resistance and collector capacity are found to be necessary. The relative advantages of linear and circular structures are considered both for base resistance and for collector capacity. Parameters, which are expected to affect the frequency behavior, are considered, including emitter depletion layer capacity, collector depletion layer capacity and diffusion transit time. Finally the parameters which might be obtainable are compared with those needed for a few typical switching applications."

The Planar Process

The development of oxide masking by Frosch and Derick [9,10] on silicon deserves special attention inasmuch as they anticipated planar, oxide-protected device processing. Silicon is the key ingredient and its oxide paved the way for MOSFET integrated electronics [22]. An account of their revolutionary development and utilization of SiO_2 as the vital foundation of today's IC industry has been described by Holonyak [22]:

> "In building our various experimental devices, we were in contact with various groups and individuals, but above all with Carl Frosch. Frosch was a consummate process chemist who was familiar with many types of processing procedures and had been working, with his technician Derick, on impurity diffusion into silicon for several years. In spite of his considerable experience, Frosch, with dry gas diffusion procedures utilizing N_2 or H_2, regularly reduced many of our silicon wafers to "cinders, " particularly at higher temperatures (≥ 1100°C).
>
> Because we had mastered building a diffused-base alloyed-emitter silicon p-n-p transistor (in spite of our problems with diffusion), one of the p-n-p-n configurations that we could explore was simply a modification of the p-n-p transistor: We could fabricate the diffused-base alloyed-emitter p-n-p on one side of a p-type substrate wafer after it first was prepared with an n-type diffused region (symmetrical) on both sides of the wafer. Either side could be chosen to form the p-n-p. The result was a p-n-p-n switch, in fact, the p-n-p-n switch of example (b) as described in [19]. (The complementary version of this exact structure, an n-p-n-p with Ga diffused into both sides of an n-type silicon wafer and then a Au-Sb emitter alloyed on one side, was later introduced at General Electric as the first commercial silicon controlled rectifier, today's thyristor. This later work was also based on our 1956 research [19].
>
> In the process of diffusing the p-type substrate wafer into an n-p-n configuration for the first stage of p-n-p-n construction, particularly in the redistribution "drive-in" phase of the donor diffusion at higher temperature in a dry gas ambient (typically ≥ 1100°C in H_2), Frosch would seriously damage our wafers. The wafer surface would be eroded and pitted, or even totally destroyed. Every time this happened the loss was apparent by the expression on Frosch's face, not to mention, on ours (N.H.). We would make some adjustments, get more silicon wafers ready, and try again.
>
> In the early Spring of 1955, Frosch commented to Holonyak, "Well we did it again," meaning the wafers were again destroyed. But then he smiled and displayed the silicon wafers – nice and green in color (in further instances also pink). He and his technician Derick had switched from a dry-gas (typically N_2 or H_2) impurity diffusion to a wet-ambient (H_2O vapor + carrier gas) diffusion, a consequence of an accident of the exhaust H_2 igniting and flashing-back into the diffusion chamber (because of gas flow fluctuations) and causing H_2O to cover, react with, and protect the silicon samples with oxide. The "wet" ambient, which was then immediately evaluated and adopted, created a protective oxide on silicon. It could be selectively removed for gaseous diffusion into the bare regions, which could then be resealed with oxide for higher temperature anneals or further diffusion. Many processing sequences could be devised for use of the protective oxide, which, of course, prevented crystal pitting and erosion. Frosch and Derick quickly found out which impurities were blocked from diffusion into silicon by the natural protective oxide (SiO_2) created in an H_2O-vapor ambient and which impurities would permeate the oxide (e.g., Ga). It was easy, once the issue of the oxide was known, to devise various schemes to diffuse into or to block impurity diffusion into silicon. The process was so flexible that planar n-type regions of any desired pattern could be prepared on a p-type substrate silicon, or the opposite, p-on-n diffused regions could be prepared on n-type silicon. All other diffusion procedures were suddenly rendered obsolete. We readily converted the Frosch-diffused silicon n-p-n into a working p-n-p-n switch [19]."

Holonyak [22] noted "what Frosch and Derick had done was to set the basis for a revolution. In fact, it is the oxide on silicon that is the basis, the vital foundation of today's IC chip, and of all of the silicon devices so critical to the electronics industry. Because of our (Holonyak) exploratory silicon device work and our involvement with Frosch, we were close observers and witnesses of his work. For example, on p. 15 of an extensive BTL memorandum [177], Frosch wrote:"

> "Thin silicon slices also were diffused with Sb for N. Holonyak for preliminary device development investigations. These were diffused for 2, 5 and 16 hours respectively at 1300°C in N_2 saturated with water vapor at room temperature. After diffusion, these slices were green in color with an excellent surface appearance. These layers were reported to have resistivities of from 10 to 20 ohms per square. The diffusion layers were reported to be uniform in thickness being 0.26, 0.39 and 0.76 mils respectively for the 3 heating times. An additional run was made for Holonyak to produce layers of somewhat higher resistivity. In this run the thin silicon slices were heated for 1 hour at 1200°C followed by 16 hours at 1300°C in N_2 saturated with water vapor at room temperature. These samples again were green in color with excellent surface appearance. These were reported to have resistivities of 45 to 90 ohms per square with a diffusion depth of 0.66 mils. The higher resistivity values obtained indicate not only a lower solubility of the Sb compound in the quartz envelope at 1200°C than at 1300°C but also the essential absence of Sb compound vapor in the carrier gas when the temperature was raised to 1300°C. Holonyak was able to produce very promising cross-point switches from some of these Sb diffusions."

Holonyak [22] continues by saying "we knew Frosch's work at first hand, and realized immediately what he and Derick had done. All of us near this work, which was just a few at first, realized its importance, but, in truth, none of us, least of all Frosch, in his exceeding modesty, projected it to its true future scale. Frosch even wrote (p. 20) the following statement in his BTL memorandum [177]:"

> "In addition to the possibilities of process simplification, the protective quartz envelope added during the heating may be useful for protecting an electrical device from atmospheric conditions. For example, the device might prove more stable if left enclosed in such a quartz envelope. However, it may not be possible to make all of the necessary electrical contacts through the quartz. In these cases some protection may be retained by the removal of a small area of the envelope for the application of the contacts."

Holonyak has summarized Frosch's innovation by "Frosch had, indeed, anticipated planar, oxide-protected device processing. He appreciated immediately the importance of the oxide. It is questionable if anyone else's contribution had as much to do with the existence of the "chip" and today's electronics as Frosch's oxide. This is easily seen by simply raising the question: *Remove the oxide, say it doesn't exist, and then what would there be? Silicon itself is, of course, the critical ingredient followed by its unique natural oxide. In some sense it could be said that Si and its technology (its oxide) "invented" the IC*," (italics entered by the author).

The benefits of SiO_2 on the surface properties of silicon were concurrently, or shortly thereafter, assessed by Attala's group [11,178]. They believed that growing an oxide under clean and controlled conditions, on a properly cleaned silicon wafer, would lead to both a reduction of surface states and passivation of the silicon surface. The planar diffused transistor developed by Hoerni [12-15] of Fairchild Semiconductor Corporation in the late 1950's pulled together a number of these strands as regards the benefits of SiO_2 and was in production by 1959. These included the concept that the SiO_2 masking layer, utilized in the fabrication of diffused silicon transistors, be left in place for the passivation of p-n junctions intersecting the surface in the case of the grown junction, alloy and diffused mesa transistors, without the necessity of growing a passivating oxide under meticulously clean conditions [179], per the insight of Hoerni [12-15] as well as ensuring a dielectric layer for supporting metallic conductor overlayers in the IC era [16]. The Si-SiO_2 diffusion technology had, in point of fact, been transferred from BTL to Shockley Semiconductor, to Fairchild Semiconductor Corporation and, from there led to the creation of Silicon Valley [180]. Numerous testimonies as to the efficacy of the planar approach have been presented [122,158,181]. Sparkes noted [159]:

> "Victor Grinich of the Fairchild Corporation presented graphs of the change with time of current gain, base-emitter voltage and cut-off current of planar transistors which were so much better than anyone had seen before that it was quite obvious that if they were genuine a real breakthrough had been achieved. After several hours' discussion with Grinich it became clear to me that the planar process was the process of the future. It was an unpalatable conclusion, since, just at that time, many companies had recently invested large sums of money in the double-diffused, the alloy-diffused or micro-diffused process with the hope of achieving a clear production run of a few years."

To more fully appreciate the significance of passsivating the p-n junction intersecting the surface in the mesa transistor such as fabricated by AT&T, one may consider Moore's assessment [122]:

> "In mesa transistors, the emitter-base junction is exposed on the top surface between the metal contacts, while the base-collector junction intersects the sides of the mesa (see Figure 3). The regions of high electric fields where the junction comes to the surface are sensitive to contamination. Contamination of the emitter-base junction can decrease the gain of the transistor dramatically. In the case of the collector junction, the breakdown voltage and leakage characteristics can change. We noted a problem that some of the transistors packaged in hermetically sealed cans in dry nitrogen showed very unstable collector junction characteristics. Breakdown voltages sometimes decreased by several tens of volts and became unstable when observed on an oscilloscope, potentially a major reliability problem. We formed a task force to try to understand and correct the problem. One of our technicians, B. Robson, carefully cut the can off one of the bad devices and examined it under a microscope. He noticed a spot of light emitted from the side of the mesa when the transistor was biased into breakdown. He shut off the power and saw a tiny particle on the side to the mesa at the point of the light emission. Carefully removing the particle and reapplying power, he found that the original high breakdown voltage was restored. The particle, evidently attracted by the high electrical field where the junction came to the surface, was causing the premature breakdown of the junction. Now we knew the cause of the low breakdown. All we had to do was eliminate all the sources of particles."

The planar process introduced extreme flexibility in the fabrication of junction transistors, since the "tooling up" to fabricate different devices involved changing the mask set, diffusion profiles and doping levels and, as appropriate, the resistivity of the starting material. The planar process additionally facilitated the fabrication of a double-diffused transistor essentially planar with the original wafer surface, without the necessity of a mesa structure. A photomicrograph of the first planar transistor is shown in Figure 4 [122]. Chin-Tang (Tom) Sah, Harry Sello and Tremere published the first quantitative analysis of the oxide masking thickness and related time and temperature processing to ensure blockage of the diffusing impurity in the oxide masked region [182]. It was especially important that these planar transistors had the base contact completely surrounding the emitter so as to eliminate any chance of the base inverting, as was the case for the original bipolar junction transistors [183,184]. With the advent of the planar technique, cut off frequencies in the range of 10 GHz could now be achieved, approaching values achieved for vacuum tubes [185]. Indeed, the exodus of Bob Noyce, Gordon Moore, Jean Hoerni, Jay Last, Julius Blank, Victor Grinich, Eugene Kleiner and Sheldon Roberts 1957 from Shockley Semiconductor Laboratory, established in 1955 [69,122], and their subsequent formation of Fairchild Semiconductor Corporation with the stated goal of fabricating double-diffused transistors for commercial gain (the original goal of Shockley Semiconductor before Shockley redirected its effort to the p-n-p-n diode [174]) was the major stimulus to the invention of the planar transistor, Silicon Valley and the development of the fledgling electronics industry. On the other hand, it should be noted that Shockley, Noyce, Moore et al. were cognizant of developments at BTL as regards SiO_2 and were quite receptive to its advantages in the fabrication of silicon junction transistors. In fact, the entire diffusion technology at BTL was made available to Shockley to help facilitate his ambition to derive a measure of success in the business world based on the transistor technology he had helped to develop. Shockley was not, however, to achieve his business goals but contributed, inadvertently, to the exodus from Shockley Semiconductor which, in conjunction with Stanford University's graduates, led to the Silicon Valley phenomenon. Indeed, Shockley has been referred to as the "Moses of Silicon Valley" [186].

established the basis for the utilization of the aluminum metallization system [22]:

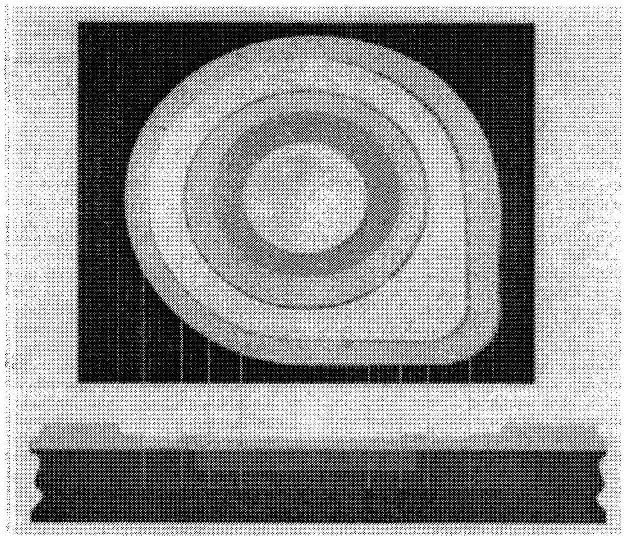

Figure 4. Photomicrograph of the first planar transistor. The diameter of the circle that forms most of the outside ring is 0.030 in. The light areas are aluminum emitter and base electrodes. (From "A Solid State of Progress," Fairchild Camera and Instrument Corporation, 1979) [122]. Reproduced by permission of the IEEE, Inc.

As noted earlier, the critical elements for the fabrication of the transistor and thryristor and, subsequently, the IC electronics era (oxidation, diffusion, photolithography, aluminum metallization and thermocompression bonding) were now all available. Jules Andrus and Bond showed that certain photoresists, when deposited on SiO_2, would protect the underlying SiO_2 during etching processes [180,187,188]. Optical exposure of the resist using contact masks in the late 1960's and early 1970's, projection masks in the middle 1970's and stepper mask methodologies beginning in the later 1980's was used to create precise window patterns (open regions) in the oxide and, therefore, precise control of diffusion areas. Aluminum metallization was utilized to form ohmic contacts to both p- and n-type material. While the former was expected due to Al being a group III dopant, the latter was achieved since the contact was subsequently identified to form a tunnel diode with the n-type silicon, the tunnel diode characteristic being linear (and, therefore, simulating an ohmic contact) for both small positive and negative voltages about the origin. Moore and Noyce received a patent for the aluminum metallization [189]. Holonyak has described the experiments conducted at BTL which

"Satisfactory Al evaporation on Si did not exist when our work started. Moll obtained permission from Tanenbaum for us to use his evaporator, and we quickly solved the problem of evaporating Al on Si, on "hot" or on "cold" Si, and were able to realize precise shallow alloyed p-type contacts or shallow "p" on "n" p-n junctions. We were able to show the various conditions under which uniform evaporated Al contacts could be realized on Si: (1) If the Si substrate was 660°C or hotter, the evaporated Al (mobile Al) nucleated at random sites that grew into larger diameter islands with more evaporated Al, and formed a discontinuous regrown region. (2) In the temperature range between the Al-Si eutectic (577°C) and the melting point of Al (660°C), uniform sticking and wetting of the Al occurred and formed continuous metallized and alloyed-regrown p-type Si without further heating. (This was nothing more than Hall's "local" liquid phase epitaxy LPE [139,140]. (3) For the Si substrate at temperature 577°C or lower, the evaporated Al merely adhered (uniformly) on the Si, and subsequently could be alloyed or could be left as a Schottky barrier. By late 1954, Goldey and Holonyak had solved the problem of metallizing Si and forming uniform shallow p-n (or n-p) junctions, or if desired, shallow ohmic contacts. Holonyak soon wrote a BTL memorandum on Al metallization and shallow junction formation on Si [190] and Goldey incorporated this material and some further results in a report published later [191]."

Contacts from the junction transistor to the header were usually made by thermocompression bonding, developed at BTL by O.L. Anderson, H. Cristensen and P. Andreasch [4]. Typically, gold (melting point 1063°C) was brought in contact with the aluminum bonding pad (melting point 660°C) in a reducing atmosphere under pressure. The gold-aluminum eutectic formed at about 350°C which, upon cooling, formed a strong, reliable bond. It was subsequently observed that a phenomenon referred to as "purple plague" often developed due to undesired reactions between the gold wires and the aluminum bonding pads at the upper range of temperatures where silicon transistors operate. The aluminum and gold formed a series of colored intermetallics (i.e., the purple plague) that ultimately caused device failure [43]. Rectification of this yield degradation was subsequently achieved by restricting the temperature range of operation (also required for the plastic packages then utilized), utilization of all aluminum systems using aluminum leads, wires and aluminum coated package connections [43]. Gold metallization systems such as the beam lead method [192] and multilevel metallization schemes [193,194] were also developed.

The benefits of the planar research in conjunction with Hoerni's insights resulted in the first meaningful description of the MOSFET device (formation of an inversion layer, i.e., enhancement mode) by Dawon (David) Kahng and Attala in 1960 [195-198], also summarized by Sah [44]. Kahng wrote an extensive BTL technical memorandum on the silicon/silicon dioxide device in 1961 [199]. Steven Hofstein and Frederick Heiman of the Radio Corporation of America (RCA) followed with an MOS IC consisting of 16 silicon n-channel MOS transistors in 1963 [200]. It should also be noted that J. Torkel Wallmark of RCA took out a patent on an FET in 1957 [201] but apparently did no further work in the field while Paul Weimer constructed an FET using CdS as the dielectric material on an insulating substrate in 1959 [202]. Although these structures were minority-carrier devices, the Metal-Semiconductor Field Effect Transistor (MESFET) and the Junction Gate Field Effect Transistor, see Figure 5 [4], first described and patented by Shockley [203,204] and built by George Dacey and Ian Ross [205] were majority carrier devices [44,206,207].

MOSFET Transistor Fabrication

The description of the oxidation process and methodologies for controlling the electrical properties of the silicon/silicon dioxide interface in the late 1950's were essential for the successful commercialization of the MOSFET and implementation of the DRAM memory era in the 1970's (see the *Integrated Circuit* section.) Deal and Grove described the oxidation kinetics of silicon [23], followed by Dennis Hess [208], Eugene Irene [209], Hisham Massoud [210] and Stanley Raider [211] and their colleagues who extended this research. Richard Williams [212] and Akos Revesz [213] also made

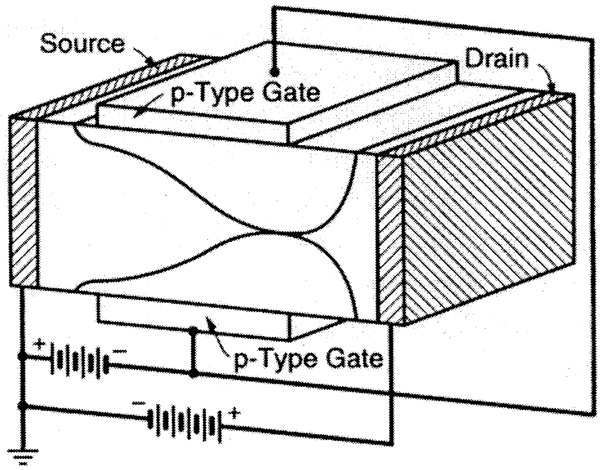

Figure 5. Schematic of the junction field-effect transistor [4]. Reproduced by permission of the IEEE, Inc.

significant contributions to the understanding of the silicon/silicon dioxide interface while George Schnable [214] advanced our understanding of a host of dielectric film deposition methodologies for Integrated Circuit (IC) applications [8].

Device reliability studies by Ed Snow, Grove, Deal and Sah at Fairchild Semiconductor identified that sodium contamination in SiO_2, introduced by the heated tungsten filament [215] for aluminum evaporation, was mobile under voltage stress, caused device parametrics to drift under operating bias and was exacerbated by increased operating temperature. The Fairchild team also observed that electron-beam evaporation did not introduce the sodium [215], thereby developing techniques for controlling the sodium and, furthermore, developed an extensive understanding of the phenomena taking place in the metal-oxide-silicon system that is basic to all modern MOSFET systems. Deal described the silicon/silicon dioxide electrical interface stability and associated effects in silicon dioxide in terms of 17 types of charge mechanisms and introduced the standard description for the charge notation associated with thermally oxidized silicon [24-30]. The reduction of mobile charges, fixed charge (Q_f) and interface state charge (D_{it}) as well as the control of the growth process (oxide thickness) was of paramount importance in ensuring threshold voltage control and the successful commercialization of MOS device products [216]. While most of the industry initially chose to fabricate PMOS devices, some companies elected to fabricate NMOS because of the higher channel mobility for electrons, compared to holes. However, positive charge control is a more serious issue in NMOS technology. Post-oxidation and post-metallization anneals were developed in the 1960's to minimize both fixed and interface charge [31,32]. The mobile charge, such as Na and K, also required stringent control.

Techniques were developed for the passivation of surface states introduced at the silicon/silicon dioxide interface during thermal processing. Pieter Balk described in 1965 the significance of a post SiO_2 anneal in a hydrogen bearing ambient [31] and a nitrogen anneal in the case of the Al-SiO_2-Si system [32] to stabilize the Si-SiO_2 interface and reduce the fixed charge, Q_f. Molecular hydrogen was suggested to anneal the surface states by bonding with the dangling silicon and oxygen bonds [31,32]. Kooi of Philips Research Labs in Eindhoven confirmed Balk's research [217]. Sah has noted that Balk's hydrogen annealing methodology has withstood the test of time for more than 30 years and is a fundamental aspect of the MOSFET IC processing methodology [44]. Sah has quoted from Balk's abstract [31] (Sah's comments are added in curly brackets):

> "The main effect of the H_2 treatment appears to be the annihilation of fast states {another name for interface states}. If these states are related to vacancies, accompanied by chemically unsaturated bonds {has been known as dangling bonds} and unpaired electrons near the interface, the H_2 {not hydrogen ion or proton as some think} annealing may be in effect the chemical saturation {now known as hydrogenation} with H atoms of these bonds at the vacancies. The low state density obtained upon steam oxidation is probably caused by hydrogen, evolved during oxidation, and retained in the oxide. The similarity in action between H_2 and Al remains as yet unexplained."

Balk clarified the unresolved issue as regards the similar benefit between a H_2 and an N_2 anneal of Al later that year in 1965 [32]. Sah [44] again quotes Balk:

> "The similarity of the annealing behavior of the electrical interface properties of Si-SiO_2 in H_2 and Al-SiO_2-Si in N_2 around 300°C suggests that the same mechanism is operative in both cases. Hydrogen released in a reaction between Al and hydroxyl groups in the oxide is proposed as the active agent in the Al-SiO_2-Si case. This model is supported by the absence of any annealing effects on 'ultra-dry' oxide."

It was suggested that residual H_2O, released from the Al during thermal processing, reacts with the hydroxyl groups to yield hydrogen. Sah further points out the beneficial effect of annealing in H_2 or forming gas (95% N_2/5% H_2) [218]:

> "Balk's hydrogen bond model of passivating and deactivating the interface traps has also been the chemical-atomic base for the characterization of the generation, annealing and charging kinetics of the interface and oxide traps due to silicon and oxygen dangling bonds."

Balk's insight was extremely important during the early 1970's, when Al was still the dominant gate electrode, before the introduction of the polysilicon gate electrode and the fabrication of the 1K and, in some cases, the 4K dynamic random access memory (DRAM)

IC. The last thermal process step (before packaging) was a 30 minute or so anneal between 400°C or 450°C to ensure sufficient reaction of the aluminum with the silicon for good contacts; the release of H_2 from the aluminum during the 450° C anneal was instrumental in passivating the interface states and recovering the desired threshold voltage. The role of hydrogen annealing and passivation in the broader context of the plasma deposited overlayer of silicon oxynitride as a seal over the entire circuit has also been discussed [219].

Dalton and Dorbek of AT&T [220] demonstrated that an overlayer of Si_3N_4 could also provide an effective seal against sodium ions; to avoid the concurrent trapping of hydrogen ions resulting in threshold voltage instabilities, silicon oxyitride was subsequently utilized. Indeed, a plasma deposited overlayer of silicon oxynitride is used as a seal over the entire circuit structure, except for the contact pads [4]. The nitride film is also very effective in reducing pinholes in the SiO_2 dielectric. Concurrently, Kerr and Don Young of IBM were developing the utilization of a phosphosilicate glass (PSG) deposited on top of the MOS gate dielectric silicon dioxide to getter the Na and K and stabilize the oxide film [221,222]. This approach was especially important for aluminum gate electrodes, utilized prior to the phosphorus doped polysilicon gate technology to be discussed below. Snow and Deal [223], followed by Pieter Balk and Jerome Eldridge [224], showed that the threshold voltage shifts of MOSFET's, induced by polarization in the PSG layer on the SiO_2 surface, could indeed be controlled. Stabilization of the surface was also beneficial for bipolar transistors for operation at low currents or high voltages where deterioration of the current gain and leakage current due to surface instabilities degraded device performance and yield.

Concurrently, Rudolf Kriegler championed an in-situ furnace gettering methodology in the early 1970's, to remove sodium as well as other deleterious contaminants such as metals and transport these mobile ions and lifetime-killing metallic impurities from the wafer to the gaseous ambient. This procedure became an especially prevalent industrial technique [225-227]. Typically, about 3% gaseous HCl or Cl_2 in the O_2 oxidation ambient [227] was utilized, similar to Robinson and Heiman [228]. Carl Osburn studied the improvement of gate oxide integrity (GOI) via these Cl methodologies [229] as part of an extensive series of analyses in the Very Large Scale Integration (VLSI) laboratory of Arnold Reisman [230]. TCE (trichloroethylene) [231] and TCA (trichloroethane) [232] were subsequently utilized with similar benefits, but without the corrosive effect of the HCl ambient on the furnace metal plumbing.

Gettering was initially believed to occur by formation of volatile metal chlorides, although the Gibbs free energy of formation of most metal chlorides was not negative. The Cl, however, was also interpreted as removing interstitials, as evidenced by the shrinkage of oxidation induced stacking faults (OISF) at sufficiently high temperature by Hiromitsu Shiraki [233], Cor Claeys [234] and their colleagues. Shih-Ming (Jimmy) Hu showed that the shrinkage (retrograde growth) of OISF at sufficiently high temperatures in the absence of HCl was dependent on both the surface orientation and ambient [235]. The early utilization of these Cl ambients was in conjunction with the oxidation ambient during high-temperature processing. The Cl incorporated in the SiO_2 also trapped and immoblized the sodium at room temperature. It was suggested by Kreigler and Osburn, furthermore, that it might be more advantageous to clean the furnace quartz tube in the presence of the Cl bearing species but not to incorporate the Cl into the SiO_2 film, per se, due to SiO_2 reliability considerations [227,229].

The subsequent capacitance-voltage (C-V) diagnostic analysis of the Si-SiO_2 interface electrical properties by Moll [236] and Terman [237], was expanded upon by Grove, Snow, Deal and Sah [238] and formulated in a set of useful charts for fabrication engineers by Zaininger and Heiman [239]. Edward Nicollian and Adolf Goetzberger's conductance analysis [240-243] quantified the description of the silicon/silicon dioxide interface electrical properties while the p-n junction under non-equilibrium conditions was described by Grove and Fitzgerald [244]. The description of oxidation kinetics by Deal and Grove [23] and the local oxidation of silicon (LOCOS) process developed by Kooi and colleagues in the late 1960's [245-247], which was instrumental in the fabrication and achievement of superior MOS IC characteristics (see Figure 6) [247], set the stage for the initiation of the MOSFET Dynamic Random Access Memory (DRAM) era in the early 1970's. The LOCOS process has been the mainstay for CMOS IC fabrication for more than 30 years and only now is shallow trench isolation seriously challenging its utilization [248].

The mesa and planar processes described above now paved the way for the fabrication of the IC by Jack Kilby (utilizing the mesa methodology) [8,33-36,43] and Robert Noyce [8,37-39,43] (utilizing the planar procedure), in 1958 (see *Integrated Circuit Beginnings* section) and, subsequently, the microprocessor [40-42].

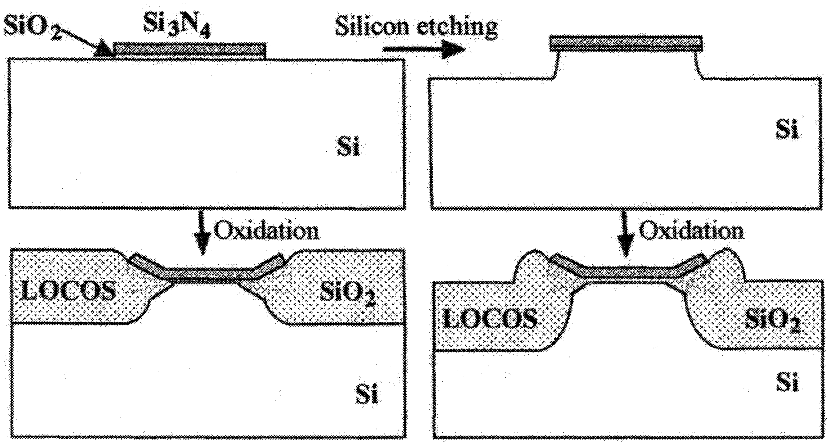

Figure 6. Conventional LOCOS procedures. A pad oxide (SiO_2) under the nitride oxidation mask relieves stress, but results in the formation of "bird's beaks" at the oxide edges. Fully recessed oxide patterns (right) exhibits complete "bird's heads" [247]. Reproduced by permission of The Electrochemical Society, Inc.

Integrated Circuit Beginnings

The challenge after implementation of the junction transistor era in the 1950s was to not only emulate the vacuum tube in as many applications as possible (without the excessive power generation and reduced operating lifetime), but to exploit the inherent advantages of solid-state electronics to new arenas. The smaller device dimensions required to achieve higher frequency operation in the junction transistor was confronted, however, with the inherent challenge of the limited power-handling capability due to the device's small size. The goal of achieving higher operating frequency and higher power-handling capability seemed to be at odds with each other. Ross described the situation as follows [4]:

> "In the meeting on that day, we were, as was frequently the case, discussing our problems in emulating the vacuum tube. R. Wallace suddenly said:
>
> "Gentlemen, you've got it all wrong! The advantage of the transistor is that it is inherently a small size and low power device. This means that you can pack a large number of them in a small space without excessive heat generation and achieve low propagation delays. And that's what we need for logic applications. The significance of the transistor is not that it can replace the tube but that it can do things that the vacuum tube could never do!"
>
> "And this was a revelation to us all. We realized that in chasing the vacuum tube, we had the wrong emphasis.... The net result was that the semiconductor community began to relax about replacing the tube and focused on developing the transistor in its own right. The transistor did eventually replace the tube in all but a few special applications, the magnetron being one outstanding example. But it took decades. In the meantime, semiconductor technology opened up important new fields that the tube could never have supported.... Having the clear goal of an application for an invention is a powerful stimulus for innovation. But frequently, the original application turns out not to be the most important."

In a similar vein, Robert Lucky has recently noted "moreover there is no *a priori* way to determine what will tip a market. It's a fundamental instance of chaos in-group dynamics. And that makes it fundamentally difficult to predict future societal behaviors in the adoption of technologies [249]."

Until the invention of the IC, electronic systems were comprised by individually connecting the various components (vacuum tubes or transistors, diodes, capacitors, resistors and inductors) together. The common feature of these endeavors was the wiring together of discrete and separately packaged device components. Of course, it was essential that these components be spaced sufficiently close so that the system propagation delay did not become the factor limiting the system speed. This required the miniaturization of the system, not just the device components. Two major system concerns surfaced which required rectification. This involved the assembly yield and reliability of a system with thousands of device components, which might be unacceptably low. Additionally, even if the device components had no errors, there would be a multitude of connections, resulting in the infamous "tyranny of numbers" [35,36,250-253].

Lester Hogan reviewed what may be the earliest attempts to rectify the "tyranny of numbers" conundrum [254]; that is, the patents filed by both Darlington [255] and Oliver [256] in 1952. Darlington and, apparently, Oliver used a grown junction transistor; both patents integrated several transistors on one piece of germanium or silicon, although they included no passive components. Geoffrey Dummer of the Royal Radar Establishment (RRE) at Malvern, England initiated his solution to the integration challenge in 1952 [257] and subsequently described his work at the Malvern Components Symposium in 1957 [258] and elsewhere [259]. Runyan and Bean [43] have quoted Dummer's 1952 status as "an integrated approach using a monolithic block comprising" [258,259]:

> "...layers of insulating, conducting, rectifying and amplifying materials, the electrical functions being connected directly by cutting out areas of the various layers."

Hogan [254] has also quoted a portion of Dummer's presentation at the 1957 meeting [258]:

> "... a transistor flip-flop with two emitter follower outputs—a total of four transistors all contained in a chip of silicon 125 mils by 375 mils. The semiconductor was doped to form a p-n-p structure and had various sections removed to leave thin bridges of material with relatively high resistances. These high-resistance paths formed the collector and emitter loads of the transistors connected to common power supply rails. Other resistors were provided by films of resistive material deposited on the surface of the silicon, while capacitors were constructed from thin metallic layers with insulators between."

Concurrently, Harwick Johnson of RCA was also developing his solution to the integration challenge [260]. Hogan [254] described Johnson's contribution:

> "As early as 1953, Harwick Johnson of RCA conceived of a complete phase shift oscillator built in a single chip of n-type germanium where p-n junctions supplied the necessary capacitance, the body resistance of the piece of germanium supplied the resistive elements and an alloy transistor at one end of the filamentary piece of germanium supplied the necessary amplification."
>
> "The significant point is that only two years after the first junction transistor was reported [reference added [66], research people were already trying to combine resistors, capacitors, (diodes) and transistors into one piece of semiconductor material in order to reduce size, to reduce the number of interconnects and to improve reliability."

The alloy junction transistors utilized by Johnson as well as the other, multi-faceted, approaches by personnel familiar with the state-of-the-art in the attempts to build an integrated circuit in the mid 1950's was perhaps best described by Dummer to be "pioneering stages" ... not capable of production [261]." Dummer's review of the work in Great Britain and Western Europe is also of interest [262].

Runyan and Bean [43] have commented about Johnson's contribution:

> "The figures included in a patent ... by Harwick Johnson ... (reference [260] added) bear a superficial resemblance to an integrated circuit. However, as expressed in the first sentence of the patent, both the discussion and claims relate only to a transistor phase-shift oscillator: 'This invention pertains to semiconductor devices and particularly to semiconductor phase-shift oscillators and devices.' Component isolation was not considered so that even if the concepts of the patent were extended to devices other than the phase-shift oscillator, the class of devices that could be made would be very limited."

It was Kilby of Texas Instruments, Incorporated, however, who filed a patent application on February 6, 1959 [8,33-36,43] explicitly "describing a concept that allowed, using relatively simple steps, the fabrication of all the necessary components of the desired circuit, both active and passive, in a single piece of semiconductor and their interconnection in situ" [43]. Kilby's initial proof of concept was a phase shift oscillator, built with about ten components, in germanium for expediency [43] on September 12, 1958. Wire bonding was utilized to interconnect the components within the chip (see Figure 7).

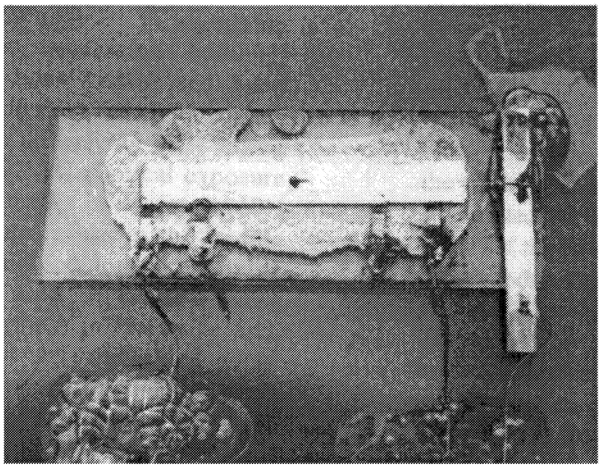

Figure 7. The first integrated circuit, a phase shift oscillator fabricated in germanium by the mesa process, invented by Jack S. Kilby of Texas Instruments in 1958, courtesy of Texas Instruments Incorporated.

A few weeks later, a flip-flop circuit was made and a patent application covering both germanium and silicon was prepared and filed (February 6, 1959). The first commercially available IC, intended for binary counter, flip-flop or shift register applications, was fabricated in silicon and announced in March, 1960 by Texas Instruments. Runyan and Bean [43] have extracted several relevant portions of Kilby's patent [33], with appropriate commentary:

> "In contrast to the approaches to miniaturization that have been made in the past, the present invention has resulted from a new and totally different concept for miniaturization.... In accordance with the principles of the invention, the ultimate in circuit miniaturization is attained using only one material for all circuit elements and a limited number of compatible processing steps for the production thereof.... "

Up to this point, the goals are perhaps not much different from those expressed in 1952 by Dummer. However, to continue with the Kilby patent:

> "In a more specific conception of the invention, all components of an electric circuit are formed in or near one surface of a relatively thin semiconductor wafer characterized by a diffused p-n junction or junctions...."

> "It is a primary object of the invention to provide a miniaturized electronic circuit wherein the active and passive circuit components are integrated within a body of semiconductor material, the junctions of such components being near and/or extending to one face of the body, with components spaced or electrically separated from one another as necessary in the circuit...."

> "Figures 1-5a (in reference 33 added by author) illustrate schematically various circuit components fabricated in accordance with the principles of the present invention in order that they may be integrated into, or as they constitute parts of, a single body of semiconductor material:"

Runyan and Bean [43] continue, "The figures and text describe bulk resistors, diffused resistors, pn junction capacitors, MOS capacitors, transistors, and diodes. In the press coverage of the March 1959 announcement of the Kilby concept, this set of standard components was stressed [263]. The patent text continues:"

> "Because all of the circuit designs described above can be formed from a single material, a semiconductor, it is possible by physical and electrical shaping to integrate all of them into a single crystal semiconductor wafer containing a diffused p-n junction, or junctions, and to process the wafer to provide the proper circuit and the correct component values...."

With the subsequent planar process patent submission by Hoerni of Fairchild Semiconductor Corporation on May 1, 1959 [15] and due to manner in which the interconnection was described in Kilby's patent [33] compared to Noyce's IC patent application, filed on July 30, 1959 [37], Noyce was actually awarded a patent before Kilby's (April 25, 1961 compared to June 23, 1964). Figure 8 is a photomicrograph of one of the first planar integrated circuits made at Fairchild Semiconductor Corporation [122]; subsequent evolutionary trends have been described for ICs [44]. The Texas Instruments IC was developed utilizing the mesa process while Intel utilized the subsequent planar process. The legal battle that ensued between Texas Instruments and Fairchild Semiconductor centered around the wording associated with the interconnection scheme. Runyan and Bean [43] have summarized the essence of the point of contention between the Kilby and Noyce patent dispute, quoting Kilby's patent [33]:

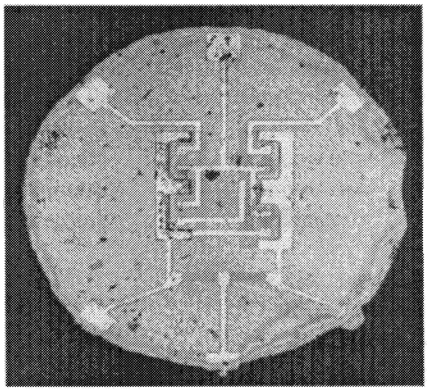

Figure 8. Photomicrograph of one of the first planar integrated circuits made at Fairchild (in silicon). This is a flip-flop circuit (with two transistors). Some of the aluminum interconnection metal has been damaged during the etching operation to form a circular chip of silicon to place into a transistor can modified to have more leads. (From "A Solid State of Progress," Fairchild Camera and Instrument Corporation, 1979) [122]. Reproduced by permission of the IEEE.

> "Instead of using the gold wires 70 in making electrical connections, connections may be provided in other ways. For example, an insulating and inert material such as silicon oxide may be evaporated onto the semiconductor circuit wafer through a mask either to cover the wafer completely except at the points where electrical contact is to be made thereto, or to cover only selected portions joining the points to be electrically connected. Electrically conducting material such as gold may then be *laid down* (italics entered by author) on the insulating material to make the necessary electrical connections."

then referencing the relevant portion of the Noyce claims [37]:

> "... an electrical connection to one of said contacts comprising a conductor adherent to *said layer* (italics entered by author) ...." The disagreement centered around whether "*laid down*" was equivalent to "adherent to." The Board of Patent Interference, ruling in Kilby's favor, asserted that it was. However, a subsequent ruling by the Court of Customs and Patent Appeals (Noyce v. Kilby; Kilby v. Noyce, decided November 6, 1969) reversed the previous rulings and allowed the Noyce claims. (The Supreme Court then refused to review the case.) Contrary to assertions by some, the ruling did not depend on whether gold could or could not be made suitably adherent to silicon oxide. The Court specifically commented on that aspect. The ruling depended on the Court's assessment of whether or not someone reading Kilby's statement would be inevitably drawn to the conclusion that the lead should be adherent."
>
> Runyan and Bean then continue [43]:
>
> "Probably the most balanced assessment of Kilby's and Noyce's relative contributions is contained in the citations of the Franklin Institute's 1966 Ballantine Medal award, which they shared. Kilby was credited for 'conceiving and constructing the first working monolithic circuit in 1958,' and Noyce for 'his sophistication

> of the monolithic circuit for more specialized use, particularly in industry.' "

Implementation of the IC concept was somewhat slow in consideration of the anticipated yield of an IC containing, say, 100-1000, transistors. That is, the reliability of a chip was anticipated to approximate the reliability of a discrete transistor degraded by a power to the number of transistors or components [4,264]. It turned out, however, that neither the yield or reliability was determined by random events degrading the transistor uniformity, per se, due to the batch method of silicon IC processing [4]. In point of fact, there tended to be large areas of a silicon wafer where the yield was close to 100% while the yield in other areas tended to zero [265]. Thus, if the IC chip area was small compared to the area on the wafer with high yielding ICs, the yield would be essentially independent of the chip size and the number of chips per wafer. With this realization, the stage was set for the IC explosion. Ross [4] has quoted Kilby in discussing the buildup of IC production [34].

> "It should be noted that one of the great strengths of the integrated circuit concept has always been that it could draw on the mainstream efforts of the semiconductor industry. It was not necessary to develop crystal growing or diffusion processes to build the first circuits, and new techniques such as epitaxy would be readily adapted to integrated circuit fabrication. Similarly, new devices such as MOS transistors and Schottky barrier diodes would be phased in as they became available. Even today, it is difficult to identify a process that is used only for integrated circuits.
>
> Another strength of the concept was that it could draw on existing circuit technology to produce a broad range of useful devices....
>
> Because of the commonality with existing processes, integrated circuits moved rapidly into a production status."

It might be appropriate, however, to note that developments in IC fabrication to enhance device performance and scaling have significantly expanded the spectrum of processes uniquely developed for IC fabrication, including, for example, self-aligned structures and the lightly doped drain (LDD) configuration (see Integrated Circuit Scaling section).

While the use of epitaxy had been an important design consideration for discrete transistors, its real impact occurred with the introduction of the bipolar IC. Prior to epitaxy, component isolation within the bipolar IC was achieved by reverse-biased p-n junction isolation techniques (introduced by Lehovec on April 22, 1959 [266,267]) such as by diffusion (consider boron) from both the front- and back-surfaces of the wafer until the diffusion fronts met in the center of the n-type wafer, leaving n-type wells in the masked regions. Runyan and Bean have reviewed the p-n junction isolation technique, including Lehovec's contribution to the interconnection methodology as well as the isolation techniques utilized by Kilby and Noyce [43]. Kenneth Bean has also described his epoch research on dielectric isolation for Bipolar ICs with Paul Gleim and Runyan [268,269]. The advent of epitaxial structures, however, offered a new design flexibility in component isolation. For example, one could utilize a thin n-type epitaxial layer (say 3 μm, for example) on a p-type substrate wafer. Component isolation could readily be achieved by boron diffusion through the epitaxial layer. Additionally, the fabrication of structures with a high concentration of dopant in the collector, to decrease transistor collector resistance, was achieved by the localized diffusion of an n-type dopant into a p-type substrate wafer prior to the growth of an n-type epitaxial layer [166-169].

Integrated Circuit Fabrication

Although the bipolar transistor exhibited better performance characteristics such as switching speed than the MOSFET transistor, the process simplicity and smaller IC chip size of the latter made it the preferred choice for implementation of leading edge design rule applications [270-272]. The first mass produced commercial MOS DRAM design was Intel's 3-transistor silicon-gate PMOS, 1K DRAM announced in 1970. Terman [273] and Hodges [274] have reviewed a number of these memory developments prior to 1972, often referred to as the small-scale integration (SSI) era (see Table 2 for an approximate taxonomy of the evolving DRAM).

TABLE 2: DRAM Process and IC Evolution (circa 1992)

Parameter	Units	ULSI	VLSI	LSI	MSI
Bits/chip	Number	$10^7 - 10^9$	$10^5 - 10^7$	$10^3 - 10^5$	$10^2 - 10^3$
Design Rule	μm	< 1	1 – 3	3 – 5	5 – 10
Power-delay product	PJ	$< 10^{-2}$	$10^{-2} - 1$	1 – 10	$10 - 10^2$
Mask levels	Number	15 – 20	8 – 15	6 –10	5 – 6
Chip area[a]	mm^2	50 – 280	25 – 50	10 – 25	10
Storage cell capacitor equivalent oxide thickness	(nm)	3.5 – 12.5	12.5 – 40	40 - 90	90 – 120
Junction depth	μm	0.04 – 0.2	0.2 – 0.5	0.5 – 1.2	1.2 – 2

a) Chip area (or more specifically, active device area) significantly impacts IC yield in conjunction with the defect density per cm^2 per critical lithographic level (for the number of critical levels).

The transition to the self-aligned aluminum gate and then the self-aligned phosphorus doped polysilicon gate adjacent to the source and drain junctions via ion implantation of the junctions, reducing the Miller capacitance [275-279] and the transition to the silicide metalization scheme [280] were significant achievements enhancing IC performance. An additional device innovation was Heiman's utilization of a four-terminal configuration for the MOSFET by applying a negative dc voltage to the substrate [281] which modified the threshold voltage of the MOSFET (increased the threshold voltage for an n-channel MOSFET) and is referred to as the "body effect." This may be seen in the threshold voltage, V_T, expression in equation 3 for an n-channel transistor (with no ion implant for V_T adjust) [282]:

$$V_T = -(Q_f/C_{OX}) + 2\phi_F + \phi_{MS} + (1/C_{OX})\,(2\,\varepsilon_{Si}\varepsilon_o\,q\,N_A)^{1/2}\,(2\phi_F + V_{BB})^{1/2} \qquad (3)$$

where:

Q_f = Fixed positive interface charge per unit area at the silicon-silicon dioxide interface. The positive charge contributes electrons to the p-type silicon at the surface, thereby making it easier to invert the p-type bulk silicon to an n-type surface inversion channel (i.e., lower V_T)

C_{OX} = Gate oxide capacitance per unit area

ϕ_F = Bulk Fermi energy relative to the the intrinsic Fermi energy (for 10^{15} holes/cm^3, ϕ_F = - 0.29 V)

ϕ_{MS} = Work function difference between the phosphorus doped polysilicon gate and the p-type silicon substrate (taking the Fermi energy in the phosphorus doped polysilicon at the edge of the conduction band, ϕ_{MS} = - 0.84 V)

$\varepsilon_{Si}/\varepsilon_o$ = Dielectric constants of silicon (11.7) and vacuum, respectively

V_{BB} = Substrate back-gate bias (V_{BB} = - 5 V for the 4K and 16K DRAM when this technique was popular). Substrate bias also reduces the junction capacitances of the source, drain and channel, an important performance advantage.

The body-effect technique also allowed determination of the bulk substrate doping, N_A, at the edge of the surface space charge region (SSCR) for comparison with that deduced by capacitance-voltage (C-V) analysis via equation 4:

$$N_A = [\,(C_{OX})\ \partial V_T/\partial(2\phi_F + V_{BB})^{1/2}]^2 \,/\, (\,2\ \varepsilon_{Si}\varepsilon_o\,q) \qquad (4)$$

Physically, the substrate near the surface is reverse biased by the negative V_{BB} body bias, thereby negating, somewhat, the influence of the electrons donated by the fixed positive charge, Q_f, at the interface and increasing V_T to the desired range. It should be noted that the $2\phi_F$ term does not scale with device scaling.

The larger number of device functions on a given MOSFET IC chip and the larger number of MOSFET IC chips for a given wafer diameter were instrumental in ensuring the eventual dominance of the MOSFET IC. Fairchild announced a 64 bit SRAM (six transistor cell design), enhancement-mode p-channel MOSFET in 1964 [44] followed by RCA's annoucement and production of an enhancement mode n-channel MOSFET, also in 1964, based on Hofstein and Heiman's research [283]. Frank Wanlass and Sah of Fairchild Semiconductor Corporation disclosed the Complementary MOS (CMOS) IC in 1963 [284,285] followed by RCA Corporation later in the year [286]. Wanlass's initial demonstration circuit, a two transistor inverter, consumed a few nanowatts of standby power, compared to milliwatts of standby power for equivalent bipolar and PMOS gates [287]. Interestingly, Wanlass utilized Heiman's back-bias methodology [281] to achieve an n-channel enhancement mode device (due to the difficulty of uncontrolled surface charges at that stage of technology to fabricate an n-channel enhancement-mode MOS transistor) to work in conjunction with the conventional PMOS enhancement-mode transistor [287].

CMOS eventually became the ultimate MOSFFET technology, since both p-and n-channel enhancement-mode transistors are normally off, drawing only quiescent power; that is, only during the switching process is significant power dissipated [286,288-290]. The DRAM memory array in CMOS ICs was fabricated in NMOS while the peripheral drivers were fabricated in CMOS. U.S. companies tended to use the same design rule for both the NMOS memory array and the CMOS peripheral devices, accentuating the latch-up phenomenon, thereby requiring the use of epitaxial structures to minimize latch-up (the coupling of an n-p-n MOS transistor with an adjacent p-n-p MOS transistor forming an n-p-n-p or p-n-p-n thrysistor [289,290]. It appears the Japanese used a larger design rule for the CMOS circuitry, thereby avoiding the use of epitaxial wafers. The Japanese approach was less costly to fabricate since polished, rather than epitaxial, silicon wafers were used. Although there were less chips per wafer due to the larger chip size, increased yields often negated the geometrical limitation. MOSTEK, Incorporated, formed in 1968, was the first semiconductor company exclusively devoted to the fabrication of MOSFET ICs. Shortly thereafter, the 256 and 1K DRAM MOSFET IC was introduced by Texas Instruments in 1970 and 1972, respectively. Intel introduced the 1K PMOSFET DRAM in 1970 [291]. Indeed, Intel's 1K p-channel (PMOS) DRAM (polysilicon gate), based on a three-transistor cell design, initiated the beginning of the MOS memory take-over of the ferrite core memory market by its implementation at computer maker Honeywell, Inc. [292].

The MOSFET IC revolution, however, really exploded when IBM chose the n-channel silicon MOSFET (NMOS), instead of the slower p-channel silicon MOSFET, for its mainframe memory computer (IBM-370/158) that was delivered in 1973. The access time of the NMOS was in the range of nsec while magnetic core memory's access time was about one µs. Intel and MOSTEK were early suppliers followed by TI in 1974. TI's design was based on the one-transistor DRAM cell structure of Bob Dennard [45] and described by others [293,294], also summarized by Sah [44]. Texas Instruments and MOSTEK utilized a single-metal-word-line/single-diffused-bit line [44,295-297], where the metal was Al and the source and drain were formed by diffusion. Texas Instruments utilized $POCl_3$ in forming the diffused source and drain. The 4K NMOS DRAM cell built by Intel was a single-poly-word-line/single-metal (Al)-bit line [44,295-297].

The 16K DRAM was announced in 1976, with three significant changes made compared to the 4K DRAM, noted by Sah [44]. These were a reduction of the design rule from the 7-8 µm regime for the 4K DRAM to about the 5 µm range for the 16K DRAM; the removal of the source diffusion which became known as the merged one-transistor DRAM cell and an overlapping double polysilicon gate, one for the sourceless transistor (the pass gate) and the second for the charge storage capacitor, thereby forming the merged one-transistor DRAM cell [44,295-297]. The wafer diameter was also subsequently increased to three-inches and, later, changes to larger diameters became commonplace when the number of DRAM chips became less than about 100 per wafer [298]. The subsequent transition to the 64K DRAM around 1979 did not result in any change in the cell design, although the design rule was reduced to the 2-3 µm range and the part was subsequently implemented on four-inch diameter wafers [298]. Four significant IC process changes were discussed by Sah [44]. These included a parallel plate storage capacitor, rather than storage in a surface inversion layer, to give a higher charge storage capacitance (about 32 fF to store $\approx 10^6$ electrons at 5V V_{DD}) from the small area of the one-transistor DRAM cell, since higher capacitance was required to reduce the soft errors due to noise electrons generated by alpha particles from the package materials of the chip, cosmic rays and other noise sources [44]; a dual dielectric for the charge storage capacitor, utilizing the higher dielectric constant of silicon nitride formed by chemical vapor deposition (CVD) on thermally grown silicon dioxide to enhance the composite storage medium's dielectric constant and to reduce pinholes in the thinner silicon dioxide (not all DRAM manufacturers utilized this option); plasma etch technology to produce steeper walls or trenches to reduce tapered structures which take up silicon real estate (chip area) and an optical wafer stepper to reduce the design rules from three to less than two microns. Rideout [296] and Chatterjee [299] have also reviewed these DRAM advances.

The 256K DRAM further reduced the design rule to the 1.5-2 μm range and introduced refractory metal silicides to reduce the interconnect wiring delay [44] and aluminum metal for double and triple polysilicon technologies. The M-bit DRAM era, initially a shrink of the original 2 μm 256K DRAM design, approached 1 μm design rules (see Table 2); more importantly, however, was the introduction of two three-dimensional (3-D) trench charge storage capacitors (see Figures 9 and 10). Sah has noted that the goal of these 3-D capacitor designs was to reduce the planar area of the storage capacitor while maintaining the storage capacitance at more than 32 fF to hold more than 10^6 electrons at a V_{DD} of 5V to limit soft errors [44]. In the stack capacitor design, multilayers of conductors (poly Si or Al) and insulators (silicon dioxide and silicon nitride) are stacked on top of the pass transistor. In the trench capacitor design, a trench is etched in the silicon and an MOS storage capacitor is fabricated in the trench, adjacent to the pass transistor which remains on the planar surface. In this case, the trench depth is about 10 μm and the spatial area is about 6-9 μm^2. Chatterjee and colleagues at Texas Instruments introduced a structure which placed the pass transistor inside the trench to further conserve silicon real estate [300,301].

The 4M DRAM era introduced the sub-micron design rule regime at 0.8 μm with 3-D storage capacitors. The types and features of storage cell designs have subsequently proliferated [44,302,303]. The decreasing design rules result in higher speed and reduced power-delay product as a result of lower capacitance and current [44]. The power-delay product is additionally reduced by reducing V_{DD} [44].

The DRAM became the test vehicle par excellence to advance the silicon IC process technology because of its repetitive memory structure. In more recent years, however, especially after the U.S. makers retreated from a significant position in the manufacture of DRAMS, their expertise in the fabrication of microprocessors has propelled the logic and microprocessor family as test vehicle drivers. Nevertheless, the DRAM continues to drive the extendibility of personal computers (PCs) vis-à-vis the memory content.

Integrated Circuit Scaling

Gordon Moore's remarkedly prescient assessment of memory component growth in 1965, initially based on bipolar and then MOS memory density, observed that a semilog graph of the number of bits on a memory IC versus the date of initial availability was a straight line, representing almost a doubling per year [50-53]. Accordingly, a quadrupling was deduced every two years (consistent with the needs of the system houses) and subsequently modified to ≈ 3 years around the mid-later 1970s and currently taken as 3-4 years based on a 1995 assessment [53]. This analysis became enshrined as

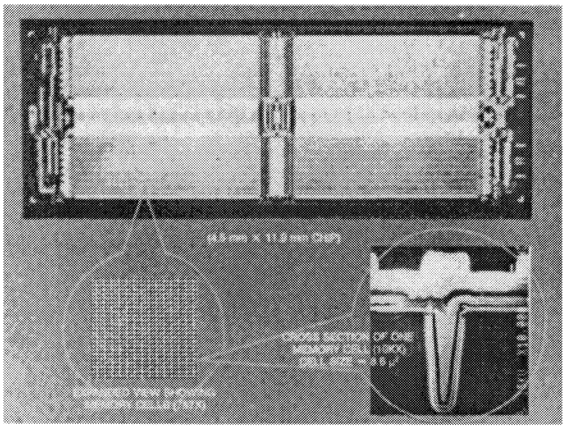

Figure 9. One Mbit CMOS DRAM chip, courtesy of Texas Instruments Incorporated.