Si Industry at a Crossroads: New Materials or New Factories?

Eugene A. Fitzgerald, Chris W. Leitz, Minjoo L. Lee, Dimitri A. Antoniadis, Matthew T. Currie

Abstract – Many trends in the silicon industry could be interpreted as the herald of the end of traditional Si scaling. If this premise holds, future performance and system-on-chip applications may not be reached with conventional Si technology extensions. We review progress towards our vision that a larger crystal structure on Si, namely relaxed SiGe epitaxial layers, can support many generations of higher performance Si CMOS and new system-on-chip functionality without the expense of significant new equipment and change to CMOS manufacturing ideology. We will review the impact of tensile strained Si layers grown on relaxed SiGe layers. Both NMOS and PMOS exhibit higher carrier mobilities due to the strained Si MOSFET channel. Heterostructure MOSFETs designed on relaxed SiGe can have multiple-generation performance increases, and therefore determine a new performance roadmap for Si CMOS technology, independent of MOSFET gate length. We also indicate that this materials platform naturally leads to incorporating new optical functionality into Si CMOS technology.

Keywords – strained-Si, SiGe, MOSFET, mobility, inverter, frequency, power.

I. Introduction

Recent data from the semiconductor industry show important trends for Si scaling [1]. It should be recognized that the 3-4 year delay in reaching peak 8-inch wafer consumption as compared to the estimated 10-15 year delay before reaching peak 12-inch wafer consumption shows that Moore's law, in some respects, has been slowing since the mid-1990's. In combination, the reduced adoption rate of 12-inch wafers, natural business cycles and the increasing cost of manufacturing facilities have created a depression in the semiconductor cycle, increasing risk and lowering return on investment. The bottom of the cycle in 1996 represented a drop of 52% in orders from the previous peak, the bottom of the 1998 cycle represented a drop of 72% from the previous

Eugene A. Fitzgerald is with the Materials Science and Engineering Department, MIT, Cambridge, MA, with the Singapore-MIT Alliance, and also with Amberwave Systems Corporation, Salem, NH.

Chris W. Leitz and Minjoo L. Lee are with the Materials Science and Engineering Department, MIT, Cambridge, MA.

Dimitri A. Anotniadis is with the Department of EECS, MIT, Cambridge, MA and with the Singapore-MIT Alliance.

Matthew T. Currie is with Amberwave Systems Corporation, Salem, NH.

peak, and the current drop so far represents a drop of about 80% from the previous peak.

In parallel to these deleterious economic factors, the scaling of the silicon MOSFET no longer delivers the end-user value that has been achieved in the past. During the prime of conventional silicon scaling, decreasing MOSFET dimension brought advantage in three areas: cost, functionality, and performance. The cost of integrated circuits was decreased by shrinking the die size of a given circuit with concurrent increases in wafer size, allowing more die to fit on a wafer. Increased functionality for the end customer was achieved since the scaling allowed more transistors to be used in a given area. Finally, the natural consequence of scaling dimension is that electrons and holes require less time to traverse the channels of smaller MOSFETs. Thus, the operational frequency of the transistor increases, resulting in higher performance for digital circuits.

The delayed deployment of larger wafers, the diminishing increase in performance for further shrinking of MOSFETs, and the 'design gap' that exists in order to create circuits that can take full advantage of higher transistor densities have all contributed to the end of the traditional silicon roadmap. However, it is inconceivable that further progress in silicon CMOS technology will not be required in order for the information age to continue. Increased functionality, performance, and lower cost are all still in demand for future applications, but a new path to achieving this in silicon technology must be found.

II. A Materials Revolution

During mainstream silicon scaling, the majority of material used in the actual silicon chip was dominated by three elements: Si, O, and Al. As we have approached the end of traditional scaling, new materials are being employed to achieve performance goals near the 0.13µm node. For complex circuits, the wiring of the silicon chip in the back-end has its own scaling limitations, i.e. without improving interconnect delay, further performance in transistor improvements will not produce the same performance increase at the circuit level. New materials solutions, such as copper interconnects and low dielectric constant insulators were introduced at the 0.13µm node to accomplish this goal. Thus, the 0.13µm node has deviated from traditional scaling in that an unprecented degree of flexibility has been introduced with respect to new material introduction. This trend can not be ignored. A clear movement toward the introduction of new elements to increase performance is occurring. Another example of this trend is the introduction of Ge into SiGe BiCMOS processes in order to enhance the I/O capability. Just a

few years ago, such an unprecedented incorporation of new materials was considered prepostorous.

In this abstract, we show that the introduction of Ge into Si-based MOSFET heterostructures enables a continual increase in MOSFET performance, leading to either faster circuits or circuits with substantially lower power consumption, both features extending the Semiconductor Industry Association (SIA) Roadmap beyond the traditional improvements. As these improvements are largely independent of device geometry, they can accelerate the performance of Si-based MOSFET technology beyond that predicted by the current Semiconductor Industry Association (SIA) Roadmap.

III. Strained or Relaxed?

SiGe is a completely miscible alloy, allowing any Ge concentration to be alloyed with Si without compound formation. Since Ge is a larger atom, the lattice constant of a SiGe alloy increases with increasing Ge content. When deposited on a Si substrate, the greater lattice constant of the SiGe alloy places the SiGe layer in a stress-strain state of biaxial compression. With enough stress, which increases linearly with Ge concentration, or if the SiGe layer thickness is great enough (the point of this transition is commonly referred to as the *critical thickness*), misfit dislocations will be introduced to relieve the stress. There are two ways SiGe layers can be implemented into a device integration scheme: 1) The entire compressive biaxial stress can be retained by limiting the Ge content and thickness of the SiGe layer (the Strained Case) 2) The SiGe layer can be completely dislocated by an engineered approach that minimizes the dislocation propagation to relevant areas for device layers (the Relaxed Case). For a review of lattice mismatched epitaxy, applications, and critical thickness see reference [2].

IV. The SiGe HBT and SiGe BiCMOS

From 1980-1990, many strained-layer devices (i.e., devices composed of lattice-mismatched semiconductors with active layer thickness below the critical thickness) were investigated. These devices avoided any problems associated with defect introduction since the layers were kept below the critical thickness. However, this often limited the enhancement in many of these devices as well. The SiGe heterojunction bipolar transistor (HBT) is such a casewhere a small concentration of Ge is incorporated into the base of a Si bipolar junction transistor. The base layers are sufficiently thin to avoid misfit dislocation introduction in the base, which would degrade the device performance. The placement of a strained SiGe layer on a Si lattice constant leads to a valence band offset between the Si and SiGe, which is the emitter-base junction in a bipolar transistor, resulting in increased gain or freqeuncy performanece.

To date, Ge has been added to commercial Si processes via the strained approach. In current leading-edge BiCMOS platforms, SiGe HBTs are implemented, but there is absolutely no improvement in the majority of the devices (i.e., the CMOS devices) on the chip . The main effect is to enhance the I/O capabilities of the chip as compared to a conventional all-Si BiCMOS chip and overall the advantages of SiGe alloys are not exploited to a great extent.

V. Relaxed SiGe: A Platform for Continual Innovation in the Si Industry

The end of the roadmap discussed in sections I and II suggest that adding functionality and performance at little increase or possibly a decrease in cost should be the goal of new material introduction. Interestingly, the introduction of copper, low dielectric constant insulator, and SiGe HBTs provides improved performance and slight increases in functionality (the I/O improvement in CMOS), but at the price of additional cost. Unfortunately, each of these introductions provides only a single performance increase, not a roadmap for continual evolution in performance. The fact that these technologies are being adopted even though they only provide incremental performance boosts is evidence that materials solutions that can increase performance, functionality, and decrease cost will be very attractive.

Relaxed SiGe on Si is a materials platform that can deliver increased performance and functionality at a reduced cost. Strained Si layers and other SiGe heterostructures deposited on top of the relaxed SiGe layers offer increased MOSFET performance equivalent to multiple generations of traditional scaling. In addition, SiGe MOSFET heterostructures can offer new functionality in MOSFETs, such as low noise figures. Such characteristics were previously only available in more exotic devices like bipolar transistors or III-V compound high electron mobility transistors. Finally, relaxed SiGe alloys on Si can be grown of sufficiently high quality such that III-V materials like GaAs and InP can be deposited on such alloys with sufficient quality for optoelectronic applications. Thus, while strained SiGe with a small amount of Ge can moderately enhance Si BiCMOS, relaxed SiGe on Si is a more universal platform with generations of improvement possible from higher performance CMOS to integration of Si CMOS and III-V optoelectronics.

VI. SiGe Alloy Implementation Map

To fully understand the difference between the application of strained SiGe and relaxed SiGe, Fig. 1 is a diagram of integrated silicon system performance vs. Ge concentration in SiGe alloys on Si. At the left side of the diagram, the lattice constant of the SiGe alloy remains close to that of Si. To avoid dislocations, little Ge can be added to silicon (i.e., limited extension to the right on this diagram). If the alloy Ge concentration is

low, a paucity of applications can benefit from this small amount of Ge. Also, the inset shows that strained SiGe alloys on Si substrates create a band structure that confines holes but not electrons. This band alignment further limits the utility of such materials. Thus, one of the few devices that can take advantage of this limited enhancement is the n-p-n heterojunction bipolar transistor which was previously discussed. Further to the right in Fig. 1, we see that many generations of CMOS improvement, including the integration of optoelectronics, are possible if the critical thickness criterion for high Ge concentrations is exceeded. For many years, all applications in III-Vs and SiGe explored only lattice constants close to that of the substrate material. However, in 1991, renewed interest in relaxed semiconductor systems led to low threading dislocation density, relaxed SiGe on Si [3]. It was quickly realized that such a platform allowed higher carrier mobilities in MOSFET configurations as well as the long term potential of integrating III-V materials [4][5][6][7][8].

Once a relaxed SiGe lattice constant is produced on Si, much higher mobilities in electrons and holes can be produced in Si-based MOSFET devices. As Fig. 1 shows in the central region, the band alignment for thin tensile-strained Si or SiGe alloys and thin compressively-strained Ge or Ge-rich alloys are advantageous for confining high mobility electrons and holes. For a given concentration of Ge in the relaxed SiGe buffer, if a lower concentration of Ge is used in a device layer, the device layer will be in tension. The tension improves the mobility and enables the confinement of electrons. If a higher concentration of Ge is used in a device layer than in the substrate, the device layer will be in compression. This compression improves the mobility of holes, and the holes are confined to the compressed layer. Thus, the relaxed SiGe buffer on Si provides a platform for the contintual enhancement of the mobility in NMOSFETs as well as PMOSFETs.

As we eventually scale to higher Ge concentrations in Fig. 1, a natural transition is the incorporation of GaAs-based materials for optoelectronics, since the lattice constant of GaAs and Ge are very close. GaAs incorporated on Si in such a fashion has been shown to be viable for LEDs [5], solar cells [9], and has been proven to host the highest minority carrier lifetime in any GaAs on Si ever produced [10]. The ability to fabricate optical links on a Si CMOS circuit using standard CMOS processing will usher in a new age of ultra-low cost board-to-board and chip-to-chip interconnections [11]. The ability to create such low-cost optics will change system architecture in many communication products, and the future markets and outcome of such a disruption is unforseen at this time.

VI. Strained Si and Ge-based Transistors

In this section, we review some of the devices that indicate how silicon CMOS can evolve using strained SiGe heterostructures (i.e., the middle part of Fig. 1). The simplest strained Si transistors incorporate a thin strained Si layer on a relaxed $Si_{0.8}Ge_{0.2}$ buffer layer, essentially the same structure used to measure the first high mobility strained Si in 1991 [4], except the top SiGe layer has been removed to simplify the structure. The schematic of the layer structure in this surface channel MOSFET is shown in Fig. 2. A Ge concentration of 20% is typically used since the electron mobility enhancement saturates at that value. Such a surface channel structure may have a lower mobility at low vertical fields (under the gate of the MOSFET) than a structure which includes a top SiGe layer, but since MOSFETs used in digital electronics operate at high vertical fields, only the high-vertical-field transport is relevant for these applications. A typical plot of mobility vs. vertical electric field under the gate is shown in Fig. 3. As one can see, in a conventional MOSFET, as the vertical field is increased, the mobility decreases. As silicon electronics has been scaled, the smaller MOSFETs have an increasing vertical field under the gate (obtained by a thinner gate oxide even though gate voltages are being scaled to lower voltages), and thus mobility in MOSFETs has been decreasing as MOSFETs have been scaled. The MOSFETs are able to operate at higher frequencies since the distance between the source and drain decrease during geometric scaling; thus, despite the decrease in mobility with scaling, increased performance was obtained during theprime of tradtional CMOS scaling.

Fig. 4 is a plot of mobility in short-flow annular NMOSFETs as a function of vertical field, for both control Si devices and the structure in Fig. 2. Fig. 4 shows that in strained Si channels, a similar roll-off in mobility is experienced with higher vertical fields. However, a very impressive characteristic is that the enhancement in electron mobility is retained, independent of vertical electric field. This behavior is important because it demonstrates the scalability of strained Si, i.e., the enhancement in electron mobility occurs independent of the generation at which it is employed. This enhancement independent of vertical field can be displayed more clearly by dividing the strained Si curve by the control Si curve, producing enhancement as a function of vertical field, as shown in Fig. 5. An enhancement as high as 80% is seen, quite a remarkable enhancement since conventional scaling, when yielding great performance gains, produced an enhancement of approximately 20-30% per generation of Si.

Fig. 1 implies that increasing Ge concentration will yield further benefit to digital electronics, yet electron mobility enhancement ceases at 20% Ge. The continued improvement will be due to enhancement of hole mobility, since increasing Ge concentration in the relaxed buffer continues to enhance hole mobility. Both electron and hole mobility enhancements are required to continually improve the speed, or decrease the power, of

digital CMOS since CMOS consists of both NMOSFETs and PMOSFETs. Fig. 6 is a graph of hole mobility extracted from MOSFET devices as a function of Ge concentration in the relaxed buffer [12][13]. Hole enhancement factors of greater than 2 were obtained with the same basic structure of Fig. 2, with higher Ge concentrations (>35%).

How far can we improve the PMOS performance using these heterostructures? We have recently created a compressive Ge channel on a 70% Ge concentration buffer, with a dislocated epitaxial Si cap layer at the surface [14]. Thus, holes in this MOSFET lie almost entirely in the compressive Ge channel, except for the part of the hole wavefunction which resides in the thin Si cap layer. This thin epitaxial layer of Si is important, since the mature SiO_2/Si gate technology can be used with this compressive Ge MOSFET. Also note that a significant facet of this MOSFET is the proof that the dislocated Si cap layer is not deleterious to the gate technology. Although it was anticipated this was the case due to the nature of dislocation electrical activity and the amorphous/crystalline interface, the data confirms this hypothesis. Fig. 7 is a plot of effective mobility enhancement in these MOSFETs vs. vertical electric field, and one can clearly see that the enhancement approaches 8! The two curves correspond to MOSFETs with two different Si cap thicknesses, 40 and 60 Å. The 40 Å cap displays a higher mobility since more of the wavefunction lies in compressive Ge and less in the top Si cap layer.

It is interesting that strained Si research, started in 1991, has revealed recently this frutiful path of SiGe MOSFET heterostructures, potentially extending all the way to pure Ge channel MOSFETs. Thus, transistor research has come full circle in the sense that the first transistors were entirely composed of Ge. Ge did not scale in manufacturing due to its poor oxide and low mechanical robustness. The advent of high quality relaxed SiGe buffers on Si substrates [3] combined with the ability to use SiO_2/Si gate technology in the Ge MOSFET [14] has potentially allowed the injection of the Ge MOSFET into the Si manufacturing infrastructure, some 50 years later.

VII. Strained Si and Ge-based Inverters

With the possibility of enhancing both the NMOSFET and PMOSFET using strained SiGe MOSFET heterostructures, we quickly realize that a CMOS inverter becomes the basic element to track for generational improvement in heterostructure CMOS. Fig. 8 is a schematic of a strained channel inverter, in which the NMOSFET and PMOSFET both have the potential for carrier mobility improvement, thus allowing the inverter to operate at higher frequency or lower power. In current Si CMOS, the electron mobility is higher than the hole mobility, and therefore to balance the inverter the PMOSFET is designed to be 2 to 3 times as wide as the NMOSFET. Although the

PMOSFET has a slightly higher capacitance due to this scaling, this increase in delay is compensated by the increased drive current and balanced inverter, thus creating an optimum typically with a ratio of widths between 2 and 3.

Fig. 9 is plot of the calculated delay time of an strained channel inverter vs. transistor gate length. Also drawn on the plot is the Si control inverter, as dictated by past and present Si scaling, and including the Semiconductor Industry Association Roadmap projection for gate lengths which have yet to be commercialized. Four curves are shown for strained heterostructure inverters, the first 3 labeled ASC1, ASC2, and ASC3 calculated for inverters based on heterostructures with increasing Ge concentration in the relaxed buffer, but with the width of the transistors unchanged. ASC4 represents an inverter based on high Ge concentrations with the width of the transistors in the inverter optimized for minimum delay. As can be seen in the figure, a change in substrate material which includes different heterostructures is very promising for future effective Si scaling.

VIII. New Materials, No New Factories

In this paper, we have presented the case that the silicon roadmap is slowing, and that new substrate materials, relaxed SiGe with strained heterostructures, offer a new path for Si CMOS technology. Relaxed SiGe on Si offers a new path since improved performance and higher functionality is obtained at lower cost. Higher performance is obtained through a fundamental increase in carrier mobility inside every MOSFET, which can be used for either higher frequency performance or lower power consumption. Fig. 10 is a plot of microprocessor frequency vs. year, based on the enhancements in inverter speed shown in Fig. 9. As one can see, significantly higher frequency performance can be expected if the transistor speed enhancement can be extracted without any other bottlenecks, such as back-end interconnect delay. Fig. 11 and Fig. 12 are plots of similar projections for power consumption. Given that interconnect delay will be introduced once again as transistor speed no longer becomes the bottleneck, and considering that the frequencies projected by Fig. 10 are not required in most current end user applications, we suspect that the use of the higher carrier mobilities to operate at lower voltages and thereby reduce power consumption dramatically (as shown in Figs. 11 and 12) will be the main use of future CMOS technology. The plethora of anticipated portable devices which require sufficient (i.e., not state-of-the-art) computing power will likely be a large driver for adoption of this higher performance roadmap. Clock frequency will no longer be the dominant figure of merit for Si CMOS as ultra-low-power heterostructure CMOS is deployed.

Higher functionality for CMOS is obtained since such low power operation combined with other low noise MOSFETs on this platform allows new system-

on-chip functionality not possible with conventional Si. Long-term, the ability to also integrate optical devices with Si CMOS in a Si CMOS manufacturing infrastructure will produce functionality in markets that can currently not be anticipated. Finally, lower cost is achieved since the infrastructure needed to manufacture products using this technology already exists, i.e. no new manufacturing equipment in new factories is needed.

Since dimensionality (gate length) is no longer delivering improved performance, cost, and functionality, what is the physical parameter on which many performance parameters will be plotted to reveal generational improvement? Based on the long path of increasing potential shown in Fig. 1, one possible conclusion is that the lattice-constant on which electronics are built is a new metric (which correlates to Ge composition). Based on the new roadmaps shown in this paper, Fig. 13 is a new roadmap based on lattice constant. The rate of deployment of lattice constant shown in Fig. 13 is consistent with the increase in performance metrics shown in Figs. 11-13.

What will this future of new materials instead of drastically new factories look like? Considering that a new 12-inch fabrication facility costs in the neighborhood of 5 Billion US dollars, and that in 2001 an estimated 63 Billion US dollars was invested in such infrastructure by world-wide semiconductor manufacturers, one must ask how such funds would be invested in this new CMOS world. First, if the funds were invested in other parts of the disaggregated industry, such as design, one can speculate that innovation at the chip level would increase drastically, especially with the new functionality introduced by the new materials platforms. However, most likely the largest effect would be a sharp decrease in silicon chip manufacturing cost, since manufacturers no longer need to invest such large funds into factories with drastically new equipment. New factories will be built, but the manufacturing equipment will be mature. In fact, in time, this suggests that fabrication equipment for semiconductor manufacturing will be commoditized, and eventually fabrication facility cost may actually decrease. Thus, if expansion occurs to increase chip volume in this commoditized world, the cost will be relatively low and the risk to expand may be lower compared to the high risk game played today. Overall, this paradigm shift to new materials should accelerate the implementation of novel complex chip technologies in end products.

IX. Conclusion

We have shown that the silicon roadmap, using the MOSFET dimension as the metric, is approaching an end. Relaxed SiGe alloys on Si substrates offer a new materials platform that can continue to enhance CMOS performance, functionality, and cost, even as conventional scaling approaches its limits. NMOS fabricated on strained Si possess electron mobilities with an enhancement of 1.8 (80% increase), and PMOS has been shown to have potential to increase hole mobilities by a enhancement factor of nearly 8. Putting these enhancements together in a digital inverter, one can enhance CMOS performance for either higher frequency performance or lower power operation far into the future. Long term, the high Ge concentrations in the buffer layers allow for the integration of high quality optoelectronic materials and devices. Thus, the lattice constant of the SiGe buffer on Si becomes the metric for the new roadmap. In a world in which new materials, not novel factories, define the technology path, factory costs eventually decrease, and capital not used in the construction of very expensive novel factories accelerates the adoption of complex chip technologies.

X. Acknowledgements

References

[1] M. O'Brian and M. Hosseini, "Semiconductor Manufacturing 2001 Midyear Review," Wit SoundView Analyst Report, July 13, 2001.

[2] E. A. Fitzgerald, "Dislocations in Strained-Layer Epitaxy: Theory, Experiment, and Applications," *Mater. Sci. Reports,* Vol. 7, p. 87, 1991.

[3] E. A. Fitzgerald, Y.-H. Xie, M. L. Green, D. Brasen, A. R. Kortan, J. Michel, Y.-J. Mii, and B. E. Weir, "Totally relaxed Ge_xSi_{1-x} layers with low threading dislocation densities grown on Si substrates," Appl. Phys. Lett., Vol. 59, No. 7, pp. 811-813, 1991.

[4] Y.J. Mii, Y.H. Xie, E.A. Fitzgerald, Don Monroe, F.A. Thiel, B.E. Weir, and L.C. Feldman, "Extremely High Electron Mobility in Si/SiGe Structures Grown by Molecular Beam Epitaxy," Appl. Phys. Lett., Vol. 59, pp. 1611-1613, 1991.

[5] E.A. Fitzgerald, Y.-H. Xie, D. Monroe, P.J. Silverman, J.-M. Kuo, A.R. Kortan, F.A. Thiel, B.E. Weir, and L.C. Feldman, "Relaxed Ge_xSi_{1-x} Structures for III-V Integration with Si and High Mobility Two-Dimensional Electron Gases in Si and High Mobility Two Dimensional Electron Gases in Si," J. Vac. Sci. Technol. B, Vol. 10, p.1807, 1992.

[6] F. Schaffler, D. Tobben, H.J. Herzog, G. Abstreiter, and B. Hollander, "High Electron Mobility Si/SiGe Heterstructures: Influence of the Relaxed SiGe Buffer Layer," Semicond. Sci. Technol., Vol. 7, p. 260, 1992.

[7]   D. K. Nayak, J. C. S. Woo, J. S. Park, K. L. Wang, K. P. MacWilliams, "High-mobility p-channel metal-oxide-semiconductor field-effect transistor on strained Si," Appl. Phys. Lett., vol. 62, no. 22, pp. 2853-2855, 1993.

[8]   J. Welser, J. L. Hoyt, J. F. Gibbons, "Temperature and scaling behavior of strained-Si N-MOSFET's," IEEE Trans. Electron Dev., vol. 40, no. 11, p. 2101, 1993.

[9]   J.A. Carlin, S.A. Ringel, E.A. Fitzgerald, and M.T. Bulsara, "High Quality GaAs Growth by MBE on Si using GeSi Buffers and Prospects for Space Photovoltaics," Progress in Photovoltaics, Vol. 8, p.323, 2000; S.A. Ringel, J.A. Carlin, J.J. Boeckl, R.J. Kaplar, and E.A. Fitzgerald, "High Lifetime GaAs on Si using GeSi Buffers and its Potential for Space Photovoltaics," Solar Energy Materials and Solar Cells, Vol. 66, p. 621, 2001.

[10]   J.A. Carlin, S.A. Ringel, E.A. Fitzgerald, M.T. Bulsara, and B.M. Keyes, "Impact of GaAs Buffer Thickness on Electronic Quality of GaAs Grown on Graded Ge/GeSi/Si Substrates", Appl. Phys. Lett., Vol. 76, p. 1884, 2000.

[11]   V.K. Yang, M.E. Groenert, S.M. Ting, G. Taraschi, C.W. Leitz, A.J. Pitera, M.T. Currie, Z.-Y. Cheng, and E. A. Fitzgerald, "Monolithic Integration of III-V Optical Links on Si using SiGe Virtual Substrates," submitted for publication, 2001.

[12]   M.T. Currie, C.W. Leitz, T.A. Langdo, G. Taraschi, D.A. Antoniadis, and E.A. Fitzgerald, "Carrier Mobilities and Process Stability of Strained Si n- and p-MOSFETs on SiGe Virtual Substrates," J. Vac. Sci. Technol. B, Vol. 19, No. 6, 2001.

[13] C.W. Leitz, to be published

[14] M.L. Lee, C.W. Leitz, Z. Cheng, A. Pitera, T. Langdo, M.T. Currie, G. Taraschi, E.A. Fitzgerald, and D.A. Antoniadis, "Strained Ge Channel p-type Metal-Oxide-Semiconductor Field-Effect Transistors Grown on SiGe/Si Virtual Substrates," Appl. Phys. Lett., Vol. 79, No. 19, 2001.

performance. Large concentrations of Ge introduced on a Si wafer require relaxed SiGe, which can host high performance Si CMOS and optical devices.
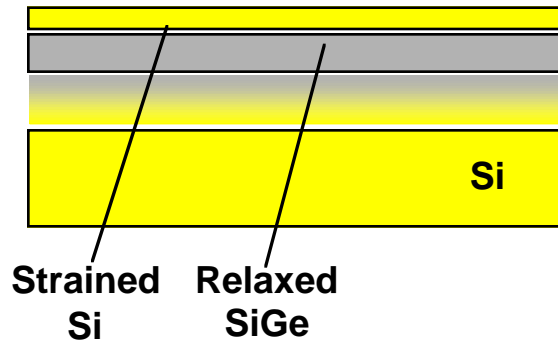

Fig. 2.   Schematic of basic structure for strained Si MOSFETs. A relaxed SiGe layer on Si is produced, followed by a final silicon layer, which is under tensile strain due to the relaxed SiGe.
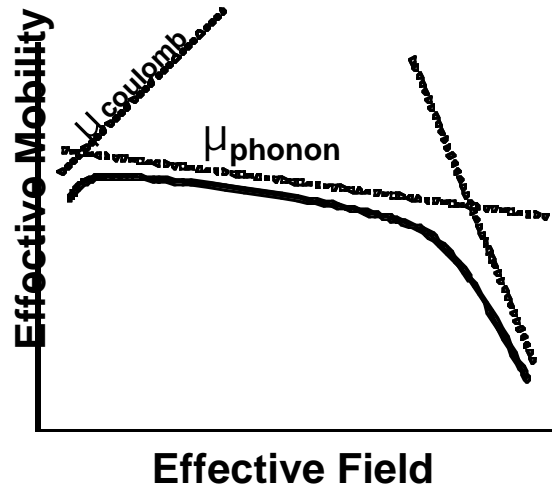

Fig. 3.   Diagram indicating the effective mobility of carriers in a MOSFET channel as a function of the vertical field.   As devices are scaled to smaller dimension, the effect is that the vertical field experienced by carriers is increased.   Thus, each generation of Si technology decreases effective mobility in the channel.
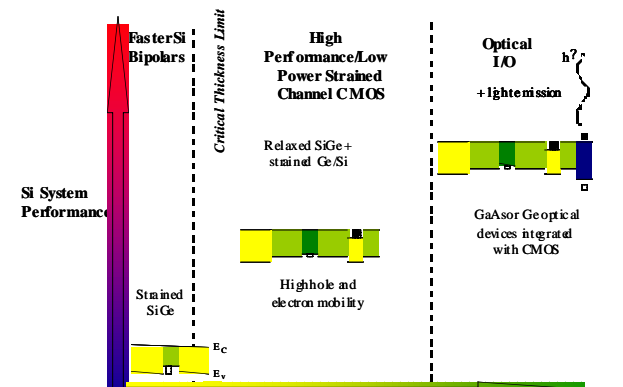

Fig. 1. Diagram indicating the qualitative relationship between Ge concentration and Si CMOS system
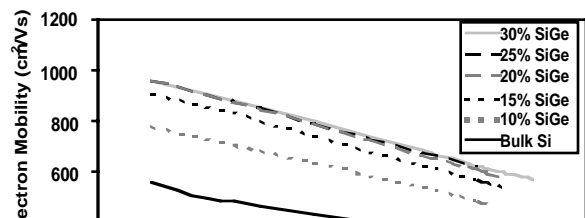
Fig. 4. Plot of effective mobility of electrons extracted from a large strained Si NMOSFETs. The curves represent the Si control (Bullk Si), as well as the mobility measured in MOSFETs with different Ge concentrations in the relaxed SiGe buffer. The increase in electron mobility saturates at a Ge concentration of approximately 20%, and the mobility enhancement is independent of vertical field.
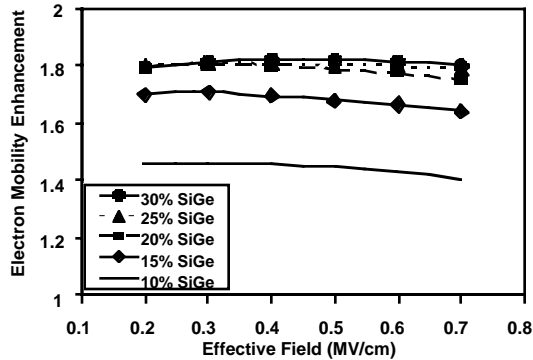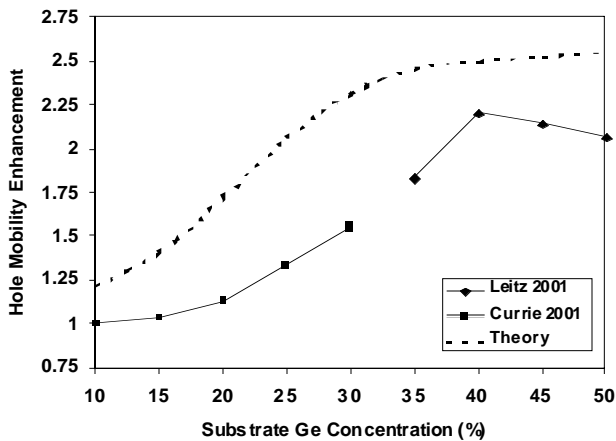
Fig. 6. Plot of hole mobility enhancement in strained Si PMOSFETs vs. Ge concentration in the relaxed SiGe buffer. A maximum enhancement of more than 2 is seen in MOSFETs with a Ge concentration >40%. The general shape of the data is in agreement with theory. The decrease in enhancement above 40% is likely due to the introduction of misfit dislocations into the channel.

Fig. 5. Plot of electron mobility enhancement as a function of vertical field.
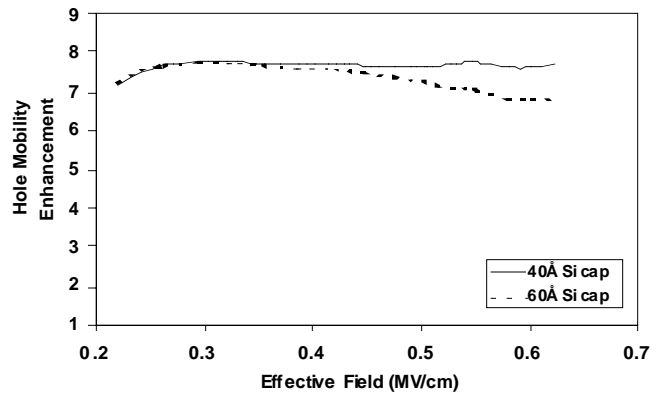
Fig. 7. Plot of hole mobility enhancement in a pure Ge compressive channel MOSFET vs. vertical electric field. The mobility enhancement nears 8 and is independent of vertical field for thin Si cap layers.
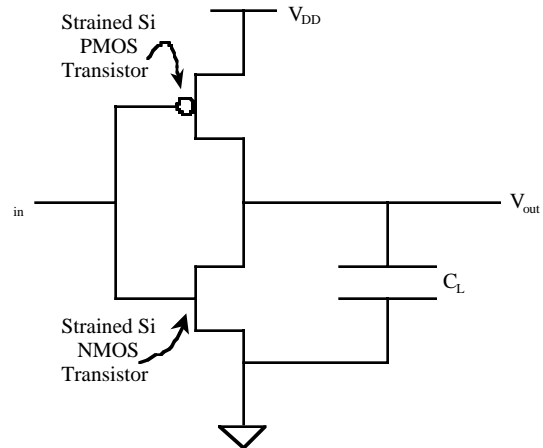
Fig. 8. Schematic of a strained Si CMOS inverter. As the strained layers increase the effective mobility in the channels of both NMOS and PMOS, the inverter delay decreases. Designing initial epitaxial structures to support a balanced enhancement over generations of strained Si CMOS technologies is important in maintaining the expected performance enhancements.
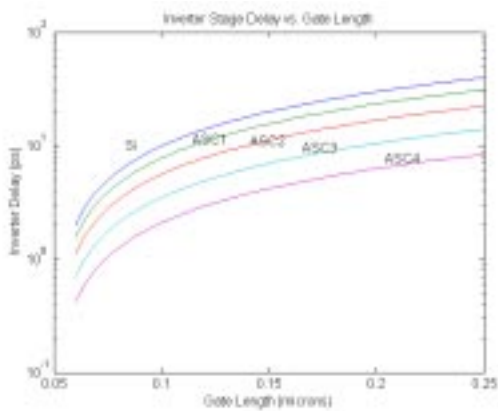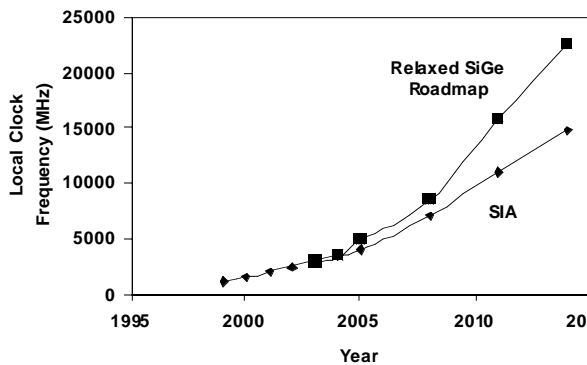

Fig. 9. A plot of possible inverter delays vs. gate length for various epitaxial strained channel structures. In general, as Ge content in the relaxed buffer is increased, performance increases. All curves keep the W/L ratio the same except for the curve labelled 'ASC4'. In this case, the W/L was adjusted to maximize performance.

]

frequency vs. time, based on the SIA roadmap and the enhancement potential of the relaxed SiGe platform.
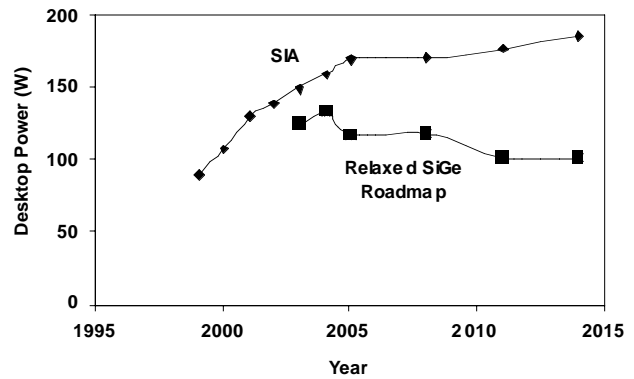

Fig. 11. Plot of expected power reduction in desktop processor based on keeping the transistor drive current the same but lowering the voltage.
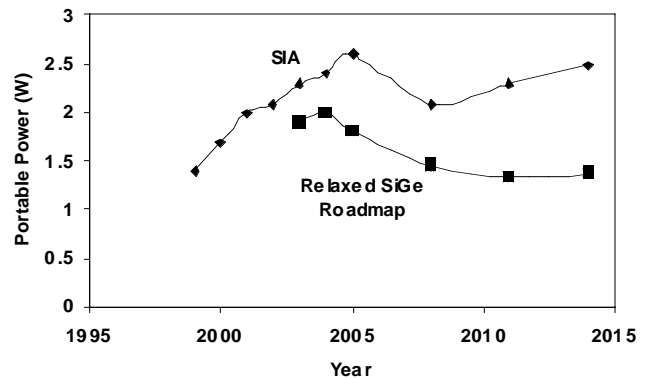

Fig. 12. Plot of expected power reduction in portable computing based on constant drive current but lower operating voltage.
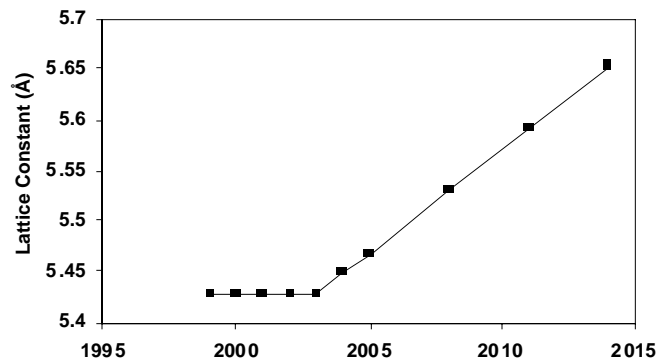
Fig. 13. Assuming the increase in performance shown in Figs. 10-12, this plot shows the necessary rate at which the lattice constant must increase vs. time. Dimension of the MOSFET is no longer the relevant parameter.