Modeling of Ge–Si Heterojunction Bipolar Transistors for Use in Silicon Monolithic Millimeter-Wave Integrated Circuits

STEPHEN A. CAMPBELL, MEMBER, IEEE, AND ANAND GOPINATH, SENIOR MEMBER, IEEE

Abstract—**Previous work on high-resistivity silicon suggests that microstrip line dielectric losses cease to be significant above 30 GHz. Silicon–Germanium heterojunction bipolar transistors now provide a well-behaved three-terminal device capable of operating at microwave frequencies, making the fabrication of silicon monolithic millimeter-wave integrated circuits a genuine possibility. The trade-offs available to operate this device at millimeter-wave frequencies are discussed, and one-dimensional calculations along with two-dimensional simulations of transistor performance are presented.**

I. Introduction

THE CHOICE of GaAs for monolithic microwave integrated circuits has been dictated by its properties: semi-insulating substrates are readily available and, three-terminal amplifying and oscillating devices are currently available to over 100 GHz. But problems in manufacture remain: the quality of the material is variable, the yields are not as high as may be expected, and the circuits are expensive. Silicon as a microwave substrate material is lossy, but it has been shown that for frequencies above 30 GHz, the "dielectric" loss due to the low resistivity (2000 to 10000 $\Omega \cdot$cm) ceases to be a problem [1]. However, apart from Schottky-barrier and p–n junction diodes, the IMPATT diode appeared to be the only active amplifying device available in this material. While IMPATT diodes currently provide the best solid-state high-power sources at the present time, they are exceedingly difficult to match in the monolithic circuit context. Thus, there is considerable reluctance to use these devices in integrated circuits. The advent of the germanium–silicon heterojunction bipolar transistor (Ge–Si HBT) provides a three-terminal device which is well behaved and capable of working at millimeter-wave frequencies. We provide an analysis to show that, indeed, these devices may be the choice for silicon-based millimeter-wave integrated circuits. In the following section, the performance of these Ge–Si HBT's is analyzed, and we show that an f_{max} of well over 100 GHz may be feasible with these devices. Thus, millimeter-wave IC's in silicon are a possible alternative to GaAs- and InP-based circuits. The advantages of silicon in device and circuit processing are obvious and will not be documented here.

Manuscript received March 27, 1989; revised July 12, 1989. This work was supported by the National Science Foundation under Grant ECS-8706913 and by the Graduate School of the University of Minnesota.

The authors are with the Department of Electrical Engineering, University of Minnesota, Minneapolis, MN 55455.

IEEE Log Number 8930957.

II. Device Structure

It has been demonstrated that molecular beam epitaxy may be used to fabricate heterojunctions of Ge_xSi_{1-x} on Si. While not lattice matched, the Ge–Si layer will grow pseudomorphically up to a critical layer thickness, at which gross stress relief will occur via the formation of dislocations [2]. This technique has been used to fabricate HBT's with modest success [3]. More recently, a vapor phase approach to the fabrication has produced exceptional results, with current gains of 400 and ideal Gummel plots remain down to a picoampere of base current [4]. It is of interest, then, to predict the microwave performance of the devices that could be fabricated from such layers. One such estimate has recently appeared in the literature suggesting that devices could be made with an f_{max} of 35 GHz and an f_T of 75 GHz [5]. However, the device modeled used 1 μm lithography and a non-self-aligned structure. In this paper we will apply submicron lithography and an advanced device structure to Ge–Si HBT's and project the device performance. An important caveat to these results is that all of the material properties of the Ge–Si layer except the band gap will be assumed to be those of silicon. We emphasize that these are first-order estimates of transistor performance.

Fig. 1 shows a super self-aligned transistor (SST), a device structure that has been developed for digital applications [6] but can be altered to fabricate analog Ge–Si npn HBT's. A 3000 Å moderately doped (10^{17} cm^{-3}) collector is first grown on a heavily doped buried layer. A thick field oxide is then grown in a local oxidation process and a 1000 Å sub poly oxide is grown. Next 1500 Å of p^{++} polysilicon is deposited and oxidized to a thickness of 1000 Å. The oxide/poly/oxide sandwich is then anistropically patterned. Base contacts 1000 Å wide are now formed through an oxide undercut followed by an undoped poly CVD plug fill and poly oxidation. Reactive ion etching then forms sidewall oxide spacers approximately 1000 Å wide. The finished emitter width is then 0.2 μm less than the patterned or "cut" dimension. Prior to growth a shallow (250 Å) isotropic recess is done to ensure proper base

0018-9480/89/1200-2046$01.00

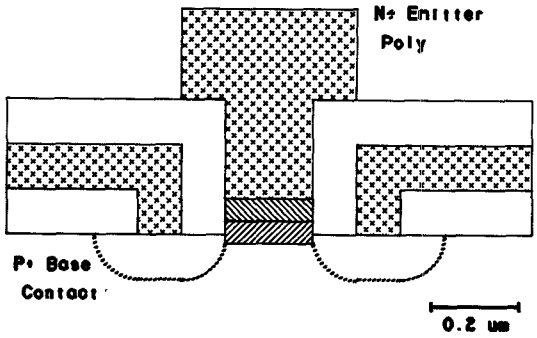


Fig. 1. SST structure to be modeled. Areas crosshatched with ×'s indicate heavily doped polysilicon; areas ruled with 45° lines are the emitter (Si) and base (Ge/Si). The base is contacted through the p^+ diffusion from the polysilicon.

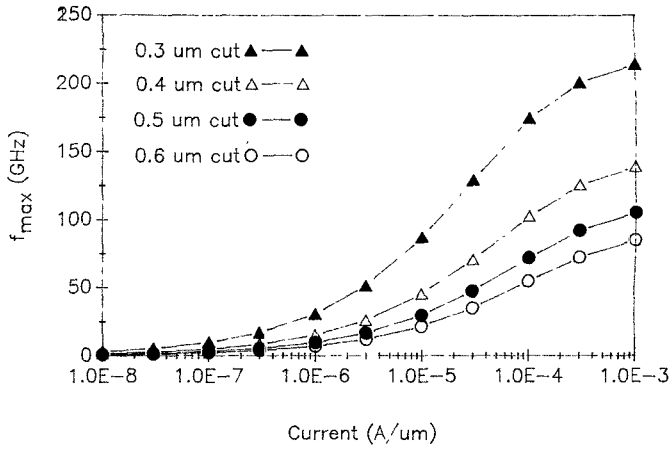


Fig. 2. Calculated maximum frequency of oscillation versus current per unit length of the transistor, with the emitter cut width varied as a parameter. Note that the final emitter width is 0.2 μm less than the emitter cut width.

contact. Using the vapor phase chemistry and the lower growth nucleation rate on the oxide, selective growth is now done to form the base and emitter regions. Finally a thicker n^{++} polysilicon is deposited and patterned to contact the emitter. The exposed poly regions may then be silicided to further reduce the parasitic resistance. The base doping is set to 5×10^{18} cm^{-3}. This represents a realistic estimate of the maximum boron doping achievable at typical Ge–Si growth temperatures. The emitter width is set to 500 Å and its concentration is also set to 5×10^{18} cm^{-3}. This represents a trade-off between dc current gain and EB capacitance. Due to the heterojunction such a device will still have β greater than 100. We assume that the transistor has an interdigitated structure with 20 fingers of 50 μm length.

The critical layer thickness as a function of the Ge mole fraction has been determined by People [7], among others. For the purpose of this calculation we shall select a base width of 500 Å and an alloy concentration of $Ge_{0.25}Si_{0.75}$, although this choice is somewhat arbitrary. The band offset of the heterointerface is a critical function of the stress in the Ge–Si film and in the emitter. The presence of a thin silicon cladding layer has been demonstrated to shift the band offset from −0.065 eV to +0.225 eV. For the purposes of the first set of calculations we shall assume that all of the discontinuity (150 mV) appears at the valence band edge. The effect of this offset is to increase the device current gain by providing an effective barrier for hole diffusion into the emitter, while minimizing the effects of hot electron injection into the base.

III. One-Dimensional Calculations

Assuming $\alpha=1$, we can now calculate f_T, the unity gain frequency, and f_{max}, the frequency at which the unilateral gain becomes unity, from

$$f_T = 1/(2\pi\tau_{EC})$$

and

$$f_{max} = \left(1/\left(16\pi^2 r_B C_C(\tau_{EC} + r_E C_C)\right)\right)^{-1/2}$$

where r_B is the base resistance, r_E is the small-signal emitter resistance, and C_C is the collector capacitance. The device transit time, τ_{EC}, can be found from

$$\tau_{EC} = \tau_E + \tau_B + \tau_C + \tau_C'$$

where τ_E is the emitter–base depletion layer charging time, τ_B is the base charging time, τ_C is the transit time of the base–collector depletion layer, and τ_C' is the RC delay of the collector. For a transistor with a heavily doped buried collector τ_C' is negligible. The value of τ_C can be found by dividing the depletion layer thickness by twice the saturation velocity. The base charging time is given by

$$\tau_B = W^2/\eta D_e$$

where W is the pinched base width and D_e is the diffusivity of the minority carrier electrons in the base. Assuming the values for silicon concentration to be dependent on diffusivity and taking $\eta=2$ for a uniform base concentration, τ_B is typically less than 0.01 ps. The emitter–base depletion layer charging time is given by

$$\tau_E = r_E\left(C_E + C_C + C_{par}\right)$$

where C_E is the emitter–base capacitance, C_C is the base–collector capacitance, and C_{par} is the total parasitic capacitance.

Fig. 2 shows a plot of the calculated f_{max} versus emitter current for the proposed super self-aligned device with several drawn strip widths. Initially, the performance improves sharply with increasing current due to the reduced emitter charging time. At higher currents the curve saturates due to the effects of the $r_E * C_C$ terms in the denominator. If we reduce the drawn stripe width and hold the sidewall oxide constant, a substantial improvement results, with the highest predicted f_{max} values well over 100 GHz. Fig. 3 shows an equivalent plot for f_T, where similar effects are seen. Due to base push-out, however, these predicted maximum values are overly optimistic. We can estimate the onset of the Kirk effect using

$$I_K = qN_C v_s A_E$$

where N_C is the collector doping density, v_s is the saturation velocity, and A_E is the emitter area. We then predict that more realistic maxima for f_{max} and f_T are 65 and 95

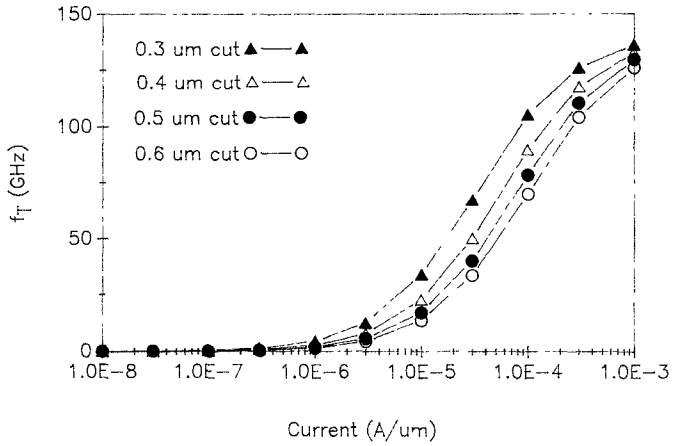


Fig. 3. Calculated unity gain frequency versus current per unit length of the transistor, with the emitter cut width varied as a parameter.

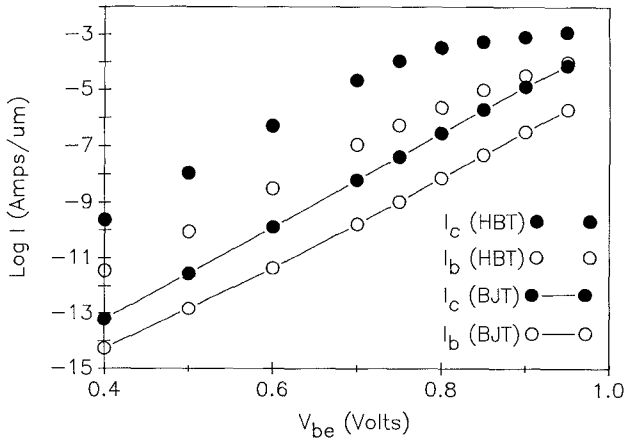


Fig. 4. Results of dc PISCES simulations for both homoepitaxial and heteroepitaxial transistors.

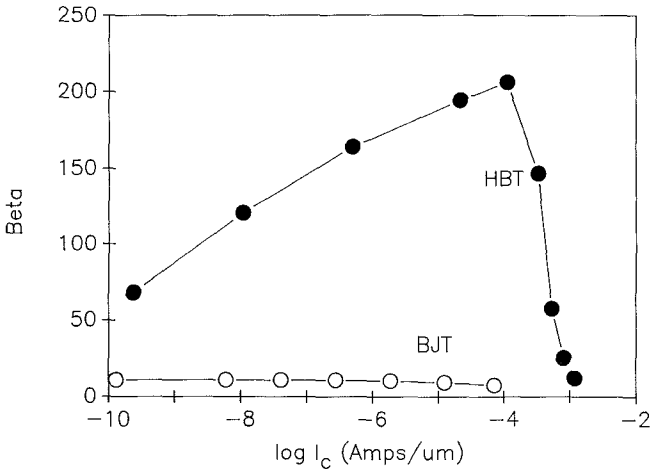


Fig. 5. Current gain of both transistors versus the collector current. The HBT shows an increase in gain with increasing collector current due to the reduced effects of leakage and generation–recombination currents, and falls off sharply at high current due to base push out.

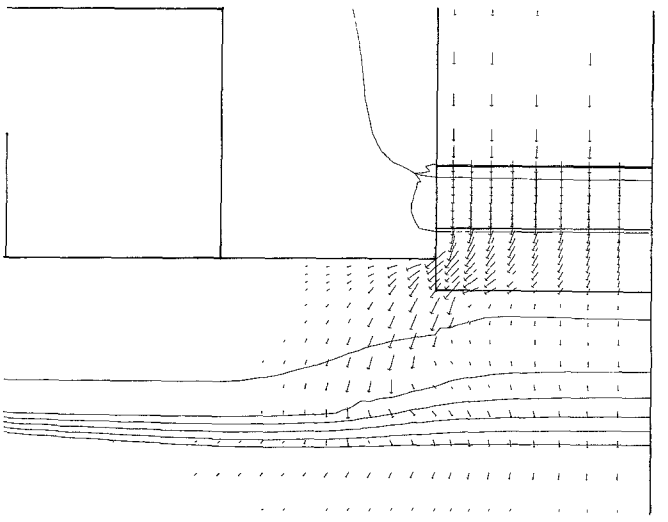


Fig. 6. Two dimensional plot showing the base contact and transistor action at high injection ($V_{ce} = 3.0$ V and $V_{be} = 0.9$ V). The arrows indicate total current flow. The contours are constant voltage surfaces at 0.1 V increments with the bottom contour representing 1.5 V. Notice the large amount of injection into the extrinsic base and the deep extension of the voltage contours into the collector.

GHz at 210 μA/μm for the 0.6 μm drawn stripe, 80 and 95 GHz at 160 μA/μm for the 0.5 μm drawn stripe, 105 and 90 GHz at 105 μA/μm for the 0.4 μm drawn stripe, and 150 and 65 GHz at 50 μm/μA for the 0.3 μm drawn stripe.

IV. Two-Dimensional Simulations

Due to the relatively simple nature of the one-dimensional calculations, a set of 2-D simulations were done using PISCES-IIB [8] to confirm the results. This program solves Poisson's and the continuity equations on a two-dimensional grid for both carrier types and applies drift and diffusion transport. PISCES does not allow the presence of more than one semiconductor. In order to model the heterostructure, the band gap narrowing parameters for p-type dopants were adjusted such that for the dopant concentration used in the base, the band gap was narrowed by an additional 150 mV. The default models for SRH and Auger recombination, and field and concentration dependent mobility were used. The structure described above was modeled with a 3.0 V collector to emitter bias.

The dc results are shown in Figs. 4 and 5 the 0.6 μm emitter cut transistor both with and without the heterojunction. Very similar results were found for the smaller devices. As predicted, current gains well in excess of 100 are obtained for the HBT, even though the effect of using band gap narrowing to simulate the heterostructure is to split the discontinuity evenly between the valence and conduction bands. This can easily be seen by calculating the ratio β(HBT)/β(BJT). The result corresponds to a valence band discontinuity of 75 mV. It is found that β declines sharply for currents above 10^{-4} A/μm. Fig. 6 shows a cross section of the active region of the 0.4 μm device with 3.0 V on the collector and 0.9 V on the base, corresponding to a collector current of 5.4×10^{-4} A/μm. Both current density vectors and equipotential lines at 0.1 V increments are shown. Base push-out is clearly seen below the collector. The base resistance does not appear to be a factor, as V_{be} is nearly uniform across the device.

To predict the ac performance a 1 mV sinusoidal signal was applied to the base and emitter contacts of the dc

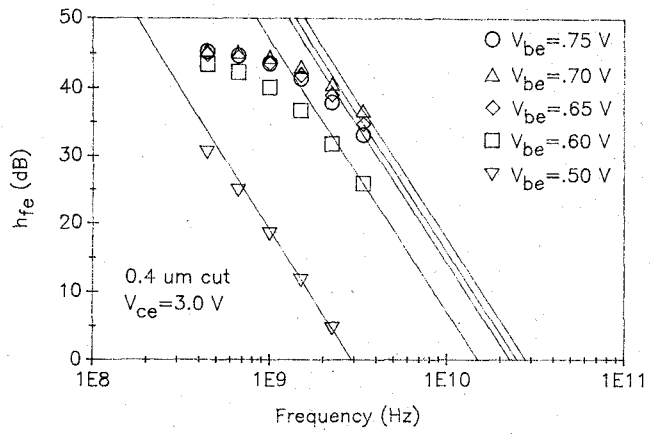


Fig. 7. Sample plot of ac results and subsequent f_T extraction for 0.4 μm emitter cut width transistor.

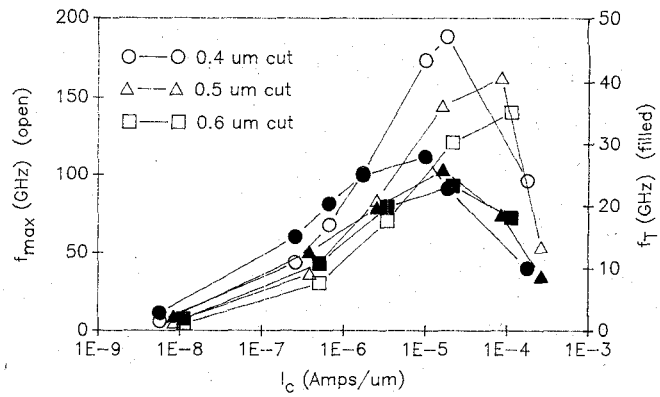


Fig. 8. PISCES predictions for f_{max} (open marks) and f_T (filled marks) as functions of collector current per unit transistor length with emitter cut width as a parameter.

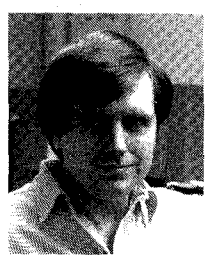
solution and the Y parameters were extracted as a function of frequency. From these Mason's gain (U) and the current gain (h_{fe}) were calculated and used to find f_{max} and f_T. PISCES was unable to converge for ac solutions on any of the structures for frequencies above 5 GHz. Thus a linear extrapolation from the last converged data point (generally 3.8 GHz) was used to calculate f_{max} and f_T. Fig. 7 shows a typical extrapolation plot. Since the data points do not always extend completely into the linear region of the plot, these results are somewhat lower than the true values, particularly for the highest frequency devices. Fig. 8 shows a compilation of the results. Due to the extrapolation problem the 0.3 μm cut HBT was not modeled. It was found that f_T increases with emitter current, as expected, although it did not reach as high a frequency as predicted by the one-dimensional calculation. The highest value for f_T was 28 GHz for the 0.4 μm cut transistor at 1.1×10^{-5} A/μm, and it was nearly independent of transistor width. The value of f_{max} was also found to increase with emitter current, with a maximum value of 170 GHz at 1.1×10^{-5} A/μm for the 0.4 μm cut device.

V. Discussion and Summary

By treating Ge–Si layers as narrow-band-gap silicon, both one-dimensional calculations and the two-dimensional PISCES simulations were done on HBT's with varying emitter widths. The results indicate that high-speed operation can be achieved with Ge–Si HBT's, although the maximum of both f_{max} and f_T occurred at lower current densities than expected from the simple model. We believe the values of f_{max} and f_T derived from the two-dimensional model to be quite conservative estimates, since only half of the heterojunction discontinuity appeared in the valence band. If the band offset were properly taken into account, substantially higher gains would result. For another 75 mV discontinuity the maximum value of f_T should be approximately 120 GHz, substantially in agreement with the one-dimensional calculation. Furthermore the doping concentrations for the devices were fixed at nonoptimal values. Additional simulations indicate that by lowering both the base and emitter concentrations, higher cutoff frequencies can be obtained. Similarly, increasing the collector concentration would allow higher current operation, further improving device performance.

References

[1] A. Rosen *et al.*, "Silicon as a millimeter-wave monolithically integrated substrate-A new look," *RCA* Rev., vol. 42, pp. 633–660, 1981.
[2] J. C. Bean, L. C. Feldman, A. T. Fiory, S. Nakahara, and I. K. Robinson, "Ge_xSi_{1-x}/Si strained layer superlattice grown by molecular-beam epitaxy," *J. Vac. Sci. Technol.*, vol. A2, pp. 436–440, 1984.
[3] S. S. Iyer, G. L. Patton, S. S. Delage, S. Tiwari, and J. M. C. Stork, "Silicon-germanium base heterojunction bipolar transistors by molecular beam epitaxy," in *Proc. 1987 IEDM* (Washington, DC) 1987, pp. 874–875.
[4] C. A. King *et al.*, "Characteristization of P-N $Si_{1-x}Ge_x$/Si heterojunctions grown by limited reaction processing." in *Proc. 1988 DRC/MRC*, 1988, paper VB-7, p. 2454.
[5] T. Won and H. Morkoc, "High-speed performance of $Si/Si_{1-x}Ge_x$ heterojunction bipolar transistors," IEEE Electron Device Lett., vol. 10, pp. 33–35, Jan. 1989.
[6] I. Antipov, "Bipolar transistor with minimized collector-to-base junction," *IEEE Trans. Electron Devices*, vol. ED-30, pp. 723–726, June 1983.
[7] R. People and J. C. Bean, "Band alignments of coherently strained Ge_xSi_{1-x}/Si heterostructures on $\langle 001\rangle$ Ge_ySi_{1-y} substrates," *Appl. Phys. Lett.*, vol. 48, pp. 538–540, 1986.
[8] M. K. Pinto, C. S. Rafferty, and R. W. Dutton, "PISCES-II—Poisson and continuity equation solver," Stanford Electronics Laboratory Tech. Rep., Stanford University, Sept. 1984.



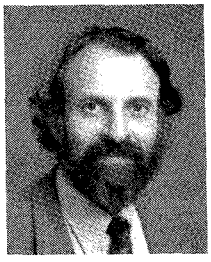
Stephen A. Campbell (M'84) was born in St. Paul, MN, in 1954. He received the B.A. degree from the College of St. Thomas in 1975 and the M.S. and Ph.D. degree in physics from Northwestern University in 1978 and 1981, respectively.

He joined the Sperry Corporation in 1981, where he worked in the field of fabrication processes and device design for digital CMOS and bipolar technologies. In 1986 he joined the Department of Electrical Engineering at the University of Minnesota, Minneapolis, where he is now active in the growth of thin silicon and germanium/silicon epitaxial layers using rapid thermal vapor phase epitaxy.

Dr. Campbell is a member of the Electrochemical Society, the Electron Devices Society of the IEEE, and the American Physical Society. He has been designated a Presidential Young Investigator by the National Science Foundation.

✠

Anand Gopinath (S'64–M'65–SM'80) was awarded the Ph.D. degree and subsequently the D.Eng. (higher doctorate) degree by the University of Sheffield, England, in 1966 and 1978, respectively.



He taught at the University College of North Wales, Bangor, Wales, until 1978, where he became Reader in Electronic Engineering, and subsequently was appointed to the Chair of Electronics at Chelsea College (now King's College), University of London in 1981. He was Visiting Scientist 1978–1979) and then Member of Technical Staff (1979–1981, 1982–1986) at MIT Lincoln Laboratory, Lexington, MA. He joined the Department of Electrical Engineering, University of Minnesota, Minneapolis, as Professor in 1986. His interests are in various aspects of guided wave structures for RF and optical devices, high-frequency, high-speed, and optical semiconductor devices, and monolithic and integrated optical circuits.