Scaling Towards 300 GHz f_T/f_{MAX} SiGe Transistors

Basanth Jagannathan, Jae-Sung Rieh, and Greg Freeman, Semiconductor Research and Development Center, IBM Microelectronics, Hopewell Junction, NY 12533 USA

basanth@us.ibm.comT Telephone: 01-845-892-8913 Fax: 01-845-892-3039

Reflecting the trend set by the microprocessors, there has been an ever-increasing need for speed in the chips powering communication systems. Wireless applications at 5GHz and wired applications at 10Gb/s are commonplace today. Consumer applications at such high frequencies are leading to a need for even higher bandwidth chips at the back planes and network access points. SiGe technology is a proven market place leader at application frequencies below 10Gb/s and there have been many recent 40Gb/s circuits demonstrated in SiGe as well [1]. From a SiGe HBT perspective there has been a constant march towards developing devices with higher f_T and f_{MAX} with many reports of devices being fabricated with f_T and f_{MAX} >150GHz [2][3]. Since the technology is generally integrated with a CMOS platform, it has the power to enable the next generation of applications that are rich in functionality at a fraction of the cost that would have been deemed feasible a few years back.

SiGe HBT technology provides a path for fabricating transistors with improved electron transport characteristics in the Si material system. Scaling the HBT technology along with the CMOS ushers-in applications that would otherwise be enabled at a later CMOS technology node or perhaps not feasible at all. Scaling the SiGe devices in the vertical dimension is made simple by the ability to grow with ease and control tailored profiles incorporating Ge, C, B in Si. Furthermore, sharing the technology elements with CMOS results in applying the advanced CMOS lateral process learning to shrink the HBT in the lateral dimension. In the three generations of IBM SiGe technology where the NPN was scaled, every generation has seen a doubling of f_T and f_{MAX} to it current value of 200GHz f_T and >250GHz f_{MAX} in the latest pre-production technology.

SiGe NPN devices in the IBM manufacturing environment have generally been implemented with an As-doped polysilicon emitter, B-doped SiGe base, a P-doped collector and an As subcollector. Over the three generations, scaling of the devices has significantly reduced base transit time and time constants associated with the collector base capacitance. This scaling was implemented by reducing the base and collector thickness and increasing its doping concentration. The collector scaling is an area of intense activity in the design of high performance NPNs. As the collector is scaled, the optimization process has to consider the effects of transient enhance diffusion (TED) of both base and collector dopants and also balance the effects of barrier formations in the collector. Since the collector is formed by implantation, the continuous vertical scaling results in complex TED interplays in both the intrinsic and extrinsic device. In the past, the extrinsic base in the high performance device was formed by implanting boron self-aligned to the emitter mandrel with a spacer separating the emitter from the extrinsic base. Lithographically scaling this device to minimize base resistance and collector capacitance could not be sustained because of the TED from the extrinsic base implant interacting with both the intrinsic boron and the phosphorus collector. In order to maintain the benefits of the lithographic scaling it became crucial to decouple the scaling of the extrinsic device from the intrinsic device. This provides the ability to scale f_{MAX} independent of f_T. In the latest generation of SiGe NPN devices at IBM, the key feature is the introduction of a raised extrinsic base which can support the existence of narrow spacer separating the emitter from the base contact (Fig. 1) to achieve a low base resistance. Fig. 2 and Fig. 3 show the AC and DC characteristics of such devices. Ring oscillators performance (Fig. 4) using these devices indicate that compared to the previous generations the stage delay is reduced as a result of significant reduction in the RB*CCb product (Fig. 5). A break-up of the components of transit time in these devices suggests that base transit time, collector base capacitance, and collector resistance can be further minimized in subsequent generations. Recent experiments at IBM have demonstrated devices with f_T and f_{MAX} approaching 300GHz. These results will be presented at the conference.

One consequence of such aggressive scaling is the designer concerns on device yield, matching, breakdown voltages, avalanche characteristics and reliability. The disposable mandrel and the spacers defining the emitter dimension and its separation from extrinsic base utilize standard CMOS processes ensuring a 3sigma variation within 10% of process targets. Process splits created to understand the effect of base doping show relative insensitivity of f_T and f_{MAX} to the variation in the pinched base sheet resistance (Fig .6) suggesting a large process window relative to this important parameter. Vbe variation for a large number of devices in the production process also indicates a large manufacturing window. Ring oscillator minimum delay distribution measured on 28 sites on the wafer is also shown (Fig. 7). The reliability and safe operating area of device operation is driven by both the rise in junction temperature and the current that the metal system can support. Junction temperature increase in these devices can be viewed quite favorably when compared with devices with similar f_T in other material systems (Fig. 8). As the transistor speed increases due to collector scaling, the peak f_T current density in the device also increases. Scaling the emitter width in proportion to the increased current density in the NPN device necessitate longer device to support the current and this provides relief with metal electromigration. This also has the beneficial effect of reducing the thermal resistance. A safe operating area encompassing the limits set by electromigration and self-heating has been determined and is shown in Fig. 9.

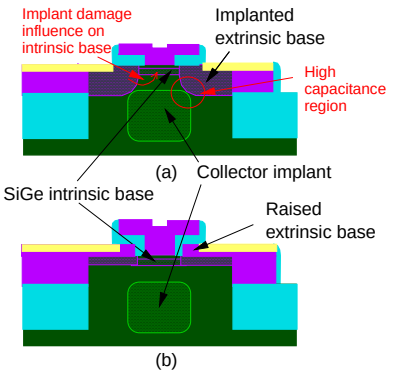

Fig. 1. (a) implanted extrinsic base device structure and (b) raised extrinsic base sructure.

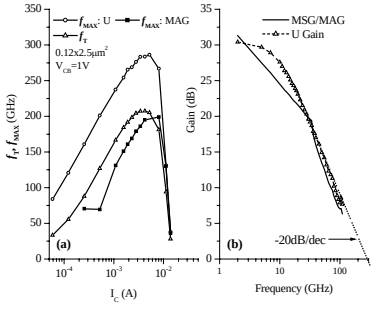

Fig. 2. f_T and f_{MAX} vs. collector current and Mason's unilateral gain (U) and maximum stable and maximum available gain (MSG/MAG) versus frequency at peak bias [2].

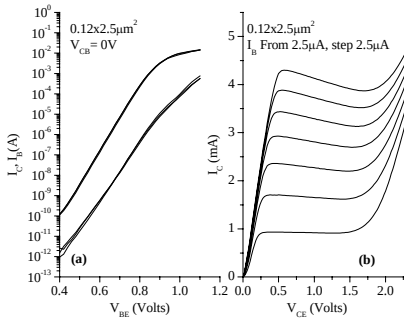

Fig. 3. Gummel and output Characteristics for 0.12μmx2.5μm 207GHz f_T device [2].

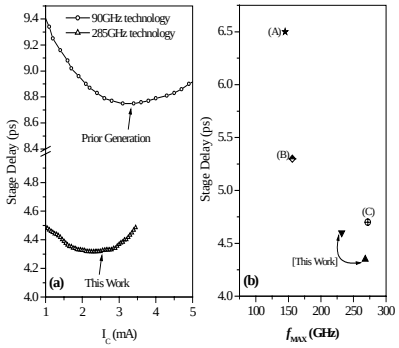

Fig. 4. (a) ECL stage delay comparison for ring oscillators fabricated in 90GHz and 285GHz f_{MAX} SiGe technologies. (b) Bipolar ring oscillator stage delay as a function of f_{MAX} (unilateral gain) from this work and recently reported data in literature [2].

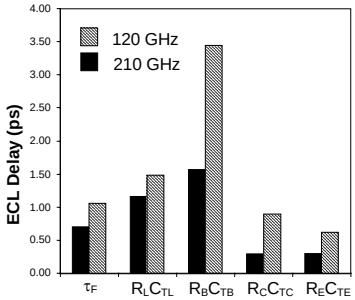

Fig. 5. Estimated ECL delay components for the 120 GHz and 210 GHz technologies. Leftmost bars are the portion from transistor transit times τ_F. The remaining four are the R*C delay components, where C_T is the weighted capacitance charged through the associated resistance.

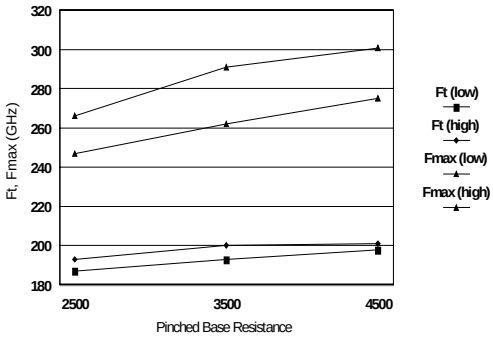

Fig. 6. f_T and f_{MAX} versus pinched-base sheet resistance R_{BI}.

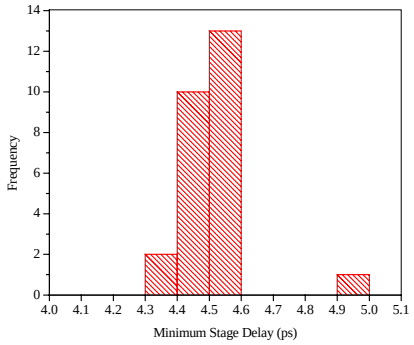

Fig.7 Ring oscillator delay of 26 sites on a wafer.

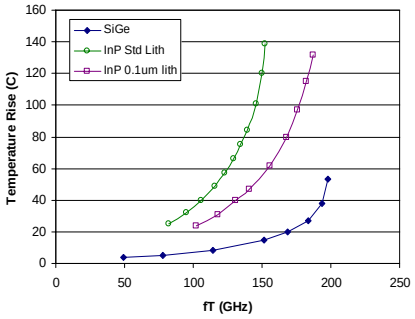

Fig. 8. Comparison of temperature rise for the 0.12x2.5μm SiGe HBT transistor and 1x3μm² InGaAs collector InP HBT transistors presented in [4]. Of the two InP transistors, the higher f_T values result from reduced parasitics from improved lithography overlay and smaller base mesa structure.

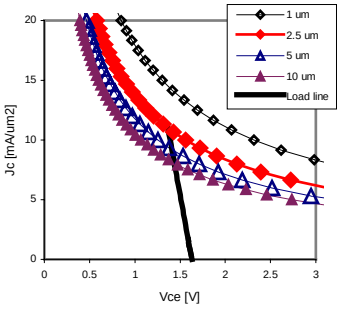

Fig. 9. Bias constraints for the metal system on a simple 1-stripe device layout, taking into account self-heating of the device as a function of bias condition and the dependence of the metal system reliability on this device temperature. All data is for 0.12μm wide emitters, with the length shown in the legend. The load line is for the 0.12x2μm² device in the ring-oscillator circuit.

References

[1] G. Freeman *et al*, IEEE J. Solid State Circuits, Vol.37, No.9, September 2002

[2] B. Jagannathan *et al*, Electron Dev. Lett., Vol.23, 2002, pp. 258-260.

[3] M. Racanelli *et al*, IEDM Tech. Dig., 2001, pp. 336-339.

[4] S. Thomas III *et al*, IEEE Trans. on Device and Materials Reliability, Vol. 1, 2001, pp. 185-189.