Loughborough University

01/05/2018

A discussion into the thermal failure modes of MOSFETs

Janith Thanthulage

[redacted]

Wolfson School of Mechanical, Electrical and Manufacturing Engineering

1. Abstract

During the acceptance testing of a CCA (circuit card assembly) in production, several failures were found. In all instances of failure, the same MOSFET part showed signs of damage which immediately indicated a possible sign of weakness in the design. Initially, it was thought that the test itself induced the failure. However, on reviewing the parts that had failed, all cases revealed that the parts had failed from thermal stress. This report explores the issues that were uncovered.

[redacted]

[redacted]

Table of Contents

2. Introduction

Metal oxide semiconductor field effect transistors (MOSFETs) are widely used in in-flight control systems. In the development of modern power MOSFETs, manufacturers have focused on the speed of switching and a continued reduction in on-resistances.

This was largely achieved through the utilization of a greatly reduced silicon die area for any given on-resistance across the drain and source (R_{DS}). Such devices can show exceptional performance in 'high power and/or high switching frequency systems (the lower the drain to source resistance, the quicker the switching between 'ON' and 'OFF' states), maximizing overall efficiency'.[2]

In the case of the [redacted] design, [redacted] n-channel MOSFET was chosen for its incredible temperature range of [redacted] in combination with its desirable transfer characteristics as seen in Figure 1.

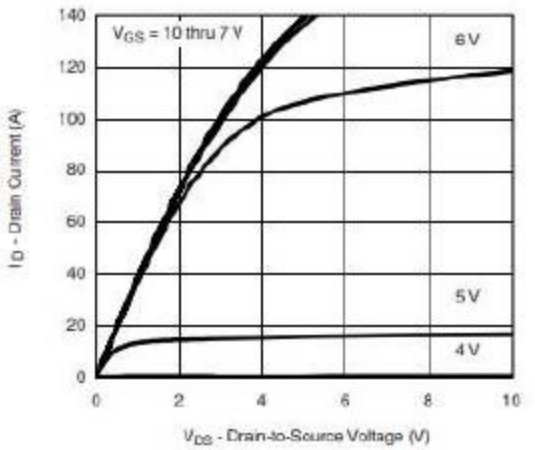

Figure 1 – [redacted] Output Transfer Characteristics

The purpose of this MOSFET was to control the level of current being circulated through the circuit via the drain-source terminals by providing a single pulse of DC (direct current) feedback through a chain of OPAMPS (inside the control circuit) to its gate, see Figure 2. This forced the MOSFET to operate solely in its linear region at a controlled drain to source voltage for the length of the pulse.

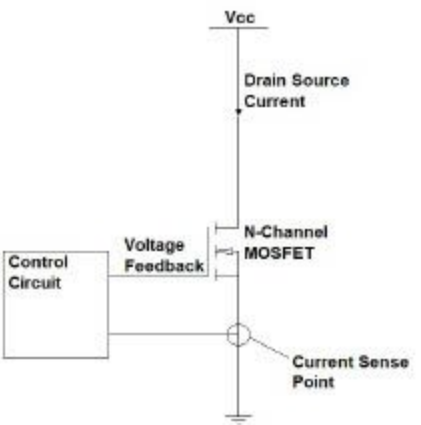

Figure 2 – Simplified MOSFET current control circuit

Its low on resistance of [redacted] and its relatively low threshold voltage [redacted] ensured a very fast switching time between OFF and ON states to be achieved with low power usage at the gate source terminal. The temperature range of the device was crucial as to meet customer requirements. The transfer characteristics ensured that the circuit was more efficient and able to function at a low generated power. Its on resistance ensured minimum timing delays between input and output

signals which was yet again a crucial factor in order to meet customer and system requirements. The above, in combination with its small size of a maximum [redacted]mm x [redacted]mm (which allowed it to easily fit onto a CCA, see Figure 3) and made the [redacted] a very desirable component.

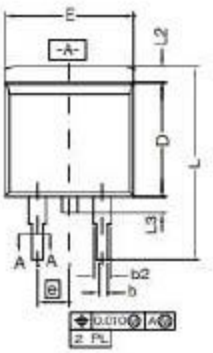

Figure 3 – Package Information Diagram of [redacted]

[redacted]

Figure 4 – Package Sizing Information [redacted]

In all simulations and analysis of the circuit with the [redacted] it was found that the operating region of the component was well within that of the specified Safe Operating Area (SOA) stated in the datasheet.

After implementation of the device in the CCA and entering production, all production level tests conducted indicated no fault in the circuit and a match between expected and test results. However, throughout production of this circuit in 2017, nine out of a total of twenty-four units failed at test. Noticeably, all of these CCAs had failed at the cold cycle of testing [redacted] No failures were captured at the [redacted] cycles of testing. At the time of discovery, the reason for this was not known, however, Section 4.6.4 discusses a likely factor. The CCA manufacturer responsible

for conducting the production level tests identified the [redacted] MOSFET to be the failed component across all failed tests. It was observed that failed devices were exhibiting a low impedance (VDS), instead of an open circuit, when VGS = 0V. This was confirmed by replacing the faulty device and retesting the CCA.

An investigation into this problem was then launched.

3. Investigation into the continued failure of [redacted]

Initially thought to be due to a recent change in test rig setup, previously accepted units were retested and returned very similar results to when tested on the previous rig setup. Additionally to this, all failed CCAs were re tested across multiple test rig setups and continued to show the same failure. The rig was also thoroughly re-characterised and matched what was expected. All of this in combination, thoroughly eliminated any possibility that the test rig was at fault.

Multiple hypothesis to the cause of the failure were then conceived:

- Failures as a result of electrostatic discharge
- Failures as a result of purchased batch problems
- Failures as a result of component material variation

Electrostatic discharge, in the context of circuit design, is the discharge of current across two charged components which then create a short between the devices. However, this possibility was eliminated on finding the same fault across multiple consequent retests of the same CCA as well as after replacing the part on multiple occasions.

The possibility of a batch problem was quickly disregarded due to the same failure found throughout the year across multiple batches. This in combination with finding that all other instances of the device used on the CCA (outside of the control circuit), functioned as expected eliminated the possibility of a manufacturer batch problem.

It was also noticed that once a CCA had passed all tests, on repetition of the same tests, the accepted CCA and MOSFET continued to pass and function as expected. In no instance of a passing unit being retested was a failure found. This lead to the hypothesis that there may be a variation in the MOSFET material rather than a degradation of the part as a result of tests that were conducted.

The investigation was then brought to the attention of the manufacturer whom then conducted further analysis on multiple instances of the [redacted] This included a close external inspection and an x-ray inspection of the devices, see Figure 5, Figure 6 and Figure 7.

Figure 5 – External inspection of the top of the device package with de soldered leads

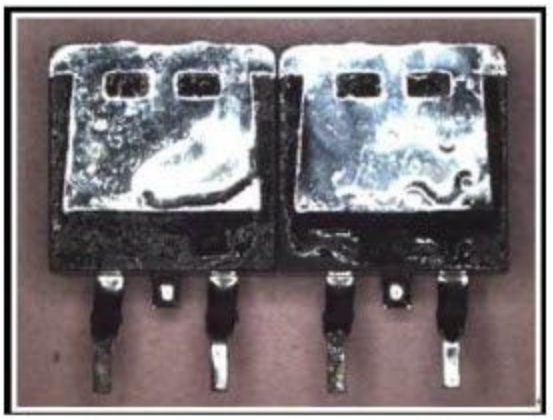

Figure 6 – External inspection of the bottom of the device package with de soldered leads

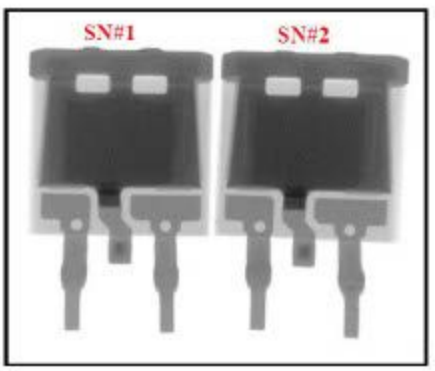

Figure 7 – X-ray inspection of device package showing no obvious anomalies

Both inspections of the component showed no obvious anomaly in the part. Consequently, these parts were then used in the CCA on which they went to both suffer failures. On returning these parts (after failure) to the manufacturer for re-analysis, electrically induced physical damage (EIPD) was observed as is clearly seen in Figure 8.

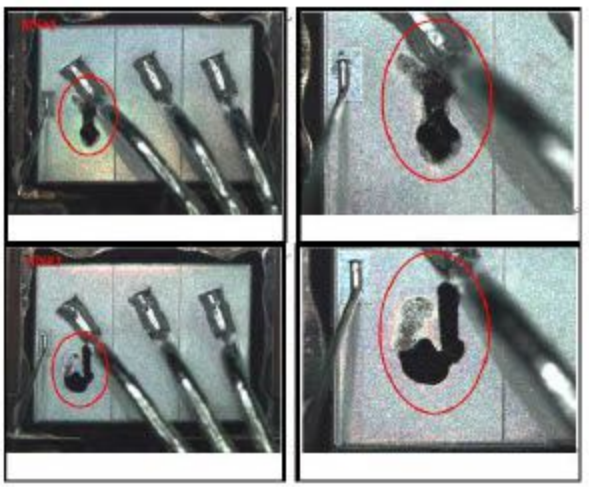

Figure 8 – Electrically induced physical damage to both components

Note: Figure 8 is of the device with its case having been de-soldered.

A further six passed MOSFETs from units that had passed were then sent to the manufacturer for further inspection. All were confirmed to have no damage associated with them. The manufacturer had suggested that this variation in results from the device may be a result of a variation in the die. Figure 9 shows the X-rays of a device that had passed production level testing where a variation in the die can be seen.

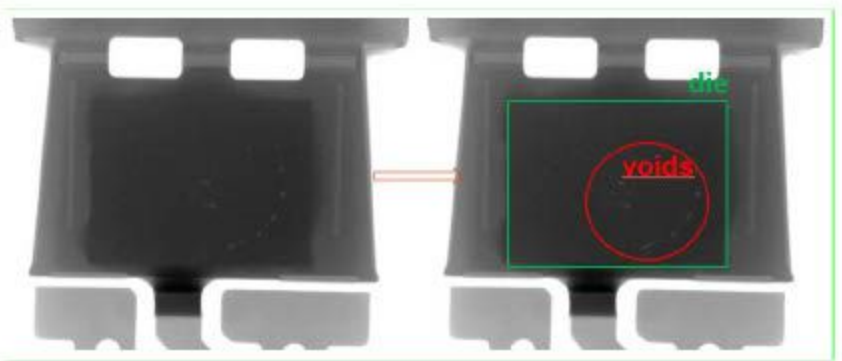

Figure 9 – X ray of [redacted] MOSFET illustrating voiding within the die of the device

With regards to failing within the advertised safe operating area of [redacted] MOSFET, it was made known that the curve had been derived through theoretical calculations for power dissipation and thermal resistance assuming switch mode operation. This is in contrast to [redacted] operation of the device in linear mode for which the manufacturer stated the device may display signs of thermal instability.

4. The Thermal Instability Phenomenon

The "thermal instability" phenomenon is the event where the generated electrical power exceeds the power that is dissipated resulting in a scenario where the component is no longer able to sustain a pulse. [2] The operation of a MOSFET in its linear mode of operation, in combination with variation in MOSFET die quality, can lead to this 'thermal instability' or 'thermal runaway'.[4]

4.1. Modes of Operation

Two key modes of operation exist in a MOSFET's. A MOSFET can either function in switch mode operation (where it is used as a basic switch to transfer between an 'OFF' and 'ON' state) or in its linear mode of operation (where it is used as a variable current device to both control and limit the current that is fed into a circuit). The region of a MOSFET's on-state characteristic one operates within, directly relates to the mode of operation being acted within. This can be seen illustrated in Figure 10. The two regions are as follows:

- Ohmic region
- Saturation region

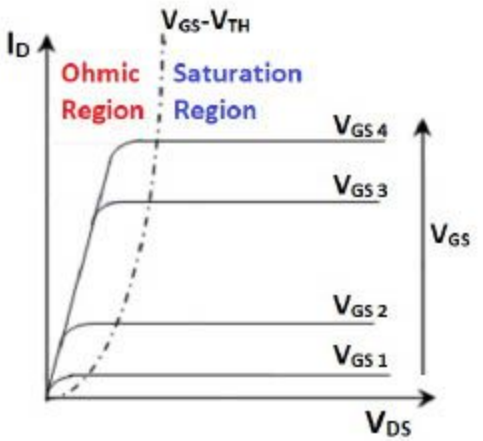

Figure 10 – General Illustration of Ohmic & Saturation Region of MOSFET characteristic

When acting as a fast transient switch (switch mode application), a higher gate-source voltage is usually applied to force it directly into its 'ON' state as quickly as possible and then removed as to switch it into its 'OFF' state. In such an application, a steeper ohmic region is extremely sought after. A steeper ohmic region (as a result of a smaller R_{DS} gradient) would lead to less 'time' a MOSFET would spend 'reaching' saturation to 'turn ON' or 'leaving' saturation to 'turn OFF', and would also reduce losses in doing so. However, when acting as a current control, a smaller gate-source voltage can be applied to force the MOSFET to operate almost solely in saturation. Due to this a smaller ohmic region is less absolutely necessary in most cases as the switching speed of the MOSFET is not usually of the most concern in such applications.

It can be noticed that the output characteristic appears linear at higher gate voltages as the MOSFET is 'fully on'; whereas at a lower gate voltages the drain current tends to shows saturation as:

$$V_{DS} > V_{GS} - V_{TH}$$

where,

- V_{DS} is the voltage across the drain and source of the MOSFET
- V_{GS} is the voltage across the gate and source of the MOSFET
- V_{TH} is the threshold voltage of the MOSFET

Note: the threshold voltage is defined as 'the minimum gate bias which can form a conducting channel between the source and drain.' [5]

4.2. Ohmic Region

When operating in this region, there is a linear relationship between a MOSFET's drain current (I_D) and its drain to source voltage (V_{DS}).

$$I_D = \frac{V_{DS}}{R_{DS-on}}$$

As demonstrated by Figure 10, as the gate voltage (V_G) becomes greater than the threshold voltage (V_{TH}), the MOSFET begins to conduct current across its drain and source terminals. The amount of current conducted then depends on the on-resistance [5] (as defined in the equation above).

4.3. Saturation Region

A MOSFET's linear mode of operation refers to the operation of a MOSFET in its current saturation region. During its saturation region, it can be seen how I_{DS} is almost entirely independent of V_{DS}, and instead is dependant directly on that of its gate-source voltage (V_{GS}). This allows the possibility to carry very high currents across a MOSFET through adjusting its gate-source voltage (V_{GS}) alone. This feature is reason why MOSFETs prove invaluable as a method of current control through using feedback from a MOSFET's drain back to its gate.

$$I_{DS} = K \cdot (V_{GS} - V_{TH})^2$$

For

$$V_{DS} > V_{GS} - V_{TH}$$

'where K is a parameter depending on the temperature and device geometry'. [3]

However, for MOSFETs acting in their saturation region, the power dissipated is much higher than what is seen when acting in its ohmic region due to the current being conducted across the component terminals. It is essential, in such cases, that the Safe Operating Area (SOA) graphs are consulted to define 'the maximum drain-source voltage (V_{DS}) and drain-source current (I_{DS}) for correct functioning of the device' [2].

The time duration a device acts within this region is also important: essentially its 'on' pulse width. Figure 11 in Section 4.4 illustrates how the power capability of a MOSFET device is directly impacted by the length of a pulse.

4.4. Application of a Safe Operating Area diagram (SOA) and its Limitations

The Safe Operating Area (SOA) diagram is crucial to review when implementing a MOSFET in a design. Forward Bias Safe Operating Area (FBSOA) graphs define the maximum value of drain to source current and voltage which would guarantee a component's safe operation when in forward bias. [5] Reverse Bias Safe Operating Area (RBSOA) graphs define this, but in reverse bias.

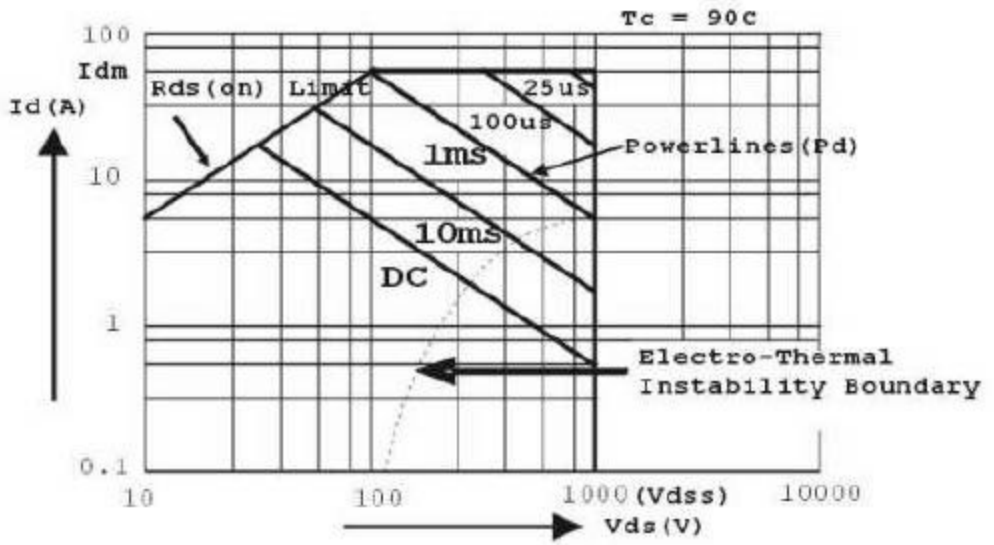

Figure 11 – Typical FBSOA graph for an N-channel Power MOSFET [3]

As illustrated in the image, it is clear how each line has been derived. The top of each line represents the pulse width. As previously stated, it can be seen how this directly impacts the area in which the component can operate. The vertical line represents the drain to source voltage limits as would be defined for each component. The 'powerlines' represents the 'maximum allowed power dissipation of the device' [3].

However as described in Section 3, the advertised SOA curves have often been calculated theoretically (in most cases from the thermal impedance curve of the device) rather than through a distribution of measured results. This often neglects the possibility of thermal runaway and can pose an issue when designing to operate in linear mode. Many manufacturer's have begun to include a boundary (typically located towards the bottom right corner) within the SOA as a restriction posed on the component when operating in its linear mode. In Figure 11, this area can be seen labelled as the 'Electro-Thermal Instability Boundary'.

4.5. Contribution of die size and die doping variation to thermal instability

It was suggested by the manufacturer that the variation in die concentration within the individual devices may be reason for the failure of some components but not others. This would also serve reasoning for components continuing to pass after having already passed tests. Through thorough retesting of passed units, as described in Section 3, enough confidence was had to ensure that a full recall of shipped units would not have to be made.

As MOSFET development has continued, the speed of switching has increased alongside a considerable reduction in on-resistance. This has largely been facilitated by a reduction in MOSFET

die area for a given on resistance. This has consequently also affected the thermal resistance of devices.

The thermal resistance of a MOSFET (R_{th}) expresses the heat transferring ability of the device: the lower the thermal resistance, the improved dissipation of heat from the component die. MOSFET Thermal resistance is known to be dependent on several factors such as area and thickness such that,

$$R_{th} \propto \frac{L}{k \cdot A}$$

where,

- k is a generic constant
- L is the thickness of the material
- A is the area of the material

Clearly, as the MOSFET die area has decreased with the development of modern MOSFETs, the thermal resistance has increased directly impeding the ability of the MOSFET to dissipate heat.

Additionally to this, material imperfections such as die voids, thermal grease cavities, gate oxide thickness and a variation in the concentration of dopants [3][2] will contribute to the local decrease of thermal conductivity. This will therefore create a temperature profile revealing small areas of the component die with a significant temperature increase relative to neighbouring die cells. This then leads to a temperature profile that is not consistent throughout a device's case and leads to the formation of 'hot spots'. A more accurate distribution of temperature across a MOSFET's die is illustrated in Figure 12. While this problem may begin for the entire component, as this thermal runaway continues, with time, the hot spot becomes smaller in size. With more power now concentrated in a smaller area, the temperature rise seen is greater and faster: the smaller the hot spot, the higher the resulting temperatures from the failure of the component.

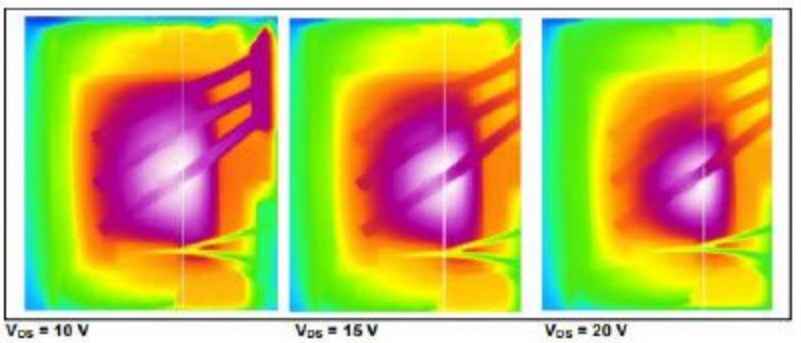

Figure 12 – Die Temperatures at different V_{DS} levels [2]

Disregarding the temperature of the three probes (as a result of the testing conducted in Ref [2]), it can be seen how the edge of the MOSFET die soldered to the mounting tab has a generally lower temperature compared to the centre of the die. [3]

The impact of voltage across the drain source terminals is immediately noticeable. The increasing of V_{DS} causes the temperature distribution at die level to become increasingly less uniform, with a clear focus in a very small die area.

It can be seen how the non-uniformity in die process may generate localized hot spots which then cause the eventual failure of the MOSFET device. This is largely due to the design of the MOSFET chip itself. 'The MOSFET chip consists of a large number of identical cells (20 to 30 million cells per square inch)'[8] connected in parallel, see Figure 13. The HEXFET is a trademark of power MOSFETs developed by International Rectifier (IR) and its illustration very clearly indicates the basic structure of a MOSFET. It can be seen how the failure of a single cell can therefore cause the failure of the entire MOSFET part as is seen in hot spot related failures where the short created by its development leads to the shorting of the entire component.

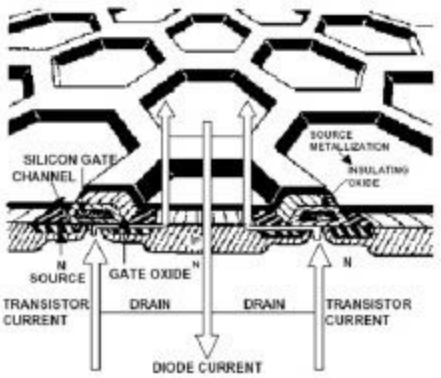

Figure 13 – The HEXFET Structure [10]

In switched mode operation, the variations in MOSFET material tend to have minimal impact. However, as seen, MOSFETs operated in their linear mode where the duration of the pulse is greater than the time it would take for heat to transfer from the junction of the MOSFET to its heat sink would be far more likely to cause damage to the component. [3] This directly relates to the "thermal instability" phenomenon: when the generated electrical power exceeds the power that the component can dissipate, the component is no longer able to sustain the pulse. [2]

In combination with evidence provided from the manufacturer (Figure 9), which indicated possible voiding in the [redacted] MOSFET die, and the impact that this variation can have (as discussed throughout this section, Section 4.5), [redacted] was able to conclude this to be a major point of reasoning for the variation in test results seen at production level testing.

4.6. Impact of Drain Current Thermal Coefficient on thermal instability

The transfer characteristic of a Power MOSFET shows the drain to source current (I_D) as a function of gate to source voltage (V_{GS}) at a fixed junction temperate (T_J). Using this curve, it is possible to characterise the thermal coefficient of drain current and hence avoid thermal instability.

4.6.1. The Zero Temperature Coefficient Point

Figure 14 illustrates the transfer characteristic of a Power MOSFET at three junction temperatures: -55 °C, 25 °C and 150 °C. It illustrates the drain to source current (I_{DS}) as a function of the voltage across the gate-source (V_{GS}) at fixed junction temperatures (T_J). The point at which they cross is commonly known as the Zero Temperature Coefficient (ZTC) point.

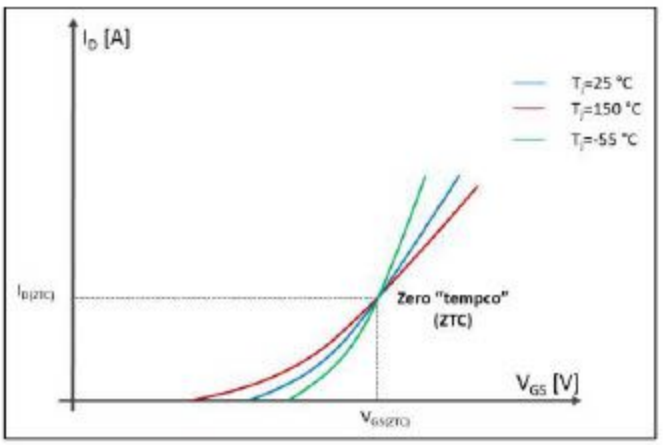

Figure 14 – General Transfer Characteristic of Power MOSFET [2]

Notice the negative temperature coefficient of the threshold voltage, $\frac{\Delta V_{GS}}{\Delta T}$:

- For $V_{GS} = V_{GS(ZTC)}$, the MOSFET's drain to source current will remain stable across all temperatures as can be seen in Figure 14.
- For $V_{GS} > V_{GS(ZTC)}$, the drain to source current of the device decreases with temperature and reaching a condition for thermal stability at that temperature.
- For $V_{GS} < V_{GS(ZTC)}$, the drain current thermal coefficient is positive and indicates a region of thermal instability. MOSFETs with smaller gain changes between temperatures here will minimise the chance of failure. The impact of a positive drain current thermal coefficient is further discussed in more detail in Section 4.6.2 below.

4.6.2. Positive Drain Current Thermal Coefficient, dI_D/dt

A positive drain current thermal coefficient suggests that as temperature increases, so does the components ability to conduct more current. However, in conducting more current, inevitably more power dissipation and consequently heat will be dissipated. Therefore, if a component was to operate at a positive drain current thermal coefficient, then as a small die zone becomes hotter than the adjacent zones, it conducts more drain current and creates more heat. This allows more current to flow due to the lower threshold voltage (see Section 4.6.4). This die area can then conduct sizeable amounts of current which, if appropriate limitations have not been implemented, will push the component to failure and cause thermal runaway.

The lower the Zero Temperature Coefficient (ZTC) point, the smaller the positive temperature coefficient area which in effect improves a device's thermal stability. [2] In addition to this it is also key that the difference in current between temperatures at a set drain-source voltage is minimal as to reduce the rate of increase in current as temperature rises. This was a key learning point taken by [redacted] in the process of finding an alternate device for the [redacted], see Section 5.

4.6.3. Relationship to MOSFET transconductance

There has been extensive analysis on the relationship between ZTC and MOSFET transconductance (gfs) when regarding MOSFET technologies and silicon characteristics. It has been found that the greater the gfs, the greater the ZTC.[6] However, as previously discussed, the greater the ZTC, the greater the area for thermal instability which consequently increases the duration that the MOSFET will operate in this unstable region during its transitions between 'OFF' and 'ON' states.

Modern Power MOSFETs, which show increasingly excellent performance in high switching-frequency and high-power applications, utilise larger and larger gfs values. This inevitably makes such modern MOSFETs less robust to the thermal runaway or hot spot phenomena for the reasons discussed.

4.6.4. Positive to Negative Drain Current Thermal Coefficient

The change of temperature coefficient from positive to negative over V_{GS} is caused by two competing effects: the behaviour of MOSFET resistance and threshold voltage over temperature.

The on resistance of a MOSFET increases as temperature increases due to a lower electron mobility in the crystal lattice[1] as further supported by Figure 15.

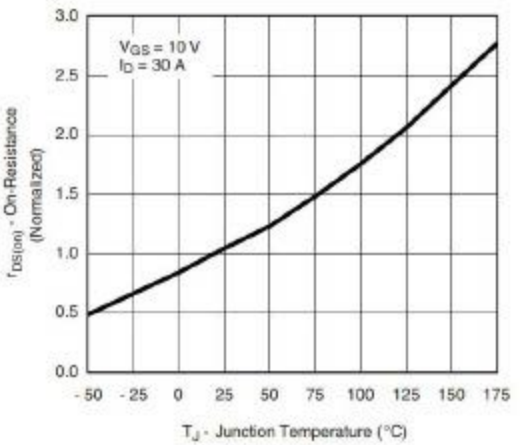

Figure 15 – On-Resistance vs Junction Temperature for t[redacted] MOSFET

The threshold voltage however, decreases as temperature rises. This is due to 'more electrons being excited into the conduction band of the MOSFET'[1].

At low temperatures, the effect of the threshold voltage decreasing over temperature will lead the current to increase as temperature increases. However, at higher temperatures, the effect of on resistance increasing over temperature will lead the source to drain current to fall over temperature.

With this characteristic of V_{th}, another factor in the failure case [redacted] realised. As all failed tests conducted by the CCA manufacturer had failed at cold, it is possible that the reduction in threshold voltage, V_{th} was sufficient enough as to not allow the component to conduct as much as under normal conditions. This would then force the component to act at a lower operating point on its output characteristic (I_D-V_{GS}) which consequently may force the component below the Zero Temperature Coefficient and into thermal instability.

5. Process in finding a replacement

Given the effect of thermal runaway on the success of the CCA, a replacement component was needed to be found. When in the process of finding a replacement, it was important to create as minimal change to the design as possible: both electrically and physically. In preserving the electrical characteristics, value changes to other components would not have to be made to account for a change the larger circuit that the MOSFET is implemented in. In preserving the physical characteristics (size and package type of the device), a re design of the tracking on the PCB (printed circuit board) of the CCA would not be needed and the alternative could directly be used on the

existing PCB layout. In doing this, a recharacterization of the CCA's weight, centre of mass and vibration stability would not be needed as the weight distribution would be preserved to when using the [redacted] MOSFET. This would then consequently ensure that the system design would not suffer further changes to account for a change in the individual CCA and hence save resources in staff, time and money in an otherwise very lengthy redesign process.

It needed to fulfil the same requirements concerning the temperature range and maximum drain to source voltage/current ratings. However, it was essential that this alternative have a far lower ZTC point relative to that of [redacted] MOSFET to avoid the same risk of thermal runaway that would lead to the same failures. This was to ensure that the area for thermal instability, where the drain current thermal coefficient is positive, is as small as possible whilst conserving the previously stated requirements. Furthermore, it was also important that the alternative be the same size as the [redacted] of the same package [redacted]. This was again in order to ensure minimum change to the design.

Ultimately, the [redacted] MOSFET was chosen. The [redacted] characteristics were incredibly similar to that of the [redacted] as can be seen in Table 1. Most notably, it had a on resistance and temperature range of [redacted] respectively. Whilst the difference in Drain-Source Voltage maximum rating is substantial, the change in its limit was decided to not have a significant impact on the application of the design. In comparison to the [redacted] the change was thought minimal in terms of its impact on the design. Additionally, the physical package and size of the alternative was the same as the [redacted] which would mean a mechanical recharacterization of the CCA would not be required. Amongst these crucial similarities however was an equally crucial difference. Consulting the Input Admittance transfer characteristic ($I_D - V_{GS}$ graph), it was immediately clear how the [redacted] had a far lower ZTC point in comparison to the [redacted]. The difference in current between characteristic curves at different temperatures can be seen to be far smaller with the [redacted] device than that of the [redacted] device which would also mean a much less dramatic effect of being within the region of thermal instability.

Table 1 – Maximum ratings of most note for the [redacted]

Parameter	[redacted]
Drain-Source Voltage	
Gate Source Voltage	
Continuous Drain Current	
Pulsed Drain Current	
Drain-Source on-state resistance	
Maximum Power Dissipation	
Operating Temperature	

The [redacted] had met all of the discussed constraints that had existed when choosing a suitable alternative to the [redacted]. To then introduce this change into the design, simulations of the circuit at worst case conditions were first conducted as was theoretical tolerance analysis. Simulation of the worst case conditions constituted running the simulation all combinations of the temperature extremes, voltage extremes and tolerance extremes of all component characteristics (threshold voltage, resistances, capacitance etc.).

In simulation of the device at worst case conditions, the electrical outputs of the CCA had remained been within 5% as in the previous design confirming there had been minimal change to the design as far as the datasheet had conveyed. In the theoretical analysis of the modification, the electrical output of the CCA had varied very little from the simulation which added to the confidence in both methods of analysis. The evidence from theoretical and simulation results, which demonstrated little variance from the original design, then gave enough confidence to transition to the testing phase.

To confirm there were no changes between designs that were not captured by the datasheet and simulation models upon which both cases of analysis relied upon, several existing CCAs were modified to have the alternative MOSFET installed and then the production tests re-conducted. Every CCA with this change had passed production level testing with good margin and consequently gave [redacted] assurance to introduce this change into their design.

6. Conclusion

This issue of thermal instability, under a MOSFETs operation in linear mode, has been well known to the automotive industry since 1997[7] due to the early introduction of fast switching MOSFET devices as they became available. As this problem emerged, numerous articles were published acknowledging the issue. However, a response from manufacturers have been slow to record this issue in datasheets and application notes for their numerous parts. Hence many parts still exist where this is not acknowledged in their respective datasheets. The aerospace industry, due to reliability concerns, have adopted these more modern MOSFETs later and have as a result begun to also encounter these problems[redacted]. [redacted] thermal runaway was found to have multiple contributing factors associated to it such as the effect of a MOSFET die's doping concentration and surface area. However, as has been discussed, the possibility for thermal runaway (when operating in a MOSFETs linear mode) can most easily be negated by paying further attention to a MOSFETs zero temperature coefficient point and choosing one with a low positive drain current thermal coefficient and zero temperature coefficient point[redacted].

7. References

References
[redacted]