RTCVD Growth and Applications of Epitaxial $Si_{1-x}Ge_x$ Alloys

James C. Sturm

Rapid thermal chemical vapor deposition (RTCVD) is a competing technology for the growth of $Si_{1-x}Ge_x$ strained-layer structures for the development of silicon-based heterojunctions. High-quality epitaxial films with growth temperatures as low as 600°C and 4.5 nm-period superlattices have been demonstrated without the use of ultrahigh-vacuum techniques. Narrow-bandgap-base heterojunction bipolar transistors made of $Si/Si_{1-x}Ge_x/Si$ have been shown to have superior characteristics compared to all-silicon homojunction devices, and extend the scaling limits of bipolar transistors. Finally, RTCVD strained $Si_{1-x}Ge_x$ films exhibit well-resolved band-edge photoluminescence—the first step on the way to developing silicon-based light-emitting devices.

INTRODUCTION

Since the beginning of the integrated circuit era some 30 years ago, silicon integrated circuits have expanded to contain over 10,000,000 transistors per chip. While integrated circuits now consist of silicon nitride, silicon dioxide, metal silicides, polyimides, aluminum, and various aluminum alloys, the actual semiconductor crystal comprising the heart of the transistors has remained silicon. Silicon has been chosen not because of its optimal electrical properties, but rather because of its excellent material processing capability. For example, stable insulating oxides, which also serve as diffusion barriers, are easily formed on its surface; crystal damage from ion implantation is easily repaired by annealing; and large-diameter, dislocation-free wafers are readily produced.

It has long been recognized, however, that far superior electrical performance compared to all-silicon "homojunction" devices might be realized if one could fabricate different parts of a transistor out of semiconductors with different bandgaps. A major roadblock in achieving such "heterojunction" transistors has been the problem of growing one compatible semiconductor on top of another without a large number of defects at the interface or in the top film.

For example, while spectacular results have been achieved by using combinations of GaAs and $Al_{1-x}Ga_xAs$ in devices in the past decade, these results have not been transferable to silicon integrated-circuit technology for two reasons. First, these compounds and most other III-V semiconductors are not compatible with silicon processing techniques—at high temperatures, these semiconductors can decompose, and their constituents can dope the silicon. Second, although their basic crystal structures are similar (diamond-like versus zincblende), the lattice constants of GaAs and $Al_{1-x}Ga_xAs$ are approximately four percent larger than that of silicon, making it difficult to grow one on top of the other. Unfortunately, there is no compatible semiconductor "lattice-matched" to silicon.

A breakthrough occurred in the 1970s and early 1980s when it was found that alloys of silicon and germanium ($Si_{1-x}Ge_x$) could be grown, within limits, lattice-matched to silicon substrates.[1-4] Although the films were in biaxial compression since the natural relaxed lattice constant of these alloys is greater than that of silicon,[5] for thin films this is the preferred growth mode. For such thin films, the energy saved by growing commensurate to the substrate to avoid misfit dislocations is larger than the compression energy of the film.

Such films are technologically attractive for several reasons. The $Si_{1-x}Ge_x$ system is much more compatible with silicon processing than GaAs; one reason is that germanium and silicon do not serve as dopants in each other. Since germanium has a bandgap smaller than silicon, the bandgap of the alloys can be adjusted by controlling the germanium fraction (x). Additionally, the strain serves to split band edge degeneracies, further reducing the bandgap so that a $Si_{0.8}Ge_{0.2}$ layer has a bandgap ~160 meV less than that of silicon.[6,7] Therefore, there now exists a silicon-based material system for developing heterojunction transistors and other heterojunction structures on silicon substrates for integrated circuits. References 1–4, 6, and 7 provide further information on the basic electrical and mechanical properties of $Si_{1-x}Ge_x$ strained alloys on silicon.

The initial growth experiments for $Si_{1-x}Ge_x$ alloys were performed by molecular beam epitaxy (MBE). This ultrahigh-vacuum (UHV) evaporation technique is well suited for research applications because of its flexibility and control, but it is not well suited for manufacturing. Chemical vapor deposition (CVD), because of its easily renewable gas sources and ability to scale to larger sizes, is a preferred thin-film growth technology for industrial applications. Two alternative CVD approaches have emerged over the past several years. One involves processing an entire boat of wafers at once in a UHV furnace environment (UHV/CVD).[8] The other generally involves only a single wafer at a time, but the wafer temperature can be varied rapidly by external lamp heating. Such rapid thermal chemical vapor deposition (RTCVD) and its application to devices such as heterojunction bipolar transistors and optical emitters is the focus of this article.

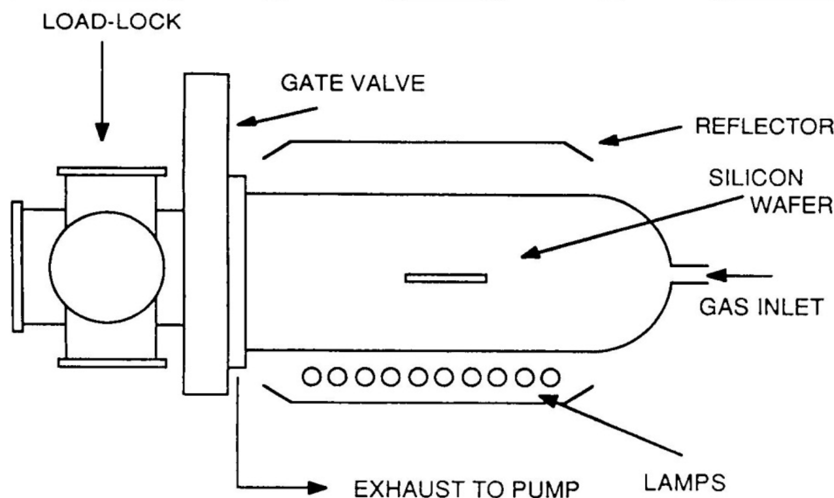

Figure 1. A schematic view of an RTCVD apparatus.

RAPID THERMAL CHEMICAL VAPOR DEPOSITION

The central feature of an RTCVD reactor is a growth chamber which consists of a quartz tube (or steel chamber with quartz windows) so that the wafer may be heated by radiation from a bank of high-intensity lamps outside the reactor (Figure 1). To enable rapid changes in the sample temperature, the wafer is typically supported on quartz pins without a susceptor. Although a horizontal gas flow is shown in Figure 1, vertical or annular flows are also used. Not shown in Figure 1 are the gas manifold and supply system or the pumping apparatus. Typical growth conditions in such a system are a gas flow of 1 vol.% dichlorosilane in a hydrogen carrier at 6 torr. Silicon is typically grown from ~700°C (giving a growth rate of 4 nm/min) to 1,000°C (0.25 μm/min). Germane (GeH_4) is added to provide a germanium source for the alloy growth, and the films are often grown at 600–650°C as described below. The growth rates can still be fairly high (>10 nm/min), however, because the germane catalyzes the silicon growth.[9,10]

The first silicon epitaxial growth experiments in a reactor like that of Figure 1 were performed by Gibbons et al.[11] and were termed limited-reaction processing (LRP). The central feature of LRP is that the reactive gas flows were first established when the reactor was cold, and then the growth reaction was switched on and off using rapid changes in sample temperature (up to 300 K/s). This differs from the conventional CVD approach, where the wafer is first brought to the growth temperature in a passive ambient gas (i.e., hydrogen) before the reactive gas flows for growth are introduced. In LRP, the gases can be switched while the wafer is cold, and subsequent temperature pulses can be changed for each layer. In the conventional approach, the reactive gas flows can be turned off while the wafer temperature is directly ramped from that for one layer to that for the next. In both approaches, the growth conditions can be optimized for each individual layer in a multilayer structure.

To examine the relative merits of the two approaches for structures on a 10 nm scale, consider a p-i-p-i superlattice with 10 nm layers (Figure 2). At 1,000°C, typical growth rates are on the order of 0.25 μm/min, so each 10 nm layer requires only 2.4 s at the growth temperature. Diffusion constants are so high at this temperature that even the simulated profile grown by LRP to minimize diffusion already had interfaces seriously degraded on a 10 nm scale. This shows that at high-temperature growth, absolute minimization of the time at the

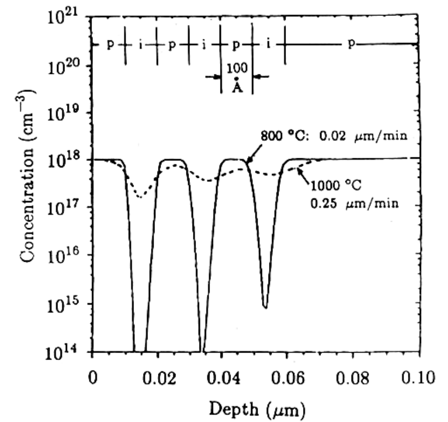

Figure 2. A simulated profile of a p-i-p-i superlattice grown at 1,000°C and at 800°C with no excess thermal exposure.[12]

PHOTOLUMINESCENCE SPECTRA

A long-sought goal of microelectronics has been a material technology for monolithically integrating the worlds of optoelectronics and very large-scale integration (VLSI) technology. This is a difficult challenge, since optoelectronics technology is based on III-V semiconductors such as GaAs and InP with direct bandgaps, and VLSI technology is based on silicon because of its manufacturability. Thus, the incompatibility and lattice mismatch problems described in the main text apply in this arena as well. Although $Si_{1-x}Ge_x$ alloys are still indirect-bandgap materials, it has been predicted that certain short-period silicon/germanium superlattice structures will exhibit "zone-folding" effects that will create a pseudodirect-gap material from the indirect-bandgap constituents and, hence, have the possibility of emitting large amounts of light.[22,23] The simplest experimental probe of a semiconductor structure for its light-emitting properties is photoluminescence—light is shone on the semiconductor to create electron-hole pairs, and one observes the spectrum emitted as the electrons and holes recombine. The photoluminescence results from such zone-folded structures[24,25] are presently somewhat controversial,[26] however, especially since some of the signals closely correspond to the known dislocation photoluminescence in silicon. Other MBE work has produced single strained films which have a strikingly large photoluminescence efficiency of ten percent.[27] The signal is very broad (~80 meV) and substantially below the bandgap (~120 meV), however; the exact transition causing this emission is not known.

The photoluminescence spectra of $Si_{1-x}Ge_x$ strained layers, quantum wells, and superlattices grown by RTCVD[28] have proved substantially different from those reported earlier by MBE. Typical spectra of a sample with multiple $Si/Si_{0.8}Ge_{0.2}/Si$ quantum wells measured at 2 K and 77 K are shown in Figure A. In the 2 K spectra, there are several well-resolved peaks which can be identified by comparison to the comprehensive work of Weber and Alonso[29] on bulk unstrained $Si_{1-x}Ge_x$ alloys. The large high-energy peak is from direct recombination of electrons and holes without a phonon; the other lower energy peaks (TA and TO) are the recombination

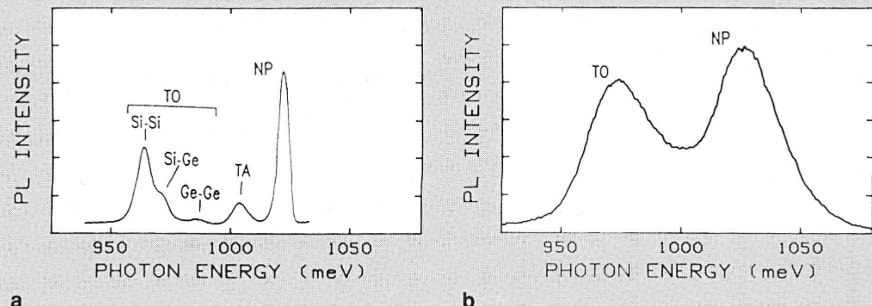

Figure A. A comparison of (a) 4 K and (b) 77 K photoluminescence (PL) spectra for a sample consisting of multiple $Si/Si_{0.8}Ge_{0.2}/Si$ quantum wells. NP = no-phonon signal.

of the same electrons and holes but accompanied by the emission of a phonon (TA or TO), so that the energy of the emitted photon is lower by the energy of the phonon. Note that the TO phonon, which represents a local mode vibration, actually splits from a single peak in silicon to three peaks in the alloy. This occurs because there are three possibilities for nearest neighbors (Si-Si, Si-Ge, and Ge-Ge). A very interesting feature is the no-phonon (NP) recombination signal. Such a signal is commonly observed in silicon at low temperatures when the excitons are bound to localization sites, but not at higher temperatures (over 20 K) when the excitons are free. In the $Si_{1-x}Ge_x$ alloys grown by RTCVD, the NP signal actually does persist to higher temperatures when the excitons are no longer bound.[28] This NP peak is still clearly evident (although thermally broadened) at 77 K (Figure A) and has been observed up to 300 K. The physical basis for NP luminescence in an indirect-gap material lies in the fact that $Si_{1-x}Ge_x$ is a random alloy. The randomness of the material breaks the perfect translational symmetry of the lattice—the arrangement of neighbors looks different from each lattice site. Therefore, the usual Bloch wave functions, which are eigenstates of momentum k, are no longer the eigenstates of the electrons in the crystal. The true eigenstates can be found by perturbation theory, which will mix different k-components into each state. Therefore, nearly any two electron and hole wavefunctions will have some optical matrix element between them, and optical transitions without phonons are allowed. A similar effect was observed in the relaxed $Si_{1-x}Ge_x$ alloys as well.[29] It should be stressed, however, that zone-folding and pseudodirect-bandgap effects are not being observed, since the superlattice periods are too big (>4 nm) and similar signals are observed in single $Si_{1-x}Ge_x$ films as well as the superlattices.

The well-resolved band-edge photoluminescence in the RTCVD films reflects their high material quality. The high lifetimes in the RTCVD films mean that the nonradiative recombination from defect sites was low enough that the carrier could radiatively recombine to give rise to a photoluminescence signal. This may be why such a signal was not observed in films grown by other techniques. Further, the narrow line width of the signal (full-width-at-half-maximum of 5 meV of the NP signal at 2 K) is comparable to the best reported for relaxed bulk alloys, indicating the good uniformity of the alloys and an absence of significant clustering. This well-resolved photoluminescence should be useful for the future study of the band structure of the alloys.

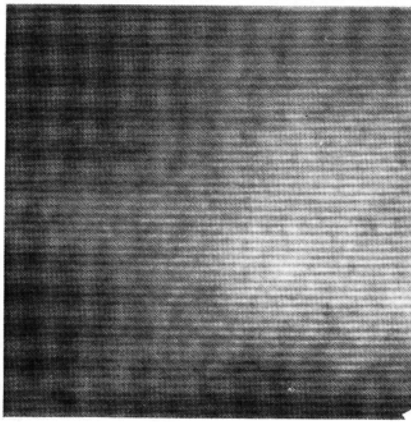

Figure 3. A cross-section transmission electron micrograph of a $Si/Si_{0.8}Ge_{0.2}$ 50-period superlattice grown by RTCVD with a period of 4.5 nm.[12]

growth temperature is important, but still not sufficient for 10 nm structures. At lower temperatures, longer growth times are required because of slower growth rates; but because the activation energies of growth are smaller than those of typical substitutional diffusers (2 eV versus 3–5 eV), more abrupt structures can be grown, as demonstrated by the 800°C simulation in Figure 2.[12] At temperatures under 800°C, the times for 10 nm layers will be on the order of one minute, so that the several seconds savings of thermal exposure made possible by LRP may not be significant. The gas switching approach also has the potential advantage under these conditions that control of rapid-temperature ramps is not required. Using the gas switching approach, well-resolved $Si/Si_{1-x}Ge_x$ superlattices have been grown with periods as small as 4.5 nm (Figure 3).[12]

As mentioned above, temperatures under 800°C are required for the growth of 10 nm structures. For $Si_{1-x}Ge_x$, even lower temperatures (<650°C) are needed to achieve metastable strained layers in excess of critical thickness and to prevent nonplanar growth modes.[3] At low temperatures, however, silicon-based CVD growth becomes exceedingly difficult because of the enhanced stability of oxygen and water on the silicon surface.[13,14] The first successful silicon epitaxial experiments in the 600°C range were performed by UHV/CVD;[8] the low gas pressures (mtorr) resulted in a very low partial pressure of any impurities (such as water vapor) that might be contained in the gas source. Initial epitaxial experiments in this range by RTCVD were performed in a reactor without a load lock, so that the reactor chamber was vented to the atmosphere each time a sample was loaded. The resulting oxygen levels in $Si_{1-x}Ge_x$ films grown at 640°C were 10^{20} cm^{-3}, two orders of magnitude higher than is generally acceptable in silicon films.[15] Subsequent experiments using a load lock to reduce water and oxygen levels showed that high-lifetime, low-oxygen films could indeed be grown in a non-UHV RTCVD reactor at 625°C.[16] Through improved gas purity and reactor design, it is now possible to grow low-oxygen films in this temperature range at atmospheric pressure.[17]

HETEROJUNCTION BIPOLAR TRANSISTORS

The heterojunction bipolar transistor (HBT) is a device which benefits exponentially from small changes in the bandgap within the device structure. The basic device consists of three semiconductor layers known as the collector, base, and emitter. Figure 4 shows the location of the conduction and valence bands of the semiconductor versus distance for an all-silicon homojunction transistor and for a narrow-bandgap-base HBT device, where the base is made of $Si_{1-x}Ge_x$ and the emitter and collector regions are made of silicon. Under normal bias conditions, electrons in the conduction band travel from the emitter over the base barrier to the collector to form collector current (I_C). Holes also travel in the valence band from the base to the emitter to form base current (I_B). Since the current gain is I_C/I_B, large collector currents and small base currents are desirable. If the bandgap of the base region is reduced and the base doping is held constant (which keeps the valence band in the base at approximately the same level), it is clear that electrons (which travel in the conduction band) will face a lower barrier as they travel from emitter to collector. To first order, the resulting increase in I_C will be by a factor of $\exp(\Delta E_G/kT)$, where ΔE_G is the bandgap reduction. Further, since the barrier faced by holes is unchanged, the base currents should be unchanged, resulting in an increased gain by the same factor. For example, for an $Si_{0.8}Ge_{0.2}$ base, the bandgap reduction of ~160 meV results in approximately a 500-fold increase in I_C at room temperature. This extra gain itself may be useful for an improved circuit, or the base doping could then be increased to reduce the base resistance. Increasing the base doping brings the gain back down, but high base resistance is a more serious limiting factor than the gain of the transistor in many circuits.

The first $Si/Si_{1-x}Ge_x/Si$ HBTs grown by any technique that exhibited a near-ideal exponential increase in collector current were grown by LRP,[15] but base currents were excessively large and limited the gain of the devices. Subsequent results in films grown by UHV/CVD or RTCVD with low oxygen levels have confirmed that ideally low base currents independent of the base composition are also possible,[18,19] in addition to the increased collector currents. An example of the effect of the narrow-bandgap base

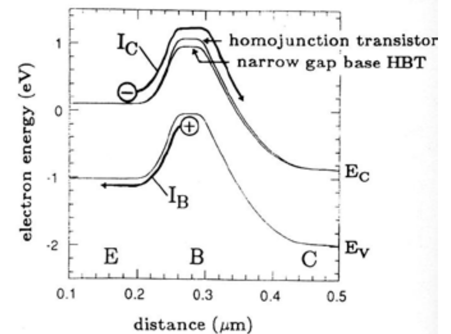

Figure 4. A band diagram showing the collector and base current components in a homojunction transistor and an HBT. E = emitter; B = base; C = collector.

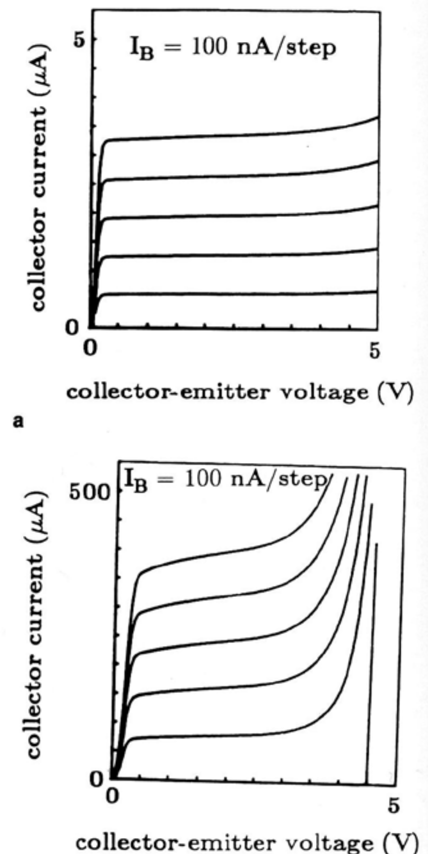

Figure 5. Common-emitter characteristics of (a) an all-silicon bipolar transistor and (b) a device of $Si/Si_{0.86}Ge_{0.14}/Si$ with similar doping levels.

on performance is shown in Figure 5, where an all-silicon device and a device with a base of $Si_{0.86}Ge_{0.14}$ but otherwise similar dopings are compared. The gain increases from 6 to 700, as expected with the narrow-bandgap base.

A related advantage of the HBTs is the possibility of low-temperature operation. The gain of conventional bipolar transistors actually degrades at low temperatures because of bandgap narrowing due to the heavy doping normally present in the emitter (resulting in an HBT but with the bandgap reduced on the wrong side

of the emitter-base junction). This can be more than compensated for in the narrow-bandgap-base HBTs, and the effect should become larger at low temperature when kT is reduced. For example, the gain of an all-silicon bipolar transistor typically decreases from 100 to <10 as the temperature is lowered from 300 K to 100 K, but the gain of an HBT with a $Si_{0.8}Ge_{0.2}$ base has been observed to increase from 1,000 to more than 10,000 over the same temperature range.[19]

Beside the possibility of increased gain (or lower base resistance), a second advantage of HBTs is the possibility of introducing a "built-in" electric field in the base by grading the germanium content (and thus the bandgap) across the base region. This electric field can speed up the electrons as they travel across the base, resulting in a faster device. Researchers at IBM have recently demonstrated $Si/Si_{1-x}Ge_x/Si$ HBTs with such graded bases which exhibit current gain at frequencies up to 75 GHz at room temperature and 96 GHz at low temperature.[20,21] These devices were actually grown by UHV/CVD, but devices with similar grading have been grown by RTCVD as well.[19] In both cases, the grading was accomplished by adjusting the source gas flows as a function of time. These results represent a more than 60 percent increase over previous silicon-based device speeds and open a new frequency performance regime to silicon integrated circuits.

ACKNOWLEDGEMENTS

The $Si_{1-x}Ge_x$ research effort at Princeton University has been supported by the Office of Naval Research, the National Science Foundation, and the New Jersey Commission on Science and Technology. The author would also like to thank C. Magee (Evans East), S. Schwarz and B. Wilkens (Bellcore), E. Fitzgerald (AT&T), L. Lenchyshyn and M.L.W. Thewalt (Simon Fraser University), and N. Rowell, J.P. Noel, J. McCaffrey, and D.C. Houghton (National Research Council, Canada) for their contributions and collaborations on various parts of the work at Princeton. Finally, the author would like to acknowledge P.M. Garone, H. Manoharan, E.J. Prinz, C.M. Reaves, P.V. Schwartz, and X. Xiao for their work at Princeton.

References

1. E. Kasper, H.J. Herzog, and H. Kibbel, *Appl. Phys.*, 8 (1975), p. 199.
2. E. Kasper and H.J. Herzog, *Thin Solid Films*, 44 (1977), p. 357.
3. J.C. Bean et al., *J. Vac. Sci. Tech.*, A2 (1984), p. 436.
4. J.C. Bean et al., *Appl. Phys. Lett.*, 44 (1984), p. 102.
5. J.P. Dismukes, L. Ekstrom, and P.J. Paff, *J. Phys. Chem.*, 88 (1974), p. 302.
6. R. People, *Phys. Rev. B.*, 32 (1985), p. 1405.
7. D.V. Lang et al., *Appl. Phys. Lett.*, 47 (1985), p. 1333.
8. B.S. Meyerson, *Appl. Phys. Lett.*, 48 (1986), p. 797.
9. B.S. Meyerson, K.J. Uram, and F.K. Legoues, *Appl. Phys. Lett*, 53 (1988), p. 2555.
10. P.M. Garone et al., *Appl. Phys. Lett.*, 56 (1990), p. 1275.
11. J.F. Gibbons, C.M. Gronet, and K.E. Williams, *Appl. Phys. Lett.*, 47 (1985), p. 721.
12. J.C. Sturm, P.V. Schwartz, and E.J. Prinz, *J. Vac. Sci. Tech. B*, 9 (July 1991), p. 2011.
13. G. Ghidini and F.W. Smith, *J. Electrochem. Soc.*, 131 (1984), p. 2924.
14. M.L. Green et al., *JEM*, 19 (1990), p. 1015.
15. C.A. King et al., *IEEE Elec. Dev. Lett.*, EDL-10 (1989), p. 52.
16. P.V. Schwartz and J.C. Sturm, *Appl. Phys. Lett.*, 57 (1990), p. 2004.
17. T.O. Sedgewick, M. Berkenblit, and T.S. Kuan, *Appl. Phys. Lett.*, 54 (1989), p. 2689.
18. G.L. Patton et al., *IEEE Elec. Dev. Lett.*, EDL-10 (1989), p. 534.
19. J.C. Sturm and E.J. Prinz, *IEEE Elec. Dev. Lett.*, EDL-12, (1991), p. 303.
20. G.L. Patton et al., *IEEE Elec. Dev. Lett.*, EDL-11 (1990), p. 171.
21. E.F. Crabbe et al., *IEDM Tech. Dig.* (1990), p. 17.
22. U. Gnutzmann and K. Clausecker, *Appl. Phys.*, 3 (1974), p. 9.
23. S. Satpathy, R.M. Martin, and C.G. Van de Walle, *Phys. Rev. B*, 38 (1988), p. 13237.
24. H. Okumura et al., *Jpn. J. Appl. Phys.*, 28 (1989), p. L1893.
25. R. Zachai et al., *Phys. Rev. Lett.*, 64 (1990), p. 1055.
26. U. Schmid, N.E. Christensen, and M. Cardonna, *Phys. Rev. Lett.*, 65 (1990), p. 2610.
27. J.-P. Noel et al., *Appl. Phys. Lett.*, 57 (1990), p. 1037.
28. J.C. Sturm et al., *Phys. Rev. Lett.*, 66 (1991), p. 1362.
29. J. Weber and M.I. Alonso, *Phys. Rev. B*, 40 (1989), p. 5683.

ABOUT THE AUTHOR

James C. Sturm earned his Ph.D. in electrical engineering at Stanford University in 1985. He is currently an assistant professor in the Department of Electrical Engineering at Princeton University. Dr. Sturm is also a member of TMS.