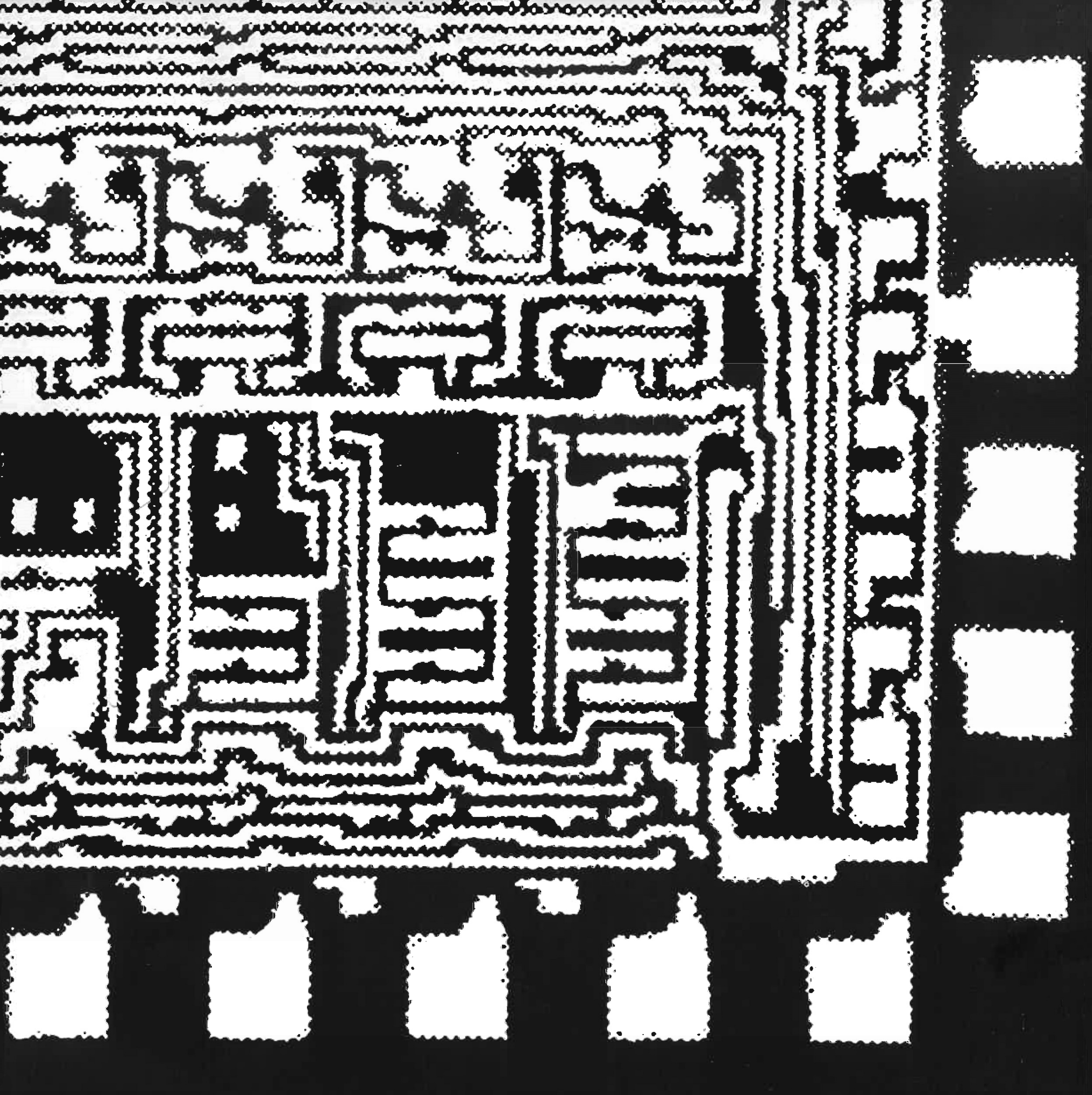


The Semiconductor Story

SCHWEBER
ELECTRONICS

Sam Kass

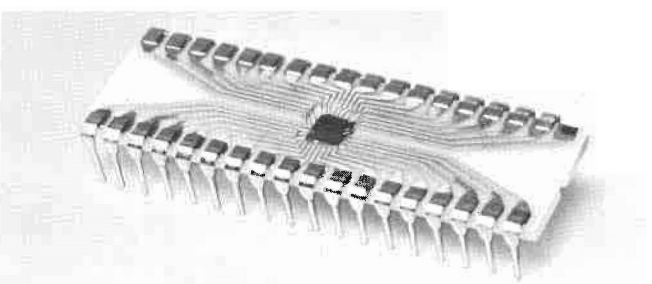
NOTE OF EXPLANATION

Over the span of two years, the story of the semiconductor was narrated in seven issues of a newsletter called "The Tip-Off," published by Schweber Electronics. It is hardly a history of the subject – seven issues of a 4-page newsletter would barely suffice as an introduction. But perhaps that is what we were doing . . . introducing the semiconductor story to an audience which is just beginning to feel its powerful effects.

Our intent is not to expand the original seven parts. To do so might deprive the first draft of the easy informality and plain speaking not considered proper for a treatise, but eminently suitable for the wide hearing we hope it will achieve. Although some new information has been added, as well as some shifts in emphasis and re-arrangement of text, hopefully to make for smoother reading, most of the changes are confined to correcting errors of typography and punctuation.

For anyone interested in pursuing the subject further – and deeper –, a bibliography is appended for each of the seven parts. Only statements considered important will merit a footnote indicating author and source. Finally we extend our appreciation for the courtesy of reprinting the illustrations from the literature of the semiconductor industry.

A condensation of Parts I, II, and III appeared in the May 1967 issue of the Student Journal of the IEEE.



The front cover is a blow-up of a LSI (Large Scale Integration) chip. The chip (¼" x ¼") is in the center of the above socket and is a Fairchild 3751 Converter, a subsystem on a chip. Three distinct blocks of logic components are built into the one chip making up a 10-Bit Digital to Analog (D/A) Converter. There are ten-flops in the Shift Register, ten more flip-flops in the Holding Register, and ten analog single pole double throw switches which are controlled by the Holding Register. The D/A Converter can convert a digital input word into an analog output voltage.

The Semiconductor Story

By Sam Kass

Contents

Library of Congress Catalog Card Number: 68-55809

FOREWORD

The growth rate of semiconductors and computers are not just coincidental. In fact, the parallel growth and developments in both these industries will, in the last third of this century, account for our greatest technological advances.

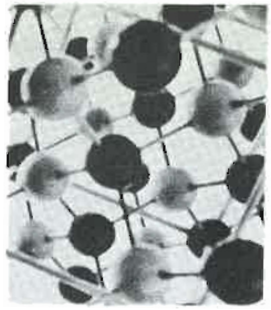
The reliability, speed and economy of computers could never have been achieved without semiconductors. Just like the child who starts teaching his parents, so today we are using computers to automatically design and make the masks for the most complicated semiconductor circuits. Computers are figuring out optimum designs in minutes, that would otherwise have taken years of man hours to perform.

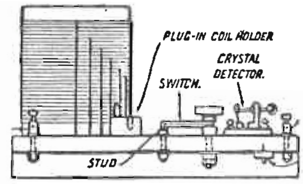
Reading this book by Sam Kass which shows the dramatic history and evolution of semiconductors, one could easily forecast the uses of solid state devices in the next twenty-five years. Wrist-watch T.V., mini-computers for shopping, all of us with card-carrying, data processing numbers for our future cashless society. Instant world information by Telstar and shopping by T.V. All this sounds exciting but what effect is all this going to have on society, on our personal lives? Is the new electronic world going to drive people apart, alienate them (young people don't even hold hands when they dance anymore) or will it make them apathetic? Exasperate them? No one knows answers now, although we have some clues. (Most people are apathetic about the war in Vietnam, a minute minority called "Hippies" seem to be the spokesmen against the establishment.)

Just as the scientists working on the first atom bomb did not realize its true effect on our lives, so the physicists who invented the transistor could not imagine its immense potential. We hope that in the same way as modern physics is today being used for peaceful purposes, so will electronics be part of building a better world – solving such pressing problems as air, water and noise pollution, faster and more transportation, improved hospital care and early detection of disease and worldwide education to help the underprivileged peoples of the world.

Will the information explosion be so voluminous that it will numb people? Will people stop fighting back because they would have to deal with machines, not people?

Seymour Schweber
President
Schweber Electronics





Crystal radio — vintage 1920.

◀ Model of atomic crystal lattice.

PART I — SETTING THE STAGE

Overture on a Familiar Theme

Solid State physics has many branches; the one best known to the public pertains to a class of solids called semiconductors. It is so well known that the phrase *Solid State Age* is often used as a rhetorical symbol for semiconductors alone, blithely ignoring such important solid state areas as superconductivity, lasers, luminescence, structural analysis, etc. Ages can be born unbaptized, and the *Solid State Age,* though born in 1900 when Max Planck postulated the Quantum Theory, was not firmly and formally tagged until the transistor burst upon the scene in 1948, trailing behind it the enormous potential which even now has not been exhausted.

The phrase *"Solid State Age"* may seem pompous; indeed the many trail-blazing inventions and breakthroughs of the 20th Century, all endowed with that patent of nobility we call an Age, may seem like a mountain of pomposities, fabled images dreamed up by public relations puppeteers. One need only look at the list of ages coined in our century – it is a long one: (to begin with the A's) the Atomic Age, the Automated Age, the Aspirin Age . . . But why not? On second thought, why should we not have a proliferation of Ages in this century? It is a matter of record that there are more scientists living and working today than in all the preceding centuries together. Nor should one be surprised that scientific research literature is growing at an exponential rate, doubling about every dozen years. Is it any wonder then that in this century an Age is born every decade and a "generation" lasts half that time?

It would be wrong of course to assume that an Age will pop up every decade on the hour like a crocus in early Spring. There is no fixed gestation period although a gestation process is involved. One can't predict a forthcoming Age by correlating the current rate of growth with past growth rates. Correlation only reveals the simple fact that the speed of change keeps growing faster, something we already knew. The fact is that Age-producing breakthroughs occur haphazardly. Yet the haphazard event is preceded by a steady accumulation of bits of knowledge that ultimately generate the final breakthrough. Let us take a backward glance at the semiconductor story as the bits of research accumulate like budding sprouts waiting for the day of harvest.

This is not being written with the engineer in mind, though he is welcome to read if he promises not to carp. Nor is it intended for that mythical individual "the nontechnical reader," since we don't think he exists. Anyone who can count digitally is technical enough for us – which takes care of the mathematics. As for the rest, all you need is the power of the imagination; whether it is fantastic, visionary, creative, or poetic, it will turn the trick for you. If there is a purpose to this project, it may well be the chance it offers to evoke the remembered pleasures and thrills of a bygone era that was fondly loved.

Wonder of Wonders: The Crystal Radio Set

The phrase Solid State is not usually associated with radio equipment before the coming of the transistor, but the fact is that the earliest radios for home use were solid state by virtue of the semiconducting crystals which they used for detection. Immediately after World War I these crystal sets became an ornament of every living room and huge 100 to 200 feet antennas, called aerials in those days, sprang up on myriad rooftops like some strange species of spider webs, hopefully strung up to achieve the maximum pickup of energy radiated by the transmitting antenna. The crystal set was designed simply to receive the audio signal wrapped up in its radio carrier, abstract the signal and discard the carrier like a torn-open envelope. Though some slight gain was afforded by other components of the crystal set, the crystal itself was entirely guiltless of any amplification.

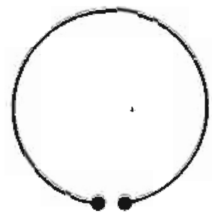
The crystal detector was so inexpensive that it seemed a shame it could not be made to amplify. This thought sparked a trail of research that finally culminated in the discovery of an amplifying diode, just as the crystal itself was the end-result of an equally long search that dated back to the middle of the 19th century. The research con-

Michael Faraday

James Clerk Maxwell

Heinrich Rudolph Hertz

HERTZ RESONATOR OR WAVE DETECTOR



(Fig. 1) Complete schematic of first radio receiver.

tinued unremittingly, if slowly, through the era of the vacuum tube, picking up added strength as the frequency range soared into the microwave region where vacuum tubes are troublesome. The transistor solved the problem of a solid state amplifying device, but its two-junction mechanism seemed to imply that a single-junction device could never be made to amplify. Eventually the persistence of man was rewarded when Dr. Leo Esaki described his "tunnel" diode in a one-page letter to the editor of the Physical Review of January 1958. Another breakthrough occurred in 1963 when J. B. Gunn discovered that a small crystal of gallium arsenide could be made to oscillate at 1,000 megaHertzes (MHz) when surrounded by a high electric field. But let us go back . . .

The Electrical Age Takes Shape

. . . all the way back . . . to Faraday's fundamental experiments with magnets in 1831, and a triple play in which three men carried the ball beginning with (1) Faraday and his magnets, to (2) Maxwell and his equations, to (3) Hertz and his radio waves. Faraday, the greatest of all experimenters, discovered that the motion of a magnet inside or close to a coil of wire induces a current in the coil. Not being a mathematician, Faraday explained his experiment by assuming that the magnet is surrounded by lines of force which cause current to flow whenever they cut across a conductor. An appreciative mathematician, Clerk Maxwell, expressed Faraday's pictorial ideas in mathematical terms. In 1864 he went a step further and suggested that electrical and magnetic changes are propagated in the form of waves, and theorized that these waves travel at the speed of light, and finally that light itself is propagated in the form of waves.

Although Maxwell's theory of light waves was generally accepted in England, the rest of Europe did not think much of it. In 1879 the Berlin Academy offered a prize to anyone who could throw light on this question. This prize led a young student named Heinrich Hertz to a study of Maxwell's equations. He promptly fell in love with them. Using amazingly simple means – an induction coil and a loop of wire – he was able to demonstrate just the sort of electrical waves predicted by Maxwell's theories. Even more remarkable, he was able to measure the speed of these waves and found them to have exactly the same speed as light. The electric waves created by Hertz in 1887 are now called radio waves. It would be an exaggeration to credit Hertz with the invention of radio, but he certainly demonstrated its possibility just as Maxwell predicted its existence. Hertz's work was interrupted by a severe and prolonged illness; and his untimely death at 37 was a great loss to science.

Hertz's experiment attracted a lot of attention from physicists, but one untutored youth built his entire life around it. A young Italian named Guglielmo Marconi became convinced that he could transmit Hertzian waves across great distances. He was amazed that others had not jumped to the same conclusion. However, his initial attempts, using Hertz's original circuits, were limited to no more than a few yards. One of his difficulties was the total inadequacy of his receiver, which is not surprising considering that Hertz had used a single loop of wire not quite closed at one end to detect his electric waves. The loop would respond, when the Hertz "transmitter" emitted a spark, by jumping a similar spark across the tiny gap. Sans antenna (as we know it), sans detector, sans any amplification, reception at more than laboratory dimensions was doomed. "Long" distance radio was temporarily blocked by the lack of a sensitive detector (see Fig. 1).

Scenario for a Research Drama

The story of the origin and development of the first practical radio detector is a classic example of a long series of lone inventors contributing crumbs of knowledge to the general scientific store until someone, dipping into the collective barrel, looks at an old idea through new eyes and new needs, and transforms the old idea into a revolutionary advance. In a sense, this is what Marconi did; but the story of the first detector is even more striking, perhaps because the list of scientists involved is as long as the cast of a Hollywood spectacular.

Time: 1850 (A nice round year to begin with.)

Edouard Branley

Place: England, France, Germany, Italy, Russia, America.

Enter Pierre Guitard, a French scientist. He reads a report about dust particles in the air to other scientists. When the air is electrified, the dust particles form into small compact clusters. He cannot explain it. Fadeout . . .

Enter S. A. Varley, an English inventor. In 1866 he is busy at work trying to protect telegraph wires as well as the poles and insulators supporting them from destruction by lightning. Guitard's observation suggests an idea to him. Varley hollows out a block of wood and fills it loosely with carbon particles and tin filings. He attaches one end of the block to the telegraph wires and the other end to ground. The idea works. How? When lightning strikes, the loose particles form a tight, compact, conducting mass which harmlessly short the flashes to ground. A minute or two later, the particles loosen enough to allow resumption of telegraph operation. Fadeout . . .

Enter David E. Hughes, a professor of music, and inventor of the microphone; born in England, educated and naturalized in America. In 1878 (the same year he invented the microphone) he performs an experiment no one pays much attention to. His experiment proves that the discharge of a Leyden jar (an old type of capacitor) does to zinc and silver filings what lightning had done to Varley's tin and carbon filings. Fadeout . . .

Enter Calzecchi Onesti, an Italian professor. In 1884 he publishes an article describing how copper filings cling together when subjected to a high voltage discharge in the immediate area of the filings. Fadeout . . .

Enter Edouard Branley, a French doctor, later Professor of Physics. Dr. Branley is interested in the way nerves carry messages to the human brain. His work leads directly to the field of electricity. In 1890, after studying the reports of his predecessors, he fills a glass tube with nickel and silver filings and plugs up each end with metal electrodes. Then he connects across his new device a battery and meter. The needle of the meter does not move – the device is apparently a poor conductor. But wait. Branley strikes a spark from a nearby battery, when lo and behold, the needle moves sharply, indicating the device is now conducting. The needle continues to read a high current until the filings are jarred loose – then no current. Very slow fadeout (Branley died at the age of 93) . . .

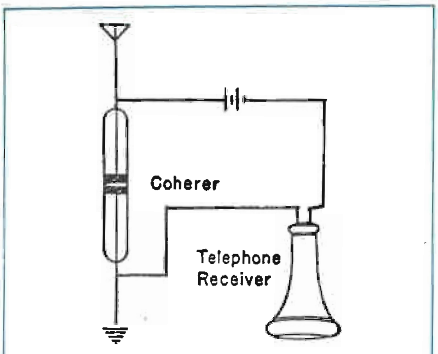
Enter Sir Oliver Lodge, eminent British scientist, first Principal of Birmingham University, knighted by the Crown. Lodge sees in Branley's device the long awaited means of detecting Hertzian (radio) waves and gives it the name by which it becomes widely known: Coherer. Lodge does more – he overcomes one of its defects by designing a clockwork mechanism which decoheres the filings automatically. In 1894 he demonstrates the sensitivity of the coherer in a public lecture at Oxford by detecting Hertzian waves over a distance of two hundred feet. Fadeout . . .

Enter Alexander S. Popoff, Russian physicist, who uses the Branley coherer to predict approaching lightning storms. He inserts an electrical bell in the coherer battery circuit and when lightning flashes occur that are still out of range of sight, the sensitive coherer is able to detect the atmospheric discharge and cause the bell to ring a warning of approaching storms. To improve the device, Popoff attaches the bell-ringing arm in such a way as to strike the coherer tube and instantaneously reset the filings for the next flash. In 1895 Popoff demonstrates his bell-ringing coherer by receiving an "audible" signal over a distance of 600 yards at a meeting of the Russian Physicist Society of St. Petersburg. Dissolving view . . .

The Stage is Set for Conclusive Experiment

Before unveiling the final tableux, let us dispose of some rival claims. The loudest claims are for Popoff because he produced the first audible signal from radio waves – which may be true, but this hardly merits the claim that he invented radio. Similar claims have been made for Sir Oliver Lodge, although he used a galvanometer rather than a bell or telegraph ticker to register the signals. Undoubtedly Lodge exhibited successful short range radio transmission,

Gugliemo Marconi



(Fig. 2) Marconi's receiver at Newfoundland.

as did Popoff a year later, but in both cases the demonstrations were of a theoretical rather than a practical nature. And while we are on this subject, something should be said for Dr. Branley since he steadfastly refused to claim anything for himself throughout his long life. His coherer, improved mechanically, but unchanged electrically, triggered the birth of modern radio.

The barrel has been filled with all manner of goodies which a clever man can put together, perhaps with a fillip of his own added, to complete a new recipe so startling as to make the world sit up and take notice. Marconi is that clever man. He took the Hertz exciter, spruced it up with more power, and the Branley coherer, evacuated the glass tube, and added his own overhead aerial and grounded rod. Very quickly he attained a distance of several miles, but this did not satisfy him. Marconi was obsessed by distance – not hundreds, but thousands of miles. (He was bitten by the DX bug, a disease he transmitted to many others until it developed into the proportions of an epidemic in the early twenties.) Gathering a staff of engineers around him, he built a more powerful transmitter, every bit of 25 kilowatts. In 1901 his signal spanned the Atlantic – a distance of 2,000 miles. The receiver he used is diagrammed in Fig. 2. Impossible, you say? But look again – there is something missing in the diagram. It does not show Marconi's aerial raised by a kite 400 feet above the earth – a typical DXer's trick.

Credit for the 'discovery' of radio is usually assigned to Marconi. There is no question about the importance of Marconi's efforts, but as Sir Robert Watson-Watt points out, it was a team effort, even though Marconi's contributions were more spectacular. In Watson-Watt's view, *". . . the four pioneers of radiotelegraphy – in alphabetic order – (are) Jackson, Lodge, Marconi, and Popoff."* The last three have been mentioned here. Admiral Jackson worked under conditions of secrecy for the Royal Navy and his accomplishments did not reach the news media although Marconi was aware of them thru personal contact with the Admiral.

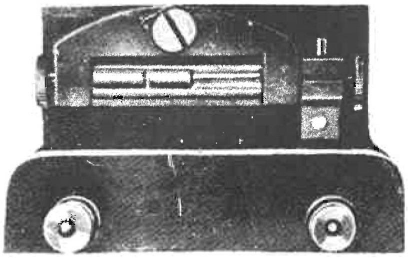
Successors to the Coherer

Having played its spectacular role in the opening act of the radio drama, the coherer gave way to improved and more consistent detection devices. Marconi himself introduced a so-called magnetic detector in 1902, and the American Prof. R. A. Fessenden came up with an electrolytic detector that same year. But it was not until 1904 that the first modern detector appeared on the scene. This was the two-element thermionic vacuum tube developed by J. A. Fleming, a member of Marconi's research staff. Two years later, Lee deForest added a third element to the vacuum tube and thus created a detector that also amplified the signal it detected and the Electronic Age was born.

Although the future belonged to the vacuum tube (as "futures" go in electronics), its high initial cost, high installation cost, and high operating cost, placed it out of reach of all but commercial installations. But in the same year, Greenleaf Pickard discovered the rectifying properties of many mineral crystals, among them galena, which became the most popular detector in home crystal sets because of its low cost and high sensitivity. It had one mechanical disadvantage – a cat's whisker was used to probe the crystal for a sensitive spot. But in return for that little trouble it was able to begin rectifying at the point of zero voltage.

Some may claim that the Semiconductor Story should begin properly with the invention of the transistor in 1948. This may be true for anyone interested strictly in the technology of their manufacture, but to begin with the decisive event – the breakthrough – is to miss the most human part of the story: the fumblings, the gropings, the feeble steps that gradually picked its halting way to success. No apologies are necessary for this pattern – they characterize all areas of human endeavor.

We are indebted to George Rostky, Editor of EEE, for information leading to the discovery of two authentic Branley Coherers. You will find them in the Museum of the City of New York on Fifth Avenue between 103rd and



(Fig. 3) The detector used by Marconi for transatlantic reception in 1901.

Greenleaf W. Pickard

104th Streets. They are displayed in the Communications gallery on the third floor. One of the coherers was donated to the museum by David Sarnoff and was originally owned and used by none other than Marconi himself. Now some of you young skeptics who don't believe the coherer ever existed, not alone worked, can go see one in the flesh.

PART II — A GAGGLE OF DETECTING DEVICES

Statement of Intent

The popular outline of the rise of radio that is usually spread before the public in our time, such as the magazine feature story and the synthetic encyclopedia article, is written to present no difficulties to the average reader. But if your interest compels you to scratch below the surface for more color, more credibility, more background, then what seemed so straightforward and simple suddenly becomes quite complex. So complex, in fact, that what we customarily read is a hodge-podge of fiction and apocrypha, garnished with a dash of truth. An example of the complicated historical threads that are still being unraveled is provided by the coherer pictured in Fig. 3. Marconi is on record as saying that this is the coherer that detected the famous first transatlantic radio signal. A contemporary worker in this field disputes this statement and claims some other form of detection was used by Marconi.[1]

So, to keep the air clear, we shall begin with a statement of intent. Readers commonly assume that the Age of the Vacuum Tube was a necessary step in the development of the radio industry. This writer 'dares' to question that assumption.

In 1945, Dr. I. I. Rabi, Nobel prize winner in physics, described the triode tube as *"so outstanding in its consequences that it almost ranks with the greatest inventions of all time."* An anthologist could fill a book with such quotations. How cruelly all these grandiose eulogies have been treated by the passage of a mere couple of decades. The vacuum tube, once said to rank with the greatest of all inventions, seems destined shortly to rest in a museum's glass cage alongside the coherer.

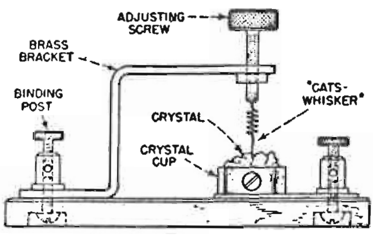
What is involved here is not simple obsolescence brought on by newer and better devices, in this instance, the transistor. The position adopted here is that the vacuum tube was an interruption in the development of solid devices which began in 1906 when Greenleaf Pickard patented what he called a "solid rectifier." Energies and resources which might have brought the transistor on the scene by 1920 were diverted toward improving the vacuum tube. Ironically the inability of the vacuum tube to perform efficiently as a microwave detector led to its downfall and a *return*[2] to the silicon point-contact diode.

The Second Detector

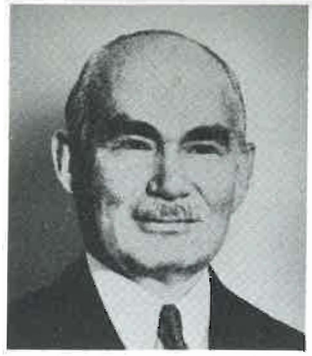
The Branley Coherer was the first detector of radio waves. It pushed radio out of the limiting confines of the laboratory into the wide world outside and gave impetus to the industry that now bears the name of Electronics. But its disadvantages weighed against it as a commercial device, the most serious flaw being its slow "recovery time." Considerable mechanical ingenuity was expended in an effort to overcome the decoherence problem, but the effort failed to achieve the degree of success required of a commercial operation. During the first dozen years of this century, inventors worked feverishly to perfect a detector good enough to replace the coherer. Their efforts in this short span of time produced a superabundance, a glut, of every conceivable device — some so fantastic as to beggar description. A sampling is listed under four main headings:

[1] *This is a modified Branley Coherer. In place of the loose filings, the device employs a tiny globule of mercury between carbon and iron electrodes. It was developed for the Italian Navy in 1899 by Thomas Tommasina.*

[2] *"The crystal detector preceded the audion and it is still the best detector when the going gets tough, as in microwaves." by Ralph Brown from foreword to William Shockley's 1950 book.*



(Fig. 4) Crystal detector employing silicon or galena.


Lee deForest

MAGNETIC DETECTORS:
Hysteresis (deForest) Magnetic Coherer (Braun)
Magnetic (Fessenden)

ELECTROLYTIC DETECTORS:
Respondor (deForest) Positive Point (Fessenden)
Tantalum (Walter) Peroxide of Lead (Brown)
Barreter (Fessenden)

CRYSTAL DETECTORS:
Silicon (Pickard) Carborundum (Dunwoody)
Galena (Braun) and 200 more oxides, sulphides, and other metallic combinations.

VACUUM TUBES:
Diode (Fleming) Audion (deForest)

MISCELLANEOUS DETECTORS:
Heterodyne (Fessenden) Bouncing Jet (Boys)
Molecular (deForest) Wave-Responsive (Shoemaker)
Flame Detector (deForest)

A reader of the thirties and forties, looking over this list of detectors, might ignore all but the vacuum tubes; the modern reader of the fifties and sixties might pause over the silicon detector of 1906; but the reader of the teens and twenties would look primarily at the list of crystal detectors, because these were the most sensitive and the least expensive. Both the popularity and the low cost resulted from the existence of over 200 different mineral combinations that could do a good job of detection and made patent rights virtually impossible to collect.

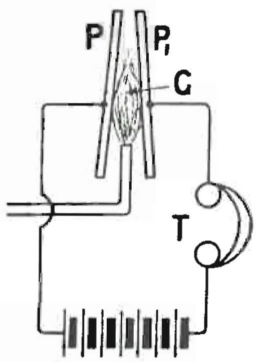
The mineral detector deserves selection as the best successor to the Branley Coherer. This second detector in one of its many mineral versions (the three noted in the above list were the most popular) dominated the radio scene from 1906 to shortly after WWI. Greenleaf Pickard began his work on mineral detectors as early as 1902 when he noticed, while testing a series of coherers, that they would occasionally work without battery current. His investigations convinced him that some of the minerals in the coherer filings had the ability to rectify radio waves. Pickard named his silicon device a "solid rectifier" which gave him the distinction of being the first to use the word "solid" in this connection. His patented device consisted of a tiny slab of pure silicon and a metal contact point whose pressure could be adjusted by means of a screw attachment surmounted by a knob (see Fig. 4).

From 1906 until after the war the crystal detector reigned supreme. Without it the army of radio amateurs and enthusiasts never would have grown to such vast proportions. Solid detectors were used on board ship and in shore installations. The audion tube was still confined to the laboratory where it continued to exhibit fine promise, but was unreliable in performance. No one knew how it worked, least of all deForest, and the "bugs" that troubled the audion remained intractable. He was forced to give up on it temporarily when his company closed down for lack of funds, and he did not take it up again until 1912. Although solid detectors dominated the field and reached some kind of peak acceptance by 1912, no significant advances comparable to the audion tube were made at a time when any kind of improvement might have been decisive for the future. There was no one around who could play deForest to Pickard's Fleming; the unveiling of the semiconductor triode was forced to wait until 1948.

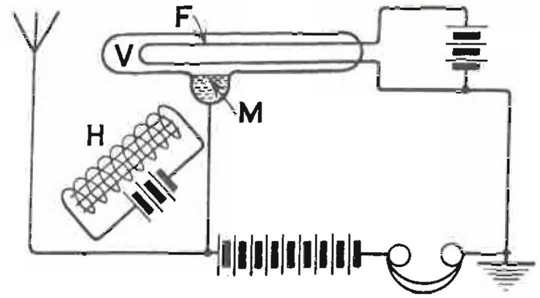
The Third Detector . . . and First Amplifier

The story usually accepted as fact is that the Edison effect led Fleming to the diode, which in turn inspired deForest to invent the triode. This is the correct chronological sequence, but we prefer the version, suggested by deForest himself, that indicates his inspiration stemmed not from the Edison effect or Fleming valve, but from another direction entirely.

Fleming had done a lot of work on the Edison effect from the moment he heard about it in 1884. In 1889 he read a paper to the Royal Society claiming that the particles flowing from filament to plate carried a charge of negative



(Fig. 5)
An arrangement similar to deForest's Flame Detector.



(Fig. 6) A gas-filled tube (V) filled with a pool of mercury (M) and a heated filament (F) electrode, forms the molecular detector patented by deForest in 1906.

electricity and could be used to convey a current in one direction only. In 1890 Fleming produced the Edison effect in air. He followed this by demonstrating the Edison effect with an electric arc. When in 1904 Marconi called in Fleming as a consultant to design a more efficient detector than the coherer, he did not have to wait long for results. Fleming invented his detector in that same year, calling his two-element diode a "thermionic valve." This valve was not considered quite so sensitive as the crystal detector, but more stable and reliable in operation.

deForest began the work which led directly to the triode in 1900. At that time he noticed a loss of illumination from the gas burner in his room whenever he switched on his spark coil. After several years of experimentation he invented his so-called Flame Detector which consisted of a gas-flame placed between two metal plates (see Fig. 5). deForest's next step was to place his flame detector inside a glass tube. Instead of the flame of a gas burner, he used an incandescent filament and added the same two plates on each side of the filament. Note that he was also working on a mercury-filled tube containing a heated filament which he called a Molecular Detector (see Fig. 6). The two-plate device was the first tube to contain three electrodes. He named it "audion," drawing the ire of Prof. Pupin who objected to the "bastard" word with its Latin prefix and Greek suffix, and was granted a patent on June 26, 1906. The third electrode and two separate battery sources increased the complexity of the circuit, creating many more circuit combinations. It quickly became obvious that the third electrode would act more efficiently if placed between the filament and plate. In his first experiment, deForest actually used a platinum plate perforated with many small holes, but changed this to a "grid" made of ordinary wire bent back and forth to reduce cost. The grid-audion was patented Jan. 15, 1907. The 1906 audion was still called a detector, but the title of the 1907 patent reads: *"Device for Amplifying Feeble Electrical Currents."* No one was quite sure how it worked (Pupin's comment was entirely philological) and it remained something of an experimenter's curiosity until 1912. While its sensitivity was superb, it was unreliable in operation because of a tendency to break into uncontrollable oscillation. Nor did the "softness" of the vacuum help matters since it prevented the use of high anode voltages.

The genesis of the audion tube from its unlikely inception as a flame detector to its eventual metamorphosis into the first grid-audion is still a masterpiece of the creative imagination no matter how poor its future may look in this Solid State Age. But there is a feature of deForest's invention which is not only still with us, but forms the basic pattern which all solid-state amplifiers are designed to fit, and this is the employment of a local source of power other than the filament supply. In deForest's words: ". . . *I was employing the high-frequency energy, not to actuate a telephone diaphragm, as Fleming had done, but to control very much larger quantities of energy from the local battery."* This remains deForest's triumph even in the Solid State Age.

It is difficult to communicate the utter indifference exhibited by the world of that day – both layman and expert – to deForest's invention. Perhaps some idea of the state of affairs can be gleaned from the comments of the government prosecutor when deForest's company was charged with using the mails to defraud in May 1912. The prosecutor's statement read: *"This is a company incorporated for $2,000,000 whose only assets were deForest's patents in a strange device which he called the audion and which device had proven worthless – even as a lamp."*(!) deForest was exonerated, but three of his associates went to jail. Hardly a just verdict, since the fraud was based on the "worthlessness" of the audion. Had any scientist or engineer of note testified to the audion's great potential value, he would have saved three men from a jail sentence and won for himself posterity's acclaim as a prophet.

1912 was deForest's last year of tough luck, financial and otherwise. He learned to stabilize the audion and realize substantial gain by re-exhausting the vacuum, and he demonstrated to the telephone company a cascaded amplifier of three stages with a gain of five per stage, and a net overall gain of 125. A.T.&T. bought the repeater (or wire) rights to the audion for $50,000, a trifling price even in those days, but deForest was on the verge of bankruptcy and needed money desperately.

In October 1913 a penniless student at Columbia University, Edwin Armstrong, was granted a patent on a regenerative (or feedback) circuit. In describing his invention, he was able to explain lucidly just why oscillation occurred, and why it had to be controlled to achieve the amazing degree of amplification of which it was capable. Looking back at the many circuits he had tried, deForest discovered one that was similar to Armstrong's, and promptly applied for a patent in March 1914 claiming prior discovery. Many years later the courts ruled in favor of deForest, but the IRE awarded its medal of honor to Armstrong and reaffirmed the award *after* the court's final decision. In 1914, A.T.&T. bought the radio rights to the triode for $90,000.

The Pros Take Over

And now the lone inventor, the amateur whose only stock in trade is his creative imagination, was brushed to one side and the professionals took over. A.T.&T., GE, Westinghouse, and RCA put their physicists, chemists, metallurgists, mathematicians, and engineers to work on their new baby – adopted of course – the triode vacuum tube, only recently born in an ill-equipped laboratory and spoon-fed to some degree of early maturity by distraught parents in great travail and worry. Teams of engineers and committees of scientists designed new bases for the vacuum tube, new glass crockery for bulbs, and metal cans for power tubes; grid after grid was added, cathodes were clothed with new materials, multiple devices enclosed in one envelope, and deForest's beloved gases were once and for all banished from the vacuum tube amplifier.

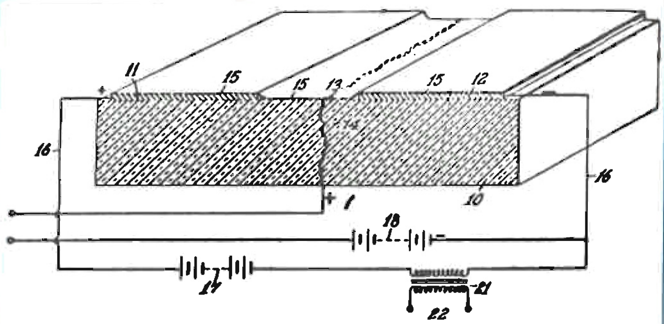
But where were the investigators of solid devices? Was anyone working to transform the solid diode into a solid triode? deForest was able to design a triode tube just two years after the diode tube came into existence, and make it perform magic in six years. Why didn't someone perform the same magic with the solid crystal diode? With millions of crystal sets in operation, and the vacuum tube triode as a guide, the appeal to the imagination of a solid triode must have been exceedingly great. Didn't anybody try to insert a third element in the crystal detector? Just what condemned us to wait until 1948? You can be assured of one thing: not one, but many lone inventors were on the trail of the semiconductor triode, but in the heyday of the vacuum tube their work was ignored, nay, suppressed. Too much money had been sunk into vacuum tube technology – the die had been cast so far as the "pros" were concerned. Anything so obviously cheap as a crystal triode would receive no hearing from the people who were baby-sitting the vacuum tube.

The Two Camps of Science

Most people who think about it, subscribe to the "two camp" picture of science; one camp called "pure" and the other "applied." Though not necessarily exclusive, it is assumed generally that the camp of "pure" science is first on the scene, and that of "applied" science follows in its wake. The purists dabble without ulterior purpose, then the appliers select for purposeful application. The "two camp" picture of science fits the history of 19th century science perfectly, as is evident when one examines the work of such dabblers as Faraday and Hertz. From a single fundamental experiment of Faraday grew the vast technology of electric power; from a single fundamental experiment of Hertz grew the equally vast technology of radio communication.[3]

Among the few idealistic phrases one meets with in the field of technology, the most curious is "pure research." It is not our intention to determine with Ivory soap precision the purity of pure research. But purity is not a feature of 20th century science. Faraday began his work at the Royal Institution of London, Hertz worked at various universities in Germany. But our schools of learning depend

[3] *Paraphrased from "The Growth of Physical Science," the section on Electricity, by Sir James Jeans.*



(Fig. 7) J. E. Lilienfeld's solid apparatus for amplifying oscillating currents. Patent first filed in 1925.

on federal grants for research work, and as Prof. Teller phrased it: *"Most of our federal expenditure is used to support applied science and the engineering developments based upon applied science."*[4] The picture of the young Einstein beginning his career as an examiner in a patent office seems to symbolize this century.[5] The largest and best equipped laboratories are interested only in improving products that have proven themselves in the market place. The lone inventor who follows the dictates of his own creative imagination is doomed to a marginal existence in our century. But that won't stop him. He is there somewhere, working in the "underground."

The Fourth Detector and Ultimate Amplifier

Bolts from the blue in scientific invention are rare. For every striking invention that hits the front page, one can always find a forerunner, a predecessor, if one digs long enough. This is the case with the semiconductor triode. Any one of the three basic ideas listed below could have developed into a solid state amplifier long before 1948 if well-equipped, well-staffed laboratories had gone to work on them promptly.

Our first exhibit comes from a book published in England in 1923.[6] The excerpt is quoted verbatim and in full in order to retain the flavor of the original.

Crystal Circuits Which Amplify

"Mention should be made of an invention of M. Lossev, of Nijni-Novogorod, who discovered that, in addition to rectifying properties, certain crystal combinations will function as amplifiers within a limited range. Several combinations have been found to possess this property, one being zincite used in conjunction with a steel point. It must not, of course, be supposed that the crystal itself magnifies — it merely serves, as a valve functions, to impress fluctuations in received oscillations on an electric current. A battery is essential, to provide a source of electrical energy. Instead, however, of having to employ a high-tension battery, a low-tension battery of a few volts is all that is necessary. The further development of this discovery may revolutionize broadcast reception, for the small first cost of a crystal, as compared with that of a valve, and the advantage of low-tension as against high-tension batteries, must instantly appeal to everybody interested in the art of wireless."

Our second exhibit is the work of Julius Edgar Lilienfeld, an American of Brooklyn, USA. It is described as a "Method and Apparatus for Controlling Electric Currents," and application for patent was filed in Canada October 22, 1925, in the USA October 8, 1926, and patent 1,745,175 was granted January 28, 1930.[7]

This is what Lilienfeld says his device will accomplish: *"The invention relates to a method of, and apparatus for, controlling the flow of an electric current between two terminals of an electrically conducting solid by establishing a third potential between said terminals; and is particularly adaptable to the amplification of oscillating currents such as prevail in radio communication. Heretofore, thermionic tubes or valves have been generally employed for this purpose; and the present invention has for its object to dispense entirely with devices relying upon the transmission of electrons thru an evacuated space . . . The invention has for a further object a simple and inexpensive amplifier not involving the use of excessive voltages, and in which no filament is present."*

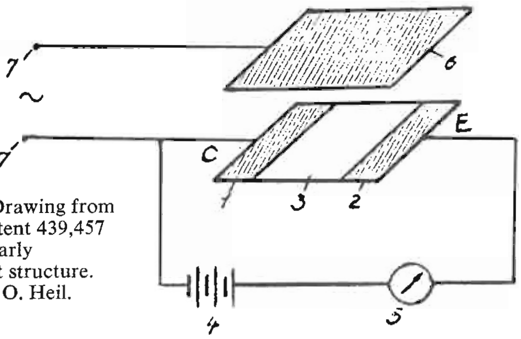
Fig. 7 is a greatly enlarged and partly sectional view of the "apparatus." 10 designates "a base member of suitable insulating material, for example, glass." 11 and 12 are a pair of conducting members such as a coating of platinum,

[4] *"The Role of Applied Science" by Edward Teller, Bulletin of the Atomic Scientists, March 1966. In this article Prof. Teller defends the role of applied science. However, he notes that: "Recently I interviewed 24 of the most promising students from various departments of MIT . . . The interviews revealed that 22 out of 24 students showed a marked preference for pure science. (This in an institution) supposed to have a particularly close connection with technology."*

[5] *In later years Prof. Einstein dismissed his job in the Swiss patent office as a "shoemaker's job." But he was not above taking out a patent himself on a metering device which could measure tiny electric discharges.*

[6] *"Wireless" by P. J. Risdon, Chapter XXII. Lossev's correct initials are O. V. The M in Risdon's report stands for Monsieur.*

[7] *The working life of a patent is 17 years. It may be extended as a result of interference proceedings.*



(Fig. 8) Drawing from British patent 439,457 showing early field-effect structure. Inventor: O. Heil.

gold, silver, or copper "which may be provided over the glass surface by well-known methods such as chemical reduction, etc." 13 is an electrode of aluminum foil which is of minimum dimensions (.0001") to reduce capacity effect, and "is secured in position by providing a transverse fracture (14) in the glass and then reassembling the two pieces." Finally, superimposed over 11 and 12 is a thin film of "uni-directional conductivity" such as a "compound of copper and sulphur."

What we see in Fig. 8 is perhaps the first recorded version of the field effect transistor. Crude though it may appear by today's standards, it still bears a remarkable resemblance to the modern versions illustrated in Part VII.

Our third and last exhibit is the work of Oskar Heil of Berlin who was granted a British patent on a version of a field effect device. Fig. 8 is described as follows:[8] *"The light area marked 3 is a thin layer of a semiconductor such as tellurium, iodine, cuprous oxide, or vanadium pentoxide; 1 and 2 designate ohmic contacts to the semiconductor. A thin metallic layer marked 6 immediately adjacent to, but insulated from, the semiconductor layer serves as the control electrode. Heil describes how a signal on the control electrode modulates the resistance of the semiconductor layer so that an amplified signal may be observed by means of its current meter 5."*

The Mysteries of Scientific Progress

When vacuum tube diodes performed poorly as detector/mixers at the microwave frequencies common to radar equipment, designers turned back a few decades to the crystal detectors which they found could do a better job than tubes. This necessitated another look at crystals – particularly the basic materials and the theory of operation. When Bell Telephone Laboratories began working on these subjects it was even theorized that a crystal amplifier might be one of the results of this investigation. Ralph Brown, then Director of Research at Bell Labs, has written of this work as follows: *"In the course of three years of intensive effort the amplifier has been realized by the invention of the device named the transistor."*[9]

Who is to say that a similar effort, begun earlier, and perhaps of longer duration, might not have developed the transistor in 1928 or 1938 rather than 1948? Perhaps the old triumvirate of vacuum tube pros – GE, RCA, Westinghouse – who are huffing and puffing all over the place trying to keep pace with the new triumvirate of transistor pros – Fairchild, Motorola, TI – (all listings strictly alphabetical) might be higher on the transistor pole if their aspirations and energies were not divided between vacuum and solid state.

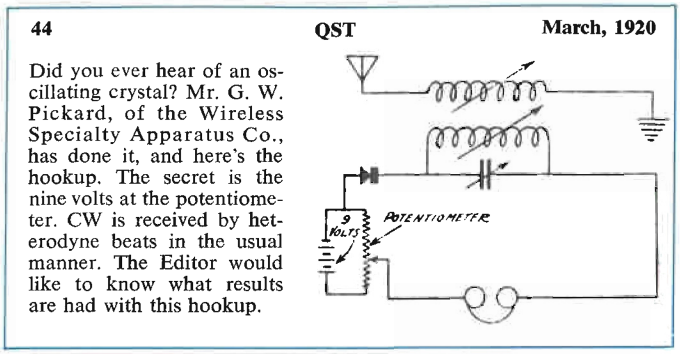
"I do not know of a single vacuum tube company or part of a company that successfully got into the transistor business.
– J. Herbert Hollomon, Assistant Secretary of Commerce for Science and Technology. June 1967.

[8] *"The Field-Effect Transistor – an old device with new promise." by J. T. Wallmark, IEEE Spectrum, March 1964.*
[9] *From foreword to William Shockley's book of 1950.*

PART III — HEROES & VILLAINS

Prologue

A story requires a hero and a villain, and *The Semiconductor Story* is no exception, – but where are we to find a hero and a villain? There is always the touchy interplay between the "pure" and applied sciences. No one with a reputation for "purity" would ever be cast in the role of villain. The reprobate in this context would be the applied scientist. But perhaps he likes it that way since the sinner is always more interesting than the angel. Well, that's not the way we see it. In our book applied science is every bit as heroic as pure science. Definitely, the applied scientist is not our villain – on the contrary, the applied-scientist-



44 QST March, 1920

Did you ever hear of an oscillating crystal? Mr. G. W. Pickard, of the Wireless Specialty Apparatus Co., has done it, and here's the hookup. The secret is the nine volts at the potentiometer. CW is received by heterodyne beats in the usual manner. The Editor would like to know what results are had with this hookup.

(Fig. 9)

Dr. William Shockley, inventor of the junction transistor.

turned-inventor is our hero. At best the pure scientist must assume a neutral position in our story because his primary interest is in a goal beyond our more limited horizons. If there is a villain, he must be the industry itself, or rather the older segment of the industry, which has shown itself content with the status quo and with the superficial refinements that make for periodic obsolescence yet leave the basic product unchanged. The villain is the sector of the industry that frowns on the big breakthrough that obsoletes the old money-making devices with better, and sometimes cheaper, superseding devices. The classic example is FM broadcasting, but almost as good a case can be made for the transistor.

State of the Art in the Dirty Thirties

In our previous issues we outlined the Semiconductor Story stepwise up to the heyday of the vacuum tube era. After its apparently complete victory, the industry abandoned all research that might lead to more efficient methods of radio detection and amplification. Of course, solid state physicists continued the "business" of pure research unaffected by the success of the vacuum tube. And marching in the vanguard of progress, as always, the "underground" inventor, spurred on by his own creative imagination, explored the semiconductor crystal in the midst of the deafening hosannas to the glories of the vacuum tube, convinced that he was on the track of something much better than the tube. And, of course, he was right. The early inventor dabbling in semiconductors required no fancy theories or intricate technologies to list the areas of solid state supremacy, to wit: the ultimate in microminiaturization this side of molecular electronics, minute power requirements permitting the use of such sources as solar batteries, no need for internal heating resulting in instantaneous operation, and lower costs. Even in 1930 there was no denying these advantages because they originate from the very nature of the basic materials that go into the making of semiconductors.

The devices dreamed up by these innovators did not make the grade because no one knew why they worked as they did. The "amplifying" crystal diode that Lossev described in the early twenties was a point contact device. So was the crystal used by Pickard as a heterodyne oscillator[1] at about the same time (See Fig. 9). Precursors of the tunnel diode? Perhaps. As Babits noted:[2] "*. . . many decades passed before the important tunnel diode was invented, but it is, nevertheless, worth mentioning (the prior art of Lossev and Pickard) because the junction transistor came from the point contact transistor.*"

Call them curiosities if you like, but they were curiosities worthy of continued study by well-equipped and staffed laboratories. Instead they were ignored by those in the "corridors of power," and survived as interesting auguries of the future, waiting for theory and technology to catch up with them – a theory that would explain their operation and a technology that could easily reproduce them.

Unhampered by the monetary considerations of 'vested interests,' theory did not wait long to come up with an explanation, – technology was another matter entirely. The theory appeared in print in 1931 and was written by Alan H. Wilson. It proposed a theoretical model of semiconductor operation based on quantum mechanics. (Note that quantum mechanics is a considerable advance over the quantum theory and took thirty difficult and brilliant years to evolve from the earlier theory.) In his summary, Wilson stated: "*When the temperature is different from (absolute) zero there will be a few excess electrons in the second energy band and a few* vacant places *in the first band, and conduction will follow.*" Compare this with Shockley's opening paragraph in his 1950 book: "*Basic to the theory of semiconductors is the idea that electrons can carry current in two distinguishable and distinctly different ways: one being called by 'excess electrons,' or 'conduction by excess electrons,' or simply 'conduction by electrons' and the other*

[1] *I am indebted to Harry E. Stockman, S.D. of Arlington, Mass. for the QST reference to Pickard's oscillating crystal diode.*

[2] *"Prior Art of Electron Devices" by Victo A. Babits in Technical Correspondence of IEEE Spectrum of December 1967.*

Dr. John Bardeen and Dr. Walter H. Brattain, inventors of the point-contact transistor.

being called 'defect conduction' or 'conduction by holes.' The possibility that these two processes may be simultaneously and separably active in a semiconductor affords a basis for explaining transistor action."

Wilson's theory of semiconductor operation was passed over in silence. *"In fact, it took about fifteen years for the full light to dawn,"* said Pearson and Brattain in 1955. They added: *"This was a case of plain lazy thinking on the part of many investigators."* Today one often sees statements claiming that "limited knowledge of semiconductor physics at that time was responsible for the failure to follow up on earlier inventions." It might be argued to better effect that lack of a materials technology was a more serious hindrance, but most of all the vested interest of the electronics industry in vacuum tubes was the wet blanket that stood in the way of rapid semiconductor progress. Or, as some people politely phrase it, ". . . interest in this early work waned as vacuum tube technology continued to advance."

To the radio industry the year 1931 seemed a particularly inappropriate time to conceive a theory of semiconductor operation. But Wilson was a physicist primarily, and the year 1931 saw the developing principles of quantum mechanics reach a peak from which it has been declining ever since.

It may be interesting to delineate ever so lightly the environment in which Alan H. Wilson worked. Although he taught in Cambridge and published his paper in London, his research was done in the *Institut fur Theoretische Physik* in Leipzig under the direction of Prof. Werner Heisenberg to whom he expresses a debt of gratitude at the end of his paper. Heisenberg published his book on "The Physical Principles of Quantum Mechanics" in 1930, but his name is more closely associated with the Uncertainty Principle which establishes the "tolerance" limits of a particle's position and velocity. For visible objects these tolerance limits are so small as to be of no importance whatsoever, but in the world of subatomic particles the tolerances are so large that calculating position and velocity at the same time verges on the impossible. To quote Max Born: *"If Gessler had ordered William Tell to shoot a hydrogen atom off his son's head by means of an alpha particle and had given him the best laboratory instruments in the world instead of a cross-bow, Tell's skill would have availed him nothing. Hit or miss would have been a matter of chance."* Heisenberg received the Nobel prize in 1932.

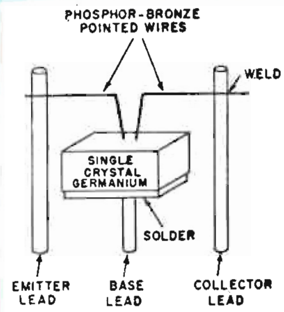
Another name that belongs in our picture of this period is Paul Dirac. He could find no satisfactory employment as an electrical engineer and decided instead to apply for a fellowship in physics at Cambridge University. In 1930 he published his paper on anti-particle physics in which he treated the vacancy left by a missing electron as a positive charge. His prediction that these "holes" existed thruout the universe aroused vigorous opposition until experiments confirmed his prediction. In fact, for every particle in the atom an anti-particle has been discovered. Paul Dirac was co-winner of the Nobel prize in 1933.

Finally there is Sir Arthur Eddington, Professor of Astronomy at Cambridge, and author of "The Mathematics of Relativity," who was preparing in Sir James Jeans' words, *"a vast synthesis that may in time explain the nature of the world we live in,"* when death intervened in 1944. To Eddington, Einstein wrote in 1925, *"I would so much like to talk with you that for this alone it would be profitable to learn the English language."* As a friend of Eddington, A. H. Wilson wrote a memorial tribute which appeared in the Cambridge Review recalling some of their conversations. This is the milieu in which Alan H. Wilson lived and worked.

In the decade between '22 and '32 one encounters in the solid state literature of the period practical and theoretical discoveries of great brilliance. In the following decade, ending in '42, follow-up achievements are few and far between.

Triumph of the Transistor

The dynamite that upset the cozy order of things in the electronics industry was provided by the second world war. A series of laboratories were established across the country



(Fig. 10) Construction of a point-contact transistor.


(Fig. 11) Silicon point contact diodes designed for video mixer and detector applications. *(Courtesy of Alpha Industries)*

in which, among other things, much work was done to improve the crystal diode. It was proving itself to be an excellent detector and mixer in radar equipment. After the war, these labs continued their operations although the wartime emphasis on device discoveries declined in favor of a stronger interest in purely theoretical research, hence the rediscovery of Wilson's theory. Within three years, in 1948, the invention of a point contact transistor, the first transistor to go into commercial production, was announced by J. Bardeen and W. H. Brattain of Bell Telephone Labs. Almost at the same time, W. Shockley applied for a patent on the junction transistor.

These two versions of the transistor differ considerably in construction though not so much in operation. The point contact device consists of two closely spaced (.0025″) metallic points, or "cat whiskers," attached to the surface of a semiconductor crystal (See Fig. 10). It is really a point contact diode with a second point contact added. That the breakthrough occurred in this type of structure is not surprising since the men involved and their associates had been working on point contact diodes since the early days of the war.

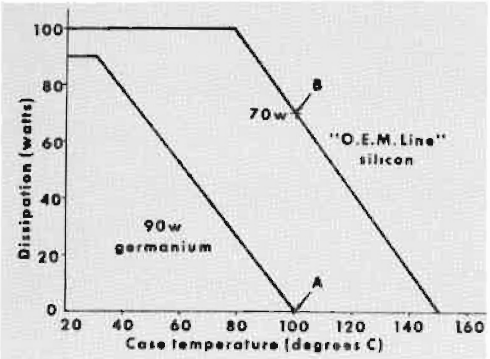
Initially the point contact transistor made some progress in the marketplace because its frequency response was higher than the first junction types. But (like similar inventions of the 20's and 30's) it proved rather difficult to reproduce because its theoretical design was unnecessarily complex. Some of these critical problems were overcome by another metal-to-crystal process developed by the Philco Corporation called the surface-barrier transistor. In this type the flat sides of the crystal are etched until two concave holes are formed, leaving an extremely thin region for the base. Then the emitter is electro-plated into one cavity and the collector into the opposite cavity. The thin base region keeps the frequency response high, but lowers the voltage breakdown rating and limits its usefulness to low power applications.

The point contact transistor, in its twentieth year, is now almost extinct, although the point contact diode (See Fig. 11) with its 50th birthday almost upon us, is still very much with us. But the first working transistor must always stand as an example of what can be done when trained men, given a proper environment, are turned loose on a problem.

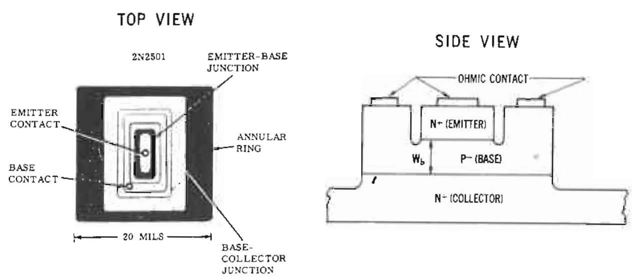
The junction transistor, in spite of its name, is not made of three solid chunks of crystal mechanically joined together. It is, in fact, a single crystal and the junctions are created by injecting the same impurity at each end of the crystal, and injecting an impurity of opposite polarity into a narrow central section. The bulk of the impurities are relatively minute, and so is the spacing between the two junctions (.001″ max.). Since most of the electrical characteristics of the transistor are dependent on its structure, the simple geometry described above has been altered significantly in the search for improved devices. But the basic relationship of the junctions remains the same.

With its inherent stability, higher power capability, and lower leakage rate, the single-crystal junction transistor did not take long to supplant the metal-to-crystal type once the problem of high frequency response was solved. Because junction type design lends itself to considerable variation, and its characteristics are strongly structure dependent, a stiff race developed among the practitioners of the art to make a better transistor than anyone else. The result of this keen competition may be seen in Table I. One product after another takes the center of the stage only to be outstripped by a later design. The chemists, the metallurgists, the ceramicists, the crystallographers, the physicists are still hard at work and the end is not yet in sight. In 1957, tabulation of transistors in the D.A.T.A. manual numbered 650 types. In 1964 this had grown to 6,500 different transistors, an increase of 1,000%.

A detailed appraisal of the following table requires some understanding of the working of a transistor. Since the transistor is not found in nature ready to operate like some crystal rectifiers and magnets, but is put together in a factory like any other mechanical contrivance, an understanding of this mechanism is also essential.



(Fig. 12) Power dissipation versus case temperature.



Two views of transistor geometry.

Transistor Type	Approx. Date	Manufacturing Process	Originator
Metal-to-Semi-conductor	1948	Point contact	Bell
	1952	surface-barrier	Philco
Junction	1950	Diffused-alloyed	G.E.
	1952	Rate-grown	G.E.
	1952	Meltback	G.E.
		Grown-diffused	T.I.
	1954	Drift	R.C.A.
	1958	Microalloy diffused (MADT)	Philco
	1959	Diffused-base mesa	Bell
	1960	Silicon planar	Fairchild
	1962	Plastic encapsulation	G.E.
	1965	Annular	Motorola

Without delving for the moment into the structural developments that play a major role in the contents of the table above, other trends may be seen at work there. One of the most important is the switch from germanium to silicon. Fig. 12 shows why. Note that where germanium is already derated to zero power at 100° C, the silicon unit can still dissipate 70 watts at the same case temperature.

If this was the entire story, germanium would disappear from the market overnight. But the mobility of hole carriers in germanium exceeds that in silicon by five to one. The electron mobility of germanium is also favorable by three to one. Therefore, at room temperature and at conservatively rated specifications, one can expect better performance from an equivalent germanium type.

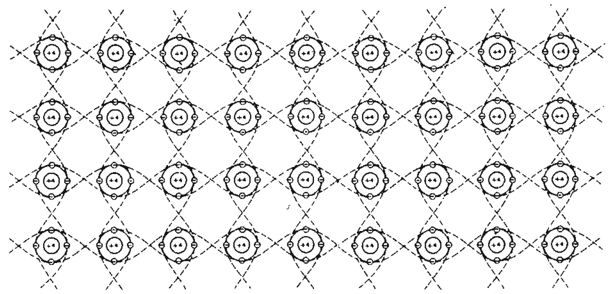
This is a classic example of the trade-off, a designing tool that the designer is not particularly proud of, but also not too proud to use. Don't expect the maximum in temperature rating, the maximum in gain, the maximum in voltage breakdown, the maximum in frequency response, etcetera, from one and the same unit. Compromise is the basis of most designs.

Another area of recent interest is the Field Effect Transistor (FET) which has been showing consistent gains in the past few years, particularly with the development of the metal-oxide-silicon (MOS) technique. The granddaddy of the FET is Lilienfeld's device patented forty years ago. Not listed in Table I are the various "active element" diodes: the tunnel (Esaki) diode, the Varactor, and the Gunn Effect diode. The ancestors of these devices are Lossev's "amplifying crystal" and Pickard's "oscillating crystal" of the early 1920's.

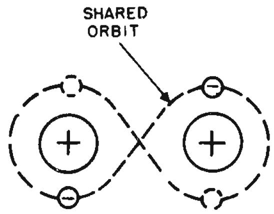
PART IV — THE MECHANISM OF A TRANSISTOR

The transistor is a mechanical contrivance manufactured from an assortment of raw materials with the aid of complicated machinery and test equipment. An understanding of the form and structure of this mechanism is a necessary prelude to an understanding of its operation. It may be that the reader has already acquired a vague picture of the composition of a typical transistor. If he imagines this vagueness is due to the fact that he is dealing with invisible particles, he would not be entirely correct. The current carrier in the vacuum tube is also invisible, yet its operation is easy to visualize because the particles flow inside a glass-enclosed vacuum containing nothing except a set of metallic elements which instigate the flow of particles and control it. Obviously, this kind of picture poses no serious problems to the imagination. But when we are told that these and other invisible entities flow thru a solid body of material which is neither completely conductor, nor completely insulator, we may be pardoned for having a little trouble with our mental pictures.

There are many similarities between the vacuum tube and the transistor; even more interesting are the dissimilarities. The structural elements of a vacuum tube may be



(Fig. 13) Two dimensional crystal lattice structure.



(Fig. 14) Covalent or electron — pair bond of atoms.

examined without the assistance of a microscope. A metal filament called the cathode is fixed in the middle of an evacuated glass tube; surrounding it is a mesh of wire called the grid; this in turn is encircled by a metal plate called the anode. This is the basic structure of the vacuum tube. Its operation is just as straight-forward as its structure. When the cathode is heated by a local source of energy it will throw off a swarm of electrons that will form a cloud around the cathode. The application of a positive voltage to the anode by a second local source will cause it to attract the electron cloud, which will pass thru the grid on its way there. The amount of electrons that reach the anode can be regulated in the manner of a valve by applying appropriate bias voltages to the grid.

The solid transistor, by comparison with the vacuum tube, seems far more complex because it contains within its solidity not only one type of subatomic particle, but a variety of such particles, each with its own personality, and all of them playing a part in its mechanism. However, it would be a mistake to over-emphasize this complexity. After all, the crystalline structure of which transistors are made is by definition an orderly arrangement of atoms. This atomic arrangement extends uniformly in all directions, repeating its pattern like wallpaper, and contains no distinct groupings of atoms or molecules. In fact, the whole crystal may be looked upon as a single molecule! The strength of the crystal pattern is so great that the presence of impurities, while causing an occasional defect in the lattice, can never destroy its basic arrangement. The impurity that finds itself trapped inside a crystal is forced to change its appearance to conform to the pattern of the crystal as best it can.

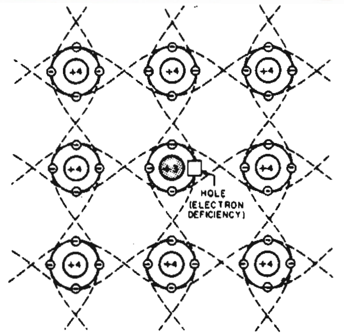
Transistor action takes place inside the crystalline pattern just as the action of a tube takes place inside the glass-enclosed vacuum. If the reader feels more at home in a vacuum, we are happy to tell him there is one in the semiconductor crystal. The distance separating the atoms of a crystal is a four hundred thousandth part of an inch (4×10^{-5} in.) and a free electron can roam about in this vacuum for perhaps an eight thousand millionth of a second (8×10^{-9} sec.) at ordinary temperatures before colliding with an electron fixed in the lattice structure. If the temperature is increased, the velocity of the free electron will also increase, possibly to the point of being able to knock a fixed electron out of the lattice so it too can roam about freely. The creation of a free electron by collision will produce two other effects. First, the departure of the fixed electron from its position will leave a "hole" in the crystal structure. Second, by losing an electron with its negative charge, the neutral state of the atom changes to a net positive charge and becomes an ion.

If we stop at this point to take an inventory of the particles found in the crystal, we find both free and fixed electrons, both neutral and ionized atoms, and last but not least, electron deficiencies called "holes." It should be added that we assume the crystal to be pure, and subject only to thermal agitation.

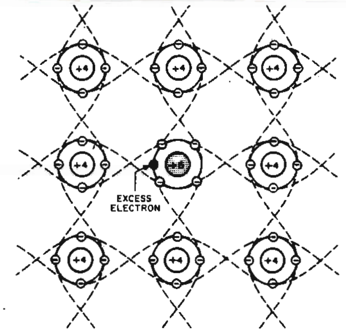
Although the manufacture of a transistor always begins with a purified crystal, transistor schematics never indicate the purity of the crystal, nor will you find the perfection of the vacuum indicated in vacuum tube schematics. The pure crystal and good vacuum are taken for granted. Creating a good vacuum was the first major achievement of vacuum tube technology. Similarly, purification of silicon and germanium by synthesis and refining was an important first step in transistor technology. It is hardly an exaggeration to say that the improvements in the refining of crystal materials speeded up the invention of the transistor. In its natural state, silicon contains about 2% of impurities. After undergoing refining, impurities may be reduced to as low as one foreign atom to every billion silicon atoms. By controlling the type of impurity, the quantity of the impurity, and the exact site of injection into a crystal of maximum purity, we have taken a giant step towards the manufacture of easily duplicated semiconductors.

It's time now to take a look at a pure crystal structure (see Fig. 13)[1]. What we see is an uninterrupted layer of

[1] *Figures 13 through 20 are from the GE Transistor Manual, 7th Edition, published 1964 by the General Electric Company and reproduced by courtesy of that company.*



(Fig. 15) Creating a hole in a crystal lattice.



(Fig. 16) Creating an excess electron in a crystal lattice.

germanium (or silicon) atoms bonded together by electron pairing. The pairing is the result of two neighboring atoms sharing the same orbit (see Fig. 14). However, even a highly refined crystal would not look this way, and if by some miracle it did, it would have to be stored at a temperature of absolute zero to retain its purity.

It is more likely that the crystal structure will reveal some "defects" such as "holes" and excess electrons as well as some ions. These defects may be caused by heat and by accidental impurities. But it is possible to create such defects intentionally, which means that we can select one crystal site to contain an abundance of holes and another site an abundance of excess electrons. This would enable us to arrange the crystal polarities in any pattern we choose. Analogy with the vacuum tube would point to a three-step pattern such as: holes-electrons-holes, or electrons-holes-electrons. When we achieve this three-step pattern in a crystal, we have created the "elements" that cause transistor action, just as cathode-grid-anode elements cause tube action.

Thermal agitation can create holes, but only as the result of taking an electron away from an atom. Sure enough this will leave behind a hole, but it will also ionize the atom. There is another way to make holes without ionization (or the removal of electrons from their atomic positions). As we can see from Fig. 13, silicon and germanium possess four electrons in their outer shell. This is balanced by an equal number of positive charges in the nucleus. Now, let us choose an atom which has only three electrons in its outer shell and inject it into the crystal. The atom will take up a position in the crystal lattice, but because the outer shell contains only three electrons, a hole appears in the pattern where an electron should be. The atom itself, however, will still be neutral because the positive charge in its nucleus matches its smaller electron family (See Fig. 15). An example of such an atom is indium.

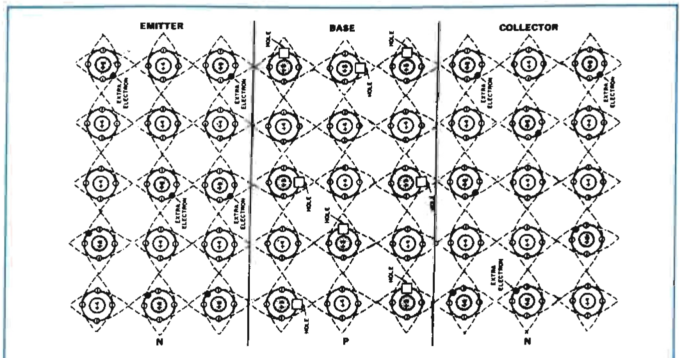
To increase the number of excess electrons in a crystal without resorting to stealing them from neutral atoms, we select an atom possessing five electrons in its outer shell and inject it into our crystal. When the atom takes up its position in the lattice, the fifth electron will become as useless as a fifth wheel on a car as far as the lattice structure is concerned (See Fig. 16). Arsenic is an example of such an atom. It is worth noting that whereas thermal agitation creates holes and excess electrons at the same time, preselected impurities create only holes or only excess electrons, but not both at once.

The injected impurities can be arranged in three consecutive areas: electrons-holes-electrons, or in reverse order, holes-electrons-holes. If we substitute the polarities of the current carriers for their proper names, the three step pattern reads negative-positive-negative, or positive-negative-positive. By using initial letters, this can be shortened to NPN or PNP for the sake of brevity. The three areas also have functional names: emitter-base-collector. Because they describe functions, these names apply regardless of the polarity of the current carrier.

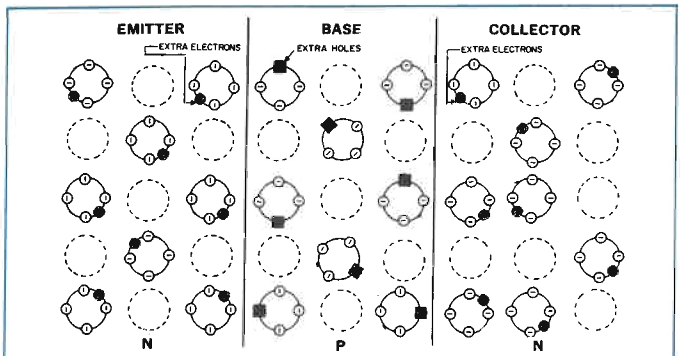
Our interest is now centered on the current carriers which exist in each of the three regions that constitute the transistor, and which we know have been for the most part deliberately introduced into the crystal. A fragment of such a transistor, in this case of the NPN type, is represented in Fig. 17. As an NPN type the majority of the current carriers in the emitter and collector regions are free electrons, whereas the majority in the base region are holes.

In the interest of clarity let us dispense with the lattice structure of Fig. 17 and indicate only the charge carriers with their associated atoms as in Fig. 18. The emitter region with its negative charges faces the positive charges in the base from one side, while with similar polarities the collector faces the opposite side of the base. The two boundary lines which separate the opposed charges are called junctions, hence the name junction transistor.

We come now to an important problem. On each side of the junction line we notice negative carriers (excess electrons) opposing positive carriers (holes). Since these charges are of opposite polarity, why don't they attract each other? Or, to be more specific, why don't the excess



(Fig. 17) NPN type semiconductor.



(Fig. 18) Charge carriers before junction formation.

electrons in both the emitter and collector areas cross the junctions which are, after all, just dividing lines, and fill the holes in the base region?

Well, the answer is they do!

But just for a split second!

Then equilibrium is restored and there is no further migration across the formed junction (See Figs. 19 and 20). What stopped the re-combination of all the holes and excess electrons? The same natural law that caused the momentary joining of electrons and holes has also operated to bring the action to a halt. This is the law which says that opposite charges attract each other and like charges repel each other.

A step by step explanation would go this way:

1. The impurities that are added to the crystal have one too many electrons in emitter and collector or one too few in the base as compared with the atoms in the crystal lattice. However, all the atoms of the impurities are electrically neutral.

2. When, at the moment of junction formation, the extra electrons are attracted to the holes in the base, this neutral condition is upset. The impurity that normally owned five outer electrons loses one and becomes ionized in the positive direction; the impurity that normally owned three outer electrons acquires one and becomes polarized in the negative direction.

3. At this moment the P side of the junction is crowded with atoms that are ionized negatively and the N side is crowded with atoms ionized positively. The junction which started out to be merely a line of demarcation between areas negatively and positively "doped," has now acquired a narrow region on each side of this line where carriers of opposite charges have been depleted in order to create a double layer of ions that can act as a barrier to prevent unlimited recombination thruout the semiconductor crystal.

The transistor is now in a state of equilibrium and will remain so, if properly stored, until external sources of energy are applied to the terminals of its three elements. These three elements and the two junctions to each of which have been added a set of depletion regions constitute the mechanism of the transistor. It may be said of this mechanism that it was implanted into the lattice-work of the crystal which serves as host to the mechanism — houses, feeds, and nourishes it.

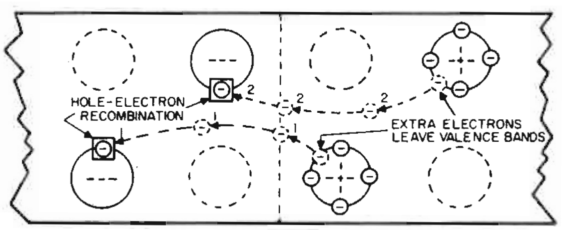
PART V — TRANSISTOR ACTION

A Closer Look at the Junctions

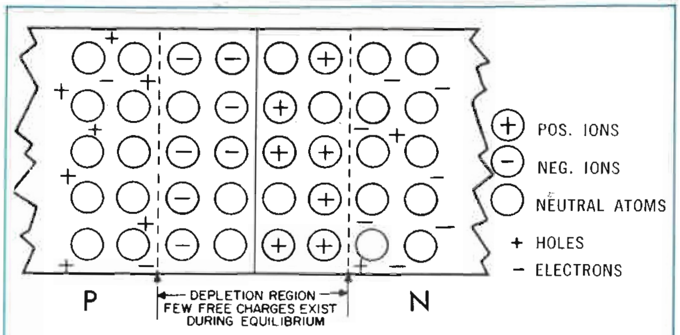
The topography of a transistor looks like a map of three contiguous continents, each containing its own flora and fauna. The continents are arranged consecutively so that they are separated by two borders. Although these borders limit the continents to a pre-determined area, they also join the continents together. And it is the joining, rather than the limiting, function that is more important — a joining in its all-inclusive sense of transfer, exchange, and union. It is at the junctions that one can best observe transistor action.

And action at the junctions begins at the moment they come into existence. This should not be surprising when one remembers that in the embryonic transistor the junctions are the only lines separating unruly mobs of oppositely charged carriers. A mad rush ensues at the junctions, and when the fun is over, we discover that almost all the charge carriers close to the junctions have been depleted by the process of re-combination. What has come into existence as a result of this initial mad rush is known as the *depletion region* (see Fig. 20).

We can list other significant changes in and around the junctions, each one important enough to warrant the assumption of its own private name. A most significant change occurs when the depleted charge carriers have been absorbed in the re-combination process by formerly neutral atoms which now become transformed into ions (see Fig. 20). The polarity of the ions will depend on whether the



(Fig. 19) PN junction at instant of forming.



(Fig. 20) PN junction at rest.

acquisition by the atoms were negative charges (electrons) or positive charges (holes). The negative ions will get their additional electron from the N-type emitter and collector; and the positive ions will lose an electron to the holes in the P-type base. The result will be two columns of ions, lined up opposite each other and of course oppositely charged as shown in Fig. 20. A similar line up will take place at the collector junction. These arrangements prevent further re-combination, and become in effect *energy barriers* (see Fig. 21a).

Still another electrical phenomenon is suggested by the existence of a negative potential on one side of the barriers and a positive potential on the other side. It can be said that the potential gradient that exists across the barrier is equivalent to a battery and the ions on each side of the depletion region create a space charge. Coupling the two terms we come up with *space charge equivalent battery* (see Fig. 21b).

We now have four names for the pair of boundaries which separate the emitter, base, and collector of the junction transistor. These names are no mere similes for essentially the same actions, although a tendency does exist to merge all the phenomena occurring at that point under the portmanteau term of P-N junction. This custom may be considered a form of engineering shorthand, which unfortunately contributes to its complexity. Each one of the four names describes a distinct facet of the junction as if one actor assumed different roles in the same play.

The *dramatis personae* of this play might be listed as follows:

Junction: A term which indicates the transistor type, and conveys information of geographical significance.

Depletion Region: An area in a semiconductor which has been depleted of active charge carriers by re-combination.

Energy Barrier: A demarcation line possessed of a restraining force capable of preventing unlimited re-combination of charge carriers.

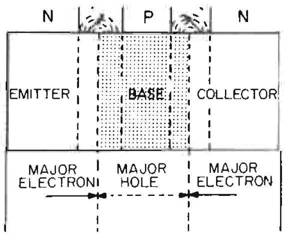
Space Charge Equivalent Battery: A phrase which describes the nature of the restraining force noted above.

The members of the *dramatis personae* hardly exhaust the cast of characters that play roles in and around the junction. There are others, perhaps not as distinguished in the annals of semiconductor literature, but waiting in the wings when needed. To play a key function in our next act, we will require one of these lesser used expressions. Its name is *resistance element,* and is defined as an element which offers resistance to a steady flow of current. Since current flow is dependent on the existence of charge carriers, the depletion region by virtue of almost a complete lack of charge carriers fits this definition like a glove. Since *resistance* can be represented by a schematic symbol in the form of a jagged line as in Fig. 22, we will employ this symbol in our drawing to represent the resistance elements at the two junctions.

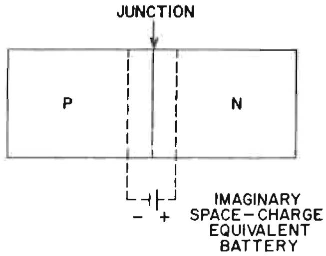
At this point, it might be in order to ask ourselves what we expect of the transistor. The usual answer in a word is: amplify. Such an answer could not be briefer — or vaguer. Amplification alone would hardly qualify as exceptional. Long before the vacuum tube, weak currents were sent over long distances by wire to energize a sensitive relay that would turn on a local power source strong enough to operate a telegraph key. We do the same thing when we manipulate a wall-switch, and with the flip of a finger turn on hundreds of watts of power. One of our oldest electrical devices is the relay which is simply a sophisticated switch employing an energized coil instead of a finger to open or close a set of contacts, and in fact the first modern computer consisted entirely of electromagnetic relays.[1]

The vacuum tube and transistor can simulate a switch or relay, but this on/off type of application is a routine accomplishment. Because it is a counting process, from zero to one and back to zero, such circuitry is referred to as digital. The full potential of the transistor is realized only when it is used as a linear amplifier where a small input is exactly reproduced at the output in amplified form. This is what the transistor can do, and the switch can't, and this is the most important function of the transistor.

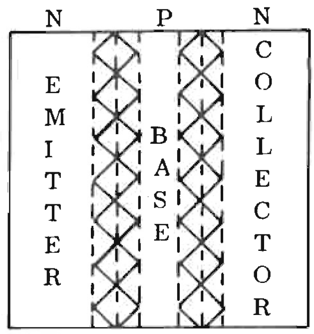
[1] *The Harvard Mark I Calculator, built in 1944.*



(Fig. 21a) Barriers balanced when hole-electron recombinations come to a near-complete halt.



(Fig. 21b) Potential gradient across space-charging region.



(Fig. 22) Balanced resistance elements at the junctions of a transistor.

The Agency by which Amplification is Produced

A single agency? Yes, just one. The reader may wonder at the singular form of the word rather than the plural. Hasn't the four-dimensional nature of the transistor with its holes, electrons, and oppositely charged ions been held up to him as a vision of complexity? So why suddenly a single agency, a single mechanism, for the production of gain? Well, because transistors achieve results the same way vacuum tubes do, by Lee deForest's method of 'relay action.' A tiny signal at the input triggers a much larger signal at the output.

Another way of looking at this agency of gain is to compare it to a lever and fulcrum where a great force is obtained when operating over a short distance of the lever at the expense of a small force operating over a long distance of the lever. In the case of the transistor, we have selected the base as the fulcrum point, the emitter resistance element as the small force (operating over the long arm of the lever), and the collector resistance element as the great force (operating over the short arm of the lever.)

In Fig. 22, both resistance elements are shown to be equal and balanced. No amplification is achieved because an equal current flowing thru equal resistances would produce the same power across each resistance load. The formula for power reads: $P = I^2R$. Therefore, the higher the value of R, the higher the value of P will be, even if I (current) remains the same.

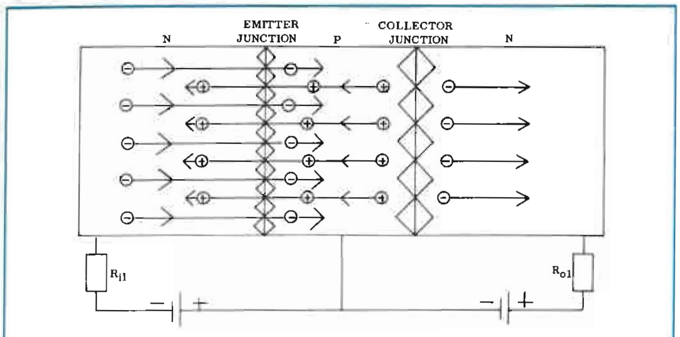
Since the value of the resistance element depends on the width of the depletion region, we must find a way of controlling the amount of real estate in the transistor that becomes subject to depletion. One way of accomplishing this aim is to utilize once again the law of charge attraction and repulsion which, you will remember, was the way the depletion regions were created in the first place. Only this time we shall have to employ external power sources as illustrated in Fig. 23.

Let us try our hand at compressing the emitter depletion region. The polarities of the external power source are applied between emitter and base in such a way that they assist the majority carriers in both emitter and base to drift toward the junction. The negative side of the battery repels the free electrons in the N-type emitter and pushes them junction-ward. The positive side of the battery repels the holes in the P-type base and pushes them junction-ward from the other side. The small space charge equivalent battery is overcome by the much larger external potential and the emitter depletion region narrows to a fine line.

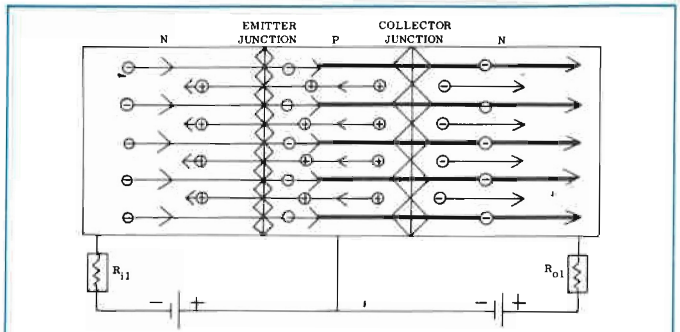
To widen the collector depletion region, we apply the polarities of still another external power source between collector and base in such a way that the majority charge carriers on both sides of the junction will be restrained from drifting toward the junction. The positive side of the battery will attract the free electrons of the N-type collector and pull them away from the junction. The negative side of the battery will attract the holes in the P-type base and pull them away from the junction on the other side. This will widen the collector depletion region and increase its resistance. Fig. 23 is a stop-motion drawing intended to depict this first act of transistor amplification.

The second and final act follows and is shown in Fig. 24 in bold-face type. Forward biasing of the emitter junction has caused a heavy flow of free electrons into the base area. Having entered the base and changed their status from majority to minority carriers, they will diffuse toward the collector junction. As they get closer, this diffusion will be speeded up because the collector junction, although reverse biased for holes, is forward biased for electrons. This will pull the electrons across the high resistance element of the collector junction.

Has amplification been affected, and if so, what is its magnitude? We can answer this question by assigning numerical values to the common base circuit of Fig. 24, and making them simple for easy calculation. Since we know the emitter resistance element to be quite low, we will also make the input resistance load (R_{iL}) low in the interest of impedance matching — 300 ohms. The emitter current, in view of the usual 3 volt bias, will be set at 10 ma. (.010 amp.). We know the collector resistance is quite high, so



(Fig. 23) Stop motion drawing of initial actions of majority charge carriers when transistor is wired into a common base circuit.



(Fig. 24) The route taken by free electrons starting at the emitter site.

to match it we will select an output load resistance (R_{oL}) of 50,000 ohms.

We can now apply the simple formulas for measuring power gain. Substituting for $P = I^2R$, the power across the emitter load works out to $.01 \times .01 \times 300$ which equals .03 watts. The power across the collector load becomes $.01 \times .01 \times 50{,}000$ which equals 5 watts. To find the power gain a ratio is set up: 5/.03 which equals to a power gain of 167.

There is one assumption in the above example that the reader might take exception to, and would be right in doing so, although the difference in the final result would be small. The 'internal' evidence would seem to indicate that the number of electrons that cross the collector junction are less in number than originally started across the emitter junction. It can be shown that the presence of a large number of minority carriers in the base region makes recombination a distinct hazard to their life span. But much fewer are lost than one might expect. Some new minority carriers are generated by collision and the base area is kept as narrow as practical so that diffusion thru the base may be accomplished faster. However, it may be more accurate to assume the current thru the collector to be 9 ma. as compared to the emitter's 10 ma. This would only lower the gain from 167 to 135.

One conclusion that may be drawn from this example is that the current gain in a junction transistor never exceeds unity and never quite makes it to unity. A second conclusion that may be stated just as clearly is that the most important element in transistor action, and the agency by which gain is produced, is the transfer of minority charge carriers across the collector resistance element. Hence the origin of the name:

TRANS (fer) + (res) ISTOR = TRANSISTOR

PART VI — THE SECOND WAVE — INTEGRATED CIRCUITS

The Short Happy Life of the Discrete Transistor

The Semiconductor Age is said to have begun in 1948 with the invention of the point-contact transistor. In our personal view, the Semiconductor Age began before the invention of the vacuum tube with the crystal detector, was sidetracked by deForest's device, and resumed its ascendancy again in 1948.

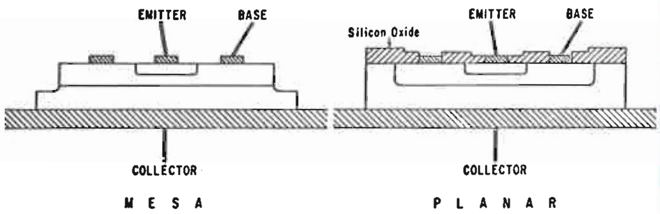
Though its potential was great, the initial point-contact transistor was too unstable to replace the vacuum tube in most applications. The junction transistor came along a year later and began making some inroads into areas formerly dominated by vacuum tubes. But in the computer field, the early junction transistor was much too slow to make a significant impact.

Yet the computer industry eyed this under-sized active element which required only trifling power with a great deal of interest. It was well known that vacuum tubes, which served as the active elements in the first large scale computers, had very grave faults. Their failure rate was so high that the equipments were more frequently off than on; their power consumption was fantastic; and their speed was slow by today's standards.

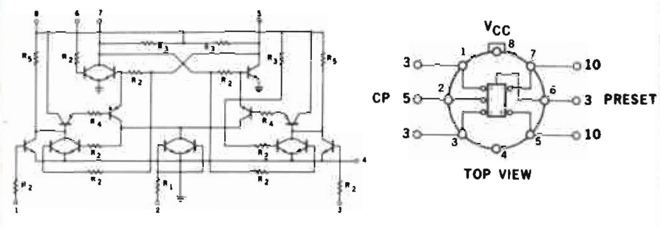
It was not until 1958, when the silicon planar[1] transistor was invented, that both high speed and reliability became available to computer designers. The result was the birth of a second generation of large scale computers, and with it the semiconductor explosion; both phenomena directly attributable to the planar process. A closer look at planar technology is a necessary prelude to understanding the next leap in our story.

The most critical areas in the transistor are the junctions. The incidence of contamination at these critical points correlates directly with the production yield. Device

[1] *Planar is a patented Fairchild process.*



(Fig. 25) Transistor structures. *(Courtesy of Fairchild Semiconductor)*



(Fig. 26) Internal circuit of Fairchild uL 923 JK Flip Flop and its logic symbol.

characteristics which are sensitive to surface exposure are reverse leakage, breakdown voltage, noise immunity, and current gain. These parameters are obviously of crucial importance, and substantial improvement in these areas would go far toward encouraging the widespread use of transistors. The planar process accomplished this feat by performing all diffusions under a passivating layer of pure silicon dioxide (glass), except for a small cut-out through which the doping impurity is introduced. It is also possible, with this process, to fabricate as many as 1,000 transistors at a time.

Fig. 25 illustrates the older mesa structure on the left with its exposed junctions, and the planar structure on the right with its junctions protected by a silicon dioxide formation.

The immediate advantages of the planar process number among them the passivation of the surface by an oxide layer at an early stage of manufacture before the junctions are formed, resulting in higher yields than was possible before, and the simultaneous fabrication of as many units as considered practical. But, even more important, it also made possible the integration of multiple elements on a single silicon crystal and pointed the way to integrated circuits.

The New Wave in Semiconductor Technology: Integrated Circuits

The first rumblings of still another sensational development in semiconductor technology appeared early in 1959. In April of that year, Texas Instruments disclosed the achievement of "integrated molecular devices." The short announcement closed with this warning: "The circuits described are strictly developmental and are not available at this time in either sample or production quantities." Before the end of the same month, the USAF awarded a $2-million contract to Westinghouse "for the investigation of microminiaturization by molecular engineering."

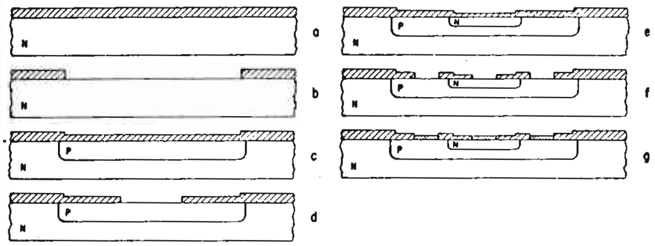
In the nine years since these announcements were made, integrated circuits (I/Cs) have grown into a $228-million market. Some experts forecast a 40% increase this year (1968), and a steady 15 to 25% upward trend for each of the next five years after this one. On the other hand, discrete transistor sales dropped 13% in 1966 from an all-time peak of $450-million, and another 2% in 1967.

No indication of headlong growth was evident when I/C technology broke out of the laboratory and into production. Some early prototypes sold for as much as $720, and it wasn't until 1961 that prices dropped below $100. This Tiffany-style selling didn't hold out much hope for a mass market. Something was needed to break the market place wide open and the necessary catalyst was provided by Fairchild. In May 1964 the industry was stunned when Fairchild cut prices drastically on a series of nine digital devices. Two of these were priced at an unheard of $2.55 each. The result was a 156% surge in I/C dollar sales in 1964 to a total for the year of $30-million. The next three years saw sales go up 92%, 87%, and 56% respectively.

Fairchild's daring maneuver succeeded for the same reasons that the 30-minute long-playing record overwhelmed the 5-minute 78 RPM record. Given a reasonably equivalent price, who would want a 5-minute record even if it was the same size? Why buy a dozen discrete components, mount and interconnect them when you can buy the same circuit completely packaged in the same space as one of the discrete components? The change in technology brought with it certain inherent advantages in both cases. The higher fidelity of the LP record is matched by the higher reliability of the I/C.

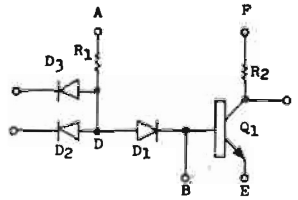
Understanding Integrated Circuits: The Building Blocks of Systems

Prof. McLuhan wrote a book called "Understanding Media: the Extensions of Man" in which he said: "The medium is the message – the content doesn't really matter." Much the same can be said for I/Cs. The content of the I/C matters very little, indeed no attempt is made to specify the individual characteristics of the active elements inside the structure of the I/C. It becomes a single unit, a 'black box,' with an input and output and the data sheet will

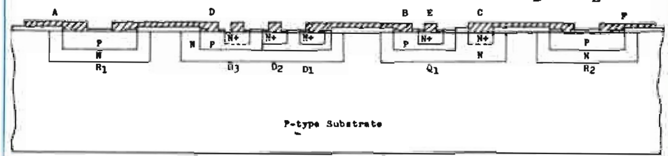


(Fig. 27) Steps in Planar transistor fabrication.

(Fig. 28a) Schematic of IC.



(Fig. 28b) Cross sectional view of IC wafer.
(Courtesy of Motorola Semiconductor)



specify only the ratings that apply to that input and output. This state of affairs is best illustrated by comparing the internal circuit of an I/C with its overall schematic symbol (see Fig. 26). In the symbol for the I/C there are no components.

If we go back to 1959, we find that most of the information presented to the general reader is slanted in such a way as to encourage the belief that the structure of the I/C departs radically from any previous technology. Here is one such description: "The solid-state circuits are made by actually rearranging the molecules in a slab of metal by electromechanical process. There are no actual "components" but only their functional equivalent." And here is another: "The specific technique involves controlled dentritic growth of crystals resulting in specific properties." Finally, here is one we can't quarrel with: "Integrated molecular electronics is the combining of the passive functions of resistance, capacitance, and inductance with the active function of amplification in a single semiconductor solid circuit." This is a good definition except for the word "molecular" which has been nonchalantly slipped in as if that explained all.

While it is true that an aura of extremely restricted dimensions is associated with anything molecular, the molecular "rearrangements" that occur in I/Cs also occur in discrete transistor manufacture. Therefore, one cannot speak of molecular rearrangement as a unique property of I/Cs. The word 'molecular' is no longer applied as an exclusive property of I/Cs, except where the word has been frozen into the name of a corporate structure, as in the case of the Molecular Electronics Division of Westinghouse like a fly in amber. Dentritic is another word with some vague connotation of crystal structures, apparently borrowed from geology where it refers to earth erosion by streams. Dentritic erosion is a far-fetched metaphor for the precision etching of I/C designs.

Since the making of I/Cs is an extension of planar technology, the simplest approach is to follow the steps in the manufacture of a planar transistor, and go on from there to the closely related structure of the I/C. The transition from a single transistor to an I/C is purely quantitative, but — as is well known — quantitative changes have a way of becoming qualitative by sheer weight of numbers.

The steps in the fabrication of a planar transistor are shown in Fig. 27.[2]

a.) Start with a suitably polished N-type wafer and cover it with an oxide layer.

b.) Mask the layer and etch the oxide off a series of circular cut-outs.

c.) Expose the wafer to the base diffusant. (Boron for NPN; phosphorous for PNP). During this operation oxygen is used so that the wafer is re-oxidized.

d.) Mask and etch again, this time over the base area.

e.) Expose to the emitter diffusant. Except for the doping level this step is similar to the base diffusion step (c).

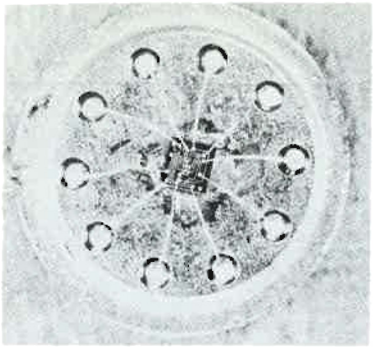
f.) Mask and etch both base and emitter for contacts.

g.) Alloy metal contacts on the etched areas.

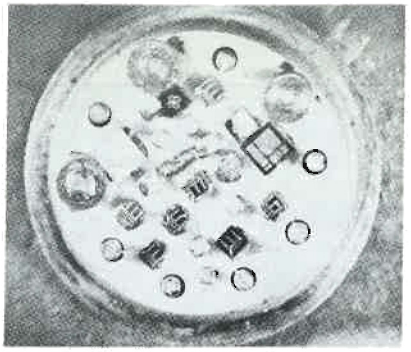
This sequence of oxidation, masking, etching, diffusion, reoxidation, etc., is also characteristic of I/C construction. Fig. 28a is a schematic of a logic gate requiring six discrete components. Fig. 28b shows a cross-sectional view of a wafer into which all six components have been fabricated and intraconnected to form a single 'discrete' I/C. The major steps in this process may mount into the dozens even for this limited circuit. Resistor R2 is made by diffusing a P-type impurity into an N-type substrate. The value of resistance is determined by the area occupied and must be defined in advance. A capacitor can be made by using the depletion region as the dielectric in a reverse biased P/N junction.

The I/C described in Fig. 29 is called monolithic because every portion of it is incorporated into a single chip of silicon. However, the problem of bringing ten leads from a tiny chip measuring 50 × 50 mils to the outside world where it will be subjected to rough usage increases the size of the final package to about a third of an inch in diameter

[2] *Condensed form Fairchild technical paper. See bibliography.*



(Fig. 29) Single block MECL gate.



(Fig. 30) Hybrid integrated circuit amplifier utilizing bipolar and unipolar transistors.

(see Fig. 29). The same size can is able to accommodate a circuit of miniature discrete components (See Fig. 30). The active elements are transistor chips, the blocks of resistors are thin film, and the capacitors are unencapsulated. This kind of unit is called a hybrid I/C.

The hybrid I/C offers some advantages when small quantities are required since the numerous masks required for monolithic construction are costly. But where large quantities are involved the repetitive nature of the planar process cannot be matched by the manual assembly and wired interconnections of the hybrid structure. Of equal importance is the high reliability of the monolithic type. Since all the elements are fashioned at the same time, their related parameters will exhibit closely matched performance statistics.

Although reliability and low cost may be important contributory reasons for the rapid rise of I/Cs, the unique feature which stamps the I/C as an archetype rather than a derivation is the promise which excited the imaginations of engineers and is now a reality, the promise that one could fabricate a complete circuit of many components in the same space formerly occupied by just one of the active elements. Imagine what this promise held out to all designers, and especially those whose efforts had produced the room-sized computers of that period and were now about to go on to something bigger, like a computer the size of a building. The size transformation promised by I/Cs induced tremor-like reactions in organizations like IBM. In no time at all, they were making I/Cs, and making them for no one else but themselves.

PART VII — LARGE SCALE WAVES

The Vacuum Tube Departs — But Its Apparition Returns

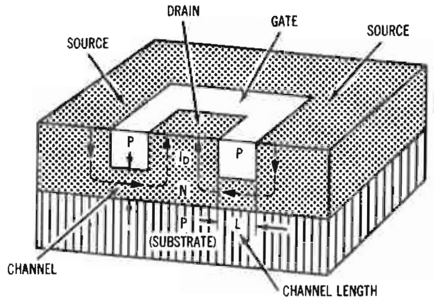
When the transistor made its bashful entrance on the electronics stage twenty years ago, circuit designers were confronted by a device capable of replacing the vacuum tube (VT) in many applications. The reader should be warned that the word "replace," as used here, is not intended to convey the idea of exact equivalency. In other words, one cannot merely insert a transistor where a VT used to operate and expect more or less the same results. If the circuit designer elects to use the transistor in place of the VT, then the entire design from A to Z must be altered to accommodate the change. No component, whether it be inductive, resistive, or capacitive, remains unaffected. Naturally, the expenditure of so much effort twenty years ago in a practically virgin territory would not have been forthcoming without compelling reasons.

And compelling reasons did exist as anyone associated with semiconductors even then was quick to realize. The reduction of power supply voltages, the elimination of excessive heating, and the fantastic shrinking in size of not only the active element, but all other components, were the compelling reasons, not to speak of such intangibles as the desire to be in the vanguard of a new technology. The infant device forged ahead on all fronts, overcoming the lethargy of the establishment mind.

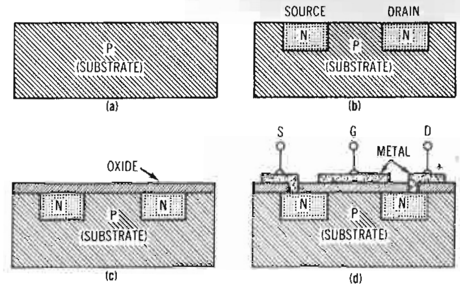
Why does this new technological animal require a whole new circuitry although it amplifies like a VT? This question can best be answered by contrasting the VT and the transistor in three important operational areas.

1.) Current flow in a VT is always "on" (assuming the filament is heated). A sufficiently high negative bias must be applied to the grid to turn it "off." The opposite is true of the transistor. Normally, current flow is "off," and must be turned "on" by applying forward bias at the input.

2.) The input and output impedances of the VT are very high; the elements of the VT are isolated from each other by a vacuum gap. In the transistor, where the elements are physically joined together as part of the same



(Fig. 31) Single-ended geometry junction FET



(Fig. 32) Development of n-channel insulated-gate field-effect transistor, enhancement mode type.

crystal, the impedances are low, or at best, moderate.

3.) A couple of volts on the VT grid can control hundreds of volts on the plate, hence the VT is referred to as a voltage amplifier. In the transistor, minority current flow across a reverse-biased junction is responsible for gain, thus qualifying it as a current amplifier.

Thus, the VT and transistor are seen to be worlds apart in operation, although both are classified as active elements. Practitioners in one field feel uneasy in the other. If one does switch over, he jettisons his former field entirely, and the switch has always been in one direction: toward transistors. Yet many of the VT parameters are held in high esteem and were given up with regret. The development of an exact solid state equivalent of the VT was certain to be welcomed.

Curiously, most of the early attempts at developing a semiconductor amplifier tried to mirror the parameters of the VT. Solid diodes had existed even before 1900; when the VT triode came along, simple analogy dictated a try at inserting a third element between the anode and cathode of the solid diode in the hope it would work like a VT. But the analogy didn't work. Improvements in theory and materials were needed before much progress in this direction could take place. Shockley again led the way in theory both before and after he was awarded the patent for the (conventional) junction transistor which acquired the name of bipolar because of simultaneous majority and minority current flow. The device that most closely parallels the operation of the VT is known as the Field Effect Transistor (FET) which, like the VT, is unipolar. It is only in the last five years that the FET has blossomed into a practical device. Increasing sales have been a major factor in sharply reducing the initially high prices. Today the FET stands out as a worthy partner of the bipolar device.

There are two basic types of FETs: the Junction FET (JFET) and the Insulated Gate FET (IGFET). This latter type is sometimes called the Metal-Oxide-Silicon FET (MOSFET). The JFET, like the VT, is "on" with zero gate voltage and can be turned "off" by supplying a sufficiently high reverse bias to the gate. The IGFET is made and operated in two different modes: Depletion or Enhancement. In the Depletion mode the IGFET is "on" with zero gate voltage and "off" when the gate is reverse biased. In the Enhancement mode, the opposite is true. Of the two, the IGFET is more popular than the JFET. The gate of the IGFET is insulated from the conducting channel by a dielectric. See Figs. 31, 32, and 33 for construction details.

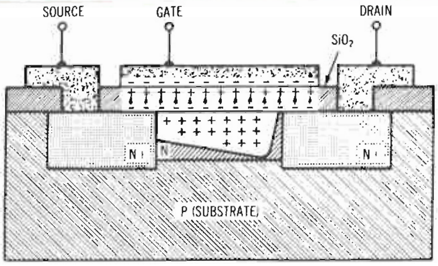
The input impedance of FETs, regardless of the type, is quite high. The gain characteristics are similar to those of pentode VTs. Yet operating voltages and power drain match those of the bipolar transistor. The size of the FET is also the same, and the chip itself is even smaller than the bipolar chip. What it adds up to for the designer is that he has available for his use, two contrasting active elements that can complement each other in several important dynamic characteristics, yet are compatible in size and power requirements.

As discrete components, the FET and bipolar devices can both be incorporated into the same circuit system without difficulty. In the manufacture of integrated circuits, the extra diffusion steps required for bipolars does create problems, but they are not insurmountable. However, bipolar ICs cannot be matched directly to MOS ICs. A matching circuit is required to convert the negative voltage levels of the FETs to the positive voltage levels of the bipolars and vice versa. Such interface circuits have been developed by Fairchild.

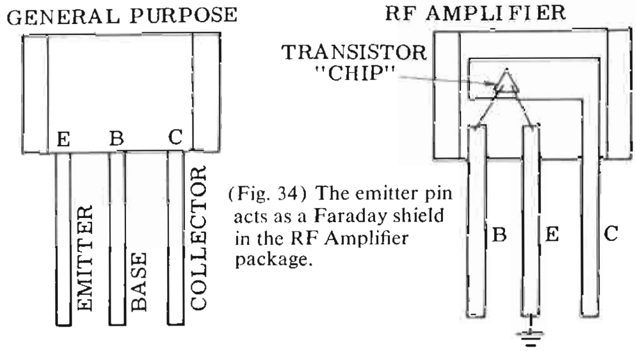
Invention by Analogy

We have seen that analogy from VT structure was of little use in the development and invention of the transistor, except in the area of inspiration. However, once the IGFET gained a foothold in the market place, invention by analogy with the VT did improve its performance.

When the early triode VT was used as an amplifier in the higher (radio) frequency bands, operation was found to be very unstable, sometimes deteriorating into outright oscillation. This difficulty was known to be caused by grid



(Fig. 33) Channel depletion phenomenon. Application of negative gate voltage causes redistribution of minority carriers in diffused channel and reduces effective channel thickness. This results in increased channel resistance.



(Fig. 34) The emitter pin acts as a Faraday shield in the RF Amplifier package.

to plate capacitance which, though small, feeds back the output signal to the input. Tricky neutralizing circuits were required to cancel this feedback, adding to the complexity of the RF stages. The problem was eventually solved by inserting a screen grid between control grid and plate. When the screen is grounded for RF, an effective (Faraday) shield is created between input and output which practically eliminates feedback capacitance and the need for neutralization.

Transistors also suffer from this problem. It is true that the transistor was born with microscopically small elements resulting in lower feedback capacitance, but this advantage has been nullified by the rise in frequency of new channels of radio communication. Neutralization, of course, has been one way of coping with this difficulty. Another way has been to change the usual pin configuration of Emitter-Base-Collector to B-E-C. With E grounded as in Fig. 34, the emitter pin acts as a Faraday shield to some extent, reducing feedback capacitance by 23%.

RCA tackled this problem by going to the solid state equivalent of the VT – the MOSFET, and by analogy with multigrid VTs inserted a second gate into the structure of the MOSFET (see Fig. 35). When grounded for RF, one gate forms a shield between the drain (output) and the other gate (input). The feedback capacitance is reduced to a point where neutralization isn't needed up to 800 MHz.

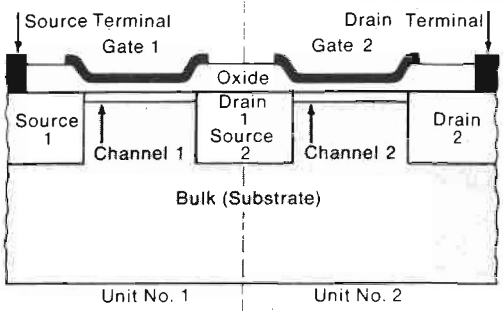
Another curious similarity that the transistor shares with the early history of the VT involves maximum voltage ratings. When first introduced, the plate was restricted to a maximum of 45 volts because vacuums were none too good in those days. Stray gases trapped inside the tube would ionize when higher voltages were applied and the resulting heavy currents would damage the tube. As vacuum technology improved, maximum voltage ratings increased.

The early planar transistor suffered a similar fate. Its maximum breakdown rating never exceeded 30 volts because of instability created by ions trapped on the surface of the transistor under the passivating oxide layer. This instability is voltage dependent and is not noticed when operating voltages are low as in ICs. At higher voltages an inversion layer (polarity reversal) occurs causing destructive "channeling" thru the semiconductor structure. This problem was licked without degrading other device characteristics by encircling the base region with a band of heavily doped material to break up the ion channels. Called the Annular process, Motorola introduced it commercially several years ago and was awarded patent rights in 1967 (see Fig. 36).

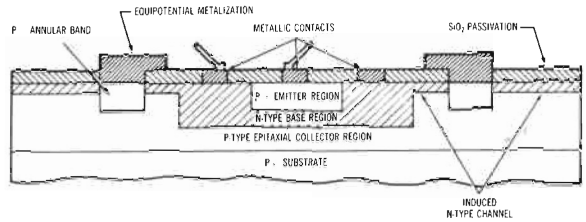
The Evolution of the Semiconductor from Component to Circuit, and Now, to System

Integrated circuits have not displaced discrete transistors so much as swallowed them. They are still being made, pretty much the same way, on the crowded IC chip as well as on the solitary transistor chip, – but with this difference: the can which used to house a single transistor now houses an integrated circuit containing a dozen or more transistors, not to mention diodes and resistors as well. In fact, it is possible to contain a thousand transistor chips inside a single TO-5 can; the trick is to interconnect them. Does it make sense to put twelve transistors on a chip and bring out thirty-six leads? Simple logic dictates that interconnections be made on the chip, and the only meaningful way of doing this is to completely assemble on one chip a circuit or sub-circuit such as a gate or flip flop. An AC-coupled J-K flip-flop with dc Set & Reset inputs and buffered outputs for counter and shift register applications up to 15 MHz, uses 16 transistors, 14 resistors, and 2 capacitors. This adds up to a total of 80 interconnections. The entire circuit when fabricated on a single chip has ten leads. This is the rationale behind integrated circuits.

But why stop at one circuit? Why not 10, 100, 1000 circuits on a chip? Well, how many circuits can you put on a chip? That would depend of course on the size of the chip, – but the larger the chip, the smaller the yield per wafer. So cost dictates as small a chip as possible. Right here, process technology is helping by introducing a stream of improvements that have been steadily reducing chip size and increasing component density. By this time it must be



(Fig. 35) Schematic representation of a dual-gate MOS field-effect transistor.



(Fig. 36) Typical PNP transistor cross-section demonstrates the surface stability afforded by Motorola's patented Annular and Field Relief Electrode semiconductor structure technology.

apparent that economics and not some dramatic new advance in technology is responsible for the trend toward Large Scale Integration (LSI). There are no insoluble problems that would prevent the building of a completely interconnected system on a single chip. In the meanwhile, the rapid rise of LSI awaits the realization that simple economics demands of the potential user that he take a good long look at it.

To take a good long look at the future state of the semiconductor species requires much more than 20-20 vision. Like any other species, it is continually evolving, but at such a rapid pace that a mere 20 years later, the point contact transistor is as much of a fossil as the Branley Coherer. Three generations of Motorola's MTTL ICs have appeared in three years, each generation an improvement on its predecessor, Since Sylvania came out with its first TTL in 1963, it has been improved 361 times. It is obvious that the superiority of one process over another can only be determined by collecting concrete test data. The accumulation of such data demands a time span of at least two years. Considering the speed of developments in the electronics field these days, one can understand the truth of the assertion: "If it works, it's obsolete."

The sales of discrete transistors have been receding as ICs take over, and LSI will no doubt affect the continued rise in sales of ICs. The semiconductor edifice with its superstructure of integrated circuits and systems gives signs of settling firmly without sustaining any serious cracks in the foundation. The actual dimensions of the "pie" that will be shared by the discrete transistor, the "discrete" IC, and the various forms of LSI is anybody's guess.

The computer industry has danced to the tune of electronoics technology for three generations. There is no reason why LSI shouldn't spawn a fourth generation, but don't look for it next month or next year. Like all generation gaps, one generation of computers is rarely compatible with those of another generation. Therefore, a time lag is inevitable while the earlier computers pay for themselves and designers get a "fix" on where they're going from here.

Semiconductors have contributed generously toward shrinking the size of 2nd and 3rd generation computers. Will this trend continue with LSI? It seems that what 4th generation designers are looking for is not a reduction in computer size, though that may be achieved; or even an improvement in performance, though that too may result; rather, they are looking for a complete change in the organization of computers – a new look.

The programming area – the heart of the software problem – is most susceptible to improvement. Since competent programmers are rare and each new generation of computers deepens the scarcity and increases the need for competence, it would seem to be a logical area to exploit. This interface between the human and the machine has been undergoing a curious race. As the machine becomes more complex, the demands on the human programmer also increases in complexity. Unfortunately, the human brain is not adaptable to the installation of ICs and LSI. However, the machine is, and the new interface must reflect back to the human not its complexity, but its simplicity of handling. It has been suggested that a computer should have a stack of "program generators" that are capable of being easily adjusted by merely following a short list of A-B-C-type instructions in simple English.

The TV-Consumer market is the only one that can be mentioned in the same breath with the computer market. But unlike the latter, it has responded only feebly to the integrated circuit. The degree of shrinking possible with ICs has less impact on the TV designer because the CRT at present dictates the length of the overall size. The only other motivation that might interest the consumer market is cost. But here the ball is entirely in the hands of the semiconductor industry. It seems likely, extrapolating past performance, that the ball will be back in the hands of the consumer product manufacturers before long. When this happens, some glimmering of the total effect of semiconductor technology on society will begin to be felt directly by the pubic at large.

BIBLIOGRAPHY:

Part I

Marconi, Pioneer of Radio by Douglas Co.
Julian Messner, Inc., 1943.

Wireless by P. J. Risdon.
Ward, Lock & Co. Ltd., 1923.

Radio & Television Almanac edited by Orrin E. Dunlap, Jr.
Harper & Bros., 1951.

Part II

History of Radio Telegraphy and Telephony by G. G. Blake.
Chapman & Hall Ltd., 1928.

Invention and Innovation in the Radio Industry by Rupert Maclaurin.
Macmillan Co., 1949.

Radio's Conquest of Space by Donald McNicol.
Murray Hill Books, Inc., 1946.

Father of Radio by Lee deForest.
Wilcox & Follett Co., 1950.

Part III

History of Radio Telegraphy and Telephony by G. G. Blake.
Chapman & Hall Ltd., 1928.

The D.A.T.A. Transistor Tabulation.
D.A.T.A., Inc.

Design Considerations for the OEM Line by Keith Suker.
Semiconductor Division, Westinghouse.

The Theory of Electronic Semiconductors by Alan H. Wilson.
Proceedings of the Royal Society of London, Vol. A133, Oct. 1931
and Vol. A134, Nov., 1931.

Thirty Years That Shook Physics by George Gamow.
Doubleday & Co., 1966.

Electrons and Holes in Semiconductors by William Shockley.
Van Nostrand Co., 1950.

History of Semiconductor Research by G. L. Pearson and W. H. Brattain.
IRE Proceedings, Vol. 43, Dec., 1955.

Semiconductors and the Transistor by E. W. Herold.
Journal of the Franklin Institute, Vol. 259, Feb., 1955.

Transistor Engineering by Alvin B. Phillips.
McGraw Hill Book Co., 1962.

Part IV

GE Transistor Manual, 7th Edition.
General Electric Company, 1964.

(Continued)

Part VI

Planar Silicon Transistors and Diodes by Jean A. Hoerni.
Copyright 1961 by Fairchild Semiconductor.

ICs — Present Status & Future Trends by I. A. Lesk.
Motorola Semiconductor, 1963.

A Progress Report on ICs.
Motorola Semiconductor, Dec., 1963.

In Electronics the Big Stakes Ride on Tiny Chips by Philip Siekman.
Fortune, June, 1966.

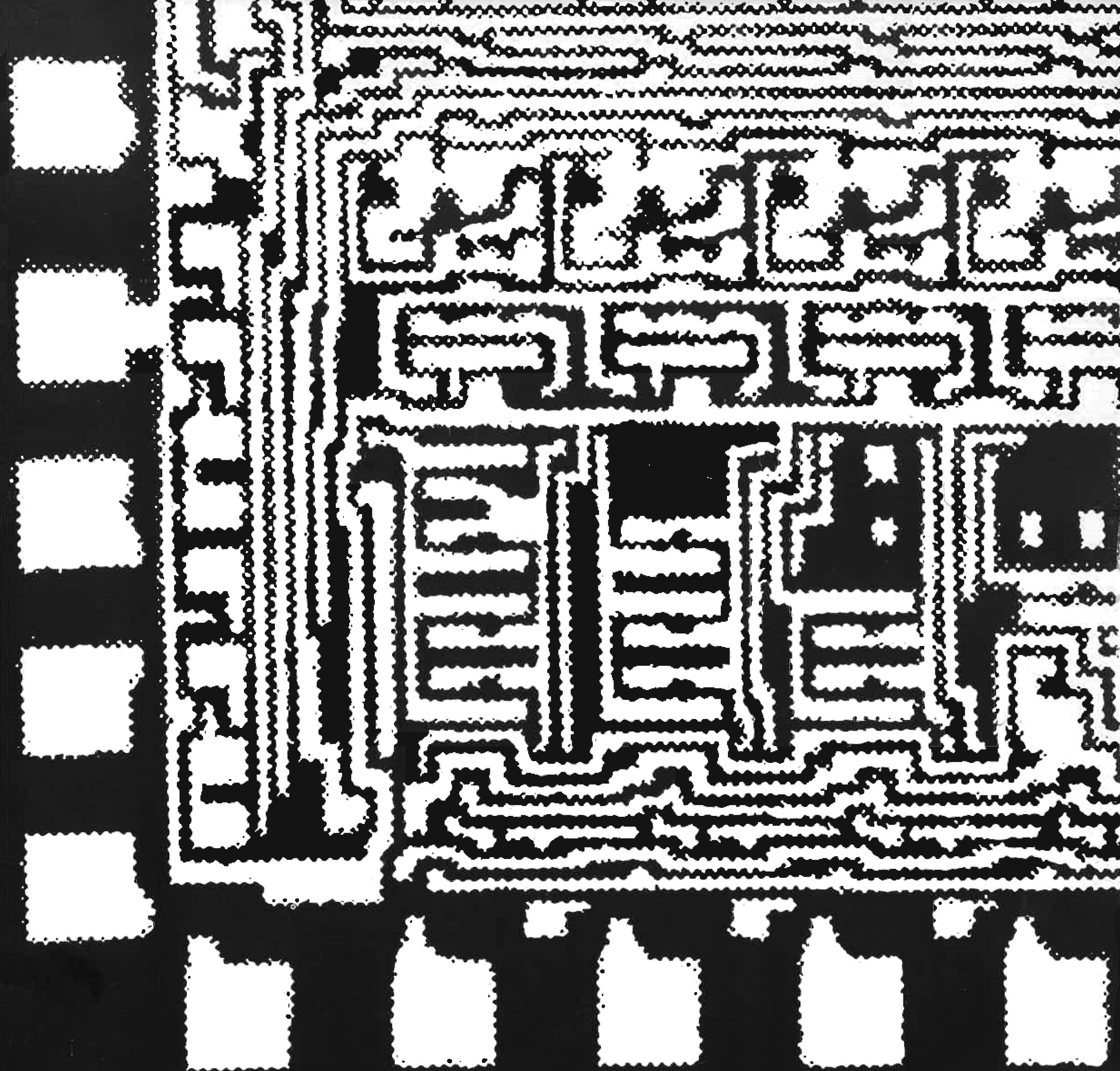
Part VII

Field Effect Transistors in Theory & Practice by James F. Kane, Donald L. Wollesen.
Published by Motorola Semiconductor Products as Application Note AN211.

Feedback Capacitance of Transistors by Donald L. Wollesen.
Published by Motorola Semiconductor Products as Application Note AN237.

Understanding & Using the Dual-Gate MOSFET by R. Dawson, R. Ahrons, N. Ditrick.
The Electronic Engineer, Sept. 1967.

Making Integrated Electronics Technology Work by Robert N. Noyce.
The IEEE Spectrum, May, 1968.



The Semiconductor Story

SCHWEBER
ELECTRONICS