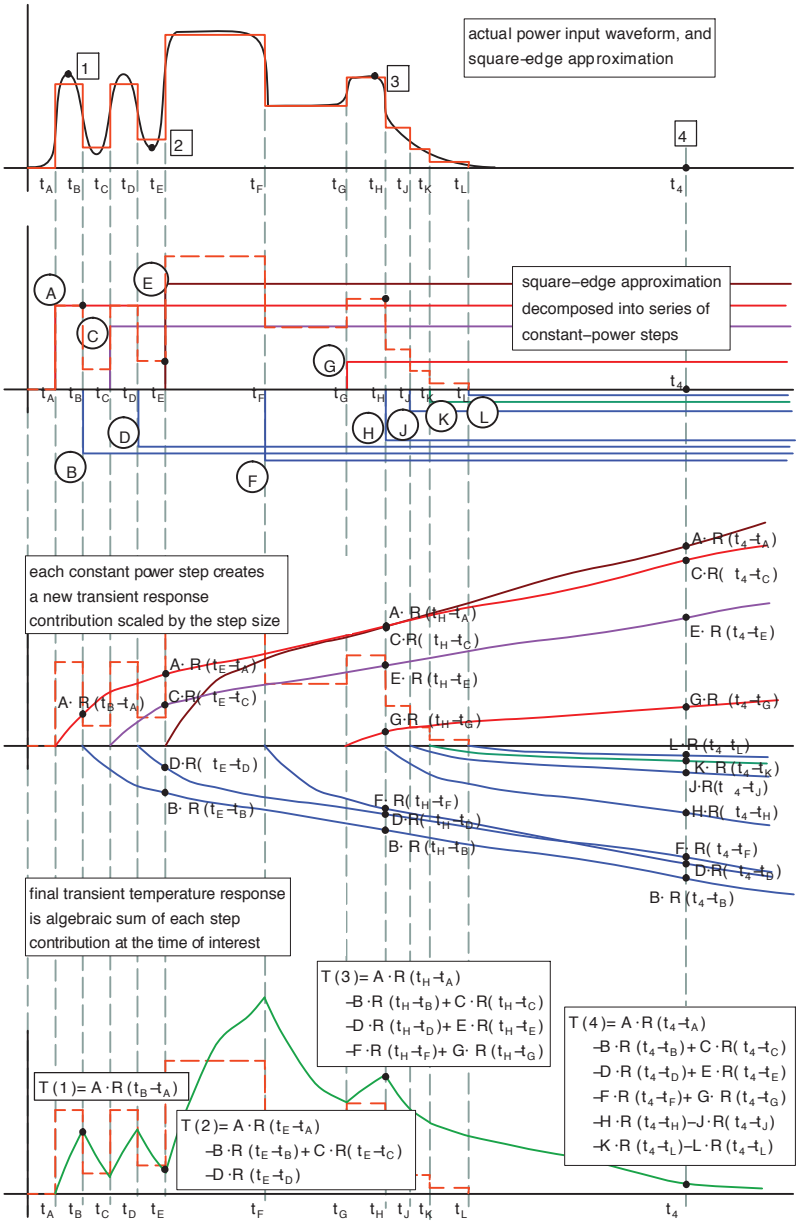


Figure 3. A Complicated Non–Periodic Power Analysis

Beginning with the upper diagram, the black trace represents the actual power input. The orange squared−off trace in that same diagram is an approximation of the actual power input, based on rectangular blocks of constant power. Usually an attempt is made to keep the total energy input correct, i.e. to make sure the area under the approximating curve is the same as that under the true power curve. The points labeled 1−4 are points at which a temperature estimate is desired. Points 1 and 3 are intended to capture the local peak temperatures, and point 2 is intended to capture the local minimum temperature. This means that the choices of the precise times for the square−edged approximation will dictate the times at which these local extrema will occur, and will not necessarily be at the exact times dictated by "real" power input waveform. In the second diagram, the approximated power input (now shown as a dashed orange line), is broken down into a sequence of steps of constant power, labeled A−L. The relative amplitudes of the power steps are the changes in power as each new step is introduced. Thus note that A, C, E, and G are positive going steps, and all the others are negative going steps. In the third figure, the transient responses due to each step of power are indicated. Each response is the basic single−pulse response scaled by the amplitude of the power step in question. In the bottom figure, the resulting temperature profile is illustrated, and the net contribution of the various pieces needed to compute the temperature (at the four points of interest) is indicated. The response at point 4 is shown mainly to illustrate that the temperature may be computed at *any* time of interest, whether or not it corresponds to a change in power; all that is needed is to include every step (in this instance, all of them) that has begun prior to the moment of interest, computing its contribution based on the elapsed time since each initiation. Clearly this could be done anywhere during the power input profile, for instance, anywhere between t_F and t_G.

One difficulty that sometimes arises, is that an $R(t)$ curve may need to be read with great precision when no such precision is apparently available. The preceding example, for instance, requires several pairs of nearby $R(t)$ values to be read accurately, so that a small difference between them may be used to calculate a temperature change. The recommended procedure is to assume that over small intervals, the transient curve can be represented by a power law (which will be a straight line segment on a log−log graph), of the form:

$$R(t) = a \cdot t^n \qquad \text{(eq. 19)}$$

Given two points far enough apart to span any particular range of interest (yet close enough together to be connected by an effectively straight line segment), the power law exponent n may be calculated using:

$$n = \frac{\log[R(t_2)/R(t_1)]}{\log[t_2/t_1]} \qquad \text{(eq. 20)}$$

Then the value a small distance ϵ from a nearby value at t may be calculated from:

$$R(t + \epsilon) = R(t)\left(1 + \frac{\epsilon}{t}\right)^n \qquad \text{(eq. 21)}$$

Yet a different approach to using the single−pulse heating curve may be possible for estimating peak temperatures of short wave trains with long gaps between the trains, but repeated periodically. For instance (refer to Figures 4 and 5), consider ten, 100 W pulses, with a 1−millisecond period and a 5% duty cycle (i.e. 0.05 ms on, 0.95 ms off); then every 100 milliseconds, repeat the same pulse train. Do this for 45 seconds. The question is, what is the highest temperature reached during this scenario?

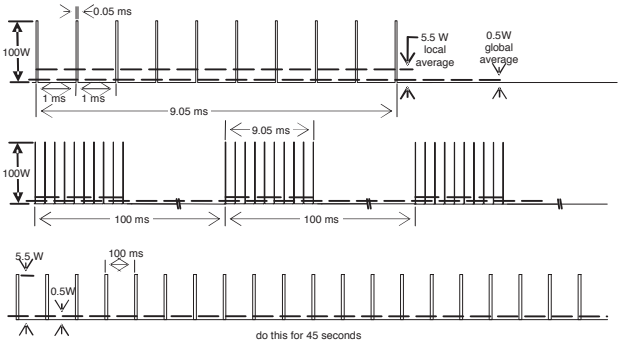


Figure 4. Periodic Short Pulse Trains Problem

Before we can answer that, we must first determine *when* this will occur, which is straightforward in this example. It will occur at the end of the 10th pulse at the end of the last pulse train at the end of the 45−second point (i.e. the end of the very, very last pulse to be applied). So now that we know *when* to find the temperature rise, we may find the *amount* of the temperature rise, which we can break down into three pieces.

Suppose we're working with the device whose transient response is given in Figure 2, on the 736 mm^2 Cu area board. First, as seen in the upper portion of Figure 4, there is an overall average power dissipation to be concerned with. Out of every 100 ms, there are 10x0.05 msec of 100 W power applied – and the power is zero the rest of the time. So we have an average power of (10*100*0.05/100) = 0.5 W, when looking at the overall waveform. This average power adds a "background" temperature rise to the system, which we thus obtain by pretending there is a constant 0.5 W of power applied for 45 s (last portion of Figure 4). Clearly in the environment we've chosen for this example, the overall system hasn't reached steady state by 45 s. But from the single−pulse curve, we *can* say that at 45 s, $R(t)$ is about 30°C/W, hence the background temperature rise will be 0.5 W * 30°C/W, or 15°C.

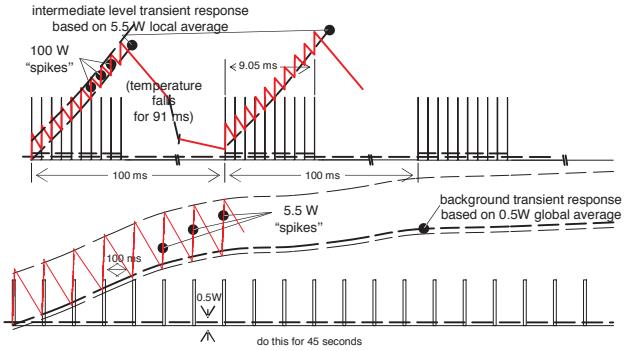


Figure 5.

Similarly, each pulse train of 10 fast pulses may be viewed as a short block of "constant" power that rides on top of this background average. Over a block of 10 fast pulses, we have an "average" power of (10*100*0.05/9.05) = 5.5 W (note that from the beginning of the first of ten pulses, to the time the 10th pulse turns off, is 9.05 ms). Between each of the 10–pulse trains, the power is off for so long (90.95 ms) that the temperature will basically fall all the way back to the "background" temperature rise, whatever that may have been at the beginning of the current train (which, again, is what ends up being about 15°C by the 45 s point in the problem). So again, reading the single–pulse curve at 9.05 ms, we find R(t) of about 3.6°C/W, so at the end of the 10th pulse within each brief train, the average temperature rise will be (5.5 – 0.5) W * 3.6°C/W, or about 18°C. (Note that we have subtracted out the 0.5 W of "background" average power, so as not to double–count it in computing the temperature rise of the average pulse train.)

Finally, we have to ask what the temperature rise of each individual 100 W pulse is around the quasi–average junction temperature of this question. We've already determined that by the end of 45 ms, 0.5 W will have translated into a 15°C rise, and another 5 W will result in a net 18°C rise between the first and last pulse of the ten, every time a 10–pulse train fires off. So what happens on that very last pulse of each ten? 94.5 W of the 100 W (having already accounted for the effect of 5.5 W), goes into "spiking" the temperature on top of whatever it was just before that pulse started. Again, from the single–pulse heating curves, R(0.05 ms) is about 0.45°C/W, so a final 94.5 W pulse adds another (94.5*0.45=) 43°C to the cumulative temperature rise. Our conclusion, therefore, is that the peak temperature at the end of the final pulse at the 45 s point in the example, will be 15+18+43=76°C above ambient. Clearly, due to the lack of precision in reading the single–pulse response chart, there is some inherent uncertainty in this result. If more accuracy than this is desired, thermal RC models may be used. These will be discussed subsequently. Before this topic is addressed, however, another application or extension of the use of transient heating curves needs to be covered.

Duty Cycle Curves

A transient heating curve may also be known as a "single pulse heating" curve. This is because it is derived from either an experiment or a model that produces the junction temperature rise in response to a sudden application of constant power – the longer the heating power, or "pulse," is applied, the hotter the junction. Clearly, the heating power could be turned off at any moment, and the temperature rise at that instant would be known – so whether the pulse is in reality turned off, or simply continued as more data is collected at longer and longer times, the resulting plot may be interpreted as being the result of a single pulse whose "width" is indicated on the x–axis. This leads us naturally to the question of what happens if a pulse is repeated periodically, rather than applied once and never again (or in any other way non periodically, as in the previous examples).

If a pulse train of square pulses, of equal width and regularly spaced, is applied to a device, it turns out that the single pulse heating curve (just described) may be transformed into a family of curves, each of which represents the peak junction temperature that will be eventually reached once the pulse train has been applied for a long enough period of time (see AND8219/D). These curves are generally called "duty cycle" curves, and are parameterized by the percent of "on" time. In Figure 6 following, the single pulse curve for the 1" pad thermal test board, shown previously in Figure 2, has been so transformed, and the resulting family of duty cycle curves is presented.

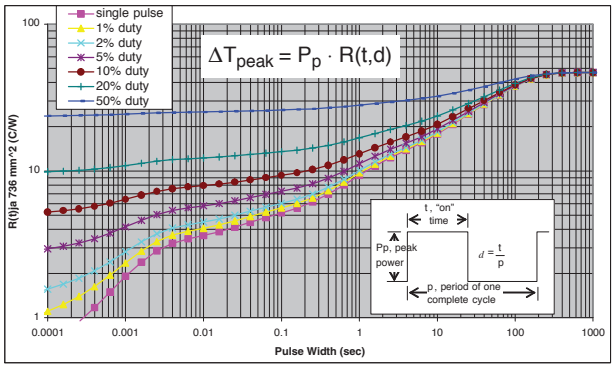


Figure 6. Typical Duty Cycle Curves

The x–axis of the chart is the individual pulse width, that is, the "on" time. So if the pulse width is "t" and the total length of a cycle (on time plus off time) is "p", then the duty cycle, d, will be the ratio of t/p (and is usually expressed in terms of percent, as shown in the figure). To read the chart, then, one figures out the percentage of "on" time for the pulse train of interest, and then looks up the appropriate transient response value for that specific "on" time (on the corresponding % duty cycle curve). This gives a value, often denoted R(t,d), in units of °C/W. Since these are square

waves, the peak power is the height of the individual pulses (note also that the average power will be d^*P_p, the duty cycle times the peak power). Note that to calculate *peak* temperature rise, one must use the *peak* power for the pulse, *not* the average power.[3]

The most common mistake made in using duty–cycle curves, is that they may only be applied for situations where the original single–pulse curve was the "correct" curve for the application environment under consideration. This means that it must end up at the correct steady state value. If the single pulse curve was for a min–pad board, for example, then *none* of the resulting duty cycle curves may be used for a 1″ pad application, regardless of how short the pulses or what the duty cycle. To understand this, it may be helpful to consider the mathematical expression typically used to derive these curves:

$$R(t, d) = (1-d) \cdot R(t) + d \cdot R(\infty) \qquad \text{(eq. 22)}$$

If d is vanishingly small, then the result is the original curve (which is clearly valid only until environmental effects come into play). For any finite d, however, *regardless* of how short the pulses of interest (i.e. "on" time), the duty cycle curve carries along a contribution from the steady state end of the original curve, i.e. $R(\infty)$.

Given the appropriate single–pulse curve, if the pulse train is periodic (even if not square), the square–wave duty cycle curves may provide a time–saving approximation. For instance, pulses that are trapezoidal or triangular in shape, partial sinusoids, etc., may be approximated by square pulses with the same total energy (i.e. area under the pulse), where the height and width of the equivalent square pulse are adjusted such that the end of the pulse coincides with the moment of peak temperature – though this itself may require some experience to judge when that is likely to occur.

Thermal RC Network Models

Thermal RC network models are an alternative way of describing the same transient thermal response previously discussed (see AND8214/D and AND8221/D). An entire transient response curve can usually be represented in just a handful of resistor and capacitor elements. If the correct computational tools are readily available, RC networks may therefore be a convenient and compact representation. Two general forms of RC networks are possible, those with grounded capacitors, and those in which the capacitors are not grounded. These will be discussed in turn.

[3] (Think *peak power* for *peak temperature*, if that helps. But to see why this must be so, think about that single–pulse curve being the equivalent of the 0% duty cycle case. For a given pulse width, if the only thing that changed was the period, you'd be staying at the same position on the x–axis while you moved from one curve to another. As the period went to infinity, you'd end up on the 0%, or single–pulse, curve; but if the power you multiplied by was the *average* power, you'd also be moving toward a zero–power average, hence the temperature rise would approach zero for a fixed pulse width. Obviously this would be incorrect – because the whole point of the single–pulse curve is to give you the actual temperature rise based on the power level of the pulse *while it's on*, so clearly you're supposed to be using the instantaneous power level, not the average power level.)

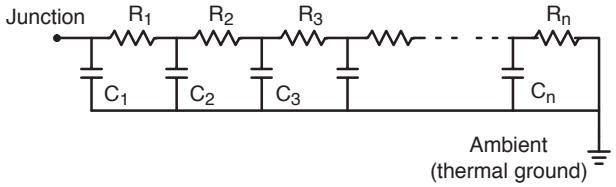


Figure 7. Grounded Capacitor Thermal Network ("Cauer" Ladder)

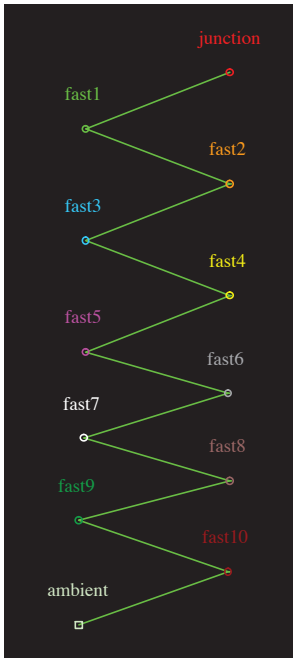


Figure 8. RC Network Schematic
Capacitors (not shown) tie each node to ground

Figures 7 and 8 illustrate a typical "grounded–capacitor" thermal ladder network. In fact, any network topology of resistors might be chosen to represent a physical thermal system (i.e. not just a linear string of resistors, but just as well a star, a bridge, or whatever). The main advantage of a grounded–capacitor network is that it derives from the fundamental heat–transfer physics. Every node in the network is connected to thermal ground through a capacitor. If a simple chain of nodes is used, it is convenient to draw the network as shown in Figure 7 because it resembles a ladder, though because the lower edge of each rung attaches directly to ground, the connections between the rungs are essentially through the resistors. Often for clarity, the capacitors are omitted entirely, in which case Figure 8 is an equivalent model. A grounded capacitor network such as shown in Figure 7 or 8 is known as a *Cauer* ladder.

Because this network derives from the real physics, there is at least a chance that experimental data from various points within the physical system can be correlated with specific individual nodes of the network model. As we move from junction to ambient, for instance, we might find physical locations corresponding to nodes in this order: silicon junction, back of silicon chip, edge of leadframe, lead (at package boundary), lead (at board interface), board (at some distance from package), and finally ambient. Of course, we may not *have* any intermediate location data to correlate with, or the intermediate data we have might not happen to land "on" a node of the model (rather, somewhere in between nodes). Also, the physical system might *not* be well represented by such a simple chain of resistors, so no correlation might be possible except at the junction itself. (This is actually quite typical, for in many environments, the heat flow follows at least two separate and distinct paths from junction to ambient, e.g. upward through the case, outward through the leads into the board, and downward through the air gap and thus directly to ambient on the back side of the board. When the heat flow is believed or known to flow along multiple parallel paths, it clearly would be better to model the system with a more complex network. Likewise, one would not expect to find a nodal correlation with physical locations if the network was willfully chosen as a simple ladder, when multiple significant parallel paths were present.) Only in the case where a single path to ground dominates heavily, would such a simple linear resistor topology be expected to yield good correlations at the intermediate nodes. Nevertheless, the point is that there *could* be such a correlation.

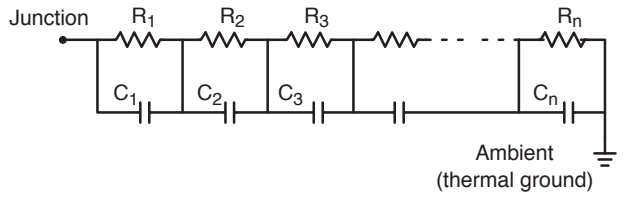


Figure 9. Non−Grounded Capacitor Thermal Ladder ("Foster" Ladder)

Non−Grounded Capacitor Networks

Contrast the grounded network of Figure 7, with the non−grounded−capacitor network of Figure 9. Figure 9 is a true "ladder" of resistors and capacitors, and is sometimes known as a *Foster* ladder. Each rung is joined to the next rung (and only to the next rung) through both the resistor and the capacitor; only the final capacitor is directly connected to thermal ground.

Difficult though it may be to grasp at first, this network has no physical basis. In the thermal/electrical analogy, a thermal capacitor is simply an element that stores energy based only on *one* temperature, that is, the temperature of the node whose thermal mass it represents. (In contrast, an electrical capacitor stores energy based on the *difference* in voltage between its two terminals.) Hence a thermal capacitor whose energy storage is based on a difference in temperature between two ungrounded nodes in a network (as is the case of most of the capacitors in Figure 9) has no physical meaning. However, there is a mathematical simplicity underlying Foster ladders. In their mathematical description, one finds that each resistor−capacitor pair contributes an "amplitude" to the overall system response, and a unique time constant associated with that amplitude. Indeed, a Foster ladder may be viewed as nothing more than a schematic of the mathematical fit to a real transient response curve. Given a transient response curve of junction temperature vs. time, a series of exponential terms consisting of amplitudes and time constants may be fit to the curve to whatever degree of accuracy is desired. (More terms usually implies a better fit.) Once done, the terms may be interpreted as an RC ladder (i.e. the Foster ladder) where each amplitude is a resistor, and each time constant is the product of its associated resistor and a capacitor in parallel with it.

Comparison and Contrast of Cauer and Foster Ladders

Clearly, the mathematical terms representing (or represented by) a Foster ladder may be added together in any order to achieve the same sum. Thus the rungs (RC pairs) of the schematic may be listed (or diagrammed) in any order and still represent the same response! Because the overall response (from junction to ambient) is immune to reordering of the individual rungs (as long as each RC pair remains a pair), the temperature that might be calculated at any other node *between* any two rungs is physically meaningless. By contrast, though a Cauer network must necessarily have a mathematical representation comprising amplitudes and time constants, one finds that *every* amplitude and *every* time constant depends on *every* resistor and *every* capacitor, in a highly complicated and algebraically intractable tangle intimately dependent on the physical location of the elements in the network.

Even so, Foster networks are typically drawn with the rungs placed in order from junction to ambient with the smallest values (i.e. fastest responding rung) at the junction end, and the largest values (i.e. slowest responding rung) at the ambient end. This is superficially similar in appearance to a typical Cauer ladder, which almost always (and necessarily) has the smallest elements nearest the junction, and the largest elements nearest thermal ground. But in the Cauer ladder, the choice is not arbitrary; rather it is imposed by the intrinsic relationship between time response and location in the model.

So where does this leave us? A grounded−capacitor model is most useful when a physically meaningful model is required, for instance, to separate the package from the environment in order to replace the environment portion of the network with a different network, representing a different environment. However, to best use a grounded−capacitor model, a circuit simulating tool is required. Of course, if a circuit simulating tool is being used for thermal calculations, any complicated, time varying, power input may be imposed on the circuit without particular cost. Finally, multiple heat source models can be built with equal facility, and again, arbitrarily complex asynchronous power inputs at any number of nodes may be managed without difficulty.

On the other hand, non−grounded−capacitor models are mathematically very simple, and quite detailed thermal calculations may be performed with spreadsheet based tools. Though only the junction has the designed correlation with physical reality in a non−grounded capacitor model, the fact remains that if the model is capable of producing the known transient response, applicable to the particular environment in question, then it may be used successfully for temperature predictions with arbitrarily complex power inputs. Even multiple input thermal models may be constructed using equivalent non−grounded networks.

It should also be mentioned that every Cauer ladder has a Foster equivalent, and vice versa. The conversion from one to the other is a non−trivial operation, but algorithms do exist for that purpose. In ON Semiconductor data sheets, generally the two equivalent networks are provided, enabling a knowledgeable customer to take advantage of the strengths of each.

Multiple−Junction Devices and Transient Response

In the preceding section, allusions have been made to multiple−input transient models. Just as with steady state descriptions of a thermal system, transient descriptions of multiple−junction devices may be constructed. If a matrix method is followed, the only difference is that every element of the matrix is a function of time. For every heat source in the device, there will be a “self heating” transient response curve, and for every other point of interest in the system (whether another heat source or a passive temperature monitoring location) an “interaction” transient response curve will exist.

Bounded by the same limiting assumptions, the principles of linear superposition and reciprocity continue to apply. That is to say, the time−varying response at any point in the system may be treated as the linear superposition of its response to each independent heat source, as if each heat source was powered individually and independently of the others. Further, the less intuitive truth of the reciprocity theorem applies in the time domain: namely, the transient response of point “A” in the network due to (constant) heat input at some other point ‘B’, will be identical to the transient response at point ‘B’, if the same amount of heat input is applied at point ‘A.’ So in the matrix description, symmetry across the main diagonal will still exist. Perhaps the most powerful implication of the reciprocity theorem is experimental: in effect, only half of the total possible interaction thermal transient responses need be measured.

Circuit Simulators

Cauer models are of little use without a circuit simulator, due to the messy algebra required to describe their mathematical response. Thus, if a Cauer model is all that is available (at least, if it consists of more than about two nodes), a circuit simulator is a must. Of course, if a circuit simulator is available, a circuit is a circuit, and it should therefore be evident that both Cauer and Foster ladders may be analyzed with equal ease. Indeed, for single−input networks, there will be no difference in the overall approach, only in the details of the network connectivity and element values.

For multiple−input networks, the Cauer network will be straightforward (see Figure 10). Recalling that a Cauer network will have been derived on some premise of having physical significance, the interactions between the various possible heat sources (and possibly passive “boundary” nodes) will be built into the topology of the network itself. Resistors and capacitors will exist that “automatically” provide the correct interaction responses due to heat inputs at every source; reciprocity and superposition are necessary consequences of the method. One simply enters the grounded−capacitor Cauer model into the simulator, with all nodes and interconnections explicit in the schematic, and the task is done.

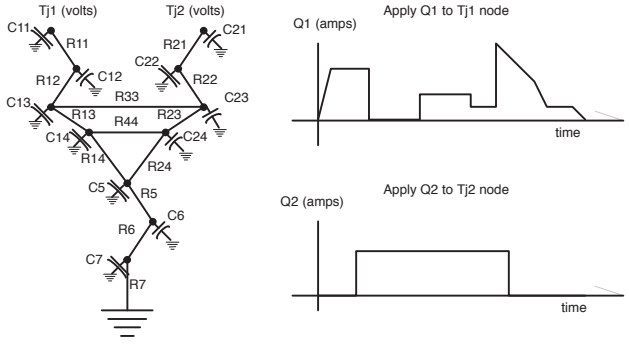


Figure 10. Implementing a Multiple Input Cauer Network in a Circuit Simulator

A multiple−source Foster model is more complicated to implement in a circuit simulator, and exactly how it is done will depend on the features of the simulator available. Since a Foster model is nothing more (in essence) than a schematic of the mathematical fit to a transient response curve, the resistors and capacitors in a particular “self heating” Foster ladder will not correlate with the resistors and capacitors in any of the interaction networks; nor will the self heating Foster ladder elements of any two heat sources have any correlation with each other, even though we may know that there is much underlying common thermal path between the two sources. Moreover, depending on how the Foster ladders were derived, even the time constants between various curves in the model may not match! (It may be observed that if “closed form” Foster ladders are derived from a Cauer model directly, at least the time constants will be shared across all the self heating and interaction curves; but if the Foster ladders are simply independent mathematical fits to response curves, from whatever source, there is no particular reason the time constants will match, unless by intentional choice.) Again, depending on how the Foster ladders are derived, there may even be “negative” amplitudes (this is, in fact, guaranteed to occur in “closed form” solutions to the interaction responses in a physically significant grounded−capacitor models.) Clearly, if negative amplitudes arise in the Foster representation, a circuit simulator must permit negative resistances (and obviously negative capacitors, since a positive time constant can only result for a negative resistor, from its product with a negative capacitor). Alternatively, the simulator must provide a programmatic method of subtracting the response of one node from another, so a negative contribution can be constructed from positive sub−circuits. Similarly, to implement a multiple input Foster model in a circuit simulator, care must be taken to intentionally create “summing” nodes that implement the principle of linear superposition between the various otherwise independent self heating and interaction heating portions of the overall model. If the circuit simulating tool does not provide features adequate for these tasks, a spreadsheet implementation will be the best alternative. Figure 11 illustrates the possible steps.

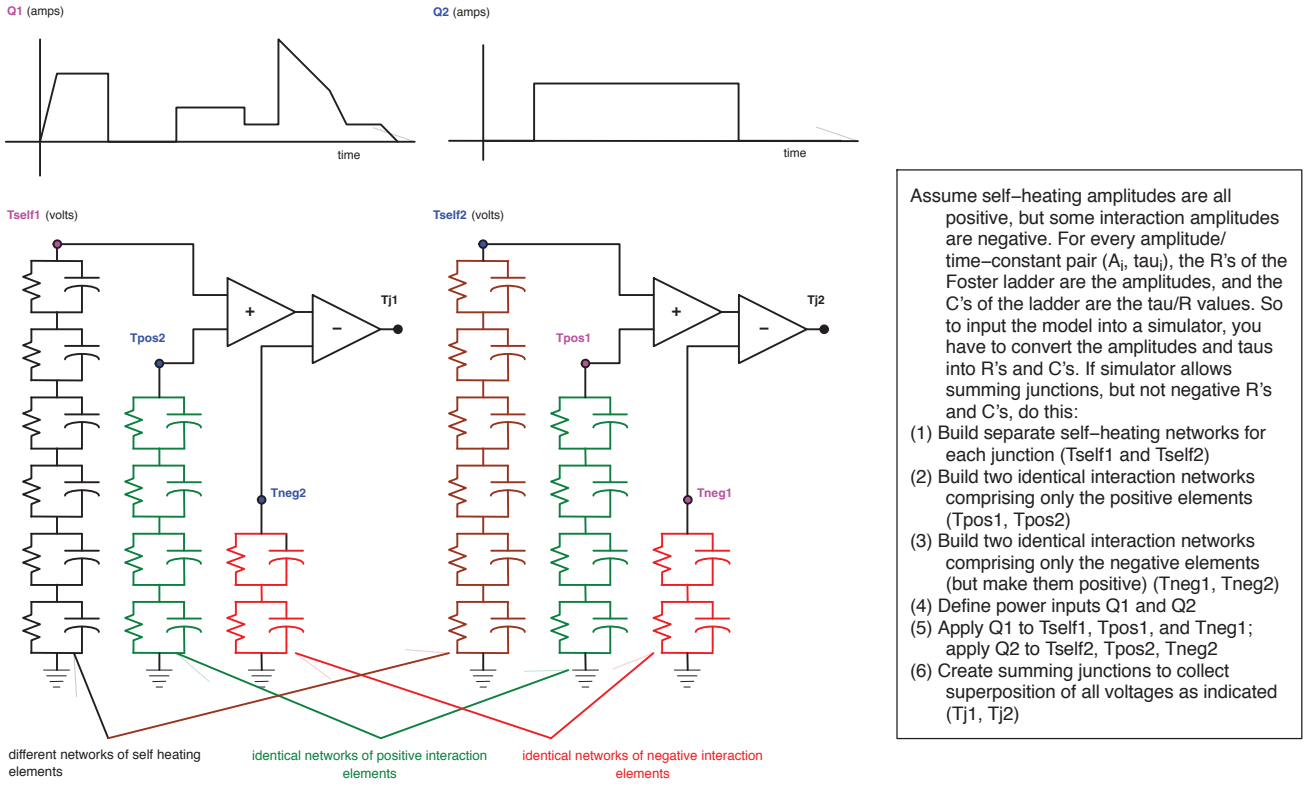


Figure 11. Implementing a Multiple Input Foster Network in a Circuit Simulator

Spreadsheet Models

As previously described, Cauer models basically require a circuit simulator, even for single−input models. For Foster ladders, however, spreadsheet tools are convenient for implementing both single input and multiple input models. This is the consequence of the mathematical simplicity of Foster models, and the triviality with which spreadsheets permit superposition to be introduced. For instance, consider the ease with which a constant power transient response for a single−input Foster ladder may be written in Microsoft® Excel. Suppose the following meanings are given to certain cells in a spreadsheet:

cell **A1** is the power level
cells **B1:B10** are the amplitudes
cells **C1:C10** are the time constants (where **C1** is the time constant for the **B1** amplitude, and so forth)
cell **D1** is the time after a constant power step begins

Then the Excel formula to calculate the temperature rise at time D1 is:

{=A1*SUM(B1:B10*(1−EXP(−D1/C1:C10)))}

Though by no means necessary, it may also be noted that by using Excel's *Name* capability and judicious use of *absolute* vs. *relative* reference notation, we can make this formula more mnemonic, and easy to copy to different locations for computing results at many different times. Amending the previous example; the mathematical expression of single−pulse heating curve in terms of Foster−type amplitudes and taus:

$$R(t) = \sum_{i=1}^{n} R_i\left(1-e^{-t/\tau_i}\right) \qquad \text{(eq. 23)}$$

Define Names

power	**$A$1**
amplitudes	**$B$1:$B$10**
taus	**$C$1:$C$10**
time	**D1**

We are now permitted the more readable formula:

{=power*SUM(amplitudes*(1−EXP(−time/taus)))}

If this formula was entered into cell E1, it could be copied down into cells E2 through E100, for instance, resulting in the time response at each of the times found in cells D2 through D100. The *Table* feature of Excel may also be used to advantage to create a table of many values from a single formula.[4]

As time varying power inputs are introduced, and multiple heat sources are introduced, it obviously gets more complicated, but remains quite manageable for a relatively limited number of inputs and time steps. The method is that illustrated earlier (example given in Figure 3), with the following embellishments: (1) the temperature at any point of interest is the superposition of the response at that point due to *all* heat sources (so just as in the steady state case, if a junction temperature is being calculated, there will typically be one self heating contribution, and multiple interaction terms); (2) a new "step" in time must be made whenever the power input changes at *any* heat source, even if there is no change in power at that instant at the point in question (because the interactions will change, even when the self heating contribution does not change).

RC Models and Short−Time Transient Response

It may be shown mathematically that when the time scale is shorter than its fastest time constant, an RC model's transient response becomes proportional to time (i.e. *linear* in time). This will not be a problem if (1) the time scale of interest is somewhat greater than the fastest time constant, or (2) if it is known that the linear response with time is appropriate for the system under consideration. However, as will be discussed subsequently, for many semiconductor devices there is a range of time (typically between about 1 microsecond and 1 millisecond) when the concept of "surface heating" closely approximates the real thermal physics. In surface heating, device transient response is

[4]To those unfamiliar with "array" formulas in Excel, the preceding example accomplishes some very powerful operations in a compact notation. First, the use of the array syntax itself (the colon as part of a cell reference) tells Excel to execute the same computation for each cell of the range in turn; thus, since there are ten cells in each of the two arrays identified, ten parallel computations result. This means that ten different terms representing ten amplitudes and time constants, all evaluated for the same time **(D1),** are calculated at once. Second, the squiggly braces {} surrounding the formula indicate that (in this case) the formula was actually entered into the spreadsheet with the *Ctrl−Shift−Enter* keystroke, rather than the ordinary *Enter* keystroke. (In other words, simply typing in squiggly braces does not accomplish the same thing at all!) This tells Excel that we wish it to actually return all the available array results in however many cells are assigned to the formula. (More typically, this "array formula" entry method is used to distribute the array results over an array of cells; for instance, if we had selected the ten cells **E1:E10** for the formula entry, typed the formula into the formula entry blank, and hit *Ctrl−Shift−Enter,* one of each of the ten individual array results would be listed in each of the ten selected cells.) Here, however, we have no need to see all ten results individually, yet we still wish to access them all even though only a single cell is the target for the formula's result. Thus, finally, we include the **SUM** function to tell Excel to add up those ten individual results, rather than report just the first one in the single cell we've picked for the location of the formula.

proportional to the square−root of time, rather than linear in time. Now a properly constructed RC model (i.e. one with time constants and amplitudes so designed) is capable of following this square−root behavior with excellent accuracy, but only for time scales greater than the model's shortest time constant. It is therefore important, whenever using an RC model, to consider whether the shortest time constant is fast enough for the needs of the analysis. For a Foster ladder, the fastest time constant is known exactly (the smallest RC product in the model). For a Cauer ladder, a good (although not exact) estimate of the fastest time constant is similarly obtained as the product of the RC pair closest to the junction. (In fact, ill−formed Cauer ladders, depending on exactly how the R's and C's appear in order, may yield actual fastest time constants several orders of magnitude faster than that of the RC pair nearest the junction; alternatively, ill−formed Cauer ladders may result in Foster amplitudes that are infinitesimal, yet with finite time constants. However, these are aberrations that do not materially affect the ability of the network to respond to square−root−of−time needs on the order of the RC pair nearest the junction.) In any case, if the shortest legitimate time constant is not smaller than the shortest time scales of interest, and especially in the microsecond to millisecond time scale, extreme caution should be taken in interpreting RC model results. If a linear model is used when a square−root model is appropriate, temperature changes as predicted by the model will occur much too slowly, and *significant* underestimates of maximum junction temperatures may result.

With that caution in mind, the following table presents RC models for the same D2pak device on two different thermal test boards (basically min−pad and 1″ pad, with some extra trace area included in the total area quantity). For each board, both Cauer and Foster networks are given. It should be emphasized that these Foster networks are in fact the exact mathematical equivalents of the corresponding Cauer networks. A number of the concepts addressed in the preceding discussions may be illustrated.

Table 1. RC Networks ("R" values are °C/W; "C" values are J/°C; "tau's" in seconds)

Drain Copper area (1 oz thick)			241 mm^2	788 mm^2		241 mm^2	788 mm^2	
(SPICE deck format)			Cauer network			Foster network		
			241 mm^2	653 mm^2	units	Tau	Tau	units
C_C1	**Junction**	**Gnd**	**6.3269E−6**	**6.3269E−6**	W−s/C	**2.9892E−7**	**2.9892E−7**	sec
C_C2	**node1**	**Gnd**	**2.9939E−5**	**2.9939E−5**	W−s/C	**4.3949E−6**	**4.3949E−6**	sec
C_C3	**node2**	**Gnd**	**8.9817E−5**	**8.9817E−5**	W−s/C	**3.8122E−5**	**3.8122E−5**	sec
C_C4	**node3**	**Gnd**	**1.9877E−4**	**1.9877E−4**	W−s/C	**2.9542E−4**	**2.9542E−4**	sec
C_C5	**node4**	**Gnd**	**1.3388E−3**	**1.3388E−3**	W−s/C	**2.3055E−3**	**2.3055E−3**	sec
C_C6	node5	Gnd	2.5099E−2	2.5099E−2	W−s/C	1.2749E−2	1.2766E−2	sec
C_C7	node6	Gnd	3.1191E−1	3.1815E−1	W−s/C	3.3747E−1	4.1823E−1	sec
C_C8	node7	Gnd	2.2054E−1	4.7830E−1	W−s/C	3.3611E+0	2.7622E+0	sec
C_C9	node8	Gnd	8.8815E−1	1.9594E+0	W−s/C	2.1614E+1	3.0643E+1	sec
C_C10	node9	Gnd	1.8889E+0	6.0036E+0	W−s/C	1.1357E+2	1.2328E+2	sec
			241 mm^2	653 mm^2		R's	R's	
R_R1	**Junction**	**node1**	**0.0578524**	**0.0578524**	C/W	**0.03814**	**0.03814**	C/W
R_R2	**node1**	**node2**	**0.173557**	**0.173557**	C/W	**0.093163**	**0.093163**	C/W
R_R3	**node2**	**node3**	**0.520671**	**0.520671**	C/W	**0.201565**	**0.201565**	C/W
R_R4	**node3**	**node4**	**1.07638**	**1.07638**	C/W	**0.936692**	**0.936690**	C/W
R_R5	**node4**	**node5**	**1.44732**	**1.44732**	C/W	**1.730444**	**1.730479**	C/W
R_R6	node5	node6	0.510799	0.510799	C/W	0.690301	0.691548	C/W
R_R7	node6	node7	2.84846	2.31584	C/W	0.333827	0.60289	C/W
R_R8	node7	node8	9.11661	4.38504	C/W	4.196175	3.230389	C/W
R_R9	node8	node9	34.2576	20.0524	C/W	6.059695	5.266272	C/W
R_R10	node9	gnd	24.9485	11.0277	C/W	60.677683	28.776447	C/W

NOTE: The boldface elements represent the part of the network most closely associated with the package; the remaining elements represent the environment. Listing the Foster rungs in ascending order of time constant provides a rough, though imperfect, equivalent, as the fast response rungs necessarily result in the most significant contributions to the short−time (hence the package) portion of the curve. As emphasized previously, however, the exact location of nodes within the Foster rungs has no direct physical significance, and any apparent correlation with the Cauer resistors is purely coincidental.

First, the fastest time constant for these networks is 2.98E−7 seconds (given exactly in the Foster *Tau* column). An approximation of this value is found by the RC product in the Cauer network closest to the junction, i.e. C_C1 times R_R1, yielding 3.66E−7 s. Second, for convenience (remember, the choice is arbitrary and does not affect the junction response whatsoever), the rungs of the Foster ladder are listed in ascending order of time constant, but it is clear that the R's do not correlate very well with the R's of the "corresponding" rungs of the Cauer networks. Third, beginning at the short time ends of the ladders (both Cauer and Foster), the models are identical between the two boards. That is to say, for single pulse heating response, only the package (which is the same) matters at first, and only after some time has passed and the heat has begun to cross from package into the board, does the environment influence the response.

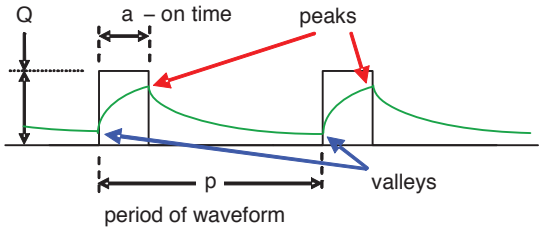


Figure 12. Basic Square Wave

Periodic Waveforms Using Foster RC Models

Square−wave duty−cycle curves have been discussed, and they are often derived from the simple formula expressed previously in Equation 22. However, given an RC model (in particular, the amplitude/time−constant Foster expression) of a single−pulse transient curve, exact closed form solutions for an infinite train of equal square pulses may be derived. We will simply present several of these solutions and illustrate how they may be applied (see AND8219/D). Given the single−pulse heating curve formulation of an n−rung RC model as indicated in Equation 23, we have the following:

$$R(a,d) = \sum_{i=1}^{n} R_i \frac{1 - e^{\frac{a}{\tau_i}}}{1 - e^{\frac{p}{d\tau_i}}}$$ peaks of simple square wave train of duty cycle d, on−time a (eq. 24)

$$Y(a,d) = \sum_{i=1}^{n} R_i e \left(1 - \frac{1}{d}\right) \frac{a}{\tau_i} \frac{1 - e^{\frac{a}{\tau_i}}}{1 - e^{\frac{a}{d\tau_i}}}$$ valleys of simple periodic square wave train (eq. 25)

Observe that the on−time, period, and duty cycle of the waveform are related through the equality $a = p \cdot d$. When the on−time is plotted on the x−axis, and the duty cycle is used as a curve parameter, Equation 24 gives us the family of duty cycle curves previously encountered in Figure 6, based on the Foster RC resistor model as fit to the original R(t) single−pulse heating curve. Indeed, if the RC model is a good fit, the duty cycle curves derived from Equation 24 will be more accurate than if they are derived from the more approximate formula of Equation 22 (with the possible exception that for very short duty cycle values, and on−times smaller than the smallest RC time constant, we may have the same limitation related to square−root of time previously discussed).

When a single pulse is repeated, (Figure 12), obviously the peaks occur at the ends of the "on" times, and the valleys occur at the ends of the "off" times (i.e., the beginning of each "on" time). Further, when only a single square pulse is repeated, it doesn't matter where the pulse is positioned within the period, if all one is concerned with is the peaks and valleys. In fact, for convenience these preceding formulas were derived under the assumption that the "on" time of each pulse begins at the beginning of each cycle.

However, if we generalize the problem slightly and allow a single square pulse to be positioned at an arbitrary point within the cycle, some more powerful formulas may be derived. For the formulas that follow, Figure 13 defines the parameters for a generalized square pulse within a period of length p. All times are relative to the beginning of a cycle.

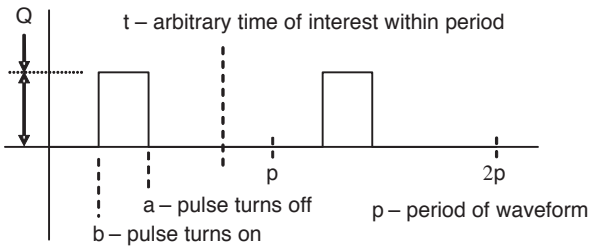


Figure 13. Generalized Square Wave

After an infinite number of identical cycles, the following three formulas describe the shape of the temperature response for the ranges indicated:

$$F(a,b,p,t) = \sum_{i=1}^{n} R_i \left(\frac{e^{\frac{a-t-p}{\tau_i}} - e^{\frac{b-t-p}{\tau_i}}}{1 - e^{\frac{p}{\tau_i}}} \right)$$ good (computable) only for $0 \le t < b$ (eq. 26)

$$F(a,b,p,t) = \sum_{i=1}^{n} R_i \left(1 + \frac{e^{\frac{a-t-p}{\tau_i}} - e^{\frac{b-t}{\tau_i}}}{1 - e^{\frac{p}{\tau_i}}} \right)$$ good (computable) only for $b \le t < a$ (eq. 27)

Note: if t = 0 and b = 0, Equation 25 results

$$F(a,b,p,t) = \sum_{i=1}^{n} R_i \left(\frac{e^{\frac{a-t}{\tau_i}} - e^{\frac{b-t}{\tau_i}}}{1 - e^{\frac{p}{\tau_i}}} \right)$$ good for $0 \le t \le p$ (computable) only for $t > a$ (eq. 28)

Note: if t = a and b = 0, Equation 24 results

For these formulas, the “computability” restriction is a practical matter arising when positive arguments appear in the exponential terms of the various numerators. Note also that these formulas describe the response curve, but the power level of the applied pulse has not yet been considered. We defer consideration of pulse power to the following formula, which now expresses the completely general superposition of any number of square pulses occurring with the same frequency, all positioned within the same cycle of period p:

$$G(t) = \sum_{k} Q_k F(a_k, b_k, p, t) \qquad \text{(eq. 29)}$$

Equation 29 now permits us to predict the “steady−state” transient behavior of any complexity of periodic power, assuming we break down the cycle into a series of square−edged pulses − a process we illustrated (see Figure 3) in an earlier example of a non−periodic waveform. By “steady state” transient response, we mean the shape of the temperature response curve for a typical cycle after an infinite number of such identical cycles has occurred. An important observation now must be made. Whereas the “peak” and “valley” temperatures for an infinitely repetitive *single* pulse can be predicted (i.e. Equation 24, 25) without knowing the details of the profile, this is not possible for a general periodic waveform, even when the waveform is a relatively simple combination of just a few square sub−pulses. Consider the following example, with the periodic power input of Figure 14 applied to the RC model given in Table 4.

Table 2. RC Model for 3−Pulse Example

tau (sec)	R [°C/W]	tau (sec)	R [°C/W]
1.00E−6	0.01104	1.00E−1	1.128566
1.00E−5	0.012806	1.00E+0	3.539626
1.00E−4	0.069941	1.00E+1	5.423616
1.00E−3	0.275489	1.00E+2	12.08694
1.00E−2	0.019806	1.00E+3	16.2933

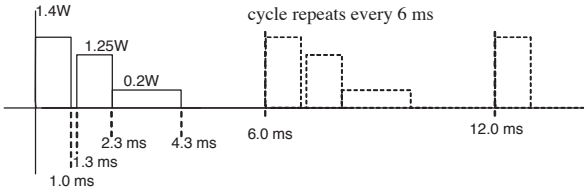


Figure 14. 3−Pulse Periodic Input

Applying Equation 26, 27, and 28 to the respective portions of each of three separate square pulses comprising the repeated pattern, and Equation 29 to superimpose their effects, we find the following temperature response:

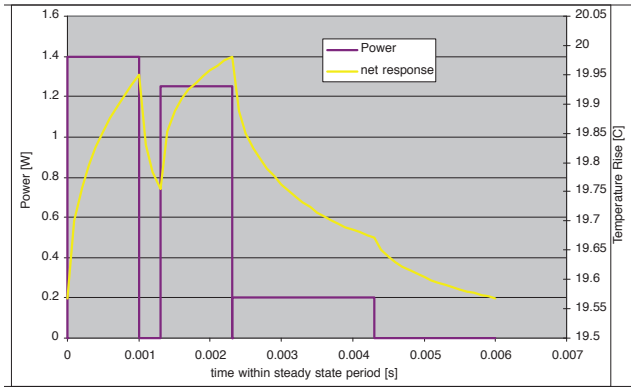


Figure 15. 3−Pulse Periodic Example Steady−State Transient Response

What makes this example particularly interesting is that the peak temperature (during a steady state cycle) occurs at the end of the *second* pulse, which has a lower power and even a small gap of zero power, between it and the higher power pulse immediately preceding it in the cycle. Knowing that the single pulse response is proportional to power, and that the peak temperature always occurs at the end of a square pulse (even when infinitely repeated), one might easily overlook the possibility demonstrated here. In other words, for a generalized periodic waveform, even when constructed of (or perhaps approximated by) just a small number of square sub−components, one does well to compute the response throughout the entire range of a cycle, not at just the “obvious” points.

Surface Heating, Square−Root of Time, and Short−Time Transient Response

In most thermal transient tests, experimental data is acquired as early as 1E−5 s (10 microseconds). However, in most cases, due to electrical switching transients, it is inconsistent between test devices out to as late as 1E−3 s. Even when measurement consistency occurs at earlier times, it is rarely reliable (meaning, consistent with expected theoretical behavior) at times earlier than 1E−4 s. Usually, in fact, measured signals corresponding to expected theoretical behavior do not occur until somewhere between 3E−4 s and 1E−3 s. The two major contributors to this correlation (or lack thereof) are electrical transient effects in the devices (from the measurement perspective), and die geometry effects (from the theoretical perspective).

More specifically, die thickness, and actual active heated area as compared to overall die size, affect the theoretical behavior. The simplest, commonly used theory for short−time thermal transient behavior is the surface heating model. This assumes constant power, 1−dimensional heat flow (i.e., flow direction is exactly perpendicular to the heated surface, with no spreading effects), and results in a surface temperature rise that is proportional to the square−root of heating time.

Because of this, it is often referred to as “sqrt(t)” heating. An important aspect of sqrt(t) heating is that on a log–log plot (refer back to Figure 2), such a heating “curve” is a straight line that rises one decade of temperature (or thermal resistance) for every two decades of time increase (i.e. a factor of 10 in temperature for every factor of 100 in time – hence the sqrt(t)). On the log–log plot this appears, therefore, as a 1:2 slope. The vertical position of this theoretical line (the intercept on the log–log plot, or the proportionality factor on a linear vertical scale) is determined by the area being heated, and the material properties of the die and whatever material adjoins the heated surface of the die. (Typically, this adjoining material is mold compound, which effectively lowers the heating rate of a silicon free surface by about 10%. However, if the silicon is covered by a copper “clip,” the short–time heating rate may be lowered by as much as 70%.) Also consequent to the sqrt(t) theory, the thinner the die, the sooner the heat reaches the back side of the silicon and thus ceases to follow the sqrt(t) model; a die of half the thickness will thus end its sqrt(t) behavior in one fourth the time. Typically, we consider that theoretical behavior should persist until about 1E–3 s for a 15 mil (380 micron) thick die, but when the thickness is as small as 10 mil (250 micron), theoretical behavior will last only 4E–4 s; for a 7–mil (180 micron) thick die, sqrt(t) will last for only 2E–4 s. Die thickness is also directly related to the other “extreme” of transient behavior, that is, how long it will be until local steady state (meaning the final temperature gradient, or maximum temperature difference between heated surface and back of silicon) is reached. All else being equal, this should take no longer than 2.5E–3 s for a 15 mil die, and 5E–4 s for a 7 mil die (as before, thickness increasing in proportion to the square root of the time required).

On the other hand, a lumped parameter RC model, due to the exponential nature of the equations describing its behavior, will always become linear in time as the shortest times are approached. Therefore an RC model will always fail to approximate a sqrt(t) behavior if taken to times smaller than its shortest time constant. Hence, as already discussed, if it is known that sqrt(t) behavior is a reasonable approximation to the actual behavior (typically in the 1E–6 to 1E–3 s range), but the RC time constants do not begin lower than this range, the sqrt(t) model should be used directly for short–pulse temperature estimates, otherwise severe underestimates of temperature changes will result. (A final caution, however, is that extremely short *repetitive* pulses require a more thorough analysis – one approach being to extend an RC model’s short–time response down into the required time scale where it accurately approximates the sqrt(t) behavior. (See AND8218/D for details.)

The following tables provide definitions and formulas useful for 1D surface heating estimates, and some typical material property values found in semiconductor packaging.

Table 3. 1D Surface Heating Formulas and Definitions

	Thermal Property
$\theta\ (t) = b\sqrt{t}$	Temperature rise at surface [°C/W]
$b = \frac{\Theta}{A}\frac{2}{\sqrt{\pi}}\frac{1}{\eta}$	Sqrt(t) proportionality[°C/W/√s]
$\eta = \sqrt{k\rho\ cp}$	Thermal effusivity [W√s/mm^2/°C]
$\eta_e = \eta_1 + \eta_2$	Effective effusivity for heating at planar interface between two materials
$\tau = \frac{L^2}{\alpha}$	Characteristic time through thickness L [s]
$\alpha = \frac{k}{\rho\ cp}$	Thermal diffusivity [mm^2/s]

where:

Q	Heat input (W)
A	Heated area (mm^2)
k	Thermal conductivity (W mm^{-1} C^{-1})
ρ	Density (kg mm^{-3})
c_p	Specific heat (J kg^{-1} mm^{-1})

Table 4. Material Properties for Short–Time Thermal Response

	Thermal Property	
Material	Diffusivity, α [mm^2/s]	Effusivity, η [W√s/mm^2/°C]
Silicon	52.7	0.0138
Typical mold compound	0.31	0.00126
Copper	111	0.0360
Gold	128	0.0281
Air	24.9	6.0E–6

IGBT Gate Drive Considerations

ON Semiconductor®

http://onsemi.com

APPLICATION NOTE

Introduction

The IGBT transistor is a much more complex structure than either a MOSFET or a Bipolar Junction Transistor (BJT). It combines features of both of these devices and has three terminals – a gate, a collector and an emitter. In terms of the gate drive, the device behaves like a MOSFET. The current carrying path is very similar to the collector−emitter path of a BJT. Figure 1 shows the equivalent device circuit for an n type IGBT.

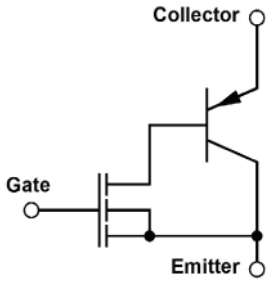


Figure 1. Equivalent Circuit

Basic Drive Understanding

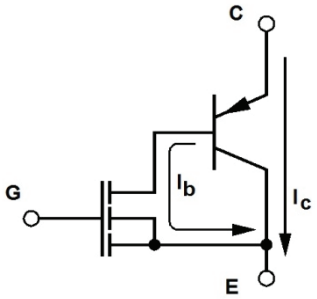


Figure 2. Turn on Current Flow

To turn a BJT on and off quickly, the gate current must be driven hard in each direction to move the carriers into and out of the base region. When the gate of the MOSFET is driven high, there is a low impedance path out of the base of the bipolar transistor to its emitter. This turns that transistor on quickly, so the faster the gate is driven high, the faster the collector current begins to flow. The base and collector current flows are shown in Figure 2.

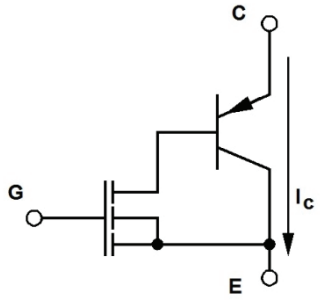


Figure 3. Turn off Current Flow

The turn off scenario is a bit different and is shown in Figure 3. When the gate of the MOSFET is pulled low, there is no current path for the base current in the BJT. The lack of base current begins the turn off process; however, for a fast turn off, current should be forced into the base terminal. There is no mechanism available to sweep the carriers out of the base, so the turn off of the BJT is relatively slow. This leads to a phenomenon called tail current since the stored charge in the base region must be swept away by the emitter current.

It should be obvious that faster gate drive dv/dt rates (due to higher gate current capability) will turn the IGBT on and off faster, but there are inherent limitations as to how fast the device can switch, especially for turn off. Due to these limitations switching frequencies are often in the 20 kHz to 50 kHz range, although in special cases they can be used in faster and slower circuits. IGBTs are generally used in high power (Po > 1 kW) circuits in both resonant and hard−switching topologies. Resonant topologies minimize the switching losses as they are either zero voltage switching or zero current switching.

Slower dv/dt rates offer improved EMI performance, when that is of concern and cause less spiking during the turn−on and turn−off transitions. This is at the expense of lower efficiency due to the slower turn on and turn off times.

Secondary Turn−on

There is a phenomenon that occurs with MOSFETs called secondary turn on. It is due to the very fast dv/dt rates on the drain voltage which can be in the range of 1000 – 10,000

V/us. Though IGBTs don't typically switch as fast as a MOSFET they can still experience very high dv/dt levels due to the high voltages used. This can lead to secondary turn on if the gate resistance it too high.

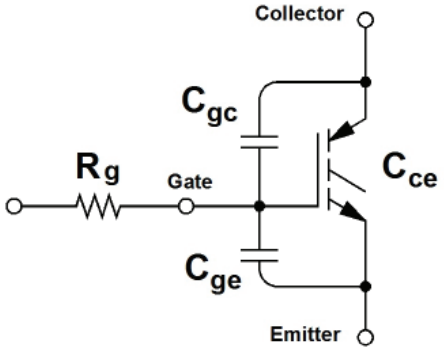


Figure 4. IGBT with Parasitic Capacitances

Under this condition, when the gate is being pulled low by the driver the device begins to turn off, but the rise in voltage on the collector generates a voltage on the gate due to the voltage divider of Cgc and Cge. If the gate resistor is too high, the gate voltage can climb high enough to turn the device back on. This causes a large pulse of power which can overheat, and in some cases destroy the device.

The limiting equation for this problem is:

$$\frac{dv}{dt} < \frac{V_{th}}{Rg \cdot Cgc} \quad \text{[Ref. 1]} \qquad \text{(eq. 1)}$$

Where:

- dv/dt is the rate of the rising voltage waveform on the collector at turn off
- V_{th} is the plateau level of the gate
- Rg is the total gate resistance
- Cgc is the gate−emitter capacitance

It should be noted that C_{iss} on the data sheet is the parallel equivalent of the Cge and Cgc capacitances.

Similarly, Rg is the series sum of the gate driver impedance, the physical gate resistor and the internal gate resistance. The internal gate resistance may be given on the data sheet. If not, it can be measured by using an LCR bridge and shorting the collector−emitter pins and then measuring the equivalent, series RC at a frequency close to the switching frequency.

The driver impedance may be found on its data sheet if it uses a FET output stage. If it is not on the data sheet, it can be approximated by taking the peak drive current at its rated V_{CC} level.

$$Rdriver \cong \frac{V_{CC}}{Ipk} \qquad \text{(eq. 2)}$$

So the maximum total gate resistance is:

$$Rg < \frac{V_{th}}{Cgc \cdot \left(\frac{dv}{dt}\right)} \qquad \text{(eq. 3)}$$

The maximum dv/dt is based on the gate drive current as well as the circuit impedances surrounding the IGBT and needs to be verified in the actual circuit if a high value resistor is used for the gate drive. Figure 5 shows turn off waveforms for three different IGBTs in the same motor control circuit. The dv/dt is 3500 V/ μs in this application.

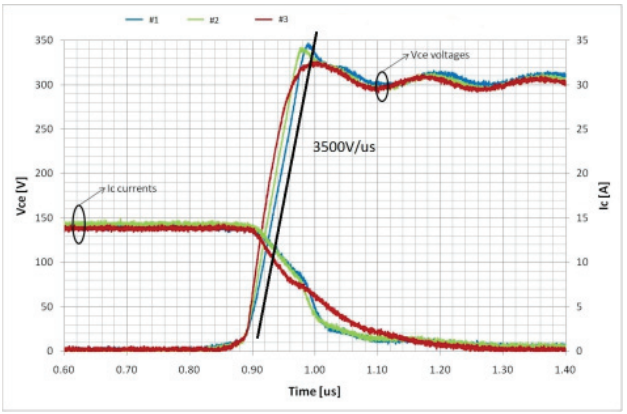


Figure 5. Turn Off Waveforms for 3 IGBTs

For this situation, IGBT #2 has a typical Cgc of 84 pF and a threshold gate voltage (at 15 A) of 7.5 V.

Using the above equation, the maximum total gate resistance for this circuit would be:

$$Rg < \frac{7.5\ V}{84\ pF \cdot \left(\frac{3500\ V}{\mu s}\right)} \qquad \text{(eq. 4)}$$

$Rg < 25.5\ \Omega$

So if the internal gate resistance is 2 Ω and the driver impedance is 5 Ω, the absolute largest gate resistor used should be 18 Ω. In practice, due to variations in IGBTs, drivers, board impedances and temperature, a maximum value of something lesser (e.g. 12 Ω) would be advisable.

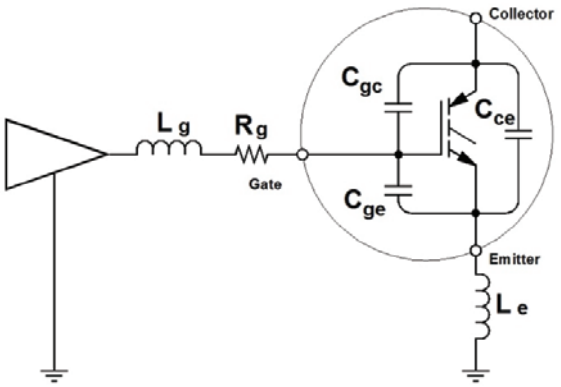


Figure 6. Equivalent Gate Drive Circuit

Gate Ringing

It may appear that eliminating an external gate resistor would give optimum high frequency performance, while assuring that secondary turn on would not occur. In some cases this may work, but it can also lead to oscillations due to the impedances in the gate drive circuit.

The gate drive circuit is a series RLC tank circuit. The capacitance comes mainly from the IGBT parasitic capacitances. The two inductances shown, come from a combination of the board trace inductance and the bond wire inductance of the IGBT and driver.

With little or no gate resistance, the tank circuit will oscillate and cause high losses in the IGBT. Enough gate resistance is required to dampen the tank circuit and therefore eliminate oscillations.

It is difficult to calculate the proper resistor as the inductance is difficult to measure. Good layout procedures are the best solution to minimizing the required, minimum gate resistance.

The path between the driver and IGBT gate should be as short as possible. This means the entire circuit path of the gate drive as well as the ground return path. If the controller does not include an integrated driver, it is much more important to locate the IGBT driver close to the gate of the IGBT than it is to locate the input of the gate driver to the PWM output of the controller. The current from the controller to the driver is very small and any stray capacitance will have much less of an effect than with the high currents and di/dt levels from the driver to the IGBT. Short, wide traces are the best way to minimize the inductance.

Typical minimum driver resistances range from 2 Ω to 5 Ω. This includes the driver impedance, external resistor value and internal IGBT gate resistance value.

Once the board is laid out, the gate resistor value can be determined and optimized.

Conclusion

Guidelines have been given for the minimum and maximum values of the gate resistor values. There is a range of values between these limits that will allow tuning of the circuit for maximum efficiency, minimum EMI or other important parameters. Designing a circuit with values safely between these extremes will offer a robust design.

References:

1. *Power Semiconductor Devices*, B. Jayant Baliga, PWS publishing Company, Boston. ISBN 0−534−94098−6

AND9054/D

IGBT Power Losses in Induction Heating Applications

ON Semiconductor®

http://onsemi.com

APPLICATION NOTE

ABSTRACT

IGBTs with blocking voltages of about 1200 V are widely used in single−ended induction heating applications. During turn−off, the high voltage impressed across the IGBT together with its residual current cause considerable turn−off losses. During the on−state of the IGBT, the power lost due to variation of its saturation voltage with load current and junction temperature adds to the total loss in power. These losses reduce the overall efficiency of these applications. Understanding the cause of these losses, and developing a reliable and relatively fast method to measure them is important especially during an IGBT design optimization process.

In this paper, the principle of induction heating together with the switching waveforms will be explained. Additionally, two methods to measure IGBT power losses will be presented.

Keywords

Induction heating, soft switching, ZVS, conduction loses, turn−off losses.

List of Symbols and Abbreviations

ZVS	Zero Voltage Switching
T_S	Switching Period
ZCS	Zero Current Switching
NPT	Non−Punch Through
$V_{CE(sat)}$	Collector Emitter Saturation Voltage
V_{CE}	Collector Emitter Voltage
D	Duty Cycle
T_S	Switching Period
E_{off}	Turn−Off Losses
E_{cond}	Conduction Losses
V_{dc}	Rectified Voltage
I_{CE}	Collector−Emitter Current
T_J	Junction Temperature

Introduction

In recent years, increases in the number of household electrical appliances have led to an increase in energy consumption per household. The cost of energy associated with using these appliances has also increased because most present day electrical appliances, such as the electric cooker, are not very efficient. As such, a considerable amount of the power obtained from the grid is not being used to do useful work. One way to maximize the amount of power obtained from the grid is to develop more energy efficient electrical appliances.

One such energy efficient device is the induction cooker. Unlike the standard household electric cooker, the induction cooker uses electromagnetic generated heat energy for cooking. This method of cooking makes the induction cooker about 25% more efficient than the electric cooker. Moreover the heat generated through induction does not heat the air around the cooker directly. As a result, there is less adverse effect on the air conditioning of the cooking environment. Additionally, induction cookers cook faster than traditional electric and gas cookers for the same input power level.

Principle of Induction Heating

Figure 1 shows a typical single−ended topology used for induction heating applications.

Magnetic energy is generated and transferred to the cooking vessel using the principle of electromagnetic induction and is transformed into thermal energy at the cooking vessel.

This principle involves rectifying the relatively low frequency ac line input voltage using an uncontrolled switching device such as a diode.

Switching the rectified voltage at a frequency between 20 kHz to 35 kHz produces a high frequency magnetic flux. The cooking vessel acts as a lossy magnetic core which converts the magnetic field into heat.

The main components used to generate and transfer this heat energy are the pan or cooking vessel, an inductor, a resonant capacitor, and the IGBT.

The geometry of the inductor winding is important in generating the magnetic field required to generate and transfer the heat.

The inductor windings are spiral in shape. The wires are wound around each other in a horizontal plane. This geometrical arrangement increases the surface area of the magnetic flux.

The concentration of these magnetic flux lines around the pan is further enhanced by using rectangular−shaped ferrite magnet bars, placed at equal intervals around the inductor windings.

The inductor windings are also multi−filar. The use of multiple small conductors minimizes skin effect and reduces the IR losses in the coil.

The inductor Lr is an air core inductor by design. The cooking vessel must be made of a magnetic material and acts as a core. At the switching frequency of the IH cooker, the thickness of the pan is much too great for an efficient core and the eddy current losses are substantial. These losses convert the magnetic field into thermal energy. This generates a great amount of heat in the pan and cooks the food.

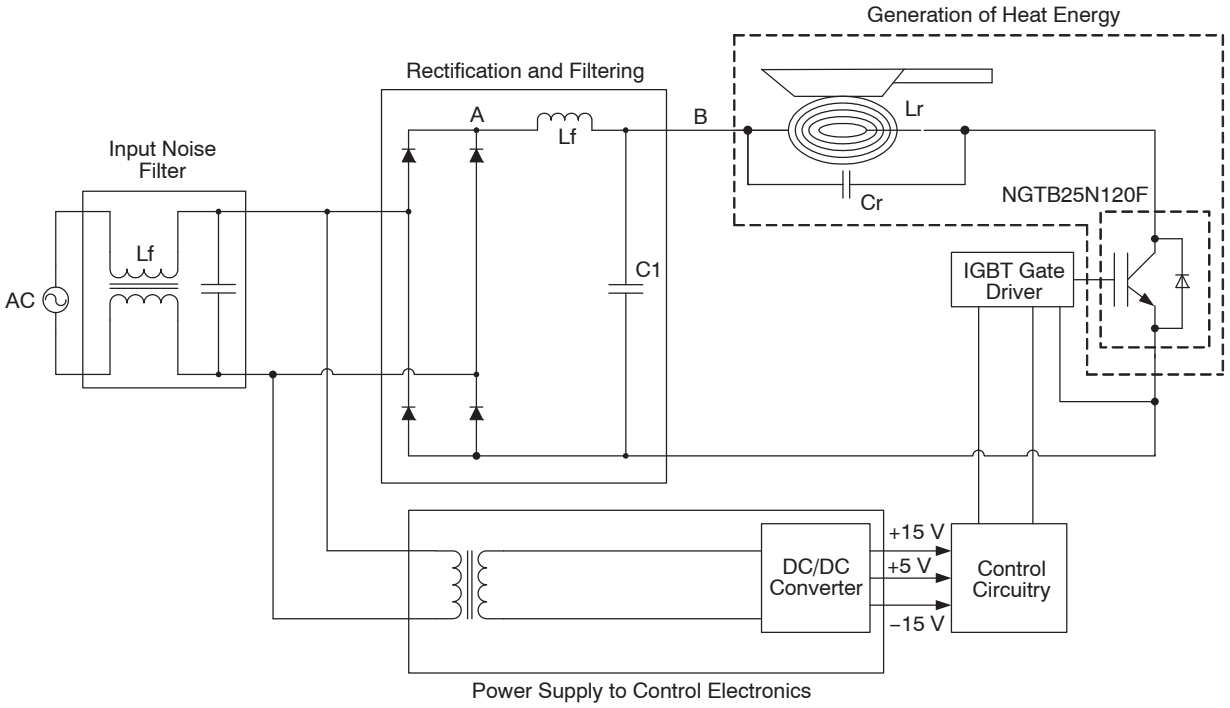


Figure 1. Block Diagram of a Single−Ended Induction Cooker

Switching Waveform Analysis

Figure 2 shows the typical switching waveforms for a single ended induction heating application. Sections of the waveform will be described based on the switching states of the co−packaged diode and the IGBT.

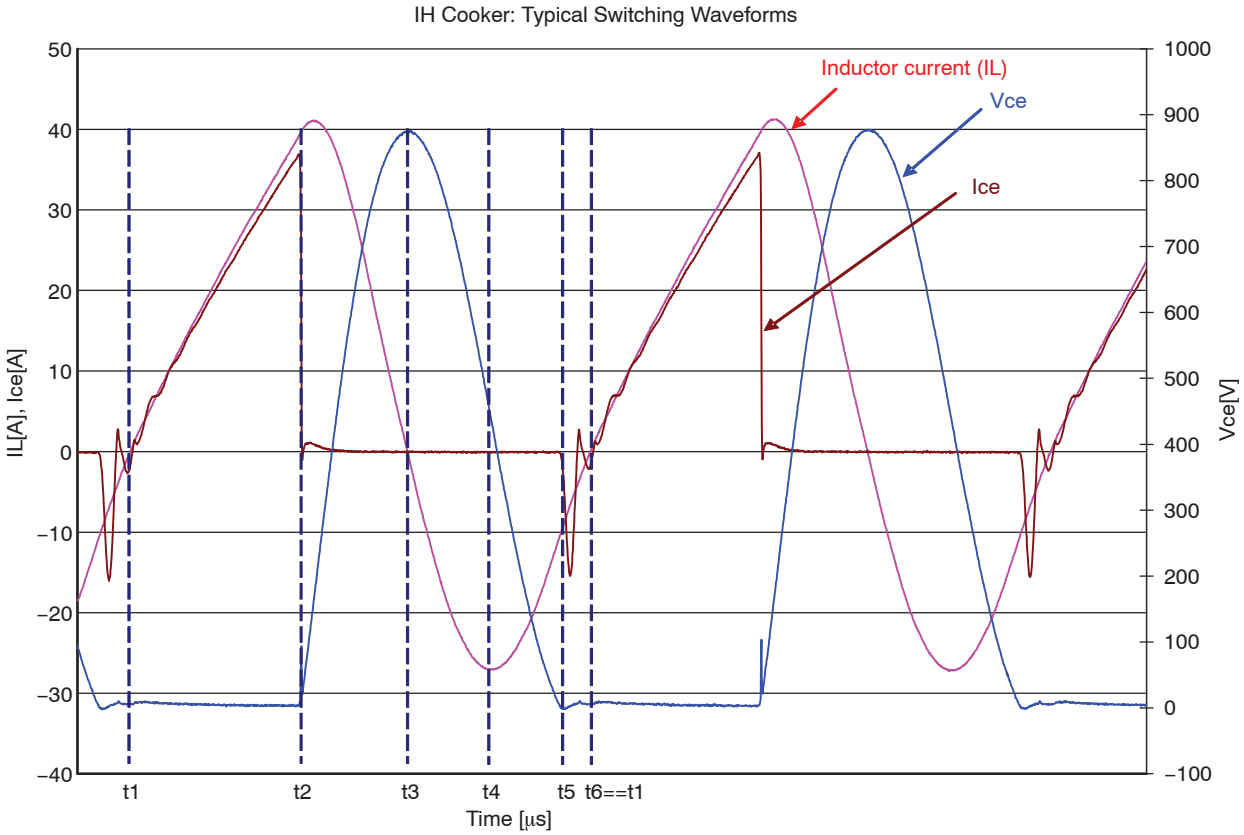


Figure 2. Typical Switching Waveforms for a Single Ended Induction Cooker

The initial switching state, t1, is defined as the time when the IGBT turns−on and the co−packaged diode is not conducting. Before t1, the co−packaged diode is conducting and V_{CE} is one diode drop negative.

Switching State 1: $t_1 < t \le t_2$: IGBT ON, Co−packed Diode OFF

At t1, the IGBT starts conducting and V_{CE} equals $V_{CE(sat)}$. The current flowing through the IGBT rises linearly as defined by Equation 1. The LC tank inductor stores energy at this point. At t2, the IGBT is switched−off.

$$V_L = V_{dc} - V_{ce} = L\frac{di_L}{dt} = L\frac{di_{CE}}{dt}$$

$$i_{CE} = \frac{1}{L}\int_0^{t_{ON}} (V_{dc} - V_{ce})dt \qquad \text{(eq. 1)}$$

$$= \frac{1}{L}(V_{dc} - V_{ce})\, t_{ON} + i_{CE}(t = 0)$$

Switching State 2: $t_2 < t \le t_5$ IGBT OFF, and Co−packed Diode OFF

When the IGBT is turned−off at t2, the energy stored in the LC tank inductor starts being transferred to the LC tank capacitor. V_{CE} rises with a slew rate of about 100 V/µs to 200 V/µs and reaches its maximum value at the instant where the inductor current crosses the zero axes. At t3, the LC tank capacitor starts transferring energy back to the LC tank inductor. The energy is being transferred to the capacitor as long as the inductor current slope is negative. Half of the cycle current is building in the inductor (from T4 to T2) and the other half (from T2 to T4) the current is reducing. At t4, V_{CE} goes lower than the dc bus voltage. The current through the LC tank reaches its maximum negative value at this point. At t5, the voltage across the inductor becomes greater than the dc bus voltage and V_{CE} becomes negative.

Switching State 3: $t_5 < t \le t_6$ IGBT OFF, and Co−packed Diode ON

When V_{CE} becomes negative, the co−packaged diode starts to conduct. To achieve ZVS, the IGBT gate signal is applied at t5. However, the IGBT does not conduct at this time. It starts conducting only when the co−packaged diode stops conducting.

This represents the full switching cycle which repeats itself beginning at switching state 1.

IGBT Power Losses in IH Applications

The total power lost in the IGBT in this application consists of turn−on, conduction, turn−off and diode losses. The contribution of the diode losses to the total power losses is negligible, and the turn−on losses can be significantly minimized, if the application employs ZVS techniques. However, ZVS is not achieved at all operating power levels of the IH cooker.

Since one end of the tank circuit is connected to the rectified input voltage, zero−state switching only occurs at power levels that resonate the tank circuit such that it reaches zero volts. Under some light load conditions, the tank circuit voltage will not reach zero volts on the collector of the IGBT and therefore zero−state switching is not achieved and the turn−on power losses will increase.

The most dominant contributors to the total power loss are conduction and turn−off losses.

Conduction Losses

The average power dissipated by the IGBT is expressed mathematically in Equation 2.

$$P_{Ave} = \frac{1}{T_S}\int_0^{T_S}\left[V_{CE}(t) \times I_{CE}(t)\right]dt \qquad \text{(eq. 2)}$$

For conduction losses, Equation 2 can be re−written as follows.

$$P_{Ave} = V_{CE(sat)}\left(t,\ I_{CE},\ T_j\right) \times I_{CE} \times D \qquad \text{(eq. 3)}$$

Equation 3 shows that conduction losses are dependent on the load current, $V_{CE(sat)}$, and the duty cycle. The value of $V_{CE(sat)}$ is not constant but varies over time. It is also dependent on the load current, and the IGBT's junction temperature.

In the IH cooker application, the control circuitry varies the duty cycle in direct proportion to the demand in cooking power. Consequently, conduction losses will therefore be highest at the highest cooking power level because all the parameters in Equation 3 will have their maximum values at the highest cooking power level.

Figure 3 shows the variation of $V_{CE(sat)}$ with I_{CE} at T_J = 67°C for a selected switching cycle. The data in Figure 3 was obtained from a commercially available IH cooker, and a clamped circuit was used to measure $V_{CE(sat)}$. This circuit clamps Vce at 10 V when the IGBT is switched−off which allows the oscilloscope to use a low volt/div setting so that V_{CE} can be accurately measured.

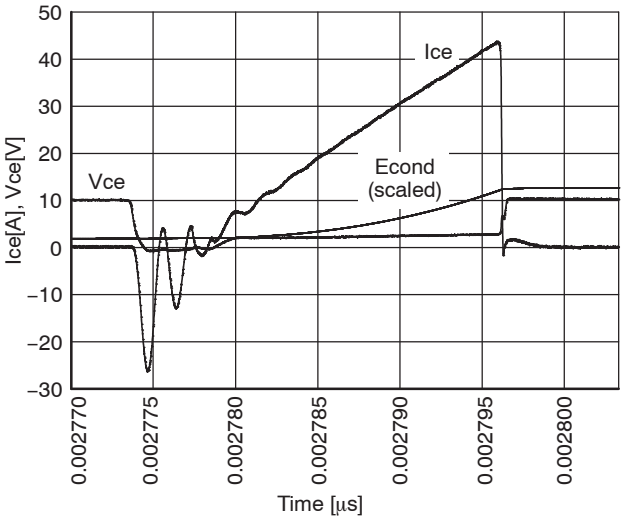


Figure 3. Variation of $V_{CE(sat)}$ with I_{CE} at T_J = 67°C

Turn−off Losses

Figure 4 shows the turn−off section of the IH cooker waveforms.

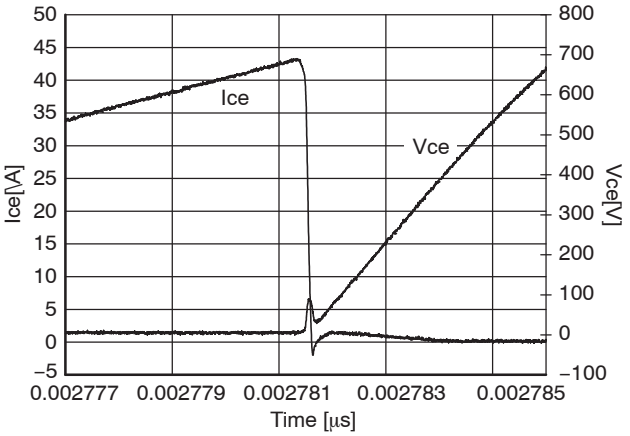


Figure 4. Measurement of Turn−off Losses using a Digital Storage Oscilloscope

These losses are influenced by the IGBT's residual current, the slew rate of V_{CE}, and the switching frequency.

The residual current results from minority carriers trapped in the drift region after the IGBT is switched−off. Factors which influence the rate of combination of these minority charge carriers include the doping concentration, the buffer layer thickness, and doping technology. The switching frequency is determined by the desired cooking power level, and the switching control algorithm of the application.

During the design, and development stages of an IGBT process, it is important to verify its performance in the target application during each development stage. This performance verification can be done by measuring the IGBT's losses in the application. If conduction and turn−off losses are to be measured separately, then a digital storage oscilloscope can be used. On the other hand, heat sink temperature measurements can be employed if measurement of the total power losses is desired.

Measuring Conduction and Turn−off Losses using a Digital Storage Oscilloscope

Accurate power loss measurement is obtained by setting−up the oscilloscope, and measurement probes correctly.

The following steps should be taken to properly set−up, and calibrate the measurement probes, as well as the oscilloscope measurement channels. This example uses channels 3 and 4 of the TDS5054B Tektronix oscilloscope as measurement inputs.

1. **Calibrating channels 3 and 4.** Disconnect all probes connected to the oscilloscope channels. Go to the main menu of the oscilloscope and select Utilities−>Calibrate channels. This process may take about 10 minutes.
2. **Setting−up the Current Probe.** Connect the current probe to channel 4 of the TDS5054B. Then connect the current probe ground terminal to the ground connector of the oscilloscope. Close the loop of the current probe and press the 'Degause' knob on the current probe. Use the scroll knob on the current probe and set the offset to as close to zero as possible. Use oscilloscope vertical waveform cursors to note the value of the offset.
3. **Setting−up the Voltage Differential Probe.** Connect the voltage probe to channel 3. Set the scale intended to be used to take measurements. Select the scale such that the voltage waveform will fill the entire oscilloscope. Use the offset knob on the voltage probe to set the offset to as close to zero as possible.
4. **Compute Energy Losses.** Equation 4 can be used to compute the energy lost during conduction and turn−off using the built−in oscilloscope math functions.

$$W(t) = \int_0^t \left[V_{CE}(t) \times I_{CE}(t)\right]dt \qquad \text{(eq. 4)}$$

Using the TDS5054B Tektronix oscilloscope, for example, this expression can be entered into the 'Math Equation Editor'. After capturing I_{CE} and V_{CE} waveforms as shown in Figures 3 and 4, the oscilloscope cursors can be placed at sections of the waveforms corresponding to the IGBT conduction, and turn−off time periods.

Next, the cursors are switched to the math function representing Equation 5. The math function then computes automatically the dissipated power within the selected waveform interval. Isolated differential probes such as the P5205 should be used to measure V_{CE}. The TCP202 current probe can be used to measure I_{CE}.

Measuring Total Losses using Heat Sink Temperature Measurements

In this procedure, a known amount of power is dissipated on the IH heat sink, using a switching device which can be operated in its linear region. The heat sink temperature rise is then measured, and its difference from ambient temperature, ΔT_{SA} is plotted against the known dissipated power. A suitable switching device to use is a Darlington Bipolar Transistor, and the test should be done with the transistor as the only device on the heat sink.

This method requires that the reference heat sink data use the same air flow as in the cooker to be able to correlate the heat sink temperature in the cooker to a known power level.

Next, the IGBT is mounted on the same heat sink in the IH cooker, and the cooker is operated under identical environmental conditions. Only the IGBT should be mounted on the heat sink. Using the measured heat sink temperature, and the gradient of the plot previously obtained, the value of the total dissipated power can be calculated.

Figure 5 shows the relation between ΔT_{SA} and the total dissipated power using the above method. ON Semiconductor's MJ11032 Darlington Transistor was used to provide this data, and the test was done in an IH cooker test environment.

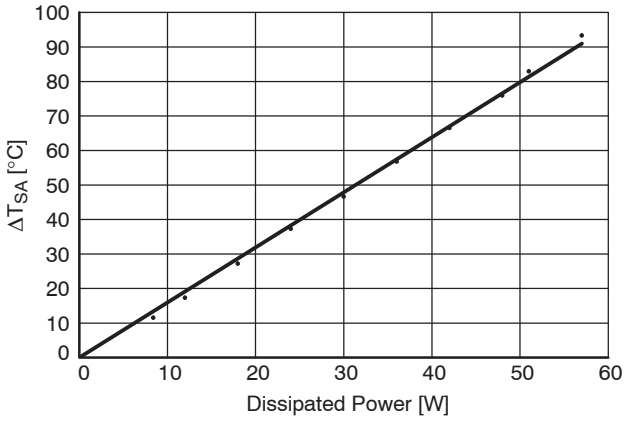


Figure 5. Variation of ΔT_{SA} with Total Dissipated Power

Summary and Conclusion

In soft−switching IH cooker applications, conduction losses and turn−off losses are the most important losses to be considered when designing an IGBT for use in these applications. Both losses contribute comparably to the overall power losses. Therefore, measuring these losses reliably provides important data which can be used to evaluate the IGBT's performance during the design process.

Two methods to measure these losses have been presented. Using a digital oscilloscope, conduction, and turn−off losses can be measured individually. Total IGBT losses on the other hand can be determined by measuring the temperature of a heat sink dedicated only to the IGBT in the application.

High Performance IGBT for Induction Heating Applications

ON Semiconductor®

http://onsemi.com

APPLICATION NOTE

Abstract

High saturation voltages associated with IGBTs using Non−Punch−Through (NPT) technology are the principal cause of high conduction losses in induction heating applications. These losses, combined with turn−off, and turn−on losses reduce the overall efficiency of IGBTs in these applications.

With Field Stop technology, lower saturation voltages can be achieved, and consequently conduction losses will be reduced. Furthermore, this technology minimizes losses that occur during IGBT turn−off by lowering the concentration of charge carriers left in the drift region. This optimization makes the IGBT ideal for use in induction heating applications.

ON Semiconductor's Field Stop IGBT, optimized for use in soft switching induction heating applications, is being presented in this paper.

Keywords

Induction heating, Field Stop, turn−off switching losses, conduction losses

List of Symbols and Abbreviation

ZVS	Zero Voltage Switching
ZCS	Zero Current Switching
NPT	Non−Punch Through
$V_{CE(sat)}$	Collector Emitter Saturation Voltage
V_{CE}	Collector Emitter Voltage
I_C	Collector Current
E_{off}	Turn−off losses
D	Duty Cycle
IH	Induction Heating
V_{dc}	Rectified dc voltage
PT	Punch Through

Introduction

IGBTs used in IH cooking applications dissipate a considerable amount of power especially when the application is operated at high power levels. In order to prevent the IGBT's junction temperature from rising above its recommended specification value, large heat sinks are usually employed.

Using large heat sinks in order to meet the maximum junction temperature requirement adds to the overall cost of production of these IH cookers. The market for induction heating applications will grow tremendously, if production costs can be reduced, and more efficient IH cookers can be developed.

Several parameters are to be considered when optimizing an IGBT. These include physical device properties such as the IGBT's thickness and die size as well as electrical properties such as saturation voltage, blocking voltage, and switching losses. Optimizing any one of these parameters comes at the cost of reducing the performance of another.

Therefore optimal trade−off values for these parameters are necessary in order to achieve the desired improvement in performance.

In this paper, the typical topologies used in IH applications will be presented. Next, the performance of ON Semiconductor's optimized IGBT in a commercially available IH cooker will be evaluated, and compared to that of several competitors.

Typical Topologies for IH Cooker Applications

Figures 1 and 2 show the power stages of the two most common topologies used in IH cooker applications.

In the half−bridge topology of Figure 1, the capacitor C2 functions as a snubber. It regulates the rate of rise of V_{CE} during turn−off, and therefore influences the energy lost during this time. A larger capacitor value causes V_{CE} to rise slowly and results in reduced turned−off energy. The current flowing through IGBT2 and the L_1C_1 resonant circuit is almost sinusoidal. During the positive half of the oscillation, IGBT1 conducts. During the negative half of the oscillation, IGBT2 conducts. The blocking voltage rating of the IGBTs is typically 600 V because they are connected directly to the rectified dc voltage, V_{dc}. Any induced voltage spike will be clamped by the IGBT co−packed diode.

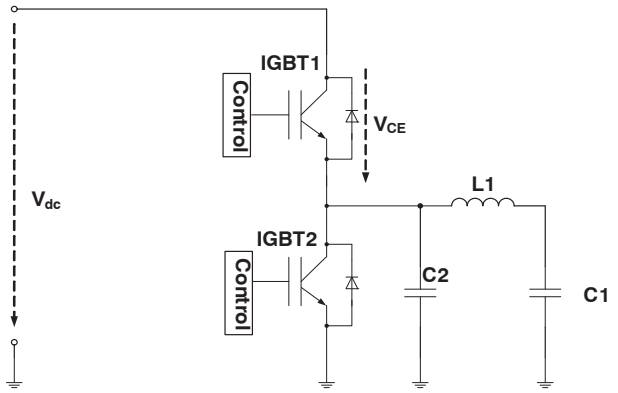


Figure 1. Half−Bridge Topology of IH Cooker Power Stage

Figure 2 shows the power stage of a single−ended topology. When the IGBT is turned−on, the current flowing through it and through the inductor rises linearly. When turned−off, the current falls with a fall time governed by the IGBT's parasitic collector−emitter output capacitance. This gives I_C an overall triangular shape. When the L_1C_1 resonant tank oscillates, the voltage across it has a sinusoidal shape. As a result, V_{CE} has a sinusoidal shape at turn−off. The amplitude of the resonant voltage for this topology can be as high as 1000 V. Therefore the blocking voltage for IGBTs used in this topology is typically 1200 V because the collector of the IGBT is connected to the resonant tank.

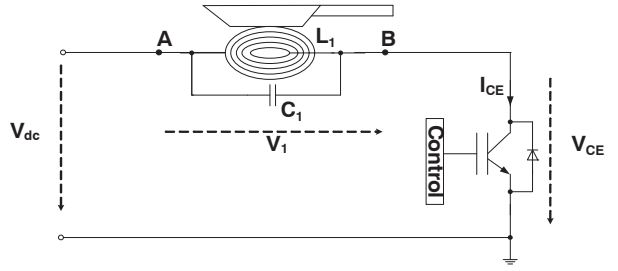


Figure 2. Single−Ended Topology of IH Cooker Power Stage

IGBT Performance Evaluation in Single−Ended IH Cooker Applications

The performance of IGBTs used in single−ended IH applications can be evaluated by measuring the amount of power dissipated during operation of the application. Since these applications employ a ZVS technique, conduction and turn−off losses are the most important losses to be considered.

Turn−on losses are negligible, especially at normal and full load conditions because the stored inductive energy in the resonant tank inductor is sufficient enough to swing the voltage at the collector of the IGBT, bringing it to zero or negative. When this happens, the co−packaged diode conducts and the voltage across the IGBT is one diode drop negative. Since power loss is the product of voltage and current, the power lost when the IGBT is turned−on at this time is very small.

In this section, the conduction and turn−off losses of ON Semiconductor's optimized IGBT, the NGTB25N120IHL, will be compared to those of several competitive devices.

The test system used for this performance evaluation is a commercially available IH cooker.

Tables 1 and 2 present a summary of the relevant test system parameters and IGBT switching parameters, respectively. Several representative switching cycles close to the peak of the input line current were used to take measurements. This procedure was repeated for each power level. In order to ensure identical switching cycles were measured for all devices being tested, information about the switching frequency of the IGBT, and the frequency of the input line voltage was used to trigger a digital storage oscilloscope at the specified switching cycle.

Table 1. SUMMARY OF TEST SYSTEM PARAMETERS

Inductor [LC tank] of IH cooker	112 μH
Capacitor [LC tank] of IH cooker	0.3 μF
Gate Resistor (NGTB25N120IHL)	15 Ω
Gate Resistor (Competitor device1)	24 Ω
Gate Resistor (Competitor device2)	18 Ω
Gate−Emitter Resistor	10 kΩ
Input Line Voltage/Frequency	220 V/60 Hz

Table 2. IGBT SWITCHING PARAMETERS AT DIFFERENT POWER LEVELS

Cooking Power Level (W)	Switching Frequency (kHz)	Duty Cycle (μs)
1000	23.6	17
1300	21.6	21
2000	19.8	24.7

Conduction losses

These losses were computed using the math function utility of a digital storage oscilloscope. For this device performance evaluation, a clamp circuit was used to obtain accurate $V_{CE(sat)}$ measurements. The losses were then measured at different load currents, where the duty cycle varies as a function of the cooking power level as shown in Table 2. Figure 3 compares the $V_{CE(sat)}$ dependence on temperature for different devices. The measurements for Figure 3 were taken at various ambient temperatures using pulsed dc current. Figure 4 shows conduction losses at different IH cooker power levels. These power levels

represent different values of the load current, and IGBT duty cycles.

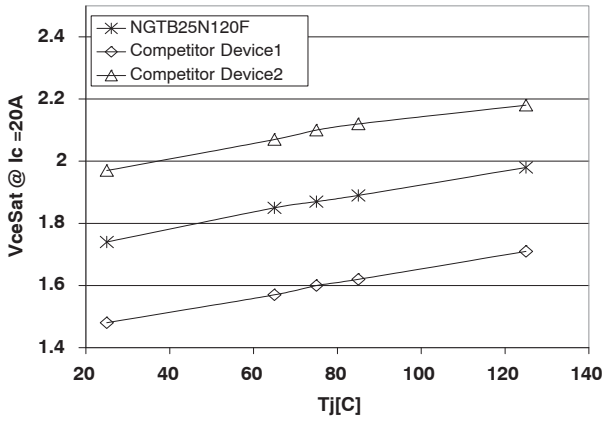


Figure 3. Variation of V_{CEsat} with IGBT Junction Temperature

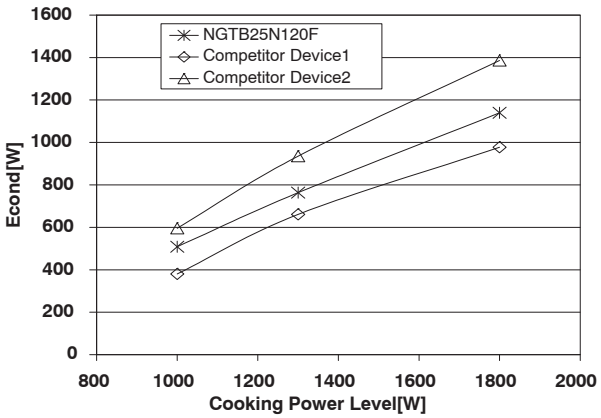


Figure 4. Conduction Losses at Different Cooking Power Levels

Switching Losses

Turn−off power loss is influenced by the rate of decay of the IGBT's residual current.

Figure 5 shows the variation of turn−off losses with IH cooker power for the NGTB25N120IHL and competitor devices 1 and 2. The turn−on time constant of the gate signal was maintained approximately the same for all of the devices by using different values for the gate resistor for each device. E_{off} is normalized using I_C in order to enable direct E_{off} comparison for all the devices.

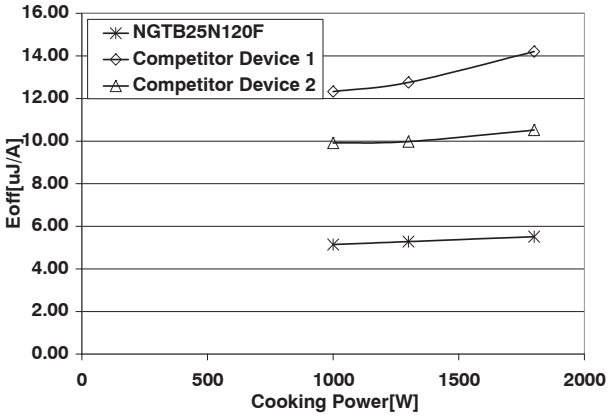


Figure 5. Variation of Turn-off losses with Cooking Power

Total Power Losses

The contribution of turn−on losses to the total power loss becomes significant at light loads. At lights loads, the LC tank inductor does not store enough inductive energy to swing the voltage at the collector of the IGBT before the IGBT is turned−on again. Consequently, the co−packaged diode never conducts and the voltage across the IGBT at turn−on is higher. However, in this evaluation, ZVS was possible at all the power levels used to test the devices. Therefore, the main contributors to the total power loss are conduction and turn−off losses.

Although these measurements were made using selected switching cycles, the validity of the measurements is reflected in the case temperature measurements. A device's temperature rise is proportional to its dissipated power. Therefore, the device with the highest total power loss also has the highest increase in case temperature as represented in Figure 6.

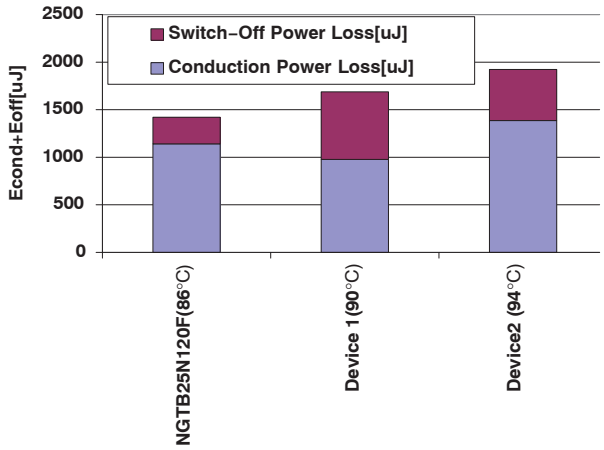


Figure 6. Variation of Total Losses (conduction and turn−off) with Case Temperature at 1800 W

Test Results Analysis

The following conclusions can be drawn from Figures 3 to 6:

1. We see from Figures 3, 4 and 6 that, the NGTB25N120IHL's $V_{CE(sat)}$ is slightly higher than that of device 1 but lower than that of device 2 of the competitor. As a result, conduction losses of the NGTB25N120IHL are slightly higher than that of device 1 but lower that that of device 2 from the competitor. Device 2 of the competitor uses NPT technology. Consequently, a higher $V_{CE(sat)}$value is expected. Device 1 of the competitor uses Shorted−Anode technology which enables lower $V_{CE(sat)}$ values but causes higher turn−off losses at higher temperatures. NGTB25N120IHL's $V_{CE(sat)}$ value represents the optimal trade−off value for minimizing both turn−off, and conduction losses.
2. Figure 5 shows that Device 1 of the competitor has the highest dependence of E_{off} on cooking power level while the NGTB25N120IHL's has the lowest turn−off losses at all power levels. Furthermore, the NGTB25N120IHL has the weakest turn−off power loss dependence on IH cooking power level. This makes it suitable for use in IH applications having a wide range of cooking power levels.
3. We see from Figure 6 that, the percentage of conduction and turn−off losses contributing to the total power loss is different for each device. However, the NGTB25N120IHL has the optimal percentage combination of both types of losses because it has the lowest case temperature of all three devices at the highest cooking power level.

Summary and Conclusions

In single−ended IH applications where ZVS techniques are employed, both conduction and turn−off losses contribute to the total IGBT power loss. The efficiency of IGBTs used in these applications can be improved by designing the IGBT to have optimal trade−off between $V_{CE(sat)}$ and switching speed.

The NGTB25N120IHL has been designed to have these optimal trade−off values, and its performance in a commercially available IH cooker has been verified, and proven to be better than that of competitor devices.

AND9059/D

Electrostatic Discharge and IGBTs

ON Semiconductor®

http://onsemi.com

APPLICATION NOTE

One of the major problems plaguing electronics components today is damage by electrostatic discharge (ESD). ESD can cause degradation or complete component failure. Shown in Table 1 are the susceptibility ranges of various technologies to ESD. As circuitry becomes more complex and dense, device geometries shrink, making ESD a major concern of the electronics industry.

Table 1. ESD SUSCEPTIBILITY OF VARIOUS TECHNOLOGIES

Device Type	Range of ESD Susceptibility (Volts HBM)
Power MOSFET	100–2,000
IGBT	4,000–8,000
Power Darlingtons	20,000–40,000
JFET	140–10,000
Zener Diodes	40,000
Schottky Diodes	300–2,500
Bipolar Transistors	380–7,000
CMOS	100–2,000
ECL (ICs)	500
TTL	300–2,500

GENERATION OF ESD

Electrostatic potential is a function of the relative position of non−conductors on the list of materials known as the Triboelectric Series, (see Figure 1 from DOD−HDBK−263). Additional factors in charge generation are the intimacy of contact, rate of separation and humidity, which makes the material surfaces partially conductive. Whenever two non−conductive materials are flowing or moving with respect to each other, an electrostatic potential is generated.

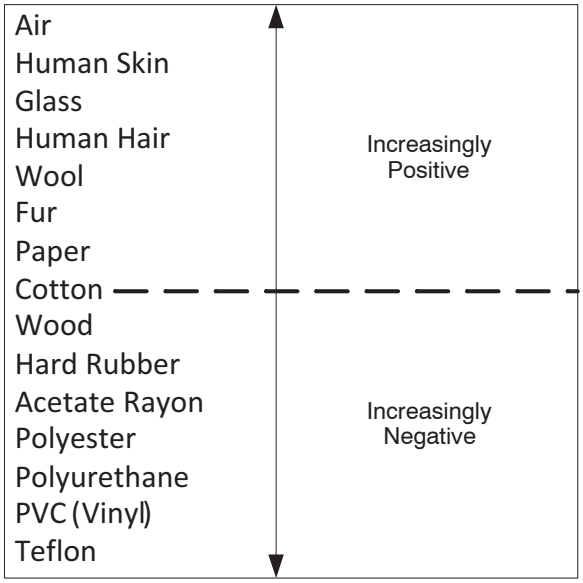


Figure 1. Triboelectric Series − A More Complete Table Appears in DOD−HDBK−263

From Figure 1, it can be seen that cotton is relatively neutral. The materials that tend to reject moisture are the most significant contributors to ESD. Table 2, also excerpted from DOD−HDBK−263, gives examples of the potentials that can be generated under various conditions. From these three Tables, it is apparent that sensitive electronic components can be damaged or destroyed if precautions are not taken, and that the necessary voltages can be easily generated.

Table 2. TYPICAL ELECTROSTATIC VOLTAGES

Means of Static Generation	Electrostatic Voltages	
	10 to 20% Relative Humidity	65 to 90% Relative Humidity
Walking across carpet	35,000	1,500
Walking over vinyl floor	12,000	250
Worker at bench	6,000	100
Vinyl envelopes for work instructions	7,000	600
Common poly bag picked up from bench	20,000	1,200
Work chair padded with polyurethane foam	18,000	1,500

ESD and IGBTs

Being MOS devices, insulated gate bipolar transistors can be damaged by ESD due to improper handling or installation. However, IGBT devices are not as susceptible as CMOS. Due to their large input capacitances, they are able to absorb more charge before being reaching the gate−breakdown voltage. Nevertheless, once breakdown begins, there is enough energy stored in the gate−source capacitance to cause complete perforation of the gate oxide. With a gate−to−emitter rating of VGE = ± 20 V maximum and electrostatic voltages typically being 100 – 25,000 V, it becomes very clear that these devices require special handling procedures. Figure 2 shows curve tracer drawings of a good device, and the same device degraded by ESD.

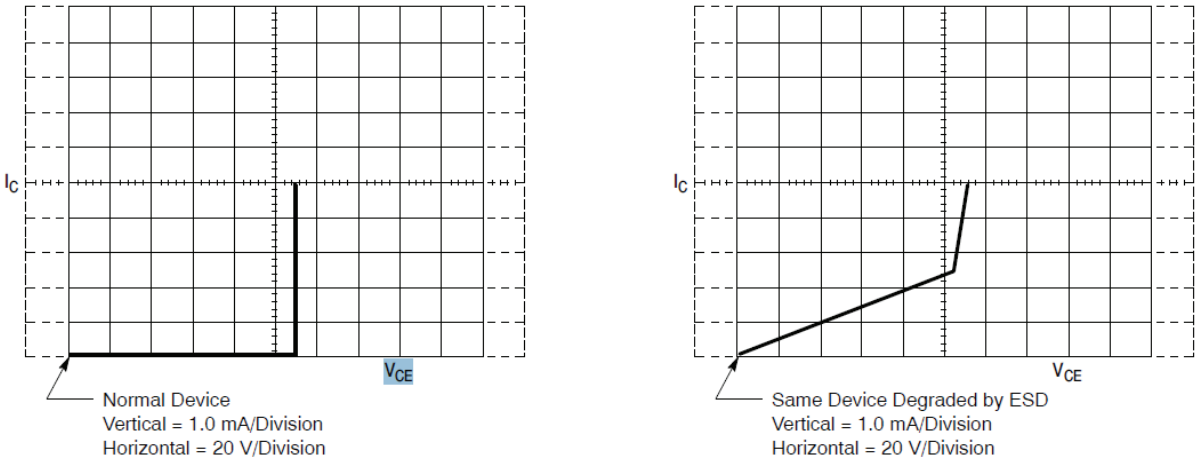


Figure 2. Curve Tracer Drawing of V_{CE} versus I_C of a Good Device (Left), and a Device in which the Gate was Damaged with ESD (Right). The Gate in this Device is Likely a Resistive Short.

Static Protection

The basic method for protecting electronic components combines the prevention of static build up with the removal of existing charges. The mechanism of charge removal from charged objects differs between insulators and conductors. Since charge cannot flow through an insulator, it cannot be removed by contact with a conductor. If the item to be discharged is an insulator (plastic box, person's clothing, etc.), ionized air is required.

If the object to be discharged is a conductor (metal tray, conductive bag, person's body, etc.), complete discharge can be accomplished by grounding it. Grounding of ESD sensitive product, such as IGBTs, must be done carefully, however. Grounding a charged IGBT with a low impedance ground will create an ESD event and could easily damage the device. Grounding of ESD sensitive devices needs to be with a high resistance path to ground, usually in the range of 1 MΩ.

A complete static−safe work station should include a grounded dissipative table top, floor mats, grounded operators (wrist straps), conductive containers, and an ionized air blower to remove static from non−conductors. Dissipative materials are those with properties between conductive and insulating and provide paths to ground in the range of a MΩ to several hundred MΩ. All soldering irons should be grounded. All non−conductive items such as Styrofoam coffee cups, cellophane wrappers, paper, plastic handbags, etc. should be removed from the work area. A periodic survey of the work area with a static meter is good practice and any problems detected should be corrected immediately. Above all, education of all personnel in the proper handling of static−sensitive devices is key to preventing ESD failures.

By following the above procedures, and using the proper equipment, ESD sensitive devices can be handled without being damaged. The key items to remember are:

1. Handle all static sensitive components at a static safeguarded work area.
2. Transport all static sensitive components in static shielding containers or packages.
3. Education of all personnel in proper handling of static sensitive components.

The maintenance of an ESD control program has been documented in the Electrostatic Discharge Association's (ESDA) ANSI/ESD S20.20−2007 and JEDEC's JESD625B.

Test Method:

ESD testing of semiconductors was first defined in military specifications such as MIL−STD−883B Method 3015.1, but today most testing of commercial product is performed using standards maintained by JEDEC or the Electrostatic Discharge Association (ESDA). These two organizations are now maintaining and developing ESD test methods jointly. The earliest and still most widely used ESD test method for semiconductors is the "human−body model" (HBM), ANSI/ESDA/JEDEC JS−001−2011. HBM consists of a network approximating the charge storage capability (100 pF) and the series resistance (1.5 kΩ) of a typical individual (Figure 3). For discrete and very low pin count devices such as IGBTs, the stress is applied between all possible pin combinations. Test results for IGBTs show that gate−oxide breakdown is most likely, and that reverse−biased junctions are about an order of magnitude more sensitive than forward−biased ones. The damage mechanism, which can be identified through failure analysis of shorted or degraded samples, is usually oxide puncture or junction meltthrough.

More recently the "charged device model" (CDM) has been found to represent the type of stress which semiconductor components experience in automated factories. CDM emulates a semiconductor device becoming charged and then discharging to a grounded, low resistance, surface. The test is usually performed using a field induction method to bring the device being tested to a high potential. A simplified schematic of the tester is shown in . The device being tested is placed in the leads up position on top of a thin insulator which sits on a metal field plate whose potential can be adjusted. The device under test is brought to a high potential by changing the potential of the field plate. The device is then rapidly grounded by touching a device pin with a grounded pogo pin. This produces a very fast, ~ 1 ns, pulse whose peak current can be from 1 to 10 of amps, depending on the size of the unit being tested. Failures from CDM are typically oxide failures rather than junction damage. The CDM test is documented in two standards, JEDEC's JESD22−C101E and ESDA's ANSI/ESD S5.3.1−2009. These two standards are very similar, but differences in the test setup and calibration modules mean that test values from the two methods while similar in magnitude are not equivalent. Efforts are underway to merge these standards.

A third ESD test method which has been used widely is the "Machine Model", (MM). The MM has a circuit diagram similar to HBM, but the capacitance is 200 pF and the 1.5 kΩ resistor is replaced by an implied inductance of about 0.8 μH. (The inductance is implied because that is what is needed to obtain the required waveform frequency with 200 pF of capacitance.) In recent years the MM has fallen out of favor because it gives very little more information than the HBM test and is not longer recommended for qualification. MM failure levels are typically about an order of magnitude lower than HBM failure levels but can range from a factor of 3 to a factor of 30.

Significance of Sensitivity Data

The Industry Council on ESD Target Levels (a collection of ESD experts from a wide ranging group of electronics companies) has done extensive studies of what ESD levels are needed to manufacture electronic systems with high yield. (White Paper 1: A Case for Lowering Component Level HBM/MM ESD Specifications and Requirements and White Paper 2: A Case for Lowering Component Level CDM ESD Specifications and Requirements) They found that if integrated circuits had HBM levels of 500 V or higher and CDM levels of 250 V and higher they can be handled in manufacturing lines with basic ESD control without yield loss. Without basic ESD control in the factory even HBM levels of 8000 V can have yield loss. Today, ON Semiconductor's IGBTs have HBM levels of 8000 V or more and CDM levels of 2000 V. With these ESD levels ON Semiconductor's IGBTs can be handled very safely in any manufacturing facility with the basic ESD controls in place required for manufacturing products containing modern integrated circuit products.

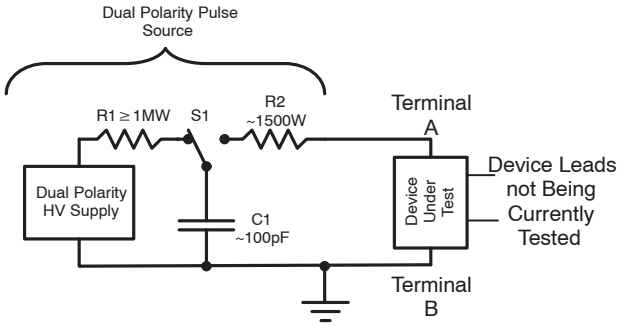


Figure 3. Simplified HBM test schematic

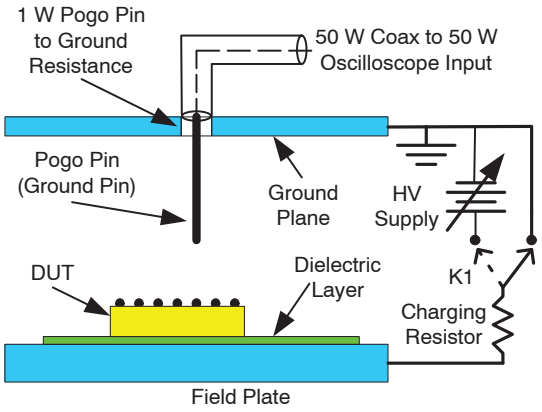


Figure 4. Simplified CDM tester schematic

AND9071/D

Noise Management in Motor Drives with IGBTs

ON Semiconductor®

http://onsemi.com

APPLICATION NOTE

INTRODUCTION

The introduction of IGBTs has enabled motor drives to move to higher switching frequencies and hence, more compact implementations. However, a side product of these advances is the increased vulnerability to the EMI/noise issues. The EMI/EMC management has always been a challenging aspect of any motor drive design and development, with a significant amount of time spent in finding empirical solutions to problems that manifest themselves during the development. The introduction of IGBTs switching at higher frequencies may complicate this task due to (a) higher switching frequencies and (b) increased proximity between components/subsystems offered by the higher level of compactness.

However, as many practicing engineers recognize, techniques exist to address EMI issues effectively. A number of these techniques are discussed in this section. The focus is on building preventive solutions into the design and layout rather than dealing with them through debugging and redesign.

A typical block diagram of the motor drive circuit is shown in Figure 1 and will be used to illustrate various techniques for EMI management.

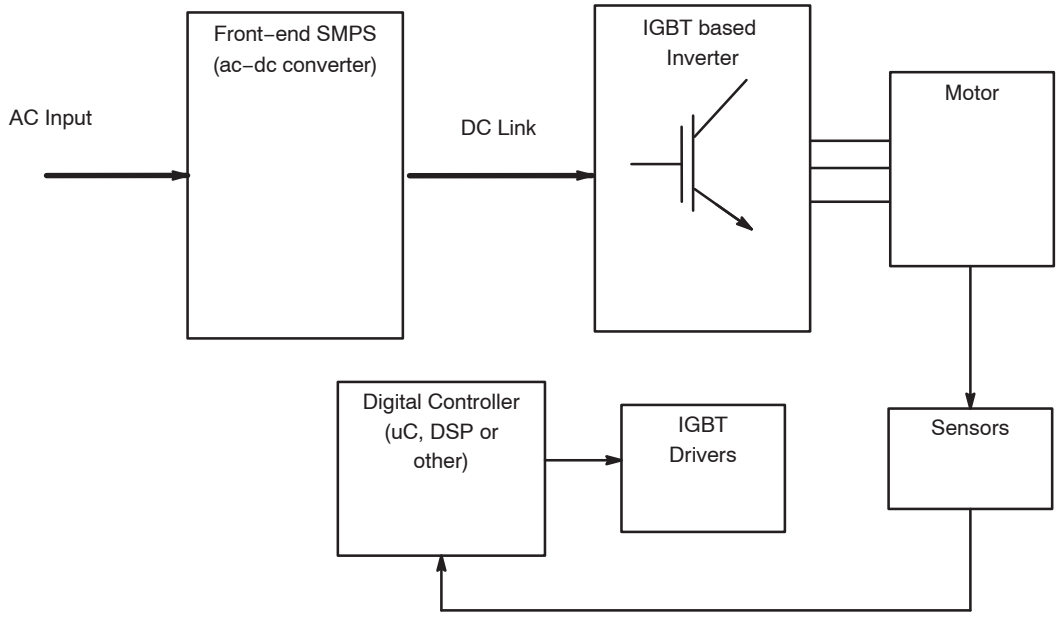


Figure 1. Typical Motor Drive System Block Diagram

As shown in this block diagram, the motor drive system contains blocks that process high power as well as those that process low level signals. Sensitivity of these blocks to EMI does differ and when they are colocated, it is important to minimize the interactions between the two types. In general, the following areas of EMI management can be identified:

1. Ensuring that digital controller circuits do not get perturbed by the high power switching noise in the inverter or the ac−dc converter.
2. Ensuring that the integrity of the sensed signals from the motor is not compromised due to the EMI generated in other blocks.
3. Ensuring that the power blocks do not get into false switching states due to presence of parasitic switching noise.
4. Ensuring that the driver (which is an interface block) acts as a proper buffer between low power circuits and high power circuits.
5. Ensuring that the motor drive circuit as a whole does not emit or conduct interference signals beyond the limits specified in the EMI/EMC compliance requirements.

The first four items above are necessary for acceptable and robust functioning of the motor drive systems, while the last one is clearly for meeting the external agency compliance requirements.

The techniques for noise management include better circuit design, relevant component choices and proper layout. Each of these will be addressed in the following sections.

CIRCUIT TECHNIQUES

Circuit design can have a profound influence on both the amount of noise produced and the susceptibility of motor drive circuits to the noisy environments in which they operate.

Front−End SMPS

When looking at the front−end ac−dc SMPS, it is important to minimize its contribution to the ambient noise. This can be done by using soft−switching techniques or at the very least, using effective snubbers to minimize the radiated EMI. Any high di/dt or dv/dt in the SMPS can lead to malfunction of more sensitive adjacent blocks such as controller, driver or sensor.

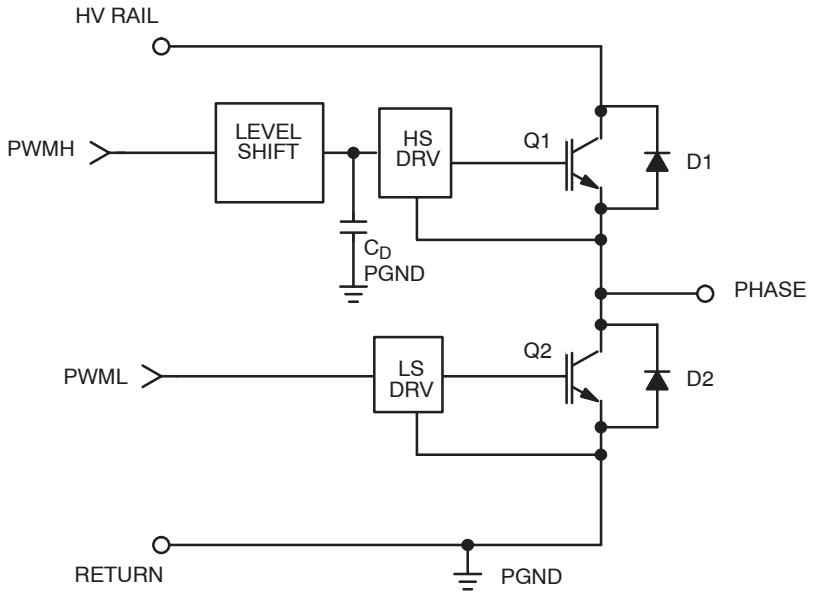


Figure 2. Typical IGBT−Based Inverter Stage (one phase)

Inverter and Driver

When considering the inverter and driver blocks, IGBTs are typically used in the high voltage applications only. In these applications, the inverter consists of three identical dual n−channel stages (each phase consisting of one high−side IGBT and one low−side IGBT, anti−parallel diodes for each and associated drive circuitry as shown in Fig. 2). Here, the challenge is to provide the level shifting for the high−side IGBT drive. The options for the drive circuit include:

- Low side discrete/integrated drivers followed by gate drive transformers – this option allows the gate drive transformer to provide the required level shifting for high−side IGBT, while the low side is coupled directly to the driver. Needs adequate delay matching and timing management to control the turn−on and turn−off of the two IGBTs. Noise susceptibility is limited with correct gate drive transformer design which would minimize the capacitance between the windings and may also require a shield.
- Use of integrated high−side, low−side driver with internal timing management – this option integrates the drive circuit into a single IC. Although the early

versions of such drivers gained some notoriety due to their noise susceptibility and also noise radiation, subsequent versions have incorporated better layout and control to overcome the noise issues.

- Use of optocouplers to transmit the gate drive signal from controller and use of local drivers for each IGBT – this option provides the best decoupling between the power and control stages. However, it comes with the price of requiring special bias and drive circuits and the need for high speed optocouplers.

While the choice amongst these options depends on the designer familiarity, BOM budget constraints and power levels, it is safe to say that good design and layout practices will allow successful implementation of any of these options in the motor drive applications.

Sensor

The sensor block can also benefit from common design practices such as filtering and buffering to improve noise immunity. In general, the sensor signals have time constants an order of magnitude below the one seen in the PWM inverter, so it is easy to eliminate the noise coupling through good low pass filtering techniques followed by use of Schmitt trigger gates to provide additional noise immunity. This is illustrated in Figure 3. The Hall sensors are in a typically noisy environment since they are close to motor windings with their PWM noise. In order to isolate the sensed position signal from the noise picked up by the Hall sensors before it reaches the controller, filtering and buffering are used. The filter consisting of R2, R3 and C2, has a time constant of 100 ns – much lower than the Hall sensor response times which are typically in micro−seconds. This is good enough to filter out the high−frequency noise spikes, while not disturbing the sensed signal. The filtered signal is next fed to a gate with a Schmitt trigger that has a relatively slow response time and built−in hysteresis. The digitization effect offered by the Schmitt trigger allows the sensed signal to be fed into the controller in a relatively noise−free manner.

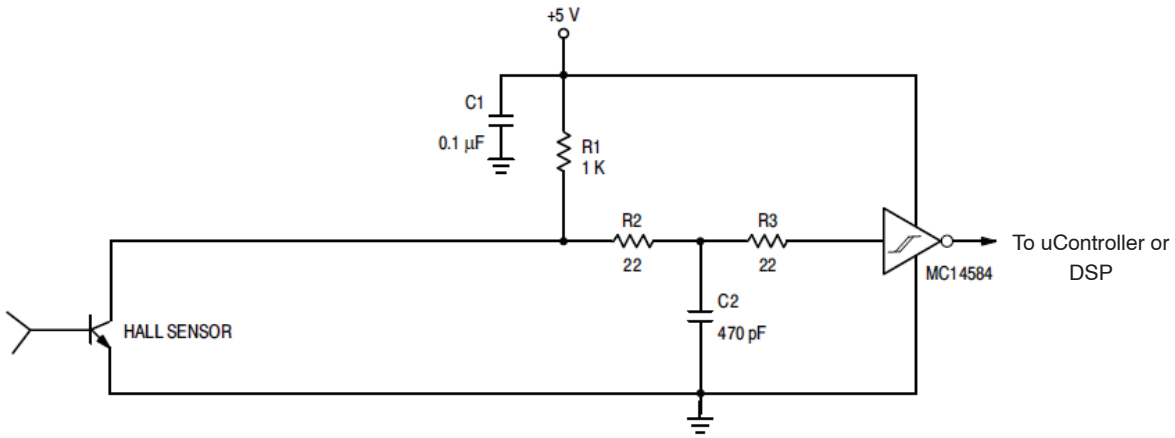


Figure 3. Processing of Hall Sensor Input for Low Noise Immunity

Controller

Finally, the controller, which may be the most noise sensitive block, also requires special attention. It needs good bypassing of critical inputs and buffering of the outputs. The immunity of the controller to noise can be further improved by firmware control where the code provides additional protection against random sequencing of drive signals caused by noise.

COMPONENT CHOICES

In addition to good circuit design techniques, appropriate component choices play an important role in EMI mitigation. Major component groups in a motor drive system are addressed in this section.

Power Semiconductors

Switching elements such as IGBTs and MOSFETs act as source of EMI and also exhibit susceptibility to it under certain conditions. To address susceptibility, it is important to choose devices which have reasonably high threshold voltages to prevent noise spikes from turning them on spuriously. Good, low impedance drive circuits are also important to prevent faulty turn−on of the switching devices.

While the noise emissions from switching devices are strictly a function of the switching speeds and can be controlled by shaping the switching intervals through the drive circuit, the built−in body diode of the MOSFET is often a major contributor to EMI. In motor drive applications, since the bidirectional switch currents are a norm, this phenomenon is more of a factor. Here, IGBTs provide a better alternative, since they do not have a body diode and an external anti−parallel diode can be independently chosen to meet the system requirements. Typically, a soft recovery diode (with low Qrr) is chosen. The reverse recovery of the high voltage diodes is also a

concern in the high side drive circuits, where a high voltage diode is used to bootstrap the bias voltage for the floating high−side bias voltage. In many integrated drivers, this diode is built−in with very snappy turn−off behavior and this leads to radiated EMI issues.

Controllers and Drivers (ICs)

Since the controllers provide the required drive signals for the switching devices, it is important to build noise immunity in these devices. Some controllers (both digital and analog) are inherently more noise immune than others. However, this fact is rarely acknowledged in public domain and it requires careful perusal of datasheets and/or application notes of a particular controller and/or prior experience designing with it to understand its vulnerability to EMI. . In controllers and drivers designed for better noise immunity, following features are generally observed:

- Physical separation of power related pins and low level signals which may be noise sensitive
- Sufficient hysteresis on any comparator inputs
- Separate analog and power ground pins
- Signal thresholds which are not too low so that noise can trigger them
- Specified noise immunity on critical pins (50 V/ns or better)
- Ability to withstand negative transient for short duration without latching up

The vulnerability can be minimized by good layout and bypassing practices and the product documentation usually provide guidance to the user regarding these practices. Drivers are meant to be buffers between the vulnerable control circuits and high−noise switching devices. However, in the absence of good design and layout practices, they can inadvertently become the conduits for the noise coupling. Similar to the controllers, the IC drivers vary widely with regards to their EMI generation and susceptibility behavior.

Magnetics

The choices for power magnetics, EMI filter and noise filtering can have a significant impact on the system EMI performance. Often, the characterization of the magnetic components is not as extensive as other components, leading to uninformed and faulty choices of magnetics in circuits. With respect to EMI, it is important to recognize that all inductors have a self−resonant frequency (SRF) beyond which they cease to be inductors. Since the function of the inductor is to typically provide a high impedance at a specified frequency, the SRF is a very critical parameter. The DC resistance of inductors provides damping which is good from an EMI point of view, but adds to the power losses. Also, depending on which section of the circuit the inductor is used in, the type of inductor may differ in terms of core material (Ferrite, powder−iron etc.) and geometries (rod, toroid, EE cores etc.). The transformers used in the front−end SMPS also play a big role in EMI generation if not properly designed. EMI shields are often used to reduce the radiated EMI and y−caps from primary to secondary are employed to provide low impedance path for common mode noise. Ferrite beads provide the quickest “band−aid” solution to the EMI problems during debugging stages, as they can be inserted easily (generally without altering the layout) and are effective in slowing down the fast edges that cause EMI. However, overreliance on these beads is not a good practice because (a) they invariably add to the losses and (b) they are not mechanically robust.

Other Passives

Capacitors are extensively used for bypassing, while a judicious combination of resistors and capacitors is used for passive snubber implementations. The choice of capacitors is important because the ESR values can impact the high frequency performance of capacitors adversely. In general, an SMT ceramic cap is the best choice for bypass filtering. Ceramic capacitors have very low ESR and an SMT package eliminates the lead inductance. The capacitor should be connected as closely as possible to the leads of the chip that it is bypassing. Typically a bypass capacitor is in the capacitance range of 0.01 uF to 1 uF.

LAYOUT

In a motor drive, layout is a critical part of the total design. Often, getting a system to work properly is actually more a matter of layout than circuit design. The following discussion covers some general layout principals, power stage layouts, and controller layouts. It is realized by practicing engineers that there is no single “correct” layout for an application. A good layout generally involves making a number of on the spot technical trade−offs that cumulatively lead to better system performance.

General Principles:

There are several general layout principles that are important to motor drive design. They can be described as five rules:

Rule 1: Minimize Loop Areas. A loop is the circuit trace path from the source of a signal (e.g. driver, FB node) to its destination (e.g. IGBT, error amp) and back to its source through the return path. This is a general principle that applies to both power stages and noise sensitive inputs. Loops are antennas. At noise sensitive inputs, the area enclosed by an incoming signal path and its return is proportional to the amount of noise picked up by the input. At power stage outputs, the amount of noise that is radiated is also proportional to loop area. A corollary of this rule is that the placement of key components that are connected is very critical and they should be placed as close to each other as possible.

Rule 2: Cancel fields by running equal currents that flow in opposite directions as close as possible to each other. If two equal currents flow in opposite directions, the resulting electromagnetic fields will cancel as the two currents are brought infinitely close together. In printed circuit board layout, this situation can be approximated by running signals

and their returns along the same path but on different layers. Field cancellation is not perfect due to the finite physical separation, but is sufficient to warrant serious attention in motor drive layouts. Looked at from a different perspective, this is another way of looking at Rule 1, i.e. minimize loop areas.

Rule 3: On traces that carry high speed signals avoid 90 degree angles, including “T” connections. If you think of high speed signals in terms of wavefronts moving down a trace, the reason for avoiding 90 degree angles is straightforward. To a high speed wavefront, a 90 degree angle is a discontinuity that produces unwanted reflections. From a practical point of view, 90 degree turns on a single trace are easy to avoid by using two 45 degree angles or a curve. Where two traces come together to form a “T” connection, adding some copper pour to cut across the right angles accomplishes the same thing.

Rule 4: Connect signal circuit grounds to power grounds at only one point. The reason for this constraint is that transient voltage drops along power grounds can be substantial, due to high values of di/dt flowing through finite inductance. If signal processing circuit returns are connected to power ground a multiple points, then these transients will show up as return voltage differences at different points in the signal processing circuitry. Since signal processing circuitry seldom has the noise immunity to handle power ground transients, it is generally necessary to tie the signal ground to the power ground at only one point. This rule can also be extended to use of ground planes. For power circuits, it is important to have either separate ground planes or ensure that the high current path on the ground plane does not traverse through sensitive signal ground areas on the same plane.

Rule 5: Use Vias very sparingly and selectively. Although vias offer an easy routing solution for complex/dense pcbs, injudicious use of them could lead to EMI and other problems. Vias are used primarily for 3 purposes:

1. To provide signal connection to/from an inner layer plane such as ground plane as well as to provide signal connection between components places on separate layers.
2. To provide alternative routing path for a trace when routing is not possible on the same layer due to presence of other higher priority traces.
3. To provide thermal relief for high current carrying paths/planes on the inner layers.

However, insertion of vias reduces the area of a plane or copper pour, adds capacitance between the vias and the adjoining signals on all layers and causes diversion in traces which could have been more directly routed. Thus, addition of vias involves trade−offs that can be made by experienced layout designers and circuit designers together during the layout.

Layout Consideration for Power Stage:

There are two overriding objectives with regard to power stage layout. First, it is necessary to control noise at the gate drives so power devices are not turned on when they are supposed to be off or vice versa. Second, it is highly desirable to minimize radiated noise with layout, where tight loops and field cancellation can reduce the cost of filters and enclosures. Looking first at the gate drive, noise management is greatly facilitated by using the source or emitter connection for each power device as a miniature ground plane for that device’s gate drive. This is particularly important for high side N − Channel gate drives, where the gate drivers have high dv/dt displacements with respect to power ground. If the power device’s source or emitter connection is used like a ground plane, parasitic capacitive coupling back to power ground is minimized, thereby increasing the dv/dt immunity of the gate drive.

To illustrate this point, let’s refer to and assume that the high−side phase output swings 300 V in 100 nsec as a result of a switching transition, and that the parasitic capacitance to power ground, Cp, is only 1 pF. Then a simple i = C(dv/dt) calculation suggests that 3 mA of charging current will flow through Cp. This 3 mA into 5.6 kΩ of node impedance is much more than enough to cause false transitions. These numbers illustrate a very high sensitivity to parasitic coupling, which makes layout of this part of the circuit very important.

In addition to viewing source or emitter connections as miniature ground planes, it is also important to keep any signals referenced to ground away from high side gate driver inputs.

Gate drive noise immunity is also facilitated by minimizing the loop area that contains the gate drive decoupling capacitor, gate driver, gate, and source or emitter of the power device. One way to do this is to route the gate drive signal either directly above or beneath its return. If the return is relatively wide (2.5 mm or greater) it forms the miniature ground plane that was previously discussed. The resulting minimum loop area minimizes capacitive coupling as well as antenna effects that inject noise at the input of the gate driver. In addition, relatively high peak gate drive currents get some field cancellation, which reduces radiated noise. The other major source of gate drive noise that causes false transitions is non−zero voltage drops in power grounds. Using opto couplers and routing each gate drive return directly to the emitter of its corresponding power device is one of the ways to provide noise immunity. For motor drives where opto couplers are not practical, taking care to minimize the inductance between power device emitters or sources is a viable alternative.

In terms of reducing the amount of noise that is produced by power stages, minimizing loop areas is a key consideration. The most important is the loop that includes the upper half − bridge IGBT drain, lower half − bridge

IGBT source, and high frequency bus decoupling cap. The idea here is to try to keep the high di/dt that is produced during diode reverse recovery in as small an area as possible. This is a part of the circuit where running traces that have equal but opposite currents directly over each other is a priority. Since the currents into and out of the decoupling cap are equal and opposite, running these two traces directly over each other provides field cancellation and minimum loop areas where they are needed most.

Figure 4 illustrates the difference between a loop that has been routed correctly and one that has not. In this figure, the solid circles represent pads, the schematic symbols show the components that are connected to the pads, and two routing layers are shown with cross−hatching that goes in opposite directions. Note that by routing the two traces one over the other that the critical loop area is minimized.

For similar reasons it is desirable to run power and return traces one directly on top of the other. In addition, if a current sensing resistor is used in the return, using a surface mount resistor is preferable due to its lower inductance. It also can be placed directly over the power trace, providing uninterrupted field cancellation from placing power and return traces over each other. Again for field cancellation, it is also desirable to run phase outputs parallel and as close as possible to each other.

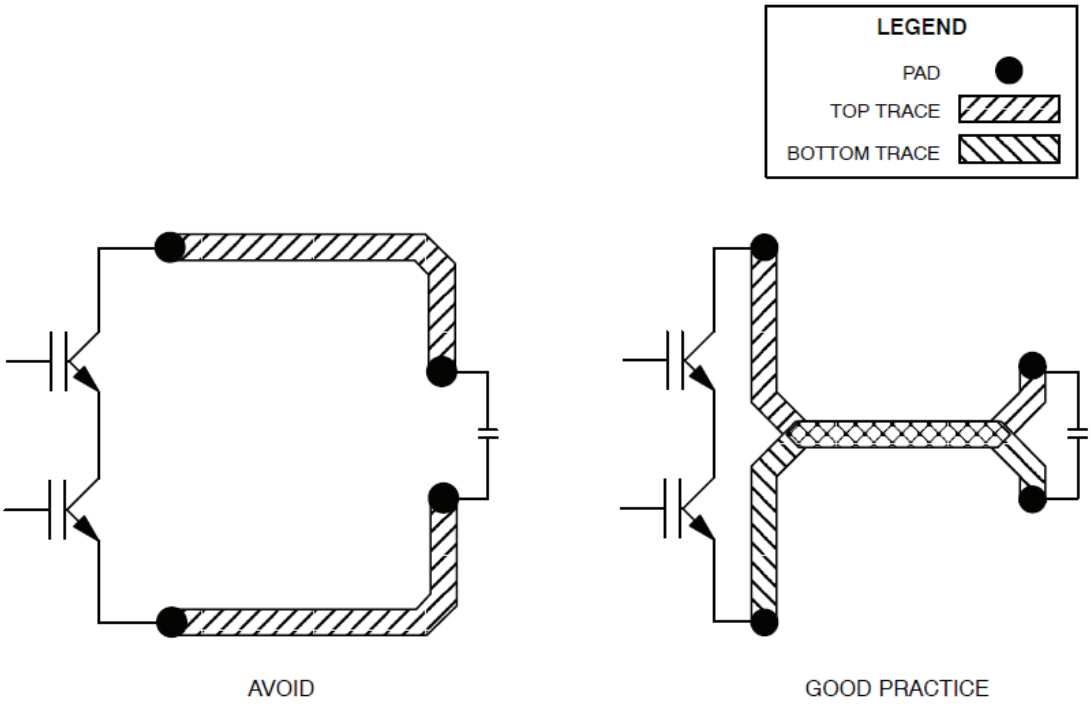


Figure 4. Minimizing Loop Areas

The power stage is the place where avoiding right angles is most important. Single traces are easy, two forty five degree angles or a curve easily accomplish a 90 degree turn. It is just as important to avoid 90 degree angles in T connections. illustrates correct versus incorrect routing for both cases.

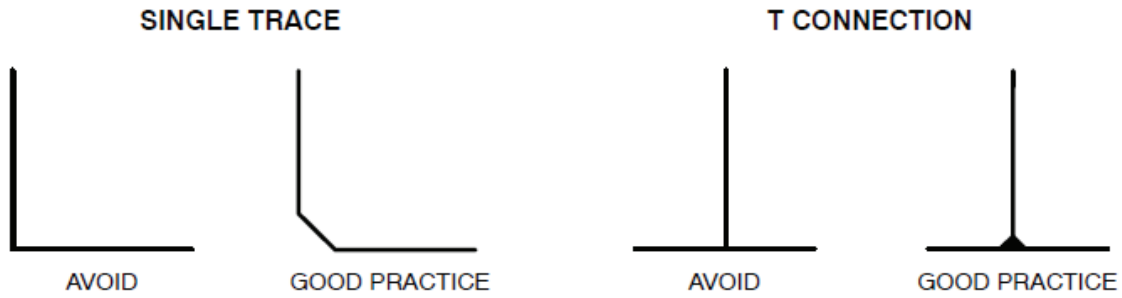


Figure 5. Routing to Avoid 90 Degree Angles

Layout considerations for Controllers:

The primary layout issue with controllers is ground partitioning. A good place to start is with the architecture that is shown in Figure 6. This architecture has several key attributes. Analog ground and power ground are both separate and distinct from digital ground, and both contact digital ground at only one point. For the analog ground, it is preferable to make the one point as close as possible to the digital converter's ground reference (VREFL). The power ground the connection should be as close as possible to the microcomputer's power supply return (VSS). Note also that the path from VREFL to VSS is isolated from the rest of digital ground until it approaches VSS.

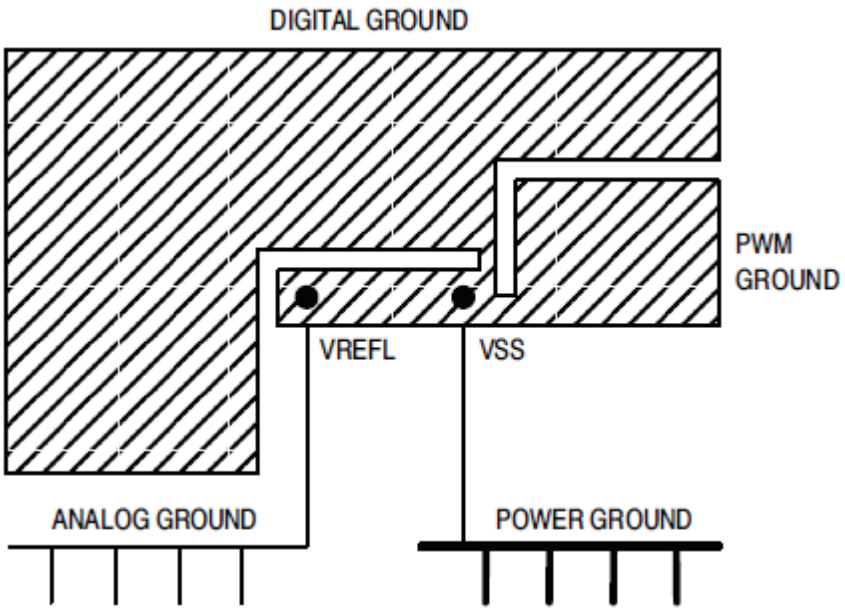


Figure 6. Ground Architecture for Controller Layout

The PWM ground is also isolated as a separate ground plane section until it approaches VSS. This is most important in systems that use optocouplers, since the current that flows through the PWM ground return will be higher than other digital return currents. If a two layer board is used, traces replace the ground planes that are shown in Figure 6. The partitioning, however, remains the same. In addition to grounding, controllers benefit from attention to avoiding 90 degree angles, since there are generally a lot of high speed signals on the digital portion of the board. Routing with 45 degree angles or curves minimizes unwanted reflections, which increases noise immunity.

CONCLUSION

For the most part, the functional architecture of motor drives is much more straightforward than some of the techniques that are required to get them to work. These challenging aspects arise from high levels of both di/dt and dv/dt that produce many noise management issues. These are systems in which a fraction of a picofarad of stray capacitance in the wrong place, a ground connection that is not carefully routed, or the absence of a functionally not so obvious component will all cause improper operation.

The most important design issues are careful attention to grounding, minimizing critical loop areas, use of series bootstrap capacitors, careful attention to power transistor transition times, and filtering sensor inputs. Additional benefits are gained by avoiding 90 degree angles in board layout and cancelling fields by routing equal and opposite current flows as close as possible to each other. As expected, consideration given to these issues up front pays off when it comes to getting a design to work right the first time.

AND9072/D

Effect of Gate-Emitter Voltage on Turn on Losses and Short Circuit Capability

ON Semiconductor®

http://onsemi.com

APPLICATION NOTE

INTRODUCTION

This application note describes some of the impacts of the gate−emitter voltage on the IGBT device performance. Unlike the MOSFETs and BJTs, the magnitude of the gate−emitter supply voltage of an IGBT has a more significant impact on the performance of the device. The magnitude of the gate−emitter voltage impacts the turn − on loss and short circuit survival capability of the devices.

BACKGROUND

While both MOSFETs and IGBTs have certain similarities in terms of a gate terminal being used to switch the device, there are operational differences that need to be considered when driving the IGBTs. A typical MOSFET's switching behaviour is minimally impacted by the gate voltage levels beyond a certain level. As a result, it makes sense to restrict the gate voltage in MOSFET drive circuit so that the total gate charge (Q_G) value is minimized. However, in case of IGBTs, there are other considerations, so when transitioning from MOSFETs to IGBTS, these considerations may require a re−evaluation of optimum gate voltage level.

IGBT Turn−on Switching Loss Considerations

There are three major areas of consideration when selecting the V_{GE} value for an IGBT.

1. V_{CE} variation – Unlike a MOSFET that gets fully enhanced at relatively low gate voltage (in most MOSFET output characteristics, there is minimal gain once V_{GS} crosses 10 V), the output/saturation characteristics of the IGBTs show continuing dependence on the V_{GE} voltage. This is best illustrated in the attached Figure 1 from an IGBT datasheet. As shown in the curves, there is a significant gain to be had in the saturation voltage (V_{CE}), as the V_{GE} value is increased beyond 11 V for same collector current value. In fact, higher current is not supportable until V_{GE} is increased above 13 V. Given high current applications, even a 0.1 V improvement in V_{CE} translates into power savings of 1 W for every 10 A of collector current.

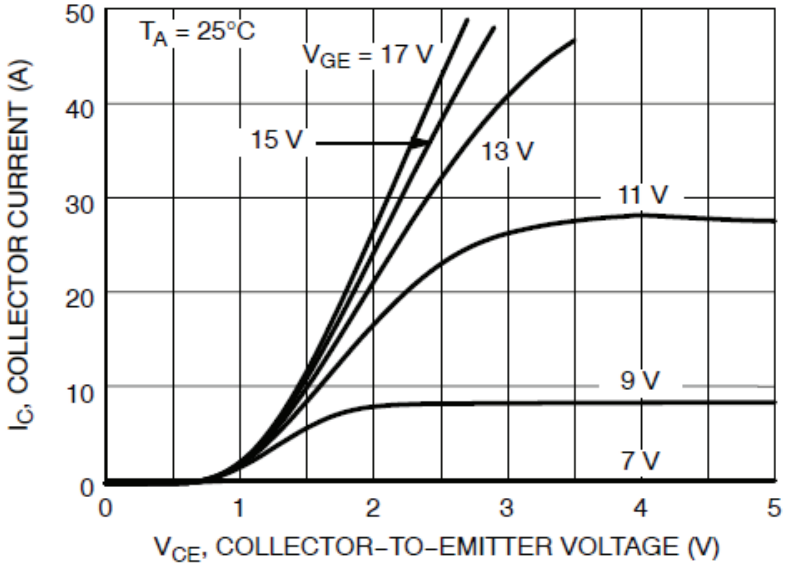


Figure 1. Output Characteristics of a 15 A, 600 V IGBT

2. Gate charge considerations – In order to fully turn−on the device in high voltage application, both MOSFET and the IGBT drivers must supply voltage beyond the plateau voltage that is needed to overcome the equivalent of the Miller effect in the device (providing the Q_{GD} and Q_{GC} charge at a constant voltage to the MOSFET and IGBT, respectively). The plateau voltage is shown in Figure 2. As shown here, the typical IGBT has a higher plateau voltage (near 10 V) than a typical MOSFET (near 5 V). There is also a higher variability in the IGBT plateau voltage as a function of collector current. As a result, it makes sense to set the IGBT V_{GE} value higher than the typical FET V_{GS} value. However, in both cases, increasing the gate voltage beyond the plateau voltage increases the value of total gate charge that has to be delivered every switching cycle. If we partition the gate−voltage vs. Q chart in three regions as shown in the figure, the first section is the Q_{GE} – from origin to the point the plateau region is reached. The next one is the Q_{GC} – which represents the plateau region and finally, the charge in the 3rd region is proportional to the actual value of V_{GE}. In this instance, every 2 V increase in V_{GE} value beyond the plateau voltage, leads to increase of about 10 nC in Q_G value. This increase leads to higher dissipation in the device as well as in the drive circuit due to the additional gate charge, but will also result in a lower V_{CE}. In that sense, it is advisable to keep the V_{GE} value always above the plateau voltage, but not too much higher than it.

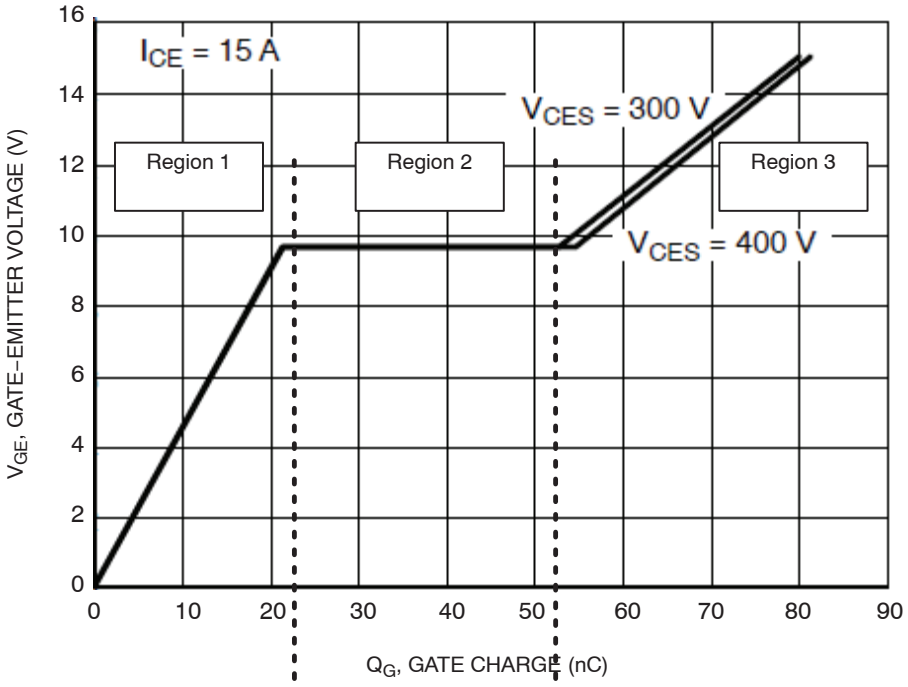


Figure 2. Gate Voltage vs. Gate Charge Characteristics (15 A, 600 V IGBT)

3. Transition time considerations – In addition to the factors described above, the turn−on time of the device is also determined by the V_{GE} value. With a given V_{GE} value, the available drive current is inversely proportional to the value of R_G. With a fixed R_G, an increase in V_{GE} results in a higher current and reduces the Q_{GE} section of the charge. This also results in shorter switching interval and significantly reduced turn−on energy – note that turn−on and turn−off energies in hard−switching applications are dominated by the collector current and voltage transitions. Many of the IC drivers are specified in terms of current drive capability, and hence increasing the V_{GE} value may have less impact in the applications where these drivers are used.

As shown in Figure 3, there is a dramatic drop in the turn−on energy, Eon, once the V_{GE} value goes above 12 V. However, there is not much difference in Eon for V_{GE} values between 15 V and 20 V. This applies for the full current range. As depicted in Figure 4, the variations in R_G also resulted in a significant variations in the Eon value. Based on these figures, it can be surmised that a high V_{GE} value (between 15 V and 18 V) and a low R_G value are the best combination for driving IGBTs.

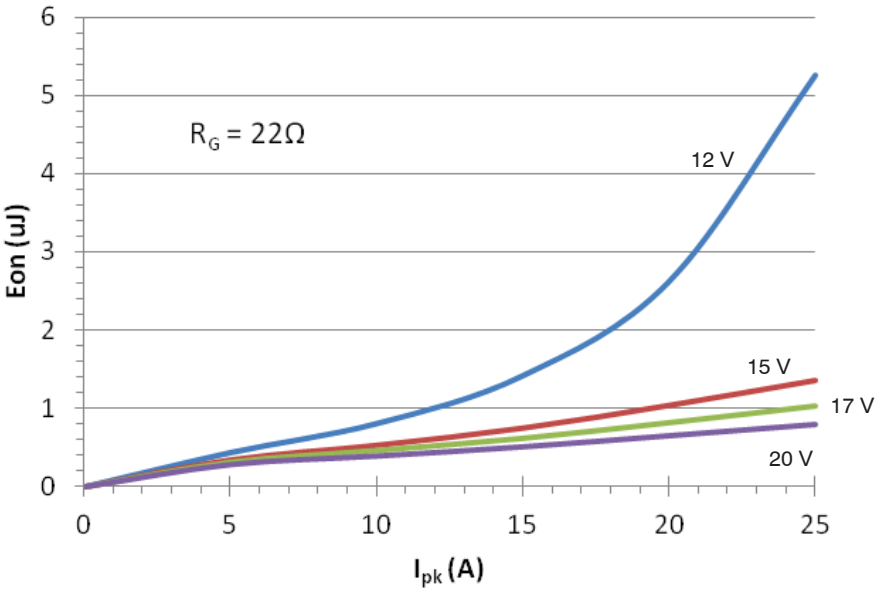


Figure 3. Turn−On Energy Characteristics (Fixed RG and varying VGE)

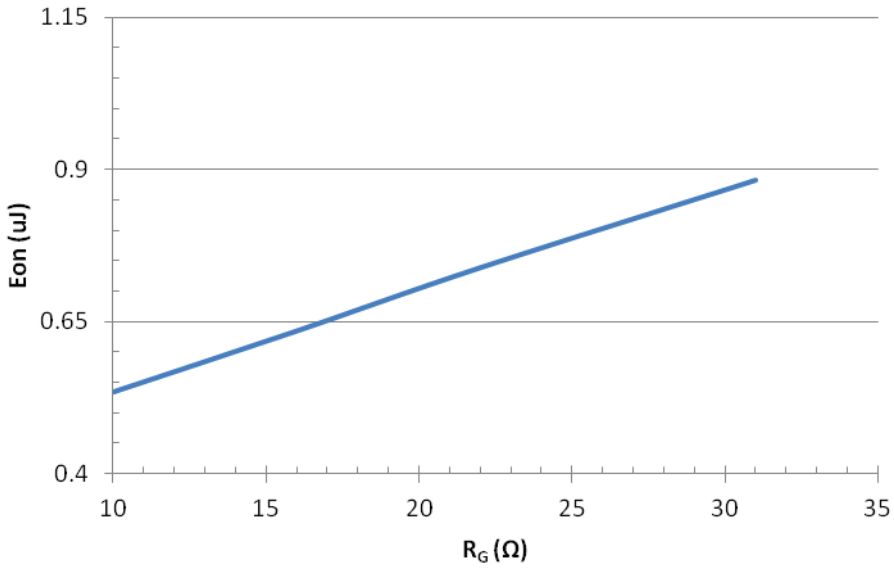


Figure 4. Turn−On Energy Characteristics (Fixed VGE and varying RG)

Based on the above three considerations, following key summaries can be drawn about the impact of VGE value on IGBT switching performance:

- It is better to use higher drive voltage, V_{GE}, (around 15 V) for IGBTs compared to the FET drive circuits
- While higher voltage helps in many ways (saturation, transition time etc.), pushing it too high can increase the gate charge requirements.
- Some margin from absolute maximum values (±20 V, typically) must be maintained to prevent any transients on the drive voltage from damaging the IGBTs.

SHORT CIRCUIT FAULT OPERATION

A major concern in inverter and motor drive applications is the ability to survive a short circuit fault condition. During the short circuit fault, the device is exposed to the supply voltage across the device, while the gate potential is at full operating value (see Figure 5). Because of the high gain characteristic of the IGBTs, the collector current will rise to some undetermined value limited by the gate − emitter voltage. During this time the device will have a large amount of energy across the device, and if the energy is beyond the capability of the device it will be destroyed due to thermal breakdown. For non−rugged devices, the large current can cause parasitic NPN bipolar transistor to turn on and cause the device to latch, wherein the gate control will be lost.

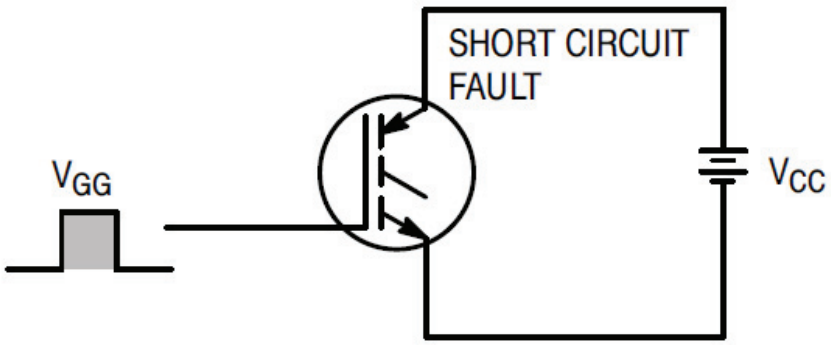


Figure 5. Equivalent Short Circuit Condition

Since an IGBT is less sensitive to second breakdown due to hot spot formulation compared to a BJT, it can survive a short circuit condition if the energy delivered to the device is maintained below some value that is tolerable to the device. An IGBT is typically specified with a guaranteed short circuit withstand time under given conditions (a typical value is 10 μs). It is expected that this withstand time allows an external circuitry to be activated and intervene to rectify the condition.

There are many different ways to protect the device from the short circuit condition for some duration. The most effective way to provide the short circuit survivability would be to inherently build current sensing capability into the device – however, that is not an option in most cases. Another method of increasing the short circuit survivability is to decrease the gate voltage when the short circuit across the device is observed. Figure 6 data shows the relationship between the gate voltage, short circuit current, and the short circuit survival time period. As shown in Figure 6, it is clear that the smaller gate voltage limits the current at lower value and increases the short circuit time duration. Thus, if more rugged performance and longer short circuit survivability is needed, it may make sense to trade−off some of the switching loss gain and reduce the V_{GE} voltage in the application.

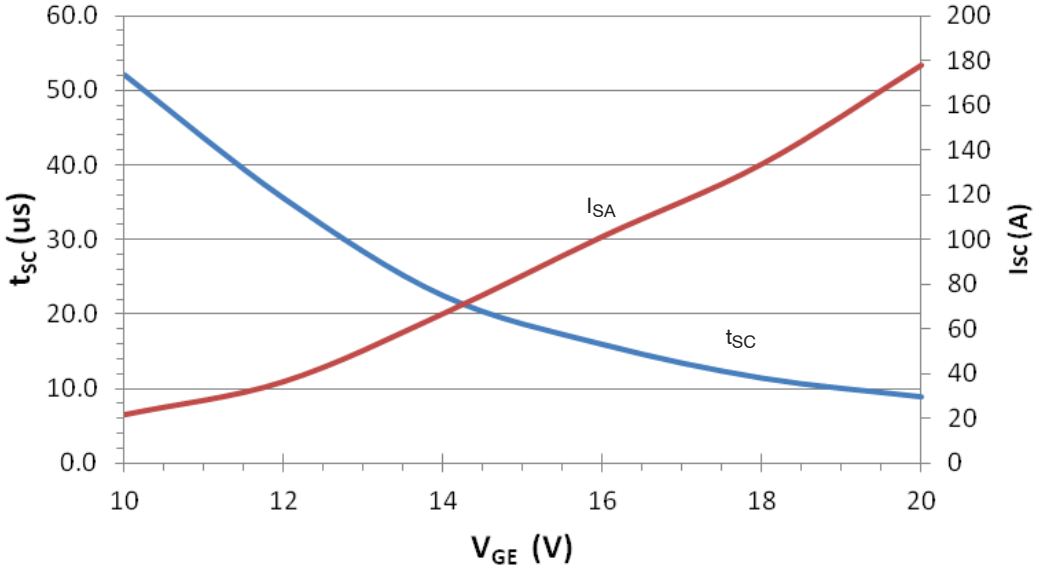


Figure 6. Short Circuit Response of IGBT

SUMMARY

IGBTs are high current and high voltage devices that offer certain benefits compared to the MOSFETs. Because of their use in high power applications, both lifetime considerations (ruggedness) and efficiency (low losses) are important. As discussed in this section, the magnitude of the gate−emitter voltage can be optimized in order to reduce the turn−on loss of the device. But on the other hand, the designer needs to understand that the high gate−emitter voltage reduces the short circuit survivability of the device. Using these two relationships and taking into consideration specific application requirements, the designer can choose the best voltage value (and the best IGBT) which will meet the design requirements.

AND9088/D

ON Semiconductor's Motor Control IGBTs and Free-Wheeling Diodes

ON Semiconductor®

http://onsemi.com

APPLICATION NOTE

ABSTRACT

Using IGBTs in motor control environments requires several attributes which can be divided in two categories. The short circuit rating and SOAs are an indication of the robustness of a device. The conduction and switching energy losses are measures of the efficiency of a device. This application note will examine ON Semiconductor's IGBTs and free−wheeling diodes and show how the losses and ruggedness issues affect the selection of an IGBT for a motor drive application. Motor control applications have several important particularities which will also be examined.

Introduction

The Insulated Gate Bipolar Transistor (IGBT) is a minority−carrier device with high input impedance and large bipolar current−carrying capability. Many designers view the IGBT as a device with MOS input characteristics and bipolar output characteristics, making it a voltage−controlled bipolar device. While the MOS input and BJT output are integrated monolithically on a single silicon die, it is often necessary to add an anti−parallel or free−wheeling diode to obtain a fully functional switch, although in some special cases the free−wheeling diode is not necessary. The diodes can be incorporated monolithically or co−packaged or as a discrete diode external to the IGBT package.

The introduction of insulated gate bipolar−transistors (IGBTs) in the mid−1980s was an important milestone in the history of power semiconductor devices. They are extremely popular devices in power electronics for medium to high power levels (a few kWs to a few MWs) and are applied extensively in dc/ac drives and power supply systems. The IGBT is suitable for many applications in power electronics, especially in Pulse Width Modulated (PWM) servo and three−phase drives requiring high dynamic range control and low noise. It also can be used in Uninterruptible Power Supplies (UPS), Switched−Mode Power Supplies (SMPS), and other power circuits requiring high switch repetition rates. IGBTs improve dynamic performance and efficiency and reduce the level of audible noise. It is equally suitable in resonant−mode converter circuits. Optimized IGBTs are available for both low conduction loss and low switching loss.

One of the more common applications of an IGBT is for use as a switching component in voltage source inverter (VSI) circuits. Generally, inverters are used in power supply and motor−control applications. The main advantages of using IGBTs in these converters are simplicity and modularity of the converter, simple gate drive, elimination of snubber circuits due to the square Safe Operating Area (SOA), low switching loss, improved protection characteristics (over−current and short circuit fault) and simpler electrical and mechanical construction of the power converters [1]. These advantages make the IGBTs very popular and promising switching devices.

Electric Drives

In contrast to grid connected ac motor drives, which are essentially constant frequency, power electronic devices (e.g. inverters), offer voltage supplies that are variable in both frequency and magnitude and are used to operate ac motors at frequencies other than the supply frequency.

Motor−control applications include mainly variable voltage, phase and frequency inverters. The purpose of a power converter is to produce a controllable voltage and frequency, and provide an ac output waveform from a DC link circuit. This DC link is often supplied by a controllable or uncontrollable AC−DC converter. Variable voltage and frequency waveforms are required to operate variable speed drives, UPS, active filters, compensators etc. which are only a few types of applications.

Pulse−width modulated, variable−speed motor drives are an application well suited for IGBTs. In this application, as shown in Figure 1 and Figure 2, IGBTs are used as the power switch to PWM the voltage supplied to a motor to control its speed, position or electromagnetic torque. Normally, the IGBT will be required to operate from a full−wave rectified line. This can require devices to have six hundred volt ratings for 230 VAC line voltage inputs, and twelve hundred volt ratings for 575 VAC volt line inputs. IGBTs that block high voltage, offer fast switching and low conduction losses, are an excellent choice for the design of efficient, high frequency drives of this type. Devices used in motor drive applications must be robust and capable of withstanding faults long enough for a protection scheme to be activated. Short circuit rated devices offer safe, reliable motor drive operation.

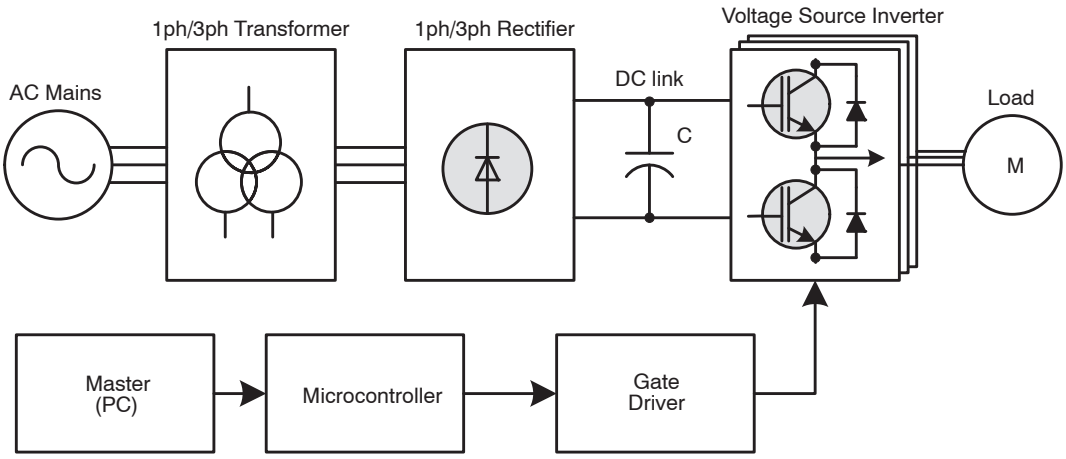


Figure 1. Typical AC Motor Drive Application

One of the most common converter topologies that is very widely used is shown in in Figure 1 and Figure 2. Figure 1 shows a general block structure of an electric drive and Figure 2 gives a more concrete view of the power topology. It consists of a three−phase bridge inverter with a diode rectifier in the front end. The rectifier (which can be single or three−phase) converts the ac input to an unregulated DC voltage. The harmonics in the DC link are filtered by an LC or C filter to generate a smooth DC voltage for the inverter. The inverter consists of three half−bridges or phase legs to generate three−phase ac for electric machines or other loads. Alternate sources for the DC power include a battery, fuel cell, or photovoltaic dc source. In all such cases, the DC voltage is usually unregulated. The battery−fed inverter drive is commonly used for electric/hybrid vehicle drives. Note that because of the diode rectifier in the front end, the converter system cannot regenerate power. The filter capacitor C sinks the harmonics from the rectifier as well as inverter sides.

According to the type of ac output waveform, these topologies can be considered to be voltage−source inverters (VSIs), where the independently controlled ac output is a voltage waveform. These structures are the most widely used because they naturally behave as voltage sources which are required by many applications, such as electric drives and are the most popular application of inverters. The typical VSI topology is shown in Figure 1 and Figure 2.

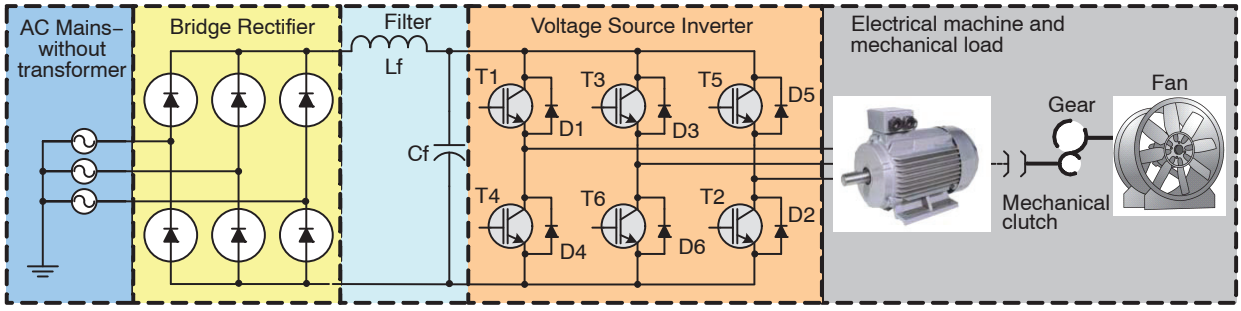


Figure 2. Common Topology of Motor Drive

Input Power Topology

The grid voltage is rectified by the line rectifier which typically consists of a diode bridge. Presently, attention paid to power quality and improved power factor has shifted the interest to more supply friendly ac−to−DC converters, e.g. a PFC topology. This allows simultaneous active filtering of the line current as well as regenerative motor braking schemes transferring power back to the mains.

It can be shown that single−phase diode rectifiers require a rather high transformer VA rating for a given DC output power. Therefore, these rectifiers are suitable only for low to medium power applications. For power outputs higher than 15 kW, three−phase or poly−phase diode rectifiers should be employed. The diode bridge rectifier shown in the Figure 2 is extremely important for generation of unregulated DC voltage. The diode rectifiers are simple, but the disadvantages are large distortion in line current and a poor displacement factor, which lead to a poor power factor. To combat these problems, various power factor correction (PFC) techniques based on active wave shaping of the line current will be proposed. The power factor controller is another application where IGBTs play an important role.

Filters are commonly employed in rectifier circuits for smoothing out the DC output voltage of the load. The typical filter capacitor is of appreciable size (2−20 mF) and therefore a major cost item. They are classified as inductor−input DC filters and capacitor−input DC filters. Inductor−input DC filters are preferred in high−power applications because more efficient transformer operation is obtained due to the reduction in the form factor of the rectifier current. Capacitor−input DC filters can provide volumetrically efficient operation, but they demand excessive turn−on and repetitive surge currents. Therefore, capacitor−input DC filters are suitable only for lower−power systems.

PWM Voltage Source Inverter (VSI)

In general, two basic types of inverters exist: Voltage−source inverters (VSI), employing a dc link capacitor and providing a switched voltage waveform, and current−source inverters (CSI), employing a dc link inductance and providing a switched current waveform at the motor terminals. CS−inverters are robust in operation and reliable due to the insensitivity to short circuits and noisy environments. VS−inverters are more common compared to CS−inverters since the use of Pulse Width Modulation (PWM) allows efficient and smooth operation, free from torque pulsations and cogging [1]. Furthermore, the frequency range of a VSI is higher and they are usually more inexpensive when compared to CSI drives of the same rating.

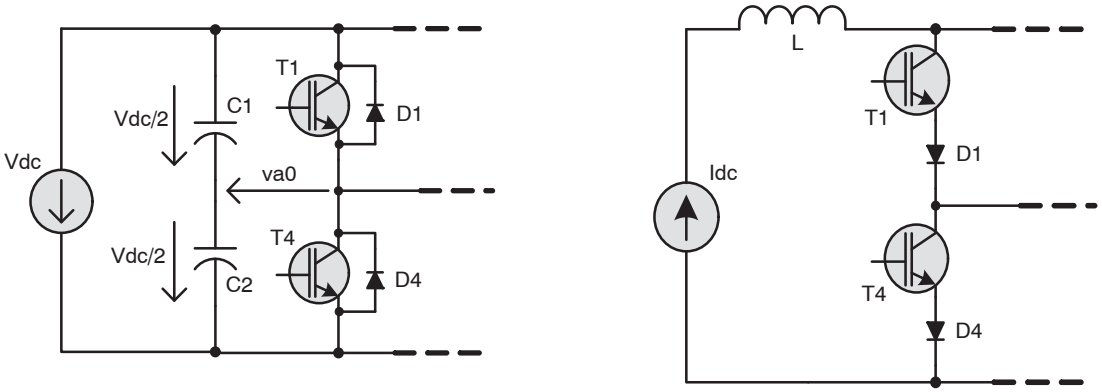


Figure 3. Voltage (left) and Current (right) Type of Half−bridge Converters

A typical voltage−source PWM converter performs the ac to ac conversion in two stages: ac to DC and DC to variable frequency ac. The basic converter design is shown in Figure 3. VSI, and as the name indicates, receives DC voltage at one side and converts it to an ac voltage on the other side. The ac voltage and frequency may be variable or constant depending on the application. In fact, the general name “converter” is given because the same circuit can operate as either an inverter or as a rectifier. A voltage−fed inverter should have a stiff voltage source at the input. This can also be obtained from a fuel cell, or solar photovoltaic array. The inverter output can be single−phase or polyphase and can have a square wave, sine wave, PWM wave, stepped wave, or a quasi−square wave at the output.

In voltage−fed converters, the power semiconductor devices always remain forward−biased due to the DC supply voltage, and therefore, self−controlled forward or asymmetric blocking devices, such as IGBTs are suitable. A free−wheeling (or anti−parallel) diode is always connected across the device to allow for reverse inductive current flow. One important characteristic of a voltage−fed converter is that the ac fabricated voltage waveform is not affected by the load parameters. The typical PWM output line and phase voltages are shown in Figure 4 and in Figure 5 along with the phase current.

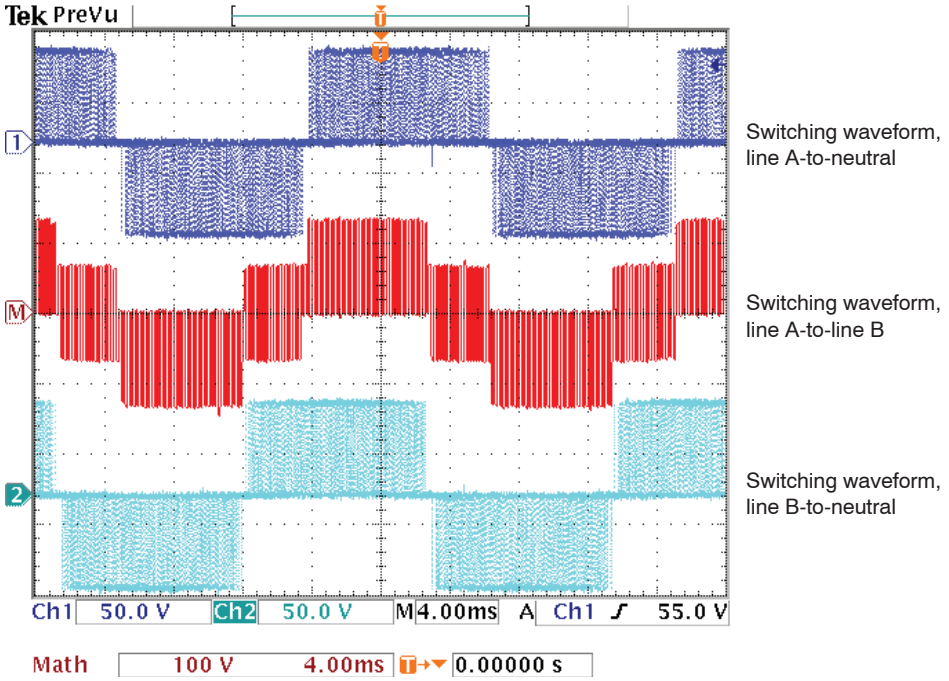


Figure 4. Phase−to−neutral and Phase−to−phase Switching Waveforms

IGBTs in Motor Drive Application

One of the critical issues in designing reliable power converters is the reliability of the power switches. The devices used in these applications must be robust and capable of withstanding faults long enough for a protection scheme to be activated. Inverters are a typical hard−switching application. In this switch−mode operation, the switches are subjected to high switching stresses and high switching power losses that increase linearly with the switching frequency of the PWM.

IGBT's have become the transistor of choice for motor control applications. The characteristics that make them favorable include their ability to pass greater current than an equivalent die size MOSFET transistor with more favorable drive schemes than bipolar transistors. In addition, IGBT's in many instances have lower conduction losses due to the V_{CEsat} when compared to the R_{DSon} of MOSFET's (this is due in part to the IGBT's loss based on collector current while the MOSFET's loss is based on the drain current squared).

An unfavorable attribute exhibited by IGBT's is the "tail time" that results from stored charge in the internal PNP transistor. The tail time is sacrificed at the expense of forward voltage drop. Devices are optimized for efficient operation in applications with regards to conduction and switching losses. The tail time issue also dictates switching speed. Motor controllers tend to operate at switching frequencies from 4 kHz to 20 kHz with 20 kHz being the preferred frequency due to the audible range of human hearing. IGBT's with tail times of less than 300 ns are being developed and these devices work well at 15 kHz. Work is under way to decrease the tail time and V_{CE} drop so that IGBT's can be operated at even higher frequencies and compete better with existing MOSFET's in motor control and power supply applications.

Since no isolation transformer is required for most motor drive applications, raising the switching frequency above the audible range has little advantage sine the filter inductance is inherent in the motor and a higher frequency will not allow for a more efficient motor design.

Anti−parallel Diode

Parallel to the power switches, free−wheeling diodes are placed across the collector−emitter terminals to conduct reverse current. These diodes are required, since switching off an inductive load current can generate high voltage peaks if a suitable path is not provided, which could destroy the power switch. The basic configuration of one inverter output phase consists of upper and lower power devices T1 and T4 (see Figure 5), and free−wheeling diodes D1 and D4. Due to the structural differences the IGBT does not have a parasitic diode like that found in a MOSFET. The optimal setup is to have the diode co−packaged with the IGBT.

A specific line of IGBTs has been created by ON Semiconductor to address this issue. These devices work very well in applications where energy is recovered to the source and are favored by motor control designers. Like the switching device itself, the anti−parallel diode should exhibit low leakage current, low forward voltage drop and fast switching speed. The diode forward drop multiplied by the average current it passes is the total conduction loss produced for the diode. In addition to the conduction losses, large reverse recovery currents can generate high switching losses. A secondary effect caused by large reverse recovery currents is generated EMI at both the switching frequency and the frequency of the resulting ringing waveform. This

EMI requires additional filtering to be designed into the circuit. By co−packaging parts, the parasitic inductances that contribute to the ringing are greatly reduced. Also, co−packaged products can be used in designs to reduce power dissipation and increase design efficiency.

For one inverter leg, Figure 5 shows the basic configuration and the inverter output voltage waveform depending on the switching state and the polarity of the current. The basic configuration of one inverter output phase consists of upper and lower power devices T1 and T4, and anti−parallel diodes D1 and D4.

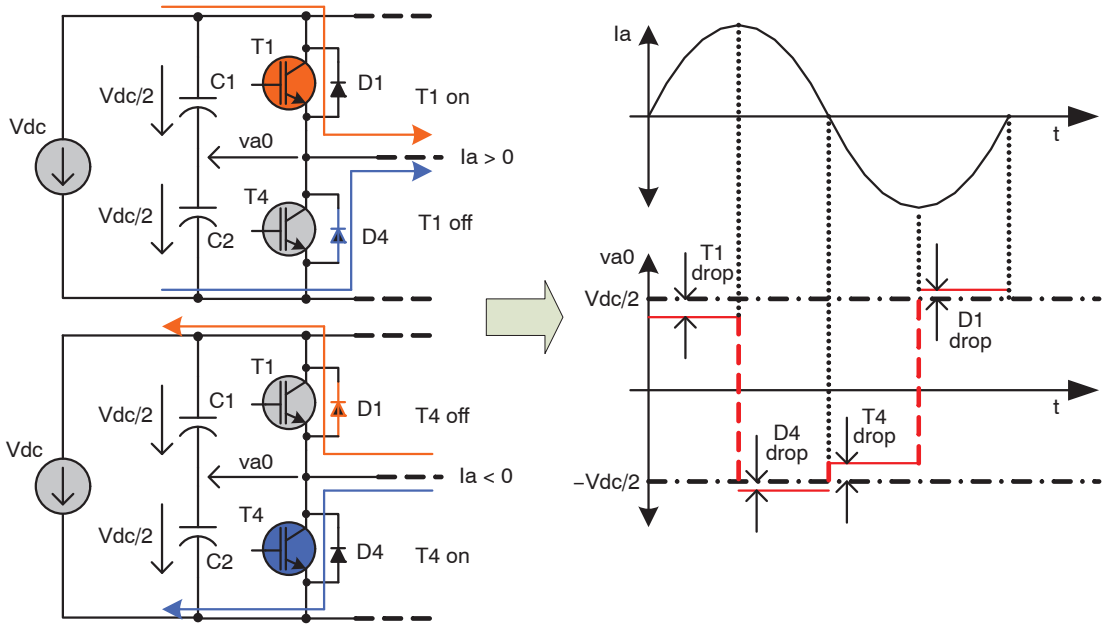


Figure 5. Fundamental Topology of a Voltage Half−bridge Inverter and Center−tapped Inverter Output Voltage. Left: Switching States and Current Direction. Right: Output Voltage and Line Current.

Pulse Width Modulation (PWM)

Usually, the on and off states of the power switches in one inverter leg are always opposite. Therefore, the inverter circuit can be simplified into three (in the case of a three−phase inverter−see Figure 2) 2−position switches. Either the positive or the negative dc bus voltage is applied to one of the motor phases for a short time. Pulse width modulation (PWM) is a method whereby the switched voltage pulse widths are varied to produce different output frequencies and voltages. A PWM produces an average output voltage value, equal to the reference voltage (scaled by a constant factor) within each PWM period.

There are various PWM schemes. Well−known among these are sinusoidal PWM, hysteretic PWM, space vector modulation (SVM) and “optimal” PWM techniques based on the optimization of certain performance criteria, e.g. selective harmonic elimination, increasing efficiency, and minimization of torque pulsation. While the sinusoidal pulse−width modulation and the hysteretic PWM can be implemented using analog techniques, the remaining PWM techniques require the use of a microprocessor.

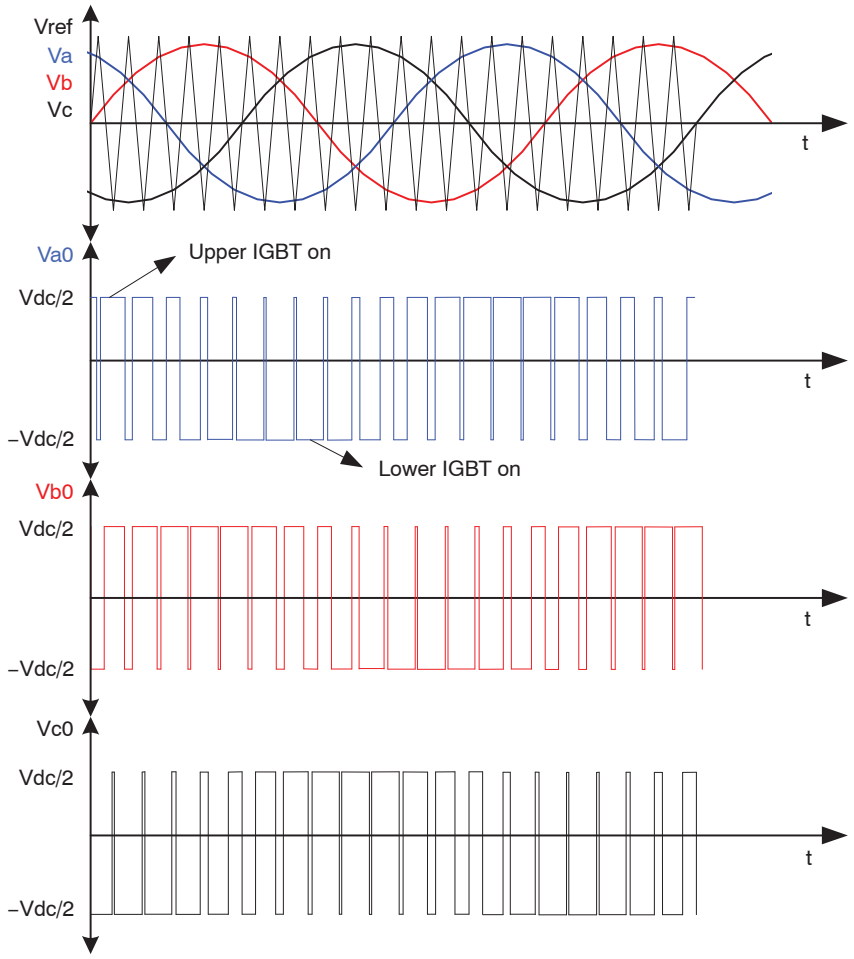


Figure 6. The Rule of Sinusoidal PWM Generation (SPWM)

As shown in Figure 6, a saw−tooth carrier wave is compared to the reference voltage for each phase to generate the fixed PWM pulses for all three phases. This modulation technique, also known as PWM with natural sampling, is called sinusoidal PWM because the pulse width is a sinusoidal function of the angular position in the reference signal. Since the PWM frequency, equal to the frequency of the carrier wave, is usually much higher than the frequency of the reference voltage, the reference voltage is nearly constant during one PWM period. Depending on the switching states, the positive or negative half DC bus voltage is applied to each phase. At the modulation stage, the reference voltage is multiplied by the inverse half dc bus voltage compensating the final inverter amplification of the switching logic into high power pulses that are applied directly to the motor windings.

Electrical Machines and Mechanical Loads

The electrical machine that converts electrical energy into mechanical energy (and vice versa) is the workhorse in a drive system. Drive systems are widely used in applications such as pumps, fans, paper and textile mills, elevators, electric vehicles, subway transportation, home appliances, wind generation systems, servos and robotics, computer peripherals, steel and cement mills, ship propulsion, etc. A machine is a complex structure electrically, mechanically and thermally. However the evolution of machines has been slow compared to that of power semiconductor devices and power electronic converters [1].

Among all types of AC machines, the induction machine, particularly the cage type, is most commonly used in industry. These machines are very economical, rugged, reliable, and are available in ranges from fractional horse power to multi−megawatt capacity. Low−power machines are available as single−phase machines but poly−phase (three−phase) machines are used most often in variable−speed drives [1]. Figure 7 shows a typical induction machine, while Figure 8 illustrates the typical, mechanical linkage from the machine to the load.

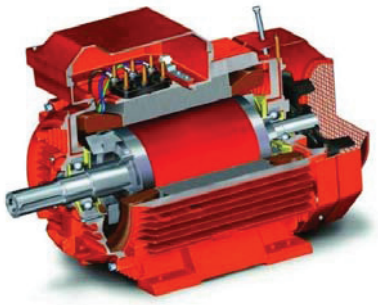


Figure 7. The Typical Induction Machine Construction

The motor is connected to a mechanical load, which may be simple or complex. Figure 8 shows an example of a mechanical load. The machine rotor with inertia J develops torque T while running at mechanical speed ω and instantaneous position θ. The motor shaft is usually connected through mechanical clutch and motor speed is adapted by a mechanical gear coupling. The secondary gear is connected to a load (in our case it is fan). Each mechanical load is characterized by a torque/speed characteristic which has, in this case, parabolic characteristics with speed ($T \approx \omega^2$).

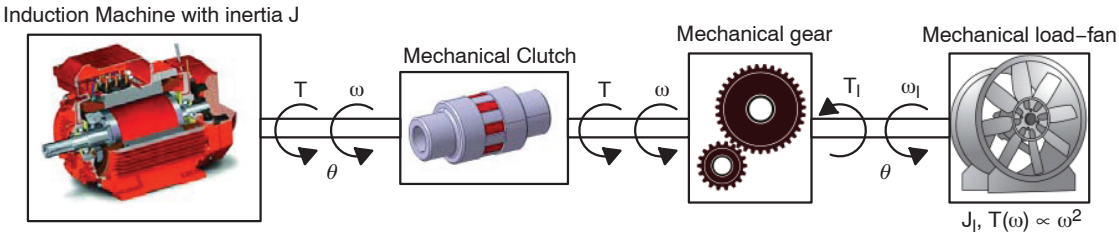


Figure 8. Simple Mechanical Load for a Motor

IGBTs for Motor Control Application

The fundamental trade−off between turn−off switching loss (Esw(off)), on state voltage drop (V_{CEsat}) and robustness in IGBT chip design is well known. Standard industrial IGBTs are typically optimized for motor drive and similar applications in which the carrier frequency is typically 10 kHz or less. For these applications conduction losses tend to dominate so the IGBT chip is primarily optimized for low V_{CEsat}. Other applications such as electric drives for white goods often require higher operating frequencies to reduce audible noise for the overall system. In these higher frequency applications dynamic losses become more dominant and often limit the maximum switching frequency of the IGBT.

Hard Switching and Switching Losses

Most motor control applications for IGBTs require “hard” switching of ohmic−inductive loads with continuous load current, i.e. the time constant of the load L/R is much larger than the cycle 1/fsw of the switching frequency [4]. The typical application of the IGBTs in motor control applications is a voltage source inverter (VSI) which is a hard−switched application.

The device waveforms (see turn−on and turn−off of an IGBT in Figure 9) of a hard−switched inverter have a number of detrimental effects, which can be summarized as follows [1]: **(a) Switching Loss** – the overlapping of voltage and current waveforms during each turn−on and turn−off switching cause a large pulse of power loss as shown in the math trace (M) in Figure 9. Since the amount of energy lost for each cycle is constant, as the switching frequency increases, the switching power losses increase.

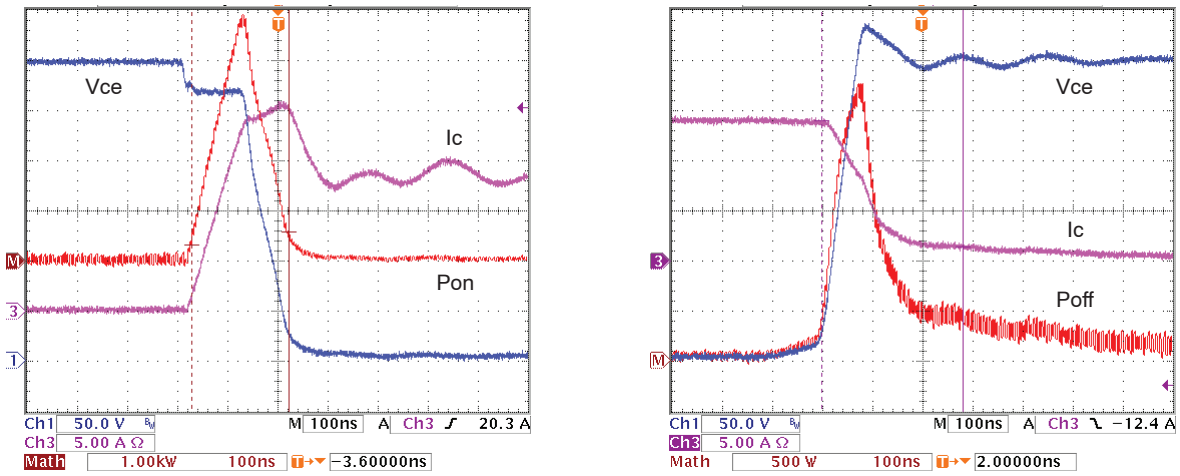


Figure 9. IGBT Turn−on and Turn−off Dynamics. Blue: Vce, Pink: Ic, Red: Ploss

The power loss sue to the switching losses is

$$P_{SW} = \left(E_{on} + E_{off}\right) \cdot f_{SW} \qquad \text{(eq. 1)}$$

An additional problem is that the cooling system is burdened due to higher losses. In fact, the PWM switching frequency of an inverter is limited because of switching loss. **(b) Device Stress –** In hard switching, the switching locus moves through the active region of the volt−ampere area which stresses the device. The reliability of the device may be impaired due to prolonged hard switching operation. This is one of the reasons for robust requirements of a power device in such applications. The SOA curve, on the data sheet, shows the limits for device operation beyond which the part may be damaged. **(c) EMI Problems –** High dv/dt, di/dt, and parasitic ringing effects at the switching transitions of a fast device can create EMI problems, which may affect the control circuit and nearby apparatus. Parasitic leakage or coupling inductance, although quite small, can be a source of EMI due to large induced (L·di/dt) voltages. Similarly, high dv/dt transients can induce common mode coupling currents (C·dv/dt) in the control circuit through the parasitic capacitance. **(d) Effect on Machine Insulation –** High dv/dt impressed across the stator winding insulation can create large displacement current (C·dv/dt), which can deteriorate machine insulation. **(e) Machine Bearing Current –** Recently, it was determined that PWM inverter drives with switching IGBT devices are known to cause a machine bearing current problem. This is due to the fast switching of IGBTs. The switching of the IGBT creates a high dv/dt source that will cause C·dv/dt current flow to the ground through machine shaft and stray capacitance of the insulated bearing. This current will tend to shorten the bearing life. **(f) Machine Terminal Overvoltage –** PWM inverters are often required to link to a machine with a long cable. The high dv/dt at the inverter output boosts the machine terminal voltage by the reflection of the high−frequency travelling wave. High−frequency ringing occurs at the machine terminals due to stray circuit parasitics. The resulting excessive overvoltage threatens the motor insulation. As you can see in optimization triangle in Figure 12, some aspects of hard−switching are inputs to the optimization process.

Switching losses are the power losses dissipated during the turn−on and turn−off switching transitions. In high frequency PWM switching, losses can be substantial and must be considered in the thermal design. The most accurate method of determining switching losses is to plot the Ic and Vce (see Figure 9 and Figure 10) waveforms during the switching transition. Multiply the waveforms point by point to get an instantaneous power waveform.

Most oscilloscopes have math functions available to perform the multiplication. This can be integrated to give the energy during the switching transistion and then multiplied by the switching frequency to obtain the power loss for that transistion.

Conduction Losses

Conduction losses are the losses that occur while the IGBT or FWD is on and conducting current. The total power dissipation during conduction is computed by multiplying the on−state voltage by the on−state current. In PWM applications the conduction loss must be multiplied by the duty factor to obtain the average power dissipated. A first order approximation of conduction losses can be obtained by multiplying the IGBT's rated V_{CEsat} by the expected average device current. In most applications the actual losses will be less because V_{CEsat} is lower than the data sheet value at currents less than rated Ic. Conduction losses of an IGBT and FWD are an important part of total losses in motor control applications and the main indicators of these losses are the V_{CEsat} and Vf parameters. Due to the relatively low switching frequency we hold the V_{CEsat} as low as possible. Very often the forward voltage drop and switching speed are traded off to enhance the short circuit capability.

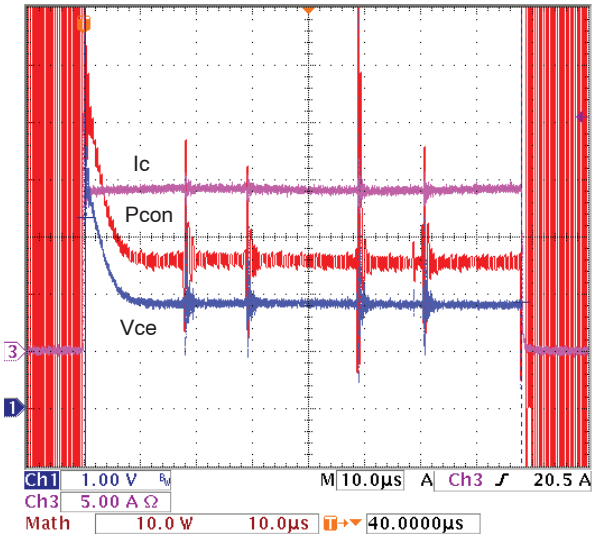


Figure 10. IGBT Conduction Dynamics
Blue: Vce, Pink: Ic, Red: Ploss

Free−wheeling diode conduction losses can be approximated by multiplying the data sheet Vf by the expected average diode current.

Process parameters of the IGBT such as threshold voltage, carrier lifetime, and the device thickness can be varied to obtain various combinations of SOA, on−state voltage, and switching time. However, there is very little overlap in the optimum combination for more than one performance parameter [2]. This is the reason that V_{CEsat} is one of the three inputs for process optimization.

FWD Importance in Hard Switching

For a long time, the importance of fast diodes had been underestimated. The performance of the IGBT switch had been impaired by the free−wheeling diodes [4]. However, ON Semiconductor free−wheeling diodes play an important

role and similar optimized processes have been implemented (see Figure 12). Firstly, from an efficiency point of view the main part of the total power loss is the conduction loss. The forward voltage drop is a very good indicator of this part of the loss so the Vf is one of the key optimization process inputs.

As soon as the IGBT is turned on, the diode will be commutated, and reverse recovery current will flow for a short period of time. During turn−on, the IGBT takes over the reverse current of the free−wheeling diode. Unfortunately, while the IGBT conducts the peak reverse current, the IGBT−voltage is still at the DC−link voltage level (see Figure 9). This is the moment of maximum turn−on losses in the IGBT. This means that the reverse recovery time directly influences IGBT turn−on. In addition a snappy shape of the reverse recovery will make the system noisy so this phenomenon has to be held smooth and soft. That means reverse recovery is another input to the optimized process as you can see in Figure 12.

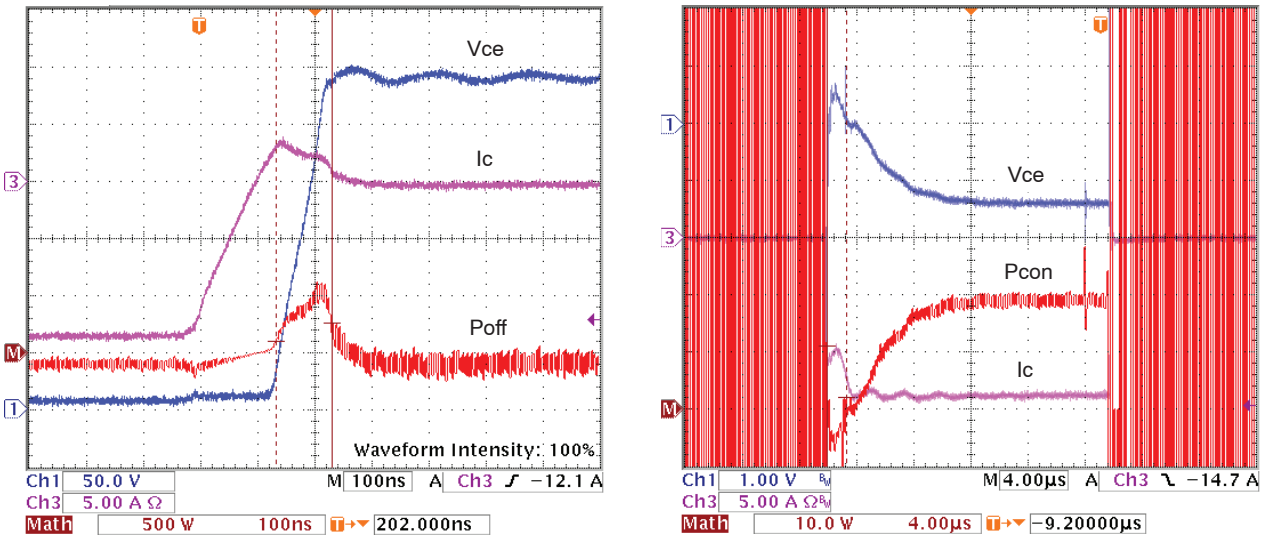


Figure 11. FWD Turn−off and Conduction Dynamics. Blue: Vce, Pink: Ic, Red: Ploss

An equally important requirement for free−wheeling diodes is dynamic ruggedness. Figure 11 shows that nearly the whole DC−link voltage is taken up by the diode, while it is still conducting a substantial tail current. If the IGBT is switched very quickly, reverse current and tail current will rise, at the same time causing a decrease of of the breakdown voltage of the IGBT, which switches over to the diode with a respectively higher dv/dt. The density of the current−carrying charge carriers will then be above the original doping density, which causes the diode to avalanche while it is turning off. This in turn reduces its breakdown voltage until the junction is depleted of carriers at which time the breakdown voltage returns to its normal level. The dynamic ruggedness is defined as the diodes ability to manage high commutating di/dt and a high DC−link voltage at the same time [4].

IGBT Robustness and Reliability

As mentioned the IGBT has to manage an almost rectangular characteristic between Vce and Ic in the case of hard switching. The SOA (Safe Operating Area) diagrams are defined as the current−voltage−power boundary within which a power switching device can be operated without destructive failure. For an IGBT, the area is defined by the maximum collector−emitter voltage Vce and collector current Ic within which the IGBT operation must be confined to protect it from damage. The robustness and reliability can be quantified by several SOA curves. There are SOA curves for switching, on−state and single pulse operation. The short−circuit capability is described by short−circuit safe operating area (SCSOA), switching capability by Forward−Biased Safe Operating Area (FBSOA), Reverse−Biased Safe Operating Area (RBSOA) and Unclamped Inductive Switching (UIS). The FBSOA is an important characteristic for applications with inductive loads and the RBSOA is important during the turn−off transient. The UIS is an important in high speed switching circuits. The simple definitions of meantioned are:

- SCSOA – describes the capability of a transistor to be controllable by its gate voltage at significant levels of collector−emitter voltage and collector current within a defined time interval.
- FBSOA – is defined as the maximum collector−emitter voltage allowable at a specified value of saturated collector current.
- RBSOA – describes the capability of a transistor to withstand significant levels of collector−emitter voltage and collector current in the turn off mode.
- UIS – the result from UIS testing is amount of energy a device can safely handle in avalange mode resulting from an unclamped inductive load.

ON Semiconductor optimizes its IGBTs for every aspect of robustness and reliability. The short−circuit capability is one of the features that is important for bridge applications because there are several circumstances under which the IGBT can see a short−circuit path from the DC voltage across the motor phase IGBTs. An IGBT must withstand this failure for time interval appropriate for its application. In addition, electric drives often work in very harsh conditions, which increase the stresses on the IGBTs. ON Semiconductor has designed its IGBTs for motor control applications and for other bridge applications with values of time periods up to 10 μs.

IGBT Trade Offs

ON Semiconductor has developed a new generation of 600 V IGBTs co−packed with free−wheeling diodes (FWD). These co−packaged products have been introduced in standard discrete packages such as the TO−247 package. The IGBTs are based on a technology platform specifically developed to serve the ever increasing requirements for motor control applications. The cell structure, fabrication process, and starting material are optimized to obtain the best possible trade−offs for this type of application. The triangles, shown in Figure 12, symbolize the trade−off processes involved in the design of IGBTs and their free−wheeling diodes. The three vertices of this triangle represent conduction losses switching losses and robustness.

While the switching losses are lowered as carrier injection efficiency is increased and the life time is reduced, the conduction losses go exactly the opposite way. The short circuit time withstanding capability is improved by incorporating design changes which reduce the short circuit current. Just as for switching losses, improvement in short circuit rating is accompanied by increase in conduction losses [6].

There are, however, design parameters which help reduce conduction losses without adversely influencing the other two. This was achieved for ON's IGBTs. Consequently ON Semiconductor's IGBTs offer low conduction losses and short circuit robustness while maintaining ultra fast switching speeds.

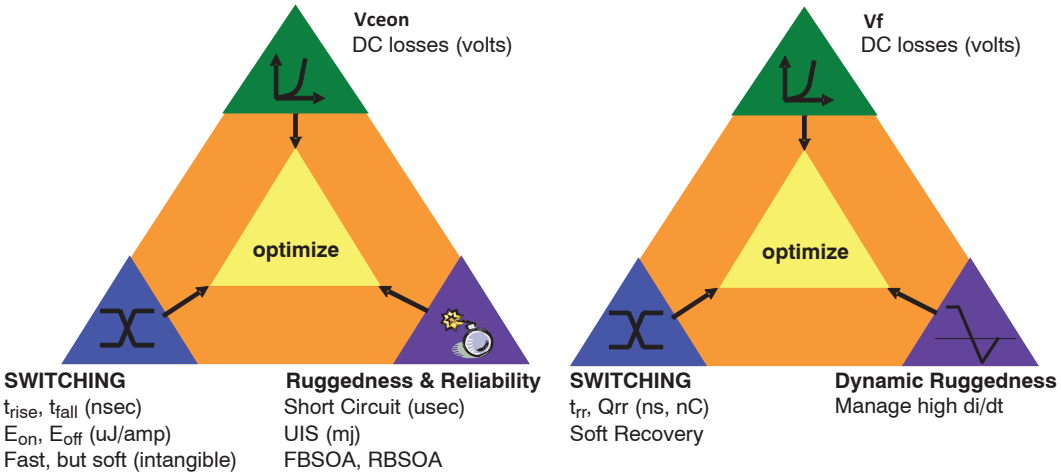


Figure 12. The IGBT and FWD Trade−off Triangle

The free−wheeling diodes are an integral part of motor drive application. ON Semiconductor has recently made tremendous strides toward perfecting the speed and softness in the recovery of these diodes. The rectifiers are optimized for fast switching speed and softer recovery throughout the current and di/dt range. These characteristics have helped reduce turn−on losses in IGBTs and generation of EMI and other transients.

Motor Drive Test set−up for Comparison of IGBT Technologies

The specific utilization of IGBTs and freewheeling diodes within a motor control system is designed to provide the inversion function from a DC Bus to a sinusoidal waveform. The AC waveform provided to the load is in the 1.0 Hz to 120 Hz frequency range, and is achieved by switching the IGBTs at PWM frequencies in the range of 4.0 kHz to 20 kHz. Typical output voltage waveforms are shown in Figure 4 and Figure 15.

Recent developments in IGBT technologies have focused on the tradeoffs of the device's forward voltage drop (V_{CEsat}) versus switching speed limitations due to device "tail time" in motor drive applications. Tail times are related to the stored charge in the IGBT's internal BJT and can be reduced at the expense of increasing the V_{CEsat} of the device. In general an increase in carriers will reduce the V_{CEsat} but slow down the switching speed. Conversly, a decrease in carriers will increase the V_{CEsat} but speed up the switching speed. Several technologies have been developed that attempt to optimize both switching times and forward voltage drop while providing rugged short circuit capability. The following work investigates the performance of several IGBT technologies in an adjustable speed motor drive

application. An off−line, 3 kW motor control test bed was used to test the performance of 600 V, 15 amp IGBT's.

3 kW Variable Speed Drive Test Set−up

A new line of IGBT devices has recently been introduced by ON Semiconductor. Devices rated at 600 V, 15 amps, are included and targeted for industrial drives of one phase 120 Vrms and/or 230 Vrms. These particular IGBT's are ruggedized for motor control applications and are specified to withstand 10 μs of short circuit current. Six of these devices were inserted into a bridge configuration in a 3−phase, 3 kW, variable speed, motor controller that operated at 10 kHz. The test system was then used to evaluate the different generations of IGBT's. Figure 13 illustrates the block structure of test system which uses a volt−Hz open−loop or closed−loop control technique, field−oriented control or direct torque control technique. The overall test bed can be divided into the motor drive system and load drive system. Both are mechanically coupled with a clutch. The energy is then transformed from electrical to mechanical energy by an induction machine. This mechanical energy is then transferred back to electrical energy by another load induction machine (generator) and then dissipated as thermal energy by an attached loading device.

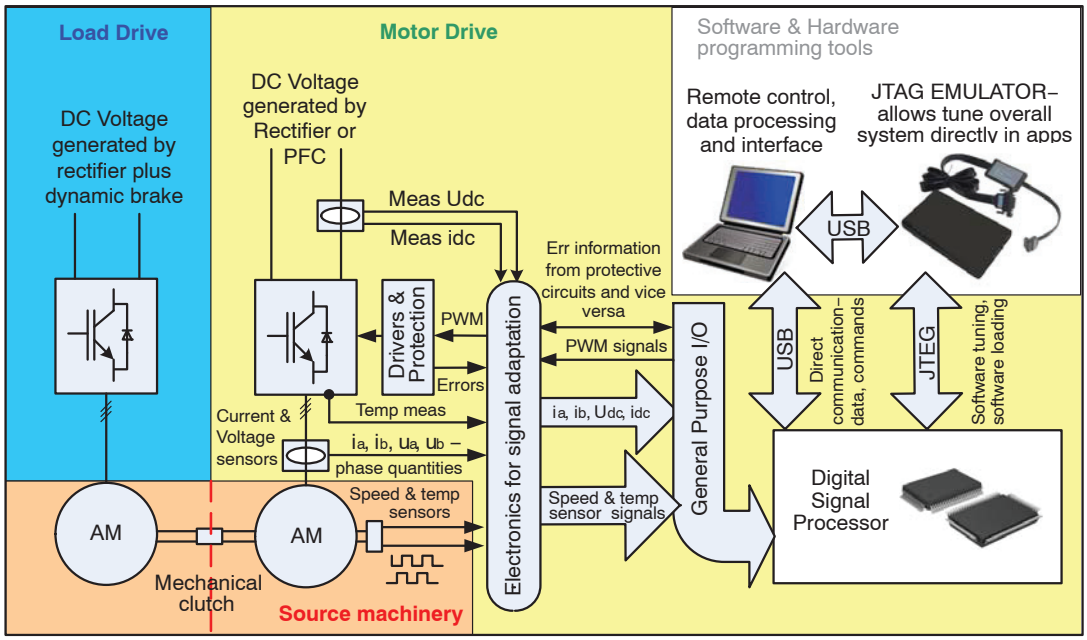


Figure 13. Block Structure of Laboratory Test Set−up

Motor drive technology has been established as a complex and multidisciplinary technology. Specialization in this area requires knowledge of power semiconductor devices, converter circuits, electrical machines and mechanics, control electronics, microprocessors and DSPs, ASIC chips, control theories, power systems, and computer−aided design and simulation techniques. Knowledge of electromagnetic interference, the passive components of such a system, and the accompanying specialized design, fabrication, and testing techniques are equally important [1].

Due to the complexity and number of sensing parameters, it is important to pay special attention to the signal processing circuits. The signals that are required to be monitored posses different forms – voltage or current and analog, digital and quadrature signal. For the proper functioning of an electric drive, multiple signals must be sensed, processed and converted for the drive to operate properly and be adequately protected [3]. The system flow block diagram is shown in Figure 13 which is also the basic structure of the test bench used in the lab for the testing of ON Semiconductor's IGBTs. Most modern motor control systems use a microcontroller or DSP to process the motor signals and control the IGBT switches.

IGBT testing in a motor control application

In order to understand the efficiency of ON Semiconductor's new IGBTs, they were installed in a three phase motor drive system and waveforms and temperatures were measured and analyzed under controlled operating conditions. Dynamic measurements include the IGBT's collector current and collector−emitter voltage values captured on an oscilloscope (see Figure 14 for more details of measurement techniques). These values were measured near the peak of the ac line current half sine wave shown in Figure 15. These data were captured by scope and then processed on a personal computer.

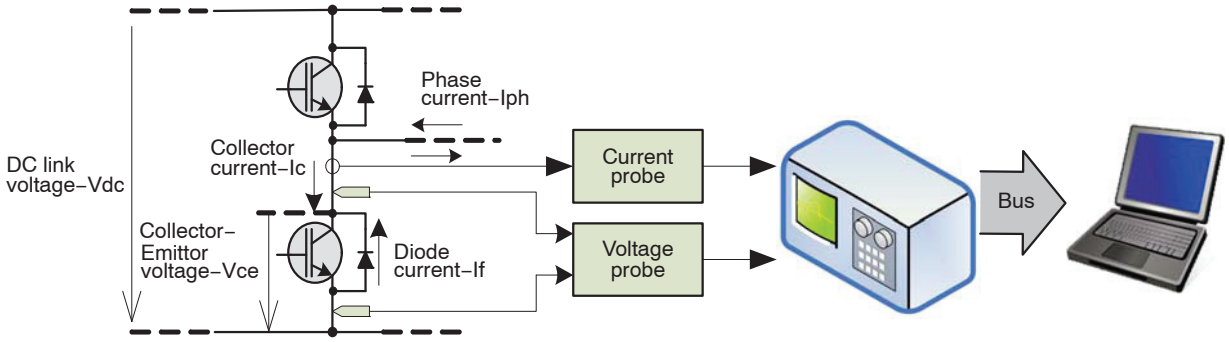


Figure 14. The Measurement Block Structure

The IGBTs are compared by loss energies computed over one switching period which means the energies represent the peak value for a single cycle switching waveform. The measurement conditions are held constant at these values: Vdc = 300 V, Ic = 14 A (these conditions correspond to approx. 3 kW output), Rg = 22 Ω, V_{GE} = 15 V, fsw = 10 kHz, Tj ≈ 100°C. The measurements were taken on the same test setup under the same test conditions for each IGBT technology.

Overall IGBT drop is difficult to measure due to the varying AC waves of the controller. However, by using the differential method proposed in Figure 14, V_{CEsat} measurements were measured. The case temperature of the IGBT was sensed by the use of an IR camera to measure device operating temperatures for comparison.

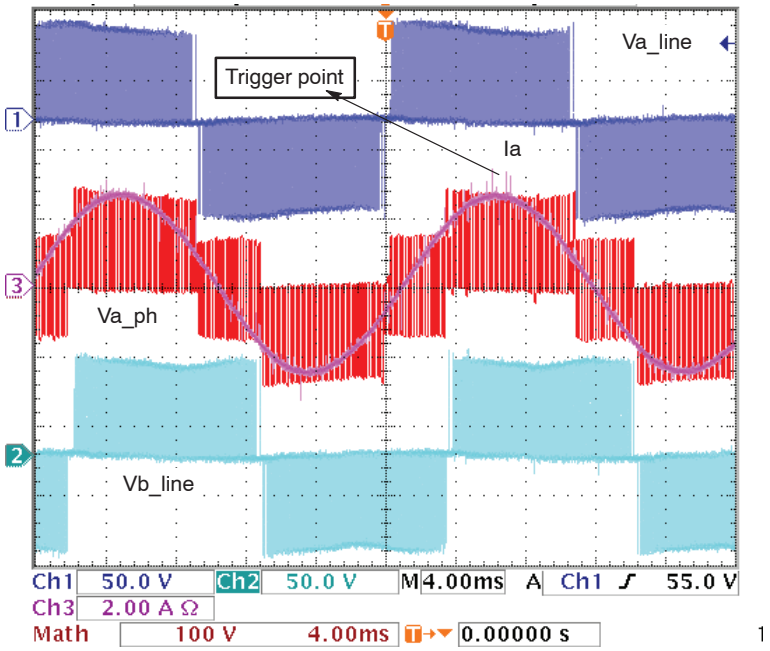


Figure 15. Output Line (blue and turquoise) and Phase Voltages (real value should be Vph/3) together with Phase Current (pink waveform).

IGBT and FWD Losses Evaluation

When using existing solid−state switching technologies, the designer must deviate from the ideal switch and choose a device that best suits the application with a minimal loss of efficiency.

One way to compare power devices such as IGBTs is power or energy loss comparison. The power losses in a power−switching device constitute of conduction losses, off−state blocking losses, turn−on switching losses, and turn−off switching losses. In practice the blocking losses are neglected.

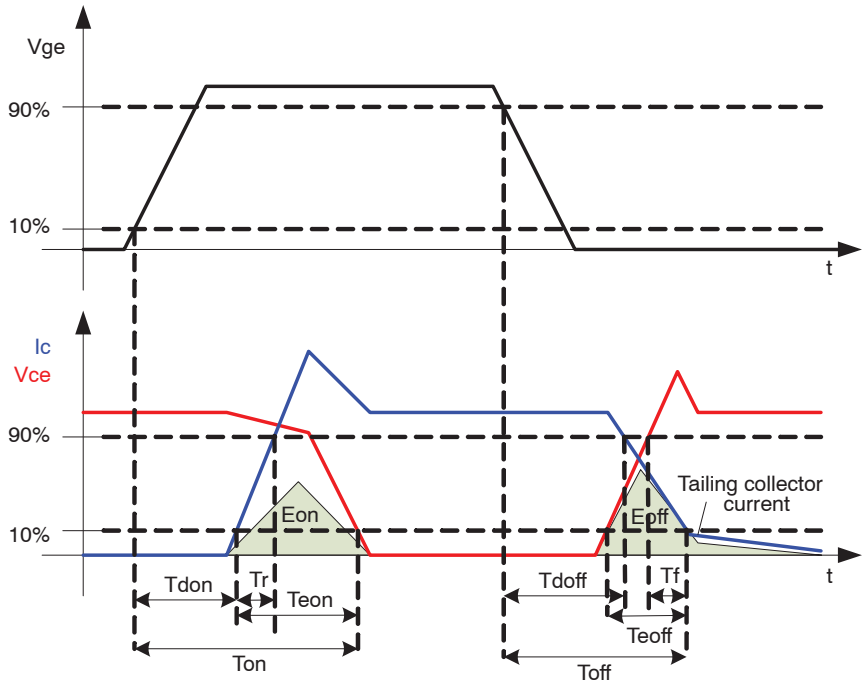


Figure 16. IGBT Current and Voltage Turn−on and Turn−off Waveform Parameters

The turn−on energy loss Eon is defined as the area under power loss waveform (P(t) = Ic(t)*Vce(t)) within the limit of 10% Ic rise to 90% Vce fall. The amount of turn−on energy loss mainly depends on the reverse recovery behavior of the diode [5]. The fundamental turn−on and turn−off parameters during turn−on and turn−off dynamics of an IGBT are depicted in Figure 16. The turn−off energy loss Eoff is defined as the area under power loss waveform (P(t) = Ic(t)*Vce(t)) within the limit of 10% Vce rise to 90% Ic fall [5].

IGBT and FWD Conduction Losses Evaluation

The conduction power loss is calculated in a straightforward manner as the product of the device current and the forward saturation voltage. The blocking loss is the product of the blocking voltage and the leakage current [1]. The conduction loss of an IGBT in VSI application doesn't depend on only the V_{CEsat} parameter but this is very good indicator of resultant conduction losses. Figure 17 shows the waveforms during the IGBT conduction time. In this picture we can see collector−emitter voltages, collector currents and the time interval where we computed the resultant V_{CEsat} voltage. Figure 17 compares resultant saturation voltages for standard planar and ON Semiconductors' trench (non−punch through) technology measured in a dynamic motor control application (this is not static measurement). The difference in saturation voltages between the two technologies is significant and can be clearly seen in this scope shot.

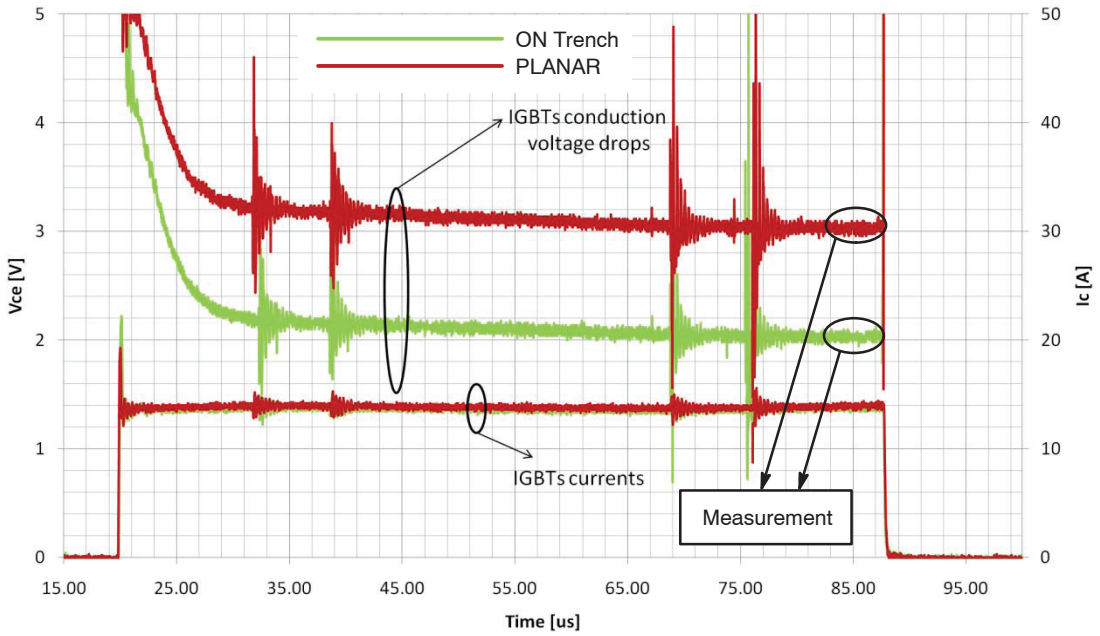


Figure 17. Conduction Dynamics Comparison of ON Semiconductor First and Second Generation Trench with Planar Technologies

This is a application measurement of V_{CEsat} at 100°C as a function of current and time for all technologies. As seen from this graph, the V_{CEsat} of the new IGBT designs is significantly reduced. The reduction in the typical value of V_{CEsat} combined with the tightening of process related parameter distribution has made it possible to spec ON Semiconductor's IGBTs more aggressively. Note that this improvement was achieved without sacrificing the fast switching speed of these motor control IGBTs.

Modern fast switching devices require fast diodes as free−wheeling diodes. With every turn−on of the switch, the free−wheeling diode is commutated from its conduction to its blocking state. These devices must be very fast but still show soft−recovery behavior and low forward voltage. Free−wheeling diodes in IGBT−converters have to cope with different requirements depending on whether they are used in a motor, a generator or a machine that can operate in either mode.

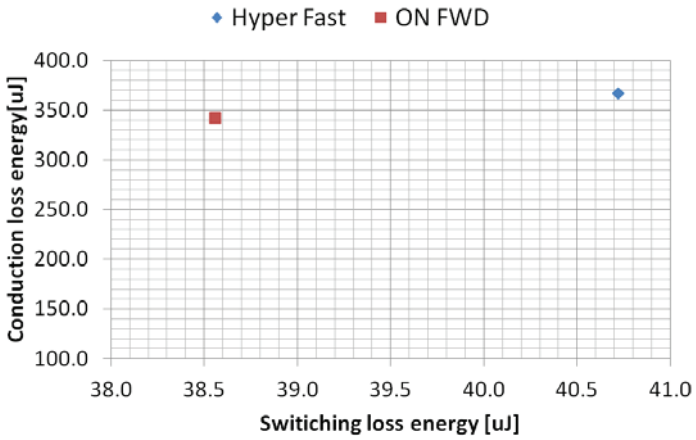


Figure 18. Trade−offs between Conduction and Switching Losses

Similarly to V_{CEsat}, the conduction losses of a free−wheeling diode (FWD) in VSI application don't depend on the V_f parameter alone but this is very good indicator of resultant conduction loss. For inverter−operation the power factor is positive (cos(θ)>0) and the power losses in the semiconductors reach their limits, if m*cos(θ) = 1, where m is modulation index. In this case maximum on−state losses and therefore, total losses in the IGBTs have been reached, whereas losses in the free−wheeling diodes are at their minimum. In the case of rectifier operation (cos(θ)<0), power losses in semiconductors reach their limits if m*cos(θ) = −1. In this case, minimum on−state losses and, therefore, total losses in the IGBTs have been reached, whereas losses in the free−wheeling diodes are at their maximum. For this reason, ON Semiconductor optimizes V_f performance together with switching performance. The performance of a free−wheel diode in a switching application (Figure 18) shows the comparison between ON Semiconductor's optimized diode with a hyper fast diode. The switching performance is important in a motor control application thus the switching losses are optimized with consideration for the voltage drop.

IGBT and FWD Switching Losses Evaluation

When passing over from the conduction to the blocking state, the internal diode storage charge has to be discharged. This results in a current flowing in the reverse direction in the diode. The waveform of this current is characterized as the reverse−recovery behavior. The typical turn−off dynamics and comparison can be seen in Figure 19, where peak current, time duration and electrical charge play important roles for optimized results. Figure 20 shows us reverse recovery of an silicon diode together with fundamental parameters during turn−off transient. The commutation slope di/dt is determined by voltage and inductance. At t_0 current crosses the zero level and at *tw* the diode starts to block. At *trrm* the reverse current reaches its maximum value. After *trrm* the current falls to leakage current. The current characteristics depend on the diode. If the current drops steeply, this is called snappy recovery behaviour. If the current drops softly, this is called soft recovery behaviour. The reverse recovery time *trr* is defined as the time between t_0 and the time, where the current reach 20% of *Irrm*. The subdivision of *trr* into *tf* and *ts* defines a quantitative value for the recovery behavior [4].

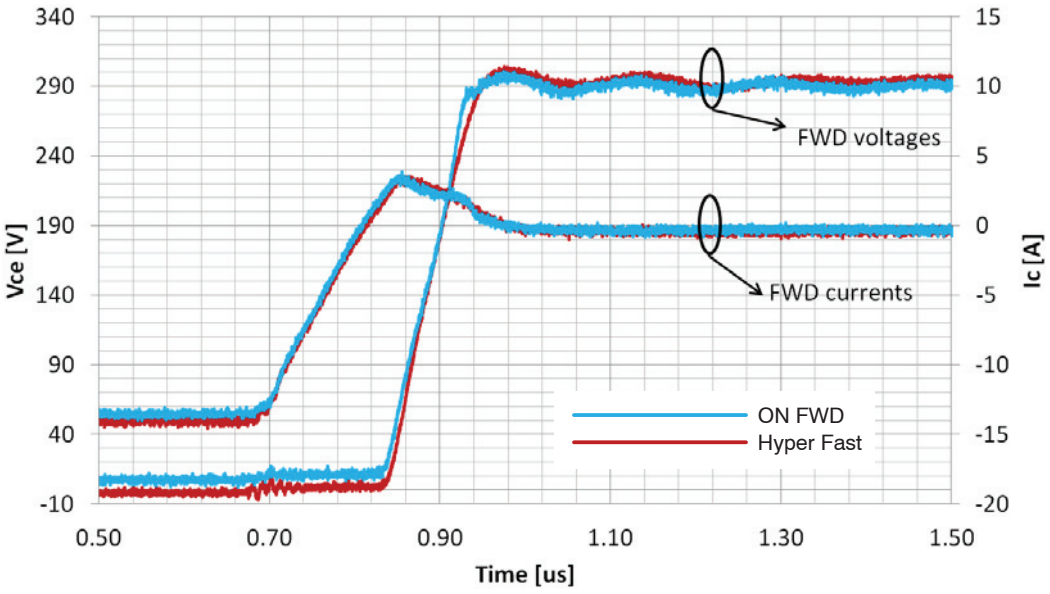


Figure 19. Reverse Recovery of FWDs

The reverse recovery behavior of a diode is the significant portion of its switching losses since its turn on losses are minimal. ON Semiconductor works to minimize the switching losses and in Figure 21 we can see the result of the optimization process.

By measuring a single switching period in a motor control application we are able to get very important information about the distribution of the power losses. The impact of the turn−on, reverse recovery and condution losses is quite apparent. The switching losses are very often the smaller part of the total power losses of a diode as can be seen in Figure 21. On the other hand turn−off phenomenon strongly influences IGBT turn−on losses.

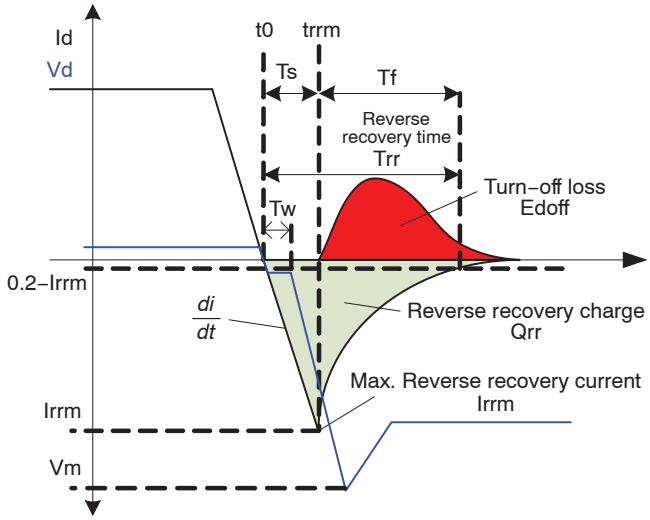


Figure 20. Current and Voltage Characteristic of the Reverse Recovery Transient of a Soft−recovery Diode and Definition of the Recovery Behaviour

Generally the reverse recovery effect is more important for its impact on the IGBT switching losses. We can define the soft recovery factor as quantitative value for the recovery behavior. This soft factor *s* can be computed by several methods, such as s = tf/ts, but the general goal is a soft waveform without a snappy recovery characteristic which means that *s* is held as high as possible. The optimization process must balance the switching speed with the soft recovery and also keep the forward drop as low as possible.

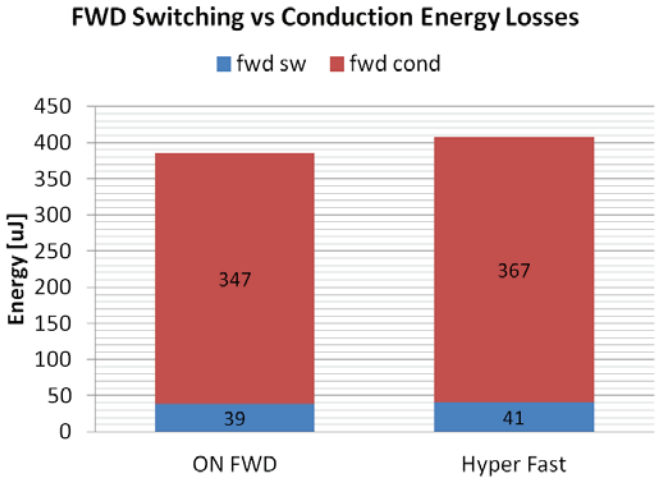


Figure 21. Switching versus Conduction Losses of FWDs

Turn−on switching losses of IGBTs are largely dependent on reverse recovery behavior of the FWD in inductive load applications. The turn−on process is complete only after the FWD has recovered and begins to block the full DC link voltage. Slower diodes, besides adding their reverse recovery current transient to the load, prolong the turn−on process. Switching times can be decreased by using a higher gate input voltage and/or lower gate resistance, thereby increasing the rate of rise of current in the IGBT and corresponding rate of fall of current in the FWD. The reduction in turn−on time is limited by the nature of the reverse recovery. Snappier recovery requires that the turn−on di/dt be reduced in order to avoid the occurrence of unsafe recovery voltage transients and generation of excessive EMI. Softer FWDs, on the other hand, allow faster IGBT turn−on. Faster and softer recovery of the FWD is therefore very important in such applications. ON Semiconductor has accomplished this by designing a series of co−packed diodes.

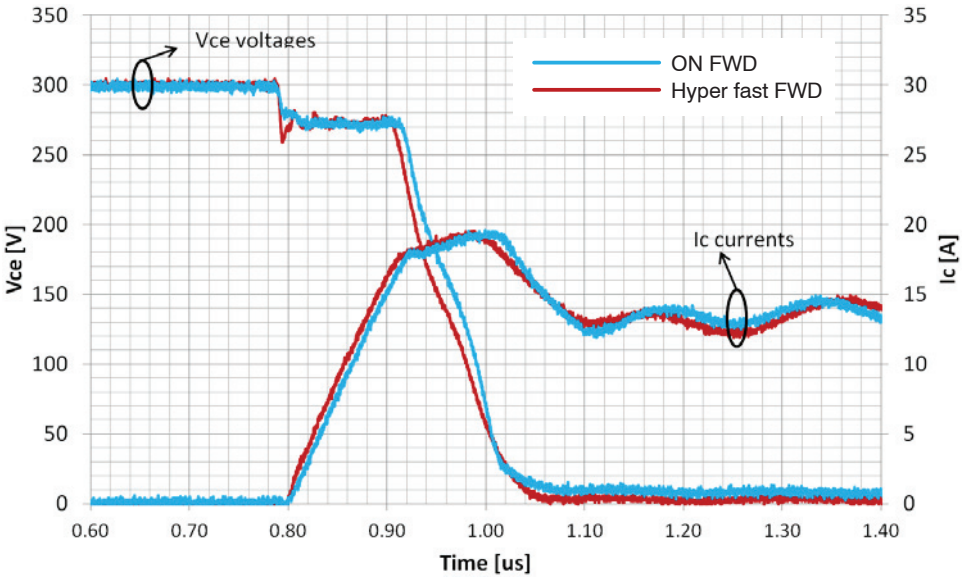


Figure 22. IGBT Turn−on

The goal of optimization is a very fast turn−on together with low reverse current peak. Figure 22 shows a comparison of diodes' affects on IGBT turn−on between ON Semiconductors' standard free−wheeling diode and a hyper−fast free−wheeling diode. These waveforms demonstrate that the hyper−fast diode responds faster and thus has a lower power loss. But we must not forget that in a motor control application the diode can operate as rectifier (in generator mode or during motor braking) and under this condition the conduction loss of diode is dominant. So, the dv/dt rate between the reverse recovery transient and its effect on the IGBT turn−on loss versus conduction loss is crucial for an efficient diode. This trade−off is apparent from Figure 18.

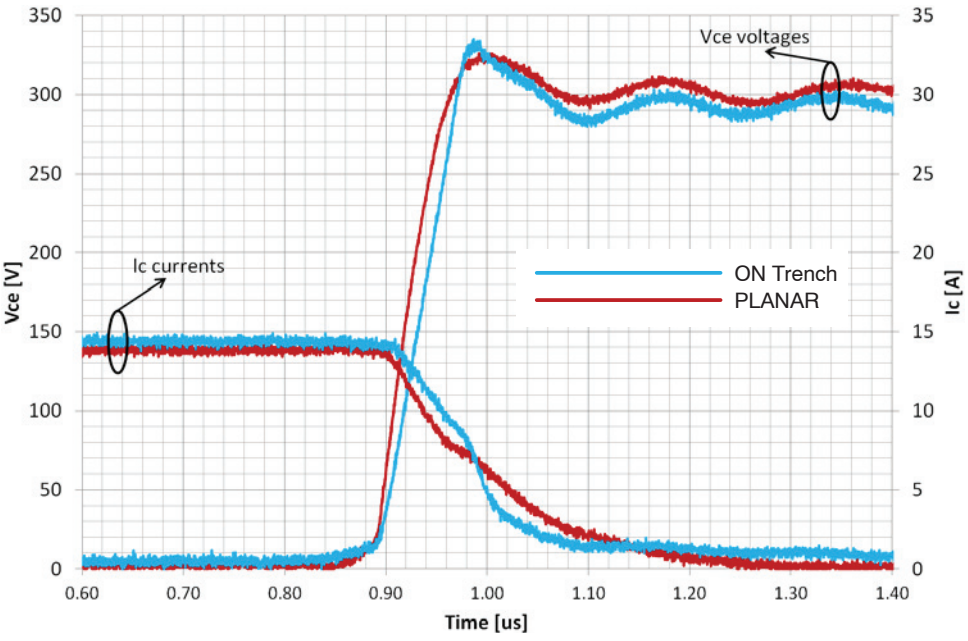


Figure 23. IGBT Turn−off

The tail current is one of the undesirable features that causes its turn−off to be slower than that of a MOSFET. While turn−on is fairly rapid, early IGBTs had current fall times in the order of microseconds. In addition, due to the long tail current the dead−time in the PWM control logic must be long enough to allow the current to go to zero. The power loss during turn−on is governed by this time interval, which is determined by the reverse recovery behavior of the power rectifier. The turn−off speed of an IGBT is limited by the lifetime of the stored charge or minority carriers in the N−drift region which is the base of the PNP transistor. The base is not accessible physically thus an external means cannot be applied to sweep out the stored charge from the N−drift region to improve the switching time. The only way the stored charge can be removed is by recombination within the IGBT. Traditional lifetime killing techniques or an N+ buffer layer to collect the minority charges at turn−off are commonly used to speed−up recombination time. Figure 23 shows comparison between ON Semiconductor's trench technology and one of the planar technology. The waveforms show that the current designs allow us to use a very short dead−time which is beneficial and suitable for higher switching speed motor control applications.

Evaluation of Total Losses

In the optimization process of the new and better IGBT or FWD it is convenient to make energy loss measurements for the IGBT and FWD separately. In addition it is desirable to divide these measurements for each area of loss within each switching time interval such as turn−on, conduction and turn−off for both the IGBT and FWD. This measurement gives us information about energy losses distribution and thus heat production inside of the package over one switching period and under a specific set of conditions.

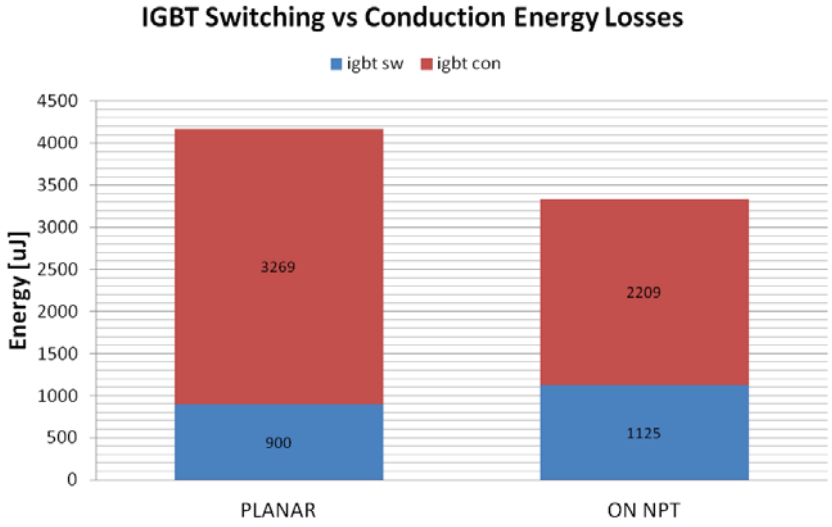


Figure 24. IGBT's Total Losses Distribution

Figure 24 shows us the total IGBT power loss division for switching and conduction losses for ON Semiconductor trench IGBTs with respective FWDs together with a convectional planar high−speed IGBTs with hyper fast diodes. The same power loss distribution but for diode is depicted in Figure 21. The measurement was done for a 10 kHz switching frequency and the conduction losses dominate; however for increasing frequencies (up to 20 kHz in motor control applications) its effect will be less and less until switching losses will dominate. If the IGBT and FWD losses are measured separately it is easy to understand the breakdown of the losses which is very useful in selecting the best device for the application. Figure 25 shows the division of the conduction and switching losses for both the IGBT and the FWD together.

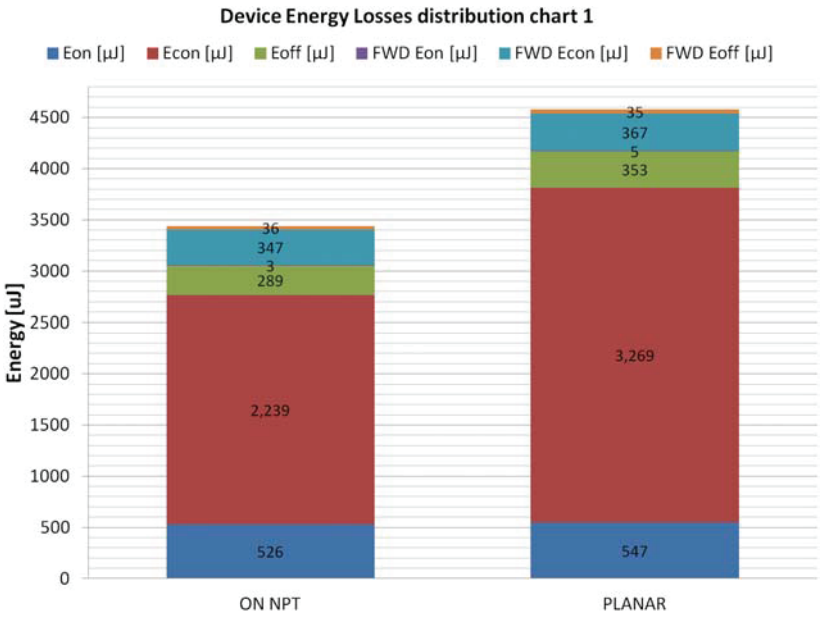


Figure 25. IGBT plus FWD Power Losses Distribution

Generaly, in power electronics, IGBTs as well as diodes operate mainly as switches. They take on various static and dynamic states within one cycle and, in addition, within one motor drive supply period. In any of these states, one power dissipation or energy dissipation component is generated, which heats the semiconductor and its environment [4]. This means that the total loss over one motor drive supply period is the sum of the losses generated over all switching periods. The electrical measurement of this total loss is very difficult, but for the comparison of the total losses we can measure the case temperature in a given system. Then in steady state conditions the thermal resistance is the only constant of proportionality abd is the same for all measurements.

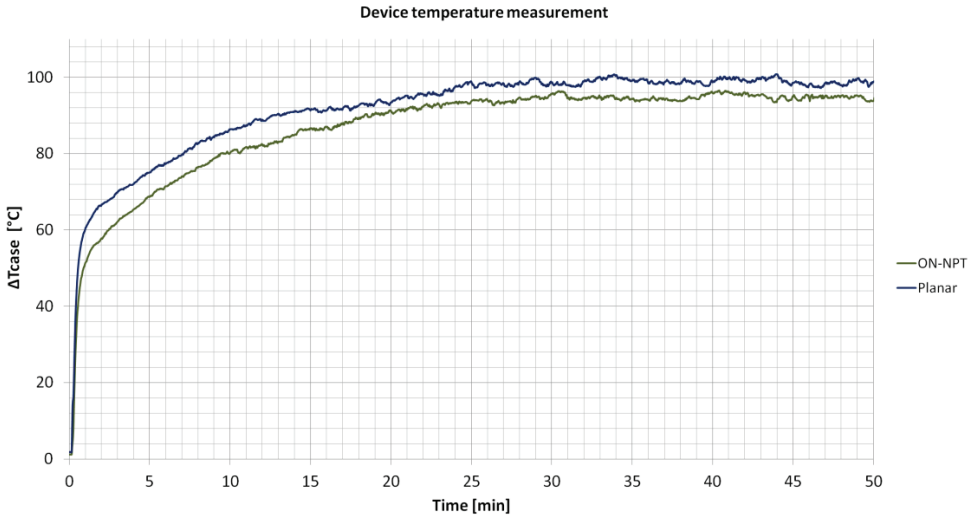


Figure 26. Steady State Temperature Measurement as Total Loss Comparison

The measurement in Figure 26 shows us the comparison of ON Semiconductor planar and trench technologies in a motor control application. This makes it easy to deduce which part generates a lower level of power loss.

REFERENCES

[1] Bose, K. Bimal. Power Elektronics and Modern Electric Drives. New Jersey: Prentice Hall, 2002. ISBN 0−13−016743−6.
[2] Rashid, H. M.. Power Electronics Handbook. Academic Press, 2001. ISBN 0125816502.
[3] Francik, O., Brandstetter, P., Rech, P., Krecek, T.: Facilities for a Digital Signals Processing of AC Drives, In: Aplied Electronics 2009, Pilsen, ISBN 978−80−7043−781−0.
[4]SEMIKRON. Application Manual [online]. c2000. Available from World Wide Web:
<http://www.semikron.com/skcompub/en/SID−1FEE9DCF−814E68FB/application_manual_2000−193.htm>
[5] IXYS, “Power Semiconductors Application Notes” IXYS Corporation, 3540 Bassett Street, Santa Clara CA 95054, and Phone: 408−982−0700.
[6] ON Semiconductor, AN1541/D “Introduction to Insulated Gate Bipolar Transistor”. 5005 East McDowell Road Phoenix, AZ 85008 USA.

AND8215/D

Semiconductor Package Thermal Characterization

Prepared by: Roger Paul Stout, PE
ON Semiconductor

ON Semiconductor®

http://onsemi.com

APPLICATION NOTE

Glossary of Symbols

$R(t)_{JA}$, $R_{\theta JA}$, $R_{\theta JMA}$, θ_{JA}, θ_{JC}	various notations indicating transient and steady state thermal resistance to ambient, case, normalized to actual power along path of interest
Ψ_{JT}, Ψ_{JL}, R_{JS}	(“junction to top”, “junction to lead”, “junction to solder”) thermal resistance parameter normalized to total package power
TSP	Temperature Sensitive Parameter
DUT	Device Under Test
TC	thermocouple

GENERAL INFORMATION

In order to measure thermal resistance of packaged semiconductors, some basic information needs to be provided. Die size, thickness and active area are used to calculate certain thermal transient characteristics of the device. Certain material properties are also necessary, specifically density, specific heat, and thermal conductivity of the primary materials in the package (encapsulant, silicon, die attach, leadframe, etc.), and from these, derived thermal transient properties of diffusivity and effusivity.

The package type is also important, surface mount or through–hole, in order to determine mounting requirements. Surface mount devices are tested on FR4 boards with minimum or 1″ pad areas. TO–220 and larger power devices are tested on a cold plate.

When measuring temperature of any power device, it is basically impossible to put a physical thermometer onto a device’s junction while under power. Instead, we must utilize some temperature “sensing” method internal to the device. For instance, in power MOSFET’s we ordinarily use the device’s inherent “body diode.” The forward–biased voltage drop of this pn junction has a very linear relationship with temperature, so, when properly calibrated, we can use it to tell us what junction temperature results from any power condition.

Thermal Parameter Test Procedure

The Temperature Sensitive Parameter (TSP)

To thermally characterize a semiconductor package, it is necessary to have a temperature sensitive parameter available (such as a diode or a resistor) within the device being tested, which can be used to measure the die surface temperature. The voltage of this TSP (in theory, at a fixed current) is measured in a calibration oven at temperatures of 25, 50, 75, 100, and 125°C. The current used is very low (typically 1.0 mA) to prevent significant self heating of the device. (In practice, a constant current supply is approximated by using a constant voltage supply and a large resistor in series with the TSP, as shown in Figure 1).

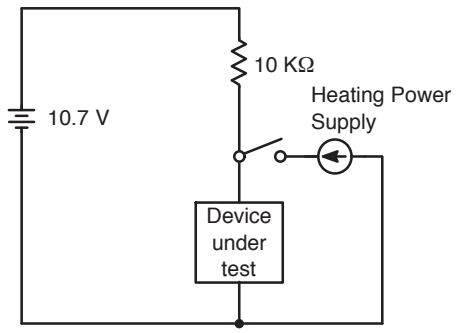


Figure 1. Basic Thermal Test Circuit

Calibration

Before calibration, a surface mount device is assembled onto a thermal test board such as the min−pad board illustrated in Figure 2, or the 1″ pad board as in Figure 3. Tab and “back of board” thermocouples are typically attached for external package temperature measurements.

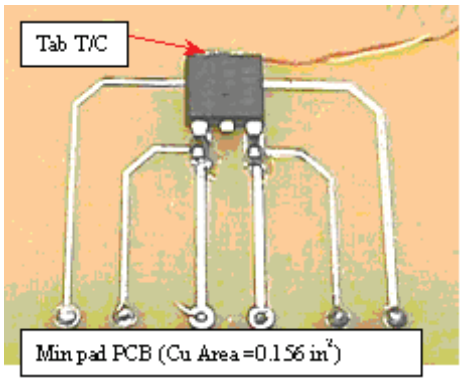


Figure 2. D²PAK on Min Pad Board

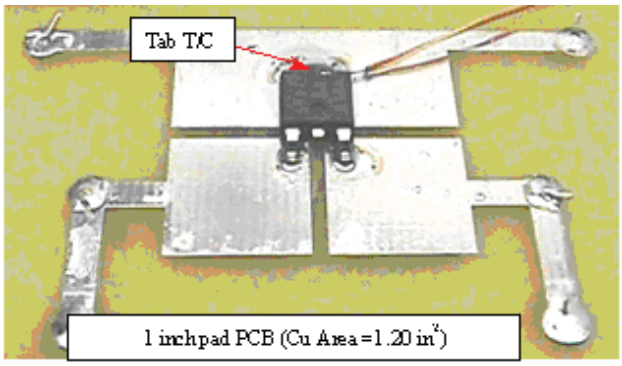


Figure 3. D²PAK on 1″ Pad Board

Devices to be characterized in a non−surface mount condition (such as socketed TO−220’s) are calibrated directly in a socket such as will be used in subsequent testing.

Thermal Test Configurations and Fixturing

After calibration, a surface mount device is put into a one cubic foot still air test chamber with the board in a horizontal position, as shown in Figure 4. An ambient thermocouple is positioned within the chamber approximately 1″ below the test board and off to the side. Generally, still air characterization is limited to about 2.0 W maximum power dissipation (and may be much lower).

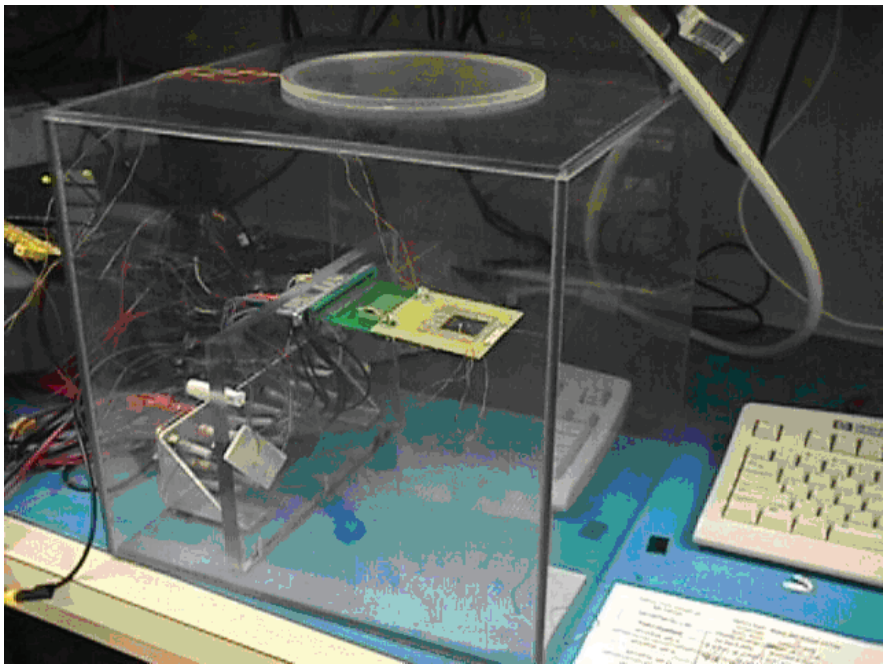


Figure 4. One Cubic Foot Still Air Chamber

For power devices requiring large amounts of power for characterization (for instance, heat−sinked TO−220’s), a coldplate (Figures 5, 6 and 7) is used with the device mechanically or hydraulically clamped to the surface. A thin layer of commercial “thermal grease” is applied at the interface between the device and the coldplate to minimize thermal contact resistance. Again, external package temperature measurements are generally made with a K−type thermocouple glued or soldered to the exposed heatsink tab of the package. Additional thermocouples are embedded within the coldplate at various locations to monitor the coldplate conditions.

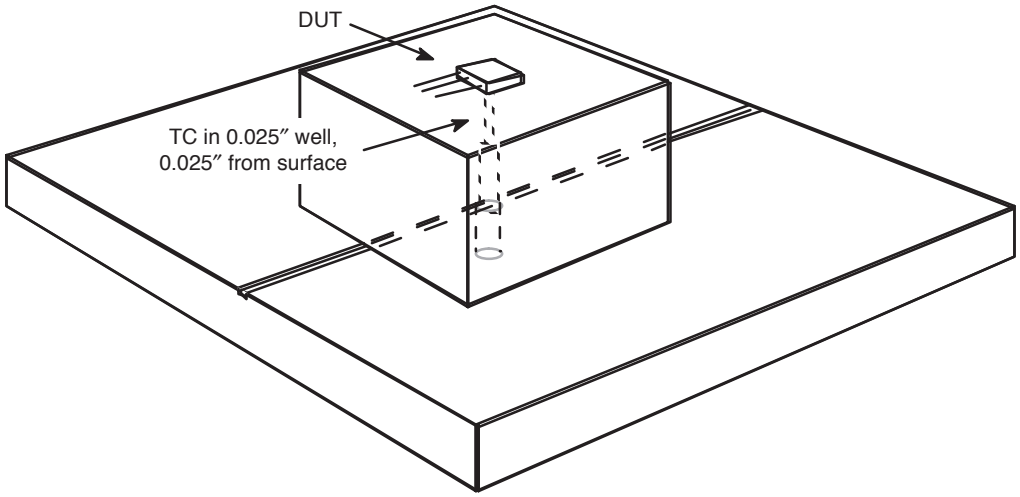


Figure 5. Oblique Schematic of Coldplate

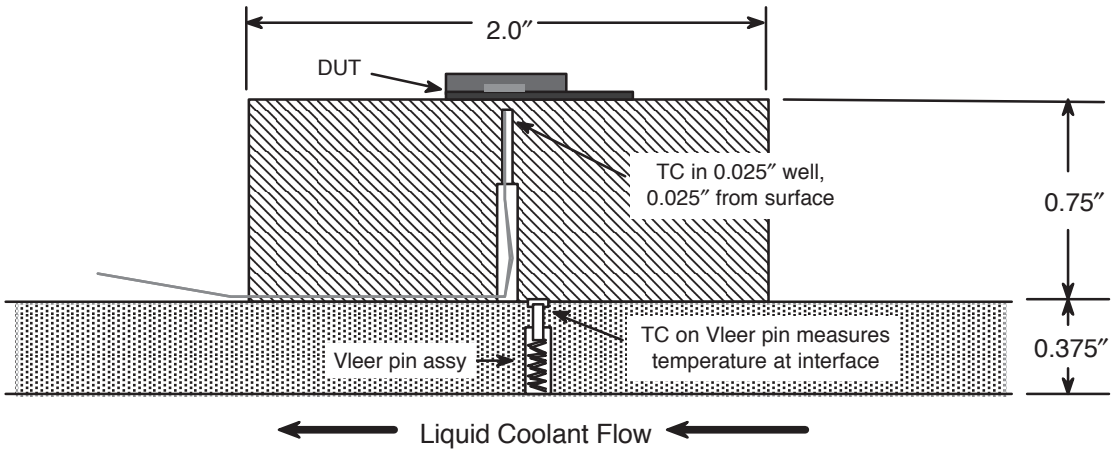


Figure 6. Cutaway Section View of Coldplate

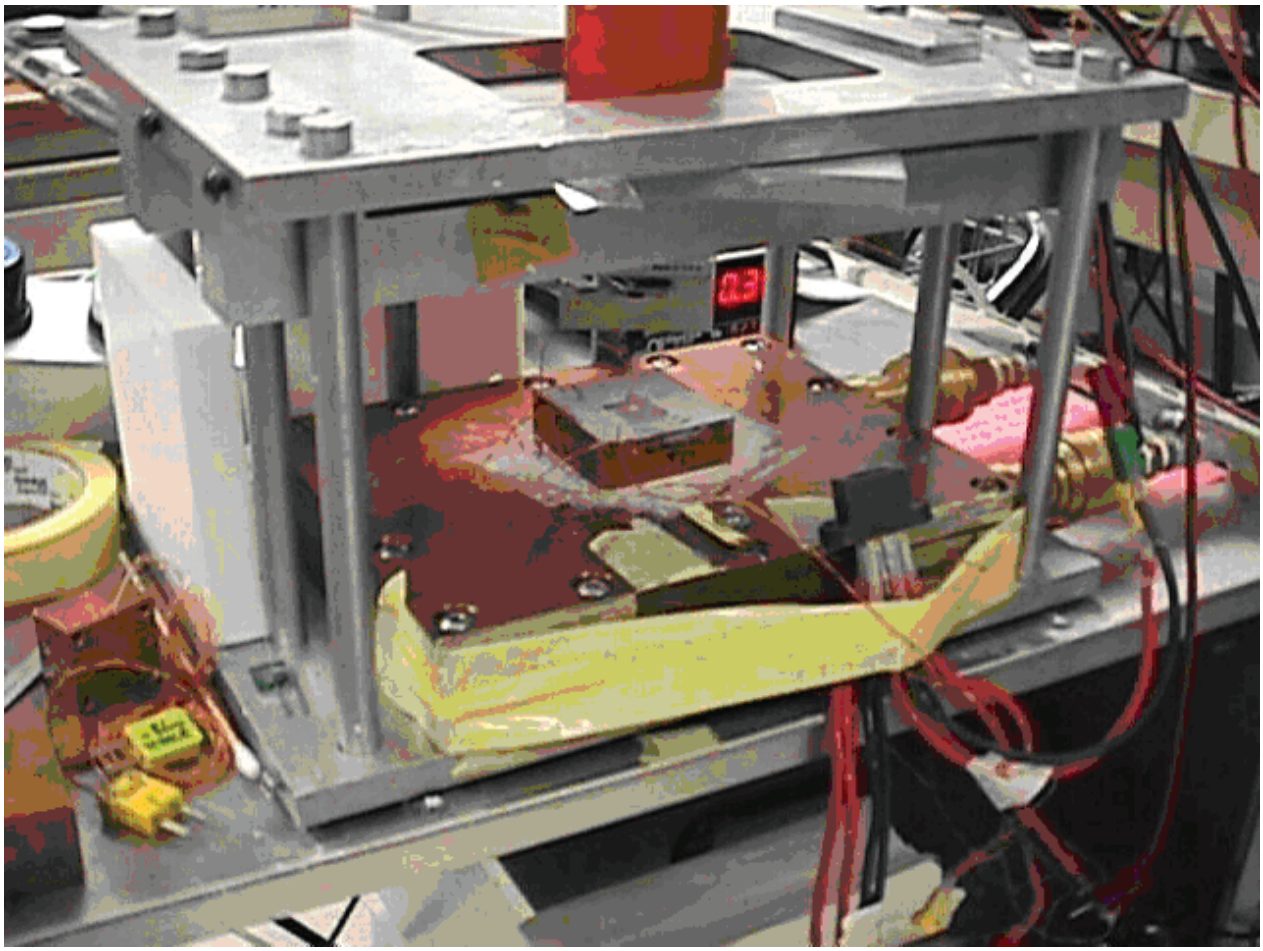


Figure 7. Photograph of Coldplate with Device on 1″ Pad Test Board

For coldplate tests, the semiconductor package is clamped directly onto a copper block 2″ square by 0.75″ thick (Figure 8). Instead of a thermocouple being in direct contact with the back side of the device under test (DUT), the “case” TC is mounted in a narrow well drilled up the center from directly underneath the DUT (see Figure 6). Over the final quarter−inch of depth, the TC well is only about 0.025″ diameter, and comes within about 0.025″ of the surface without breaking through. In other words, this copper block is a close approximation to a pristine semi−infinite heatsink design, though to get there we have sacrificed our ability to measure the case temperature by direct contact. Instead, we must rely on the embedded TC to get as “close as possible” to the case temperature, assuming negligible temperature gradient through the thermal grease and intervening 0.025″ copper. This (almost) solid copper block is in turn clamped against the upper surface of the cold plate (see Figure 6). The Vleer−pin now presses the thermocouple against the underside of the copper block. The TC embedded within the copper block is an approximation of the “case” temperature, while the Vleer−pin TC serves as a “distant” temperature reference. Indeed, measured values of “thermal resistance” for these reference TC’s were only 0.23°C/W, “case” TC to coolant (with standard deviation 0.02°C/W), and 0.03°C/W, Vleer to coolant (with standard deviation 0.01°C/W).

Most high power devices tested on the coldplate are also tested (at much lower power) in natural convection conditions where the device is not specifically heatsinked at all, except by virtue of the high−power test socket in which it is held, typically as shown in Figure 9.

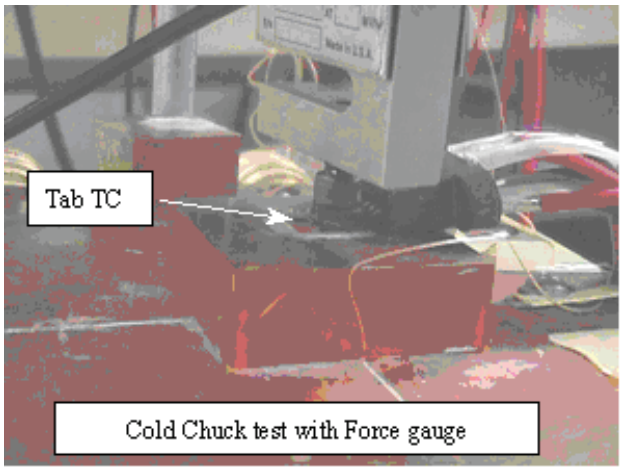


Figure 8. TO−220 Clamped to Coldplate

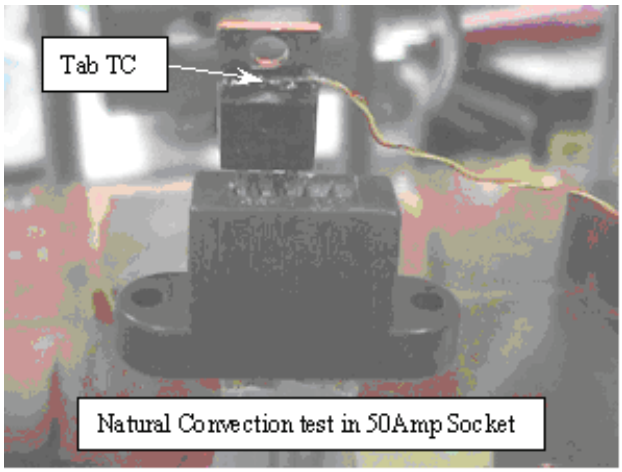


Figure 9. TO−220 in Socket for Free Convection Test

Thermal Test Procedure − Basic Test Method

The basic test method is to heat the device with constant power input, and allow it to reach thermal equilibrium. Heating is accomplished preferably by using the primary heat dissipating structure as occurs in normal device operation. (Likewise, preferably, the same structure serves the purpose of the TSP, so that temperature information is available wherever the heating happened.) This may be the collector−emitter channel in a bipolar device, the source−drain channel in a MOS device (often reverse biased to utilize the so−called "body diode"), the anode−cathode channel in a thyristor, an ESD diode on an output driver of a logic device, etc.

Once heated steady state is reached, "hot" values are recorded for all thermocouples, power is switched off, and the device junction temperature is recorded as the entire system relaxes back to ambient. Power switching is synchronized between the data−logging computer (which controls the test) and a custom−designed switching box. The switching box switches from fully on to fully off in about 10 μs. The first reliable temperature readings (that is, stable voltage readings not obviously contaminated with electrical switching noise) on most junctions are generally not seen until about 250 μsec after the switching box switch is triggered; however, this is somewhat device dependent. These voltage readings are converted to temperatures using the calibration data previously recorded. (As suggested in Figure 1, in many cases the "measurement" circuit is never actually disconnected during heating phase; but in any case, during the measurement phase, the same circuit as was used during calibration is once again used. There are also rare situations when a TSP completely electrically isolated from the heating circuit is available, in which case temperature measurements may indeed be made simultaneous to heating. Obviously, however, this precludes the possibility of measuring the temperature exactly *at* the heat source.)

Transient Results

General Background

The temperature of the junction at the moment of switching off the heating supply is the steady state junction temperature. It is an unfortunate reality that this temperature can never actually be measured (that is, immediately and instantaneously), due to the electrical transients which must settle out before a trustworthy temperature signal can be detected. Much of the art in thermal characterization is therefore the method used in deducing how much the temperature may have dropped during this initial unmeasurable period. In any case, temperature data is collected from the earliest possible moment, as the device cools down to a final steady−state (unpowered) at ambient. The cooling data can be converted to a transient heating curve and an RC network can be determined.

Transient results require a considerable amount of post−processing in order to get the most useful information. First of all, there is the normal amount of experimental variability encountered in any laboratory setting: intrinsic instrument resolution, etc. In semiconductor package measurements, there are also experimental variations attributable to specific aspects of the typical device and environment. For instance, short−time scatter due to die attach irregularities; intermediate time scatter due to board−level soldering, gap between package and board, etc.; long−time scatter due to test board variation (trace thickness) and random convection noise.

Second, as previously stated, the very nature of the measurements precludes ever knowing a "true" initial temperature at the heat source itself. Since junction temperatures in general cannot be measured (at low current) simultaneous to heating the device (at high current), the switching time required to change current opens the door for an unknown amount of cooling to occur before the first reliable temperature measurement can be made. (There is an alternative "heating curve" test method that at least permits the experimentalist to know the starting ambient temperature to any desired degree of precision. An

equivalent problem exists, however, because still the junction cools by an unknown amount, each and every time the heating pulse is discontinued for the eventually required measurement.)

Minimizing Scatter

In the “cooling curve” method, the starting steady−state temperature is fundamentally a function of the entire system thermal resistance (i.e. test board variations, thermal grease variations, clamping pressure variations, etc.), let alone individual package manufacturing variations. So even though each of five samples of a certain device might be expected to have very similar transient performance for the first several milliseconds, in practice all their transient cooling curves will begin at different temperatures. (In contrast, the “heating curve” method, by definition, guarantees a common, well defined ambient starting condition which is not at all a function of the test fixturing.) Nevertheless, for the sake of analysis, we’d like to artificially bring each set of these cooling curves together to begin at a common value, and allow the intrinsic system influences to add up to differences in the data towards the “cold,” or steady state, end of the curve. Because our primary interest is in the thermal performance of the package itself, the best approach is to shift the curves (of a single data set representing a single configuration) relative to each other so as to minimize the scatter between them over some designated time range. Typically the range selected is from the earliest possible instant (end of switching noise) out through the end of “die level” influences (based on theoretical expectations arising from die properties). Figures 10 and 11 illustrate this process.

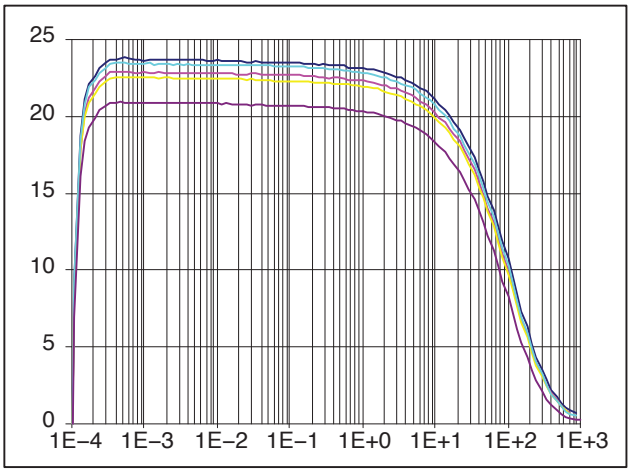


Figure 10. Raw Cooling Curves

For a more detailed explanation, see AND8216/D.

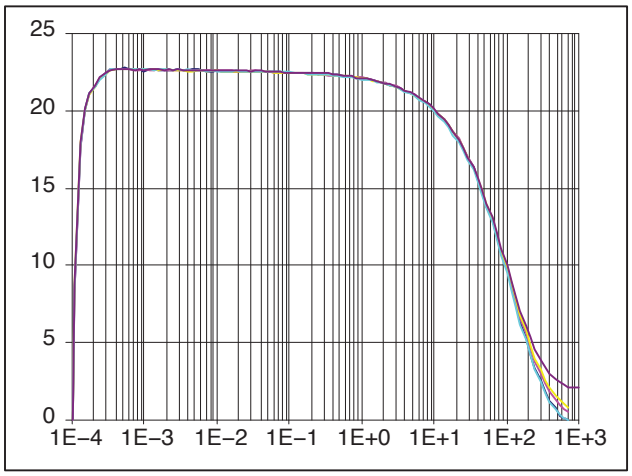


Figure 11. Scatter Minimized from 0.5−2.0 msec

Unmeasurable Short−time Characteristics

To handle the initial “unmeasurable” period in the transient cooling curve, MIL standards suggest that a sqrt(time) assumption be used to back−extrapolate from the first reliable measurements to time zero. (This arises from the closed−form mathematical solution to a uniform heat flux applied to a semi−infinite domain.) We find that if we use a linear regression technique to examine the form of our data between, say, 250 μsec and 1.0 msec, the “correlation coefficient” is only so−so for an assumed sqrt(t) function (say 0.94, where 1.00 is a perfect fit). In fact, over this timescale, the correlation is no worse for our real data if we assume a perfect exponential behavior rather than a sqrt(t), or even a straight−line linear extrapolation. On the one hand, this may be simply because our measurement resolution is not good enough to discriminate the form with this degree of accuracy. On the other hand, it could be because the actual form is neither sqrt(t), nor exponential, nor linear. Consider that these MIL standards were written several decades ago when silicon die thicknesses were undoubtedly much larger, and the sqrt(t) approximation was good out to several milliseconds (assuming uniform surface heating). A 10−mil thick die, however, can only be assumed to exhibit sqrt(t) behavior for about 400 μsec – departing from this idealized behavior at about the same timescale at which the first reliable temperatures can be measured. Compounding matters is that the heat may not be fairly approximated as a surface flux (especially as the die gets thinner) – in fact certain silicon technologies minimize their electrical resistance properties by dissipating large fractions of their thermal energy through the entire thickness of the die. Last, but certainly not least, is the question of whether the “temperature sensitive parameter” being calibrated in the first place is actually at the surface of the die at all, or whether it covers so large an area (or even volume of material) that it is measuring some sort of average temperature, rather than the idealized “peak” junction temperature! So for all these reasons, it is really not clear what should be done to extrapolate back to time zero. Ordinarily we go ahead and use the sqrt(t) fit, recognizing its

limitations, but also that it tends to be conservative with respect to power distributions which are not actually surface heating. One can calculate the sqrt(t) constant based on a simple, closed form solution to one−dimensional surface heating of a semi−infinite solid:

$$\theta(t) = \frac{2}{A\sqrt{\pi k \rho c_p}}\sqrt{t} = b\sqrt{t} \qquad \text{(eq. 1)}$$

where

$$b = \frac{2}{\sqrt{\pi}\, A \eta} \qquad \text{(eq. 2)}$$

the quantity $\eta = \sqrt{k\rho c_p}$ is known as the thermal effusivity. (eq. 3)

A = active area, k = thermal conductivity, ρ = density, c_p = specific heat.

This sqrt(t) behavior is followed, within a couple of percent, for about 40% of the “characteristic time” for the continuous medium in question (in this case, through the thickness, L, of the silicon die). The characteristic time is given by:

$$\tau = \frac{L^2}{\alpha} \qquad \text{(eq. 4)}$$

where the quantity $\alpha = \frac{k}{\rho c_p}$ is known as the thermal diffusivity. (eq. 5)

As one may surmise from the preceding discussion, the decision making process for handling the “unmeasurable” time period of an experimentally generated thermal transient response curve, is somewhat subjective. It is worthwhile, therefore, to present some additional guidelines and rules of thumb which are considered during this process.

1. Calculate the characteristic time[1] (Equation 4) of a hypothetical piece of silicon of the thickness of the die. For the temperature ranges of interest to us, our standard value for the thermal diffusivity of silicon is 0.0122 msec/mil^2 (in units convenient for the sake of this discussion). A 10−mil thick die thus has a 1.22 ms characteristic time.
2. Calculate the sqrt(t) proportionality (Equation 2) for the full die area, and also for the active die area. Using our standard silicon properties, the thermal effusivity (Equation 3) is 0.0138 W√s /mm^2/°C. Thus, for example, using a full die area of 3.0 x 3.0 mms, you obtain a value for b (Equation 1 and 2) of 9.1°C/W/√s. If the active area is only 2.0 x 2.0 mils, b would be 20.6°C/W/√s.
3. Using the 40% rule of thumb, identify the latest time at which the sqrt(t) proportionality might be expected to be valid. For a 10−mil thick die, that would be 0.5 ms. Compare the die−thickness time scales with the time at which the experimental data is believed to be good, to see whether the actual die properties should have anything to do with it or not. (Obviously “good” is somewhat circular reasoning, but for instance we keep clear of obvious electrical noise in the data, like when it appears that the junction is getting hotter after we kill the power, rather than getting cooler! Clearly the data can’t be correct until it at least looks like it’s getting cooler, and probably not even then.)

(a) If the data looks clean all the way back into the range of 40% of tau, then we hopefully can reconcile the actual slope of the curve with a theoretical sqrt(t) curve which is based on active die area (if it’s in the 40% range, die thickness hasn’t come into play yet). For 15−mil thick die this is not unusual.

(b) On the other hand, if it looks like the data is clean back into somewhere between 40% and 100% of tau, we want to suppose that actual die−level thermal characteristics are driving the data, but it isn’t so clear whether the theoretical sqrt(t) curve means as much. In this case, we’ll play a visual curve matching game trying to convince ourselves that we can make a smooth blending of the sqrt(t) curve into the real data, etc., but see item (4).

(c) Worst case is (like for particularly thin die) when our data doesn’t look clean even as early as 100% of tau, or maybe it looks clean but doesn’t very much look like the theoretical sqrt(t) at all.

[1]It should be noted in passing that this is not a “time constant” in the strict mathematical sense of an exponential function decay time (because it isn’t), but it does relate to how long it takes the heat to pass through a slab of material of the indicated thickness, under the following conditions: surface heating, one−dimensional heat flow through the slab, thermal ground on the “destination” side of the slab.

4. Active area and other factors affecting the 40% rule – the theoretical sqrt(t) curve is based on the assumption of uniform surface heating, and 1–D heat flow. In the real world, obviously that never happens. However, silicon having a much higher thermal conductivity than the mold compound which typically surrounds it on all but the back side, the 1–D approximation is reasonable, at least if 100% of the surface area of the die was heated. But that's rarely the case. Among other things, you've got wirebond pads that may have active silicon underneath ("bond over active" technology), so the heat loss *upwards* is not uniformly small; second, wirebonds may be over *inactive* areas, and there's usually some dead zone around the perimeter to allow for minor variations in saw–cutting the die apart, etc. One can also look at how far apart active zones are (sometimes there are multiple heat sources, especially in large–scale analog devices), and how far away these active regions are from the edges of the die. Finally, one must also look at the width of the heated region as compared to the die thickness. Following a 45–degree spreading angle (for the lack of any better rule), see if the heat gets to the full die dimensions before it gets through the die, etc. The competing effects are that if the heated area is large compared to the thickness, one expects the sqrt(t) proportionality based on active area to resemble that of the actual curve; but if the distance to the edges is small, one expects to reach something more like a sqrt(t) based on the total area by a "tau". On the other hand, if the percent area is small, and if the width of the heated area is small compared to the die thickness, the situation begins to more closely resemble a point source of heat, and the sqrt(t) behavior will never occur in the first place. (The same thing goes for volumetrically distributed heat sources, which never follow a sqrt(t) either.) For small die with relatively large pads and dead zones, it's very subjective. Even so, the short–time problem can be bounded with a sqrt(t) based on active area as the upper bound, and a sqrt(t) based on 100% area as a lower bound. And the earlier in time the data is "clean," the closer together come these two bounds.
5. One must consider the presence of thermally significant materials on the top side of the die. Indeed, we've already glossed over the typical situation, where mold compound encapsulates the die. Though a moderately poor thermal conductor, the effusivity of a typical compound reduces the effective sqrt(t) constant of the surface heating model by about 10%. So the numbers quoted earlier are, strictly speaking, applicable to a "bare" die situation more so than to an encapsulated die. In bond–over–active technology, the wire itself (particularly the ball bond resting on the bond pad) may need to be considered, especially when the die is very small. Likewise, many power devices have a "clip" of some sort (usually copper) in lieu of wires, soldered directly over the active area of the silicon. These materials change the theoretical sqrt(t) proportionality; however, the same % area considerations (meaning, now, the % coverage of the other material) described above still come into play. If the clip is thinner than the die, however, it affects the initial sqrt(t), yet its effect dies out within its own "tau" (which is based on its thickness and alpha), probably long before the data ever became "good." So there still may be a brief period where the die–only sqrt(t) shows up.
6. Finally, in special cases, the heat source may be better modeled as a volumetric heat distribution rather than surface heating, in which case the initial transient response is proportional to linear time rather than sqrt (t).

Finite Element Models (FEM)

If any of these particularly difficult experimental situations arises, whether due to complex geometries, multiple material combinations, or simply when the experimentally unmeasurable period extends well past the expected sqrt(t) behavior based on silicon thickness, we typically supplement the measured data with a finite–element model (FEM) at the die level. Such a model generally includes the die attach material and leadframe immediately adjacent to the silicon. For computational reasons, mold compound above the die is generally omitted from such model, but more thermally conductive materials will be included (especially "clips"), if present.

Last but not least, a die–level FEM is also very useful for adjusting/interpreting short–time data when the TSP is known to be separate from the heated region, or when the practical exigencies of measurement preclude heating of the desired region at all. (In this latter case, you heat what you can in order to measure the overall package characteristics, and rely on the model to fill in what you want to know at the die level.) The following figure illustrates the situation of a very small active area as compared to overall die dimensions, and asymmetrically located on the die surface. No simple short–time model will match the peak temperature transient profile in such a device, so the FEM gives us the best means of predicting the initial transient heating with which to augment the measurements.

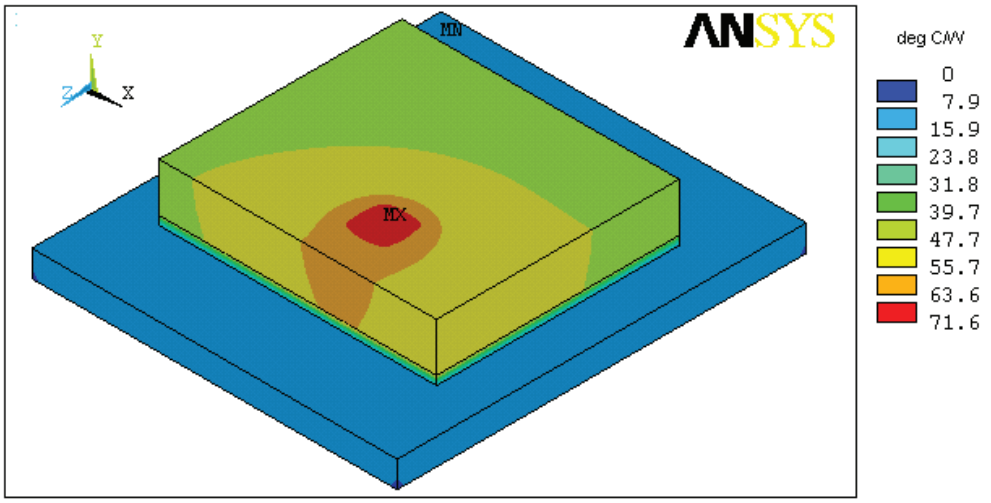


Figure 12. Finite Element Model of Quad−Op−Amp Die on Leadframe, One Output Heated

Steady State Values

Once the initial “unmeasurable” temperature change at the die level has been determined, the overall steady state junction−to−ambient (or junction−to−coldplate) values come from adding the deduced short−time temperature change to the averaged net value from the cooling curves. Additional steady state values for junction−to−board, junction−to−tab, etc. are also calculated from available thermocouple measurements. Clearly, as the “thermal distance” between the junction and the thermocouple in question increases, the overall variability in the steady state result will increase as it includes progressively more of the sources of experimental variability (die attach, board mounting, board/convection variations). In many cases, certain points in the cooling curve may be identified with physical location in the test configuration. For instance, a pronounced “bend” in the cooling curve at about five seconds in a 1″ pad test (Figure 13) corresponds to the heat “arrival” at the edge of the 1″ pad. A bend in the coldplate curve at about one second (Figure 14) corresponds to heat “arrival” at the perimeter of the 2″ copper block. Knowing these characteristics of the test fixturing helps correlate steady state results with the transient results.

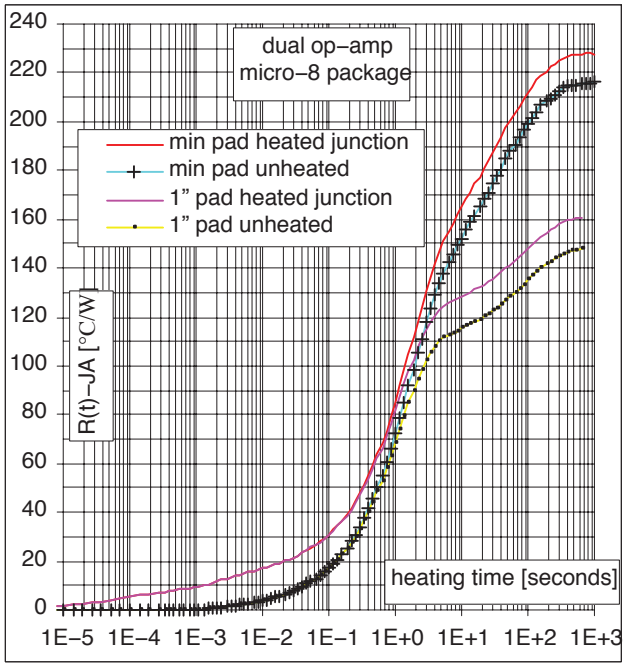


Figure 13. Min Pad vs. 1″ Pad Heating Curves

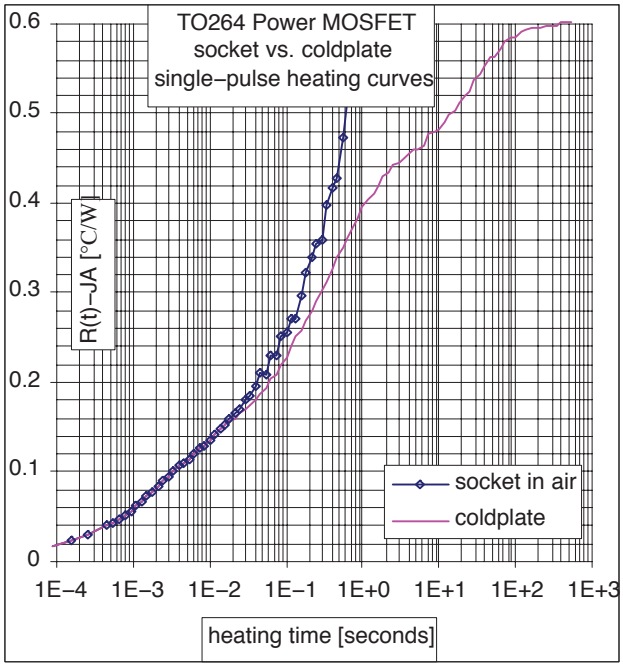


Figure 14. Socket vs. Coldplate Heating Curves

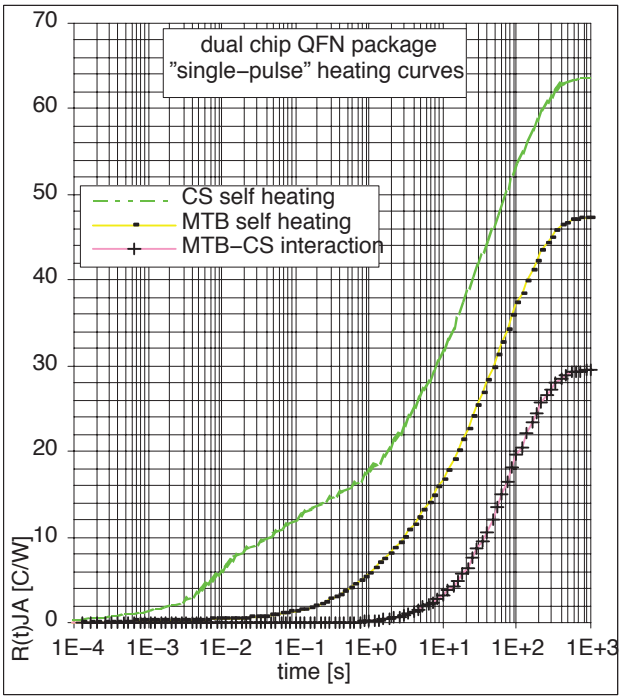


Figure 15. Heating Curves for Two−Chip Device

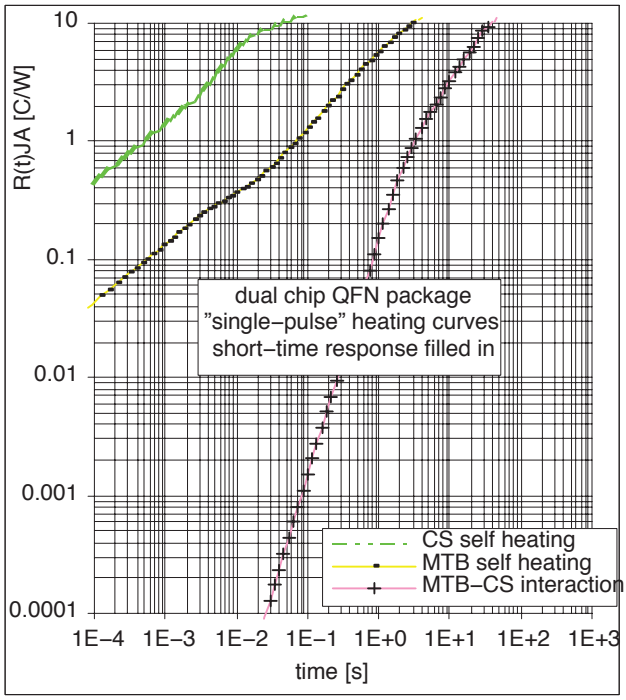


Figure 16. Short−Time Response Filled In

Thermal RC Network Analysis

The complete transient curve for a device (model or theory blended with measurements) is often used for a thermal RC network analysis. Proprietary software provides a best fit model to the data (usually within a fraction of a percent of mean error between RC network and input curve). Any network topology may be selected, but usually a simple "ladder" network of resistors with capacitors tied to ground is chosen (such as in Figure 17). Multiple−junction devices are typically modeled with branched ladder networks, and the input transient curves (examples in Figures 13, 15) must include "self heating" and "interaction heating" for all the junctions represented in the model. Figure 18 depicts a four−input RC thermal network, where only the nodes and connecting resistors are indicated, the (grounded) capacitors being implied.

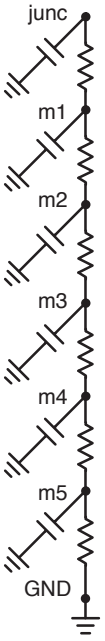


Figure 17. Simple RC "Ladder"

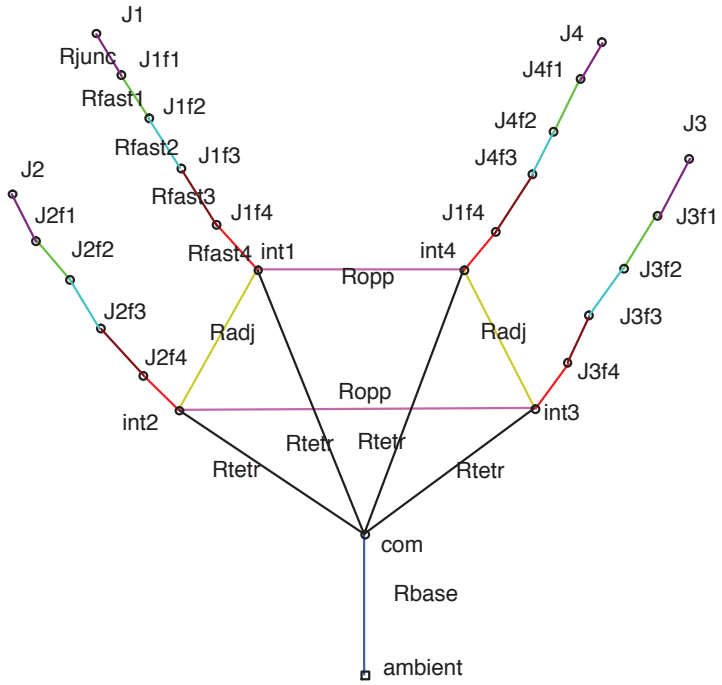


Figure 18. Four Input Thermal RC Network, Capacitors Implied

For the specific non−symmetric dual−chip device whose heating curves were shown in Figures 15 and 16, we illustrate in Figure 19 the RC network used to fit the data; here the capacitors are implied. Figure 20 shows the same network again, only with the capacitors explicitly included. Note that each variable−temperature network node connects to ground through a capacitor.

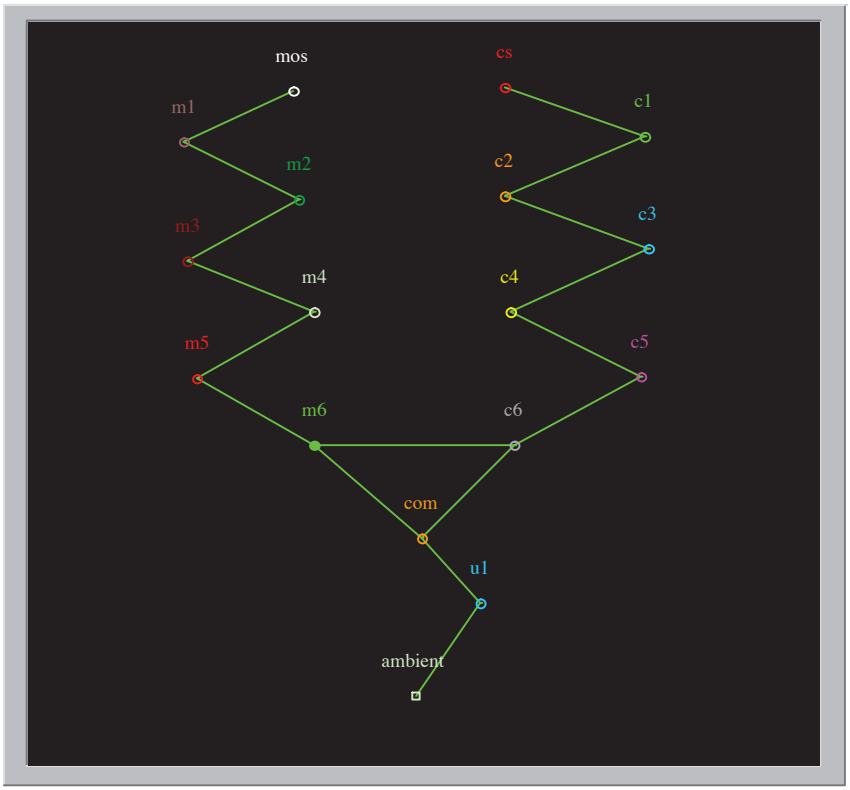


Figure 19. Two Input Thermal RC Network, Grounded Capacitors Implied

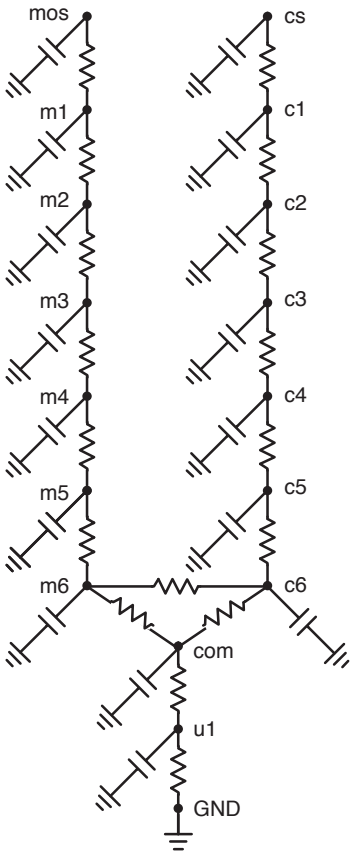


Figure 20. Explicit Capacitor Version

Grounded vs. Non−Grounded Capacitor Networks

It is important to understand that Thermal RC models which have any presumed physical significance at all, that is, where nodes in the model bear some correlation with physical locations in the thermal system, must utilize grounded capacitors. Even so, the thermal behavior of the heat−input junction of the network can be described by a mathematical equation (or set of equations) which resemble an RC network whose capacitors are not tied to ground, but are connected between nodes in parallel with the resistive links. In this purely "mathematical" version of the network, amplitudes of the individual circuit "elements" correspond to resistances, and the time constants of the mathematical terms may be interpreted as RC products; thus the capacitances can be "deduced" from the time constants, though neither they, nor the resistors, have any particular correlation with physical locations in the actual thermal system.

To try to clarify this point (which is *not* what Figures 19 and 20 were illustrating), the following table (Table 1) provides the R and C values corresponding to the data in Figures 15 and 16 and the network shown in Figures 19 and 20. Table 2 provides the mathematical time constants and amplitudes corresponding to the set of equations which describe the thermal performance of the two junction nodes, both their isolated "self heating" and their "interaction heating." Note that the R's in Table 1 and the amplitudes of Table 2 bear little resemblance to each other.

Table 1. A SPICE Compatible Thermal RC Model

Element Name	Node 1	Node 2	Element Value
C_C1	gnd	cs	9.21909E−6
C_C2	gnd	c1	4.36252E−5
C_C3	gnd	c2	1.30876E−4
C_C4	gnd	c3	1.75727E−4
C_C5	gnd	c4	8.87879E−4
C_C6	gnd	c5	2.06324E−2
C_C7	gnd	c6	3.25284E−1
C_C8	gnd	mos	9.59163E−5
C_C9	gnd	m1	4.53881E−4
C_C10	gnd	m2	1.36164E−3
C_C11	gnd	m3	4.08493E−3
C_C12	gnd	m4	1.66362E−2
C_C13	gnd	m5	6.00462E−2
C_C14	gnd	m6	4.60781E−1
C_C15	gnd	com	3.03315E+0
C_C16	gnd	u1	2.51716E+1
R_R1	cs	c1	3.97030E−2
R_R2	c1	c2	1.19109E−1
R_R3	c2	c3	3.57327E−1
R_R4	c3	c4	2.59622E−1
R_R5	c4	c5	8.51408E+0
R_R6	c5	c6	6.55035E+0
R_R7	c6	com	3.92892E+1
R_R8	mos	m1	3.81609E−3
R_R9	m1	m2	1.14483E−2
R_R10	m2	m3	3.43448E−2
R_R11	m3	m4	1.03035E−1
R_R12	m4	m5	2.57613E−1
R_R13	m5	m6	5.90459E+0
R_R14	m6	com	2.40044E+1
R_R15	c6	m6	5.31662E+1
R_R16	com	u1	1.72692E+1
R_R17	gnd	u1	4.35939E+0

Table 2. Mathematical Model – Roots and Amplitudes for $R(t) = \sum_{i=1}^{n} \left(\sum_{j=0}^{m} A_{ij} t^j \right) e^{r_i t}$

		MTB40N10E Self Heating		CS5342 Self Heating		Interaction	
i	r_i	A_{i0}	A_{i1}	A_{i0}	A_{i1}	A_{i0}	A_{i1}
1	–3.34538E+6	–2.51600E–3	2.88500E–3	–2.61750E–2	–3.00200E–2	–3.76038E–46	–9.55148E–41
2	–2.27628E+5	–6.15500E–3	–1.12100E+0	–6.36000E–2	1.16007E+1	–4.70539E–32	–9.49738E–28
3	–4.66437E+4	–5.60013E–17		–3.34200E–2		7.92060E–26	
4	–2.25273E+4	–1.83010E–2		1.00724E–15		–6.31528E–23	
5	–9.77316E+3	2.79361E–17		–4.13157E–1		3.43633E–20	
6	–2.18916E+3	–6.57940E–2		–4.03874E–16		–6.66750E–16	
7	–2.29490E+2	–1.77952E–1		–6.91961E–14		6.02476E–10	
8	–9.99632E+1	2.51105E–15		–7.68731E+0		–3.73658E–8	
9	–7.42153E+0	–1.50758E–7		–6.73210E+0		1.00700E–3	
10	–2.42893E+0	–4.24632E+0		–1.89000E–4		–2.83180E–2	
11	–1.62443E–1	–5.32689E+0		–1.09936E+1		7.65256E+0	
12	–1.02279E–1	–7.89576E+0		–8.32867E+0		–8.10933E+0	
13	–1.83258E–2	–1.31004E+1		–1.30787E+1		–1.30896E+1	
14	–6.96478E–3	–1.61600E+1		–1.61463E+1		–1.61531E+1	
15	0.00000E+0	4.70001E+1		6.35032E+1		2.97268E+1	

Notes for Table 2: In this specific case, $n = 15$ (number of "roots", including the constant) and $m = 1$ (maximum root multiplicity less one). The "roots" here are the coefficients in the exponents of the exponential function, so called due to the method of solution by which they are obtained (being the roots of a polynomial, the determinant of the system of equations representing the transient response of the RC network). As such, they are related to the "time constants" of the exponential responses of the network, through the expression $\tau_i = -\frac{1}{r_i}$, hence the equation can be rewritten

$$R(t) = \sum_{i=1}^{n} \left(\sum_{j=0}^{m} A_{ij} t^j \right) e^{-\frac{t}{\tau_i}}.$$

All nodes of a particular network share the same set of time constants, but the amplitudes of the terms corresponding to each time constant differ from node to node. In certain cases, roots are repeated, from which arise the higher order ($j > 0$) terms t^j. Note that most of the higher order A_{i1} coefficients are exactly zero (hence absent from the table); some others are effectively zero and the associated term may therefore be omitted if convenient. (For instance, terms $i = 1$ through $i = 8$ for the "interaction" equation are certainly negligible.) Also, in determining which higher order terms are negligible, it may be useful to note that in general, the maximum of any term $y_{ij} = A_{ij} t^j e^{r_i t}$, is $y_{max} = A_{ij} \tau_i^j e^{-j}$, and it occurs at time $t = j\tau_i$.

Short–time Limits of RC Networks

A significant limitation of RC networks is that for times shorter than the approximated RC time constant of the junction node (which, for a well defined network, will be the fastest node), transient response falls off linearly (going back in time) rather than according to the previously described sqrt(t) constant. This means that if the RC network model is exercised for power fluctuations on the order of (or faster than) this minimum response time, a thermal RC network will vastly underestimate the sqrt(t) response demanded by surface heating theory. (Interestingly, if the RC model is extended to such short times and implied physical dimensions as correspond to locally volumetric heating, *volumetric* heating theory demands *linear* temperature increases with time. In other words, carried to times too short for the *surface* heating model to apply, the RC network model actually becomes "correct" again. However, this requires that one have an input transient curve which is known to exhibit the correct behavior over the entire time scale of interest, whether this be linear at the shortest times, or sqrt(t) at the shortest times. (Observe, for instance, the 1:2 slope in the 1E–4 to 1E–3 s range of the responses of the heated junctions in Figure 16.) The RC response will not magically become linearly correct just because it *fails* to follow the sqrt(t) where that is appropriate. The point, here, is that one must be cautious in the use of an RC network model which is not designed to provide sqrt(t) behavior, over a time scale in which sqrt(t) behavior is expected. If one must work in this time scale without a suitable RC network, Equation 1 should be used directly in lieu of the RC model. For more information, see also AND8218/D.

For further information on Thermal Resistance Measurements:

1. MIL−STD−883E−1012.1, Test Method Standard, Microcircuits, Thermal Characteristics.
2. EIA/JEDEC, JESD24−3, 24−4, 51, 51−1 through 51−8.
3. R.R. Tumala, E.J. Ramaszewski, A.G. Klopfenstein, *Microelectronics Packaging Handbook,* 2nd Ed., Chapman & Hall, New York, 1997.
4. “Basic Semiconductor Thermal Measurements,” Application Note #AN1570, ON Semiconductor, <www.onsemi.com>.
5. Y.L. Xu, R.P. Stout, D.T. Billings, “Electrical Package Thermal Response Prediction to Power Surge” ITHERM, May 2000.
6. R.P. Stout, D.T. Billings, “How to Extend a Thermal−RC−Network Model to Respond to an Arbitrarily Fast Input,” IEEE CPMT−A Journal, Dec. 1998.
7. Michael Pecht, *Handbook of Electronic Package Design,* Marcel Dekker, New York, 1991.
8. Tony Kordyban, *Hot Air Rises and Heat Sinks (Everything You Know About Cooling Electronics is Wrong),* ASME Press, 1998 (ISBN 0−7918−0074−1).

Predicting the Effect of Circuit Boards on Semiconductor Package Thermal Performance

Prepared by: Roger Paul Stout, PE
ON Semiconductor

ON Semiconductor®
http://onsemi.com

APPLICATION NOTE

Abstract

In some situations, a circuit board may be approximated by varying regions of roughly axisymmetric geometry, and a cascaded two−port model may be easily constructed and solved analytically. Each axisymmetric segment of the model is a two−port (in thermal terms, a circular fin lacking the traditional adiabatic boundary condition at the outer radius), hence an arbitrarily complex (axisymmetric) board model is represented by a cascaded two−port network. The overall network is driven by a heat source at its inner radius, and some outer radius boundary condition; the two−port concept naturally separates the interior conduction and surface convection properties from the temperature and heat flux boundary conditions. Using this scheme, temperature and heat flow may be easily determined at every position within the model, thus providing necessary information on the interactions between locations within the model − permitting yet more complex analyses of a circuit board as a multiple heat source system. Application of the axisymmetric two−port method is made to thermal characterization of semiconductor devices, including the relationship between so called "min pad" and "1−inch pad" device characteristics. The model is also compared to other experimental data, where the "best fit" of the model parameters shows a reasonable correlation with the expected physical values of the experiment.

Key Words

thermal analytical model axisymmetric two−port

Glossary of Symbols

A	cross−section conduction area (m^2)
A, B, C, D	two−port transmission parameters
a, b	general purpose constants
G	thermal conductance ($W \cdot C^{-1}$)
h	film coefficient ($W \cdot C^{-1} m^{-2}$)
$I_p(z)$	modified Bessel function of the first kind
$K_p(z)$	modified Bessel function of the second kind
k	fin (board) conductivity ($W \cdot C^{-1} m^{-1}$)
m	system parameter (m^{-1})
q	heat flow (W)
Q	total device power dissipation (W)
r	radial coordinate (m)
T	temperature variable (C)
T_∞	bulk fluid temperature (C)
T	two−port transmission matrix
t	fin (board) thickness (m)
z	dimensionless radius

Greek Symbols

Ψ	temperature rise/device power dissipation ($C \cdot W^{-1}$)
θ	temperature rise above bulk fluid temperature (C)

Subscripts and Superscripts

i, j, k, o	board region or boundary designations
b, s, e	beginning, spreader, ending boundary designations
p	order of Bessel function

INTRODUCTION

In the semiconductor package thermal characterization business, it is ultimately desirable to accurately describe device performance in an actual application environment. Obviously this performance is the result of a combined package and system, thus one must know both the package and the system characteristics in order to answer the question. The reality is that the interaction between a package and its environment may be quite complex (e.g., multiple and significant parallel heat paths, and temperature dependent behavior). Nonetheless it is convenient, and in many cases adequate for first−order estimates, to consider that a uniquely identifiable and clean boundary separates the package from its environment (which, for the purposes of this discussion, will be simply a "board"). From the package's perspective, the board simply absorbs heat (power), and, all else being equal, responds by attaining a predictable temperature at the package/board interface.

From the board's perspective, the package is simply a heat source, the location of the package defining a "boundary" of the board. The power injected into the board by the package constitutes a boundary condition of the model, and the local temperature varies in response. Ambient provides a second boundary condition for the board. It is a fixed temperature, and the ultimate sink for all the power injected by the package. So for a board with a single package present, there is an input heat source and associated temperature, and an output heat "sink" with a second associated temperature.

A circuit board may therefore be thought of as a thermal resistance between the package and ambient, even though it is distinguished from a traditional thermal resistance in that it "leaks" heat to convention as it flows from the input end to the output end.

Two–Ports

In the electrical realm, the external view of a network generally consists of terminals, often grouped in pairs. A terminal pair, or port, is then represented as having a potential, or voltage, between the two terminals of the port, and a current flow into (and return from) the network at that port. A network may have any number of ports, but if it may be described using two ports, it is classified as a two–port network, or more simply, a "two–port." The literature is rich in theory and applications of electrical two–ports, in particular, passive, linear systems which will be the focus here (1,2).

Using the conventional thermal–electrical analogy, thermal "potential" is temperature, corresponding to voltage; thermal "flow" is power, corresponding to current. A circuit board having a single point heat input, and a single output boundary condition, is, therefore, a thermal two–port.

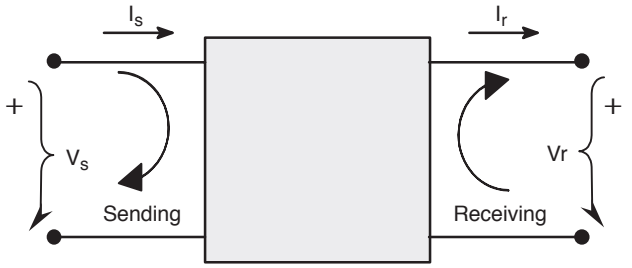


Figure 1. Electrical Two–Port

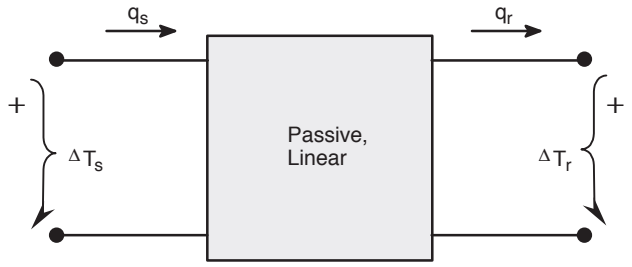


Figure 2. Thermal Two–Port

Note that two–port theory does not require that flow out the receiving port equals flow into the sending port. Rather, the model describes an overall relationship between the four quantities associated with the two ports, as shown in Equation 1, and represents the internal system by way of the *transmission parameters* of the two–port: A, B, C, and D.

$$\begin{Bmatrix} \Delta T_s \\ q_s \end{Bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{Bmatrix} \Delta T_r \\ q_r \end{Bmatrix} \quad \text{(eq. 1)}$$

The transmission matrix, then, is defined as:

$$\mathbf{T} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \quad \text{(eq. 2)}$$

The nature of a two–port is to permit a continuum of boundary conditions at each port, trading off potential for flow, and vice versa. Further, the trade off at one port influences the conditions at the other port. For instance, an adiabatic boundary at the receiving port ($q_r = 0$) translates into particular expressions for both potential and flow at the sending end, purely in terms of the potential at the receiving end:

$$\begin{aligned} \Delta T_s &= A \cdot \Delta T_r \\ q_s &= C \cdot \Delta T_r \end{aligned} \quad \text{(eq. 3)}$$

A zero potential at the receiver ($\Delta T_r = 0$) results in different relationships:

$$\begin{aligned} \Delta T_s &= B \cdot q_r \\ q_s &= D \cdot q_r \end{aligned} \quad \text{(eq. 4)}$$

Alternatively, values may be specified at the sender's port, and corresponding expressions derived for the receiver, by inverting Equation 1:

$$\begin{Bmatrix} \Delta T_r \\ q_r \end{Bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix}^{-1} \begin{Bmatrix} \Delta T_s \\ q_s \end{Bmatrix} \quad \text{(eq. 5)}$$

It turns out that for a properly formulated two–port (1,2), the determinant of the transmission matrix will be unity (which also implies that the four transmission parameters are not completely independent), i.e.:

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = 1 \quad \text{or} \quad AD–BC = 1 \quad \text{(eq. 6)}$$

Equation 5 becomes:

$$\begin{Bmatrix} \Delta T_r \\ q_r \end{Bmatrix} = \frac{1}{AD–BC} \begin{bmatrix} D & –B \\ –C & A \end{bmatrix} \begin{Bmatrix} \Delta T_s \\ q_s \end{Bmatrix}$$

hence: (eq. 7)

$$\begin{Bmatrix} \Delta T_r \\ q_r \end{Bmatrix} = \begin{bmatrix} D & –B \\ –C & A \end{bmatrix} \begin{Bmatrix} \Delta T_s \\ q_s \end{Bmatrix}$$

Simple Boards as Two−Ports

It may now be seen that a simple circuit board, a "leaky" thermal resistance, having a single package as a heat source and a single ambient thermal ground, fits the two−port model. The sending port is the package (injecting heat into the system), and the receiving port represents the thermally distant environment, perhaps the edges of the board or even farther away. The driving potential at the sending port is the temperature differential between the package/board interface and some convenient reference temperature. The potential at the receiving port is a similar temperature differential between some location in the thermal system being modeled, and some other convenient reference temperature.

In many semiconductor applications, it is sensible to reference all temperatures to some common ambient. In particular, both the sending potential (the package/board interface temperature), and the receiving potential (the board perimeter temperature), should be referenced to ambient. Indeed, it will be seen that this is a necessary step in creating a thermal two−port representing a circuit board. It is *not* necessary that the receiving temperature itself *be* ambient; if this happens to be so, it is the special case where the receiving potential is zero (analogous to the short−circuit behavior of an electrical two−port).

What then should be the receiving port, that is, what temperature and location in a circuit board thermal model should be selected as the receiving port? To make sense mathematically, the temperature must be an isotherm of the system. That is to say, since a single scalar value is being used in the model to represent the temperature at some location in the system, it simultaneously represents *all* the points in the system sharing that same temperature. To make sense physically, the requirement is to identify an isotherm that remains an isotherm over the operating range of interest in the system (which is not to say that the temperature of the isotherm is constant, but rather that its shape remain fixed.) In real world systems, we may have to compromise. Whatever isotherm is chosen as the receiving port will influence the particular transmission parameters for that model. Our goal should be to choose a useful one.

Axisymmetric Board

Consider the very simple model of a circuit board shown in Figure 3. What is particularly helpful about the axisymmetric model is that by definition the outer edge of the board will be an isotherm, hence an acceptable choice as the receiving port of the two−port. In this model, there is an axisymmetric heat source at the inner radius (the package), and an exterior circular perimeter of the board at which we will specify both a temperature rise above ambient and possibly a heat flow from the edge. Between the inner and outer radii, board properties are uniform, and heat is lost continuously to convection, characterized by a constant film coefficient. Note that if the heat loss at the outer radius is zero, we have a conventional circular fin, whose solution may be found in any number of references (3,4). We are interested here, however, in the more general case where the outer edge is neither adiabatic nor a fixed temperature, i.e., it is a thermal "port."

The governing equation for this system (5,6) is:

$$\frac{d^2T}{dr^2} + \frac{1}{r}\frac{dT}{dr} - \frac{2h}{kt}(T - T_\infty) = 0 \qquad \text{(eq. 8)}$$

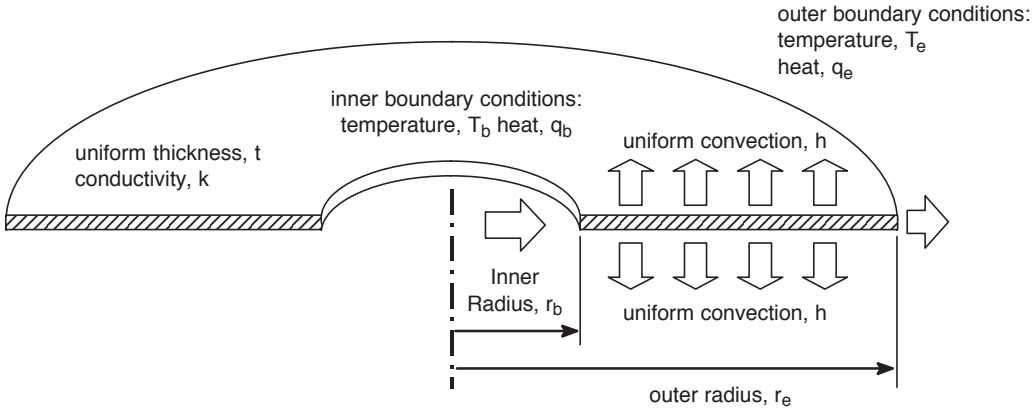


Figure 3. Basic Problem Geometry

With the following changes of variables,

$$\theta = T - T_\infty \quad \text{(eq. 9)}$$

$$m = \sqrt{\frac{2h}{kt}} \quad \text{(eq. 10)}$$

$$z = mr \quad \text{(eq. 11)}$$

Equation 8 becomes:

$$\frac{d^2\theta}{dz^2} + \frac{1}{z}\frac{d\theta}{dz} - \theta = 0 \quad \text{(eq. 12)}$$

Note that with this definition of θ, all temperatures in the model (throughout the domain, inclusive of the two boundaries) become thermal potentials, now referenced to the common ambient. Being a second order, homogeneous ordinary differential equation, the general solution of Equation 12 may be written (7) as follows:

$$\theta = aI_0(z) + bK_0(z) \quad \text{(eq. 13)}$$

where a and b are arbitrary constants, and $I_0(z)$ and $K_0(z)$ are the *modified Bessel functions* of the first and second kinds, respectively. (Additional detail may be found in the Appendix.) Since we also are interested in flow related boundary conditions for the two−port, we must utilize the basic Fourier heat conduction equation, that is:

$$q = -kA(r)\frac{dT}{dr} \quad \text{(eq. 14)}$$

which relates local temperature gradients (changes in potential) to local heat flow. Equation 14 was in fact central to the derivation of the original governing Equation 8. Again using the change of variables from r into z, we have:

$$\begin{aligned} q &= -k \cdot t \cdot 2\pi r \cdot \frac{d\theta}{dz}\frac{dz}{dr} \\ &= -2\pi ktmr\theta' \\ &= -2\pi ktz\theta' \end{aligned} \quad \text{(eq. 15)}$$

Equation 15 thus leads to a second expression in the two arbitrary constants (refer again to the Appendix for additional detail on the Bessel functions), specifically:

$$q = -2\pi ktz[aI_1(z) - bK_1(z)] \quad \text{(eq. 16)}$$

Axisymmetric Board as a Two−Port

We may now put the axisymmetric board solution into the form of a two−port. Observe that together, Equations 13 and 16 constitute a system of two equations in the two unknowns a and b, given appropriate boundary values for the potential and flow.

In matrix form, these equations become:

$$\begin{Bmatrix} \theta_i \\ q_i \end{Bmatrix} = \begin{pmatrix} I_0(z_i) & K_0(z_i) \\ -G_iI_1(z_i) & G_iK_1(z_i) \end{pmatrix} \begin{Bmatrix} a \\ b \end{Bmatrix} \quad \text{(eq. 17)}$$

where:

$$G_i = 2\pi ktz_i \quad \text{(eq. 18)}$$

We thus can solve for a and b given a potential and flow at any position within the domain. For instance, if specified at some z_i, then:

$$\begin{Bmatrix} a \\ b \end{Bmatrix} = \begin{pmatrix} I_0(z_i) & K_0(z_i) \\ -G_iI_1(z_i) & G_iK_1(z_i) \end{pmatrix}^{-1} \begin{Bmatrix} \theta_i \\ q_i \end{Bmatrix} \quad \text{(eq. 19)}$$

Now if the domain is defined over a range from z_1 to z_2, inclusive, then Equation 19 is true for the two endpoints z_1 and z_2, that is:

$$\begin{Bmatrix} a \\ b \end{Bmatrix} = \begin{pmatrix} I_0(z_1) & K_0(z_1) \\ -G_iI_1(z_1) & G_iK_1(z_1) \end{pmatrix}^{-1} \begin{Bmatrix} \theta_1 \\ q_1 \end{Bmatrix} \quad \text{(eq. 20)}$$

and also:

$$\begin{Bmatrix} a \\ b \end{Bmatrix} = \begin{pmatrix} I_0(z_2) & K_0(z_2) \\ -G_iI_1(z_2) & G_iK_1(z_2) \end{pmatrix}^{-1} \begin{Bmatrix} \theta_2 \\ q_2 \end{Bmatrix} \quad \text{(eq. 21)}$$

We can therefore eliminate a and b between Equations 20 and 21, leaving an expression for the boundary conditions at one end of the domain in terms of those at the other end of the domain:

$$\begin{Bmatrix} \theta_1 \\ q_1 \end{Bmatrix} = \begin{pmatrix} I_0(z_1) & K_0(z_1) \\ -G_iI_1(z_1) & G_iK_1(z_1) \end{pmatrix} \begin{pmatrix} I_0(z_2) & K_0(z_2) \\ -G_iI_1(z_2) & G_iK_1(z_2) \end{pmatrix}^{-1} \begin{Bmatrix} \theta_2 \\ q_2 \end{Bmatrix} \quad \text{(eq. 22)}$$

Recall now Equation 1:

$$\begin{Bmatrix} \Delta T_s \\ q_s \end{Bmatrix} = \begin{bmatrix} A & B \\ C & D \end{bmatrix} \begin{Bmatrix} \Delta T_r \\ q_r \end{Bmatrix} \quad \text{(eq. 23)}$$

It should be evident that Equation 22 is the two−port representation of the axisymmetric board model in accordance with Equation 1. Finally, the transmission matrix for the two−port is the following two by two matrix product:

$$\mathbf{T}_{ij} = \begin{pmatrix} I_0(z_i) & K_0(z_i) \\ -G_iI_1(z_i) & G_iK_1(z_i) \end{pmatrix} \begin{pmatrix} I_0(z_j) & K_0(z_j) \\ -G_jI_1(z_j) & G_jK_1(z_j) \end{pmatrix}^{-1} \quad \text{(eq. 24)}$$

or

$$\mathbf{T}_{ij} = \begin{bmatrix} A_{ij} & B_{ij} \\ C_{ij} & D_{ij} \end{bmatrix} \quad \text{(eq. 25)}$$

where the subscript pairs indicate that it applies over the domain spanning boundaries i and j. Certainly Equation 24 could be expanded to provide expressions for each of the four transmission parameters explicit in Equation 25 – see the Appendix – but there is no particular benefit in doing so. Indeed, presuming that a computerized tool of some sort will be used to perform actual calculations, there is less opportunity for error in implementing the form of Equation 24 directly, given its symmetry and relative clarity. (Refer again to the Appendix). Observe, also, that the symmetry of Equation 24 demonstrates that simply exchanging boundary subscripts corresponds exactly to obtaining the reverse transmission matrix, i.e.:

$$\mathbf{T}_{ji} = \mathbf{T}_{ij}^{-1} \quad \text{(eq. 26)}$$

In other words, so long as our bookkeeping is consistent, we can build our transmission matrices symbolically in whichever “orientation” is desired (small radius to large, or vice versa). The sending port need not be the heat source at all, unless this is convenient for the problem at hand.

Single–Zone Thermal Test Board

The Appendix shows how the axisymmetric transmission matrix of Equation 24 may be used to derive the following expression for the temperature distribution in a circular board with an adiabatic edge. This may also be found under the guise of a “circular fin” in references such as (3) and (7). In Equation 27, the b subscript represents the base of the fin (the sending port), and the e subscript the adiabatic outer radius of the fin (the receiving port):

$$\frac{T-T_\infty}{T_b-T_\infty} = \frac{K_1(mr_e)I_0(mr) + I_1(mr_e)K_0(mr)}{K_1(mr_e)I_0(mr_b) + I_1(mr_e)K_0(mr_b)} \quad \text{(eq. 27)}$$

When a real thermal test board is considered, none of the “constants” may actually be known accurately, especially considering that we’re using this axisymmetric model to approximate a rectilinear geometry (and even then, rectangular circuit boards are generally far from uniform in material properties, thanks to buried irregular metal planes and actual circuit traces everywhere). Thus, if temperatures are measured at known distances from a heat source, the best choices for T_b, r_b, r_e, and m may be better left as a statistical “best fit” problem. Our main interest here is whether or not the temperature profile derived for a uniform axisymmetric circuit board, bears any resemblance at all to a real temperature profile of a real board. If so, we can learn much about the trade–offs of board properties (size, conductivity, convection coefficient) holding the package itself as a constant.

For this comparison, it is useful to rephrase Equation 27 in terms of our usual “normalized” temperatures where everything is related to the total power dissipation of the heat source in question, Q (that is, the heat input at the inside radius r_b), and the bulk convecting fluid temperature T_∞. For simplicity in the best fit process, we have chosen the following form of the solution (see Appendix):

$$\frac{T-T_\infty}{Q} = c\left[I_0(mr) + \frac{I_1(mr_e)}{K_1(mr_e)}K_0(mr)\right] \quad \text{(eq. 28)}$$

where c, m, and r_e are the independent fit parameters. (Following JEDEC terminology (8–10), this quantity would be known as Ψ_{BA}, or psi–BA, the board–to–ambient temperature difference normalized by package power dissipation.)

Figure 4 illustrates an actual thermal test situation where two components of interest were mounted on a customer’s board, providing an excellent opportunity to compare the analytical board model Equation 28 to a real application. This circuit board had a fairly continuous embedded power plane, so its thermal conductivity would be expected to be reasonably uniform over a large area. Each device (a two–channel TMOS driver) could be heated at either end, and several thermocouples were placed on the board as indicated in the figure. (Not shown is a TC on the back side of the board.) Between the four heat sources and eight TC’s, approximately 30 different samples of temperature rise vs. distance were available.

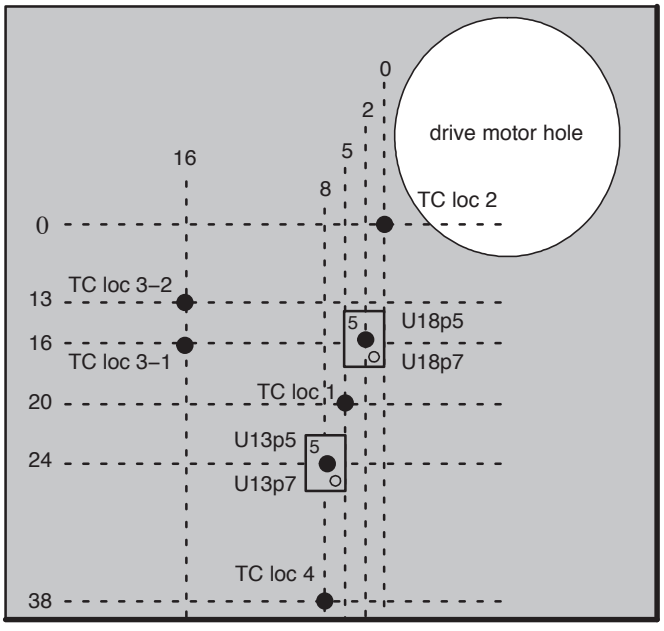


Figure 4. Board Diagram of Heat Sources and TC Locations

Two different test configurations were measured. Figure 5 shows the results (normalized data vs. the best–fit board models) where both sides of the test board were exposed to free convection conditions. Figure 6 shows the results where only one side was exposed to convection, and the other blocked from convection with a thick layer of low density insulating foam. The best fit parameters could, if desired, be compared with the theoretical values based on the known properties, or conversely, used to estimate what the material and convection properties must be in order to give those parameter values. The main point to be made here is that this “physics based” axisymmetric model, of board temperature variation with distance from heat source, is clearly a very reasonable approach, even for a real application board with rectangular geometry and other non–ideal attributes.

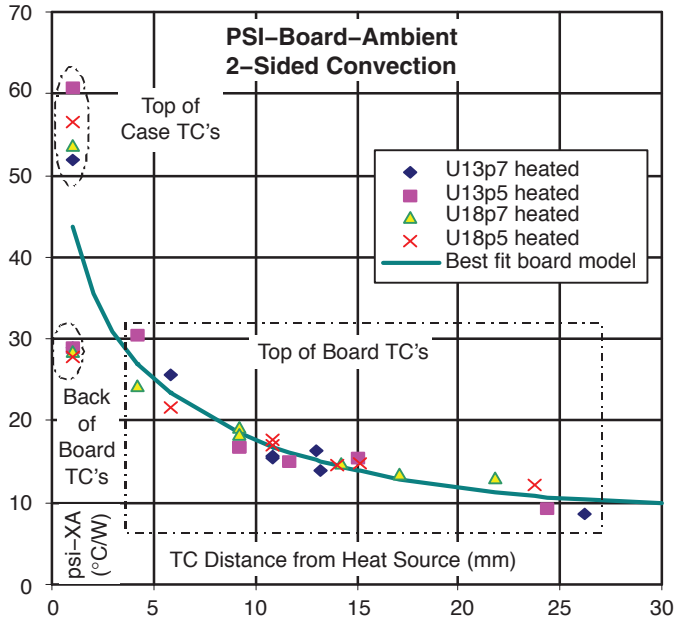


Figure 5. Closed–Form Best Fit to 2–Sides Convection Data

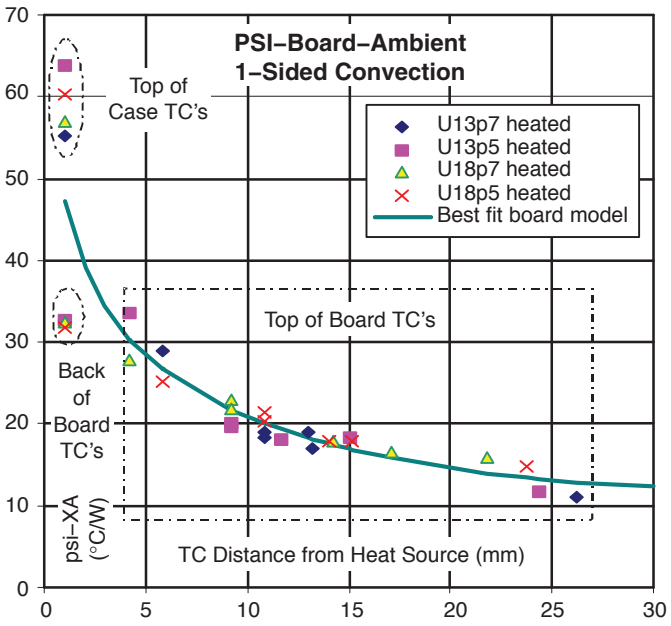


Figure 6. Closed–Form Best Fit to 1–Side Convection Data

Multiple−Zone Two−Port Board Model

In the semiconductor component manufacturer's thermal characterization business, commonly encountered board designs are the so−called "1−inch pad" test boards (Figure 7).

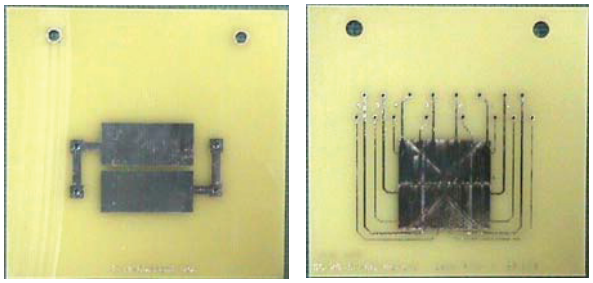


Figure 7. 1−Inch Pad Test Boards for 2 and 8−Lead Devices

In these designs, a plated−on copper heat spreader surrounds the immediate vicinity of the test device (generally mounted at the center of the heat spreader). The spreader is itself much smaller than the entire test board. Thus the effective thermal conductivity of the board changes radically from an inner zone to an outer zone. Even with the simplifying assumption of axisymmetry, the "brute force" approach to this problem would be to return to the original governing Equation 8, solve it for two separate zones with continuity conditions at their interface, and apply appropriate boundary conditions at the overall inner and outer radii.

The two−port approach is far more elegant. Buried within the transmission parameters of Equation 23 is all the internal behavior of the domain, including the details of the physics, the geometry, and even the nature of the coordinate system. The ports are the boundary conditions. When two−ports are cascaded, each interface is by definition an application of continuity (equal potential, conservation of flow).

For instance, to represent the 1−inch pad test board, we may most simply write just two transmission matrices, multiply them through, and apply the desired overall boundary conditions:

$$\begin{Bmatrix} \theta_b \\ q_b \end{Bmatrix} = \mathbf{T}_{bs} \cdot \mathbf{T}_{se} \cdot \begin{Bmatrix} \theta_e \\ q_e \end{Bmatrix} \qquad \text{(eq. 29)}$$

where here we have labeled the boundary of the spreader area as interface s, so the inner domain $(b-s)$ is covered by the 1−inch heat spreader, and the outer domain $(s-e)$ is plain circuit board material. (Using the axisymmetric formulation, we assume there to be a circular isotherm separating the plated zone from the bare zone. A sensible choice for the radius of boundary s is that which gives the same area as the square heat spreader.) We may apply a different film coefficient to each zone, if desired, but at the very least, the effective conductivity of the board will be different in the two regions due to the plating on the inner region.

If we don't care about the temperatures anywhere in between the base and the end, Equation 29 is all there is to it. For example, we can now easily explore the first of two commonly asked questions in semiconductor packaging, namely what is the difference in thermal performance on a "min pad" vs. a 1−inch pad for a particular package? (A min−pad board has no metal except the minimum traces and pads necessary to mount the device and access it electrically.) Intuitively, we know that the min−pad performance is a strong function of the size of the package, and that 1−inch pad performance is much better than min−pad performance. Equation 29 gives us a simple, yet physics−based answer to this question. Note also that if both inner and outer regions are given the same (unplated) board material properties, Equation 29 applies equally well for the min−pad board.

Let us rephrase Equation 29 in terms more specific to our immediate interest:

$$\begin{Bmatrix} T_b - T_\infty \\ Q \end{Bmatrix} = \mathbf{T}_{bs} \cdot \mathbf{T}_{se} \cdot \begin{Bmatrix} T_e - T_\infty \\ 0 \end{Bmatrix} \qquad \text{(eq. 30)}$$

As before, T_∞ is the bulk fluid temperature (i.e., ambient). T_b is the board temperature, found at the inner domain boundary radius r_b (that is, the inner boundary of the 1−inch copper plated region). Q is the total package power dissipation, which is input at r_b.

The inner radius itself, r_b, represents the package "size". To get the most use out of the results, ultimately we will want to correlate specific package geometries of interest to this "generic" package size parameter. For example, in a standard dual inline type package, r_b might best be represented as half the distance from the leads on one side of the package to the leads on the other side of the package; whereas for a soldered heatsink type device, the best choice might be the radius giving the same area as the actual heatsink. In any event, r_b is a fundamental variable in this analysis, implicit in $\mathbf{T}_{bs}$.

Finally, T_e is the board temperature at the outer radius of the plain circuit board material domain, r_e. (To be consistent, its value will be the radius that gives the same area as the actual total area of the test board.) Though we don't as yet have a value for T_e, we have assumed in Equation 30 that the exterior radius of the test board is adiabatic (meaning the outer radius heat flow is expressly zero).

Introducing our subscript notation to indicate the domain of application:

$$\begin{bmatrix} A_{be} & B_{be} \\ C_{be} & D_{be} \end{bmatrix} = \mathbf{T}_{bs} \cdot \mathbf{T}_{se} \qquad \text{(eq. 31)}$$

Using this, Equation 30 becomes:

$$\begin{Bmatrix} T_b - T_\infty \\ Q \end{Bmatrix} = \begin{bmatrix} A_{be} & B_{be} \\ C_{be} & D_{be} \end{bmatrix} \begin{Bmatrix} T_e - T_\infty \\ 0 \end{Bmatrix} \qquad \text{(eq. 32)}$$

or, expanding back into two individual equations,

$$T_b - T_\infty = A_{be}(T_e - T_\infty) \qquad \text{(eq. 33)}$$

$$Q = C_{be}(T_e - T_\infty) \quad \text{(eq. 34)}$$

Thus, the normalized temperature rise at the adiabatic external board radius is:

$$\frac{T_e - T_\infty}{Q} = \frac{1}{C_{be}} \quad \text{(eq. 35)}$$

and the normalized temperature rise at the inner radius, Ψ_{BA} is:

$$\frac{T_b - T_\infty}{Q} = \frac{A_{be}}{C_{be}} \quad \text{(eq. 36)}$$

For quantitative results, the procedure is to compute the system transmission matrix from the individual domains' **T**'s. From that, extract just the two elements necessary to compute Equation 36. Figure 8 is a plot for three cases of interest, a min–pad board, and a 1–inch copper heat–spreader board with two different thicknesses of copper (1 oz and 2 oz). The horizontal axis is named *package radius*, and is the board inner radius r_b.

A second commonly asked question is, for a given package, how does the thermal performance vary with copper area? Equation 32 again contains the answer, if we now hold the package size as a constant, and vary the intermediate domain boundary, r_s. Though not explicit when we presented Equation 36, this intermediate boundary radius appears in the second half of $\mathbf{T}_{bs}$, and the first half of $\mathbf{T}_{se}$. (In Figure 8, it was a constant corresponding to the "radius" of the 1–inch square pad.) Figure 9 shows a sample of these alternate possibilities arising from Equation 36.

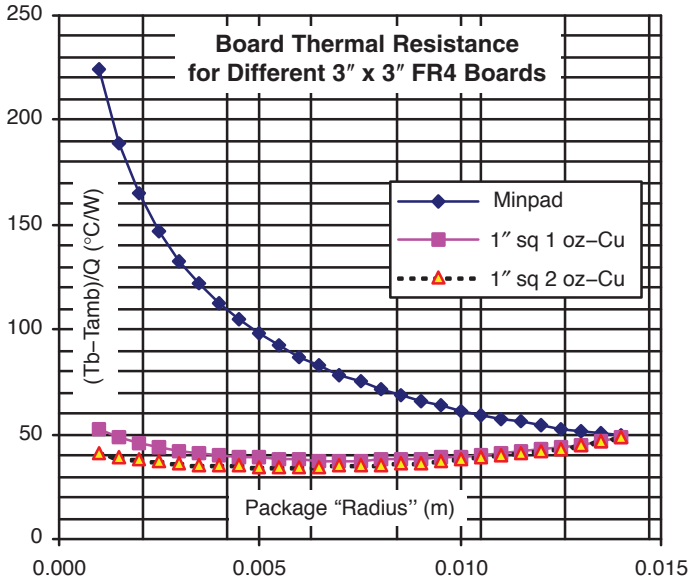


Figure 8. Board Thermal Resistance, Varying Package Size

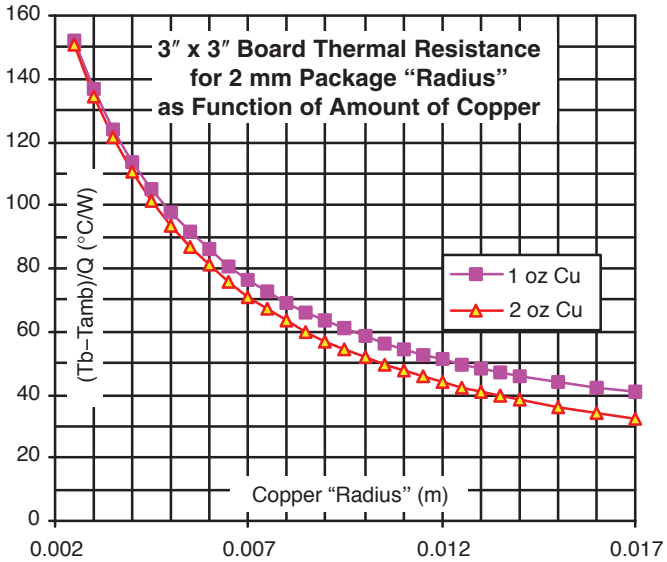


Figure 9. Board Thermal Resistance Varying Copper Size

Solution for Internal Temperature Profile

At the outset, it was suggested that a two−port model was ideally suited for a system−level view of the test board, that is, overall performance without requiring internal solution details. However, the two−port board model in fact can be used to predict the temperature profile within the various zones. (Indeed, this was how Equation 27 is derived in the Appendix.) Specifically, temperature rise as a function of distance from the heat source, is precisely the "interaction strength" that one device (heat source) has on other nearby devices. It mainly depends on the thermal characteristics of the board, and has practically nothing to do with particular package characteristics. Adding even more motivation, the so called "reciprocity theorem" (1,2) allows us to determine the interaction strength of locations in geometries which would otherwise be intractable to analyze directly. For example, the axisymmetric problem in the left of Figure 10 we can solve using the techniques outlined here, whereas to obtain a complete closed−form solution to the one on the right, with an eccentric heat source, is extremely difficult, if not impossible. Yet the reciprocity theorem says that the response at the center of the board to the eccentric heat source, will be identical to the response of the eccentric position to a central heat source.

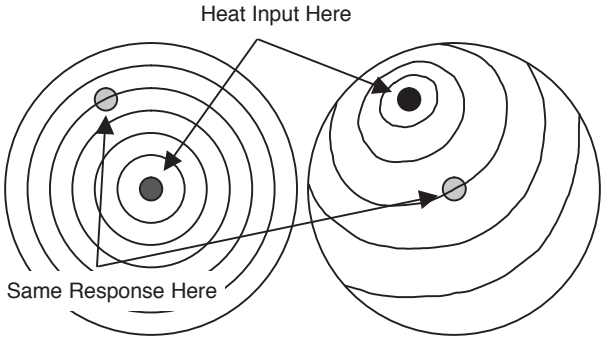


Figure 10. Illustration of the Reciprocity Theorem

To solve for the internal temperature profile in the axisymmetric board model, we note again that any number of two−ports may be cascaded, resulting in but a more complex two−port. It still has but four transmission parameters, though it may be extremely tedious to write them explicitly in symbolic form. With automated computational tools, however, a lengthy string of two−ports is effectively no different than a finite element model. In the case of our axisymmetric board, we could, for instance, model arbitrarily varying convection, conduction, or thickness as a function of radius. To predict the internal temperature profile of our 1−inch pad board, however, it suffices to expand our existing system to use four zones, as illustrated in Figure 11.

How does this help? As we noted earlier in the two−zone model of the 1−inch pad board, when the material properties are the same in the two zones, the model corresponds overall to the min−pad board, or single−zone problem. Clearly, then, if the two zones on either side of a boundary have the same properties, then the specific position of the boundary is irrelevant to the overall behavior of the model (shown rigorously in the Appendix).

On the other hand, if we solve for the potential and flow at the internal boundary, we obtain different values, depending obviously on exactly where the boundary is placed. Thus, a variable radius boundary, within a zone of constant properties, serves as a probe for the temperature and flow within that region. For example, suppose we want the temperature profile within the inner, copper plated region of the 1−inch board model. We construct a new transmission matrix:

$$\mathbf{T}_{r_i e} = \mathbf{T}_{r_i s} \cdot \mathbf{T}_{se} \qquad \text{for } r_b \le r_i \le r_s \quad \text{(eq. 37)}$$

where $\mathbf{T}_{se}$ is exactly as before, but now $\mathbf{T}_{r_i s}$ is based on the 1−inch pad domain properties with a variable inner radius r_i. The following parallel to Equation 32 results:

$$\begin{Bmatrix} T(r_i)-T_\infty \\ q(r_i) \end{Bmatrix} = \begin{bmatrix} A_{r_i e} & B_{r_i e} \\ C_{r_i e} & D_{r_i e} \end{bmatrix} \begin{Bmatrix} T_e-T_\infty \\ 0 \end{Bmatrix} \quad \text{(eq. 38)}$$

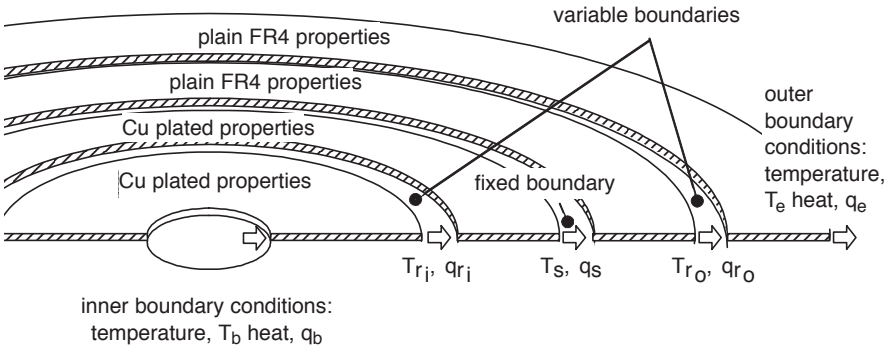


Figure 11. Multiple Regions of Annular Geometry

What we do *not* want now is:

$$\frac{T(r_i)-T_\infty}{q(r_i)} = \frac{A_{rie}}{C_{rie}}$$

because Q, upon which our normalization in Equation 32 is based, is definitely not the same as $q(r_i)$. (The reason is that some of the device power has been dissipated by convection in the area between r_b and the radius we're trying to probe.) Rather, from Equation 38 we obtain:

$$T(r_i)-T_\infty = A_{rie}(T_e-T_\infty) \qquad \text{(eq. 39)}$$

and putting this together with Equation 33, the correct expression is:

$$\frac{T(r_i)-T_\infty}{Q} = \frac{A_{rie}}{C_{be}} \qquad \text{for } r_b \le r_i \le r_s \qquad \text{(eq. 40)}$$

Similarly, we have in the outer region:

$$\frac{T(r_o)-T_\infty}{Q} = \frac{A_{roe}}{C_{be}} \qquad \text{for } r_s \le r_o \le r_e \qquad \text{(eq. 41)}$$

where we find A_{roe} directly from what previously was the T_{se} transmission matrix, but now evaluated at a variable inner radius r_o, rather than the fixed 1−inch "square" spreader radius r_s. Equations 40 and 41 have been used to generate the results shown in Figure 13 for the same three boards previously analyzed in Figure 8, and two of which were analyzed also in Figure 9. Observe, for example, that the value of 47°C/W appears in all three figures for the specific case of the 1 in−1 oz board at a package (or board inner) radius of 0.002 m. (In Figure 9, this value is found at a copper radius of 0.014 m, which is the circular radius giving an area of 1 sq. in. and was the spreader radius used to generate Figures 8 and 13).

Figure 12 shows a 1−inch−pad thermal test board with an "SMB" body−style package, and eight thermocouples placed at various locations on the board, four TC's along a diagonal from the device to the corner of the board, and four on a line through the center of the package and parallel to the board (and 1−inch pad). TC's were soldered to the pad locations, and glued at the FR4 locations. The SMB has a body size of approximately 0.0033 x 0.0043 m, and its two leads are 0.005 m apart.

For the purposes of Figure 14, the SMB was modeled as a centered, axisymmetric heat source of radius 0.003 m. Ψ_{BA} was calculated for each TC based on total heat dissipation of the device. Measurements were made at both 0.75 W and 1.5 W and averaged for the chart. The only TC reading that does not have excellent agreement with the model is the one at the corner of the 1−inch pad, where its actual distance is apparently larger than its "effective" distance from the heat source. In fact, it is very close to the perimeter of the pad, so it might be that a better measure of "distance" on a square spreader such as this is its relative position as compared to the circularized spreader radius; this would shift it to just left of the pad edge in Figure 14, where it obviously would be in much better agreement with the axisymmetric model.

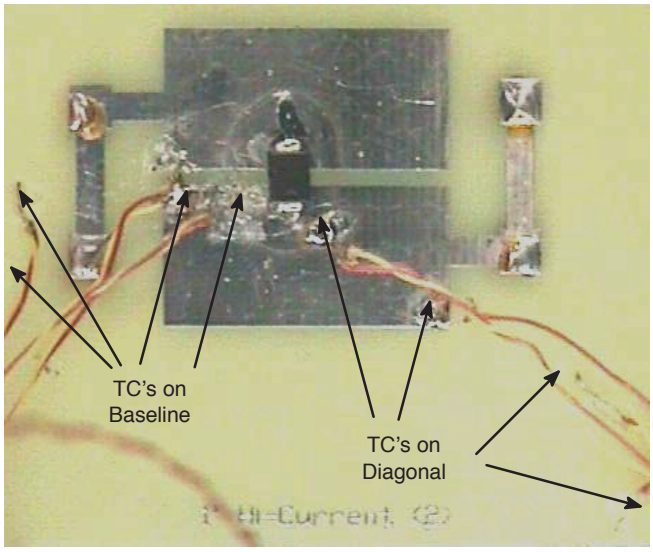


Figure 12. SMB and TC's on 1−inch Pad Thermal Test Board

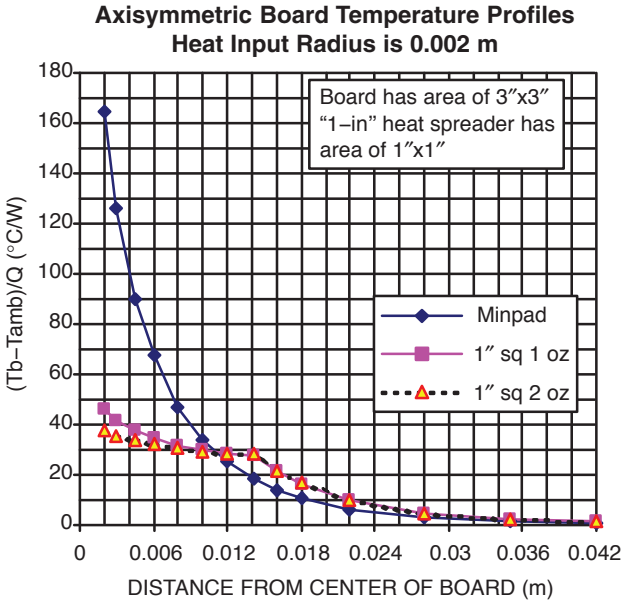


Figure 13. Temperature Profile Over Board Radius

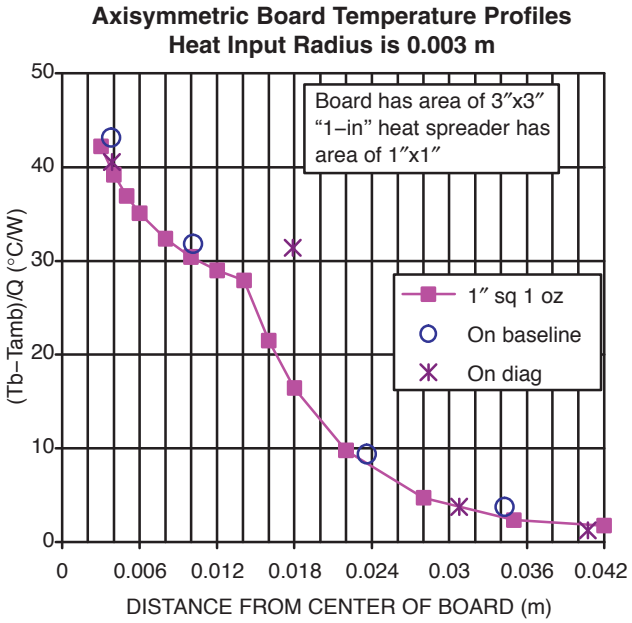


Figure 14. Axisymmetric Board Profile vs. Actual 1″ Pad Data

Limitations

Aside from the obvious limitations of imposing the simplifying axisymmetric assumptions onto far more complicated real geometries, the main limitation not discussed to this point is the assumption that 100% of the device power is deposited into the board at the inner boundary of the axisymmetric model. In reality, of course, some device power will be lost by convection and radiation directly into the ambient environment from the exposed surface of the package; similarly, some will be lost from the (presumably continuous and exposed) far side surface of the board beneath the package (and therefore inside the nominal inner boundary radius). In many cases, this amount of heat loss is a negligible fraction of the total. When it is not, the two–port board model provides the true thermal resistance of the board’s path, facilitating the construction of a more complete model that accounts for these additional heat losses in parallel to it. This would be approach when, for example, the package has a locally applied heatsink that diverts a significant portion of the package power directly into the air above the package, thus bypassing the board. The two–port solution provides the thermal resistance of the board to be modeled as a parallel resistance to the heatsink path.

Final Discussion – Why Bother?

Anyone with a little mathematical knowledge can fit a polynomial curve to the data of Figures 5, 6, 13, or 14. Anyone, even with no mathematical knowledge, but with a modern spreadsheet program, can fit any number of types of curves to the same data. Only a physics based method, however, can provide the correct form of the curve to which that data *should* be fit, and thus with any justification extrapolated beyond the endpoints, or to somewhat different conditions. It is remarkable indeed, and a testimony to the validity of the preceding assertion, that the simple 2D axisymmetric two–port developed in this paper does so well in fitting actual data from real life, 3D non–axisymmetric situations. That it may be implemented in a spreadsheet program with a few simple formulas (see the Appendix), makes it well worth the effort in deriving. All the more so, given that it may be easily extended, using the same simple computational tools, to more complex systems, limitations notwithstanding.

It has been suggested that because one needs a computer to readily implement the axisymmetric thermal two–port, there is little value in utilizing it in preference to a more sophisticated (and therefore less approximate) computerized methodology, for example, a full finite element or finite difference thermal modeler. Unfortunately, many end users of semiconductor packages, especially of the “commodity” type, have no more sophistication than a common spreadsheet program, nor are they willing or able to invest in a more specialized tool. Likewise, manufacturers of commodity semiconductor devices are often asked to provide very simple, quick responses to questions on thermal performance, without sufficient information to accurately define the end application at all. There is a place, therefore, for a modeling method which accounts for the actual physics of a combined convection/conduction thermal system using the most mundane of computer tools, with very little investment in time and detail. It may require the knowledge of a thermal specialist, but it need not require other specialized resources.

SUMMARY AND CONCLUSIONS

A simple axisymmetric board model (the equivalent of a “circular fin”) was used as the foundation for a two–port network thermal analysis. It was shown to match reasonably well certain situations of actual thermal test board data. Multiple two–ports cascaded together enabled complex models of multiple zones of constant property domains to be constructed. This two–port network method was then used to quickly construct parameterized predictions of board–ambient thermal characteristics using the simplest possible idealization of a semiconductor package, that is, a heat source of a certain size. It also was shown that the two–port method may be used to predict temperature gradients within a board. This, in turn, permits rough estimates of component interactions in more complex thermal environments.

ACKNOWLEDGMENTS

The author would like to acknowledge the dedicated efforts of the staff of the ON Semiconductor Thermal Characterization Lab, specifically Ron Lawson, without whom we would have no test equipment with which to perform high quality thermal transient cooling tests, nor data to analyze.

APPENDIX

Modified Bessel's Equation

A particular form of Bessel's equation:

$$\frac{d^2\theta}{dz^2} + \frac{1}{z}\frac{d\theta}{dz} + \left(-1-\frac{p^2}{z^2}\right)\theta = 0 \qquad \text{(eq. A1)}$$

has as its general solution the modified Bessel functions of the first and second kinds, usually denoted $I_p(z)$ and $K_p(z)$ respectively. Comparison of this equation with the axisymmetric governing Equation 12:

$$\frac{d^2\theta}{dz^2} + \frac{1}{z}\frac{d\theta}{dz} - \theta = 0$$

shows that our problem is the special case where $p = 0$, hence the two linearly independent solutions for our second order problem will be $I_0(z)$ and $K_0(z)$.

The derivatives of the modified Bessel functions are given in general as:

$$\frac{dI_p(az)}{dz} = aI_{p+1}(az) + \frac{p}{z}I_p(az) \qquad \text{(eq. A2)}$$

$$\frac{dK_p(az)}{dz} = -aK_{p+1}(az) + \frac{p}{z}K_p(az) \qquad \text{(eq. A3)}$$

Thus the specific derivatives of our solutions will be:

$$\frac{dI_0(z)}{dz} = I_1(z) \quad \text{and} \quad \frac{dK_0(z)}{dz} = -K_1(z) \qquad \text{(eq. A4)}$$

We thus arrive at the general solutions for potential and flow given in the body of the paper as Equations 13 and 16.

Expansion of Axisymmetric Transmission Matrix

Given the axisymmetric transmission matrix, with $G_i = 2\pi k t z_i$ (see Equation 24):

$$\mathbf{T}_{ij} = \begin{pmatrix} I_0(z_i) & K_0(z_i) \\ -G_i I_1(z_i) & G_i K_1(z_i) \end{pmatrix}\begin{pmatrix} I_0(z_j) & K_0(z_j) \\ -G_j I_1(z_j) & G_j K_1(z_j) \end{pmatrix}^{-1}$$

recall:

$$\det\begin{bmatrix} a & b \\ c & d \end{bmatrix} = ad-bc \quad \text{and} \quad \begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$$

thus Equation 24 may be expanded into the more explicit form:

$$\mathbf{T}_{ij} = \frac{\begin{pmatrix} I_1(z_j)K_0(z_i) + I_0(z_i)K_1(z_j) & \frac{I_0(z_j)K_0(z_i)-I_0(z_i)K_0(z_j)}{2\pi k t z_j} \\ 2\pi k t z_i[I_1(z_j)K_1(z_i)-I_1(z_i)K_1(z_j)] & \frac{z_i}{z_j}[I_1(z_i)K_0(z_j) + I_0(z_j)K_1(z_i)] \end{pmatrix}}{I_1(z_j)K_0(z_j) + I_0(z_j)K_1(z_j)} \qquad \text{(eq. A5)}$$

Now the Wronskian of $I_p(z)$ and $K_p(z)$ states (11):

$$I_{p+1}(z)K_p(z) + I_p(z)K_{p+1}(z) = \frac{1}{z} \qquad \text{(eq. A6)}$$

hence Equation A5 becomes, more simply:

$$\mathbf{T}_{ij} = \begin{pmatrix} z_j[I_1(z_j)K_0(z_i) + I_0(z_i)K_1(z_j)] & \frac{I_0(z_j)K_0(z_i)-I_0(z_i)K_0(z_j)}{2\pi k t} \\ 2\pi k t z_j z_i[I_1(z_j)K_1(z_i)-I_1(z_i)K_1(z_j)] & z_i[I_1(z_i)K_0(z_j) + I_0(z_j)K_1(z_i)] \end{pmatrix} \qquad \text{(eq. A7)}$$

Equation A6 also allows us to quickly demonstrate the expected identity for the properly formed two–port:

$$\det \mathbf{T}_{ij} = \frac{z_i[I_1(z_i)K_0(z_i) + I_0(z_i)K_1(z_i)]}{z_j[I_1(z_j)K_0(z_j) + I_0(z_j)K_1(z_j)]} = 1$$

Variable Radius within Constant Property Region

Consider the cascaded transmission matrices based on Equation 24 for two zones, b and e, separated by boundary s.

$$\mathbf{T}_{bs} = \begin{pmatrix} I_0(z_b) & K_0(z_b) \\ -a_1 z_b I_1(z_b) & a_1 z_b K_1(z_b) \end{pmatrix} \begin{pmatrix} I_0(z_s) & K_0(z_s) \\ -a_1 z_s I_1(z_s) & a_1 z_s K_1(z_s) \end{pmatrix}^{-1} \quad \text{(eq. A8)}$$

$$\mathbf{T}_{se} = \begin{pmatrix} I_0(z_s) & K_0(z_s) \\ -a_2 z_s I_1(z_s) & a_2 z_s K_1(z_s) \end{pmatrix} \begin{pmatrix} I_0(z_e) & K_0(z_e) \\ -a_2 z_e I_1(z_e) & a_2 z_e K_1(z_e) \end{pmatrix}^{-1} \quad \text{(eq. A9)}$$

where $a_x = 2\pi k_x t_x$.

The overall transmission matrix $\mathbf{T}_{be}$ is thus:

$$\mathbf{T}_{be} = \begin{pmatrix} I_0(z_b) & K_0(z_b) \\ -a_1 z_b I_1(z_b) & a_1 z_b K_1(z_b) \end{pmatrix} \begin{pmatrix} I_0(z_s) & K_0(z_s) \\ -a_1 z_s I_1(z_s) & a_1 z_s K_1(z_s) \end{pmatrix}^{-1} \cdot \begin{pmatrix} I_0(z_s) & K_0(z_s) \\ -a_2 z_s I_1(z_s) & a_2 z_s K_1(z_s) \end{pmatrix} \begin{pmatrix} I_0(z_e) & K_0(z_e) \\ -a_2 z_e I_1(z_e) & a_2 z_e K_1(z_e) \end{pmatrix}^{-1} \quad \text{(eq. A10)}$$

but if the properties of the two regions are the same, then $a_1 = a_2 = a$, and the two matrices in the center are their own inverses. They thus cancel out, yielding:

$$\mathbf{T}_{be} = \begin{pmatrix} I_0(z_b) & K_0(z_b) \\ -a z_s I_1(z_b) & a z_s K_1(z_b) \end{pmatrix} \begin{pmatrix} I_0(z_e) & K_0(z_e) \\ -a z_e I_1(z_e) & a z_e K_1(z_e) \end{pmatrix}^{-1} \quad \text{(eq. A11)}$$

which is, of course, what we would have had for the entire region b–e having uniform properties in the first place.

Adiabatic Fin Derived from Two–Port Solution

Equation (A7) may be written for the whole domain from b to e, and also for an internal region of variable radius, out to e. Thus:

$$\mathbf{T}_{be} = \begin{pmatrix} z_e[I_1(z_e)K_0(z_b) + I_0(z_b)K_1(z_e)] & \frac{I_0(z_e)K_0(z_b) - I_0(z_b)K_0(z_e)}{2\pi kt} \\ 2\pi kt z_e z_b[I_1(z_e)K_1(z_b) - I_1(z_b)K_1(z_e)] & z_i[I_1(z_b)K_0(z_e) + I_0(z_e)K_1(z_b)] \end{pmatrix} \quad \text{(eq. A12)}$$

$$\mathbf{T}_{re} = \begin{pmatrix} z_e[I_1(z_e)K_0(z) + I_0(z)K_1(z_e)] & \frac{I_0(z_e)K_0(z) - I_0(z)K_0(z_e)}{2\pi kt} \\ 2\pi kt z_e z[I_1(z_e)K_1(z) - I_1(z)K_1(z_e)] & z_i[I_1(z)K_0(z_e) + I_0(z_e)K_1(z)] \end{pmatrix} \quad \text{(eq. A13)}$$

Similarly, applying the adiabatic condition at z_e to (A12) yields, for the temperature equation:

$$T_b - T_\infty = z_e[I_1(z_e)K_0(z_b) + I_0(z_b)K_1(z_e)](T_e - T_\infty) \quad \text{(eq. A14)}$$

and for the flow equation:

$$Q = 2\pi kt z_e z_b[I_1(z_e)K_1(z_b) - I_1(z_b)K_1(z_e)](T_e - T_\infty) \quad \text{(eq. A15)}$$

Applying the adiabatic condition at z_e to (A13) yields:

$$T(z) - T_\infty = z_e[I_1(z_e)K_0(z) + I_0(z)K_1(z_e)](T_e - T_\infty) \quad \text{(eq. A16)}$$

and:

$$q(z) = 2\pi k t z_e z [I_1(z_e)K_1(z) - I_1(z)K_1(z_e)](T_e - T_\infty) \quad \text{(eq. A17)}$$

Now the ratio of (A14) to (A16) yields the conventional adiabatic circular fin relationship between temperature at any radius in terms of the temperature rise at the base of the fin:

$$\frac{T(z)-T_\infty}{T_b-T_\infty} = \frac{z_e[I_1(z_e)K_0(z) + I_0(z)K_1(z_e)](T_e-T_\infty)}{z_e[I_1(z_e)K_0(z_b) + I_0(z_b)K_1(z_e)](T_e-T_\infty)} = \frac{I_1(z_e)K_0(z) + I_0(z)K_1(z_e)}{I_1(z_e)K_0(z_b) + I_0(z_b)K_1(z_e)} \quad \text{(eq. A18)}$$

which, with the change of variable $z_i = mr_i$, was stipulated in the body of the paper as Equation 27.

To put the local temperature rise in terms of device power dissipation, simply divide (A15) into (A16), thus:

$$\frac{T(z)-T_\infty}{Q} = \frac{z_e[I_1(z_e)K_0(z) + I_0(z)K_1(z_e)](T_e-T_\infty)}{2\pi k t z_e z_b[I_1(z_e)K_1(z_b) - I_1(z_b)K_1(z_e)](T_e-T_\infty)} = \frac{1}{2\pi k t z_b}\frac{I_1(z_e)K_0(z) + I_0(z)K_1(z_e)}{I_1(z_e)K_1(z_b) - I_1(z_b)K_1(z_e)} \quad \text{(eq. A19)}$$

constant terms may be collected as follows:

$$\frac{T(z)-T_\infty}{Q} = \frac{1}{2\pi k t z_b}\frac{\frac{I_1(z_e)}{K_1(z_e)}K_0(z) + I_0(z)}{\frac{I_1(z_e)K_1(z_b)}{K_1(z_e)} - I_1(z_b)} = c\left[\frac{I_1(z_e)}{K_1(z_e)}K_0(z) + I_0(z)\right] \quad \text{(eq. A20)}$$

where, also as used in Equation 28 in the main body:

$$c = \frac{1}{2\pi k t z_b}\frac{1}{\frac{I_1(z_e)K_1(z_b)}{K_1(z_e)} - I_1(z_b)} \quad \text{(eq. A21)}$$

Spreadsheet Implementation of Two−Port Model

Perhaps the most universally accessible, computerized mathematical tool is the Microsoft® Excel spreadsheet program. The axisymmetric thermal two−port is easily implemented using simple formulas, as suggested below. Probably the only difficulty the casual Excel user might encounter is that the modified Bessel functions are hidden away in the "Analysis Toolpak," which takes a (one time only) extra step to access. (Specifically, one goes to the **Tools: Add−Ins** menu, and checks the **Analysis Toolpak** checkbox.) Note that in the following examples, items in uppercase indicate built−in Excel functions.

First, variables may be defined with mnemonic names such as mval1, rad1, etc., according to the region of the board model in question. This obviously facilitates a parametric model, where changes in geometry (board thickness, radii of interest), and thermal properties (convection coefficients, conduction properties, etc.) are built in to these variables' own formulas.

Second, the symmetry of the transmission matrix, as seen in Equations 24 and 25, lends itself to visually "clean" coding into subgroups of four cells each, each of which is also assigned a mnemonic name, for instance:

Zone 1 Matrix at r1, "Izone1"	
=BESSELI(mval1*rad1,0)	=BESSELK(mval1*rad1,0)
=−cval1*rad1*BESSELI (mval1*rad1,1)	=−cval1*rad1*BESSELK (mval1*rad1,1)

Zone 1 Matrix at r2, "rzone1"	
=BESSELI(mval1*rad2,0)	=BESSELK(mval1*rad2,0)
=−cval1*rad2*BESSELI (mval1*rad2,1)	=−cval1*rad2*BESSELK (mval1*rad2,1)

The Excel formula to create the zone−1 transmission matrix is then:

{=MMULT(lzone1,MINVERSE(rzone1))} (eq. A22)

which must, using normal Excel methods, be properly entered comprising four cells of its own (and presumably with its own mnemonic name, say tzone1). Without going into more detail here, the expression shown as Equation A22 is called an "array formula," so indicated by the curly braces surrounding the expression. Besides the mnemonic name for the zone−1 transmission matrix, each of its four individual cells might also be named according to its position as one of the four standard transmission parameters, for instance tzone1A, tzone1B, etc.

If several zones are needed for the model (for instance, four zones are needed to compute the graph of Figures 13 and 14), each zone would have its own separate variables and cells defined in exact similarity to that shown above for zone−1. Then, the overall transmission matrix would be computed with another array formula, for instance:

{=MMULT(tzone1,MMULT(tzone2,MMULT(tzone3,tzone4)))} (eq. A23)

If desired for subsequent formulas, as would be needed for Equations 40 and 41 as an overall transmission parameter of the entire system, thus Equation 40 might be represented by:

=tzone4A/tsysC (eq. A24)

REFERENCES

1. H.H. Skilling, Electrical Networks, John Wiley and Sons, 1974.
2. L. Weinberg, Network Analysis and Synthesis, McGraw Hill Book Company, Inc., 1962.
3. E.R.G. Eckert & R.M. Drake Jr., Heat and Mass Transfer, McGraw Hill, 1959.
4. J. VanSant, Conduction Heat Transfer Solutions, Lawrence Livermore National Laboratory, Livermore, CA, 1980.
5. H.S. Carslaw & J.C. Jaeger, Conduction of Heat In Solids, Oxford Press, 1959.
6. J.P. Holman, Heat Transfer, 3rd Ed., McGraw Hill, 1972.
7. H. Wayland, Complex Variables Applied in Science and Engineering, Van Nostrand Reinhold Company.
8. EIA/JEDEC Standard JESD51−2, Integrated Circuits Thermal Test Method Environmental Conditions – Natural Convection (Still Air), Electronic Industries Alliance, December 1995.
9. EIA/JEDEC Standard JESD51−6, Integrated Circuit Thermal Test Method Environmental Conditions – Forced Convection (Moving Air), Electronic Industries Alliance, March 1999.
10. EIA/JEDEC Standard JESD51−8, Integrated Circuit Thermal Test Method Environmental Conditions – Junction−to−Board, Electronic Industries Alliance, October 1999.
11. M. Abramowitz, I. Stegun (eds), Handbook of Mathematical Functions, Dover Publications, Inc., 9th Printing, Dec. 1972.

Sales and Design Assistance from ON Semiconductor

ON Semiconductor Distribution Partners

AMSC Co.	www.amsc.co.jp	(81) 422 54 6622
Arrow Electronics	www.arrow.com	(800) 777-2776
Avnet	www.em.avnet.com	(800) 332-8638
Daiwa Distribution Ltd.	www.daiwahk.com	(852) 2341 3351
Digi-Key	www.digikey.com	(800) 344-4539
EBV Elektronik	www.ebv.com/en/locations.html	(49) 8121 774-0
Fuji Electric Co.	www.fujiele.co.jp	(81) 3 3814 1411
Future & FAI Electronics	www.futureelectronics.com/contact	1-800-FUTURE1 (388-8731)
KH Electronics Inc.	www.khelec.com/kor	(82) 42 471 8521
Marubun	www.marubun.co.jp	(81) 3 3639 5630
Mitsui Electronics Inc.	www.btel.co.jp	(81) 3 6403 5900
Mouser Electronics	www.mouser.com	(800) 346-6873
Newark/Farnell	www.farnell.com/onsemi	(800) 4-NEWARK
Promate Electronic Co.	www.promate.com.tw	(886) 2 2659 0303
Segyung Britestone Co.	www.britestone.com	(82) 2 3218 1511
Serial Microelectronics, HK	www.serialsys.com.hk	(852) 2790 8220
Taewon Inc.	www.taewon.net	(82) 2 6679 9000
Tokyo Electron Device Co.	www.teldevice.co.jp	(81) 45 443 4000
World Peace Industries Co.	www.wpi-group.com	(852) 2365 4860
WT Microelectronics Co.	www.wtmec.com	(852) 2950 0820
Yosun Electronics	www.yosun.com.tw	(886) 2 2659 8168

INTERNATIONAL

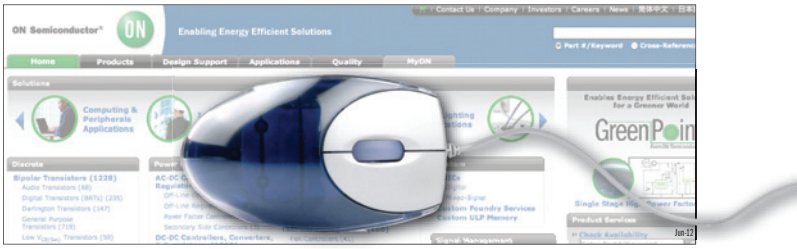
GREATER CHINA	Beijing	86-10-8577-8200
	Hong Kong	852-2689-0088
	Shenzhen	86-755-8209-1128
	Shanghai	86-21-5131-7168
	Taipei, Taiwan	886-2-2377-9911
FRANCE	Paris	33 (0)1 39-26-41-00
GERMANY	Munich	49 (0) 89-93-0808-0
INDIA	Bangalore	91-98-808-86706
ISRAEL	Raanana	972 (0) 9-9609-111
ITALY	Milan	39 02 9239311
JAPAN	Tokyo	81-3-5817-1050
KOREA	Seoul	82-2-2190-3500
MALAYSIA	Penang	60-4-6463877
SINGAPORE	Singapore	65-6442-1226
SLOVAKIA	Piestany	421 33 790 2450
UNITED KINGDOM	Slough	44 (0) 1753 70 1676

AMERICAS REP FIRMS

Alabama	Huntsville	e-Components	(256) 533-2444
Brazil	Countrywide	Ammon & Rizos	(+55) 11-4688-1960
California	Bay Area	L2	(408) 453-5000
Canada	Eastern Canada	Astec	(905) 607-1444
	Western Canada	Sifore	(503) 977-6267
Connecticut	Statewide	Paragon Electronic Systems	(603) 645-7630
Florida	Statewide	e-Components	(888) 468-2444
Georgia	Atlanta	e-Components	(888) 468-2444
Illinois	Hoffman Estates	Stan Clothier Company	(847) 781-4010
Indiana	Fishers	Bear VAI	(317) 570-0707
Iowa	Cedar Rapids	Essig & Associates	(319) 363-8703
Kansas	Overland Park	Stan Clothier Company	(913) 894-1675
Maine	Statewide	Paragon Electronic Systems	(603) 645-7630
Maryland	Columbia	Third Wave Solutions	(410) 290-5990
Massachusetts	Statewide	Paragon Electronic Systems	(603) 645-7630
Mexico	Countrywide	Ammon & Rizos	(+55) 11-4688-1960
Michigan	St. Joseph	Bear VAI	(440) 526-1991
Minnesota	Eden Prairie	Stan Clothier Company	(952) 944-3456
Missouri	St. Charles	Stan Clothier Company	(636) 916-3777
New Hampshire	Statewide	Paragon Electronic Systems	(603) 645-7630
New Jersey	Statewide	S.J. Metro	(516) 942-3232
New York	Binghamton	TriTech - Full Line Rep	(607) 722-3580
	Jericho	S.J. Metro	(516) 942-3232
	Rochester	TriTech - Full Line Rep	(585) 385-6500
North Carolina	Raleigh	e-Components	(888) 468-2444
Ohio	Brecksville	Bear VAI Technology	(440) 526-1991
Oregon	Portland	SiFore Technical	(503) 977-6267
Puerto Rico	Countrywide	e-Components	(888) 468-2444
Rhode Island	Statewide	Paragon Electronic Systems	(603) 645-7630
Vermont	Statewide	Paragon Electronic Systems	(603) 645-7630
Washington	Bellevue	SiFore Technical	(425) 990-4701
Wisconsin	Evansville	Stan Clothier Company	(608) 882-0686
	Oconomowoc	Stan Clothier Company	(608) 882-0686

For a comprehensive listing of ON Semiconductor Sales Offices, please visit:

www.onsemi.com/salessupport



PUBLICATION ORDERING INFORMATION

LITERATURE FULFILLMENT:
Literature Distribution Center for ON Semiconductor
P.O. Box 5163, Denver, Colorado 80217 USA
Phone: 303-675-2175 or 800-344-3860 Toll Free USA/Canada
Fax: 303-675-2176 or 800-344-3867 Toll Free USA/Canada
Email: orderlit@onsemi.com

N. American Technical Support: 800-282-9855 Toll Free USA/Canada.
Europe, Middle East and Africa Technical Support:
Phone: 421 33 790 2910
Japan Customer Focus Center
Phone: 81-3-5817-1050

ON Semiconductor Website: **www.onsemi.com**

Order Literature: http://www.onsemi.com/orderlit

For additional information, please contact your local Sales Representative