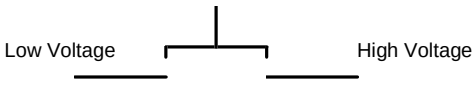
SOI Technology: IBM's Next Advance In Chip Design

I- Introduction

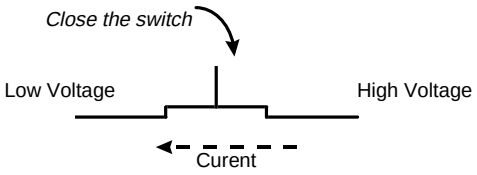
As with IBM's leadership manufacturing microchips using copper interconnect technology, the company has now announced what it believes to be the first commercially-viable implementation of silicon-on-insulator (SOI). For more than three decades, scientists have been searching for a way to enhance existing silicon technology to speed computer performance. This new IBM success in harnessing SOI technology will result in faster computer chips that also require less power -- a key requirement for extending the battery life of small, hand-held devices that will be pervasive in the future. SOI is a major breakthrough because it advances chip manufacturing one-to two-years years ahead of conventional bulk silicon. The following provides a step-by-step look at the developments leading up to the development of SOI technology including: a basic description of what a transistor is, a description of how chips are made, the need for improving the way chips are made, the elements of SOI technology and the impact of SOI technology.

II- Transistors And Chips

What is a transistor? A transistor is the electronic version of a switch. When it is on, it allows current to flow, and when it is off, it will stop current from flowing. It is exactly the same as the everyday switch that we use to turn lights on and off, except instead of using our finger, the switch is turned on and off by electricity.



Switch is Open



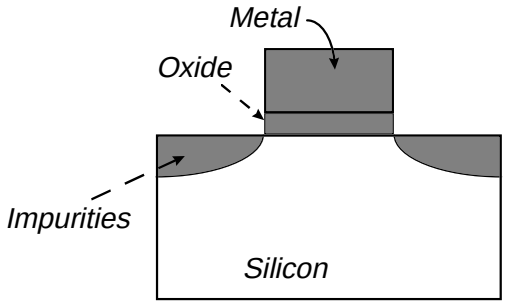
Switch is Closed

The switch that is used to turn a light on and off is very big, very slow, and carries a lot of current. The transistor or "electronic switch" is usually very small, very fast, and carries less current. There are two basic type of transistors or electronic "switches" -- bipolar and MOS or metal-oxide-semiconductor.

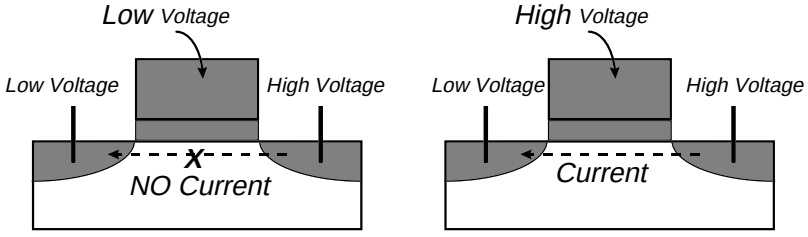
Bipolar switches were widely used from the 1960's to the 1980's in computer chips for mainframe computers because performance was the key factor. Bipolar chips usually consume more power, are bigger, and are harder to manufacture but were faster than MOS chips until about 5 years ago. Bipolar switches are gradually being eliminated from use in chips and IBM, in fact, has stopped making mainframe computers with bipolar transistors in them.

MOS transistors are used in the vast majority of semiconductors. MOS transistors are smaller, faster, consume less power, and cheaper to manufacture. IBM just announced its G5 family of mainframe computers containing MOS transistors that are as fast as the fastest mainframes containing bipolar transistors. The advantage to MOS technology is the chips can be built a lot smaller and are cheaper to manufacture. Next generation MOS mainframes will be faster than any bipolar mainframes. MOS switches are currently used in making IBM servers and almost all PCs and electronic products on the market.

What does an MOS transistor look like? MOS transistors actually look very much like a switch: they consist of a piece of metal on an oxide that is placed on a piece of silicon (semiconductor). Pure silicon does not conduct electricity. But if enough of the right kind of chemical "impurities" are added to the silicon, it will conduct electricity.



On the piece of silicon that is on either side of the metal, known as a "gate" area, chemical "impurities" are added to make the silicon conduct electricity. If current is transmitted to the metal, the silicon will conduct electricity. The "impure" silicon will be "attached" to the two conducting areas on either side of the metal and electricity will flow. Silicon, then, can act both as a conductor and a non-conductor, which is why it is called a "semi-conductor".



Switch Open (i.e. no current) Switch closed (current)

What is a chip? Chips are basically made of thousands to millions of transistors packed into a small piece of silicon. The transistors are wired together to perform a specific function. Traditional wiring uses aluminum. In chips made using more advanced technology, they are wired together using copper. The number of transistors on a chip, the speed of each transistor, and the delay passing electricity through each transistor and each transistor's metal interconnect determine how fast the entire chip operates.

Three of the most common kind of chips include: microprocessors, memory and logic chips. Microprocessors, the "brains" of a computer or other electronic device, carry out most calculations and operations. Memory chips are used to store data and computer programs that are executed. There are usually two kinds of memory found in a computer: static memory (SRAM), which is fast, but not as dense and dynamic memory (DRAM), which, is slower but much denser. The forth kind of chip on the computer is a collection of logic chips: these are the chips that control the operation of the "bus" on the computer, the disk drive, and many other operations.

Since the speed and the number of transistors on a chip are critical in determining the performance of computers, most leading edge microprocessors and dynamic memory chips on the market today usually use the most advanced technology available when they are manufactured. It is fair to assume that any leading chip technology under development today will appear in the most advanced computer microprocessors within 1-2 years. The copper metal technology, announced by IBM last year, is currently being shipped to customers.

What is CMOS? CMOS stands for Complementary Metal Oxide Semiconductor. In the traditional MOS switch, if "high" voltage is applied to the metal "gate", it will conduct electricity. If low voltage is applied, then the switch will be open and current flow will stop. Up to 15 years ago, only this traditional switch was in use in chips. Since then, another kind of MOS switch has been developed. This second kind of switch, operates with the opposite polarity charge of a traditional MOS, "complementing" the traditional MOS switch. In this switch, if a low voltage is applied to the metal "gate", the switch will close and current will flow. Likewise, if high voltage is applied, then it will open and the current will stop. Almost all microprocessor, memory and support chips use both kinds of MOS switch, and thus are CMOS-based.

III- Miniaturization And Performance

What is scaling? Scaling simply refers to the drive to continue making the transistor, and the technology that connects transistors together, smaller. As a transistor becomes smaller, it becomes faster and can conduct more electricity. It also consumes less power when it switches on or off or when electricity passes through it. Finally, the smaller a transistor becomes, the cost of producing each transistor goes down and more of them can be packed on a chip. It is worth mentioning that the critical dimension of a MOS switch, first introduced in the IBM PC 15 years ago, has gone down from 10 microns to 0.2 microns in leading chips today. In that time, a chip's electrical frequency rate has also increased by more than 50 times, and the number of the transistors on the chip has gone up by a factor of 20.

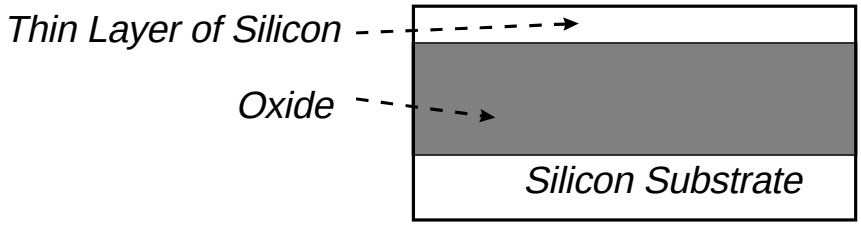
The single, biggest driving force behind growth in the semiconductor, computer, networking, consumer electronic and software industries in the last half century has been the continuous scaling or miniaturization of the transistor. By making the transistor smaller, faster, lower powered and cheaper, more of them can be packed on a chip so it performs more functions faster and at lower power. Computers and other electronic devices have become smaller, more portable, cheaper, easier to use, and more accessible to everyone. As long as we can make the transistor faster and smaller, make the wiring that attaches them less resistive to electrical current and make each chip denser, the digital revolution will continue to accelerate into the 21st century.

Why the need for faster performance? The increased speed and capability of computers has had enormous impact on our society. The Internet, advanced software applications, speech recognition, advances in telecommunications and the many services that we take for granted all are made feasible only because the performance of microchips has increased over time. And judging from what is in development in the research labs of many companies and universities, the increased performance of chips will continue to fundamentally change the way we live in the future.

How do we make chips faster? A number of techniques can be used to make chips faster. One way is to make the transistor smaller. But as transistors are made smaller and faster, delays through the wiring also become more frequent, which limits the speed of the transistor. Copper wiring, first introduced by IBM, aims to reduce the wiring delay of the chip. Another technique to speed up chips is to use an alternative "faster" semiconductor.

IV-Silicon on Insulator

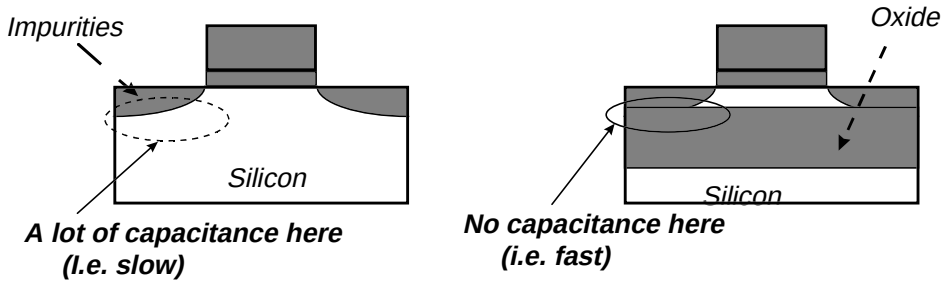
What is silicon-on-insulator? One faster semiconductor substrate that has been under active consideration for the last 30 years has been silicon-on-insulator (SOI). SOI refers to placing a thin layer of silicon on top of an insulator such as silicon oxide or glass. The transistors would then be built on top of this thin layer of SOI. The basic idea is that the SOI layer will reduce the capacitance of the switch, so it will operate faster.



What is capacitance? Capacitance is the ability of a structure to store electrical charge. Every structure that can pass electricity through it has capacitance associated with it. The MOS transistor or switch is one such structure. Every time the transistor is turned on, it must first charge all its internal capacitance before it can begin to conduct electricity. It takes a relatively long time to charge and discharge compared to the actual time that it takes to turn the transistor on and off. Developing a way to reduce the internal capacitance of a transistor, it would cause it to operate much faster.

How Does SOI Reduce Capacitance? One of the areas that can store charge in a MOS switch is the area between the impurities added to a chip's silicon and the silicon substrate itself, which is free of impurities. That area is called the "junction capacitance". If a thin layer of an insulator, such as glass, is placed between the impurities and the silicon substrate, the junction capacitance will be eliminated and the MOS transistor will operate faster.

This is the basic reason why many research groups have been trying to develop SOI technology. SOI technology results in a faster MOS transistor. The ultimate goal has been to use SOI as the substrate for mainstream CMOS technology used in the manufacturing of microprocessor chips that power computers and other emerging electronic devices. This is why IBM's announcement is so significant. While others, including IBM, have been successful in developing SOI technology, IBM is the first to be able to apply it in building fully functional mainstream microprocessors, the most complex type of chip.

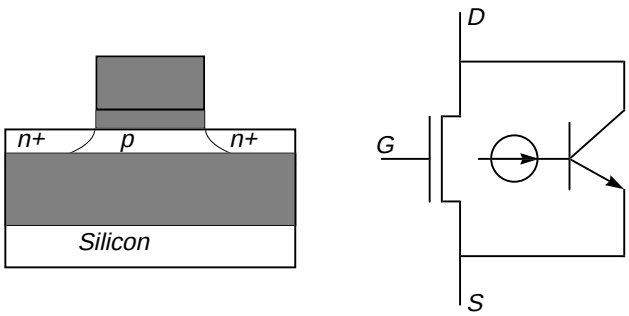


V- Overcoming SOI's Hurdles

What have been some of the challenges in developing SOI? The silicon film used for making MOS transistors is perfect crystalline silicon. The insulator oxide layer on the other hand is not crystalline. It is very difficult to make perfect crystalline silicon on oxide or silicon with other insulators since the insulator layer's crystalline properties are so different from pure silicon. If perfect crystalline silicon is not obtained, then defects will find their way onto the SOI film. This could degrade the MOS switch severely both in terms of speed and leakage through the switch.

One of the major areas of research has been trying to obtain "perfect" SOI layers. One insulator found to be perfectly crystalline and whose properties are close to those of silicon is sapphire. While silicon-on-sapphire has been one of the SOI layers actively investigated, SOI film utilizing it has still been found to be defective and silicon-on-sapphire was never able to be successfully used as an insulating material.

Even in cases where the quality of the SOI layer has been fine for making single chips or small circuits, the MOS transistor placed on the SOI layer has continued to pose serious challenges. When a MOS transistor is placed on SOI, it is also placed next to a parallel bipolar transistor. The bipolar transistor can turn on when the MOS switch passes current through itself, i.e. base current will be supplied through the base impact ionization. This reaction of the bipolar transistor to current passing through the nearby MOS transistor gives rise to a number of unwanted effects that complicate



MOS Transistor on SOI *Equivalent Devices (MOS & Bipolar)*

the operation of the chip's circuit and its modeling. Significant effort has been devoted in the last three decades in characterizing, modeling, eliminating or at least reducing, and controlling the unwanted effects caused by the bipolar device.

One solution to the bipolar transistor dilemma that has been under strong consideration over the last two-to-three years was to use very thin layer SOI films (less than 0.1 micron) called "fully-depleted films." IBM's approach has been different, and it uses slightly thicker films (greater than 0.15 micron), called "partially-depleted films."

Perhaps one of the biggest obstacles facing adoption of SOI as a mainstream technology has been steady progress in traditional bulk CMOS technology over the years. While much attention has been devoted to perfecting SOI, silicon bulk technology has shown consistent advances.

VI- IBM's SOI Approach

Why didn't IBM give up on SOI? At its core, IBM is a technology company. As a leader in the information technology industry, IBM is committed to developing the best technology for use in its products.

Leading edge silicon technology is a major element of IBM's leadership. The company's investment in silicon technology has resulted in major breakthroughs. IBM has led the development of one-transistor cell memory, quarter-micron CMOS technology and the use of copper to replace aluminum in producing faster, better-performing, lower-power chips. Today, SOI is a key component of IBM's goal to provide customers the best CMOS technology in the industry.

IBM's Microelectronics Division first began looking at SOI in the early 1970's. As part of that pioneering work, IBM first developed the "bonded" method of making SOI films. Initially discontinued in the mid-1970's, the team responsible for developing IBM's leading edge bulk silicon technology at the company's Research Division took a second look at SOI in 1989. A team at Research worked closely together with the bulk team and was instrumental in the rapid progress of SOI technology at IBM.

After five years of significant progress on SOI material, device, characterization, and some early circuit work, the decision was made to move the SOI technology development to a "manufacturing-like" development line at the Advanced Silicon Technology Center (ASTC) in East Fishkill, New York. move to ASTC was particularly important, since it placed SOI CMOS technology under the same rigor that the bulk technology was going through. The move to ASTC also allowed SOI technology access to a low-defect manufacturing line that was instrumental in SOI material development as well as many advanced microprocessor and memory designs, modeling and circuit design infrastructure, and methodology for evaluating reliability of SOI devices and circuits.

Today, SOI technology continues to go through the same qualification process as bulk CMOS technology, with the same rigorous attention to detail that is part of all technology development at IBM's Microelectronic Division.

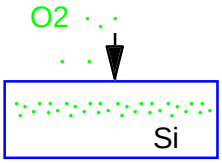
How did IBM finally make SOI a Feasible Alternative? A number of factors led to IBM's ability to make SOI a mainstream chip making technique. One important factor is the slowing of the rate of progress in CMOS technology, making the extra performance offered by SOI even more valuable. Secondly, designers will use whatever performance is given to them by the developers of a technology. A faster technology in use by designers translates into faster microchips and, ultimately, faster computers in the marketplace.

But probably the single most important factor in SOI becoming a mainstream technology is the breadth and quality of work done at IBM while developing it. The most important breakthrough, among many, was the recent demonstration of fully functional microprocessors and large static random access memory chips utilizing SOI. This work has convinced many skeptics that after three decades SOI is real technology that can be used to design the next generation of chips at profitable levels and achieve a two year performance gain over conventional bulk silicon technology.

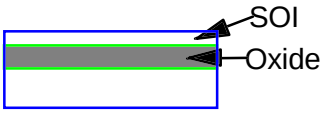
How is the SOI material made? There are a number of methods under development inside and outside IBM for making SOI material and wafers. Presently, IBM is manufacturing its own SOI wafers using what is known as the SIMOX or "Separation by Implantation of Oxygen" technique. In this method, oxygen is implanted in very heavy doses and the wafer is annealed at a high temperature until a thin layer of SOI film is formed. In addition, IBM continues to focus on new methods to reduce the SOI film defects.

During the course of this work, IBM is the only company that has provided evidence proving that there is no difference in yield between bulk and SOI wafers.

Implant Oxygen:



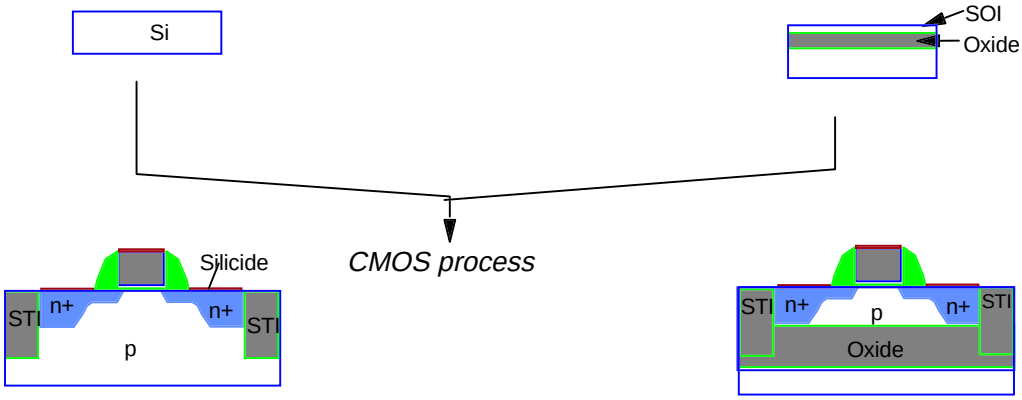
Anneal damage:

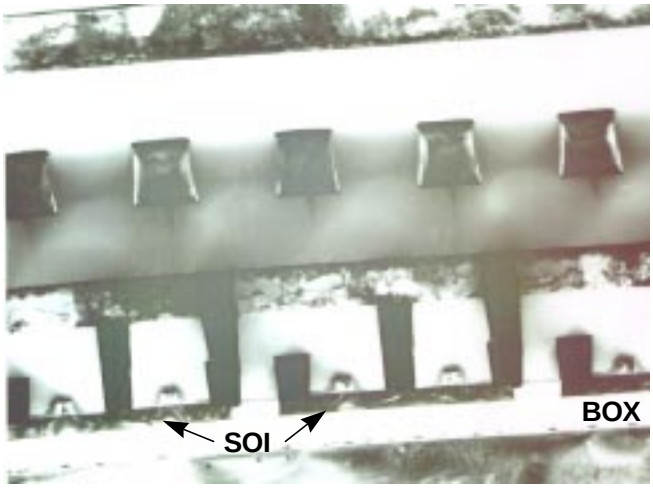


How is the SOI Transistor Applied To The Circuit? Once the SOI film is made, putting the transistor on the SOI film is straightforward. It will basically go through the same process as a similar bulk CMOS wafer. There are minor differences in the details of the process, but it uses the exact same lithography, tool set, and metalization. At IBM we are in the final stages of qualifying a 0.22 micron CMOS on SOI and at the initial stages of development of 0.15 micron CMOS on SOI technology.

Bulk CMOS:

SOI CMOS:

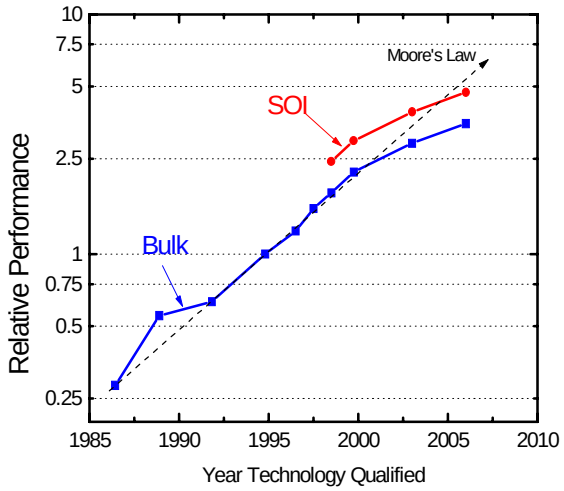


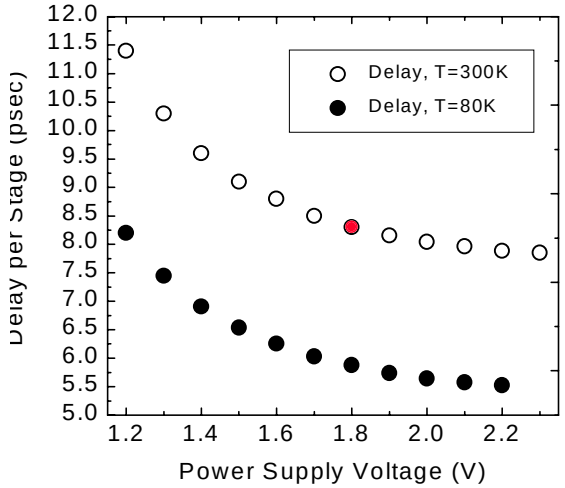


Additionally, when designing the SOI chip, IBM put significant effort to minimize the effect of the bipolar transistor in parallel with the MOS transistor. Minimizing the bipolar effect would make the SOI chip “bulk-like” and would allow circuit designers to reuse many familiar circuit techniques and macros that have been developed in bulk CMOS technology.

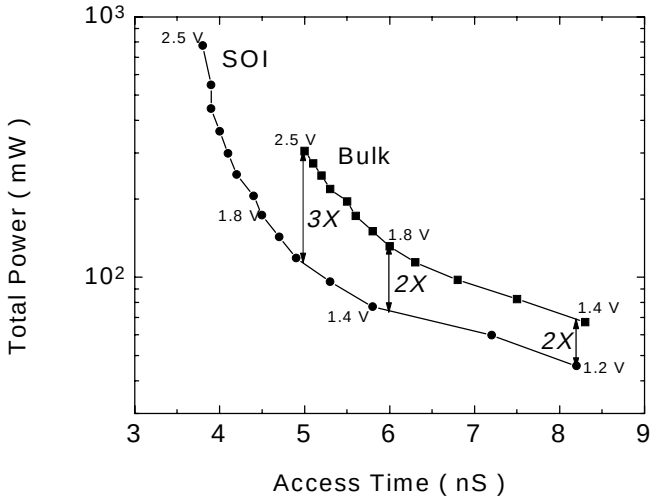
VII- Benefits of SOI

Performance: IBM has built and tested SOI-based chips that have produced 20-25% cycle time and 25-35% improvement over equivalent bulk CMOS technology. This is equivalent to about two years of progress in bulk CMOS technology. On Moore’s chart, SOI will cause a jump in the performance road-map, and will compensate for some of the expected loss of bulk technology performance improvement in the next few years. The sources of increased SOI performance are elimination of area junction capacitance and elimination of “body effect” in bulk CMOS technology. Body effect is an MOS bulk effect, which results in lower current, and lower performance, in bulk MOS transistors, if they are placed in a certain configuration (i.e. if they are stacked, of if the sources of the MOS devices are not grounded).





Low Power: Although this paper focuses on SOI performance issues, SOI has excellent capability as a low power technology. One disturbing trend over the last few years, as people have been using more advanced technologies, has been that the power of microprocessors has been going up. For example, the 486 chip of several years ago consumed less than five Watts, while a Pentium consumes about 10 Watts, and a 400-MHz Pentium II has peak power consumption of about 28 Watts. Increased power seriously limits the use of microprocessors, especially in mobile applications. IBM replaced bipolar transistors with MOS transistors due to the high consumption power of the bipolar transistors. Dropping the voltage is very effective in reducing chip power. The ability of SOI as a low power source originates from the fact that SOI circuits can operate at low voltage with the same performance as a bulk technology at high voltage. As ASIC libraries for SOI are developed, SOI will have a tremendous impact on applications where low power is needed, such as portable and wireless applications. As an example, measured on a 4 Mb SRAM. At the same performance as bulk CMOS, SOI can reduce the chip power by 1.7-3X (depending on the switching factor of the devices).



*Note: Performance characteristics displayed above are actual figures derived in lab tests using SOI-based chips

Soft Error Rate: One of the early benefits of SOI has been the reduction in soft-error rate. Soft error rate refers to upset of data in the memory by cosmic rays and background radioactive material. In fact one of the first early

applications of SOI has been in memories for space application, since the memories built on SOI were perceived to be more resistant to soft error rate. As chips are getting smaller and the voltages are dropped, soft error rate is slowly becoming a major concern in server and mainframe chips. As an example, computers that do not guard band for SER, have a higher failure rate if they are operated in high altitudes. IBM's early studies indicate that as CMOS is scaled into 0.18 micron range and voltages are dropped, SOI indeed has much lower soft error rate than bulk CMOS.

VIII- Summary

IBM has shown its chip technology leadership by becoming the first to implement copper interconnect technology. IBM's announcement of a mainstream SOI technology is equally significant. SOI technology improves performance over bulk CMOS technology by 25-35%, equivalent to two years of bulk CMOS advances. SOI technology also brings power use advantages of 1.7-3 times. IBM is currently working with many circuit designers and product groups that are designing with SOI technology. The company expects SOI will eventually replace bulk CMOS as the most commonly used substrate for advanced CMOS in mainstream microprocessors and other emerging wireless electronic devices requiring low power.

# # #