

Process Technology for Silicon Carbide Devices

Docent seminar by Carl-Mikael Zetterling

March 21st, 2000

Welcome to this Docent seminar on Process Technology for Silicon Carbide Devices

Actually an alternative title might have been Process Integration ..., since the focus of this talk is on putting all the process steps together, and on the devices.

Since this is a docent seminar, the intended audience is not necessarily experts in the SiC field. I have aimed at an audience familiar with semiconductor devices in for instance silicon.



First I think it is important to address the question of WHY we should be interested in SiC.

Then we can have a look at HOW we make devices.

Last, we will see WHICH devices have been made, and if they are any good.



I will start assuming you know practically nothing abot SiC, and tell you what it is and some of its properties.

Then we will look at each property and see the advantage for devices.



Silicon carbide is made up of equal parts silicon and carbon.

Both are period IV elements, so they will prefer a covalent bonding such as in the left figure. Also, each carbon atom is surrounded by four silicon atoms, and vice versa.

This will lead to a highly ordered configuration, a single crystal, such as in the right figure.

(The crystal is polarized, meaning that we can identify a silicon face and a carbon face, each having atoms with one free bond.)

However, whereas silicon, or GaAs has only one crystal structure, SiC has several.



Since there is exactly one silicon atom for each carbon atom, lets think of this as a unit, symbolized by this ball.

The most efficient way to pack balls is in this hexagonal fashion, in crystallographical terms called Hexagonal Close Packing.

This layer of balls labeled A represent a double layer of silicon and carbon atoms.

The next layer has to be displaced according to the covalent bonding scheme, but we find there are two choices, indicated by B and C respectively.

The rule is A can be followed by B or C, but not A. The same goes for B and C.



It turns out that there exists several stable stacking orders with a long term order throughout a large sample.

This figure shows four of them:

3C, 2H, 4H and 6H. The number corresponds to the number of double layers of Si and C before the pattern is repeated. For instance, 4H repeats ABAC ABAC etc.

Of these, it is 4H and 6H which are of interest technologically since large wafers can be made in this material, and hence used for device production.

We will look at manufacture later, now lets look at some properties of this semiconductor material.

ERRATA: 6H is stacked ABCACB, incorrect fig.

VETENSKA OCH KONST HÖGSKOLAN	Kungl TEKNISKA HÖGSKOLAN KUNGL					
at 300 K	Si	GaAs	4H/6H-SiC	GaN		
Eg (eV)	1.12	1.4	3.0 <mark>-3.2</mark>	3.4		
Ec (MV/cm)	0.25	0.3	<mark>2.2-</mark> 2.5	3		
μn (cm²/Vs)	1350	8500	100 -1000	1000		
Er	11.9	13	10	9.5		
Vsat (cm/s)	1x10 ⁷	1x10 ⁷	2x10 ⁷	3x10 ⁷		
λ (W/cmK)	1.5	0.5	3 - 5	1.3		
© 2000 Carl-Mikael Zetterling 7						

This table compares four semiconductors: silicon, gallium arsenide, silicon carbide and gallium nitride. The first two you probably know already.

I include gallium nitride here since in some respects it is perhaps a better material than SiC. It is also of interest to combine GaN with SiC as we will se in a later slide.

The big difference is the energy bandgap. Our standard semiconductors have almost three times smaller bandgaps than the wide bandgap materials SiC and GaN. However, it is probably the ten times larger critical field for breakdown which makes the biggest difference.

There are no large differences in the other parameters, except the high mobility of GaAs.



Lets investigate how the critical field influences device performance. First some device basics:

The idea with a semiconductor device is that they can either block a voltage, or conduct a current with low power loss, ie an ideal switch.

To block a voltage, there has to be a depletion region inside the device. It is perhaps easiest to see this in a pn junction like here. The figure shows the electric field as a function of distance. Most of the depletion region spreads into the lowest doped part of the junction. We can calculate the voltage supported by this depletion region from this formula, which is the area under the graph.

So ten times Ec means ten times shorter W, which should translate to lower on-resistance.



When the device is not blocking, it should conduct large currents with small voltage drop (power loss).

To calculate the on-resistance we need to know the doping in the low doped material where the depletion region extends. As we calculate this, we find that ten times higher Ec leads to one hundred times higher doping. This is good news since higher doping means lower resistance.

The on-resistance will depend on the critical field cubed, and two material parameters: permittivity and mobility. But Ec is more important.

Now lets look at a plot of the on-resistance versus the designed breakdown voltage.



This figure shows Si, and 4H and 6H SiC. GaAs is a factor 12 better than Si GaN is a factor 2 better than SiC

For most power devices the current will be conducted through the substrate. This adds some resistance since the mobility and the amount of doping is limited.

Normally calculated per unit area, correct name is specific on-resistance. This is because normally we are interested in designing for a certain current density, 100 to 1000 A/cm2 for power devices. The voltage drop is then easily calculated, and the limit is say 10 volts, corresponding to 10 mohmcm2.

Seems low, explanation is conductivity modulation. The figure is valid for majority carrier devices only.



At high current densities so many minority carriers are injected, that the conductivity is no longer determined by the doping but by the minority charge. Depending on the minority carrier lifetime, a smaller or larger part of the drift region will be conductivity modulated. In this case, showing a SiC simulation, 0.1 us cuts the on-resistance to one half, whereas 1 us reduces it even more. This means the on-state losses are very small.

However, this minority carrier charge has to be removed before the device can block a voltage, and this removal results in a switching loss. The total loss is on-state loss plus switching frequency times switching loss. Tradeoff! High lifetime for low frequency and low lifetime for high frequency.

KUNGL TEKNISKA HÖGSKOLAN HIgh frequency devices								
	$\tau = \frac{W}{2v_{sat}} \qquad \qquad C \propto \varepsilon_r$							
	at 300 K	Si	GaAs	4H/6H-SiC	GaN			
	Ec (MV/cm)	0.25	0.3	<mark>2.2-</mark> 2.5	3			
	Er	11.9	13	10	9.5			
	Vsat (cm/s)	1x10 ⁷	1x10 ⁷	2x10 ⁷	3x10 ⁷			
© 2000 Carl-Mikael Zetterling 12								

The delay time in a high speed device is limited by transport through the depleted region, see equation.

From previous slide we know W is 10 x shorter. Table also shows vsat is 2-3 times higher. Either we can operate at higher frequency, or for same frequency we get 10x voltage.

10x voltage means higher output impedance, which is easier to match in the amplifier.

Permittivity 25 % lower, means lower capacitance, lowers RC time constants 25 %

Semi-insulating SiC allows even lower loss devices.



The first advantage of wide bandgap materials was actually considered the possibility of high temperature electronics. If we calculate the intrinsic concentration as function of temperature, we find it is exponentially dependent on the energy bandgap.

This figure shows why germanium (0.67 eV) is no longer used. When the intrinsic concentration ni is larger than the doping level, temperature effects can be seen in the device.

SiC is good for high temperature circuits (amplifiers) or sensors which need to be in hot places, eg combustion engine monitoring. Main obstacle is not only the semiconductor material but also all contacts etc have to withstand high temp for long time.



The temperature rise in a device is proportional to power and thermal resistance. The highest thermal resistance will limit the amount of power for a certain temperature difference, since all thermal resistances are in series.

Si and GaAs: Rth of solder is limiting. 3x thermal conductivity means more heat spreading, hence larger area and lower effective thermal resistance.

For the same current density, SiC device will be 10 times smaller volume because Ec is 10 x and W is 10x smaller. This is especially true in lateral high frequency devices. This will result in higher device temperatures, but it means packaging is a key issue.

3D thermal simulations are necessary to ensure proper heat conduction in packages.



It is clear that silicon will always be used for most integrated circuits. It is questionable if there will be any highly integrated circuits at all in SiC. A high temperature amplifier needs less than 100 devices.



The wide bandgap around 3 eVallows photodetectors in the UV range, and possibly solar blind (not sensitive to sun-light) which makes missile detection possible in daylight

Blue LEDs was the first SiC product, even though it has an indirect bandgap. Now GaN or InGaN alloys are used for blue and green with very high efficiency. White LEDs (phosphor in lens) is also a major product for GaN.



The number one reason is the high critical field for breakdown, which is good for all devices!



Now on to the second section.

HOW do we make SiC devices?

The order is roughly the same as the manufacturing order.

First we need a wafer, with epitaxial layers.

We need selective doping, we might need to etch mesas, we need to isolate devices from each other, and we need metal contacts to package the device.

This is mostly a overview, since each of these are a subject for a lecture or more each. What I want to show is which constraints are put upon device manufacture by the process steps when adapted to SiC.



First of all we need a SiC wafer. The controlled growth of SiC was not achieved until 10 - 15 years ago. When we look at the process we can understand why:

SiC powder is heated by a RF field coupled to this graphite crucible to 2300 C.The powder sublimes (Si2C, SiC2 etc). The seed crystal has a lower temperature and the SiC vapor will condense to form a boule. Growth rate 1 mm/h

Temperature uniformity around the wafer will limit size of boule. It is difficult to dope to high levels. Vanadium can be used to achieve semi-insulating material. Wafer sizes from 25 mm to 75 mm, at 2000 \$/wafer. Most limiting are the micropipes, which are (defects) holes through wafer, 100/cm2.



Chemical Vapor Deposition, CVD is used for growth of doped epitaxial layers of SiC on the sawed and polished SiC wafers. Wafer doping 1e17-1e18. Epi doping 1e14-1e20.

Temperature 1400-1600 C, growth rate 2-4 um/h

Uses silane and propane.

Doping from nitrogen or trimethylaluminum.

Figure shows a cold wall reactor, an enclosed graphite susceptor is also possible (hot wall).



Heteroepitaxy is the growth of other materials on SiC. Materials of interest are the III-V nitrides, AlN, GaN and InN, or alloys of them. This materials system allows bandgap engineering in the range 2 to 6 eV with a small change in lattice parameter. Especially AlN is closely matched to SiC. Since GaN wafers can not be made yet, other substrates with good thermal and electrical conductivity is of interest for InGaN LEDs and lasers. However, here we will mainly consider these materials for heterojunctions between GaN and SiC.



Dopants in SiC: Al and B for p, N and P for n

Epitaxy allows uniform doping entire wafer, for selective doping we use diffusion or implantation.

For silicon diffusion can be performed at temperatures below 1200 C, which allows the successful use of silicon dioxide masks.

Solid state diffusion of dopants into SiC needs temperatures of 1800 C and higher. Graphite masks may be one possibility.

Ion implant causes damage, which is difficult to remove. High T during implant (700 C) and high T anneal (1200-1700 C) is necessary.

One advantage for SiC is the low diffusion. A narrow profile stays narrow, see figure.



Wet etching of SiC can be done with molten KOH (500 C), but no masks are known.

Dry etching with flourine species (CF4, SF6 etc) forms volatile SiF4, C is removed by sputtering.

A good mask for deep etching is Ni or Al, otherwise photoresist can be used.



Thermal oxidation of SiC is slower than silicon, but silicon dioxide is still formed. High temperatures are necessary (1100 C) and about 50% SiC is consumed. For thick oxides deposited oxides are used.

Use is for gate dielectrics in MOSFETs, surface passivation etc



Metal contacts are needed for all devices. Normally these should be Ohmic as opposed to rectifying (Schottky) contacts. A Schottky barrier is the result of a metal work function different from the fermi level (dashed line) in the semiconductor. Since it is not always possible to find a metal with the correct work function, which doesn't get pinned in the bandgap, the other way to achieve an ohmic contact is by heavy doping of the semiconductor along with a high temperature anneal to alloy the metal with the semiconductor.

The same metal can be used for both ohmic and Schottky, only different doping. Schottky gates are used in MESFETs.

Long term stability at high temperature is an issue.



As we have seen the process steps are conceptually the same as for silicon processes.

The main difference is perhaps that most steps are performed at higher temperature or energy.

Nothing is impossible, that is we have not seen a process step that would halt all device research. In comparison, diamond devices have no n-type dopant, GaN has no wafers, and several semiconductors have no natural, stable oxide for isolation.

Some would say say price and size, but the main obstacle for large scale production of devices is still the material quality.



Finally we will look at some different devices made in SiC, and compare their performance to the theoretical limits.

Lets take these by the number of terminals:

2, 3 and 4 respectively.

The last thing will be to look at the economy of things.



There are two main ways to make pn-diodes in SiC.

1. Ion implantation is similar to how silicon diodes are made. 2. Mesa etching of grown epilayers has been used since p-type implantation is not always successful in SiC.

The depletion refion which will support the voltage is marked W in this and following slides. Blue represents p-type SiC since p-type wafers are blu-gray, and green represents n-type since 6H n-type wafers are green.

A device like this will have problems for more than about 1 kV blocking voltage, due to field crowding at corners. The electric field at these points will be higher than the critical field before the field is high in the bulk. The solution is to use termination.



This figure shows four popular methods to terminate the junction. The idea is to 1. Avoid sharp corners in the field and 2. Spread out the electric field on the surface of the device where pn junction meets air.

JTE adds additional implants of lower dose outside the main implant.

Ar implant (without annealing the damaged area) creates a resistive region, where the leakage current will help spread the field.

Field plates spread the field by capacitive coupling.

Field rings are implanted circles outside the anode, which are floating, ie not electrically connected.



The simplest device we can make is actually the Schottky rectifier, which needs no implantation, only one type of doping. By selecting the metal correctly we get rectification.

It actually has a few advantages over pn diodes:

There is no reverse recovery since the metal can not inject minority carriers, so low loss for high frequency switching (up to 1 THz).

The forward voltage drop can be lower than for the pn diode, if the pn diode has little conductivity modulation. This assumes the same voltage drop in the blocking region. The built in voltage of the pn diode is proportional to the bandgap, hence close to 3 V for SiC, whereas the barrier height is normally chosen to around 1 V.



The Schottky has lower breakdown voltages than pn diodes, since tunneling occurs for high fields at the metal junction. What we really want is a device with low forward voltage drop but the high blocking voltage of a pn diode. This device is called a Junction Barrier Schottky device, or JBS.

It is made by combining implanted p-regions with the contact area.

In forward operation the current is dominated by the Schottky, but in reverse the pn-junctions block and the depletion region created shields the metal contact from high fields.

Normally these are not operated in high injection for SiC, so they are as fast as Schottkys.



Time for a comparison of 2-terminal devices. This figure shows theoretical on-resistance for depletion region and measured differential on-resistance versus blocking voltage.

As figure of merit Vb² / Ron is used, which shows how close to theoretical line we are.

PN junction has 1-2 V higher voltage drop than Schottky because of built-in voltage (for same onresistance and current level)

We also see how high injection changes things. The CREE pn diode improves 40 x when operating in high injection, 5 kA/cm2, lifetime 100 - 400 ns, TED nov 99.

Mostly Schottky, most relevant (maj carr devices).



Now to high voltage switches. MOSFETS operate by a isolated gate. With a positive voltage we create an inversion channel of electrons in the p-type material, which conducts the current.

Once again, two designs are possible, ion implanted and mesa etched. (meaning grow epi, then etch away)

To be able to block higher voltages, the drain to gate distance should be larger than source to gate, so that there is room for the depletion region..

(35 um between drain and gate)

Lateral MOSFETs are also possible as high temperature devices for ICs.



If we want high blocking voltages and large currents at the same time, a vertical approach can be better in terms of current per cm2 of SiC. This is also done in silicon and the designs arecopied from Si.

However, DMOS in Si refers to the double diffusion of dopants which control the gate length. In SiC, this is done as double implantation, so it is sometimes called DIMOS. Channel is lateral, and then vertical.

The UMOSFET also has a vertical channel, and the potential for better channel length control. Made by etching and ion implantation. Vertical current. Difficult make good gate oxides on etched surface.

Idea is that Rdrift should dominate, however Rchannel is also large!



The channel conductance, which is reversely proportional to the channel resistance, can be calculated from this relation. Z is channel width and L is channel length. Ueff is the effective channel mobility and CxV is the channel charge. By measuring channel resistance we can calculate the effective inversion mobility for experimental devices. In Si we get approximately 50 % of the bulk mobility, which we understand as the increased scattering at the oxide interface due to the attracting gate field. In SiC it looks like it is a factor 5 to 10 lower, but this may actually not be the correct way to understand this. If the channel charge in SiC is fixed to 80 %, the mobility will look 5x too small. A solution is to use accumulation channel, called a ACCUFET.



Another issue with SiC MOSFETs is the reliability. We have to make sure the oxide field is not too large. In the on-state it is calculated from gate voltage and oxide thickness, and it is over the source region this occurs. In off-state it is the field of the reverse biased pn-junction, which is multiplied by the ratio of the permittivities (E field should be continuous). Since the field close to breakdown is ten times higher in SiC than silicon, this is a bigger problem, since the field is multiplied by a factor 2.5. In silicon fields are generally 1-2 MV max, compared to breakdown at 10-12.



Another candidate for a switch is the bipolar transistor. Normally to support high voltage it is vertical, made by epi and mesa etching. This is also a candidate for high frequency operation. The emitter can inject electrons via the base to the collector, and this is the main current. To switch on, we need a base charge of holes, and a base current to supply them since a few leak into the emitter. The figure of merit, the current amplification factor beta, is the ratio of these. The transistor is improved if we reduce the leakage of holes, and this can be achieved with a wider bandgap emitter, called HBT. Also, beta is improved with less recombination in base, achieved by long minority carrier lifetime in comparison to transport time, ie base width, these 2 extra key parameters.



This figure shows a comparison for 3 terminal SiC devices, on resistance versus breakdown voltage.

The comparison is relevant since MOSFETs are all majority carrier devices. They are less succesful than Schottkys in achieving theoretical on resistance, which is due to the problem of channel resistance (mobility). There are very few bipolar results, but rumor says CREE has a 1800 V, 10.8 mohmcm2 bipolar transistor.

Note that quite a few designs have been tested, mainly vertical UMOS and DMOS both inversion channel and accumulation channel, but the highest voltage is laterally, 2.7 kV.



Now we will look at some high frequency devices in SiC. The first is the junction field effect transistor. The dual gates (front and back) control the channel resistance by varying the reverse bias and depleting and finally shutting off the channel.

This may not be the best design, since the gate length is difficult to make very short.

Current is majority carriers in n-channel, so it has higher mobility than MOSFET.



A better design is perhaps the MESFET, which uses a reverse biased Schottky junction as gate to control the channel resistance.. Submicron gatelengths can routinely be achieved. Thanks to semi-insulating wafers, the parasitic drain capacitance is reduced, and newer designs also use via holes for the source (rather than bonding wire) to reduce the parasitic source inductance. This figure shows mesa etched source and drain, and a recessed gate. Ion implanted source and drain are of course also possible. Gate is placed closer to source to increase breakdown voltage, depletion region W spreads from drain. These devices have achieved impressive amounts of power per mm gate width.

Even better is the HFET.



The heterojunction field effect transistor, also called MODFET or HEMT, is not really SiC. However, improved devices have been made with SiC rather than sapphire substrates in terms of amount of power per area, thanks to better thermal conductivity of SiC.

Operation is as follows: The wide bandgap material creates an inversion channel of electrons. The mobility is very high due that channel can be made undoped. The channel charge comes from the AlGaN and the strong piezoelectric effects of GaN.

Gate bias changes the band bending and can remove or enhance the channel charge.



There is also a vertical device for high frequency. It can not operate at as high frequency, but total power can be larger. This is the Static Induction Transistor, also known as a Permeable base transistor. These operates as a vertical JFET or MESFET (or triode for vacuum electronics people).

Submicron lithography and etching is needed, as well as technology to connect a large number of these fingers.



Here is the comparison of high frequency devices. There is no straightforward figure of merit to compare these, since operation frequency, amount of power, pulsed or continuous operation makes a big difference. Here is an example of some recent devices, plotted as power / mm gate width versus frequency at which this was measured. There are two things to understand from this:

HFETs > MESFETS > SITs etc

There is a tradeoff between power and frequency

It is not clear how cooling is handled at these high power levels, ten times that of silicon or GaAs. Ten times the power at the same thermal resistance will cause a temperature rise of ten times.



Finally lets say something about thyristors. This is the power switch of choice for silicon high power, since low on resistance can be achieved through conductivity modulation. The n- drift region will be depleted from the anode in reverse blocking, and depleted from the p base in forward blocking. When charge is injected from the gate in forward operation, we can get the anode and emitter to both inject electrons and holes of concentrations much higher than the doping (and hence reduce resistance). To turn off, the gate has to remove this charge so that the depletion region can reform.

For SiC this needs longer minority carrier lifetimes than are available, and larger defect free areas than what is routinely available. No comparison.



Instead lets look at the economy of SiC devices versus device size. Using a simple model for the yield of working devices versus device area and defect density / cm2, we get this plot. For small devices, 2-3 mm chip length (microwave), more than 50 % yield can be achieved, and for very small such as LEDs 0.5 mm square there is production. However, for power devices needing many mm, this is not yet true. This is all due to micropipes.



Many devices were shown, there are a few more.

There is a trend of vertical use for power (large area) and lateral for high frequency (lithography to determine gate length and parasitics).

Breakdown voltages are close to theoretical for diodes, but not MOSFETs. The MOSFET mobility is the obstacle, too high channel resistance.

Thermal modeling will probably be important due to the high power densities.

Production economy is still questionable.



Before the conclusions, here are some references online from which I have borrowed some figures.

VETENSKAP OCH KOST HÖCSKOLAN	CONCLUSIONS				
WHY SIC?					
Excellent Properties!					
HOW SiC?					
Higher T than Silicon!					
WHICH SiC?					
Small	Area for Economy!				
© 2000 Carl-Mikael Zetterling					

In conclusion, I hope I have been able to convince you that SiC has excellent properties, that higher temperatures are needed for the processing, and that great devices can be made, but I will give no guarantees on the economy of it all.

Thanks for your attention!