On the Performance and Scaling of Symmetric Lateral Bipolar Transistors on SOI

Tak H. Ning, *Life Fellow, IEEE,* and Jin Cai, *Senior Member, IEEE*

***Abstract*—The performance potential and scaling characteristics of thin-base SOI symmetric lateral bipolar transistors were examined using 1-D analytic equations for the currents and capacitances. The device can operate at collector current densities > 100 mA/μm^2, and it scales similarly to CMOS in terms of density. The physical base width is scalable to less than 20 nm. Multiple devices of different specifications can be integrated on a chip. A sample design is shown to have f_T > 200 GHz, f_{max} > 1 THz, V_A >4 V, and a self gain of 60. A balanced design is shown to have 350-GHz f_T and 700-GHz f_{max}, V_A of 2.4 V, and a self gain of 20. These results are superior to those reported for 32 nm SOI CMOS. The results suggest a need to rethink bipolar circuit design. They also suggest opportunities for novel bipolar and BiCMOS circuits. The devices in high-speed Si-base bipolar circuits operate at about 1.0 V. The path toward 0.5 V bipolar circuits is to use semiconductors with smaller bandgap, such as Ge.**

***Index Terms*—Bipolar transistors, complementary bipolar transistors, lateral bipolar transistors, SOI devices.**

I. Introduction

A UNIQUE feature of the recently reported [1] complementary thin-base symmetric lateral bipolar transistors on SOI, illustrated in Figs. 1(a) and 1(b), is the absence of a lightly doped collector. The lateral transistors have no deleterious base push out effect. Unlike vertical bipolar transistors, there is no rapid performance drop off at high current densities. The lateral transistors operate equally fast in forward-active and reverse-active modes. In this paper, the frequency response of this device is examined in detail. Specifically, its f_T, f_{max}, and V_A are calculated for a wide range of device parameters. The transistor self gain, i.e., the product of transconductance g_m and output resistance r_0, are examined. The device design tradeoffs are discussed. The device size and size scaling are discussed and compared to CMOS. Scaling of the physical base width as well as the factors limiting base width scaling are discussed. Opportunities for significant power reduction in designing conventional bipolar circuits in the novel lateral transistors are illustrated. Finally, a path for reducing the operating voltage for bipolar circuits toward 0.5 V is offered.

Manuscript received October 4, 2012; revised November 25, 2012; accepted December 5, 2012. Date of publication January 29, 2013; date of current version February 26, 2013. The review of this paper was arranged by Editor K. Shenai.

The authors are with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY 10598 USA (e-mail: ningth@us.ibm.com; jincai@us.ibm.com).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JEDS.2012.2233272

II. Device Structure and Properties

For modeling purposes, the npn device structure in Fig. 1 is represented by the schematics in Fig. 2, where W_E represents the effective distance between the emitter ohmic contact and the E–B junction, L_E is the emitter stripe length, and W_{E-C} is the separation between the n+ emitter and the n+ collector. The emitter and collector area is $A_E = L_E \times T_{si}$. The parasitic resistances and the relationship between device terminal voltage and junction voltage are as indicated in the figure.

A. One-Dimensional Transport

When a lateral npn transistor is turned on with a voltage V_{BE}, its base current flows vertically down from the base terminal and then turns and flows in the intrinsic base horizontally toward the emitter. The vertical base current flow causes a vertical *IR* drop between the top (p+/p interface) and the bottom (p/BOX interface) of the intrinsic base, causing V'_{BE}(top) to be larger than V'_{BE}(bottom). When this voltage difference is larger than kT/q, there is appreciable current crowding, with the local current density appreciably larger near the top than near the bottom. An upper bound for this voltage difference is [2]

$$V'_{BE,top} - V'_{BE,bottom} < R_{Sbi}(0) J_B(V'_{BE}) T_{si}^2/2 \qquad (1)$$

where $R_{Sbi}(0)$ is the intrinsic-base sheet resistivity at $V'_{BE} = 0$, and $J_B(V'_{BE})$ is the operating base current density. Since this is an upper bound, we can assume that current crowding is negligible if the RHS of (1) is less than kT/q.

As an illustration, consider a device with $R_{Sbi}(0) = 5\,\text{k}\Omega/\square$, $T_{si} = 60$ nm (which is typical silicon thickness for PDSOI CMOS), and operating at $J_B = 1$ mA/μm^2. The RHS of (1) gives 9 mV, which is about 0.35 kT/q. That is, the transistor currents can be described by 1-D transport equations. In this paper, 1-D transport is assumed.

B. Current Capability in Low-Injection Operation

When a bipolar transistor is operated at low currents, its current increases with voltage as $\exp(qV'_{BE}/kT)$. As the current increases, at some point the local minority-carrier density becomes larger than the majority-carrier density. Beyond that point, the dependence of current on V_{BE} degrades. For a vertical transistor, this "current degradation" point is determined by the collector, which is the most lightly doped region. For a symmetric lateral transistor, the degradation point is determined by the base, which is the most lightly doped region.

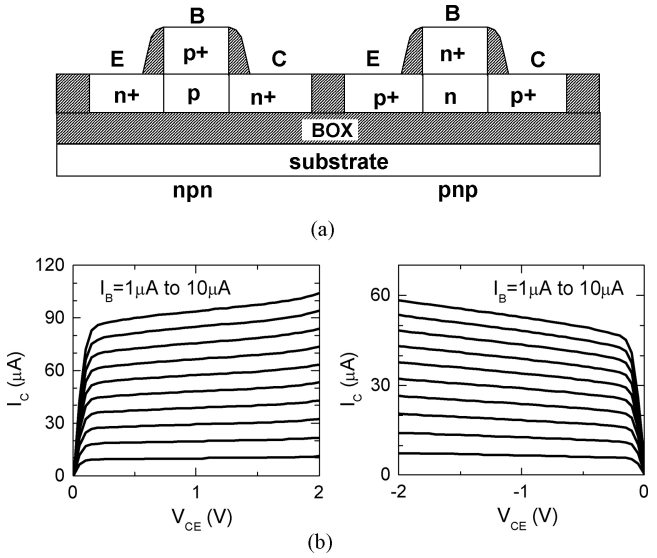


Fig. 1. (a) Schematic illustrating the structure of complementary thin-base symmetric lateral bipolar transistors on SOI (after [1]). (b) Examples of measured I–V characteristics of symmetric lateral bipolar transistors on SOI. $L_E = 1\,\mu$m. Left: npn. Right: pnp (after [1]).

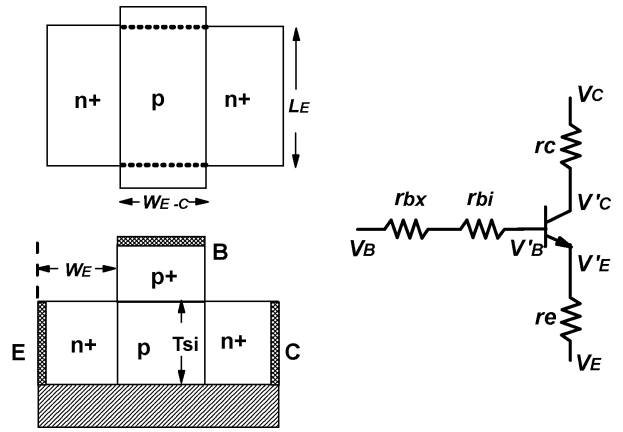


Fig. 2. Schematic of the device structure top view (top left) and cross sectional view (bottom left), and device junction and terminal voltages (right) for modeling purposes.

Below the degradation point, a transistor operates in the so-called low-injection regime where the currents are determined entirely by the device doping profiles.

With E–C symmetry, the low-injection currents are [2]:

$$I_C = I_{C0}(e^{qV'_{BE}/kT} - 1) - (I_{C0} + I_{B0})(e^{qV'_{BC}/kT} - 1), \quad (2)$$

and

$$I_B = I_{B0}(e^{qV'_{BE}/kT} - 1) + I_{B0}(e^{qV'_{BC}/kT} - 1), \quad (3)$$

with

$$I_{C0} = A_E q D_{nB} n_{ieB}^2 / W_B(V'_{BE}, V'_{BC}) N_B, \quad (4)$$

and

$$I_{B0} = A_E q D_{pE} n_{ieE}^2 / W_E N_E, \quad (5)$$

where n_{ieB} and n_{ieE} are the effective intrinsic carrier densities in the base and the emitter, respectively, D_{nB} and D_{pE} are the diffusion coefficients for electrons in the base and holes in the emitter, respectively, N_B and N_E are the doping concentrations in the base and in the emitter, respectively, and W_B is the width of the quasi-neutral base. Equations (2) to (5) are valid for V_{BE} satisfying the low-injection criteria

$$(n_{ieB}^2/N_B)\exp(qV'_{BE}/kT) < N_B. \quad (6)$$

Let us consider a device with $N_B = 2 \times 10^{19}\,\text{cm}^{-3}$ and $W_{E-C} = 20\,\text{nm}$. The maximum voltage satisfying (6) is $V'_{BE} = 1.039\,\text{V}$. The corresponding collector current density J_C is $200\,\text{mA}/\mu\text{m}^2$. In other words, a symmetric lateral transistor can operate at J_C as large as $200\ \text{mA}/\mu\text{m}^2$, yet its currents can still be described by (2) to (5).

The realizable J_C in practice will depend on the emitter series resistance r_e of the device. With the emitter and collector doping concentrations, N_E and N_C, greater than $1 \times 10^{20}\,\text{cm}^{-3}$, r_e is limited by contact resistance. In modern CMOS technology, silicide processes yielding contact resistivity $< 1 \times 10^{-8}\,\Omega\text{-cm}^2$ are common [3]. For a contact resistivity of $1 \times 10^{-8}\,\Omega\text{-cm}^2$, a transistor with $L_E = 1\,\mu$m will have $r_e = 50\ \Omega$ if the silicide contact width is 20 nm.

C. Device Capacitances and Intrinsic-Base Resistance

The width W_B of the quasi-neutral base is modulated by the junction voltages through the relationship

$$W_{E-C} = W_B(V'_{BE}, V'_{BC}) + W_{dBE}(V'_{BE}) + W_{dBC}(V'_{BC}) \quad (7)$$

where W_d is the junction depletion layer width. Voltage-modulation effect is small for a large base width, but not negligible for base widths of nanometer dimension. In this paper, voltage-modulation effects were accounted for in a self-consistent manner in all the calculated results.

As long as N_B is small compared to N_E and N_C, we have

$$W_{dBE}(V'_{BE}) = \sqrt{2\varepsilon_{si}(\psi_{bi} - V'_{BE})/qN_B}, \quad (8)$$

and

$$W_{dBC}(V'_{BC}) = \sqrt{2\varepsilon_{si}(\psi_{bi} - V'_{BC})/qN_B}, \quad (9)$$

where ψ_{bi} is the diode built-in potential. With both the emitter and collector degenerately doped, the Fermi levels in the emitter and collector are pinned at their band edges, and the built-in potential is given by

$$q\psi_{bi} = E_g/2 + kT\ln(N_B/n_i), \quad (10)$$

where E_g is the bandgap energy and n_i is the intrinsic carrier density.

The B−E diode depletion layer capacitance is

$$C_{dBE,tot}(V'_{BE}) = A_E\varepsilon_{si}/W_{dBE}(V'_{BE}), \quad (11)$$

and the B–C diode depletion layer capacitance is

$$C_{dBC,tot}(V'_{BC}) = A_E\varepsilon_{si}/W_{dBC}(V'_{BC}). \quad (12)$$

The B–E capacitance C_{BE} and the B–C capacitance C_{BC} are

$$C_{BE}(V'_{BE}) = C_{dBE,tot}(V'_{BE}) + L_E C_{BE,fringe} \quad (13)$$

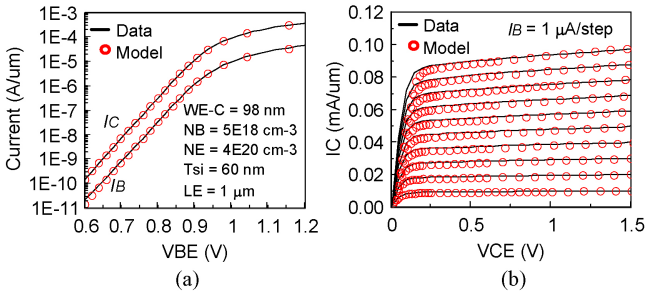


Fig. 3. Example of fitting to measured (a) Gummel plots at $V_{BC} = 0\,\mathrm{V}$ and (b) output characteristics, using $r_{bx} = 400\,\Omega$, $r_e = r_c = 450\,\Omega$, and $W_E = 45.5\,\mathrm{nm}$.

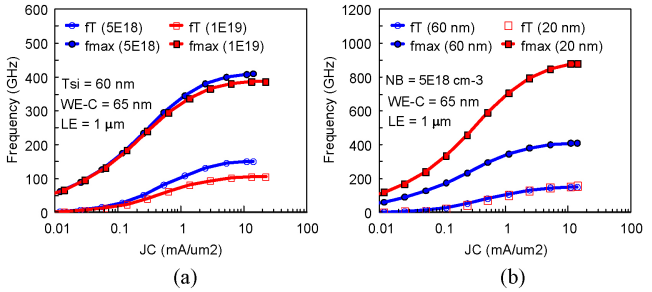


Fig. 4. (a) Calculated f_T and $f_{\max}$ for $T_{si} = 60\,\mathrm{nm}$ and two N_B values. (b) Calculated f_T and $f_{\max}$ for $N_B = 5\times10^{18}\mathrm{cm}^{-3}$ and two values of T_{si}. $V_{BC} = -1\,\mathrm{V}$, and the resistances assumed are $r_{bx} = 20\,\Omega$ and $r_e = r_c = 100\,\Omega$.

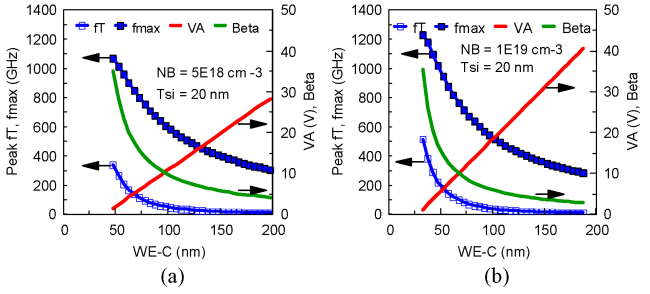


Fig. 5. Calculated dependence on W_{E-C} for (a) $N_B = 5\times10^{18}\mathrm{cm}^{-3}$ and (b) $N_B = 1\times10^{19}\mathrm{cm}^{-3}$, both at $V_{BC} = -1\,\mathrm{V}$, using the same resistances as in Fig. 4. The peak frequencies correspond to the largest V_{BE} satisfying (6).

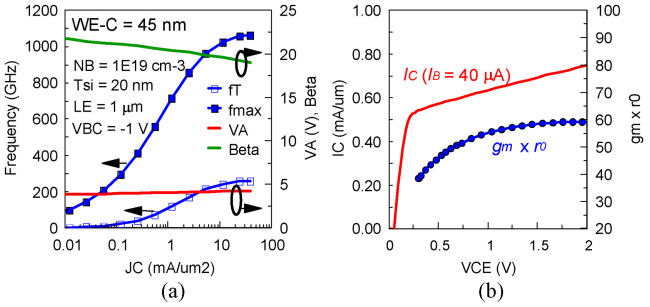


Fig. 6. (a) Calculated frequencies, early voltage, and current gain for a device with $W_{E-C} = 45\,\mathrm{nm}$, $N_B = 1\times10^{19}\mathrm{cm}^{-3}$, and $T_{si} = 20\,\mathrm{nm}$. (b) Calculated output current for the same device at $I_B = 40\,\mu\mathrm{A}$ and the corresponding self gain $g_{\mathrm{m}}\times r_0$.

and

$$C_{BC}(V'_{BC}) = C_{dBC,tot}(V'_{BC}) + L_E C_{BC,fringe}, \tag{14}$$

where $C_{BE,fringe}$ and $C_{BC,fringe}$ are the fringing capacitances per unit emitter stripe length. For CMOS, the fringing capacitance is typically 0.08 fF/μm. The same value is assumed here.

The four delay times or transit times representing the diffusion capacitance caused by stored minority charge are [2]:

$$\tau_E = I_B(V'_{BE}, V'_{BC})W_E^2/3I_C(V'_{BE}, V'_{BC})D_{pE} \tag{15}$$

for the emitter delay time in a shallow-emitter transistor,

$$\tau_B = W_B^2(V'_{BE}, V'_{BC})/3D_{nB} \tag{16}$$

for the base delay time,

$$\tau_{BE} = W_{dBE}(V'_{BE})/2v_{sat} \tag{17}$$

for the B–E space-charge-region delay time, and

$$\tau_{BC} = W_{dBC}(V'_{BC})/2v_{sat} \tag{18}$$

for the B–C space-charge-region delay time, where v_{sat} is the electron saturation velocity. The corresponding emitter diffusion capacitance is

$$C_{DE}(V'_{BE}, V'_{BC}) = \tau_F q I_C/kT, \tag{19}$$

where $\tau_F = (\tau_E + \tau_B + \tau_{BE} + \tau_{BC})$ is the forward transit time. If the B–C diode is also forward biased, there is a corresponding collector diffusion capacitance C_{DC}.

The intrinsic-base resistance r_{bi} is

$$r_{bi} = (T_{si}/3L_E)\rho_B/W_B(V'_{BE}, V'_{BC}), \tag{20}$$

where ρ_B is the base resistivity. The total base resistance is

$$r_b = r_{bx} + r_{bi}, \tag{21}$$

where r_{bx} is the extrinsic-base resistance.

III. Modeling the Device Currents

Equations (2)–(5) were used to model the measured device currents. The unified bandgap narrowing model [4] is used to calculate n_{ieB} and n_{ieE}. The parameters W_E, W_{E-C} (if not known), and the resistances r_e, r_c and r_{bx} can be determined from fitting the Gummel plots. Fig. 3 shows good agreement between modeled and measured currents for a device with $W_{E-C} = 98\,\mathrm{nm}$. The large r_e, r_c, and r_{bx} in the data are due to the non-optimal device layout and silicide process used in fabricating the devices.

IV. Projected Device Performance Results

The transistor performance was evaluated by using the current equations (2) to (5), the capacitance equations (11) to (14), the transit time equations (15) to (18), and the base resistance equations (20) and (21) to calculate the frequencies f_T and $f_{\max}$, and the Early voltage V_A from the commonly used expressions [2]:

$$(2\pi f_T)^{-1} = \tau_F + kT(C_{BE} + C_{BC})/qI_C + C_{BC}(r_e + r_c), \tag{22}$$

$$f_{\max} = \sqrt{f_T/8\pi r_b C_{BC}}, \tag{23}$$

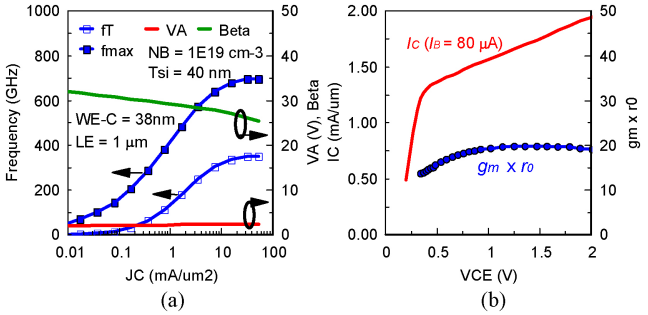


Fig. 7. (a) Calculated frequencies, early voltage, and current gain for a device with W_{E-C} = 38 nm and T_{si}= 40 nm. The target was a device with peak f_T of 350 GHz and peak $f_{\max}$ of 700 GHz. (b) Calculated output current for the same device at I_B = 80 μA and the corresponding self gain $g_m \times r_0$.

and

$$V_A + V_{CE} = qN_B W_B(V'_{BE}, V'_{BC})/C_{dBC}(V'_{BC}). \quad (24)$$

We assumed $r_e = r_c = 100\,\Omega$-μm, which is consistent with modern CMOS technology. The assumed $r_{bx} = 20\,\Omega$-μm is reasonable for metal contact to extrinsic base located directly above the intrinsic base. The assumed T_{si} of 20–60 nm is consistent with advanced SOI CMOS technology.

Fig. 4(a) shows that for a given W_{E-C}, f_T and $f_{\max}$ decrease as N_B is increased, due to an increase in W_B and junction capacitance. Fig. 4(b) shows that reducing T_{si} increases $f_{\max}$ due to r_{bi} reduction with T_{si}. Fig. 5 shows the peak f_T, peak $f_{\max}$, V_A and current gain as a function of W_{E-C} for two N_B values. It shows a strong tradeoff between peak-frequency performance and V_A. The results for a device with W_{E-C} = 45 nm, $N_B = 1 \times 10^{19}$ cm^{-3} and T_{si} = 20 nm are shown in Fig. 6. It shows a device with peak $f_T > 200$ GHz, peak $f_{\max} > 1$ THz, $V_A > 4$ V, and current gain > 19.

V. Example of Design to Specification

The design in Fig. 6 has peak $f_{\max} > 1$ THz, but the peak f_T is only 260 GHz. For applications needing a better balance between f_T and $f_{\max}$, W_{E-C} can be reduced to increase f_T and T_{si} can be increased to reduce $f_{\max}$. The reduction in V_A and self gain can be minimized by increasing N_B, if desired. Fig. 7 shows that f_T of 350 GHz and $f_{\max}$ of 700 GHz can be achieved in this approach. The values shown in Fig. 7 are significantly higher than those reported for 32 nm SOI CMOS [5].

VI. Density and Scaling

The layout of lateral bipolar is similar to that of CMOS. One difference could be in the placement of metal contact to the extrinsic base. In CMOS, the metal contact to the gate is located away from the inversion channel region. In bipolar, if r_{bx} is a concern, metal contact to the extrinsic base should be located over the intrinsic base, not away from the intrinsic base.

A. Density Comparison with CMOS

For a targeted device current, a bipolar device takes less area than a CMOS device. As an example, consider a device carrying 1 mA. A bipolar transistor with T_{si} = 60 nm operating at J_C of 100 mA/μm^2 needs an emitter length (L_E in Fig. 2) of only 170 nm. A typical advanced CMOS device would need a channel width of 500–1000 nm.

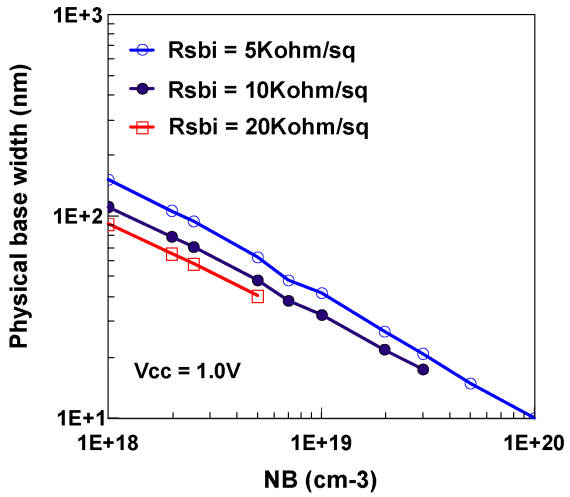


Fig. 8. Physical base width W_{E-C} for a lateral bipolar transistor. W_{E-C} is calculated at $V_{BE} = V_{BC} = 0$. The R_{Sbi} values are for zero bias across B–E and B–C diodes.

The silicide area on the emitter and collector scale similarly to those for the source and drain as long as r_e remains acceptable. If r_e becomes too large, the width of the silicide region should be increased to lower r_e. If the scaled "spacer" on the extrinsic base becomes too thin and leads to unacceptably large base current, "raised source/drain" techniques common used in scaled CMOS may be used. The penalty is increased B–E and B–C fringing capacitances.

B. Physical Base Width Scaling

The physical width W_{E-C} is scalable with increased N_B. The collector current is proportional to the intrinsic-base sheet resistivity R_{Sbi}. If current gain is adequate, designers typically aim for lower R_{Sbi} for smaller r_b and higher performance. R_{Sbi} of about 5 k$\Omega/\square$ is typically used in high-performance designs and R_{Sbi} of 10–20 $\Omega/\square$ in designs where higher current gain or higher breakdown voltage is desired. R_{Sbi} is related to W_B by

$$R_{Sbi}(V'_{BE}, V'_{BC}) = \rho_B / W_B(V'_{BE}, V'_{BC}). \quad (25)$$

Fig. 8 is a plot of W_{E-C} as a function of N_B, showing W_{E-C} scalable to less than 20 nm at N_B in the mid 10^{19} cm^{-3} range.

C. Factors Limiting Physical Base Width Scaling

Note that Fig. 8 shows no design points at $N_B > 5\times10^{18}$ cm^{-3} for a design with $R_{Sbi}(0) = 20\,\text{k}\Omega/\square$. This is due to the fact that the corresponding W_B in (25) becomes fully depleted when the B–C diode is reverse biased in standby mode. In standby mode, $V_{BE} = 0$ V and $V_{BC} = -V_{cc}$, and W_{dBC} is widened by the reverse bias. In a vertical transistor, $N_C < N_B$ and W_{dBC} widens into the collector. However, in a symmetric lateral transistor, $N_C > N_B$ and W_{dBC} widens into the base. If W_B is not wide enough to absorb this depletion

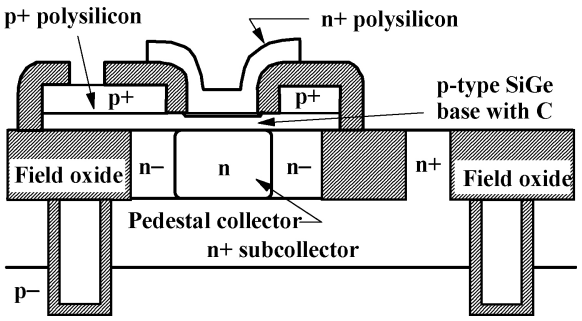


Fig. 9. Schematic cross sectional view of a modern vertical bipolar transistor. When base push out occurs and/or when the transistor is in saturation, minority charge is stored primarily in the n and n⁻ collector regions.

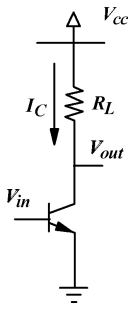


Fig. 10. Circuit schematic of a bipolar inverter with resistor load.

layer widening, the transistor suffers E–C punch through and the design is not acceptable. Similarly, as indicated in Fig. 8, there is no acceptable design at $N_B > 3\times10^{19}\,\mathrm{cm}^{-3}$ for a device with $R_{Sbi}(0) = 10\,\mathrm{k\Omega}/\square$. For a device with $R_{Sbi}(0) = 5\,\mathrm{k\Omega}/\square$, W_{E-C} can be scaled to 10 nm at N_B about $1\times10^{20}\,\mathrm{cm}^{-3}$. For such large N_B, N_E and N_C should be increased toward $1\times10^{21}\,\mathrm{cm}^{-3}$. Band-to-band tunneling in such heavily doped diodes will limit base width scaling.

VII. Rethink Bipolar and its Opportunities

Fig. 9 is a schematic cross section of a typical modern vertical transistor. During operation, minority charge is stored in the emitter, the base, and the collector. The collector stores the most charge by far because of its light doping and large physical volume. In saturation, when the B–C diode is forward biased, or when operated in reverse-active mode with the B–C diode performing the switching, the large amount of minority charge in the collector makes a vertical transistor very slow. A golden rule in circuit design using vertical bipolar has been not to let a transistor go into saturation during operation. Vertical bipolar circuits that operate in reverse-active mode, such as I^2L (Integrated Injection Logic), have speeds limited to somewhat less than 1 ns [6].

The unique characteristics of symmetric lateral bipolar transistors, with no base push out and equal speed in forward-active and reverse-active modes, suggest a need to rethink bipolar circuits and circuit opportunities offered by the technology.

A. Conventional Resistor-Load Bipolar Circuits

Many bipolar circuits, such as ECL (emitter-coupled logic) and CML (current-mode logic), employ resistors as loads. The voltage drop across a load tends to drive the transistor connected to the load into saturation. Designers use larger V_{cc} to avoid saturation, resulting in larger power dissipation.

As an illustration, consider the inverter shown in Fig. 10. We assume both V_{in} and V_{out} have the same swing ΔV. If saturation is not an issue, V_{cc} can be ΔV. However, if forward biasing of the B–C diode is to be avoided completely, the minimum V_{cc} is $2\Delta V$. This example suggests that the voltage across an inverter element in a bipolar circuit can be reduced by up to 50% if the circuit is designed using symmetric lateral transistors.

B. Power Reduction from Transistor Size Consideration

It is apparent from comparing the vertical transistor in Fig. 9 with the lateral npn transistor in Fig. 1 that the CMOS-like layout leads to large reduction in device area. For the same L_E, a lateral transistor is probably about 3x smaller. Furthermore, as discussed in Section II-B, a lateral transistor can operate at $J_C > 100\,\mathrm{mA}/\mu\mathrm{m}^2$ without performance degradation. Vertical transistor typically has performance peaking at $< 40\,\mathrm{mA}/\mu\mathrm{m}^2$ [7], [8]. Therefore, for a given emitter-stripe width (T_{si} in lateral transistor), a lateral transistor can have L_E that is 2.5x shorter. For a given device current, a lateral transistor could be more than 7x smaller in size compared to a vertical transistor.

A smaller transistor size means shorter wires and smaller device and wire capacitances. Since $C\Delta V/I$ is a measure of speed, the current needed to achieve a target speed is reduced as C is reduced. This current reduction translates directly into reduction in power for a circuit designed in lateral transistors.

C. Re-evaluation of Integrated Injection Logic

From the mid to the late 1970's, there was a lot of excitement in the VLSI industry about the prospect of integrated injection logic. I^2L is by far the densest circuit. It uses small size devices, and requires one pnp per gate for current injection and only one npn per fan-out. Thus an inverter with FO = 3 takes only four transistors. The npn transistors in an I^2L circuit operate in the reverse-active mode. As a result, even with advanced self-aligned vertical Si-base bipolar technology, I^2L has minimum delays not much below 1 ns [6]. This speed limitation, together with the rapid progress in CMOS scaling in the early 1980's, caused the demise of I^2L.

Since the symmetric lateral bipolar transistors operate equally fast in forward-active and reverse-active modes, there should be no speed degradation due to reverse-active-mode operation. The performance of I^2L in lateral bipolar was simulated using the terminal-oriented model [9], for a FO = 3 circuit. The npn has T_{si} = 25 nm, L_E = 50 nm and W_{E-C} = 32 nm. During operation, the maximum pnp current is $I_{Cpnp,\max} = I_{C0pnp}\exp(qV_{cc}/kT)$. The corresponding pnp base current is $I_{Bpnp,\max} = I_{B0pnp}\exp(qV_{cc}/kT)$. In standby, the power dissipation depends on the input. For a chain of identical circuits, with one stage driving the next, and using npn and pnp having the same base current characteristics, an upper bound for the average standby power per circuit is $V_{cc}[I_{Cpnp,max}+2(1+\mathrm{FO})I_{Bpnp,max}]/2$. The power supply voltage V_{cc} is simply V_{BE} of the pnp injector. The simulated delays

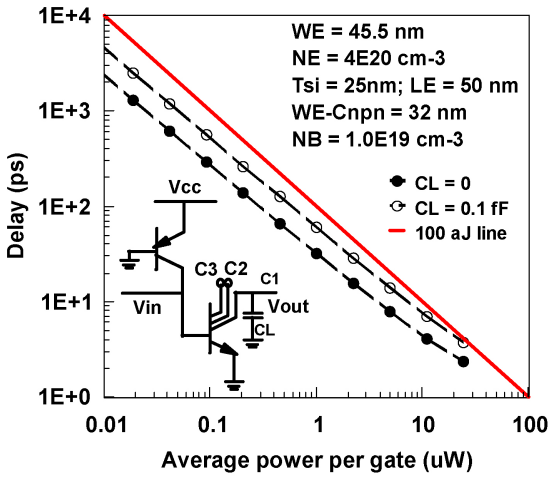


Fig. 11. Simulated delay versus upper bound estimate of average standby power dissipation for an I^2L gate with FO = 3. The npn transistors have W_{E-C} = 32 nm, T_{si}= 25 nm, and L_E = 50 nm.

are shown in Fig. 11. The results are consistent with those obtained by simple area scaling from the data obtained using 2.5 μm vertical bipolar [6]. The minimum delay reported for vertical bipolar was 0.8 ns. The minimum delay in Fig. 11 is 2.4 ps. Fig. 11 indicates a power-delay product $<$ 100 aJ. The reasons for the small power-delay product are the very small device capacitance and the small V_{cc} values.

The power-delay product in Fig. 11 is well within the same range as 2-way NAND gates built with speculative "post-CMOS" devices [10]. It is therefore reasonable to include I^2L in lateral bipolar in any future evaluation of "post-CMOS" logic devices. This should be done at the circuit macro and chip levels.

D. Complementary Bipolar Circuits

The concept of CMOS-like complementary bipolar (CBipolar) circuits has been around for a long time [11]. The basic building block is a CBipolar inverter shown in Fig. 12. It operates with a voltage V_{cc} equal to the V_{BE} for one transistor. In operation, the transistors go into full saturation with B–C forward biased at V_{cc}. As explained earlier, vertical bipolar devices are simply not suitable for operation in saturation, and hence there has been little development of CBipolar circuits.

The lateral devices shown in Fig. 1 are ideal for CBipolar because there is no penalty operating in saturation. The performance of a CBipolar circuits remains to be evaluated. Nonetheless, the density discussion in Section VI-A suggests that CBipolar could have density advantage over CMOS.

E. Novel BiCMOS Circuits

The fabrication process flow for lateral bipolar is similar to that for CMOS [1], suggesting the possibility of integrating lateral pnp and npn with SOI CMOS. This in turn suggests the possibility of novel circuits consisting of one or more CMOS devices and one or more CBipolar devices. If needed, two silicon thicknesses may be used, one for CMOS and one for lateral bipolar. One such circuit is the SRAM cell illustrated in Fig. 13, consisting of two cross-coupled CBipolar inverters

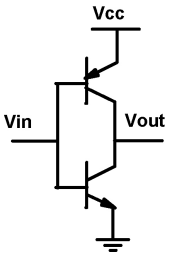


Fig. 12. Circuit schematic of a complementary bipolar inverter.

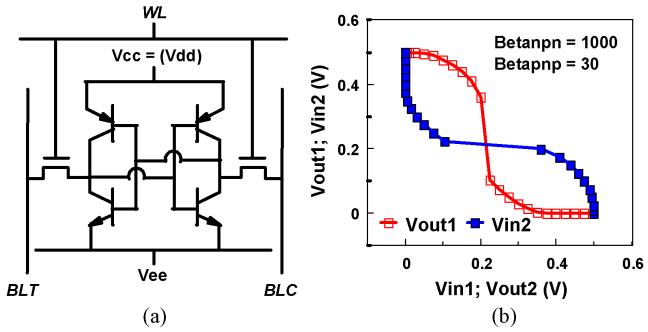


Fig. 13. (a) Proposed SRAM cell using cross-coupled complementary bipolar inverters as a memory element for low voltage applications. (b) Modeled butterfly curves for the bipolar inverters, indicating high noise margin even when there is large mismatch in current gain (after [1]).

as the memory element and two nFETs as access devices. The exponential voltage dependence of the bipolar currents leads to large noise margins in the butterfly curves. The example shows that even for a large imbalance (more than 30x) in the npn and pnp current gains, the noise margin remains large. This example is just one indication of the opportunities for circuit innovation when both complementary bipolar devices and CMOS devices are available on the same chip.

VIII. Path to Lower-Voltage Operation

Bipolar transistors are typically designed to operate at their maximum useful current densities. This approach gives the densest circuit and smallest device and wire capacitances. For a Si-base transistor to operate at J_C of 100 mA/μm^2, V'_{BE} is 1.0 V, suggesting a minimum V_{cc} value of 1 V. The exact V_{cc} value depends on the circuit and on the required design margin.

The quantity $(n_{ieB})^2$ in (4) and (6) is proportional to $\exp(-E_{gB}/kT)$, where E_{gB} is the bandgap of the base region. Therefore, the collector current has the form

$$I_C \propto \exp[(qV'_{BE} - E_{gB})/kT]. \tag{26}$$

As long as $(qV'_{BE} - E_{gB})$ is the same, the collector current density remains about the same when a transistor is built using a different semiconductor having a different bandgap.

As an example, consider using GeOI instead of SOI. The bandgap is 1.12 eV for Si and 0.66 eV for Ge. For a given device design, a Ge-base transistor operates at a voltage 0.46 V lower than a Si-base transistor. Therefore, if a circuit using Si-base transistors operates at V_{cc} of 1 V, the same circuit

performance can be realized at V_{cc} of about 0.54 V using Ge-base transistors, leading to almost 4x reduction in power.

IX. Discussion and Conclusion

The thin-base symmetric lateral bipolar transistor on SOI represents a major breakthrough in bipolar technology. It is an ideal bipolar device structure, has no base push out effect, and operates equally fast in forward-active and reverse-active modes. It is similar to CMOS in layout and scaling.

Compared with CMOS, "gate patterning" for lateral bipolar is probably less challenging because W_{E-C}, being the sum of W_B and the two adjacent depletion layers, is relatively large. It would be challenging to obtain a device with both peak $f_{\max}$ near THz and $V_A > 10$ V. However, the CMOS-like process suggests that devices having various W_{E-C} and N_B, and hence various characteristics, can be made available on the same chip, e.g. from a device with $f_{\max}$ of about 1 THz and V_A of about 5 V to a device with $f_{\max}$ of 300 GHz and V_A of 40 V. Such on-chip device variety makes symmetric lateral bipolar on SOI a versatile technology for THz applications.

Compared to vertical bipolar, lateral bipolar offers large device area reduction and enables bipolar circuits to be operated at lower voltages, up to 2x lower in some cases. The combination of voltage reduction and area reduction suggests a need to rethink and re-evaluate bipolar circuit design. With npn and pnp being available with CMOS on the same chip, there should be exciting opportunities for innovation in bipolar circuits and BiCMOS circuits. With Si as the semiconductor, a lateral bipolar transistor operates at about 1.0 V. With Ge as the semiconductor, the operating voltage can be reduced to about 0.5 V.

References

[1] J. Cai, T. H. Ning, C. D'Emic, K. K. Chan, W. E. Haensch, J.-B. Yau, and D.-G. Park, "Complementary thin-base symmetric lateral bipolar transistor on SOI," in *Proc. IEDM*, 2011, pp. 386–389.

[2] Y. Taur and T. H. Ning, *Fundamentals of Modern VLSI Devices*, 2nd ed. Cambridge, MA: Cambridge Univ. Press, 2009.

[3] Z. Zhang, F. Pagette, C. D'Emic, B. Yang, C. Lavoie, Y. Zhu, M. Hopstaken, S. Mauer, C. Murray, M. Guillorn, D. Klaus, J. Bucchignano, J. Bruley, J. Ott, A. Pyzyna, J. Newbury, W. Song, V. Chhabra, G. Zuo, K.-L. Lee, A. Ozcan, J. Silverman, Q. Ouyang, D.-G. Park, W. Haensch, and P. M. Solomon, "Sharp reduction of contact resistivities by effective Schottky barrier lowering with silicides as diffusion sources," *IEEE Electron Device Lett.*, vol. 31, no. 7, pp. 731–733, Jul. 2010.

[4] D. B. M. Klaassen, J. W. Slotboom, and H. C. de Graaff, "Unified apparent bandgap narrowing in n- and p-type silicon," *Solid-State Electron*, vol. 35, no. 2, pp. 125–129, 1992.

[5] S. Lee, J. Johnson, B. Greene, A. Chou, K. Zhao, M. Chowdhury, J. Sim, A. Kumar, D. Kim, A. Sutton, S. Ku, Y. Liang, Y. Wang, D. Slisher, K. Duncan, P. Hyde, R. Thoma, J. Deng, Y. Deng, R. Rupani, R. Williams, L. Wagner, C. Wermer, H. Li, B. Johnson, D. Daley, J. O. Plouchart, S. Narasimha, C. Putnam, E. Maciejewski, W. Henson, and S. Springer, "Advanced modeling and optimization of high performance 32 nm HKMG SOI CMOS for RF/analog SoC applications," in *Proc. Symp. VLSI Tech.*, 2012, p. 135.

[6] D. D. Tang, T. H. Ning, S. K. Wiedmann, R. D. Isaac, G. C. Feth, and H. N. Yu, "Subnanosecond self-aligned I^2L/MTL circuits," in *Proc. IEDM*, 1979, pp. 201–204.

[7] R. Krithivassan, Y. Lu, J. D. Cressler, J.-S. Rieh, M. H. Khater, D. Ahlgren, and G. Freeman, "Half-terahertz operation of SiGe HBTs," *IEEE Electron Device Lett.*, vol. 27, no. 7, pp. 567–569, Jul. 2006.

[8] S. Böck, H. Schäfer, H. Knapp, K. Aufinger, M. Wurzer, S. Boguth, T. Böttner, R. Stengl, W. Perndl, and T. F. Meister, "3.3 ps SiGe bipolar technology," in *Proc. IEDM*, 2004, pp. 255–258.

[9] H. H. Berger and S. K. Wiedmann, "Terminal-oriented model for merged transistor logic (MTL)," *J. Solid-State Circuits*, vol. 9, no. 5, pp. 211–217, Oct. 1974.

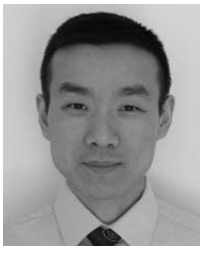
[10] K. Bernstein, R. K. Cavin, III, W. Perod, A. Seabaugh, and J. Welser, "Device and architecture outlook for beyond CMOS switches," *Proc. IEEE*, vol. 98, no. 12, pp. 2169–2184, Dec. 2010.

[11] H. H. Berger and S. K. Wiedmann, "Complementary transistor circuit for carrying out Boolean functions," U.S. Patent 3 956 641, May 1976.

Tak H. Ning (M'75–SM'81–F'87–LF'10) received the Ph.D. degree in physics from the University of Illinois at Urbana-Champaign, Urbana, in 1971.

He joined the IBM Thomas J. Watson Research Center, Yorktown Heights, NY, in 1973. He has made contributions to various areas of silicon devices and technology, including bipolar, CMOS, DRAM, silicon-on-insulator, EEPROM, and hot-electron effects.

Dr. Ning was a recipient of several awards, including the Electrochemical Society 2007 Gordon E. Moore Medal, the 1991 IEEE Jack A. Morton Award, and the 1989 IEEE Electron Devices Society J. J. Ebers Award. He is a member of the National Academy of Engineering and a fellow of the American Physical Society. Since 1991, he has been an IBM Fellow.



Jin Cai (M'01–SM'10) received the Ph.D. degree from the University of Florida, Gainesville, in 2000.

He is currently a Research Staff Member with the IBM Thomas J. Watson Research Center, Yorktown Heights, NY.

Dr. Cai is an Associate Editor for IEEE Electron Device Letters.