ADVANCED BIPOLAR TECHNOLOGY FOR THE 1990's

IBM Research Division, Thomas J. Watson Research Center,
P.O. Box 218, Yorktown Heights, NY 10598; (914) 945-3433

ABSTRACT

The advent of low temperature epitaxy processes provides a new degree of freedom for bipolar device scaling. This paper describes new vertical scaling concepts and process technology elements required for advanced scaled bipolar (NPN and PNP) devices which will be the core of high-performance application-specific bipolar, BiCMOS, or complementary Bipolar/BiCMOS logic and memory chips. In particular, we address key issues such as transit time reduction by SiGe base band-gap engineering, junction field/capacitance control by using lightly-doped emitter (LDE) and collector (LDC) concepts, lateral scaling (reduction of parasitic R and C) by advanced self-aligned structures and trench isolations, and liquid-nitrogen temperature (LNT) operation. Challenges for future BiCMOS and complementary bipolar/BiCMOS process technologies are examined.

I. INTRODUCTION

The decade of 1990's marks a new era of advanced high performance bipolar technology. Double-poly self-aligned device structures with ion-implanted intrinsic profiles and trench isolation have recently become the industry standard [1-4]. Yet there are many new opportunities and challenges for scaling the vertical profiles of future advanced bipolar devices, e.g., thin base formation, emitter depth/interface and current gain control, collector profile, junction breakdown and capacitance control, and band-gap engineering. Low temperature epitaxy (LTE) techniques [5-7] together with reduced thermal processing facilitate thin base formation and band-gap engineering, pushing the cut-off frequency of Si bipolar transistors toward the 100 GHz range [8, 9]. Moreover, a coordinated vertical and lateral scaling becomes even more important for thin-base transistors and calls for novel self-alignment and process integration schemes.

II. VERTICAL SCALING

II.1. Thin Base Formation and Band-Gap Engineering

Conventional vertical scaling of bipolar transistors for ECL circuits [10] calls for the reduction of base width without emitter-collector punch-through and the increase of collector doping to suppress the Kirk effect. Very thin base (≤80 nm) and high cut-off frequencies (fy ≥ 50 GHz) have been achieved by low energy ion implantation and scaled polysilicon emitter junction depth [11]. Further reduction in transit time is possible, however, only with (i) more abrupt doping profiles by LTE for very thin base widths, and (ii) band-gap engineering in the base. SiGe band-gap engineering in the base reduces both emitter charge storage time, τe, due to higher current gain, and base transit time, τb, due to a quasi-electric field (~20 KV/cm) from alloy grading (Fig.1 and Fig.2). A cut-off frequency (fy) up to 70 GHz and an ECL delay of 24.6 ps [12] have been realized in a self-aligned HBT with graded-band-gap SiGe base (Fig.3 to Fig.6).

II.2. Junction Field and Capacitance Control

The high doping concentrations in the base and collector required for scaled bipolar transistors lead to high capacitances and high electric fields which result in performance degradation, junction leakage and low breakdown voltages. These problems can be overcome by introducing a lightly doped emitter (LDE) [12, 13] and a lightly-doped collector (LDC) [14] using LTE techniques as shown in Fig.7. These lightly-doped layers help reduce junction electric fields, in a way similar to the lightly-doped drain (LDD) concept for MOSFET's.

LDE reduces the tunneling current, reverse leakage current, and increases the breakdown voltage of the E-B junction (Fig.8). It also lowers the E-B capacitance, giving a higher cutoff frequency over a wide range of current densities (Fig.9).

In a Si BJT with LDE, there will be an increase in the base transit time due to the retarding field from the retrograded boron profile in the LDE layer. However, there is very little transit time penalty for the SiGe HBT where the quasi drift field in the graded SiGe base overcomes the retarding field.

LDC reduces the impact ionization (multiplication factor, M-I) for a given cut-off frequency as seen in Fig. 10. The penalty in B-C junction transit time due to LDC is minimal since carriers travel at saturation velocity or even higher speed.

LDC therefore improves BVCEO with very little fy penalty (Fig. 11). It is also clear from Figs. 10 and 11 that a properly designed SiGe HBT with band-gap engineering in the base shows improvement in fy over the Si BJT without sacrificing the multiplication factor or BVCEO. Even though the impact ionization rate of a bulk SiGe material is expected to be higher, there is clear evidence (Monte Carlo simulation and experimental results [15]) that impact ionization is a non-local effect which depends mostly on the carrier energy, and therefore, the presence of a thin SiGe layer in the base has little or no noticeable effects on the impact ionization and breakdown voltage.

II.3. Scaled Polysilicon Emitter

Advanced bipolar transistors also require scaling of the emitter size (emitter stripe width) and junction depth without degrading the current gain, gm, emitter resistance, and yield. Process challenges include the doping and activation of the emitter poly with reduced thermal budget (especially if the emitter window topography is severe), the control of the
poly/mono Si interface quality and E-B junction depth, and the ability to ensure a good emitter contact metallurgy. It has been shown [16] that a non-uniform emitter junction depth and excess leakage current can result from perimeter depletion and emitter plug effects in narrow and shallow polysilicon emitters with significant topography. An in situ doped emitter process will alleviate the emitter doping concern while also allowing reduced thermal cycles in the subsequent steps to minimize the emitter junction depth. It may present additional process complexities in a complementary bipolar process. There has also been significant progress in wide-gap emitters such as SiC [17]. A wide-gap emitter can reduce emitter charge storage and 

Generic LTE related issues include material quality, stability, defect density, doping profile control and reproducibility, and selectivity, etc. Reduced thermal processing is required to maintain the thin base profile and device performance.

IV. COMPLEMENTARY BIPOLAR AND BiCMOS

Significant breakthroughs have also been made in Si PNP transistors (Fig.17) with 

The challenges of integrating high performance NPN, PNP, and CMOS devices include low-resistance, dual buried layers without autodoping, the optimization of collector epi and well thicknesses, thermal cycle compatibility, process complexity (Fig. 18), cost, etc. Modular processes and judicious trade-offs among various devices/elements according to each specific application are necessary for high performance bipolar/BiCMOS and complementary circuits at affordable cost.

V. CONCLUSIONS

We have proposed and illustrated new scaling concepts for advanced high performance bipolar technologies in the 1990's and beyond. New opportunities and challenges offered by low temperature epitaxy and band-gap engineering remain to be explored and exploited for high speed ULSI applications.

REFERENCES

For linear grading ($\Delta E_{\text{G}} > kT$):

$$
\frac{\tau_{E}(\text{SiGe})}{\tau_{E}(\text{Si})} = \frac{\beta(\text{SiGe})}{\beta(\text{Si})} = \frac{R_{\text{bl}}(\text{SiGe})}{R_{\text{bl}}(\text{SiGe})} \cdot \frac{e^{-\Delta E_{\text{G}}/kT}}{\Delta E_{\text{G}}} \cdot \frac{kT}{\Delta E_{\text{G}}}
$$

Fig. 1. Bandgap engineering using a graded SiGe base.

Fig. 2. Simulated delay components and cut-off frequency showing the reduction of the base transit time ($\tau_b$) and emitter charge storage time ($\tau_E$) with SiGe base.

Fig. 3. SIMS profile of a self-aligned NPN SiGe-base HBT.

Fig. 4. Gummel plots of self-aligned NPN Si BJT and SiGe-base HBT showing collector current increase in SiGe HBT.

Fig. 5. Measured cutoff frequency of self-aligned SiGe and Si NPN transistors.

Fig. 6. Measured average gate delay of self-aligned SiGe ECL circuits. A minimum delay of 24.6 ps was achieved with tighter lithographic ground rules (Lith-2).
Fig. 7. Schematic NPN doping concentration profile showing the lightly doped emitter (LDE) and lightly-doped collector (LDC).

Fig. 8. Measured E-B diode characteristics showing reduced forward and reverse leakage currents using LDE.

Fig. 9. Measured cutoff frequency improvement over a wide range of collector current densities using LDE.

Fig. 10. Measured $f_t$ and avalanche multiplication factor (M-1) showing the benefit of LDC.

Fig. 11. Measured and reported $f_t$ vs. $BV_{CEO}$ tradeoff showing the benefit of LDC and SiGe base.

Fig. 12. Measured $f_t$ improvement of SiGe-base HBT with 45 nm base width. A 40% improvement in peak $f_t$ is obtained from 298K to 85K.
Fig. 13. Measured average gate delay of a SiGe ECL circuit as a function of temperature, with 28 ps at 84K. Also shown for comparison are reported gate delays of conventional ion-implanted Si ECL circuits.

Fig. 14. Measured average gate delay of a SiGe ECL circuit with reduced emitter area showing a 112 fJ power-delay product at 84K.

Fig. 15. Improved power-delay using a butted emitter (solid shape) structure with shallow trench field isolation (hatched area). Ideal device characteristics and 90 fJ power-delay product were achieved.

Fig. 16. Schematic cross section of mesa self-aligned epi-base device structure (MSST [18]).

Fig. 17. Advances in PNP performance (cutoff frequency)

Fig. 18. A rough comparison of the process complexities of advanced BiCMOS, complementary bipolar, and complementary BiCMOS based on a high-performance NPN core technology. Sub-structures include buried layers, isolation, well implants, and collector reach-throughs. Super-structures include the steps for NPN, PNP, and CMOS structure definition and doping.