Understanding Power MOSFET Saturation Operation Capability

AND90187/D

A power MOSFET can be operated in four different operating modes: as a switch, called ohmic or linear[1] operating mode, in diode conduction mode, in avalanche mode, and in saturation mode. MOSFET saturation mode is illustrated in figure 1 below, which shows a power MOSFET Id vs Vds and Vgs curves (this plot displays first quadrant operation only, so diode condition is not shown and with Vds truncated so avalanche operation is not shown). MOSFET saturation mode is defined as operation with high Vds, specifically where Vds > (Vgs − Vth). In saturation operation, device power is given by Vds x Id, not by Id ^ 2 x Rds(on).

Power MOSFET saturation operation is important to understand for several reasons. First, saturation operation occurs in nearly every type of application utilizing MOSFETs. There are many applications where the application function depends on MOSFET saturation operation. These include linear voltage regulation, linear DC motor control, class A and AB audio amplification, current in−rush control, (e.g., hot swap), active clamp control for unclamped inductive loads, and airbag “safing” FET operation. And what may not be realized is that a power MOSFET operates in saturation during each switching transition, as the device Vds transits thru the Miller plateau region. Second, at high current and high Vds operation, device power is much higher than in linear switch−mode operation and therefore thermal management may be an important design consideration. Third, a MOSFET operating in saturation may be susceptible to thermal runaway, leading to device failure at a supposedly safe device power level. It is this third point which is the focus of this application note: to describe MOSFET thermal instability in saturation operation and to understand how to utilize power MOSFET data sheet safe operating (FBSOA) plot data.

Thermal stability can first be described in general terms considering a system containing some general device that dissipates power and some methodology for the system to sink heat generated by the device power dissipation. Consider the three diagrams in figure 2[2]. The green curve in each plot is the same; this represents the system capability to sink heat: the slope of the line is given by the reciprocal of the system thermal resistance, or dQ / dTj, where Q is used to represents device power and Tj represent device junction temperature. The intersection of the system line and the x−axis represents the initial device junction temperature (at Q = 0), labeled as Tx. The device line in each curve shows the device power dissipation as a function of Tj. In the first plot, at left, the device line slope is less than the system line. In this case the system can absorb more heat than the device can generates, and a stable thermal operating point is possible. In the middle plot, the device line slope is greater than the system line. In this case, device power increases faster than the system can pull heat form the device and thermal runaway results. The right plot shows a non−linear device power function, which is more likely in real world applications, where it is possible both for thermal stable operation and thermal runaway operating points.

Thermal runaway for a MOSFET almost always results in device destruction. Figure 3 shows the typical result, a burn spot on the die commonly labeled as “EOS” (electrical overstress) failure. What happens in a thermal runaway event is that the hottest spot on the die surface, determined by the physical boundary conditions of the die and package designs, such as source clip placement, reaches intrinsic temperature. At this point the hot spot area becomes a conductor and the resultant high current density increases the temperature in the area to levels that device structures begin to melt together. This results in a vertical “filament” thru the die and electrically measures as a drain to source to gate low ohmic short. Why this happens is illustrated in figure 4. This shows the transconductance curve of a power MOSFET, which is the plot of Id(sat) vs Vgs, for a device operating in saturation at some specified Vds value. Typically, curves are shown for three different junction temperatures, however some plots may include data for two junction temperatures.

One obvious attribute of this plot is that the curves representing different junction temperature values intersect at a single point. This means that Id(sat) is independent of junction temperature at some Vgs value, referred to as Vgs(ZTC), where ZTC stands for “zero temperature coefficient). The reason this happens is that the MOSFET saturation current function (equation 1) includes two main temperature dependent variables: threshold voltage, Vth, and mobility, μ_0. The general forms of mobility and threshold as functions of junction temperature, Tj, are listed as equations 2 and 3. For a MOSFET in saturation at higher Vgs, the mobility term dominates and at lower Vgs, the delta Vgs to Vth term dominates, creating the crossing transconductance functions with respect to temperature.

1. In practice, engineers often refer to saturation operation as “linear mode”; this because they consider the MOSFET operation in an application, such as linear regulator function, not the operating mode of the MOSFET device itself.
2. Refer to **onsemi** application note AND8223/D “Predicting Thermal Runaway” for complete discussion on this topic.

January, 2023 − Rev. 0
Publication Order Number:
AND90187/D

Another way to visualize this is shown in figure 5, which plots Id(sat) as function of Tj for increasing Vgs values. In this example, Id(sat) is constant with respect to Tj at Vgs(ZTC) = ~5.2 V. At Vgs values above this, Id(sat) has a negative temperature coefficient with respect to Tj and for Vgs values lower the Vgs(ZTC) the Id(sat) temperature coefficient is positive.

$$Idsat \cong \frac{\mu_o \cdot Cox \cdot W}{2 \cdot L} \cdot (Vgs - Vth)^2 \quad \text{(eq. 1)}$$

$$\mu(T) = \mu_o \cdot \left(\frac{T}{T_o}\right) - \gamma \quad \text{(eq. 2)}$$

$$Vth(T) = Vth_o \cdot (1 - m \cdot (T - T_o)) \quad \text{(eq. 3)}$$

It is when the Id(sat) temperature coefficient is positive that thermal runaway is possible. Referring to the general discussion about thermal instability, if the device power is increasing as Tj increases at a rate greater than the system can sink the added heat, the device junction temperature will runaway and device failure results. This can only happen if the device is operating in saturation and operating at Vgs values below Vgs(ZTC), where the Idsat temperature coefficient is positive. Simply said thermal runaway can occur when the device generates more power (Pgen) than the system can dissipate (Pdis). Equations 4 and 5 define Pgen and Pdis respectively and if we define thermal instability where Pgen > Pdis, then this function can be described by equation 6. From this equation it is clear thermal instability can occur only if the Id(sat) temperature coefficient is positive.

$$Pgen = Idsat(T) \cdot Vds \quad \text{(eq. 4)}$$

$$Pdis = dT / r(t) \quad \text{(eq. 5)}$$

$$Pgen > Pdis \rightarrow Idsat(T) \cdot Vds > dT / r(t) \rightarrow$$

$$\rightarrow \frac{Idsat(T)}{dT} \cdot Vds > 1 / r(t) \quad \text{(eq. 6)}$$

Thermal instability can be visualized if we measure dTj as a function of increasing power for a power MOSFET device operated in saturation. If a device is operating with a stable thermal operating point, then dTj as a function of power should measure as a linear function. That is, we expect dTj = Pd / r(t). One method to make such a plot is to measure a device in saturation at a constant Vds value and increment current, measuring dTj response for some given pulse duration. We then plot dTj as a function of Id and expect a linear response if the device is not in thermal instability region. Figure 6 shows two examples of such plots each at different Vds values and different pulse durations. In both plots we can observe initially (at lower Id values) the linear thermal response expected. However, at some Id (i.e., power) value, the dTj response increases much more than expected, indicating the onset of thermal instability operation. In the right−hand plot, the device reached failure; one can observe the rapid dTj/Id for the measurements just prior to failure — this is thermal runaway.

The physical mechanism of thermal runaway can be observed by taking device to failure and then perform physical analysis of the failed units. As discussed earlier, the power MOSFET, if subject to thermal runaway failure, will fail at a position that correlates to the hottest spot on the die. This spot (note that some refer to thermal runaway as “hot−spotting failure) is determined by the physical construction of the device. Refer to figure 7 which shows two different constructions using the same package (SO8FL) and same silicon die. The only difference between the two is the source clip design. The hottest spot on the die correlates to the area near that of highest front (surface) metal current density. In the left design, there is a larger section of die active area not covered by the source clip compared to the right design. These devices were taken to failure in saturation operation (thermal runaway). The failure location on the die shifts to the left for the smaller clip design, correlating to an area near highest front metal current density. This result shows that thermal instability is not only function of die technology and electrical attributes, but also the thermomechanical design of the total device.

To measure saturation capability requires a circuit like that blocked out in figure 8. The circuit needs to drive the MOSFET such that a constant drain current, Id, and constant drain to source voltage, Vds, is achieved for the full pulse width. A standard current sink circuit is utilized to control both Id and pulse duration. A separate bias circuit forces constant Vds whilst affording gate to source voltage, Vgs, to vary during the pulse duration as needed to maintain the constant Id and Vds target values. Such circuitry is available commercially, although maximum allowed power and minimum pulse duration can be limiting to a full characterization for the device. In these cases, manual bench top set−ups are employed. Standard characterization methodology is to set a Vds value, and increment Id until failure, and repeat for multiple Vds values and pulse durations. The collection of Id(fail) values for a given pulse duration can be fitted to a function. This technique is illustrated in the right−hand side of figure 8. The Id(fail) data typically fits to a power function of the form: Id = constant x $Vds^{-\alpha}$, where the exponent magnitude, α, is >1. This function clearly indicates that device power capability decreases significantly as Vds increases, another indication of the thermal instability effect (refer to equation 6).

Although pulsed saturation capability can ideally be measured for any pulse duration, including up to steady state operation, there are practical limitation to consider. First, as apparent from equation 6, the application thermal system affects saturation operation capability. For typical power MOSFET packages and associated application thermal management (PCB and heat sink design), the system external thermal boundaries only appreciably affect heat flow for pulse duration greater than approximately 10 ms. Therefore, especially on more recent datasheets, pulsed saturation capability is published for pulse durations ≤10 ms. Note that if longer pulse durations are specified, the

associated external thermal boundary conditions for the device in the system environment should be specified. On the short pulse duration side, it is difficult to measure higher current capability at pulse durations less than ~300 ms on commercial equipment, and perhaps <~100 ms on manual setups. The issue is the current waveform becomes trapezoidal in shape at shorter pulse duration and high current, due to stray inductance in the test circuit and slew rate control to avoid significant current overshoot.

Figure 9 displays measured pulsed saturation capability (pulse duration = 1 ms in this example) comparing planar technology to second, third, and fourth generation trench technologies. The y−axis is current density (normalized to die active area) so that multiple technology and die area data can be included in the same plot. Also shown on the plot is a normalized constant power curve, which is the expected safe power capability of the device for a 1 ms duration power pulse, assuming a given dTj (in this example dTj = 150°C). There are two important observations to consider on this plot. One, is that the power function of planar is less steep compared to trench technology data (lower exponent value). Because of this, the other important consideration is that the planar Id(fail) data remains greater than the calculated safe thermal capability, whereas the trench technology Id(fail) data measure less than the safe calculated thermal capability at higher Vds voltages. This is an important observation when discussing FBSOA plot formation later in this paper.

Power MOSFET pulsed saturation capability is included in the data sheet FBSOA (forward bias safe operating area) plot (refer to figure 10). The FBSOA plot, included on MOSFET data sheets since production inception in late 70s to early 80s, is based on the power bipolar transistor active region SOA data sheet plots. These plots were designed to show device wire bond collector current capability, thermally limited power capability, and power limitation due to secondary breakdown. The power MOSFET FBSOA plot is designed for similar information, including pulsed current limit (in ohmic mode), absolute maximum Vds operation, thermal limit in saturation operation, and saturation operation limited by thermal instability. The power MOSFET FBSOA plot also includes limitation due to Rds(on), which is simply the slope on the curve that represents the maximum Rds(on) value at maximum junction temperature. Note that the safe operating area for ohmic operation (i.e., limited by Rds(on)) is *above* the Rds(on) maximum line on the plot).

If we focus on the pulsed saturation portion of the FBSOA plot curves, we find that the plot indicates limitation due to thermal capability and limitation due to thermal instability. That is, the pulsed saturation capability for a given pulse duration is a curve with a bend or knee in the Id, Vds operating space. This was not always the case. Historically, the pulsed saturation capability for a power MOSFET device was simply the calculated safe constant power capability. This was because for early MOSFET planar technologies and in some case first generation trench technologies, the pulsed saturation capability of such devices measured greater (by sufficient margin) than the calculated safe thermal capability (refer to figure 9) for the devices. Therefore, publishing the safe calculated thermal capability was sufficient for the FBSOA plot. As later generation, much higher gain trench MOSFET technologies were introduced it was found that measured saturation capability fell below the safe calculated thermal capability for devices. Therefore, empirical pulsed saturation data had to be included in the FBSOA plot.

To form a pulsed saturation capability curve on the FBSOA plot the safe calculated thermal capability is determined. Then measured Id(fail) data is collected and then de−rated. Both sets of data when plotted in the Id, Vds space, will intersect. It is this intersection of the calculated constant power curve and the de−rated measured Id(fail) data that forms the curve “knee”. Thus, the “knee” indicated in a pulsed saturation capability curve has no physical meaning (refer to figure 10).

The power MOSFET FBSOA plot has limits to its utilization in real world application analysis. Most obvious is the FBSOA plot, for practical reasons, provides data for only a few pulse durations. A user can at least understand a range of pulsed saturation capability for some device, but if data for a specific pulse duration is required, it must be measured. Note that power MOSFET thermal instability behavior is difficult to physically model and such behavior is not included in device simulation models. Less obvious is that the power MOSFET saturation power function for most real−world applications is not square−wave, yet the FBSOA plot provides data only for a square−wave power function.

Figure 11 shows expected power functions for real−world situations where a power MOSFET is operated in saturation, compared to the standard square−wave power function utilized in device characterization. These include active clamp inductive load switch−off, in−rush current control, and inductive load hard switched transitions (e.g., in some DC/DC power supply applications). Most real−world application saturation operation results in triangular power functions. In general, these functions have greater SOA compared to the square−wave function. Thus, the data sheet FBSOA plot data could be considered as worst−case analysis for most applications. However, application analysis based on data sheet FBSOA plot may result in a conservative device selection. If possible, it is better practice to compare actual device performance in the application circuit and operating conditions to make an optimal MOSFET selection.

Power MOSFET device saturation capability often must be considered in application fault analysis. That is, standard thermal analysis of the fault mode operation may indicate no expected issue with MOSFET device performance, but in practical testing the device fails. This is commonly due to unexpected thermal runaway failure. Consider a short circuit fault, where the MOSFET switch is connected through a low ohmic path either to battery positive or battery

negative (depending on lead configuration) resulting in much greater drain current (and associated power) whilst the device remains activated, and the short fault persists. In modern systems, circuits are designed so that a short circuit fault is recognized, and action taken to switch off the MOSFET switch or switches often occurs within 5−10 μs, and in some cases less than 3 μs. Thus, even though instantaneous power may be quite large (perhaps tens of kilowatts), the energy is low (due to short fault duration) and in general failure of the MOSFET due to pure thermal consideration does not occur.

Consider figure 11, which shows oscilloscope traces for a short circuit fault condition (of duration = ~12 μs) for a second−generation trench technology device. In this application example the MOSFET is configured low side, so the fault is a short to battery. Important to consider in this case is that the gate drive pull down resistance is quite large, at ~400 ohms; this results in the relatively slow switch−off observed in the scope traces. The first observation is the device survives the 12 μs short circuit pulse at high drain voltage and current, where power peaked at nearly 10 kW. The second observation is that the device does eventually fail, indicated by the rapid increase of drain current off scale and the corresponding rapid decrease of drain voltage (the device shorts out). The further and key observation is that the failure occurs *well* after the device is deactivated, when gate voltage is at or close to zero volts and device power is close to zero. The slow switch−off requires the device to operate in saturation for some time, and before the device can completely shut−off (where Id = 0), the conditions are met such that the device goes into thermal runaway and fails short. In this example, another run was made with the pull−down resistance reduced significantly to 10 ohms. In this case, for the same short circuit fault, the device survived. This is because the time duration in saturation at switch−off was significantly reduced.

A power MOSFET operates to some degree in saturation in nearly every application function. A power MOSFET may also unexpectedly operate in saturation in response to some system fault event. If the power MOSFET is operating in saturation with Vgs voltage < Vgs (ZTC) voltage, then depending on other operating conditions the device may suffer thermal runway and fail. For this reason, it is important to understand pulsed saturation capability characterization and associated limitations to typical published FBSOA data.

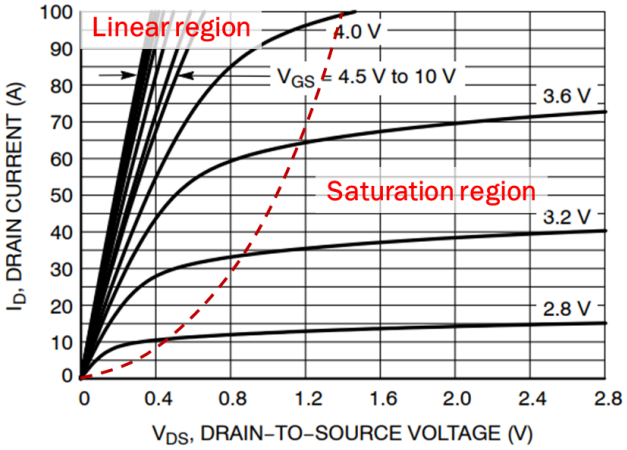

Figure 1. Typical Power MOSFET Forward Biased Id, Vds Function Showing the Delineation of the Linear and Saturation Operation Regions

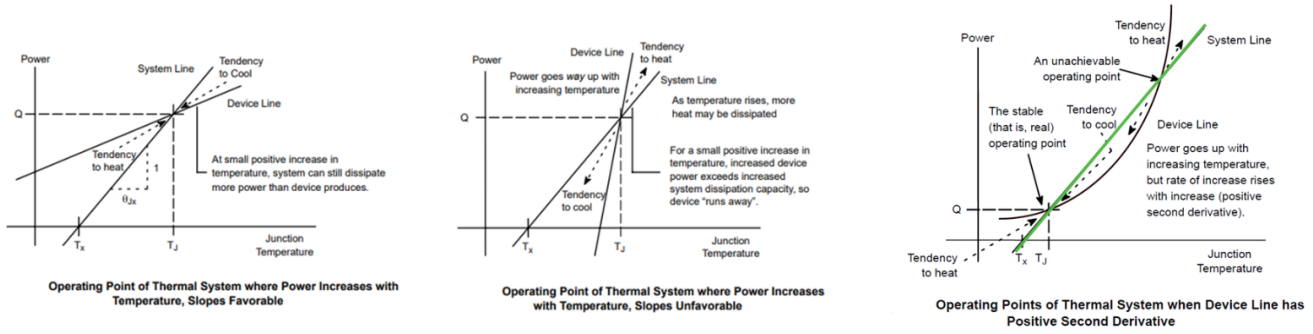

Figure 2. General Representation of Device Power as Function Temperature and System Capability to Sink Power and Relationship to Stable or Unstable (Thermal Instability) Operating Points

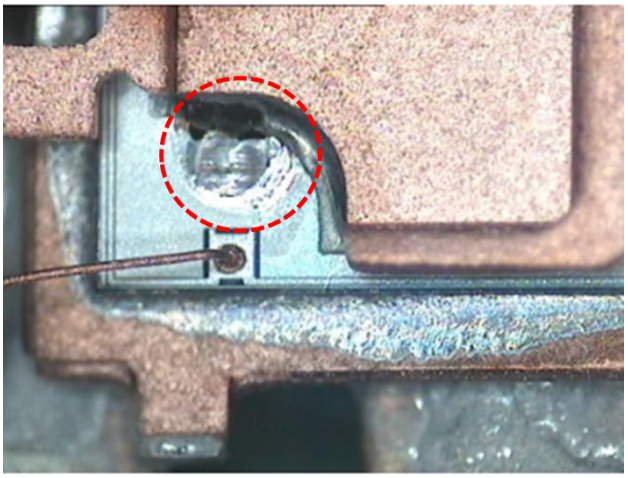

Figure 3. Photograph of a Power MOSFET Device (after De−capsulation) that Failed During Saturation Operation (Thermal Runaway). The Failure Site is Circled in Red.

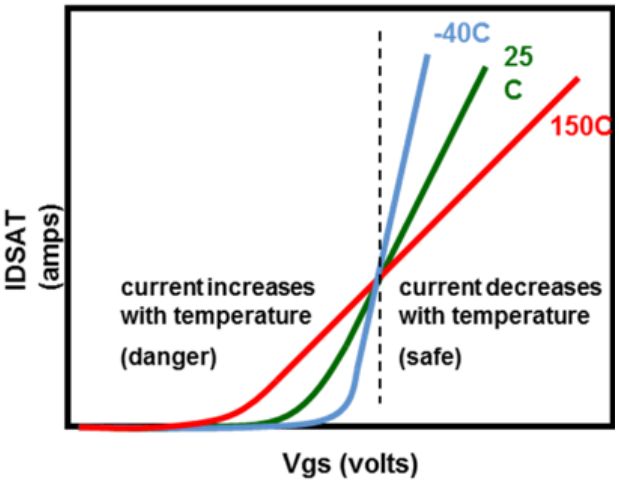

Figure 4. A Generic Representation of Power MOSFET Transconductance Curves at Different Junction Temperatures. The Vgs Value Where the Curves Intersect is Known as Vgs(ZTC).

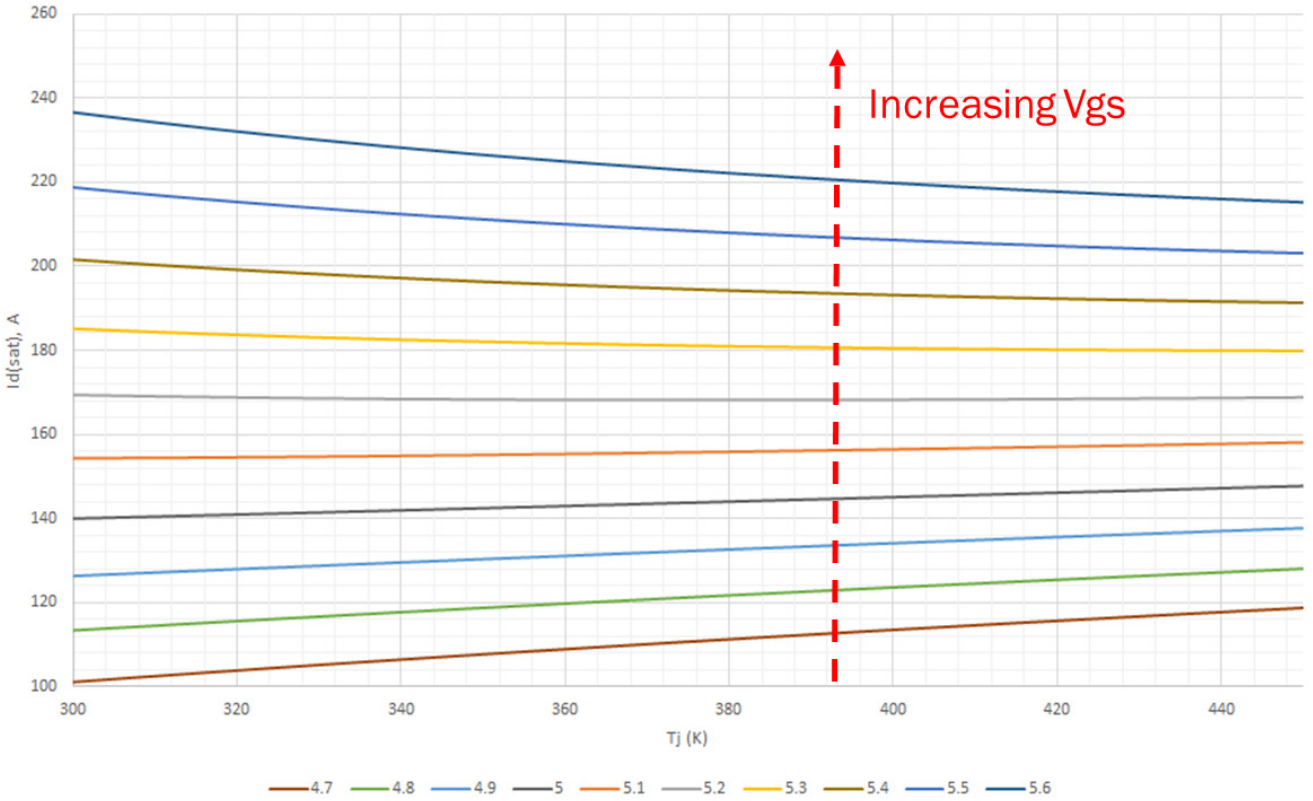

Figure 5. An Example Data Set Showing Idsat as a Function of Temperature for Increasing Vgs Bias. This is Another Representation of Vgs(ZTC), Which in This Case is at Vgs = ~5.2 V, Where Idsat is Independent of Temperature.

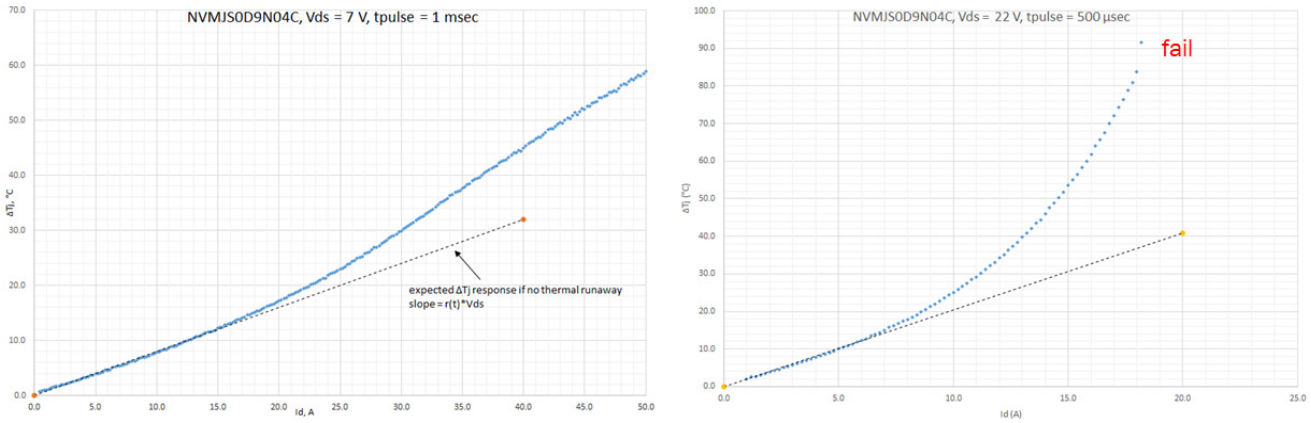

Figure 6. Visualization of Thermal Instability Behavior. Initially Delta Tj Increases Linearly with Applied Power, but as the Device Enters Thermal Instability Operation, Delta Tj Increases More Rapidly Than the Expected Linear Response.

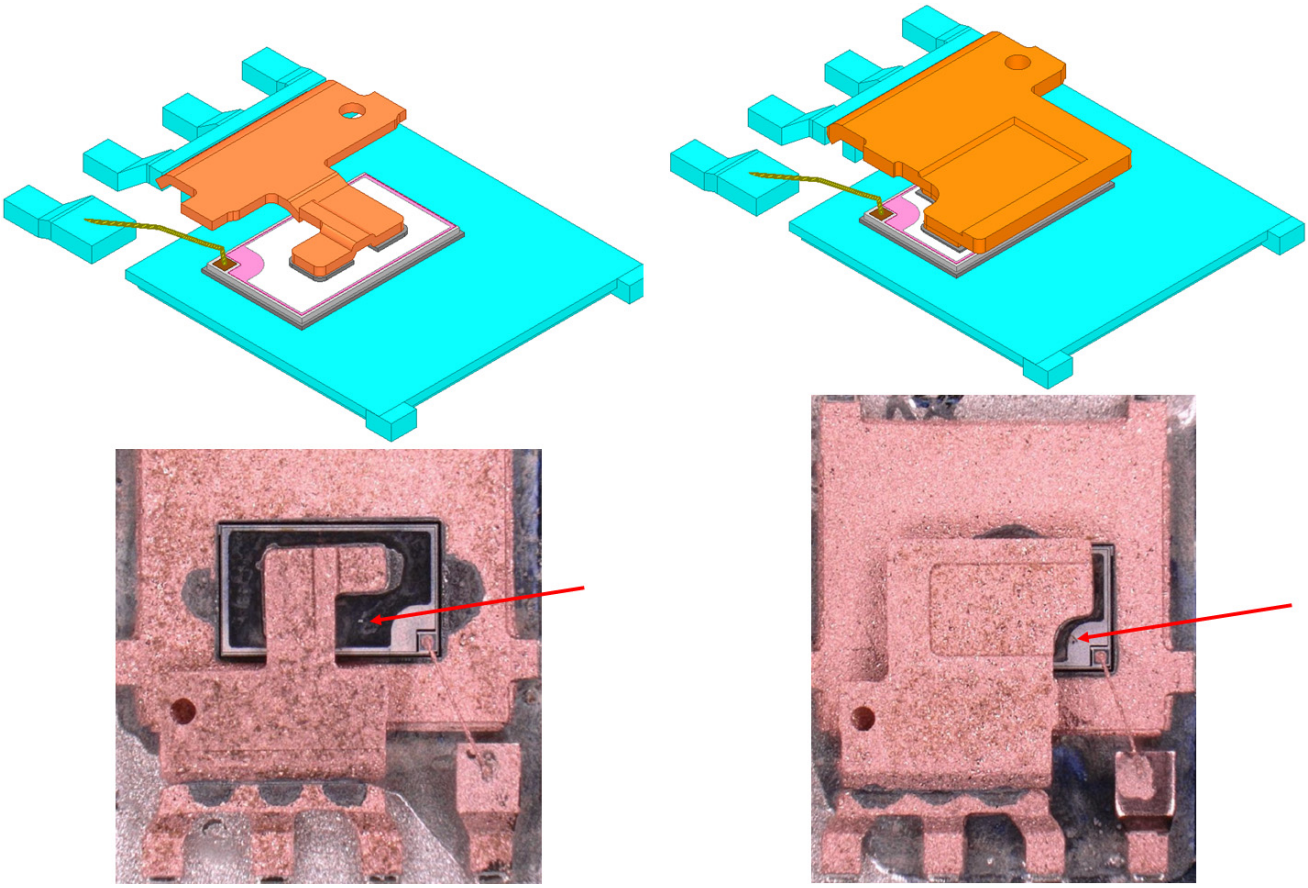

Figure 7. Comparison of Thermal Runaway Failure Site Location (Indicated by Red Arrows) for Two Device Utilizing Same Die and Same Package but Different Source Clip Designs

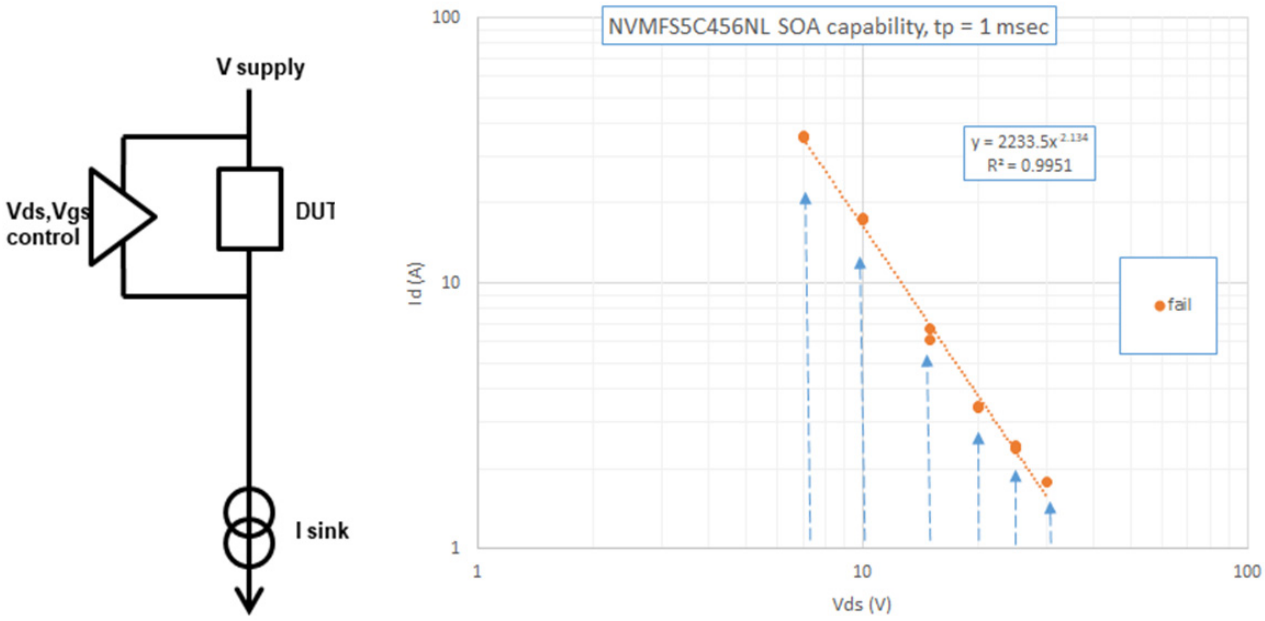

Figure 8. Left Side Shows the General Block Diagram for a Circuit to Evaluate Power MOSFET Pulsed Saturation Operation Capability. The Right Plot Illustrates the Procedure: For a Fixed Pulse Duration, Id is Incremented to Failure Point for Different Vds Values.

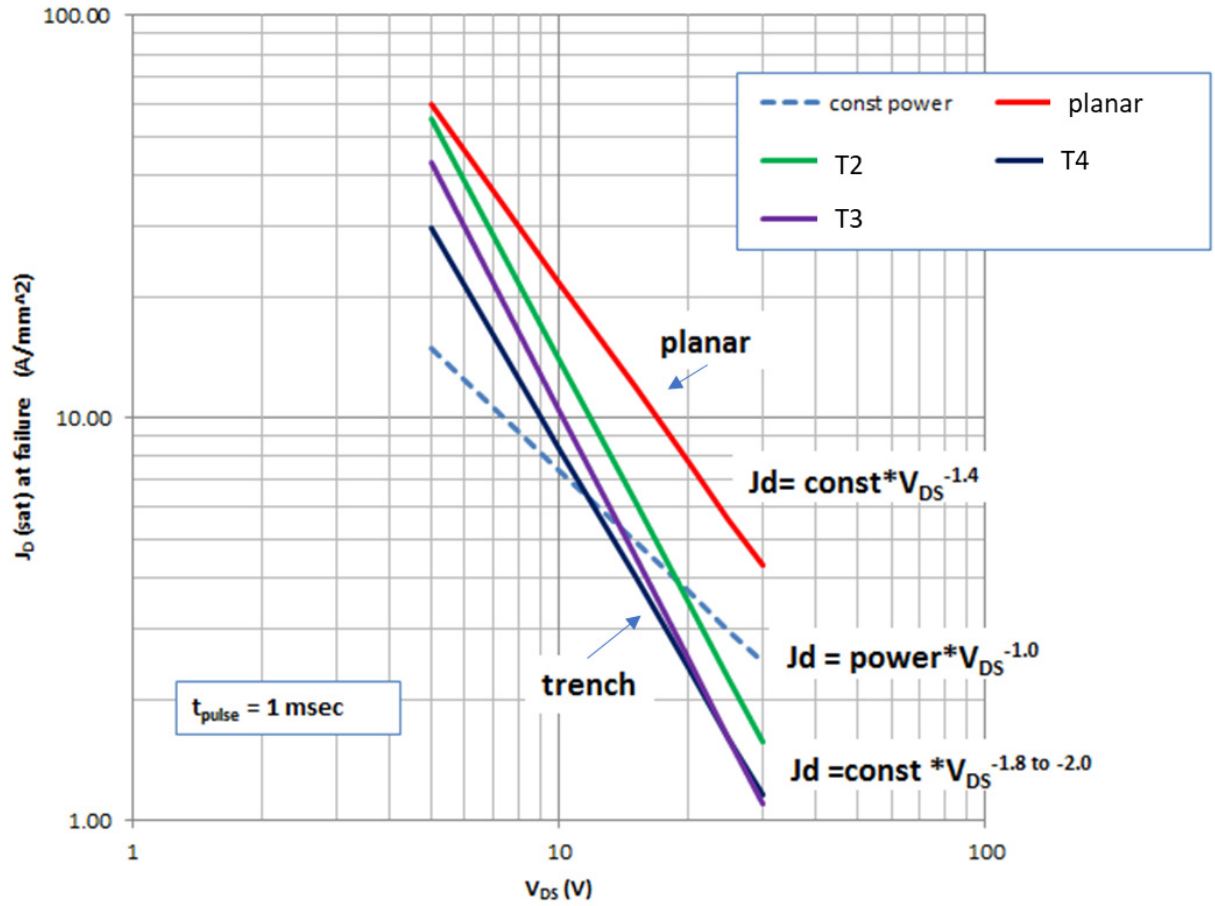

Figure 9. Typical Planar and Trench Technology Pulsed Saturation Capability (at Failure) Compared to Calculated Safe Thermal Capability (Constant Power Given by dTj / r(t))

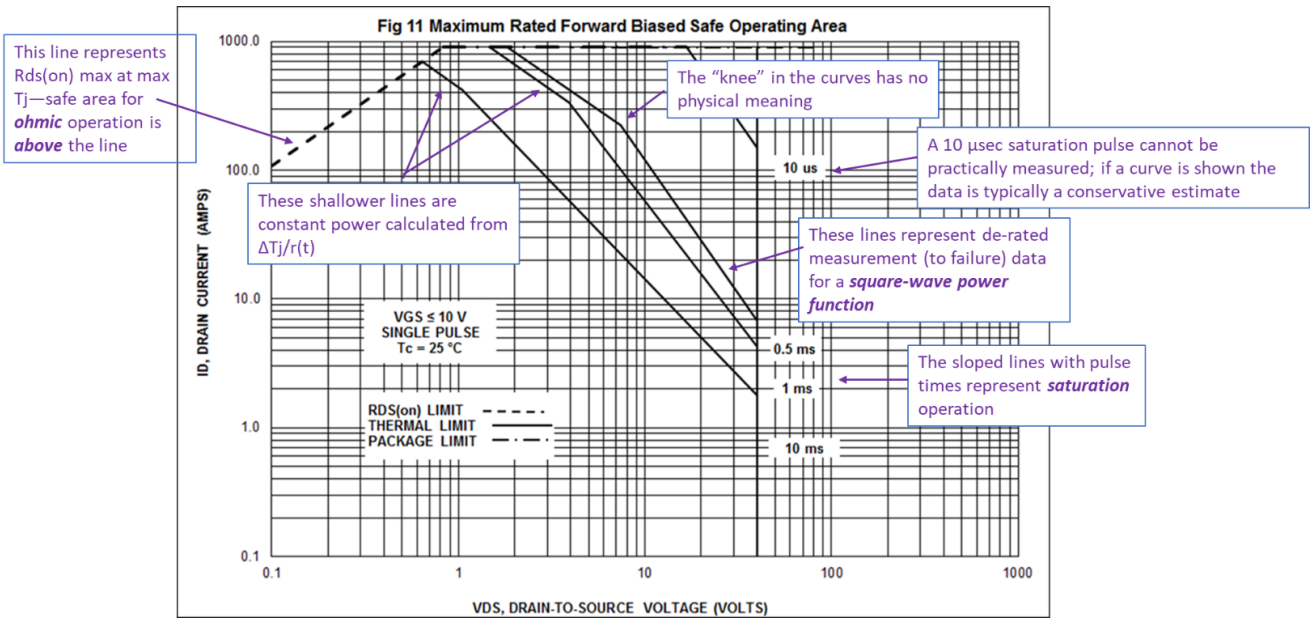

Figure 10. A Typical Power MOSFET FBSOA Plot

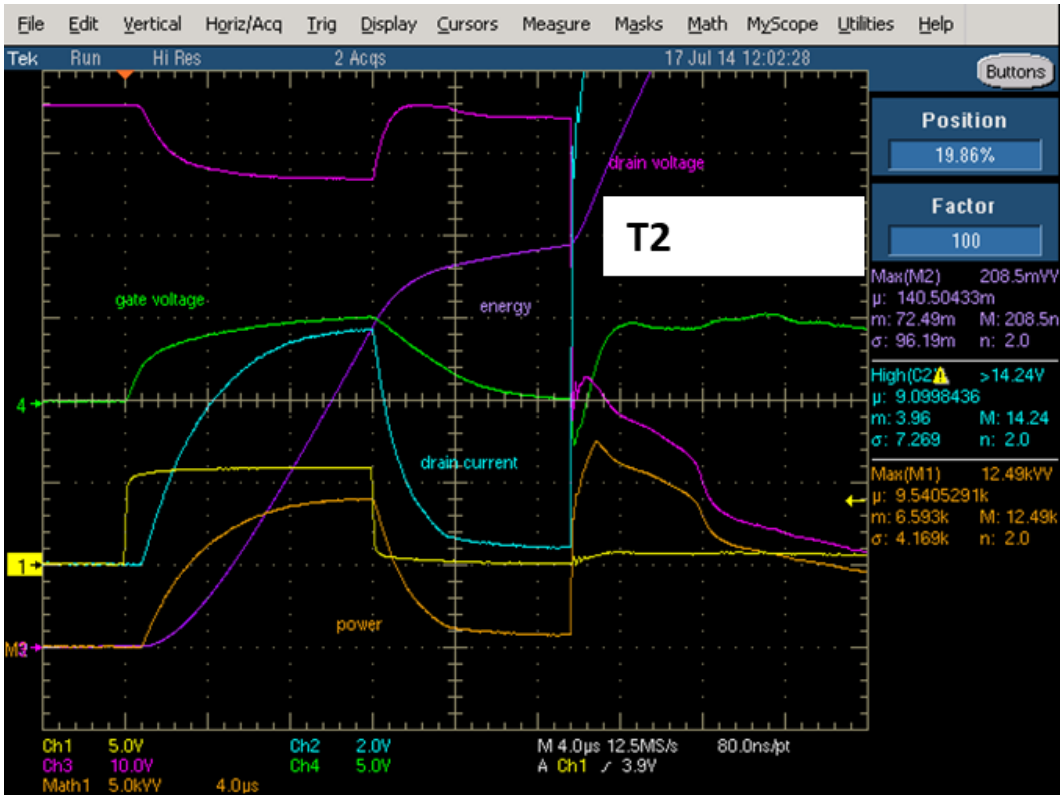

Figure 11. Scope Traces for a Low−side Power MOSFET Subjected to Low Ohmic Short to Battery Positive. Note the Device Fails Well after Gate is Commanded Off in Response to the Fault.

PUBLICATION ORDERING INFORMATION

LITERATURE FULFILLMENT:
Email Requests to: orderlit@onsemi.com

onsemi Website: www.onsemi.com

TECHNICAL SUPPORT
North American Technical Support:
Voice Mail: 1 800−282−9855 Toll Free USA/Canada
Phone: 011 421 33 790 2910

Europe, Middle East and Africa Technical Support:
Phone: 00421 33 790 2910
For additional information, please contact your local Sales Representative

◊