Evolution of Silicon Power Devices and Challenges to Material Limit

Akio Nakagawa

Abstract  The author first briefly reviews recent success of MOS gate power devices. The main objective is to predict, for the first time, the silicon limit characteristics of IGBTs for its on-resistance and SOA. The author also proposes ideal gate drive in order to realize the ultimate limit of high speed switching of MOS gate power devices. The results lead to new FOM, characterizing the high speed switching capability of various power devices.

Silicon devices have still great potential, competing with emerging new material devices.

I. EVOLUTION OF POWER DEVICES

Power devices have evolved so rapidly that 3.3kV IGBTs have even replaced 4.5kV GTOs, which was developed in the late 80’s for traction control of bullet trains. Figure 1 and 2 show application fields of power devices in 1997 and 2005, respectively. The distinguished difference of the two figures is that most of the applications of GTO and BTr have been occupied by IGBT and its module.

Figure 3 shows the evolution of high voltage large current power devices in Toshiba. The lifetime of GTOs was as short as only 12 years. Nowadays, MOS gate devices are predominantly used in almost all of the application fields, including LDMOS in power ICs, MOSFETs for low voltage and medium voltage applications and IGBTs for high power applications.

New material SiC and GaN devices are being developed in order to break through the silicon limit. In the mean time, super junction devices were proposed and developed in 1998[1]. The super junction MOSFETs already broke through the so called silicon unipolar device limit in the voltage range from 200V to 700V and significantly enhanced the potential of silicon devices.

Another recent remarkable advancements are high speed trench power MOSFET, intelligent power module(IPM) and power IC technologies. Trench MOSFET switching speed has been greatly improved since 1999 in order to meet the requirement of high efficiency and high di/dt of Voltage Regulator Modules for CPUs. The details are described in Section III. Progress in IPM was already reviewed in Ref.[2]. Power IC technologies are classified into two categories. One is high voltage SOI power ICs[3] for monolithic DC motor control ICs and PDP flat panel display drivers. Figure 4 shows first 500V 1 and 3 ampere SOI one chip inverter ICs for DC motor control in 1994 and 1999[4]. The other is low voltage power ICs[5]. Figure 5 shows feature size trends in BCD power IC. Fine design rule is now often required in system power ICs for mobile equipments such as cell phones and automotive field.

A.Nakagawa is with Semiconductor Company, Toshiba Corporation, 580-1, Horikawa-Cho, Saiwai-ku, Kawasaki, 212-8520, Japan. E-mail: akio.nakagawa@toshiba.co.jp
II. IGBTs

A. Brief History of IGBT

Concept of IGBTs was first described in the patent by Becke et al.[7]. The actual fabrication was reported by Baliga et al. in 1982[8]. Since then numerous papers were published to improve the device characteristics, such as first switching speed[9,10] and large current capability[11].

In early development stage, IGBTs suffered from the latch-up of the parasitic thyristors and the poor current capability. Non-latch-up IGBTs, satisfying the concept of Becke, were demonstrated for the first time, in 1985[12]. Figure 6 shows the first demonstration of short circuit capability of non-latch-up IGBTs.

IGBT electrical characteristics have been steadily improved. Figure 7 shows major technology achievements in IGBT history. However, there is no prediction of IGBT limit characteristics attributed to silicon material. The author, for the first time, predicts the silicon limit of IGBT.

B. Theory for silicon limit of IGBTs

This section proposes a theory to achieve the lowest forward voltage drop in IGBTs and proposes a new trench gate IEGT/IGBT, realizing the theoretical limit.

The adopted assumption is asymmetrical conduction: “all of the current flows by electrons.” Holes contribute only to the conductivity modulation. From the assumption of no hole current flow, the following equations are valid under the high injection condition:

\[ J_p = qD_p \frac{\partial p}{\partial x} - q\mu p E = 0 \quad \text{-----Eq.(1)} \]

\[ J_n = J_{total} = 2 \times qD_n \frac{\partial n}{\partial x} \quad \text{-----Eq.(2)} \]

\[ E = \frac{kT}{q} \frac{1}{n} \frac{\partial n}{\partial x} \quad \text{-----Eq.(3)} \]
The situation is satisfied by assuming that the carrier density distribution is approximately a linearly decreasing function from cathode to anode. The current density(J) - voltage(V) relation of the proposed IGBT can be derived by integrating Eq.(2) with $D_n = \frac{a}{n+b}$, $a=3.24e18$, $b=9.39e16$ and appropriate boundary conditions[13].

$$V_F = \frac{2kT}{q} \ln \left( \frac{1}{n_i} \left\{ \left( \frac{QJ}{qD_n} + b \right) \exp \left( \frac{JW}{2qa} \right) - b \right\} \right) + R_{on}J$$

**Fig.8** V-I curve comparison between proposed theory and conventional 600V IGBTs.

The induced electron density is greater than $1x10^{17}$ cm$^{-3}$, and effectively blocks the hole current flow, realizing electron injection efficiency of more than 0.9. The proposed narrow mesa IGBT realizes a low forward voltage even with the p-emitter of very low injection efficiency. The details will be published in ISPSD 2006[14].

Figure 10 compares the on-resistance of the proposed IGBT with state of the art devices. The proposed IGBT successfully reduces its on-resistance to below SiC limit for over 1.5kV.

**Fig.10** The proposed IGBT, denoted as “IGBT limit,” is compared with state of the art devices. Predicted IGBT limit surpasses so called SiC limit for over 1.5kV range.

C. Design for Large Electrical Short-Circuit SOA

Achieving a large safe operating area is one of the big concerns for IGBT development. Short-circuit SOA is especially important for motor control application. In this section, the author shows a theoretical basis that IGBTs have a potential of infinitely large SOA. In fact, in 1996, Hagino et al. reported very high critical power density of 2MW/cm$^2$[15] for short-circuit withstand capability. However, no theory has been presented, so far, how to design such large short-circuit SOA in IGBTs.

It is generally a good assumption in PTIGBT that the ratio of the hole current density($J_p$) over the total current density($J$) does not change throughout the high field region in the n-base. The ratio is equal to the anode efficiency $\gamma$ if the high field reaches the n-fuffer.

In the present paper, the anode efficiency $\gamma$ is defined as the ratio of the hole current over the total current at the n-base n-buffer junction, being identical to the product of p-emitter injection efficiency $\gamma_{PE}$ and transport factor in the n-buffer $\alpha_T$. The electron and hole densities can be calculated by the following equations in the high field region.

$$\gamma = J_p/J, \quad p = J_p/qv_h, \quad n = J_p/qv_e,$$

where $p$ and $n$ denote hole and electron densities, $v_h$ and $v_e$ denote hole and electron saturation velocities, respectively. The net charge in the high electric field region $\rho$ is given
by Eq.(5) with the donor density \( N_D \).
\[
\rho = N_D + J_S \gamma \left[ \frac{1}{\varphi_n} \ln \left( 1 + \frac{p-n}{N_D} \right) \right] - \frac{\gamma J}{q} \quad \text{-------Eq.(5)}
\]
\[
\chi = \frac{\varphi_n}{\gamma} \left( \frac{1}{\varphi_n} \ln \left( 1 + \frac{p-n}{N_D} \right) \right) \quad \text{(high field case)} \quad \text{-------Eq.(6)}
\]

If \( \gamma \) is lower than \( \gamma_c \), the second term (mobile charge) in Eq.(5) is negative. The net charge \( \rho \) decreases as the current density \( J \) increases, and eventually changes its sign when \( J \) exceeds the critical current density \( J_c \): \( J_c = qN_D/(1-\gamma/V_e - \gamma/V_h) \) \quad \text{-------Eq.(7)}

Once the net charge becomes negative, the peak high electric field appears at the n-base n-buffer junction. It should be noted that the electric field is uniform in the n-base when \( J = J_c \).

Figure 11 shows the electric field build-up in the n-base n-buffer junction in a 1200V IGBT as \( J \) increases. Avalanche breakdown will take place when the peak electric field in the n-base n-buffer junction exceeds the critical value \( E_c \). This phenomenon is very similar to the second breakdown in npn bipolar transistors.

SOA locus can be predicted by calculating the \( p^n \) diode breakdown voltage, assuming that the impurity concentration of the n-region is the same as \( p \) given by Eq.(5). It is also assumed that the breakdown occurs at the critical peak field \( E_c \) of 1.8x10^5 V/cm. Figure 12 compares analytical results and TCAD results of 600V thin wafer PTIGBTs with a low \( \gamma \). The arrows indicate the breakdown voltage points, caused by the high electric field at the n-base n-buffer junction. Good agreement is seen for high current density region. For low current density region, estimation of breakdown voltage based on the constant critical field, \( E_c \), is not adequate.

It is predicted that high short-circuit withstanding voltage can be obtained if the saturated collector current is approximately the same as the critical current density, \( J_c \), determined by Eq.(7). This is because the flat uniform electric field distribution is realized in the n-base, and the electric field magnitude can be minimized and the impact ionization can also be minimized.

The value of \( J_c \) simply increases as \( \gamma \) approaches \( \gamma_c \).

Thus, analytically predicted short-circuit SOA increases enormously as \( \gamma \) approaches \( \gamma_c \) as shown in Fig.13. Electrical SOA can be sufficiently large if the device is designed so that the adequate \( \gamma \) is realized.

Fig.12 TCAD results are compared with analytical theory, shown by solid line. Dotted line indicates the locus, determined by TCAD, where significant impact ionization at n-base n-buffer junction starts to occur.

Fig.13 Analytically predicted SOA locus increases significantly as \( \gamma \) approaches \( \gamma_c \). The dotted line shows the locus \( J = C/[W_N \exp(bW_N/V)] \), under which no significant impact ionization occurs if \( \gamma \) is optimized.

Maximum SOA locus can be predicted also by calculating the impact ionization current under the assumption that an optimum \( \gamma \) is chosen and that the flat and uniform electric field is realized in the n-base. The maximum SOA is defined as the area where the impact ionization current density, \( J_{\text{imp}} \) is limited to below a constant value.

\[
J_{\text{imp}} = \int J_{\alpha_n} \exp(-\frac{b}{E}) \, dx = (1-\gamma) J \alpha_n W_N \exp(-\frac{bW_N}{V}) < \text{Const}
\]

where \( W_N \) denotes the n-base width and the impact ionization by hole current is ignored. If the electric field is constant and sufficiently small, the integral is easily
evaluated. The SOA boundary locus is expressed as:

$$J = C/[W_n \exp(-bW_n/V)],$$

----------Eq.(8)

where $W_n$ is the n-base width, $C$ and $b$ are constants. In Fig.13, the dotted line shows the SOA locus given by Eq.(8) where $I_{\text{imp}}$ is assumed to be approximately 200A/cm$^2$, which corresponds to the generation rate of $2.5 \times 10^{12}$cm$^3$ For example, a point of 500V and 10$^4$A/cm$^2$ is within the SOA locus and 5x10$^6$W/cm$^2$ power dissipation can be allowed. More concrete device design based on TCAD will be presented in ISPSD2006[14].

![Image](image1.png)

Fig.14 Toshiba’s roadmap for high speed MOSFETs

### III. POWER MOSFET

#### A. Recent advancement in MOSFET

Since 1999, switching speed of power MOSFETs has been greatly improved for the application of VRM (Voltage Regulator Module) for CPUs. The FOM of $R_{dsQ\text{gd}}$ is conventionally adopted for high speed MOSFETs as design guide.

Figure 14 shows the Toshiba’s roadmap of 30V power MOSFET. $R_{dsQ\text{gd}}$ was improved from 160mΩnC in 1999 to 30mΩnC in 2005. The buck converter efficiency was improved from 85.5% in 2000 to 90% in 2004.

As the on-resistance of MOSFETs decreases, it is recognized that the package impedance itself occupy a large part of the total on-resistance. Recently, new packages adopt metal ribbons in place of bonding wires, shown in Fig.15.

![Image](image2.png)

Fig.15 Low impedance package using Aluminum ribbon.

#### B. Multi-chip Module

If one try to pursue higher efficiency of synchronous buck converters, it is recognized that (1) reduction in parasitic inductances of the power stage circuits, (2) prevention of self-turn-on of low-side MOSFET, and (3) dead-time optimization are equally important, as compared with the improvement of MOSFETs, themselves.

![Image](image3.png)

Fig.16 Analyzed circuit of buck converter

![Image](image4.png)

Fig.17 Influence of Parasitic inductances on converter efficiency

Figure 17 shows the influence of each parasitic inductance on buck converter efficiency[17]. The each parasitic inductance is defined in the circuit in Fig.16. The most influential one is the high side MOSFET source inductance, $L_{\text{HS}}$. If the MOSFET is turned-on, the drain current increase rate, $dI/dt$, induces the voltage drop in the parasitic inductance $L_{\text{HS}}$. The voltage applied by the gate driver circuit is the sum of the actually applied MOSFET gate voltage and the voltage drop in the inductance $L_{\text{HS}}$. The high $dI/dt$ reduces the actually applied gate-source voltage, resulting in the delayed turn-on. In order to realize the first switching-on, the parasitic source inductance should be minimized. The dotted line in Fig.16 should be adopted for the gate driver ground connection.

The other parasitic inductances increase voltage spike in the switching transients of high side MOSFET and increases the power loss of the high side MOSFET.
The parasitic inductances include the ones inside the package and the ones in the PCB board. In order to reduce the parasitic inductances and resistances, multi-chip module was introduced. Figure 18 shows MCM, called DrMOS, proposed by Intel. Three chips of high-side and low-side MOSFETs and the driver circuit are mounted in the single package, thus minimizing the parasitic impedances. DrMOS improves converter efficiency by 2 or 3% as shown in Fig.19, even if the same rated MOSFET chips are used.

![Fig.18 MCM (DrMOS)](image)

C. Future technology for Multi-Chip-Module

In the conventional gate drive circuit, switching speed is determined by \( \frac{Q_{sw}}{I_g} \).

\[
P_{loss} = R_{on} I_{D}^2 + I_D V_D \frac{Q_{sw}}{I_g} f + \frac{1}{3} Q_{dv} V_D f + Q_G V_G f \]

--- Eq.(9)

where 1\(^{st}\), 2\(^{nd}\), 3\(^{rd}\) and 4\(^{th}\) terms show on-state loss, switching loss, main junction capacitance loss and gate charge loss, respectively. The main junction capacitance loss, \(Q_{dv}V_D/3\), is added. Here, \(Q_{dv}\) denotes output charge, \(Q_{un}\). The coefficient is 1/3 not 1/2. The reason is described in Appendix.

If the power loss is determined by the first two terms in Eq.(1), the product of \(R_{on}\) and \(Q_{sw}/I_g\) can be used reasonably as figure of merit (FOM) for MOSFETs.

If the gate drive circuit impedance is assumed to be very low and if the value of \(Q_{sw}/I_g\) is negligibly small, the 2\(^{nd}\) term in Eq.(9) disappears and the switching loss is determined only by the main junction capacitance and the 3\(^{rd}\) term expresses the switching-off loss as shown in Eq.(10).

\[
P_{loss} = R_{on} I_D^2 + \frac{1}{3} Q_{dv} V_D f + Q_G V_G f \]  --- Eq.(10)

Again, if the gate loss assumed to be ignored compared with the first two terms, \(R_{on}Q_{dv}\) is regarded reasonably as new FOM[6].

Figure 21 shows the MOSFET switching-off simulation results using the very low impedance gate drive. The gate and the source is shunted with a 1m\(\Omega\) resistor. The turn-off time is only 2ns, which corresponds to the \(Q_{dv}V_D/3\) value. There is no plateau in the gate voltage waveform, originating from \(Q_{dv}\) value. The switching-off loss can be minimized by the low impedance gate drive[18].

The low impedance gate drive is especially effective for MCM with low voltage MOSFETs or monolithic
solution, because much faster switching of 2nsec will be realized by MCM or ICs. It should also be emphasized that the self-turn-on of the low side synchronous MOSFET can be prevented by the low impedance gate drive.

The low impedance gate drive proposed in this section is explained in the following way: If the channel inversion layer still remains and conducts electron current after the depletion layer is formed in the main junction in the turn-off transient, joule loss occurs in the depletion layer. This makes the major switching loss and is expressed by the 2nd term in Eq.(9). The concept proposed in this section is that the MOS gate channel should cease before the main junction starts to recover so that no joule loss occurs in the switching-off transient but just the main junction capacitance is charged or the main junction recovers. The charged main junction capacitance is discharged and joule loss occurs in the turn-on transient.

It should be noted that the switching-off loss, \(Q_{\text{str}}V_D/3\), does not depend on the magnitude of the drain current. This is the distinguished difference from the conventional switching, whose switching time depends on the \(Q_{\text{sd}}\) value, which increases as the drain current increases.

It should be emphasized that it is difficult to reduce the turn-on loss even by the proposed gate drive method. This is because the drain current starts to flow through the main junction depletion layer immediately after the channel inversion layer is formed. This makes joule loss until the device forward voltage becomes low enough. The device forward voltage is determined by the outside circuit condition. Thus, the total power loss is expressed by Eq.(11).

\[
P_{\text{loss}} = R_{\text{on}}f_D^2 + \frac{1}{3} Q_{\text{str}}V_Df + p_{\text{parasitic}} f + Q_{C}V_Df \quad \text{-----Eq.(11)}
\]

Figure 22 compares the efficiency of DCDC converters for the two cases, where conventional gate drive circuit and the extremely low impedance gate drive circuit are used. In the ideal gate drive condition, the efficiency will improve and achieve more than 90% at 20A output current even if the same MOSFETs are used. These results imply that efficiency in DC-DC converters is still expected to be improved in future. Thus, low voltage trench MOSFETs will be still mainstream for these applications.

It is definitely important to reduce parasitic stray inductances in the power stage circuits and to adopt voltage clamping method in order to reduce voltage spikes caused by the high speed switching. The combination of low impedance gate drive and MCM or monolithic IC technology will provide the solution. A good method to realize the ideal gate drive circuit of very low impedance is to integrate the driver circuit within the power MOSFET chip itself. This can be easily realized using lateral MOSFET and BCD power IC technology. High speed switching can be easily realized in the integrated solution as seen in Fig.23.

\[
NFOM = R_{\text{on}}Q_{\text{str}} = T_{\text{sw}}V_F \quad \text{-----Eq.(12)}
\]

\(Q_{\text{str}}\) is stored carrier quantity between the drain and the source. For MOSFETs, \(Q_{\text{str}}\) is equivalent to \(Q_{\text{ds}}\). Switching time, \(T_{\text{sw}}\), is expressed as follows:

\[
T_{\text{sw}} = \frac{Q_{\text{str}}}{I_D}, \quad V_F = R_{\text{on}}I_D \quad \text{-----Eq.(13)}
\]

Fig.22 Prediction of converter efficiency using ideal gate drive

D. New FOM

In this section, we introduce a new figure of merit(NFOM) [6] based on the discussions in the previous section.

\[
NFOM = R_{\text{on}}Q_{\text{str}} = T_{\text{sw}}V_F \quad \text{-----Eq.(12)}
\]

\(Q_{\text{str}}\) is stored carrier quantity between the drain and the source. For MOSFETs, \(Q_{\text{str}}\) is equivalent to \(Q_{\text{ds}}\). Switching time, \(T_{\text{sw}}\), is expressed as follows:

\[
T_{\text{sw}} = \frac{Q_{\text{str}}}{I_D}, \quad V_F = R_{\text{on}}I_D \quad \text{-----Eq.(13)}
\]
$Q_{\text{str}}$ and $R_{\text{on}}$ are represented by the following equations, assuming ideal $R_{\text{on}}$ and the applied voltage being near the breakdown voltage of the device:

$$Q_{\text{str}}=\varepsilon E_c$$  \hspace{1cm} \text{Eq. (14)}

$$R_{\text{on}}=4V_{\text{BD}}^2/\varepsilon iE_c^3$$  \hspace{1cm} \text{Eq. (15)}

$$\text{NFOM}=4V_{\text{BD}}^2/\varepsilon iE_c^2=4V_{\text{BD}}^2/BHFOM$$  \hspace{1cm} \text{Eq. (16)}

It can be shown that NFOM is closely related to the BHFOM under special assumptions. NFOM can be defined specifically for each device including bipolar device by using the stored carriers, $Q_{\text{str}}$, and the on-resistance. This feature is the distinguished difference from the BHFOM.

It should be noted that $Q_{\text{str}}$ depends on the operating condition just like $Q_{\text{str}}$. Equation (14) assumes that the applied voltage is the same as the breakdown voltage.

As for 30V silicon MOSFET, device simulation shows that the turn-off time is expected to be 2 nsec by ideal gate drive. This value coincides with the value calculated from $Q_{\text{str}}/I_D$.

Figure 24 shows the comparison of NFOM among silicon devices and new material devices. NFOM of IGBTs depends on the design of the devices. The squares show the NFOM of the IGBTs described in Section II. The circles show high speed IGBTs, having flat carrier density distribution. Super-junction MOSFETs, shown by circles, show the ideal simulation results. SiC MOSFETs, shown by triangles, are plotted, using the product of the reported on-resistance and $\varepsilon E_c$. NFOM values of GaN FETs, shown by squares, are calculated by multiplying the reported on-resistance and two-dimensional electron gas density of $1 \times 10^{13} \text{cm}^{-2}$.

For less then 100 V, NFOM of conventional silicon MOSFET is even superior to any devices shown in the figure and have the potential of fastest switching speed among the devices, although gate loss is not considered. The reported SiC MOSFET is not far better than silicon MOSFETs, because the currently available on-resistance is not sufficiently low from the view point of NFOM.

**APPENDIX**

The stored energy in the main junction, $P_{\text{loss}}$, in the turn-off transient can be estimated assuming step junction approximation and inductive load, where the drain current, $I_D$, keeps in the same level within the turn-off transient. Using the voltage, $V$, as a function of depletion layer width, $d$, the final result is easily derived in the following:

$$V = \frac{nD_0 d^2}{2\varepsilon} \quad I_D \Delta t = \varepsilon N_D \Delta d, \quad V = \frac{I_D^2}{2qN_D} d^2, \quad V_D = \frac{I_D^2}{2qN_D} t^2,$$

where $t$ denotes the switching time.

$$Q_{\text{str}} = \frac{I_D^2}{(2qN_DdV_D)^2} \quad P_{\text{loss}} = \frac{1}{6} I_D V_D dt = \frac{I_D^2}{6qN_D} t^2 = \frac{1}{3} Q_{\text{str}} V_D$$

 Acknowledgement

The author would like to thank Vice-President K. Tani, Technology Executives M. Hidestima, K. Murakami and Dr. K. Morizuka for their support for this work, and Mr. T. Kawano, Mr. K. Nishitani, Mr. M. Yamaguchi, Mr. N. Yasuhara, Mr. Y. Kawaguchi, Mr. K. Nakamura and Mr. S. Ono for their contribution to this article.

**REFERENCES**