ERICSSON ⧉

Inside the RF Power Transistor

Prepared by Ted Johansson, Dr. Tech.
Process and Device Design, Business Center RF Power,
Ericsson Components AB, Kista, Sweden

Introduction

The purpose of this application note is to show some of the chip level design considerations and technical details for Ericsson's bipolar RF Power transistors. Whether you are an amplifier designer, applications engineer, technical manager or just generally interested, these devices often appear to contain quite a lot of magic inside, and can require even more magic to use! This is partially true, if we define magic as something that is not easily calculated, or in these days, simulated on a computer.

The bipolar RF power transistors are among the oldest silicon devices.

High-frequency transistors were first fabricated in germanium in the late 1950s but were soon replaced by silicon bipolar transistors at the beginning of the 1960s and have since then dominated the RF-power area. In 1971, H. F. Cooke described the theory and design of these transistors in an often-cited paper [1]. Much of his text is still applicable on today's devices. For cellular radio, the bipolar transistors are totally dominating and can deliver great performance up to and beyond 2 GHz with good stability, availability and price.

Cooke wrote in his article: "Not only are they difficult to build, but most of the more desirable parameters have been optimized." He was right about the difficulties in building the devices, but luckily, he was wrong about the optimization! Detailed knowledge in semiconductor device physics, device fabrication, RF & microwave theory and practice, and a lot of hands-on work are required, but it is still possible to improve performance, more than 25 years after Cooke's statement.

The improvements have mostly come from advances in the process and fabrication techniques. Progress in CMOS IC technology, photolithography, dry etching, metallization techniques for III-V technology, process and yield control has been utilized. The RF power transistors are now fabricated using 4" or 6" wafers with very high yield. Computer tools are available for process and device simulations. For accurately predicting the RF large signal performance, the tools still need to be improved, particularly when considering a complete packaged de-
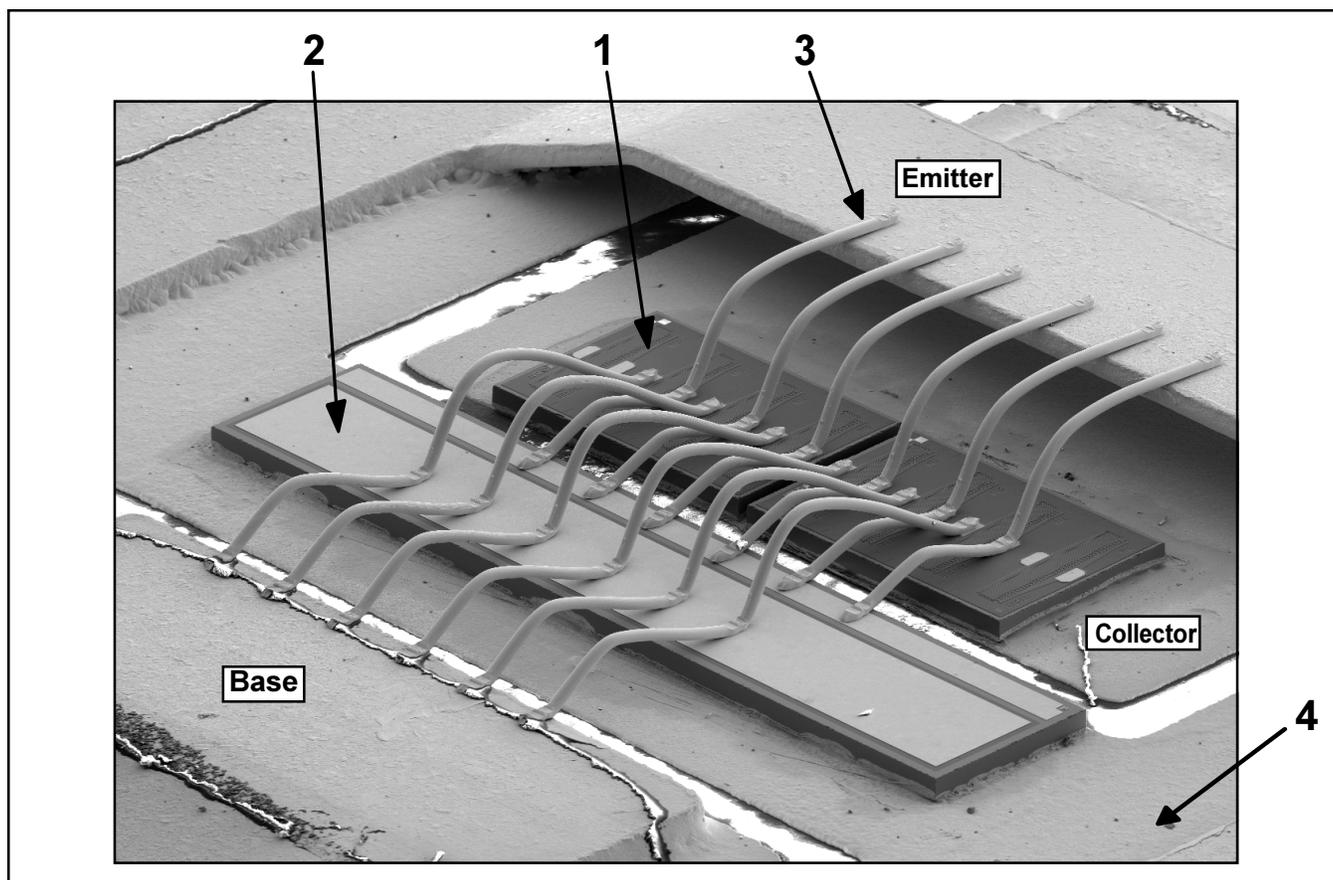
Figure 1. Inside Ericsson's PTB 20105 bipolar RF power transistor.

vice with bond wire geometry and internal matching capacitors.

RF power devices have not been paid much attention to from the scientific/academic point of view since the 1970s, except for the advances in III-V power device technology. Considering the evaluation of computing power, simulations tools and process technology, and the broad commercial interest, RF power has now a great research potential for the study of new structures, thermal behavior, linearity, bonding and matching, packaging and amplifier design.

Today, we are not close to any fundamental limits for the RF power devices. They will continue to evolve as the high-volume applications in the telecommunication area continue to go to higher frequency bands. There is potential to further improve the silicon for power applications in the 2 to 3 GHz range. The excellent price/performance, stability and availability of the silicon RF power devices will guarantee continued growth in the wire-less area.

Inside the package

To illustrate some device considerations for 1 to 2 GHz device technology, let us take a closer look at what is actually inside a transistor package!

Figure 1 shows an open transistor, the PTB 20105 (925-960 MHz, 20 W transistor, typical gain of 10 dB) as seen by Scanning Electron Microscopy, SEM. The magnification is 27 times.

The following parts can easily be identified:

(1) The silicon chips.

In this particular transistor, two transistor chips (1.4 x 0.9 mm, 55 x 35 mils), containing 4 transistor cells each, are used in parallel (actually only 3+3 transistor cells are connected). The chips are mounted very precisely, close to the edge of the metallization area, to minimize critical bond wire lengths, and to ensure the uniformity of the electrical characteristics.

The silicon chips have been lapped to a thickness of around 120 µm (<5 mils) from its initial thickness of 525 µm (21 mils) to lower the thermal resistance to the package.

The design considerations for making a good transistor at the chip level are discussed in the next sections.

(2) The chip capacitors for internal matching.

An input chip capacitor is connected to the input side (base), while no output matching was necessary for this device. A highly doped (n+) silicon is used for the chip capacitor, which is of MIM (metal-insulator-metal) type, with one "metal" being the highly doped substrate silicon and the other metal being a similar gold metallization as used on the transistor chips. The insulator is silicon dioxide or silicon dioxide + silicon nitride.

The chip capacitor has also been lapped to 120 µm, same height as the transistor chips, to ease the bonding.

(3) The bond wires.

The emitter wires go to a "bridge" that is connected to the emitter flange of the transistor package and the socket. In the other direction, the wires are connected to the substrate contact (grounded) on the chip capacitor.

The base wires first land on the capacitor and are then bonded to the base flange of the package.

The geometry of the bond wires is very critical to the device performance and bonding has to be done with great precision and repeatability.

(4) The package.

The package is usually gold plated Cu/W. Since the bipolar transistor chips have their collector (connected to the supply voltage) on the backside, steps must be taken to isolate the chip mounting area from the grounded package by using an isolating material, while providing a low thermal resistance through the package. Beryllium oxide (BeO) has the unique properties required and is used in most isolated packages today. (The BeO in figure 1 is visible as the reflecting areas ("water"), because, being nonconductive, it charges and reflects the electrons used as "light" in the SEM.)
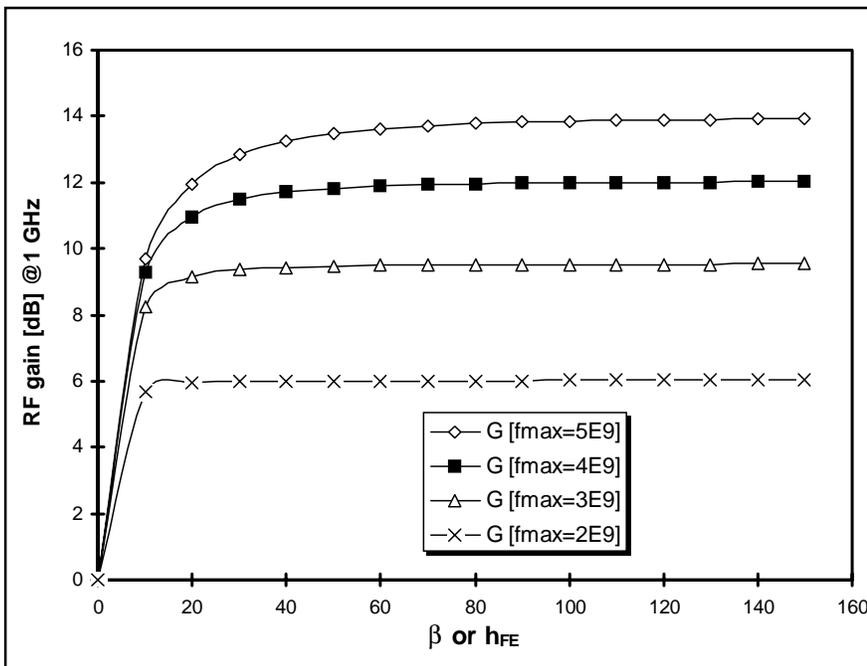
Figure 2. The RF gain at 1 GHz as a function of the DC gain (h_{FE}) and f_{max}.

ERICSSON ≋

Device considerations

During the device design, a number of critical considerations have been made on the chip level.

A few of them will be discussed in this section.

Breakdown voltage, epi selection

The design of transistors intended for high power applications follows a different path than transistors for small signal analog or digital circuits with supply voltages of 5 V, 3.6 V or lower. A high collector-base breakdown voltage (BV_{CBO}) is desired, because the devices need to be operated at high supply voltage (26 V) to achieve high output power. Furthermore, a switching transistor's collector side experiences up to two times the supply voltage during large-signal operation with inductive load. To support this high voltage, the collector epitaxial region (epi) must be thick, typically 6 to 8 µm, to avoid reach-through type of breakdown, and have a resistivity value (doping level) typically in the 1 to 2 Ωcm range, set by the avalanche breakdown of the junction.

The final selection of epi thickness and resistivity also affects base push out or Kirk effect (higher doping gives higher saturated output power), gain (collector-base capacitance), high frequency characteristics (f_T), resistive power loss ($P_{loss} = IR^2$), ballasting (see ballasting section, below), and linearity of the device.

Because of junction curvature, the theoretical breakdown voltage of the collector-base junction cannot be reached. However, by careful geometrical layout and the use of collector depletion rings or graded diffusions at the device edges, the sharp peaks of the electrical field can be reduced and breakdown voltage raised near its theoretical limit.

An empirical formula for the relationship of breakdown voltages in the transistor and current gain, β or h_{FE}, states [2]:

$$BV_{CEO} = \frac{BV_{CBO}}{\sqrt[n]{\beta}}$$

where BV_{CEO} is the collector-emitter breakdown voltage with open base, BV_{CBO} is the collector-base breakdown voltage with open emitter, and n is an empirical constant, usually between 2.5 and 4.5, related to the nature of the BC-junction breakdown.

This formula shows that for a given epi doping and device design (constant n), BV_{CEO} and β are directly correlated: higher β gives lower BV_{CEO}. If n can be improved, by different doping profile tricks or other means, BV_{CEO} can be increased for a given epi resistivity and β.

Data sheets often specify BV_{CER} instead of BV_{CEO}. A small resistor is then connected between the base and emitter. If the resistor is small enough,

BV_{CER} will approach BV_{CES}, which is close to BV_{CBO}.

High frequency power gain

The power gain is usually the parameter most limited by the frequency, since the devices normally are operated at the frequency limited portion of the characteristics with a (theoretical) 10 dB/decade degradation.

The gain at RF can be described by the following relationship [3]:

$$G(f) \approx \frac{G_0}{\sqrt{1 + G_0^2 \left(\dfrac{f}{f_{max}} \right)^4}}$$

where G_0 is the zero-frequency gain (β or h_{FE}) and f_{max} is the maximum oscillation frequency, or the frequency where the power gain is equal to 1.

A plot of h_{FE} versus G is shown in figure 2 for different f_{max} values at f = 1 GHz. From this plot it can be concluded that a high f_{max} and a not too low β are detrimental for a good RF power gain. If β is higher than required for a constant RF gain, the device will be sensitive to parasitic oscillations, and will have a decreased BV_{CEO} (see previous section). High β also usually means high base resistance, which will lower the f_{max} (see below) and consequently the RF power gain.

An approximate expression for f_{max} is:

$$f_{max} \approx \sqrt{\frac{f_T}{8\pi R_b C_{bc}}}$$

where f_T is the transition frequency, or the frequency where the current gain is equal to 1, R_b is the base resistance and C_{bc} is the collector-base capacitance. f_T is a function of delay and charge times in the vertical C/B/E structure of the transistor. It is dominated by the base delay, which is proportional to the square of the base width.

Ballast resistor **Active area** **Cross section in figure 4**

Base

Emitter

Backside: Collector
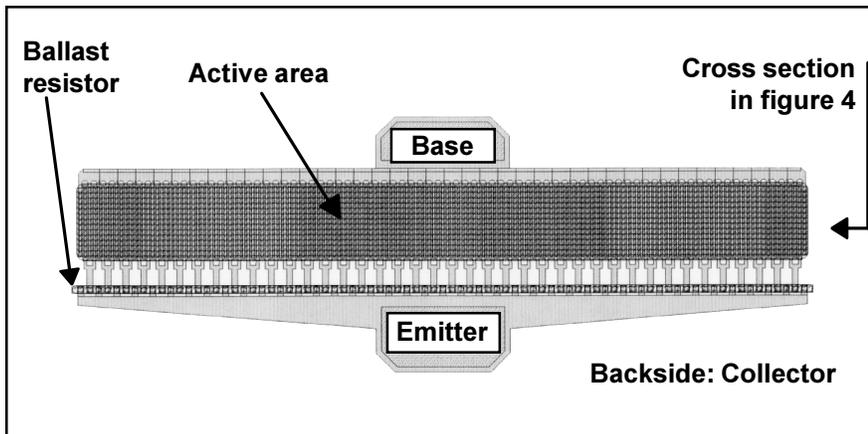
Figure 3. Interdigitated cell layout for RF power transistor .

3

High power gain at RF could be achieved by having a high f_T and a similar high f_{max}, but this is generally hard to accomplish, considering the delay times and the high supply voltage (the "Johnson limit"). In addition, a thin base (high f_T) means more base resistance (lower power gain) and/or higher β (decreased BV_{CEO}) and tighter process control.

Instead, devices are designed to have an f_T not very much higher than the operating frequency (1.5 to 2 times), and then f_{max} is maximized (R_b and C_{bc} minimized). f_{max}/f_T ratios of 2 to 3 can usually be obtained. As a comparison, f_{max}/f_T ratios for low-voltage high-frequency transistors are usually less than one.

Reducing the base resistance (higher base doping and lower parasitic resistances) and reducing collector-base capacitance (mostly by lower parasitics) is the secret to high power gain. R_b and C_{bc} are both influenced by vertical dimensions (doping profiles, isolation thickness) and lateral dimensions (layout). All major and minor contributions to these two parameters must be considered and actions taken to minimize them to achieve good RF gain.

Transistor layout for high gain and output power

Through the years, a number of layout concepts have been in use [1], but for cellular applications today, the inter-digitated cell layout in figure 3 is totally dominating. For the common-emitter configuration, this layout offers the best performance with respect to RF gain, power/current distribution and area utilization.

The most critical parasitic capacitance for the gain is between collector and base. In a vertical RF power transistor, the silicon substrate is connected to the collector. The device area consists of a single diffused base tub, which area must be minimized for a given power requirement to obtain best possible gain.

The base metal also contributes to collector-base capacitance, and is therefore as narrow as possible. The emitter metal area, which must feed high currents, has been tapered toward the ends of the cell to save some additional collector-emitter capacitance.

The bond pads do not have square form; instead they are rectangular, similar to the "foot-print" of the bonding wire. To further reduce capacitance, the pad corners have been rounded.

The collector is connected to the back-side of the chip, so only two types of bond pads, emitter and base, need to be bonded using wires.

By dividing larger cells into smaller units and spreading them over the silicon surface, the total thermal resistance can be lowered, but at the cost of increased interconnect metallization capacitance. In practice, the thermal improvements are moderate with spread-out layouts. Ericsson's high-power 2 GHz layouts use very large interdigitated cell layouts with several hundreds of parallel emitter fingers.

High output power is achieved by paralleling several cells on a chip and several chips in one package. In the PTB 20105 transistor in figure 1, each chip has four cells, and in total, two chips and six cells are used. The cells can be interconnected on the chip level to achieve better current balance and the chips connected by using inter-chip bond wires.

Emitter ballasting resistor, ruggedness

Also indicated in figure 3 is the *emitter ballast resistor*.

As the power increases in the transistor, the junction temperature also increases. If the current is not perfectly distributed over the active transistor area, *hot spots* will appear, which very quickly lead to thermal runaway and catastrophic failure. To ensure even current distribution, resistors are added in series with the emitter fingers. One resistor can be connected to an individual emitter finger or a small group of fingers. Any undesired increase in the current through a particular emitter or group of emitters will be limited by the resistor (effectively, V_{BE} will be decreased) [4]. This is called emitter ballasting and can be implemented as NiCr resistors, diffused resistors or polysilicon resistors.

Ericsson's transistors use diffused resistors, which are easy to fabricate and control. The resistor value can easily be optimized for different products by changing one implantation dose. Drawbacks are larger CE capacitance contribution compared to polysilicon resistors, and the risk of lower BV_{CES} values if the resistor doping profile is too shallow or too deep.

The most common cause of device failures are high current standing wave ratios (VSWR) which occur at high power levels during load mismatch. The resistance of the epitaxial silicon limits high currents and increases the device's ability to survive during load mismatch; the RUGGEDNESS is improved. The additional epi resistance is called collector
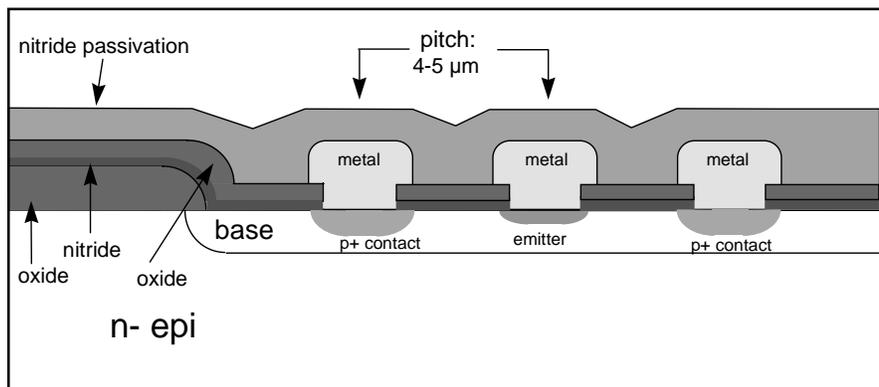
Figure 4. Cross-section of device structure for 1 GHz technology.

nitride passivation

pitch:
4-5 μm

metal   metal   metal

base

p+ contact   emitter   p+ contact

nitride

oxide   oxide

n- epi

ERICSSON ⚡
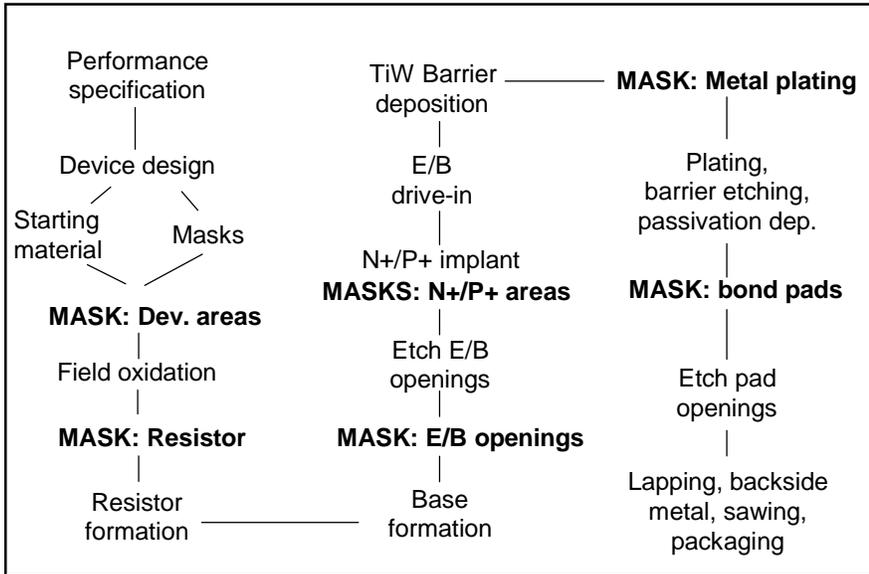
Figure 5. Basic design and process flow for a bipolar RF Power transistor.

ballasting. It increases the saturated V_{CE} and reduces the voltage swing over the collector.

Adding ballasting has one drawback—it lowers the output power or the gain. Which is the best method for increasing the ruggedness, emitter or collector ballasting? The high VSWR during load mismatch creates large temperature variations. A general limitation of the collector current (collector ballasting) is less efficient to prevent the occurrence of hot spots leading to catastrophic failures. Therefore, the distributed limitation (emitter ballasting) is the preferred method for increasing the ruggedness.

Basic technology for 1 GHz

Figure 4 shows a schematic cross-section for the device structure in use for up to 1 GHz. A typical interdigitated cell layout was shown in figure 3, with the position of the cross section in figure 4 marked with an arrow. Common silicon planar technology is used. A principle design and process flow is shown in figure 5.

At high current levels, only the periphery of the emitter areas is delivering power because of the current crowding effect. The emitter perimeter is therefore made very large and the fingers very narrow. The base contact fingers are placed between the emit-

ters, thus reducing the extrinsic base resistance from the base contact to the emitter edge, to increase RF gain. The metallization extends on top of the emitter and base contact areas. Very little parasitic resistance is added compared to other layout principles.

There is no interdevice isolation; the whole silicon substrate is the collector, which means that no circuit functions can be designed using a single chip in this technology. Emitter and base are contacted using the front side bond pads.

There are various methods to obtain high breakdown voltage at the device edges. For Ericsson's 1 GHz technology, a classical method with three floating depletion rings is used. The

rings have the same doping as the base region, but are not connected. This method requires no extra process steps, only changes in the mask layouts.

The "pitch," defined as the emitter-base opening periodic distance, is 4-5 µm for a 1 GHz device, and the emitter and base openings are typically 1 to 1.5 µm wide.

The very high currents and narrow metallization lines complicate the use of aluminum (Al) for metallization. Electromigration, the wear-out phenomenon at high current densities in thin metallization lines, will eventually cause device failures and limit the long-term reliability of the devices.

Instead, metallization consists of a single layer 1-2 µm thick gold (Au), which is very resistant to electromigration. The metal can be sputtered and etched/ion milled or electroplated using a resist mask. All Ericsson RF power transistors are electroplated. A titanium-tungsten (TiW) layer is used under the gold to ensure excellent reliability of the metallization system and adhesion to the underlying layers. A silicide (preferably platinum silicide, PtSi) is often used in the emitter and base openings to lower the contact resistance and thus the parasitic base resistance.

Basic technology for 2 GHz

As the demands for higher operating frequencies have grown, a structure similar to figure 6 has been brought
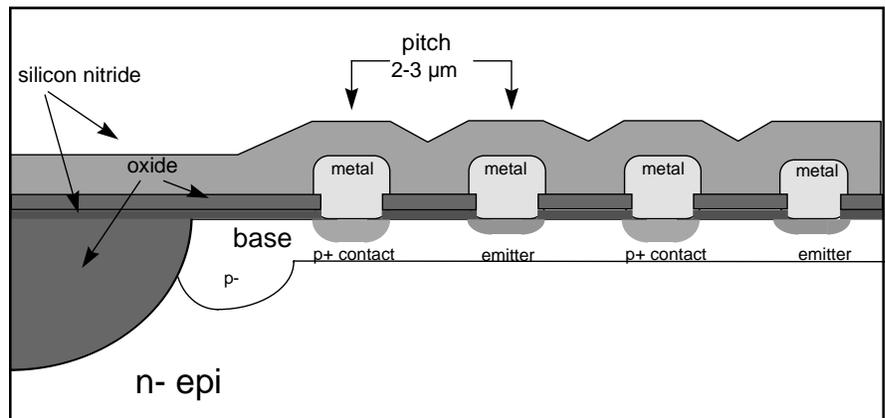
Figure 6. Cross-section of device structure for 2+ GHz technology.

into use for PDC (1500 MHz), PCS (1900 MHz), etc. and for new frequency bands above 2 GHz.

To reduce parasitic capacitance from the metallization to the substrate to sustain RF gain, the thickness of the field oxide isolation has been increased from around one micron to at least three microns.

Such a thick isolation can be achieved by increasing the initial thermally grown oxide or depositing a thick isolation layer, and then an etch to open the device areas. Another method is to use a LOCOS isolation, similar to a CMOS device but with much larger oxide thickness.

However, to successfully open the small emitter/base contacts and metalize the narrow openings, reducing the topology is necessary. Instead, various forms of fully recessed local oxidation can be applied with an almost planar structure as a result. The recessed structure is obtained by depositing a nitride/oxide layer as for a LOCOS structure. The nitride/oxide is then etched outside the device areas. The etching is continued up to one micron into the silicon substrate. After a long oxidation, that can be accelerated by the use of high-pressure oxidation (HIPOX), and a planarization to reduce the "bump" at the oxide/device area edge, the almost perfectly planar structure in figure 6 is obtained.

To further reduce parasitic R and C to increase gain, the pitch has been scaled from 4 to 5 µm for the 1 GHz device to half that value. The emitter and base openings are typically 0.7 to 1.2 µm wide in this technology.

Another change to reduce parasitic capacitance, a different edge breakdown method, was applied for this structure, namely junction termination (JTN). By increasing the doping near the device edges (by implantation and diffusion), the collector-base breakdown was increased to very close to the theoretical limit. The JTN occupies less area than the guard rings used for the 1 GHz devices, thus lowering the C-B capacitance.

Similar to high-speed low-voltage bipolar transistors, the use of polysilicon emitters can improve the performance, mainly by higher transistor emitter efficiency, resulting in higher f_T and f_{max}. However, the improvements appear to be quite modest for this type of device and the problems associated with manufacturing stability and repeatability are well known. Ericsson's 2 GHz technology does not use polysilicon emitters, but will when the pitch is shrunk below 2 µm.

References

[1] H. F. Cooke, "Microwave Transistors: Theory and Design", Proc. IEEE, Vol. 59, No. 8, p. 1163, 1971.

[2] S. M. Sze, "Physics of Semiconductor Devices", 2nd Ed., p.151, John Wiley & Sons, Inc., 1981.

[3] R. Allison, "Silicon Bipolar Microwave Power Transistors", IEEE Trans. Microwave Theory & Techniques., Vol. MTT-27, No. 5, p. 415, 1979.

[4] R. P. Arnold, D. S. Zorglu, "A Quantitative Study of Emitter Ballasting", IEEE Trans. El. Dev., Vol. TED-21, No. 7, p. 385, 1974.