Basic and Advanced Theory on Power Semiconductor Devices

Akio Nakagawa

Nakagawa Consulting Office, LLC.

1

Chapter 1 Introduction & IGBT

1. Introduction
2. Evolution of IGBT
3. Silicon limit characteristics of IGBT
4. Basics of IGBT Design
5. SOA of IGBT
6. MOSFET-mode IGBT
7. Design Issue of FS IGBT
8. IGBT Cell Pattern Design

Chapter 2　Power MOSFET

1. Fundamentals of power MOSFETs
2. Recent topics in low voltage power MOSFETs
3. High Speed Power MOSFET
   ―Ideal switching in power MOSFET
4. Super Junction MOSFET

Chapter 3 pin diode

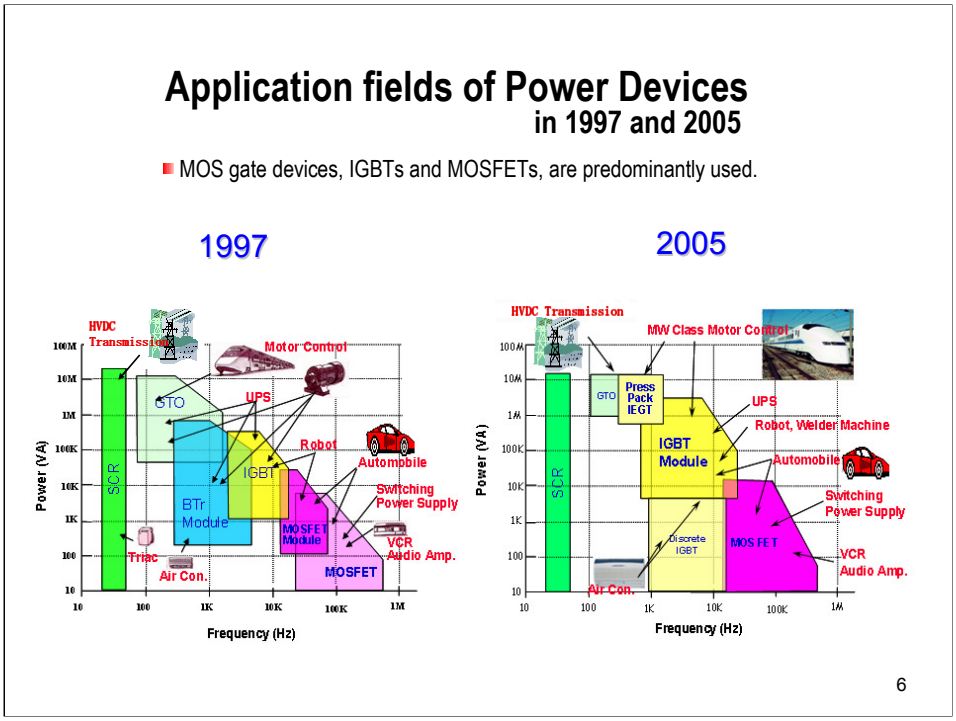
1. Basics of pin diode
2. Analytical and TCAD analysis of 200V high speed pin diode
3. Discussions on injection efficiency of bipolar transistor

Appendix

1. Basics of Semiconductor
2. Basic Device Equations
3. Ionization Integral
4. Bandgap Narrowing & Fermi Statistics

Outline

1. Introduction
2. Evolution of IGBT
3. Silicon limit characteristics of IGBT
4. Basics of IGBT Design
   Basic operation, Turn-off characteristics,
   Effect of JFET, FS-IGBT, Turn-on characteristics,
   Reverse recovery of Diodes
5. SOA of IGBT
6. MOSFET-mode IGBT
7. Design Issue of FS IGBT
8. IGBT Cell Pattern Design



This figure shows the application fields of power devices in 1997 and 2005.

There are many applications for power devices. These are high voltage DC transmission, Motor control, Automobile, Power Supply, Home appliances and so on.

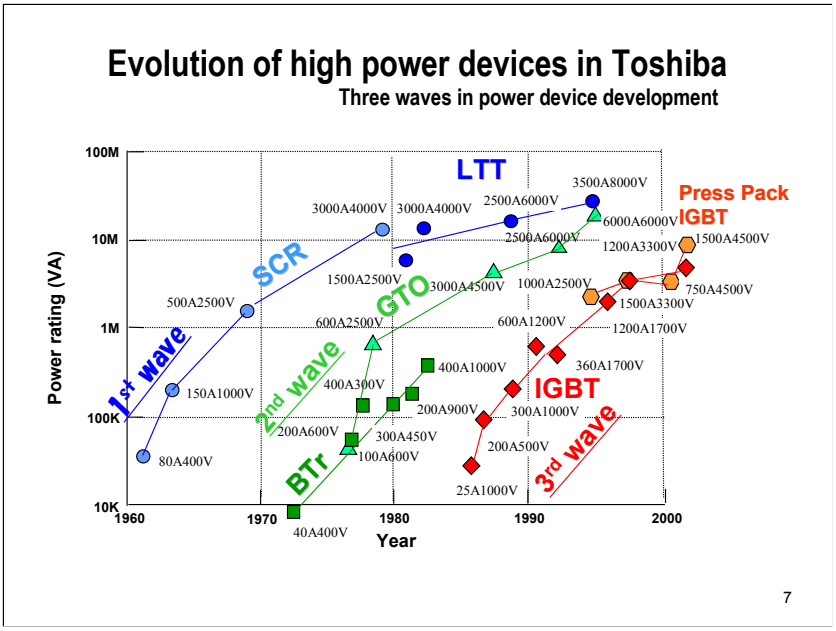
The clear difference between the two figures is that GTO and BTr were replaced by IGBTs and its module.

This big change occurred only in 8 years.

Now, MOS gate devices, IGBTs and MOSFETs, are predominantly used for many applications.

IGBTs are used for applications of large power and relatively lower frequency.

Power MOSFETs are used for high frequency applications.



This figure shows the evolution of high power devices.
There were three waves in the power device development.

The first wave is the development of thyristors.
The large power thyristor such as 3500A and 8000V was developed.

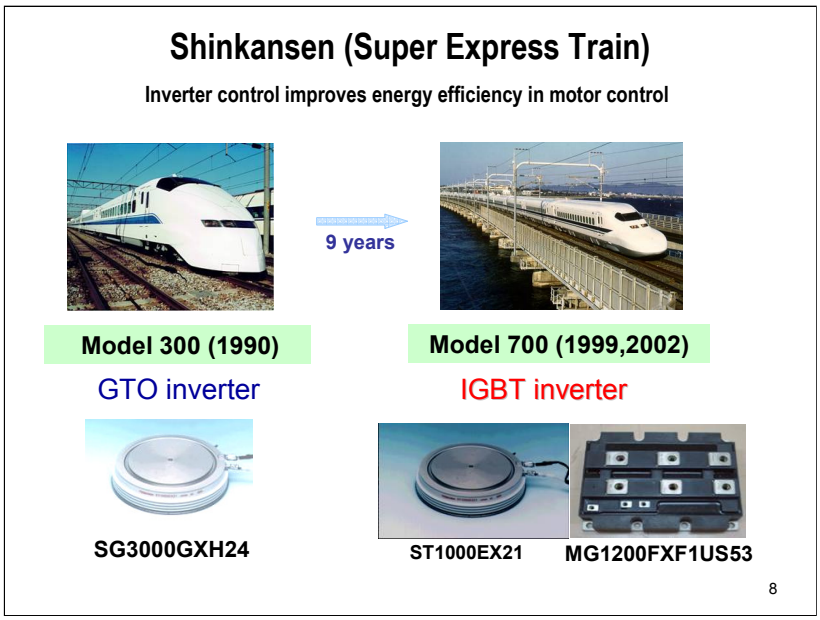
The second wave is the development of current control devices, such as bipolar transistors and GTOs.

The disadvantage of current control devices is the large gate current and the long switching time.

For example, in order to turn-off 3000A of GTO current, 1000A of gate current is required. We need another amplifier just for the gate current.

IGBT was developed in order to remove the disadvantages of current control devices.

Now, IGBTs are major devices in high power applications.



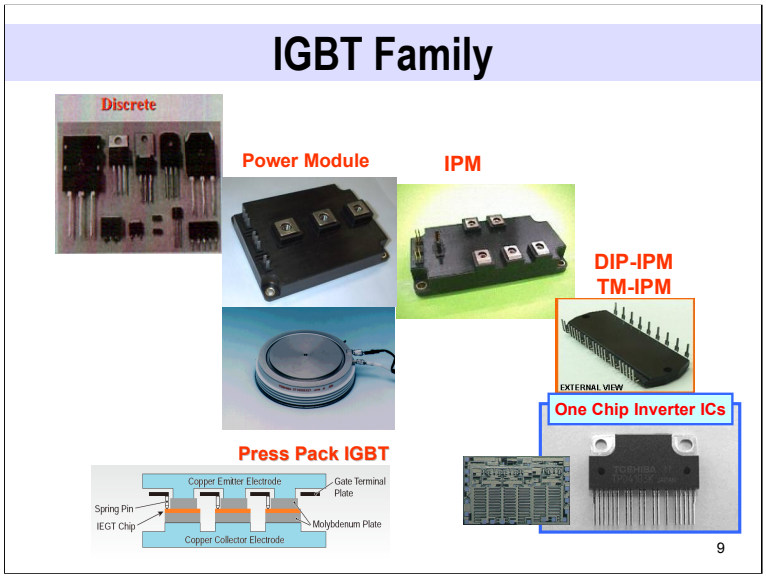
This shows Shinkansen, Super express trains.

Inverter control significantly improves energy efficiency of motor control.

The first inverter system was introduced to the Sinkansen in 1990, using GTOs.

Only 9 years later, Shinkansen was controlled by IGBT inverters.

This was really an epoch-making event in IGBT history.



This figure shows IGBT family.

Discrete IGBTs, Power Modules, Press pack IGBT, Intelligent power module, Transfer mold IPM, and one chip inverter IC.

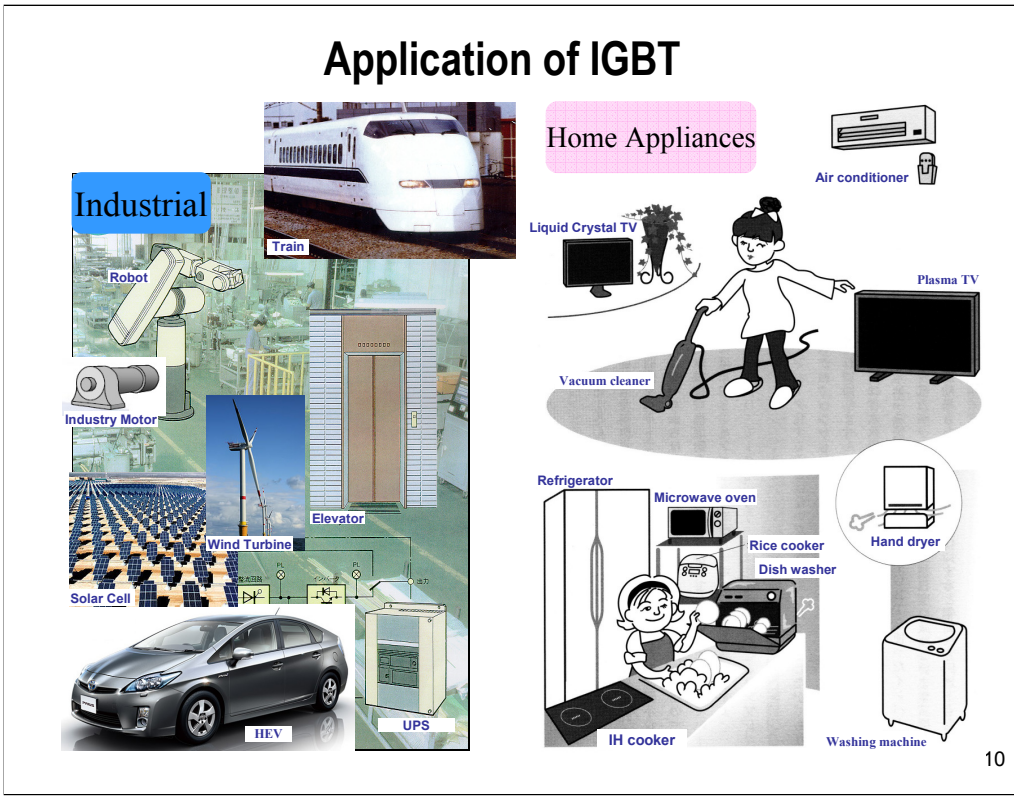
There are variety of IGBTs developed for various applications.

Power Modules contain multiple IGBT chips, operating in parallel and realize a large current capability.

Press pack IGBT also contains multiple chips operating in parallel. The two electrodes directly contact the both sides of the chip and conduct current and cool the chip from the both sides.

IPM integrates IGBTs and its driver and protection circuits in a same package.

One chip inverter integrates inverter circuits in a single chip.



This figure shows application of IGBTs.

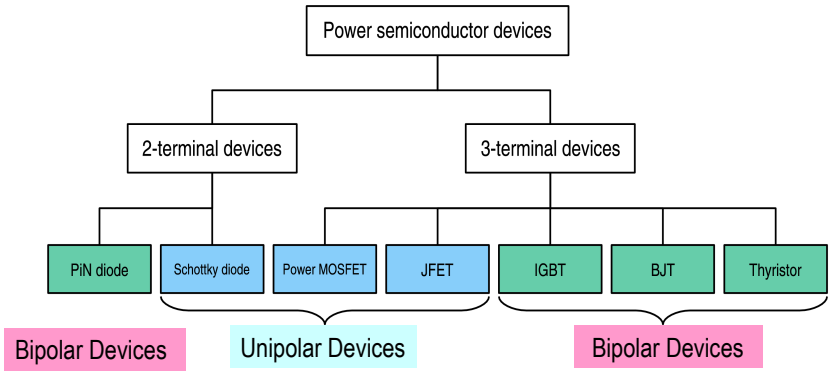
There are many applications.

For industry, Robots, Elevators, UPS, Bullet trains, and hybrid vehicles.

And, also, there are many applications for home electric appliances.

Air conditioners, Washing machines, IH cookers, Microwave ovens, Refrigerators, and so on.

Classification of Power Devices



Unipolar device: Only electrons conduct current. High speed switching.
Bipolar Devices: Electrons and holes conduct current. Large current, high voltage devices.

11

This figure classifies major power devices.

Power devices are divided into 2 terminal devices and 3 terminal devices.

As for 2 terminal devices, there are pin diodes and Schottky diodes.

3 terminal devices are switching devices.

There are power MOSFETs, Junction FETs, IGBTs, Bipolar Transistors, and thyristors.

These are further classified into two categories.

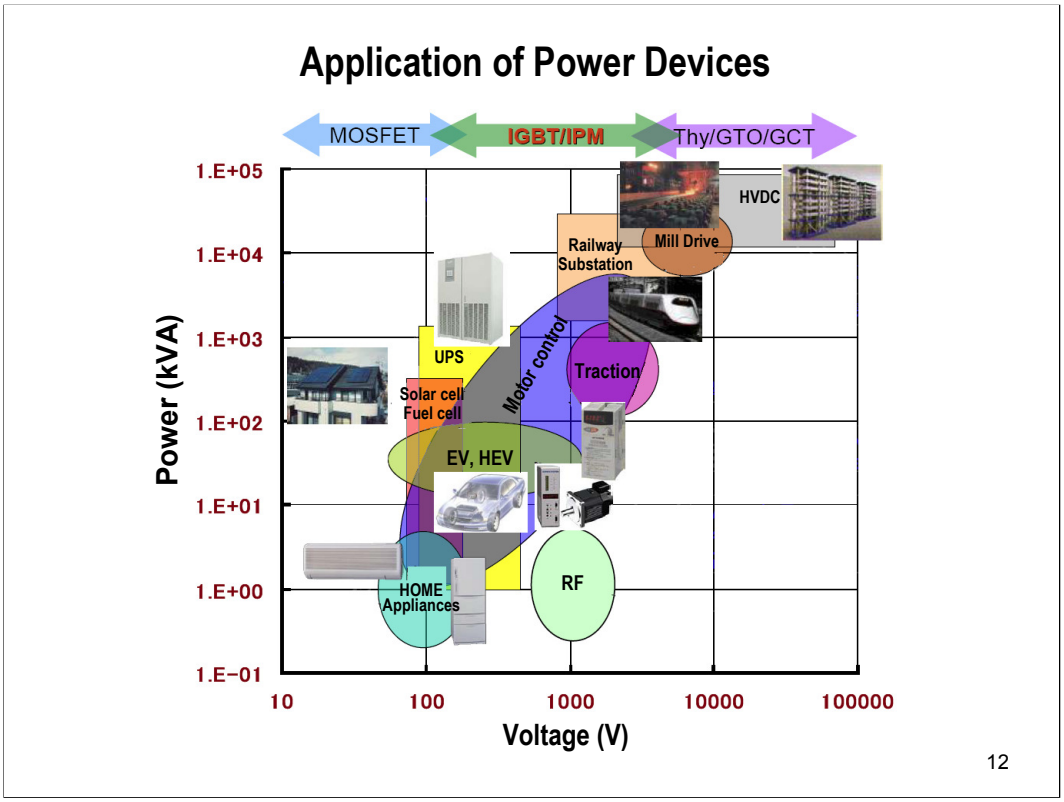
One is unipolar devices.

The other is bipolar devices.

Unipolar devices only use electrons to conduct current.

High speed switching is realized in unipolar devices.

Bipolar devices use electrons and holes to conduct current. Large current high voltage devices can be realized in bipolar devices.

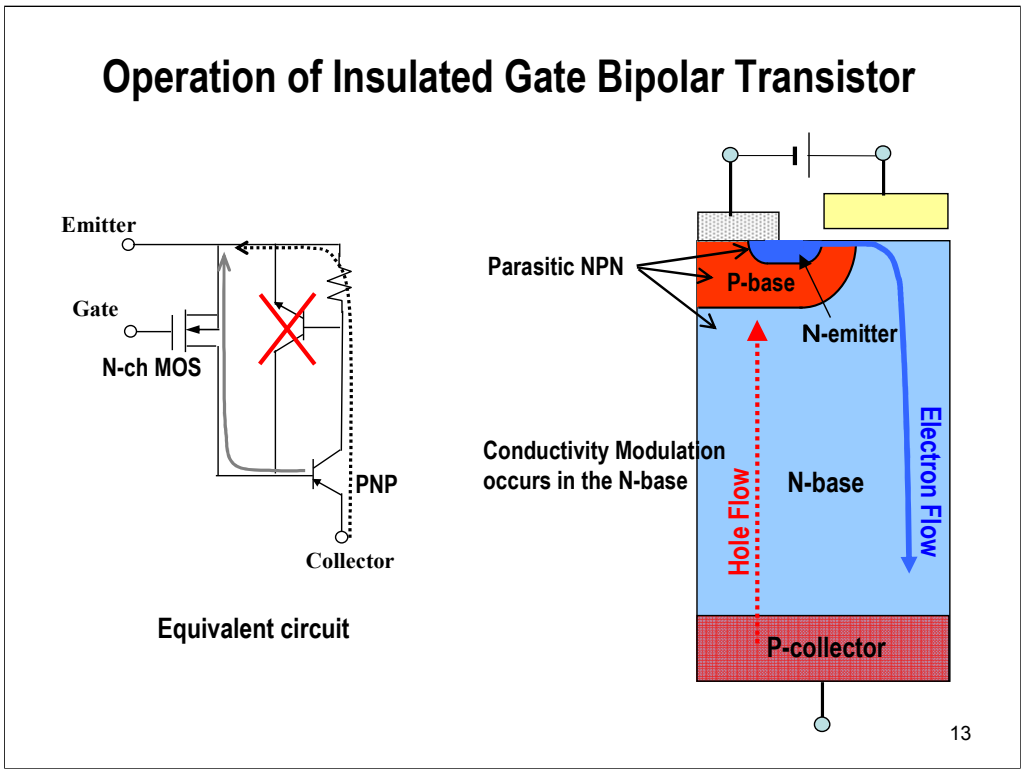


This figure shows application of power devices.

Power MOSFETs are used for low power high frequency application.

Most of medium and high power applications are covered by IGBTs.

For huge power applications, Thyristors, GTOs or GCTs are used.



I will show the basic operation of IGBTs.

The equivalent circuit is shown in the left figure.

The N-ch MOSFET supplies the base current of PNP transistor and drives the PNP transistor.

The current gain of the PNP Tr is small.

Thus, the major current is supplied by the N-ch MOSFET, the base current of PNP transistor.

This figure shows the cross section of IGBT.

If positive bias is applied to the gate, the surface of the P-base is inverted, and the channel is created in the surface.

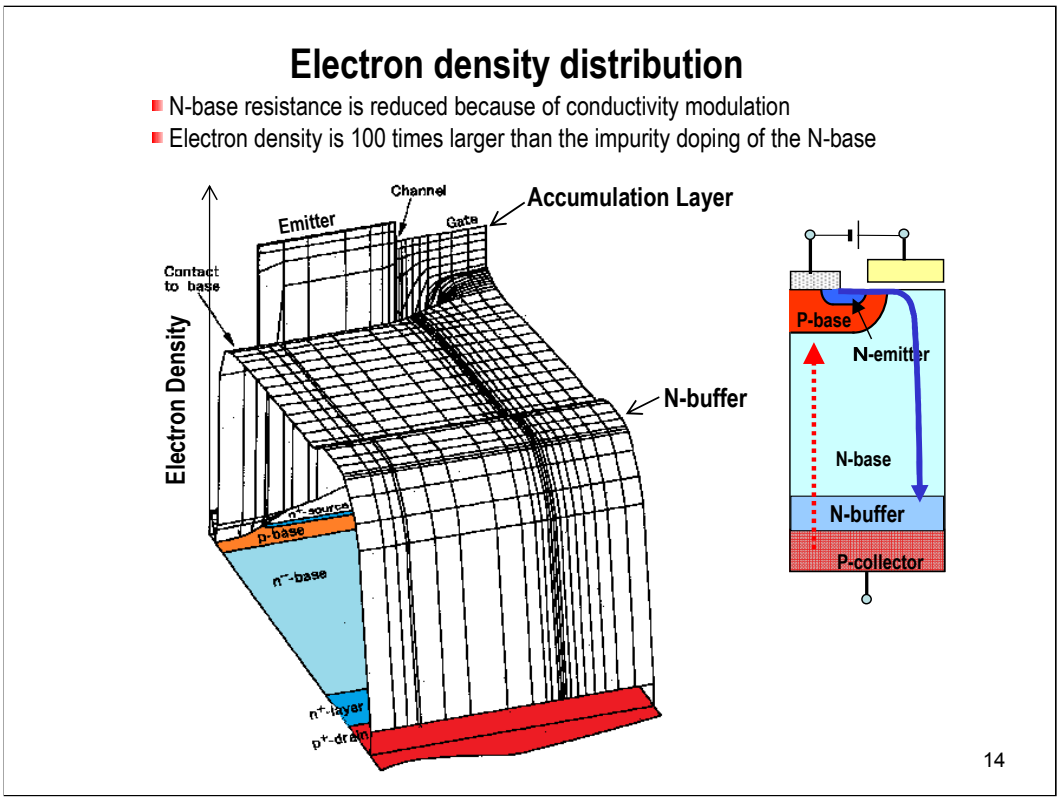
Electrons are injected into the N-base from the channel, and forward bias the collector PN junction.

Holes are injected from the collector and are collected by the P-base.

P-collector N-base and P-base constitute a PNP transistor.

There is a parasitic NPN transistor, consisting of N-emitter, P-base and N-base. The emitter and the P-base is short-circuited by the emitter electrode.

So, the operation of NPN transistor is prevented.



This figure shows electron density distribution.

Vertical axis shows electron density in logarithmic scale.

This is Emitter, Channel, Accumulation layer, N-buffer.

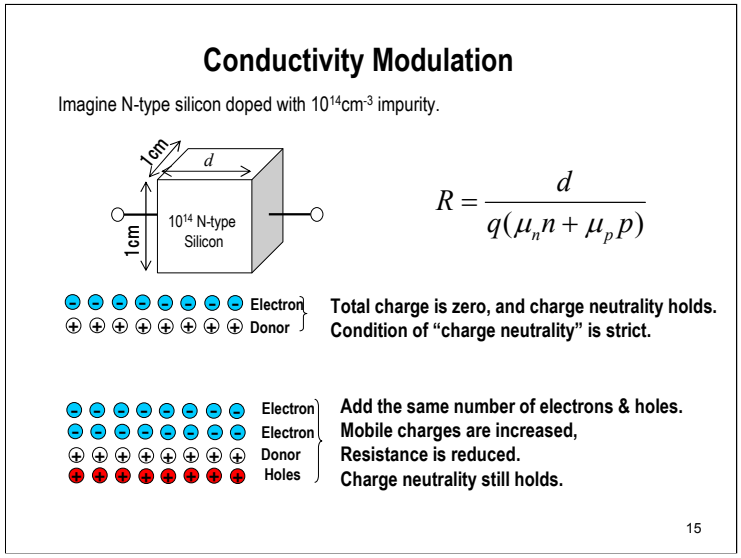
Electrons are injected from the channel, and holes are injected from the p-collector.

A large number of electrons and holes are accumulated in the N-base.

The electron density is the level of $10^{16}/cm^3$

This is 100times larger than the impurity doping of the N-base.

Thus, the N-base resistance is reduced $1/100^{th}$ of the original value.



Here, I interpret conductivity modulation.

Please imagine N-type silicon. The impurity concentration is 10^{14}.

The number of electrons per $1cm^3$ is 10^{14}, the number of donor is also 10^{14}.

Electrons have negative charges and donors have positive charges. The total charge is zero and the charge is neutral. This charge neutrality condition is very strict.

If the charge neutrality is lost, a large voltage appears between the terminals.

Now, we add the same number of electrons and holes.

For example, we introduce 10^{14} of electrons and the same number of holes. In this condition, the charge neutrality is still kept.

The number of mobile electrons become double and we have also holes.

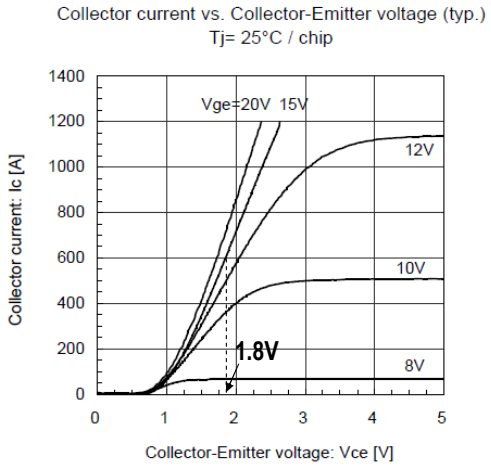
The resistance is inversely proportional to the number of electrons and holes. Thus, the resistance can be decreased with keeping the charge neutrality condition.

This is conductivity modulation.

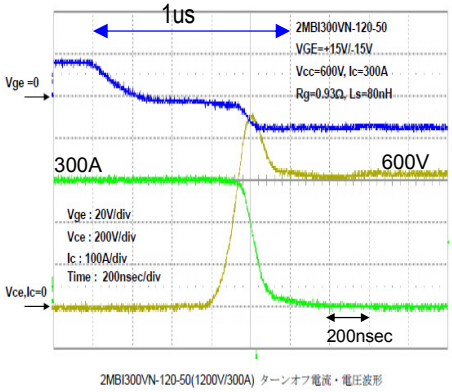
If electron number is increased 100 times with keeping charge neutrality, the resistance of the silicon can be decreased to one 100^{th} of the original value..

Typical electrical characteristics of IGBT

I-V relation of 1200V 600A IGBT



Turn-off waveforms of 600V 300A



2MB1600VN-120-50

16

This figure shows typical electrical characteristics of IGBT.

The rating is 1200V 600A device.

This shows current voltage characteristics.

The horizontal axis is voltage and vertical axis is current.

The forward voltage at 600A is 1.8V, when V_G=15V.

This figure shows typical turn-off waveforms.

This is collector current, collector voltage, and the gate voltage.

Time scale is 200nsec/Division.

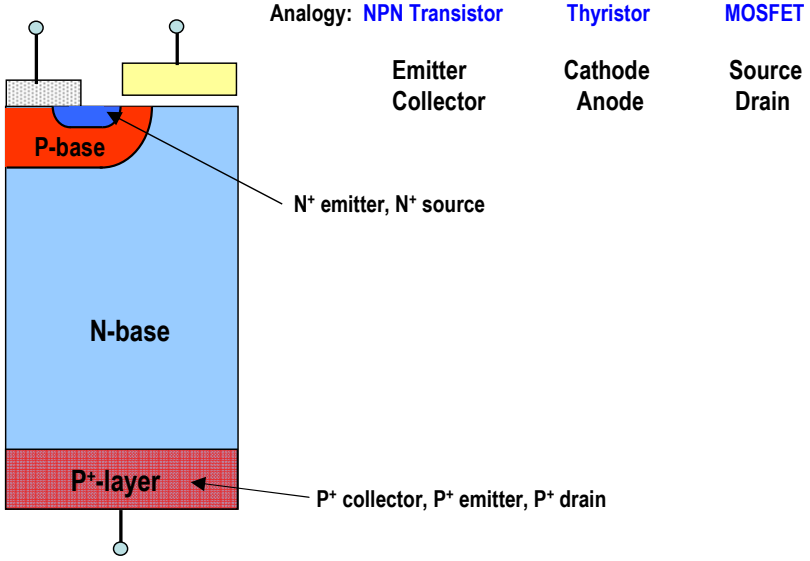
300A is turned off within 1us, and 600V is applied.

The fall time to turn-off 300A is less than 200nsec and is very fast.

2. Evolution of IGBTs

from non-latch-up IGBTs to FS-IGBTs

Ambiguity in terminology

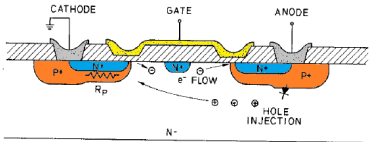


18

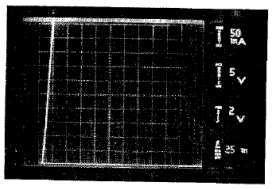
Don’t be confused.

First IGBT-mode operation found in lateral devices.

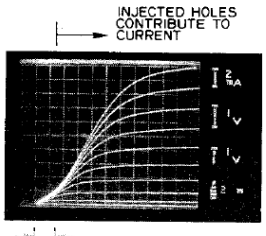
In 1978, B.W. Scharf and J.D. Plummer reported in ISSCC, for the first time, that four layer MOS gate thyristor operates as transistor in low current level.



Two DMOS Transistors as Bidirectional Switch



Thyristor operation in large current



Transistor operation in low current

A MERGED DEVICE based upon double-diffused MOS (DMOS) technology and combining the MOS and thyristor families has been developed, resulting in an insulated gate triac structure applicable to areas now served by current-controlled PNPN switches. The device – MOS-controlled Triac (TRIMOS) – may be integrated with other MOS components, for use in crosspoint switching, output stages and power control.

Although it is a single regenerative device, TRIMOS can be viewed, as can I^2L, as several merged conventional components, MOSFETs, BJTs, and resistors. Such a partitioning gives rise to a circuit which can be analyzed by a nonlinear circuit analysis program to provide physical insight into the three modes of TRIMOS operation.

Figures 1 and 2(a), cross section and photomicrograph of TRIMOS, show that the device is formed by merging two high-voltage DMOS transistors[1] around a common drain. Contact is made to the source and diffused channel of each DMOS, forming symmetrical anode and cathode contacts, and to the shared gate metal, forming the TRIMOS control electrode.

With the cathode grounded and the gate held below the positive DMOS threshold voltage, the P^+N^- junction at the cathode end blocks any applied positive anode voltage, holding the switch off up to its breakdown voltage; 200V at present.

For gate potentials above threshold, there are three distinct regions of operation. In the low-level realm, anode potentials of less than about 1.5V allow both DMOS channels to become inverted. Both transistors are in their linear regions and all the anode-to-cathode current is carried by electrons at the surface. The device exhibits the low on-resistance and I-V characteristics of two short channel (2.5μ) DMOS transistors in series.

The intermediate level of operation occurs for increasing anode bias which causes the P^+N^- anode junction to become forward biased as indicated in Figure 1, serving as the emitter

operation.

As the PNP collector current increases with anode or gate potential, its flow through the pinched resistor R_p raises the potential of the cathode's P region beneath the gate and begins to turn on the vertical NPN transistor inherent in the DMOS structure. This NPN, which can be identified in Figure 4, forms with the PNP a four-layer diode which regeneratively switches when ($\alpha_{NPN} + \alpha_{PNP}$) equals unity. In its on state, TRIMOS exhibits a dynamic resistance of less than 10Ω and can pass currents on the order of amperes; Figure 2(a).

Control of the switching point by the $I_{CPNP}R_p$ product has been demonstrated by fabrication of devices with anode switching currents varying from tens of microamps to hundreds of milliamps by varying the geometrical layout. By shunting R_p with a switch (Figure 4), the TRIMOS may be switched out of its on state or inhibited from triggering. This type of shunt switch has been realized by an MOS transistor fabricated adjacent to the TRIMOS. Without such a bypass structure, a TRIMOS typically has turn-on and turn-off times on the order of 200ns and its single pulse dv/dt capability exceeds 1000V/μs.

As the discussion of operation has indicated and as Figure 4 illustrates, several bipolar and MOS transistors and resistors can be identified within the TRIMOS structure. Analysis of the circuit formed by these components has resulted in a model for operation below regeneration. A circuit analysis program containing sophisticated models of both bipolar and MOS devices must be used to obtain accurate simulation of device characteristics. This has been done on the Mini-MSINC[2] program using a DMOS model developed earlier[3] and an integral charge-control bipolar model[4,5]. The comparison of measured and modelled characteristics in Figure 3 delineates the two regions of operation and shows quantitative agreement over a rather large operating region. The apparent increase in g_{fs} for $V_{DS} > 1.5V$ as the lateral PNP turns on is visible in both the experimental and simulated curves.

There are several advantages to the discretized model. First, it corresponds to the physical structure and aids in an intuitive understanding of the different modes of TRIMOS operation since the individual building blocks are familiar devices. Secondly, the parameters required for the model are those routinely measured on the BJT and MOSFET. Many of these can be measured directly from the TRIMOS component of interest and the others may be inferred from ordinary test structures or process characteristics. Lastly, process variations affecting the component para-

rather than one dimensional discrete devices.

*Project performed under NIH Grant No. 1 P01 GM17940-5 and NSF Grant No. ENG74-18419.

[1]Plummer, J.D. and Meindl, J.D., "A Monolithic 200-V CMOS Analog Switch", *IEEE J. Solid State Circuits*, Vol. SC-11, p. 809-817; Dec., 1976.

[2]Young, T.K. and Dutton, R.W., "Mini-MSINC — A Minicomputer Simulator for MOS Circuits with Modular Built-In Model", *IEEE J. Solid State Circuits*, Vol. SC-11, p. 730-732; Oct., 1976.

1976.

222 • 1978 IEEE International Solid-State Circuits Conference

MOS-Thyristor can be operated as transistor, if thyristor action is not triggered

First IGBT operation was found in lateral devices.

In 1978, B.W. Scharf and J.D. Plummer reported in ISSCC that a four layer MOS gate thyristor operates as a transistor in low current level.

They used two DMOS transistors. One is used as cathode, the other is used as anode.

The upper figure shows the current voltage relation in large current region.

The device operate as a thyristor.

The lower figure shows the current voltage relation in low current level.

This shows that the device can operate as a transistor if thyristor is not triggered.

They showed that MOS-Thyristor can be operated as a transistor, if thyristor action is not triggered.

Vertical IGBT was first reported by Baliga et al. in 1982.

- He showed the device can operate as a four layer MOS-gate transistor
- Switching characteristics were very poor.

THE INSULATED GATE RECTIFIER (IGR):

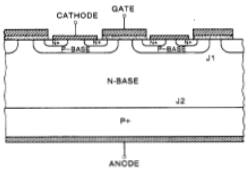
A NEW POWER SWITCHING DEVICE

B.J. Baliga, M.S. Adler, P.V. Gray, R.P. Love

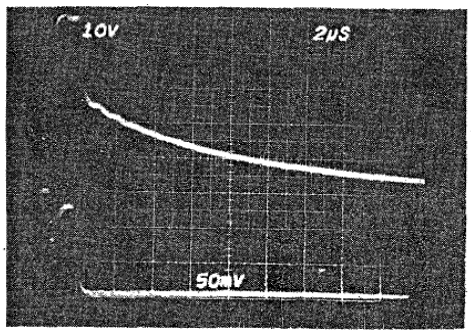
General Electric Company
Corporate Research and Development Center
Schenectady, NY

Nathan Zommer

Intersil Inc.
Cupertino, CA



Switching time >18μsec



20

Vertical IGBT was first reported by Baliga in 1982.

He showed that the device can operate successfully as a four layer MOS-gate transistor.

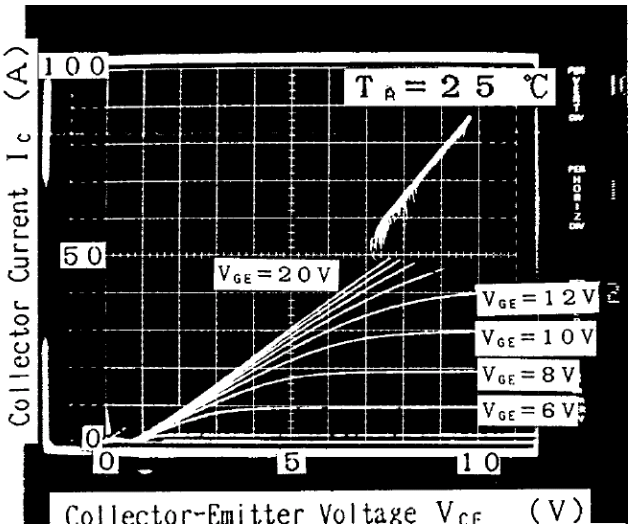
However, the switching characteristics of the reported device was very poor.

The switching speed was very slow, the fall time is almost 18usec.

The device latched up, if current density exceeded a few 100A/cm^2.

Measured Current Voltage Characteristics of early stage IGBTs

- If I_C exceeded 50A, latch-up occurred, and the device lost the control of the gate.
- Early stage IGBTs were easily destroyed because of latch-up.
- It was thought nearly impossible to prevent latch-up in 1983.



I-V characteristics of IGBTs in the early stage

21

This shows the measured current-voltage characteristics of early stage IGBTs.

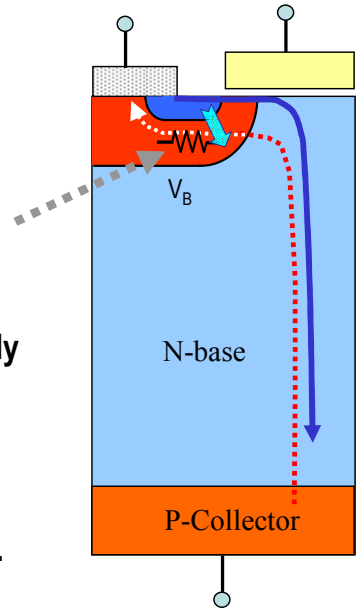
If the collector current exceeded 50A, latch-up occurred, and the device lost the control of the gate.

Early IGBTs were easily destroyed because of latch-up.

It was thought almost impossible to prevent latch-up in 1983.

What is Latch-Up

Hole current flows in the P-base. When voltage drop in the p-base exceeds the built-in voltage of PN junction, electrons are directly injected into P-base, resulting in thyristor action.
Once this occurs, the current cannot be controlled by the gate.

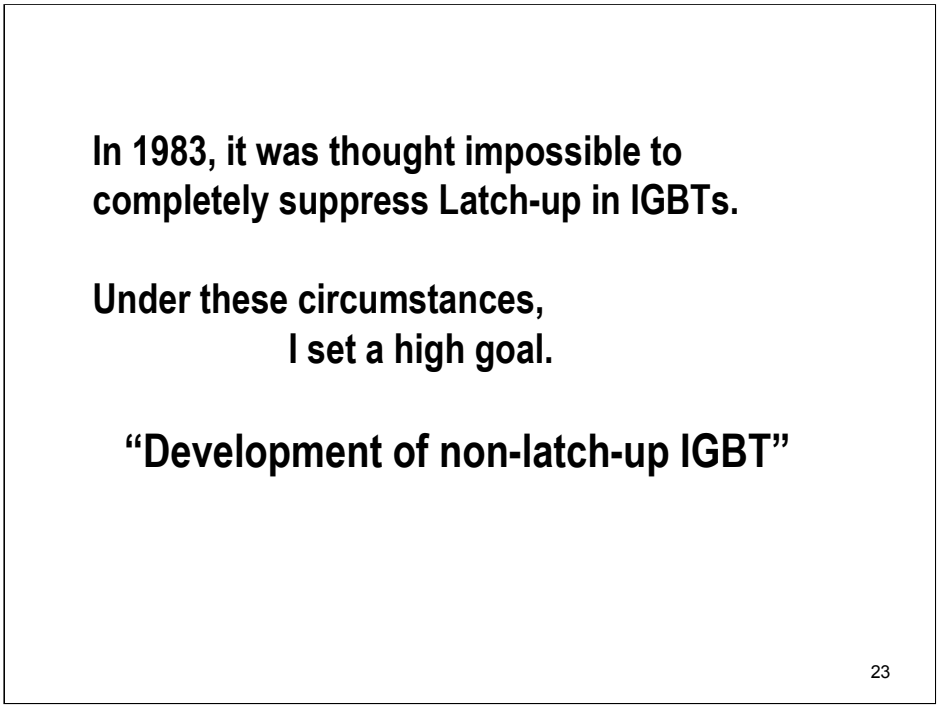


22

What is latch-up?

Hole current, coming from the collector, flows in the p-base under the N+ emitter.
If the voltage drop in the p-base exceeds the built-in voltage of the PN junction, the NPN transistor is activated and electrons are injected directly into the P-base.

The current is amplified both by PNP and NPN transistor action, and the current flows in thyristor action.
Once the thyristor action occurs, the current cannot be controlled by the gate.

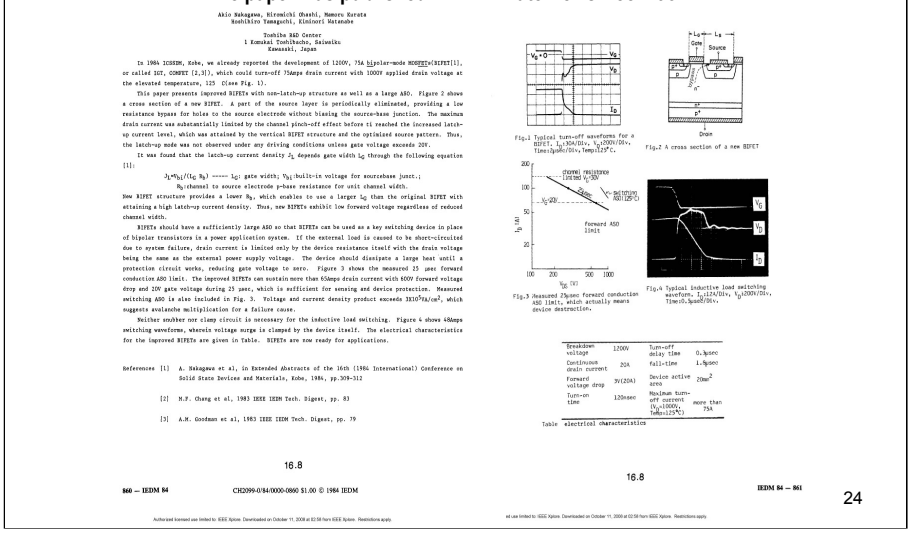


In 1983, it was thought nearly impossible to completely suppress latch-up in IGBTs.

Under these circumstances, in 1983, I set a very high goal, that was development of non-latch-up IGBT.

Non-Latch-Up IGBTs were successfully developed in 1984
The developed devices never latched-up when V_G < 20V.

- The paper was published in IEDM Late News Dec. 1984



Akio Nakagawa, Hiromichi Ohashi, Mamoru Kurata
Yoshihiro Yamaguchi, Kinlouri Watanabe

Toshiba R&D Center
1 Komukai Toshibacho, Saiwaiku
Kawasaki, Japan

In 1984 ICSSEM, Kobe, we already reported the development of 1200V, 75A bipolar-mode MOSFET(=BIFET)[1], or called IGT, COMFET [2,3]), which could turn-off 75Ampe drain current with 1000V applied drain voltage at the elevated temperature, 125 (See Fig. 1).

This paper presents improved BIFETs with non-latch-up structure as well as a large ASO. Figure 2 shows a cross section of a new BIFET. A part of the source layer is periodically eliminated, providing a low resistance bypass for holes to the source electrode without biasing the source-base junction. The maximum drain current was substantially limited by the channel pinch-off effect before ti reached the increased latch-up current level, which was attained by the vertical BIFET structure and the optimized source pattern. Thus, the latch-up mode was not observed under any driving conditions unless gate voltage exceeds 20V.

It was found that the latch-up current density J_L depends gate width L_G through the following equation [1];

$J_L \propto V_{bi}/(L_G R_B)$ ----- L_G: gate width; V_{bi}: built-in voltage for sourcebase junct.;
R_B: channel to source electrode p-base resistance for unit channel width.

New BIFET structure provides a lower R_B, which enables to use a larger L_G than the original BIFET with attaining a high latch-up current density. Thus, new BIFETs exhibit low forward voltage regardless of reduced channel width.

BIFETs should have a sufficiently large ASO so that BIFETs can be used as a key switching device in place of bipolar transistors in a power application system. If the external load is caused to be short-circuited due to system failure, drain current is limited only by the device resistance itself with the drain voltage being the same as the external power supply voltage. The device should dissipate a large heat until a protection circuit works, reducing gate voltage to zero. Figure 3 shows the measured 25 μsec forward conduction ASO limit. The improved BIFETs can sustain more than 65Ampe drain current with 600V forward voltage drop and 20V gate voltage during 25 μsec, which is sufficient for sensing and device protection. Measured switching ASO is also included in Fig. 3. Voltage and current density product exceeds $3X10^6 A/cm^2$, which suggests avalanche multiplication for a failure cause.

Neither snubber nor clamp circuit is necessary for the inductive load switching. Figure 4 shows 65Ampe switching waveforms, wherein voltage surge is clamped by the device itself. The electrical characteristics for the improved BIFETs are given in Table. BIFETs are now ready for applications.

References [1] A. Nakagawa et al, in Extended Abstracts of the 16th (1984 International) Conference on Solid State Devices and Materials, Kobe, 1984, pp.309-312

[2] M.F. Chang et al, 1983 IEEE IEDM Tech. Digest, pp. 83

[3] A.M. Goodman et al, 1983 IEEE IEDM Tech. Digest, pp. 79

16.8

860 — IEDM 84

CH2099-0/84/0000-0860 $1.00 © 1984 IEEE

Fig.1 Typical turn-off waveforms for a BIFET. I_D:30A/Div, V_D:200V/Div, Time:1μsec/Div, Temp:125°C.

Fig.2 A cross section of a new BIFET

Fig.3 Measured 25μsec forward conduction ASO limit, which actually means device destruction.

Fig.4 Typical inductive load switching waveform. I_D:12A/Div, V_D:200V/Div, Time:0.5μsec/Div.

Breakdown voltage	1200V	Turn-off delay time	0.4μsec
Continuous drain current	20A	Fall-time	1.4μsec
Forward voltage drop	3V(20A)	Device active area	$20mm^2$
Turn-on time	120nsec	Maximum turn-off current (V_D=1000V, Tc=125°C)	more than 75A

Table electrical characteristics

16.8

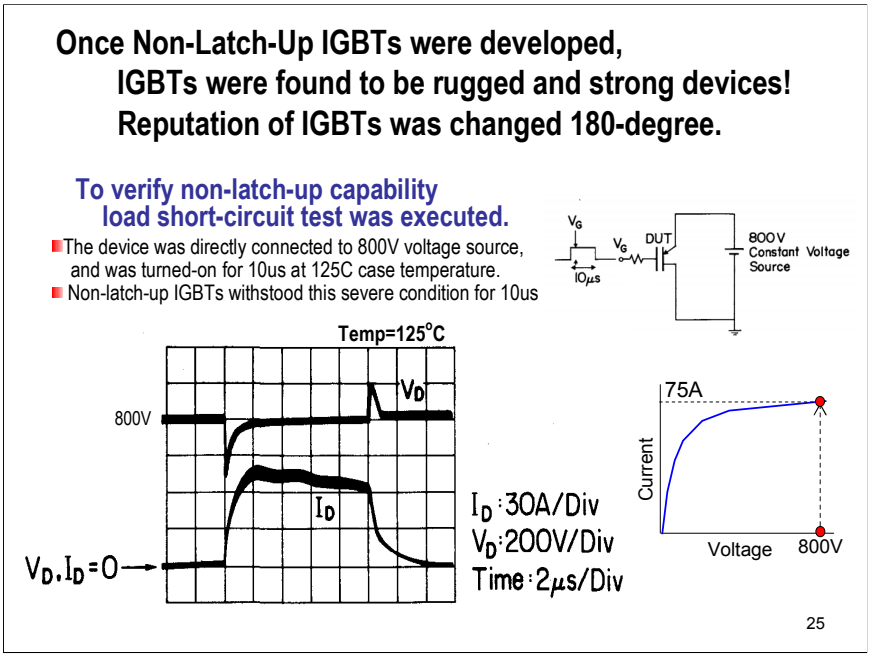
IEDM 84 — 861

24

In 1984, I successfully developed non-latch-up IGBTs and completely suppressed the latch-up of the parasitic thyristor.

My paper of non-latch-up IGBT was published as a late News paper in International Electron Device Meeting in 1984.

The developed device never latched-up under any driving conditions if the gate voltage was below 20V.



Once Non-Latch-Up IGBTs were developed,

IGBTs were found to be very rugged and strong devices!

Reputation of IGBTs was changed 180-degree.

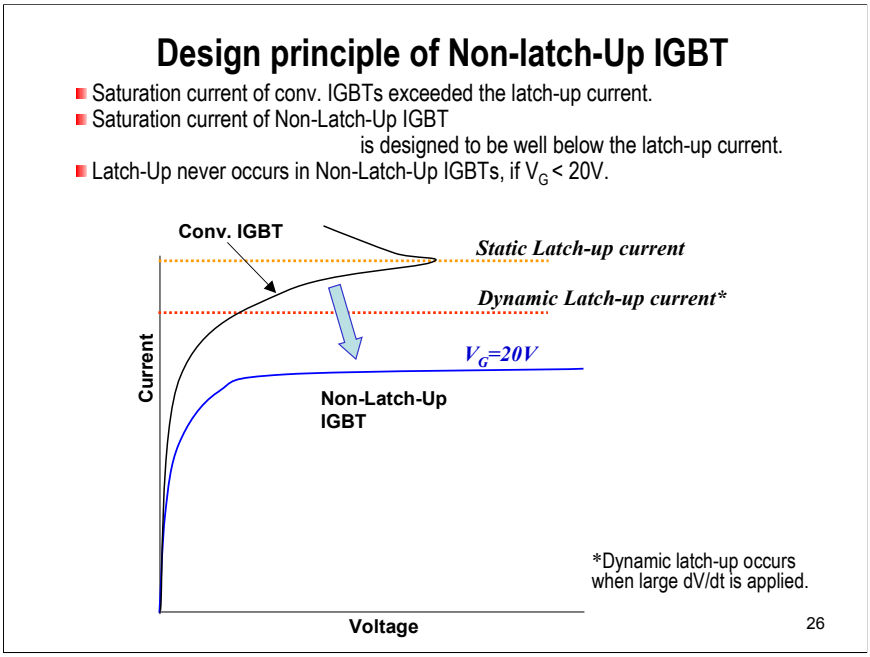
In order to test the non-latch-up capability, I executed world first load short circuit test.

The device was directly connected to a 800V voltage source, and the device was turned-on for 10us at 125C case temperature.

The device operating point changes from this off-state to this on-state.

75A current flowed in the device, and the forward voltage became 800V, the same voltage as the voltage source.

Non-latch-up IGBTs could withstand this severe condition for 10us.



This shows the design principle of non-latch-up IGBT.

This is typical current voltage curve of conv. IGBT and this is non-latch-up IGBT.

The saturation current of conv. IGBTs exceeded the latch-up current.

The saturation current of non-latch-up IGBTs is designed to be smaller than the dynamic latch-up current.

Latch-Up, theoretically, never occurs.

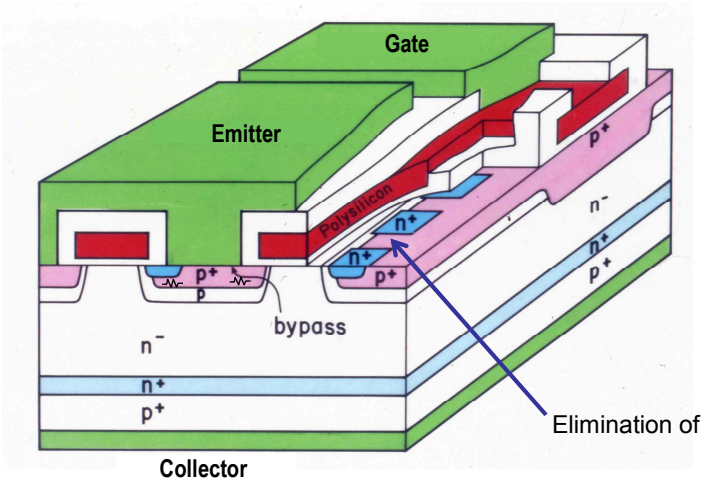
This is very simple and natural design principle.

However, in early development stage, this was never considered in actual IGBT design.

This design principle became de facto standard, and is widely used now.

Non-Latch-Up IGBT structure

- Part of n^+ layer is periodically eliminated to reduce the saturation current
- p^+ diffusion is extended into the channel for low resistance hole bypass in the area where n^+ is eliminated.



27

This figure shows the structure of non-latch-up IGBTs.

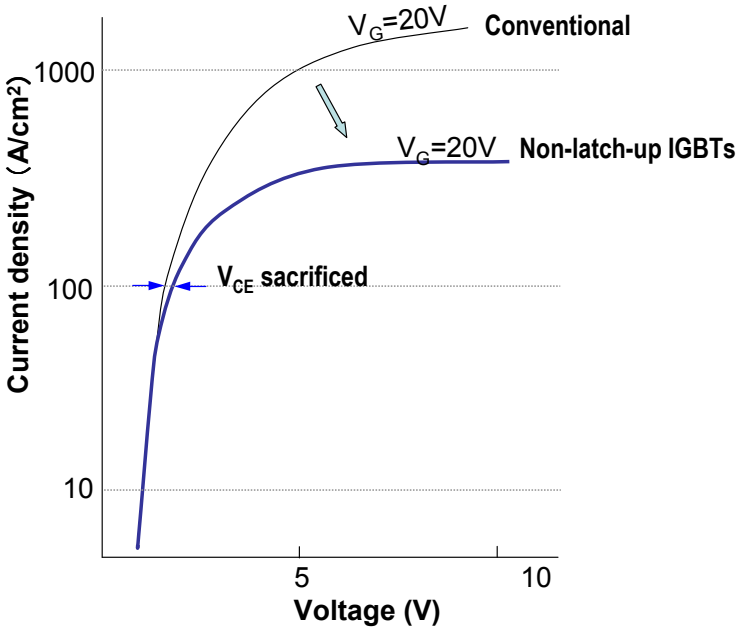
This is …..

In order to reduce the saturation current, portions of N+ source layers were periodically eliminated to reduce the width of the channel.

P+ diffusion is extended into the channel region to create a low resistance hole bypass in the area where N^+ is eliminated

Comparison of I-V curves between conventional IGBTs and non-latch-up IGBTs

- I_{SAT} for 20V gate voltage was decreased to around 400A/cm^2
- V_{CE} at rated current was somewhat sacrificed.
- It was important to demonstrate Non-latch-Up IGBTs.



28

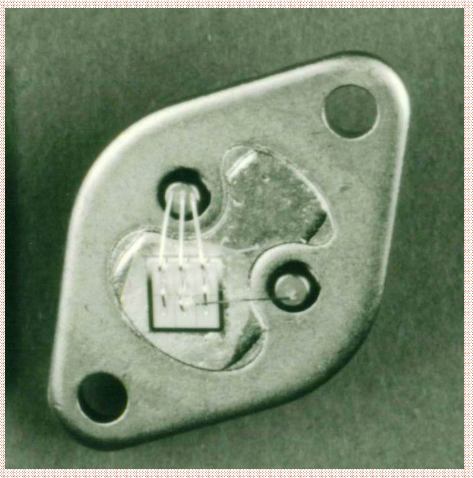
This figure compares I-V curves of conventional IGBTs and non-latch-up IGBTs.

The saturation current for 20V gate voltage was decreased to around 400A/cm^2.

This somewhat sacrificed device VCE at the rated current: 100A/cm^2.

However, it was more important to demonstrate non-latch-up characteristics.

First Non-Latch-Up IGBT in 1984.



2010 IEEE William E. Newell Power Electronics Award

For development of non-latch-up IGBTs

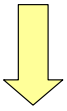
29

This photo shows first non-latch-up IGBT, fabricated and mounted on TO-3 package in 1984.

I received IEEE William E. Newell Power Electronics Award in 2010 for development non-latch-up IGBTs.

2nd high goal: In 1989

We were confident that
BTr would surely be replaced by IGBTs!
Next step was displacement of GTO
by new MOS gate device.



V_{CE} of High Voltage IGBT was high.
Breakthrough technology was necessary
to realize high voltage MOS Gate devices.

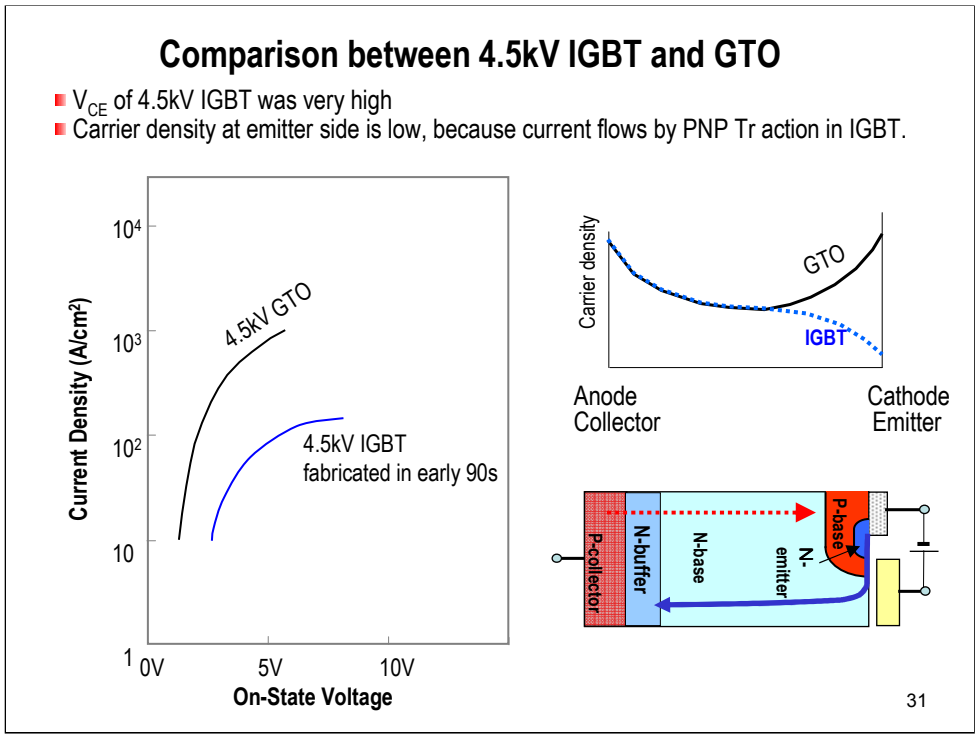
30

In 1989, we were confident that BTrs would surely be replaced by IGBTs.

At that time, our another dream was to replace GTO with a new MOS gate device.

It was very difficult to develop high voltage IGBTs, because IGBT V_{CE} rapidly increased as the breakdown voltage increased.

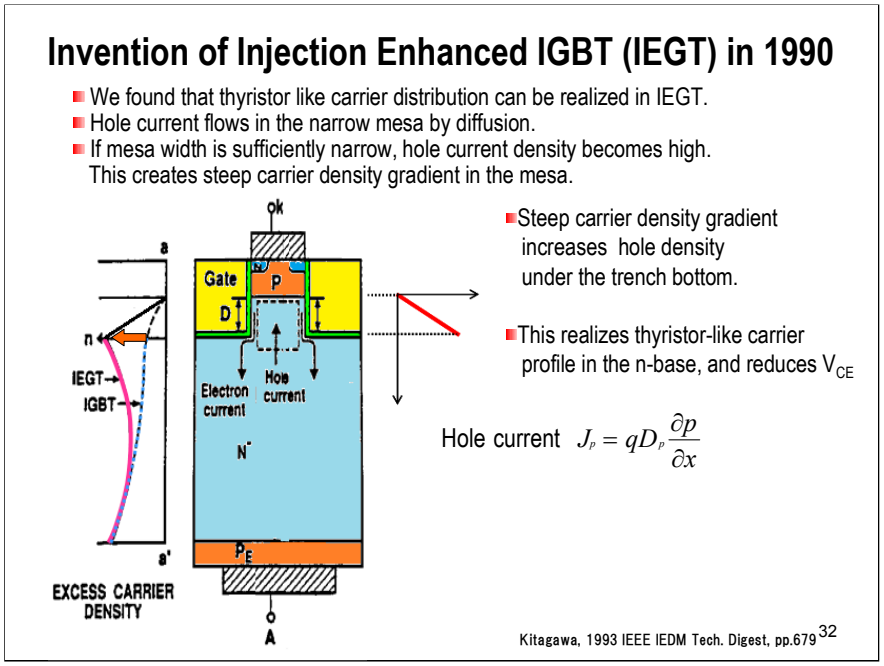
A breakthrough technology was necessary to realize high voltage IGBTs.



This figure compares current-voltage relations between 4.5kV IGBT and GTO.

VCE of IGBT is higher than GTO

The reason is that carrier density at the emitter side is low, because current flows by PNP transistor action in IGBT.



We invented injection enhanced IGBT or IEGT in 1990.

In trench gate IGBTs, hole current flows in this narrow mesa region by diffusion.

Electron current flows in the accumulation channel, which is created on the side wall.

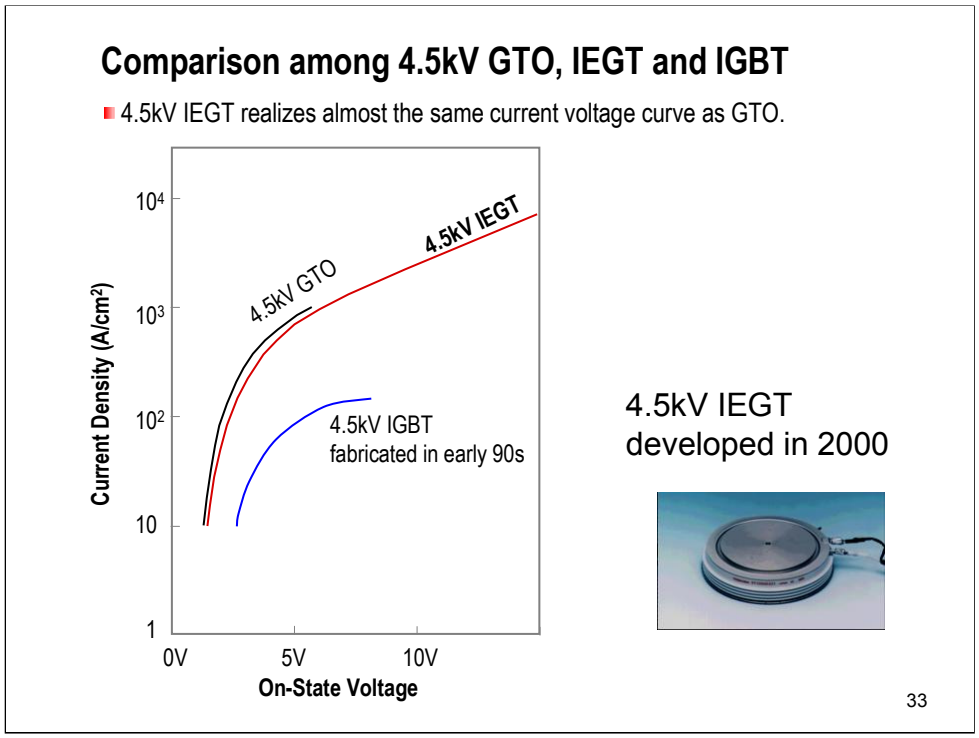
If the mesa width is sufficiently narrow, hole current density becomes high in the Mesa region.

This creates steep carrier density gradient in the mesa region.

This steep carrier density gradient increases carrier density under the trench bottom.

The carrier density in the n-base becomes high on the both sides, on the anode side and on the cathode side.

This carrier distribution is similar to that of thyristor, and low VCE can be realized in high voltage IGBTs.



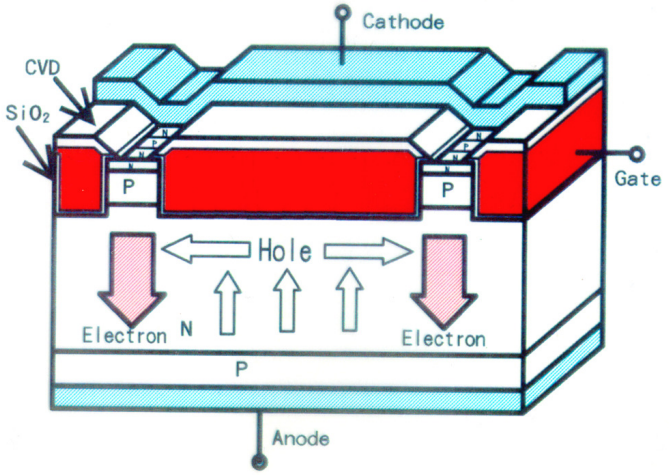
This figure compares current-voltage relations among 4.5kV IGBT, IEGT and GTO.

4.5kV IEGT realizes almost the same current voltage relation as GTO.

In 2000, 4.5kV IEGT was successfully developed.

Proposed IEGT

- Wide and deep trench gates are necessary for high voltage IEGTs



34

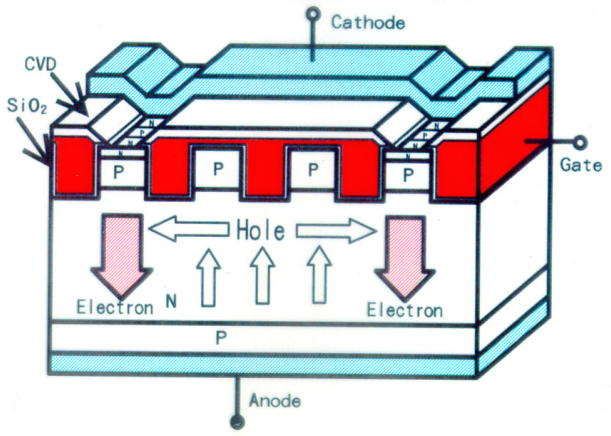
This figure shows the IEGT structure.

We need wide and deep trench gates and narrow mesa for high voltage IEGTs.

However, it is very difficult to actually fabricate wide and deep trench gates.

Practical Structure

- This structure can be realized, using the process of conventional trench IGBTs.
- Some of P-bases are not connected to the emitter and are kept electrically floating.



35

This figure shows practical IEGT structure.

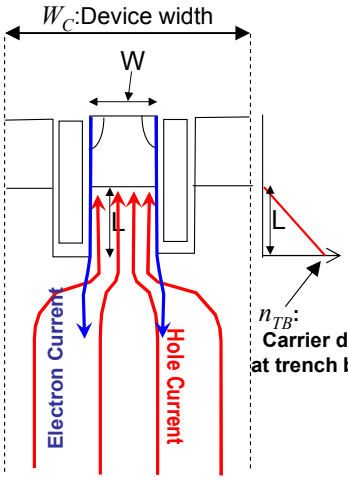
The structure can be fabricated, using the fabrication process of conventional trench IGBTs.

The difference is that this and this p-base are not connected to the emitter electrode, and are kept electrically floating.

This structure shows the same electrical characteristics as the original structure.

Analysis of Injection Enhancement Effect

- derive the expression of trench bottom carrier density, n_{TB}
  - Electron current flows through accumulation layer on the side wall.
  - Hole current flows in the mesa region. Electron current is zero in the mesa region.



J_{hole} : hole current density in N-base

$$J_{h,Mesa} = \frac{W_C}{W} J_{hole} = \frac{W_C}{W} \gamma J_C \quad \text{...Eq.(1)}$$

γ : Ratio of Hole current / Total current

J_C : Collector current density

$$J_{e,Mesa} = 0 = -qD_e \frac{n_{TB}}{L} + q\mu_e nE = \mu_e(-kT\frac{n_{TB}}{L} + qnE) \text{...Eq.(2)}$$

$$J_{h,Mesa} = qD_h \frac{n_{TB}}{L} + q\mu_h nE = \mu_h(kT\frac{n_{TB}}{L} + qnE) \quad \text{...Eq.(3)}$$

Using Eq.(2), Eq.(3) reduces to

$$J_{h,Mesa} = 2qD_h \frac{n_{TB}}{L} \quad \text{...Eq.(4)}$$

Because Eq.(1) = Eq.(4), we have:

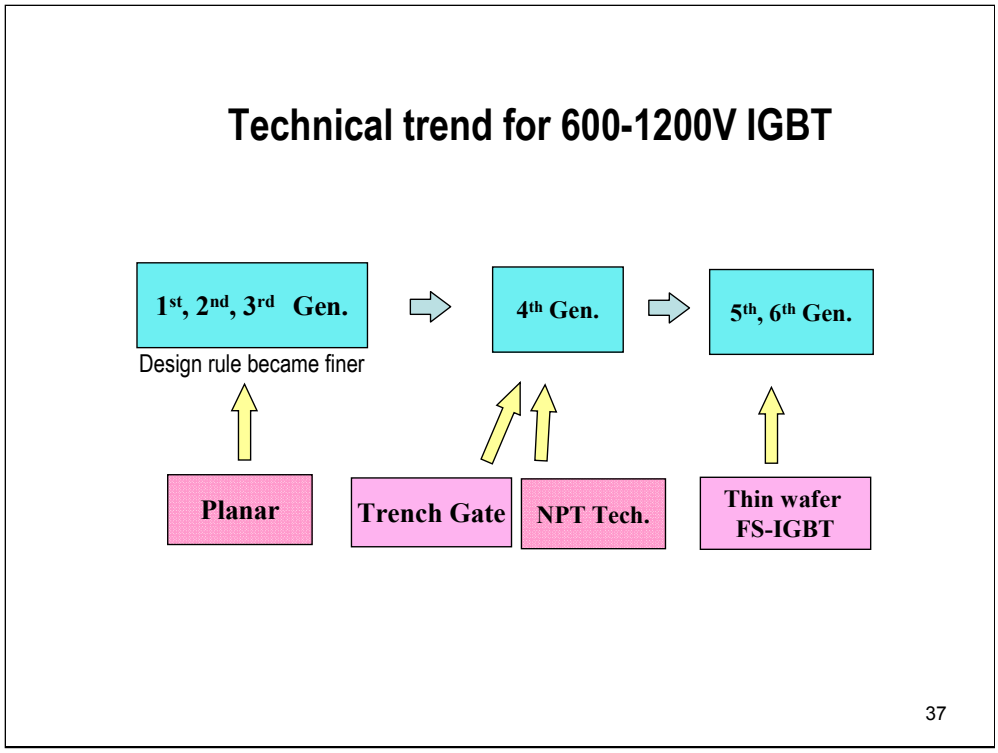
$$n_{TB} = \frac{W_C L}{W} \frac{\gamma J_C}{2qD_h} \propto \frac{W_C}{W} L \quad \text{: Design parameter}$$

- **IE effects can be increased by increasing the value of LW_C/W**

36

Now, we analyze the injection enhancement effect.

IE efffect depends on the parameter $W_C L/W$



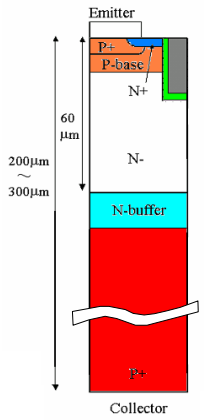
This figure shows technical trends for 600V and 1200V IGBTs.

1st 2nd and 3rd generation adopted planar technology and the design rule became finer.

NPT technology and trench gate was introduced in the 4th generation.

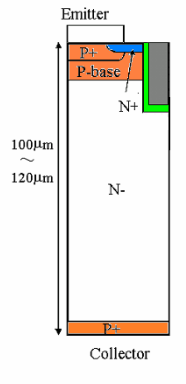
Recently, thin wafer technology was introduced for the new generation IGBTs.

PT-IGBT and NPT-IGBT



600V PT-IGBT

Fabricated in high resist. Epi on P^+ Sub.
Lifetime control is necessary because of high γ of P^+ emitter.
N-buffer stops depletion layer



600V NPT-IGBT

Low dose p-emitter realizes high speed
No lifetime control
Thick N-base is required
Widely used in the late 90's

38

Now, I introduce two kinds of IGBT structures.

First one is 600V punch-through IGBT.

IGBT is formed in the high resistivity epitaxial layer on the thick P+ substrate.

Because injection efficiency of the P+ substrate is very high, lifetime control is necessary to obtain high switching speed.

N-buffer stops the depletion layer. Thus, the thickness of the N-base can be reduced. This is good to achieve lower on-state voltage.

NPT IGBT uses low dose p-type layer as anode p-emitter.

Lifetime control is not necessary because low dose p-emitter automatically realizes high switching speed.

NPT-IGBT uses relatively thick N-base because there is no N-buffer layer.

NPT IGBTs was widely used in the late 90's.

In 1998, 3rd high goal: "Realize 1.0V on-state voltage in IGBT"

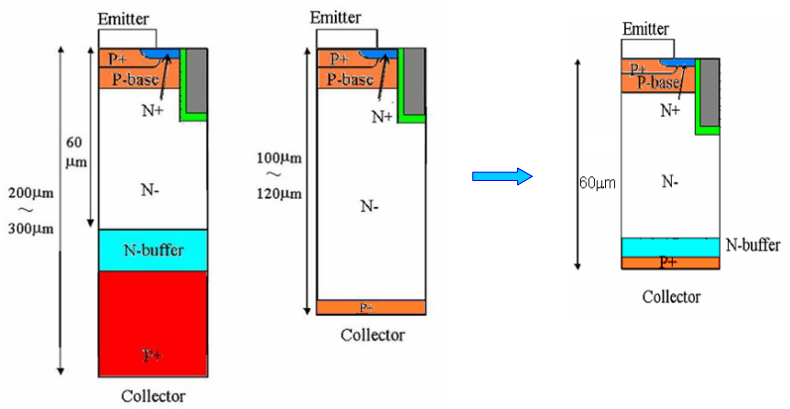
- NPT-IGBT was widely used in 1990s.

We thought that

NPT-IGBT should not be a final goal!

NPT IGBT is good because of low dose p-emitter

"N-buffer + low dose p-emitter" should be final goal.



39

In 1998, we introduced another goal: that is to realize 1.0V on-state voltage in IGBT.

NPT IGBTs were widely used at that time.

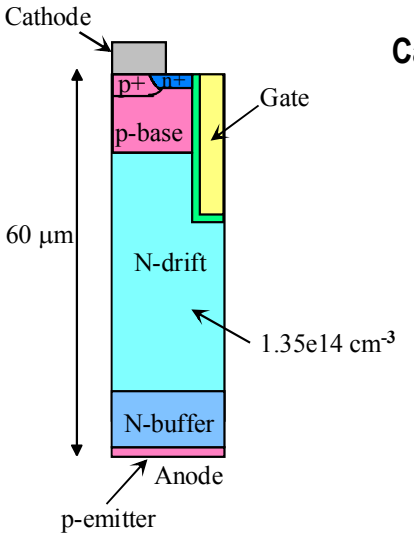
We thought

NPT-IGBT should not be a final goal!

NPT IGBT is good because NPT IGBT uses low dose p-emitter.

We thought combination of N-buffer and low dose p-emitter should be final goal.

We calculated 600V thin-wafer IGBT(FS-IGBT) in 1998



Calculated Condition:

- Total wafer thickness
  ➔60 μm
- N-buffer layer
  ➔ 600V breakdown voltage
- low dose p-emitter
  ➔High speed switching
- High carrier lifetime
  ➔10μs

40

So, we calculated electrical characteristics of 600V thin wafer IGBT structure.

We assumed:

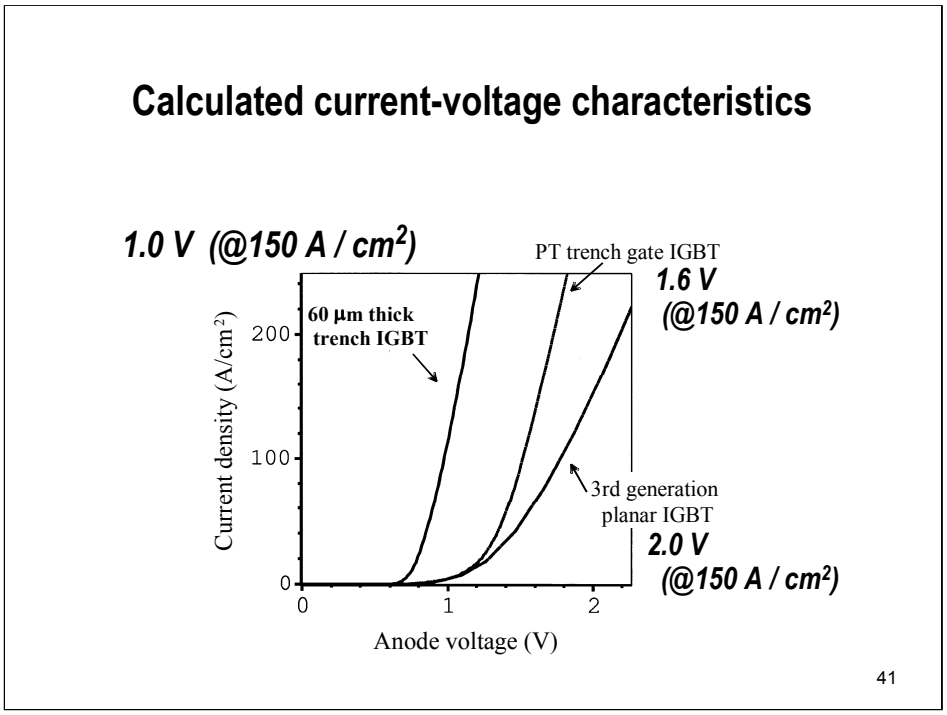
・Total wafer thickness of 60 μm.

・N-buffer layer to achieve 600V breakdown voltage.

・adopt low dose p-emitter for high speed switching

・High carrier lifetime of 10us.

We found that low on-state voltage of 1.0V was possible for this structure.



This figure shows the calculated results.

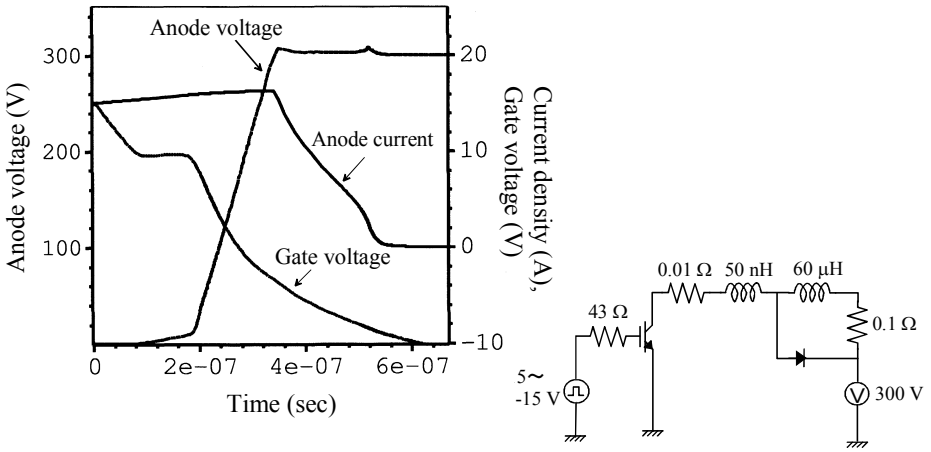
The on-state voltage of the thin wafer IGBT was almost 1V for 150A/cm2.

On the other hand, the on-state voltage of conventional trench gate IGBT was 1.6V,

And 3rd generation planar IGBT was 2.0V

Turn-off characteristics

t_f = *170 nsec* sufficiently good!



42

This shows the turn-off characteristics. Calculated fall time was 170ns for the thin wafer IGBT.

First thin wafer IGBT(FS-IGBT) paper
published in Toshiba Review in Nov., 1999

ダイオードより低いオン電圧を持つ600 V系トレンチゲートIGBTの設計

Potential of 600 V Trench Gate IGBT Having Lower On-State Voltage Drop than Diodes

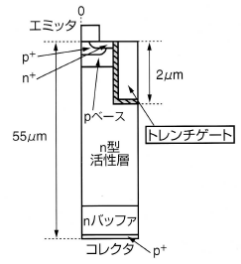
末代　知子　MATSUDAI Tomoko
中川　明夫　NAKAGAWA Akio

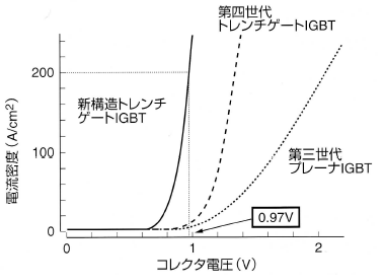
近年，パワースイッチングデバイスの主力であるIGBT（Insulated Gate Bipolar Transistor）において，オン電圧とスイッチングスピードとのトレードオフを大幅に改善できるトレンチ型のゲートを持つ構造が，幅広い定格電圧系で広く取り入れられている。

低損失化の観点から更なる低オン電圧化を目指し，素子耐圧600 V系の素子において薄層基板と微細設計ルールを用いた次世代のトレンチゲートIGBT構造を提案，検討した。高速のスイッチングスピードを維持したまま，オン電圧を耐圧600 Vのダイオードよりも低い，1 V以下にまで下げられる可能性を，シミュレーションにより示した。

The trench gate structure has recently been widely introduced both for low-voltage 600 V insulated-gate bipolar transistors (IGBTs) and high-voltage 4.5 kV IGBTs. The reason is because of the significant improvement that has been achieved in the tradeoff between device on-state voltage and switching speed.

This paper outlines the ultimate limit design and characteristics of a next-generation trench gate IGBT, using a very thin substrate and a finer design rule in the trench gate IGBT. The proposed IGBT will realize a forward voltage drop even lower than that of a 600 V diode while retaining fast switching speed.





We published the results, first in Toshiba Review in Nov. 1999 in Japanese.

2nd thin wafer IGBT(FS-IGBT) paper

published in IPEC 2000 in April, 2000

New 600 V Trench Gate Punch-Through IGBT Concept with Very Thin Wafer and Low Efficiency p-emitter, having an On-state Voltage Drop lower than Diodes

Tomoko Matsudai, *Kozo kinoshita and **Akio Nakagawa

Discrete Semiconductor Div., Semiconductor Company, Toshiba Corporation
**Information System Center, Toshiba Corporation*
***Advanced Semiconductor Devices Research Laboratories, Toshiba Corporation*
1 Komukai Toshiba-cho Saiwai-ku, Kawasaki 210-8583, Japan
Tel: +81-44-549-2602 Fax: +81-44-549-2883
E-mail: tomoko.matsudai@toshiba.co.jp

Abstract : A vertical trench gate 600 V punch through IGBT(Insulated Gate Bipolar Transistor) structure with a new concept of thin substrate wafer with a low dose n-buffer is proposed, for the first time, to realize an excellent trade-off relation between the device on-state voltage and the switching speed. This paper outlines the ultimate limit design and characteristics, using a very thin wafer in trench gate IGBTs with the transparent p+ emitter and the n-buffer layer. The realized forward voltage drop is predicted to be even lower than that of 600 V diodes with retaining a fast switching speed.

INTRODUCTION

In the vertical devices, the trench gate structure has been widely introduced both for low voltage 600 V IGBTs and high voltage 4.5 kV IGBTs[1,2,3]. The reason is because of the significant improvement that has been achieved in the trade-off relation between a device on-state voltage and a switching speed. It has been widely believed that a lower on-state voltage can be attained by electron injection enhancement from the MOS channel into the drift layer. This leads to an

In the present paper, we successfully introduce the idea of very thin wafer punch-through IGBTs, for the first time, and numerically study the potential of 600 V IGBTs, having an on-state voltage drop lower than diodes. We propose very thin substrate wafer with a low dose n-buffer and the transparent p+ emitter. The proposed IGBT will realize a on-state voltage drop that is even lower than that of 600 V diodes with retaining fast switching speed! The merits of proposed design are that the lifetime control is not necessary and the shallow trench gate is sufficient.

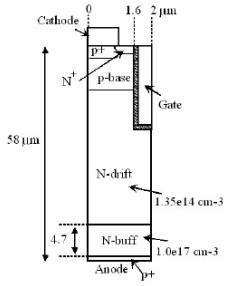


Fig. 1 Cross-sectional view of the simulated trench gate IGBT

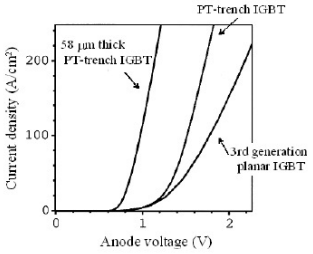


Fig. 3 Simulated forward current-voltage characteristics for conventional IGBTs and 58 μm thick trench gate IGBTs

44

And, then, in English, in IPEC2000, in April, 2000.

1200V FS-IGBT from Infineon in June, 2000

- First experimental results **ISPSD 2000**
- Established 100um wafer technology is utilized

The Field Stop IGBT (FS IGBT) –
A New Power Device Concept with a Great Improvement Potential

T. Laska[1], M. Münzer[2], F. Pfirsch[1], C. Schaeffer[3], T. Schmidt[4]

[1]Infineon Technologies, Balanstraße 73, D-81541 München
[2]eupec, Max-Planck-Straße 5, D-59581 Warstein
[3]Infineon Technologies EZM, Siemensstraße 2, A-9500 Villach
[4]Infineon Technologies OHG; Siemensstraße 2, A-9500 Villach

***Abstract*—By a vertical shrink of the NPT IGBT to a structure with a thin n⁻ base and a low doped field stop layer a new IGBT can be realized with drastically reduced overall losses. Especially the combination of the field stop concept with a trench transistor cell results in the almost ideal carrier concentration for a device with minimum on state voltage and lowest switching losses.**

I. Introduction

In recent years both IGBT concepts of PT (Punch Through) and NPT (Non Punch Through) seem to have been improved vertically to their optimum by innovations of the buffer structure and the lifetime killing process (1,2) on the PT side as well as reducing wafer thickness on the NPT side (3). Also the transistor cell structures were modified by minimizing the planar cells (4) or implementing trench cell geometry (2,5,6,7).
But this must not be the end in progress on IGBT devices, as some drawbacks of both PT and NPT IGBT still did remain: The PT IGBT has an unneccessary high carrier concentration at the back resulting in undesired high turn off current and losses or extremely high lifetime doping leading to a rather high on state voltage. In contrast, the NPT IGBT has the desired low carrier concentration at the back, but the n⁻ layer has to be rather thick due to its triangular electrical field in case of blocking condition. This rather thick n⁻ layer results in higher static and dynamic losses than necessary if the NPT structure were thinner. So both structures are not yet ideal.

advantages of the NPT concept of the low efficient emitter and the high carrier lifetime should not be given up. This is possible by implementing a field stop layer with a very low dose not influencing the low dose p emitter of the NPT IGBT but high enough for stopping the electrical field under blocking conditions (8). So it is possible to shrink the thickness of the NPT structure by 1/3, e.g. a 1200V Field Stop IGBT (FS IGBT) can be made with a thickness of 120µm instead of 175µm. In this case the field stop layer has a doping of only 10^{15} to $10^{16}cm^{-3}$ totally different from typical buffer layers in PT IGBTs. The latter have to act not only for stopping the electrical field but also for reducing the enormous p emitter efficiency in PT IGBTs and therefore have dopings of more than $10^{17}cm^{-3}$ at thicknesses of 10µm or more (figure 1).

	PT-IGBT	NPT-IGBT	FS-IGBT
p-emitter	very high efficient	low efficient	low efficient
n⁻-layer	thin	medium	thin
additional n-layer	buffer layer = highly doped -to reduce the very high emitter efficiency -to stop the electrical field	no	field stop layer = weakly doped -to stop only the electrical field
carrier lifetime	low (lifetime killing)	high	high

Fig.1: Comparison of the different IGBT concepts

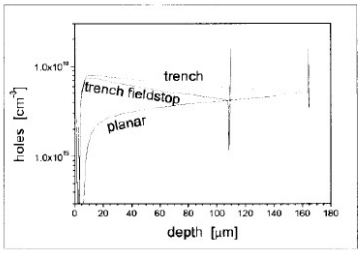


Fig. 7: Simulated carrier concentration of planar NPT, trench NPT and trench FS IGBT

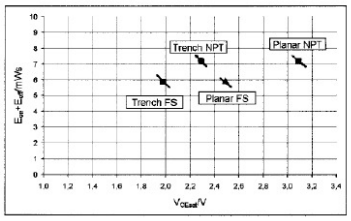
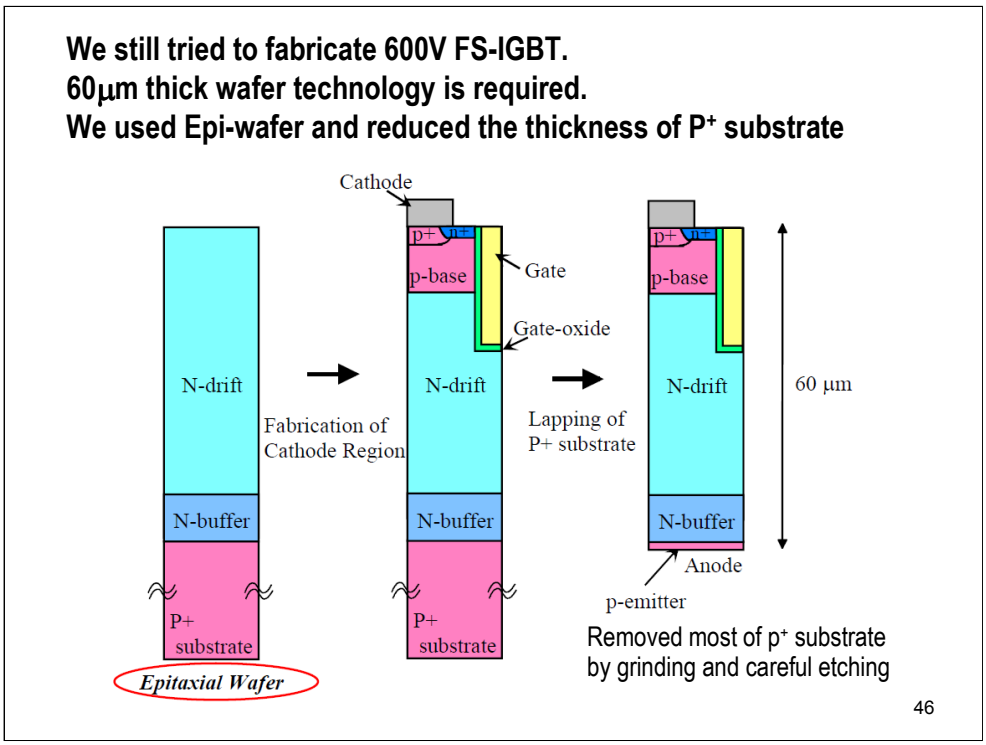


Fig. 6: Trade off diagramm switching losses versus on state voltage (at 125°)

Infineon published first experimental results, in ISPSD in June, 2000.

Their device is 1200V IGBT, and thus, they could use well established 100um thick wafer technology.



We still tried to fabricate 600V thin wafer IGBT.

We needed 60um thick technology.

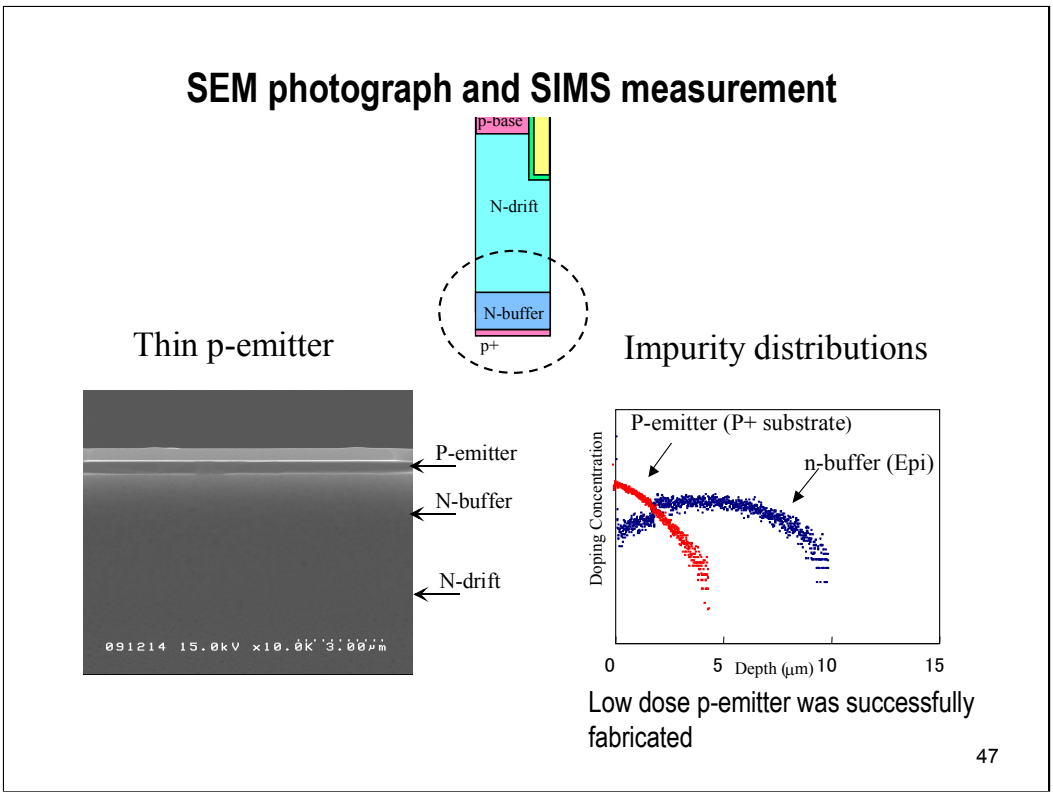
60um is too thin, and we did not have such thin wafer technology.

So, we finally used a trick.

We first fabricated this IGBT using conventional epitaxial wafer.

We removed most of the thick p+ substrate by grinding and careful etching technique.

And finally we got very thin p-emitter.



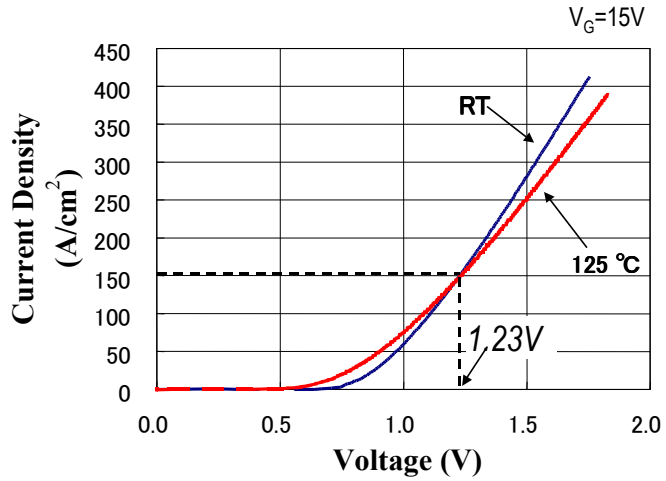
These figures are the SEM photograph

and the SIMS measurement result of the anode region.

This is the SIMS measurement of the impurity distribution in the anode region.

This very low dose p-emitter was successfully fabricated.

Measured forward current-voltage characteristics



48

This figure shows the measured current-voltage characteristics of the fabricated 600V thin wafer IGBT.

This blue line shows the curve for the room temperature,

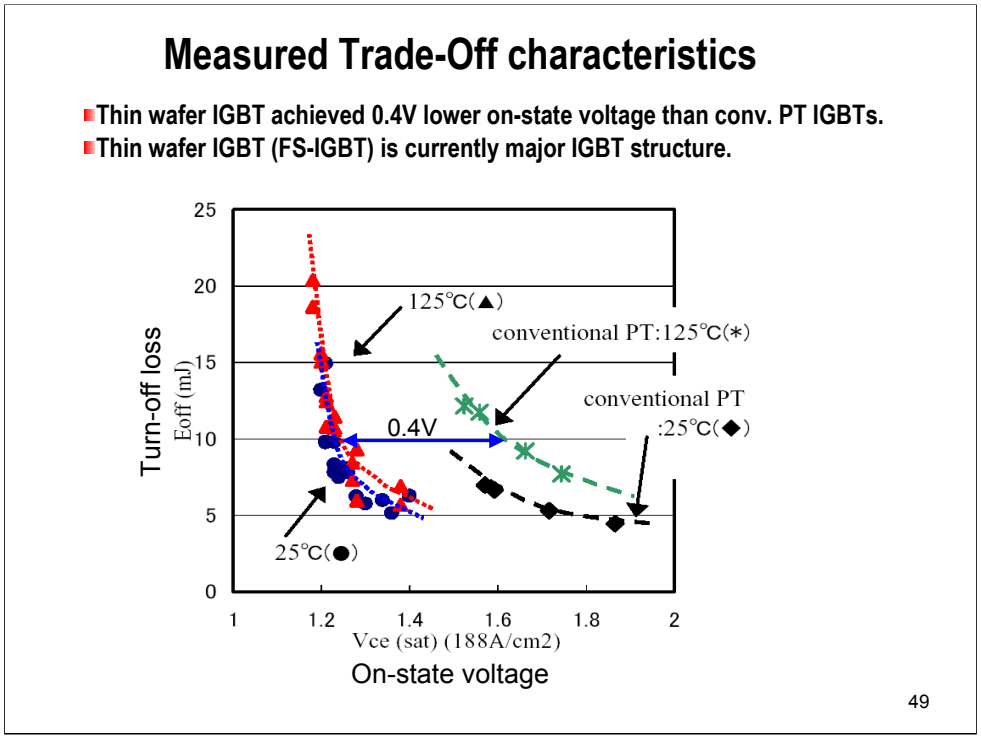
and, the red line shows the curve for 125 C.

The gate voltage is 15 V.

The V_{CE} of the fabricated IGBT

is 1.23 V for 150 A/cm^2 at room temperature.

It was found that the very thin wafer IGBT with n-buffer and a low dose p-emitter is effective in reducing the V_{CE} of trench gate IGBTs.



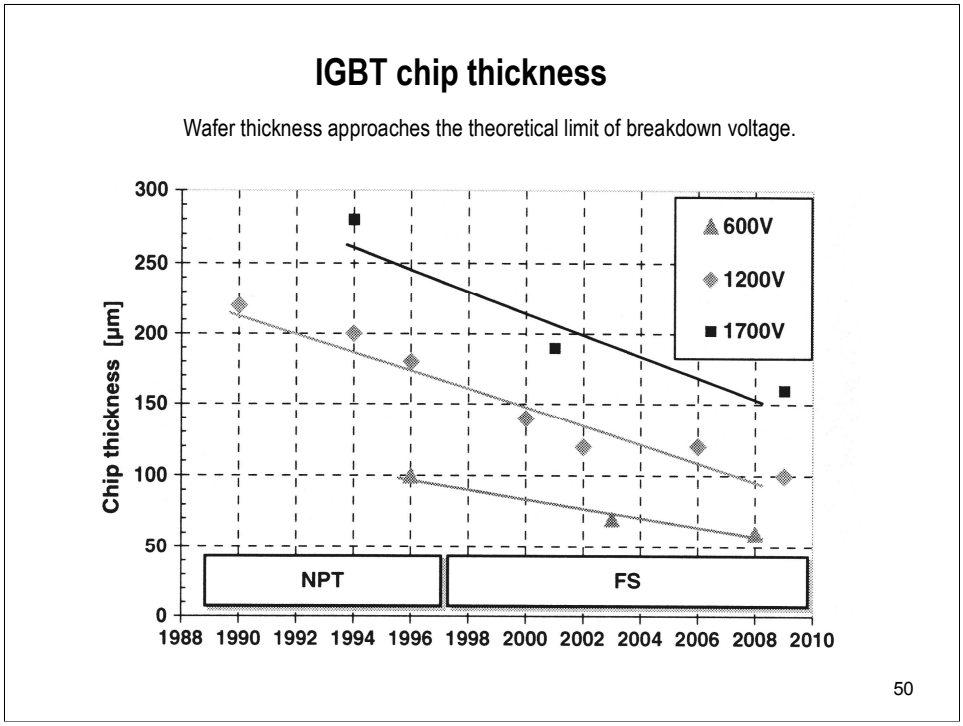
This figure shows the measured trade-off relation between the device on-state voltages and the turn-off losses.

The typical characteristics of conventional PT-IGBT are plotted in this figure as a comparison.

The excellent trade-off relation for 600 V IGBT was obtained both for 25 °C and 125 °C.

Thin wafer IGBT achieved 0.4V lower on-state voltage than conv. PT IGBTs.

And, this device structure, or FS-IGBT, is currently major IGBT structure.

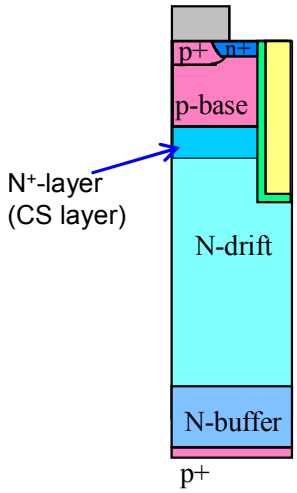


This figure shows wafer thickness of Infineon's IGBT as a function of year.

The thickness almost approaches the theoretical limit of breakdown voltage.

CSTBT (Carrier Stored Trench Gate IGBT) in 1996

- High impurity concentration N layer (CS) is formed inside the trench.
- The CS layer works as a barrier for holes.
- CSTBT achieves the same effect as IEGT



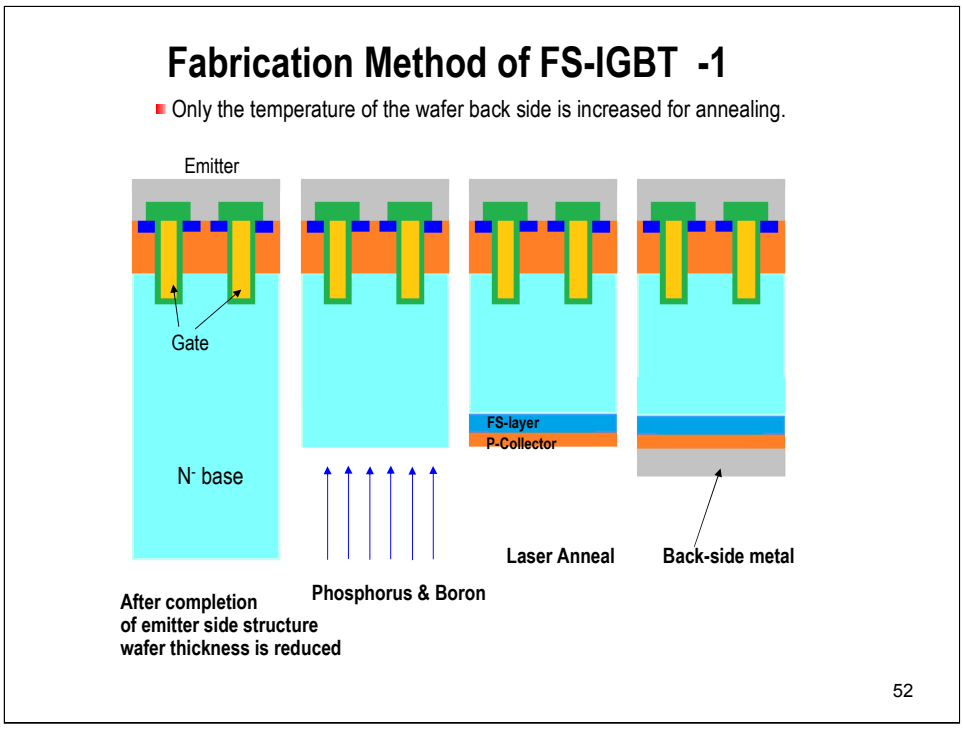
51

Carrier stored Trench gate IGBT was proposed by Mitsubishi electric 1996.

High impurity concentration N layer called “Carrier Storage layer” is formed inside the trench.

The CS layer works as a barrier for holes.

Thus, CSTBT has the same effect as IEGT.



Now I would like to talk about fabrication method of FS-IGBT.

First, cathode side fabrication process is completed.

After completion of cathode side structure, wafer thickness is reduced.

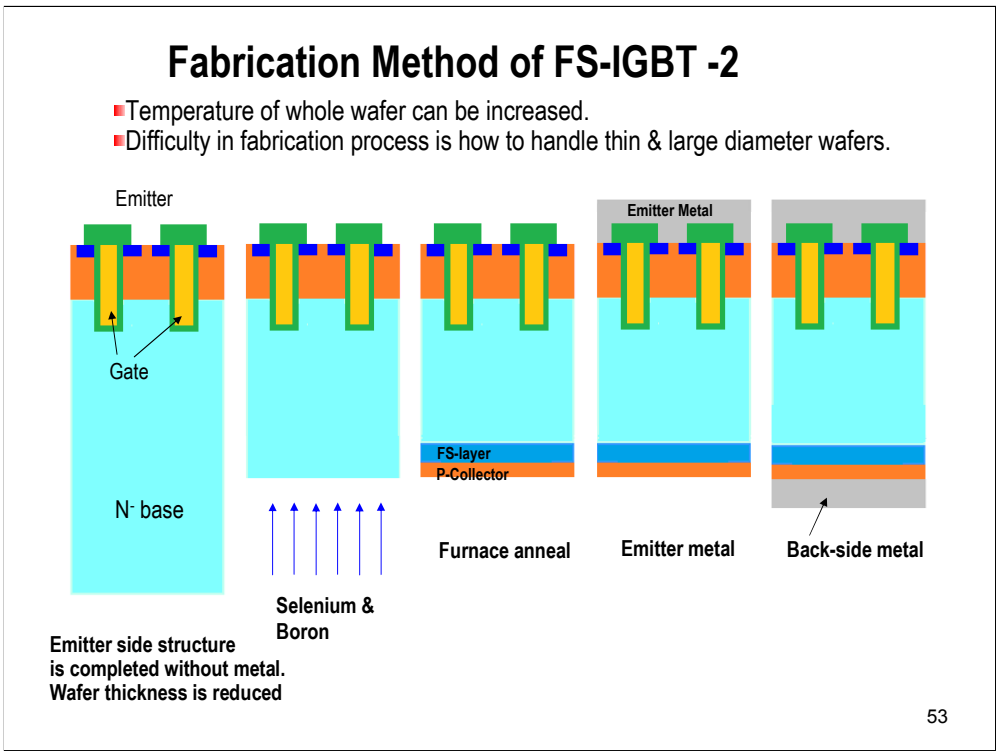
Then, Phosphorus and boron are implanted.

Laser is used to anneal and activate impurities in the back side surface.

As there is a metal layer on the emitter side of the wafer, the temperature of the whole wafer cannot be increased.

Only the temperature of the back side surface of the wafer can be increased in a short time period.

At last, back-side metal is formed.



This is another fabrication method.

First, the emitter side structure is completed without cathode metal.

The wafer thickness is reduced.

Then, selenium and boron are implanted and annealed at high temperature furnace.

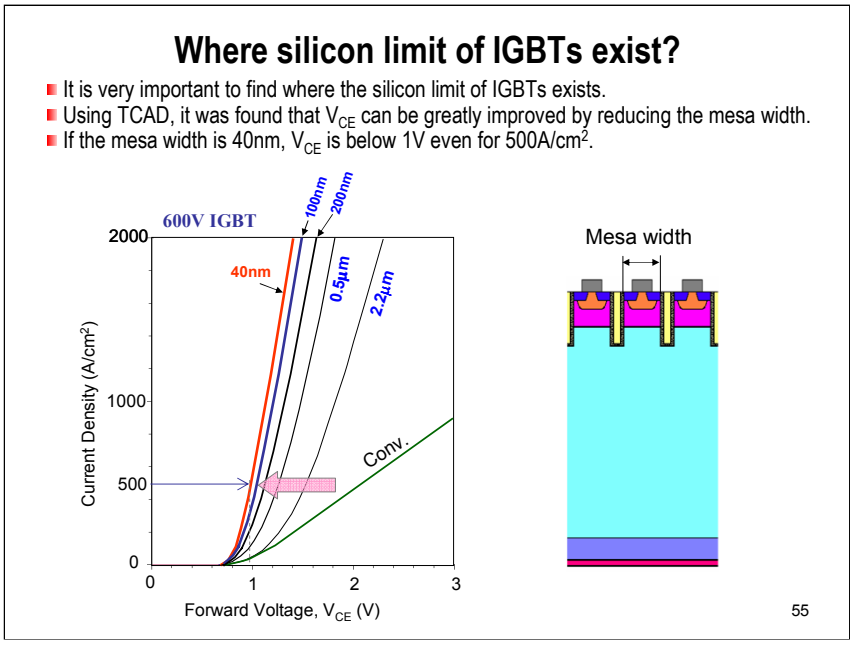
There is no metal layers, temperature of the whole wafer can be increased.

Finally, emitter and collector metals are deposited and patterned.

The difficulties in fabrication process of FS-IGBTs are how to handle thin large diameter wafers.

Cracking and chipping wafers and the wafer warpage are the issues we have to handle.

3. Silicon limit characteristics of IGBT



Now, It is very important to find where the silicon limit of IGBTs exists.

We found by using TCAD that forward voltage can be greatly improved by reducing the mesa width.

The mesa width is defined as the distance from trench to trench.

If the mesa width is very narrow such as 40nm,

the forward voltage is less than 1V even for a very high current density of 500A/cm^2.

Derive analytical I-V curve of ideal silicon limit IGBT

$\mu_{electron} > \mu_{hole}$

If all current flows by electrons, this gives the lowest forward voltage!

Now, we assume the following:

(1)N-base is in high injection condition.
n=p

(2) n_1 denotes the electron density at the collector-side of N-base and n_W denotes the electron density at the emitter-side.

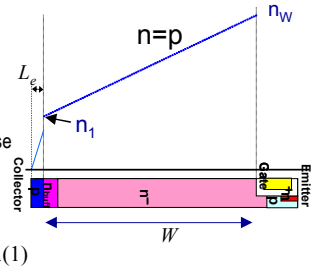
(3) All the current flows by electrons, thus, hole current is zero.

$$J_p = -qD_p\frac{\partial n}{\partial x} + qn\mu_p E = \mu_p(-kT\frac{\partial n}{\partial x} + qnE) = 0 \quad ...\text{Eq.(1)}$$

From Eq.(1) $E = \frac{kT}{q}\frac{1}{n}\frac{\partial n}{\partial x}$ ...Eq.(2)

(4) Using Eq.(2), electron current is given by Eq.(3), :

$$J = qD_n\frac{\partial n}{\partial x} + qn\mu_n E = \mu_n(kT\frac{\partial n}{\partial x} + qnE) = 2qD_n\frac{\partial n}{\partial x} ...\text{Eq.(3)}$$



56

Now, I derive analytical current-voltage relation of ideal silicon limit IGBT.

Generally, the electron mobility is greater than hole mobility.

So, if all the current flows by electron, this gives the lowest forward voltage.

Now, we assume the following:

(1) N-base is in high injection condition, and, thus, n=p holds.

(2) n1 denotes the electron density at the collector side of the N-base and nW denotes electron density at the emitter side.

(3)Because all the current flows by electrons, hole current should be zero.

From eq.(1), the electric field in the N-base is given by Eq.(2).

(2) then, electron current is given by this equation.

Here, as the electric field is given by Eq.(2), the electron current is given by double the diffusion current.

(1) Next, obtain the voltage drop in the n-base: V_i

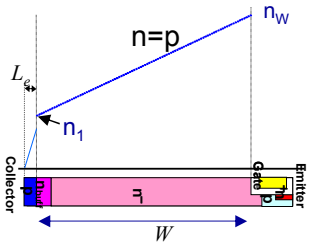
Cite Eq.(2) again.

$$E = \frac{kT}{q}\frac{1}{n}\frac{\partial n}{\partial x} \quad \text{...Eq.(2)}$$

Integrate E over the entire N-base, from x=0 to x=w.

$$\int_0^w E dx = V_i = \frac{kT}{q}\int_{n_1}^{n_W}\frac{1}{n}dn$$

$$V_i = \frac{kT}{q}\ln\left(\frac{n_W}{n_1}\right) \quad \text{...Eq.(4)}$$



P-collector, N-base and Accumulation layer form a pin diode.
The external junction voltage, V_j is given by Eq.(5), which will be derived in Chapter 3.

$$V_j = \frac{kT}{q}\ln\left(\frac{n_1 n_W}{n_i^2}\right) \quad \text{...Eq.(5)}$$

The voltage drop of this pin diode, V_{pin} is given by Eq.(6).

$$V_{pin} = V_i + V_j = \frac{2kT}{q}\ln\left(\frac{n_W}{n_i}\right) \quad \text{...Eq.(6)}$$

57

Next, we obtain the voltage drop in the n-base: Vi

cite Eq.(2) again.

Integrate E over the entire N-base, from x=0 to x=w.

Then, Eq.(4) is obtained.

Vi is the voltage drop in the n-base.

P-collector, N-base and Accumulation layer forms a pin diode.

The external junction voltage, *Vj* is given by Eq.(5).

(Equation (5) will be derived, later.

We simply use this relation here.)

Then, the voltage drop of this pin diode, V_{pin} is given by Eq.(6).

(2) In this slide, we obtain the value of n_W

Cite Eq.(3), again.

$$J = 2qD_n \frac{\partial n}{\partial x} \quad \text{.... Eq.(3)}$$

D_n is approximately expressed, using Eq.(7).

$$J = 2q \frac{a}{n+b} \frac{\partial n}{\partial x} \quad \text{... Eq.(3.1)}$$

$$D_n = \frac{a}{n+b} \quad \text{...Eq.(7)} \qquad (a = 3.7e18cm^{-1}s^{-1},\ b = 9.39e16cm^{-3})$$

Integrating Eq.(3.1) from x=0 to W, we have Eq.(8).

$$\frac{J}{2qa} \int_0^W dx = \int_{n_1}^{n_W} \frac{1}{n+b} dn$$

$$\frac{JW}{2qa} = \ln\left(\frac{n_W + b}{n_1 + b}\right) \quad \text{... Eq.(8)}$$

Solve Eq.(8) for n_W.

$$n_W = (n_1 + b)\exp\left(\frac{JW}{2qa}\right) - b \quad \text{... Eq.(9)}$$

58

In this slide, we obtain the value of n_w.

We show again equation(3).

The diffusion coefficient Dn is approximated by equation (7).

integrate Eq.(3.1) from x=0 to w.

We get Eq.(8).

Solve Eq.(8) for n_w, we have Eq.(9).

Eqs.(5) & (9) are cited again.

$$V_{pin} = \frac{2kT}{q}\ln(\frac{n_W}{n_i}) \quad ...Eq.(5)$$

$$n_W = (n_1 + b)\exp(\frac{JW}{2qa}) - b \quad ...Eq.(9)$$

Substituting Eq.(9) for n_W in Eq.(5), we have

$$V_{pin} = \frac{2kT}{q}\ln[\frac{1}{n_i}\{(n_1 + b)\exp(\frac{JW}{2qa}) - b\}] \quad ...Eq.(10)$$

In p-collector, J_n is expressed as:

$$J = J_n = \frac{qD_E n_1^2}{Q} \quad ...Eq.(11)$$

Solve Eq.(11) for n_1.

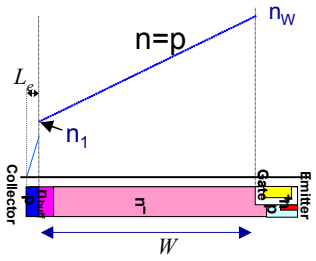
$$n_1 = \sqrt{\frac{QJ}{qD_E}} \quad ...Eq.(12)$$

Substitute Eq.(12) for n_1 in Eq.(10).

$$V_{pin} = \frac{2kT}{q}\ln[\frac{1}{n_i}\{(\sqrt{\frac{QJ}{qD_E}} + b)\exp(\frac{JW}{2qa}) - b\}]$$

The forward voltage of IGBT is, then, given by adding the voltage drop in the channel.

$$V_F = V_{pin} + V_{ch} = \frac{2kT}{q}\ln[\frac{1}{n_i}\{(\sqrt{\frac{QJ}{qD_E}} + b)\exp(\frac{JW}{2qa}) - b\}] + R_{ch}J \quad ...Eq.(13)$$



59

We show equation(5) and (9) again.

Substitute equation (9) for n_w in equation(5).

We have equation (10).

By the way, total current J is equal to electron current, Jn.

In the p-collector, electron current flows by diffusion.

As the P-collector dose is small, the electron current is given by equation (11).

(The equation (11) will be derived later in diode section.)

Then, solve equation(11) for n1.

We have equation(12).

Substitute eq.(12) for n1 in equation(9).

Then, we have pin diode voltage, Vpin.

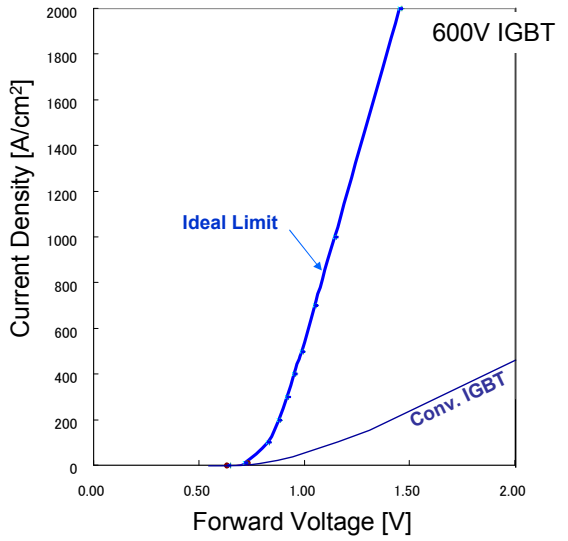
The voltage drop of ideal IGBT is given as the summation of pin diode voltage, V_{pin}, and the voltage drop in the channel, V_{ch}.

Vch is given by the product of channel resistance and the current density J.

Equation (13) is the current voltage relation of ideal IGBT.

Analytical I-V of Ideal IGBT Limit

$$V_F = \frac{2kT}{q}\ln[\frac{1}{n_i}\{(\sqrt{\frac{QJ}{qD_E}} + b)\exp(\frac{JW}{2qa}) - b\}] + R_{ch}J \quad \text{...Eq.(13)}$$

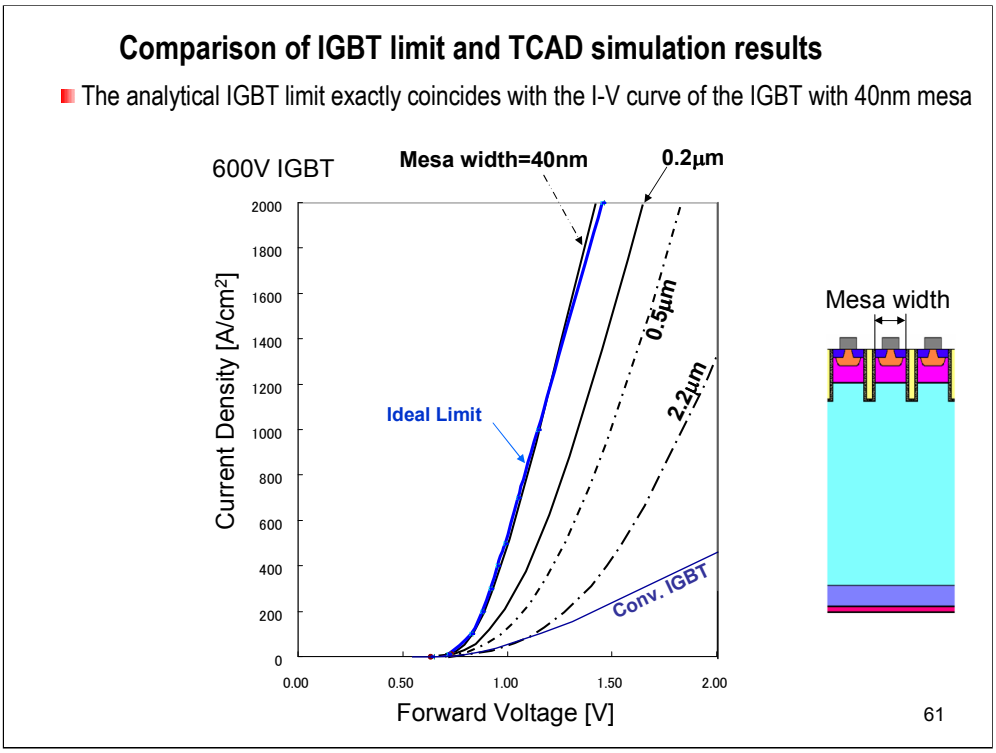


60

This figure shows the current-voltage curves of the ideal IGBT limit.

Conventional IGBT current voltage curve is also shown for comparison.

It is easily seen that the ideal IGBT limit is far better than conventional IGBT.

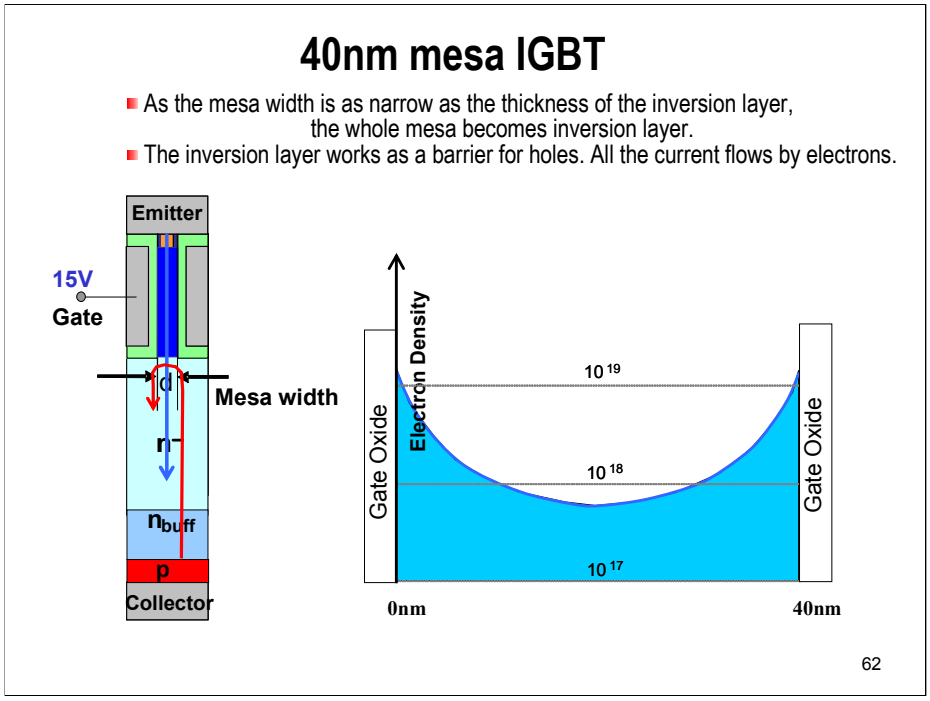


This figure compares the IGBT limit with TCAD simulation results.

The numbers show the mesa width.

The ideal IGBT limit exactly coincides with the I-V curve of the IGBT with 40nm mesa,

Ideal IGBT limit is difficult to realize because the very narrow mesa is required.



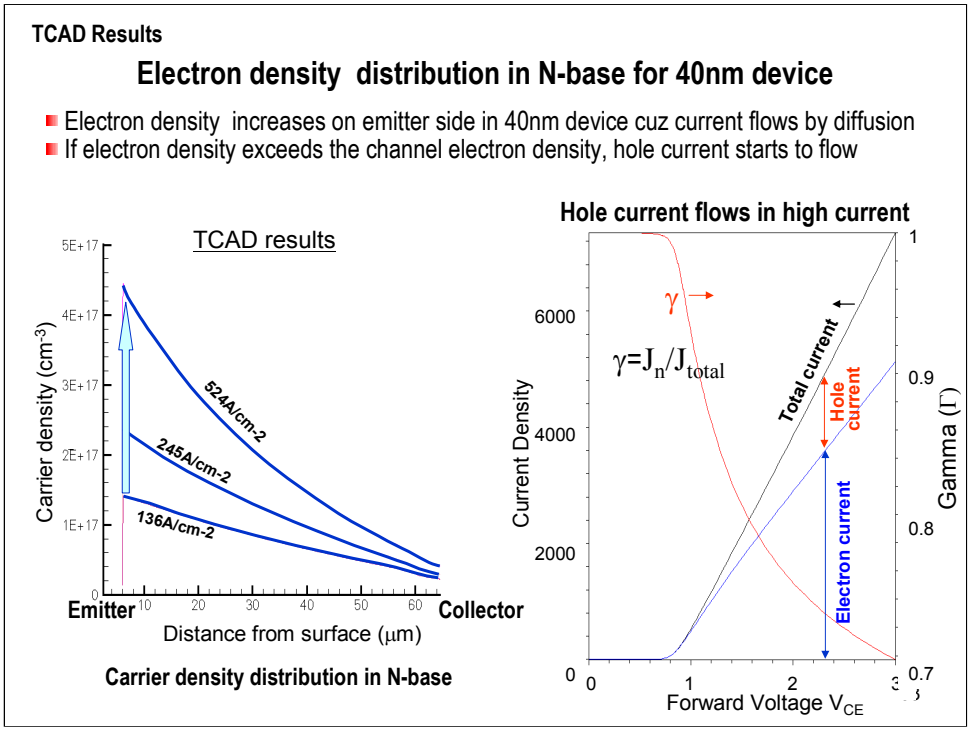
Now, I'll talk about 40nm mesa IGBTs.

If the mesa width is as narrow as the thickness of the inversion layer,

the whole narrow mesa becomes inversion layer and a high concentration electron density layer appears.

This inversion layer works as a barrier for holes and prevents hole current flow.

All of the current flows by electrons for the very narrow mesa IGBTs.



This figure shows the carrier density distribution in the n-base.

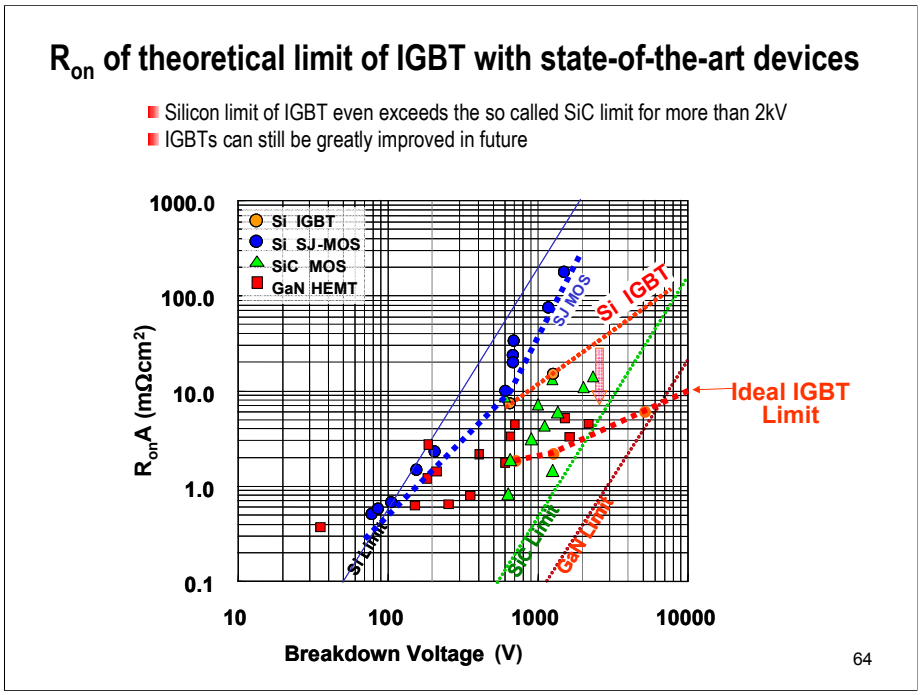
The carrier density on the emitter side in the n-base increases rapidly as current density increases because current flows by diffusion.

If the carrier density exceeds the induced channel electron density in the mesa, hole current starts to flow and the electric field in the n-base increases.

The right figure shows the total current, electron current and electron injection efficiency as a function of forward voltage.

The black line shows total current and blue line shows electron current. This portion shows the electron current magnitude and the this portion shows the hole current magnitude.

The hole current starts to flow as the total current exceeds 500A/cm^2, and the electron injection efficiency rapidly decreases.



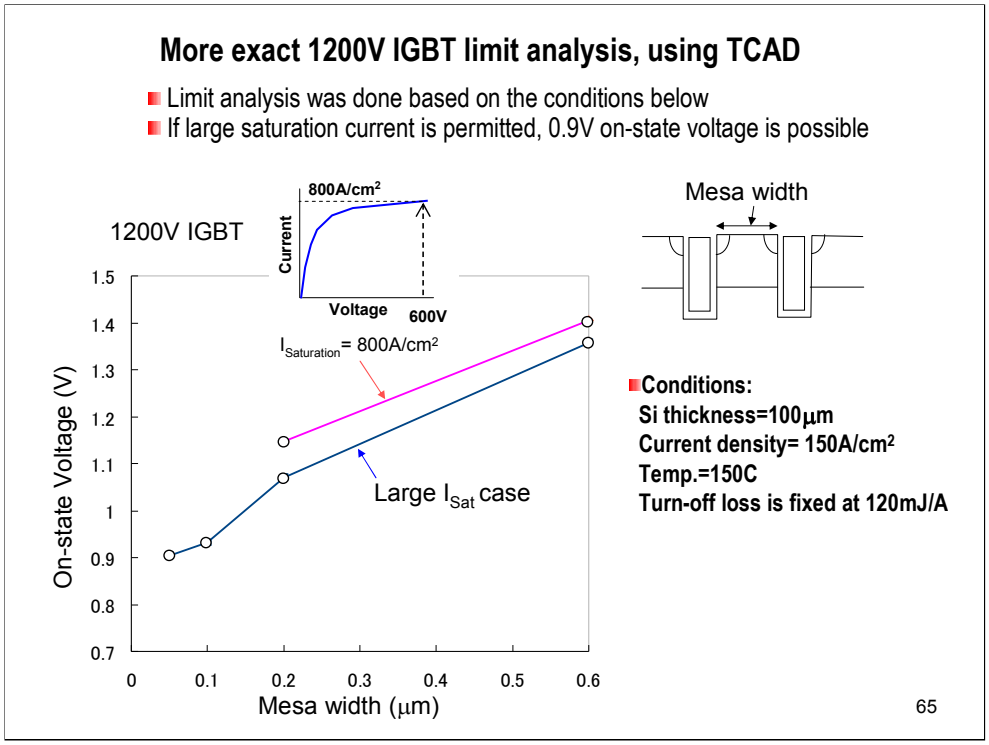
This figure shows the on-resistance of theoretical limit of IGBT together with various state-of-the-art devices.

The vertical axis shows on-resistance and horizontal axis is breakdown voltage.

The blue line shows silicon super-junction MOSFET.

IGBT, SiC MOSFET and GaN HEMT.

The silicon limit of IGBT even exceeds the so called SiC limit for more than 2kV.



I carried out more exact 1200V IGBT limit analysis, using TCAD.

The conditions are:

Silicon wafer thickness=100um

Current Density =150A/cm2

Temp.=150C

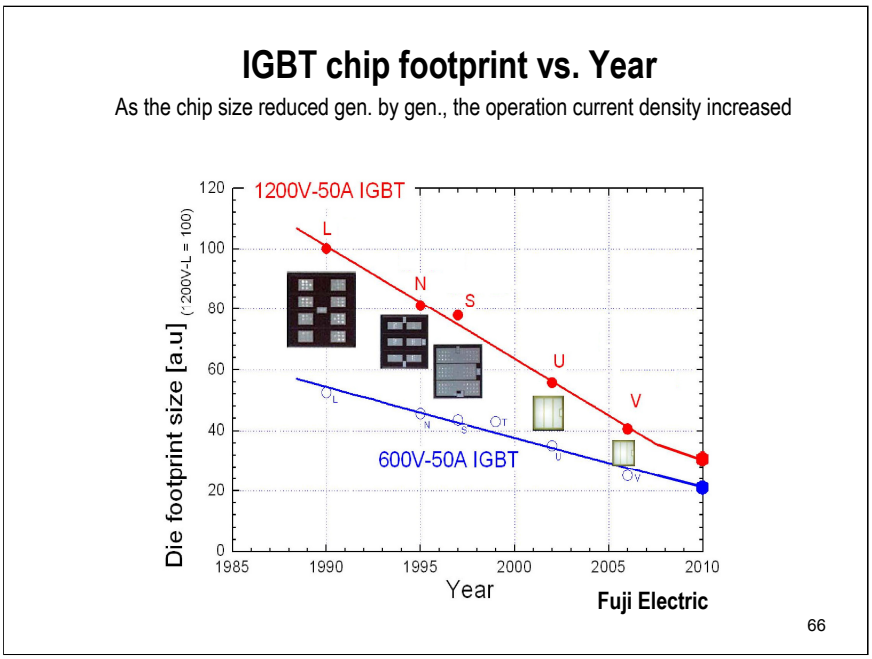
Turn-off loss is adjusted to be 120uJ/A

The red line is the on-state voltage as a function of the Mesa width when the saturation current density is fixed at 800A/cm2.

The lowest on-state voltage is around 1.15V.

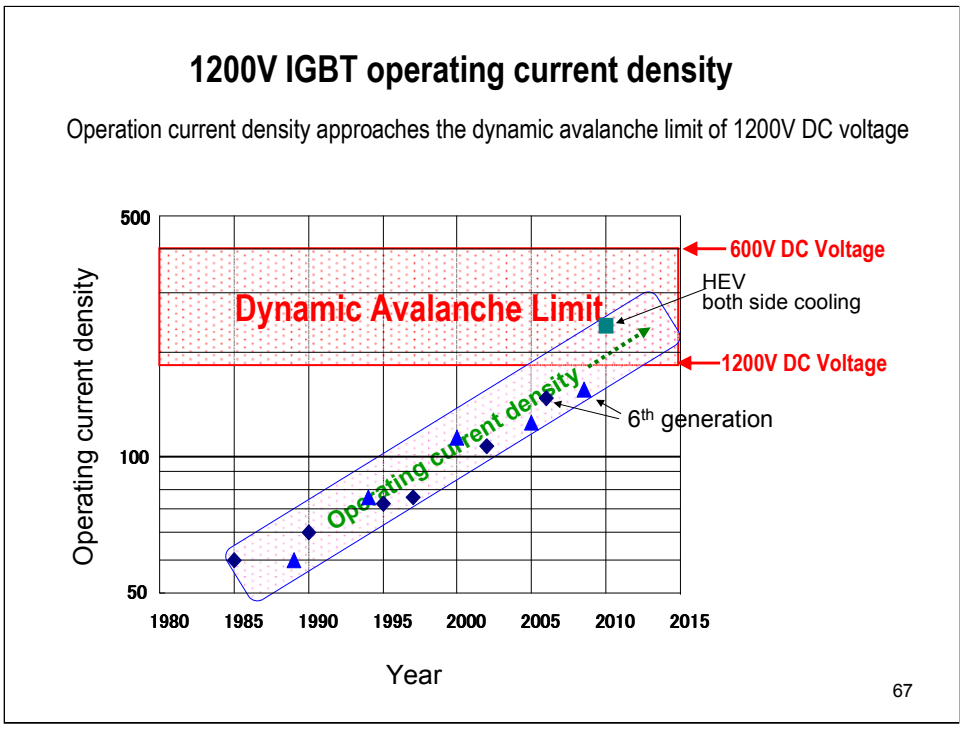
The black line is the case for large saturation current.

If large saturation current is permitted, 0.9V on-state voltage is possible.



IGBT chip size reduced by generations.

This shows the case of Fujielectric. As the chip size reduced, the operation current density increased.



This shows operation current density for 1200V IGBTs as a function of year.

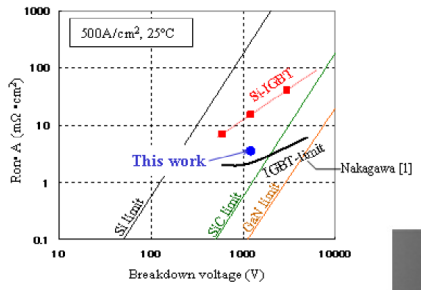
The operation current density increased by generations.

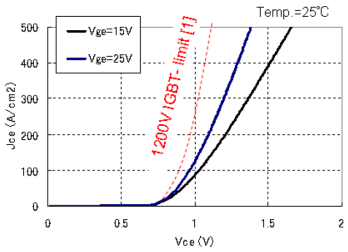
Now, the operation current approaches the dynamic avalanche limit of 1200V DC voltage.

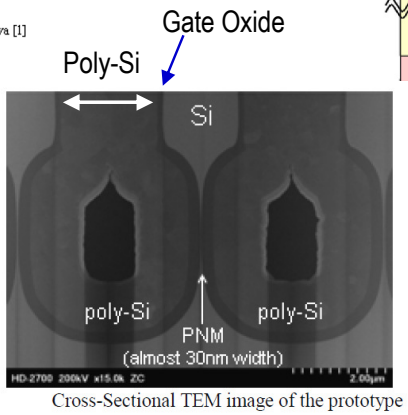
Further increase in the operation current will needs special care.

Denso developed "Partially Narrow Mesa IGBT" ISPSD2012

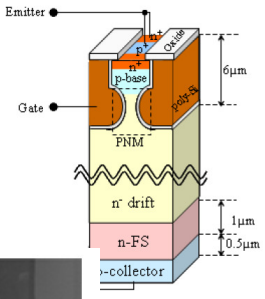
- Trench gate is widened in the bottom portion, and narrow mesa can be realized
- Denso actually fabricated 30nm mesa IGBT and realized low Ron close to the IGBT limit.







Cross-Sectional TEM image of the prototype



68

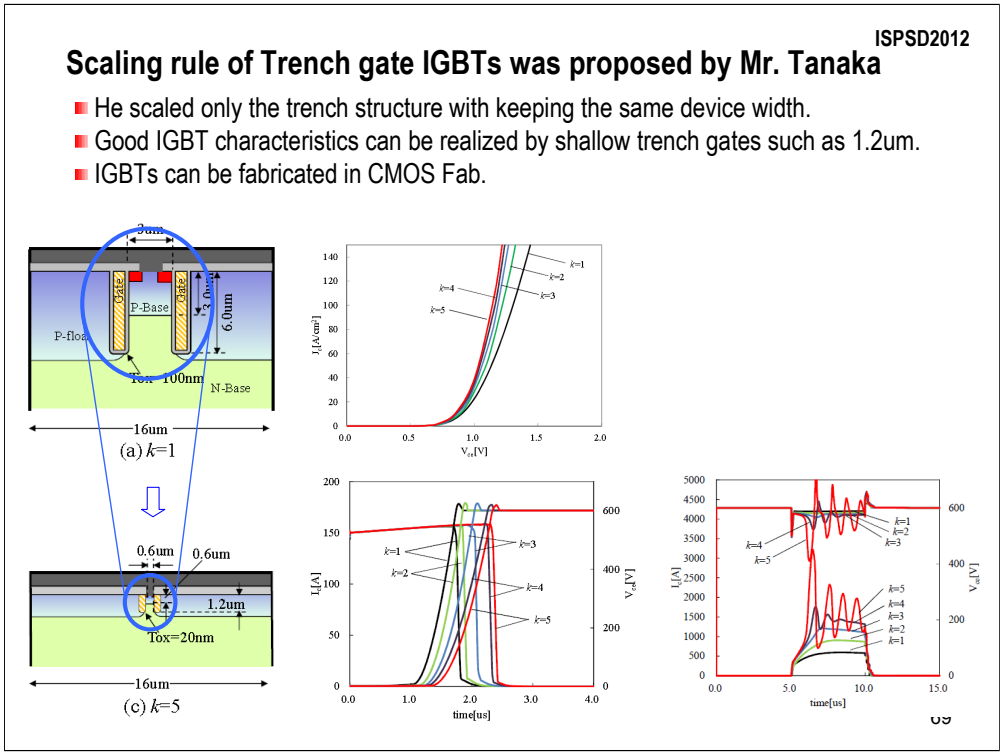
Recently, Denso developed a unique trench gate structure.

This trench gate is widened in the bottom portion, and, thus, narrow mesa can be easily realized.

They call this device partially narrow mesa IGBT.

Denso actually fabricated 30nm wide mesa IGBT.

They realized very low on-resistance close to the limit of 1200V IGBT.



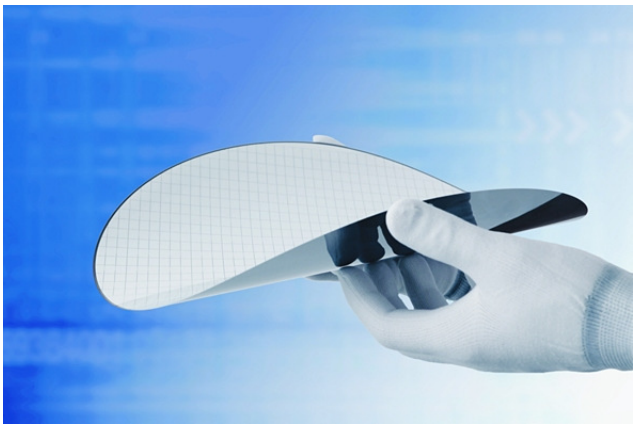
Tanaka-san recently proposed a scaling rule of IGBTs.

According to the scaling, good IGBT characteristics can be realized by very shallow trench IGBTs.

For example, the trench depth is as shallow as 1.2um.
This paper shows that good IGBTs can be fabricated in CMOS fabrication facility.

30cm CMOS Fab for Power Devices

contributes to better performance and low cost.



70

Infineon recently announced that it already introduced 30cm CMOS fab for power device fabrication.

CMOS fab will surely contribute to increasing device performance and also to reducing chip cost.

Patent of IGBT

1972 Yamagami -- He invented the basic structure of IGBT
(He filed patent only in Japan)

1978 J.D. Plummer discovered " IGBT mode operation in thyristor"
and was granted a patent.

1980 Hans Becke invented basic idea of IGBT.
He claimed "**no thyristor action occurs under any device operating conditions**"
Becke's patent eventually became basic IGBT patent after development of non-latch-up IGBTs!!!

1984 Nakagawa invented the design concept of Non-Latch-Up IGBT.
Saturation current ＜ Latch-up current

71

Now, I'd like to talk about the patent of IGBTs.

IGBT device structure was first invented by Yamagami, in 1972. He filed his patent only in Japan.

In 1978, Prof. Plummer discovered IGBT-mode operation in thyristor, and got a patent of IGBT-mode operation in Thyristor.

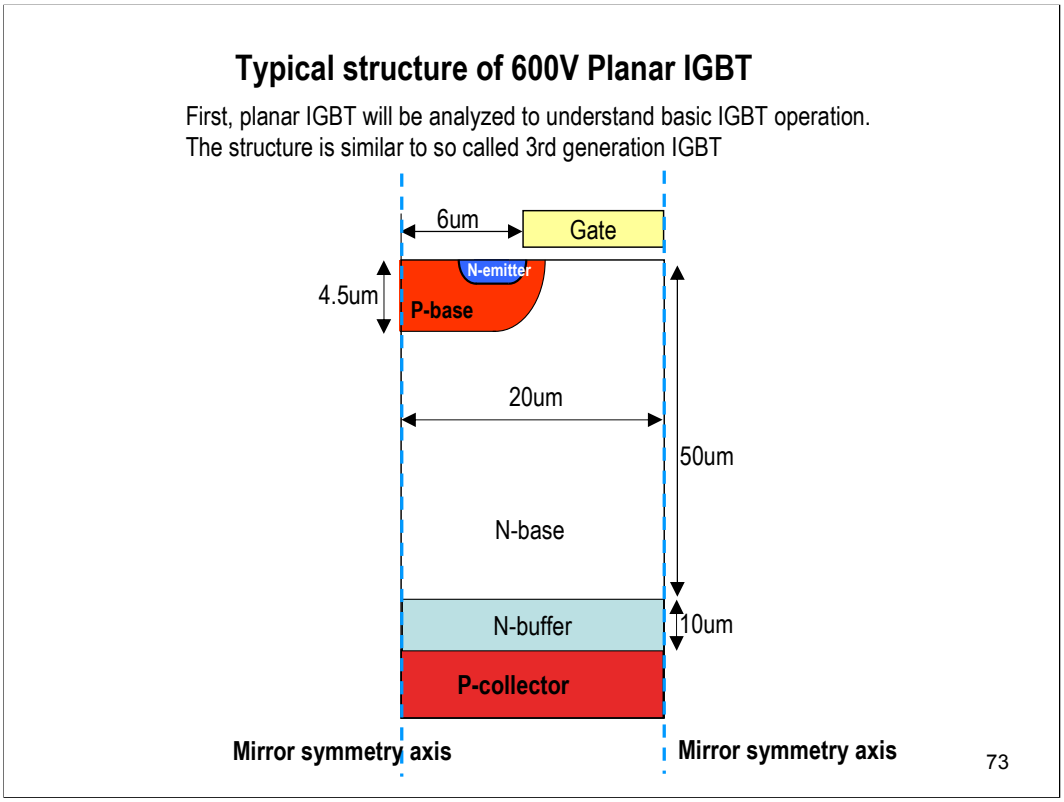
Becke tried to get a patent of IGBT, but he failed because Plummer already discover the IGBT-mode operation in Thyristor. So, in his claim, he included the restriction "no thyristor action occurs under any device operating conditions."

Becke's patent eventually became the basic IGBT patent, because I developed non-latch-up IGBT.

Baliga's IGBT latched-up, so his IGBT is the invention of Plummer, not the IGBT of Becke.

4. Basics for IGBT Design

Next, I talk about basic IGBT design and the mechanism of its operation.



This shows typical structure of 600V IGBT.

First, planar IGBT will be analyzed to understand basic IGBT operation.

The structure is similar to so called 3rd generation IGBT.

For the time being, I will analyze this structure.

The device half cell width is 20um.

The half of P-well width is 6um.

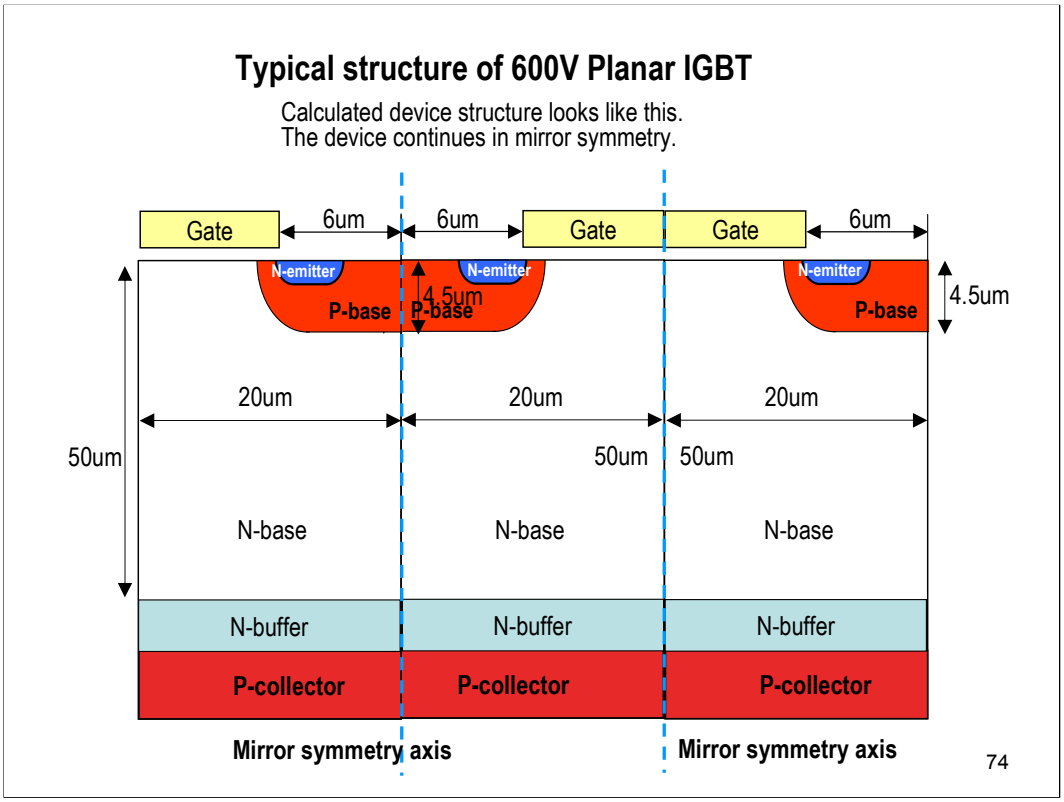
The P-well diffusion depth is 4.5um.

The thickness of the N-base is 50um.

The thickness of the N-buffer is 10um.

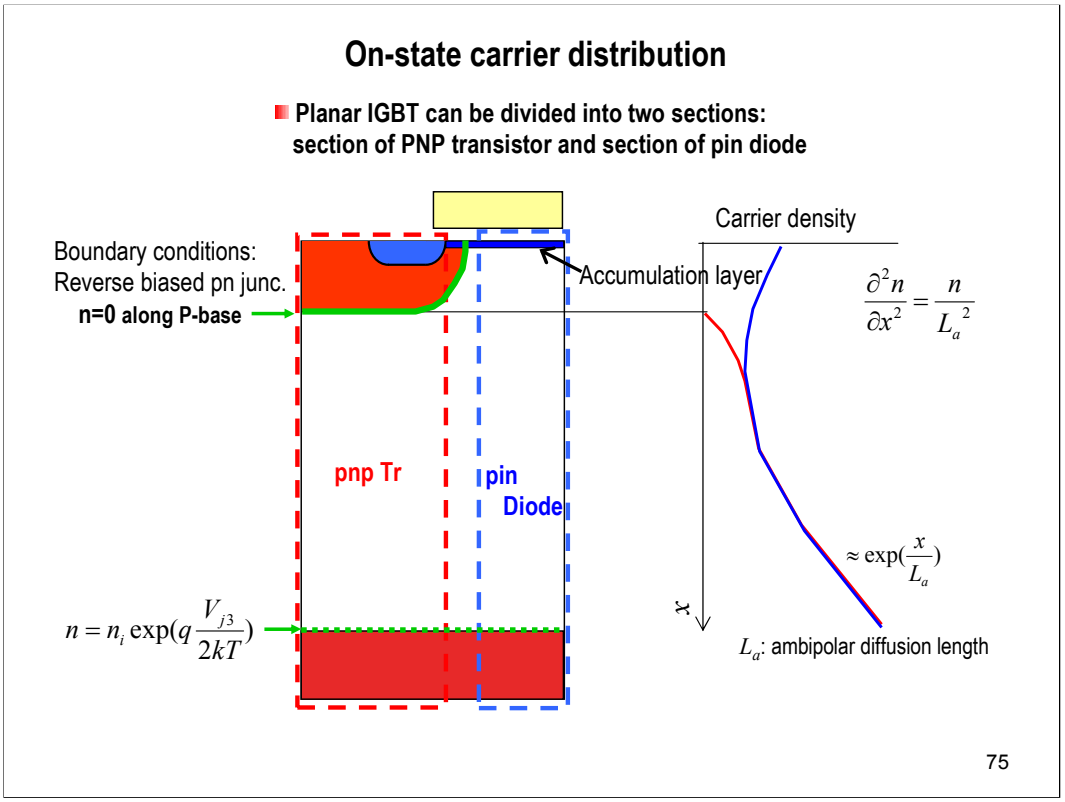
The both device sides are mirror symmetry axis

So, the actual device structure is shown in the next slide.



The calculated device structure looks like this.

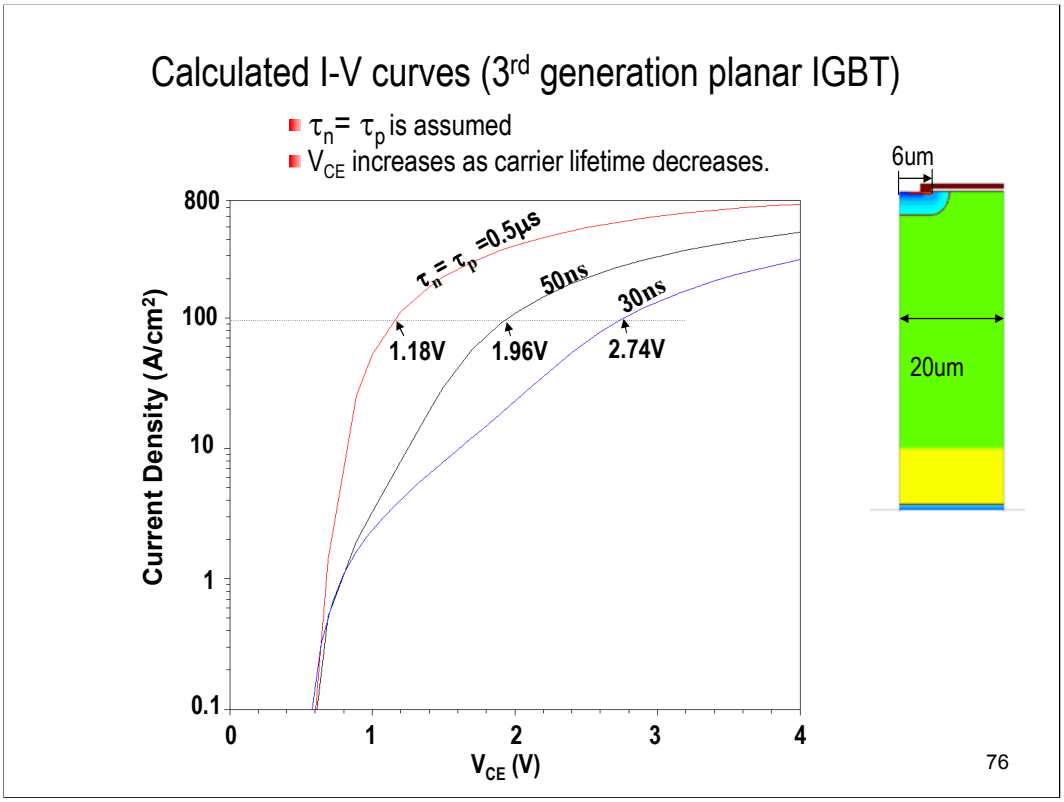
The device continues in mirror symmetry.



The device can be divided into two sections: a section of PNP transistor and a section of pin diode.

In the PNP transistor section, the P-base N-base junction is reverse biased, and the carrier density at this junction is zero. The carrier density decreases exponentially from the P-collector N-base junction toward the P-base and is approximately in proportion to $\exp(x/L_a)$, where L_a is the ambipolar diffusion length.

On the other hand, in the pin diode section, electrons are injected from the accumulation layer and holes are injected from the P-collector. Thus, the carrier density distribution in the pin diode section is similar to that of the pin diode. The carrier density is high both at emitter side and also at the collector side.



This is the calculated I-V relation of the planar IGBT.

For simplicity, electron and hole lifetimes are set to be the same.

The horizontal axis is VCE, and vertical axis is current density.

The parameters are electron and hole lifetimes.

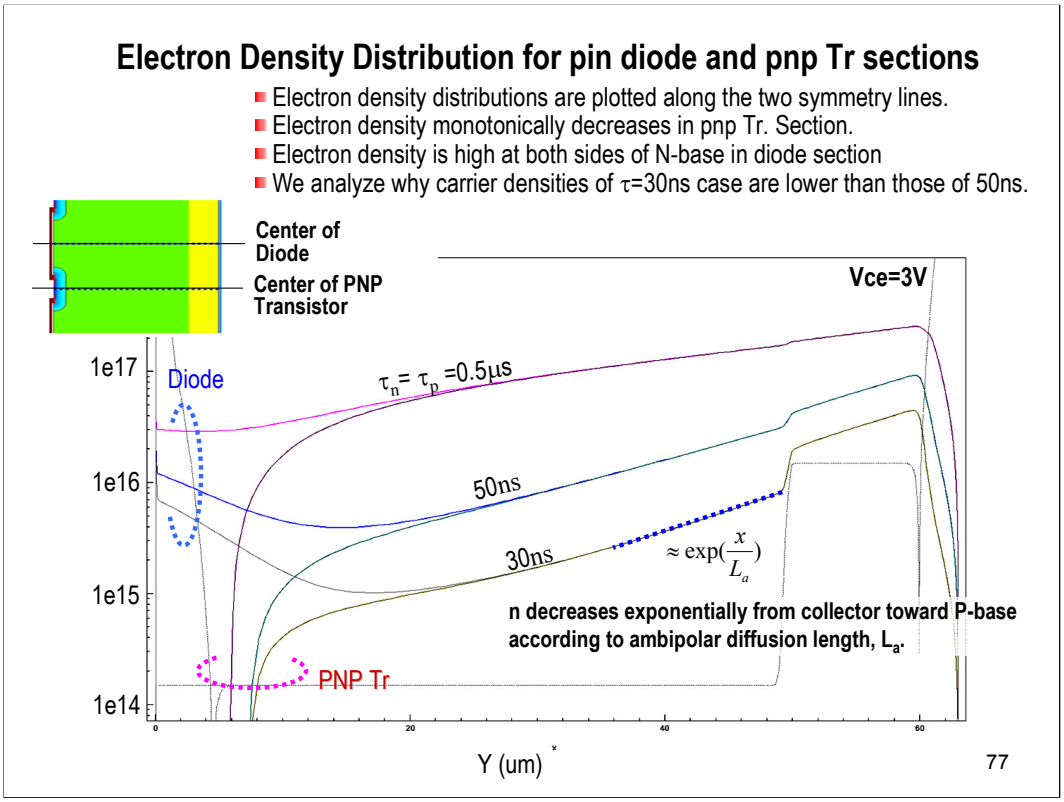
The collector emitter voltage increases as the carrier lifetime reduces.

For example, V_{CE} is 1.18V when τ=0.5us.

V_{CE} increases to 2.74V if τ is reduced to 30ns.

The device structure is the same as the one I already mentioned before.

The device half cell width is 20um. The half of P-well width is 6um. The P-well diffusion depth is 4.5um.



This figure shows electron density distributions for pin diode and pnp transistor sections.

Electron density distributions are plotted along the two symmetry lines.

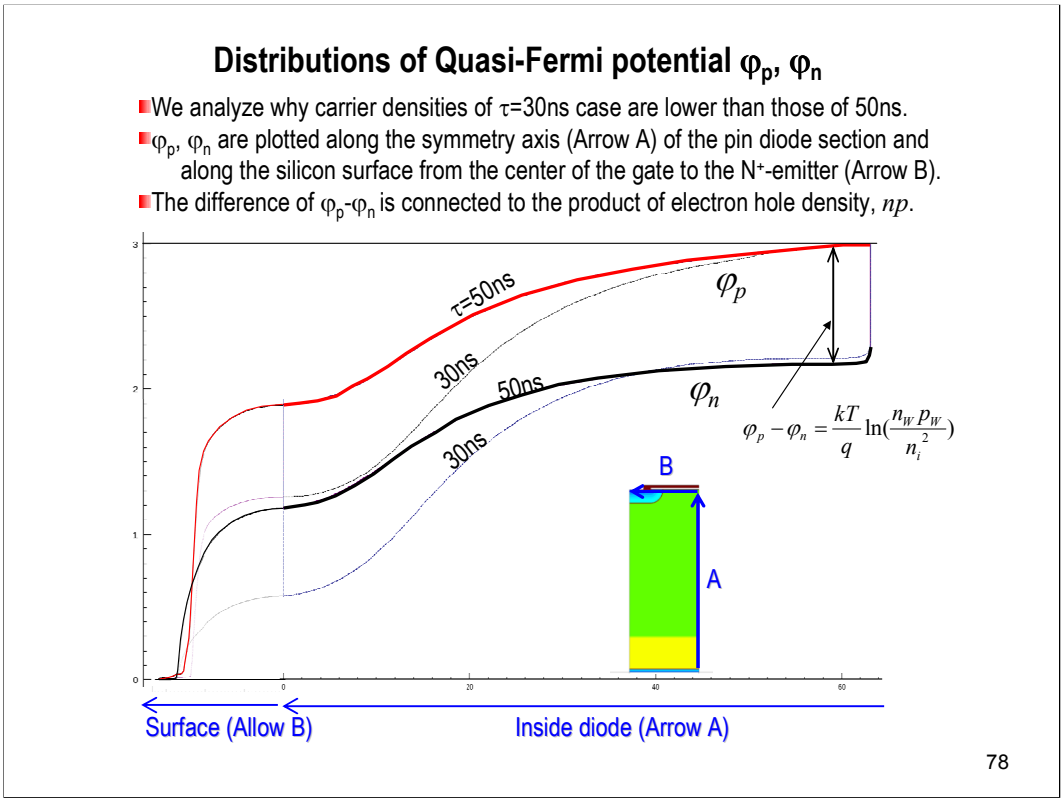
One is the line, the center of the PNP transistor, crossing the center of the P-base. The other is the center of Diode section, crossing the center of the gate electrode.

This shows the electron density distribution of the PNP transistor section. The electron density monotonically decreases from the collector toward the P-base. The decreasing rate depends on the ambipolar diffusion length.

This shows the electron density distribution of the diode section. The electron density is high at the both sides of the N-base, namely, at the N-base under the gate and at the collector-side of the N-base.

When the carrier lifetime is as low as 30nsec., the carrier density is very low in the N-base, lower than that of 50nsec case.

In the next slide, I will analyze this in more detail.



This figure shows electron and hole quasi-Fermi potentials.

We will analyze why carrier densities of τ=30ns are lower than those of 50ns.

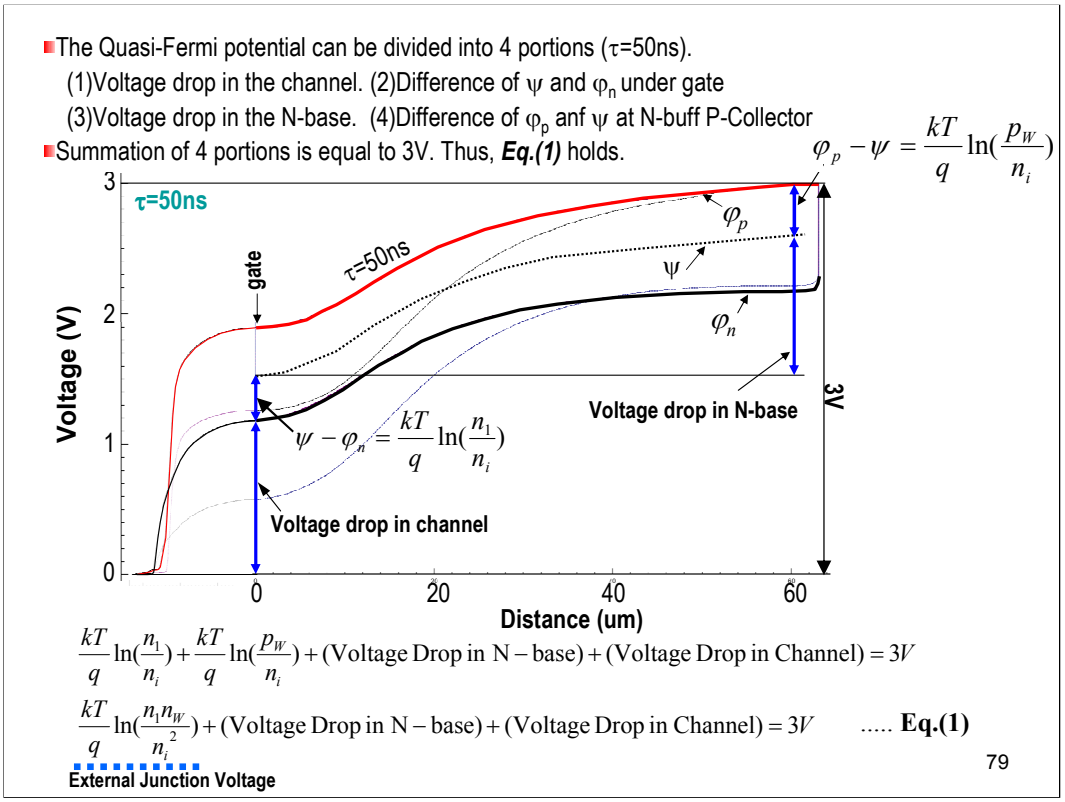
The electron and hole quasi-Fermi potentials are plotted along the symmetry line, Arrow A of the pin diode section, and along the silicon surface from the center of the gate to the N^+-emitter (Arrow B).

This portion corresponds to the Arrow A.

And this portion corresponds to the Arrow B.

This is P-collector, and this position is the center of the gate and this portion is the N emitter.

The difference of φ_p-φ_n is connected to the product of electron hole density.

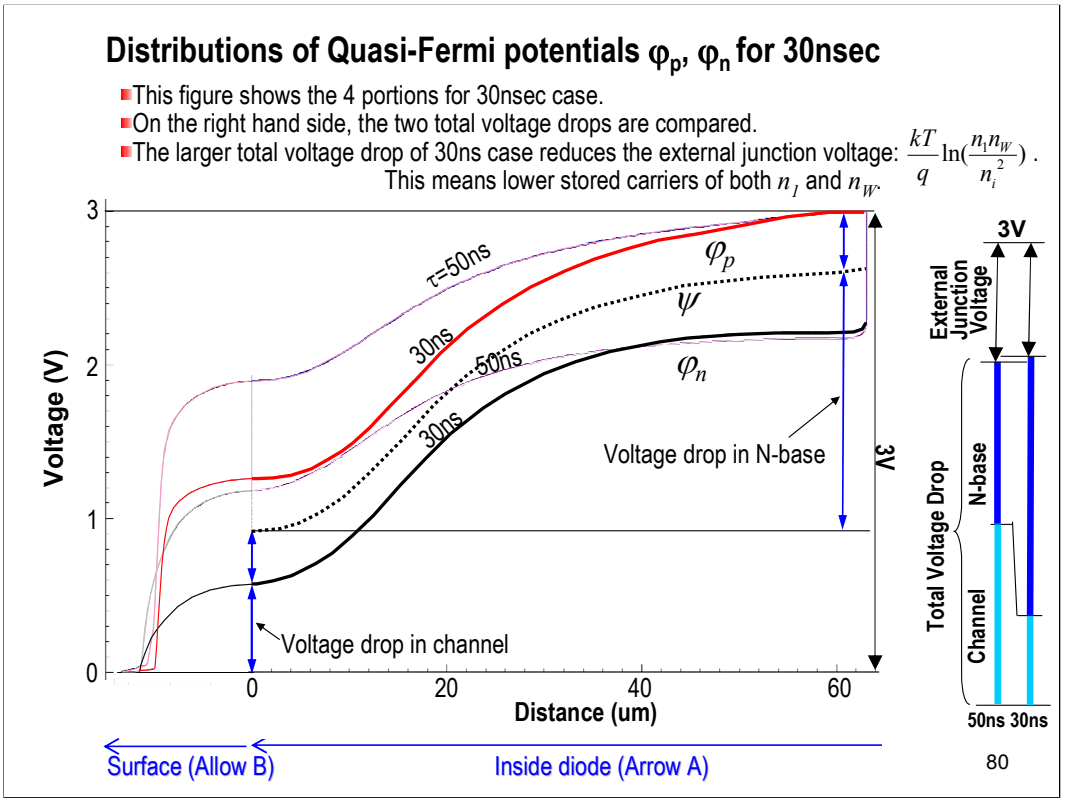


For simplicity, we disregard the effect of the N-buffer. The ψ is assumed to be in the center of the two quasi-Fermi potentials. We first treat the case of τ=50ns.

The quasi-Fermi potentials can be divided into 4 portions. First one is the voltage drop in the channel. The second is the difference between ψ and φ_n. This is related to the electron density n_1 in the surface. The 3rd one is the voltage drop in the N-base. The final one is the difference between φ_p anf ψ. This is related to hole density p_W at the collector side of the N-base.

The quasi-Fermi potential difference is 3V from N^+-emitter to P^+-collector. The first and the second terms can be merged into one. Thus, Eq.(1) holds. Here, n_1 denotes the carrier density under the gate. n_W denotes the carrier density at the collector side of the N-base. We call this term as “external junction voltage.”

The summation of 3 components is just equal to 3V.
If the voltage drop inside the device becomes large, the external junction voltage have to decrease, in other words, the product of carrier densities at both sides, n_1*n_w, have to decrease.



Here, I show quasi-Fermi potentials of τ=30ns case.

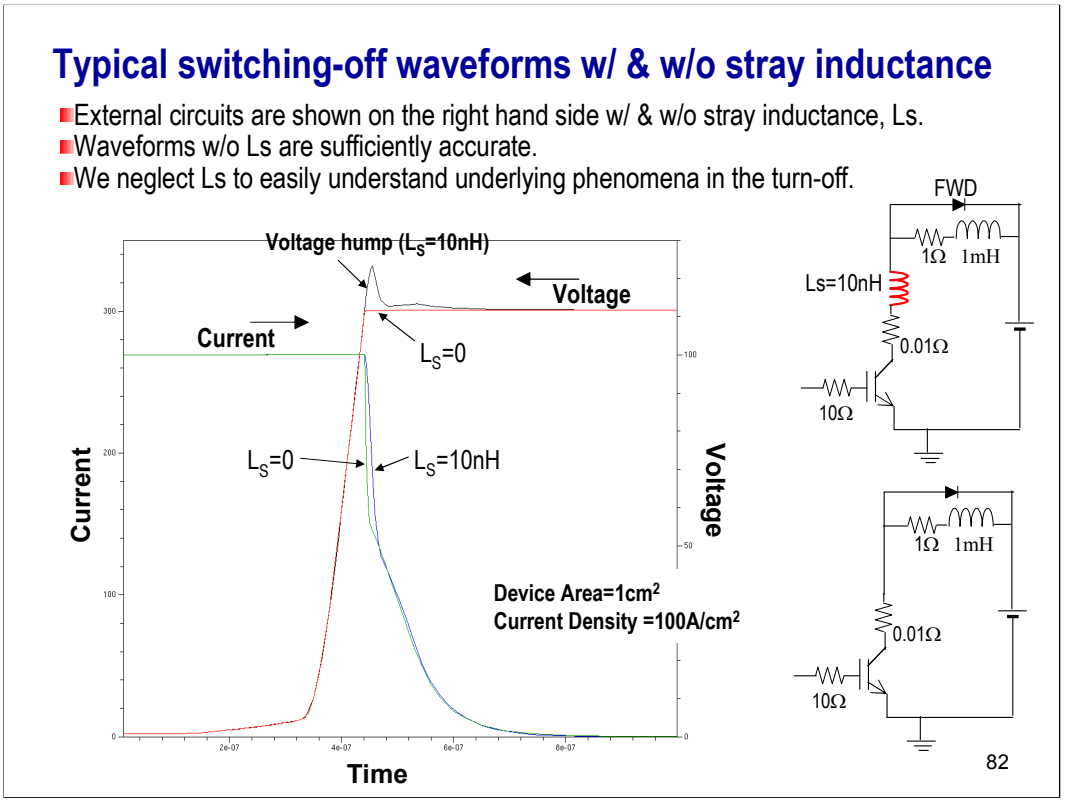
Again, the quasi-Fermi potentials are divided into the 4 components…

On the right hand side, the total voltage drops between 50ns and 30ns cases are compared. The total voltage drop of 30ns case is slightly larger than that of 50ns case.

Thus, the junction voltage of 30ns case is lower than that of 50ns case. This reduces the carrier densities, n1 and nw.

Namely, the carrier density, n1, under the gate and the carrier density, nw, at the collector-side of the N-base are smaller than those of 50ns case.

Switching characteristics of IGBTs



From now on, I would like to talk about switching of IGBTs.

This figure shows typical switching-off waveforms with and without Ls. The used circuit is shown here. The main load is 1mH inductance.

In an actual circuit, a stray inductance of, for example, 10nH exists. The figure compares the calculated waveforms with and without 10nH stray inductance.

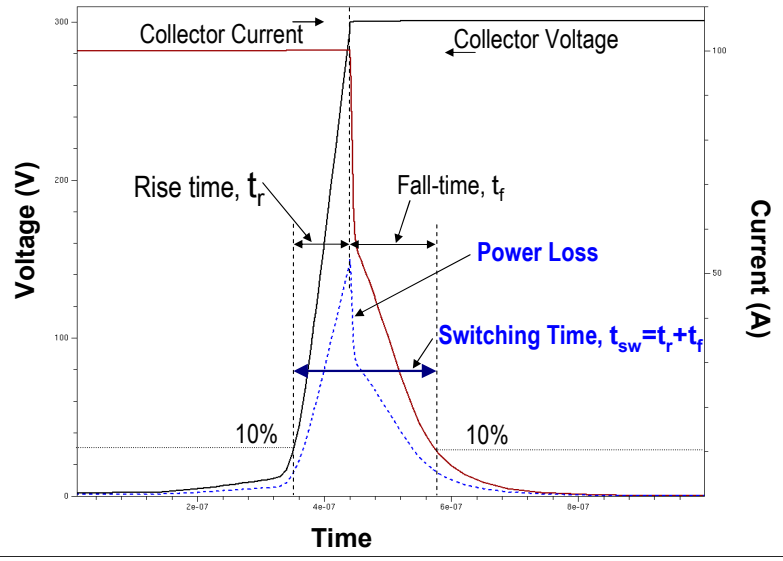
Sufficiently accurate switching waveforms still can be obtained even if we neglect the stray inductance.

The two waveforms are similar, although a voltage hump appears if Ls is included.

We neglect the stray inductance in order to easily understand the underlying phenomena in the turn-off transient.

Definition of Switching Time, t_{sw}

- Power loss becomes large when the collector voltage exceeds 10% of the applied voltage.
- Power loss is large until current decreases below 10% of the initial current.
- Switching time is defined as the time from 10% of the applied voltage to 10% of the initial current



83

This slide shows the definition of switching time.

This is typical switching waveform of an IGBT.
This is collector current and this is collector voltage.

The dotted line shows the power loss.

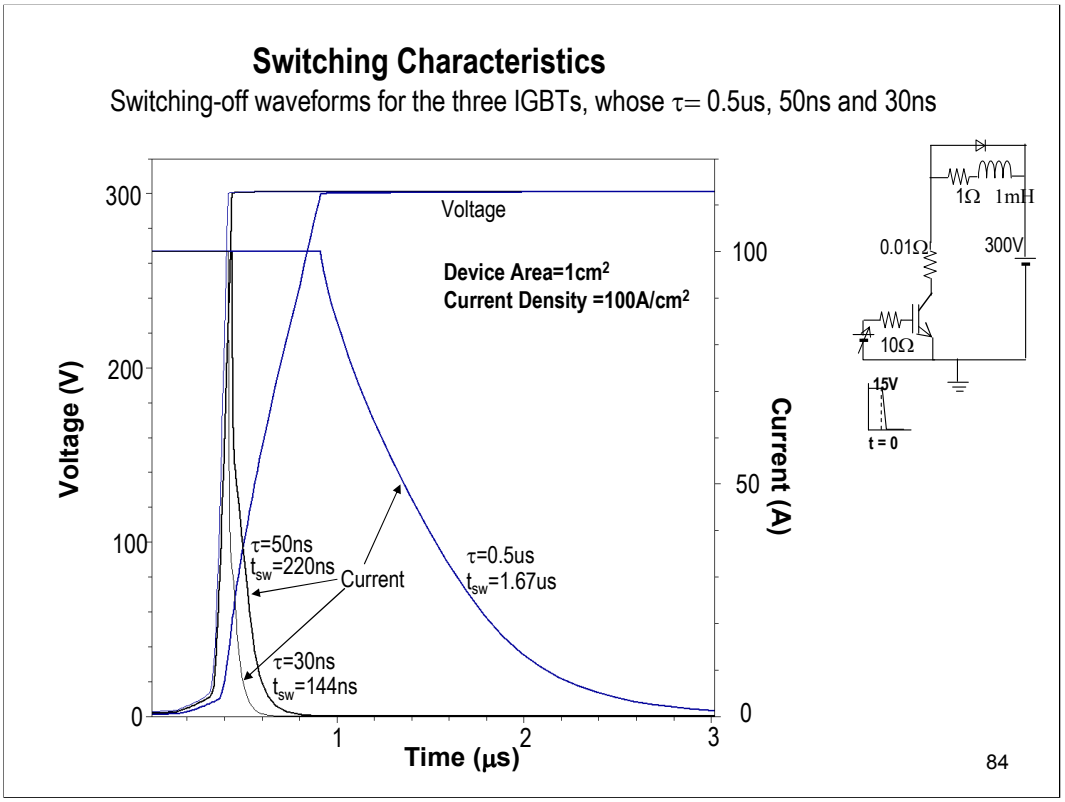
The power loss becomes large when the voltage exceeds 10% of the applied voltage. The power loss is large untill the collector current becomes less than 10% of the initial value.

Thus, the switching time is defined, in this text, as the time from 10% of the applied voltage to 10% of the initial current as shown in the figure.

Usually, voltage rise time is defined as the time from 10% of the applied voltage to 90% of the applied voltage.

Similarly, current fall time is defined as the time from 90% of the total current to 10% of the total current.

The switching time in this talk is the summation of tr and tf.



This figure shows three switching-off waveforms of the three IGBTs. The carrier lifetimes are 0.5us, 50ns and 30ns.

The used circuit is shown here. The main load is 1mH inductance. The internal resistance of the inductor is 1Ω. A free wheeling diode is connected between the two terminals of the inductance. A gate resistance, 10Ω, is used.

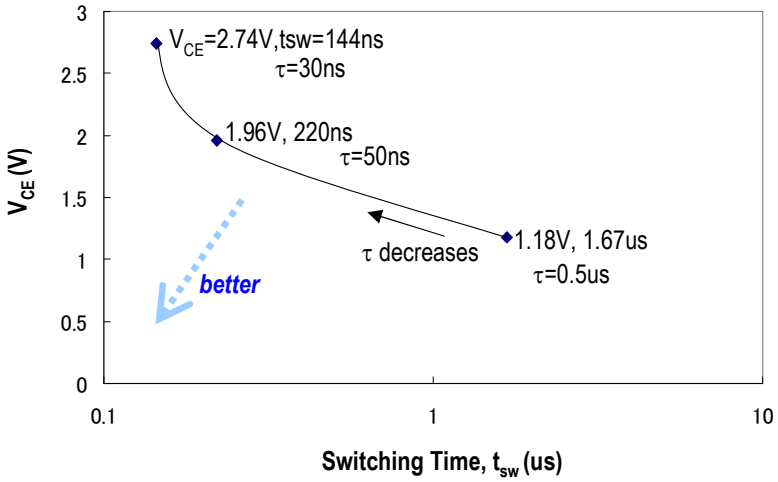
The device total area is 1cm^2, and the current density is 100A/cm^2.

Thus, the load current of 100A is turned off.

In the next, we will analyze these waveforms.

Trade-off relation between V_{CE} and switching time

- As carrier lifetime decreases, switching time improves but V_{CE} increases.
- It is difficult to simultaneously improve both. This relation is called "**trade-off**."
- In order to improve the trade-off, usually we need new techniques or new devices.



85

This shows trade-off relation between collector-emitter voltage, V_{CE}, and switching time for planar IGBTs.

If we decrease the carrier lifetime , the switching time decreases, but, V_{CE} increases.

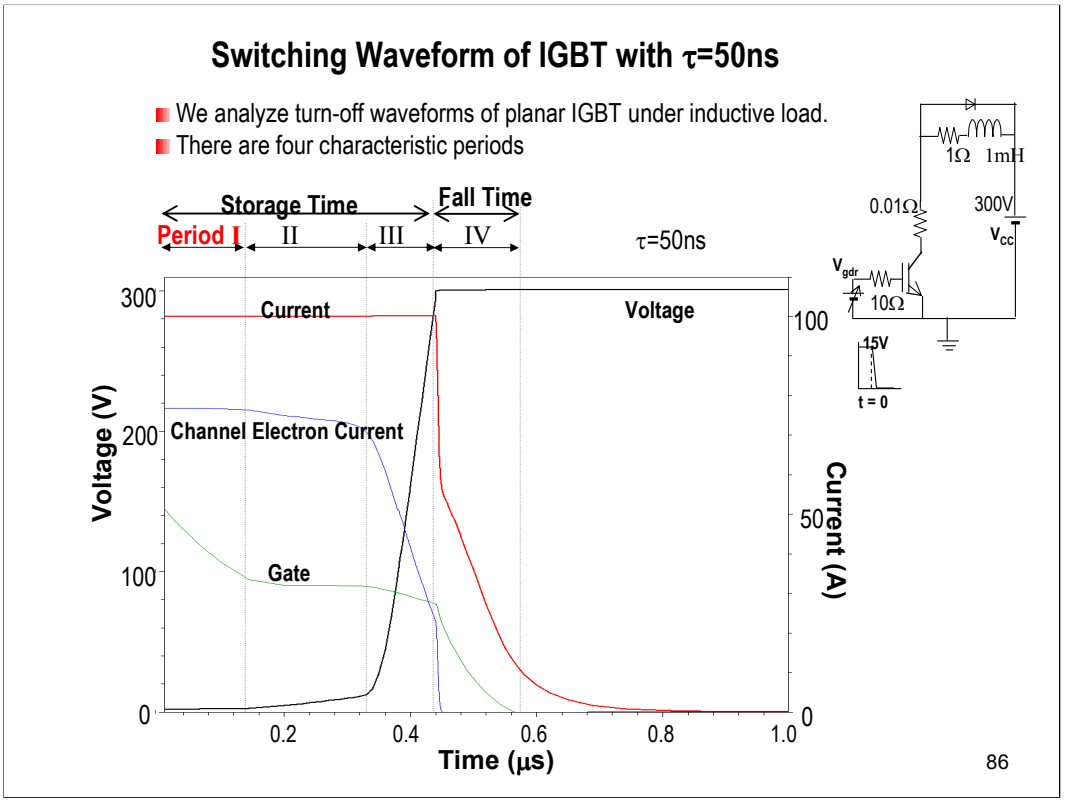
For example, ……..

Generally, it is quite difficult to simultaneously improve switching time and V_{CE}.

This relation is called trade-off.

If we can shift the curve toward the lower left corner, then, we can say that the trade-off is improved.

In order to improve the trade-off relation, usually we need to develop new technique, or new devices.



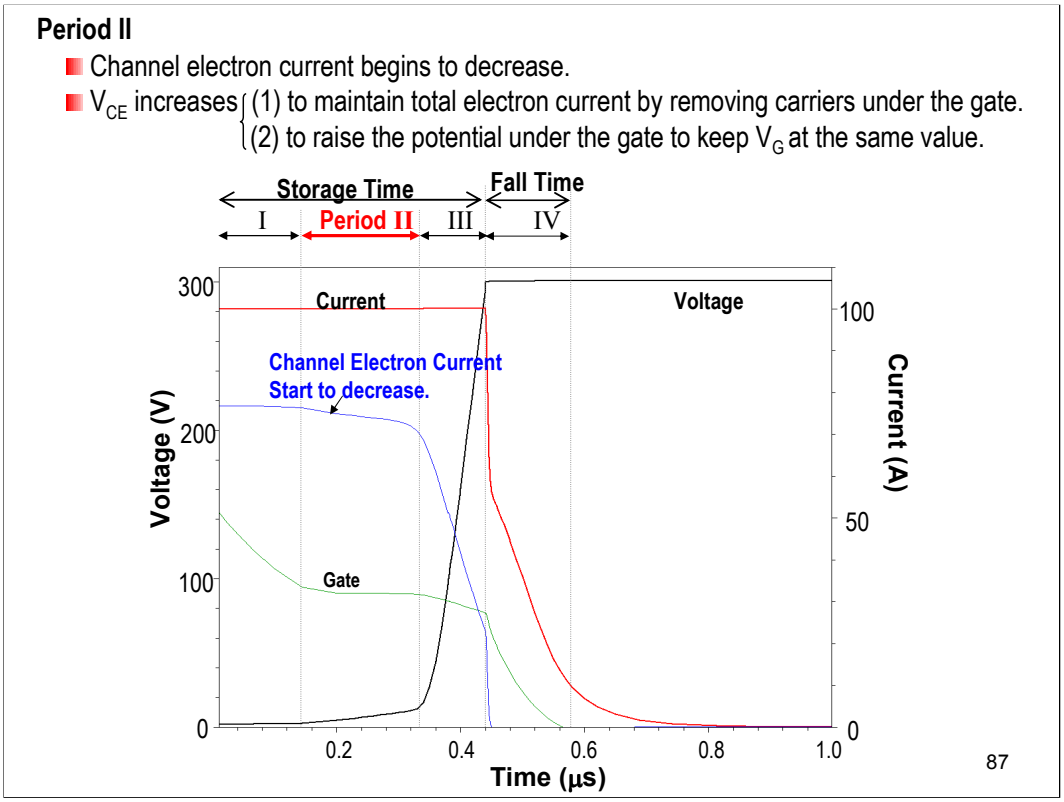
This figure shows the switching waveform of IGBT.

The carrier lifetime of IGBT is 50nsec.

We will analyze the switching-off waveforms of planar IGBT under the inductive load.

There are four characteristic periods.

In the period I, the gate voltage decreases, but, almost the same channel current flows and no manifest change occurs.



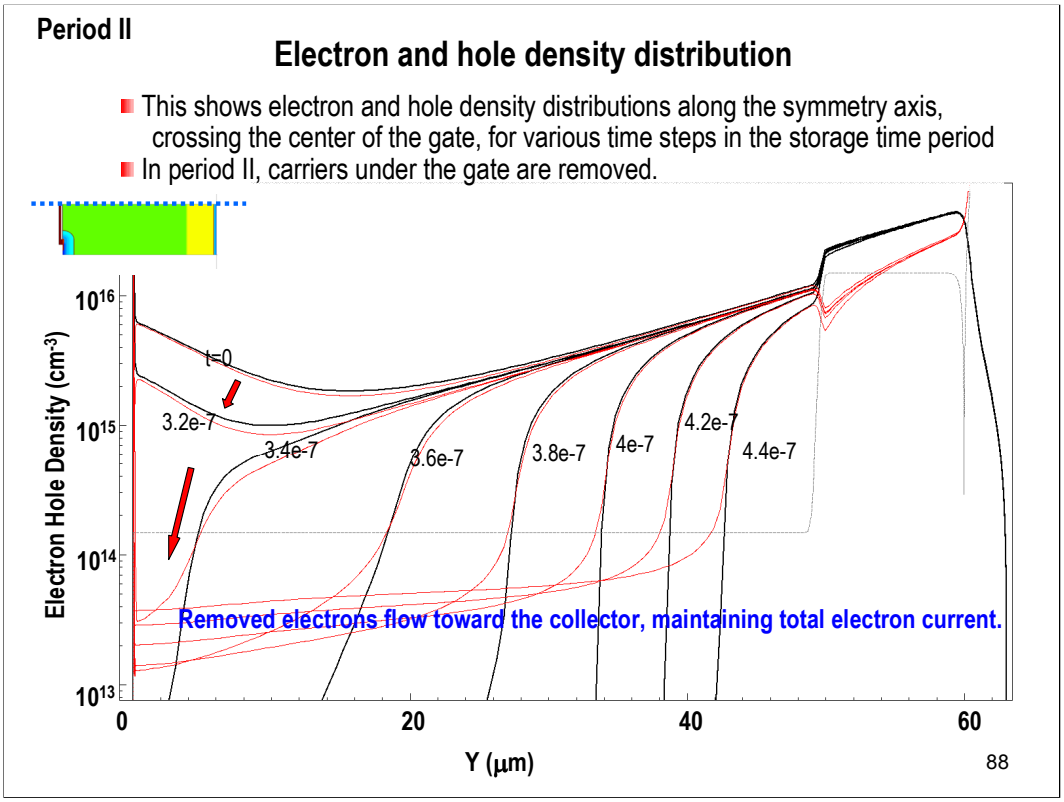
In the period II, the channel electron current begins to decrease.

V_{CE} increases,

(1) In order to maintain total electron current by removing the stored carriers under the gate.

(2) to raise the potential under the gate so as to keep the gate voltage at the same value.

We will interpret this in the next slide.

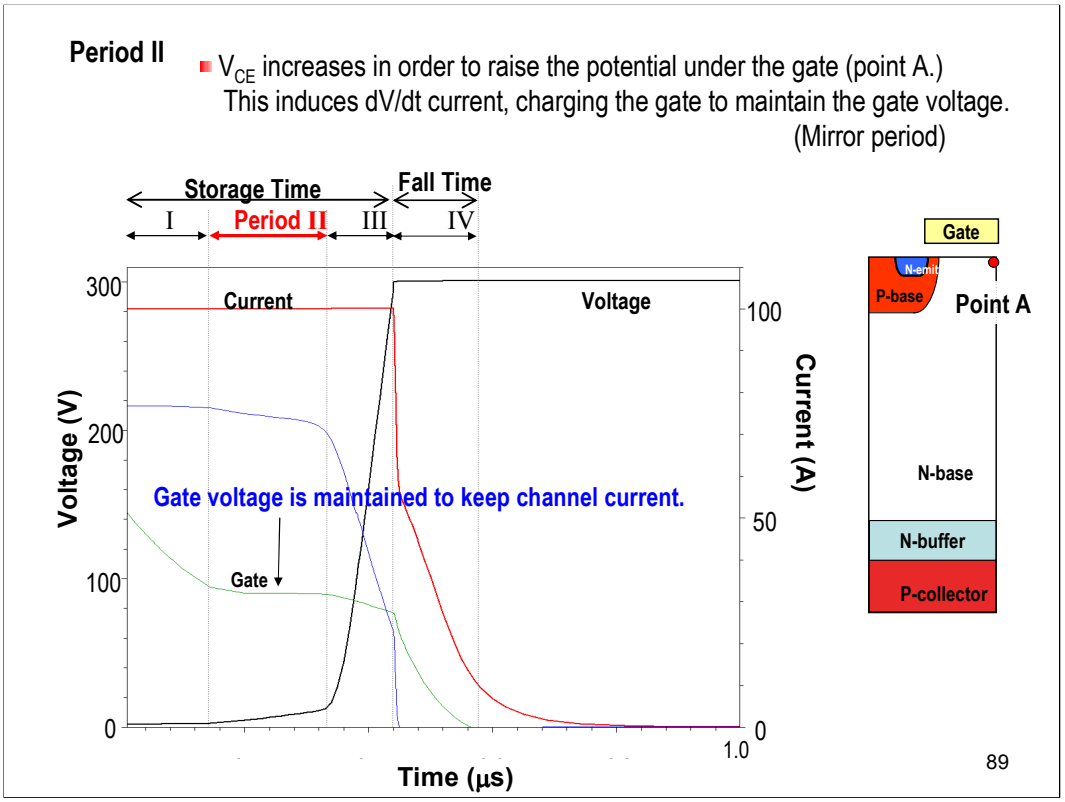


This shows electron and hole density distributions along the symmetry axis, the center of pin diode section.

The numbers show time steps in the storage time period.

In the period II,
carriers under the gate are removed, and a depletion layer appears under the gate.

Removed electrons flow toward the collector, and create electron current. The total electron current is maintained at the same value, although the channel electron current decreases.

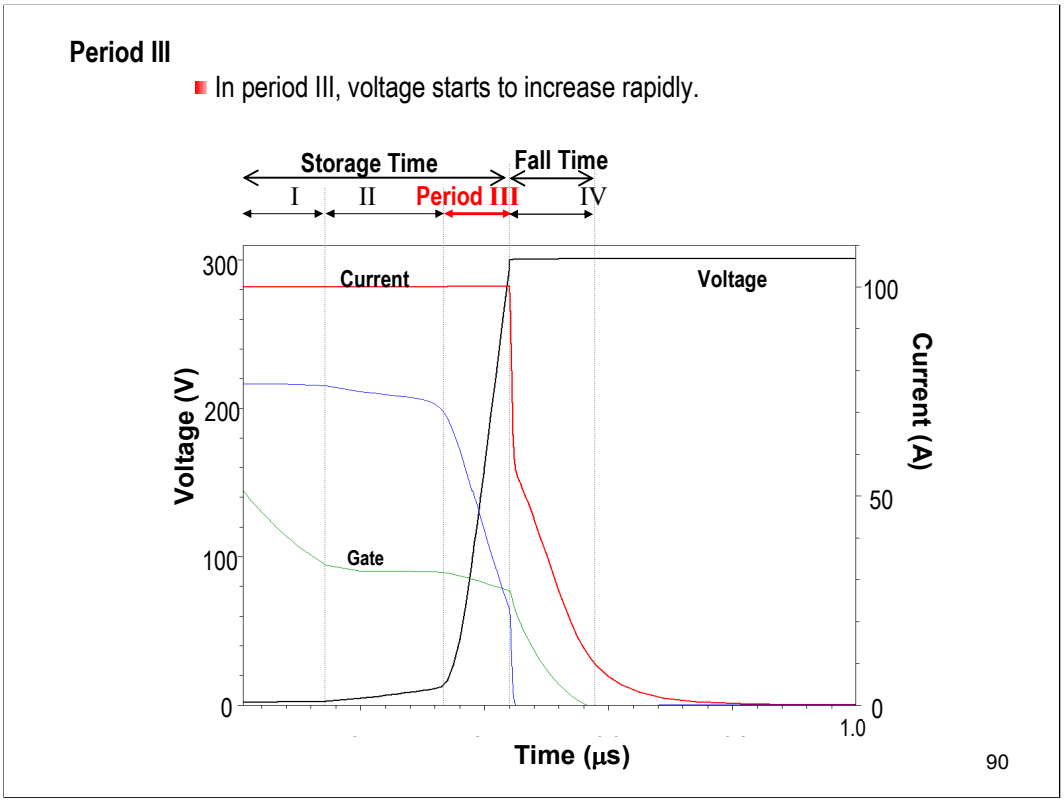


In the period II

The collector-emitter voltage, V_{CE}, increases in order to raise the potential under the gate oxide, for example, the potential at Point A.

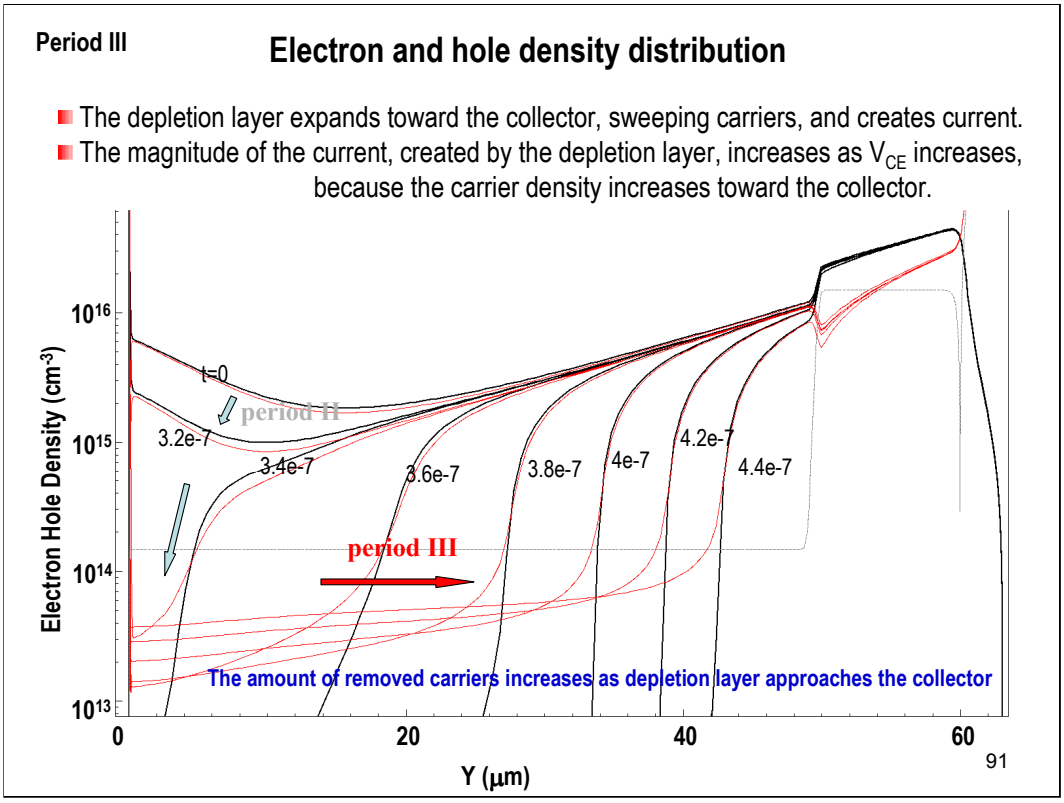
The potential increase at the point A induces dV/dt current though the gate oxide, charging the gate to maintain the gate voltage.

The gate voltage remains almost at the same value to keep the channel electron current.



In the period III

The collector-emitter voltage, V_{CE}, start to increase rapidly.

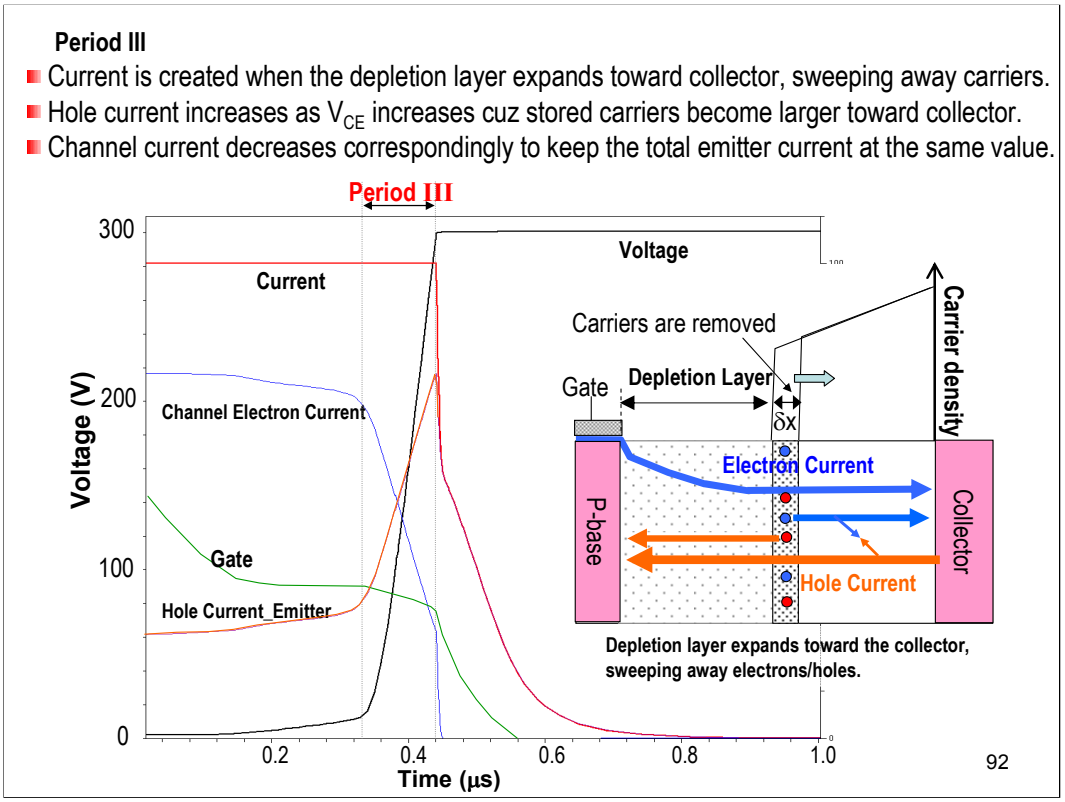


This shows electron and hole density distributions, again.

The numbers show the time step.

In the period III,
The depletion layer develops and expands toward the collector, sweeping away carriers and creates current.

The magnitude of the current created by the depletion layer expansion increases as V_{CE} increases because the stored carrier density increases toward the collector.



In the period III,

Current is created when the depletion layer expands toward the collector, sweeping away carriers in the front edge of the depletion layer.

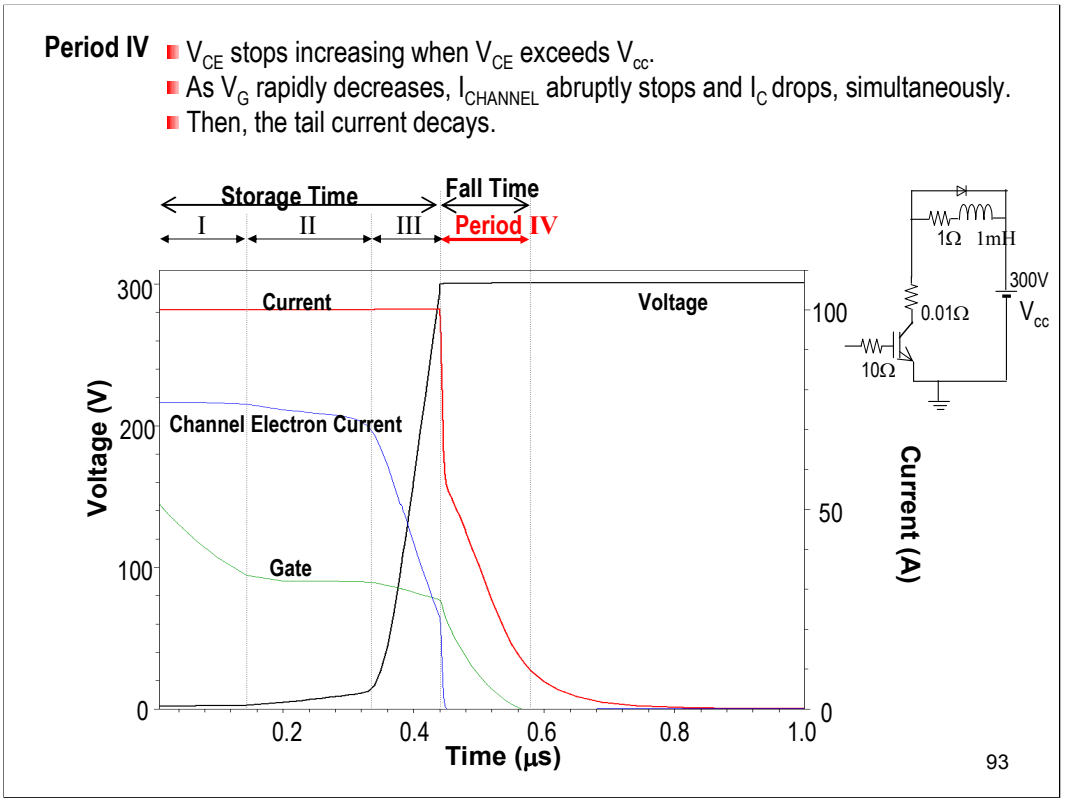
This is depletion layer. The carriers of this portion are removed and creates current.

The removed holes move toward emitter and become hole current flowing into the emitter.

The hole current increases as the collector voltage increases because the stored carrier density becomes larger toward the collector.

The summation of the channel electron current and the hole current makes the total current. The channel current decrease correspondingly to keep the total emitter current at the same value, because the hole current increases.

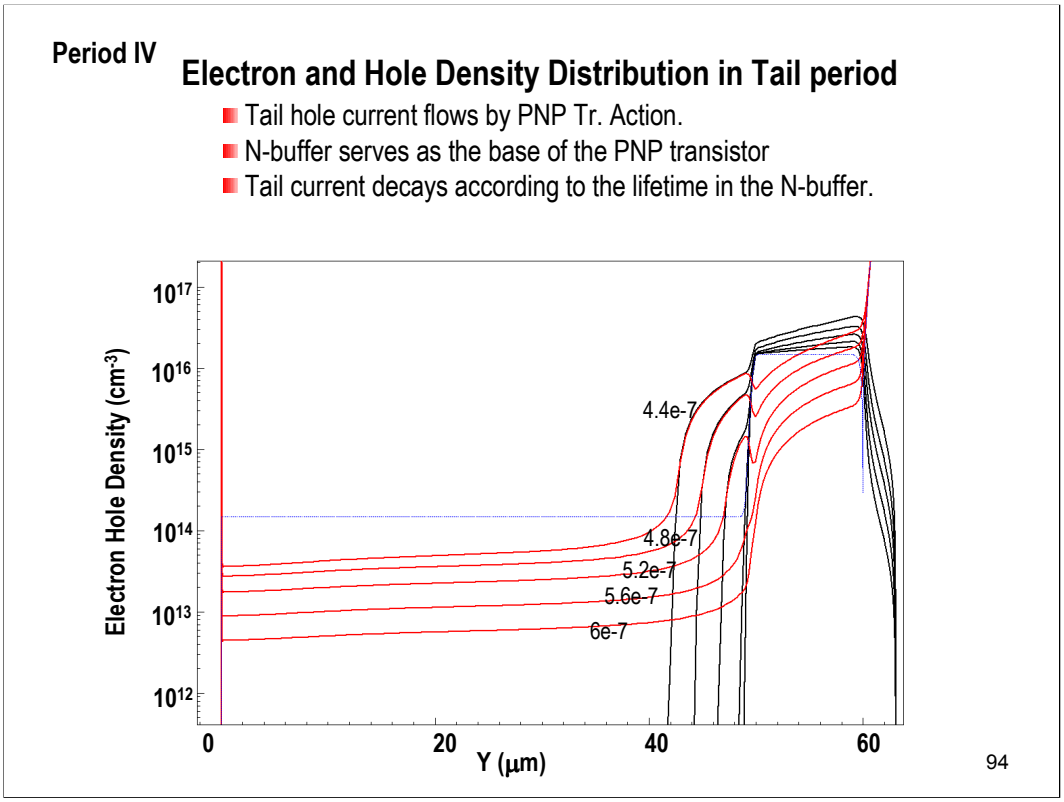
Thus, the gate voltage decreases in the period III.



In the period IV, the collector voltage stops increasing when the voltage exceeds 300V, the voltage of external voltage source, Vcc.

The gate voltage rapidly decreases and the channel electron current stops abruptly. This induces the rapid decrease in the collector current in the beginning of the period IV.

After that, the tail current flows.

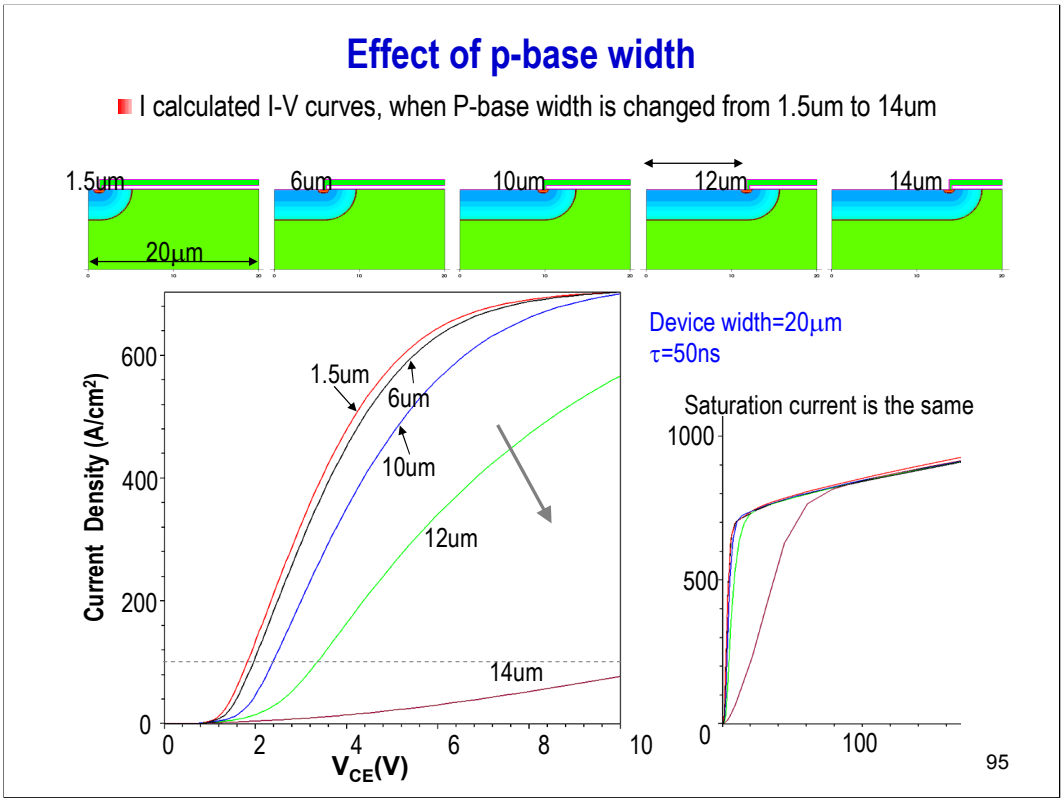


This shows the hole and electron density distribution in the tail period.

The tail current flows by the PNP transistor action.

The N-buffer serves as the base of the PNP transistor and the depletion layer acts as the collector of the PNP transistor.

The tail current decays according to the effective carrier lifetime in the N-buffer.



In this slide, I calculated I-V curves of various planar IGBTs, when P-base width is changed from 1.5um to 14um.

The carrier lifetime is 50ns.

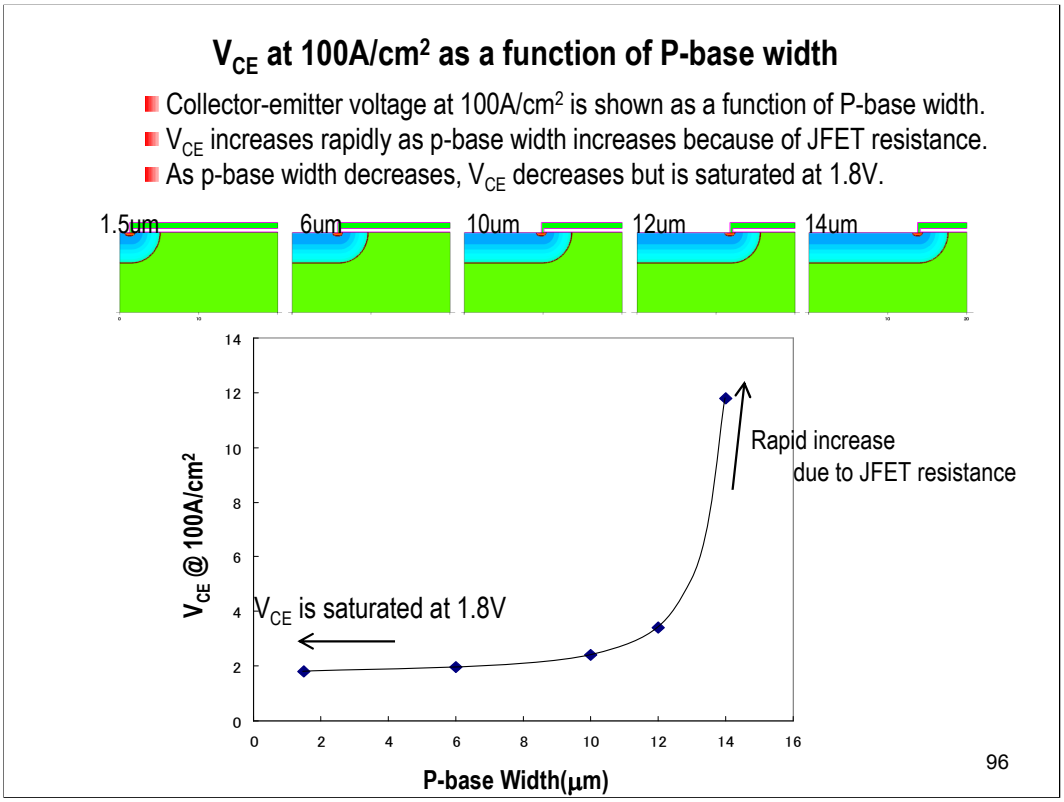
The device width is kept 20um so that the poly-silicon gate width is changed accordingly.

The total channel width is kept at the same value, because the device width is the same.

Thus, the saturation current for a high applied voltage is not changed.

The device characteristics changes very much, as shown in the figure.

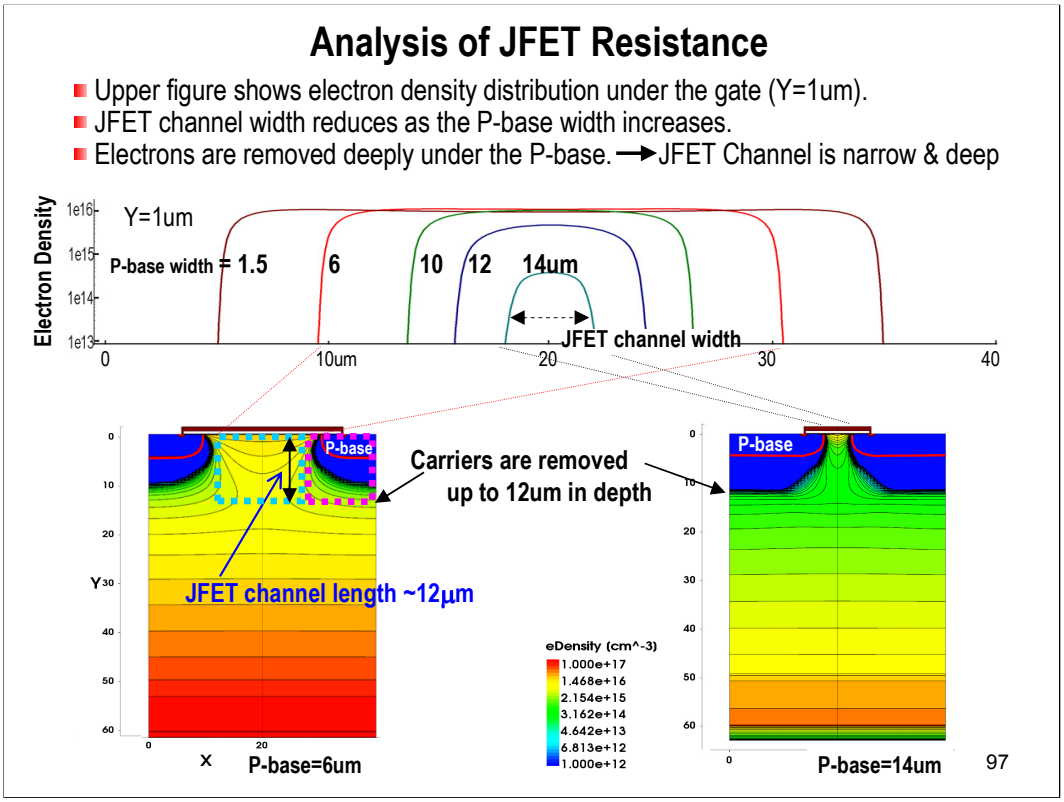
Especially, the V_{CE} (collector emitter voltage) increases considerably, when the p-base width is 14um.



This figure shows V_{CE} at 100A/cm^2 as a function of P-base width.

V_{CE} rapidly increases as the P-base width becomes more than 12um because of JFET resistance.

On the other hand, V_{CE} decreases but is saturated at the value of 1.8V as the P-well width decreases.



We will analyze JFET resistance.

The upper figure shows the electron density distribution under the gate on the line of Y=1um.

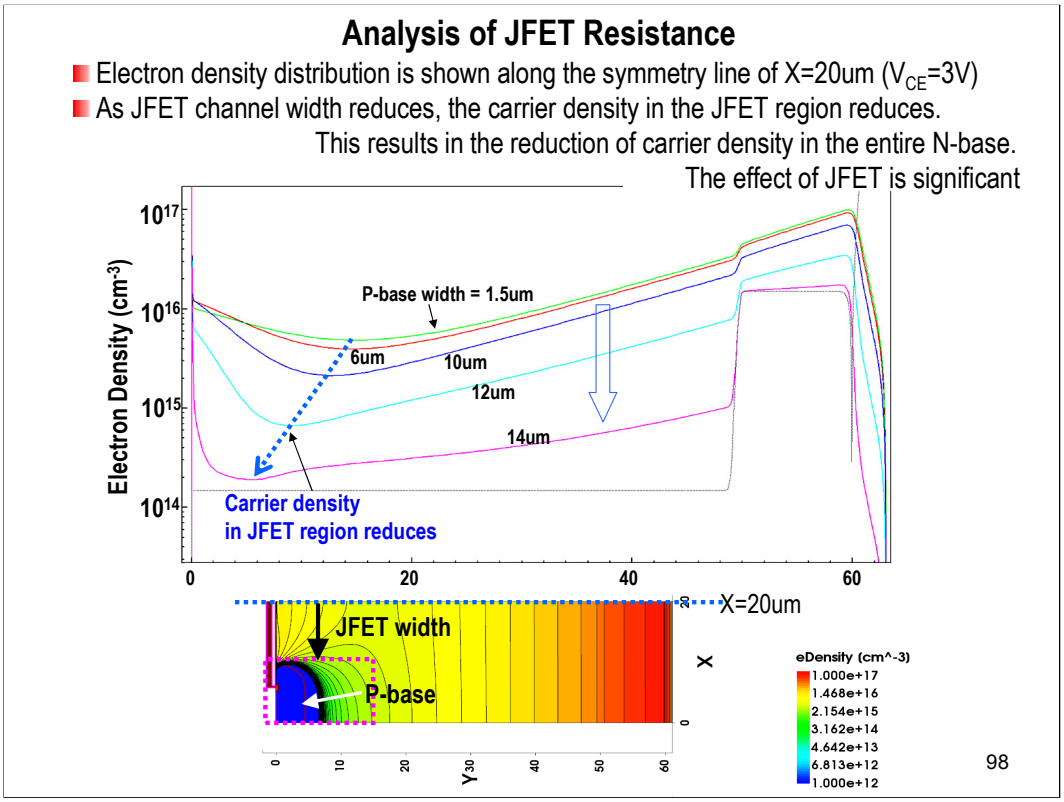
JFET channel width reduces as the P-base width increases.

Moreover, the electrons are removed deeply under the P-base.

The p-base depth is 4.5um as shown by the red line. However, the electron density is decreased up to 12um in depth.

The effective JFET channel length is approximately 12um, from the surface to 12um in depth, and is longer than the depth of P-base diffusion.

Thus, the JFET channel is narrow and long.

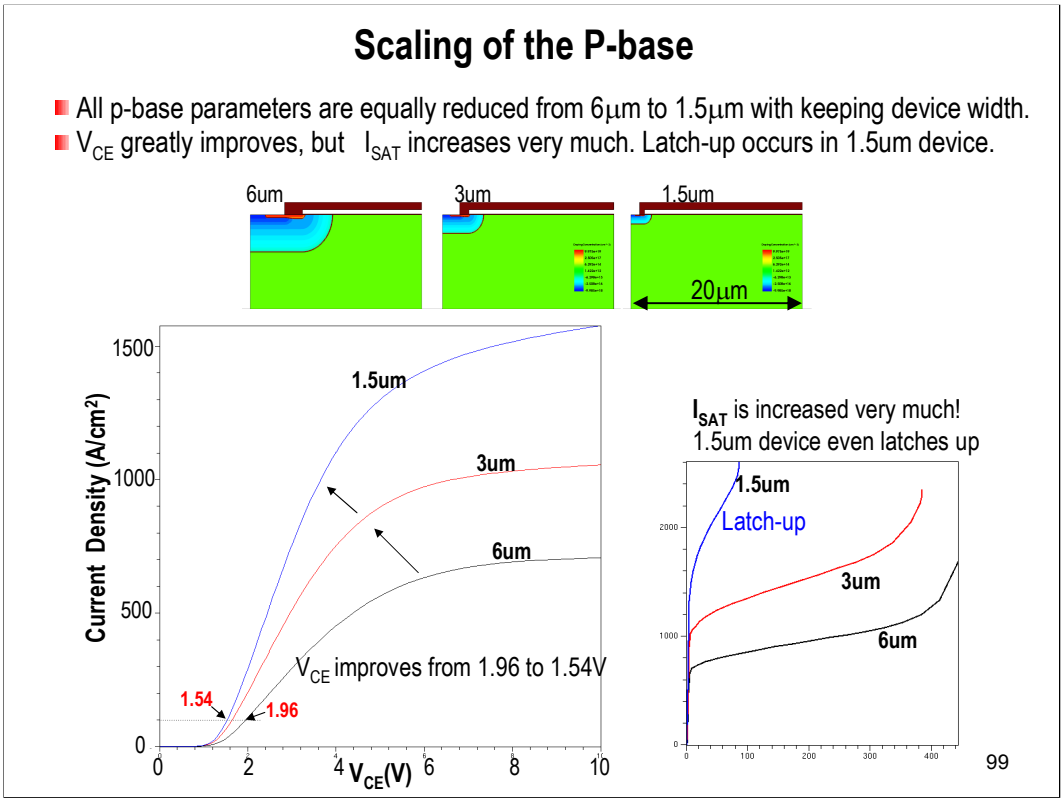


This figure shows the electron density distribution along the line of X=20um, the center of the gate.

As the JFET channel width reduces, the carrier density in the JFET region is reduced.
This results in the reduction of carrier density in the entire N-base.

As for IGBT of 14um P-base, the electron density in the N–base is lower than 10^{15} throughout the N-base.

The effect of JFET is significant.



In the previous slides, we reduce only P-base width with keeping all the other parameters.

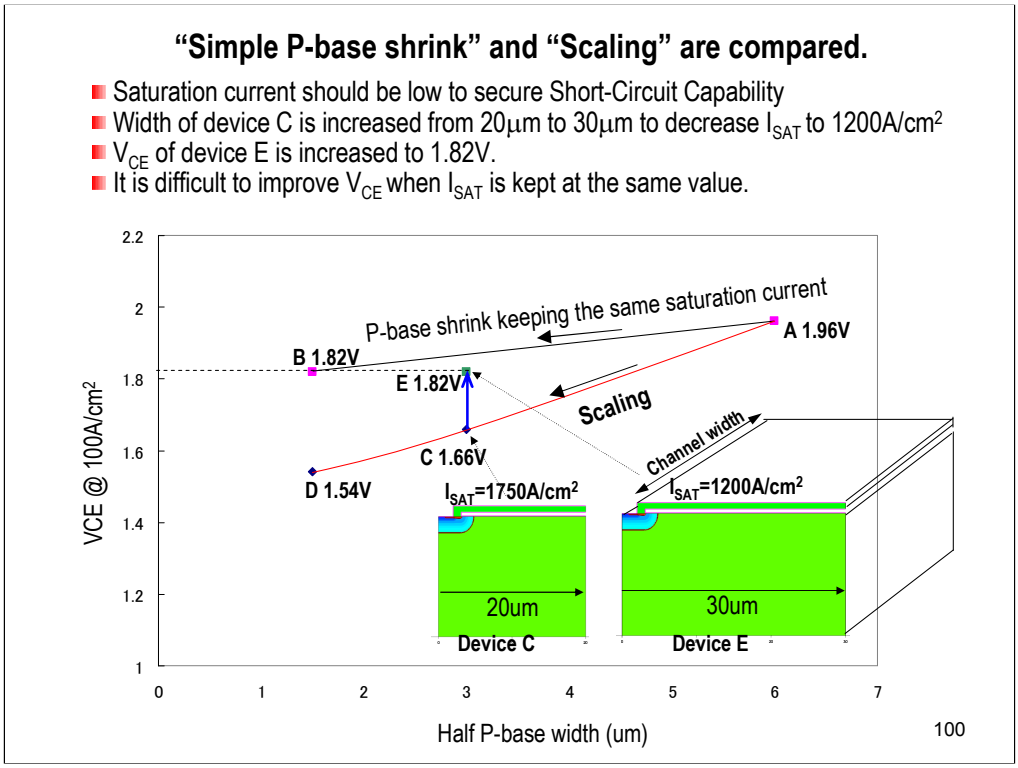
Here, all the P-base parameters are equally reduced from 6um to 1.5um with keeping the same device width of 20um.

If we scale down the P-base from 6um to 1.5um, V_{CE} improves from 1.96V to 1.54V at 100A/cm^2 current density.

The saturation current density also increases very much, if we scale down the P-base because the channel length reduces.

The 1.5um device even latches up at high current density.

We need to reduce saturation current.



This figure shows V_{CE} vs. P-base width.

The upper line A-B shows the change when we shrink the P-base width with keeping all the other parameters at the same value.

The lower line ACD shows the change when we scale the P-base.

The device C shows very high saturation current density of 1750A/cm^2.

The saturation current should be low enough to secure short-circuit capability.

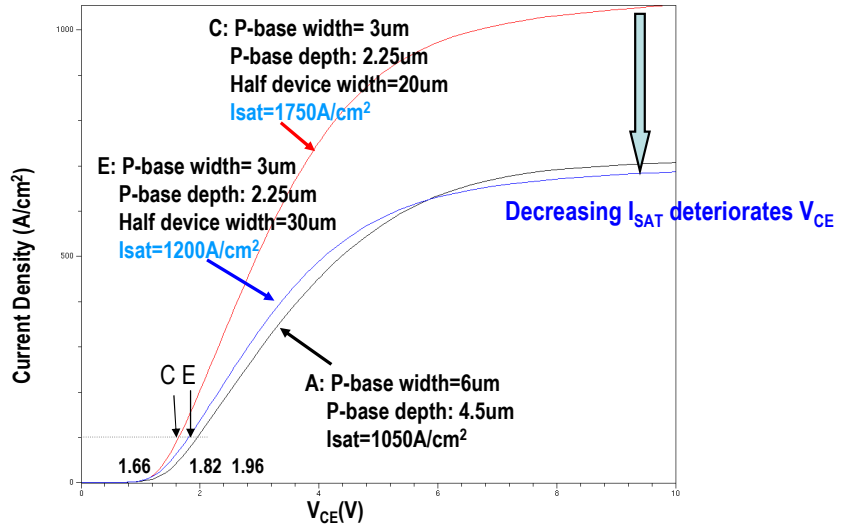
The device width of device C is increased from 20um to 30um in order to decrease the saturation current to 1200A/cm^2 by reducing the total channel width for unit device area of 1cm^2.

We name the new device E. But, the V_{CE} of device E increases to 1.82V, which is the same as device B,

It is very difficult to improve V_{CE} with keeping the same saturation current.

Decreasing I_{sat} deteriorates V_{CE}

- Device A is the basic device: P-base width: 6um, depth: 4.5um. Isat=1050A/cm^2
- Device **C** is scaled device: P-base width: 3um, depth: 2.25um. Isat= 1750A/cm2
- Device **E** is the same as device C, but the width of the device is increased to 30um. Isat=1200A/cm^2
- Decreasing the saturation current deteriorates V_{CE}. **Isat** is one of the important design parameters.



101

This figure shows the calculated I-V curves of device A , C, and E.

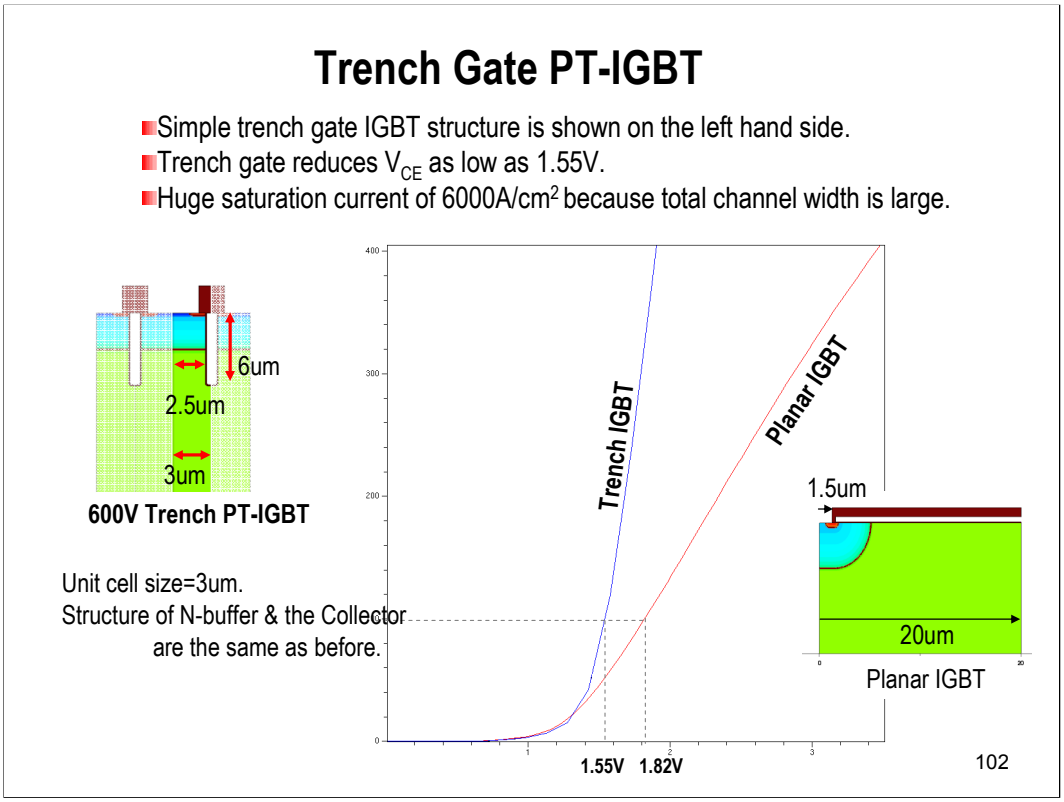
Device A is the basic device structure: P-base width of 6um and the depth of 4.5um. Isat = 1050A/cm2.

The device C is a scaled device: P-base width of 3um and the depth of 2.25um. Isat is 1750A/cm^2.

The device E is a scaled device: P-base width of 3um and the depth of 2.25um, but the half width of the device is increased to 30um. Isat = 1200A/cm2.

If we decrease the saturation current, this deteriorates V_{CE}.

Saturation current density is one of the important design parameters, which decisively influence device characteristics.



From now, I'd like to talk about trench gate IGBTs. Simple trench gate IGBT structure is shown on the left hand side.

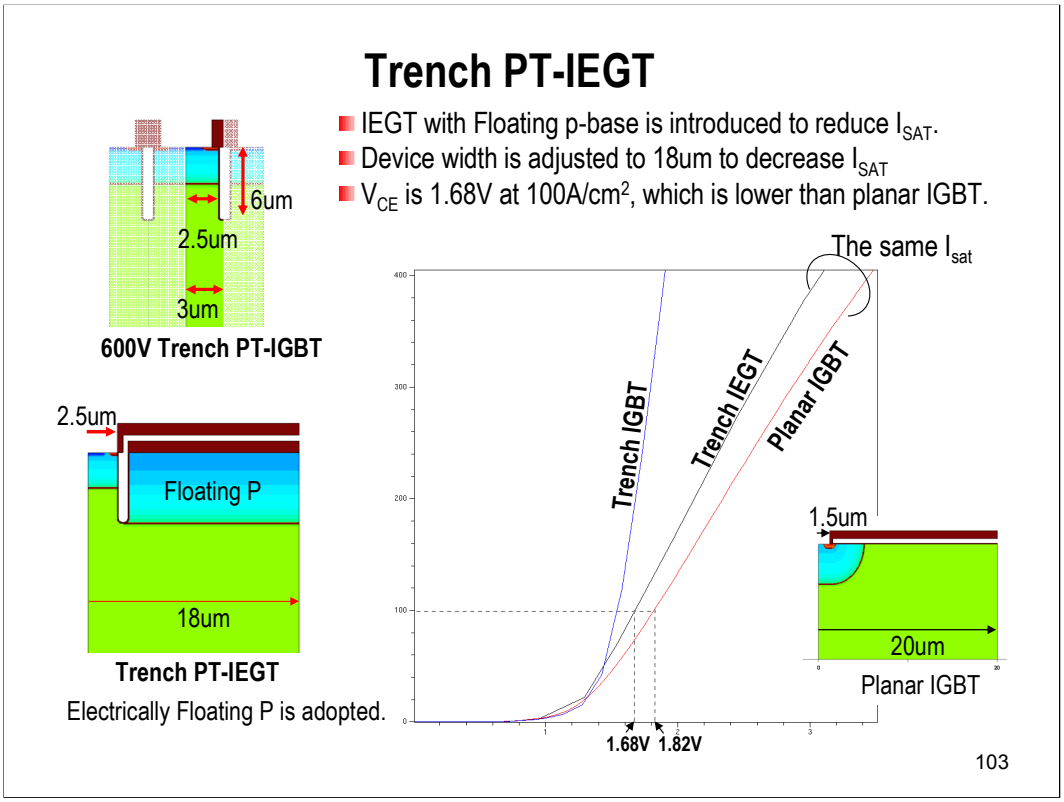
The mesa width is 5um. The trench depth is 6um. The unit cell size or the device width is 6um.

The N-buffer and the collector structure is the same as the planar IGBTs.

The collector emitter voltage at 100A/cm^2 is very good, as low as 1.55V.

However, the saturation current density of Trench IGBT is huge as high as 6000A/cm^2.

This is because the total channel width for 1cm^2 is very large, compared with planar IGBTs because the unit cell size is only 6um.



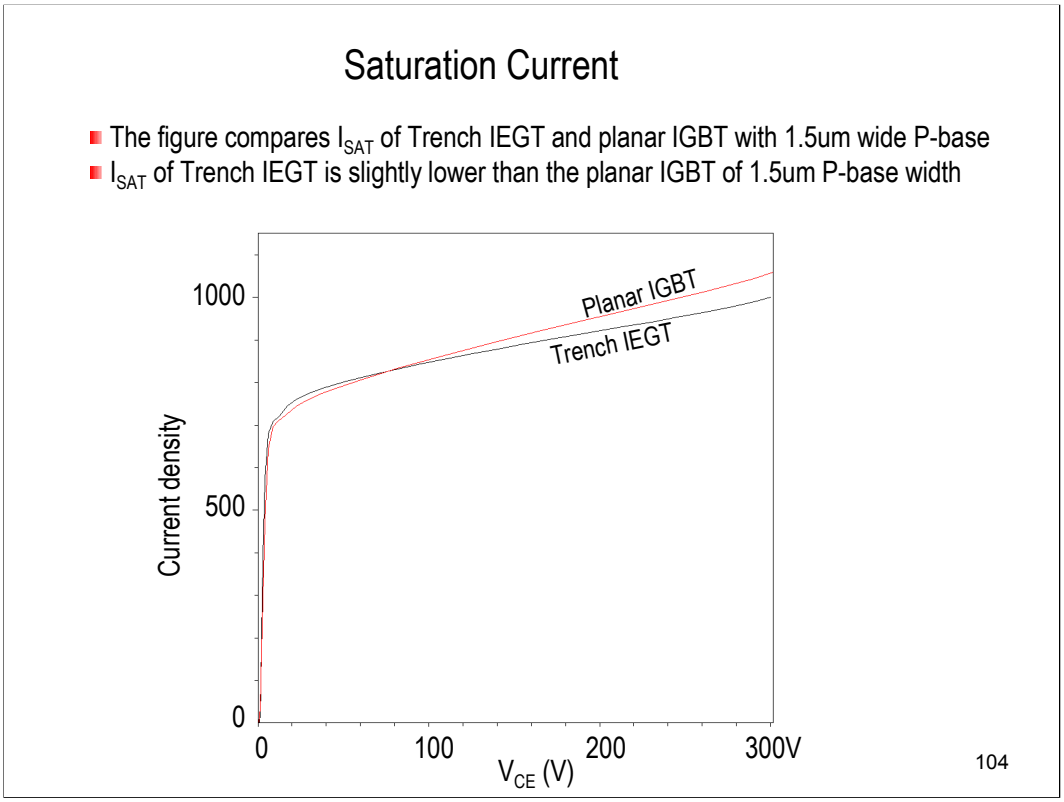
In order to reduce the saturation current to the level of planar IGBTs, we introduce IEGT structure.

The IEGT structure, we adopt, is shown at the lower left.

We introduces a deep P-type floating layer between the trench gates.

The device width is adjusted to 18um to decrease I_{SAT}.

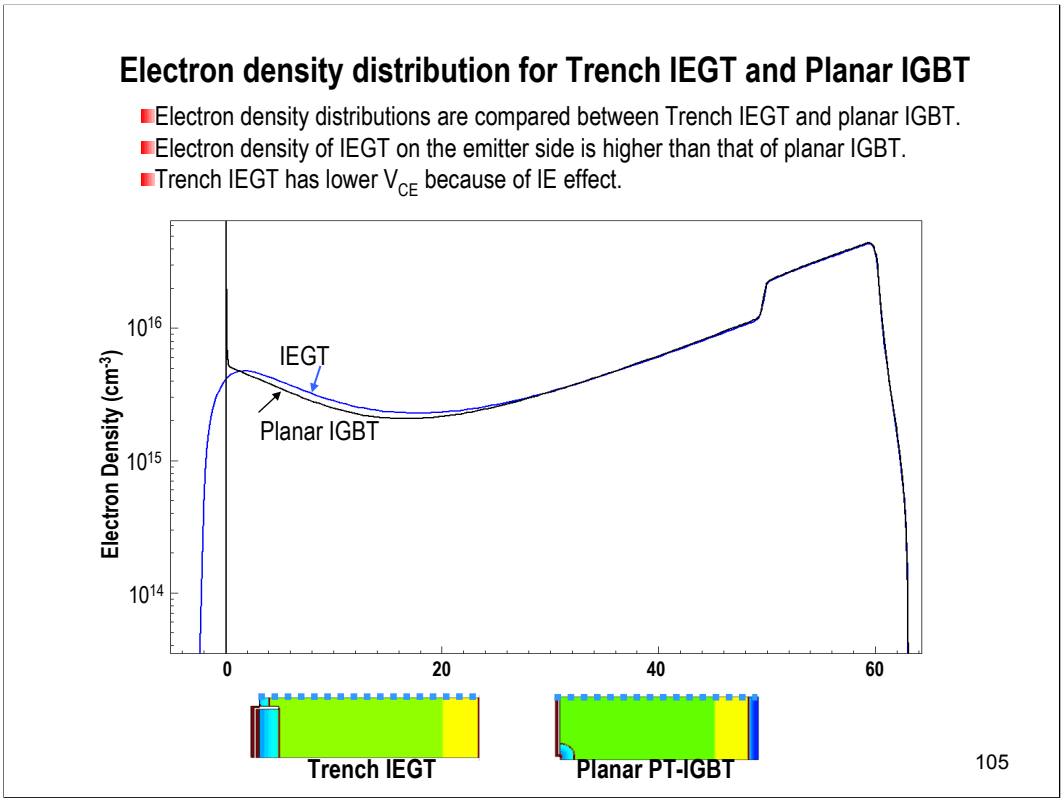
The calculated V_{CE} is 1.68V at 100A/cm^2, which is lower than planar IGBT.



The figure compares the saturation current of Trench IEGT and planar IGBT with 1.5um wide P-base.

The saturation current density of Trench IEGT is slightly lower than the planar IGBT of 1.5um P-base width.

The saturation currents of the two devices are almost the same.



Electron density distributions are compared between trench IGBT and IEGT.

Electron density of IEGT on the emitter side is higher than that of planar IGBT.

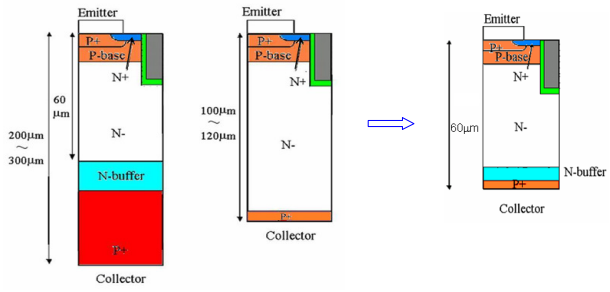
This is because the electron density of the trench IEGT is increased by IE effect.

Moreover, in planar IGBT, the holes on the emitter-side are reduced because the P-base extracts carriers.

This is the reason why V_{CE} of trench IEGT is lower than that of planar IGBT.

Evolution of IGBT Structure

- In late 90's, NPT-IGBT structure without N-buffer were widely utilized.
- NPT IGBT was not a final solution.
- Recently, FS-IGBT was developed and n-buffer was utilized again to reduce V_{CE}.
- FS-IGBT is the combination of NPT-IGBT and N-buffer.
- So, we skip NPT-IGBT, and talk about FS-IGBT in the next slide.



PT-IGBT NPT-IGBT FS-IGBT

106

So far, I talked about PT-IGBT, using epitaxial layer.

In late 90's, NPT IGBT structure without N-buffer were widely utilized.

However, NPT IGBT was not a final solution.

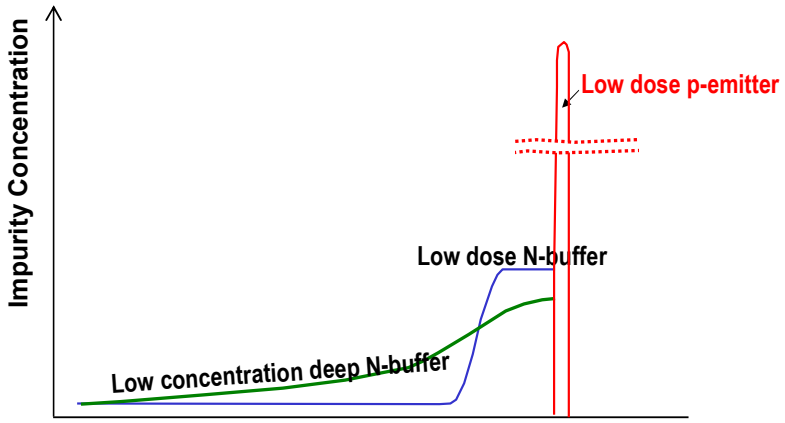
Recently, FS-IGBT was developed and n-buffer was utilized again to reduce the V_{CE}.

FS-IGBT is the combination of NT-IGBT and N-buffer.

So, we skip NPT-IGBT, and talk about FS-IGBT in the next slide.

Field Stop IGBT (Thin Wafer IGBT)

- Field stop IGBT or Thin wafer IGBT was introduced in 2000
- Switching speed is controlled by the dose of P-collector.
- Thin wafer is utilized to fabricate FS-IGBTs.
- N-buffer & P-collector are formed by I.I. and annealing.
- Carrier lifetime is high throughout the N-base



107

Field stop IGBT or Thin wafer IGBT was introduced in the year 2000.

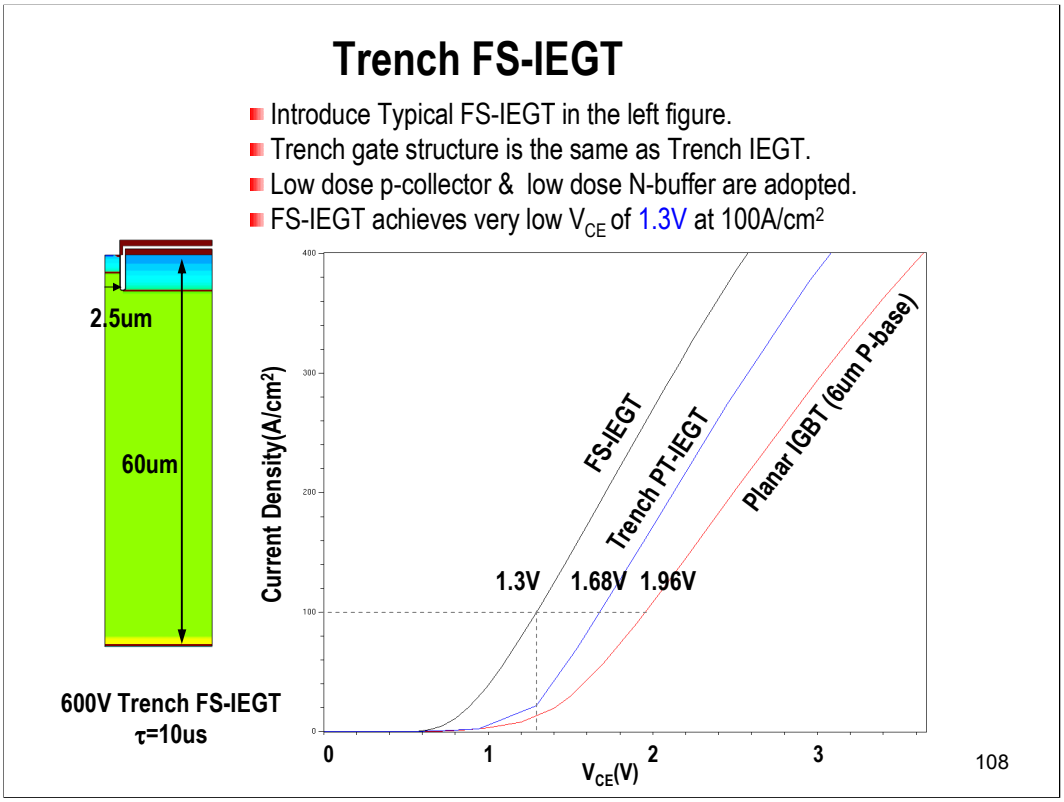
After that, field-stop IGBT has been a major structure of IGBTs.

Switching speed is controlled by the dose of the P-collector.

Thin wafer is used to fabricate FS-IGBTs.

N-buffer & P-collector is formed by ion-implantation and annealing.

Carrier lifetime is high throughout the wafer.



In this figure, I introduce typical 600V FS-IEGT structure.

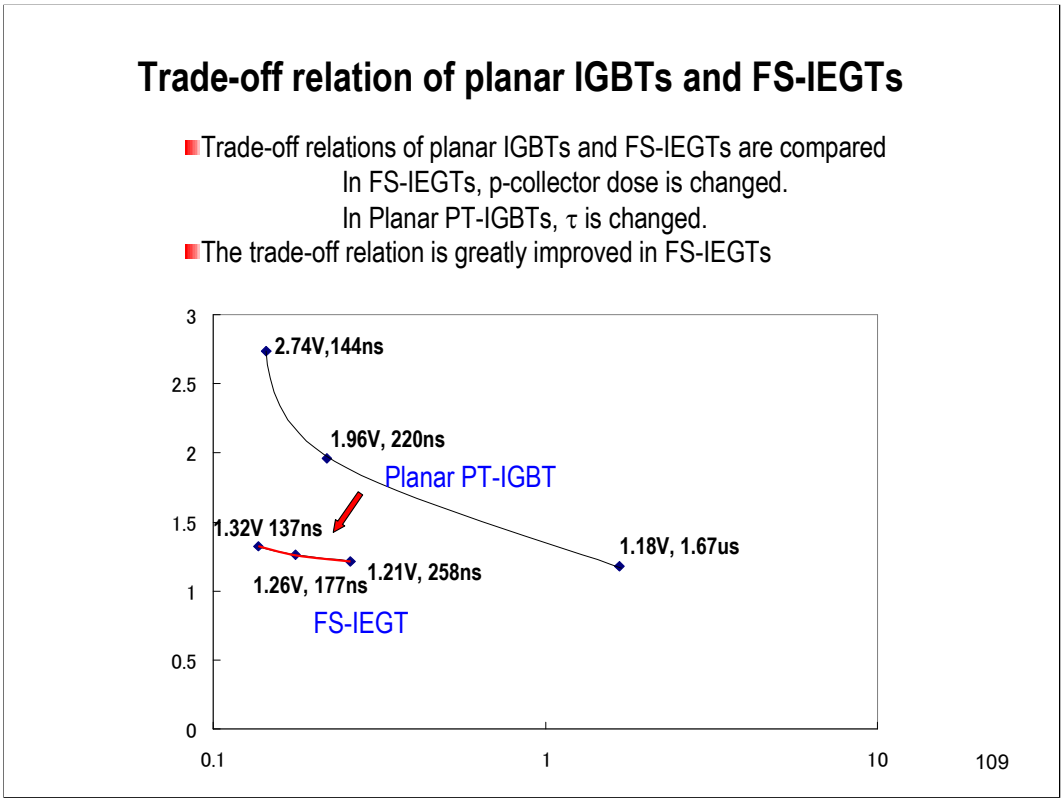
The trench structure on the emitter-side is the same as that of trench PT-IEGT.

Wafer thickness is 60um. The half mesa width is 2.5um.

On the collector-side, the low dose N-buffer and the low dose P-collector are adopted.

The figure compares the I-V curves of FS-IGBT, Trench PT-IEGT and planar IGBT.

FS-IGBT achieves very low on-state voltage, V_{CE}, as low as 1.3V at 100A/cm2.



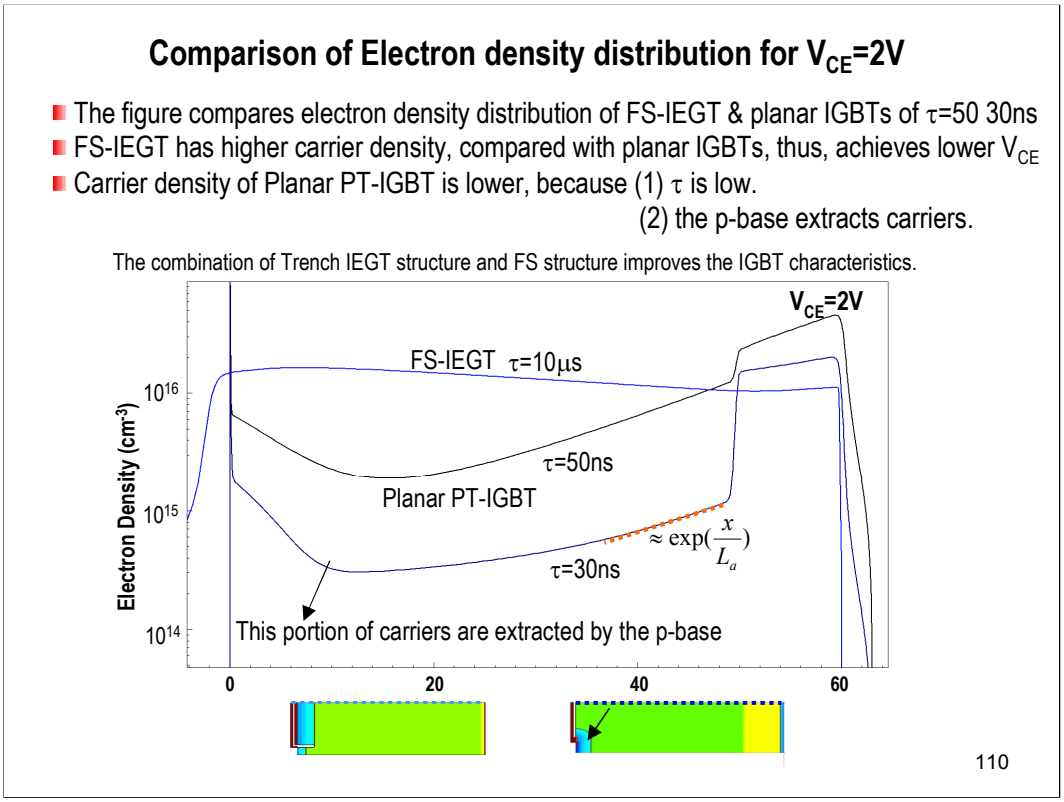
In this figure, I compare the trade-off relations of planar IGBTs and FS-IEGTs.

In planar IGBTs, carrier lifetime is changed to change the switching speed.

In FS-IEGTs, the dose of the collector P-layer is changed to vary the switching speed.

For example, VCE=1.32V and switching time is 137ncec.

The trade-off relation is greatly improved in the FS structure.



This figure compares the electron density distributions of FS-IEGT and two planar IGBTs with carrier lifetime of 50ns and 30ns.

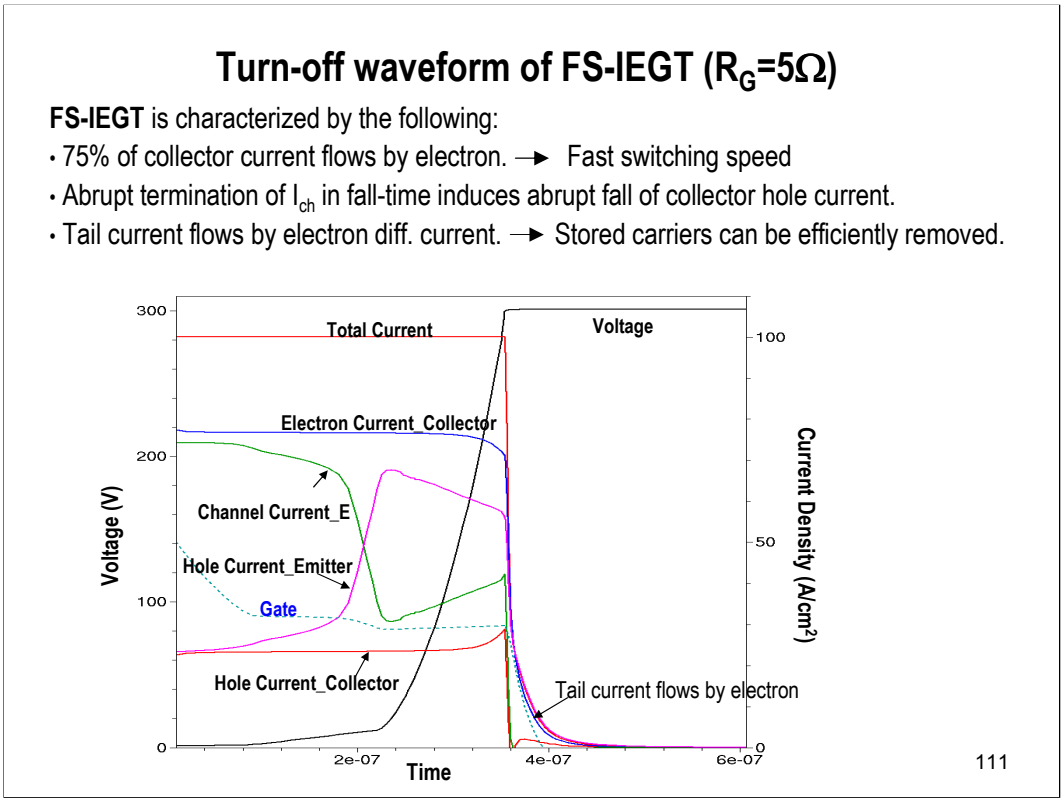
Carrier density of FS-IEGT is significantly higher in the N-base, especially on the emitter-side, compared with that of planar-IGBT.

Carrier density of planar IGBT is low, because

(1) The carrier lifetime is low in planar IGBTs.
So, the carrier density decreases exponentially according to the ambipolar diffusion length.

(2) The P-base extracts holes and reduces the carrier density around the P-base.

The combination of Trench IEGT structure on the emitter-side and FS structure on the collector-side improves the IGBT characteristics.

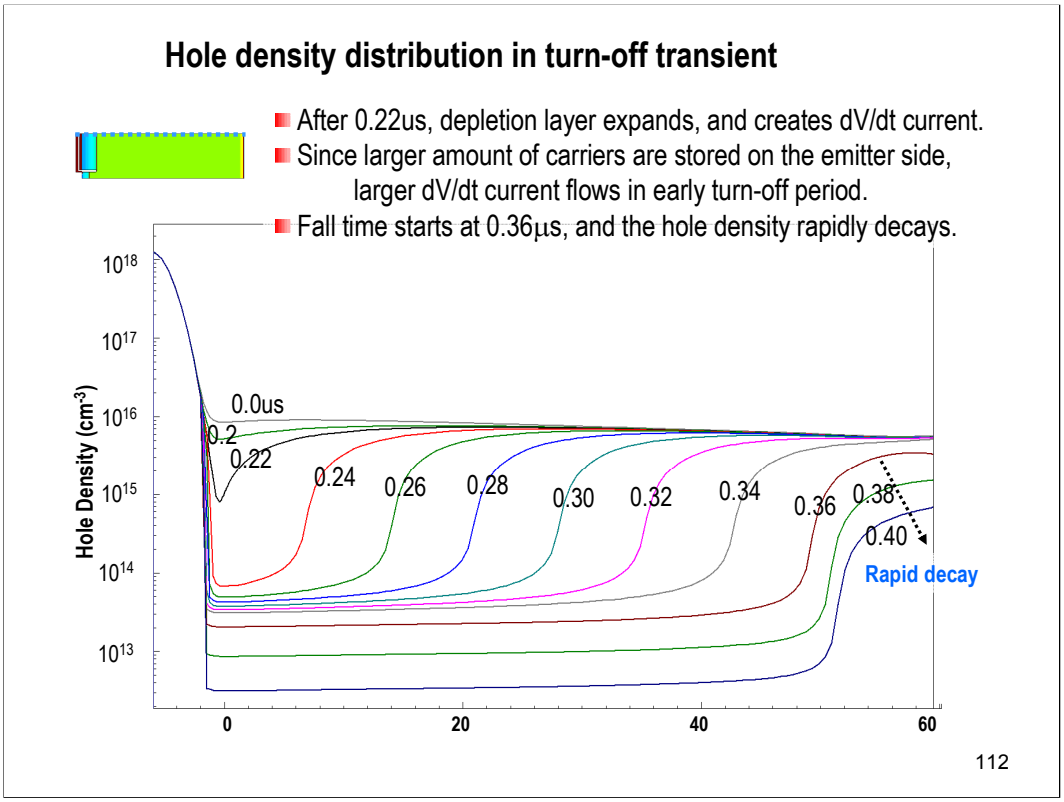


This figure shows the typical switching-off waveforms of FS-IEGTs.

FS-IEGT is characterized by the following:
75% of collector current flows by electron.
The electron current can be directly turned-off by the gate, so, fast switching speed is generally realized.
(In FS-IEGTs, the collector P-layer is so-called transparent, most of the electron current reaches the collector electrode without recombination.)

The channel electron current is abruptly terminated in the fall time. This abrupt termination of channel electron current induces abrupt fall of collector hole current in the fall-time.

Tail current flows by electron diffusion current.
So, the stored carriers in the N-base can be efficiently removed.



This figure shows hole density distribution in the turn-off transient.

After 0.22us, the depletion layer expands, and creates dV/dt current.

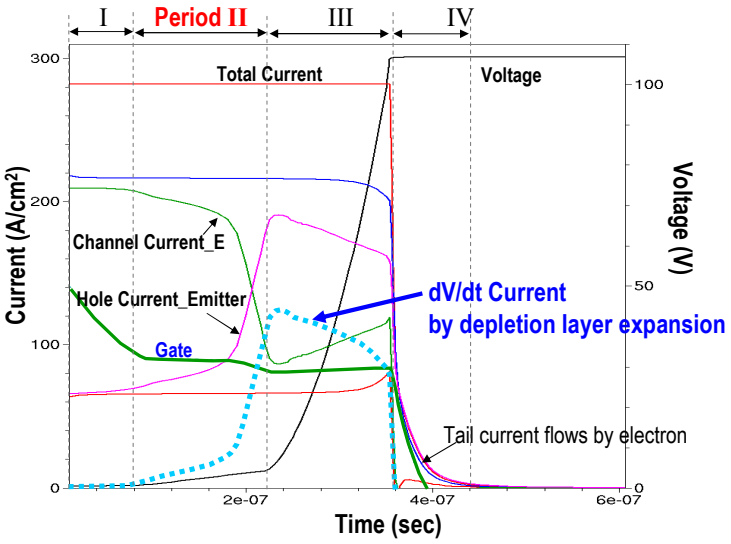
Since larger amount of carriers are stored on the emitter side, larger dV/dt current flows in early turn-off period.

Fall time start at 0.36us and the hole density on the collector-side rapidly decays.

This is because electron diffusion current mostly flows in the tail period and electrons easily escape from the N-base into the P-collector.

Turn-off waveform of FS-IEGT (R_G=5Ω)

- Turn-off waveforms can be divided into four periods
- In period II, I_{ch} starts to decrease, and V_{CE} increases to keep V_G at constant.
- But, the current created by depletion layer greatly increases the hole current in period II, V_G starts to decrease already in period II.



113

Turn-off waveforms can be divided into the four period.

This is similar to planar PT-IGBTs.

In period I, the gate voltage simply decreases.

In the period II, the channel electron current start to decrease, and the collector voltage increases to keep gate voltage at the same value and to maintain electron channel current.

However, in FS-IGBTs, dV/dt current created by the depletion layer becomes very large already in the period II as shown by the broken blue line.

This dV/dt current greatly increases the hole current.

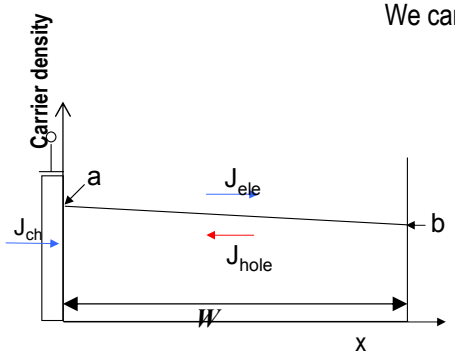
Thus, the gate voltage start to decrease already in the period II to decrease channel electron current.

This is one of big differences between planar PT-IGBTs and FS-IGBTs

Simple Analytical Turn-off Model of FS-IGBTs

- **Help to understand the current created by depletion layer!**

We can create simple analytical turn-off model of FS-IGBT because we can neglect recombination.



Carrier profile is simply approximated as a line:

$$n = \frac{b-a}{W}x + a, \quad n = p$$

High injection condition

Total current is given as summation of J_{ele} and J_{hole}.

$$J_{total} = J_{ele} + J_{hole}$$

Following eq. holds.

$$J_{ele} = J_{ch}$$

P-emitter injection efficiency is defined as:

$$\frac{J_{hole}}{J_{total}} = \alpha : \text{p} - \text{emitter Injection efficiency}$$

114

In this slide, I introduce a simple IGBT turn-off model.

This model helps to understand the current created by depletion layer.

We can create simple analytical turn-off model of FS-IGBT because we can neglect recombination.

In FS-IGBTs, the carrier distribution is approximated as a linear line.

Total current J_{total} is given as the summation of J_{ele} and J_{hole}.

The electron current is the same as the channel current.

P-emitter injection efficiency is defined as this equation:

$J_{hole}/J_{total}=\alpha$

It is assumed that the channel current, J_{ch}, stops before rapid V_{CE} increase.

The same amount of electron current is created by depletion layer after J_{ch} stops.

It is assumed that depletion layer expand dx for the time step dt. This creates charge $J_{ch}dt$.

$$J_{ch}dt = qndx \qquad ...(1)$$

$$J_{ele} = J_{ch} = (1-\alpha)J_{total}$$

In the depletion layer, all current flows by holes in the saturated velocity, v_s :

$$\frac{dE}{dx} = -\frac{q}{\varepsilon}(N_D + \frac{J_{total}}{qv_s})$$

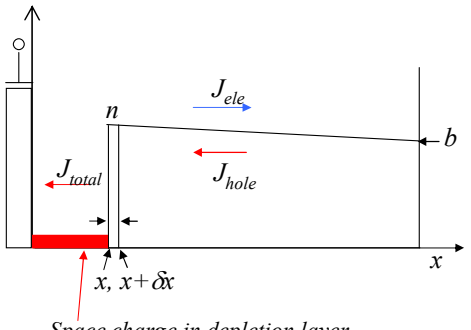
Integrate dE/dx twice, we have:

$$V = -\int_0^x Edx = \frac{1}{2}(\frac{qN_D}{\varepsilon} + \frac{J_{total}}{\varepsilon v_s})x^2 \quad ...(2)$$

Carrier density is given as:

$$n = \frac{b-a}{W}x + a \qquad ...(3)$$

Solve (1)(2) & (3), using Excel.



Space charge in depletion layer

115

It is assumed that the channel current stops before the rapid increase of VCE.

The same amount of electron current is created by the depletion layer after the channel current disappears.

We assume that the depletion layer expand dx for the time step dt. This creates the charge *Jchdt*.

Thus, Eq.(1) holds.

In the depletion layer, all current flows by holes in the saturated velocity, *vs.*

Thus, the Poisson equation holds.

Integration of the Poisson eq. twice gives Eq.(2).

The carrier density is given by Eq.(3)

We can solve the three equations.

Solution procedure

$n = -\frac{a-b}{W}x + a$ ...(3) Substitute Eq.(3) for n in Eq.(1).

$qndx = J_{ch}dt$ ...(1)

Then, we have

$$[-(a-b)x + aW]dx = \frac{WJ_{ch}}{q}dt$$

Integrate

$$-\frac{(a-b)}{2}x^2 + aWx = \frac{WJ_{ch}}{q}t$$

$$\frac{(a-b)}{2}x^2 - aWx + \frac{WJ_{ch}}{q}t = 0$$

The solution : $x = \frac{aW - \sqrt{a^2W^2 - 2(a-b)WJ_{ch}t/q}}{(a-b)}$, $J_{ch} = (1-\alpha)J_{total}$

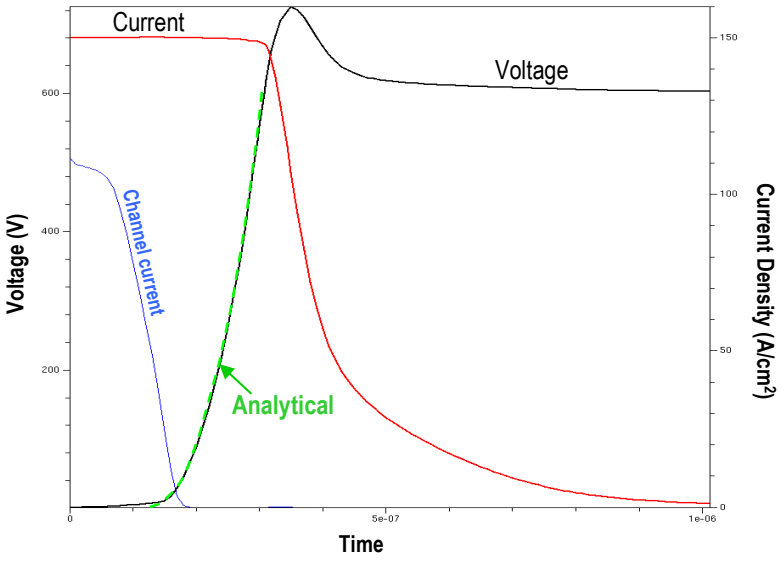
$V = \frac{1}{2}(\frac{qN_D}{\varepsilon} + \frac{J_{total}}{\varepsilon v_s})x^2$ ...(2) Substitute for x in Eq.(2).

116

This shows the solution procedure:

Comparison of TCAD and Analytical Solution

The figure compares TCAD simulation and the analytical model with parameters below:



a=2.0e16
b=1.3e16
J=150A/cm^2
α=0.26
W=120um

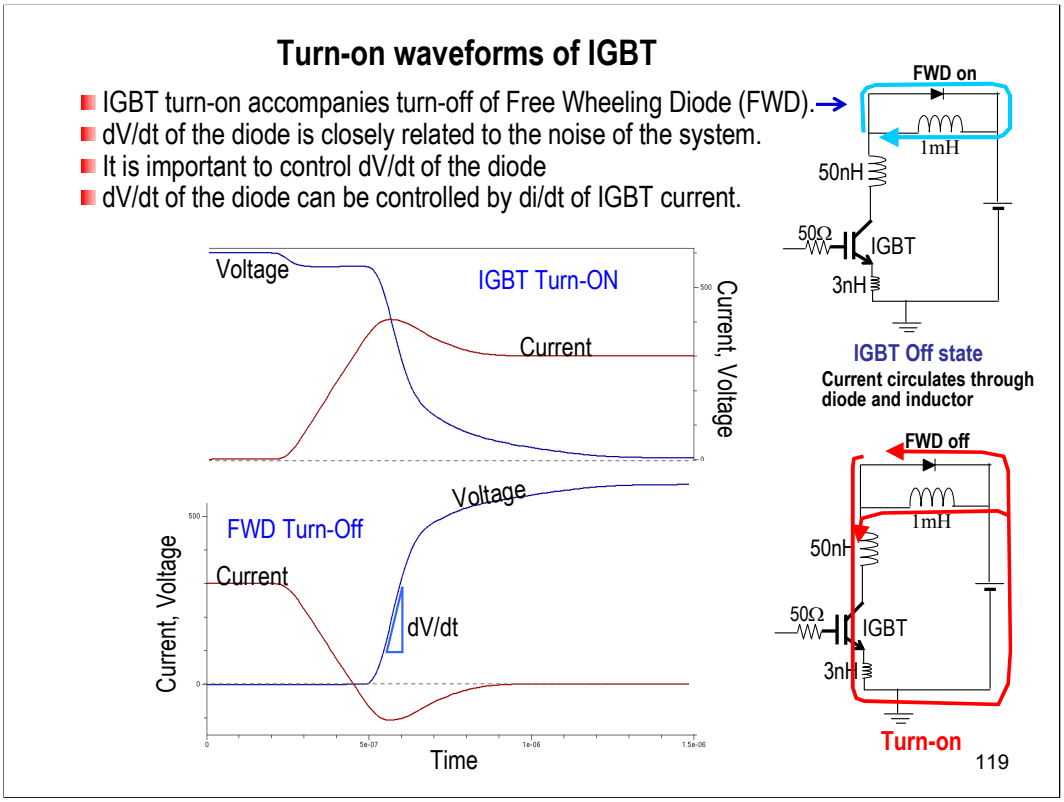
117

This figure compares the analytical model and TCAD simulation.

The agreement is quite good.

The used parameters are given in the slide.

We focus on Turn-On Characteristics of IGBT and Diode Recovery



IGBT turn-on accompanies the turn-off of Free wheeling diodes.

The inductor tries to keep the current at the same value.

The current circulates among the inductor and the diode while the IGBT is off-state.

When the IGBT is switched-on, the inductor current flows into the IGBT and the diode is reverse-baiased.

These are IGBT turn-on and diode recovery waveforms.

The dV/dt of the diode is closely related to the noise of the system.

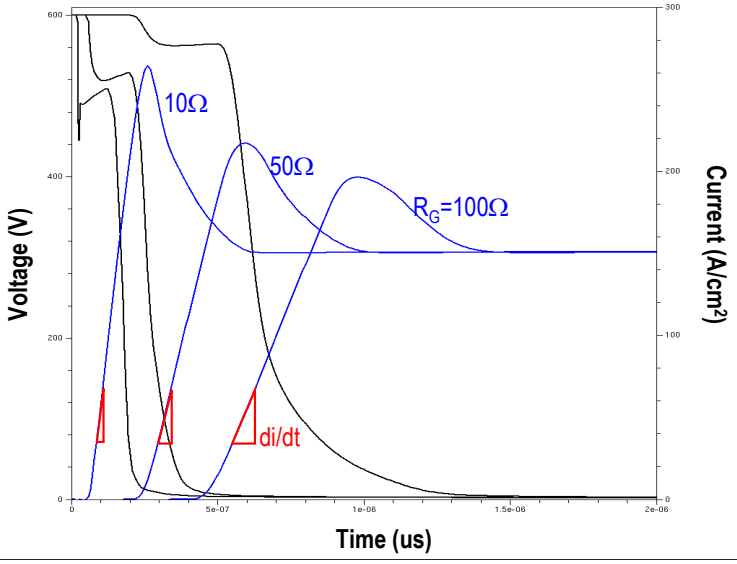
So, the control of the dV/dt of the diode is strictly important.

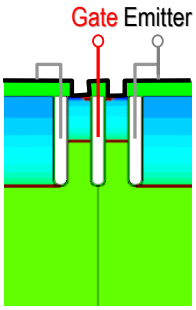
The dV/dt of the diode can be controlled by the di/dt of IGBT current.

Thus, the di/dt control capability of IGBT is important.

di/dt controllability of Dummy Gate IGBT

- di/dt of the IGBT current should be controlled by R_G
- The figure shows the turn-on waveforms of the IGBT when R_G= 10, 50 and 100Ω
- di/dt is well controlled by the gate resistance, R_G.





Dummy Gate IGBT
Poly-Si of outer trenches is electrically connected to emitter electrode

120

Now, I'd like to discuss the di/dt control capability of IGBTs.

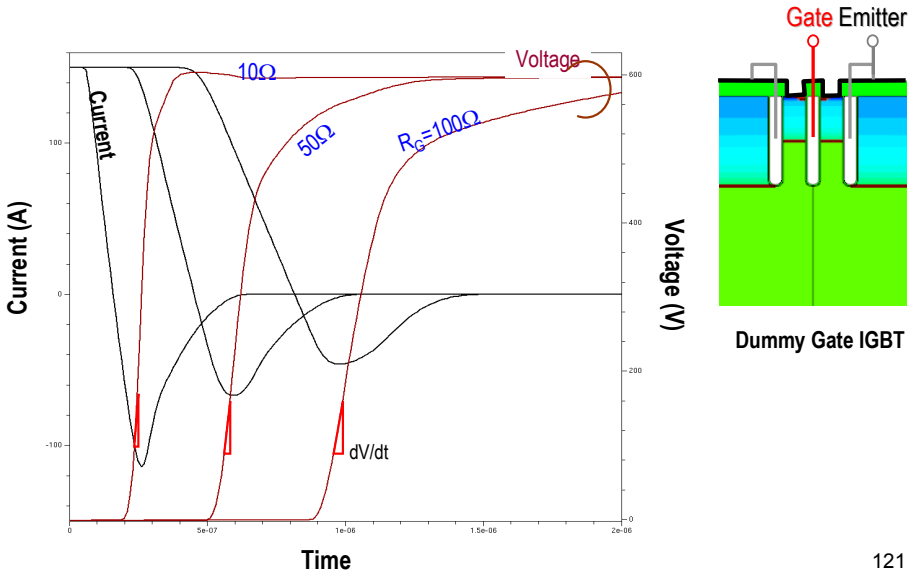
The di/dt of the current can usually be controlled by the value of the gate resistance.

This figure shows the turn-on waveforms of Dummy Gate IGBT when the gate resistance of IGBT is 10Ω, 50Ω and 100Ω.

The di/dt is successfully well controlled by the gate resistance.

FWD Recovery

- The figure shows the turn-off waveforms of FWD, controlled by Dummy Gate IGBT
- The diode dV/dt is excellently controlled by di/dt of Dummy Gate IGBT
- It is very important to control di/dt of IGBT current.



121

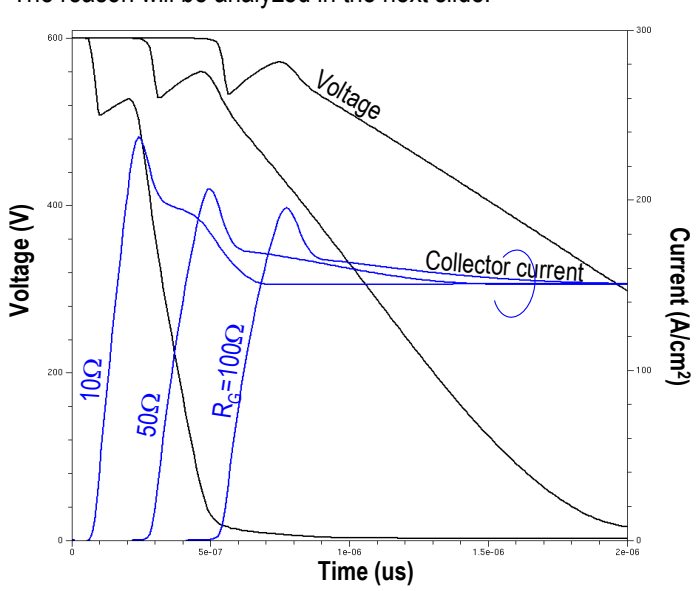
This shows the turn-off waveforms of FWD, which is controlled by Dummy Gate IGBT

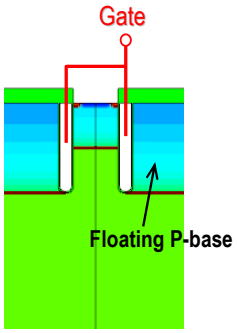
The diode dV/dt is excellently controlled by di/dt of Dummy Gate IGBT.

It is very important to control the di/dt of IGBT current since the current flows commonly in IGBTs and FWDs.

di/dt controllability of Floating P-base IEGT

- This shows the turn-on waveforms of IGBTs with Floating P-base.
- If there is a Floating P-base adjacent to the gate, di/dt cannot be controlled by R_G.
- The reason will be analyzed in the next slide.





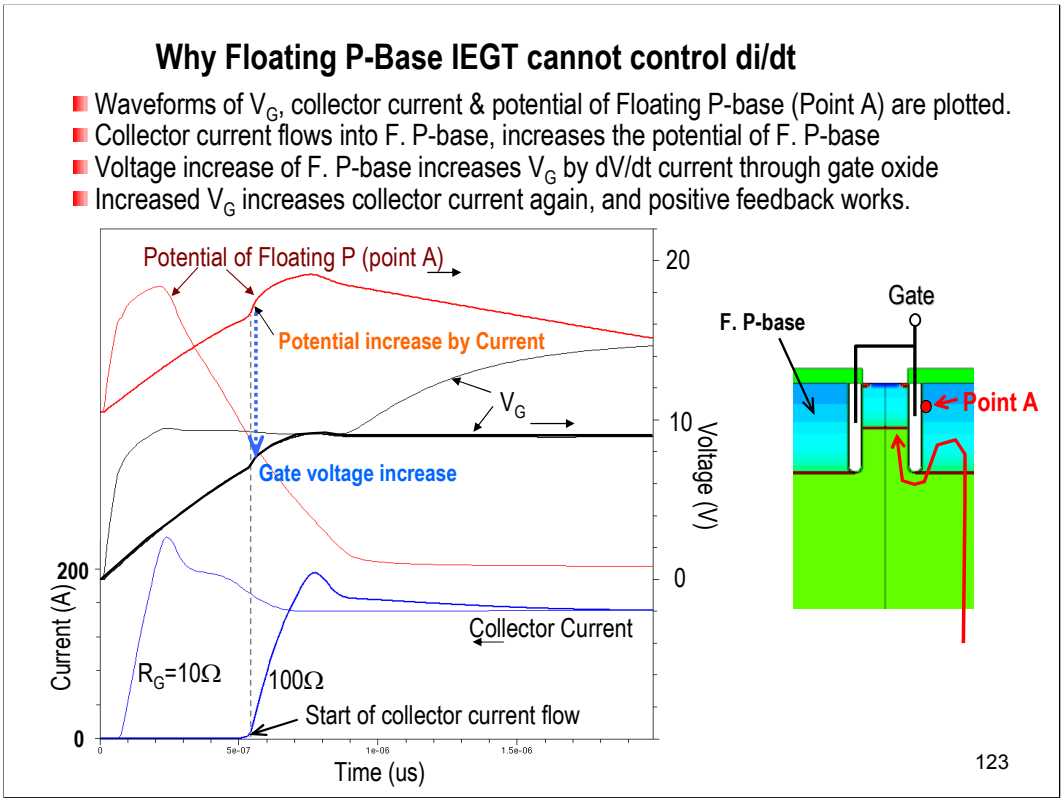
122

This shows the turn-on waveforms of IGBTs with floating P-base.

If there is a floating P-base adjacent to the gate electrode, the di/dt cannot be controlled by R_G.

Even if we increase the gate resistance, the di/dt of the current cannot be controlled effectively if we use IGBTs with floating P-base.

The reason is shown in the following figure.



This figure shows waveforms of the gate voltage, the collector current as well as the potential at the point A, which is located in a floating p-base and adjacent to the gate oxide.

When the gate voltage increases and exceeds the threshold voltage, the collector current starts to flow.

The collector current flows into the floating P-base and increases the potential of the floating p-base.

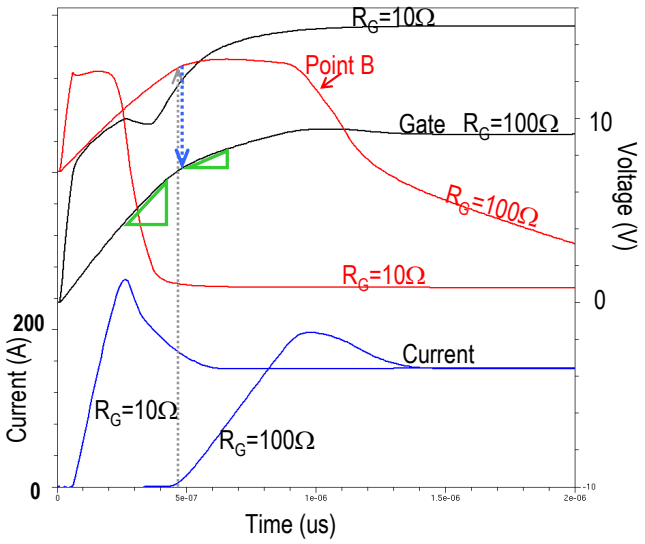
This also increases the gate voltage by the dV/dt current through the gate oxide.

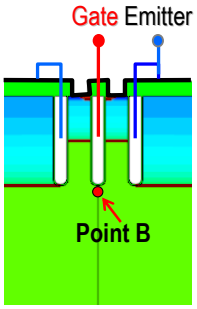
The increase in the gate voltage increases the collector current again.

This positive feedback prevents the control of di/dt by the gate resistance.

Dummy Gate IGBT

- The dummy gate is electrically connected to emitter, thus blocks the influence of floating P
- Voltage of point B does not increase even if collector current flows
- di/dt can be controlled by R_G without the influence of Floating P-base





Dummy gate is connected to Emitter

124

In the dummy gate IGBT, the poly-silicon of the outer trenches is electrically connected to the emitter electrode.

Thus, the influence of the floating p-base is blocked by the outer trenches.

We need to consider only the influence of the potential under the gate electrode, namely the potential of the point B.

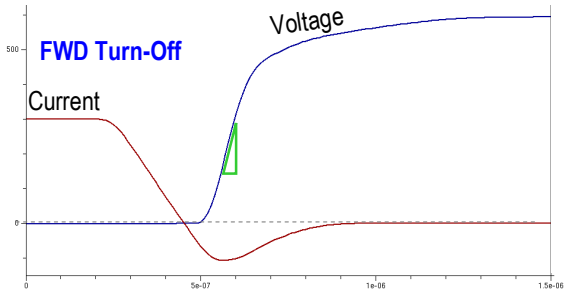
The potential of point B does not increase but stays in the same value even if the collector current flows.

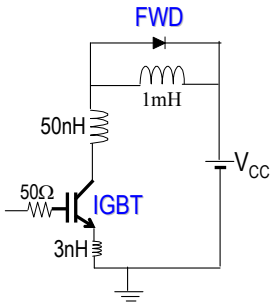
Thus, the gate voltage does not rapidly increase after the collector current flows.

In the dummy gate IGBT, di/dt can be controlled by the gate resistance because the influence of the floating P-base is blocked.

Design of Diode structure

- System noise is closely related to dV/dt of diode recovery.
- Diode dV/dt depends also on the structure of diodes.
- Introduce diode design concept of **"soft recovery"** in the next slide.





125

The system noise is closely related to the dV/dt in diode recovery.

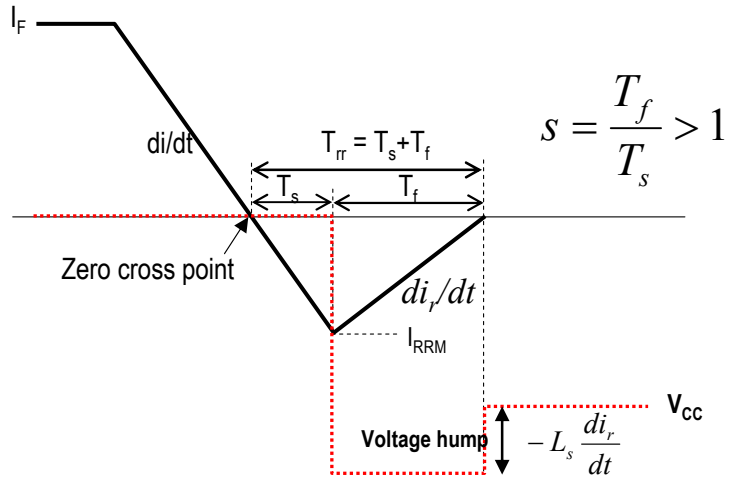
So, the controlling the dV/dt in diode recovery is very important.

Although the dV/dt of diodes is controlled by IGBT, the dV/dt also depends on the structure of diodes.

I' like to introduce the concept of soft recovery in the next slide.

Diode Recovery Waveform

- Typical diode turn-off waveform is illustrated.
- Recovery time, T_{rr} is defined below.
- Softness factor S is defined as T_f / T_s.
- di_r/dt in the time period T_f creates voltage hump, thus should be reduced.



126

This shows typical diode recovery characteristics.

Initiaily, current IF flows. Then, the current decreases at a rate of di/dt, and crosses zero and reaches reverse peak current , IRRM.

Then the current decreases and reaches zero.

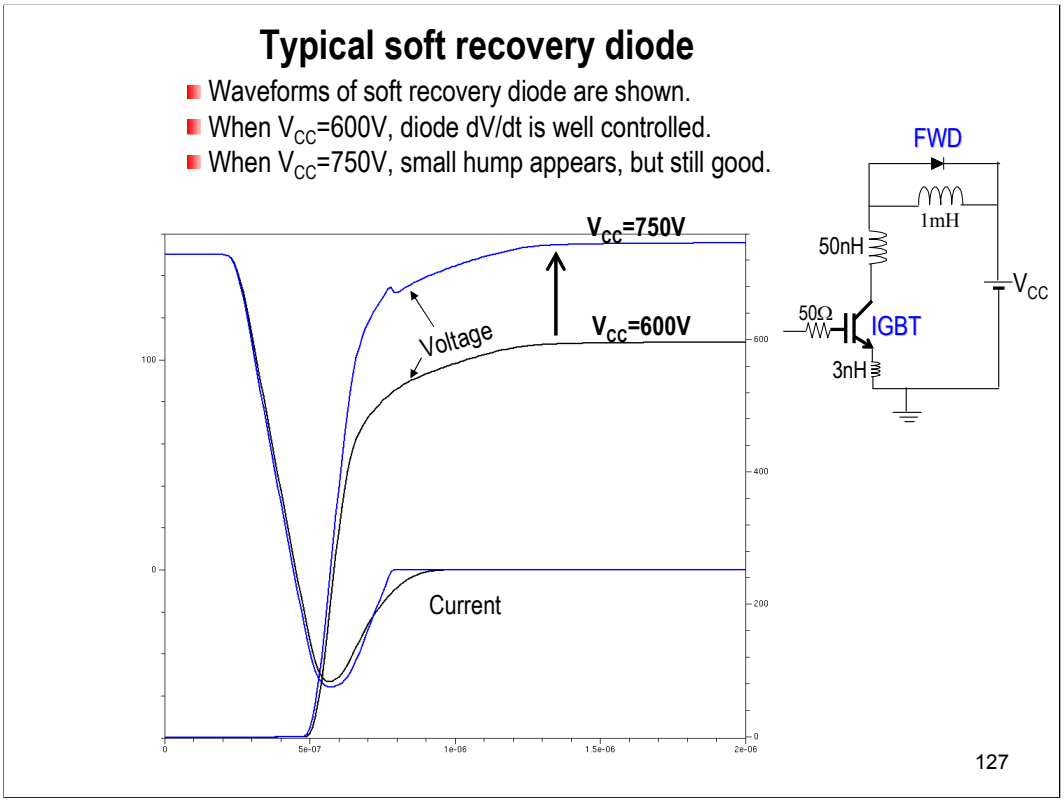
Reverse recovery time t_{rr} is defined as the time from the first zero cross point and the 2nd zero cross point.

Softness factor, S, is defined as S=Tf/Ts.

The di_r/dt in the time period Tf creates the voltage hump in the reverse recovery.

So, the di_r/dt in the time period Tf should be small.

This means that the softness factor S should be large.



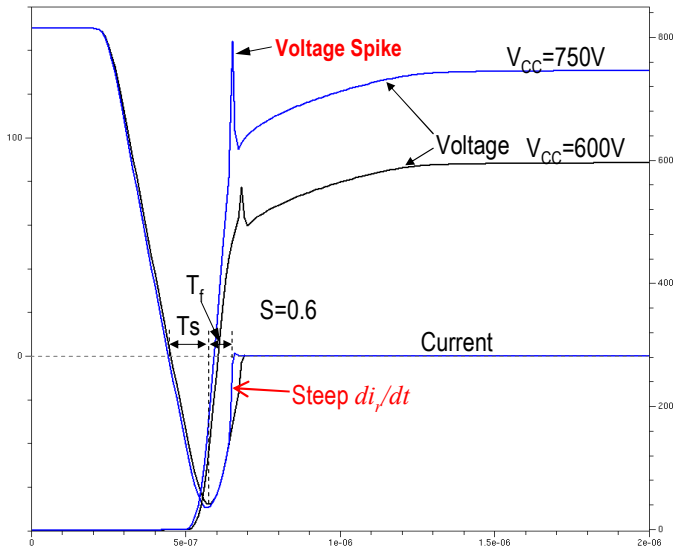
This is a typical soft recovery diode.

If the applied voltage is 600V, reverse recovery characteristics are very smooth.

When the applied voltage is increased to 750V,
the voltage waveform is still acceptable, although a small hump appears.

Example of Snappy diode

- Voltage spike appears because a steep di_r/dt appears in the current waveform.
- Softness factor=0.6



128

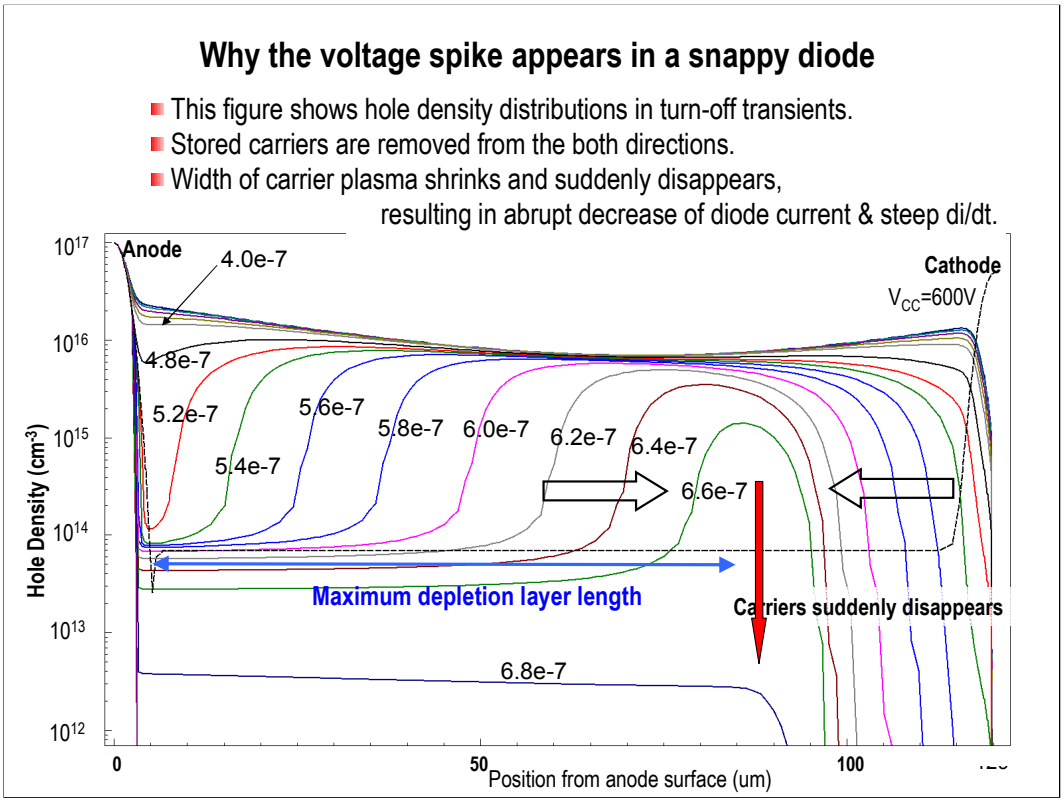
This is an example of a snappy diode.

There is a voltage spike in the reverse recovery waveform.

This voltage spike is created by a steep $\boldsymbol{di_r/dt}$ in the recovery current.

The voltage spike becomes very large when the applied voltage is increased to 750V.

The softness factor of this diode is 0.6.



This slide shows why the voltage spike is created in a snappy diode.

The figure shows hole density distributions in turn-off transients

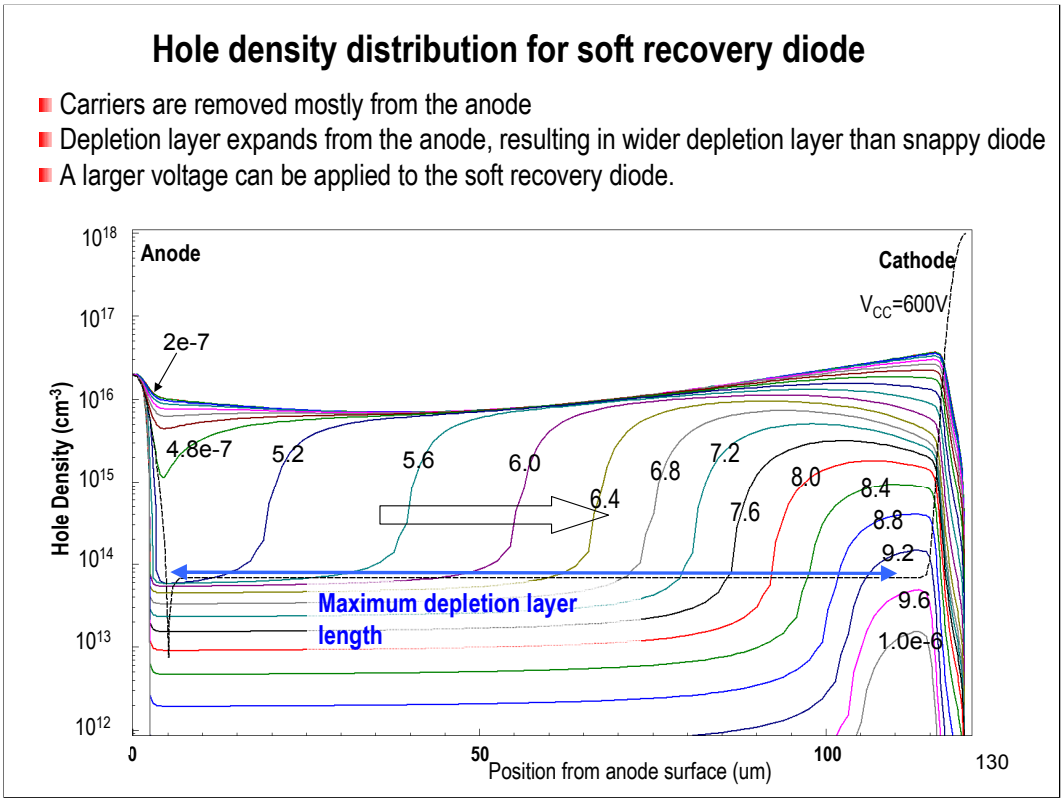
Stored carriers are removed from the both directions, namely from the anode and from the cathode.

The width of the carrier plasma shrinks and finally suddenly disappears.

This causes abrupt decrease of diode current and a steep di/dt.

The steep di/dt induce a large voltage spike.

Thus, this diode design should be avoided.



This figure shows hole density distribution of soft recovery diode in the turn-off transients.

The depletion layer expands mostly from the anode.

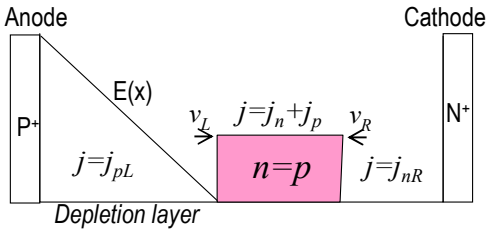
The carriers are not removed from the cathode.

The final length of the depletion layer is longer than that of the snappy diode when the plasma disappears.

This means that a larger voltage can be applied to the soft recovery diode, compared with the snappy diode.

1-d Analysis of Carrier Plasma Shrink

- Analyze how carrier plasma is removed in pin diode
- Assume simple case: carrier distribution is flat.
- Assume plasma shrinks in velocity v_L at the left edge and v_R at the right edge



$J_n\ J_p$ flows only by drift inside the plasma.
Thus, is expressed in terms of total current j:

$$j_p = \frac{\mu_p}{\mu_n + \mu_p} j \quad \text{...Eq.(1)}$$

$$j_n = \frac{\mu_n}{\mu_n + \mu_p} j \quad \text{...Eq.(2)}$$

In the depletion layer, all current is carried by holes. Using hole current, j_{PL}, we have:

$$j = j_{pL} = j_n + j_p \quad \text{...Eq.(3)}$$

At the left edge of plasma, hole current increases by the amount Δj_p

$$\Delta j_p = j_{pL} - j_p = j_n = \frac{\mu_n}{\mu_n + \mu_p} j \ldots \quad \text{Eq.(4)}$$

131

Now, we analyze how carrier plasma is removed in the diode.

We assume a simple case, where carrier distribution is flat. We further assume that plasma shrinks in velocity *vL* at the left edge of plasma and *vR* at the right edge of plasma.

As carrier profile is flat, electron hole current flows only by drift. The ratio of hole current and electron current is mobility ratio.

Thus, the equation (1) and (2) hold.

In the depletion layer, all of the current flows by holes.

The hole current in the depletion layer, j_{pL} is equal to the total current. Thus, Eq.(3) holds.

At the left edge of the carrier plasma, hole current increases by the amount, $\Delta j_p = j_{pL} - j_p$. This is equal to J_n from Eq.(3). Thus, Eq.(4) holds.

Cite Eq.(4) again.

increment $\Delta j_p = j_{pL} - j_p = j_n = \frac{\mu_n}{\mu_n + \mu_p} j \ldots$ *Eq.*(4)

The amount of carriers, which is removed from plasma in the time step, dt, is $\Delta j_p dt$

$$\Delta j_p \cdot dt = j_n \cdot dt = \frac{\mu_n}{\mu_n + \mu_p} j \cdot dt$$

This amount is equal to the amount, $qndx$, which is removed by depletion layer expansion:

$$\Delta j_p \cdot dt = j_n \cdot dt = \frac{\mu_n}{\mu_n + \mu_p} j \cdot dt = q \cdot n \cdot dx \ldots \text{Eq.(5)}$$

Then, v_L is expressed as:

$$|v_L| = \frac{dx}{dt} = \frac{j_n}{qn} = \frac{\mu_n}{\mu_n + \mu_p} \frac{j}{qn} \approx 0.75 \frac{j}{qn} \quad \ldots \text{Eq.(6)} (\mu_n \approx 3\mu_p)$$

Analogously, v_R is given as:

$$|v_R| = \frac{j_p}{qn} = \frac{\mu_n}{\mu_n + \mu_p} \frac{j}{qn} \approx 0.25 \frac{j}{qn} \quad \ldots \text{Eq.(7)}$$

132

The amount of carriers, which is removed in the time step, dt, is $\Delta j_p dt$, and is equal to $j_n dt$.

This amount is equal to the amount, qndx, which is removed by the increase in the depletion layer width, dx. Thus, Eq.(5) holds.

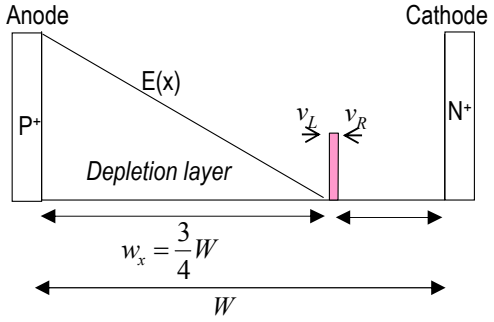
Now, the velocity of plasma front , v_L, is given by dx/dt, and Eq.(6) holds.

In the same way, at the right edge of carrier plasma, Eq.(7) holds.

The maximum depletion layer width, w_x, before the plasma disappears, is given by the velocity ratio, and is expressed as:

$$w_x = \frac{v_L}{v_L + v_R} W \approx \frac{3}{4} W \quad \text{...Eq.(8)}$$

Thus, if carrier plasma is flat, the depletion layer length is 3/4 of i-region width, W.



133

The maximum depletion layer width, *wx* , before the plasma disappears, is given

by the velocity ratio, and is expressed by Eq.(8).

Thus, if carrier plasma is flat, the depletion layer length is ¾ of the i-region width.

Non-uniform Carrier Plasma

- Extend the theory to non-uniform carrier plasma
- Assume carrier density n_L for the left edge of plasma and n_R for the right edge.

If the non-uniformity is not large, we can neglect diffusion current inside plasma.
Eqs. (1)and (2) still hold.

$$j_p = \frac{\mu_p}{\mu_n + \mu_p} j \quad \text{...Eq.(1)}$$

$$j_n = \frac{\mu_n}{\mu_n + \mu_p} j \quad \text{...Eq.(2)}$$

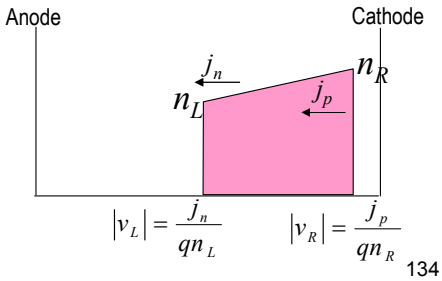
Then, we can derive similar Eqs.(9) and (10) at the both edges of plasma.

$$|v_L| = \frac{j_n}{qn_L} \quad \text{...Eq.(9)}$$

$$|v_R| = \frac{j_p}{qn_R} \quad \text{...Eq.(10)}$$

Then, we have

$$\frac{v_R}{v_L} = \frac{\mu_p}{\mu_n} \cdot \frac{n_L}{n_R} \quad \text{...Eq.(11)}$$



134

We extend the theory to non-uniform carrier plasma.

Assume carrier density n_L for the left edge and n_R for the right edge.

If the non-uniformity is not large, we can neglect diffusion current in the carrier plasma.

Still, Equations (1)and (2) hold.

On the both edges of plasma, we assume that Eqs.(3), (4) hold.

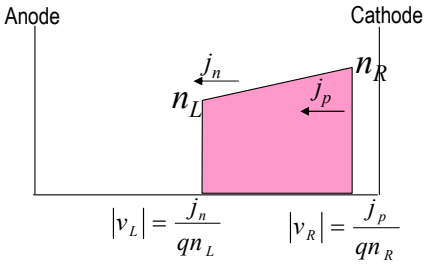
Then, the plasma front velocity ratio is given by Eq.(5).

Cite Eqs.(9),(10) and (11).

$$|v_L| = \frac{j_n}{qn_L} \quad \text{...Eq.(9)}$$

$$|v_R| = \frac{j_p}{qn_R} \quad \text{...Eq.(10)}$$

$$\frac{v_R}{v_L} = \frac{\mu_p}{\mu_n} \cdot \frac{n_L}{n_R} \quad \text{...Eq.(11)}$$

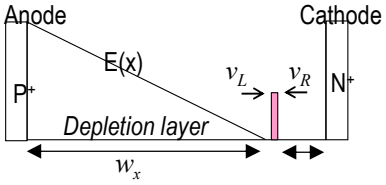


Then, the maximum depletion layer width (plasma vanishing point) is given by Eq.(12).

$$w_x = \frac{v_L}{v_L + v_R} W = \frac{1}{1 + \frac{\mu_p}{\mu_n}\frac{n_L}{n_R}} \cdot W = \frac{1}{1 + \frac{\mu_p}{\mu_n}\eta} \cdot W,$$

$$\eta = \frac{n_L}{n_R} \quad \text{...Eq.(12)}$$

$$if\ \eta = 1/3, w_x = 0.9W$$



It is important in diode design that cathode side n_R should be larger than anode side n_L

135

We cite Eq.(3) (4) and (5) again.

The maximum depletion layer width or plasma vanishing point w_x is given by Eq.(6).

If e-ta η=1/3, the depletion layer expands 90% of the i-region.

It is important in diode design that cathode side *nR* should be larger than anode side *nL*

5. SOA of IGBT

Device failure

Device failure is a complicated issue. Causes for device failure usually involve both thermal and electrical issues and, thus beyond the scope of this text. We focus only on electrical issues.

Safe Operating Area

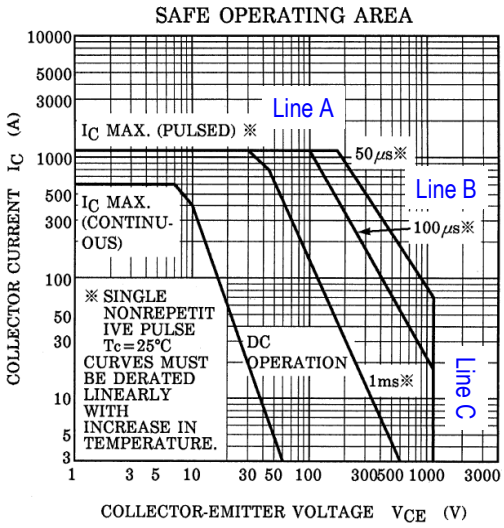
Safe operating area is defined as the area where the device can be operated safely as long as the power dissipation is kept within the thermal constraints such that the temperature rise does not exceed the maximum junction temperature.

Thermal Limit --- complicated issue

Electrical Limit --- theoretically manageable

Aged deterioration --- outside the scope of our text

Forward Biased SOA of 1200V 600A IGBT



MG600Q1US51

Line A limits the maximum allowable current. This is related to the reliability of materials.

Line C is the limit that is imposed by the maximum static breakdown voltage

Line B is the limit when devices are in the state of both high voltage and high current. The limitation comes mostly from power dissipation and impact ionization induced device failure.

This will be discussed in the following.

138

This shows an example of forward Biased SOA of 1200V 600A IGBT.

The line A limits the maximum allowable current.

The maximum current is limited by various reliability issues.

The line C is the limit that is imposed by the maximum static breakdown voltage.

Line B is the limit line when devices are in the state of both high voltage and high current.

This situation is encountered in the turn-on transients.

The limitation comes mostly from power dissipation and impact ionization induced device failure.

This will be discussed in the following.

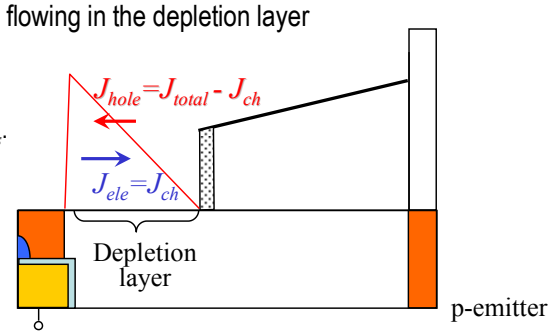
Maximum limit of Forward Biased Safe Operating Area

- Introduce a simple model of FBSOA
- Assume current is flowing in the depletion layer

Current flows in saturated velocity, v_s.

$$J_{hole} = qv_{hole,s}p \quad \text{....Eq.(1)}$$

$$J_{ele} = qv_{ele,s}n \quad \text{....Eq.(2)}$$



Obtain space charge density ρ:

$$\rho = N_D + p - n = N_D + \frac{J_{hole}}{qv_{hole,s}} - \frac{J_{ele}}{qv_{ele,s}} \quad \text{...Eq.(3)}$$

We assume that $J_{ele} = J_{ch}$ $v_{hole,s} = v_{ele,s} = v_s$

$$\rho = N_D + \frac{J_{total} - J_{ch}}{qv_{hole,s}} - \frac{J_{ch}}{qv_{ele,s}} = N_D + \frac{J_{total}}{qv_s} - \frac{2J_{ch}}{qv_s} \quad \text{....Eq.(4)}$$

139

Here, we introduce a simple model of forward SOA for IGBTs.

First, we assume that current is flowing in the depletion layer.

In the depletion layer, we assume that holes and electrons travel in their saturation velocites: $v_{hole,s}$ $v_{ele,s}$.

The electron current and the hole current are expressed as Eq.(1) and (2).

The space charge density is given by Eq.(3).

Electron current, J_{ele}, must be equal to channel current, J_{ch}.

And, we assume that saturation velocities of electrons and holes are the same.

Then, Eq(4) holds.

Static breakdown voltage for an abrupt pn junction is given by this equation:

$$V_B = 60\left(\frac{E_G}{1.1}\right)^{\frac{3}{2}}\left(\frac{N}{10^{16}}\right)^{-\frac{3}{4}} \quad \text{...Eq.(5)} \quad \text{from Sze}$$

It is assumed that N can be replaced with ρ in Eq(5).

$$V_{BD} = 60\left(\frac{E_G}{1.1}\right)^{1.5}\left(\frac{N_D + J_{total}/qv_s - 2J_{ch}/qv_s}{10^{16}}\right)^{-3/4} \quad \text{.....Eq.(6)}$$

Eq.(6) gives the maximum limit of FBSOA. If the device exceeds this limit, large dynamic avalanche generation takes place.
If $J_{ch}=J_{total}/2$, V_{BD} does not depends on the current density J_{total}.
This means FBSOA is infinite.

FBSOA of IGBT is usually very large, because the electron channel current, J_{ch}, exists and is distributed uniformly by the MOS channel.

140

The static breakdown voltage for an abrupt pn junction is given by this equation, using the donor density.

We assume this equation is valid for more general cases.

We substitute ρ for the donor: N in Eq.(5).

The breakdown voltage, V_{BD}, is given by Eq.(6).

If we assume J_{ch} is equal to $J_{total}/2$, V_{BD} does not depends on current density J_{total}.

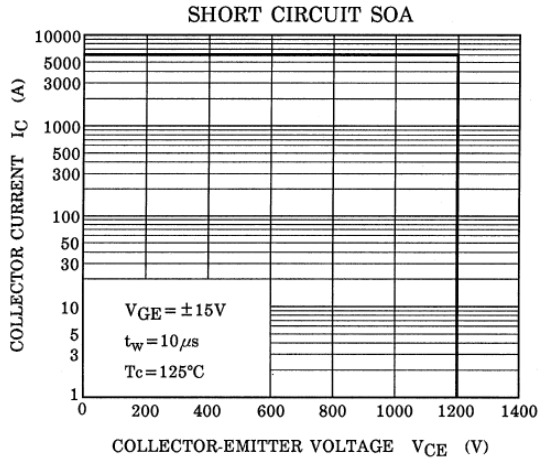
This means that Forward SOA can be very large, almost infinite in such IGBTs.

In general, It is true that FBSOA of IGBT is usually very large, because electron current is distributed uniformly by the MOS channel.

A typical example is the case of the load short-circuit.

Short-Circuit SOA

Short-circuit SOA is a special case of FBSOA.
If the pulse width is as small as 10us, very large SOA is assured.



141

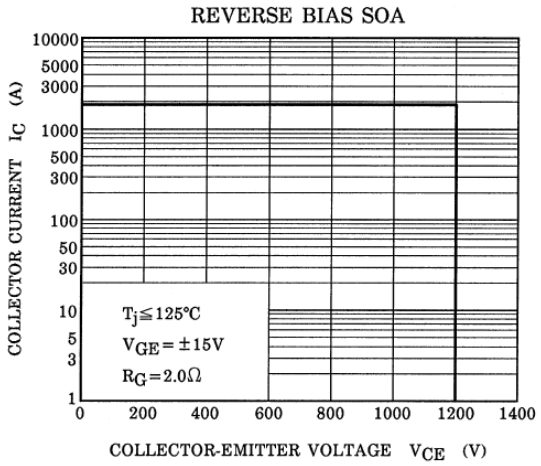
This shows an example of short-circuit SOA of 600A IGBT.

Short-circuit SOA is a special case of Forward biased SOA.

If the pulse width is as small as 10us, very large SOA is assured.

Reverse Bias SOA

Usually, square SOA is assured for RBSOA.
RBSOA is smaller than short-circuited SOA.
The reason is given in the next slide.



142

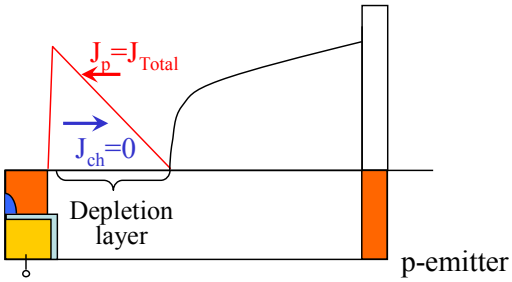
Usually, square SOA is assured for RBSOA.

RBSOA is smaller than short-circuited SOA.

The reason is given in the following.

Electrical RBSOA of IGBT

- We assume that the channel electron current stops in the early stage of turn-off
- RBSOA is smaller than FBSOA because there is no channel current.



$$\text{Space charge}\quad \rho = N_D + \frac{J_{Total}}{qv_s} \qquad ...Eq.(1)$$

$$V_{BD} = 60\left(\frac{E_G}{1.1}\right)^{1.5}\left(\frac{N_D + \frac{J_{Total}}{qv_s}}{10^{16}}\right)^{-3/4} \qquad ...Eq.(2)$$

143

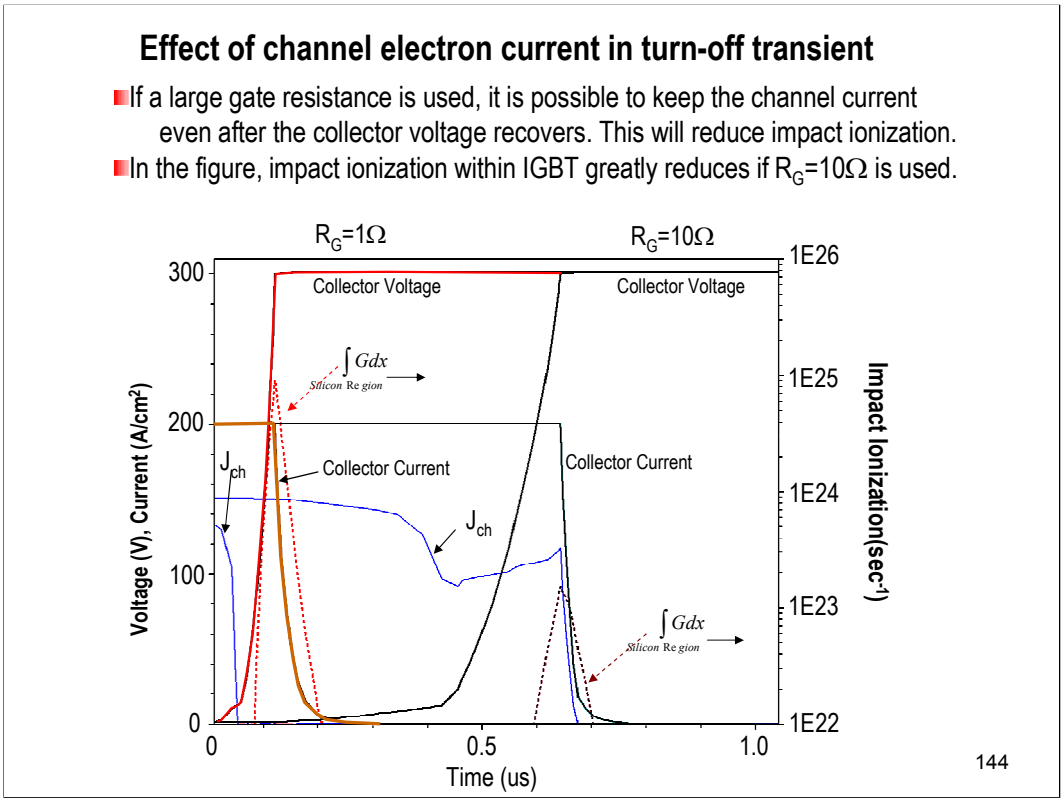
We assume that the channel electron current stops in the early stage of turn-off.

In the depletion layer, all of the current is carried by holes.

The space charge is given by Eq.(1).

Thus, the breakdown voltage by impact-ionization is given by Eq.(2).

As there is no contribution of channel electron current, the electrical RBSOA, is smaller than that of FBSOA.



In deriving RBSOA, we neglect the channel electron current.

However, if we use a large gate resistance, it is possible to keep the channel current even after the collector voltage recovers.

This figure shows the example.

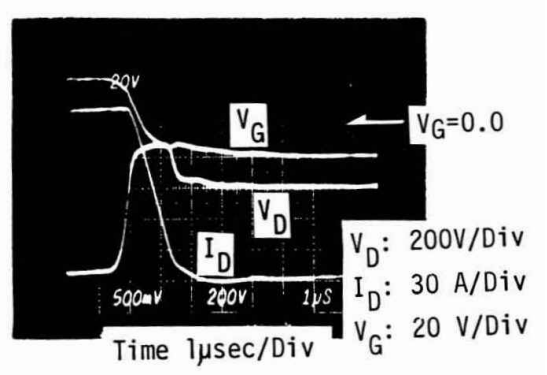
When the gate resistance is as small as 1Ω, channel current stops before collector voltage recovers.

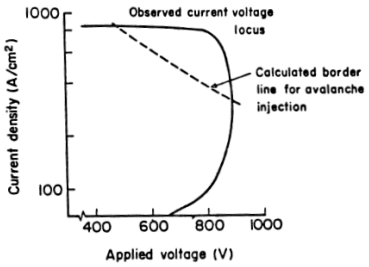
When we increase the gate resistance up to 10Ω, then the channel electron current occupies 50% of the total current.

Then, the total amount of impact ionization occurring inside the device decreases drastically and expands SOA.

It should be noted, however, that the total power loss increases since the turn-off time increases.

Expanded SOA is often observed in the turn-off transients of IGBTs,
Especially, effective in high voltage IGBTs





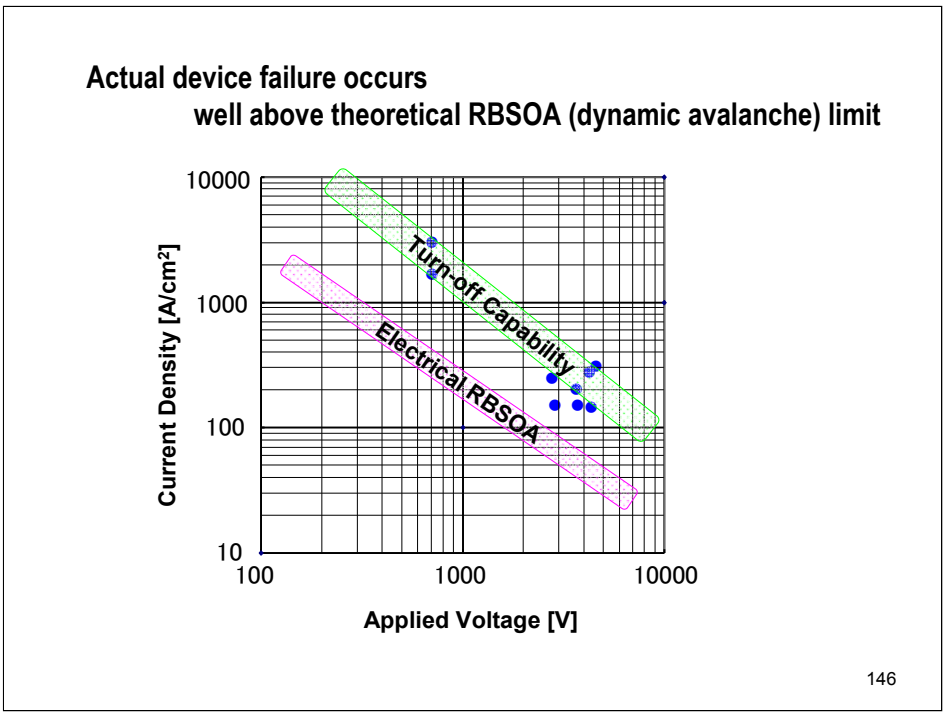
Turn-off waveforms of non-latch-up IGBTs in 1985

145

This expanded SOA is often observed in the turn-off transients of IGBTs.

This figure shows the turn-off waveforms of IGBTs in 1985.

The observed current-voltage locus exceeded the predicted dynamic avalanche limit of RBSOA.



This figure shows actual device destruction points as a function of current and voltage.

Device failure does not occur even if the current voltage locus exceeds the RBSOA limit or the dynamic avalanche limit.

The actual failure occurs well above the avalanche limit.

6. MOSFET-mode IGBT

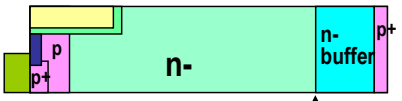
147

Next, I will introduce MOSFET-mode IGBT.

Definition of MOSFET-mode Operation

- First, introduce the concept of anode efficiency, defined at n-base n-buffer junction

Anode efficiency $\gamma \equiv \frac{J_p}{Total\ J} = \frac{J_p}{J_n + J_P} = \gamma_{PE}\alpha_T$



γ is defined at this point!

- MOSFET-mode operation is defined as:

Anode efficiency $\gamma \equiv \frac{J_p}{J_n + J_P} \leq \frac{\mu_p}{\mu_n + \mu_p} \equiv \gamma_{MOS} \approx 0.25$

In MOSFET-mode operation, hole current occupies only less than 25%.
Most of the current flows by electrons in the p-collector.
The collector p-layer works like electron absorber rather than hole emitter.

148

First, we introduce the concept of anode efficiency.

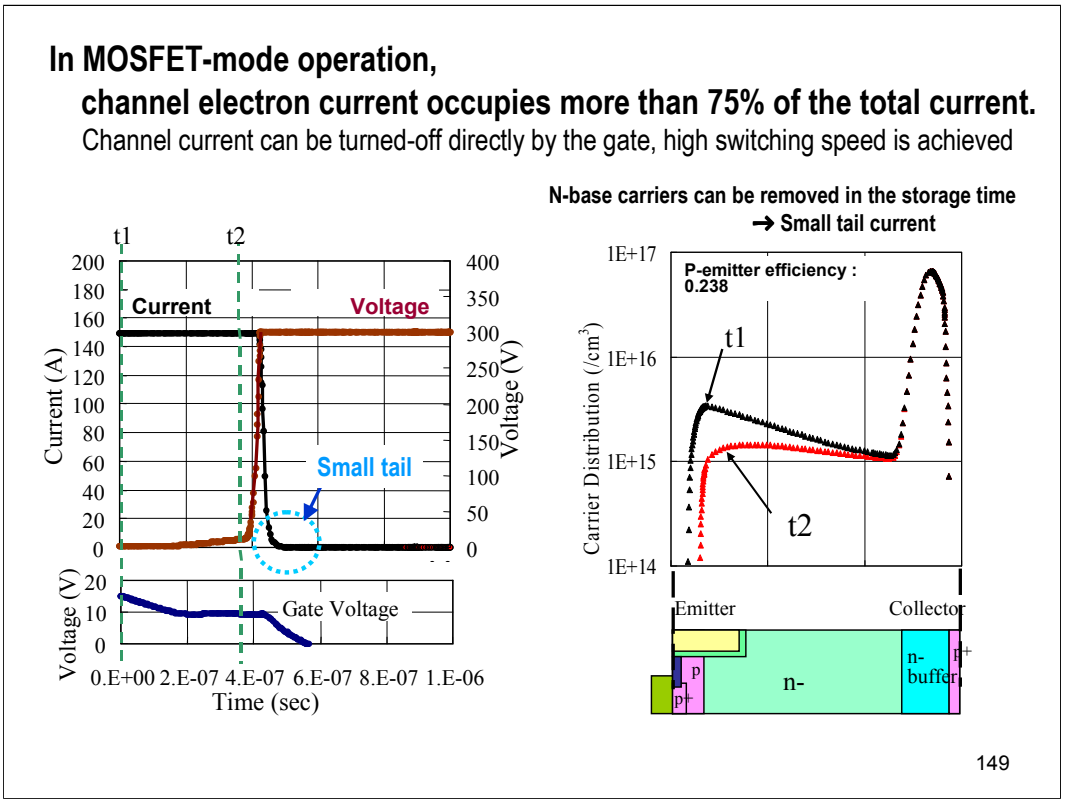
The anode efficiency is defined as hole current over the total current at the n-base n-buffer junction.

Anode efficiency can be expressed as the product of p-emitter injection efficiency γ_{PE} and transport factor in the n-buffer α_T.

MOSFET-mode operation is defined in such a way that the anode efficiency γ is less than γ_{MOS}, which is equal to:

$\mu_p/(\mu_n + \mu_n)$.

In the MOSFET-mode operation, most of the current flows by electrons. More than 75% of total current flows by electrons even in the collector. Thus, the collector p-layer works as more like electron absorber rather than hole emitter.



In the MOSFET-mode operation, the channel electron current occupies more than 75% of the total current.

The channel electron current can be turned-off directly by the gate, high switching speed is achieved.

The left hand side figure shows the turn-off waveforms.

The switching speed of MOSFET-mode IGBT is very fast and tail current is small.

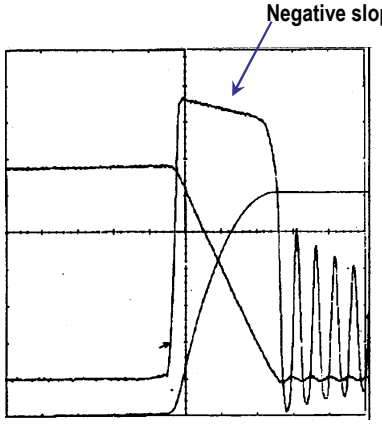
The right hand figure shows the carrier distributions for the time step t1 and t2. A large amount of carriers are stored on the emitter side at the initial time step.

The most of the carriers are removed at the time step t2 before the fall time. Thus, the tail current is small.

Sustaining waveforms (Unclamped Inductive Switching)

- Sustaining waveforms of MOSFET-mode IGBTs are very similar to MOSFET
- Self-clamped voltage decreases as the drain current decreases

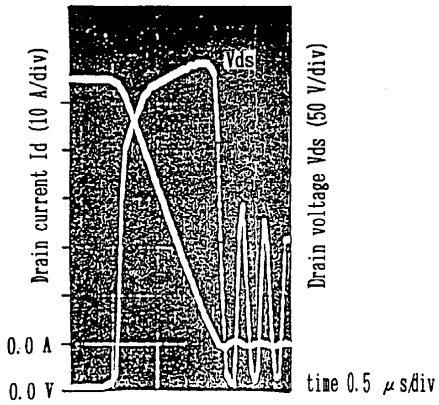
MOSFET-mode IGBT



Time: 0.4μs/Div

- Self-clamped voltage decreases as the drain current decreases

Conv. IGBT



- Self-clamped voltage increases as the drain current decreases

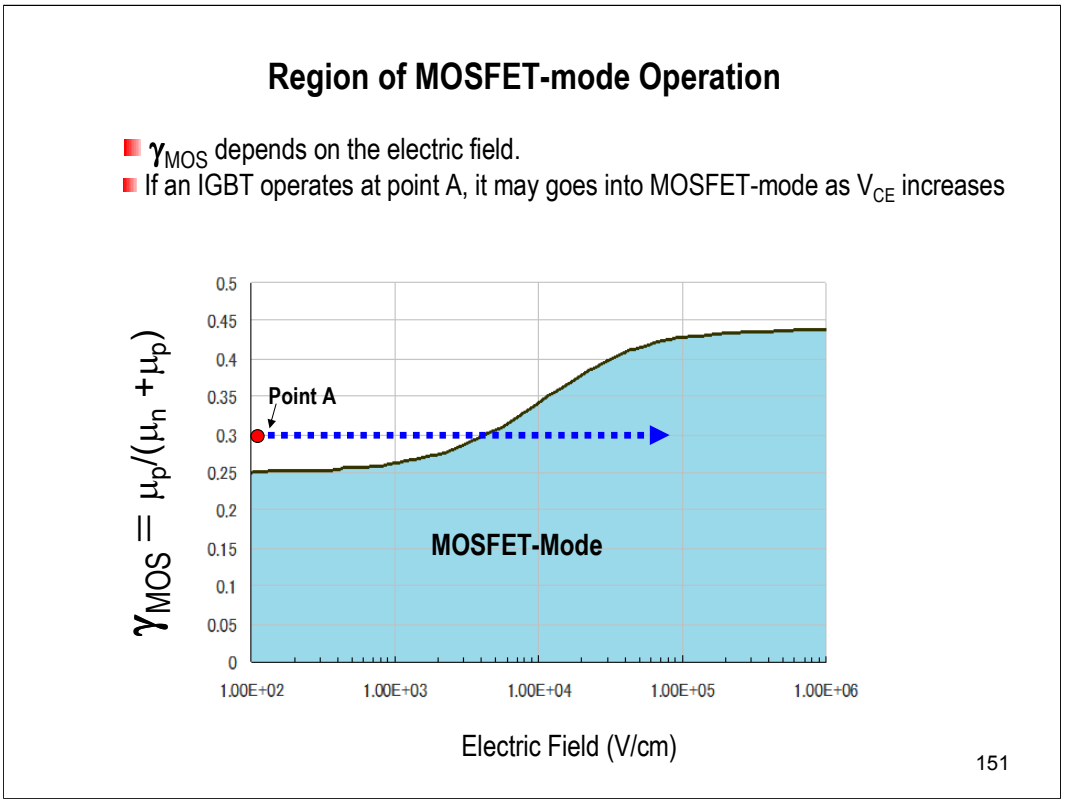
150

Another interesting characteristics for the MOSFET-mode operation is the waveforms of sustaining operation.

Sustaining waveforms of MOSFET-mode IGBTs are very similar to those of MOSFET.

In conventional IGBTs, the self-clamped voltage value increases as the drain current decreases.

On the contrary, the self-clamped voltage decreases as the drain current decreases in the MOSFET-mode operation.



This figure shows the region of the MOSFET-mode operation.

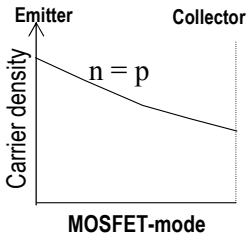
γ_{MOS} value depends on the electric field.

This is because the mobility μ is a function of the electric field.

If an IGBT operates at point A, it may goes into MOSFET-mode as VCE increases because the electric field inside the device increases.

MOSFET-mode operation in low voltage

- The figure illustrates the carrier density distribution in the MOSFET-mode operation
- Carrier density simply decreases from emitter toward the collector side.

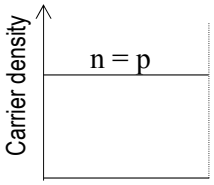


Definition of MOSFET-mode:

$$\frac{J_p}{J_n + J_p} \le \frac{\mu_p}{\mu_n + \mu_p}$$

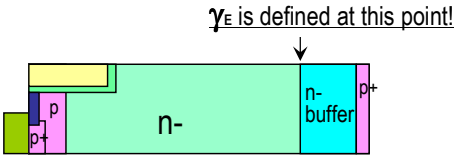
$$\Vert\!\Vert$$

$$\frac{J_n}{J_n + J_p} \ge \frac{\mu_n}{\mu_n + \mu_p}$$



$$\frac{J_n}{J_n + J_p} = \frac{\mu_n}{\mu_n + \mu_p}$$

The carrier distribution is similar to that of injection enhanced IGBTs.
IE effect is required to achieve low V_{CE} in MOSFET-mode operation.



152

This slide illustrates the carrier density distribution in the MOSFET-mode operation.

The carrier density simply decreases from the emitter to the n-buffer in the MOSFET-mode IGBTs.

The carrier distribution is similar to that of injection enhanced IGBTs.

Actually, IE effect is required to achieve low on-state voltage in MOSFET-mode operation.

Short-circuit SOA for MOSFET-mode IGBT

- From now, analyze short-circuit-SOA of MOSFET-mode IGBTs
- We assume that a high electric field appears in the entire n-base
- First, obtain Space charge density, ρ, and critical current density, J_{cri}, for SOA prediction.

◆ Space charge ρ is given, using saturation velocity v:

Anode efficiency γ is defined at the n-base n-buffer junction.

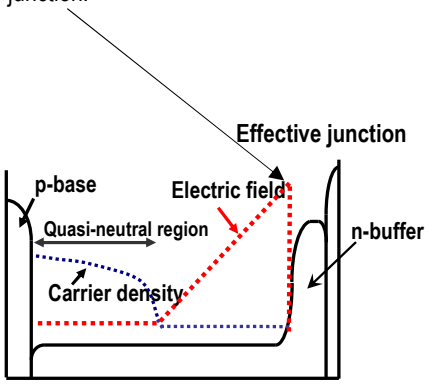
$$\begin{cases} \gamma \equiv \dfrac{J_p}{J} = \dfrac{J_p}{J_n + J_P} \\ p = \dfrac{J_p}{qv_h} = \dfrac{\gamma J}{qv_h} \qquad n = \dfrac{J_n}{qv_e} = \dfrac{(1-\gamma)J}{qv_e} \end{cases}$$

$$\rho = N_D + p - n = N_D + \frac{\gamma J}{qv_h} - \frac{(1-\gamma)J}{qv_e}$$

$$= N_D + \frac{J(v_h + v_e)}{qv_h v_e}(\gamma - \gamma_{MOS}) \qquad \text{...Eq.(1)}$$

$$\gamma_{MOS} = \frac{v_h}{v_h + v_e} \cong 0.43$$

$$\gamma - \gamma_{MOS} < 0 \quad \text{in MOSFET-mode operation}$$



153

We analyze short-circuit SOA of MOSFET-mode IGBT.

We assume MOSFET-mode IGBT is short-circuited and high electric field appears in the entire n-base

First, we obtain space charge density, ρ, and critical current J_{cri}.

Anode efficiency, γ, is defined as the ratio of the hole current over the total current at the n-base and n-buffer junction.

The electron and hole density in high field region is given by this equation, using the saturation velocities, v_e and v_h.

ρ is given by the Eq.: $\rho = N_D + p + n$……..

In MOSFET-mode operation, from the definition of MOSFET-mode, the second term of Eq.(1) is negative

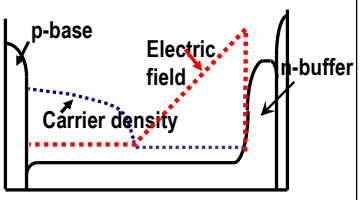
Cite Eq.(1), again.

$$\rho = N_D + \frac{J(v_h + v_e)}{q v_h v_e}(\gamma - \gamma_{MOS}) \qquad \text{...Eq.(1)}$$

$\gamma - \gamma_{MOS}$ = negative in MOSFET - mode operation

◆ Critical current density, J_{cri} is defined by setting **Eq.(1)=0**, and solving for J:

$$J_{cri} = \frac{qN_D}{\gamma_{MOS} - \gamma} \frac{v_h v_e}{v_h + v_e}$$



When $J = J_{cri}$, Space charge $\rho = 0$

When $J > J_{cri}$, Space charge ρ become negative,

⟶ A high electric field appears in the N-base-N-buffer junction

154

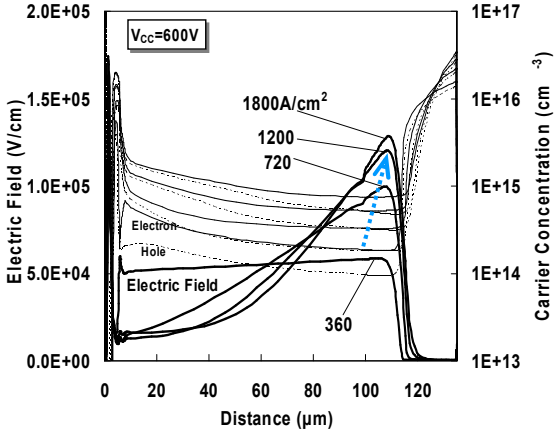
We can define critical current density, J_{cri}, by setting Eq.(1)=0 and solve for J.

When J = J_{cri}, the space charge, ρ, is zero and the electric field is uniform in the n-base.

When J exceeds Jcri, the space charge, ρ, becomes negative, a high electric field appears in the N-base-N-buffer junction.

TCAD Simulation & Analytical Diode Model

- The left figure shows the simulated electric field of MOSFET-mode IGBTs when V_{CE}=600V.
- Electric field at N-base N-buffer monotonically increases as the current density increases
- The electric field can be calculated, using the diode model in the right figure.
- The point is to set **impurity density of high resistance layer = Eq.(1).**



Diode model

Impurity concentaration

$$\rho = N_D + \frac{J(v_h + v_e)}{q v_h v_e}(\gamma - \gamma_{MOS})$$

...Eq.(1)

155

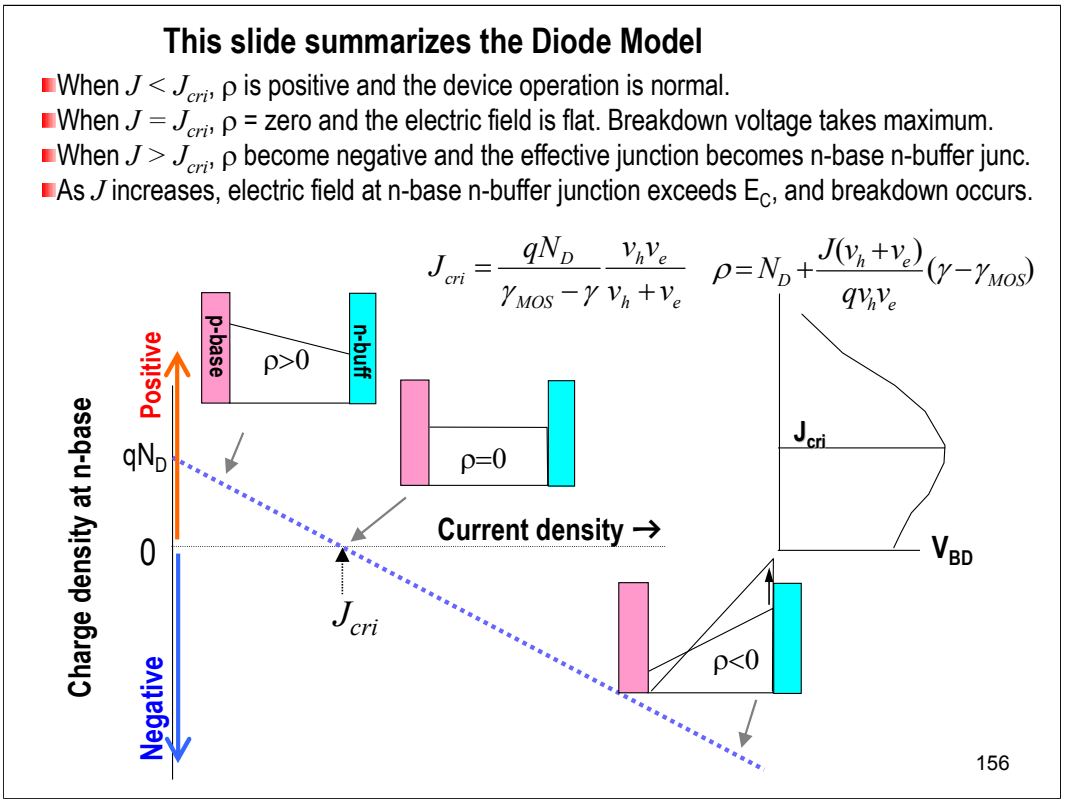
The left figure shows the simulated electric field distribution of 1200V MOSFET-mode IGBTs when 600V V_{CE} is applied.

The electric field is almost flat when the current density is 360A/cm^2.

The electric field at the N-base-N-buffer junction monotonically increases as the current density increases.

The electric field distribution can be analytically calculated, using a simple diode model of Fig.(2).

The point is that the impurity concentration of the high resistance layer is given by Eq.(1).



This slide summarizes the diode model.

When the current density $\boldsymbol{J < J_{cri}}$, n-base charge, ρ, is positive and the device operation is normal.

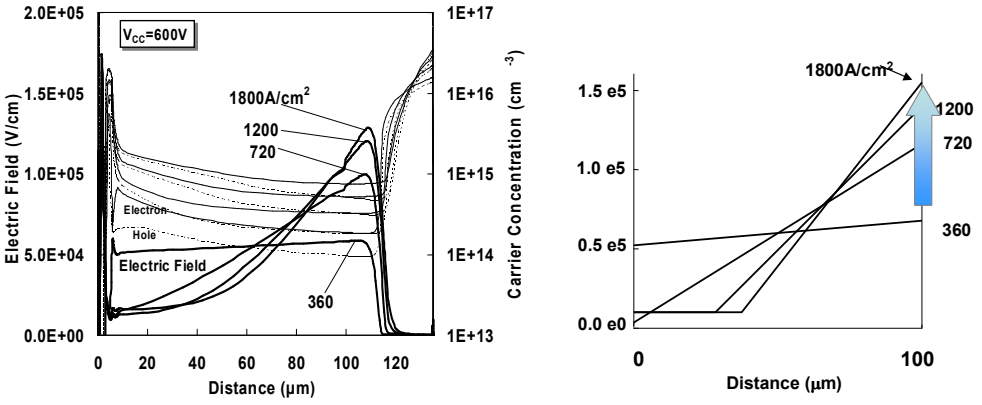
When $\boldsymbol{J = J_{cri}}$, the n-base charge becomes zero and the electric field is flat. The device breakdown voltage takes the highest value.

When $\boldsymbol{J > J_{cri}}$, the n-base space charge become negative and the effective junction become the n-base n-buffer junction.

As the current density further increases, the peak electric field at the N-base N-buffer junction increases monotonically, and the device breakdown will take place.

Electric field can be calculated by the Diode Model

- This figure compares TCAD results and the analytical Diode Model.
- The agreement between TCAD & Diode Model is quite satisfactory, although it's simple.
- The Diode Model can be used to predict the safe operating area of MOSFET-mode IGBTs.



Simulated **Analytical Diode Model**

157

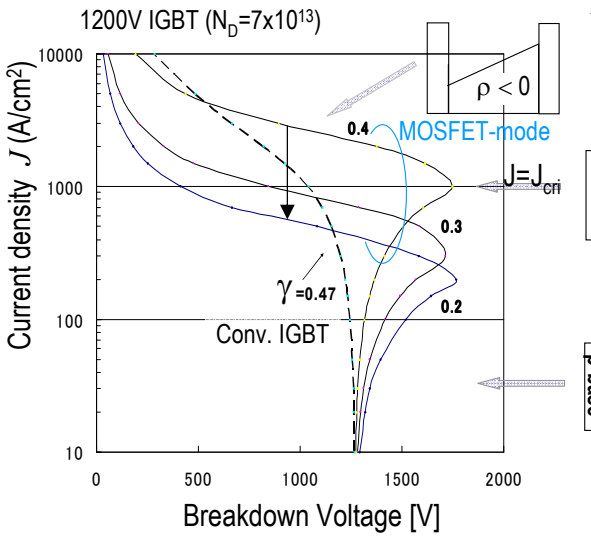
This figure compares the TCAD results and the analytical diode model.

The agreement between the two models is quite satisfactory, if we consider the simplicity of the model.

Analytical model can be used to predict the safe operating area of MOSFET-mode IGBTs.

SOA can be predicted by the Diode Model

- Breakdown voltage of the Diode is plotted as a function of current density.
- As anode efficiency γ decreases, allowable maximum current density decreases.
- Theoretically, maximum current density becomes infinite, when $\gamma = \gamma_{MOS}$ (=0.43).



Anode efficiency $\gamma \equiv \frac{J_p}{Total\ J} = \frac{J_p}{J_n + J_P}$

$\rho = N_D + \frac{J(v_h + v_e)}{q v_h v_e}(\gamma - \gamma_{MOS})$

158

SOA can be predicted by the diode model.

This figure shows the calculated safe operating area.

The breakdown voltage of the diode is plotted as a function of current density.

This relation can be regarded as SOA.

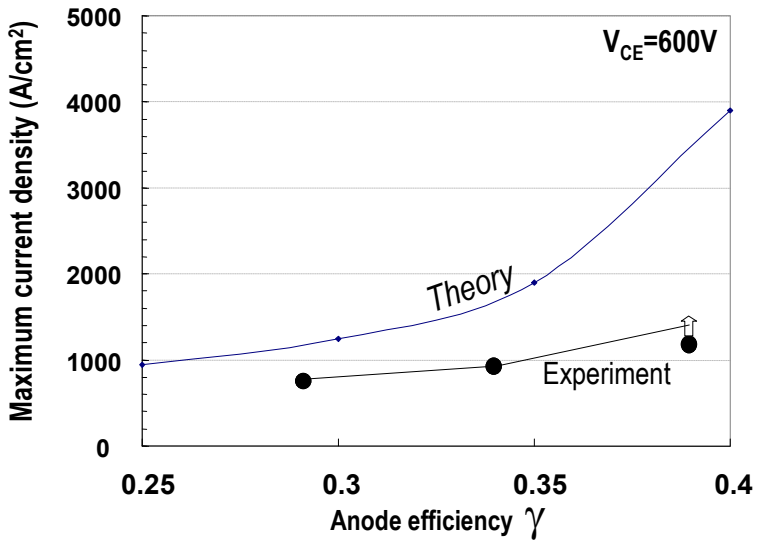
As anode efficiency γ decreases, allowable maximum current decreases.

Theoretically, maximum current becomes infinite, when γ is equal to 0.43.

The broken line shows the typical SOA of conventional PT-IGBTs.

Short-Circuit SOA

- This figure compares theory and experiments.
- The maximum current density for V_{CE}=600V decreases as γ decreases.
- The same tendency was confirmed by experiments.



159

This figure compares theory and experiments.

The maximum current density decreases as γ decreases.

The same tendency was confirmed by experiments.

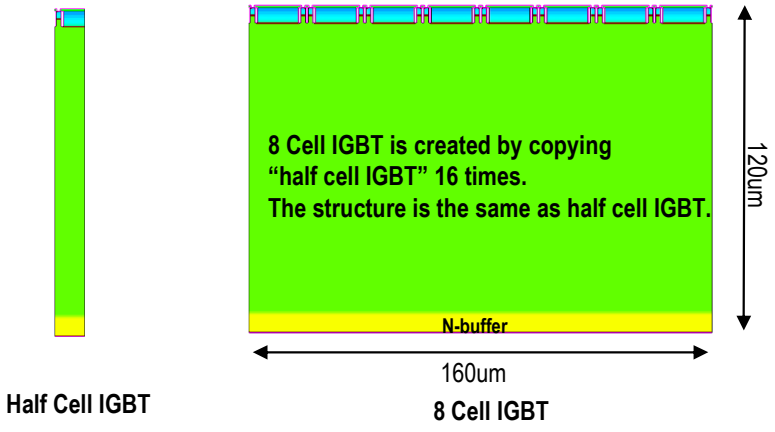
This is one of the direct verification of the present theory.

Destruction mechanism of MOSFET-mode IGBTs

160

TCAD analysis of MOSFET-mode IGBT

- It was experimentally found that short-circuit SOA of MOSFET-mode IGBTs is deteriorated.
- Detailed investigation will be shown, using large scale TCAD simulation.
- Half cell IGBT and 8 cell IGBT are simulated and compared.
- P-collector and N-buffer are set such that the anode efficiency is less than 0.27
- Self-heating is considered.
- Thermal resistance of 0.3K/W is set between collector electrode and heat-sink



161

It was experimentally observed that short-circuit capability is deteriorated in MOSFET-mode IGBTs.

I’ll shows the more detailed analysis, using large scale TCAD simulations.

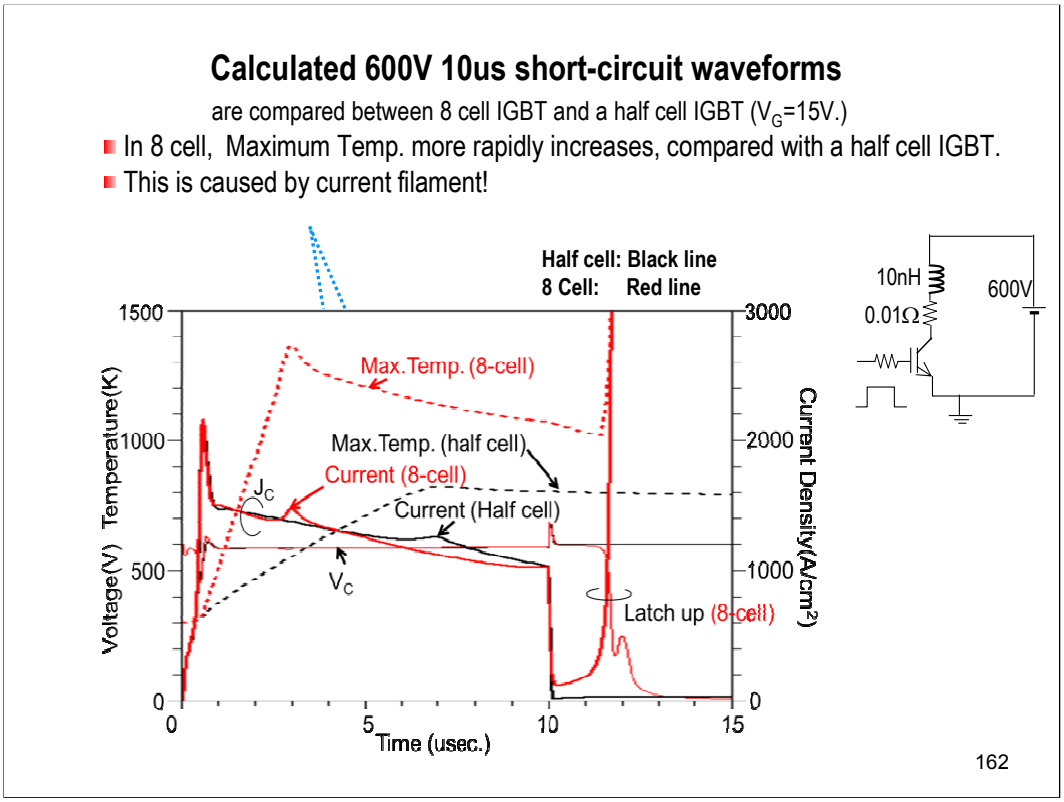
Half cell IGBT and 8-cell IGBT are simulated and compared. In 8 cell IGBT, doping profiles and the MOS channel structures are completely homogeneous.

The N-base thickness is 120um. The lateral device size of 8 cell IGBT is 160um.

The P-collector and the N-buffer are set such that the γ is less than 0.27 at the rated current density to realize MOSFET-Mode operation.

Device simulation is performed with taking into account self-heating. The thermal resistance of 0.3K/W is set between the anode electrode and the heat-sink.

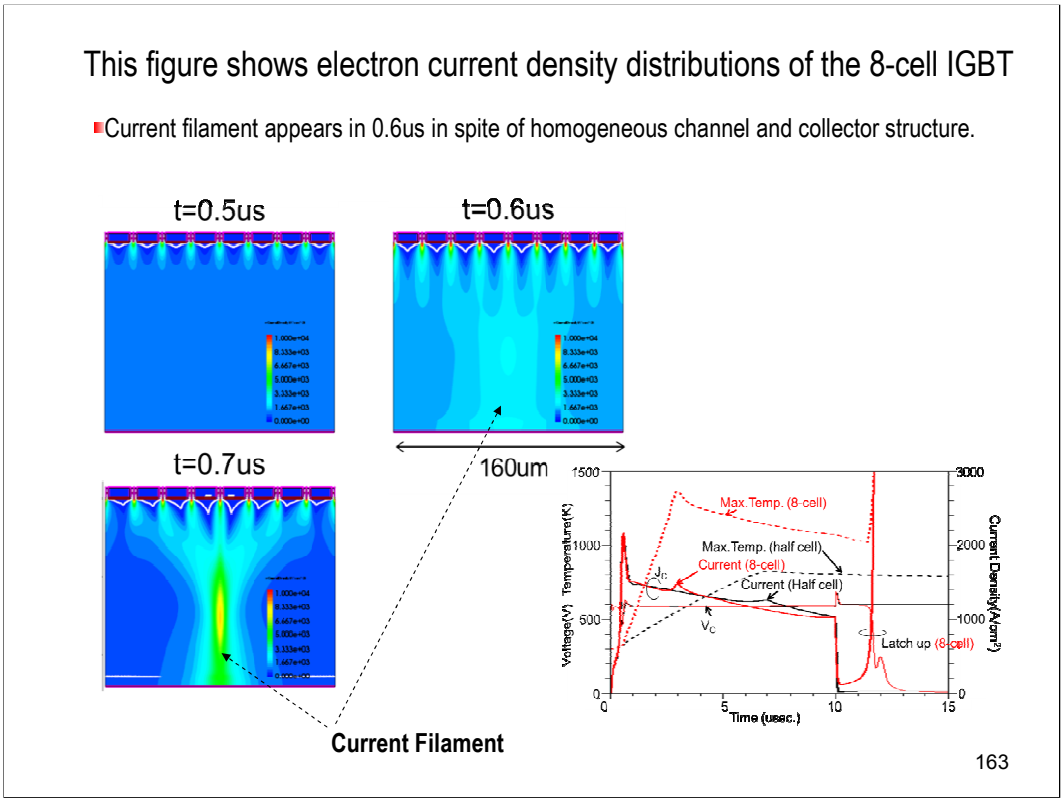
The heat-sink temperature is set at 300K.



This figure compares the 600V 10us short-circuit waveforms between 8 cell IGBT and a half cell IGBT when VG=15V.

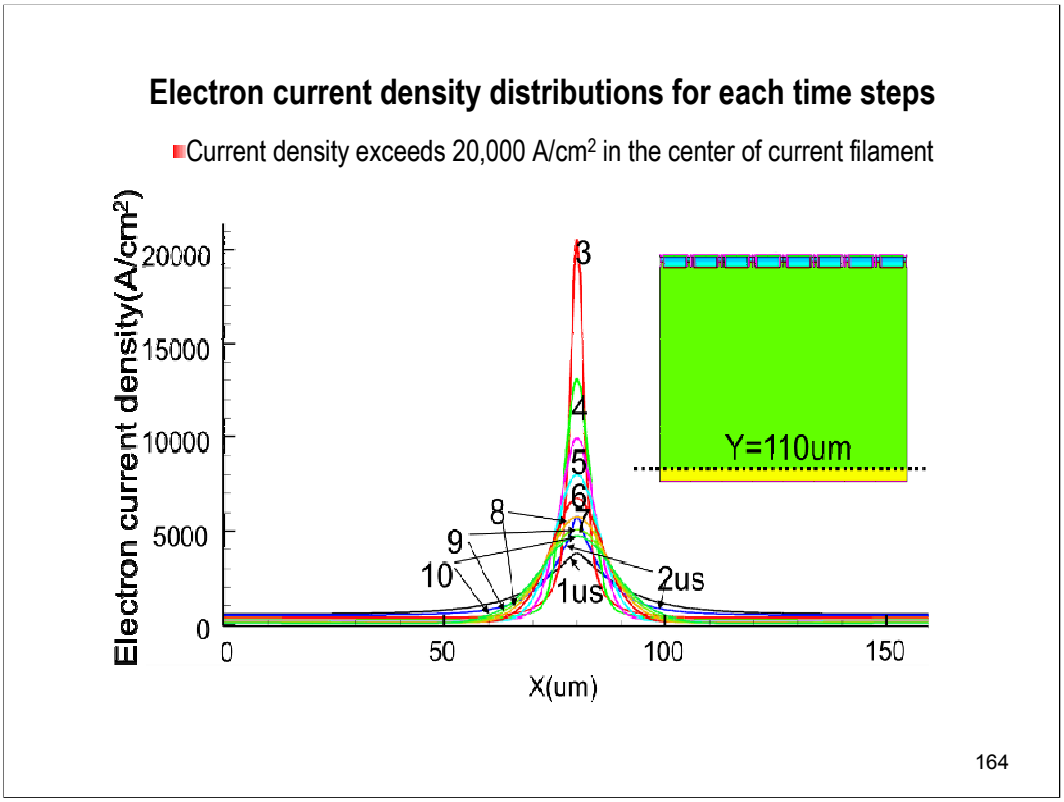
In the 8 cell, the maximum temperature more rapidly increases, compared with the result of a half cell IGBT.

This is caused by current filament.



This figure shows electron current density distributions of the 8-cell structure for the time steps 0.5us, 0.6us and 0.7us when V_G=15V.

Current filament appears in 0.6us in spite of homogeneous channel and anode structure.

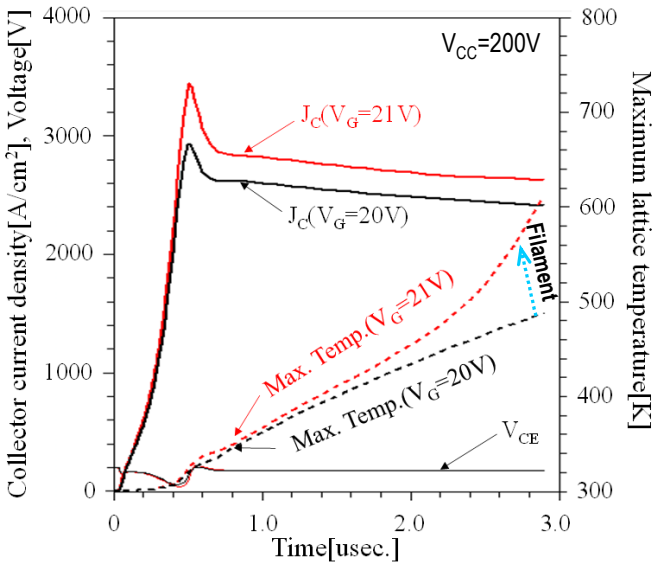


This figure shows the electron current density distributions for every micro seconds.

It is seen that the current density exceeds 20,000A/cm^2 in the center of current filament.

Filament appears suddenly when V_G exceeds a certain value.

- Max. Temp. rises linearly when V_G=20V.
- Max. Temp. begins to increase rapidly when V_G=21. This is because filament appears.



165

It was found that current filament appears suddenly when the gate voltage exceeds a certain value.

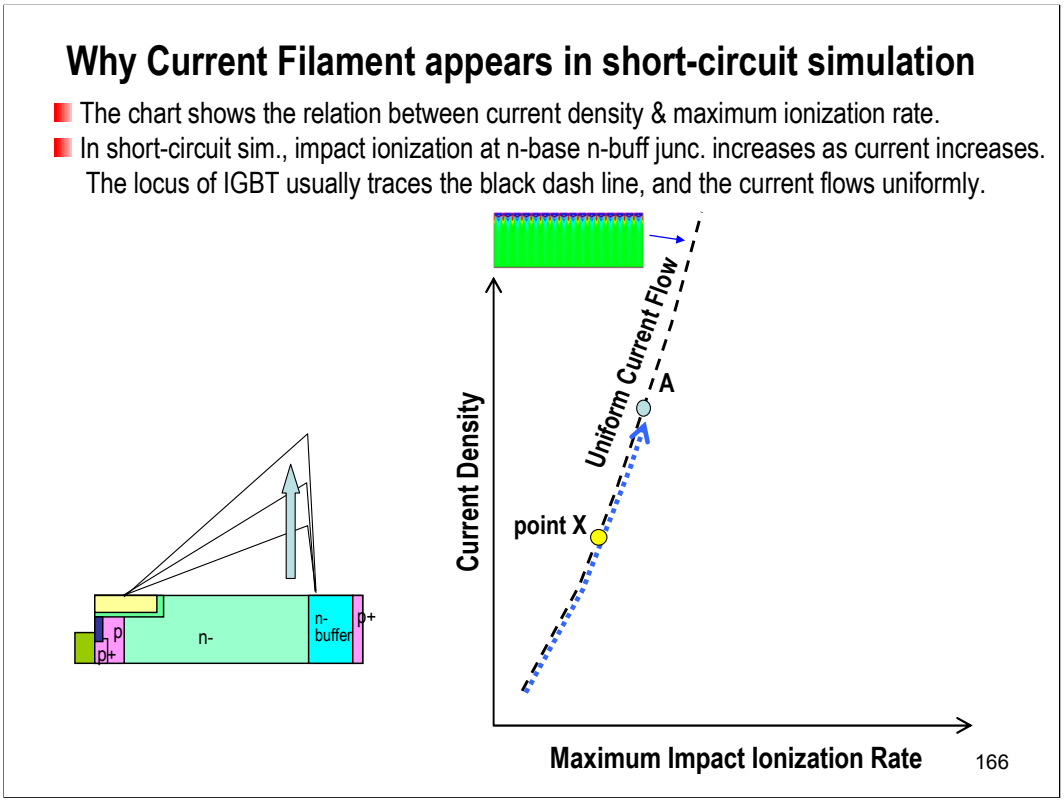
In this case, the current filament appears when the gate voltage exceeds 20V.

The maximum temperature rises linearly when V_G=20V.

However, the maximum temperature begins to increase more rapidly when V_G=21.

This happens because current filament appears.

We shows why current filament appears in the next slide.



We show why current filament appears in short-circuit simulation, using this chart.

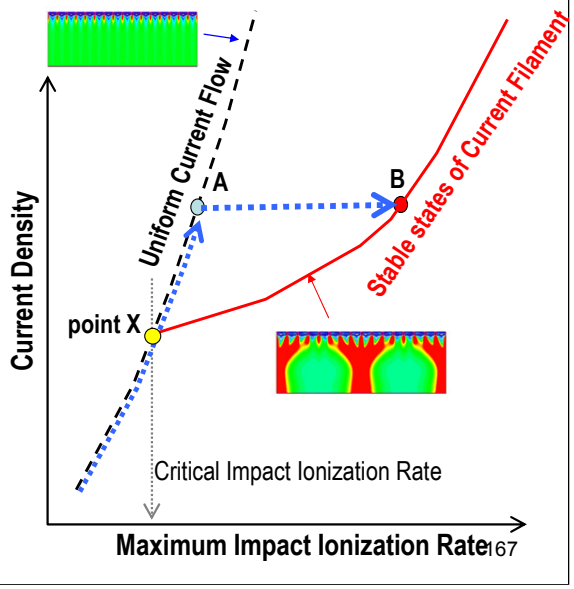
The chart shows the relation between the collector current density and the maximum impact ionization rate inside the device.

In short-circuit simulation of MOSFET-mode IGBTs, as the collector current increases, impact ionization at the N-base N-buffer junction increases. The locus of IGBT in the simulation, usually traces the black line.

Why Current Filament appears in short-circuit simulation -2

- We found that **stable states of current filaments** (red line) exist beyond point X
- If locus of IGBT crosses point X, the locus moves from A to B and filament appears.
- Below point X, there is no state of filament, thus, current flows uniformly.

- Current filament appears, because filament state is more stable.
- If current filament appears, impact ioniz. rate within current filament increases enormously, the locus of IGBT moves from A to B.



167

We found that stable states of current filaments (red line) exist beyond point X. Below point X, there is no state of filament, thus, current flow uniformly

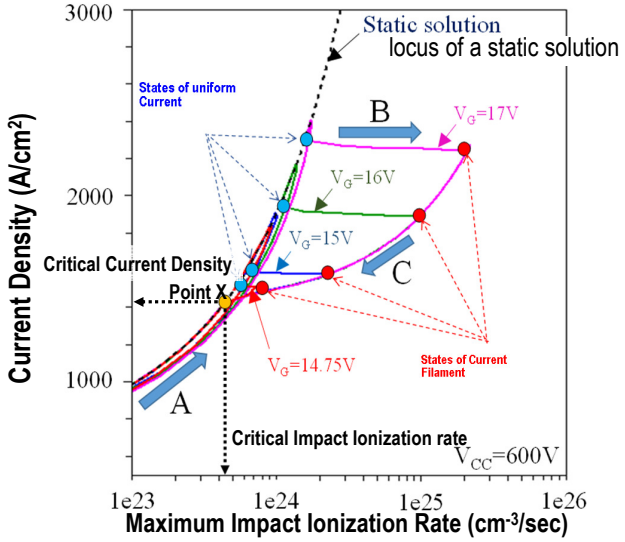
If the locus of IGBT crosses point X, current filament appears, because there exist stable states of current filament, the locus of IGBT moves horizontally from A to B.

If the current filament appears, the current density increases enormously within the current filament. The impact ionization rate within the current filament increases very much, although the device collector current does not change very much.

Thus, the system locus moves horizontally, from A to B, when a current filament appears.

Actual locus when short-circuit simulations are performed

- Solid line shows the locus, which IGBT traces when we execute the short-circuit simulations.
- When V_G is 17V, for example, in short-circuit sim., IGBT traces the locus of the pink line.
- If the locus exceeds **critical current density** or **critical impact ionization rate** of point X, current filament appears and device failure occurs.
- Critical current density or critical impact ionization rate is defined as the value at point X.



168

This shows actual locus when short-circuit simulations are performed.

The vertical axis shows current density.

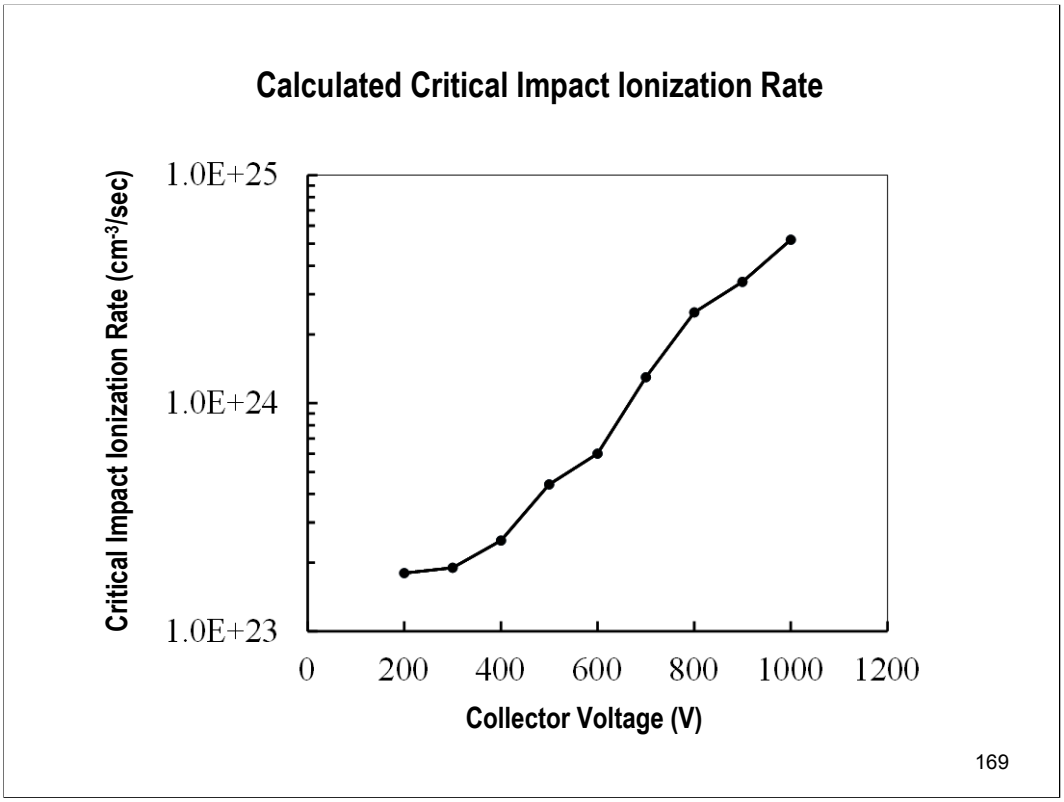
The horizontal axis shows the maximum impact ionization rate inside the device.

The solid lines show the locus, which IGBT traces when we execute the short-circuit simulations.

For example, when we switch-on the IGBT without any loads and raises the gate voltage to 17V, the locus of IGBT traces the pink line.

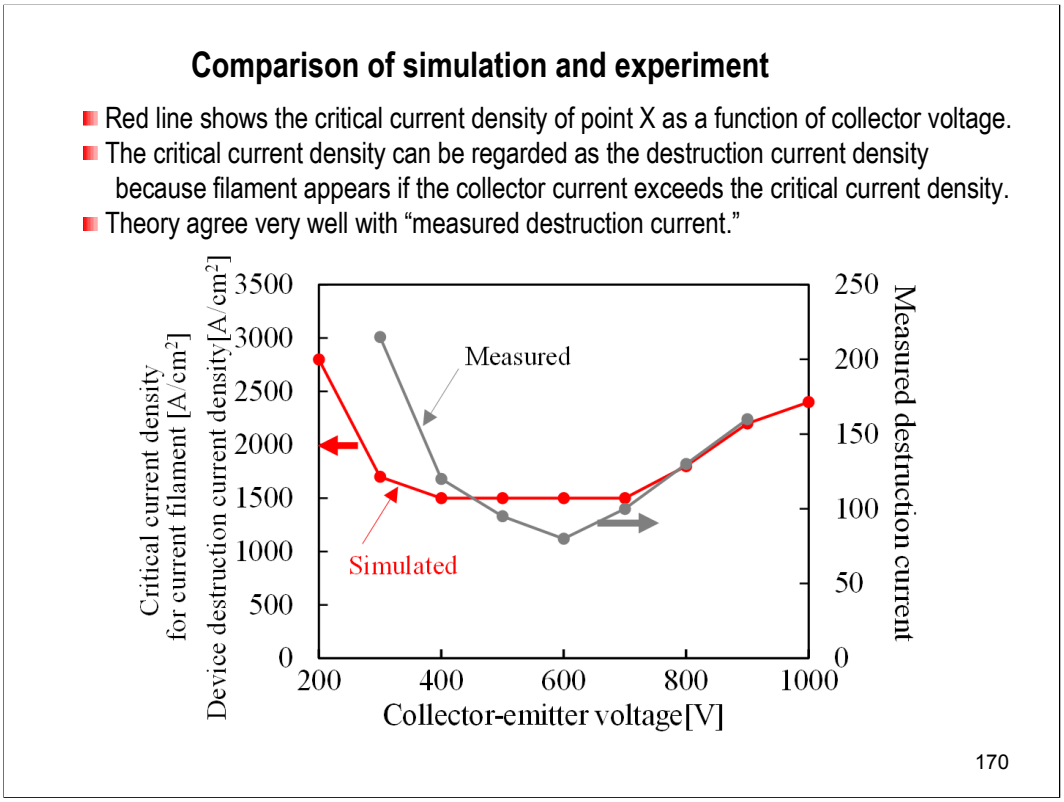
If the locus exceeds critical current density or critical impact ionization rate of the point X, current filament appears and device failure occurs.

Critical current density or critical impact ionization rate is defined at point X



This figure shows the critical impact ionization rate.

If the impact ionization rate exceeds this critical value, current filaments appear.



This figure compares the simulation and experiment.

The red line shows the critical current density as a function of collector-emitter voltage.

When the collector current density exceeds the critical current density, current filament appears.

Once the current filament appears, device destruction immediately occurs.

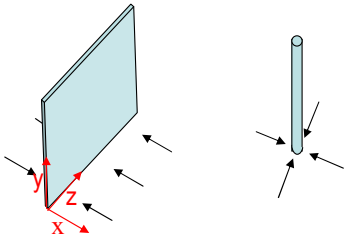
So the critical current density can be regarded as the destruction current density.

The black line shows the experimental results of the measured destruction current as a function of collector voltage.

The two lines well agree with each other.

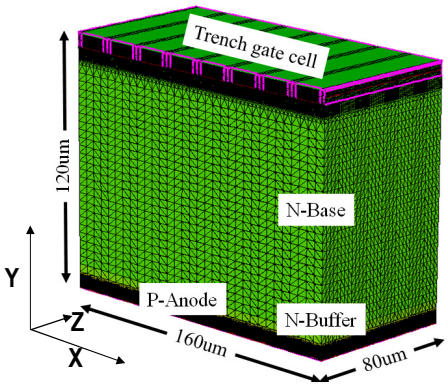
Three Dimensional Simulation of Short-Circuit

- In 2D simulation, a current filament occurs in a line.
- In actual cases, current filament occurs in a point.
- In order to exactly investigate current filament, 3D simulations were executed.



2 dimensional

3 dimensional



3-d device structure for 3-d simulation
Device size in Z-direction is 80um.

171

In the previous simulations, we used a two dimensional simulator.

In two dimensional simulators, we solve device equations in only x and y coordinates. So, everything in z-direction is the same and homogeneous.

In two dimensional simulator, a current filament occurs in a line, as shown in the left figure.

But, in actual cases, current filament occurs three dimensionally, or in a point, as is shown in the figure on the right hand side.

In order to exactly investigate current filament, 3D-simulations are performed.

On the right hand side, I shows the 3-d device structure, we used for 3-d simulation.

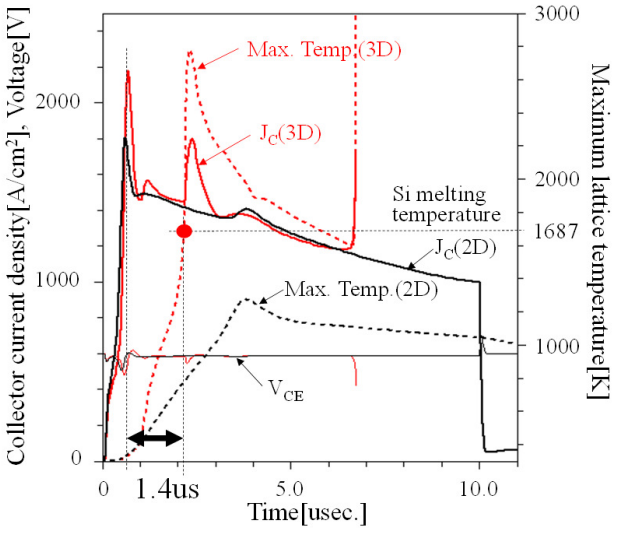
The structure of the device is the same as 2-d case.

The depth of the device is assumed to be 80um.

This is the typical meshes we used in the simulations.

Comparison of Three and Two Dimensional Simulations

- The maximum temperature increases quite rapidly in the 3-d calculations.
- The temperature reaches the melting point in only 1.4 usec after current filament appears.
- Device destruction will occur approximately in 1.5usec. once a current filament appears.



172

This figure compares Two and Three Dimensional Simulations

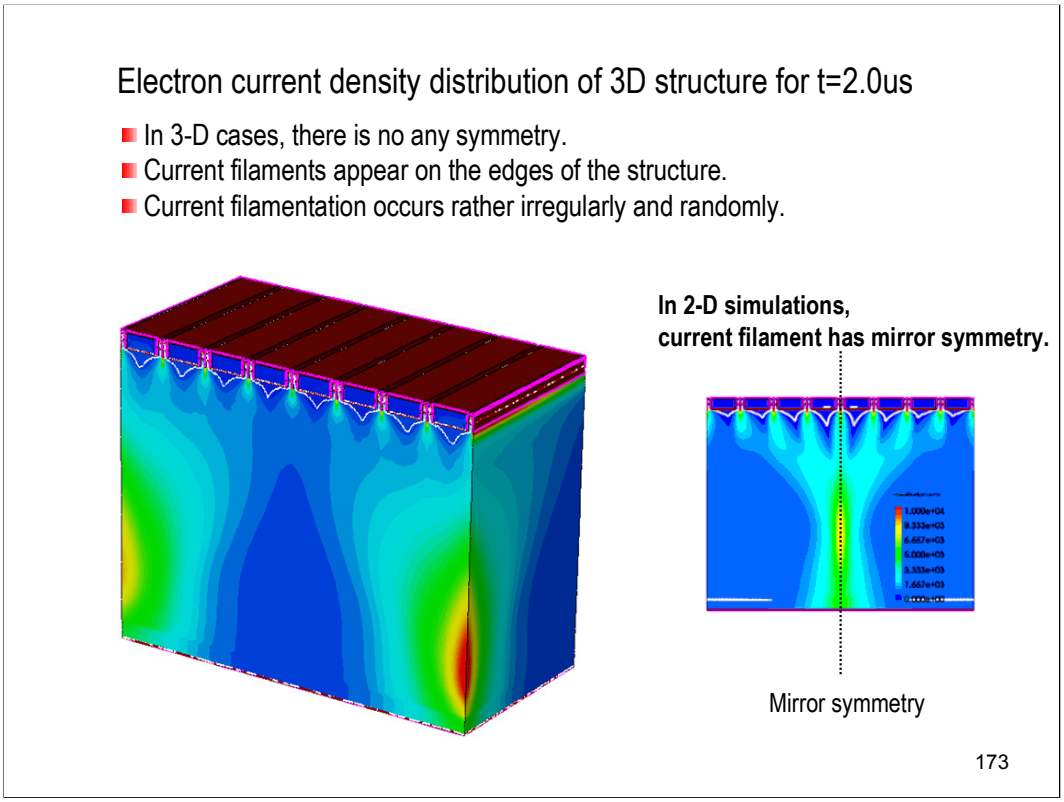
The red line shows the 3-D results.

And the black lines shows the 2-D cases.

Current filamentation occurs in 3-dimensionally, and the maximum temperature increases quite rapidly in the 3-D calculations.

The temperature reaches the silicon melting temperature of 1687K in only 1.4 usec after the current filament appears.

This suggests that the device destruction will occur approximately in 1.5usec. once a current filament appears.



This figure shows the electron current density distribution of the 3D structure for t=2.0us.

It is seen that current filaments appear on the edges of the structure.

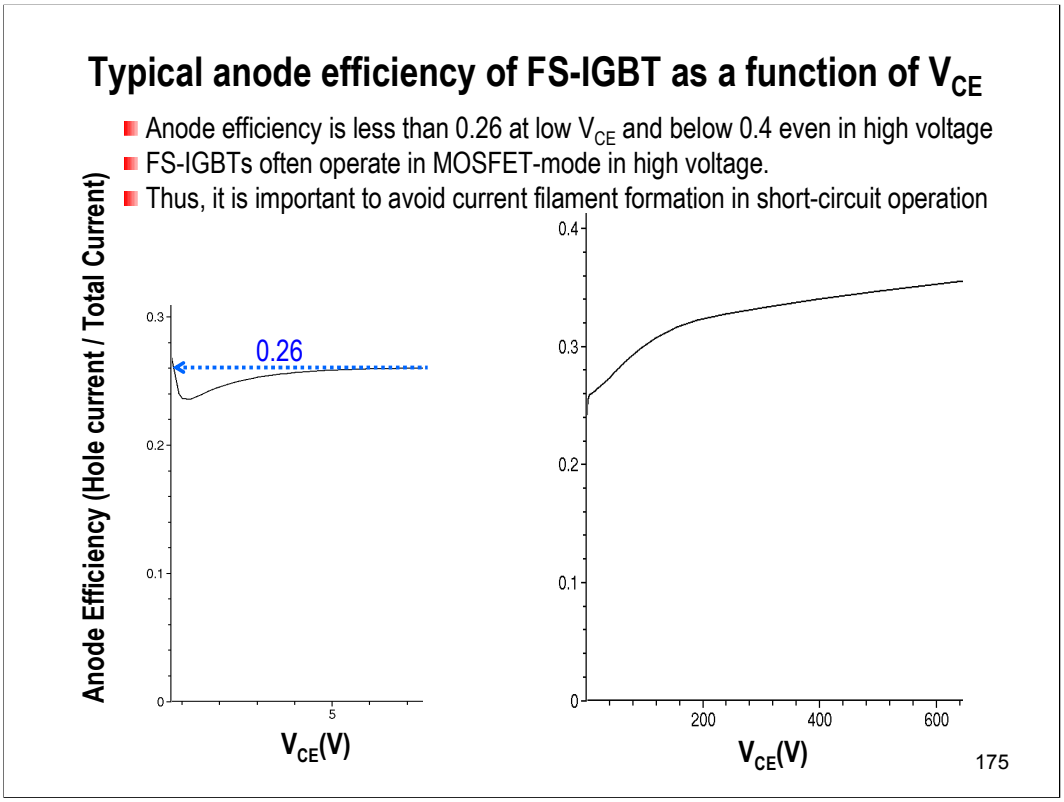
In two dimensional simulations, current distribution is in mirror symmetry.

The symmetry axis is a line crossing the center of the device.

However, in 3-d cases, there is no any exact symmetry.

It seems, in 3-d cases, that current concentration occurs rather irregularly and randomly.

7. Design Issue of FS IGBT



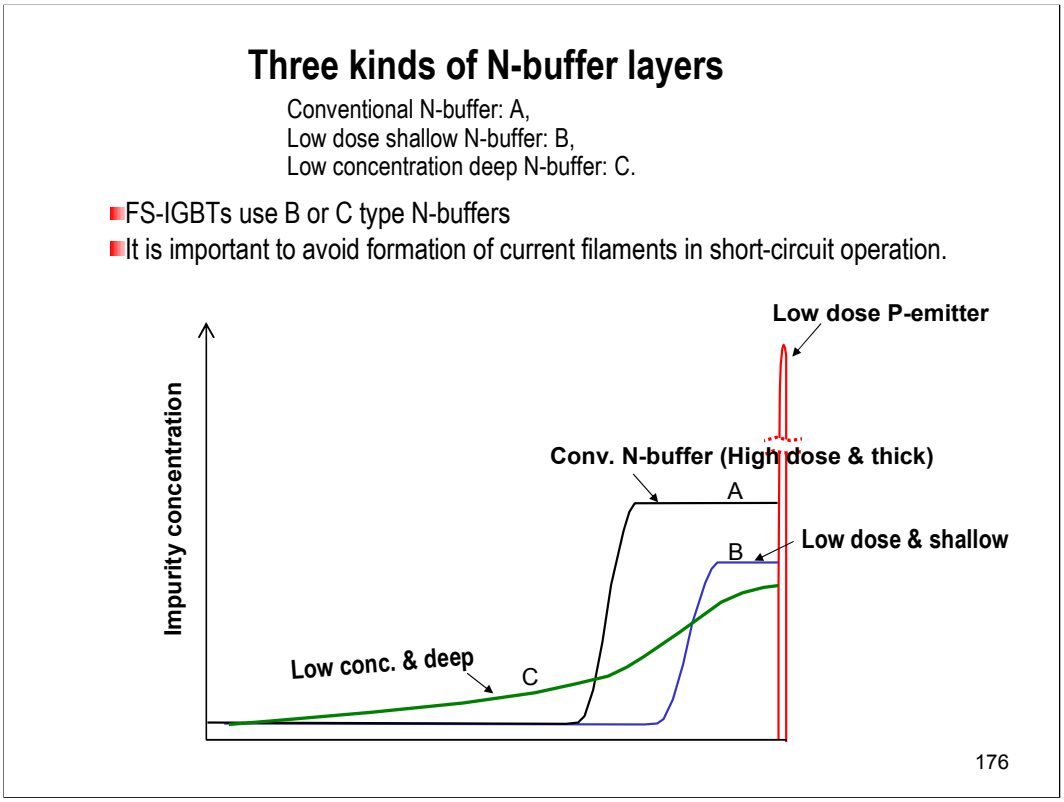
This figure shows typical anode efficiency of an FS-IGBT as function of V_{CE}.

The anode efficiency in low voltage range is 0.26.

In high voltage region, the anode efficiency is still less than 0.4.

This means that FS-IGBT operates in MOSFET-mode at least in high voltage region.

So, it is very important to avoid current filament formation in short-circuit operation.



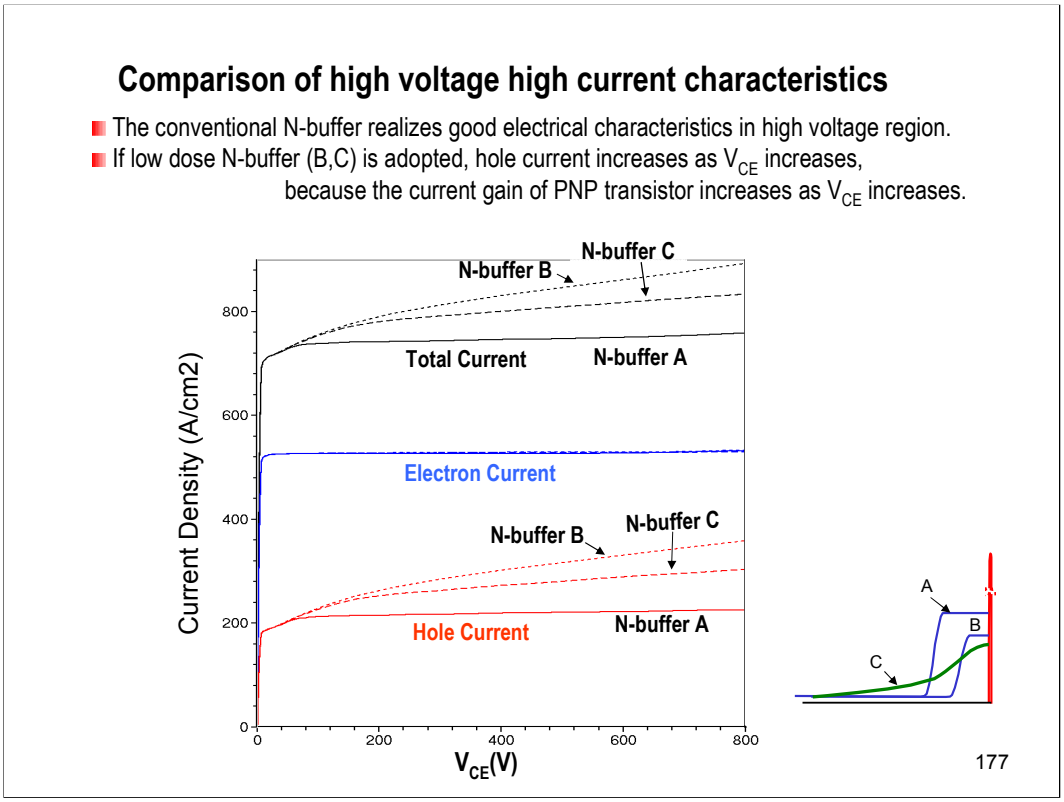
This figure shows three typical N-buffers: (A) conventional N-buffer, (B) low dose shallow N-buffer, and (C) low concentration deep N-buffer.

Conventional N-buffer (A) has high impurity concentration and wide n-buffer.

On the other hand, the total dose of N-buffer B and C is relatively small.

FS-IGBTs use B or C type of N-buffer layers.

In FS-IGBTs, it is very important to avoid the formation of current filaments in load short-circuit operation.



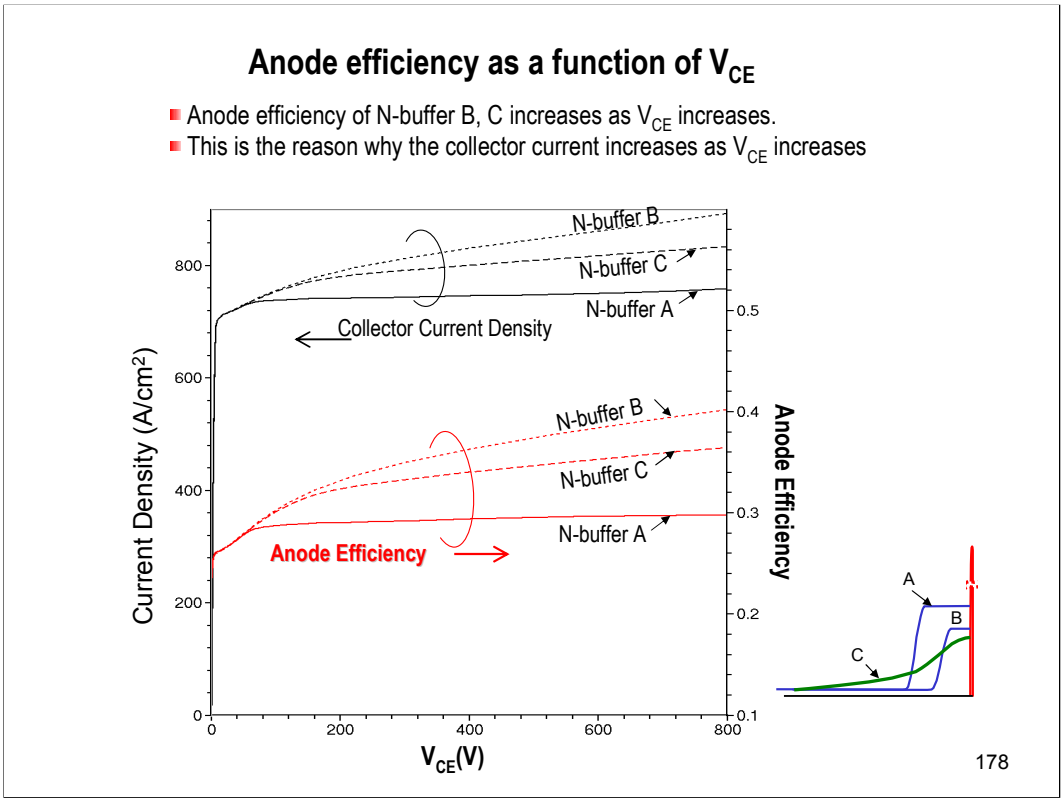
This figure compares high current high voltage characteristics of three IGBTs, which have the three different N-buffers, A, B, and C.

The conventional N-Buffer realizes very good electrical characteristics in high voltage region.

The collector current dose not depend on the collector voltage, and the collector current is almost the same even if the collector voltage is increased.

On the other hand, if the low dose N-buffer is adopted, hole current increases as the collector voltage increases.

This is because the current gain of the PNP transistor increases as the collector voltage increases.



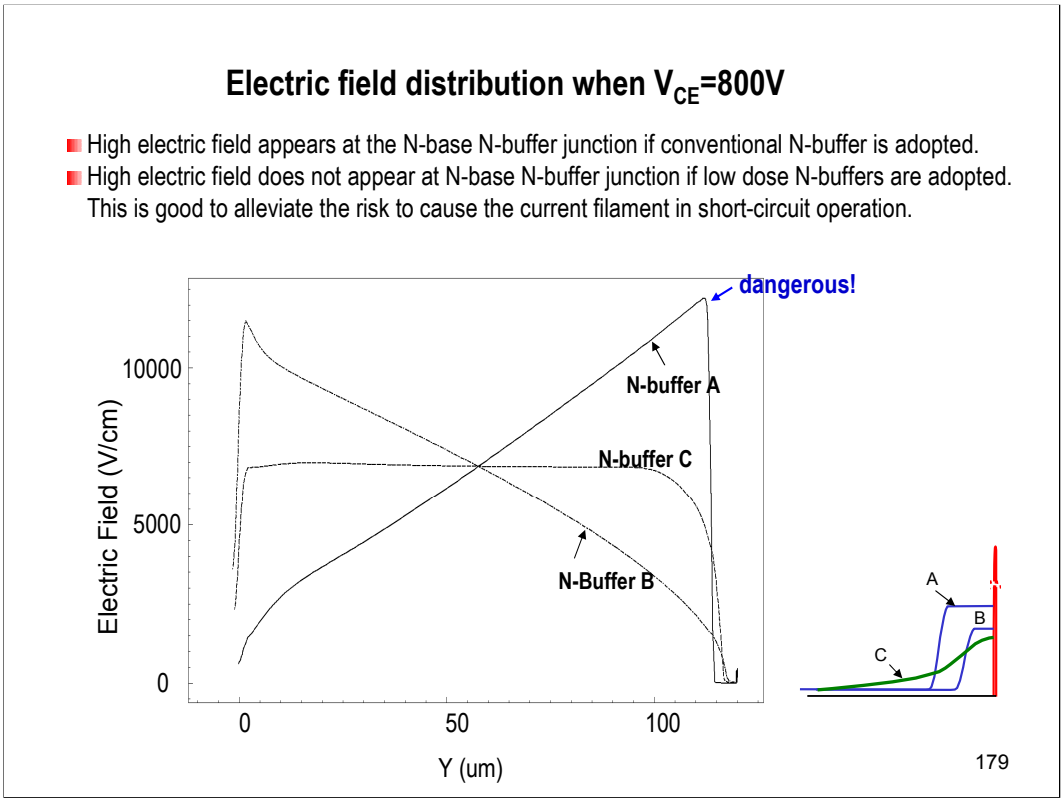
This figure shows the anode efficiency as a function of V_{CE}.

In low dose N-buffer, anode efficiency increases as the applied voltage increases.

The definition of the anode efficiency is the ratio of hole current over total current at the N-base N-buffer junction.

The anode efficiency coincide with the PNP transistor gain when the whole N-base is depleted.

The increase in anode efficiency is the reason why the collector current increases as the V_{CE} increases.



This figure shows the electric field distribution when V_{CE} is 800V.

High electric field appears at the N-base-N-buffer junction in IGBTs if a conventional N-buffer is adopted.

However, a high electric field at the N-base N-buffer junction does not appear in FS-IGBTs if low dose N-buffers are adopted.

This is good to alleviate the risk to cause the current filament in short-circuit operation.

This is one of the reasons why we use low dose N-buffers in FS-IGBTs.

8. IGBT Cell Pattern Design

180

In this section, I talk about how to increase latch-up current density when one execute actual IGBT design.

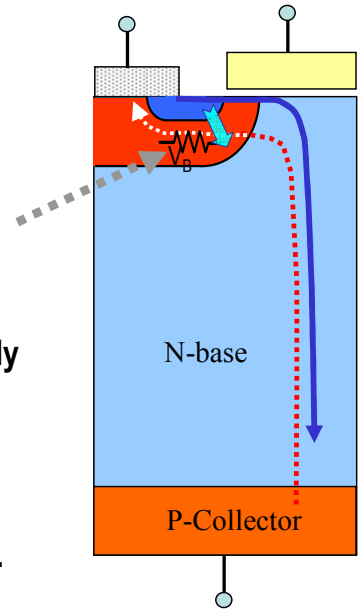
How to suppress latch-up in planar IGBTs

Underlying philosophy is still good for modern IGBT design.

Latch-Up:

Hole current flows in the P-base. When voltage drop in the p-base exceeds the built-in voltage of PN junction, electrons are directly injected into P-base, resulting in thyristor action.

Once this occurs, the current cannot be controlled by the gate.

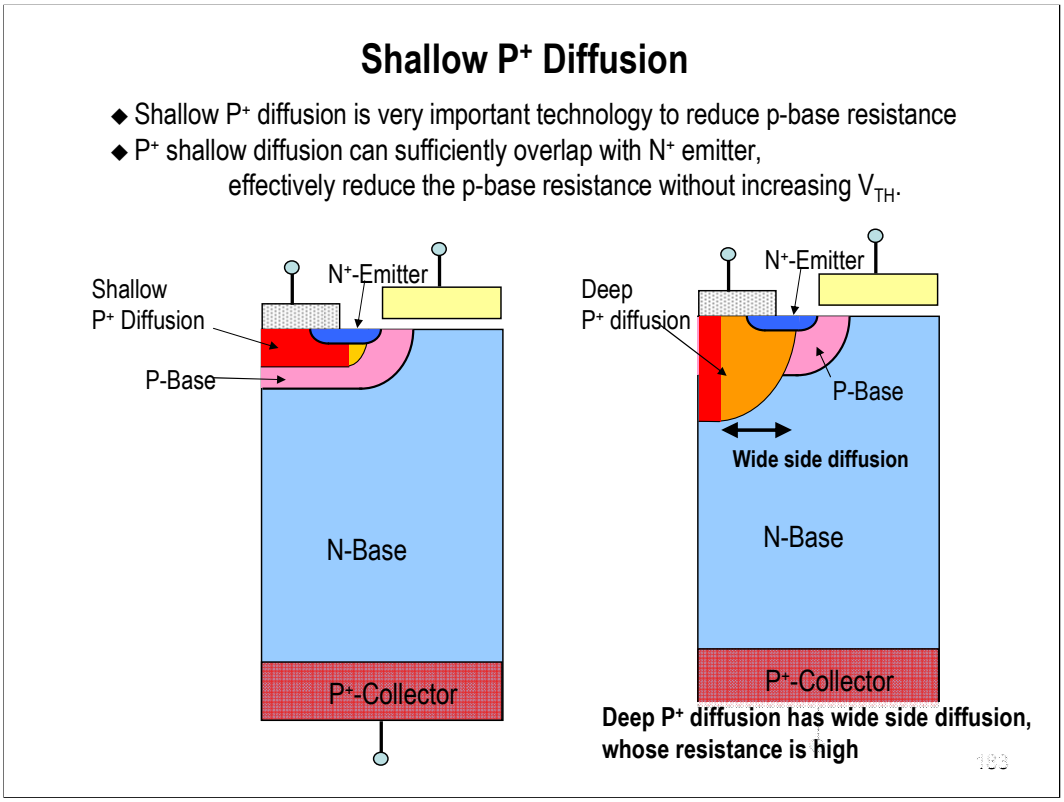


181

The ideas to suppress latch-up in IGBTs are:

(1) Decrease P-base resistance under the N-emitter

(2) Suppress the local hole current crowding into a small number of cells.



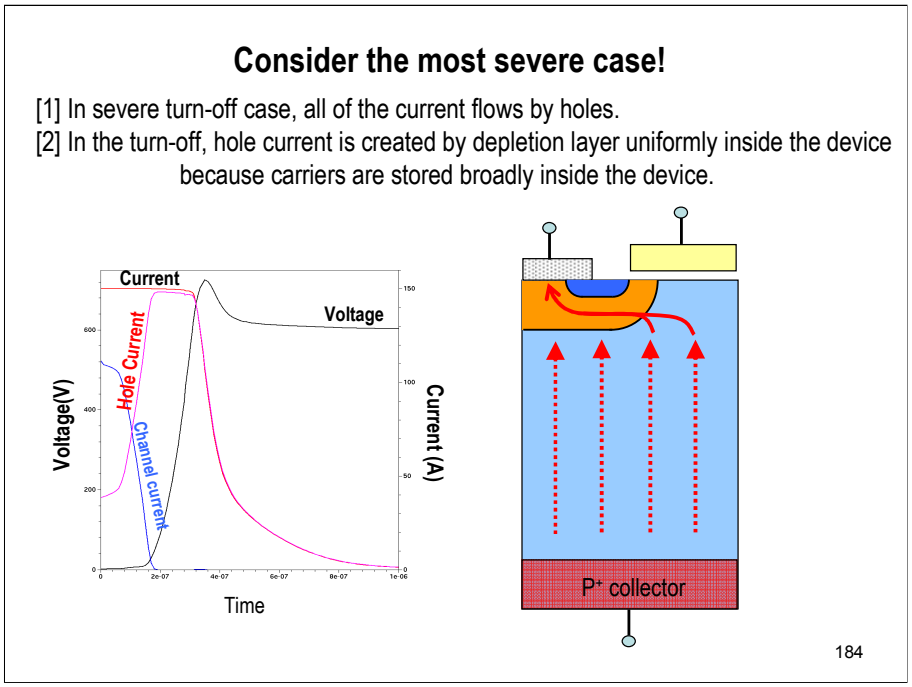
Shallow P^+ diffusion is very important technology to reduce the p-base resistance under the N^+ source.

People tend to use a deep P+ diffusion to reduce the p-base resistance, as shown in the figure on the right hand side.

However, a deep P^+ diffusion has to have a wide side diffusion, which is the only portion that can diffuse under the N^+ source without increasing the threshold.

Impurity concentration of the side diffusion is low and thus, cannot decrease the p-base resistance effectively.

On the contrary, the shallow P^+ diffusion has a narrow side diffusion region, and thus, the P^+ shallow diffusion can sufficiently overlap with N^+ emitter, and thus, effectively reduce the p-base resistance under the N^+ layer.



In executing IGBT design, it is important to consider the following conditions:

[1] In severe turn-off case, all of the current flows by holes.

[2] In the turn-off, hole current is created by depletion layer uniformly inside the device

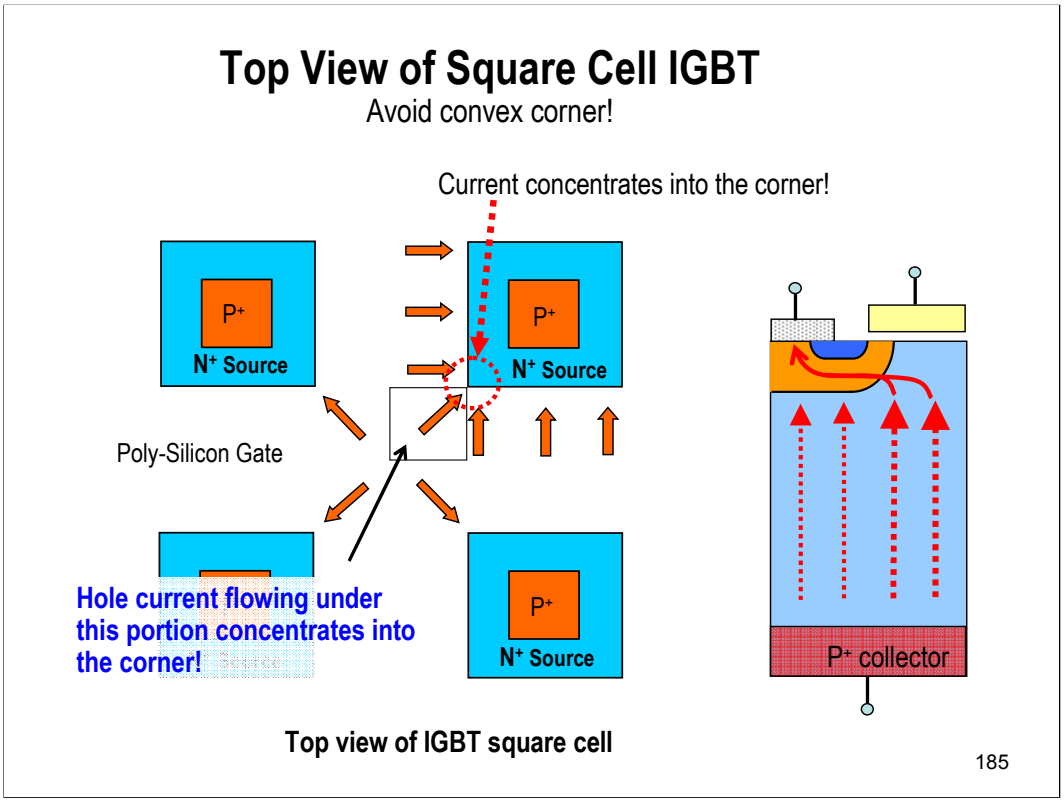
because carriers are stored broadly inside the device.

The figure on the left shows turn-off waveforms, when the gate resistance is very small. The channel electron current stops in the early stage of the turn-off. So, all of the current is carried by holes in the turn-off. In the p-base, it should be assumed that most of the current flows under the N^+ source.

Another point is, in the turn off transient, most of the current is created by the depletion layer expansion.

The stored carriers exist almost anywhere inside the device. Thus, the hole current flows almost uniformly when V_{CE} is recovering.

We need to consider these facts to design the devices.



Now, we consider square IGBT cells.

This slide shows the top view of square cell IGBT.

This is N+ source, P+ contact, and this white region is gate poly-silicon.

In the turn-off transient, hole current flows from p+ collector toward the gate poly-silicon,

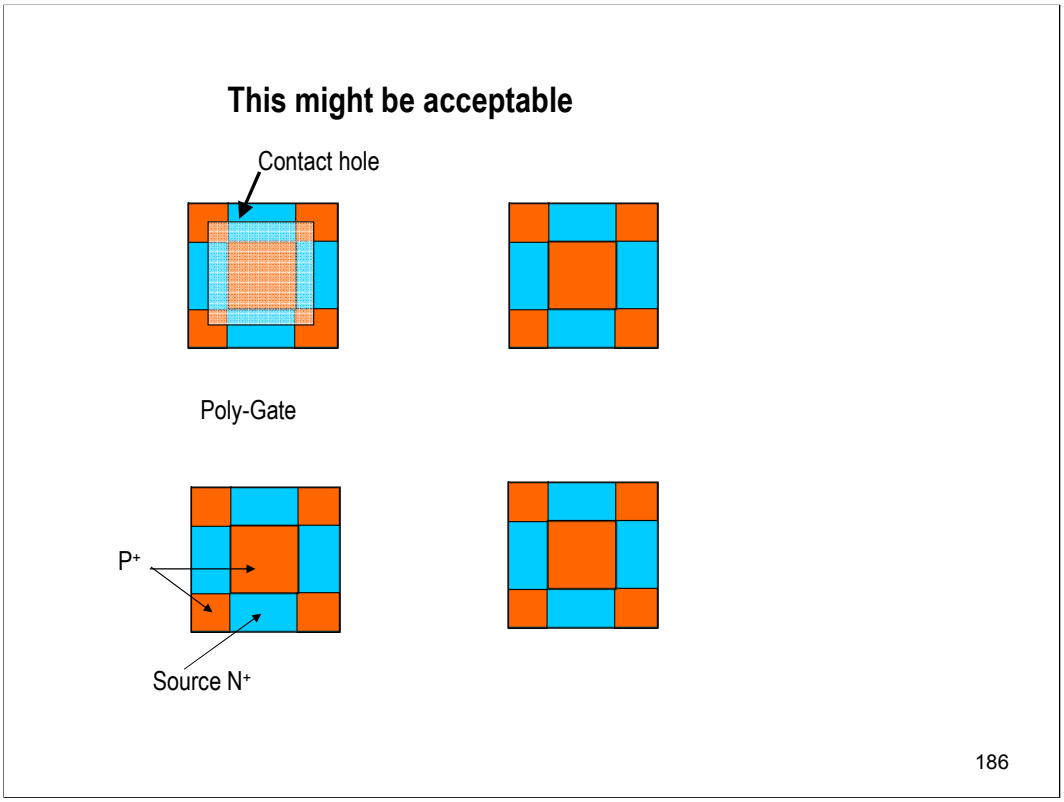
And then, flows into the P-base.

We pay attention to this portion.

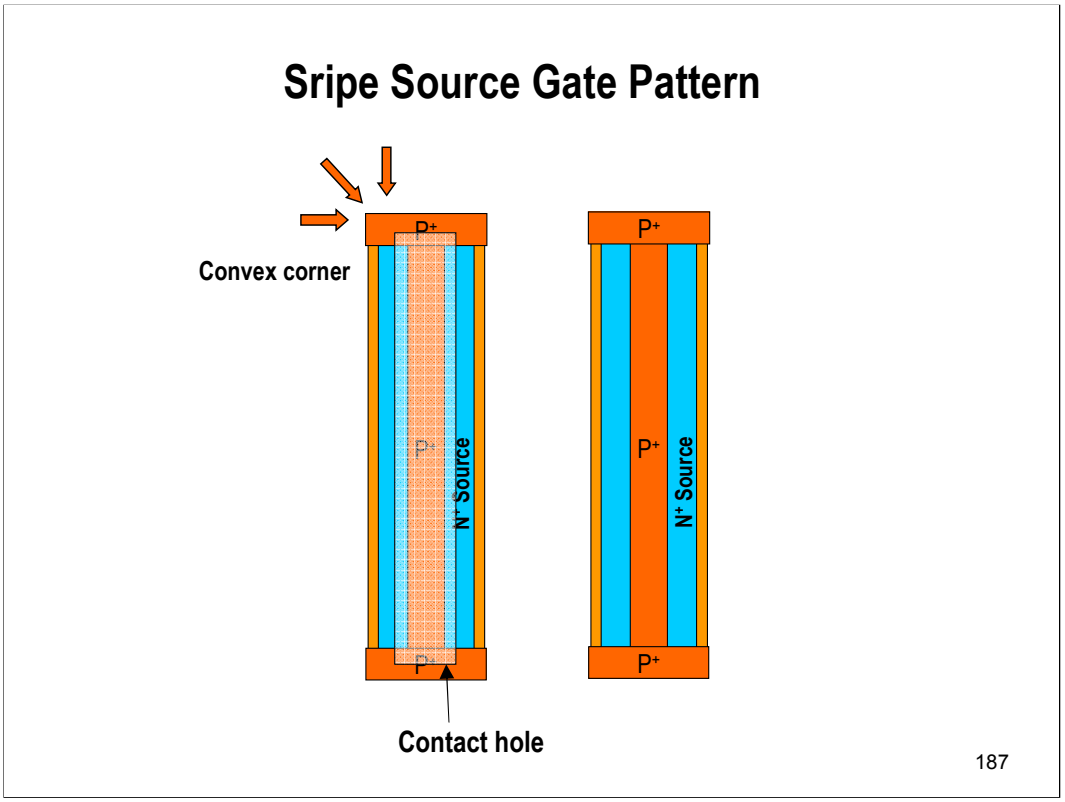
The hole current flowing up toward this portion flows into this particular corner.

So, hole current concentrates into this corner, and latch-up may occur in this corner.

If there is convex corner in the cell geometry, it is dangerous.



If you want to use square cells, it is better to eliminate N+ source in the corners and make P+ diffusion in the corner.



Stripe source gate pattern is good.

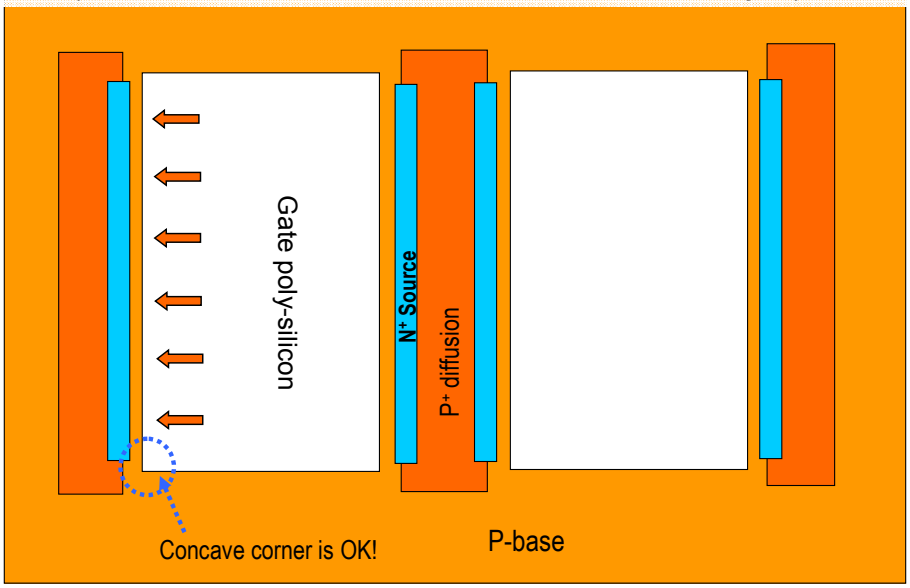
But, we need to adequately treat the ends of the stripe cell.

There is a convex corner.

We can make P+ diffusion in the corner to eliminate the danger.

Sripe Source Gate Pattern is Good

Frequently used technique is to connect the ends of the stripe cells, using P-type diffusion



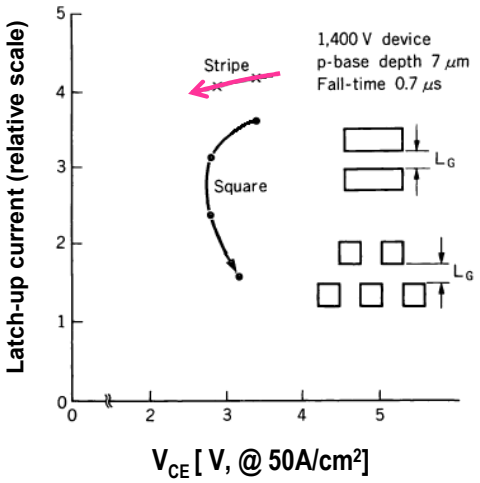
188

The frequently used technique in IGBT cell design is to connect the ends of the stripe cells, using P-type diffusion layers.

There are concave corners, but, the concave corners have no problem.

Stripe Cell & Square Cell

- This figure compares Stripe cell and Square cell
- Latch-up current of Stripe cell is larger than that of Square cell
- V_{CE} can be decreased by increasing L_G if Stripe cell is adopted.



189

This figure compares stripe cell and square cell.

Vertical axis shows latch-up current and horizontal axis shows collector emitter voltage.

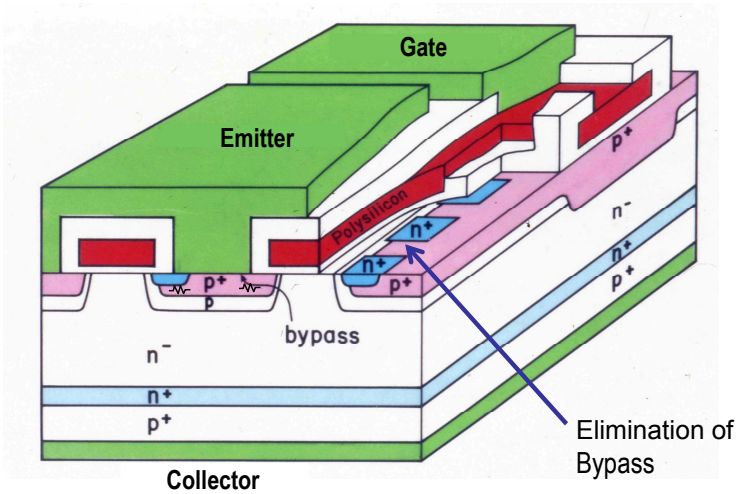
If we adopt square cell and when we increase the gate width L_G, V_{CE} decreases but, the latch–up current rapidly decreases.

On the other hand, for stripe cells, latch-up current is always larger than that of square cell.

If we adopt stripe cells, we can increase the gate width in order to decrease V_{CE} without significantly deteriorating latch-up current.

Hole Bypass

- **Hole bypass increases latch-up current and simultaneously decreases I_{SAT}**
- N^+ layer is periodically eliminated. Region where N^+ is eliminated is called "bypass."
- The shallow P^+ is extended into the channel region in the bypass region.
- Bypass creates a low resistance path from the N-base to the Emitter electrode



190

Now, I interpret the idea of hole bypass.

Hole bypass is the technique to increase latch-up current and simultaneously decreases the saturation current.

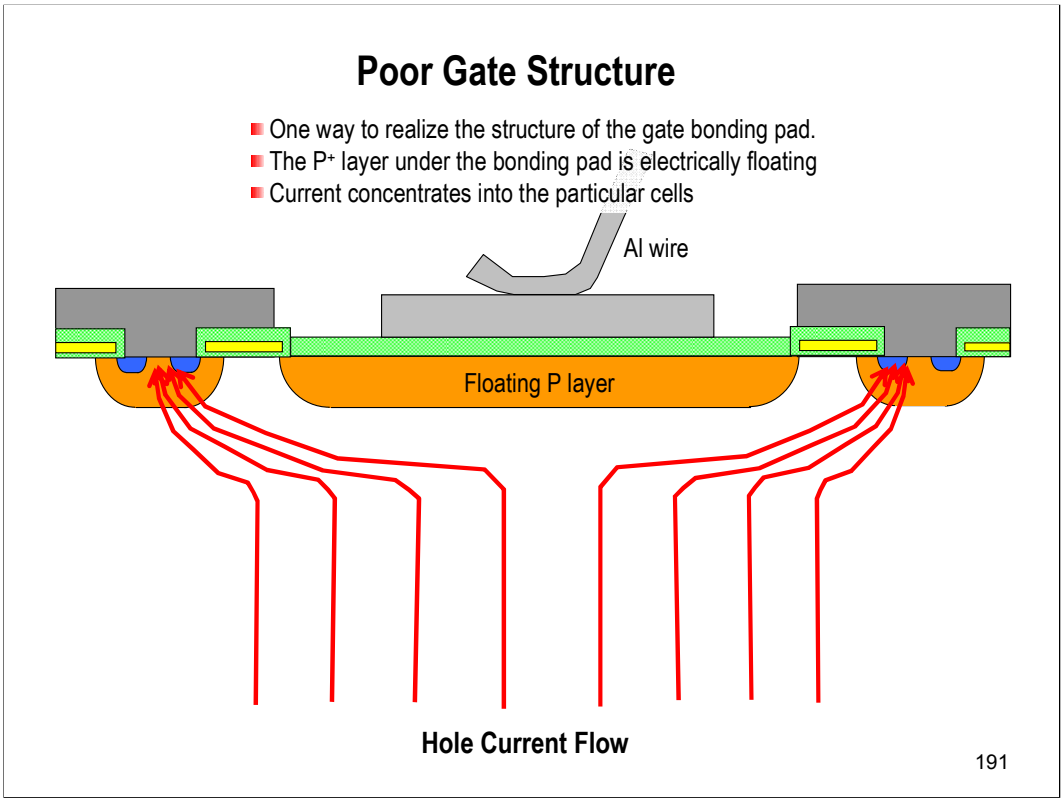
Thus, this is good to realize non-latch-up IGBTs.

This figure shows the structure.

The N^+ emitter layer is periodically eliminated to reduce the saturation current.

The region where N^+ layer is eliminated is called "bypass". In the bypass, the shallow P^+ layer is extended to the channel region to reduce the p-base resistance.

The bypass creates a low resistance path from the N-base to the emitter electrode. The bypass effectively reduces the p-base resistance and increases latch-up current.



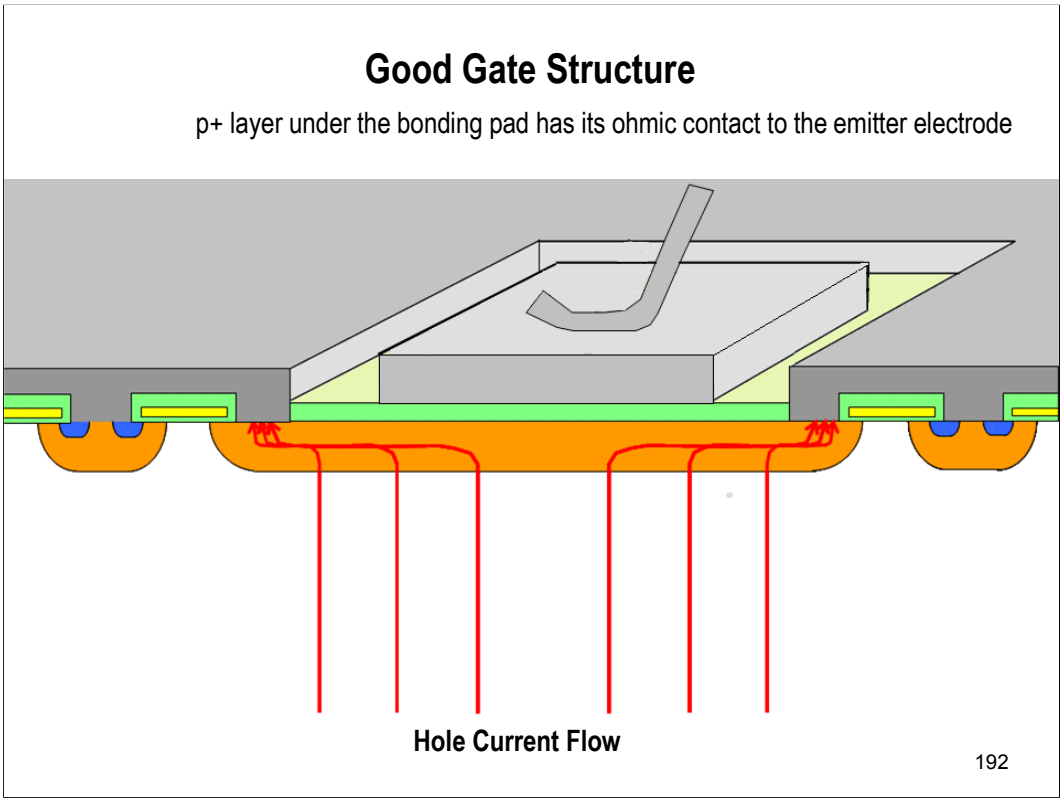
The last thing I would like to mention is the gate structure.

This is one way to realize the structure of the gate bonding pad.

The P+ layer under the bonding pad is electrically floating.

Thus, the hole current flowing toward the bonding pad flows into the nearest IGBT cell.

So, the current concentrates into this particular cell, which is located nearest to the bonding pad.



This slide a good gate structure.

p+ layer under the bonding pad has its ohmic contact to the emitter electrode.

Thus, we can avoid the current concentration.

Fin

Power MOSFET

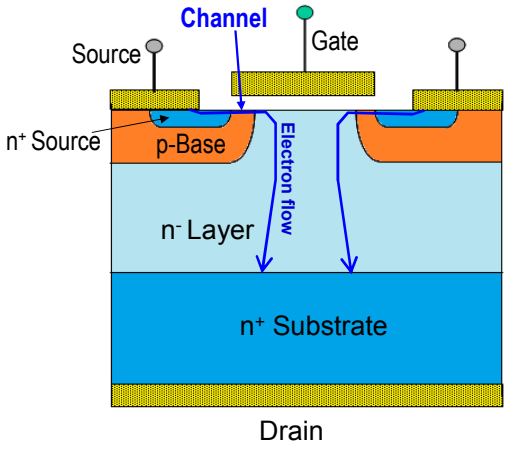
1. Fundamentals of power MOSFETs
2. Recent topics in low voltage power MOSFETs
3. High Speed Power MOSFET
   — Ideal switching in power MOSFET
4. Super Junction MOSFET

2

1. Fundamentals of power MOSFET

Double Diffusion MOSFET (DMOSFET)

- This is typical structure of Double Diffusion MOSFET.
- Only electrons conduct current. Majority carrier device.
- High speed switching is basically realized.



4

This is typical structure of Double Diffusion MOSFET.

It's called DMOS for abbreviation.

I talked about IGBT, so far.

MOSFET is more basic device, compared with IGBTs.

IGBT uses P^+ substrate.

MOSFET uses N^+ substrate.

Only electrons conduct current.

MOSFET is majority carrier device.

So, high speed switching is basically realized.

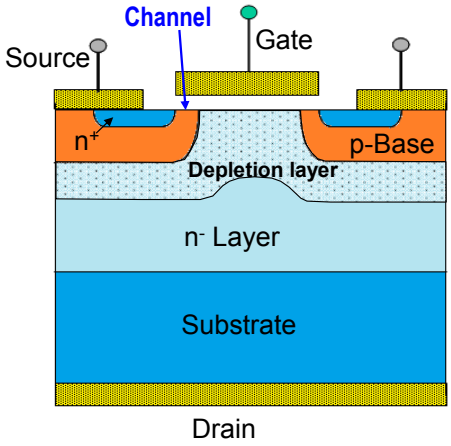
This is N^+ source, P-base, high resistance drain and N^+ drain.

If we apply a positive bias to the gate, the surface of the p-base is inverted, and a channel layer is formed

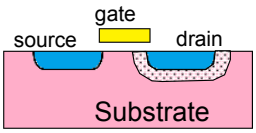
Electrons are injected from the channel into the high resistance drain, and current flows.

DMOSFET (Double Diffusion MOSFET)

- Distinguished feature of DMOSFET is that the depletion layer expands in n^- drain layer.
- In NMOS, the depletion layer expands in the channel and the channel plays two roles: switching current & sustaining voltage
- In DMOS, the two roles are assigned to two different layers.
- Using high resistance n^- drain layer, a high voltage capability can be achieved in DMOS.



Channel plays two roles in NMOS.
One is switching current.
The other is to sustain voltage



NMOS

5

High voltage power MOSFET withstand a high voltage such as 100V, 500V or more.

NMOS withstand only several volts.

Distinguishing feature of power MOSFET is that the depletion layer expands in the high resistance drain layer.

In low voltage NMOS, the depletion layer expand in the channel region. The channel plays two roles. One is switching the current. The other is to withstand the applied voltage.

In DMOSFET, these two roles are assigned to two different layers, the channel and the drain.

Using high resistance n^- drain layer, a high voltage capability can be achieved in DMOS.

Diffusion Self-Alignment(DSA)

- Diffusion self alignment was developed in 1970.
- DMOS uses this technology to form the channel region.

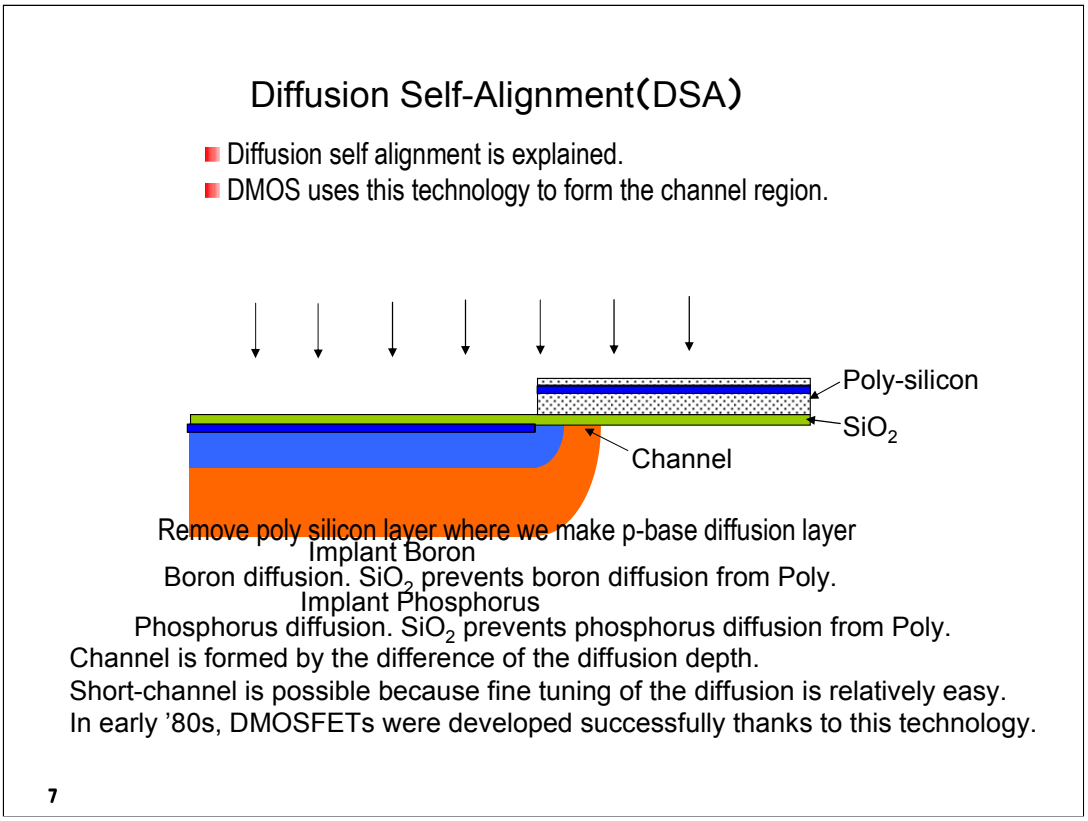
First, form 100nm silicon dioxide, and deposit 500nm poly silicon layer

6

Here, I interpret the concept of diffusion self alignment.

DMOS uses this technology to form the channel region.

First, form 100nm silicon dioxide, and deposit 500nm poly silicon layer.



Then, remove poly silicon layer where we make p-base and n-source diffusion layer.

Then, introduce boron by ion implantation.

Boron is introduced into silicon layer, where poly-silicon layer is removed.

Boron is blocked where poly-silicon layer exist.

Then, by thermal annealing, boron diffusion layer are formed.

Borons, trapped inside the poly silicon, stays inside the poly-silicon layer, because silicon dioxide blocks the boron diffusion.

Then, again we execute phosphorous ion implantation.

Phosphorous atoms are introduced in silicon where diffusion layer will be formed.

After thermal treatment, shallow n+ layer is formed.

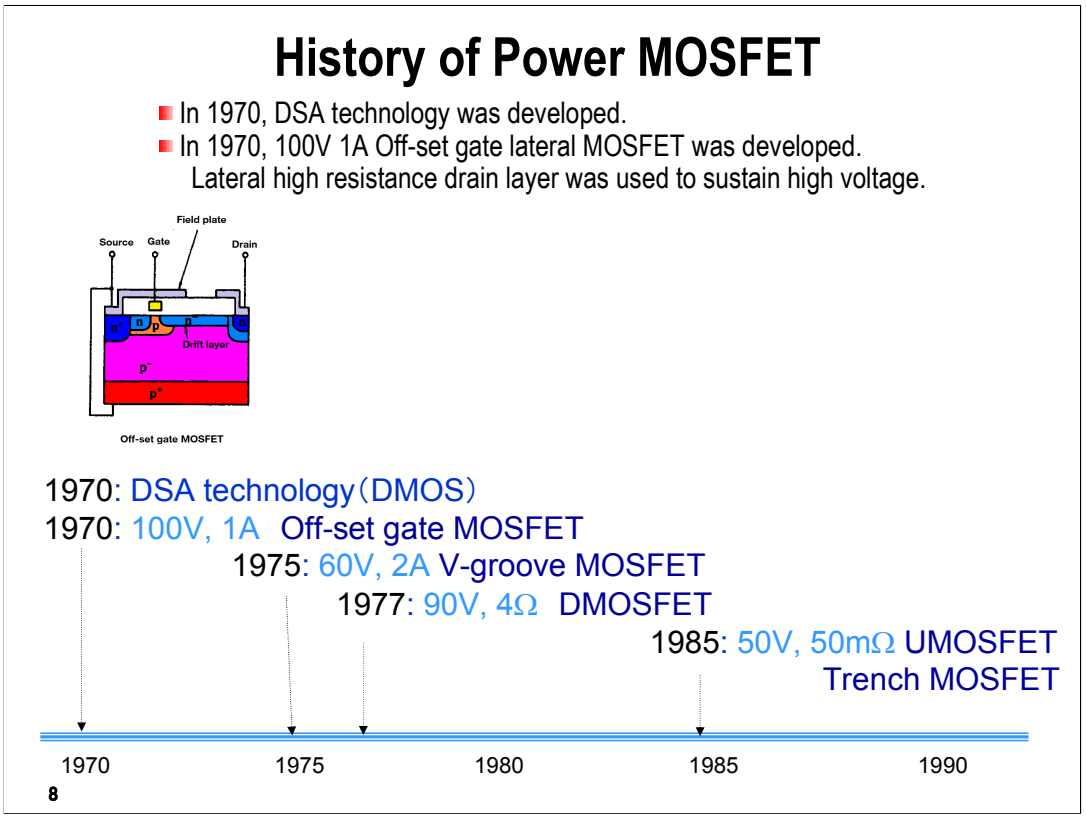
Again, phosphorous atoms, trapped inside the poly, stays inside the poly.

Finally, we have channel region under the silicon dioxide layer.

The channel is formed by the difference of the diffusion layer depth.

Thus, short-channel is easily formed because the fine tuning the diffusion depth is relatively easy.

In early '80s DMOSFETs were developed successfully thanks to this tecnology.



This figure shows the history of power MOSFETs.

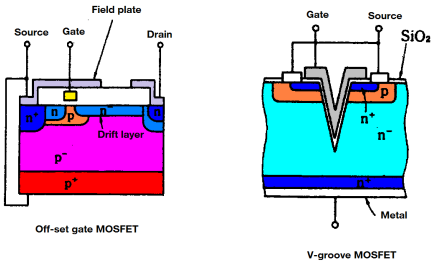
In 1970, DSA technology were developed.

In 1970, lateral high voltage MOSFETs were developed.

Lateral high resistance drain layer is used to sustain high voltage.

History of Power MOSFET

- In 1975, V-groove MOSFET was developed using anisotropic etching.
  The disadvantage of this structure was the sharp edge of the V-groove.



1970: DSA technology（DMOS）
1970: 100V, 1A Off-set gate MOSFET
1975: 60V, 2A V-groove MOSFET
1977: 90V, 4Ω DMOSFET
1985: 50V, 50mΩ UMOSFET
Trench MOSFET

1970 1975 1980 1985 1990

9

In 1974, vertical power MOSFET using V-groove was developed and commercialized.

V-grooves are easily fabricated by using anisotropic etching.

The channel is formed along the V-groove.

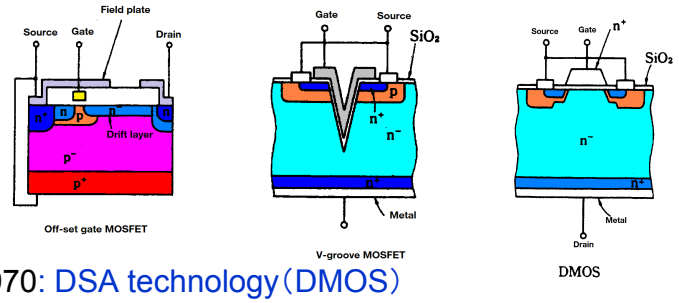
The disadvantage of this structure is the sharp edge of the V-groove.

A high electric field appears in the sharp edge,

And, thus, a high breakdown voltage is difficult to achieve in V-groove MOSFETs.

History of Power MOSFET

- DMOS power MOSFET was developed in 1977.
  DMOS became the major technology of power MOSFET.



1970: DSA technology（DMOS）
1970: 100V, 1A Off-set gate MOSFET
1975: 60V, 2A V-groove MOSFET
1977: 90V, 4Ω DMOSFET
1985: 50V, 50mΩ UMOSFET
Trench MOSFET

1970 1975 1980 1985 1990

10

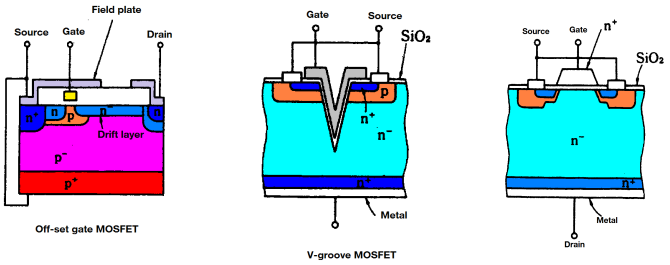
DMOS power MOSFET was developed in 1977.

DMOS became the major technology of power MOSFET.

The disadvantage of the V-groove was successfully overcome.

History of Power MOSFET

- In 1985, trench MOSFET was developed.



1970: DSA technology（DMOS）
1970: 100V, 1A Off-set gate MOSFET
1975: 60V, 2A V-groove MOSFET
1977: 90V, 4Ω DMOSFET
1985: 50V, 50mΩ UMOSFET
Trench MOSFET

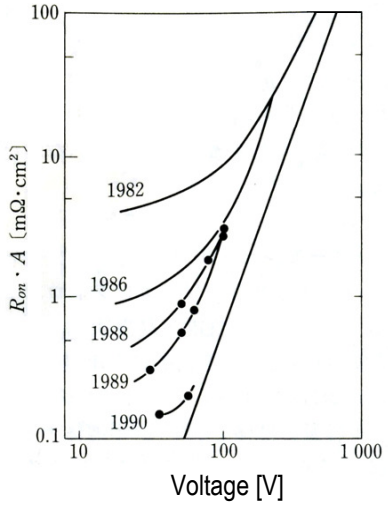
1970 1975 1980 1985 1990

11

In 1985, first trench MOSFET was developed.

Improvement in MOSFET On-resistance

The R_{on} improvement was extensively done in 1980s.
In 1990, the MOSFET on-resistance approached the silicon limit.



12

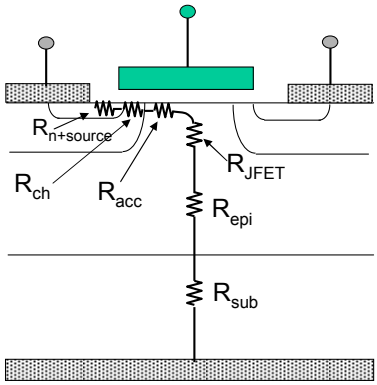
This slide shows the history of improvement in MOSFET on-resistance.

The improvement was extensively done in the 1980s.

In 1990, the MOSFET on-resistance approached the silicon limit by using trench gate MOSFETs.

MOSFET on-resistance

- On-resistance consists of many components.
- $R_{n+source}$, R_{sub} is usually negligible.



13

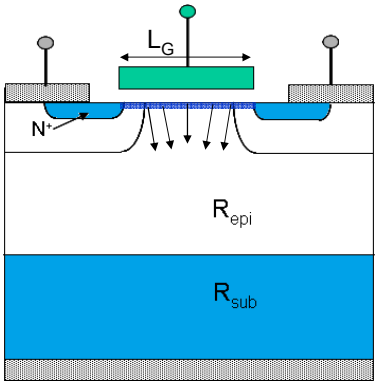
MOSFET on-resistance consists of many components.

These are the resistance of N^+ source layer, the channel resistance, accumulation layer resistance, JFET resistance, epi-layer resistance and substrate resistance.

N^+ source layer resistance and the substrate resistance are relatively small compared with the other resistances especially in high voltage MOSFETs.

MOSFET on-resistance

- Accumulation layer is formed in the surface of n^- layer
- Electrons are injected from the accumulation layer, if it exists.



14

Accumulation layer is formed in the surface of n- layer

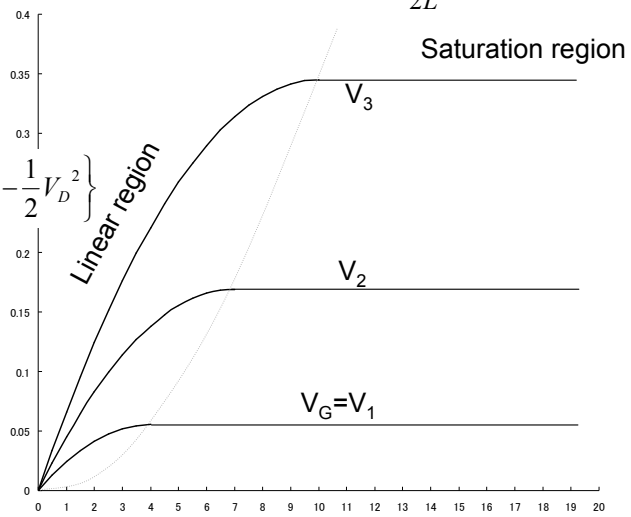
Electrons are injected from the accumulation layer if it exists.

Long channel theory of MOSFET

- The channel length of power MOSFET is relatively long in order to sustain high voltage.
- Theory of long channel is useful to describe power MOSFET electrical characteristics

$$I_D = \frac{\mu C_{ox} W}{2L}(V_G - V_{TH})^2$$

$$I_D = \frac{\mu C_{ox} W}{L}\left\{(V_G - V_{TH})V_D - \frac{1}{2}V_D{}^2\right\}$$

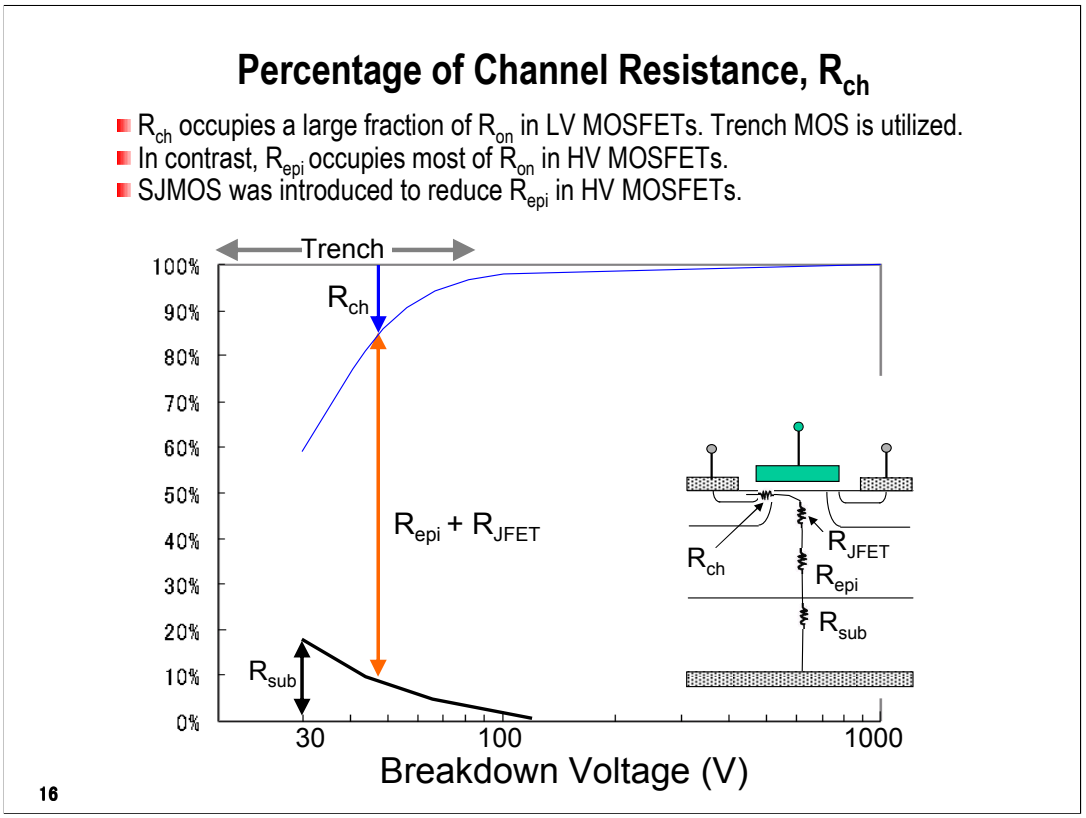


15

The channel electrical characteristics is very important to determine the power MOSFET characteristics.

In power MOSFETs, channel length is relatively long to sustain a high voltage, theory of long channel is useful to describe power MOSFET electrical characteristics.

In this slide, I show familiar expressions of current voltage relation for the linear region and the saturation region.



This chart shows the percentages of channel resistance and epi-layer resistance in the total MOSFET resistance.

Channel resistance occupies a large fraction of total resistance in low voltage MOSFETs, especially less than 60V. Trench MOSFET is utilized in low voltage MOSFETs.

The channel resistance in high voltage MOSFET over 500V only occupies a small fraction of the total resistance. So, planar MOSFET technology is still frequently used in high voltage power MOSFETs.

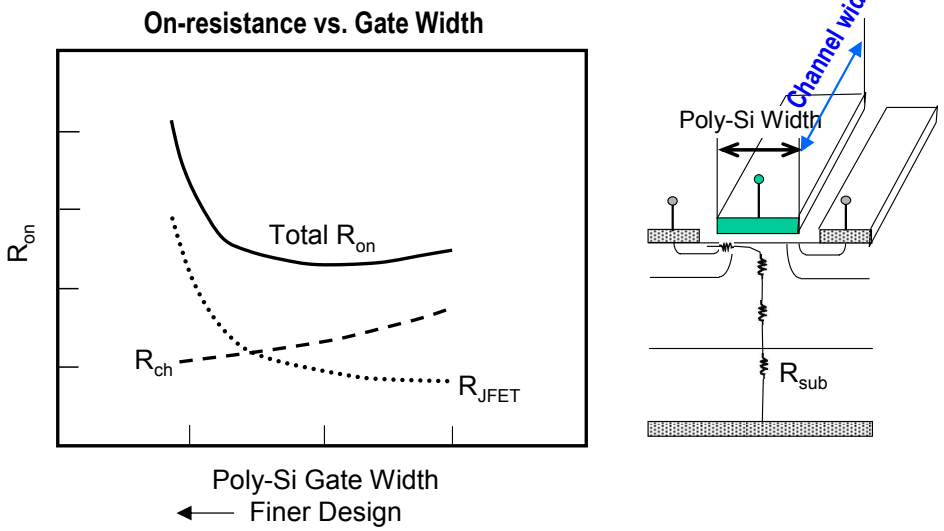
The epitaxial layer resistance occupies most of the MOSFET resistance in high voltage MOSFETs.

So, high voltage MOSFET on-resistance faces silicon material limitation.

This is why super junction technology was introduced in high voltage MOSFETs.

Limitation of R_{on} reduction in planar MOSFET

- In the 90s, R_{on} was extensively reduced.
- In planar MOSFETs, it was found finer design does not necessarily reduce R_{on}
- Finer design reduces R_{ch}. But, R_{JFET} increases as poly-silicon gate width reduces.
- There is a limitation in Ron as far as the planar MOSFET technology is concerned.



17

In the 1990s, great efforts were done to reduce MOSFET on-resistance, especially in low voltage MOSFETs.

It was found that introducing finer design does not necessarily reduce MOSFET on-resistance.

Finer design can decrease the channel on-resistance by reducing the cell pitch.

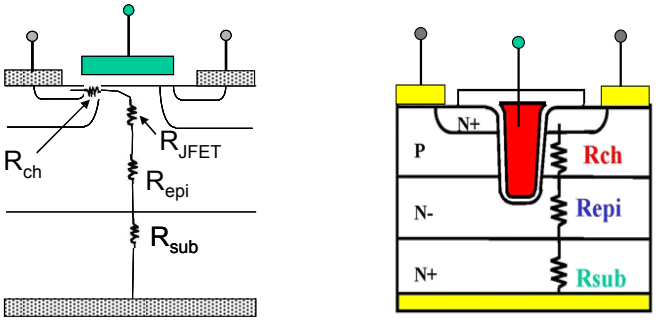
However, if the poly-silicon gate width is excessively reduced, JFET resistance increases and the total on-resistance also increases.

There is a trade-off between the JFET resistance and the channel resistance.

Thus, there is a limitation in attainable on-resistance as far as the planar MOSFET technology is concerned.

Trench Gate

- Trench technology was introduced to breakthrough the situation.
- Trenches penetrate n^+-source, p-base and reaches N^- epi-layer.
- There is no JFET. Thus, fine design simply reduces the on-resistance.
- In the 1990s, R_{on} was further reduced by the trench technology.



18

In the 1990s, trench technology was introduced to breakthrough the situation.

This figure compares planar and trench technology.

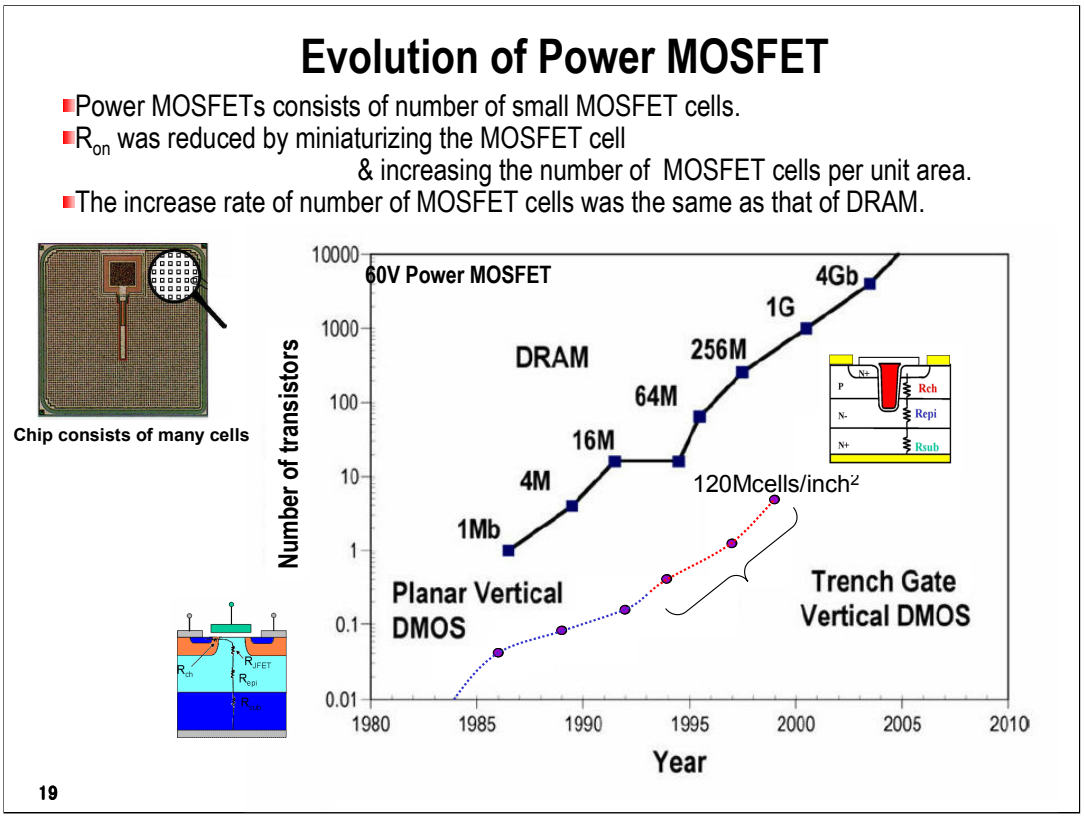
In the trench MOSFET, trenches are formed, penetrating the n^+-source and the p-base layers and reaching the high resistance n-layer.

Inversion layers are formed on the surface of the p-base along the trench side walls.

There is no JFET in the current path.

Thus, reducing the trench to trench distance simply reduces the on-resistance.

The on-resistance of the low voltage MOSFETs was further reduced in the 1990s by introducing the trench technology.



Power MOSFETs consists of number of small MOSFET transistors.

In the past, MOSFET on-resistance is continually reduced by introducing finer design rule from CMOS technology.

Miniaturizing the unit MOSFET and paralleling a large number of unit transistors is the way to reduce the MOSFET on-resistance.

This figure shows number of transistors, integrated in a unit area, as a function of year.

The figure also compares the power MOS technology with the DRAM technology.

In the 80s and early 90s, power MOSFETs were fabricated by planar technology.

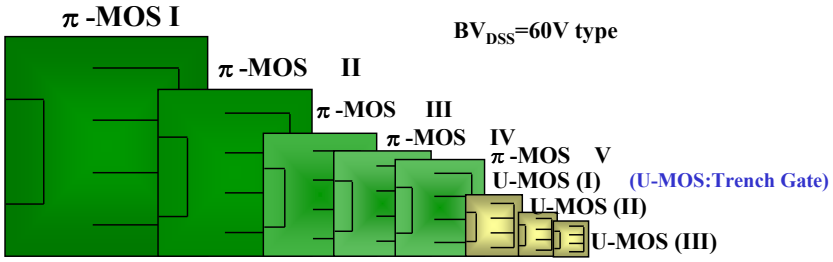
After 1994, trench technology was introduced.

Increase rate in the number of transistors integrated in the unit area of power MOSFETs is almost the same as that of DRAM.

The number of cells exceeded 100M/square-inch in 1999.

Evolution of 60V Power MOSFET

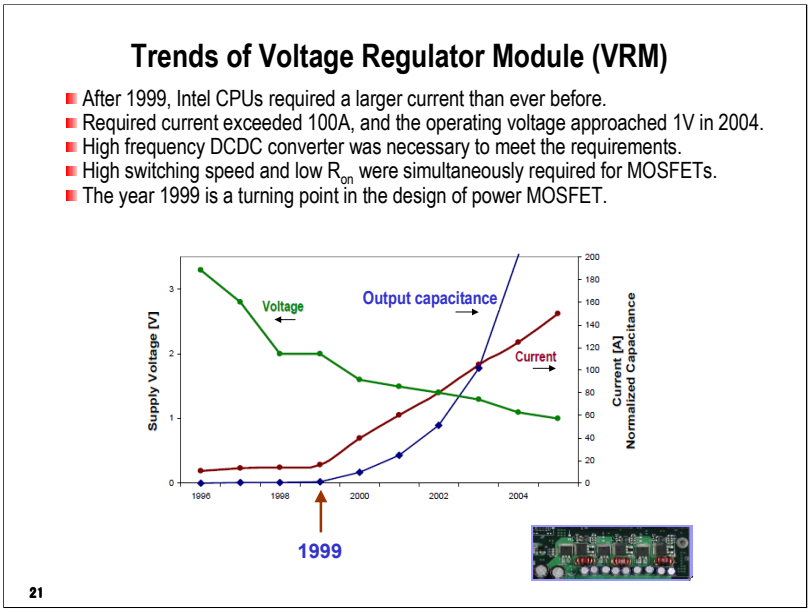
- R_{on} in 1999 was decreased to 3.5% of that of the first MOSFET product in 1982.
- The chip size was also decreased to just 4% of the first chip.
- 1999 is a special year in the MOSFET history.



	π-MOS I	π-MOS II	π-MOS III	π-MOS IV	π-MOS V	U-MOS(I)	U-MOS(II)	U-MOS(III)
Chip Area	1	0.6	0.33	0.25	0.2	0.08	0.05	0.04
$R_{DS(ON)}$ (mΩ・cm^2)	12.7	7.6	4.2	3.2	2.4	1	0.6	0.45
Cell Density (Mcells/Inch^2)		0.2	1	2	4	10	30	120
Mass Production	1982	1984	1986	1989	1992	1994	1997	1999

20

This figure shows evolution of 60V MOSFET.
The on-resistance drastically decreased to 3.5% of that of the first MOSFET product in 1982.
The chip size was also decreased to just 4% of the first chip.
The year 1999 was a kind of turning point in the MOSFET technology.



This figure shows the trend of Voltage regulator module for micro-processor.

The year 1999 is a special year.

After 1999, the CPU required a larger current than ever before and CPU operating voltage reduces below 2V.

The required current exceeded 100A, and the CPU operating voltage approached 1V in 2004.

Requirement for DCDC converter became stringent and High frequency switching DCDC converters are necessary to meet the rapid current change of the load.

The required output capacitance of DCDC converters increased rapidly after 1999.

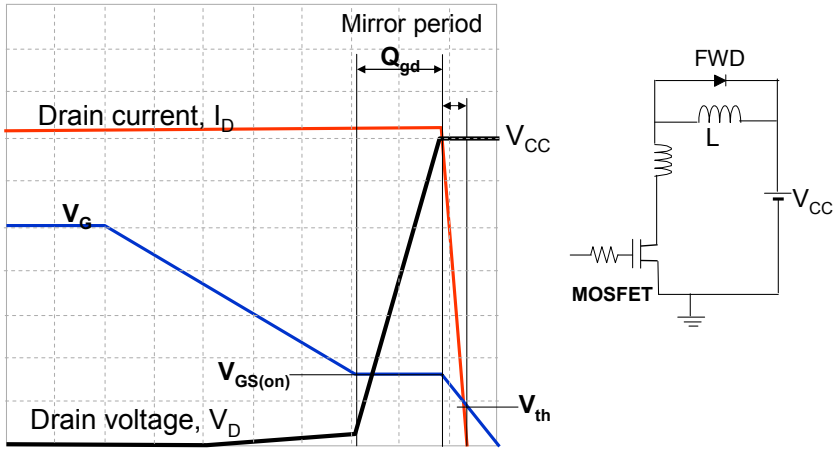
Requirements for power MOSFETs became also severe.

Higher switching speed capability and lower on-resistance were simultaneously required.

The year 1999 is a turning point in the design of power MOSFET.

Switching waveforms and definition of Q_{gd}

- First, V_G decreases to V_{GS}(on). Then, V_D rapidly increases and mirror period starts.
- In the mirror period, V_D rapidly increases in order to keep I_D & V_G at constant value.
- The gate charge, extracted in the mirror period, is defined as Q_{gd}.
- After V_D exceeds V_{CC}, FWD is forward biased and V_D is clamped at V_{cc}.
- Then, V_G decreases and I_D rapidly decays.



22

In this slide, I show typical switching waveforms under inductive load and the definition of Qgd.

The gate voltage initially simply decreases to V_{GS}(on). After that, the drain voltage rapidly increases in order to keep the drain current at the same value.

This period is called as “mirror period.”

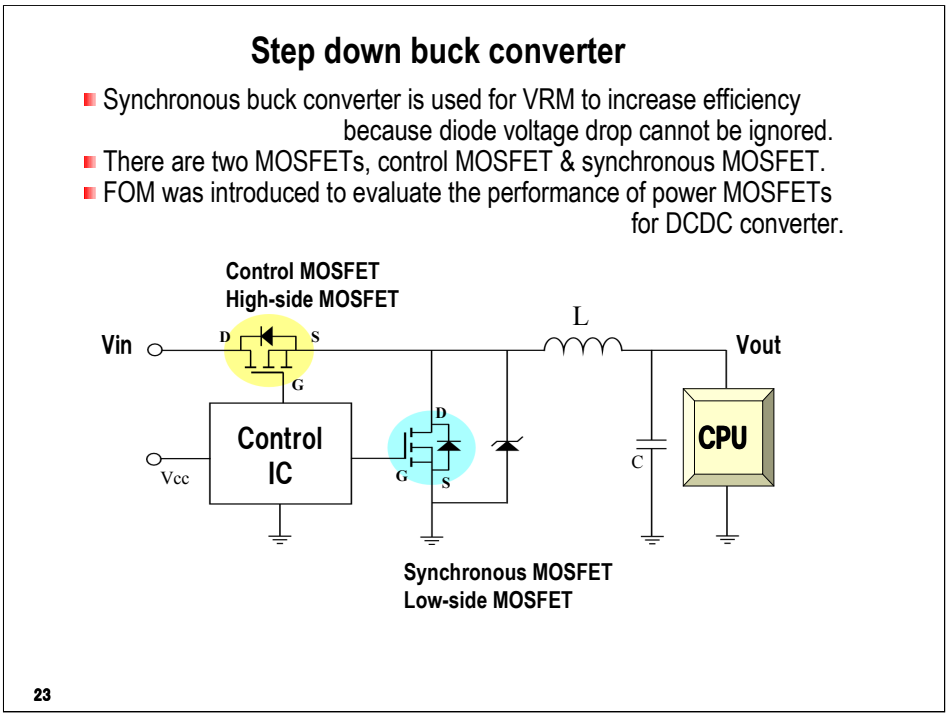
In the mirror period, the gate voltage is also kept constant. The gate charge is always extracted from the gate even in the mirror period.

The gate charge, extracted in the mirror period, is defined as Qgd.

After the drain voltage exceeds the external voltage source, the free wheeling diode is forward biased and clamp the drain voltage.

When the drain voltage is clamped, the gate voltage start to decrease, and the drain current also decreases.

When the gate voltage goes below the threshold voltage, the drain current becomes zero.



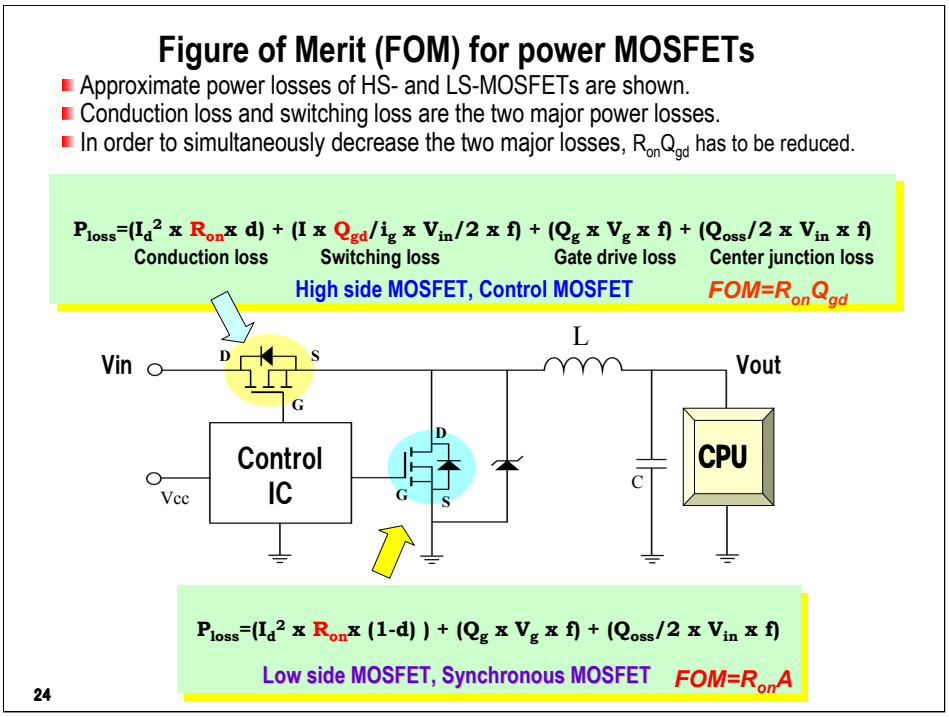
This figure shows step down buck converter.

Synchronous buck converter is used in Voltage Regulator Modules to increase efficiency, because the diode voltage drop cannot be ignored.

There are two MOSFETs. One exists at high voltage side, and is called high-side MOSFET or control MOSFET, which actually switches the current.

The other is located at low voltage side, is called low side MOSFET, or synchronous MOSFET, that rectifies current.

FOM was newly introduced to evaluate the performance of power MOSFETs for the DCDC converter.



I'll introduce Figure of merit for power MOSFETs.

This shows the approximate power losses, generated in the two power MOSFETs.

The first term is conduction loss.

2nd is switching loss.

3rd is gate loss.

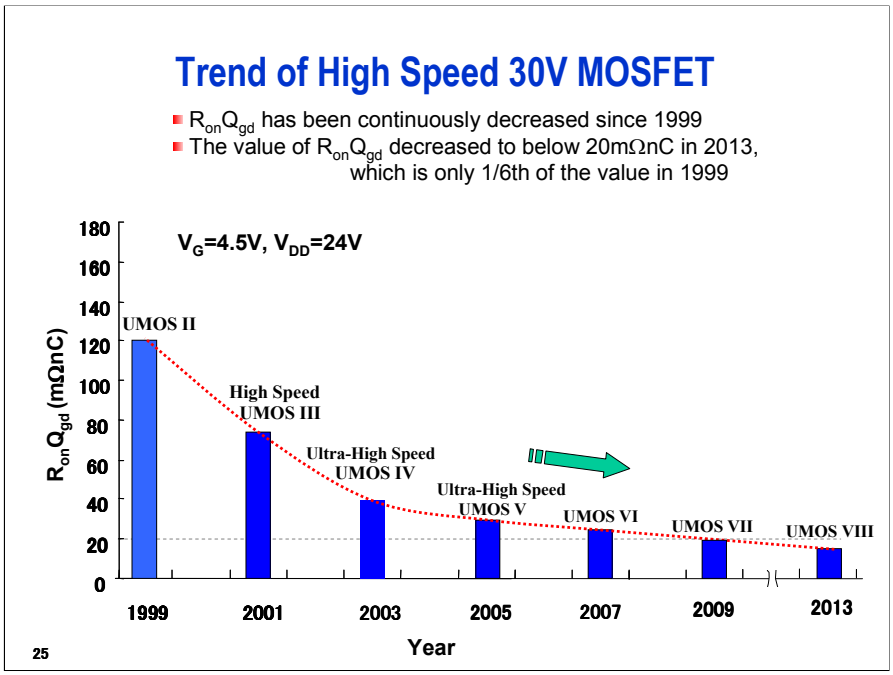
The forth is center junction capacitance loss.

The first two terms are the major losses.

In order to reduce both losses simultaneously, the product of RonQgd has to be reduced.

Thus, the figure of merit for high-side MOSFETs is $R_{on}Q_{gd}$.

Anologously, figure of merit for low side MOSFETs is RonA.

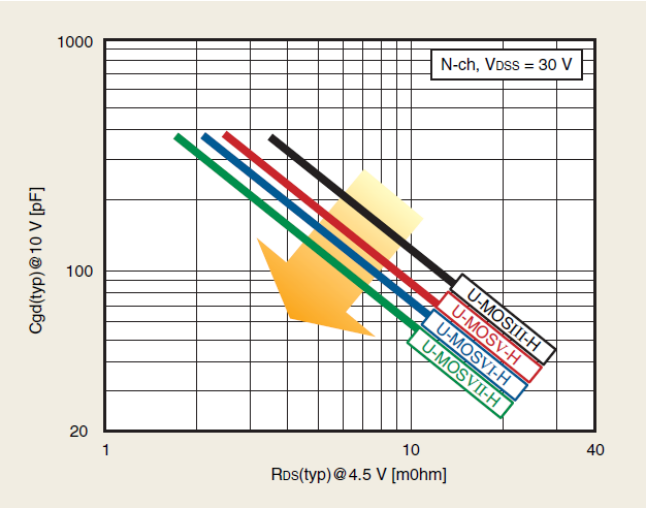


After 1999, the RonQgd value of high speed MOSFET has been continuously decreased.

The value of RonQgd decreased below 20mΩnC, that is only 1/6th of the value in 1999.

Trade-off between C_{GD} and R_{on}

In order to reduce $R_{on}Q_{gd}$, both R_{on} and C_{GD} should simultaneously be reduced.

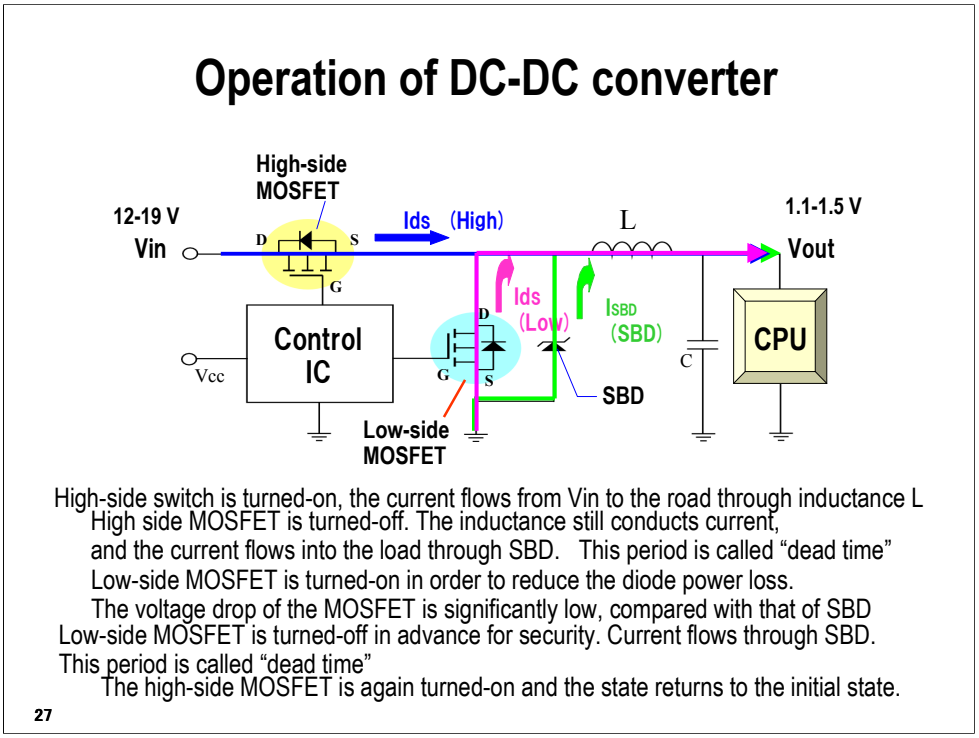


26

In order to reduce RonQgd value, both the on-resistance and the gate-drain capacitance should be reduced.

This figure show the trade-off curve between C_{GD} and on-resistance for a series of MOSFET generations.

From generation to generation the trade-off is improved.



This figure briefly interprets the operation of DCDC converter.

First, high-side switch is switched-on and the current flows from Vin to the load through inductance L.

Energy is stored in L in this period.

Then, the high side MOSFET is turned-off before turning-on the low side MOSFET. The inductance still conducts current, and the current flows into the load through SBD. This period is called “dead time.”

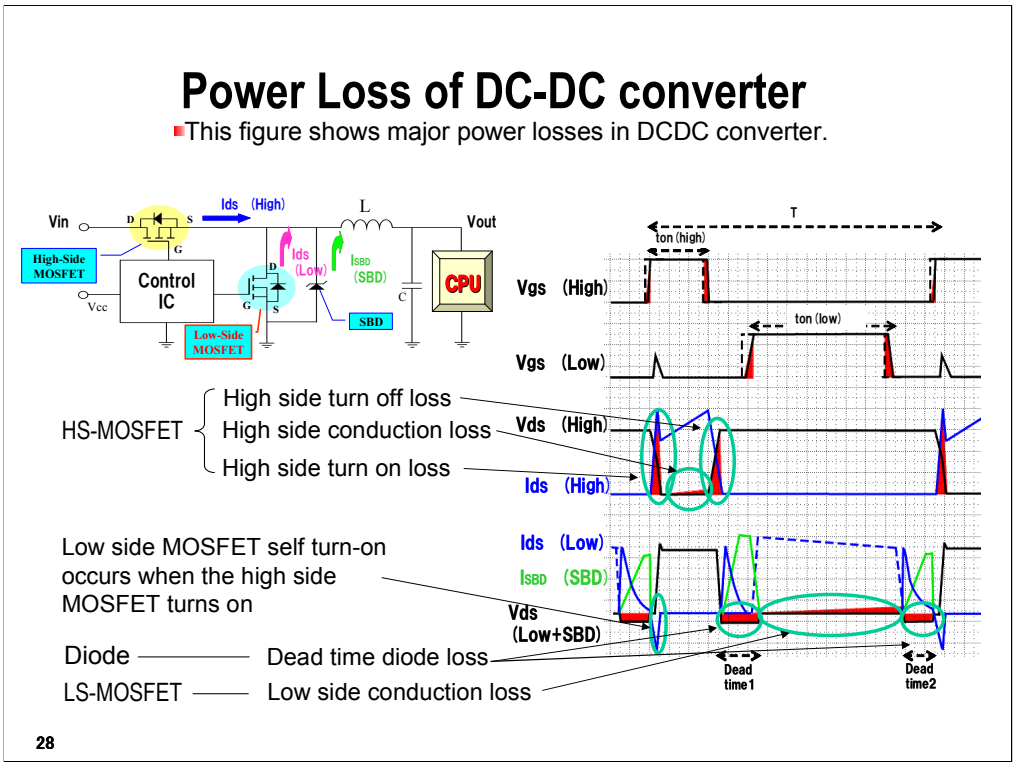
Then, the low-side MOSFET is turned-on in order to reduce the power loss of the SBD. As the on-resistance of the MOSFET is very low, the voltage drop of the MOSFET is significantly low, compared with that of SBD.

Then, low-side MOSFET is turned-off in advance in order to turn-on the high-side MOSFET.

For security reason, there is a time lag between switching-off the low-side MOSFET and turning-on the high-side MOSFET. This time lag is called dead-time.

In the dead-time period, the current flows through the SBD.

And, then, the high-side MOSFET is again turned-on and the state returns to the initial state.



This figure shows major power losses in DCDC converter.

For high-side MOSFET, major loss is high-side MOSFET turn-on loss, conduction loss, and turn-off loss.

For low-side MOSFET, conduction loss is the major loss.

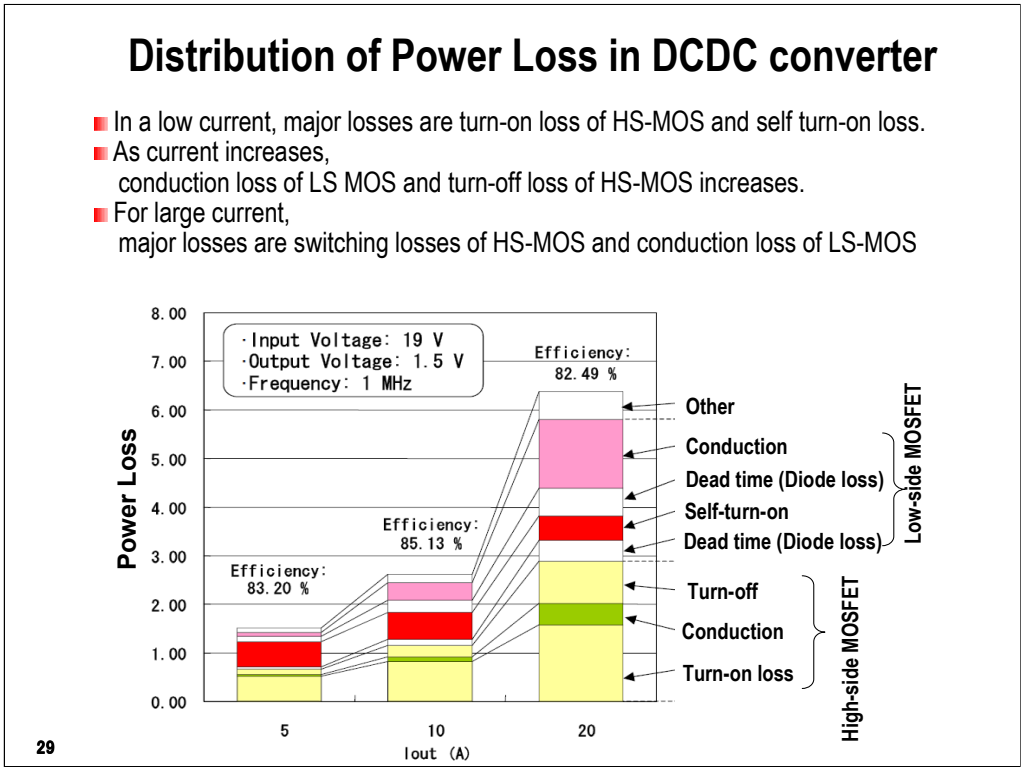
For diode, two dead time losses.

Extra loss, I should mention, is self turn-on loss in low side MOSFET.

The low-side MOSFET unintentionally turns-on for a short time period, when high-side MOSFET is turned-on.

This occurs, because a large dV/dt is imposed to the low-side MOSFET.

We will discuss this problem in detail, later.



This shows distribution of power loss in DCDC converter.

In a low current level, major losses are turn-on loss of high-side MOSFET and self turn-on loss.

As current increases, the conduction loss of low-side MOSFET and the turn-off loss of the high-side MOSFET increases.

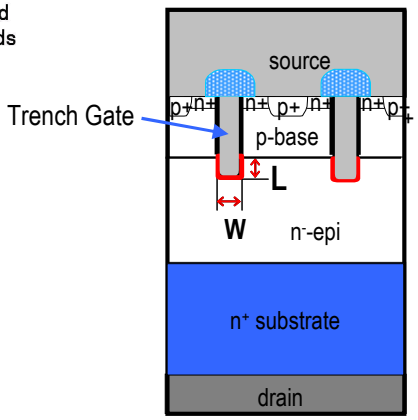
For large current level, all the components of power loss increases.

Major losses are switching power losses of high-side MOSFET and the conduction loss of low-side MOSFET

Reducing C_{GD} is important because Q_{gd} determines t_{sw}

- C_{GD} depends on the penetration depth, L, and the width, W, of trench gate
- Precise control of the depth and the width of the trenches is required.

Ciss=Cgs+Cgd
Coss=Cgd+Cds
Crss=Cgd



30

It is very important to reduce the gate drain capacitance

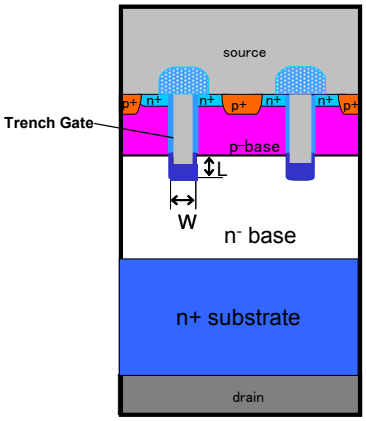
Because Q_{gd} determines switching speed.

The gate drain capacitance depends on the penetration depth of the trench, L, and the width of the trench, W.

It is important to precisely control the depth and the width of the trenches.

Reducing C_{GD} of Power MOSFET

One method is to thicken the bottom oxide of the trenches.
This reduces C_{gd} without increasing R_{on}.



Thick bottom oxide

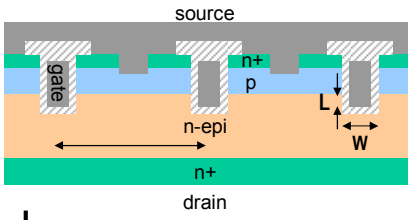
31

Various measures are proposed to reduce the gate drain capacitance.

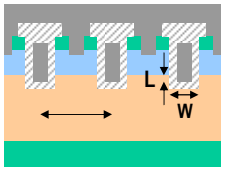
One method is to thicken the bottom oxide of the trenches. This reduces Cgd without increasing Ron.

Finer Design

- Finer design is good to improve the device performance.
- Simple reduction of the size of the devices improves R_{ch} but increases C_{GD}
- The trench penetration depth and the width have to be reduced simultaneously.



Smaller trench to trench distance: Lower $R_{on}A$, but, large C_{gd}
Smaller penetration depth: Smaller C_{gd}
Smaller trench width : Smaller C_{gd}



32

Finer design is very good to improve the device performance.

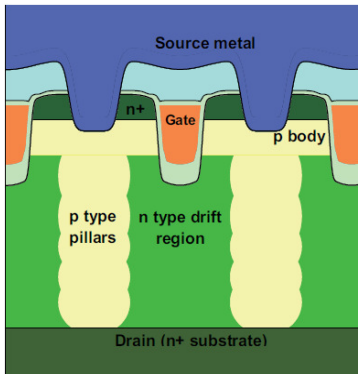
However, simple reduction of the size of the devices improves on-resistance but increases the gate drain capacitance.

This does not results in the reduction in RonQgd.

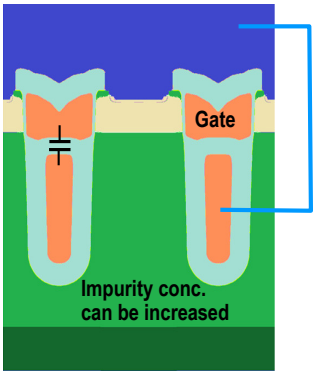
We have to keep in mind that the simultaneous reduction in the gate penetration depth and also the reduction in the width of the gate are required.

Super Junction and Split Gate MOSFETs

- SJ and Split gate are frequently used to improve $R_{on}Q_{gd}$
- SJ simply reduces R_{on}.
- In Split gate, upper poly is used as gate, lower poly is connected to the emitter.
- Lower poly works as field plate. The structure is effective for low C_{GD} & low Ron.



Super Junction



Split Gate

33

Recently, Super junction structure and split gate structure are frequently used for high speed MOSFETs.

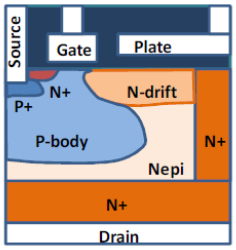
SJ simply reduces Ron.

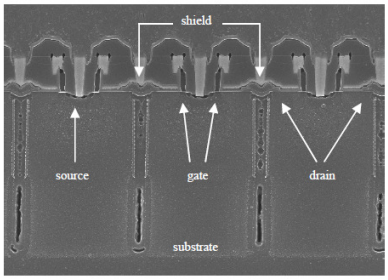
Regarding the split gate structure, the gate ploy-silicon is divided into two portions. This poly-silicon is used as gate, the other is electrically connected to the emitter and works as field plates. The epi-layer resistance can be decreased because of the field plate. Thus, the on-resistance can be decreased. The trench bottom capacitance works as gate source capacitance instead of gate drain capacitance.

The split gate structure is effective for reducing on-resistance and simultaneously reducing the gate drain capacitance.

Lateral MOSFET is effective for low $R_{on}Q_{gd}$

- Drain N^+ layer is connected to the backside N^+ substrate to make a vertical device!
- The Lateral device is good for low C_{gd}, and achieved lowest R_{on} Q_{gd} as shown in the figure.





ISPSD 2012

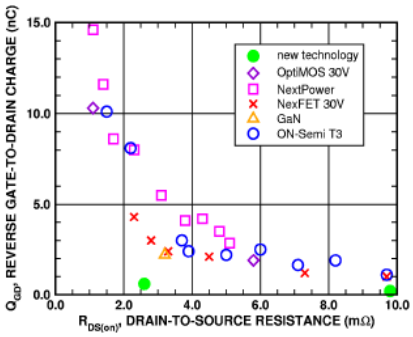


Figure 8. Comparison of the Miller charge (Q_{GD}) to other FET technologies [9-13] including OptiMOS™ (Infineon), NextPower™ (NXP), NexFET™ (Texas Instruments), GaN (IR and Efficient Power Conversion), and T3.

On Semiconductor

34

It was found that Lateral MOSFET structure is especially effective for reducing RonQgd.

The surface drain N^+ layer is electrically connected to the backside N^+ substrate to make vertical device.

The Cgd is very low in lateral MOSFETs.

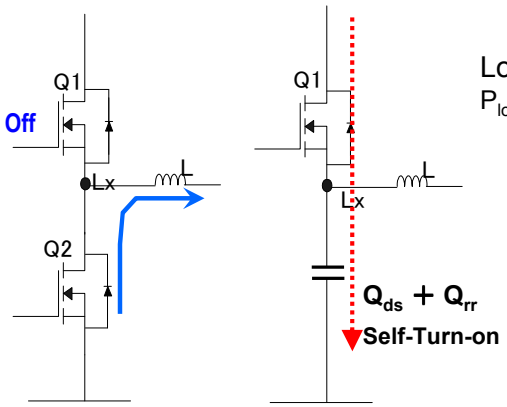
In the figure on the right-hand side, trade-off between Qgd and Ron is compared among various devices.

It is reported that lateral MOSFETs have the lowest RonQgd value and even better than GaN.

2. Recent topics
in low voltage power MOSFETs

Problems in MOSFET Turn-on

- When HS-MOS is off-state, current flows in the diode and carriers are stored.
- High-side MOSFET turn-on induces (1)diode recovery loss, Q_{rr} and (2)center junction capacitance loss of LS-MOS, Q_{ds} & (3)self-turn-on loss.
- Self-turn-on is analyzed in the next slide.



Loss of Low side MOSFET

$P_{loss} = (Q_{rr} + Q_{ds} + \text{Self-On-Charge}) * V_{in}$

36

When the high side MOSFET is in the off-state, the current flows in the diode and the carriers are stored in the diode.

Then, the high-side MOSFET turns on.

The diode is reverse biased and the power loss occurs in the diodes.

When the voltage is applied across the diode,

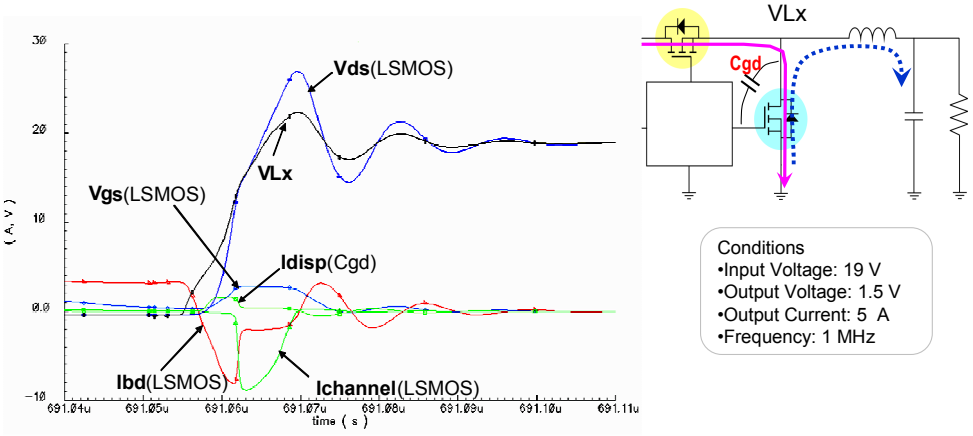
the capacitance of the center junction of the low side MOSFET is then charged.

So, the power loss depends on the reverse recovery charge of the diode, Q**rr**, and the charge of the low side MOSFET center junction, Q_{ds}.

In addition, if self turn-on of the low side MOSFET occurs, additional power loss occurs.

Calculated waveforms of self-turn-on

- Turn-on of HS-MOS causes reverse recovery current (**Ibd**) of body diode of LS-MOS
- As the diode recovers, a large dV/dt is applied across the LS-MOS (**Vds**).
- dV/dt of LS-MOS (**Vds**) induces dV/dt current (**Idisp**) through C_{GD}, and charges the gate
- When the gate voltage ($\mathbf{V_{gs}}$) exceeds V_{TH}, Self-Turn-On occurs (**Ichannel**).



37

This shows the calculated waveforms of the self turn on of the low side MOSFET.

Turn-on of the high side MOSFET causes the reverse recovery of the body diode of the low side MOSFET.

The reverse current of the body diode is shown by the red line.

As the body diode is recovered, the voltage is applied across the low side MOSFET.

This is shown by the dark blue line, Vds.

The large dV/dt of the drain voltage induces dV/dt current through the drain-gate capacitance,

and charges the gate capacitance.

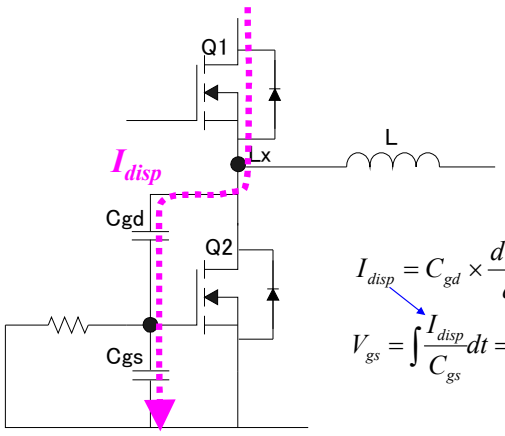
This increases gate voltage, and when the gate voltage exceeds the threshold voltage, the self turn-on occurs.

The gate voltage shown by blue line, exceeds the threshold, and the channel current flows in the low side MOSFET.

This is shown by the green line.

Method to prevent Self-turn-on

- The displacement current(I_{disp}), flowing through C_{gd}, is given by $C_{gd}*dV_{ds}/dt$. I_{disp} charges C_{gs}. The gate voltage(V_{gs}) depends on the ratio C_{gd}/C_{gs}.
- MOSFET design for smaller C_{gd}/C_{gs} is required to prevent self-turn-on.



$$I_{disp} = C_{gd} \times \frac{dV_{ds}}{dt} \quad \text{...Eq.(1)}$$

$$V_{gs} = \int \frac{I_{disp}}{C_{gs}} dt = \int \frac{C_{gd}}{C_{gs}} \cdot \frac{dV_{ds}}{dt} dt = \int \frac{C_{gd}}{C_{gs}} dV_{ds} \quad \text{...Eq.(2)}$$

38

I talk about the method to prevent self-turn-on.

This figure shows the dV/dt current path, flowing through Cgd.

The displacement current is expressed by Eq.(1).

The current charges the gate-source capacitance.

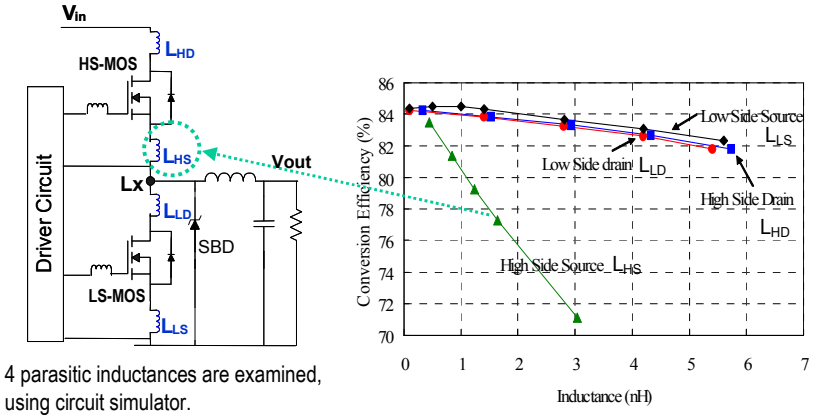
Thus, the gate voltage is expressed by Eq.(2).

Gate voltage depends on the ratio of Cgd/Cgs.

Thus, MOSFET design for small ratio of Cgd/Cgs is effective to prevent self-turn-on.

Influence of parasitic inductances

- It is said that parasitic inductances in power stage influence converter efficiency.
- 4 parasitic inductances are examined, using circuit simulator.
- Inductance of source bonding wire (HS-MOS) greatly reduces efficiency.



4 parasitic inductances are examined, using circuit simulator.

39

This figure shows a typical buck converter.

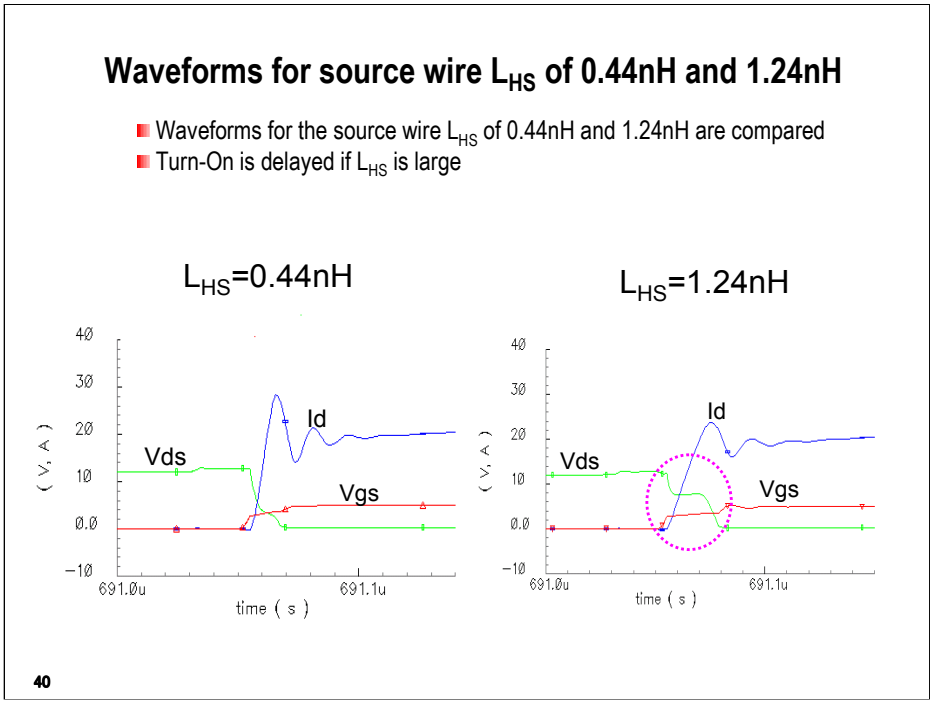
It is often said that the parasitic inductances in the power stage circuits greatly influence the converter efficiency.

So, we examined the four parasitic inductances, using the circuit simulator.

This figure shows the results.

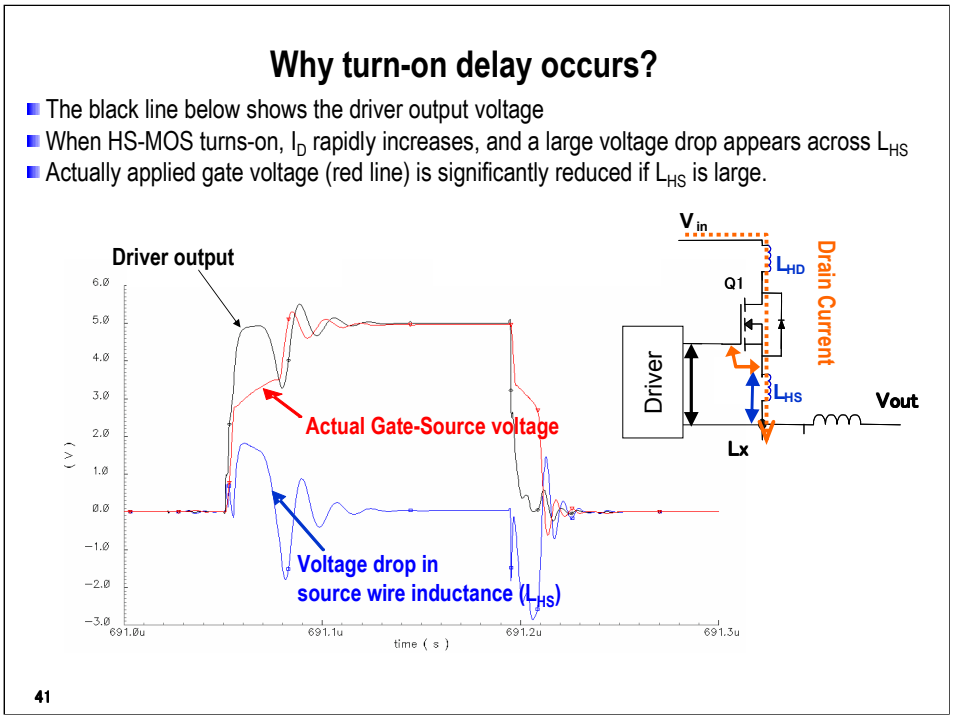
We found that the inductance of the source bonding wires of the high side MOSFET greatly influences the converter efficiency.

Other parasitic inductances also reduce the converter efficiency.



These figures compare the switching-on waveforms between the source wire inductances of 0.44nH and 1.24nH.

If the source inductance is large, turn-on time simply increases and switching loss also increases.



Why does the turn-on delay occurs?

The black line shows the driver output voltage.

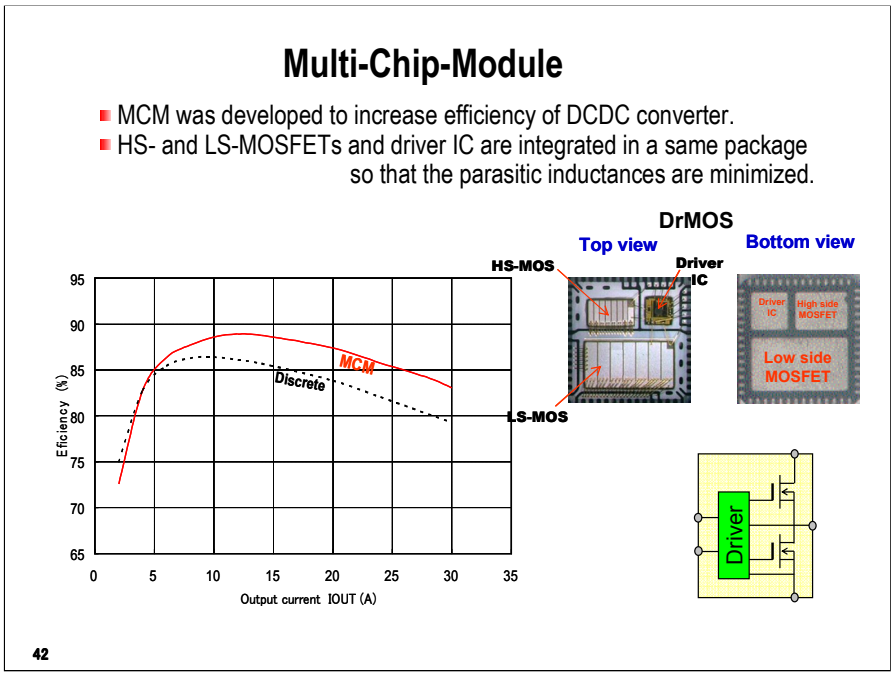
The red line is the actually applied gate-source voltage.

When the MOSFET is turned-on, the drain current rapidly increases, and the large di/dt, induces a large voltage drop in the source wire inductance L_{HS}.

The blue line shows the voltage drop in the source wire inductance.

The actually applied gate voltage is significantly reduced by the voltage drop in the source wire inductance.

This delays the turn-on of the MOSFET.



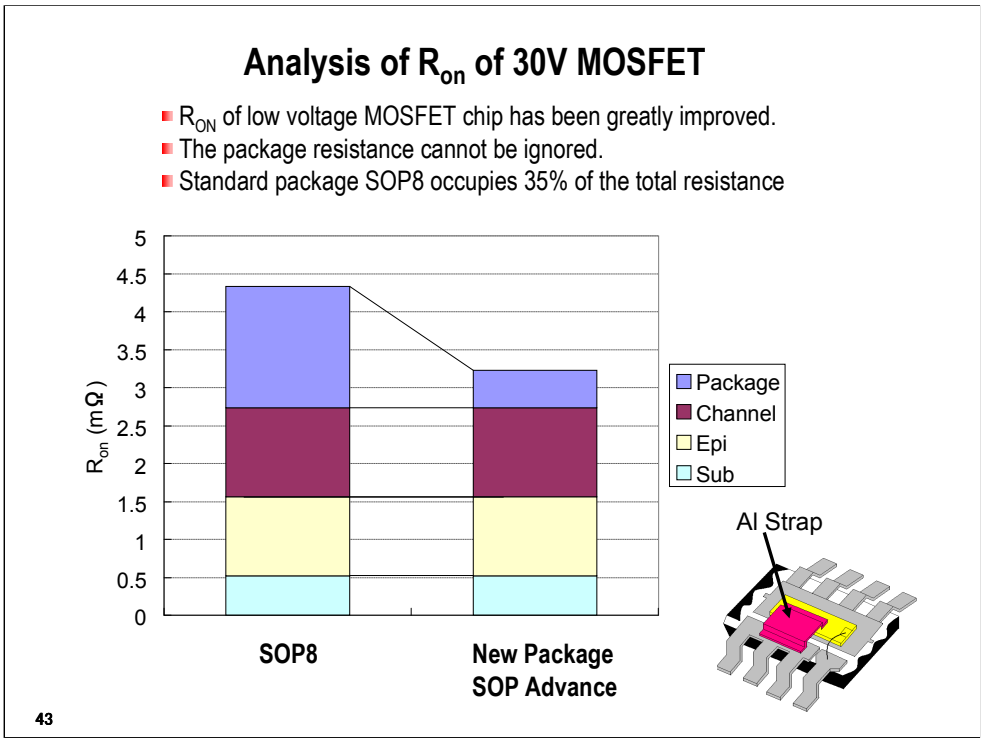
Multi-chip module was developed in order to increase the efficiency of DCDC converter.

High side and low side MOSFETs and driver IC are integrated in a same package so that the parasitic inductances are minimized.

By using the MCM, converter efficiency can be increased, compared with discrete MOSFETs.

Parasitic inductance values can be optimized by using MCM solution.

This is an example of MCM, called Driver MOS, proposed by Intel.



Recently, the on-resistance of low voltage MOSFET chip has been greatly improved, and the resistance of the package cannot be ignored.

This figure illustrates the distribution of on-resistance of 30V high speed MOSFET.

It is seen that the package resistance of SOP8 occupies 35% of the total resistance.

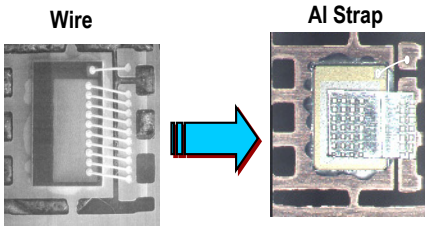
This is unreasonable.

Many MOSFET manufacturers developed new packages that has reduced package resistance.

This new package reduces the package resistance to 1/3rd of the original value.

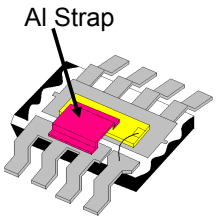
Aluminum Strap

New package uses flat aluminum ribbons instead of bonding wires



Features

- Low Resistance & Inductance
- High reliability
- Low thermal resistance



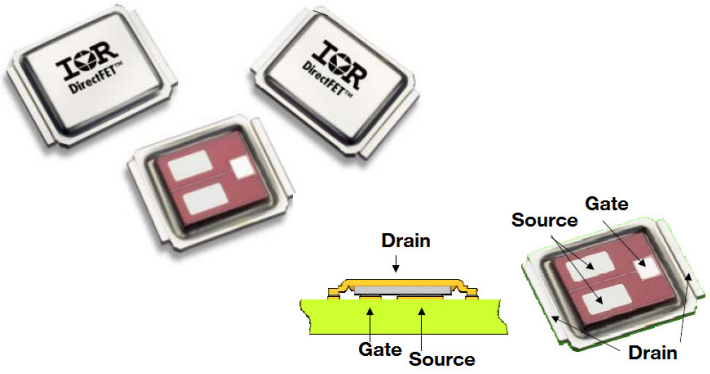
44

The new package introduced new bonding wire called aluminum strap.

This aluminum strap uses flat aluminum ribbons instead of bonding wires, reducing the wire resistance and also the wire inductance.

DirectFET

IR introduced a new package, called DirectFET.
MOSFET chip is exposed to air and is directly mounted on PCBs by solder.

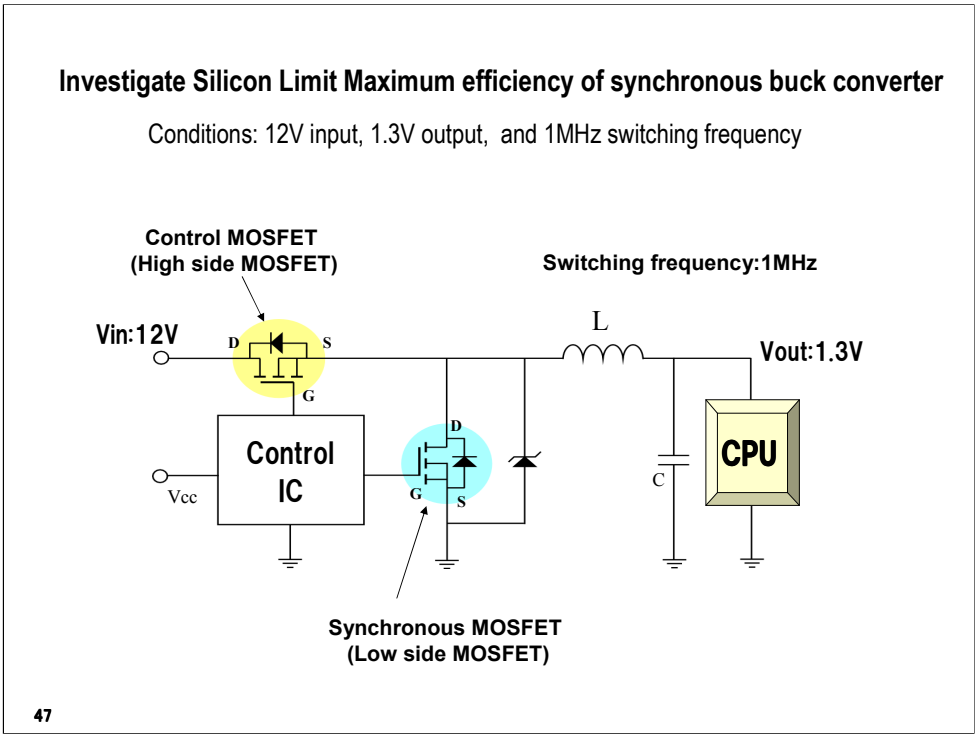


45

IR introduced a new package, called directFET, where the MOSFET chip is exposed to air and is directly mounted on PCBs by solder.

3. High Speed Power MOSFET
—Ideal switching in power MOSFET

Next, I'll talk about power MOSFETs.



From now on, I will investigate the silicon limit maximum efficiency of the synchronous buck converter.

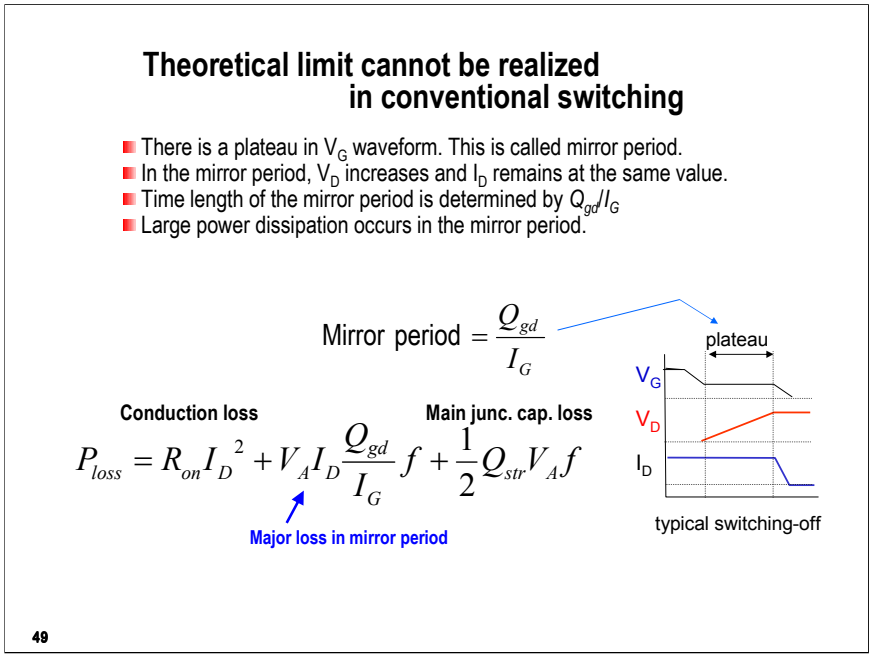
The conditions are 12V input voltage, 1.3V output voltage, and 1MHz switching frequency.

Theoretical Silicon limit of Switching Speed :

$$t_f = \frac{\text{Stored Charge } (Q_{str})}{\text{Drain Current } (I_D)}$$

48

The theoretical limit of the MOSFET switching speed is expressed by this equation: total stored charge in the MOSFET divided by the drain current.



In the conventional switching, the theoretical limit cannot be realized.

This shows the typical switching-off waveforms.

There is a plateau in the gate voltage waveform.

This plateau is so-called mirror period.

In the mirror period, the drain voltage increases and the drain current remains in the same value.

So, a large power dissipation occurs in the mirror period.

This shows the power loss.

First term is conduction loss.

2nd term is switching loss in the mirror period.

The time length is given by Q_{gd}/I_G.

This is the major loss.

The last one is the loss due to the main junction capacitance.

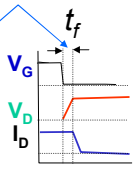
Ideal MOSFET Switching

- If we can supply large gate current, mirror period substantially disappears.
- Major losses are conduction loss and main junction capacitance loss
- Theoretical limit of switching speed is realized in the main junction capacitance loss.

Theoretical limit:

$$t_f = \frac{Q_{str}}{I_D}$$

$$\text{Mirror period} = \frac{Q_{gd}}{I_G} \cong 0$$



Conduction loss	Main junc. cap. loss	
$P_{loss} = R_{on} I_D^{\,2}$	$+\frac{1}{2} Q_{str} V_A f$	Q_{str}: Stored charge in main junc.
$= R_{on} I_D^{\,2}$	$+\frac{1}{2} \frac{Q_{str}}{I_D} I_D V_A f$	

→ **Low Impedance gate drive is a key technology to supply a large gate current to eliminate mirror period.**

50

If we can supply large gate current, mirror period substantially disappears.

And the switching speed approaches the theoretical limit.

In this case, the power loss is the steady state power loss and the main junction capacitance loss.

Theoretical limit of switching speed is realized in the main junction capacitance loss,
because the loss is expressed in this way.

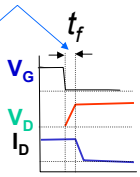
The low impedance gate drive is the key technology to supply a large gate current and to eliminate mirror period in the switching period.

Ideal MOSFET Switching

- Total power loss is the sum of the conduction loss and main junction capacitance loss.
- This takes a minimum value, when the two terms are equal.
- $R_{on}Q_{str}$ should be **new FOM** for ideal switching case.

Theoretical limit:

$$t_f = \frac{Q_{str}}{I_D} \qquad \text{Mirror period} = \frac{Q_{gd}}{I_G} \cong 0$$



Steady state loss **Main junc. cap. loss**

$$P_{loss} = R_{on}I_D{}^2 \qquad + \frac{1}{2}Q_{str}V_A f$$

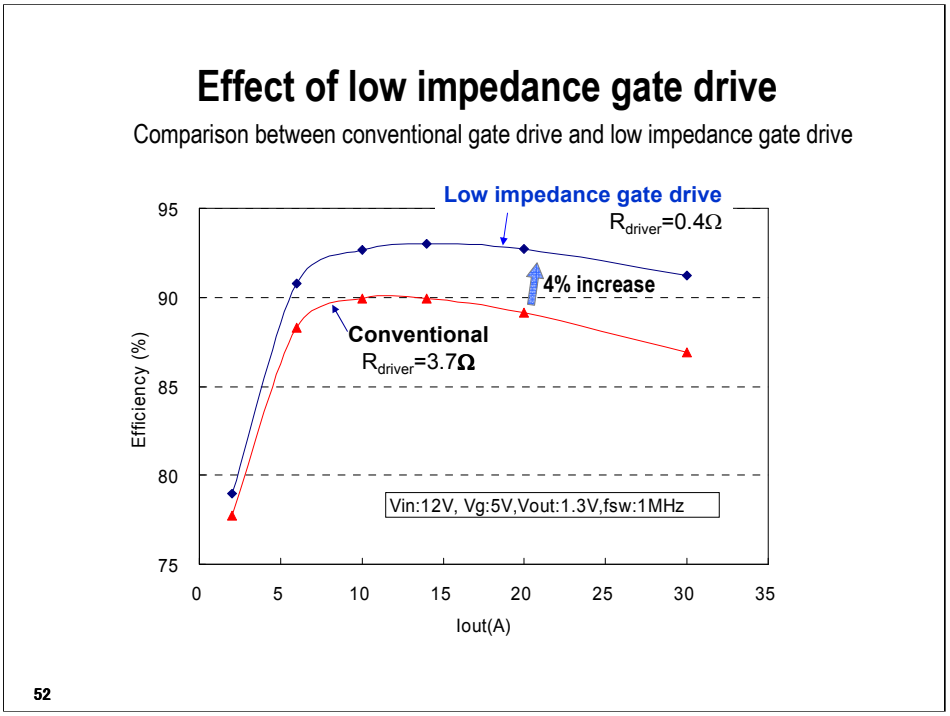
$$= R_{on}I_D{}^2 + \frac{1}{2}Q_{str}V_A f \geq 2\sqrt{R_{on}Q_{str}\frac{1}{2}I_D{}^2 V_A f} \qquad a+b \geq 2\sqrt{ab}$$

New FOM

51

The total power loss is the sum of the conduction loss and main junction capacitance loss. This takes a minimum value, when the two terms are equal to each other, as shown by the inequality equation.

The power loss depends on the value of $R_{on}Q_{str}$. Thus, $R_{on}Q_{str}$ can be the new FOM for the ideal switching case.



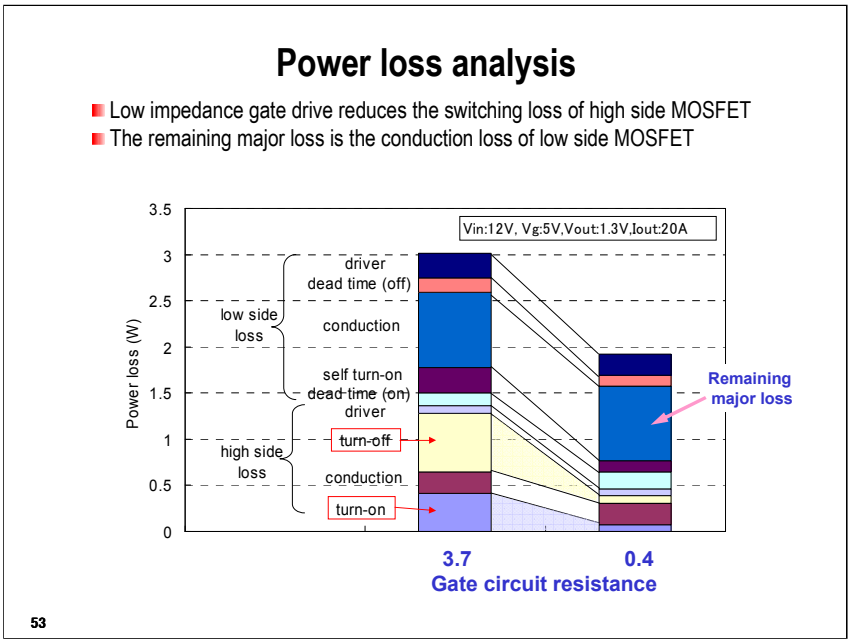
This figure compares conventional gate drive and the proposed low impedance gate drive.

Converter efficiency is shown as a function of output current.

In the conventional gate drive, the typical gate circuit impedance is, for example, 3.7Ω.

If the gate circuit resistance is as low as 0.4ohm, we can increase the efficiency by 4%, and the efficiency exceeds 90% even at 30A output current.

These results imply that efficiency in DC-DC converters can be greatly improved by the low impedance gate drive.

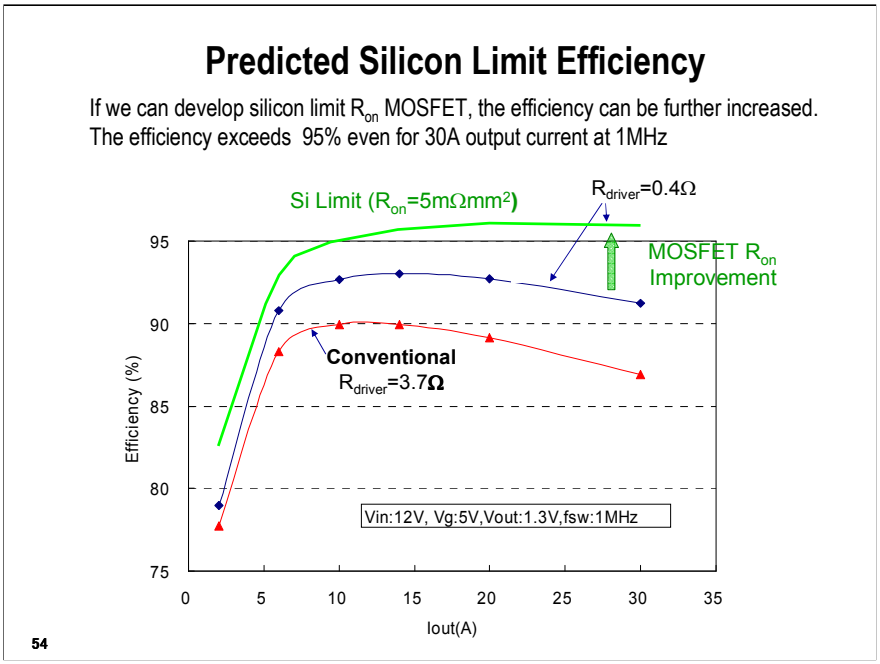


This figure shows power loss analysis.

This portion shows the power losses of the low side MOSFET.

And this portion shows the losses of the high side MOSFET.

It is seen that turn-off loss and turn-on loss of the high side MOSFET are greatly reduced when the low impedance gate drive is used.
Now, the remaining major loss is the conduction loss of low side MOSFET.

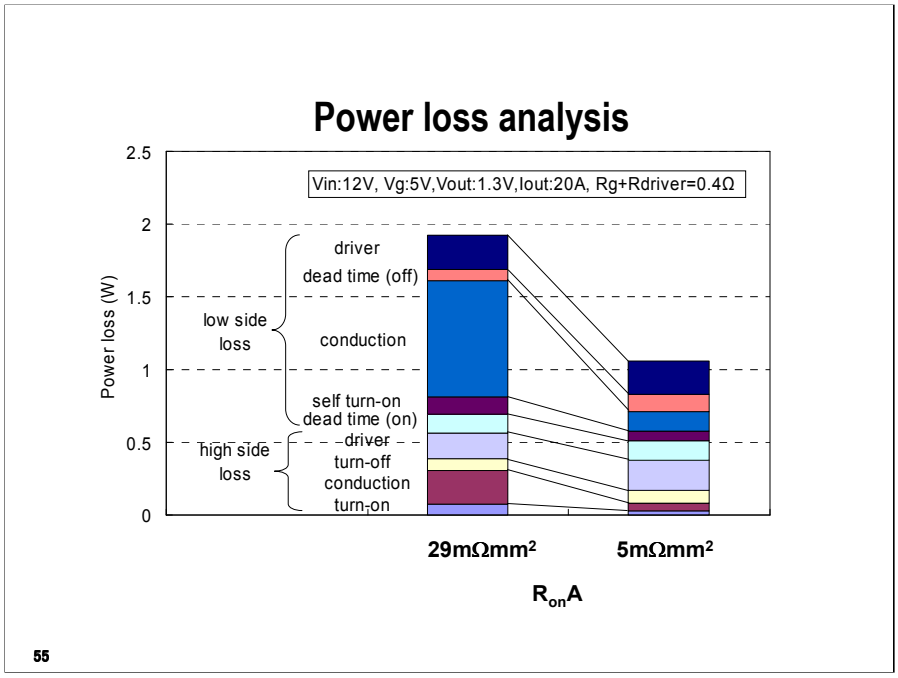


If we can develop silicon limit MOSFET, the efficiency can be further increased.

The silicon limit on-resistance for 30V MOSFET is $5m\Omega mm^2$.

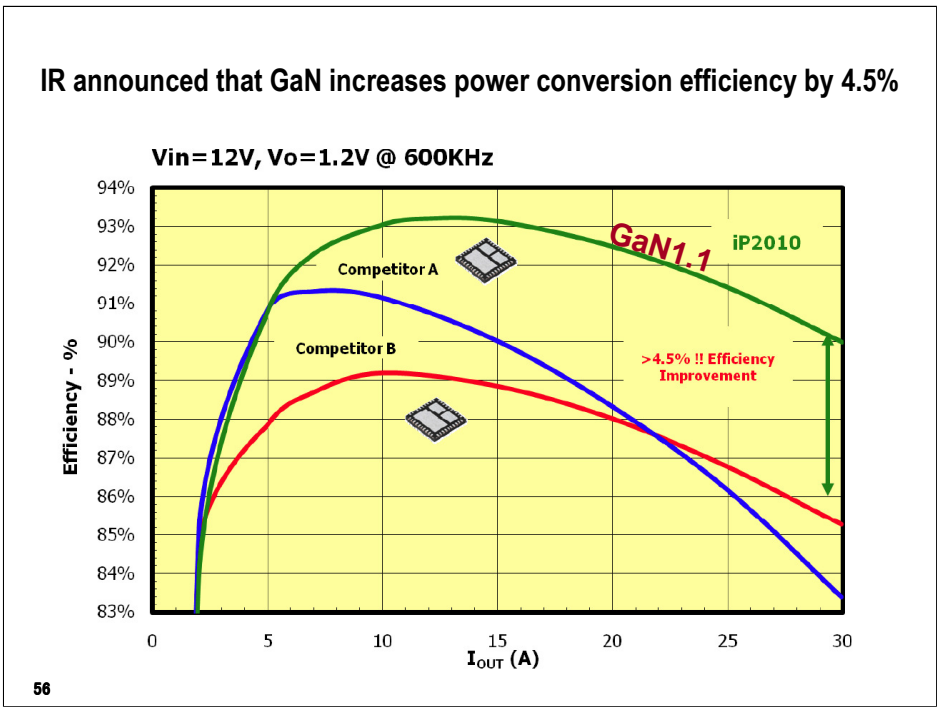
If we can use the silicon limit MOSFET, the efficiency exceeds 95% even for 30A output current at 1MHz.

This is our prediction of silicon limit converter efficiency.

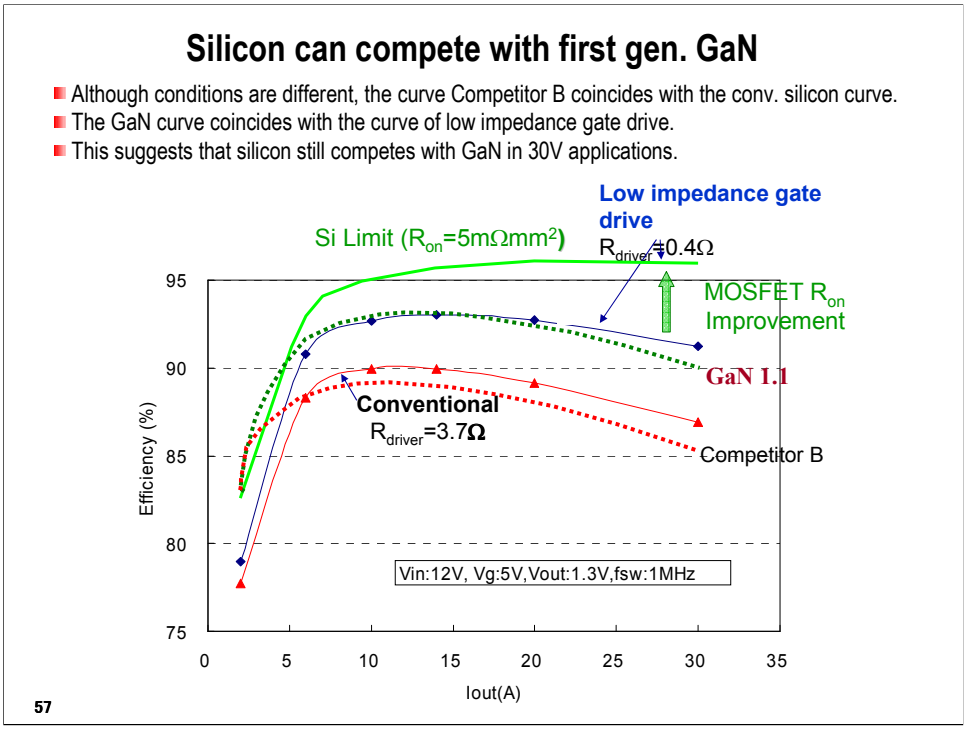


This figure shows power loss distribution.

Now, the conduction loss can be greatly reduced by using the low on-resistance, silicon limit MOSFET.



IR recently anounced that GaN increases power conversion efficiency by 4.5%, compared with Silicon.



For rough estimation, I plotted GaN FET efficiency curve together with our data.

The curve Competitor B coincides with our conventional silicon curve,

And, GaN FET curve coincides with the curve of low impedance gate drive.

Currently available GaN efficiency can be realized by improving silicon device.

Thus, silicon still can compete with currently available GaN FETs.

New FOM

- R_{on} relates to voltage drop and Q_{str} relates to switching speed.
- NFOM is a measure to compare devices under the condition of same switching speed.

$$NFOM = R_{on}Q_{str} = R_{on}J \cdot Q_{str}/J = V_F T_s$$

$$Q_{str}=Q_{oss}$$

For ideal case:

$$Q_{str}=\varepsilon E_C$$

$$R_{on}=4V_{BD}^2/\varepsilon\mu E_C^3$$

$$NFOM=4V_{BD}^2/\mu E_C^2$$

58

I'd like to interpret in more detail about new FOM.

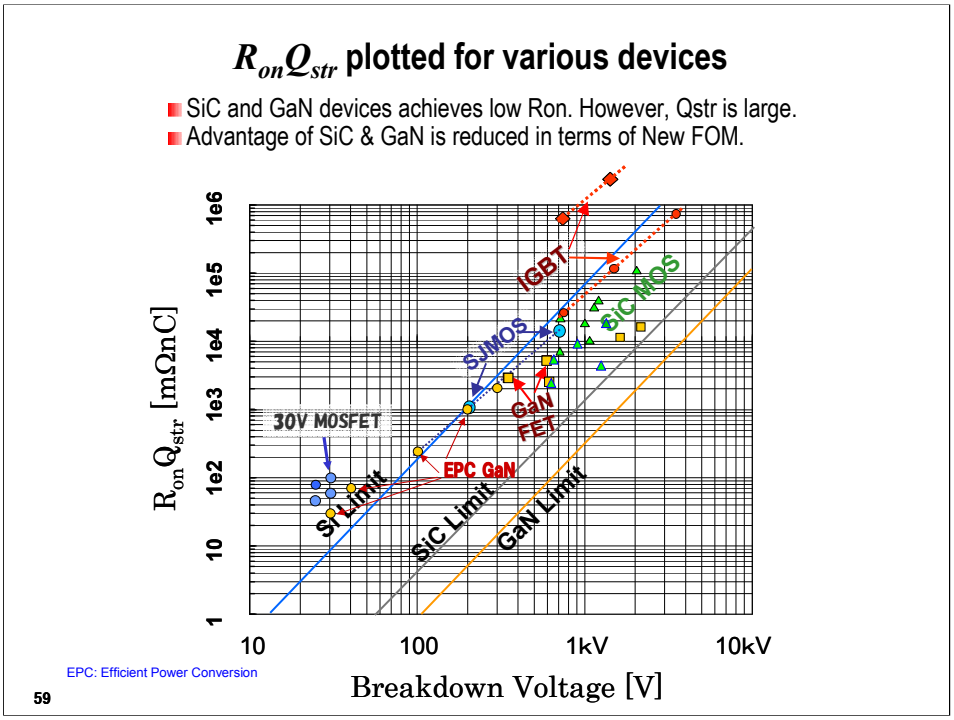
Ron relates to voltage drop and Qstr relates to switching speed.

If we multiply R_{on} by J and divide Q_{str} by J, then, we get forward voltage: VF and switching speed: Ts.

NFOM is a measure to compare devices under the condition of same switching speed.

For ideal cases, Qstr and Ron are expressed as follows:

(ε: permittivity μ: permeability)



This figure shows New FOM for various devices.

SiC and GaN devices achieves low Ron. However, Qstr is large.

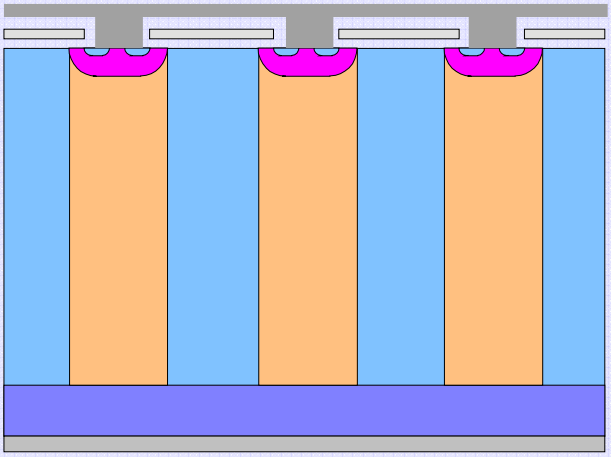
Advantage of SiC & GaN is reduced in terms of New FOM.

For 30V range, conventional silicon MOSFETs are still competing with GaNFETs.

For high voltage region, GaNFETs are better than SJMOSFET.

SiC are definitely better than silicon IGBTs.

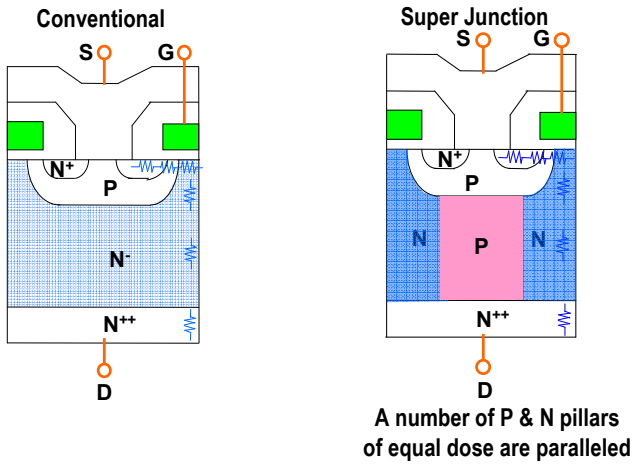
4. Super Junction MOSFET
Cool MOS



60

Super Junction MOSFET

- In SJ MOSFETs, a number of N and P pillars are alternately formed.
- N and P pillars are completely depleted and sustains a high voltage.
- Impurity concentration of the pillars is higher than that of the conv. N^- epi layer.



61

Recently, super junction MOSFET is paid much attention, because it achieves lower on-resistance.

This figure compares the conventional MOSFET and the Super Junction MOSFET.

In the Super Junction MOSFET, a number of n and p layer pillars are alternately formed.

N and p pillars are completely depleted and sustains a high voltage.

A low on-resistance can be obtained because the impurity concentration of the pillars is higher than that of the epi layer of conventional MOSFET.

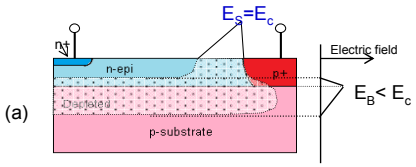
Resurf technology

N-epi is formed on high resistance p-substrate.

(a) If n-epi doping is high, a high electric field appear at n-epi p^+ junction. → Surf. Break Down

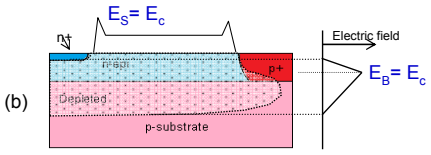
(c) If n-epi doping is low, a high electric field appear at n^+ n^- junction. → Surf. Break Down

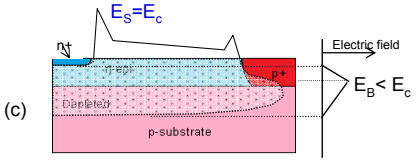
(b) If n-epi is adequately doped, the surface n-epi is uniformly depleted, vertical junction breakdown is realized.

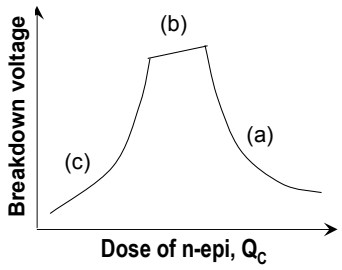


Breakdown occurs at E_C.

Optimum dose: $Q = \varepsilon E_c/q$







62

Resurf technology is frequently used in power IC technology.

This is very much related to the present super junction technology.

The structure is shown in this slide for a diode case.

A high resistance n-epi layer is formed on the high resistance p-substrate.

A positive bias is applied between the n+ and the p+ layer.

If the dose of the n-layer is high, a high electric field appears in the surface between n-epi p+ junction.

Premature junction breakdown occurs before n-epi layer is completely depleted.

On the other hand, the dose of the n-epi is low, a high electric field appear in the n+ n- junction.

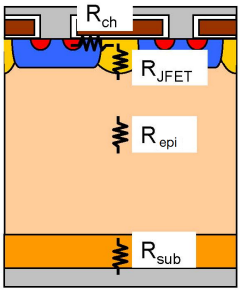
If the dose of the n-epi is adequate, the surface n-epi is completely depleted, and the vertical junction breakdown is realized and, thus, the breakdown voltage is high.

The adequate dose of the junction is give by the equation.

On-resistance of High Voltage MOSFET

Epi-layer resistance occupies a large fraction of total resistance in HV MOSFET

SJ MOSFET reduces R_{epi}



Voltage	100V	600V	1000V
Channel Resistance R_{ch}	0.05mΩcm² (2.6%)	1mΩcm² (1%)	1mΩcm² (0.3%)
JFET Resistance R_{JFET}	0.46mΩcm² (24%)	17mΩcm² (18%)	39mΩcm² (14%)
Epi Layer Resistance R_{epi}	1.4mΩcm² (72%)	76mΩcm² (81%)	240mΩcm² (86%)
Substrate Resistance R_{sub}	0.03mΩcm² (1.4%)	0.1mΩcm² (0.1%)	0.1mΩcm² (0.03%)
Total $R_{on}A$	1.9mΩcm²	95mΩcm²	280mΩcm²

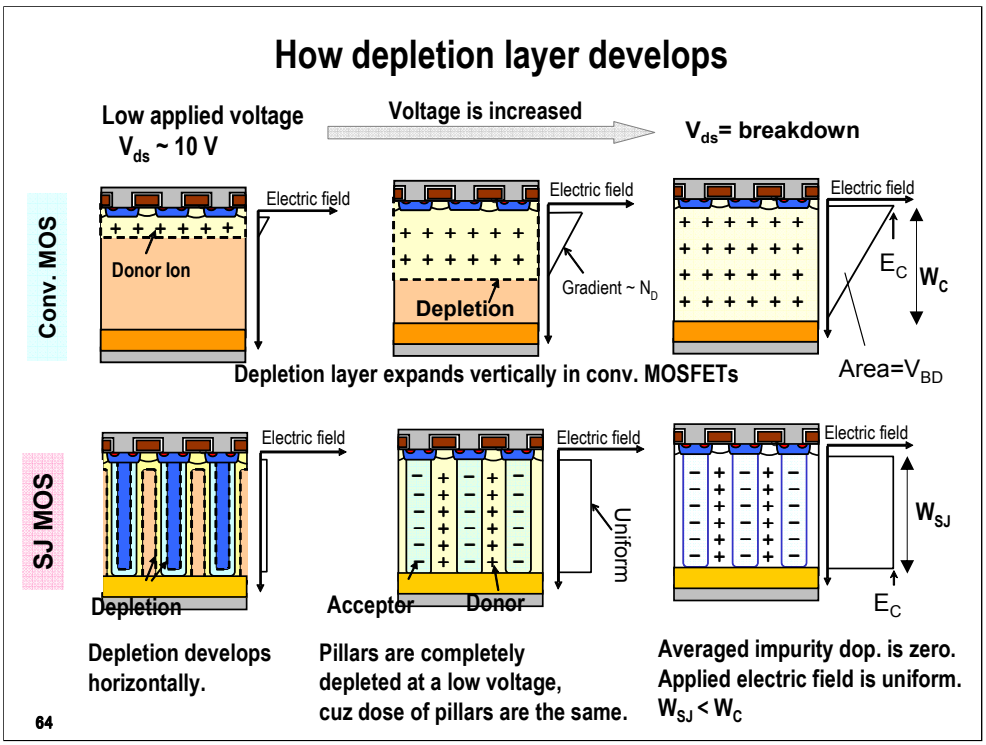
63

This figure shows the on-resistance of high voltage MOSFET.

epi-layer resistance occupies a large fraction of total resistance.

It is very important to reduce this resistance.

Super junction MOSFET is good to reduce the epi-layer resistance.



This figure shows how depletion layer develops in conventional MOSFET and Super junction MOSFET.

In conventional MOSFET, depletion layer expands vertically as the applied voltage increases.

A high electric field appears at the center junction and the shape of the electric field is triangular.

On the other hand, in super junction MOSFETs, initially, the depletion layer expands horizontally.

As the dose of the P and N pillars are equal, all the pillars are completely depleted by a low voltage.

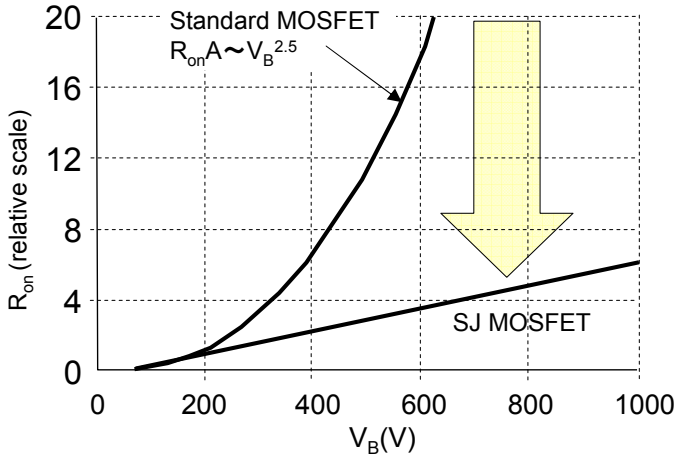
The averaged impurity doping of the pillars are zero, the applied electric field is uniform.

If a further voltage is applied, the magnitude of the electric field increases uniformly.

Thus, the required length of the pillars are shorter than that of the N-base of the conventional MOSFET.

Voltage dependence of R_{on}

- R_{on} of conventional MOSFET increases in proportion to $\mathbf{V_B^{2.5}}$.
- In SJ-MOS, the breakdown voltage depends on the length of the pillars.
- Thus, R_{on} of SJ-MOS increases linearly with breakdown voltage.



65

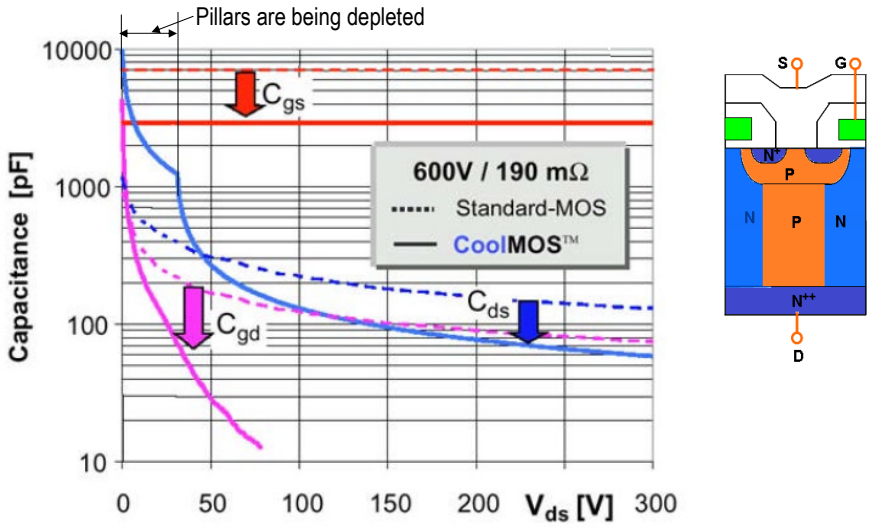
The on-resistance of conventional MOSFET increases in proportion to the 2.5th power of breadown voltage.

In SJMOSFET, the breakdown voltage can be increased by increasing the length of the pillars.

Thus, the on-resistance increases only linearly with breakdown voltage.

Capacitances

- **Cgs** is small because the chip size is small.
- **Cds** is initially large, cuz the area of pn junction Is large, then becomes small when depleted.
- **Cgd** is initially large, cuz impurity of N-pillar is large, then rapidly decreases when depleted.



66

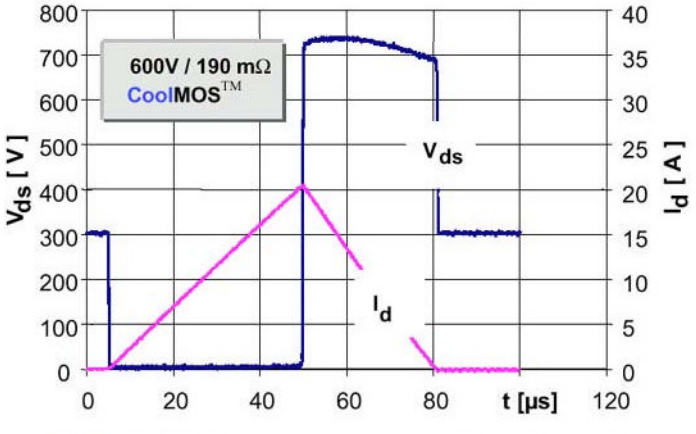
This shows the characteristics of capacitances.

Cgs is small because the chip size is reduced.

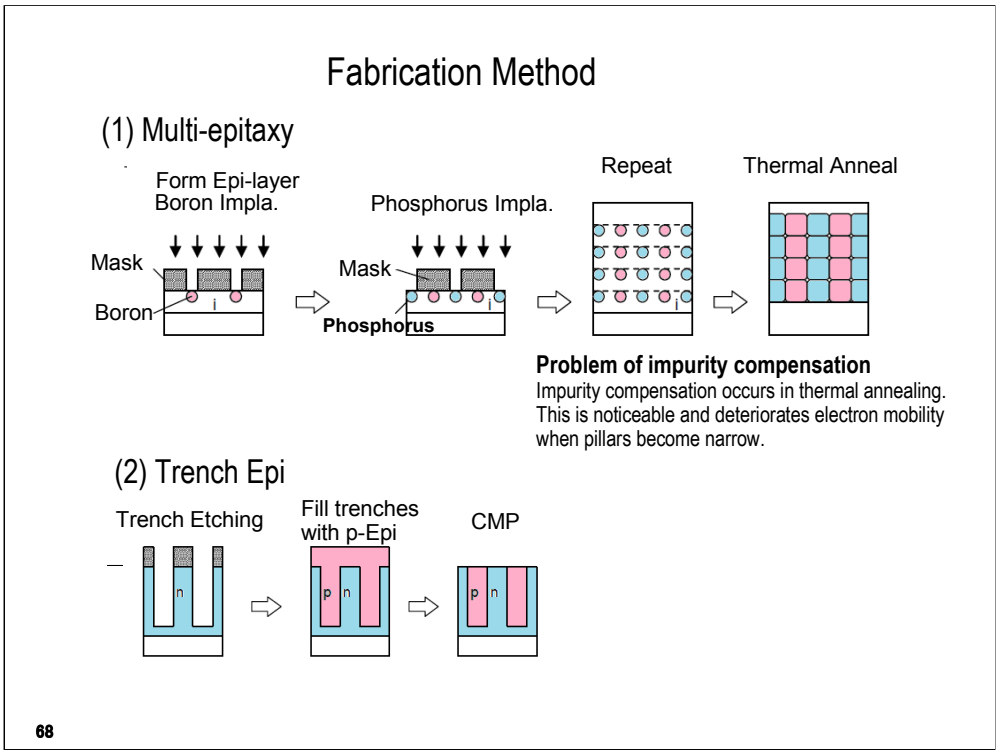
Cds is initaialy large, because area of pn junction Is large. The pillar is initially depleted horizontally, and then vertically, thus Cgd is rapidly decreased.

Cgd is initaialy large, because the impurity of the N-pillar is large, but is decreased, when the pillar is depleted.

Sustaining (UIS) capability of SJMOS



67



This figure shows typical manufacturing methods of super junction MOSFETs.

First one is called multi-epitaxy.

First, form an epi-layer. And using one mask, implant boron. After that, implant phosphorus, using another mask.

Repeat this process several times.

Finally, execute thermal annealing to activate and diffuse impurities.

Second one is called trench epi.

First, prepare n type sillicon layer.

Make deep trenches.

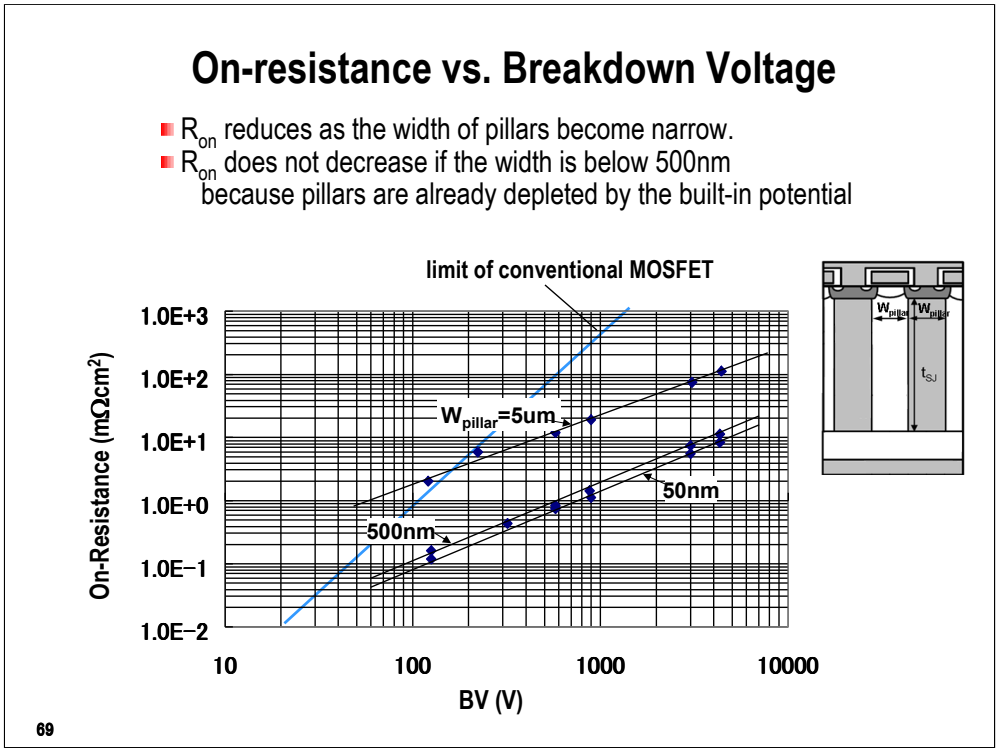
Then, fill trenches with boron doped silicon, using epitaxy.

Finally, make surface planarization using CMP.

Initially, multi-epitaxy was utilized.

However, recently, trench epi is frequently used because impurity compensation occurs in multi-epitaxy, and this reduces the mobility of electrons.

If one try to make SJMOS with narrow pillars, the effect of impurity compensation cannot be neglected.

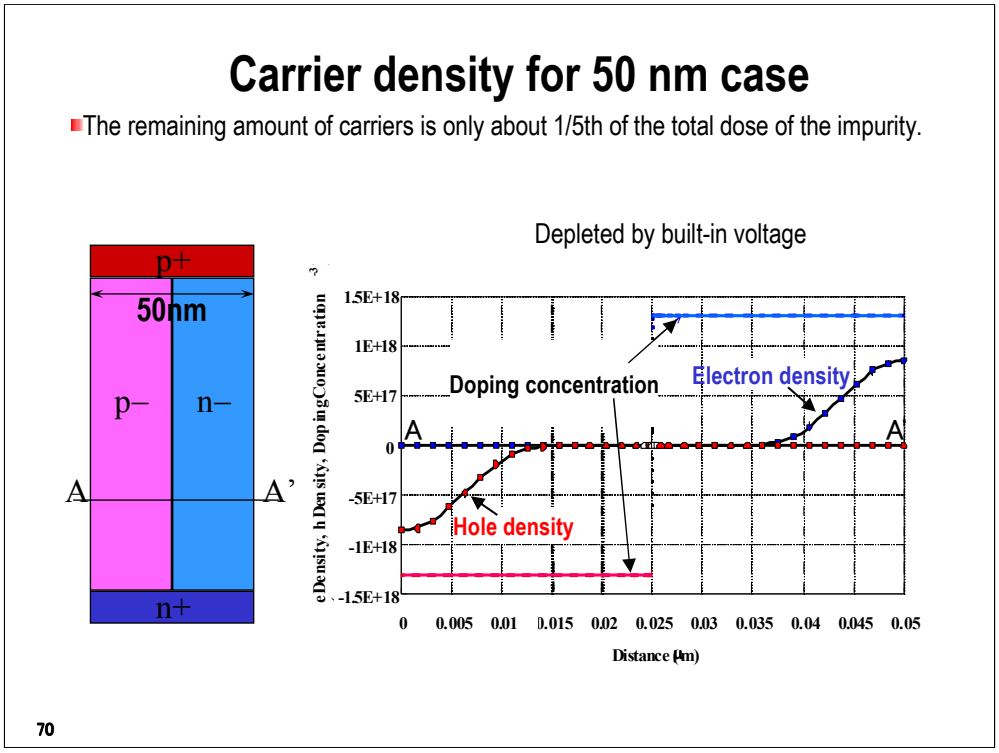


This figure shows calculated on-resistance as a function of breakdown voltage with pillar width as a parameter.

The device on-resistance reduces significantly as the p and n pillar width become narrow.

However, the on-resistance does mot decrease if the pillar width is less than 500nm.

This is because the p and n pillars are depleted by the junction built-in potential if the pillar width is excessively narrow.

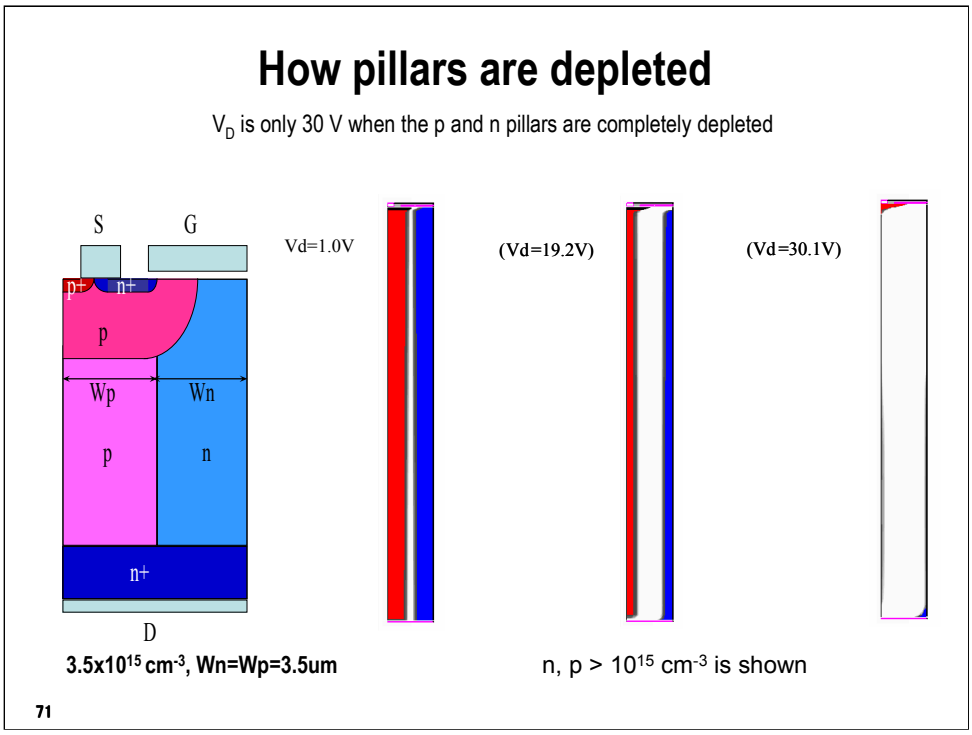


This figure shows the carrier density and the impurity doping concentration when the pillar width is 50nm.

This line is the doping density and this line is the carrier density.

It is shown that the amount of the carrier is only about one fifth of total dose of the impurity.

So, the resistance cannot be decreased.



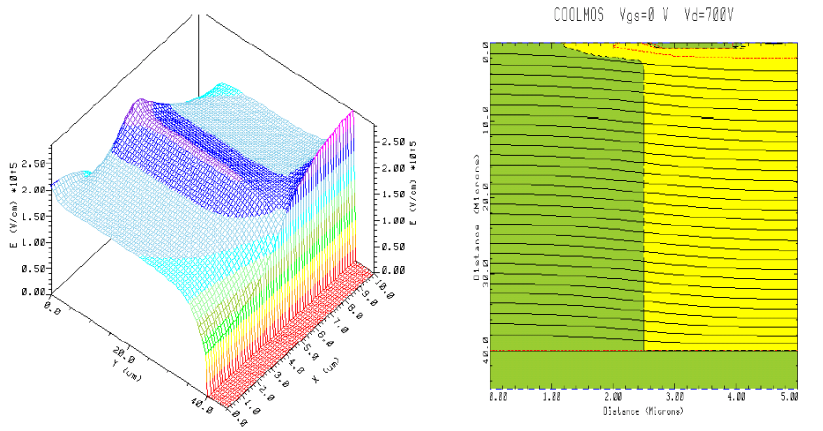
This figure shows how pillars are depleted when a voltage is applied.

The white region shows the depletion layer and the red and blue portions show the area where the carrier density is more than 1e15 cm^{-3}.

The applied drain voltage is only 30 V when the p and n layers are completely depleted.

Calculated electric field distribution

- It is necessary to deplete pillars horizontally.
  Thus, the electric field at the boundary of the pillars is higher than the other region



Zhang 2000

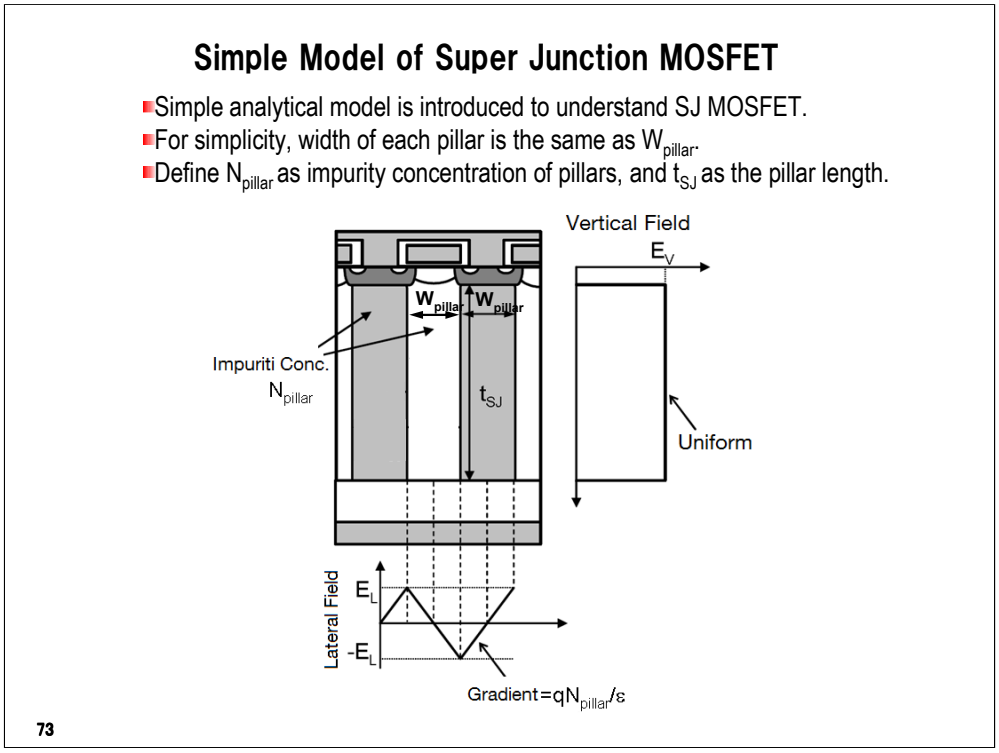
72

This figure shows simulated electric field distribution.

It is necessary to deplete pillars horizontally.

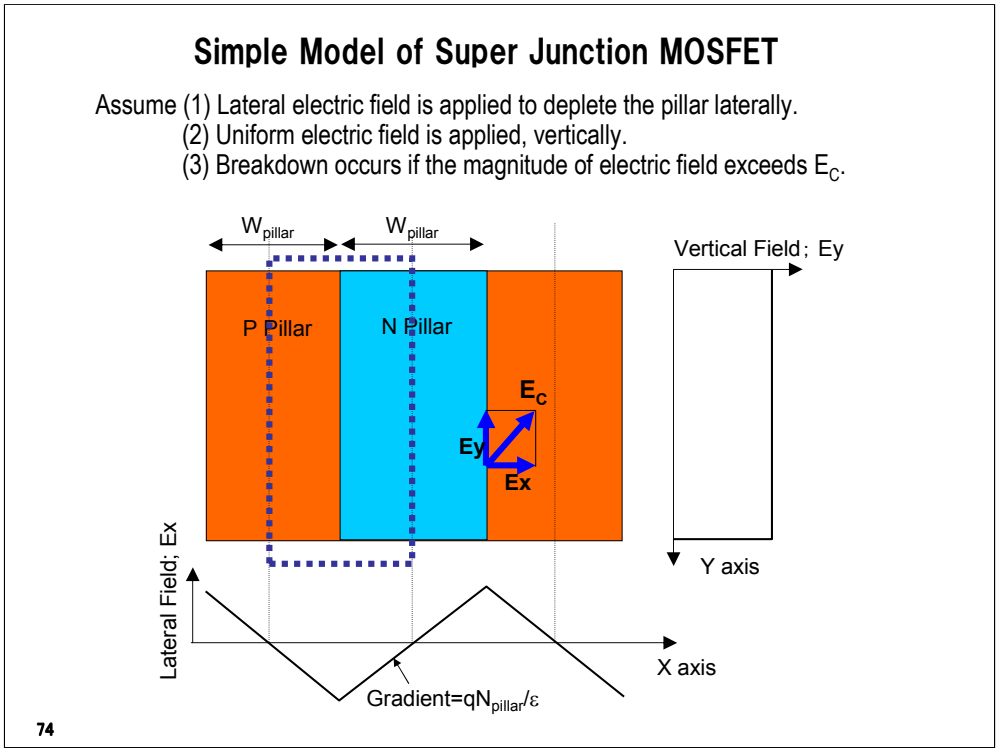
Thus, the electric field at the boundary of the pillars is higher than the other region.

The electric field is almost uniform in the vertical Y-direction.



Now, I introduce a simple analytical model to understand the SJMOS in more detail.

First, we define N_{pillar} as the impurity concentration of pillar, W_{pillar} as the width of pillar.

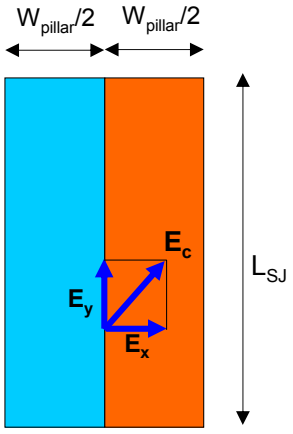


Lateral electric field is applied to deplete the pillar laterally.

For vertical direction, uniform electric field is applied.

It is assumed that breakdown occurs if the magnitude of electric field exceed the critical field, Ec.

Obtain Breakdown Voltage when $E_{Max}=E_c$



It is assumed that the device is in the edge of breakdown.
Total dose Q_{dose} is given by Eq.(1).

$$Q_{dose} = N_{pillar} W_{pillar} \quad \text{...Eq.(1)}$$

$$\varepsilon E_x = qN_{pillar}\frac{W_{pillar}}{2} = q\frac{Q_{dose}}{2} \quad \text{...Eq.(2)}$$

$$E_y = \sqrt{E_c^2 - E_x^2} = \sqrt{E_C{}^2 - (\frac{qQ_{dose}}{2\varepsilon})^2} \quad \text{...Eq.(3)}$$

Breakdown voltage V_B is given as:

$$V_B = E_y t_{SJ}$$
$$= t_{SJ}\sqrt{E_C{}^2 - (\frac{qQ_{dose}}{2\varepsilon})^2} \quad \text{...Eq.(4)}$$

R_{on} is given by:

$$R_{on} = 2\frac{t_{SJ}}{q\mu N_{pillar}} \quad \text{...Eq.(5)}$$

75

Now, we obtain the breakdown voltage of SJ diode.

We assume that the device is in the edge of breakdown.

Total dose of the pillar, Qdose, is Npillar*Wpillar.

The lateral electric field, εEx, to deplete the pillar, is given by q*Npillar*Wpillar.

The magnitude of the electric field must be E_C at the breakdown. Thus, the vertical field is calculated as Eq.(3).

As the electric field is uniform in vertical direction:

The breakdown voltage is given by Ey*L_{SJ}. as is given by Eq.(4).

On-resistance is easily calculated by assuming that all the pillars were N pillar. Then, the layer were uniformly doped with the doping concentration, Npillar.

And the real Ron is double the calculated resistance.

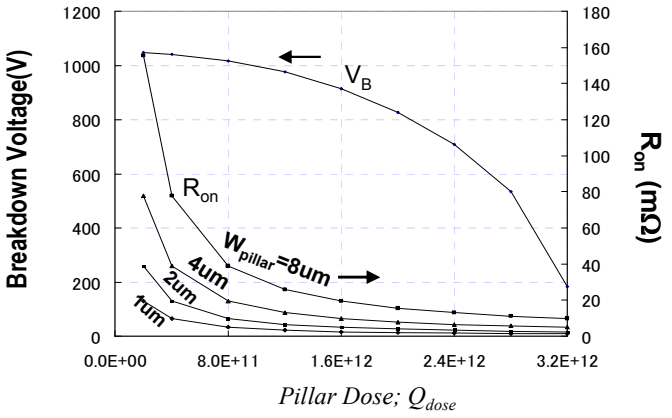
Breakdown Voltage & R_{on}

- V_B and R_{on} [Eq.(4) and (5)] are calculated as a function of Q_{dose} under the condition E_C=2.5x10^{5}V/cm, t_{SJ}=42um.

$$V_B = t_{SJ}\sqrt{E_C^{\,2} - \left(\frac{qQ_{dose}}{2\varepsilon}\right)^2} \quad \text{...Eq.(4)}$$

$$R_{on} = 2\frac{t_{SJ}}{q\mu N_{pillar}} \quad \text{...Eq.(5)}$$

$$N_{pillar} = \frac{Q_{dose}}{W_{pillar}}$$



q	ε	μ	E_C	W_{pillar}	t_{SJ}
1.60E−19	1.04E−12	1350	2.5E+05	parameter	42μm

76

VB and Ron [Eq.(4) and (5)] are calculated as a function of Q_{dose} under the condition E_C=2.5x10^{5}V/cm, t_{SJ}=42um, using the Eq.(4) and (5).

There is a trade-off between breakdown voltage and Ron.

Obtain optimized $R_{on}A$

$$R_{on}A = \frac{2t_{SJ}}{q\mu N_{Pillar}} \quad \text{...Eq.(5)}$$

$$V_B = t_{SJ} \cdot \sqrt{E_C^{\,2} - E_x^{\,2}} \quad \text{...Eq.(6)}$$

$$E_x = \frac{qN_{Pillar}W_{pillar}}{2\varepsilon} \quad \text{...Eq.(7)}$$

} Already derived.

Substitute Eq.(6),(7) for t_{SJ}, N_P in Eq.(5).

$$R_{on}A = \frac{V_B W_{pillar}}{\varepsilon\mu E_x \sqrt{E_C^{\,2} - E_x^{\,2}}} \quad \text{...Eq.(8)}$$

execute $\frac{\partial(\mathrm{R_{on}A})}{\partial \mathrm{E_x}} = 0$ to obtain optimum $\mathrm{E_{x,opt}}$: $E_{x,opt} = \frac{E_C}{\sqrt{2}}$ ...Eq.(9)

substitute Eq.(9) for $\mathrm{E_x}$ in Eq.(8), we have

$$(R_{on}A)_{opt} = \frac{V_B W_{pillar}}{\varepsilon\mu E_x \sqrt{E_C^{\,2} - E_x^{\,2}}} = \frac{2V_B W_{pillar}}{\varepsilon\mu E_C^{\,2}} \quad \text{...Eq. (10)}$$

Substitute Eq.(9) for $\mathrm{E_x}$ in Eq.(6),

$$t_{SJ} = \sqrt{2}\,\frac{V_B}{E_C} \quad \text{...Eq.(11)}$$

77

We showed already that RonA is given by Eq.(5).

Breakdown voltage VB is given by Eq.(6).

Here, Ex is given by Eq.(7).

Now, deleting t_{SJ} and N_P from Eq.(5), using Eq.(6) & (7), RonA is given by Eq.(8).

The optimum Ex is obtained by differentiating Eq.(8) by Ex.

Optimum Ex is given by Eq.(9).

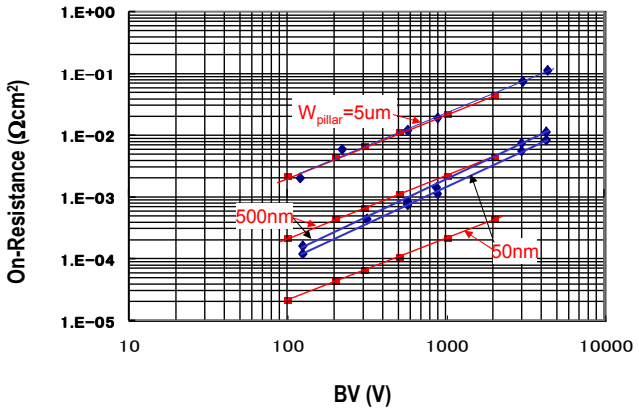
Thus, optimum Ron is given by Eq.(10)

t_{SJ} is given by solving Eq.(6) for t_{SJ} and using Eq.(9).

Analytically predicted R_{on} vs. Breakdown Voltage

- Eq.(10) is plotted (red points) and compared with TCAD results (black points).
- For 50nm case, analytical model does not include the depletion due to built in potential.
- Agreement is quite satisfactory.

$$(R_{on}A)_{opt} = \frac{2V_B W_{pillar}}{\varepsilon \mu E_C^{\,2}} \qquad \text{...Eq. (10)}$$



78

We have derived this analytical expression, Eq.(10).
In this figure, Eq.(10) is plotted (red points) and compared with TCAD results (black points).

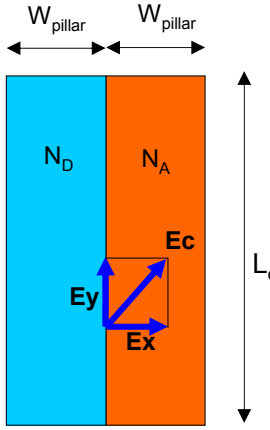
The red points show analytical results, and black points show TCAD results.

Although we use very simple analytical model, the results agree very well with TCAD.

For 50nm case, analytical model does not include the depletion due to the built in voltage.

Thus, the results do not agree with TCAD.

Breakdown voltage degradation due to charge imbalance



If $N_D \neq N_A$, there is charge imbalance.

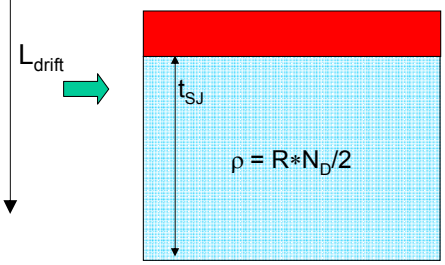
Assume imbalance rate R in such a way that

$$N_D=(1+R)N_{pillar},\ N_A=N_{pillar}.$$

The averaged impurity concentration ρ is given as:

$$\rho*(2W_{pillar}) =(1+R)N_{pillar}W_{pillar} – N_{pillar}W_{pillar} \quad \ldots Eq.(12)$$

$$\rho= R*N_{pillar}/2 \quad \ldots\ldots\ldots Eq.(13)$$



79

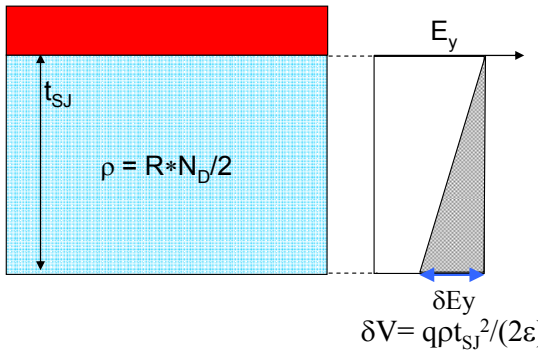
In actual cases, there is charge imbalance.

If the impurity concentration, N_D, of N pillar is higher than that of P pillar by R,

Then, the average impurity concentration, ρ, of the drift layer is given by Eq.(12).

The pillar layers behaves like a resistance layer, whose impurity concentration is r, given by eq.(13).

Voltage reduction δV is calculated in the following.



$\delta V = q\rho t_{SJ}^2/(2\varepsilon)$

The electric field, δEy, which is created by the charge ρ is given as:

$\varepsilon(\delta Ey) = q\rho\, t_{SJ}$ (Gauss's Law)
$\delta Ey = q\rho\, t_{SJ}/\varepsilon$

Voltage reduction is given by shaded triangle area
$\delta V = (q\rho\, t_{SJ}/\varepsilon)*t_{SJ}/2 = q\rho t_{SJ}^2/(2\varepsilon)$

80

Voltage reduction δV is calculated in the following.

The electric field, δEy, which is created by the charge ρ is given as:

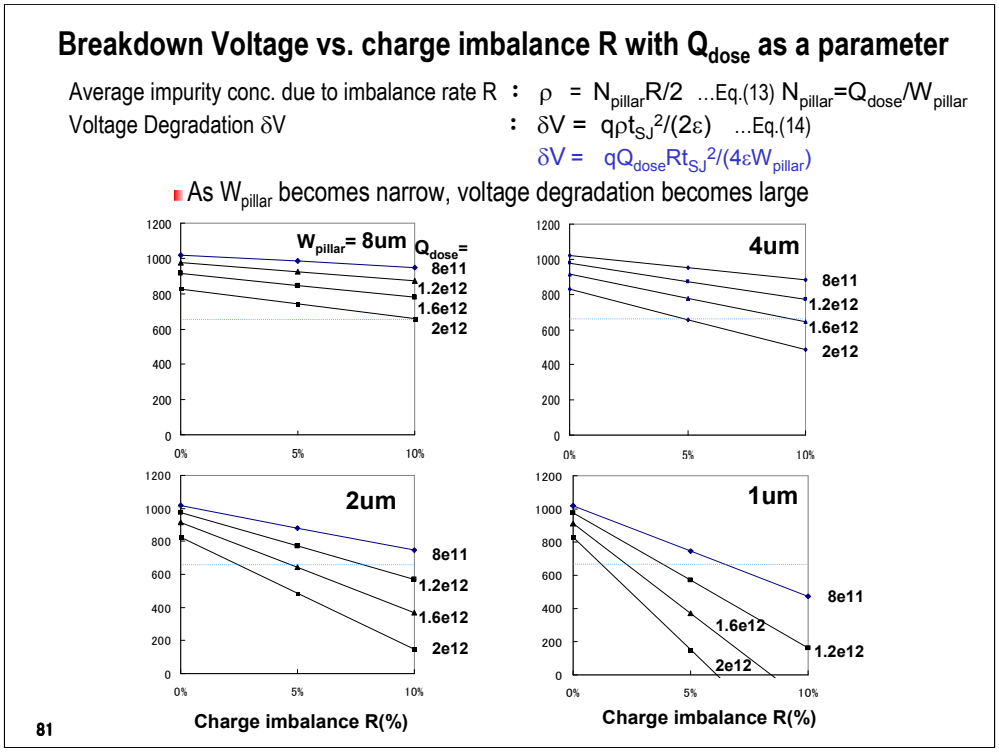
$\varepsilon(\delta Ey) = q\rho\, t_{SJ}$

$\delta Ey = q\rho\, t_{SJ}/\varepsilon$

Voltage reduction is given by shaded triangle area

$\delta V = (q\rho\, t_{SJ}/\varepsilon)*t_{SJ}/2 = q\rho t_{SJ}^2/(2\varepsilon)$

This induces voltage degradation: $q\rho t_{SJ}^2/(2\varepsilon)$

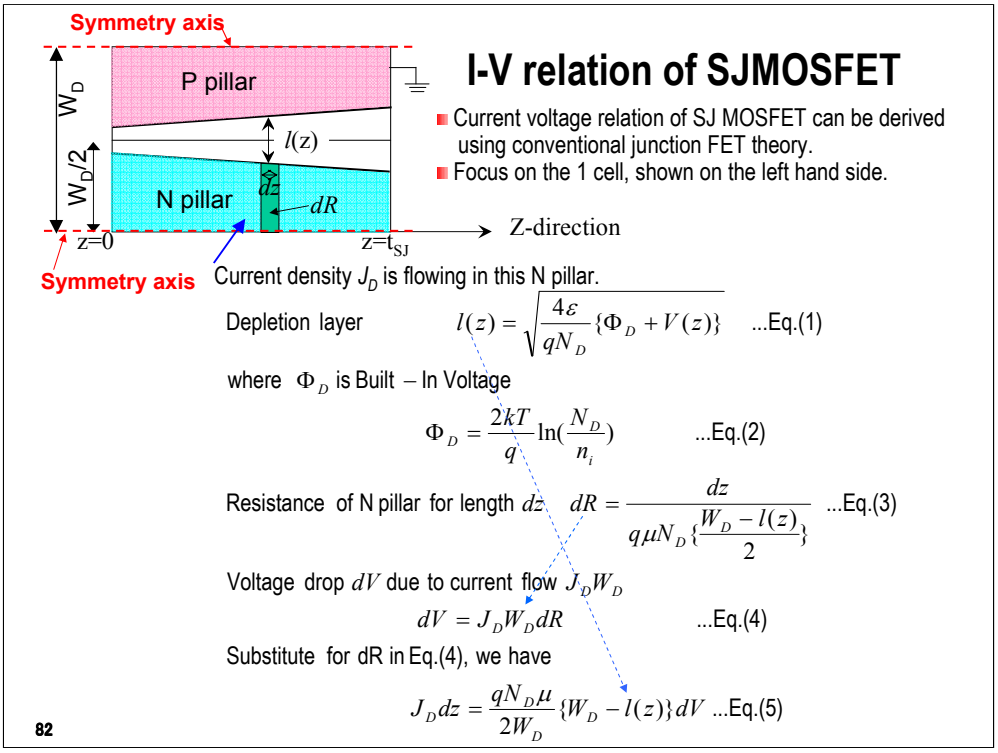


These figures shows the breakdown voltage as a function of charge imbalance rate R with total dose of the pillar as a parameter.

As the width of the pillar decreases, voltage degradation becomes large.

In other words, as the pillar width becomes narrow, the on-resistance can be reduced, however, the effect of charge imbalance becomes large.

So, the impurity concentration must be more precisely controlled.



Current voltage relation of SJ MOSFET can be derived using conventional junction FET theory.

We need to solve only this small area, because of symmetry.

We assume current density is J_D.

A part of N pillar is depleted because of the voltage drop V(z).

The deletion layer width l(z) is given by the equation (1).

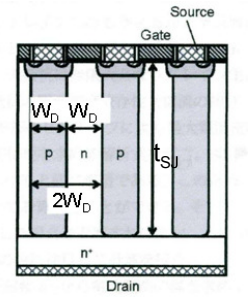
Where Φ(fhi) is the built-in potential of the junction.

The resistance of the N pillar for the length dz is given by Eq.(3).

The magnitude of the current flowing in this pillar is $J_D W_D$.

The voltage drop dV is given as Eq.(4).

Inserting Eq.(3) into Eq.(4), Eq.(5) is derived.



Cite Eq.(5) again [$l(z)$ is replaced by Eq.(1)].

$$J_D dz = \frac{qN_D\mu}{2W_D}\left(W_D - \sqrt{\frac{4\varepsilon}{qN_D}\{\Phi_D + V(z)\}}\right)dV \quad \text{...Eq.(5)}$$

Integrate Eq.(5) from z=0 to t_{SJ} (V=0 to V_D).

$$J_D = \frac{1}{R_{on}A}[V_D - \frac{2}{3\sqrt{V_P}}\{(V_D + \Phi_D)^{\frac{3}{2}} - \Phi_D^{\frac{3}{2}}\}] \quad \text{...Eq.(6)}$$

$$R_{on}A = \frac{2t_{SJ}}{qN_D\mu} \quad \text{...Eq.(7)}$$

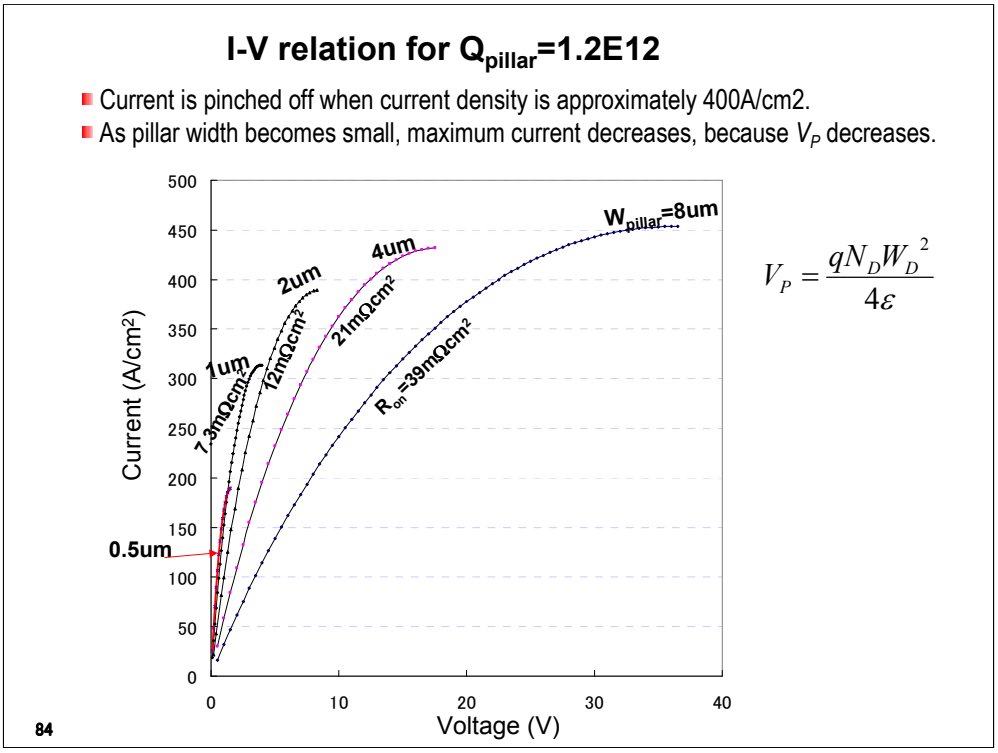
$$V_P = \frac{qN_D W_D^{\,2}}{4\varepsilon} \quad \text{...Eq.(8)} \quad \text{Pinch-off voltage}$$

83

Integrating Eq.(5), from z=0 to z=t_{SJ}, we obtain the final equation (6).

This is the analytical current-voltage relation.

Vp is the pinch-off voltage.



I-V relation when Qpillar=1.2E12.

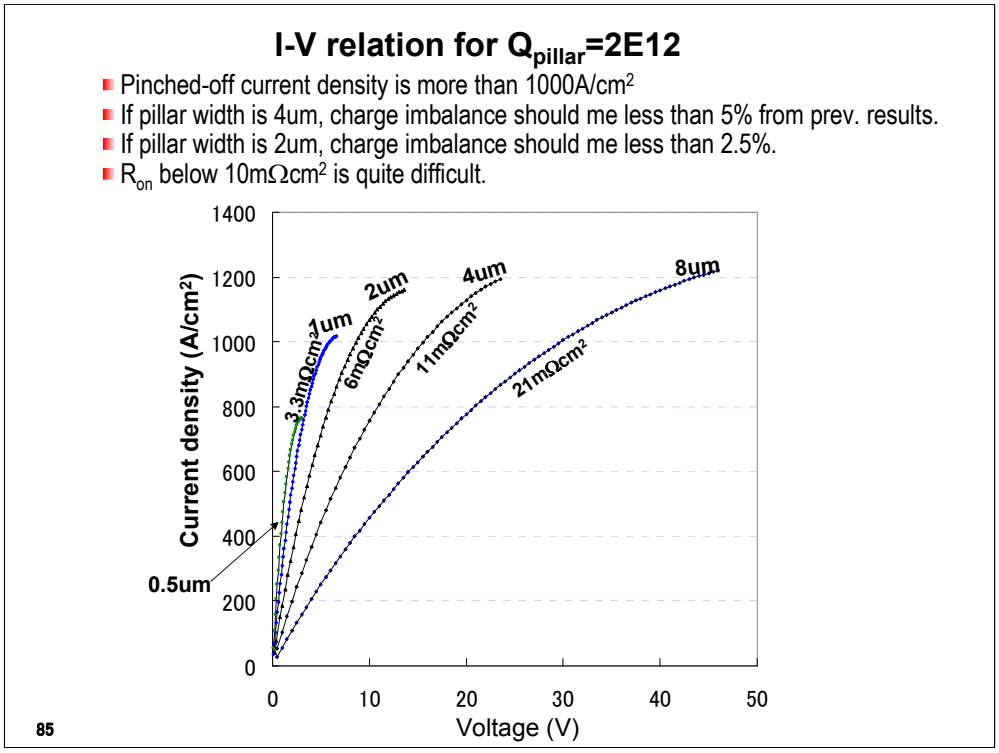
Current is pinched off when current density is approximately 400A/cm^2.

As pillar becomes small, maximum current decreases.

This is because the pinch-off voltage decreases, as the width decreases.

If the pillar width is 2um, on-resistance 12mΩ is obtained.

However, charge imbalance is severe. In order to keep more than 600V, control of charge imbalance should me less than 7% from the previous figure.



I-V relation when Qpillar=2E12.

Current density that is pinched off increases more than 1000A/cm^2.

The control of charge imbalance becomes severe.

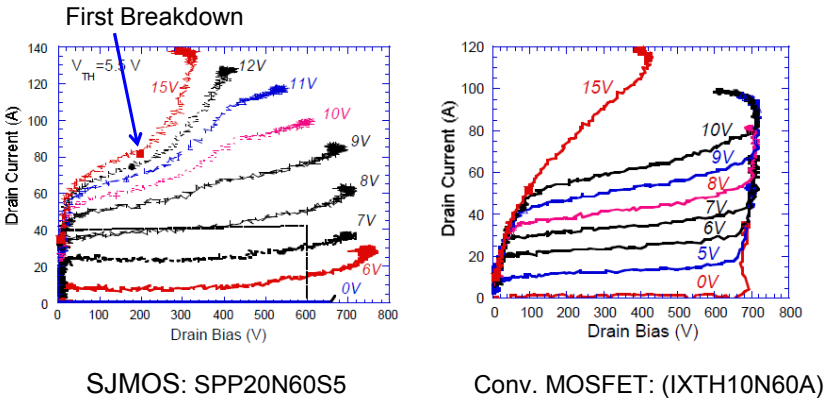
If the pillar width is 4um, the control of charge imbalance should me less than 5%.

If the pillar width is 2um, the control of charge imbalance should me less than 2.5%

Although we use simple analytical model, realizing the on-resistance below 10mΩcm2 is pretty difficult.

Measured FBSOA

- In SJ MOSFET, there is another new breakdown in high current density region.
- This is not seen in conventional MOSFETs.



SJMOS: SPP20N60S5

Conv. MOSFET: (IXTH10N60A)

86

This figure shows the measured current voltage curve of super junction MOSFET called “CoolMOS”.

I also show current voltage relation of conventional MOSFET for comparison.

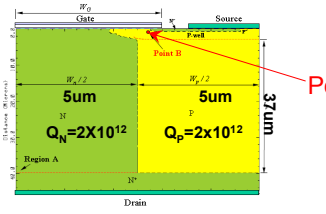
In super junction MOSFET, there is another new breakdown called “first breakdown” in the high current density region.

This is not seen in conventional MOSFETs.

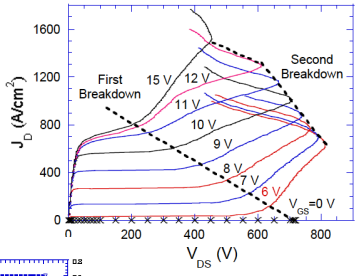
Next, we will analyze this mechanism.

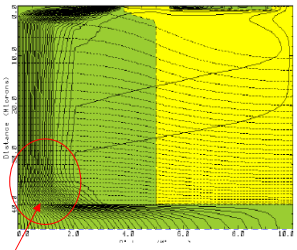
Analysis of FBSOA

- As in N pillar, current flows in saturated velocity, the net space charge is given by Eq.(1).
- Since the negative charge is created, the charge balance is not maintained anymore.
- (1)A large impact ionization takes place and first breakdown occurs.
- (2)If current further increases, second breakdown occurs because of NPN transistor action.

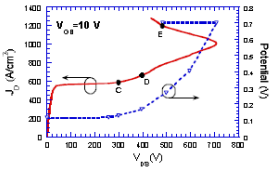


Point B





(1)Impact ionization due to charge imbalance caused by electron current



(2) V-I relation (V_G=10V, red line)
Voltage at point B (blue line)
npn action occurs since voltage Exceeds 0.7V

(1)First Breakdown :
positive charge is compesated due to electron current

$$\rho_n = q(N_D - \frac{J_D}{qv_s}) \quad \text{...Eq.(1)}$$

87

ISPSD'2000 B.Zhang

This was first analyzed by B.Zhang in 2000.

When a high current flows in the super junction MOSFET and a high voltage drop occurs simultaneously, current flows in the saturated velocity in the N pillar. The negative space charge is created by the electron current flow.

The net negative charge is, thus, given by Eq.(1).

Since the negative charge is created in the N-pillar, the charge balance is not maintained anymore.

And a large impact ionization takes place and the drain current increases.

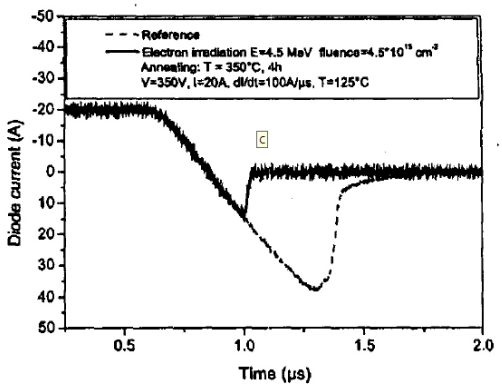
This is called first breakdown.

If the current density further increases, this finally trigger the action of parasitic npn transistor.

And, finally second breakdown takes place.

Problem of inner diode

- When the inner diode is reverse biased, extremely large reverse recovery current flows.
- The reverse current includes the stored carriers as well as current depleting of all the pillars
- After depleting pn junction, the current stops abruptly. SJ MOS diode is snappy.
- Recently, lifetime killer was introduced to improve the diode.



*Schmitt, 2002

88

There is many issues associated with super junction MOSFETs.

As super junction MOSFETs are fabricated by a large number of pillars, there is a large amount of carriers stored inside the pn junction, formed by pillars.

If super junction MOSFETs is reverse biased, the inner diode is forward biased and the current flows.

And then, when the diode is reverse biased, a large reverse recovery current flows.

This reverse recovery current is very large, because all the pillars are depleted simultaneously.

The reverse current includes the stored excess carriers as well as the charges depleting of all the pillars.

As the current, depleting pn junction, stops abruptly, the reverse recovery characteristics of the inner diode is snappy.

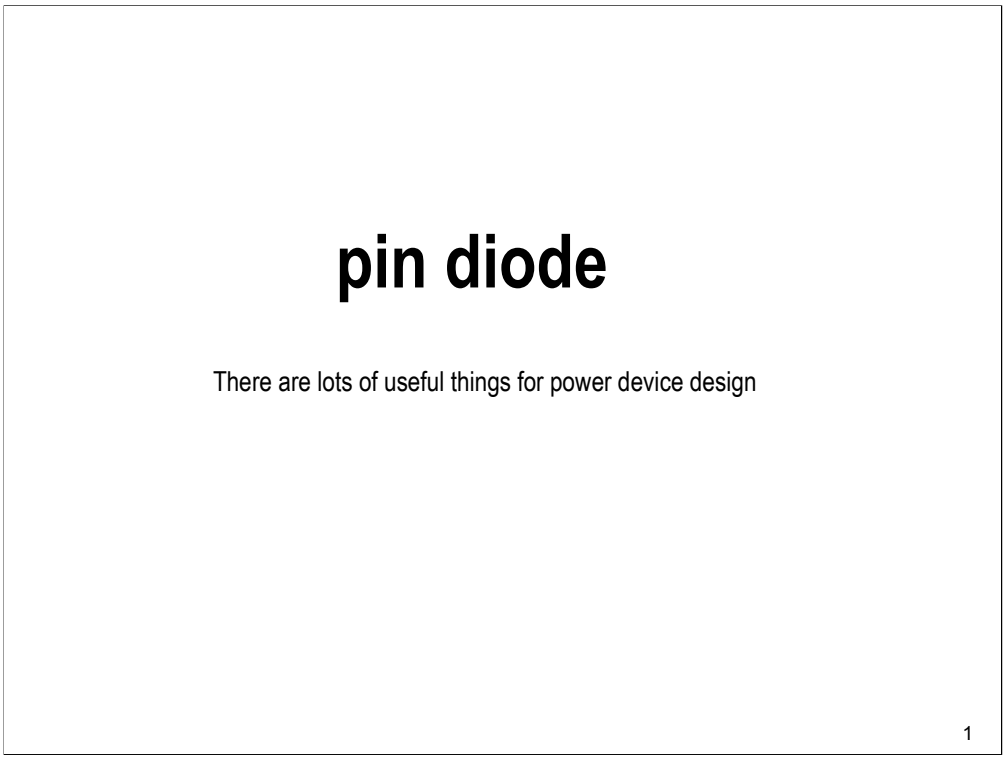
So, super junction MOSFET is not adequate for inverter application.

Recently, a lot of efforts have been done to improve the inner diode.

One of the method is to use lifetime killer.

This figure shows the effect of electron irradiation.

Reverse current can be reduced by using electron irradiation.



From now on, I will talk about pin diode.

Pin diode is very basic and simple structure device.

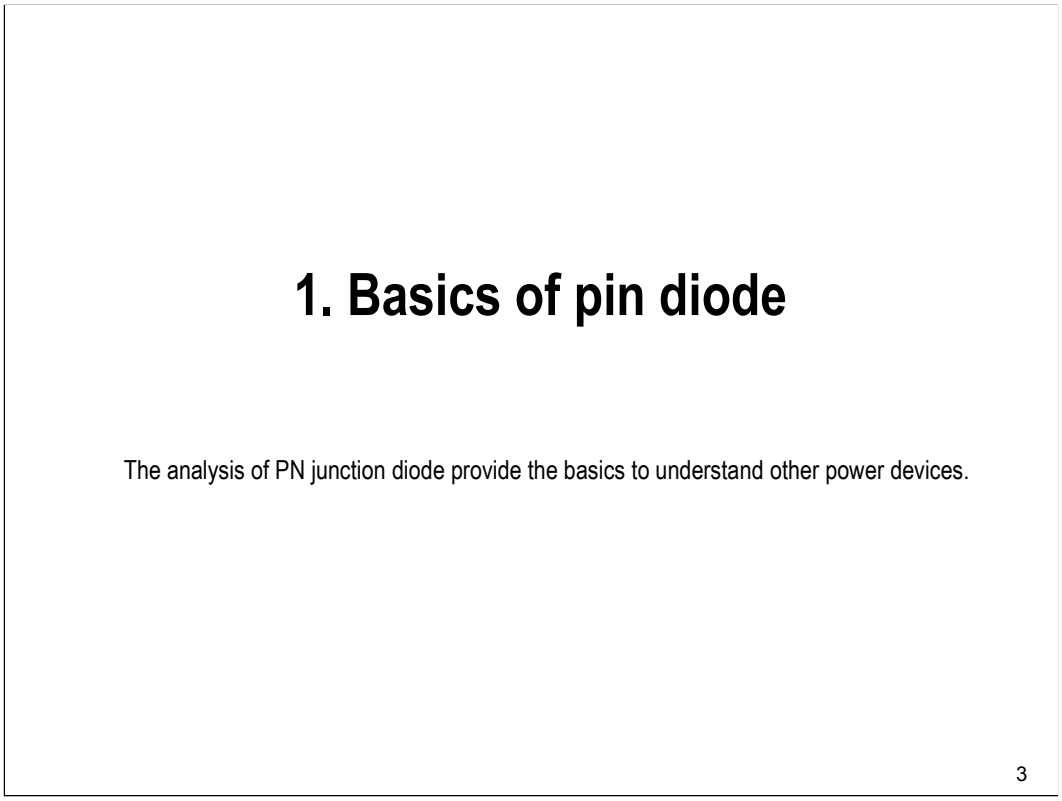
However, there are lots of useful things for power device design.

Chapter 3 pin diode

1. Basics of pin diode
2. Analytical and TCAD analysis of 200V high speed pin diode
3. Discussions on injection efficiency of bipolar transistor

Appendix

1. Basics of Semiconductor
2. Basic Device Equations
3. Ionization Integral
4. Bandgap Narrowing & Fermi Statistics



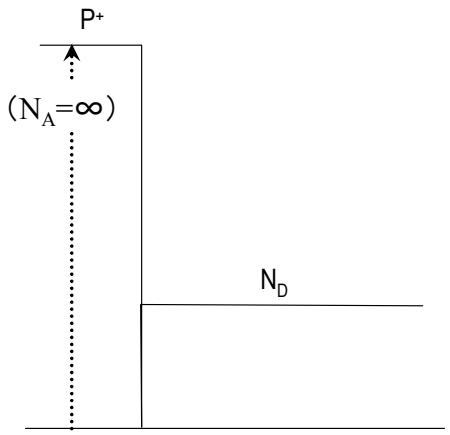
I talk about the electrical characteristics of PN diodes.

The analysis of PN junction diode provide the basics to understand other power devices.

Breakdown voltage for abrupt junction diode

- We first derive breakdown voltage of abrupt P^+N^- junction.

(1)Abrupt junction approximation



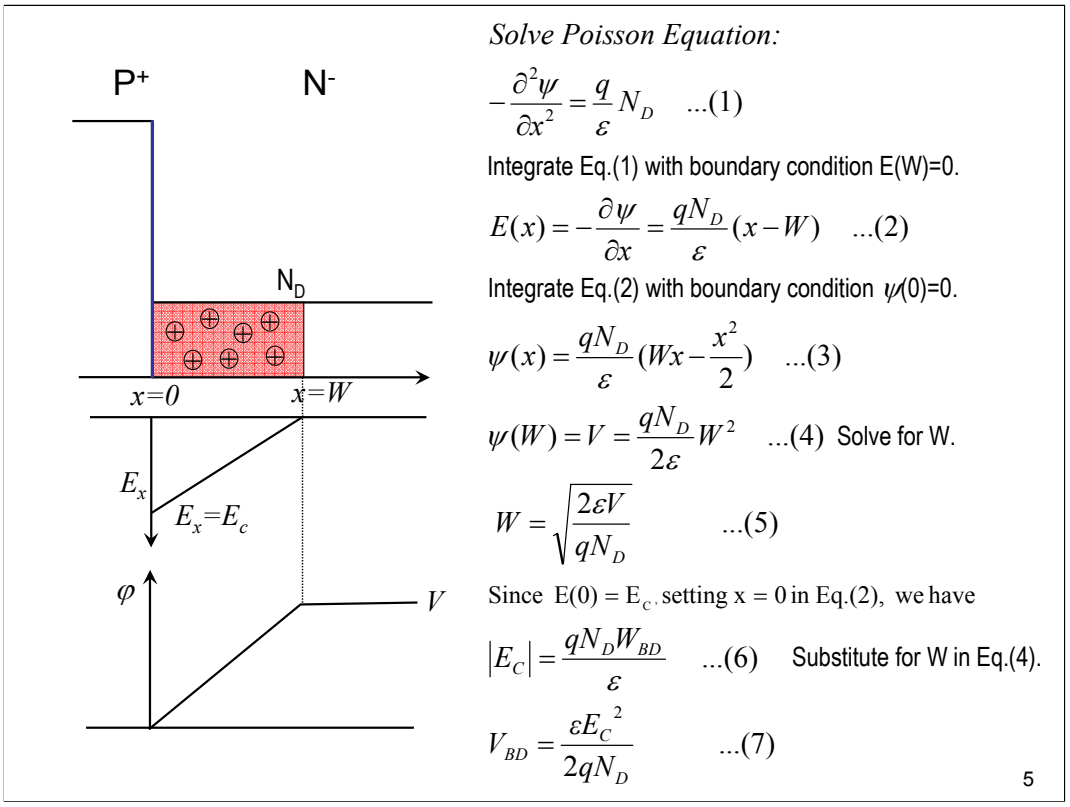
Notations:

N_D: impurity concentration of N^- layer
E_C Critical electric field when Junction breakdown occurs
W depletion layer width
W_{BD} depletion layer width when Junction breakdown occurs
V_{BD} Breakdown voltage
ψ Electric potential

4

First, I talk about breakdown voltage of PIN diodes.

We assume abrupt pn junction.



We solve Poisson equation for the abrupt P^+ N^- junction:

This is Poisson equation.

Integrate Eq.(1) with respect to x, and the electric field is obtained.

The boundary condition is that the electric field is zero at x=W.

Integrate Eq.(2) again, and then, we get static potential psi(x).

The boundary condition is that Phi =0 at x=0.

When x=W, psi(W) is equal to the applied voltage, V, and, thus, we have Eq.(4).

Solving Eq.(4) for W, we have Eq.(5), where the depletion width, W, is expressed as a function of applied voltage V.

When device breakdown occurs, E at x=0 is equal to the critical electric field, E_C, Eq.(6) holds.

W_{BD} is the depletion width when the device breakdown occurs.

Deleting W from Eq.(4), using Eq.(6), we have Eq.(7), where V_B is expressed as a function of critical electric field, E_C.

Applying ionization integral to the abrupt junction

We use effective ionization coefficient, α_{eff}:

$$\alpha_{eff} = \alpha_n = \alpha_p = 1.8 \times 10^{-35} E^7 \qquad \text{...Eq.(8)}$$

$$E(x) = -\frac{\partial \psi}{\partial x} = \frac{qN_D}{\varepsilon}(x - W) \qquad \text{...Eq.(2)}$$

Substituting for E in Eq.(8), Ionization integral is given as:

$$\int \alpha_{eff} dx = \int_0^W 1.8 \times 10^{-35} \left(\frac{qN_D}{\varepsilon}|x - W|\right)^7 dx = 1 \qquad \text{…Eq.(9)}$$

The solution of this equation gives the critical depletion layer width at breakdown

$$W_{BD} = 2.67 \times 10^{10} N_D^{-7/8} \qquad \text{…Eq.(10)}$$

Using Eq.(4), V_{BD} is given by:

$$V_{BD} = 5.34 \times 10^{13} N_D^{-3/4} \qquad \text{…Eq.(11)}$$

Using Eq.(6), critical electric field at breakdown is:

$$E_C = 4010 N_D^{1/8} \qquad \text{...Eq.(12)}$$

Other similar expressions for Breakdown voltage

$$V_B = 60\left(\frac{E_G}{1.1}\right)^{\frac{3}{2}}\left(\frac{N}{10^{16}}\right)^{-\frac{3}{4}} \qquad \text{...(13) from Sze}$$

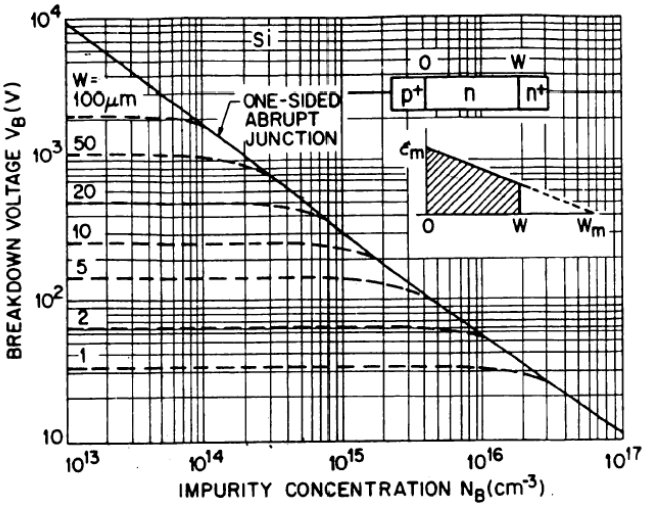
$$V_B = \frac{1}{2}\left(\frac{8}{B}\right)^{\frac{1}{4}}\left(\frac{\varepsilon}{qN}\right)^{\frac{3}{4}} = 563 \cdot \left(\frac{N}{4\times 10^{14}}\right)^{-\frac{3}{4}} \qquad \text{...(14) from Lutz}$$

($B = 2.107 \times 10^{-35}\ cm^6/V^7$)

7

Breakdown Voltage for Punch-through PIN diode

If the resistivity of the i-region is high enough, the depletion layer reach N^+ region.
The approximate breakdown voltage is obtained by calculating the area of the shaded region
The V_{BD} depends mostly on the thickness of the i-region, if the resistivity is high enough.
I shows the theoretical results from Sze's book.



8

If the resistivity of the i-region is high enough, the depletion layer extend throughout the i-region.

The approximate breakdown voltage can be obtained by calculating the area of the shaded region assuming the maximum electric field is the critical electric field, E_C, given by Eq.(12).

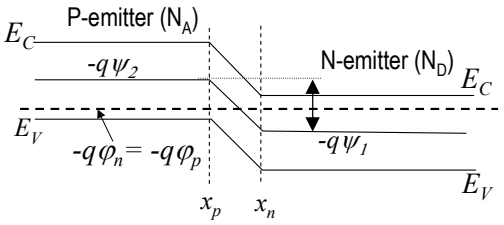
The breakdown voltage depends mostly on the thickness of the i-region, if the resistivity of the i-region is high enough.

I shows the theoretical results from Sze's book.

This is very useful chart in order to confirm the ideal breakdown voltage of the pin diode.

Built-in Potential of PN junction

If p- and n-type silicon make a pn junction there is a potential difference across the junction. The potential difference is called "built-in potential."



$$n = n_i \exp\frac{q}{kT}(\psi - \varphi_n) \quad \text{...Eq.(15)}$$

$$p = n_i \exp\frac{q}{kT}(\varphi_p - \psi) \quad \text{...Eq.(16)}$$

Quasi-Fermi potential is constant throughout the device in thermal equilibrium.

$set\ \ \varphi_p = \varphi_n = \varphi_0$, then we have

$$\text{In N - emitter, } \ n = N_D = n_i \exp\frac{q}{kT}(\psi_1 - \varphi_0)$$

$$\text{In P - emitter, } \ p = N_A = n_i \exp\frac{q}{kT}(\varphi_0 - \psi_2)$$

Multiply the above two equations :

$$\text{Built} - \text{in Potential} = \psi_1 - \psi_2 = \frac{kT}{q}\ln\frac{N_D N_A}{n_i^2}$$

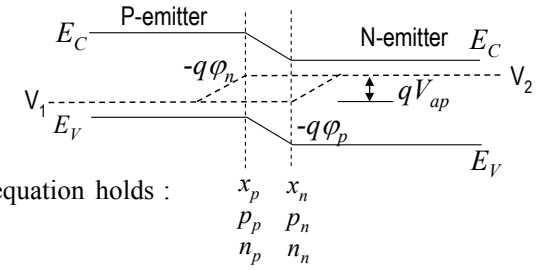
9

Basic relations for forward biased PN junction

- Transition region is the region that the potential is changing from the value of the P/N emitters.
- Within the transition region [x_p, x_n], φ_p and φ_n are assumed to be constant.
- In P/N emitter, quasi-Fermi potential of majority carrier is const.
  and is equal to the voltage of the ohmic contact.

$$n = n_i \exp\frac{q}{kT}(\psi - \varphi_n) \quad \text{...Eq.(15)}$$

$$p = n_i \exp\frac{q}{kT}(\varphi_p - \psi) \quad \text{...Eq.(16)}$$



In the transitio n region, the following equation holds :

$$q(\varphi_p - \varphi_n) = V_{ap} \quad \text{...Eq.(17)}$$

At the edges of the transitio n region :

$$p_n \cdot n_n = n_p \cdot p_p = n_i^2 \exp(\frac{q}{kT}(\varphi_p - \varphi_n)) = n_i^2 \exp(\frac{qV_{ap}}{kT}) \quad \text{...Eq.(18)}$$

$$p_{n0} \cdot n_n = n_{p0} \cdot p_p = n_i^2 \quad \text{...Eq.(19)} \quad \text{0 denotes the value in thermal equilibrium}$$

$$n_p \cdot p_p = n_i^2 \exp(\frac{qV_{ap}}{kT}) = n_{p0} \cdot p_p \exp(\frac{qV_{ap}}{kT}) \rightarrow n_p = n_{p0} \exp(\frac{qV_{ap}}{kT})$$

10

First, I'd like to mention basic relations in pn-junction.

The transition region is defined as the region that the potential is changing from the value of the P- or N-regions.

Within the transition region of the pn-junction, the two quasi-Fermi potentials are assumed to be constant and are equal to the values at the edges of the N- or P-regions, x_p x_n.

Inside the N-emitter or the P-emitter, the quasi-Fermi potentials of the majority carriers are constant and is equal to the voltage of the ohmic contact.

Thus, the equation (17) holds.

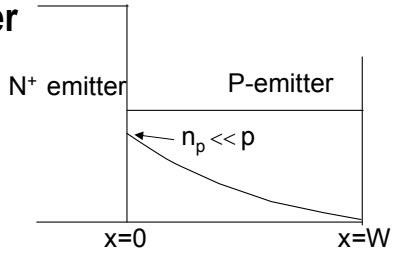
The product of the carrier densities n and p within the transition region of the junction [x_p,x_n] is given by the Eq.(18).

The product in the thermal equilibrium is given by Eq.(19).

Using Eq.(19), the value n_p is expressed by the product of the thermal equilibrium value n_{p0} and the factor $\exp(qV_{app}/kT)$.

Minority carrier current in P-emitter

N^+P junction is forward biased
and electrons are injected into P-emitter.
Low injection level is assumed: $n << N_A = p$



$$q\frac{\partial n_p}{\partial t} = \frac{\partial J_n}{\partial x} - qR \qquad \ldots\text{Eq.(20)}$$

$$\frac{\partial n_p}{\partial t} = D_n\frac{\partial^2 n_p}{\partial x^2} - \frac{n_p - n_{p0}}{\tau_n} = 0 \qquad \ldots\text{Eq.(21)}$$

Consider steady state case $\frac{\partial n_p}{\partial t} = 0$

$$\frac{\partial^2 n_p}{\partial x^2} - \frac{n_p - n_{p0}}{L_n^{\,2}} = 0, \text{ where } L_n = \sqrt{\tau_n D_n} \qquad \ldots\text{Eq.(22)}$$

Boundary Conditions
$$\begin{cases} n_p = n_{p0}\exp(\frac{qV_{ap}}{kT}) & (at\ x = 0) \quad \ldots\text{Eq.(23) [from Eq.(18,19)]} \\ n_p = n_{p0} & (at\ x = W) \quad \ldots\text{Eq.(24)} \end{cases}$$

$$n_p = n_{p0} + n_{p0}(\exp(\frac{qV_{ap}}{kT}) - 1)\frac{\sinh((W - x)/L_n)}{\sinh(W/L_n)} \qquad \ldots\text{Eq.(25)}$$

11

I will talk about the minority carrier current in the P-emitter.

N^+P junction is forward biased and electrons are injected into the P-emitter.

We assume low injection level .

And, thus, $n << N_A = p$.

The electron continuity equation is given by Eq.(20).

As electrons flow by diffusion, Equation (20) reduces to Eq.(21).

In the Eq., the subscript, p means that the variable is associated with the P-emitter.

In the steady state, the derivatives with respect to time is zero.

Thus, the equation we solve is Eq.(22).

The solution is given by Eq.(25), based on the boundary conditions: Eqs.(23) and (24).

Electron current flows by diffusion at the junction edge x=0, and is given by Eq.(26).

$$J_n = qD_n \left.\frac{\partial n_p}{\partial x}\right|_{x=0} \qquad \text{...Eq.(26)}$$

Substituting Eq.(25) for n_p in Eq.(26), the electron current at the junction x=0, is given by

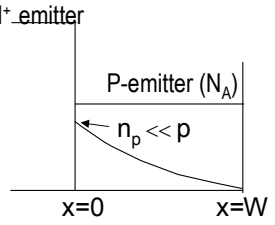
$$J_n = -\frac{qD_n n_{p0}(\exp(\frac{qV_{ap}}{kT}) - 1)}{L_n \tanh(\frac{W}{L_n})} \qquad \text{...Eq.(27)}$$

$$\left[\sinh(x) = \frac{e^x - e^{-x}}{2}, \cosh(x) = \frac{e^x + e^{-x}}{2}, \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}\right]$$

12

Substituting Eq.(25) into Eq.(26), the electron diffusion current at the junction x=0, is given by Eq.(27).

Thick p-emitter case: W/L_n>>1



Considering W/L_n >>1, Eq.(25) reduces to Eq.(28).

$$n_p = n_{p0} + n_{p0}(\exp(\frac{qV_{ap}}{kT}) - 1)\exp(-\frac{x}{L_n}) \quad \text{...Eq.(28)}$$

$$J_n = qD_n \left.\frac{\partial n_p}{\partial x}\right|_{x=0} = -\frac{qD_n n_{p0}}{L_n}(\exp(\frac{qV_{ap}}{kT}) - 1) \quad \text{...Eq.(29)}$$

Eq.(29) is expressed by Eq.(30).

$$J_n = -J_{ns}(\exp(\frac{qV_{ap}}{kT}) - 1) \quad \text{...Eq.(30)}$$

$$J_{ns} = \frac{qD_n n_{p0}}{L_n} = q\frac{n_i^2 D_n}{N_A L_n} \quad \text{...Eq.(31)}$$

◆We derive another useful Eq.(33). By setting x=0 in Eq.(28), we have

$$n_p(x=0) = n_{p0}\exp(\frac{qV_{ap}}{kT}) \quad \text{...Eq.(32)}$$

Using Eq.(32), Eq.(29) reduces to

$$J_n = -\frac{qD_n}{L_n}(n_p(0) - n_{p0}) \approx -\frac{qD_n n_p(0)}{L_n} \quad \text{...Eq.(33)}$$

13

When the P-emitter is very thick, we can assume W/Ln>>1.

Then, Eq.(25) reduces to Eq.(28).

The electron diffusion current is given by Eq.(29).

The electron current is also expressed by simple equations (29) and (30).

Next, we derive another useful expression.

We will use this expression, afterwards.

At x=0, n_p (n sub p) is given as Eq.(32).

Then, Eq.(29) is expressed also by Eq.(33), neglecting the 2nd term.

Analogously, the hole current density in the N^+ emitter is given by:

$$J_p = -J_{ps}(\exp(\frac{qV_{ap}}{kT}) - 1) \quad \text{...Eq.(34)}$$

$$J_{ps} = q\frac{D_p p_{n0}}{L_p} = q\frac{D_p n_i^2}{L_p N_D} \quad \text{...Eq.(35)}$$

Total current density is given by:

$$J = \left|J_{ps}\right| + \left|J_{ns}\right| = J_s(\exp(\frac{qV_{ap}}{kT}) - 1) \quad \text{...Eq.(36)}$$

$$J_s = qn_i^2(\frac{D_p}{N_D L_p} + \frac{D_n}{N_A L_n}) \quad \text{...Eq.(37)}$$

This is the ideal I-V relation of pn junction diode.
J_s is called as "saturation current."

14

Analogously, the hole current density in the N+ emitter is given by Eq.(34) and (35).

Final total current density of the pn junction is given by Eq.(36).
This is what we call ideal I-V relation of junction diode.

Very thin emitter case: W<<L_n

Considering W<<L_n, Eq.(25) reduces to Eq.(38).

$$n_p = n_{p0} + n_{p0}(\exp(\frac{qV_{ap}}{kT}) - 1)(1 - \frac{x}{W}) \quad \text{...Eq.(38)}$$

$$J_n = -\frac{qD_n n_{p0}}{W}(\exp(\frac{qV_{ap}}{kT}) - 1) = -\frac{qD_n n_i^2}{Q_A}(\exp(\frac{qV_{ap}}{kT}) - 1) \quad \text{...Eq.(39)}$$

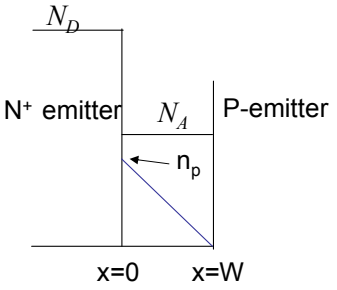
$Q_A = N_A W$: Total Dose $N_A n_{p0} = n_i^2$

---

by setting x=0 in Eq.(38), we have

$$n_p(x=0) = n_{p0}\exp(\frac{qV_{ap}}{kT}) \quad \text{...Eq.(40)}$$

$$J_n \approx -\frac{qD_n n_p(0)}{W} \quad \text{...Eq.(41)}$$



15

Now, we treat the other extreme case of very thin P-emitter.

We can assume that the width, W, is far less than the diffusion length, Ln.

Electron density is now given by Eq.(38) in stead of Eq.(25).

The electron current density is given by Eq.(39).

The current is now in proportion to the inverse of the width of the emitter.

In other words, we can say that the current is inversely proportion to the dose of the emitter.

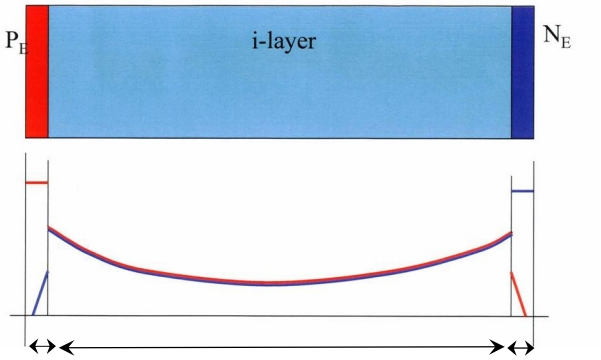
In the thin emitter case, we again derive another current equation.

The electron density at the junction, n_p(x=0) is given by Eq.(40).

The electron current Jn can now be written by Eq.(41), using n_p(x=0).

Here, we use W in place of diffusion length Ln.

Next, we analyze high voltage PIN diodes



$$V_F = V_j(p^+n) + V_i + V_j(nn^+)$$

p^+n Junction Voltage — Voltage drop in i-region — nn^+ Junction Voltage

16

Next, we treat high voltage pin diodes.

Pin diode is simple, but still have many issues to solve.

Basic equation in high injection level

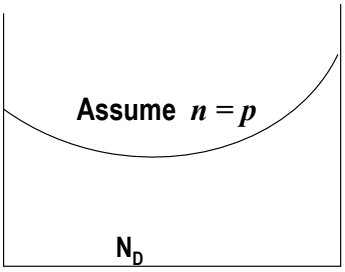
- We derive basic equation, which is valid in high injection level.
- In high injection condition, $n = p$ >> N_D is assumed.

$$J_n = qD_n \frac{\partial n}{\partial x} - q\mu_n n \frac{\partial \psi}{\partial x} \quad \ldots \text{Eq.(42)}$$

$$J_p = -qD_p \frac{\partial p}{\partial x} - q\mu_p p \frac{\partial \psi}{\partial x} \quad \ldots \text{Eq.(43)}$$

$$\frac{\partial J_n}{\partial x} = -\frac{\partial J_p}{\partial x} = qR \cong q\frac{n}{\tau_n + \tau_p} \quad \ldots \text{Eq.(44)}$$

$$D_n = \frac{kT}{q}\mu_n \; ; \; D_p = \frac{kT}{q}\mu_p \quad \ldots \text{Eq.(45)}$$



Delete ψ by executing Eq.(42)*μ_p - Eq.(43)*μ_n.
Then, make derivative with respect to x, and use Eq.(44), we get Eq.(46).

17

We derive basic equation, which is valid in high injection level.

In high injection condition, n = p >> ni can be assumed.

Delete ψ by executing Eq.(42)*μ_p - Eq.(43)*μ_n.

Then, make derivative with respect to x, we get Eq.(46), which is shown in the next slide.

Basic equation in high injection level -2

$$\frac{\partial^2 n}{\partial x^2} = \frac{1}{L_a^{\,2}} n \qquad \text{.....Eq.(46)}$$

$$L_a = \sqrt{D_a \tau_{hi}}, \qquad \text{.....Eq.(47)}$$

$$D_a = \frac{2 D_n D_p}{D_n + D_p}, \qquad \text{.....Eq.(48)}$$

$$\tau_{hi} = \tau_n + \tau_p \qquad \text{.....Eq.(49)}$$

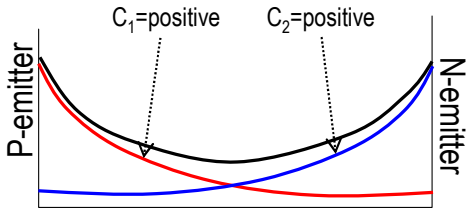
L_a is called "ambipolar diffusion length", τ_{hi} is high injection level lifetime
Eq.(46) is useful equation, governing carrier distribution in high injection condition, where $n=p > N_D$ is valid.

General solution

$$\frac{\partial^2 n}{\partial x^2} = \frac{1}{L_a^{\,2}} n \qquad \text{.....Eq.(46)}$$

General solution:

$$n = C_1 \exp(-\frac{x}{L_a}) + C_2 \exp(\frac{x}{L_a}) \qquad \text{.....Eq.(47)}$$



pin Diode

19

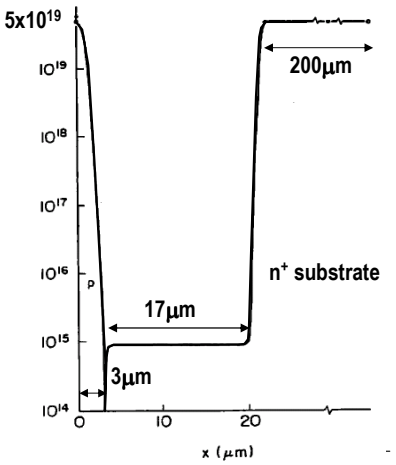
General solution of Eq.(46) is given by Eq.(47).

2. Analytical and TCAD analysis of 200V high speed pin diode

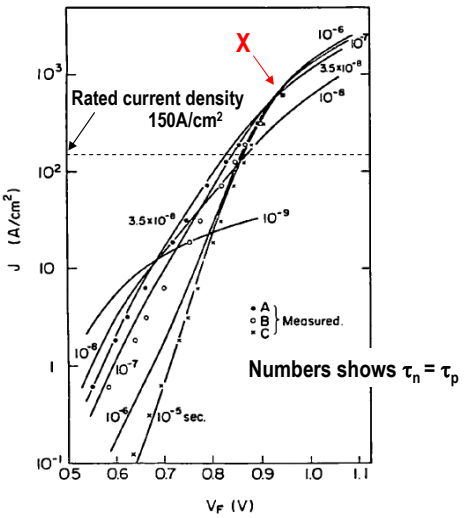
In this section, I present two methods to reduce diode forward voltage drop and reverse recovery time, simultaneously.

200V pin diode profile & J-V Curves

- Left figure shows analyzed 200V pin diode. (20μm epi layer on 200μm n^+ substrate.)
- The right figure shows the calculated and measured J-V relation.
- There is a cross point X, where most of the J-V curves crosses at this point.
- If J is lower than the cross point X, forward voltage decreases as τ decreases.



Calculated 200V pin diode impurity profile



Calculated & Measured J-V

21

The left figure shows the impurity concentration of 200V pin diode.

The diode has 200um thick substrate and 20um epi layer.

3um p-type diffusion layer is formed in the surface.

The resulting i-region is 17um.

The right figure shows the calculated and measured current voltage relation.

A, B, C show the measured results.

The numbers show the carrier lifetime, which is used in the simulator.

Electron and hole lifetimes are assumed to be the same.

The rated current density is 150A/cm^2.

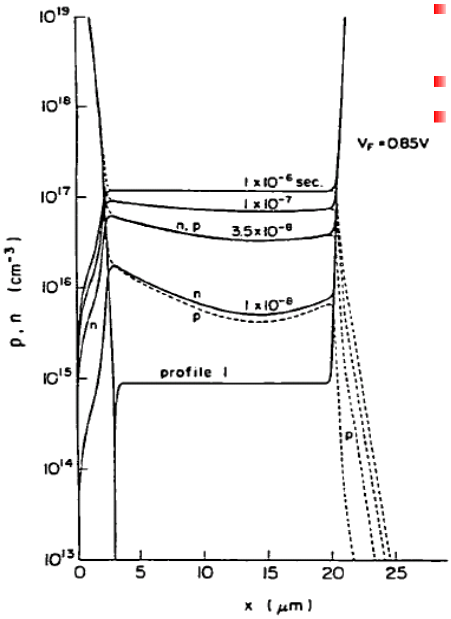
There is a cross point X, where most of the I-V curves crosses at this point.
If the current density is greater than the current density of the cross point X, the forward voltage simply increases as the carrier lifetime decreases.

On the other hand, the forward voltage decreases as the carrier lifetime decreases, if the current density is lower than the current density of the cross point X.

Although the forward voltage initially decreases as the carrier lifetime decreases, it increases again if the carrier lifetime is excessively decreased.

The minimum forward voltage is achieved when the carrier lifetime is approximately 35nsec.

Carrier density distributions



- Carrier density distributions are shown for various τ when V_F=0.85V
- Carrier distribution is almost flat, if τ >35ns
- Carrier density rapidly decreases, if τ becomes lower than 35ns

Electron & hole density distributions

22

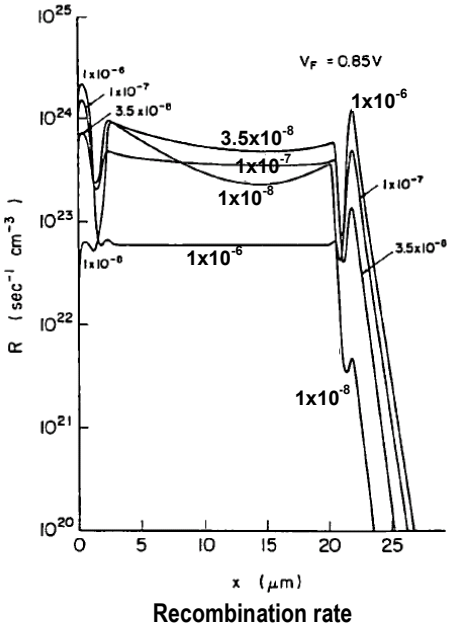
The figure on the left hand side shows carrier density distributions for various lifetimes when the forward voltage, V_F, is 0.85V.

The carrier distribution is almost flat when the carrier lifetime is large.

The carrier profile is still almost flat, when the carrier lifetime is 35ns,

When the carrier lifetime further decreases below 35ns, the carrier density rapidly decreases.

Distribution of Recombination Rate, R



Recombination rate

- Total area of R is the same as the total current

$$\frac{\partial J_n}{\partial x} = qR \quad J_n(x=0)=0,\ J_n(x=W_{ca})=J$$

$$\int_0^{W_{ca}} \frac{\partial J_n}{\partial x}\,dx = J = \int_0^{W_{ca}} qR\,dx$$

- When τ is high,
  R in i-region is small, and
  R in the two emitters is large.
  This means that current flows mostly by diffusion.
- As τ decreases,
  R in i-region becomes large, and
  R in two emitters becomes small.
  Current flows mostly by recombination current.
- R in i-region becomes maximum when τ~35ns

23

The figure on the right hand side, shows the distribution of the recombination rate.

It should be noted that the total area of the recombination rate is the same as the total current.

When the carrier lifetime is high, the recombination in the i-region is low, and the recombination rate in the two emitters is high.

This means that current flows mostly by diffusion.

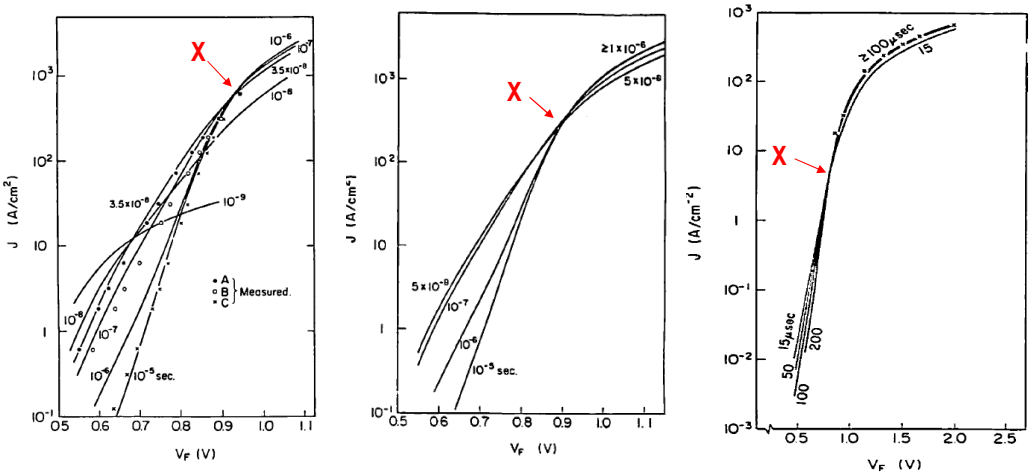
As carrier lifetime decreases, the recombination in the i-region increases, and the diffusion current in two emitters decreases.

The recombination in the i-region becomes maximum when τ=35ns.

The recombination rate decreases in all the regions when the carrier lifetime is further decreased.

J-V relations of pin diodes for i-region width of 17um, 24um and 420um

- The current density of the cross point X decreases as i- region width increases
- Current density of point X for 420um pin diode is only 5A/cm^2
- In high voltage diodes, V_F increases at the rated current density when τ decreases.



W_i=17μm 24μm 420μm

24

These figures compares current-voltage relations of three pin diodes with different i-region width, 17um, 24um and 420um.

The interesting thing is that the current density of the cross point decreases.

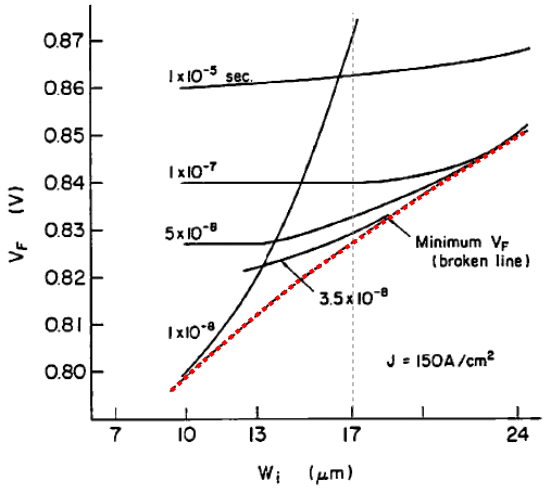
The current density of the cross point for 420um pin diode is only 5A/cm^2.

In high voltage diodes, the forward voltage increases at the operating current density when the carrier lifetime decreases.

There is an optimum τ for minimum V_F

- The figure shows forward voltage vs. i-region width with τ as a parameter.
- Red line shows the minimum V_F as a function of i-region width.
- Minimum V_F is achieved when τ takes an optimum value.

  Example: When W_i = 17um, optimum τ~30ns, minimum V_F = 0.826V.



25

This figure shows the forward voltage vs. i-region width with carrier lifetime as a parameter.

The red dashed line shows the minimum forward voltage as a function of i-region width.

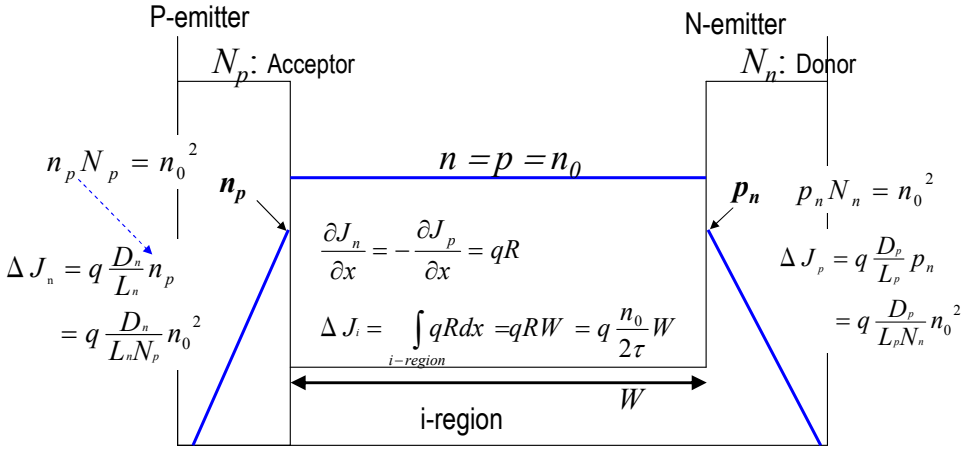
Minimum forward voltage is achieved when τ takes an optimum value.

For example, if i-region width is 17um, the minimum forward voltage is 0.826V, the optimum τ=30ns.

If the carrier lifetime is decreased to 10ns, the forward voltage becomes very high, more than 0.86V.

Analytical analysis to achieve low V_F
by maximizing recombination current

- We assume that (1) carrier density is flat in the i-region and (2) decreases as τ decreases with keeping the flat profile. $n = p = n_0$
- Across the junction, quasi Fermi potentials are kept constant. $n_p N_p = n_0^2$
- Diffusion current ΔJ_n ΔJ_p are given by Eq.(33) of Section 1. $\boldsymbol{qD_n n_p / L_n}$
- In the i-region, recombination current ΔJ_i is given by qRW.



From now on, I analytically analyze the pin diode.

First, I assume that carrier density profile is flat in the i-region.

I assume carrier density decreases with keeping the flat profile as the carrier lifetime decreases.

The electron and the hole density is the same as n_0 throughout the i-region.

2nd, across the junction, quasi-Fermi potentials are kept constant.

This means that pn product is the same across the junction.

At the junction edge of the p-emitter, electron hole product is n_{p0}*N_p.

At the adjacent i-region, the electron hole product is n_0^2.

Then, n_{p0} is given by this equation.

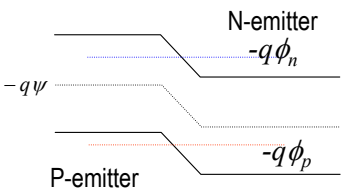
Using eq.(33), diffusion current, ΔJ_n, is given by this equation.

The similar equation is obtained in the N-emitter.

In the i-region, current flows by recombination.

Recombination current, ΔJ_i, is given by qRW, by integrating the current continuity Eq.

Define "Effective Impurity Concentration"



In the transition region,

$$np = n_{i0}^{\ 2} \exp\left\{-\frac{q}{kT}(\phi_n - \phi_p)\right\} = const$$

If bandgap narrowing is considered,
the conserved quantity is $[\varphi_n - \varphi_p]$ *or* $[np/n_i^2]$

$$\frac{np}{n_i^{\ 2}} = \exp\left\{-\frac{q}{kT}(\phi_n - \phi_p)\right\} = const$$

Introduce h & Effective Impurity Concentration

$$h \equiv \frac{n_i}{n_{i0}} \quad N_{eff} \equiv \frac{N}{h^2}$$

$$\frac{np}{n_i^{\ 2}} = \frac{N_D}{n_i^{\ 2}} p_n = \frac{N_D}{n_{i0}^{\ 2} h^2} p_n = \frac{N_{D,eff}}{n_{i0}^{\ 2}} p_n$$

n_{i0} is the intrinsic concentration without bandgap narrowing.

- **If effective Impurity concentration N_{eff} is used,**
  we can treat semiconductor as if there were no bandgap narrowing

27

Conventionally, np product is constant across the junction.

However, if bandgap narrowing is considered, the conserved quantity is Fermi-level difference or np/n_i^2.

We introduce h, which is the ratio ni/ni0.

Then, the conserved quantity is written like this.

We introduce effective impurity doping, defined as this.

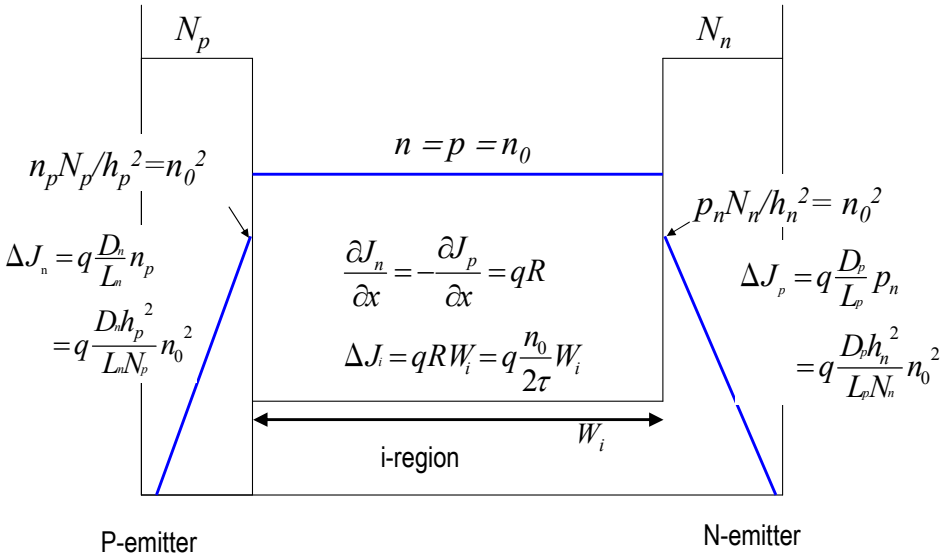
Then, we finally get this conserved quantity.

Apparently, there is no parameters relating to bandgap narrowing.

If we use effective doping, we can treat the system as if were were no bandgap narrowing.

Analytical analysis Including Bangap Narrowing

We introduce effective impurity doping.



28

Now, we use effective impurity doping Np/h^2, Nn/h^2.

Just replace Np by Np/h^2 and Nn by Nn/h^2.

Then, the same equation can be used.

Total current is given by $\Delta J_n + \Delta J_p + \Delta j_i$

$$\Delta J_n = q\frac{D_n h_p^{\,2}}{L_n N_p} n_0^{\,2} \qquad \text{...Eq.(1)}$$

$$\Delta J_p = q\frac{D_p h_n^{\,2}}{L_p N_n} n_0^{\,2} \qquad \text{...Eq.(2)}$$

$$\Delta J_i = q\int_{i-region} R\,dx = q\frac{n_0}{2\tau} W_i \qquad \text{...Eq.(3)}$$

$$R = \frac{pn - n_i^{\,2}}{\tau_p(n + n_i) + \tau_n(p + n_i)} = \frac{n_0^{\,2} - n_i^{\,2}}{\tau(n_0 + n_i) + \tau(n_0 + n_i)} \cong \frac{n_0}{2\tau}$$

Total current J is given by

$$J = q\left(\frac{D_n h_p^{\,2}}{L_n N_p} + \frac{D_p h_n^{\,2}}{L_p N_n}\right) n_0^{\,2} + q\frac{n_0}{2\tau} W_i \qquad \text{...Eq.(4)}$$

29

The total current is given by the summation of the P-emitter diffusion current, the N-emitter diffusion current and i-region recombination current.

These are written by Eq.(1) (2) and (3).

Recombination current is given by integrating generation rate, R, over the i-region.

The generation rate, R, is written in this form.

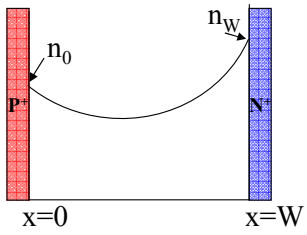
Here, as p=n, and tn=tp, the equation can be simplified as this.

Thus, the total current is expressed by Eq.(4).

Derive Junction Relation

$$V_F - V_i = V_j = \frac{kT}{q}\ln\left(\frac{n_o n_W}{n_i^2}\right)$$

$n = p$ is assumed



$$n = n_i \exp\frac{q}{kT}(\psi - \varphi_n)$$

$$p = n_i \exp\frac{q}{kT}(\varphi_p - \psi)$$

Assume n = p in i – region, the following eq. holds :

$$\psi(0) - \varphi_n(0) = \varphi_p(0) - \psi(0) \quad \text{...Eq.(5)}$$

$$\varphi_p(W) - \psi(W) = \psi(W) - \varphi_n(W) \quad \text{...Eq.(6)}$$

If no voltage drop inside the P^+ ,N^+ layers, then

$$\varphi_p(0) - \varphi_n(W) = V_F \quad \text{...Eq.(7)}$$

Voltage drop in i - region :

$$V_i = \psi(0) - \psi(W) \quad \text{...Eq.(8)}$$

Adding Eq.(5)and (6), and using Eqs.(7)and (8),

$$\psi(0) - \varphi_n(0) + \varphi_p(W) - \psi(W) = V_F - V_i \quad \text{...Eq.(9)}$$

30

In this slide, I'd like to derive basic junction relation.

We assume high injection condition i-region,

and thus, assume that electron and hole densities are equal to each other in i-region.

Eq.(5) and (6) hold from the assumption.

If the voltage drop in P+ and N+ emitter is neglected,

(The hole quasi-Fermi potential in P-emitter) – (the electron quasai-Fermi potential in N-emitter) is equal to the applied voltage, V_F.

Vi is the voltage drop in the i-region.

Vi is given by the static potential difference between the both ends of the i-region.

Adding eq.(5) and (6), and using Eq.(7) and (8),
Eq.(9) is obtained.

Cite Eq.(9) again.

$$\psi(0) - \varphi_n(0) + \varphi_p(W) - \psi(W) = V_F - V_i \quad \text{...Eq.(9)}$$

Multiplying Eq.(9) by $\frac{q}{kT}$ and, then, take exponetial of Eq.(9):

$$n_0 p_W = n_i^2 \exp\frac{q}{kT}(V_F - V_i) \qquad \text{...Eq.(10)}$$

Since $p_W = n_W$,

$$V_F - V_i = \frac{kT}{q}\ln(\frac{n_o n_W}{n_i^2}) \qquad \text{...Eq.(11)}$$

Finished, go back to our analysis!

31

Here, I again show the Eq.(9).

Multiplying Eq.(9) by the factor q/kT, and then take exponential of Eq.(9), we have Eq.(10).

Since $p_W = n_W$, we can replace p_W with n_W.

Then, we get the final equation (11).

$$V_F - V_i = \frac{kT}{q}\ln(\frac{n_o n_W}{n_i^2}) \qquad \text{...Eq.(11)}$$

Since $n_0 = n_W$

$$n_0 = n_i \exp\frac{q}{2kT}(V_F - V_i) \qquad \text{...Eq.(12)}$$

Next, we will obtain expression of V_i.
As all current flows by drift in the i-region, we have

$$J = (q\mu_n n + q\mu_p p)E = 2q\mu_i n_0 \frac{V_i}{W_i}\ , \qquad \text{...Eq.(13)}$$

where $\mu_i = \frac{1}{2}(\mu_n + \mu_p)$ $E = \frac{V_i}{W_i}$ ...Eq.(14)

$$J = q(\frac{D_n h_p^{\ 2}}{L_n N_p} + \frac{D_p h_n^{\ 2}}{L_p N_n})n_0^{\ 2} + q\frac{n_0}{2\tau}W_i \qquad \text{...Eq.(4)}$$

Equating Eq.(13) with Eq.(4), we have

$$V_i = \frac{W_i}{2\mu_i}[(\frac{D_n h_p^{\ 2}}{L_n N_p} + \frac{D_p h_n^{\ 2}}{L_p N_n})n_0 + \frac{W_i}{2\tau}] \qquad \text{...Eq.(15)}$$

32

From Eq.(11), we obtain Eq.(12) because n_0=n_W.

Next, we will obtain Vi.

In the i-region, all current flows by drift.

Thus, Eq.(13) holds.

Here, n=p=n0, and the electric field is given by Vi/W.

We also introduce average mobility, μ_i.

The total current is also given by adding all diffusion current and recombination current as shown in Eq.(4).

Equating Eq.(13) with Eq.(4), we have Eq.(15).

Neglect diffusion current in comparison with recombination current

$$V_i = \frac{W_i}{2\mu_i}\left[\left(\frac{D_n h_p^{\,2}}{L_n N_p} + \frac{D_p h_n^{\,2}}{L_p N_n}\right) n_0 + \frac{W_i}{2\tau}\right] \cong \frac{W_i^2}{4\mu_i \tau} \qquad \text{...Eq.(15)}$$

Substituting Eq.(15) for V_i in Eq.(12), we have

$$n_0 = n_i \exp\frac{q}{2kT}\left(V_F - \frac{W_i^2}{4\mu_i \tau}\right) \qquad \text{....Eq.(16)}$$

Substituting Eq.(16) for n_0 in Eq.(4), we get the final current-voltage relation:

$$J = q\left(\frac{D_n h_p^{\,2}}{L_n N_p} + \frac{D_p h_n^{\,2}}{L_p N_n}\right) n_i^{\,2} \exp\frac{q}{kT}\left(V_F - \frac{W_i^2}{4\mu_i \tau}\right) + q\frac{W_i}{2\tau}\, n_i \exp\frac{q}{2kT}\left(V_F - \frac{W_i^2}{4\mu_i \tau}\right) \qquad \text{....Eq.(17)}$$

I-V relations can be drawn, using the following typical values:

$$\left.\begin{array}{l} D_n = D_p = 1.5 cm^2 \sec^{-1}, \quad W_i = 17\ \mu m \\ \mu_i = 500, \quad h_p = h_n = 6, \quad N_P = N_N = 5 \times 10^{19} cm^{-3} \end{array}\right\}$$

33

In general, diffusion current can be neglected in comparison with the recombination current.

So, Eq.(15) is written in this way.

Substituting Eq.(15) for *Vi* in Eq.(12), we have Eq.(16).

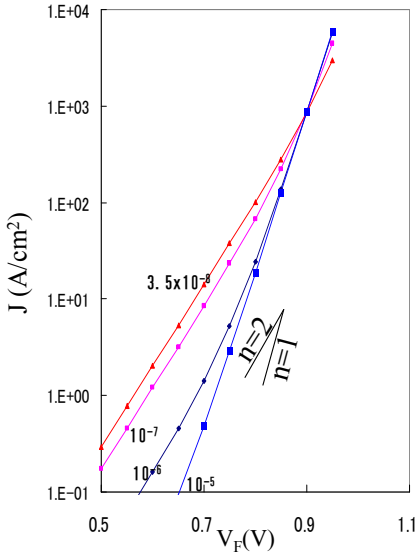
Substituting Eq.(16) for n0 in Eq.(15), we get the current-voltage relation Eq.(17).

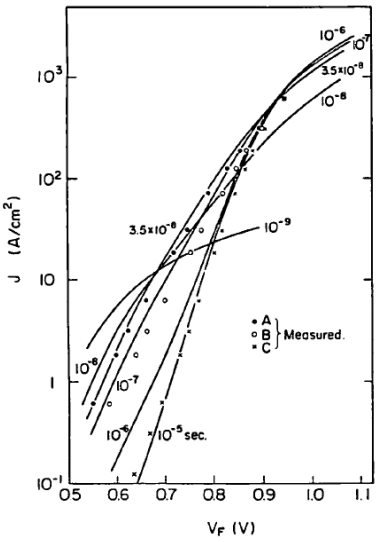
We can draw I-V relations using the following typical values.

Comparison of analytical results & TCAD

- Left figure, analytically calculated results using Eq.(17), agrees well with right figure.
- If n=2, current flows by recombination.
- If n=1, current flows by diffusion.

$$n-value: J \propto \exp(\frac{q}{nkT}V_F)$$





34

The left hand side figure is the analytically calculated results using Eq.(17) & Excell.

The right hand side figure shows the TCAD simulation results and experiment.

The agreement between the two figures is excellent.

In the left figure, I show the gradient of n-value.

n-value is the number, which appear in the denominator in the parenthesis.

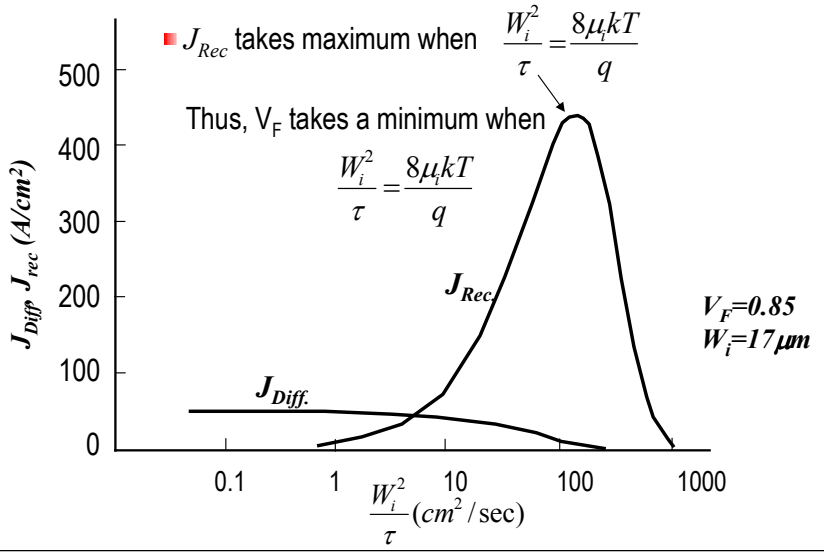
If the n-number is 2, this means current flows by recombination.

If the n-number is 1, the current flows by diffusion.

Total current consists of recombination, J_{rec} and diffusion currents, J_{Diff}

$$J_{Diff} = q(\frac{D_n h_p^{\ 2}}{L_n N_p} + \frac{D_p h_n^{\ 2}}{L_p N_n}) n_i^{\ 2} \exp\frac{q}{kT}(V_F - \frac{W_i^2}{4\mu_i \tau}) \quad J_{Rec} = q\frac{W_i}{2\tau}\, n_i \exp\frac{q}{2kT}(V_F - \frac{W_i^2}{4\mu_i \tau})$$

- J_{rec} and J_{Diff} are illustrated as a function W_i^2/τ



35

The current Eq.(17) can be divided into two components: diffusion current and recombination current.

This figure shows the diffusion current density and recombination current density as a function of W_i^2/τ.

The recombination current takes maximum at $W_i^2/\tau = 8\mu_\iota kT/q$.

Thus, forward voltage takes a minimum when the carrier lifetime $\tau = qW_i^2/8\mu_\iota kT$

Obtain the Current Density of Cross Point X

The point X can be obtained by equating J_{Diff} (high τ) with $J_{Rec}(\tau_{opt})$:

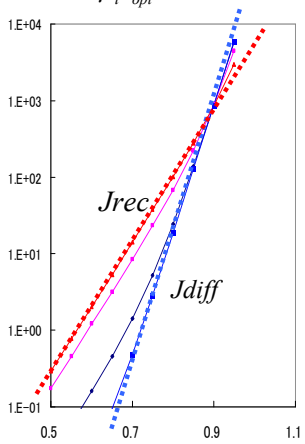
$$J_{Diff}(\tau = \infty) = J_{Rec}(\tau_{opt})$$

$$q\left(\frac{D_n h_p^{\,2}}{L_n N_p} + \frac{D_p h_n^{\,2}}{L_p N_n}\right) n_i^{\,2} \exp\frac{q}{kT} V_F = q\frac{W_i}{2\tau_{opt}}\, n_i \exp\frac{q}{2kT}\left(V_F - \frac{W_i^2}{4\mu_i \tau_{opt}}\right)$$

$$\tau_{opt} = \frac{qW_i^2}{8kT\mu_i} \qquad \text{...Eq.(18)}$$

$$J_{cross} = \left(\frac{kT}{q}\right)^2 \frac{16q\mu_i^{\,2}}{e^2 A W_i^2} \qquad \text{...Eq.(19)}$$

J_{cross} is inversely proportional to W_i^2
J_{cross} is very small for high voltage diode.



36

Now, we will obtain the current density of the cross point X.

This current density can be obtained by equating diffusion current for high lifetime with the recombination current at optimum lifetime.

The solution is given by Eq.(19).

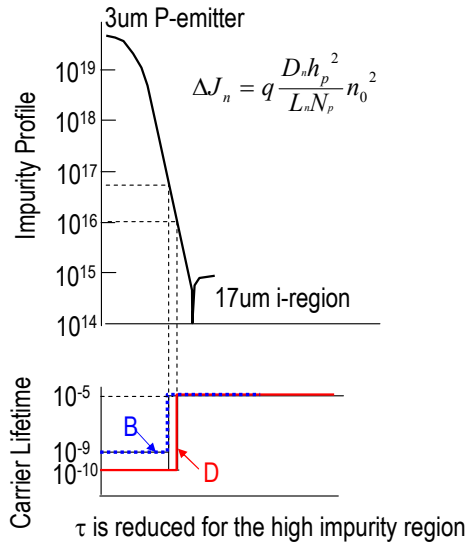
The current density of the cross point is inversely proportion to square of i-region width.

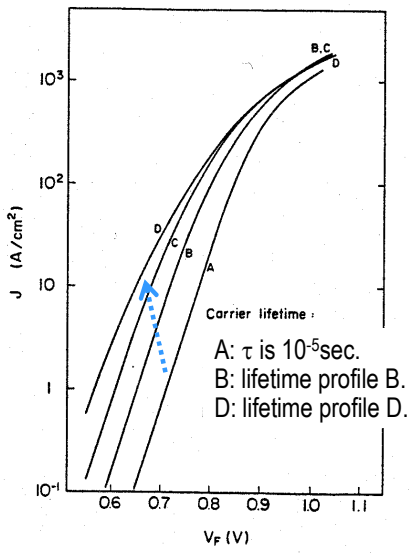
If the i-region width becomes 10 times thicker, the cross point current density becomes 100 times smaller.

Thus, the forward voltage simply increases when carrier lifetime is reduced in high voltage diodes.

Another method to achieve low V_F by maximizing J_{diff}

- Decrease τ for the high impurity region of P-emitter to enhance J_{diff}
  Ex. Profile D means that τ is reduced to 10^{-10}sec for the region where impurity conc.> 10^{16}
- Diffusion current is greatly enhanced for the lifetime profiles B, D as shown in right figure.





37

Next, I talk about the method to decrease the diode forward voltage by increasing the diffusion current.

One method is to decrease the carrier lifetime for high impurity region of P-emitter.

The left hand side figure shows the carrier lifetime profile used for TCAD simulation.

For example, lifetime profile D means that the carrier lifetime is reduced to 10^{-10}sec for the region where impurity concentration is greater than 10^{16}.

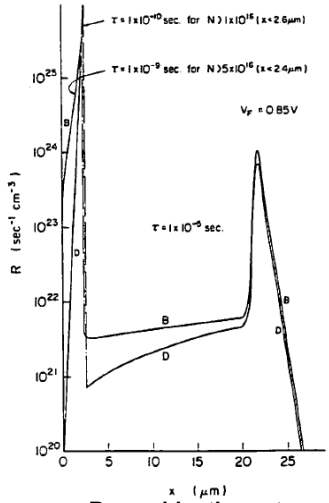
The calculate I-V curves is shown in the figure on the right hand side.

The line A shows the I-V curve when the carrier lifetime is high.

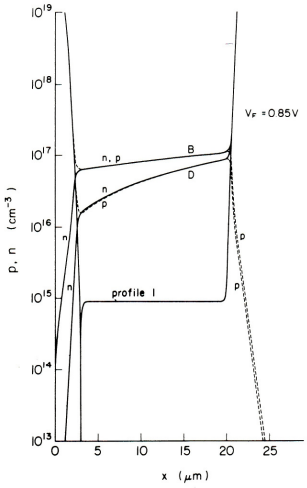
The line D shows the I-V curve when we use the carrier lifetime profile D.

We can successfully reduce the forward voltage of the diode by reducing the carrier lifetime for the high impurity region of the p-emitter

- Left figure shows recombination rate.
- In the P-emitter, recombination rate is markedly increased, creating a large electron J_{diff}
- Electron diffusion current becomes large and occupies a large fraction of the total current
- A large carrier density gradient appears in the i-region (right figure)
  because of large electron diffusion current.



Recombination rate



Carrier Distribution

38

The left hand side figure shows recombination rate.

In the P-emitter, where the lifetime is reduced, the recombination rate is markedly increased.

This directly shows that the diffusion current is enhanced.

Thus, electron diffusion current occupies a large portion of the total current.

As a large portion of the total current is carried by electrons in the i-region, a large gradient in the carrier density distribution appears in the i-region, as shown in the figure on the right hand side.

Analyze the case of enhanced electron diffusion current

- Assume (1) Both emitters have constant impurity Np, Nn.
  (2) P-emitter is quite thin so that electron diffusion current is greatly enhanced. Hole current can be neglected, thus, all the current is carried by electrons.
  (3) Carrier distribution in the i-region is linearly graded. (define n_1 and n_w)

Electron diffusion current is given by Eq.(20) for the thin P-emitter.

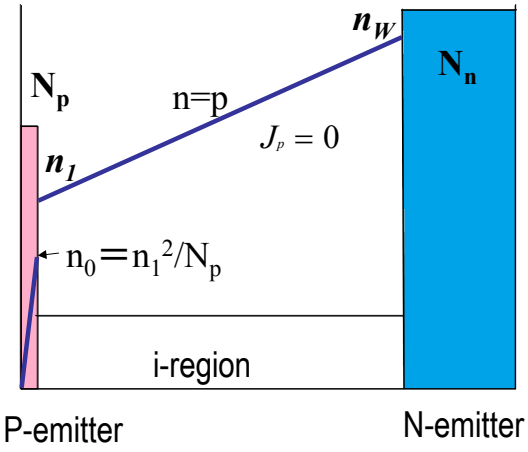
$$J_n = J = \frac{qD_n}{W} n_0$$

$$= \frac{qD_n}{W}\frac{n_1^2}{N_p} = \frac{qD_n}{Q} n_1^2 \quad \text{...Eq.(20)}$$

$$Q = N_p W \quad \text{...Eq.(21)}$$

Solve Eq.(20) for n_1 :

$$n_1 = \sqrt{\frac{QJ}{qD_n}} \quad \text{...Eq.(22)}$$



39

Now, we again analyze the case of the enhanced electron diffusion current.

We assume that the impurity concentration of both emitter is constant, and the thickness of the P-emitter is very thin.

We assume that electron diffusion current is enhanced in the P-emitter and all the current is carried only by electrons.

We also assume that the carrier distribution in the i-region is linearly graded.

The carrier density in the P-emitter side of the i-region is assumed as n_1, and the density in the N-emitter side of the i-region is n_W.

The electron diffusion current is expressed by Eq.(20) for the thin emitter, using the thickness, W.

Eq.(20) is expressed also by using n_1.

The $N_P W$ is the same as the dose of the emitter.

Thus, we use total dose, Q, in place of $N_P W$.

We solve Eq.(1) for n_1. n1 is expressed by Eq.(22).

In the i-region, all the current is carried by electrons, and $J_p=0$

$$J_p = qD_p \frac{\partial p}{\partial x} - q\mu_p pE = \mu_p(-kT\frac{\partial n}{\partial x} + nE) = 0 \quad \text{…Eq.(23)}$$

Solving Eq.(23) for E, *we have*

$$E = \frac{kT}{q}\frac{1}{n}\frac{\partial n}{\partial x} \quad \text{….Eq.(24)}$$

J_n is equal to the total current, J, and is given by

$$J_n = J = \mu_n(kT\frac{\partial n}{\partial x} + nE) = 2 \times qD_n \frac{\partial n}{\partial x} \quad \text{…Eq.(25)}$$

Integrating Eq.(24) over the i-region, we obtain the voltage drop in i-region, V_i,

$$V_i = \int_0^W Edx = \int_{n_1}^{n_W} \frac{kT}{q}\frac{1}{n}dn = \frac{kT}{q}\ln(\frac{n_W}{n_1}) \quad \text{…Eq.(26)}$$

40

Cite Eqs.(11)(22)(25) and (26), again.

$$V_F - V_i = \frac{kT}{q}\ln(\frac{n_1 n_W}{n_i^2}) \text{...Eq.(11)}$$

$$n_1 = \sqrt{\frac{QJ}{qD_n}} \quad \text{...Eq.(22)}$$

$$J_n = J = 2 \times qD_n \frac{\partial n}{\partial x} \text{...Eq.(25)}$$

$$V_i = \frac{kT}{q}\ln(\frac{n_W}{n_1}) \quad \text{...Eq.(26)}$$

Combining Eq.(11) & Eq.(26), we have

$$V_F = \frac{2kT}{q}\ln(\frac{n_W}{n_i}) \quad \text{...Eq.(27)}$$

D_n can be approximately given by:

$$D_n = \frac{a}{n+b} \quad \text{...Eq.(28)}$$

Integrating Eq.(25) over i-region, using Eq.(28)

$$\frac{J}{2q}\int_0^W dx = \int_{n_1}^{n_W} \frac{a}{n+b} dn \text{...Eq.(29)}$$

$$\frac{JW}{2qa} = \ln(\frac{n_W + b}{n_1 + b}) \quad \text{...Eq.(30)}$$

$$n_W = (n_1 + b)\exp(\frac{JW}{2qa}) - b \text{...Eq.(31)}$$

Substituting Eq.(31)& Eq.(22) for n_W, n_1 in Eq.(27), we have:

$$V_F = \frac{2kT}{q}\ln[\frac{1}{n_i}\{(\sqrt{\frac{QJ}{qD_n}} + b)\exp(\frac{JW_n}{2qa}) - b\}]$$

...Eq.(32)

41

We show Eq.(11)(22)(25) and Eq.(26), again.

Combining Eq.(11) & Eq.(26), we have Eq.(27).
D_n can be approximated by Eq.(28).

Integrating Eq.(25) over i-region, using Eq.(28), we have Eq.(29).
Then, we have Eq.(30).
We solve Eq.(30) for n_W.

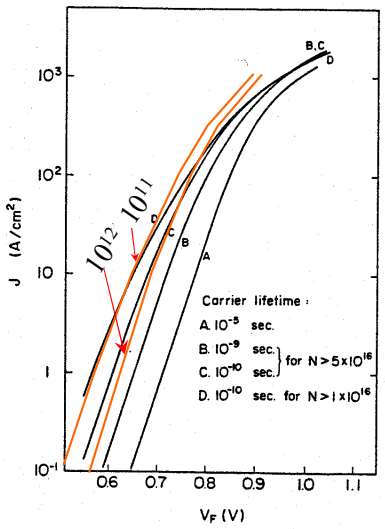
Substituting Eq.(31), Eq.(22) for n_W, n_1 in Eq.(27), we finally get Eq.(32).

Analytical results are compared with TCAD

- Eq.(32) is plotted in the figure (orange lines.)
- Two results are quite similar, although the methods are different.
  This is because the both results enhance the electron diffusion current.

$$V_F = \frac{2kT}{q}\ln[\frac{1}{n_i}\{(\sqrt{\frac{QJ}{qD_n}}+b)\exp(\frac{JW_n}{2qa})-b\}]$$

...Eq.(32)



V-J relation from Eq.(32)

42

Here, I compare analytical results and TCAD results.

Although the assumption is different, two results are quite similar, because the electron diffusion current is enhanced in both calculations.

3. Discussions on injection efficiency of bipolar transistor

It will be understood in this section that injection efficiency depends on Gummel number.

First, derive basic equations for minority carrier current

First, consider electron current in P-region

$$j_n(x) = -q\mu_n n_p \frac{d\varphi_n}{dx} \quad \text{...Eq.(1)}$$

$$Assume \quad \frac{d\varphi_p}{dx} \approx 0 \quad \text{....Eq.(2)}$$ Gradient of quasi-Fermi potential of majority carriers is zero.

$$j_n(x) = q\mu_n n_p \frac{d}{dx}(\varphi_p - \varphi_n) \quad \text{...Eq.(3)}$$

$$\varphi_p - \varphi_n = \frac{kT}{q}\ln(\frac{n_p p_p}{n_i^2}) \quad \text{...Eq.(4)}$$ Substitute Eq.(4) for $\varphi_p - \varphi_n$ in Eq.(3)

$$j_n(x) = q\mu_n n_p \frac{kT}{q}\frac{n_i^2}{n_p p_p}\frac{d}{dx}(\frac{n_p p_p}{n_i^2}) = qD_n \frac{n_i^2}{p_p}\frac{d}{dx}(\frac{n_p p_p}{n_i^2}) \quad \text{...Eq.(5)}$$

Analogously, hole current in N - region is given by:

$$j_p(x) = -qD_p \frac{n_i^2}{n_n}\frac{d}{dx}(\frac{n_n p_n}{n_i^2}) \quad \text{...Eq.(6)}$$

44

In this slide, I derive basic device equations for minority carrier transport.

We consider electron current in P-region.

The hole carrier density is always equal to the impurity dose NA.

The derivative of quasi-Fermi potential of the majority carrier density is considered to be zero.

Otherwise, it creates a large majority carrier current.

Then, equation (3) holds.

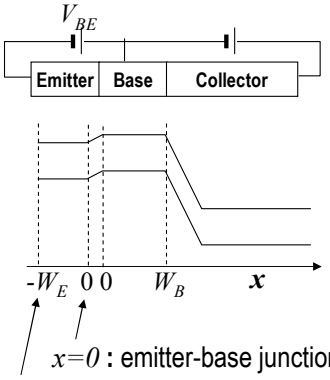
We have the familiar relation equation (4).
Substitute Eq.(4) for φp-φn in Eq.(3).

Now, we have equation (5).

Analogously, we can derive equation (6).

Injection efficiency of N-emitter
Assumption: No recombination in the emitter

Bipolar transistor is analyzed. The results can be applied to pin diodes



Cite Eq.(6) again.

$$j_p(x) = -qD_p \frac{n_i^2}{n_n} \frac{d}{dx}\left(\frac{n_n p_n}{n_i^2}\right) \quad \text{...Eq.(6)}$$

In N-emitter, n_n can be replaced by $n_{n0}(=N_E(x))$ in Eq.(6):

$$j_p(x) = -qD_p \frac{n_i^2}{n_{n0}} \frac{d}{dx}\left(\frac{n_{n0} p_n}{n_i^2}\right) \quad \text{.....Eq.(7)}$$

Boundary condition at $x = -W_E$ is given as:

$$j_p(x = -w_E) = -q(p_n - p_{n0})_{x=-w_E} \cdot S_p \quad \text{.....Eq.(8)}$$

Divide Eq.(7) by $qD_p n_i^2/n_{n0}$, and then integrate Eq.(7).

$$j_p \int_{-W_E}^{0} \frac{n_{n0}}{qD_p n_i^2} dx = -\frac{n_{n0} p_n}{n_i^2}\bigg|_{x=0} + \frac{n_{n0} p_n}{n_i^2}\bigg|_{x=-w_E} \quad \text{.....Eq.(9)}$$

$$p_n(0) = p_{n0}(0)\exp\left(\frac{qV_{BE}}{kT}\right) \quad \text{.....Eq.(10)}$$

45

In this slide, I derive equations for injection efficiency.

I consider bipolar transistor, because this treatment conventionally be used for bipolar transistors.

I assume there is no recombination in the N-emitter.

Namely, τ=infinite.

The results can be used for other devices such as pin diodes, etc.

First, In N-emitter, n_n can be replaced by $n_{n0}(=N_E(x))$ in Eq.(6), we have Eq.(7).

Boundary condition at $x = -xE$ is given by Eq.(8).

Dividing Eq.(7) by $\boldsymbol{qD_p n_{ie}^2/n_{n0}}$, and then integrating Eq.(7) from x=-w_E to 0, we have Eq.(9).

As usual we can derive Eq.(10).

$p_n(0)n_n(0)=n_i^2\exp(qV_{BE}/kT)$

$p_n(0)N_E(0)=p_{no}(0)N_E(0)\exp(qV_{BE}/kT)$

$p_n(0)=p_{no}(0)\exp(qV_{BE}/kT)$

Using Eqs.(8) and (10), Eq.(9) is reduced to:

$$j_p \int_{-w_E}^{0} \frac{n_{n0}}{qD_p n_i^{\,2}} dx = -\exp(\frac{qV_{BE}}{kT}) + 1 - \frac{n_{n0}(-w_E)}{n_i^{\,2}(-w_E)qS_p} j_p$$

Then, neglect 1, we have:

$$\approx -\exp(\frac{qV_{BE}}{kT}) - \frac{n_{n0}(-w_E)}{n_{ie}^{\,2}(-w_E)qS_p} j_p \quad \text{.....Eq.(11)}$$

Solve Eq.(11) for Jp:

$$j_p = \frac{-\mathrm{q}\cdot\exp(\frac{qV_{BE}}{kT})}{\int_{-w_E}^{0} \frac{n_{n0}}{D_p n_i^{\,2}} dx + \frac{n_{n0}(-w_E)}{n_i^{\,2}(-w_E)qS_p}} \quad \text{...Eq.(12)}$$

$$|j_p| = \frac{\mathrm{q}\cdot\exp(\frac{qV_{BE}}{kT})}{\int_{-w_E}^{0} \frac{N_E(x)}{D_p n_i^{\,2}} dx + \frac{N_E(-w_E)}{n_i^{\,2}(-w_E)qS_p}} \quad \text{...Eq.(13)}$$

In the ohmic contact, $S_p = \infty$

$$|j_p| = \frac{\mathrm{q}\cdot\exp(\frac{qV_{BE}}{kT})}{\int_{-w_E}^{0} \frac{N_E(x)}{D_p n_i^{\,2}} dx} \quad \text{...Eq.(14)}$$

Using n_i/n_{i0} instead of n_i, we have Eq.(15).

46

I show equation (11), again.

We neglect 1, then the Eq.(11) is simplified.

Solve Eq.(11) for Jp, then, we have Eq.(12).

In the ohmic contact, the surface recombination is infinite.

Then , we get final equation (14) for hole current.

$$|j_p| = \frac{qn_{i0}^2 \cdot \exp(\frac{qV_{BE}}{kT})}{\int_{-W_E}^{0} \frac{N_E(x)}{D_p(\frac{n_i}{n_{i0}})^2} dx} \quad \text{...Eq.(15)}$$

If we neglect bandgap narrowing and the doping dependence of D_p, we get the simple form.

$$|j_p| = \frac{qn_{i0}^2 D_p \exp(\frac{qV_{BE}}{kT})}{\int_{-W_E}^{0} N_E(x)dx} = \frac{qn_{i0}^2 D_p \exp(\frac{qV_{BE}}{kT})}{G_E} \quad \text{...Eq.(16)}$$

$$G_E \equiv \int_{-W_E}^{0} N_E(x)dx \quad \text{...Eq.(17)} \quad \text{.....Gummel Number}$$

$$Injection\ Efficiency = 1 - \frac{j_p}{j_{total}} = 1 - \frac{qn_{i0}^2 D_p \exp(\frac{qV_{BE}}{kT})}{j_{total}} \frac{1}{G_E} \quad \text{...Eq.(18)}$$

Injection efficiency depends on Gummel Number

47

I show Eq.(15) again. here.

If the denominator $D_p n_i^2$ does not depends on *x,* we get the simple form Eq.(16).

We define Gummel number.

Gummel number is usually defined as the integration of the impurity of emitter or base.

The same as the total impurity dose.

The minority current is proportion to the inverse of the Gummel number.

Injection efficiency of the emitter is calculated as this.

So, if you would like to increase the injection efficiency, increase in the total dose of the emitter is effective.

Please pay attention to the assumption to derive this eauation.

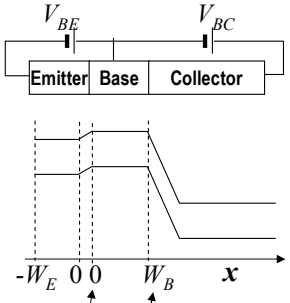
(1)No reconbination in the emitter.

(2)Low injection condition.

Collector Current

- Using almost the same procedure as j_p, we can derive Eq.(19) for j_n in the p-base.

$$j_n = -\frac{qn_{i0}^2 \cdot \exp(\frac{qV_{BE}}{kT})}{\int_0^{w_B} \frac{N_B}{D_n(\frac{n_i}{n_{i0}})^2} dx} \quad \text{...Eq.(19)}$$



$x=w_B$: base collector Junc. depletion layer edge

$x=0$: emitter-base junc.

If we neglect bandgap narrowing

$$|j_n| = \frac{qn_{i0}^2 D_n \exp(\frac{qV_{BE}}{kT})}{G_B} \quad \text{...Eq.(20)}$$

$$G_B \equiv \int_0^{w_B} N_B(x)dx \quad \text{...Eq.(21)} \quad \text{.....Base Gummel Number}$$

The current gain of bipolar transistor is expressed in the form:

$$\text{Current gain} = \frac{j_n}{j_p} = \frac{D_n G_E}{D_p G_B} \quad \text{.....Eq.(21)}$$

48

In the same procedure, we can derive electron current Eq.(19) in the base of the bipolar transistor.

Here, we also introduce base gummel number.

The current gain of the npn bipolar transistor is now expressed in this form.

The current gain is in proportion to the ratio of emitter gummel number over base gummel number.

Appendix

1. Basics of Semiconductor
2. Basic Device Equations
3. Ionization Integral
4. Bandgap Narrowing & Fermi Statistics

1. Basics of Semiconductor

Electron density

$$n = \int_{E_C}^{\infty} \rho_n(E - E_C) P_e(E) dE = \int_{E_C}^{\infty} \frac{\rho_n(E - E_C)}{1 + \exp\frac{E - E_{Fn}}{kT}} dE$$

$$\cong \int_{E_C}^{\infty} \rho_n(E - E_C) \exp\frac{E_{Fn} - E}{kT} dE$$

Boltzmann statistics approximation is used, above.

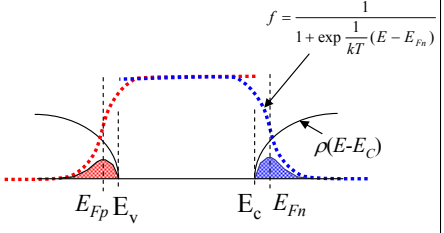
A new variable $E' = E - E_C$ is defined.

$$= \exp\frac{E_{Fn} - E_C}{kT} \int_0^{\infty} \rho_n(E') \exp(-\frac{E'}{kT}) dE'$$

Electron density n is given as:

$$n = N_C \exp(-\frac{E_C - E_{Fn}}{kT})$$

$$N_C = \int_0^{\infty} \rho_n(E) \exp(-\frac{E}{kT}) dE$$



ρ_n(E): Density of states of conduction band
N_C, N_V: effective density of states
n_i: intrinsic carrier concentration
=1.4x10^{10}cm^{-3}
E_{Fn}, E_{Fp}: Quasi-Fermi Energy
E_i: Fermi Energy of intrinsic semiconductor
ψ: Electric potential = $-E_i/q$
φ_n: Electron quasi-Fermi potential = $-E_{Fn}/q$
φ_p: Hole quasi-Fermi potential = $-E_{Fp}/q$

50

Electron density is given by this integral.

ρ is the density of states of conduction band.

The probability P(E) that the energy level ρ(E) is occupied by an electron, is given by Fermi-Dirac distribution function.

We do not discuss the details about this in this lecture.

We will clarify the definitions of basic equations.

Now, we introduce Boltzmann statistics as an approximation.

And, new variable E' is introduced.

This portion can be outside the integral.

Then, electron density is given by this familiar expression.

Here, Nc, called effective density of states, is defined by this equation.

Now, it is clear, we use Boltzmann statistics to derive the equation.

Hole density

$$p = \int_{E_V}^{-\infty} \frac{\rho_p(E_V - E)}{1 + \exp\frac{1}{kT}(E_{Fp} - E)} dE$$

Using the same procedure,
hole density p is given by the following equation:

$$p = N_V \exp(-\frac{E_{Fp} - E_V}{kT})$$

$$N_V = \int_0^{\infty} \rho_p(E') \exp(-\frac{E'}{kT}) dE$$

N_C=2.8x10^{19}, N_V=1.04x10^{19} for Silicon

Fermi energy of intrinsic semiconductor, E_i, is obtained by equating $n=p$:

$$n = p = N_C \exp(-\frac{E_C - E_F}{kT}) = N_V \exp(-\frac{E_F - E_V}{kT})$$

$$E_i = E_F = \frac{E_C + E_V}{2} + \frac{kT}{2}\ln(\frac{N_V}{N_C})$$

Intrinsic carrier concentration:

$np = n_i^2$ (In thermal equilibrium, $E_{Fn}=E_{Fp}$)

$$n_i = \sqrt{N_C N_V} \exp(-\frac{E_G}{2kT}) = N_C \exp(-\frac{E_C - E_i}{kT}) = N_V \exp(-\frac{E_i - E_V}{kT})$$

52

Fermi-energy of intrinsic semiconductor is given by equating n=p.

Fermi energy depends on the ratio of effective density of states.

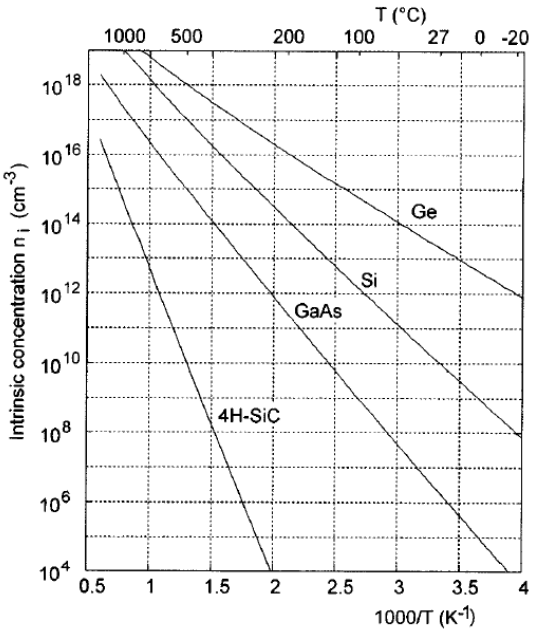
The second term is small, the Fermi energy lies in the middle of the gap.

In equilibrium, the product of electron density and hole density is constant.

n_i is called intrinsic carrier concentration and is given by these equation.

It depends on bandgap and temperature.

Intrinsic Carrier Density



ni=

53

This shows ni for various semiconductor:

***n* and *p* can be expressed using n_i**

$$n = N_C \exp\left(-\frac{E_C - E_{Fn}}{kT}\right) = N_C \exp\left(-\frac{E_C - E_i}{kT}\right)\exp\left(\frac{E_{Fn} - E_i}{kT}\right)$$

$$= n_i \exp\left(\frac{E_{Fn} - E_i}{kT}\right)$$

$$p = N_V \exp\left(-\frac{E_{Fp} - E_V}{kT}\right)$$

$$= n_i \exp\left(\frac{E_i - E_{Fp}}{kT}\right)$$

54

n and p can be expressed in more familiar expression by using n_i and intrinsic Fermi energy E_i.

We introduce quasi-Fermi potentials: φ_n, φ_p and electro-static potential: ψ.

$$define \quad \psi \equiv -E_i / q$$
$$\varphi_n \equiv -E_{Fn} / q$$
$$\varphi_p \equiv -E_{Fp} / q$$

$$n = n_i \exp \frac{q}{kT}(\psi - \varphi_n)$$
$$p = n_i \exp \frac{q}{kT}(\varphi_p - \psi)$$
$$np = n_i^2 \exp \frac{q}{kT}(\varphi_p - \varphi_n)$$

55

Finally, we derive practically useful equations, using three potentials.
We define quasi-Ferimi potentials: φ_n, φ_p and electro-static potential: ψ.

And then, we get the following three equations.

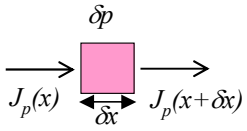
It is useful to know that pn product is the function of the difference of Fermi quasi-potentials.

2. Basic Device Equations

Current continuity equations (conservation of charges)

$$\frac{\partial J_n}{\partial x} = qR + q\frac{\partial n}{\partial t}$$

$$-\frac{\partial J_p}{\partial x} = qR + q\frac{\partial p}{\partial t}$$



$$[-\frac{J_p(x+\delta x)}{q} + \frac{J_p(x)}{q}]\delta t = [\delta p + R\delta t]\delta x$$

Net number of holes flowing into the box
=(increased holes + recombined holes) in the box

Current equation

$$J_n = qD_n\frac{\partial n}{\partial x} - q\mu_n n\frac{\partial \psi}{\partial x}$$

$$J_p = -qD_p\frac{\partial p}{\partial x} - q\mu_p p\frac{\partial \psi}{\partial x}$$

Poisson equation

$$-\varepsilon\frac{\partial^2 \psi}{\partial x^2} = q(N_D - N_A + p - n)$$

R: recombination rate
J_n: electron current density
J_p: hole current density
q: elementally electric charge
ε : electrical permittivity
N_D: ionized donor density
N_A: ionized acceptor density

56

Now, basic equations for device electrical characteristics are given in this slide.

First two equations are current continuity equations.

This is based on conservation of charged particles.

We consider 1-dimensional case for simplicity.

Imagine small box, whose length is dx.

The number of charges that flowing out is *Jp(x+dx).*

The number of charges that flowing in is *Jp(x).*

So, the net number that flowing in this box is given by the left hand side of equation.

This number should be equal to the increased number of holes in the box and the recombined holes in the box.

This means that number of charged particles is conserved.

The next two equations are current equations.

The current is composed by two components, that are diffusion current and drift current.

The final equation is poisson equation, that determines electro-static potential.

Diffusion current

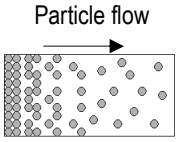
If there is spatially different concentration in mobile charges, they diffuse from a region of high concentration to a region of low concentration.
The particle flow density *F* is given by

$$F = -D\frac{\partial C}{\partial x}$$

,where proportionality factor D is the diffusion constant.
If they have electric charge the particle flow, F, becomes electrical current, J.

$$J_n = qD_n\frac{\partial n}{\partial x}$$

$$J_p = -qD_p\frac{\partial p}{\partial x}$$



Drift current

Electrons and holes in silicon are thermally excited and moves like free particles in random direction, colliding with lattice atoms. The current resulting from a large number of carriers is zero if there is no electric field.
If an electric field applied, charges experience the electric force *qE* and is accelerated between collisions with the lattice atoms. Charges gain drift velocity v_d, and is given by:

$$v_d = \mu E$$

The proportionality factor μ is called mobility.
The drift velocity yields electron and hole current, J_n and J_p,

$$J_p = q\mu_p pE$$
$$J_n = q\mu_n nE$$

where μ_n and μ_p are electron and hole mobilities.

58

Current Equations

Using Einstein relation, current equation can simply be expressed using quasi-Fermi potentials:

$$D_n = kT\mu_n$$
$$D_p = kT\mu_p$$

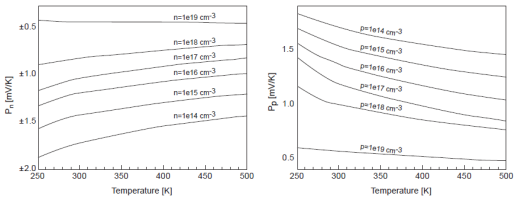
$$J_n = qD_n\frac{\partial n}{\partial x} - q\mu_n n\frac{\partial \psi}{\partial x}$$
$$= -q\mu_n n\frac{\partial \phi_n}{\partial x}$$
$$J_p = -qD_p\frac{\partial p}{\partial x} - q\mu_p p\frac{\partial \psi}{\partial x}$$
$$= -q\mu_p p\frac{\partial \phi_p}{\partial x}$$

Under ***non-isothermal condition*** (Thermodynamic Model)**:**

$$J_n = -q\mu_n n(\frac{\partial \phi_n}{\partial x} + P_n \frac{\partial T}{\partial x})$$

$$J_p = -q\mu_p p(\frac{\partial \phi_p}{\partial x} + P_p \frac{\partial T}{\partial x})$$

Thermoelectric Power: P_n, P_p



TEPs P_n (*left*) and P_p (*right*) as a function of temperature and carrier concentration

T.H. Geballe and G.W. Hull, "Seebeck Effect in Silicon," Physical Review, vol.98,no.4, pp.940–947, 1955.

Lattice Temperature equation:

$$\frac{\partial}{\partial t} c_L T - \frac{\partial}{\partial x}(\kappa \frac{\partial T}{\partial x}) = -\frac{\partial}{\partial x}[(P_n T + \phi_n) J_n + (P_p T + \phi_p) J_p]$$
$$-(E_C + \frac{3}{2}kT)\frac{\partial J_n}{\partial x} - (E_V - \frac{3}{2}kT)\frac{\partial J_p}{\partial x}$$
$$+ qR_{net}(E_C - E_V + 3kT)$$

Mobility

$$v_d = \mu E$$

$\mu_n = 1400 cm^2/Vs$ high resistivity silicon.

$\mu_n = 450 cm^2/Vs$

- Mobility is defined as the proportionality coeff. between drift velocity and electric field.

  The mobility depends on various factors such as the doping level of impurity, magnitude of the electric field as well as roughness of the surface.

- The difference of the electron and hole mobility

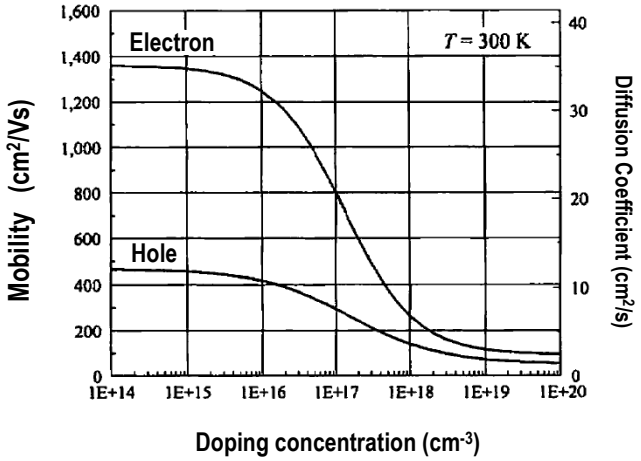
(1) is related to the anti-symmetry in the electrical characteristics between electrons and holes.

(2) is related to the fact why n-channel devices are predominantly used in power MOSFETs and IGBTs.

(3) is also largely related to the safe operating area.

Doping Dependence of Mobility

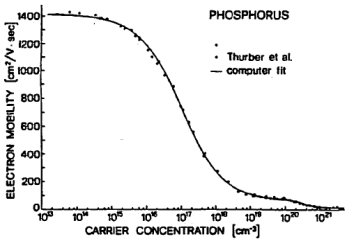
The mobility depends on the doping level of impurity.

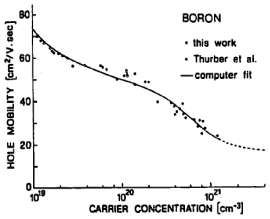


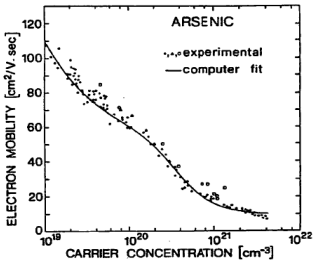
62

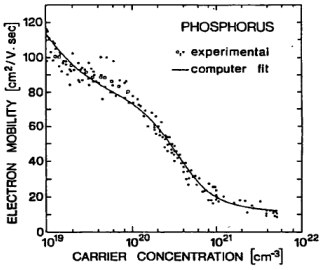
Masseti model for doping dependent mobility, μ_{dop}

The precise empirical expressions are very useful to accurately predict the electrical characteristics of electron devices in TCAD.









63

$$\mu_{dop} = \mu_{min1} \exp\left(-\frac{P_c}{N_{A,0} + N_{D,0}}\right) + \frac{\mu_{const} - \mu_{min2}}{1 + ((N_{A,0} + N_{D,0})/C_r)^{\alpha}} - \frac{\mu_1}{1 + (C_s/(N_{A,0} + N_{D,0}))^{\beta}}$$

$$\mu_{const} = \mu_L \left(\frac{T}{300\text{K}}\right)^{-\zeta}$$

Symbol	Electrons	Holes	Unit
μ_L	1417	470.5	cm^2/Vs
ζ	2.5	2.2	1
μ_{min1}	52.2	44.9	cm^2/Vs
μ_{min2}	52.2	0	cm^2/Vs
μ_1	43.4	29.0	cm^2/Vs
P_c	0	9.23×10^{16}	cm^{-3}
C_r	9.68×10^{16}	2.23×10^{17}	cm^{-3}
C_s	3.43×10^{20}	6.10×10^{20}	cm^{-3}
α	0.680	0.719	1
β	2.0	2.0	1

64

Carrier Carrier Scatering
(Conwell-Weisskopf model)

- When the high level of conductivity modulation occurs,
  the mobility is degraded also by carrier to carrier scattering.
- The carrier to carrier scattering depends on the magnitude of electron and hole product.

$$\frac{1}{\mu} = \frac{1}{\mu_{other}} + \frac{1}{\mu_{eh}}$$ Matthiessen's rule

$$\mu_{eh} = \frac{D(T/300\,\mathrm{K})^{3/2}}{\sqrt{np}}\left[\ln\left(1 + F\left(\frac{T}{300\,\mathrm{K}}\right)^{2}(pn)^{-1/3}\right)\right]^{-1}$$

Symbol	Parameter name	Value	Unit
D	D	1.04×10^{21}	$\mathrm{cm^{-1}V^{-1}s^{-1}}$
F	F	7.452×10^{13}	$\mathrm{cm^{-2}}$

65

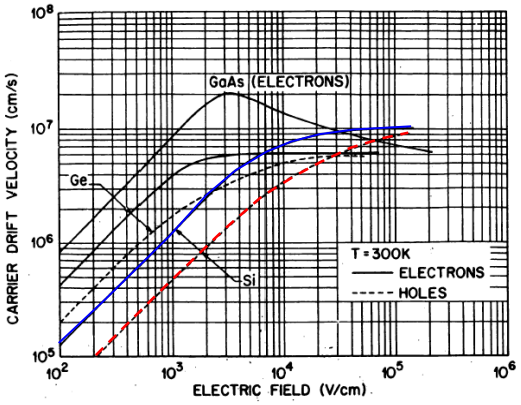
When the high level of conductivity modulation occurs, the mobility is degraded also by carrier to carrier scattering. The carrier carrier scattering depends on the magnitude of the product of electron and hole.

Here, I introduce Conwell and Weisskopf model.

Velocity Saturation

Drift velocity depends on E and saturates when E > 10^4V/cm.

The mobility model should correctly reproduce this phenomena.



66

The drift velocity depends on the magnitude of electric field.

The drift velocity saturates at high electric field more than 10^4V/cm.

The mobility model should correctly reproduce this phenomena.

In TCAD, high field mobility is expressed as a modification of low field mobility, μ_{low}

$$\mu(E) = \frac{\mu_{low}}{\left[1 + \left(\frac{\mu_{low} E}{v_{sat}}\right)^{\beta}\right]^{1/\beta}}$$

$$\beta = \beta_0 \left(\frac{T}{300K}\right)^{\beta_{exp}}$$

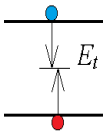
Symbol	Electrons	Holes	Unit
β_0	1.109	1.213	1
β_{exp}	0.66	0.17	1

67

In TCAD, high field mobility is expressed as a modification of low field mobility, μ_{low} as shown in this slide.

SRH recombination

Electrons and holes recombine via a deep level in the bandgap. This process is dominant mechanism in indirect semiconductors such as silicon.
The recombination rate is given by the following form.



$$R = \frac{np - n_i^2}{\tau_p (n + n_i \exp(\frac{E_t - E_i}{kT})) + \tau_n (p + n_i \exp(\frac{E_i - E_t}{kT}))}$$

SRH lifetime has a component, τ_{dop}, which has doping dependence, is expressed in the form:

$$\tau_{dop} = \frac{\tau_0}{(\frac{N}{N_{ref}})^{\gamma}}$$

SRH lifetime τ_n, τ_p is given by the following equation:

$$\frac{1}{\tau_{n,p}} = \frac{1}{\tau_{\max}} + \frac{1}{\tau_{dop}}$$

$\tau_{\max}$ is the lifetime associated with crystal defects or heavy metals

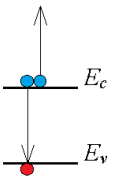
Symbol	Parameter name	Electrons	Holes	Unit
τ_0	tau0	1×10^{-5}	3×10^{-6}	s
N_{ref}	Nref	$5x10^{17}$	$5x10^{17}$	cm^{-3}
γ	gamma	1	1	1

69

Auger recombination

Auger recombination is a process that the energy released by the recombination of electron-hole pair is transferred to a third electron or hole. This process determines carrier lifetime for the region whose doping level is greater than 10^{18} or in the high injection level whose carrier density is greater than 10^{18}.

$$R = (C_n n + C_p p)(np - n_i^2)$$



$$C_n(T) = (A_{A,n} + B_{A,n}(\frac{T}{T_0}) + C_{A,n}(\frac{T}{T_0})^2)[1 + H_n \exp(-\frac{n}{N_{0,n}})]$$

$$C_p(T) = (A_{A,p} + B_{A,p}(\frac{T}{T_0}) + C_{A,n}(\frac{T}{T_0})^2)[1 + H_p \exp(-\frac{p}{N_{0,p}})]$$

Symbol	A_A [$cm^6 s^{-1}$]	B_A [$cm^6 s^{-1}$]	C_A [$cm^6 s^{-1}$]	H [1]	N_0 [cm^{-3}]
Parameter name	A	B	C	H	N0
Electrons	6.7×10^{-32}	2.45×10^{-31}	-2.2×10^{-32}	3.46667	1×10^{18}
Holes	7.2×10^{-32}	4.5×10^{-33}	2.63×10^{-32}	8.25688	1×10^{18}

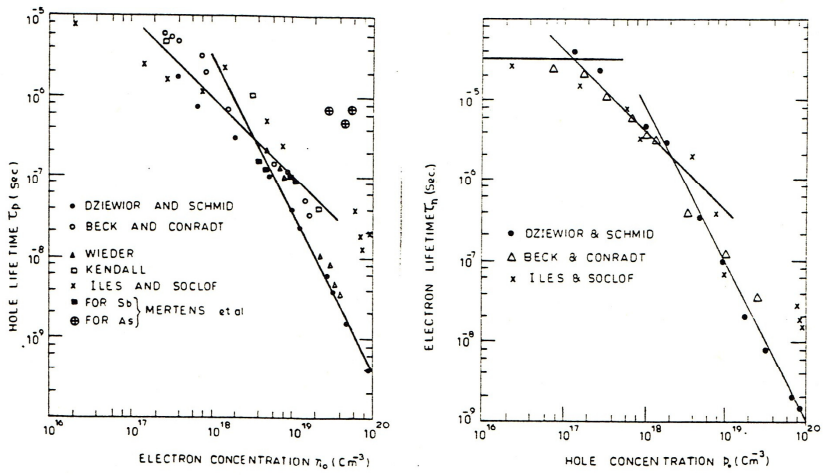


Figure 5.3: Concentration dependence of carrier lifetimes as reported by different authors. After Tyagi *et al.* [223].

Impact Ionization (Avalanche Generation)

- **Impact ionization is important phenomenon that governs device breakdown as well as safe operating area.**
- If the electric field is sufficiently high, the carriers gain enough energy so that they collide with lattice atoms and excite electron hole pairs. This process is called impact ionization.

The electron hole pair generation rate is given by:

$$G = \alpha_n n v_n + \alpha_p p v_p = \frac{1}{q}(\alpha_n |J_n| + \alpha_p |J_p|) \qquad v_n,\ v_p : \text{drift velocity}$$

$$\alpha_n = \alpha_{n0} \exp(-\frac{b_n}{|E|}) \qquad \alpha_{n0} = 7 \times 10^5 / cm,\ b_n = 1.23 \times 10^6 V / cm$$

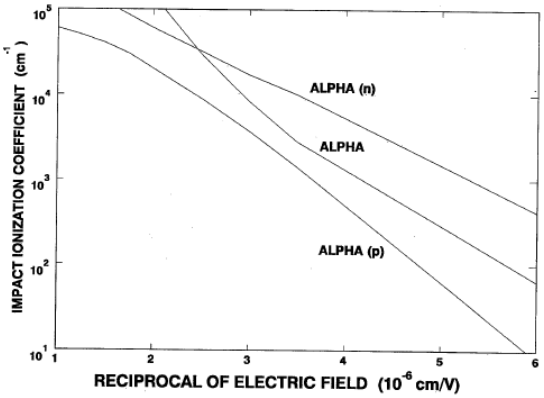
$$\alpha_p = \alpha_{p0} \exp(-\frac{b_p}{|E|}) \qquad \alpha_{p0} = 1.6 \times 10^6 / cm,\ b_p = 2 \times 10^6 V / cm$$

α_n (the electron ionization coefficient) is defined as the number of electron-hole pairs generated by an electron per unit distance traveled.
α_p is analogously defined for holes.

■Approximate expression of ionization coefficient is useful for deriving analytical expression of breakdown voltage,

This is given by:

$$\alpha_{eff} = \alpha_n = \alpha_p = 1.8 \times 10^{-35} E^7$$



73

3. Ionization Integral

Junction breakdown voltage can be evaluated using the ionization integral.

- Assume a depletion layer of width W [x=0,W], where carriers are generated by Impact Ioniz.
  Hole current I_{p0} is incident at the left hand side (x=0) of the depletion region.
  Hole current I_p will increase within the depletion layer and reaches a value $M_p I_{p0}$ at x=W.

Total current I:

$I=I_p+I_n$ ..... Eq.(1)

Suppose small region [x, x+dx]:

Number of holes entering this region is $I_p(x)/q$,

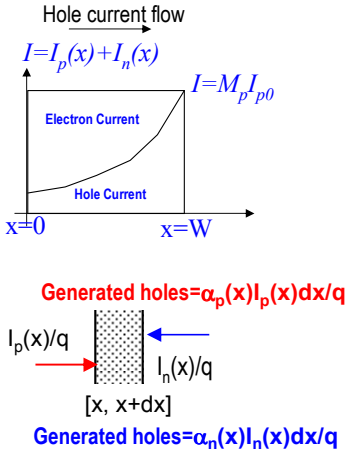
generate electron hole pairs of $\alpha_p(x)I_p(x)dx/q$,

Number of electrons entering the region is $I_n(x)/q$,

generate electron hole pairs of $\alpha_n(x)I_n(x)dx/q$.

The hole current increment in the region is:

$dI_p=\alpha_p I_p dx+\alpha_n I_n dx$ ..... Eq.(2)



74

Using $I=I_p+I_n$ (Eq.(1))

$$\frac{dI_p}{dx} = (\alpha_p - \alpha_n)I_p + \alpha_n I \text{ ..... Eq.(3)}$$

The solution of Eq.(3) is given by Eq.(4) with boundary condition $I_p(0)=I/M_p$, using Appendix I.

$$I_p(x) = I\left\{\frac{1}{M_p} + \int_0^x \alpha_n \cdot \exp\left(\left(-\int_0^{x'}(\alpha_p - \alpha_n)dx''\right)\right)dx'\right\}\exp\left(\int_0^x (\alpha_p - \alpha_n)dx'\right) \quad \text{....Eq.(4)}$$

As no electron enters from x=W, Eq.(5) holds.

$$I_p(W) = I, \quad \text{..... Eq.(5)}$$

Set $x = W$ in Eq.(4) and divide Eq.(4) by $\exp\left(\int_0^W (\alpha_p - \alpha_n)dx'\right)$, we have

$$I\exp\left(-\int_0^W (\alpha_p - \alpha_n)dx'\right) = I\left\{\frac{1}{M_p} + \int_0^W \alpha_n \cdot \exp\left(\left(-\int_0^{x'}(\alpha_p - \alpha_n)dx''\right)\right)dx'\right\} \quad \text{....Eq.(4a)}$$

Thus, the following equation holds.

$$\frac{1}{M_p} = \exp\left(-\int_0^W (\alpha_p - \alpha_n)dx'\right) - \int_0^W \alpha_n \cdot \exp\left(\left(-\int_0^x (\alpha_p - \alpha_n)dx'\right)\right)dx \quad \text{.....Eq.(6)}$$

75

Eq.(7) can be obtained, using Appendix II.

$$1-\frac{1}{M_p}=\int_0^W \alpha_p \cdot \exp\left(-\int_0^x (\alpha_p-\alpha_n)dx'\right)dx \quad \text{.... Eq.(7)}$$

The avalanche breakdown is defined as the voltage where M_p becomes infinity.
The breakdown voltage is given by the ionization integral:

$$\int_0^W \alpha_p \cdot \exp\left(-\int_0^x (\alpha_p-\alpha_n)dx'\right)dx=1 \qquad \text{.... Eq.(8)}$$

Similarly, the following equation holds for electron initiated avalanche process:

$$\frac{1}{M_n} = \exp(-\int_0^W (\alpha_n - \alpha_p)dx') - \int_0^W \alpha_p \cdot \exp\left((-\int_x^W (\alpha_n - \alpha_p)dx') \right)dx \qquad \text{.....Eq.(9)}$$

$$1 - \frac{1}{M_n} = \int_0^W \alpha_n \cdot \exp\left((-\int_x^W (\alpha_n - \alpha_p)dx') \right)dx \quad \text{.... Eq.(10)}$$

$$\int_0^W \alpha_n \cdot \exp\left((-\int_x^W (\alpha_n - \alpha_p)dx') \right)dx = 1 \qquad \text{.... Eq.(11)}$$

If we define effective ionization coefficient α_{eff}, assuming $\alpha_{eff}=\alpha_n=\alpha_p$
The ionization integral reduces to the simple expression:

$$\int_0^W \alpha_{eff} dx = 1$$

Appendix I (From Sze p.99)

$$\frac{dy}{dx} + Py = Q$$

The standard solution of the above equation is given by:

$$y = \left[\int_0^x Q \cdot \exp\left(\int_0^x P dx' \right) dx + C \right] / \int_0^x P dx'$$

Appendix II

The following relations generally holds.
Equation (8) can be obtained from Eq.(6) using Eq.(A1).

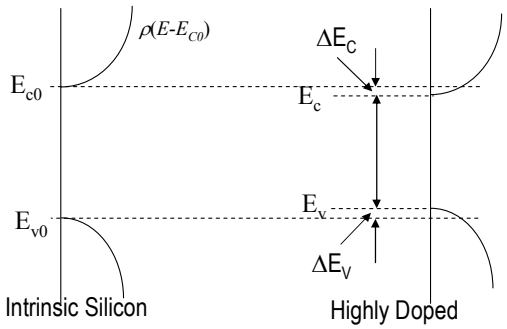
$$\int_0^W f(x) \cdot \exp\left(\left(-\int_0^x f(x') dx' \right) \right) dx = 1 - \exp\left(-\int_0^W f(x) dx \right) \quad \text{...Eq.(A1)}$$

$$\int_0^W f(x) \cdot \exp\left(\left(-\int_x^W f(x') dx' \right) \right) dx = 1 - \exp\left(-\int_0^W f(x) dx \right) \quad \text{...Eq.(A2)}$$

4. Bandgap Narrowing & Fermi Statistics

If the semiconductor is highly doped, the bandgap is reduced.
This will affect the electrical characteristics of various devices.

Injection efficiency of highly doped emitter layers becomes lower, if the bandgap narrowing is included.



We derive the basic current transport equations, in order to estimate the effect.

79

If the semiconductor is highly doped, the bandgap is reduced.

This will affect the electrical characteristics of various devices.

For example, the emitter layers are usually highly doped, and emitter injection efficiency of bipolar transistors or diodes are lower than the values, which are estimated without the effect of the bandgap narrowing.

In order to estimate the effect, we derive the basic current transport equations.

We introduce a parameter, w_n for electrons, which is defined in Eq.(1)

$$n = \int_{E_C}^{\infty} \frac{\rho(E - E_C)}{1 + \exp\frac{1}{kT}(E - F_n)} dE \equiv N_C \exp\frac{1}{kT}(F_n - E_C + qw_n) \quad \text{----- Eq.(1)}$$

$$E_C = E_{C0} - \Delta E_C \quad \text{----- Eq.(2)}$$

E_C: Conduction band edge energy

ρ(E-E$_C$): Density of states of conduction band

ΔE_C: Reduction of conduction band-edge energy from intrinsic semiconductor

w_n denotes the effect of Fermi statistics and ΔE_C shows the effect of bandgap narrowing.

80

First, we define the conduction band-edge, E_C, and introduce a parameter, w_n for electrons. The w_n value are defined in Eq.(1)

ρ(E-EC) means density of states of the conduction band, and ΔE_C shows the energy reduction of the conduction band-edge energy from the band edge energy E_{C0} of the intrinsic semiconductor.

Similarly, w_p is defined for holes:

$$p = \int_{E_V}^{-\infty} \frac{\rho_p(E_V - E)}{1 + \exp\frac{1}{kT}(F_p - E)} dE \equiv N_V \exp\frac{1}{kT}(E_V - F_p + qw_p) \quad \text{----- Eq.(3)}$$

$$E_V = E_{V0} + \Delta E_V \quad \text{----- Eq.(4)}$$

E_V: valence band edge energy
ΔE_V: Valence band edge shift,

Next, Consider thermal equilibrium case: $F_n = F_p$.

$$pn = n_i^2 = N_C N_V \exp\frac{1}{kT}(E_V - E_C + qw_n + qw_p)$$

$$= N_C N_V \exp\frac{1}{kT}(E_{V0} + \Delta E_V - (E_{C0} - \Delta E_C) + qw_n + qw_p)$$

$$n_i = n_{i0} \exp\frac{1}{2kT}(qw_n + qw_p + \Delta E_C + \Delta E_V) \quad \text{----- Eq.(5)}$$

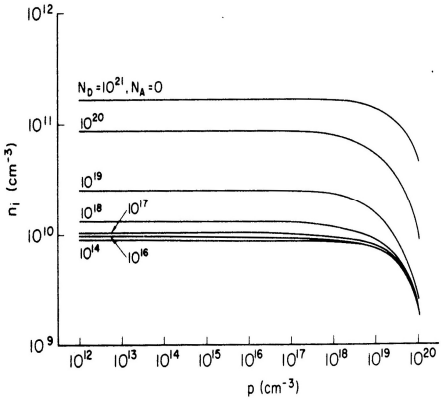
n_i indicates effective intrinsic carrier concentration.

$n_{i0} = \sqrt{N_C N_V} \exp -\frac{1}{2kT}(E_{C0} - E_{V0})$: intrinsic carrier concentration without bandgap narrowing

81

n_i dependence on carrier injection level

The figure shows the dependence of n_i on injected minority hole density, p.
[N-type silicon with N_A=0, $n=N_D+p$ (charge neutrality) is assumed]



n_i vs. hole density in N-type Silicon when N_A=0, $n=N_D+p$

As excess hole density, p, increases over 10^{19}, n_i rapidly decreases because of the effect of Fermi statistics.

The effect of Fermi statistics reduces n_i value as if the bandgap were widened.

Experimental measurement of the bandgap narrowing

has been done by many authors. Most of them measured the value of $n_{i,equi}$ in thermal equilibrium, using current transport equations based on Boltzmann statistics.
(One of typical measurements was done by J.W. Slotboom and H.C. de Graaff[Slot1])

It should be noted that the measured values, $n_{i,equi}$, includes both the effect of bandgap narrowing and the effect of Fermi statistics.

$n_{i,equi}$ can be used if injected carrier density is below 10^{19}.
This is satisfied in almost all cases.

[Slot1] J.W. Slotboom and H.C. de Graaff, "Measurements of Bandgap Narrowing in Si Bipolar Transistors," Solid-State Electronics, vol.19, no. 10, pp.857–862, 1976.

83

Experimental measurement of the bandgap narrowing has been done by many papers. Most of them measured the value of ni,equi using current transport equations based on Boltzmann statistics.

(One of typical measurements was done by J.W. Slotboom and H.C. de Graaff[Slot1]) and has extensively been used in TCAD. The measured values correspond to ni,equi in Eq.(5), thus, includes the effect of bandgap narrowing and, simultaneously, the effect of Fermi statistics. If one use the measured values, ni,equi, one should keep in mind that ni,equi includes the effect of Fermi statistics. ni,equi can be used from previous figure if injected carrier density is below 1019, which is satisfied in almost all cases.

Current equations including bandgap narrowing & Fermi Statistics

$$J_n = -q\mu_n n \frac{\partial \varphi_n}{\partial x} \quad \text{----- Eq.(6)}$$

$$J_p = -q\mu_p p \frac{\partial \varphi_p}{\partial x} \quad \text{----- Eq.(7)}$$

$$F_n = -q\phi_n \quad \text{----- Eq.(8)}$$

$$F_p = -q\phi_p \quad \text{----- Eq.(9)}$$

Taking derivative of Eq.(1), and using Eq.(6), the following electron current flow equation can be derived. (see next slide)

$$J_n = -\mu_n kT \frac{\partial n}{\partial x} - q\mu_n n \frac{\partial}{\partial x}(\psi + \frac{\Delta E_C}{q} + w_n)$$

$$J_p = -\mu_p kT \frac{\partial p}{\partial x} - q\mu_p p \frac{\partial}{\partial x}(\psi - \frac{\Delta E_V}{q} - w_p)$$

Now, we define ΔE_G as follows:

$$\Delta E_G \equiv \Delta E_C + \Delta E_V + (w_n + w_p)/q = kT \ln(\frac{n_i}{n_{i0}})$$

$$n = \int_{E_C}^{\infty} \frac{\rho(E - E_C)}{1 + \exp\frac{1}{kT}(E - F_n)} dE \equiv N_C \exp\frac{1}{kT}(F_n - E_C + qw_n)$$

$$= N_C \exp(-\frac{E_{C0} - E_i}{\mathrm{kT}}) \exp(\frac{F_n - E_i - E_C + E_{C0} + qw_n}{kT})$$

$$= N_C \exp(-\frac{E_{C0} - E_i}{\mathrm{kT}}) \exp(\frac{F_n - E_i + \Delta E_C + qw_n}{kT})$$

$$= n_{i0} \exp(\frac{F_n - E_i + \Delta E_C + qw_n}{kT})$$

$$= n_{i0} \exp(\frac{-q\varphi_n + q\psi + \Delta E_C + qw_n}{kT})$$

$$= n_{i0} \exp\frac{q}{kT}(-\varphi_n + \psi + \frac{\Delta E_C}{q} + w_n)$$

E_{C0}: band edge energy of intrinsic silicon

Replace ψ simply by $\psi + \Delta E_C/q + W_n$, then you'll get the electron current equation including bandgap narrowing.

Practical method including Bandgap Narrowing & Fermi Statistics

We can determine the value of n_i usually only by experiment.
Thus, it is difficult to determine each values of ΔE_C, ΔE_V, w_n and w_p.
An good approximation is:

$$\Delta E_C + qw_n = \Delta E_V + qw_p = \Delta E_G/2 = \frac{kT}{2}\ln(\frac{n_i}{n_{i0}})$$

Thus, the final current transport equations are:

$$J_n = -\mu_n kT\frac{\partial n}{\partial x} - q\mu_n n\frac{\partial}{\partial x}(\psi + \frac{\Delta E_G}{2q})$$

$$J_p = -\mu_p kT\frac{\partial p}{\partial x} - q\mu_p p\frac{\partial}{\partial x}(\psi - \frac{\Delta E_G}{2q})$$

The effect of Fermi statistics is included in the measured parameters of Bandgap Narrowing.

FIN