Chemical Coat Helps Semiconductor Prospects

An irksome feature of the surface of gallium arsenide limits the performance of transistors in microcircuits, but new chemical treatments offer hope for relief

Although the semiconductor gallium arsenide is superior to silicon in several respects, it has the misfortune of having no native oxide that can uniformly coat its surface with an insulating layer while protecting against the formation of electrically active imperfections. In contrast, the marvelous match between silicon and silicon dioxide greatly facilitates the multistage process by which microcircuits are manufactured layer by layer. The lack of a comparable oxide does not by any means preclude gallium arsenide microcircuits, but ways around the problem either add to the manufacturing cost, thereby making the material less competitive for anything but specialized applications, or result in transistors with some less than optimal characteristics.

There is hope for doing better, however. New results from the IBM Yorktown Heights Laboratory and from Bell Communications Research (Bellcore), while short of a cure, show that the surface of gallium arsenide can be either photochemically or chemically treated to remain free of defects, at least for a short time. So far, the findings have not been translated into an improved manufacturing process, but they do give some hints on how to go about doing so. For example, neither the IBM treatment nor the Bellcore treatment generates good insulating layers on the gallium arsenide surface, but they suggest that any future process must both remove excess arsenic from the surface and tie up the surface bonds to prevent oxidation processes that generate the arsenic.

The problem with the gallium arsenide surface is well understood, but its origin has been a controversial subject for many years. In a semiconductor, there is a gap between two energy bands of electron quantum states. The lower energy valence band is almost completely filled with electrons, whereas the higher energy conduction band is mainly unoccupied. Only electrons in the conduction band and electron vacancies or holes in the valence band can carry current in the semiconductor. And the numbers of electrons and holes, which need not be identical, are determined by the electrochemical potential for electrons (the Fermi energy), which usually lies in the energy gap.

The next step is to find ways that are also compatible with the requirements of microcircuit technology.

In contrast to the silicon surface, the Fermi energy at the gallium arsenide surface is strongly "pinned" at a certain value toward the middle of the energy gap unless special measures are taken. To prevent pinning, the surface must be generated and kept in an ultrahigh vacuum or covered by a high-quality layer of another material in such a way that the atoms across the interface are in registry (epitaxial layer). Why Fermi energy pinning causes a difficulty is illustrated by the most popular type of silicon transistor, the metal-oxide-silicon field-effect transistor (MOSFET).

In the silicon MOSFET, current flows between the source and drain electrodes through a channel near the surface. A voltage applied to a gate electrode opens and closes the channel by changing the Fermi energy and the number of current carriers there. A thin silicon dioxide insulating layer separates the silicon channel from the metal gate electrode, hence the name of the device. Because the silicon-silicon dioxide interface is electrically inactive, the Fermi level there is free to respond to the gate voltage. This permits a small input signal to the gate electrode to turn a large output signal on and off, so that the MOSFET is both a switch and an amplifier. A similar structure built from gallium arsenide does not amplify a signal, however, because the pinned Fermi level does not move much when the gate voltage is changed.

Gallium arsenide microcircuits are a growing industry, so the surface problem is not fatal for many applications. One way around the difficulty is to remove the oxide layer between the gate electrode and the channel. The resulting transistor structure is a metal-semiconductor field-effect transistor or MESFET. An energy barrier (Schottky barrier) that forms between the gallium arsenide and the metal plays a similar role as the oxide insulator in the MOSFET. Pinning of the Fermi level at the interface between the gallium arsenide and the metal still occurs, however, and has the effect of fixing the height of the energy barrier at a much lower than desired value, a value that is the same for a wide variety of metals.

In a so-called enhancement mode MESFET, for example, the gate voltage to turn on the MESFET is the same as the height of the energy barrier. A low barrier means that the input signal to the gate electrode must be controlled very carefully to avoid accidentally turning the MESFET on or off at the wrong time. This imposes stringent requirements on the fabrication process to ensure that all the transistors on a chip have identical properties, making their manufacture more expensive. Moreover, circuits have to be designed to accommodate the lower voltages. "That's why digital circuit people love MOSFETs instead," says Jerry Woodall of IBM.

Why the Fermi energy is pinned at the gallium arsenide surface remains a controversial issue. Because the Fermi energy is always pinned at the same value, for a long time researchers assumed that an atomic imperfection intrinsic to the surface was responsible. In a material with many electrons in the conduction band, for example, each imperfection site would trap an electron at the surface, depleting the conduction band at the surface of electrons and moving the Fermi energy there toward the energy of the surface quantum state associated with the imperfection. (Quantum states whose energies are well above the Fermi energy are empty; those whose energies are well below the Fermi energy are occupied by electrons; and those whose energies are near the Fermi energy are partially occupied.)

It is no longer believed, however, that the Fermi level at the gallium arsenide surface is intrinsically pinned. For example, in the late 1970s, William Spicer of Stanford University and several co-workers suggested that the imperfection responsible for fixing the barrier height at metal–gallium arsenide junctions was introduced by the metal. The Stanford researchers reached their conclusion partly from the observation that a gallium arsenide surface prepared by cleaving a crystal in ultrahigh vacuum exhibited no signs of Fermi energy pinning until a considerable time had passed. Moreover, nowadays researchers are investigating a large number of "advanced" gallium arsenide transistor structures that use epitaxial layers of aluminum gallium arsenide for several purposes, including protection of the gallium arsenide surface. The epitaxial growth process, which is slow and expensive, may not be suitable for mass production, however.

Also in the late 1970s, Adam Heller and several colleagues at AT&T Bell Laboratories demonstrated that it was possible to move the surface Fermi energy of gallium arsenide away from the middle of the energy gap with a chemical treatment. The discovery turned up in the course of an investigation of photoelectrochemical solar cells involving gallium arsenide and selenium-based electrolytes. The investigators found they could substantially enhance the conversion efficiency of their cells from 9 to 12% by dipping the gallium arsenide into a ruthenium-containing solution immediately before inserting it into the electrolyte.

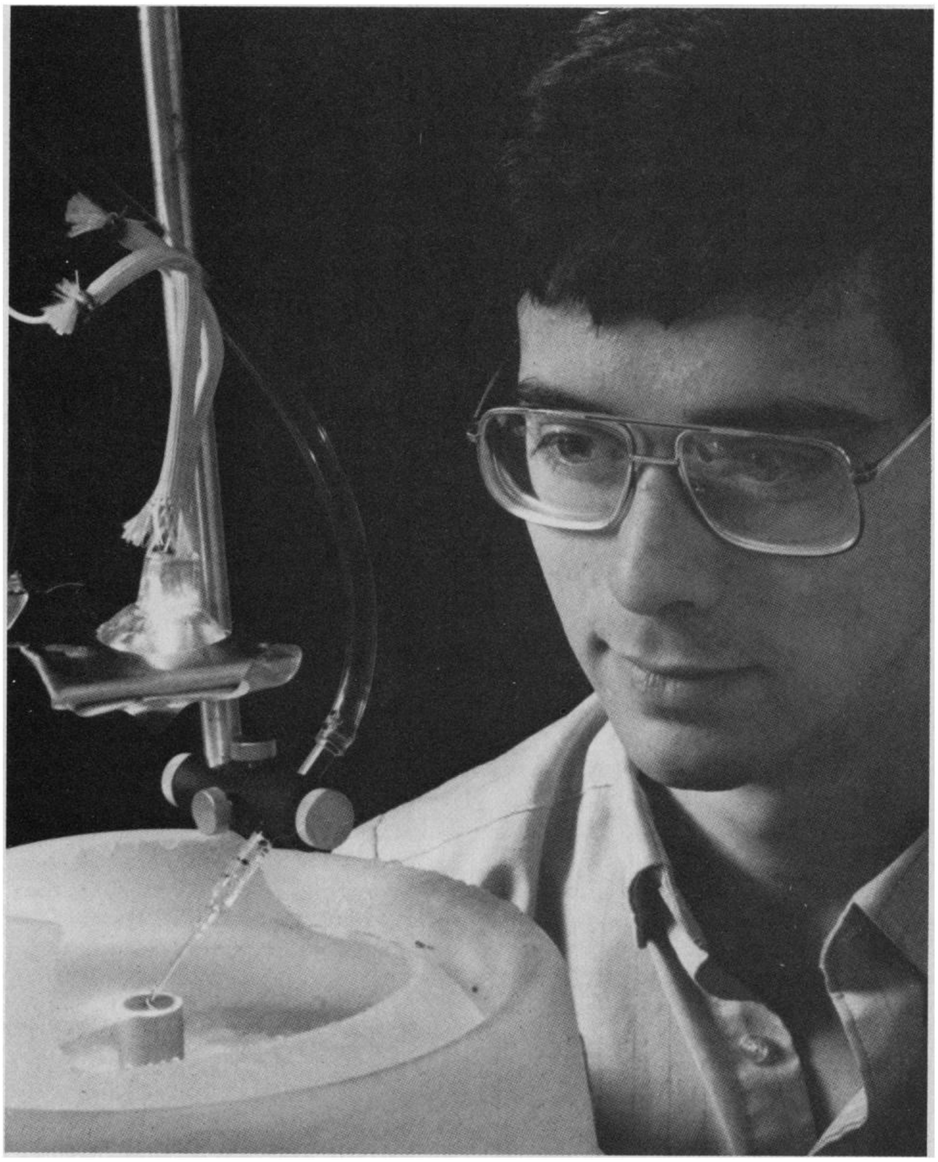
In a solar cell, Fermi energy pinning at the surface degrades the conversion efficiency by means of surface recombination. In this process, holes created by the absorption of light are rapidly trapped at surface quantum states and recombine with electrons there rather than migrating across the cell and contributing to the photovoltage. The effect of the ruthenium was therefore to reduce the surface recombination. Subsequent experiments specifically measured a surface recombination that was reduced by a factor of 10 over that of nonruthenium-treated gallium arsenide.

In retrospect, the Bell Labs experiments support a model for Fermi energy pinning that was introduced in 1981 by John Freeouf of IBM and Woodall. From a variety of published data on metal–gallium arsenide barriers, the researchers concluded that the Fermi energy pinning was due to microclusters of elemental arsenic on the surface. In this model, the pinning is not due to an atomic imperfection, intrinsic or extrinsic. Instead, the surface Fermi level is fixed by the comparative values of the work function (energy to remove an electron from the surface) of arsenic and the electron affinity of gallium arsenide. Nonetheless, there are surface quantum states associated with the presence of arsenic on gallium arsenide that cause surface recombination.

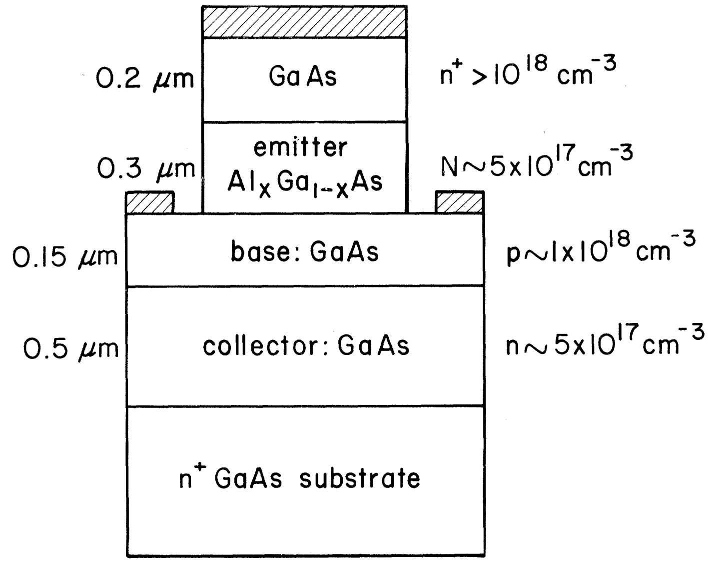
According to Heller, the present interpretation of the Bell Labs experiments is that the selenium solution dissolved the gallium oxide that forms on all gallium arsenide surfaces when the material is exposed to air or aqueous solutions. Arsenic oxide is unstable and reacts with gallium arsenide to form more gallium oxide and elemental arsenic. The ruthenium reacted with the arsenic to form a stable compound. With the absence of elemental arsenic, the surface Fermi energy was no longer at the middle of the energy gap. However, it was now pinned at an energy in the conduction band. Pinning at the conduction band does not hamper solar cell operation, but it is not suitable for MOSFETs.

Further experimental evidence for the extrinsic nature of Fermi energy pinning came from G. Horowitz of the Solar Photochemistry Laboratory in Thiais, France, and his associates who studied gallium arsenide in a variety of electrolytes. They found that the barrier height at the liquid–gallium arsenide interface varied strongly with a parameter that characterizes electrolytes called the redox potential. In other words, the Fermi energy was not fixed at one value.

The first unpinned gallium arsenide surface outside an ultrahigh vacuum environment was demonstrated by Woodall's group at IBM. Stephen Offsey (who was a visiting student from the Massachusetts Institute of Technology), Allan Warren, and their colleagues showed that it was possible to unpin the Fermi energy at the gallium arsenide surface for short times in air by a photochemical treatment. After cleaning the surface and mounting the sample on a standard microelectronics fixture called a photoresist spinner, the investigators exposed the surface to a stream of deionized water and light from a laser. In this experiment, the laser light served two purposes. It initiated a photochemical reaction at the gallium arsenide surface and excitated electrons from the valence to the conduction band. The fluorescence emitted when the photoexcited electrons and holes recombined directly served


IBM

Unpinning the surface. *Water spins onto the illuminated surface of gallium arsenide and washes away arsenic and arsenic oxide, leaving a gallium oxide–covered surface.*



Bell Communications Research

Bipolar transistor. *Surface recombination at the emitter-base interface that causes a low transistor gain is reduced by coating with a sodium sulfide layer.*

as an indirect measure of Fermi level pinning because the surface recombination associated with the pinning short-circuits the direct recombination and reduces the fluorescence intensity.

In fact, the IBM investigators observed a considerable increase in the fluorescence intensity from gallium arsenide subjected to their photochemical treatment. Moreover, they observed intensity increases for gallium arsenide containing either free electrons (*n*-type) or holes (*p*-type), indicating that the Fermi energy was truly unpinned and had not just moved away from the middle of the energy gap to another energy, as in the Bell Labs experiments. Unfortunately, the beneficial effects of the treatment were not long lasting in air, as indicated by a roughly exponential decay of the fluorescence effiency with a time constant of 20 to 30 minutes. However, the surface was stable in an inert enviroment.

The IBM explanation for these results is that, during photochemical oxidation, the water-soluble arsenic oxide and arsenic was washed from the surface, leaving a gallium oxide layer. But the gallium oxide does not fully protect against reoxidation of the surface. Reoxidation gradually generates more elemental arsenic by the two-step process invoked by Heller at Bell Labs, causing a repinning of the Fermi energy with time. Two independent groups at the Aerospace Corporation subsequently reported evidence consistent with this explanation.

More recent spectroscopy studies by Peter Kirchner and his IBM colleagues support the contention that the unpinned gallium arsenide surface is largely free of elemental arsenic and its oxides. A slightly modified procedure was used to unpin the surface Fermi energy. When cleaned and then sprayed with deionized water (see photo), the surface had a visually observable blue oxide film. Auger spectroscopy, which shows what elements are present, indicated only about 2% arsenic in any chemical form on surfaces treated in this way. X-ray photoelectron spectroscopy, which gives information about the chemical state of elements, suggested that the surface is 90% gallium oxide (Ga_2O_3). The gallium oxide film produced in this way was not a good insulator, however, so this treatment is not sufficient for the fabrication of gallium arsenide MOSFETs.

Last month, the Bellcore group reported that a strictly chemical technique may generate a gallium arsenide surface that is almost as free from Fermi energy pinning as the interface between gallium arsenide and aluminum gallium arsenide. The technique was developed from an idea of Bellcore's Claude Sandroff. Once again, surface recombination was used as the measure of Fermi energy pinning, but the investigators used a new apparatus developed by Bellcore's Eli Yablonovitch dubbed the laser-pumped carrier lifetime bridge. A pulse of laser light excites free electrons and holes in the gallium arsenide, which then decay by both direct and surface recombination. A radiofrequency apparatus monitors the change in the conductivity at the surface as the photoexcited carriers decay.

Yablonovitch, Sandroff, and their coworkers compared the surface recombination rates deduced by this means for gallium arsenide with an untreated surface, a photochemically treated surface, an aluminum gallium arsenide–covered surface, and a surface on which thin films of $Na_2S{\cdot}9H_2O$ had been deposited by spraying it on gallium arsenide mounted on a spinning table. They found that the surface recombination rate for the surface treated by the latter method was 1/25 that for untreated surfaces and was only about twice as high as that for the aluminum gallium arsenide–covered surface.

In all cases, the surface had been cleaned by chemical etching prior to further treatment. According to the Bellcore researchers, after etching the usual gallium and arsenic oxides and elemental arsenic formed. During the chemical treatment, arsenic oxide dissolved in the alkaline sulfide solution, while arsenic reacted with sulfur to form another soluble complex. Finally, sulfur atoms bonded to surface gallium atoms to generate a surface resistant to further oxidation.

Surfaces treated by this means were in fact quite resistant to degradation. Heating the gallium arsenide to temperatures as high as 350°C, leaving it in air for up to 19 hours, and washing it in water all caused only a slight increase in the surface recombination. Unfortunately, the films are ionic conductors and have poor mechanical properties, so they do not themselves constitute a full solution to the problem of passivating gallium arsenide surfaces for MOSFET-type devices.

But Sandroff and several Bellcore coworkers have shown that passivating surface films generated by this method have beneficial effects on a different type of transistor, a so-called heterostructure bipolar transistor. A bipolar transistor has three parts: an emitter, a base, and a collector. A small current injected into the base causes a large current between the emitter and collector, another form of amplification. In contrast to MOSFETs, which rely primarily on either electrons or holes to carry current, a bipolar transistor requires both *n*- and *p*-type regions.

As the transistor is miniaturized and the surface-to-volume ratio increases, surface recombination on the periphery of the emitter-base region (rather than at the interface itself) increasingly becomes a problem, resulting in a low amplification (low gain). The Bellcore group showed that coating this area with a $Na_2S{\cdot}9H_2O$ film a few tenths of a micrometer thick increased the gain of their transistors by a factor of 60.

All in all, the IBM and Bellcore findings demonstrate and confirm that the Fermi energy at the surface of gallium arsenide is not inevitably pinned and that there are ways to unpin it. The next step, on which activity is already focused, is to find ways that are also compatible with the requirements of microcircuit technology. ■

ARTHUR L. ROBINSON