Silicon takes the strain

by Luke Collins, consultant, UK

The semiconductor industry is investigating a new way of making the transistors that are the basic building blocks for almost all integrated circuits (ICs).

This way promises to squeeze more device performance out of current manufacturing processes, or to get the same device performance using less power. The technique is known as 'strained silicon' and has become a hot topic in semiconductor research in the past two years. In 2003 eight papers on strained silicon were presented at the International Electron Devices Meeting, a key research conference. Last year there were 16.

"What strained silicon does is allow you to go faster without going smaller," said Dr. Rona Belford, president of strained silicon research company Belford Research. This may present some relief for an industry that is under pressure from the exponentially rising costs of its manufacturing processes.

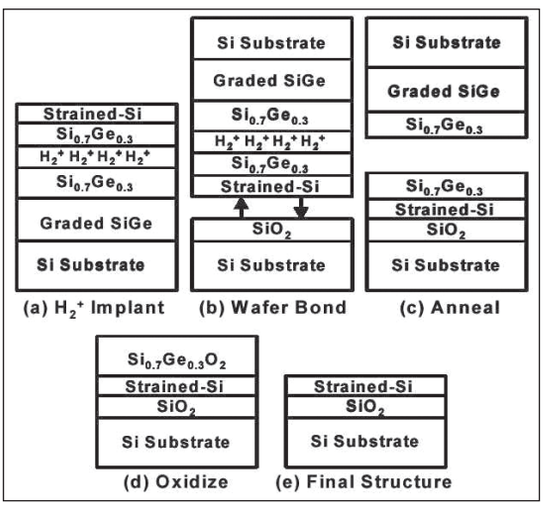

Fig. 1: SSOI fabrication process.

Channel changer

Strained silicon techniques are being applied to one of the most fundamental aspects of semiconductor technology, the conduction channel of the field effect transistors (FETs) used in CMOS IC processes. This channel sits between the source and drain of the FET, with a 'gate' electrode sitting over it controlling the conduction of the channel.

Two major factors control how fast a transistor can switch: the channel length from source to drain; and the speed at which charge carriers move through the semiconductor material. Improving lithographic technologies have shrunk channel lengths to the order of tens of nanometres. Strained silicon can improve the mobility of charge carriers in the channel, further boosting performance.

Dr. Maynk Bulsara, co-founder and chief technology officer of AmberWave Systems, a strained-silicon wafer technology development company, quantifies the improvements that strained silicon can deliver: "At the circuit level you can get a 34% reduction in power consumption at the same speed, or a 17% increase in speed at the same power."

Fearful symmetry

So how does strained silicon work?

Semiconductors have electronic properties that lie between insulators and conductors. Low-energy electrons in silicon form covalent bonds between the atoms in a lattice, leaving few electrons free to aid conduction. But electrons at higher energies are free to move through the lattice, existing in what is called the conduction band.

Pure silicon forms a crystalline structure that resembles a cube with an atom at each vertex. In trying to understand why straining the lattice eases the flow of charge carriers it is tempting to think of electrons as 'bullets' and the atomic lattice as a grid of 'particles'. In this model, straining the silicon moves the 'particles' apart and so makes it easier for the 'bullets' to pass through.

According to Ken Rim, a member of the research staff at strained silicon pioneers IBM Research, this is a flawed model. Strained silicon needs to be understood through quantum mechanics, which says that electrons exist at discrete energy levels and in a number of states at each level.

He says that in unstrained silicon, there are six discrete states in which electrons can exist at the lowest energy level in the conduction band. Each state has a different momentum and electrons can move between them with a small input of energy, for example from a quantised packet of acoustic or thermal energy (a phonon) given off by the lattice.

When the lattice is strained, its physical symmetry is broken and with it the electronic symmetry. The lowest energy level of the conduction band is split, with two of the six

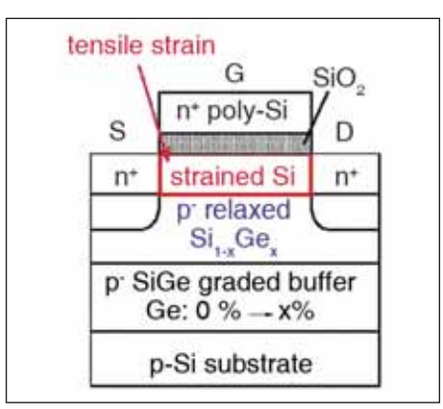

Fig. 2: Typical structure of strained-Si MOSFET.

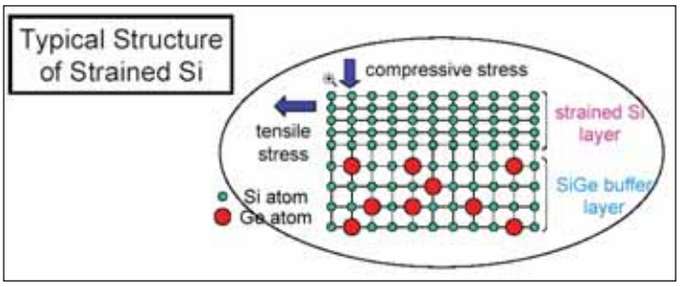

Fig. 3: Typical structure of strained Si.

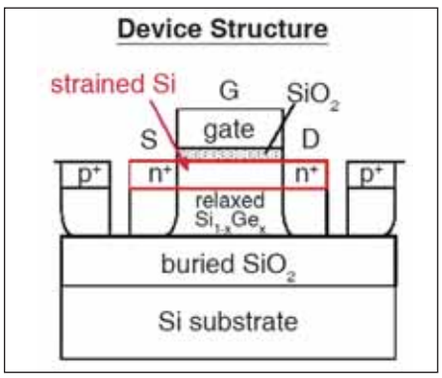

Fig. 4: Features of strained SOI MOSFET.

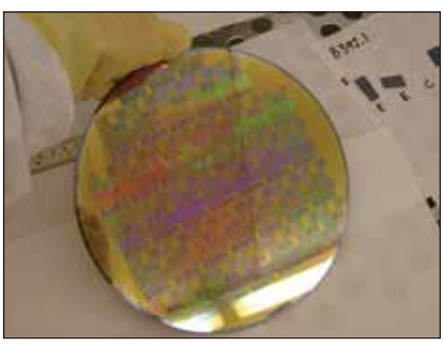

Fig. 5: AmberWave strained silicon wafer.

original states dropping to a lower energy level and four rising to a higher energy level. Now it is much more difficult for electrons to be 'scattered' between the lowest energy states by a phonon, because there are only two states to occupy.

"Whenever electrons scatter it randomises their motion," said Rim. "It's like a resistance which keeps electrons from maintaining a momentum in one direction." Reducing scatter increases the average distance an electron can travel before it is knocked off course, so increasing its velocity in the conduction direction.

The second quantum-mechanical effect that enhances conduction in strained silicon is a reduction of the electron's 'effective mass', a measure of how much it will accelerate in a given field.

"In solid structures there's an interplay between the crystalline structure of the lattice and the nature of the electron," said Rim. Distorting the lattice through strain can distort the interaction between it and electrons in a way that reduces their effective mass.

The upshot of straining silicon, then, is to increase the average distance that an electron travels before it is knocked off course and to reduce its effective mass so it is accelerated more by a given field.

Two-way stretch

There are a number of ways to induce strain in silicon. The key is to make the process repeatable, compatible with existing manufacturing steps and able to withstand any thermal cycles they demand.

There are various types of strain, each with different effects. Strain can be applied in one, two or three dimensions. Biaxial strain is widely used to strain lattices in the xy plane of conduction, causing a related negative strain (compression) in the z direction. Some researchers are working on 'uniaxial' strain, where the majority strain is in one direction and the other two dimensions adjust to match. Stretching or compressing silicon induces different effects, as does the relationship between the direction of strain and of conduction.

"I strain silicon uniaxially and have found a great difference in the preferential mobility along, or at right angles to, the electron travel direction," said Belford.

Belford Research is working on a number of ways of mechanically inducing strain in thin silicon layers. In one approach devices are made on a standard wafer, and then the device layer is separated from the bulk material, thinned down to a membrane thickness and glued to another substrate. That substrate is then bent to induce very large mechanical strains in the thin silicon layer attached to it. In another approach a wafer is heated until it expands slightly and then bonded to a cold wafer. When the two cool, the size mismatch induces strain.

AmberWave Systems is researching ways of making strained silicon wafers that can be used in standard wafer fabs. It aims to develop its technology to the point where it can be licensed to a big wafer producer or a chip maker. It is already working with AMD, which makes x86 processors, and UMC, the world's second largest foundry.

The AmberWave approach relies on the fact that the lattice of a silicon-germanium (SiGe) alloy is slightly larger that that of pure silicon. So AmberWave starts with a silicon wafer and grows a SiGe crystal lattice on top of it. At the interface between the two the concentration of germanium (Ge) is kept small so the lattices match. As more layers of SiGe are laid down, the Ge concentration is increased and dislocations are induced into the lattice to ease the formation of the larger SiGe lattice.

Eventually a graded SiGe layer with a lattice constant around 1% greater than that of Si has been formed. A new 'relaxed' layer of SiGe is grown on top of this graded layer to terminate any of its dislocations and 'set' its lattice constant. The top surface is then highly polished to allow a high-quality and very thin film of silicon to be grown. The resultant Si layer is strained because it continues the SiGe lattice spacing.

The strain that is induced in the Si layer can be adjusted by varying the amount of Ge used in the SiGe layer, and may need to be greater to improve hole mobility than it is to improve electron mobility.

Rim said: "Strain has a greater effect on electrons because it changes the band splitting with a greater effect. With enough strain, hole performance can exceed that of electrons."

He says a 20% Ge concentration will cause around a 0,8% strain, which will improve electron mobility. Going to a 35 – 40% Ge concentration will induce a 1,2% strain in the Si, which will start to have a stronger effect on hole mobility.

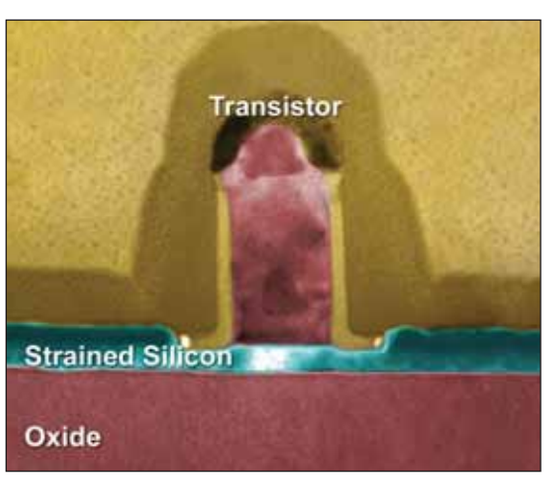

Fig. 6: IBM's strained silicon on insulator transistor.

Feeling isolated

Many researchers are now trying to combine the performance gains of strained silicon with the advantages of building devices on insulating substrates – so- called silicon on insulator techniques. SOI reduces the capacitance between the source and drain and the bulk material under it, allowing devices to switch faster or at the same speed for less energy. It also cuts current leakage.

Both AmberWave and IBM Research are working on a technique to produce strained silicon on insulator (SSOI) wafers. A strained silicon wafer is produced using the usual process, and then hydrogen is implanted into the relaxed SiGe layer at a uniform depth to form a shear plane.

A second silicon wafer with a layer of oxide on top is bonded face to face with the strained silicon wafer. The wafers are then separated by annealing them so the first wafer splits along the hydrogen bond layer. The second wafer then carries a layer of relaxed SiGe over the strained Si layer, which sits on an insulating oxide layer. Polishing the SiGe layer leaves an SSOI wafer. AmberWave's early results suggest the electron mobility in SSOI will be 40 – 50% greater than in standard SOI.

Getting real

Strained silicon has been seen as offering intriguing possibilities for future improvements in standard CMOS processes. But last year Intel shocked the strained silicon research community by announcing it would apply strained silicon techniques to its next generation of production microprocessors. It used last December's International Electron Devices Meeting to reveal how.

Intel's process does not rely on strained silicon wafers. Instead it uses a standard wafer and induces strain within each device through clever process tricks. For pMOS transistors Intel selectively deposits SiGe in the source and drain regions of the device, which create a uniaxial compressive strain along the channel. This results in what Intel calls a 'dramatic' increase in hole mobility of more than 50%.

For nMOS devices, Intel has applied what it describes as a "highly tensile" silicon nitride capping layer to the whole transistor. The stress from this layer is transferred to the nMOS channel through the source and drain regions to create a tensile strain in the channel. This improves the device's performance by 10%. Although the process allows separate optimisations for p- and n-type devices, strained silicon researchers have their concerns about its manufacturability.

Rim at IBM Research said: "Process-induced techniques are much more problematic to control because you are introducing strain individually, so the impact of process variation can be dramatic. Very small variations can create control issues and performance variations."

Belford is concerned about the thermal cycling performance of the nitride capping layer used to strain the nMOS devices: "They'll always have problems with [cycling] things again and again."

But Intel remains sanguine. In its IEDM paper it says the techniques are being incorporated in the manufacturing technology it is now ramping into high volume for its next generation of microprocessors.

It seems an industry that has been under pressure for the past three years is now learning to live with strain as well.

Acknowledgement

The original version of this article was published in the December 2003 issue of IEE Review, the members' magazine of the Institution of Electrical Engineers (www.iee.org) and is republished with permission.

Contact Luke Collins,
lukecollins1@compuserve.com □