FUTURE TRENDS IN SEMICONDUCTORS

By Robert N. Noyce
Fairchild Semiconductor Corporation
Mountain View, California
May 22, 1961

Predicts micro miniaturization
Planar IC invented in '59

The development of semiconductor devices has affected the electronics industry profoundly in the last decade, and we may expect the changes in the industry to continue through the next ten years. It is very dangerous to make predictions in any technical field since no one is wise enough to foresee the discoveries that will be made. If he were, he would also be able to make those discoveries. Consequently, predictions which have been made have always been too conservative or they have been dismissed as science fiction, and not realistic.

However, the seeds of some developments have been planted already, and I believe that the extrapolation of trends which we have seen already will provide a conservative estimate of changes in electronics which will occur in the future.

I would like to start by following the history of the semiconductor devices in electronics, leading up to a description of the present state of the art. Then I would like to mention some recent developments which may provide the basis for more rapid change in the future.

The semiconductor diode was one of the first elements used in electronics, in the form of a point-contact diode in which the semiconductor material was a crystal of lead sulfide, or galena. This element was displaced by the vacuum diode shortly thereafter, only to be considered again when the frequencies used could not be detected efficiently with the vacuum diode. Consequently, starting about 1940, and thereafter, intensive work was undertaken to develop semiconductor diodes for use as microwave detectors. At the same time, a great effort was made to understand the mode of operation of the semiconductor diode, and in particular, the anomalous effects which would not be explained by the existing theory. It was in the investigation of one of these effects--extraneous surface leakage currents--that Drs. Bardeen and Brattain discovered the first semiconductor device exhibiting power gain: the point contact transistor. Because of the impact that this development might have when refined, the phenomenon was the subject of intensive study, during which the theory of P-N junctions was developed by Dr. Shockley. Out of the study of P-N junctions came the invention of the junction transistor. However, another two years passed after the theoretical demonstration that such a device could be made before the techniques for making such a device were refined to the extent that it could be done.

In the early 1950's a battle ensued between those who thought the junction transistor too difficult to make, and therefore impractical, and those who thought the point contact transistor too difficult to understand, and therefore impractical. But as has happened many times, that which was understood theoretically could be refined and developed. Thus, today the point contact transistor is insignificant in comparison to the junction transistor, and techniques for making the junction transistor which were thought impossible ten years ago are in common usage today.

You all know that the junction transistor consists in general of a bit of semiconductor material including regions of differing impurity concentrations, a P region, a N region, and another P region or conversely, an NPN configuration, with connections made to each of these regions. The design of transistors has consisted of determining the optimum impurity densities, dimensions, and lifetimes in these three regions and the grading of impurity densities between them, consistent with the limitations imposed by the manufacturing techniques used for making the device to the theoretical design. A great amount of effort has been expended in both areas, and as new designs were shown theoretically to be advantageous, the technology for achieving the design was developed.

For use in electronics, the transistor had to serve the frequency range of interest. The early transistors were most severely limited in application by their low frequency response, due to their large dimensions. These dimensions were not really large, with base widths of the order of one tenth of a millimeter, but they are large compared to those which are in use today. The early techniques for controlling these dimensions were essentially mechanical. In the grown junction transistor, the crystal was grown in one conductivity type, an impurity of the opposite conductivity type was added to the melt, and growth was continued, then another impurity was added and growth continued. Alloy junction transistors are made using a slice of semiconductor lapped to about 0.1 mm thickness, then reducing the thickness a controlled amount by dissolving the semiconductor in a molten metal bearing the desired impurity, allowing this impurity to be included as the melt recrystallized. The electrochemical etching technique is the most accurate of these mechanical methods, consequently the frequency response of transistors made by this technique is the highest of any made by mechanical means.

The early theoretical designs showed, however, that dimensions much smaller than those which could be achieved by mechanical means would be required if the frequencies of over 100 megacycles were to be served. Consequently, other techniques were explored, and of these, diffusion has proved to be the most practical. In use of diffusion for transistor manufacture, the dimensions of critical importance, in particular the base thickness, are controlled by the diffusion of impurities in the solid material at elevated temperatures, rather than by mechanical methods. Since the laws of diffusion are well known, the dimensions may be controlled very accurately by controlling the time and temperature at which diffusion is done. Base widths of one or two microns are commonly used in manufacturing transistors today, giving gain at frequencies of several hundred megacycles.

The use of diffusion techniques has been adapted to the three techniques previously mentioned, resulting in the grown-diffused transistor, the pos alloy diffused transistor, and the electrochemically machined diffused base transistor. Because these are adaptions of previous techniques some design limitations still exist in these methods of manufacture. These design limitations are largely eliminated in the double-diffused, or diffused base-diffused emitter transistor. Commercial versions of such a transistor with gain at 1 kilo-megacycle are available, and laboratory transistors have been made which will oscillate at 3 kmc.

Using diffusion technology, the active regions of the transistor can and have been made extremely small. Because of mechanical handling problems, however, a slice of the semiconductor which is thick compared to the few microns of material in the active regions of the transistor must be included for mechanical support. This adds appreciably to the resistance in series with one of the terminals of the transistor, usually in the collector lead. The recent introduction of epitaxial growth has made it possible to reduce and nearly eliminate this resistance. In this process, a layer of pure semiconductor is grown on a high conductivity substrate, and the entire transistor formed in the thin layer of epitaxially grown material. I think that this technique will be used to a much greater extent in the future, since it results in higher performance devices.

In terms of performance, transistors are now available to replace the function of the vast majority of vacuum tube applications. They now have higher gain band-width products than normal vacuum tubes. The transistor will not yet replace some special vacuum tubes, such as very high power tubes, and those operating on an electron inertia principle, such as travelling wave tubes or magnetrons at microwave frequencies. Semiconductor devices are approaching these frequencies, and through special application techniques can be used at low power in the microwave regions. I will discuss these shortly.

I should like to discuss the advances which have been made in the reliability of semiconductor devices as well, since this is of importance to some of the extensions which I believe will occur in the semiconductor field. Transistors were expected to contribute to the reliability of electronic equipment soon after they were introduced, but they have done so only after full understanding of their causes of failure. Transistors were expected to have longer life than vacuum tubes because there was no inherent failure mode recognized such as the evaporation of barium from the cathode of a vacuum tube, which limits its average life to 10,000 hours of operation. In addition, since the transistor is small and rugged, mechanical failures such as the shorting of elements in the vacuum tubes were expected to decrease. Semiconductor devices have lived up to these expectations, although the early ones did not. The cause of failure of early transistors was largely due to changes in the condition of the semiconductor surfaces. The importance of impurities on the surface of the semiconductor is perhaps best demonstrated by pointing out that a layer of impurities a single atom thick on the surface will contain more impurity atoms than the entire active volume of the device would contain by design. However, it was soon found that with extreme cleanliness and care in sealing the transistor in hermetic enclosures, this mode of failure could be greatly reduced. Since it is difficult to assure that a good seal has been made without testing, the most reliable transistors have been completely leak tested. Mechanical failures have been found as well, and weaknesses have been designed out. Normal commercially available transistors of good quality have been operated in large computer systems with mean times between failures of one to ten million transistor hours. This represents an enormous improvement over what would have been possible prior to the introduction of the transistor.

Here I would like to mention further development which has made possible another great improvement in the reliability of transistors. This is the planar technique, in which the transistor is sealed in its own skin at the beginning of the manufacture, and consequently never exposed to contaminants in manufacture or use. In making transistors by this technique, the first step is to grow a layer of silicon dioxide, or quartz on the surface of a crystal of hyperpure silicon. This forms the closure for the finished transistor of an impervious, and very high resistivity insulator, atomically bound to the surface of the semiconductor. Thus the P-N junctions are never exposed to the atmosphere. In addition, since the surface properties of the semiconductor are stable, they may be adjusted to the ideal surface potential, making the transistor have more nearly theoretical performance.

The great impetus for development in the semiconductor electronics field was the discovery of minority carrier transport, i. e. the diffusion of holes in N-type materials, and electrons in P-type materials. Although there have been other phenomena used in semiconductors to obtain amplification, minority carrier diffusion has received by far the most attention.

In the past few years, two other phenomena have received a great deal of attention. The first is the parametric amplifier, and the second the tunnel diode discovered by Esaki.

A parametric amplifier may be made using any element which displays a non-linear reactance. The two phenomena which have been used are reactances using a magnetic material which saturated partially in the range of operation, and the variable capacitance of the semiconductor diode. The variable capacitance of the semiconductor diode, if the diode is properly designed, has low enough losses that it may be used at microwave frequencies as an amplifier, and for harmonic generation up to as high as 100 kilomegacycles.

The tunnel diode depends on quantum mechanical tunnelling across the forbidden energy gap of the semiconductor. It has a negative resistance characteristic from D C to over 100 kilomegacycles if properly designed, and may be used for oscillators, and negative resistance amplifiers in the microwave frequency range, or as a switching element in high speed logic circuits.

In both of these devices, the active region is extremely thin, about 100 Angstroms in the tunnel diode to perhaps 1000 Angstroms in the parametric diode, and their high frequency response results from this very small active region. They are both two terminal devices, with the input and output leads being the same in actual use. Although this presents no problems in their application as oscillators, frequency dividers, or frequency multipliers, it does present problems in making stable broad-band amplifiers, and this is not as easily done as with the transistor in which the input lead and the output lead are separate and distinct. The devices themselves are quite simple to make. The only fabrication problem in their manufacture is to make them of small area, and to reduce as far as possible the high frequency resistive losses. Since the active region is very thin, the capacitance per unit area is large, and in order to match impedances to practical microwave circuitry, the area must be made small in order to keep the reactive impedance small. The series resistive losses occur in the extraneous semiconductor material outside the active region of the device, simply because we have no methods of handling materials as thin as they should be. The use of epitaxial growth for the fabrication of these devices will improve their high frequency performance, however.

I believe that in the future we will see the extension of the tunnelling principle to other devices. Early experimental transistor-like devices have been made using a tunnelling principle for carrier injection rather than a P-N junction. Although there is still much development to be done, this principle seems to hold promise of useful three terminal devices in the future.

Although germanium and silicon have been the semiconductors most widely used in transistors and diodes, a great many other semiconductors are used in photo devices, and research is being carried on to find other materials for transistors. Gallium arsenide has received the most interest recently.

This material has a larger energy gap than germanium, and consequently may be operated at higher temperatures than germanium. It also has higher carrier mobilities than silicon, so that for the same dimensions will have higher frequency capabilities than the silicon device of the same dimensions. However, the technology of using the material is far behind that of germanium and silicon, and for that reason, I do not believe that it will come into widespread use for transistors in the near future. On the other hand, I do think that it will be used more and more in the simple devices, that is, high speed diodes, parametric diodes, and tunnel diodes.

In transistors, I believe that there will be a shift from germanium to silicon. Since the mobility in germanium is higher than in silicon, it was expected that germanium would always have higher frequency capability than silicon. However, when trying to get the ultimate frequency response from the transistor, it is necessary to operate at high current densities to increase the emitter cutoff frequency, and at high collector voltages to decrease the collector capacity, and thus operate at high power densities. Since the silicon device will operate at higher temperatures, higher power densities are possible in silicon than in germanium. In addition, the newer diffusion techniques seem more applicable to silicon and thus today, we find the highest frequency transistors being fabricated of silicon instead of germanium.

The early transistor fabrication techniques were simpler to use in germanium than in silicon, and the resulting germanium transistors were less expensive than silicon. For that reason, silicon was used only where germanium could not be used. This situation is changing; the silicon is more reliable, and will do the job better at the same cost. Consequently, I expect the usage of transistors to change heavily to silicon in the near future.

The use of semiconductor devices has resulted in a great reduction in size and weight of electronic equipment, and has made possible much more complex electronic equipment because of higher component reliability. We realize, however, that this is just a beginning of this trend, and there is a great deal of interest in exploring the continuation of this trend through microminiaturization, integrated circuits, and eventually to molecular engineering. The motivation for this exploration is several-fold; first, we can see utility for even more complex electronic equipment if such equipment can be made economically. Second, an improvement in the reliability is needed for such equipment to be useful. Since the interconnections between components is now a major failure mode, reliability improvement should occur in reducing the number of interconnections through integrated circuitry. Third, in high speed logic circuitry we have progressed to the point that the delay caused by signal transmission time from one part of the circuit to another limits computation rates, and this can be reduced only by making the entire circuit smaller. Fourth, in this space age, the weight of electronic equipment which is put into orbit determines the cost of putting it there.

Because of the need for integrated microcircuits, this is a very active field for research and development today. Although there may eventually be other solid state devices which will displace the transistor, in the near future, the transistor will be the principal active element used in microcircuits. I would guess that there is more effort being expended in microminiaturization than any other advanced semiconductor development today.

The approaches to microminiaturization have been many and varied. The first approach which has been in progress for several years is that of shrinking the physical size of each individual component, and then assembling microcircuits from microcomponents. This approach has been directed primarily to the objective of making circuitry small. In some modifications, all components are made in a standard module, so that they may be assembled automatically and packed more densely than components of varied shapes. I regard this as an early interim approach. However, it may be that many of the techniques learned will carry over to the more sophisticated integrated circuits made later. The assembly of microcomponents does not appear to represent an overall cost savings, however, for savings in assembly costs, if any, will probably be used up in higher component cost, since the individual component is forced into an unnatural configuration.

The second approach being taken is to fabricate all circuit modules through evaporated films of various materials to make resistors, capacitors, and, hopefully, other active elements. For the most part, those using this approach are now using transistors for the active gain elements in the circuit. This use of transistors for the gain elements counters the advantages of lower cost, since hand assembly is needed at this point in assembling the circuit. It may be possible, in the future, to use other active elements which may be fabricated by evaporation, so that the entire circuit may be completed at once. Devices which are candidates for fabrication by evaporation are superconducting devices like the cryotron, devices employing quantum mechanical tunnelling between superconductors, transistor-like devices employing tunnelling for carrier injection, and thin-film magnetic amplifiers. None of these devices has been developed to a point where it is certain that the device will be practical when the overall system requirements are considered. For that reason, transistors will probably continue to be used in this approach to microminiaturization.

A third approach to the microminiaturization of electronics has been to fabricate the integrated circuit from a semiconductor material, using the semiconductor with configurations of P-N junctions and resistivities to serve the functions of transistors, diodes, capacitors, and resistors. In the work we have done, a slice of semiconductor material is fabricated with different areas of N and P type material by diffusion techniques to make the elements necessary for the circuit function to be performed. These are isolated electrically, although they are all in the same physical piece of material. This slice of semiconductor material is oxidized, to provide an insulating surface, and holes are etched in the oxide to provide electrical access to the components. Then the components are interconnected by evaporating metal conductors on the oxide surface of the desired pattern. Finally, the circuit thus made is packaged and thereafter treated as an individual component. We are now working with circuit modules containing only four to ten transistors; at this size there is about an 80 percent reduction in the total number of leads external to the component. I believe these modules will become larger in the future.

This approach does not yet represent the ultimate in microminiaturization, but it does offer advantages in cost and reliability as well as in size and weight. It has been estimated that the individual transistors could be made on a wafer for less than 1/2 lira each, prior to interconnecting. Not all of them will be good, unfortunately. However, if suitable methods of redundancy can be worked out, it may be possible to use these components in far more complex circuitry at a reasonable cost.

The fourth approach to integrated microcircuits is that which I shall call molecular engineering. In this approach, the identity and function of the individual component is lost and a new component is defined by the circuit function which it performs. Electronics has progressed in the past because of the ability to synthesize complex functions from simple components; in the molecular engineering approach this can no longer be done. Consequently, it is an endeavor which calls for great invention and insight. Some such devices have been invented and made. I believe that a complete family of these devices will never exist to do all the functions which our industry will need. Consequently, as such devices do become available, they will be incorporated into more conventional circuit configurations, and that synthesis of circuit complexes will proceed as it has in the past, starting from simple elements.

In the past ten years, solid state electronics and the semiconductor industry has become as fundamental to the electronics industry as the vacuum tube was before the invention of the transistor. It has made possible much more complex electronic equipment than was possible previously because of better reliability. With the exception of such devices as travelling wave tubes, the frequency performance of transistors now matches or surpasses that of normal vacuum tubes.

With the development of the technology applicable to silicon transistors, I believe that silicon will displace more and more of the transistor market because of better performance, reliability, and competitive costs. Other materials will be used for specialized semiconductor devices, but will not in the near future be used for transistors.

New basic phenomena such as quantum mechanical tunnelling and new principles of use such as the parametric amplifier will extend the use of semiconductor devices into the microwave range, and may result in new types of active amplifying devices.

I believe that the most apparent change which will occur is the progress in integrated microcircuits. The approach will be first to digital circuitry, in which there is a great deal of iteration of identical circuits. To take full advantage of the possibilities, new principles of redundancy and statistical design will have to be developed.