Power Electronics Semiconductor Devices
Edited by Robert Perret

Chapter 1

Power MOSFET Transistors

1.1. Introduction

Before 1930, and thus a long time before the origins of the semiconductor transistor, J.E. Lilienfield was granted a patent for an electrostatic effect device allowing a current control like a MOSFET function. Thanks to planar technology, MM Khang and Atalla discovered in June 1960 the metal oxide semiconductor (MOS) structure, shown in Figure 1.1. This immediately provided the possibility of building:

– integrated circuits;

– large input impedance amplifier circuits;

– high frequency amplifiers.

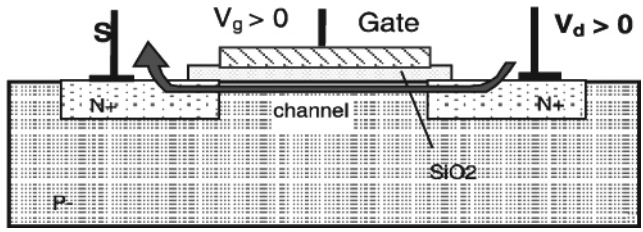


Figure 1.1. *Theoretical structure of a planar MOS*

Chapter written by Pierre ALOÏSI.

However, this planar structure is not able to simultaneously meet the following demands by power applications:

– high voltage operation,

– high current control.

Various research groups tried to improve this technology. The first trials to obtain a high voltage power MOSFET were based on improvements to lateral structures; see Figure 1.2. Due to the technological limitations of lateral structures for electrical field and current density, it was obvious that the vertical structure was the correct technology.

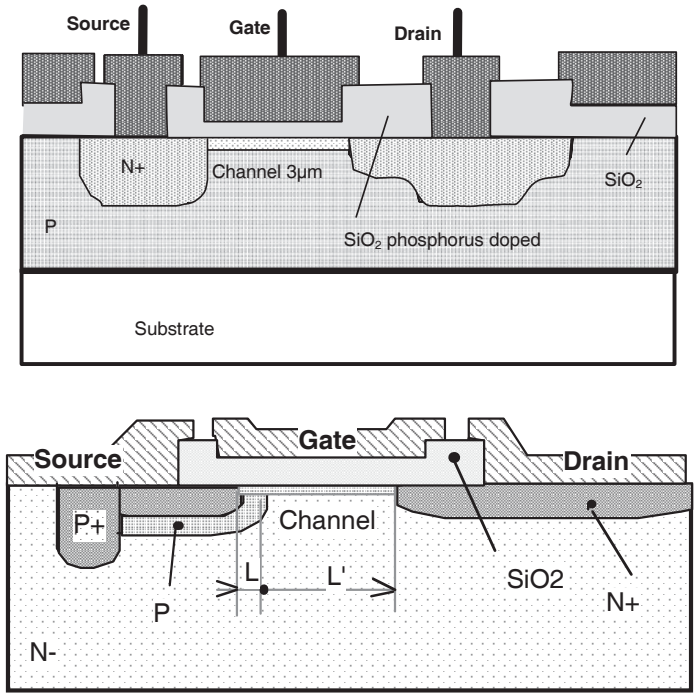


Figure 1.2. *DAWSON and lateral DMOS structures*

Removing the drain electrode from the silicon surface, the current density can be increased and the electrical field of the device is now independent of the channel length.

The first vertical structure was built around a V groove, or a truncated V, etched into the silicon using an anisotropic chemical, as seen in Figure 1.3a for the VVMOS and Figure 1.3b for the VUMOS. However, due to the high electric field at the point of the V or truncated V, this technology was replaced by a new VDMOS technology, with a double diffused channel vertical structure, as seen in Figure 1.4.

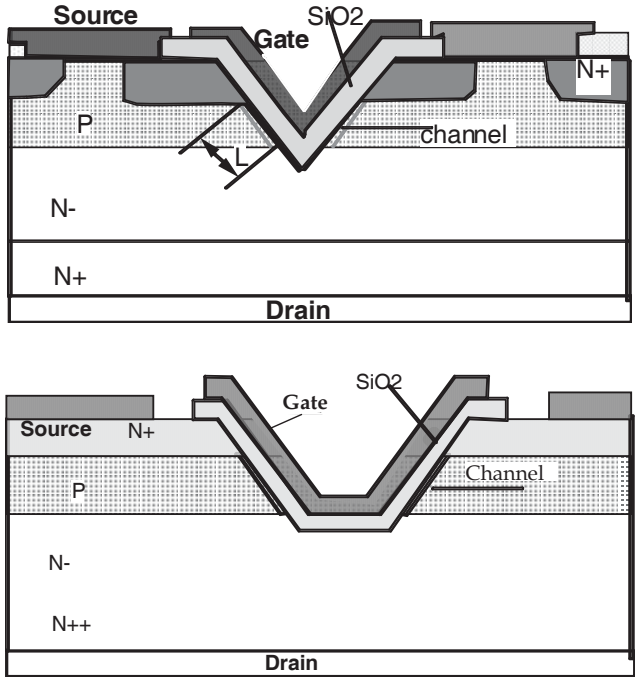


Figure 1.3. *VVMOS and VUMOS*

This vertical structure allows us to build the gate over the drift zone N-. This way, the die is also metallized, as well as the field plate, and enables a better thermal resistance.

Using a simple diffusion, including several masks, the polysilicon gate is built in order to make the source windows self-align, and to find a better compromise between the blocking voltage V_{dss} and the ON resistance R_{dson}.

The drawbacks are a higher interconnection capacitance and a bigger resistance for gate access, which increases the switching times and decreases the frequency performance.

The first designer of this P device was Yoshida from Hitachi in 1976, followed in 1979 by International Rectifier with the N-channel HEXFET, Siemens with the SIPMOS and Motorola with the TMOS.

Today, the VDMOS structure remains just about the same as the initial device, with a few adjustments.

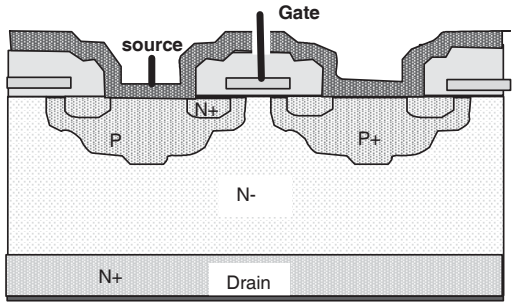


Figure 1.4. *VDMOS structure*

MOSFETs are unipolar devices where current transportation is carried out by majority carriers (electrons or holes). Thus, the expected current modulation from stored charges – which is the main phenomenon of a bipolar device – does not occur. Therefore, switching speeds are very fast and independent of the temperature. They are limited only by internal capacitances, which are charged and discharged when the device is turned on and off. The temperature coefficient of the internal resistance, R_{dson}, is positive: this leads to an easy paralleling. While bipolar transistors are driven by quite a large base current, the high input impedance of MOSFET allows a low energy gate drive. MOSFETs have a very good robustness during overloads, and the lack of secondary breakdown allows a large safe operating area. However, all of these good characteristics are in contrast with the large internal resistance of medium and high voltage devices.

Technological progress has allowed a large market for MOSFETs, mainly in low voltage segments of the market such as the automotive or telecoms industry. New technologies, such as “Superjunction”, increased the possibilities for medium voltage segments like the domestic market (240 V AC), thanks to internal resistances in the range of 0.3 Ω in 500 V devices, encapsulated in standard epoxy packages.

1.2. Power MOSFET technologies

1.2.1. *Diffusion process*

VDMOS structure is made with a lot of cells. Each of them is like a tiny independent parallel connected MOSFET.

The process of making them starts with a N+ substrate (doped at 10^{19} cm^{-3}), which is the drain. Then an epitaxial layer is grown (doped from 10^{14} cm^{-3} to 10^{16} cm^{-3}), which becomes the drift region where each cell is produced. For that, three diffused layers are produced: a P well (doped at 10^{16} cm^{-3}), then a P+ for the channel (doping 10^{18} cm^{-3}) and two N+ (doping 10^{19} cm^{-3}) for the source. To make a P-channel MOSFET, change all P and N regions to their opposites. The gate starts with an oxide that is very well controlled in thickness (around 100 nm) over which a polycrystalline silicon is laid down. The gate is over two P adjacent cells (channels) with the N- drift region in between, at the end a new oxide is deposited in order to insulate the gate from the source metallization, this becomes a link between all the cells and recovers all the device; only a gate pad is produced inside. The cell shape can be square or hexagonal, as seen in Figure 1.5.

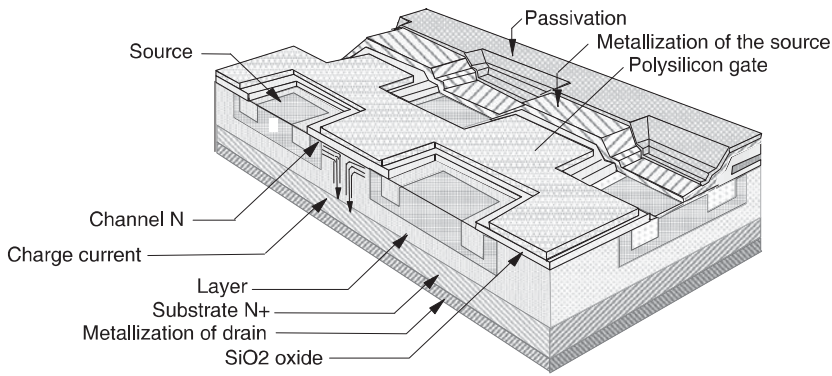


Figure 1.5. *Cellular MOSFET*

The main process steps are carried out as follows (see Figure 1.6):

– an epitaxial layer is grown on a 300 μm thick N+ wafer. Its thickness depends on the MOS voltage;

– a thick silicon dioxide SiO_2 is deposited over the die in which cell windows are opened to diffuse the P well and N+ source;

– etching and P+ channel implantation;

– thick oxide is removed except for the periphery, gate oxide is grown and polycrystalline silicon is deposited for gate metallization;

– gate oxide and polysilicon gate are etched to open cell windows. Boron for the P well is implanted and driven to make all the well. Thick oxide is grown on the die;

– cells window is again opened on the oxide and N+ sources are diffused;

– polysilicon gate is insulated by a SiO_2 deposition, the gate pad is opened for connection;

– source metallization over the die, contact pads for source and gate are opened;

– oxide is spread over the die for insulation. Metallization of drain back side occurs.

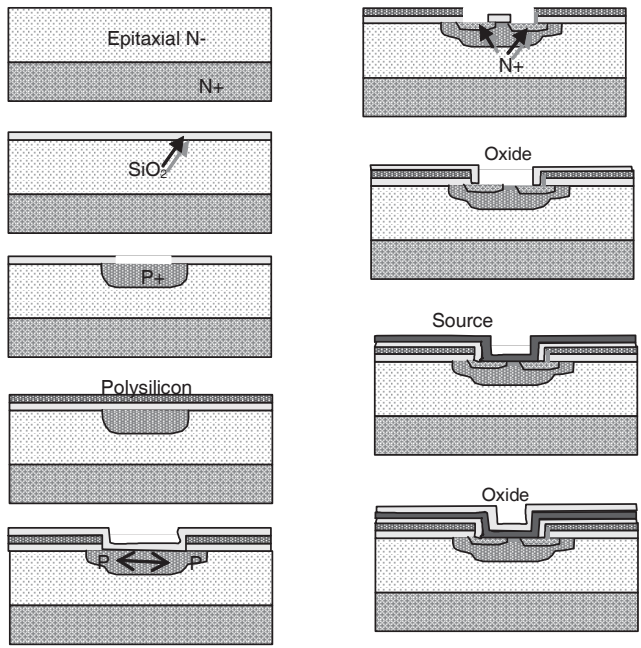


Figure 1.6. *Power MOSFET process*

1.2.2. *Physical and structural MOS parameters*

1.2.2.1. *Vertical structure*

If we examine the vertical structure of a MOSFET in more detail, we notice that it is made up of a N+N-PN+ parasitic bipolar transistor, in which the collector, emitter and base are formed by the drain, source and P channel. In order to avoid any parasitic transistor being turned on, base and emitter are short-circuited by the source metallization, but it remains a parasitic bipolar diode where the drain is the anode and the source is the cathode (see Figure 1.7), so the power MOSFET cannot sustain any reverse voltage.

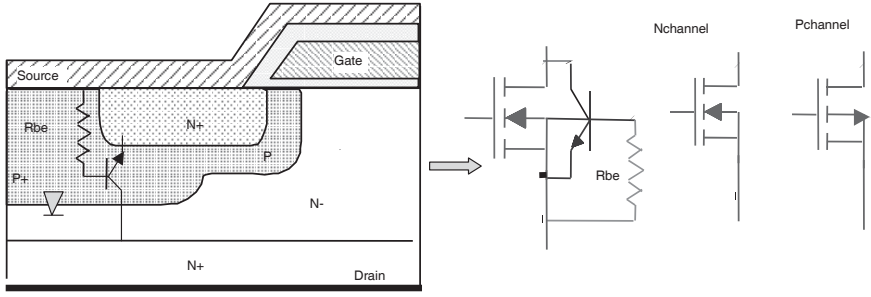
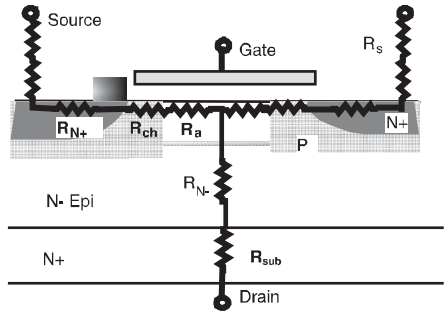


Figure 1.7. *Parasitic bipolar transistor and symbols*

1.2.2.2. *Upper side technological choices*

Power MOSFET is divided into two parts: the N- drift zone which sustains the electric field and the upper part including the gate, source and channel. This part controls the switching times of the power MOSFET. This part is very important for the internal resistance of a low voltage power MOSFET; see Figure 1.8. The main technological choices for the cells are the geometry and size; it will be the same for the P-well, the channel, the gate and the source.



R_{DSon}		
	VDS = 30V	VDS = 600V
RS	7%	0.5%
RN+	6%	0.5%
Rch	28%	1.5%
Ra	23%	0.5%
RN-	29%	96.5%
Rsub	7%	0.5%

Figure 1.8. *MOSFET internal resistance distribution*

1.2.2.2.1. Geometry of the cells

The rule for the channel resistance is very well known

$$R = pL/A$$

where ρ is the material resistivity, L is the channel length and A is the channel section. As the channel depth is constant, the channel resistance is governed by channel length and channel width. Thus for the same die size, channel resistance is lower when cells are optimal in terms of minimum channel length L and maximum perimeter Z = 4R; see Figure 1.9.

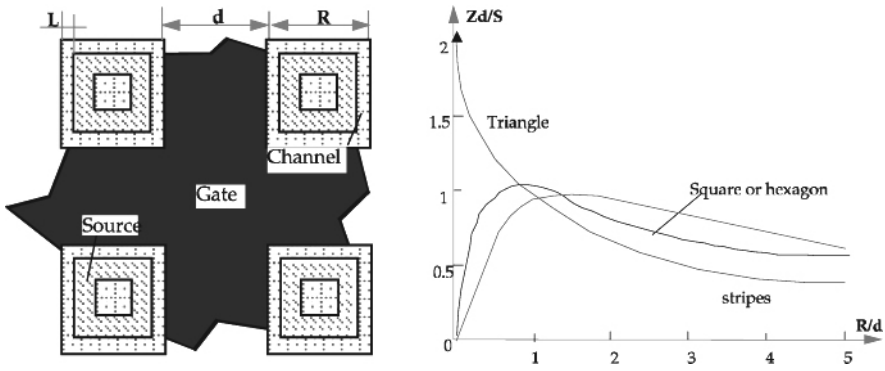


Figure 1.9. *Optimal shape of cells*

Figure 1.9 shows quality factor versus cell size; that optimal shape of the cell is given when R = d, source width = distance between two cells. Cell density is also an important parameter for lower channel resistance; this explains why the challenge for a low voltage power MOSFET is the maximum number of cells per surface unit. For medium and high voltage devices the number of cells per surface unit is not so important because the part of channel resistance in the R_{dson} is very low, and on the other hand, in high voltage devices the distance between two cells must be sufficient in order to avoid any JFET parasitic element.

1.2.2.2.2. P well choice

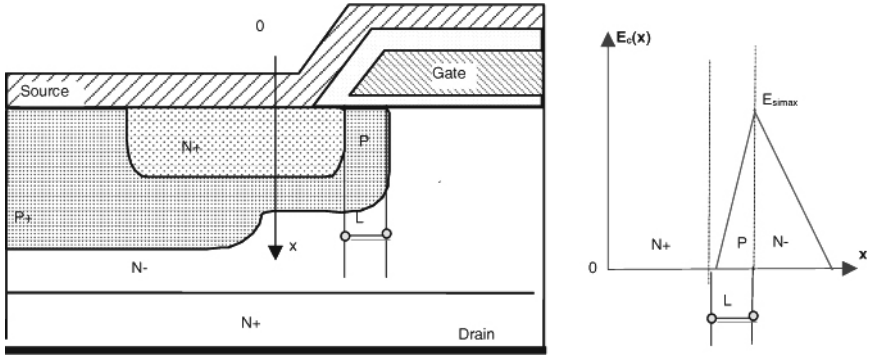


Figure 1.10. *Drain-source punch through*

When a voltage is applied between the drain and the source, a space charge $E_c(x)$ is spread out between the N- drain and P channel; see Figure 1.10. Thus, the doping level in the well must be high enough in order to avoid any field at the N+P source-well junction, if not, a punch through will come at this junction, For example, if a uniform doping is made inside the P well, a minimum channel length L is designed for the critical field E_{simax}.

$$L_{\min} > \frac{\varepsilon_{Si} E_{Si\max}}{qN_A}$$

P doping of the well can be increased in order to reduce the channel length, to reduce channel resistance. However, this increases the threshold voltage, and the voltage level of drive may be too high for the existing integrated drivers.

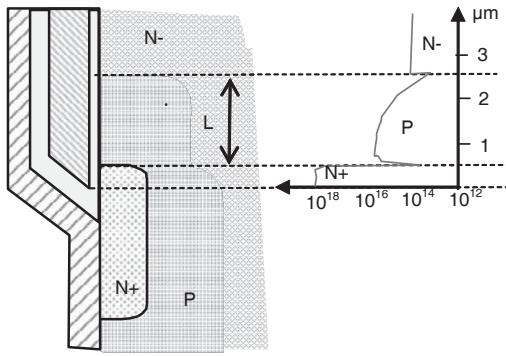


Figure 1.11. *Well doping profile*

From a practical point of view, channel length is defined by lateral doping drives of N+ source and P channel. This way, very short channels may be obtained. At the beginning, length L was around 10 µm; nowadays, length L is less than 1 µm. Normally, the doping level of N+ is much higher than the P level of the well, and punch through is automatically avoided; see the doping profile in Figure 1.11.

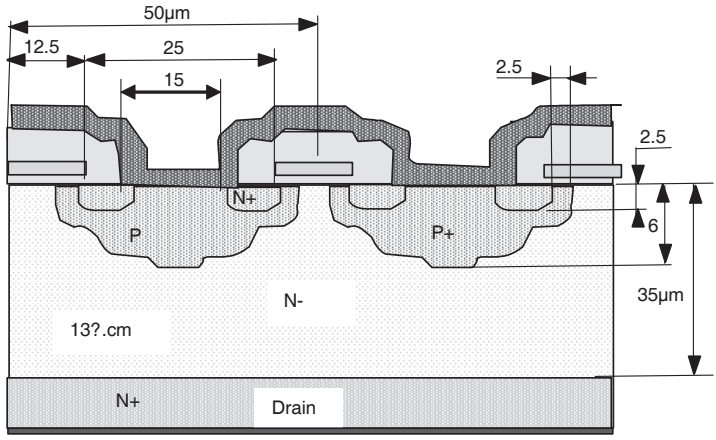


Figure 1.12. *400 V, 5 A MOSFET size*

As an example, Figure 1.12 shows the dimensions of a medium voltage power MOSFET, with 25 µm square cells, and a 35 µm N- drift zone, with a resistivity of 13 Ω.cm, in order to sustain a 400 V blocking voltage.

1.2.2.2.3. Gate realization

Gate characteristics must include:

– a low access resistance,

– a low leakage current,

– a great stability versus time,

– a high breakdown voltage,

– a low input capacitance.

The gate must also be easy to make.

The first gates were made of aluminum. This material has a lot of the desired qualities, except the fact that it sometimes contains particles of sodium, which in turn could create reliability problems.

Another possibility was to use molybdenum for gate manufacturing, but its cost is quite high.

Final choices remain monocrystalline silicon, but it is too resistive and expensive, and polycrystalline silicon, which is less resistive than monocrystalline,

but its resistance remains 3,000 times higher than that of aluminum, and 50 times higher than that of molybdenum. However, this resistance has no significance for the switching times of a current power MOSFET, and a gate may be N+ doped, during source diffusion, in order to be very well controlled.

Gate oxide thickness d_{ox} determines not only the maximum gate voltage but also the switching times by its capacitance. Maximum gate voltage is given by:

$$V_{GSmax} = d_{ox}. E_{oxmax}$$

where E_{oxmax} is the oxide maximum electric field, around 750 V/µm.

For a 75 V maximum gate voltage, a 100 nm oxide thickness is sufficient. Gate oxide capacitance per square centimeter, versus its thickness, is given by:

$$C_{ox} = \varepsilon_{ox}/d_{ox}$$

where ε_{ox} is the oxide permeability, which is $\varepsilon_{si}/3$ (around 3.5 10^{-13} F/cm). Cox is reduced when dox increases. Thus, a capacitance per square centimeter from 35 nF to 3.5 nF may be obtained for a gate thickness from 100 nm to 1 µm. Oxide capacitance is the main part of the MOSFET gate-source capacitance C_{gs}, and it may be obtained using:

$$C_{GS} = A_{ox}. \varepsilon_{ox} /d_{ox}$$

where A_{ox} is the oxide covered area, which may be between 50% and 80% of the die area. For example, C_{gs} per silicon square centimeter varies between 17.5 nF and 28 nF for d_{ox} = 100 nm and between 1.75 nF and 2.8 nF for d_{ox} = 1 µm, according to the oxide coverage. To decrease the input capacitance, d_{ox} must be increased, but a thick oxide leads to a large threshold gate voltage, while V_{th} is proportional to d_{ox}. Additionally, channel transconductance is decreased. A typical d_{ox} that is currently used is around 100 nm.

1.2.2.2.4. Source choice

The doping level of the N+ source is matched in order to control the channel length. This is also important for access resistance. Source metallization allows an easy paralleling of cells and an overall die coverage for good temperature spreading.

1.2.2.3. *Sustaining static drain source voltage V_{DSS}*

1.2.2.3.1. Drift region N-

V_{DSS} is proportional to the N- region thickness. For a given V_{DSS}, an optimization process is conducted between N- doping level and N- thickness, W_v. Avalanche, ionization and multiplying effects must also be taken into account. For voltage strength, a N- MOSFET structure is similar to a P+N-N+ diode structure. Figure 1.13 shows the electric field inside a MOSFET when it is polarized by a maximum drain-source voltage VDSS. In other words, when the electric field rises, the silicon critical field $E_{simax} = E_{Cmax}$.

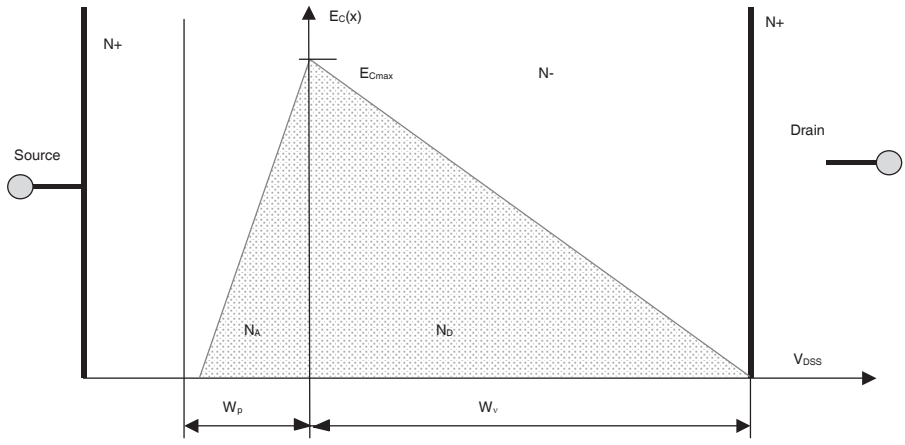


Figure 1.13. *Electric field inside the MOSFET drift region N-*

If the doping concentration inside the N- layer is constant, the electric field decreases linearly from the PN- junction to the N-N+ drain. Assuming an "abrupt junction", electric field $E_C(x)$, thickness of the W_{ZD} zone and the applied voltage V_{app} are linked together by the following equations:

$$E_C(x) = \frac{2V_{app}}{W_{ZD}^2}(W_{ZD} - x)$$

and $$W_{ZD} = \left(\frac{2\varepsilon_{si}.V_{app}}{qN^-}\right)^{1/2} = \left(\frac{1.3\cdot 10^7 V_{app}}{N^-}\right)^{1/2}$$

When the critical field is affected, silicon breakdown appears. At this time, the electron and hole ionization integral is equal to one, and carriers are produced by atom ionization in the space charge area. The current is increased by avalanche multiplication, M_p and M_n factors diverge to infinity:

$$1-\frac{1}{M_p}=\int_0^{W_{ZD}}\alpha_p \exp\left(-\int_0^x(\alpha_p-\alpha_n)dx'\right)dx=1$$

$$1-\frac{1}{M_n}=\int_0^{W_{ZD}}\alpha_n \exp\left(-\int_0^x(\alpha_n-\alpha_p)dx'\right)dx=1$$

where αp and αn are the ionization coefficients for holes and electrons, which vary exponentially with the electric field. They are given by the following equations:

$$\alpha_n = a.\exp(-b/E) \qquad \text{and} \qquad \alpha_p = 0.344\ a.\ \exp(-b/E)$$

where $a = 1.6.10^6$ (cm^{-1}) and $b = 1.65.10^6$ ($V.cm^{-1}$). Analytical integration of the two equations is very difficult, and solutions are found numerically. In order to obtain an analytical expression for the breakdown voltage, two simplifying hypotheses may be used. The first assumes equal ionization coefficients. Thus:

$\alpha_n = \alpha_p = \alpha_i$ (Mn = Mp =M) and α_i may be analytically expressed.

The second simplifying hypothesis uses the same multiplication terms for the preceding equations. Analytical expression for α_i is difficult and may be approximated by:

$$\alpha_I = 3.3\cdot10^{35}E(x)^7 \qquad (cm^{-1})$$

This equation is the best compromise for voltages over 400 V.

This way, the criterion for maximum voltage determination becomes:

$$1-\frac{1}{M}=\int_0^{W_{ZD}}\alpha_i dx=1$$

for $V_{app} = V_{DSS}$ and $W_{ZD} = W_\nu$

Thus: $\alpha_i = 4.2.10^{-33} \frac{V_{DSS}^{\ 7}}{W_\nu^{14}} (W_\nu - x)^7$

and V_{DSS} is given by:

$$V_{DSS} = 5.5.10^4 \, W_\nu^{6/7}$$

Breakdown can now be obtained versus doping level N_D and N- thickness W_ν:

$$V_{DSS} = 4.26 \; 0.10^{13} \, N_D^{-3/4}$$

$$W_\nu = 2.94. \; 10^{-2} \, V_{DSS}^{7/6}$$

Thus, silicon resistance per unit area in the W_ν drift zone is:

$$R_\nu = 8.2.10^{-9} \, V_{DSS}^{2.5}$$

For example, if N_D doping level is $5.10^{14} cm^{-3}$, V_{DSS} could be around 400 V with a drift zone thickness of 32 µm. Then, silicon resistance per unit area is: $0.026 \; \Omega.cm^2$.

For a doping level $N_D = 10^{14} \; cm^{-3}$, V_{DSS} rises to 1,350 V with a thickness of 130 µm, and silicon resistance per unit area is $0.55 \; \Omega.cm^{-2}$.

In the previous analysis, the drift zone is completely empty when the E_{simax} electric field is raised. For the same maximum voltage, drift zone thickness may be reduced in a different way: the "punch-through" technique consists of doping the drift zone in such a way that the maximum electric field is lower than the critical E_{simax} field limit, when the drift zone is completely empty. In this case, the space charge spreads in the N+ drain, and the voltage is maximum when the critical field is raised (see Figure 1.14). If the ratio between maximum field and minimum field is expressed by a variable such as:

$$E_{Cm} / E_{CM} = 1 - \alpha$$

Then, the voltage at $E_{CM} = E_{simax}$ is:

$$V_{dsopt} = 0.5(E_{CM} + E_{cm}) W_{\nu opt} = 0.5(2 - \alpha) \, E_{simax} \, W_{\nu opt}$$

If the target for V_{DSSopt} is the value without the "punch-through" technique (α = 1), the following condition is requested:

$$V_{DSopt} = 0.5(2 - \alpha)E_{simax}.\ W_{vopt} = V_{DSS} = 0.5\ E_{simax}.\ W_{v}$$

This leads to: $W_{vopt} = W_{v}/(2 - \alpha)$

The electric field slopes during avalanche with and without the "punch-through" technique are:

$$\frac{E_{Si\max}}{W_v} = \frac{qN_D}{\varepsilon_{Si}} \quad \text{and} \quad \frac{\alpha E_{Si\max}}{W_{vopt}} = \frac{qN_{Dopt}}{\varepsilon_{Si}}$$

From these equations, the following characteristics are obtained:

$$N_{Dopt} = \alpha(2 - \alpha)N_D \qquad (cm^{-3})$$

$$V_{DSSopt} = 5.5.10^{14}\ [(2 - \alpha)W_{vopt}]^{6/7} \qquad (V)$$

$$V_{DSSopt} = 4.26 \cdot 10^{13} \left[\alpha(2-\alpha)\right]^{3/4} N_{vopt}^{\ -3/4} \qquad (V)$$

$$W_{vopt} = 2.94 \cdot 10^{-2} \frac{1}{2-\alpha} V_{vopt}^{\ 7/6} \qquad (\mu m)$$

$$R_{vopt}^{\ *} = 8.2 \cdot 10^{-9} \frac{1}{\alpha(2-\alpha)^2} V_{DSSopt}^{\ 2.5} \qquad (\Omega.cm2)$$

Designers generally use α = 0.75. For example, for a 400 V device, the N- area is 26 µm deep with a doping of $4.7.10^{14}$ cm^{-3}, resistance per unit area is: 0.022 $\Omega.cm^{-2}$, leading to a 18% saving compared to the case without the punch-through technique.

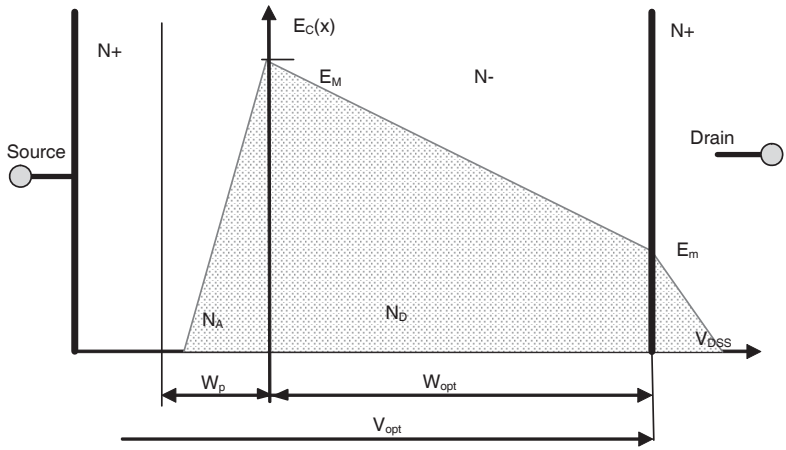


Figure 1.14. *"Punch through" electric field inside the N- MOSFET drift region*

1.2.2.3.2. Gate oxide and drift zone N- interface

When the MOSFET is blocked, the gate is connected to ground or negatively polarized. Drain source voltage and possibly the negative gate voltage are spread out between the drift zone and the gate oxide. In the case where no charges are stored in the gate oxide and in the gate silicon oxide interface, Figure 1.15 shows the electric field in the oxide and in the drift zone. In the silicon oxide interface we obtain:

$$\varepsilon_{ox}.E_{Cox} = \varepsilon_{si.}\, E_{CN}$$

When the field $E_{Cox} = 3E_{CN-}$ raises the critical value for silicon, the previous equality $E_{CN-} = E_{simax}$ becomes:

$$\varepsilon_{ox}.E_{Cox} = \varepsilon_{si.}\, E_{simax}$$

which gives:

$$E_{Cox} = 3\, E_{simax}$$

As there is a ratio of three between the dielectric constants of the two materials ($\varepsilon_{si}/\varepsilon_{ox} = 3$), the critical field of the oxide ($7.5.\ 10^6$ V/cm) is higher than the critical field in silicon, and the ratio E_{oxmax}/E_{simax} may be between 9 and 30 for a doping between 5.10^{17} cm^{-3} and 10^{14} cm^{-3}. This is much higher than 3. Therefore, breakdown

occurs only in silicon. With a doping level of $1.6.10^{14}$ cm^{-3}, and for a sustaining voltage around 1,000 V, the maximum electric field is $E_{simax} = 2.5.10^5$ V/cm when a 1,000 V voltage is applied between drain and source. Voltage inside the oxide is:

$$\Delta V_{ox}\ E_{oxmax}.d_{ox} = 3E_{simax}.d_{ox}$$

This is a 7.5 V voltage (d_{ox} = 100 nm, V_{GSmax} = 75 V).

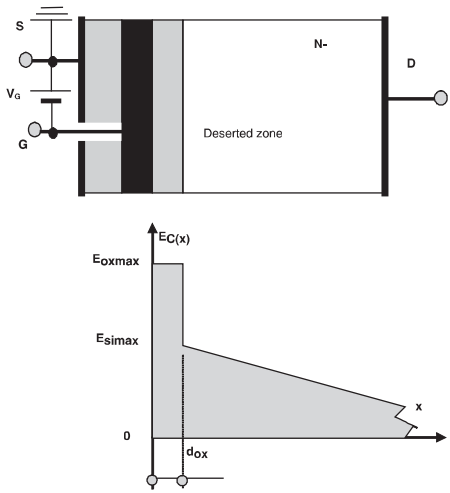


Figure 1.15. *Interface between gate oxide and deserted N- zone*

In the conduction state, before switch-off, the N- region under the gate oxide is a neutral region (rather than an accumulation region). When the voltage increases during switch-off, this neutral region becomes a space charge region. The two space charges at the PN- junction should be interpenetrated before the V_D voltage goes beyond the value of:

$$E_{oxmax.}\ d_{ox} \quad \text{(see Figure 1.16).}$$

This phenomenon is called the "gate shielding effect". Thus, a maximum distance appears between two wells: d_{max}. For example, for a junction 5 µm deep and an oxide thickness of 100 nm, d_{ox} becomes 34 µm for $N_D = 2.10^{14}$ cm^{-3} and 25 µm for $N_D = 4.10^{14}$ cm^{-3}.

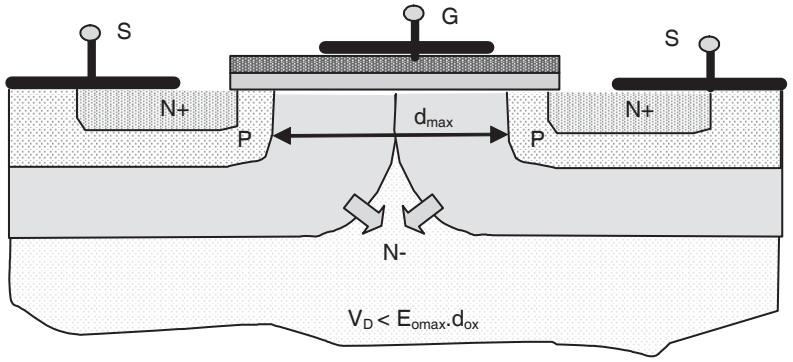


Figure 1.16. *Gate shielding effect*

1.2.2.3.3. Peripheral sustaining voltage

The silicon critical field is around 20 V/cm, thus a 1,000 V voltage may be sustained by a silicon thickness of 100 µm. However, at the atmospheric pressure, the air breakdown field is around ten times less. Therefore, a 500 µm distance in the air is requested to sustain this voltage, and the surrounding silicon must be large enough in order to keep the breakdown inside the silicon. Thus, in a silicon MOSFET, techniques such as "field plate", "guard ring", "pocket", etc. are used; see Figure 1.17.

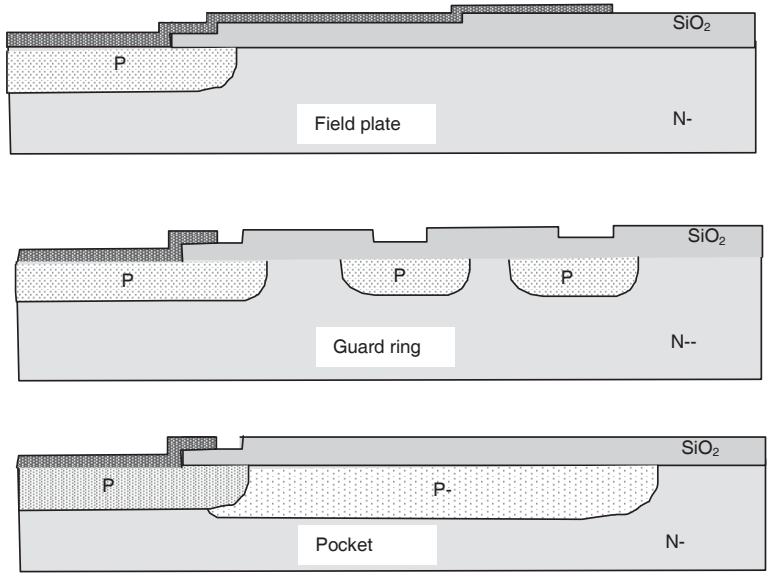


Figure 1.17. *Peripherals examples*

1.2.3. *Permanent sustaining current*

In the power device, the permanent sustaining current is mainly thermally limited. The drain-source current density is the same as the channel current density. The latter, with an adequate gate voltage, is generally much larger than the nominal values given by the manufacturer. During permanent operation, charge in the channel is given by:

$$Q_{ns} = \Delta V_{ox}\ L.2.\ \varepsilon_{ox}/d_{ox}$$

This gives the channel resistance:

$$R_{channel} = \frac{L^2}{\mu_{ns} Q_{ns}} = \frac{d_{ox} L}{\mu_{ns} \Delta V_{ox} Z \varepsilon_{ox}}$$

A channel may include the following characteristics: length: 2 µm, width: 10 m per cm^2, square cells of 10 µm, oxide thickness: 100 nm, channel resistance: $R_{channel}$: $9.5.10^{-4}\ \Omega.cm^{-2}$. This makes a current density of 525 $A.cm^{-2}$, if a 0.5 V voltage drop in the channel is acceptable. This value is much higher than current densities well known by power electricians ($\leq$100 $A.cm^{-2}$ for a low voltage MOSFET and around 50 $A.cm^{-2}$ for a power MOSFET over 600 V). Figure 1.18 shows a power MOSFET in a non-insulated package, such as a TO220. Losses in the die increase with

temperature and the thermal flux flows in the direction from die to die bonding, thermal spreading layer and package environment, a heat sink for example. Losses in the on state are:

$$P = R_{dson}.I^2_{DS}$$

The die temperature is given by

$$T_j = P\,(R_{th(si)} + R_{th(sol)} + R_{th(hs)}) + T_{sur}$$

where $R_{th(si)}$, $R_{th(sol)}$ and $R_{th(hs)}$ are the three sheet thermal resistances, which are given by the following equations:

– for silicon: R_{th} (si) = d(si) /σ_{th}(si).A(si)), thus 0.036°C/W per square centimeter, with a thickness d_{si} = 300 µm and $\sigma_{th(si)}$ = 0.83W/°C.cm;

– for soldering: R_{th} (sol) = d(sol) /σ_{th}(sol).A(sol) thus 0.03°C/W per square centimeter, for a thickness d_{sol} = 100µm and $S_{sol} = S_{si}(\sigma_{th(sol)})$ = 0.33 W/°C.cm for lead;

– for thermal spreading layer: R_{th} (TSL) = d(TSL) /σ_{th}(TSL).√A(TSL), thus 0.036°C/W per square centimeter, for a thickness d(TSL) = 2 mm A(TSL) = 2Si and σ_{th}(TSL) = 3.9 W/°C.cm for copper.

The amount of the three resistances is called “junction to case thermal resistance”:

$$R_{th(j\text{-}c)} = R_{th(si)} + R_{th(sol)} + R_{th(TSL)}$$

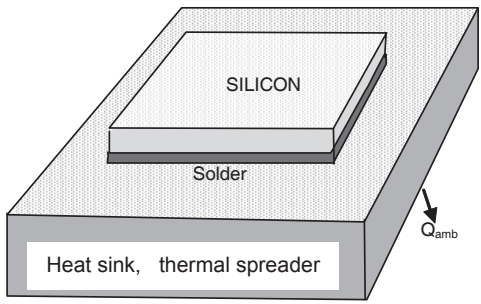


Figure 1.18. *Power MOSFET die stack*

We can see that for one square centimeter of silicon, the thermal limit is around 0.1°C/W. For a power MOSFET with a voltage over 400 V, we can accept an internal resistance R_{Dson} quite the same as that of the drift zone, $\tilde{R_v}$. The continuous current per square centimeter of silicon becomes:

$$J_{DS} = \sqrt{\frac{T_j - T_{env}}{R_{th(j-c)} R_{Dson}}} = \sqrt{\frac{T_j - T_{env}}{R_{th(j-c)} 8,2.10^{-9} V_{DSS}^{2.5}}}$$

with $R_{th(jh-c)}$ in °C/W and R_{DSon} in $\Omega.cm^{-2}$.

If the maximum temperature of silicon is T_j = 150°C, and if the ambient temperature is T_{amb} = 50°C, current density cannot be over 144 $A.cm^{-2}$, for a 500 V power MOSFET, and 60 $A.cm^{-2}$ for a 1,000 V power MOSFET. This calculated value of thermal resistance Rth(j-c) is always under the true value due to some defects in the links of the three layers. Now, electric insulation between silicon and external elements is always requested in power electronics. This insulation increases the thermal resistance, the previous values for the overall thermal resistance are always optimistic and the current density cannot be over 60 $A.cm^{-2}$ for a 500 V power MOSFET, and 20 $A.cm^{-2}$ for a 1,000 V device.

As the source connection of a power MOSFET is made of aluminum wires, the diameter and the number of these wires also make a continuous current density limitation. The semiconductor industry uses 250 µm diameter aluminum wires, for a maximum current of 10 A.

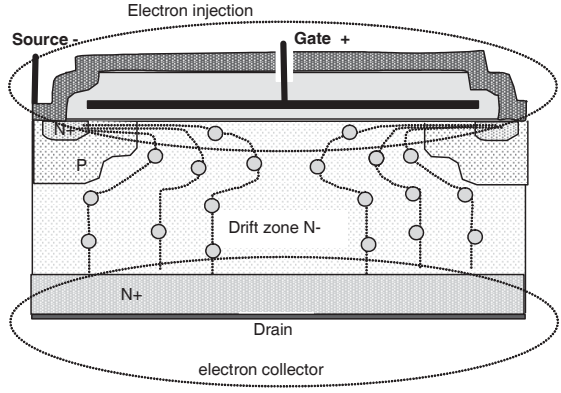


Figure 1.19. *Power MOSFET operation principle*

1.3. Mechanism of power MOSFET operation

1.3.1. *Basic principle*

As power MOSFETs are unipolar devices, only one carrier type operates: electrons for N-type power MOSFETs, and holes for P-type power MOSFETs. Electron injection is performed by N+ source and channel is driven by the gate. When $V_{DS} > 0$, injected electrons move through the drift zone and raise the N+ zone and the drain. When the gate stops the electron injection, it turns off the MOSFET. Figure 1.19 shows this principle. Thus, the channel is the main part of the power MOSFET drive.

1.3.2. *Electron injection*

Channel studies were undertaken early in the development of power MOSFET. Today, its mechanism is very well known. When the gate-source voltage is greater than a V_{th} value (a minimum value which creates a channel inversion), free electrons rise to the silicon surface in the P-well, called the inverting channel. This inversion modulates the channel resistivity. When V_{GS} is high, resistivity decreases. Figure 1.20 shows two possible operations. The voltage between channel ends is V(L).

First, saturation operation occurs when: $V_{GS} - V_{th} >> V(L)$, as depicted in Figure 1.20(a).

This is the channel state when the power MOSFET is conducting. The equation for this case is:

$$I_{Cchannel} = \frac{Z}{L}\mu_{ns}\frac{\varepsilon_{ox}}{d_{ox}}(V_{GS} - V_{th})V(L) - \frac{1}{2}V(L)^2$$

where μ_{ns} is the electron mobility at the silicon surface.

The threshold voltage is approximated by:

$$V_{th} = \phi_{ms} + 2\phi_{Fi} + \frac{d_{ox}}{\varepsilon_{ox}}\sqrt{4qN_A\varepsilon_{Si}\phi_{Fi}}$$

where Φ_{ms} is the difference between the metal and the semiconductor working functions, N_A is the channel doping, Φ_{Fi} is the distance between the Fermi level and the intrinsic semiconductor Fermi level, which is:

$$\Phi_{Fi} = U_T.\text{Log } N_A/ni$$

where U_T is a thermo dynamic unit (26 mV at 25°C).

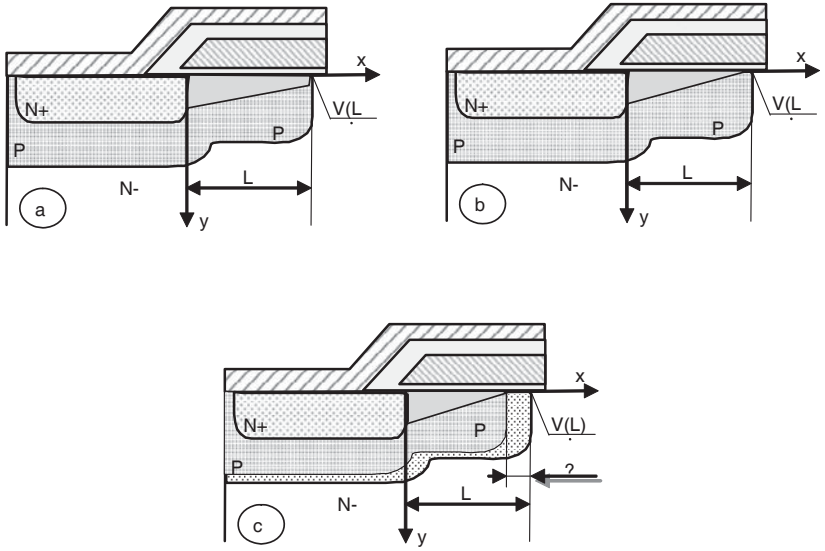


Figure 1.20. *Electron injection:*
(a) saturated operation; (b) beginning of pinch off;
(c) desaturated operation

We can see that the threshold voltage V_{th} increases with the channel doping, but also with the gate oxide thickness.

First, desaturation operation occurs when: $V_{GS} - V_{th} < V(L)$, depicted in Figure 1.20(c).

Starting from a conduction state with an I_D current in the channel, if gate source voltage slowly decreases, the charge density in the channel decreases and the channel resistance increases, in turn increasing the voltage drop in the channel. When the channel inversion leads to the disappearance of the channel ends, we obtain a channel pinch off (see Figure 1.20(b)). If the gate voltage is even more reduced, pinch off moves towards the source, and a deserted region, or space charge, is settled at the channel ends. This is the desaturation operation. Channel voltage is quite equal to the space charge ΔL. If the channel doping is not important, and if the gate oxide thickness is thin, the pinch off voltage is: $V(L) = V_{channel} = V_{GS} - V_{th}$.

After the channel pinch off, channel electron flux only depends upon the gate voltage V_{GS}. This maximum current is:

$$I_{channel} = \frac{Z}{2L}\mu_{ns}\frac{\varepsilon_{ox}}{d_{ox}}(V_{GS} - V_{th})^2$$

If ΔL/L is not negligible, as in the case of a short channel, the channel current is a function of voltage V/(L) in the form of:

$$I_{channel} = \frac{Z}{2L}\mu_{ns}\frac{\varepsilon_{ox}}{d_{ox}}(V_{GS} - V_{th})^2 \cdot \frac{L}{L - \Delta L}$$

where $\frac{Z}{L}\mu_{ns}\frac{\varepsilon_{ox}}{d_{ox}}$ is usually represented by K_p, the transconductance parameter.

1.3.3. *Static operation*

This operation may be the ON state or the OFF state. The ON state is the conduction of an I_D current, which is not channel driven, but imposed by the load. So, the number of electrons at input is equal to the number of electrons at drain output. If the requested charge density for the current is small compared to the doping density, $J_D/qv_n << N_D$, the drift zone is like a R_D resistance, and the power MOSFET shows a V_{DS} drop in voltage (see Figure 1.21). If the drain N+ and the source N+ resistances are neglected:

$$V_{DS} = I_D . R_{DSon} = I_D . (R_n + R_a + R_{channel})$$

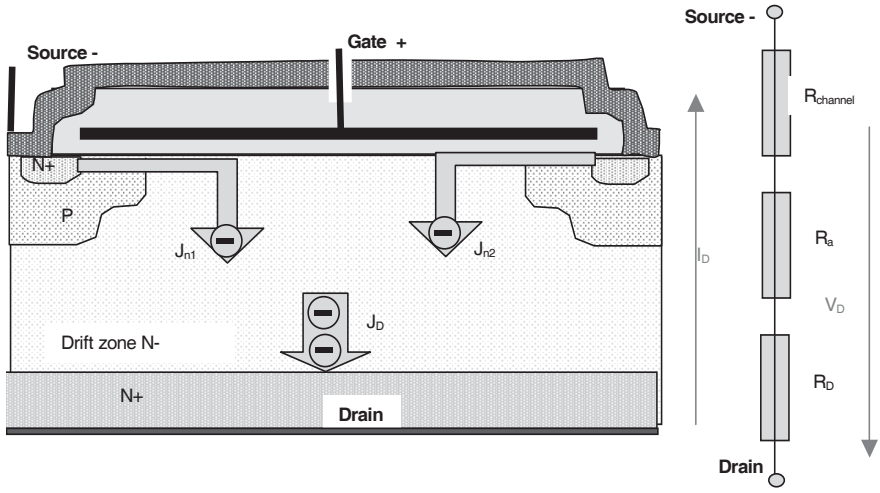


Figure 1.21. *Conduction operation:* $J_{D1} + J_{D2} = J_D$

Channel resistance depends upon the physical characteristics of the material used, on channel geometry and also on gate voltage. The relationship may be written as:

$$R_{channel} \approx \left[\frac{Z}{L} \mu_{ns} \frac{\varepsilon_{ox}}{d_{ox}} (V_{GS} - V_{th}) \right]^{-1}$$

Access resistance, R_a, and drift zone resistance, R_v, are two phenomena able to change their theoretical values. This is achieved by narrowing the access region and restricting current lines at the drift region entry. We can see, as shown in Figure 1.22, that electrons, after running in the channel close to the silicon surface, turn at 19 degrees in the drift zone. The sum of the voltage drop, in the channel, in the lateral region (weak, thanks to the accumulated region), and in the access region, build up a space charge in the access zone, and, consequently, the pathway for electrons is narrowed. After that, electrons are injected into the drift zone and go out through the drain. We can see that the current pass at the entry level of the drift zone, is narrower than at the output level. Due to these two parasitic effects, the effective resistances R_a and R_v increase up to 20%.

This R_a resistance may be written as:

$$R_a = \lambda \cdot \frac{1}{q\mu_n N_D} \frac{2h}{Zd}$$

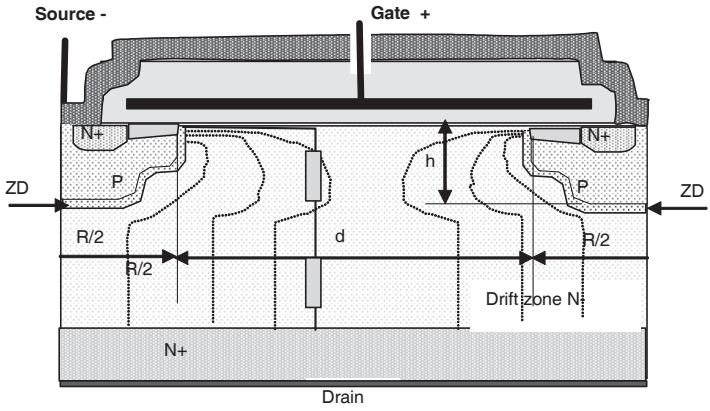


Figure 1.22. *Resistances R_a and R_v increase due to the narrowing pathway*

This may be considered as an N- silicon bar with a cross-section Zd/2, a length h and a doping level N_D increased by a factor λ according to cell geometry and gate voltage. This factor is greater than one, and increases when the doping level increases, and decreases with P+ depth and gate voltage V_{GS}.

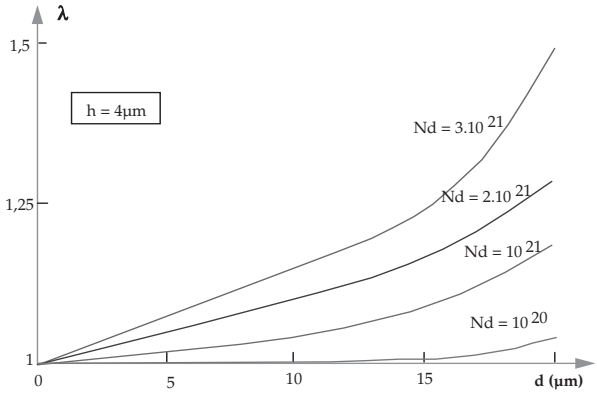
Thus, λ is high for a low voltage power MOSFET, and λ is close to one for a high voltage power MOSFET (as shown in Figure 1.23(a)).

The drift zone resistance is:

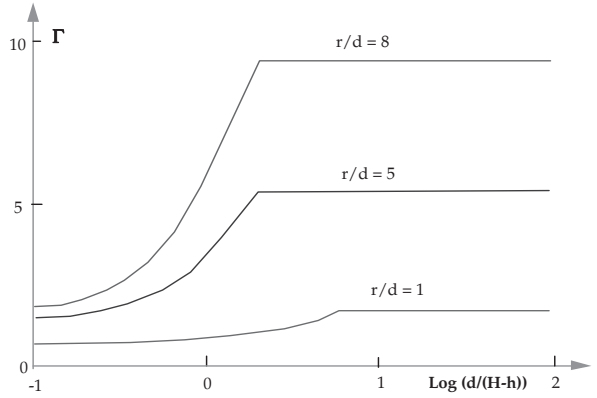
$$R_{vV} = W_v / q\ \mu_n\ N_D\ A(si)$$

The actual value for this resistance is given by: $R_v = \Gamma . R_{vV}$

The Γ factor represents the geometrical factors. It is larger for a low voltage power MOSFET, according to Figure 1.23(b).



(a)



(b)

Figure 1.23(a) and (b). *Resistances Ra and R_v versus geometrical factors*

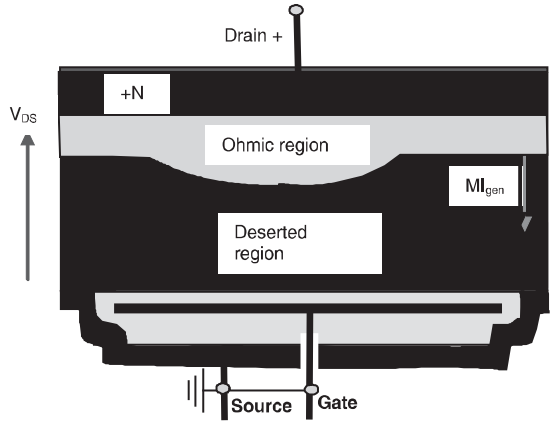
In the OFF state, the external power circuit imposes a supply voltage between the drain and source of the MOSFET. Drift zone N- includes two regions: a deserted space charge region W_{ZD}, which sustains the voltage, and a resistive region, which is the remaining part of the drift region without a field, depicted in Figure 1.24(a).

Very few mobile carriers ($p << n << n_i$) are located in the deserted space. The field generates some electron-hole pairs, and consequently a generative current. This is the main part of the leakage current of the MOSFET when the die temperature

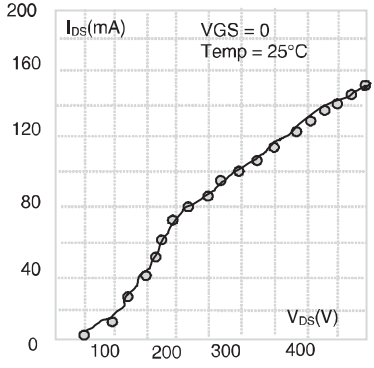
remains under about 170°C. If the generative electron-hole pair is represented by a generative rate τ_{gen}, the leakage current is:

$$J_D \text{ leakage} = q.n_i W_{ZD}/\tau_{gen}$$

Leakage increases with the applied voltage, through space charge W_{ZD}, and increases with temperature through the intrinsic concentration n_i, which is closely related to temperature. Figure 1.24(b) shows the leakage current versus the applied voltage.



(a)



(b)

Figure 1.24(a) and (b). *Blocked MOSFET, leakage current versus V_{DS}*

1.3.4. *Dynamic operation*

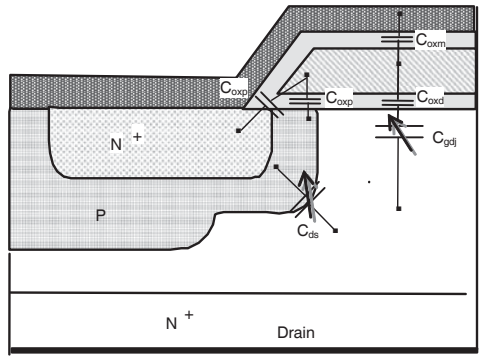
The dynamic operation's main function is to commutate the power electronics system. Power MOSFETs are much faster than bipolar transistors, where a conductivity modulation takes place. Only a very small quantity of charges are involved, corresponding to the establishment and discharge of the space charge in the drift region. Consequently, the dynamic operation is only dominated by capacitive effects, as seen in Figure 1.25. Two types of capacities are included: constant capacities and the non-linear capacities. Constant capacities are: C_{oxm}, due to oxide and source metallization; C_{oxd}, the access capacity; C_{oxp}, the channel capacity; and C_{oxs}, the P+ source capacity.

The three last capacities can be easily calculated using the following equations:

$$C_{oxd} = 0.5\ L_{acc}.Z.\varepsilon_{ox}/d_{ox},\ C_{oxp} = 0.5\ L.Z.\varepsilon_{ox}/d_{ox} \text{ and } C_{oxs} = L_{rec}.Z.\varepsilon_{ox}/d_{ox}$$

where L_{rec} is the gate length over the source N+.

(a)



(b)

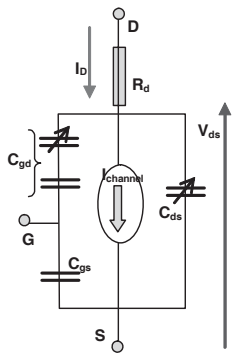


Figure 1.25. *(a) Various internal capacities and (b) equivalent circuit*

The drift zone is divided into two regions: the space charge region, and the ohmic region. The non-linear capacities highlight the space charge variation: C_{gdj} for the access deserted area L_{acc}, and C_{dsj} for the P+N- L_{ds} deserted area. The non-linear capacities are described by the following equations:

$$C_{gdj} = \frac{1}{2} L_{acc} \sqrt{\frac{q\varepsilon_{si} N_D}{2V_{ZD(acc)}}} \quad \text{and} \quad C_{dsj} = \frac{1}{2} L_{ds} \sqrt{\frac{q\varepsilon_{si} N_D}{2V_{ZD(ds)}}}$$

where $V_{ZD(acc)}$ and $V_{ZD(ds)}$ are the space charge voltages in the L_{acc} and L_{ds} parts of drift region N-. They are equivalent to:

$$V_{ZD(acc)} = V_{DS} - I_D . R_{acc} - V_{GS} \qquad \text{and}$$

$$V_{ZD\,(ds)} = V_{DS} - I_D . R_{ds} + \Phi\Phi_{PN}$$

where Φ_{PN} is related to the P+N junction in thermo dynamic equilibrium.

Due to the difference between input and output currents in silicon, the power MOSFET involves an enlargement or a narrowing of the space charge in this silicon. Thus, there is a differrence between the $I_{channel}$ current in the channel, and the I_D current imposed by the load.

When the drain current is larger than the channel current, $\Delta I_N = I_D - I_{channel} > 0$, more electrons are released from the drain than are input by the channel. Thus, an enlargement of space charge occurs and V_{DS} increases.

Now, when $\Delta I_N = I_D - I_{channel} < 0$, more electrons come from the channel input than are released by the drains output. Thus, a narrowing of space charge occurs and V_{DS} decreases. The following formula describes these phenomena:

$$I_D - I_{channel} = C_{gdj} . dV_{ZD(acc)}/dt + C_{dsj} . dV_{ZD(ds)}/dt$$

The overall gate-source capacitance is quasi-constant and equal to:

$$C_{gs} = C_{oxm} + C_{oxp} + C_{oxs}.$$

The drain-source capacitance C_{ds} is only related to the space charge in the drift region and in the P+N junction. It is obviously non-linear. The gate-drain capacitance is made by a series association of capacitances: $C_{gd} = C_{gdj} . C_{oxd}/(C_{gdj} + C_{oxd})$. The neutral (or ohmic) region of the drift region is divided in two parts, L_{acc} and L_{ds}, represented by two resistances, R_{acc} and R_{ds}, according to Figure 1.25.

In the channel, for behavioral studies, a quasi-stationary approach is sufficient in the dynamic case. The channel current versus time is given by:

$$I_{channel}(t) = 0 \qquad V_{gs} - V_{th} \leq 0$$

$$I_{channel}(t) = \frac{Z}{L}\mu_{ns}\frac{\varepsilon_{ox}}{d_{ox}}(V_{GS}(t) - V_{th})V(L,t) - \frac{1}{2}V(L,t)^2 \qquad V_{gs} - V_{th} > V(L,t)$$

$$I_{channel}(t) = \frac{Z}{2L}\mu_{ns}\frac{\varepsilon_{ox}}{d_{ox}}(V_{GS}(t) - V_{th})^2 \qquad V_{gs} - V_{th} \leq V(L,t)$$

The channel voltage drop, $V_{channel}(t)$, is close to V(L,t) before pinch off. It remains constant and equal to V_{gs} (t) – V_{th} after pinch-off, because the deserted region of the drift zone sustains the main part of the voltage $V_{DS}(t)$.

Neglecting bi-dimensional effects, such as dissimilarities of the two parts of the neutral region ($R_{acc} = R_{ds} = R_d$), neglecting the access region, the parasitic elements like N+N-PN+ transistor and the anti-parallel diode, the equivalent circuit representing the dynamic operation of power MOSFET is depicted in Figure 1.25(b).

In this simplified schematic, the voltage V(L,t) becomes V_{DS} for the channel voltage, when it is under $V_{GS}(t) - V_{th}$. The resistance R_d in $\Omega.cm^{-2}$ is given by:

$$R_d = \frac{1}{q\mu_n N_D} \cdot (W_V - \sqrt{\frac{2\varepsilon_{Si}V_{DS}}{qN_D}})$$

Most of the data sheets give the input capacitance C_{iss}, the output capacitance C_{oss} and the transfer capacitance C_{rss}.

They are measured by a guarded capacitor measurement device, versus voltage V_{ds}, at 1MHz. The guard is made in order to avoid any interference from the other two capacitances, when one capacitance is being tested, as shown in Figure 1.26.

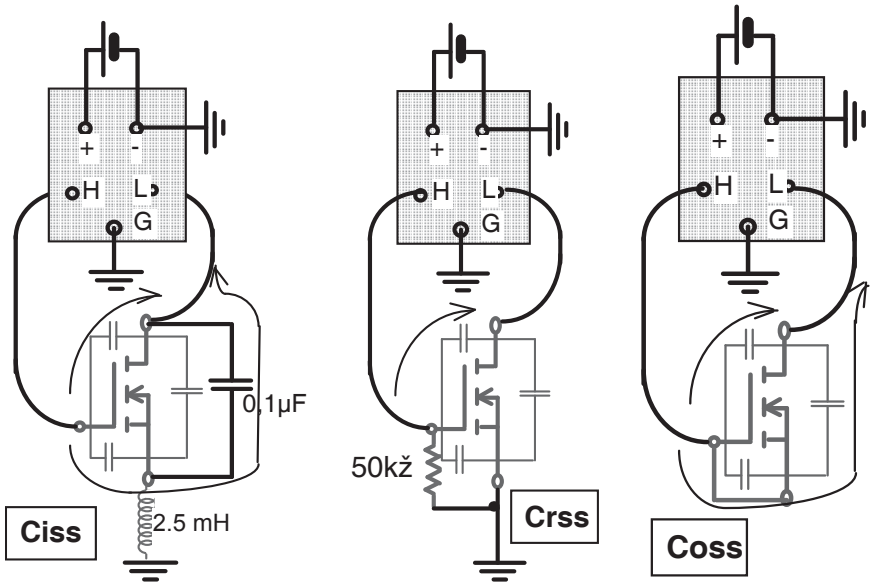


Figure 1.26. *Measuring MOSFET capacitance*

From these measured capacitances, the three capacitances of the equivalent circuit may be calculated as:

$$C_{gd} = C_{rss}, \qquad C_{gs} = C_{iss} - C_{rss} \qquad \text{and} \qquad C_{ds} = C_{oss} - C_{rss}$$

Figure 1.27 shows the capacitances versus V_{ds} for a 400 V, 14 A power MOSFET. We can see that the C_{gs} capacity is constant, while the C_{gd} and C_{ds} capacitances are strongly related to V_{DS}.

If $V_{DS} \approx V^*_{DS}$ is accepted, the following equation can be derived from the equivalent circuit (shown in Figure 1.25(b)):

$$(C_{ds} + C_{gd}).(dV_{DS}/ dt) - C_{gd}\,(dV_{GS}/ dt) = I_D - I_{channel}$$

$$\text{and} \quad (C_{gs} + C_{gd}).(dV_{GS}/ dt) = C_{gd}\,(dV_{DS}/ dt) + i_g$$

These two equations give a simplified one-dimensional model of a power MOSFET in a dynamic regime. More complicated equivalent circuits exist, but this one is suitable for most switching circuit operations.

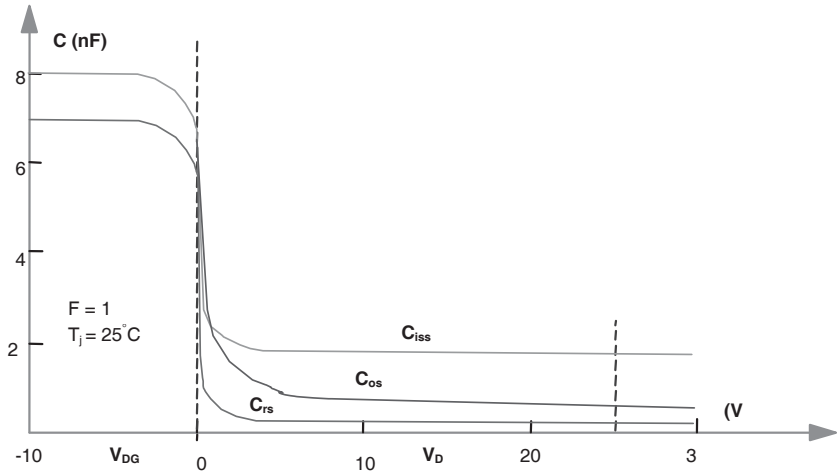


Figure 1.27. *Capacitances versus drain voltage V_{ds}, for a 400 V, 14 A, power MOSFET*

1.4. Power MOSFET main characteristics

Figure 1.28 shows a 500 V, 7 A, 8 Ω MOSFET, where output characteristic $V_D = f(I_D)$, versus gate voltage V_{GS}.

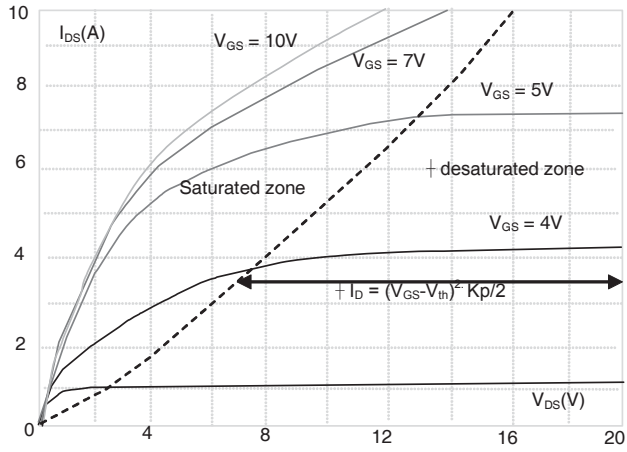


Figure 1.28. *Output characteristic of 500 V, 7 A, 8 Ω power MOSFET*

This chart shows two operating zones: a saturated or resistive zone and an active or linear zone, separated by a dotted line, which corresponds to the maximum current given by the channel, according to the gate voltage V_{gs}.

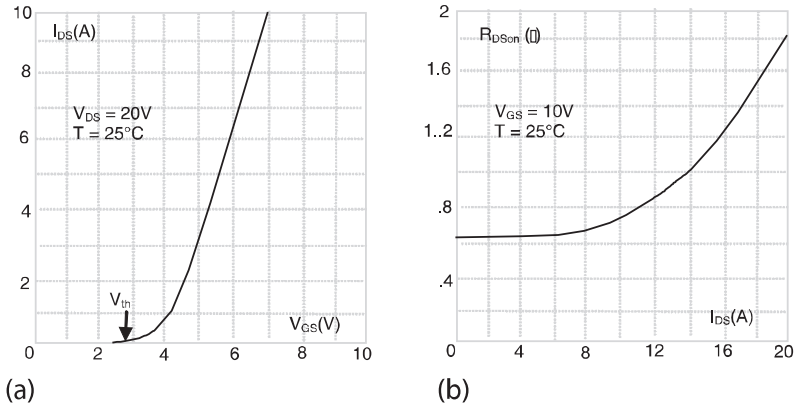


Figure 1.29. *Transfer characteristics and $R_{DSON} = f(I_D)$ for a 500 V, 7 A, 8 Ω MOSFET*

Figure 1.29(a) shows the transfer characteristics I_{DS} versus V_{GS} in the linear zone, with a constant V_{DS} and over V_{GS}-V_{th}. Figure 1.29(b) shows the internal resistance V_{DS}/I_{DS} in the saturated zone, with a constant voltage drive greater than V_{th}. Figure 1.30 depicts the gate charge for a constant drain current.

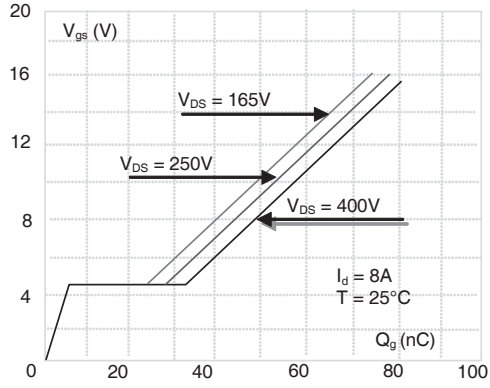


Figure 1.30. *Gate charge of a 500 V, 7 A, 8 Ω power MOSFET*

1.5. Switching cycle with an inductive load

Switching with an inductive load is very common in power electronics.

Figure 1.31(a) shows a power circuit where the inductive load is represented by a perfect current source in parallel with a free-wheeling diode D_{FW}, and driven by a power MOSFET. The gate charge is driven by a perfect voltage source $+V_g$ or $-V_g$, through a resistance R_g. The MOSFET equivalent circuit is given in Figure 1.31b. Figure 1.32 shows various waveforms and Figure 1.33 explains the various phases.

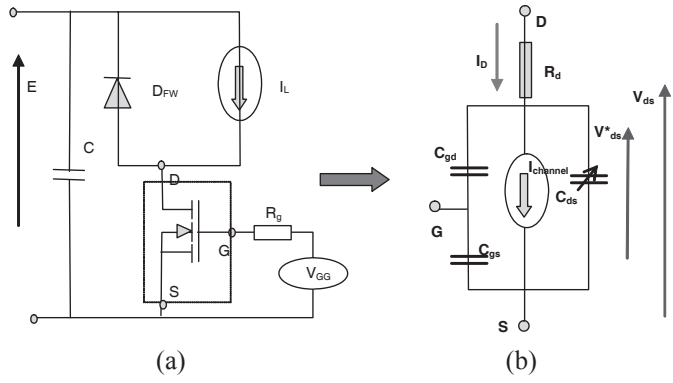
1.5.1. *Switch-on study*

1.5.1.1. *Delay time at switch-on*

At time t = 0, the drive signal $V_{GG}(t)$ changes from zero or a negative voltage to a positive voltage $+V_{GG}$. The gate is charged from the beginning, as indicated by the following exponential function:

$$V_{GS}(t) = V_{GG}\left[1 - 2e^{-\frac{t}{R_g(C_{gs}+C_{gd})}}\right] \approx V_{GG}\left[1 - 2e^{-\frac{t}{R_g C_{gs}}}\right] < V_{th}$$

The gate current $i_g(t)$ decreases from an initial value $2V_{GG}/R_g$. This current $i_g(t)$ includes two parts: the main part charges C_{gs}, and a negligible other part, compared to the first one, discharges C_{gd}. The drain current $I_D(t)$ is zero until the gate voltage $V_{GS}(t)$ raises the threshold voltage V_{th}. The supply voltage remains unchanged.



(a) (b)

Figure 1.31. *Power circuit test and MOSFET equivalent circuit*

1.5.1.2. *Current rise time $t_{r(on)}$*

Drain current $I_D(t)$ starts when the threshold voltage V_{th} is rise. After that, $I_D(t)$ rises following the next law, when gate voltage $V_{GS}(t)$ increases:

$$I_D(t) = (V_{GS}(t) - V_{th})^2 \mu_{ns}.Z.\varepsilon_{ox}/ (2L.d_{ox})$$

Gate voltage $V_{GS}(t)$ increases in the same way as during the delay time:

$$V_{GS}(t) = V_{GG}\left[1 - 2e^{-\frac{t}{R_g(C_{gs}+C_{gd})}}\right] \approx V_{GG}\left[1 - 2e^{-\frac{t}{R_g C_{gs}}}\right] > V_{th}$$

Drain current increases until the current given by the circuit, $I_{ch} + I_{RM}$, rises. During this time, voltage $V_{DS}(t)$ remains at the same value E, and the voltage $V^*_{DS}(t)$ is almost $V_{DS}(t)$. Because the voltage drop inside R_d is tied, the drain current $I_D(t)$ is negligible. When the current I_D is equal to $I_{ch}+I_{RM}$, the following equation is given for the gate voltage $V_{GS}(t)$:

$$I_{ch} + I_{RM} = (V_{GS(Miller)on} - V_{th})^2 \mu_{ns}.Z.\varepsilon_{ox}/ (2L.d_{ox})$$

where $V_{GS(Miller)on}$ is Miller voltage at switch-on. Its value is high when $I_{CH} + I_{RM}$ is large.

1.5.1.3. *Voltage fall time $t_{v(on)}$*

When the load current $I_D(t)$ raises $I_{ch} + I_{RM}$, gate voltage still increases, and so $I_{channel}(t)$ tries to exceed the $I_D(t)$ value imposed by the load. A new relationship occurs between drain current and channel current: $\Delta I = I_{channel}(t) - I_D(t) > 0$. The C_{gd} and C_{ds} capacitances begin to discharge, and $V_{DS}(t)$ decreases. The fall in $V_{DS}(t)$ is given by the following equation:

$$\frac{dV_{DS}(t)}{dt} = \frac{-\Delta I + \frac{C_{gd}}{C_{gd}+C_{gs}} i_g}{C_{ds} + C_{gd} - \frac{C_{gd}^{\ 2}}{C_{gd}+C_{gs}}} \approx \frac{-\Delta I}{C_{ds}+C_{gd}}$$

Here the voltage drop inside R_d is neglected.

Capacitances C_{ds} and C_{gd} are weak at the beginning of the fall in $V_{DS}(t)$, so the initial speed of $dV_{DS}(t)/dt$ is very high, $V_{DS}(t)$ falls progressively, then the

capacitances are increasingly loaded, so the speed $V_{DS}(t)/dt$ becomes smaller and smaller.

When $V_{DS} = V_{GS} \approx 10V$, the capacitances increase substantially and dV_{DS}/dt is very small compared to the static conditions. In the end, gate current $i_g(t)$ is mainly used to charge the Miller capacitance C_{gd} while $V_{GS}(t)$ is quite constant. This phenomenon is called the "Miller effect". When $dV_{DS}(t)/dt = 0$, the Miller effect disappears and $V_{gs}(t)$ can increase up to V_{GG}. Thus:

$$V_{GS}(t) = V_{GG}\left[2e^{-\frac{t}{R_g(C_{gs}+C_{gd})}} - 1\right] > V_{GS(Miller)on}$$

1.5.2. *Switch-off study*

1.5.2.1. *Switch-off delay time $t_{d(off)}$*

Switch-off process is the reverse of the switch-on process. Drive voltage $V_{GG}(t)$ instantaneously goes from $+V_{GG}$ to a ground voltage $V_{GG} = 0$, or a negative voltage $-V_{GG}$. Gate voltage decreases as an exponential function given hereafter, and the gate current $i_g(t)$ increases from an initial value $-2V_{GG}/R_g$.

$$V_{GS}(t) = V_{GG}\left[2e^{-\frac{t}{R_g(C_{gs}+C_{gd})}} - 1\right] > V_{GS(Miller)off} > V_{th}$$

This delay phase occurs until the channel current $I_{channel}(t)$ is able to compensate the load current I_L. When the gate is discharged in such a way that the channel current cannot drive the load current, $V_{DS}(t)$ seriously increases, and so the switch-off delay time ends.

During this delay, the drain current is the load current $I_D(t) = I_L$, which also occurs in the conduction state.

If the gate voltage V_{GS} is defined, when $I_{channel} = I_L$, by $V_{GS(Miller)off}$, the following equation may be written:

$$I_{ch} = (V_{GS(Miller)on} - V_{th})^2 \mu_{ns}.Z.\varepsilon_{ox}/ (2L.d_{ox})$$

where $V_{GS(Miller)off}$ represents the Miller voltage level at switch-off, as $V_{GS(Miller)on}$ for switch-on.

1.5.2.2. *Voltage rise time $t_{v(off)}$*

After the delay time, gate voltage tries to decrease, and $I_{channel}(t)$ has the tendency to be under the load current I_L. So, as for the switch-on, a potential inequality appears between $I_{channel}$ and the load current $I_L = I_D(t)$: $\Delta I = I_{channel}(t) - I_D(t) < 0$. C_{gd} and C_{ds} are discharged and V_{DS} can increase. The rate of increase is defined by:

$$\frac{dV_{DS}(t)}{dt} = \frac{-\Delta I + \frac{C_{gd}}{C_{gd}+C_{gs}} i_g}{C_{ds}+C_{gd}-\frac{C_{gd}{}^2}{C_{gd}+C_{gs}}} \approx \frac{-\Delta I}{C_{ds}+C_{gd}}$$

Note: the voltage drop inside R_d is neglected.

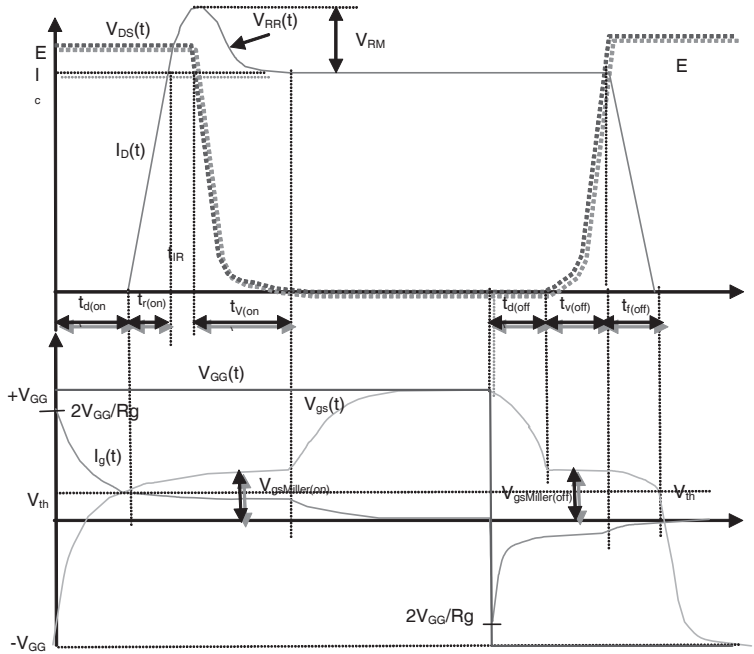


Figure 1.32. *Inductive load switching waveforms*

At the beginning of the V_{DS} voltage rise time, C_{gd} and C_{ds} are large and $dV_{DS}(t)/dt$ is a gradual increase in speed. When $V_{DS}(t)$ increases, capacitances become weaker and the $dV_{DS}(t)/dt$ becomes increasingly larger.

During this time, the C_{gs} discharge current is compensated for by the C_{gd} charge current, maintaining quite a stable V_{gs}: this same Miller effect occurs during switch-on.

The switch-off $t_{v(off)}$ voltage rise time stops when $V_{DS}(t) = E$, the supply voltage.

1.5.2.3. *Fall time current $t_{f(off)}$*

When $V_{DS}(t) = E$, $dV_{DS}(t)/dt = 0$, and the Miller effect disappears.

Thus, the gate can be normally discharged. The load current can now decrease until the gate voltage goes down to zero:

$$I_D(t) = I_{channel}(t) = (V_{GS}(t) - V_{th})^2 \mu_{ns}.Z.\varepsilon_{ox}/(2L.d_{ox})$$

$$V_{GS}(t) = V_{GG}\left[2e^{-\frac{t}{R_g(C_{gs}+C_{gd})}} - 1\right] < V_{GS(Miller)off}$$

Tables 1.1 and 1.2 sum up the various states of $V_{DS}(t)$, $I_{DS}(t)$ and $V_{GS}(t)$ during one switching cycle on inductive load.

For this analysis, the inductive parasitic effects were neglected. Parasitic inductances are not only in the power circuit, but also in the drive loop. Thus, the waveforms given by Figure 1.32 are slightly inaccurate.

Due to C_{gd}, the Miller effect is all the time in opposition to any gate voltage variation. At switch-on, V_{DS} decreases, dV_{DS}/dt is negative and a current, i_{CGD}, is extracted from the gate. At switch-off, V_{DS} increases and injects a current, i_{cgd}, into the gate.

If the gate current is very large, the Miller effect is quite negligible, and the V_{GS} "plateau" during switch-on and switch-off may disappear.

The equation for $dV_{DS}(t)/dt$ is still the same but $V_{GS}(t)$ varies during $t_{v(on)}$ and $t_{v(off)}$, according to the following laws:

$$(C_{gs} + C_{gd}).\, dV_{GS}/dt = C_{gd}.\, dV_{DS}/dt + i_g$$

$$\text{and} \qquad R_g.i_g + V_{GS} = V_{GG}$$

<table>
<tr><td colspan="4">SWITCH-ON</td></tr>
<tr><td rowspan="2">$t_{d(on)}$</td><td>$V_{DS}(t) = E$</td><td colspan="2">$I_D(t) = 0$</td></tr>
<tr><td colspan="3">$V_{GS}(t) = V_{GG}\left[1 - 2e^{-\frac{t}{R_g(C_{gs}+C_{gd})}}\right] \approx V_{GG}\left[1 - 2e^{-\frac{t}{R_g C_{gs}}}\right] < V_{th}$</td></tr>
<tr><td rowspan="2">$t_{r(on)}$</td><td>$V_{DS}(t) = E$</td><td colspan="2">$I_D(t) = \frac{Z}{2L}\mu_{ns}\frac{\varepsilon_{ox}}{d_{ox}}(V_{GS}(t) - V_{th})^2$</td></tr>
<tr><td colspan="3">$V_{GS}(t) = V_{GG}\left[1 - 2e^{-\frac{t}{R_g(C_{gs}+C_{gd})}}\right] \approx V_{GG}\left[1 - 2e^{-\frac{t}{R_g C_{gs}}}\right] > V_{th}$</td></tr>
<tr><td rowspan="2">$t_{V(on)}$</td><td colspan="3">$\frac{dV_{DS}(t)}{dt} = \frac{-\Delta I + \frac{C_{gd}}{C_{gd}+C_{gs}} i_g}{C_{ds} + C_{gd} - \frac{{C_{gd}}^2}{C_{gd}+C_{gs}}} \approx \frac{-\Delta I}{C_{ds}+C_{gd}}$</td></tr>
<tr><td>$I_D(t) = I_{Ch} + I_{RR}$</td><td colspan="2">$V_{GS} = V_{gs(Miller)on}$</td></tr>
<tr><td colspan="4">ON STATE</td></tr>
<tr><td></td><td>$V_{DS} = I_{Ch} \cdot R_{Dson}$
$= I_{Ch} \cdot (R_v + R_a + R_{channel})$</td><td>$I_D(t) = I_{Ch}$</td><td>$V_{GS} = +V_{GG}$</td></tr>
</table>

Table 1.1. *Various parameters variation during switch-on for an inductive load*

<table>
<tr><td colspan="4">SWITCH-OFF</td></tr>
<tr><td rowspan="2">$t_{d(off)}$</td><td colspan="2">$V_{DS} = I_{Ch} \cdot R_{Dson}$
$= I_{Ch} \cdot (R_v + R_a + R_{channel})$</td><td>$I_D(t) = I_{Ch}$</td></tr>
<tr><td colspan="3">$V_{GS}(t) = V_{GG}\left[2e^{-\frac{t}{R_g(C_{gs}+C_{gd})}} - 1\right] > V_{GS(Miller)off} > V_{th}$</td></tr>
<tr><td rowspan="2">$t_{V(off)}$</td><td colspan="3">$\frac{dV_{DS}(t)}{dt} = \frac{-\Delta I + \frac{C_{gd}}{C_{gd}+C_{gs}} i_g}{C_{ds} + C_{gd} - \frac{C_{gd}{}^2}{C_{gd}+C_{gs}}} \approx \frac{-\Delta I}{C_{ds}+C_{gd}}$</td></tr>
<tr><td colspan="2">$I_D(t) = I_{Ch}$</td><td>$V_{GS} = V_{gs(Miller)off}$</td></tr>
<tr><td rowspan="2">$t_{f(off)}$</td><td>$V_{DS}(t) = E$</td><td colspan="2">$I_D(t) = \frac{Z}{2L}\mu_{ns}\frac{\varepsilon_{ox}}{d_{ox}}(V_{GS}(t) - V_{th})^2$</td></tr>
<tr><td colspan="3">$V_{GS}(t) = V_{GG}\left[2e^{-\frac{t}{R_g(C_{gs}+C_{gd})}} - 1\right] < V_{GS(Miller)off}$</td></tr>
<tr><td colspan="4">SWITCH-OFF STATE</td></tr>
<tr><td colspan="2">$V_{DS}(t) = E$</td><td>$I_D(t) = 0$</td><td>$V_{GS} = -V_{GG}$</td></tr>
</table>

Table 1.2. *Various parameters variation during switch-off for an inductive load*

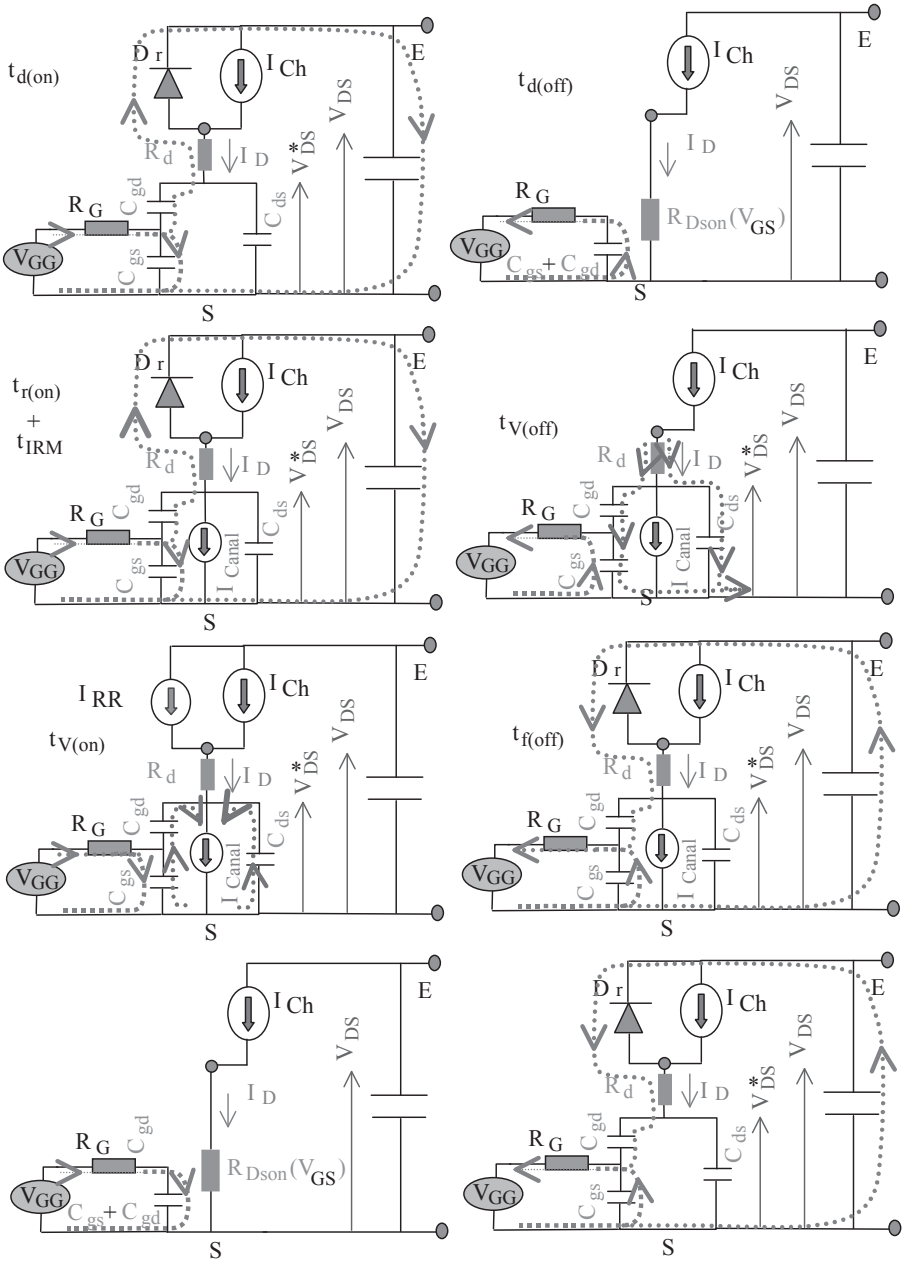


Figure 1.33. *Currents through capacitances Cgs, Cgd and Cds, during one switching cycle, on inductive load, with hard commutation*

1.6. Characteristic variations due to MOSFET temperature changes

The power electronics thermal aspects are of as much importance as the pure electrical aspects, mainly because electrical parameters change with temperature. In a power MOSFET, several parameters are sensitive to the die temperature:

– Kp transconductance, because the surface electrons mobility μ_{ns} decreases according to the following law:

$$\mu_n (T_j)/\mu_n(T_{jo}) = (T_j/T_{j0})^{-1.5}$$

– drift zone resistivity $\rho_v = 1/ q.\mu n.ND$. Due to the electron volume mobility reduction, μ_n increases with temperature according to:

$$\mu_n (T_j)/\mu_n(T_{jo}) = (T_j/T_{j0})^{-2.5}$$

– threshold voltage V_{th} decreases by 6 mV/°C, due to Fermi potential;

– power MOSFET in conduction is characterized by its internal resistance R_{DSon}, which is the sum of three resistances:

- channel resistance, R_{ch},
- access resistance, R_{acc}
- and drift zone resistance, R_v

For low voltage power MOSFET (<100 V), channel resistance, R_{ch} is the most important. Drift zone resistance, R_v is, however, of greater importance for medium and high voltage power MOSFETs (>400 V). Thus, the theoretical resistance R_{dson} versus temperature, from the mobility evolution, can be given by:

$$R_{DSon} (T_J) = R_{DSON}(T_{jo}) (T_j /T_{jo})^{\alpha}$$

where the temperature is given in Kelvin degrees, and where $\alpha = 1.5$ for low voltage MOSFET (<100 V) and $\alpha = 2.5$ for medium and high voltage power MOSFETs (>400 V).

The previous equation may also be written:

$$R_{DSon} (T_j) = R_{DSon} (25°C). (1+ \alpha. \Delta T/ 300)$$

where ΔT, in °C, is the increase of temperature from 25°C. This equation really shows the positive temperature coefficient of power MOSFETs.

Figure 1.34 shows characteristics versus temperature for a 500 V power MOSFET. We can see that between 25°C and 125°C, R_{DSon} doubles while V_{th} reduces by 25%.

For the transfer characteristic, two parameters are in opposition when temperature increases: decreases V_{th}, contrasting the rise of dI_D/dT_j. For the same V_{GS}, reduction of the transconductance parameter K_p has the opposite effect. Thus, dI_D/dT_{jh} includes two reverse behaviors in temperature, versus I_D current.

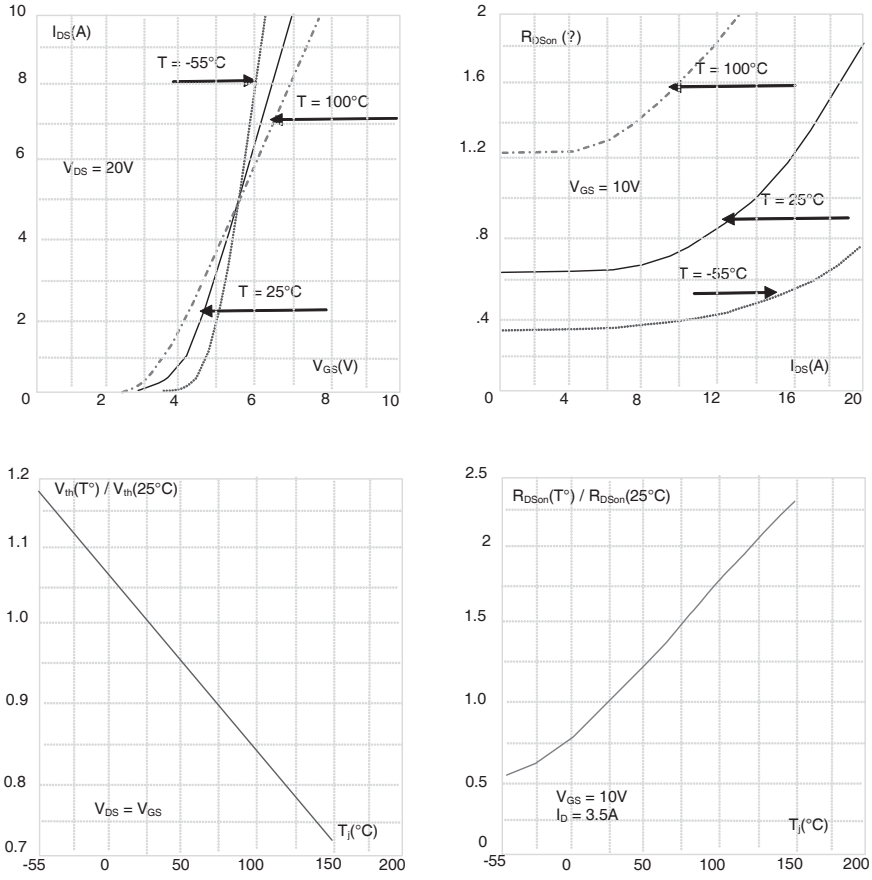


Figure 1.34. *500 V, 7 A, 8 Ω MOSFET characteristics versus temperature*

Switching characteristics show very few variations versus temperature, because V_{th} has a minimal reaction, and the power MOSFET capacitances are insensitive to

temperature. Variation speeds $dI_D(t)/dt$ and $dV_{DS}(t)/dt$ do slow down with temperature, because the parameter of transconductance Kp is reduced.

1.7. Over-constrained operations

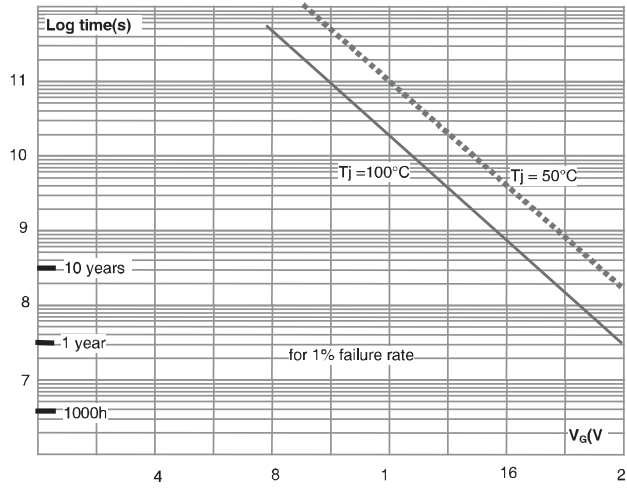


Figure 1.35. *MOS lifetime versus gate voltage V_{GS}*

1.7.1. *Overvoltage on the gate*

Despite a maximum gate voltage of 70 V for an oxide thickness of 100 nm, manufacturers specify a gate voltage V_{GS} limited to 20 V.

Two reasons justify this limitation. On the one hand, Figure 1.35 shows that power MOSFET lifetime decreases substantially when V_{GS} increases. On the other hand, V_{GS} over 20 V gives nothing more for the power MOSFET. So, an overvoltage protection must be set between the gate and the source.

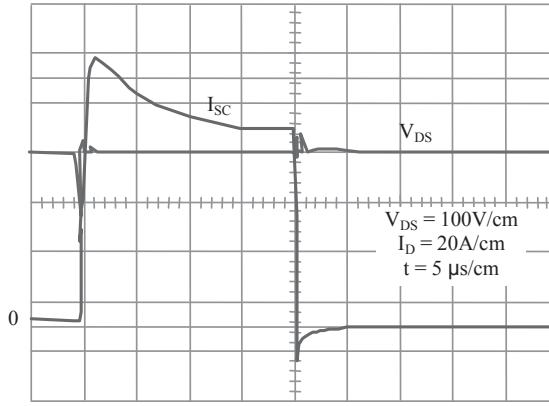


Figure 1.36. *Short circuit on a 500 V, 14 A, 0.4 Ω MOSFET*

1.7.2. *Over-current*

Short circuit is the most common and the most constraining overload, because the device must sustain both the supply voltage and a large drain current I_D in concert. If the current density in the die is considered uniform, and if the critical field in silicon is the cause of the failure, then the electric field is reversed in the N-N+ junction, and the theoretical current density versus short circuit voltage is given by:

$$J_D = q.N_D v_{ns} + v_{ns}.\varepsilon_{si}.E_{simax}^2/2V_{DS}$$

where v_{ns} is the saturation speed of the electrons (10^7cm/s).

If a doping level of $N_D = 4.10^{14}$ cm^{-3} is applied on a 500 V power MOSFET, the critical field E_{simax} is $2.7.10^5$ V/cm, and the critical current density is around 1745 A.cm^{-2} for a voltage of 70% of the maximum voltage (1415 A.cm^{-2} at 100% of V_{DSS}). This value is at least 30 times higher than the nominal current density, which is very far away from the normal short circuit current with a gate voltage under 20 V. So, from an electrical point of view, power MOSFETs can easily sustain a short circuit current with a gate voltage under 20 V, and a supply voltage equal to V_{DSmax}.

Figure 1.36 shows a short circuit with V_{GS} = 15 V for a 500 V, 14 A, 0.4 Ω power MOSFET, including two dies of 25.5 mm^2 each, combining to create 51 mm^2, and so a current density of 220 A.cm^{-2}.

The time the power MOSFET can sustain the short circuit current is only limited thermally. In Figure 1.36, we can see that the short circuit current decreases with time (or increase of temperature), because the channel current decreases with the surface mobility of electrons, which decreases with temperature.

With some simplifications, the approximate dissipated energy in silicon is:

P_{sc} = 350 V x 90 A = 31.5 kW, according to the thermal impedance chart specified for this device.

In Figure 1.37, the maximum temperature rise at the end of the short circuit is:

$$T_j = Z_{th}.P_{sc} + T_c = r(t).\ R_{th(j\text{-}c)}.P_{sc} + T_c = 0.01 \text{ x.7 x } 31.5.10^3 + 25 = 245.5°C,$$

which is not very far away from the critical temperature of silicon:

$$300°C \text{ for } N_D = 4.10^{14} cm^{-3}.$$

When the short circuit current duration is short (<50 µs), we can consider that the energy is adiabatically dissipated inside the small die. The volume of silicon involved corresponds to the space charge volume: W_{ZD} x S_{Si} where W_{ZD} is a function of the short circuit voltage, thus giving:

$$W_{ZD} = \sqrt{\frac{2\varepsilon_{si} V_{DS}}{qN_D}}$$

The maximum temperature in silicon may be estimated as a function of the dissipated energy W_j as shown below:

$$T_j = \frac{W_J}{k_{Si} W_{ZD} S_{Si}} + T_c \qquad \text{with} \qquad W_J = \int_0^{t_{cc}} I_D(t) V_{DS}(t) dt$$

where K_{SI} is the volume calorific capacity of silicon, around 1.767 J $°C^{-1}$. cm^{-3}.

With the previous MOSFET (S_{SI} = 51 mm^2, Z_{ZD} = 34 µm, W_j = 0.63 J), the short circuit current makes a temperature at the short circuit end of 231°C, which is the same as the previous result.

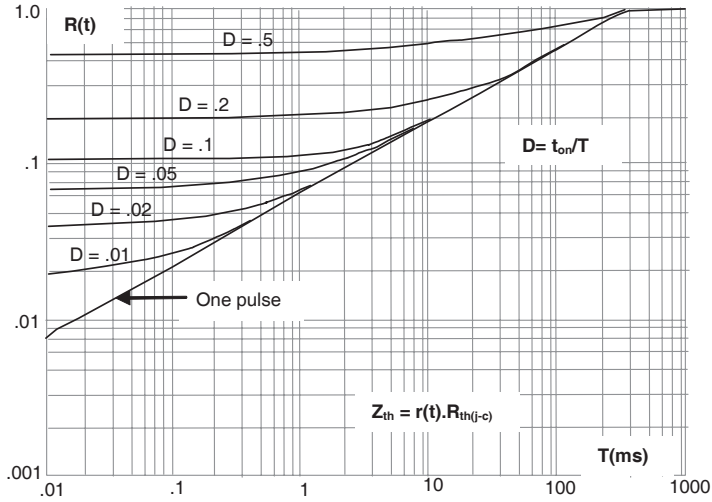


Figure 1.37. *Transient thermal resistance for 500 V, 14 A, 0.4 Ω MOSFET*

1.7.3. *Avalanche sustaining*

Avalanche occurs when the supply voltage connected on the device is over its maximum allowable voltage, which is generally slightly higher than the specified maximum voltage given in the data sheet, shown in Figure 1.38. From a physical point of view, avalanching means the beginning of a non-controlled process, which is more or less a failure activation.

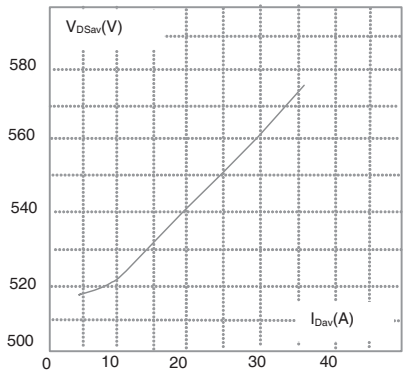


Figure 1.38. *Breakdown voltage versus drain current for a 500 V, 36 A, 0.13 Ω MOSFET*

Nevertheless, if the avalanching current remains under a value which does not lead to too great an increase of silicon temperature, this phenomenon may be reversible. The reason for this is the absence, in a power MOSFET, as a unipolar conduction device, of any second breakdown voltage. In other words, there are no hot spots.

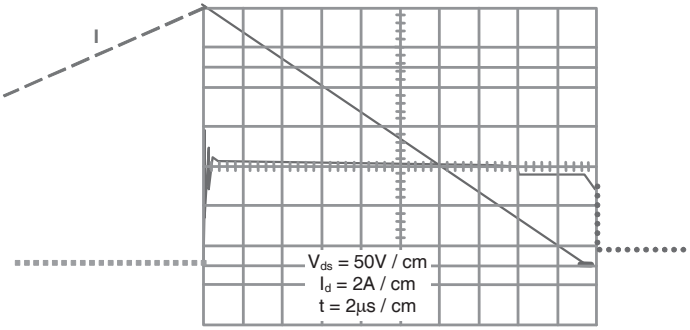


Figure 1.39. *Avalanche test on a 100 V, 12 A, 0.33 Ω MOSFET*

In the data sheets, the avalanche characteristic is shown as a maximum energy that the device is able to sustain at switch-off. The test is done with an inductance of few millihenries, without a free-wheeling diode, depicted in Figure 1.39. This figure shows the device can sustain an avalanche during 20 µs, with 12 A and 130 V, thus an energy of W = 0.016 J. Avalanche phenomenon is dangerous for power semiconductors due to possible lifetime degradation. So, this manufacturer's characteristic should be considered only as an accidental overvoltage strength.

1.7.4. *Use of the body diode*

Power MOSFETs have a structural body diode which may be used as a free-wheeling diode.

It should be interesting to use this diode in a lot of power applications, like ZCS (Zero Current Switch), in soft switching or in hard switching applications, where this diode can be used as a free-wheeling diode. Unfortunately, failures may occur and some limitations must be introduced.

Figure 1.40 shows the diode switch-off and the voltage application. Sometimes, when the speed of application of the supply voltage is over a few kV/µs, a failure appears inside the device. Even today, a true explanation for this phenomenon is still unknown. Maybe, the parasitic bipolar transistor inside the MOSFET is the cause of the failure, but some experiments do not seem in accordance with this explanation.

A large electric field, with a huge gradient, is applied inside the die: this leads to constraints applied on the peripheral of the die, and on the lithography. This could also play a role in explaining failure.

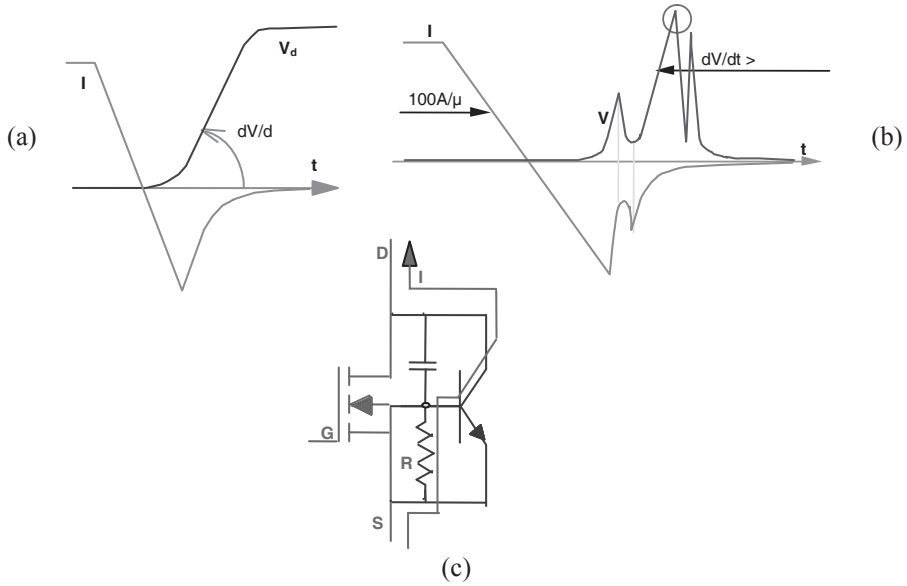


Figure 1.40. *Body diode recovery;*
(a) good recovery; (b) MOSFET fails; (c) parasitic bipolar transistor

1.7.5. *Safe operating areas*

If only the physics of silicon is considered, and if the maximum junction temperature is respected in the die, the theoretical current density is much higher than the load current, even in short circuit. So, in terms of current density in the device, the probability of failure is very low. Practically, due to technological aspects like current density dispersion between cells, packaging, and maximum temperature of various elements, allowable maximum current density is much lower. So, various safe operating graphs are specified in order to guarantee the device lifetime. Two of them are shown in Figure 1.41: the Forward Bias Safe Operating Area (FBSOA) and the Switching Safe Operating Area (SSOA). The same maximum pulsed current, I_{DMmax}, is specified in these two charts: this maximum value is mainly limited by contact devices like bounding wire diameter, metallization and thermal resistance of the encapsulated device. I_{DMmax} is around two to six times the nominal current, and is weaker than the short circuit current. In the

following graph, we can see that the I_{DMmax} is 4.3 times the nominal current, with a current density of 118 A.cm^{-2}. Nevertheless, this device sustained around 160 A.cm^{-2} during short circuit.

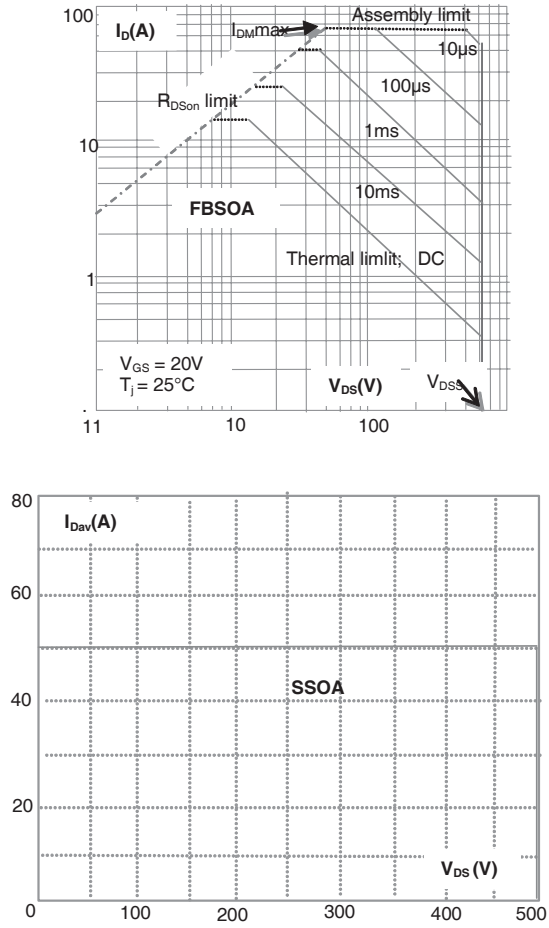


Figure 1.41. *500 V, 14 A, 0.4 Ω MOSFET safe operating areas*

For the FBSOA, four limits can occur:

The maximum sustaining voltage, V_{DSS}. Normally, this is specified slightly under the avalanche voltage of the die, but a voltage applied over this value is not good for device lifetime.

The junction temperature limit, which is generally between 150°C and 175°C. This limit decreases when the pulse duration (Δt) increases. Maximum dissipated energy in the die is:

$$W = U.I.\Delta t$$

and is constant for a same die size.

The maximum current, I_{DM}, also varies with the pulse width, due to the same thermal limitations: requested time for the thermal energy to reach the wires during short pulses, and maximum thermal capacity of the elements.

R_{DSon} limit varies because the voltage drop on the MOSFET is limited by: I_D. R_{DSon}.

1.8. Future developments of the power MOSFET

Voltage drop of power MOSFET is defined by R_{DSon}, and dynamic performances by the gate oxide. Thus, the main target, for designers, is to reduce internal resistance and capacitance, keeping ruggedness of the MOSFET at least at the same level.

In a low voltage power MOSFET, the main part of R_{DSon} is set by the channel and access resistances. Thus, cell density should be increased. However, a JFET then appears between adjacent cells, which limits the available current when cells are too close. The technological limit is around 45,000 cells/mm^2. Thus, a vertical gate is the current solution for channel resistance reduction, depicted in Figure 1.42.

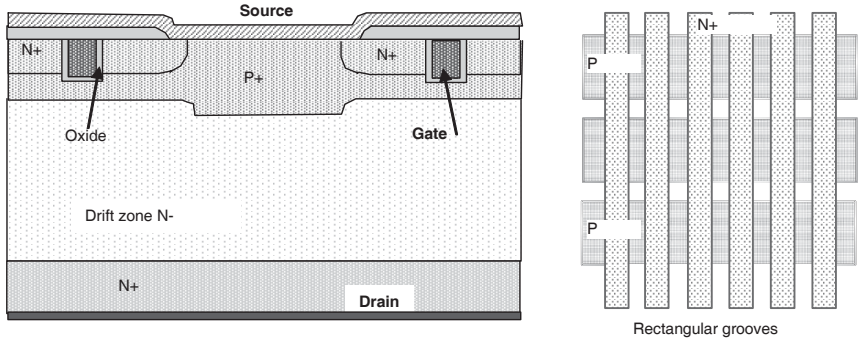


Figure 1.42. *Trench gate MOSFET*

RMOS from Matsushita, and RMOS, or Trench Gate from General Electric, provided the solution: a vertical way for current avoids any access resistance, and the short distance of the gate oxide close to N- makes a low $C_{gd,}$ through C_{oxd} reduction. Gate grooves cannot be as close as could be expected, as P+ contact needs enough place. An enhancement was made by crossing gate grooves, N+ sources and P well. The gate groove is approximately 3 µm in width, with an approximate 5 µm spacing around. Substrate resistance is around 10% of the total resistance of a low voltage MOSFET: its thickness is designed in order to insure the mechanical resistance of the wafer. A new technology, called "thin wafer", could provide further improvement of R_{DSon}.

For a high voltage power MOSFET, R_{DSon} is mainly due to the N- drift zone. Its resistivity is tied to the sustaining voltage, so there are few possibilities for resistance improvement.

Recent research into bipolar diodes led to the discovery of a new technology called "superjunction", which may dramatically improve drift zone resistance. The trade names for these new MOSFET are "Coolmos" or "MDmesh". P+ and N+ islands are introduced into the N- region in order to decrease the R_{DSon}, for the same breakdown voltage, as shown in Figure 1.43.

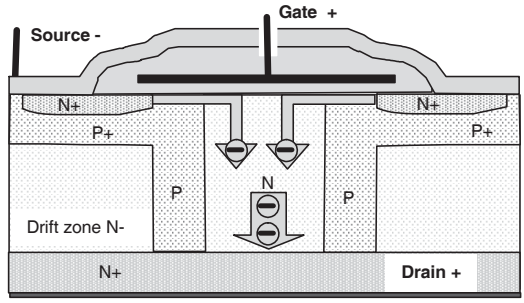


Figure 1.43. *"Superjunction" MOSFET*

At the beginning of switch-off, space charge starts between the N+ and P+ islands, and at the completion of switch-off this space charge spreads out into the total drift zone. The electric field is not only in the vertical part, but also in the horizontal part: this leads to a tri-dimensional structure. Thus, N- drift zone resistivity is no longer tied to $V_{DSS}^{2.5}$. This new technology divides the R_{DSon} of medium and high voltage power MOSFETs by an approximate factor of five. This makes them competitive compared to 500 V IGBT, but also compared to 1,000 V IGBT. The R_{Dson} of a 1,000 V MOSFET may be in the range of 0.2 Ω.cm, at 100°C. This means a voltage drop of around 2 V, with a drain current of 10 A.

The main drawback is the cost of such a technology, where P+ and N+ islands must be introduced in the N- zone.

In a high voltage power MOSFET, the Miller capacitance C_{gd} may also be reduced by decreasing the gate oxide capacitance C_{oxd} over the N- drain. For this, oxide thickness over the N- zone is increased. This technology requires sufficient space between cells, which can only be made within high voltage devices, as shown in Figure 1.44.

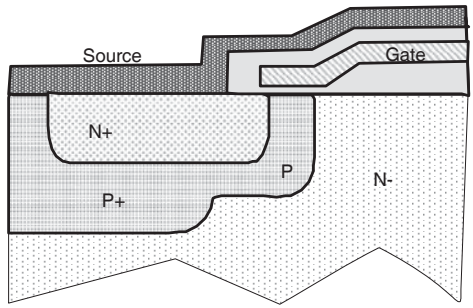


Figure 1.44. *"Up gate" MOSFET*

1.9. References

[ALO 87] ALOISI P., "Savoir utiliser les MOS de puissance", *Electronique de puissance*, vol. 15, June 1987.

[ALO 86] ALOISI P., Power switch, Motorola Inc., 1986.

[ARN 92] ARNOULD J., MERLE P., *Dispositifs de l'électronique de puissance*, Hermès, 1992.

[BAL 07] BALIGA B.J., *Modern Power Devices*, John Wiley & Sons, 1987.

[BUS 99] BUSATTO G., PERSIANO G.V., IANNUZZO F., "Experimental and numerical investigation on MOSFET's failure during reverse recovery of its internal diode", *IEEE Transactions on Electron Devices*, vol. 46(6), 1999.

[DAV 85] BLACKBURN D., "The turn off failure of power mosfet", *IEEE Power Electronic Specialists Conference*, Toulouse, 1985.

[ERI 80] ERIKSON, BEHEN, MIDDLEBROOK, "Characterization and implementation of power MOSFET in switching converter", *Proceedings of Powercon* 7, March 1980.

[FIS 97] FISCHER K., SHENAI K., "Electrothermal effects during unclamped inductive switching (UIS) of power MOSFET's", *IEEE Transactions on Industry Applications*, vol. 44(5), 1997.

[GAM 80] GAMBOA ZUNIGA M., MOS de puissance, relaxation thermique, PhD Thesis, University of Toulouse, 1980.

[GHA 85] GHARBI M., La tenue en tension et le calibre en courant du transistor MOS vertical dans la gamme des tensions (300 V à 1 000 V), Thesis, University of Toulouse, 1985.

[GHA 87] GHANDHI S.K., *Semiconductor Power Devices*, John Wiley & Sons, 1987.

[GRA 89] GRANT D.A., GOWAR J., *Power MOSFET Theory and Applications*, John Wiley & Sons, 1989.

[GRO 67] GROVE A.S., *Physics and Technology of Semiconductor Devices*, John Wiley & Sons, 1967.

[GUE 79] GUEGAN G., Etudes des propriétés dynamiques du transistor MOS à canal vertical, PhD Thesis, University of Toulouse, 1979.

[INT 95] INTERNATIONAL RECTIFIER, HEXFET deseigner's manual, section application notes, *HDM-1 vol. III*, USA, 1995.

[KAL 89] KALKEN T., "Overcurrent protection for power MOSFET", *EPE*, Aachen, 1989.

[LOR 98] LORENZ L., MARZ M., "CoolMOS^{TM} – A new approach towards high efficient power supplies", *Proceedings of PCIM*, 1998.

[MOT 96] MOTOROLA, Power MOSFET transistor data, section 1, theory and application, USA, 1996.

[NAP 91] NAPIERALSKA M., Modélisation du transistor V.DMOS pour simulation de circuits en électronique de puissance, PhD Thesis, INSA, Toulouse, 1991.

[NAP 91] NAPIERALSKA M., TRANDUC H., CORDONNIER C.E., BERRY J.P., ROSSEL P., "Power MOSFET's library builder switching circuits simulation and design", *Proceedings of the EPE MADEP Conference*, Florence, September 1991.

[ROU 88] ROUSEL P., MAIMOUNI R., BELABADIA M., TRANDUC H., CORDONNIER C.E., Bairanzade M., "Power MOSFET models for switching circuit", *Journal de physique*, vol. 49(9), 1988.

[SGS 85] SGS, Power MOS devices, technical notes, Italy, 1985.

[SIE 88] SIEMENS, SIPMOS data, technical information, Germany, 1988.

[SZB 81] SZE S.M., *Physics of Semiconductor Devices*, John Wiley & Sons, 1981.

[SZB 85] SZE S.M., *Semiconductor Devices Physics and Technology*, John Wiley & Sons, 1985.

[TIE 82] TIEN PHAN PHAM, Le compromis entre la résistance à l'état passant et la tenue en tension dans le transistors MOS de puissance, Thesis, University of Toulouse, 1982.

[TRA 81] TRAN DUC H., "The on-state résistance versus the breakdown voltage", *Proceedings of the ESSDERC*, Toulouse, 1981.

[WIL 87] WILLIAMS B.W., *Power Electronics, Devices, Drivers and Applications*, John Wiley & Sons, 1987.