13)[153-156] and their colleagues were just two of an extensive array of personnel studying oxygen in CZ silicon, to be further discussed below.

The disappearance of the upper yield point between approximately 1100-1300°C was interpreted as a result of the formation of SiO_x precipitates due to the supersaturated oxygen content.[148] Above 1300°C, however, the SiO_x structures dissociate as the degree of super-saturation decreases and the oxygen returns to its atomic form, with the concurrent recovery of both the upper and lower yield points. Detailed studies of the correlation of the precipitated oxygen on both the upper and lower yield point in CZ silicon were pursued by Koji Sumino[121,157-160] and Yojiro Kondo[161] and their colleagues. The functional dependence of the yield point on the dislocation speed was found to be related to the charge state of the dislocation kink and, therefore, dependent on the Fermi energy (*i.e.*, the doping level).[162] The comparative values of silicon's critical shear stress at different temperatures and varying oxygen precipitate concentrations have been summarized.[1,121,163-165] The critical stress increases with both reduced wafer processing temperature and reduced oxygen precipitation. Alternatively, oxygen precipitation can reduce the critical stress for subsequent plastic deformation by the generation of internal stresses around the precipitates. The oxygen microstructure and extent of precipitated oxygen was shown to be dependent on both material and IC thermal processes as well as equipment issues, such as the insert rate of the boat into the furnace (set a given dwell temperature), wafer spacing, thermal design of the furnace, etc.[1,121,165,166] The dislocation structure was noted to preferentially form on the concave side of a bowed wafer by Bernard Leroy, Lawrence Dyes and Howard Huff (Fig. 14) and their colleagues.[164,167] Excessive oxygen precipitation can also preclude effective lithographic printing due to uncontrolled wafer warpage as shown by Hirofumi Shimizu and colleagues.[168,169] The increased ultilization of rapid thermal annealing (RTA) requires attentiveness to minimize global alignment errors as well as to monitor warpage so as to minimize global alignment errors in lithography. Site flatness degradation may not be as severe, and may be corrected for by site-to-site realignment. Although a number of these observations were for significantly different wafer diameters and oxygen content, the comments were generically applicable to 200 mm and may be useful guidelines for the approaching 300 mm diam era.

In view of the importance of the oxygen content for internal gettering (IG) (see Gettering section below) and considering the influence of uncontrolled oxygen content in plastically deforming and warping the wafer, the comprehension of oxygen in silicon became very critical. Detailed investigations of the mechanisms and concentration of oxygen incorporated in the grown CZ crystal were implemented[1,121,149,165,170-174] The oxygen microstructure and extent of precipitated oxygen was pursued by numerous investigators, covering the oxygen range as high as 2.0×10^{18} cm^{-3}, corresponding to 40 ppma (though typically 1.5×10^{18} cm^{-3}, corresponding to 30 ppma) to as low as $2\text{-}5 \times 10^{17}$ cm^{-3}, corresponding to 4-10 ppma.[1,121,149,165,170-174] The importance of concurrently reducing the carbon concentration to less than 1×10^{16} cm^{-3}, 0.2 ppma to minimize its catalytic effects on oxygen precipitation was also noted.[175,176]

Precipitated oxygen enhanced plastic deformation in CZ wafers, compared to float-zone (FZ) silicon (utilized for power devices), where the oxygen content is typically less than 10^{16} cm^{-3}.[172] Small clusters of nitrogen in FZ crystals, however, were shown to be even more effective as regards work-hardening compared to oxygen in CZ crystals.[172,177] Nevertheless, FZ material has not replaced CZ material as the dominant material for IC fabrication, probably as a result of the extensive experience already in place with the utilization of CZ material, the wide installed base of IC processes consonant with CZ material and the larger wafer diameter available with CZ, compared to FZ, material.

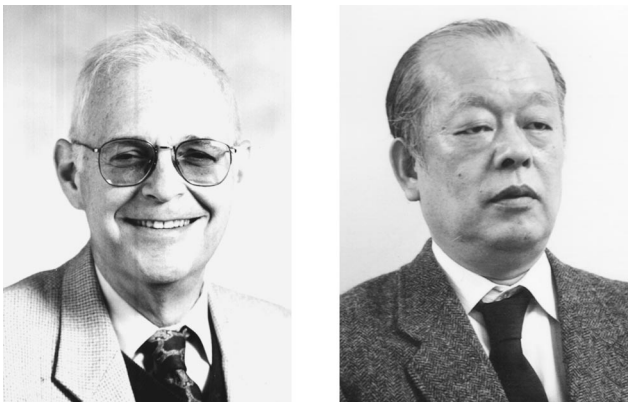

Figure 14. (*left*) Howard Huff and (*right*) Shin-ichiro Takasu.

Wafer Mechanics

The commercial introduction of dislocation-free crystals in the mid-1960s clarified the differences among dislocations; grown-in during crystal growth, introduced into the crystal during the mechanical conversion of the crystal into polished wafers, and generated in the wafer during epitaxial deposition and IC fabrication. A silicon wafer prepared for microelectronics applications requires a polished, flat upper surface free from residual edge mechanical damage and edge chips, etc. (due to the shaping processes in the conversion from the crystal to a polished wafer) and minimal residual polish damage, chemical contamination, and particulates, as well as exhibit controlled surface microroughness.[1,121,178,179] It was experimentally found that extreme care is required to maintain the optimally prepared silicon wafer during device and IC fabrication. It was noted as early as 1966 by Schwuttke that ''one should handle wafers as little as possible and use extreme care when it is necessary.'' [108] Today's 300 mm diam silicon wafers have both a polished wafer back-surface and edge. The formation of dislocations was shown to be an extremely process- and structure-sensitive parameter. Designing a wafer from a process-structure-property perspective, in order to avoid plastic deformation during IC fabrication, was usefully employed to assess, monitor and control the above variables in the multidimensional wafer and IC process space.[1,121,165]

Radial thermal gradients developed within silicon wafers as a result of nonuniform heating during the approach to (or cooling during the descent from) the epitaxial or furnace process temperature, the insert and dwell temperatures (and strain rate due to the boat insertion/withdrawal speed) and the furnace ramp rate to the processing temperature during the furnace process. The location of wafer support points and wafer spacing in conjunction with the wafer diameter, type of boat material, and related equipment and process parameters were also critical. The furnace boat shape and the material contacting the wafers also significantly influenced the thermal gradient within the wafers. A completely enclosed boat considerably reduced the thermal gradient across the wafer by reducing the rate of heat transfer to or from the wafers.[180] The portion of the wafer near the contact point with the rails of the boat can exhibit a significant temperature gradient and, therefore, activate dislocation sources. This gradient is the result of both local heat flow nonuniformities due to differing thermal conductivities between silicon and the quartz glass boat[165] as well as differences in their coefficients of thermal expansion. The latter source acts when the silicon wafer sticks to the quartz glass boat. Plastic deformation has been noted to be more difficult to initiate at these points of contact by the utilization of silicon carbide or silicon, rather than quartz glass, boats due to the more favorable matching of the coefficients of thermal expansion.[165]

The mathematical assessment of plastic deformation during IC fab oxidation processes was initiated by Shih-Ming (Jimmy) Hu

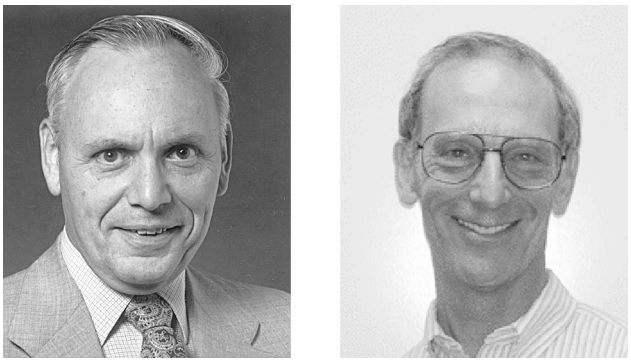

Figure 15. (*left*) Jim Amick and (*right*) Phil Tobin.

Figure 16. (*left*) Takao Abe and (*right*) Eicke Weber.

(Fig. 34), Electronics Division 1985 Awardee, who studied the conditions under which wafers would plastically deform, including the role of the wafer diameter, wafer thickness and wafer spacing.[181] An extensive literature was subsequently developed,[1] including the research of Kenji Morizane and Paul Gleim[182] and applications to rapid thermal processing (RTP) by Mehrdad Moslehi,[183] and Martin Schrers and colleagues.[1,184] The pioneering studies of Runyan in 1959 showed that large diameter, 150 mm wafers could indeed be fabricated by the Teal-Little methodology,[185] and the procedure has morphed to 300 mm at the present time. The combined influence of thermal and gravitational stresses for 300 mm diam wafers has been extensively discussed in the 1990s and, at this time, the engineering design issues have been shown to be solvable by Huff and Randal Goodall in conjunction with Robert Nilson and Stuart Griffiths as well as by Shimizu and colleagues.[1,160,165,186–188]

The mathematical formulation describing the thermal stress in silicon wafers during epitaxial deposition was initiated by Howard Huff and colleagues[167,189] and extended by Jan Bloem and colleagues.[190] Significant improvements in epitaxial process technologies were subsequently described by Mac Robinson, Jon Rossi, Rick Wise, and their colleagues.[191-193] In light of these developments to control plastic deformation, the experimental observation by Leroy, Huff, and Dyer and their colleagues that plastic deformation could also occur in the central regions of epitaxial wafers, where no macroscopic radial gradient existed, was quite interesting.[164,167] This was interpreted as a result of the entrapment of a hot region within a colder annular zone in a concave shaped wafer.[164,167,194] Further examination of the front- and back-surface of the wafer indicated that dislocations moved out of the wafer and preferentially piled up on the front (*i.e.*, concave) shaped wafer surface or near the periphery area due to its larger (compressive) stress compared to the convex surface.[164,167,194] This observation was clarified via a detailed study of the warpage of silicon wafers, induced by oxygen precipitation in wafers processed in furnaces, without the complication of a back-surface epitaxial susceptor contact.[164]

The mechanical properties of silicon, including studies on the change in wafer shape due to process-induced warpage during wafer heating and cooling (observed to initiate at the wafer periphery and center, respectively), were summarized in the classic article by Shinichiro Takasu (Fig. 14), based on his research at the VLSI Technical Research Association in Japan from 1976-1979.[195] Minimizing the wafer warpage was shown by Shimuzu and colleagues to be facilitated by reducing the insertion/withdrawal dwell temperature of the wafers to/from the furnace to less than about 850°C[168,169] and by Phil Tobin (Fig. 15) and colleagues who illustrated the interactive effects of oxygen precipitation, wafer warpage, wafer spacing, and the push/pull temperature for critical steps.[196] The wafer bow was also modified not only by thermal processes or oxygen precipitation during IC fabrication but by unbalanced front and back surface film stresses resulting from the formation and selective removal of various films.[197] Experimental procedures were derived in the late 1960s and early 1970s to control the activation of dislocations and subsequent wafer warpage when the boat was manually introduced into the furnace operating temperature, although the usefulness of dislocations for gettering purposes was also noted.[1,121,165] Introduction of automatic ramp programs in the 1970s resulted in an extensive literature,[1,121,165] indicating that 10°C/min insertion rates and 5°C/min cool rates were generally effective procedures to minimize the onset of plastic deformation for 75 mm diam wafers and a wafer spacing of 125 mils.

Point-Defect Dilemma

The introduction of dislocation-free Si wafers removed a significant source of bulk gettering sites which would have immobilized metals spuriously introduced during IC processing. The metals now preferentially segregated within the strained SCR of the device, often in conjunction with oxidation induced stacking faults, OISF,[198-200] and dislocation arrays introduced during IC fabrication, as described by Jim Amick (Fig. 15), ECS President in 1994-1995 and ECS Fellow.[200] This resulted in the degradation of their electrical characteristics by modifying the electric field distribution in the p-n junction[41,121,142] as well as degrading DRAM refresh characteristics.[201] K.V. Ravi, author of *Imperfections and Impurities in Semiconductor Silicon*,[202] Shyam Murarka, ECS Fellow, 1987 Callinan Awardee, and 2001 Electronics Division Awardee and colleagues[203,204] and others have extensively described the degraded electrical characteristics of p-n junctions[142,205-207] and gate oxide integrity (GOI),[208-213] due to OISFs metallic precipitates and metallic decorated structural defects such as dislocations and OISFs. Indeed, the model of p-n junction degradation due to imperfections in the SCR, metallic precipitation at LOCOS edges, trench structures and ion-implantation induced defects at spacer edges bordering SCRs, as well as on stacking faults in the p-n junction, was generally found to be more detrimental than the generation-recombination (g-r) effect of the metals, per se, as discussed by Bernd Kolbesen and Horst Strunk.[1,121,214] This phenomenon, noted earlier, was described as the Point-Defect Dilemma by Erhard Sirtl.[124]

Clarification began with the realization by a number of personnel that dislocations grown-in during crystal growth behaved differently than dislocations introduced during device or IC fabrication. For example, dislocations introduced during crystal growth were generally saturated with oxygen atoms, a major contaminant in silicon during CZ growth.[117,121] Accordingly, the direct electrical effects anticipated with dislocations were often not observed, as discussed earlier.[121,122] A moderate density of dislocations, however, did yield beneficial results in device and IC performance,[1,121,125-128] although extensive grown-in dislocations did exhibit deleterious effects in both the fabrication and characteristics of transistors.[1,113,121]

Over and above the point-defect dilemma, an increased understanding of the incorporation of point defects into the solidifying silicon crystal was also being obtained.[172,215-219] The grown crystal,

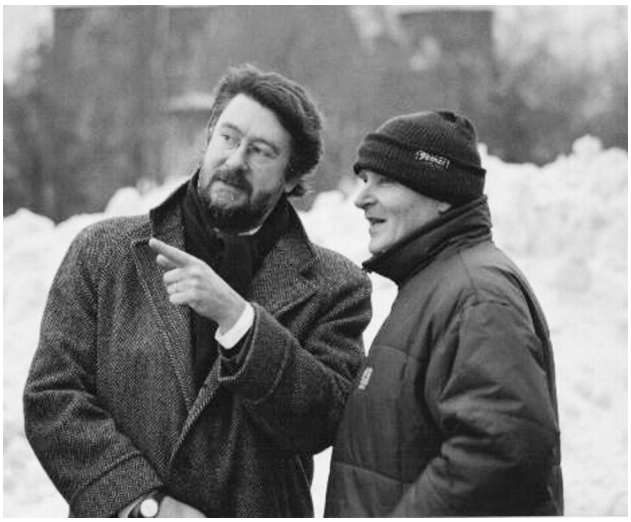

Figure 17. (*left*) Bob Falster and (*right*) Vladimir Voronkov.

sectioned along the direction of crystal growth, was observed by Takao Abe (Fig. 16) to generally exhibit a spatial demarcation between vacancy-rich and interstitial-rich regions, depending on the crystal growth parameters.[172,215-219] This demarcation region was observed as *OISF* generated during subsequent thermal processes by Eric Dornberger and von Ammon.[220] The degradation of GOI due to OISF, generally decorated by metals, was noted earlier.[208-213] Vacancy-rich CZ material was also shown to result in GOI failures in 20 nm oxides, due to the agglomeration of the vacancies into macroscopic voids and related structural defects which can lead to thinning of the gate oxide as well as other intermediate structures.[2,221] Gate oxide integrity in the sub-2 nm oxide regime, however, appears to be independent of COPs and metals,[2] limited instead by intrinsic phenomena such as direct tunneling. The identification of vacancy agglomerates with crystal originated pits (COPs)[222-224] and flow pattern defects (FPDs),[225] including their modeling with the crystal growth parameters, was also clarified.[226-228] Post-growth annealing in a hydrogen ambient has been utilized for annihilating near-surface COPs.[229]

Concurrently, Kolbesen and Dolf de Kock and their colleagues[230-232] showed that swirl defects in FZ silicon (referred to as *A* defects) were interstitial, rather than vacancy, dislocation loops and by their detailed models, showed that the *A* and *B* defects were significantly different. That is, the *A* defects were generated by the agglomeration of silicon interstitials. The *B* defects were suggested to originate from dropletlike agglomerates of interstitials by Jun-ichi Chikawa and colleagues,[233] which subsequently converted to the *A* defects, although the experimental conditions by which the former deduction was made does not appear to truly correspond to the conditions present during conventional crystal growth. Nevertheless, these studies on FZ material yielded a generic insight into crystal growth which became helpful for subsequent CZ studies.[232]

Tuning the crystal-growth parameters was subsequently shown by Vladimir Voronkov (Fig. 17) and later, in conjunction with Robert Falster (Fig. 17), to result in either vacancy-rich or interstitial-rich material.[234,235] Interstitial-rich material was found to result in an improvement in GOI, although the IC DRAM yield might decrease because the IC processes utilized had been previously developed for the case of vacancy-rich material. Changing from a vacancy-rich material to an interstitial-rich material (for the same IC process flow) presumably facilitates the formation of interstitial dislocation loops which, if localized in the device SCR and decorated by metallic impurities, could lead to leakage currents degrading IC performance and yield. Re-engineering the IC process flow to be consistent with interstitial-rich material may, therefore, be required. The crystal-growth conditions resulting in either a vacancy-rich or interstitial-rich material was found to be dependent on the ratio of V/G, as a function of radius, where V is the microscopic crystal growth rate and G is the thermal gradient at the crystal-melt interface. The parameter, V/G, was shown by Voronkov and Falster to be the deciding factor as to which of the two species survives and, hence, which type of defects develop in the growing crystals. When $(V/G)(r)$ is greater than a critical number (1.34×10^{-3} cm^2/min-K or 2.2×10^{-3} cm^2/min-K), vacancy-rich material results, as described by von Ammon and Eric Dornberger[132] or Hourai[236] and their colleagues, respectively. When $(V/G)(r)$ is less than the critical number, interstitial-rich material predominates.[132,236] Numerous discussions about the validity and relevance of the V/G parameter have been going on for more than a decade and Voronkov's approach is widely accepted. As most of the thermophysical properties cannot be directly measured, however, an extensive effort has been carried out to theoretically calculate V/G, including the detailed mechanism wherein the interstitial or vacancy are dominate as well as the formation of the OSF ring by Abe Brown, Talid Sinno, Ryuichi Habu, and their colleagues.[217,237,238] Nevertheless, Abe and co-workers have noted the greater importance of the thermal gradient, per se, at the crystal-melt interface in controlling the formation of point defects.[215-218,239] In any case, the controlled use of after-heater heat shields and the growth of a crystal such that small numbers of both vacancies and interstitials are formed (*i.e.*, no point-defect clusters) is preferred.

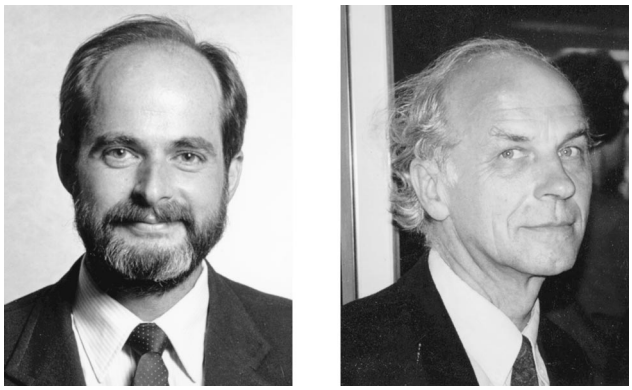

Figure 18. (*left*) Carlton Osburn and (*right*) Wolfgang Schroter.

Gettering

Background.—The diffusion, solubility and electrical activity of transition metals in silicon have been extensively studied by Eicke

Figure 19. (*left*) Ulrich Gosele and (*right*) Werner Kern.

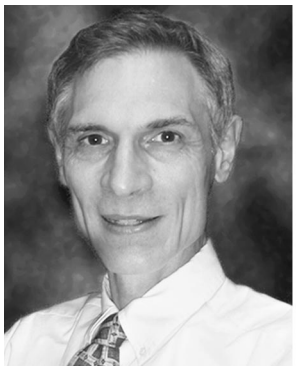


Figure 20. (*left*) Alain Diebold and (*right*) Linus Kimerling.


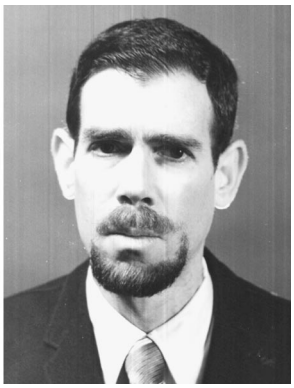

Figure 21. (*left*) George Rozgonyi and (*right*) Ted Kamins.

Weber (Fig. 16) and Wolfgang Schroter (Fig. 18) in conjunction with their colleagues as well as other personnel[1,240–242] and summarized by Klaus Graff.[243] Weber has shown that the transition metals exhibit a significant component of interstitial, as well as substitutional, solubility. This results in a mixed diffusion process, with a fast interstitial component (interstitialcy mechanism). The 3d transition metals in silicon, such as iron and chromium, mainly occupy and diffuse via interstitial sites. They accordingly exhibit high diffusion coefficients. Cobalt, Ni and Cu diffuse sufficiently fast, even at 27°C, that these elements could transfer from interstitial sites onto dislocation sites as well as form clusters and precipitates. Copper, however, appears to be especially different form the other 3d metals and has recently been extensively reviewed by Andrie Istratov and Weber, illustrating the insights and changes in our interpretation of experimental Cu diffusion data in silicon over the last 40 years.[244] The 4d metals in silicon, such as Ag, exhibit a solubility on substitutional sites comparable to interstitial sites and, therefore, diffuse slower than the 3d transition metals. The 5d metals, such as Au and Pt, exhibit a dominant substitutional solubility. The 5d metals accordingly take a longer time until their substitutional solubility is attained throughout the wafer. Size effects and electronegativity differences between the impurity and host atoms are also important considerations and must be taken into account for these analyses.

The U-shaped diffusion profile of Au is clear evidence that there is also a fast diffusing interstitial component for the 5d metals. The lack of sufficient interstitial sinks in dislocation-free silicon limits the kinetics of the transfer of Au onto substitutional sites during diffusion. The diffusion mechanism in the case where the impurity occupies both substitutional and interstitial sites was expanded from the conventional mechanism wherein an interstitial metal combines with a vacancy to occupy a substitutional site. The new phenomenon, identified by Urlich Gosele (Fig. 19), Electronics Division 1999 Awardee, involved an interstitial metal transferring to a substitutional site ejecting an interstitial and was referred to as the kick-out mechanism, (alternatively a substitutional metal combining with an interstitial wherein the metal transferred to the interstitial site).[245-247] Masaki Aoki showed that the temperature dependent solubility and diffusion coefficients, taking into account kinetic considerations during cool-down from high-temperatures, are especially important for transition metals such as Fe in determining their state of aggregation at room temperature.[248]

Metallic impurities can affect the electrical resistivity of silicon if they are present in sufficiently large enough concentrations (much greater than 10^{15} cm^{-3}). Metal concentrations in the range of 10^{15} cm^{-3} are generally more important as traps or g-r centers which can degrade a number of device characteristics. Today's high-quality materials, however, typically have much less than 10^{12} metals/cm^3, corresponding to significantly less than 10^{10} metals/cm^2. Process-induced contamination is often the most catastrophic effects of metals due to their localized coagulation or precipitation on defects in p-n junctions or near the channel region of an MOS device. Metallic contamination by Fe, Ni, and Cu in the range of 10^{11}-10^{13} cm^{-2} during ion implantation, dry etching, and plasma CVD, however, is much more likely to be introduced onto the wafer surface during wafer preparation or IC fabrication. In fact, Paul Schmidt and colleagues showed that metallic impurities can also be introduced from the furnace per se, let alone from handling tools, carriers or other tools utilized in the fab.[249,250] The surface metals diffuse into the wafer bulk during subsequent thermal processing. The surface and bulk descriptions, however, cannot be simply transposed via the wafer thickness due to surface-bulk segregation during thermal processing. In view of this, prescriptions are under development to determine the equivalent bulk concentration from a given surface concentration during subsequent thermal processing. The comprehension of the redistribution of metals between the surface and bulk during IC thermal processing, dependent on surface cleaning methodologies (and sequence) is receiving increased attention and is especially important as regards lifetime doping. An extensive methodology has developed for the removal of surface metals by cleaning with suitable solutions of high-purity chemicals, as initiated by Werner Kern (Fig. 19), ECS Fellow and DS&T Callinan Awardee in 1972, editor of the influential *Handbook of Semiconductor Wafer Cleaning Technology* and colleagues.[251]

A set of defect/diagnostic tools for IC process evaluation has been reviewed by George Rozgonyi.[252] The sensitivity of surface diagnostic techniques varies, but was shown by Alain Diebold (Fig. 20) co-editor of the NIST series on *Characterization and Metrology for ULSI Technology*, and colleagues to range between 10^8-10^{10} atoms/cm^2 for a variety of transition metals.[253] A useful qualitative diagnostic technique to detect the presence of metals was accomplished by monitoring the etch pit density with a large concentration of metals present, conventionally referred to as "haze" as measured by the modulated optical reflectance technique.[254] More quantitative results were available by monitoring the carrier lifetime and electron beam induced current (EBIC). The utilization of EBIC techniques have been useful to assess the electrical activity of the various microstructures fabricated to getter the undesired impurities from the device active regions. Junction capacitance and current spectroscopy procedures were pioneered by Chin-Tang Sah and colleagues.[255] Deep level transient spectroscopy (DLTS) has become a major diagnostic tool to identify the energy levels of the various chemical impurity-structural defects in the majority-carrier half of the band-gap. L. C. Kimerling (Fig. 20), Electronics Division 1995 Awardee, in conjunction with Janet Benton and colleagues have developed an extensive catalog of the energy levels for these structural-chemical complexes in the band-gap of silicon, based on the DLTS methodology.[256,257]

Gettering provides a mechanism to remove these process-induced metal contaminants from active device regions and has been correlated with significantly improved device performance, yield and reliability.[1,121] These benefits were observed in p-n junction device characteristics,[258-260] oxide leakage current and breakdown,[261]

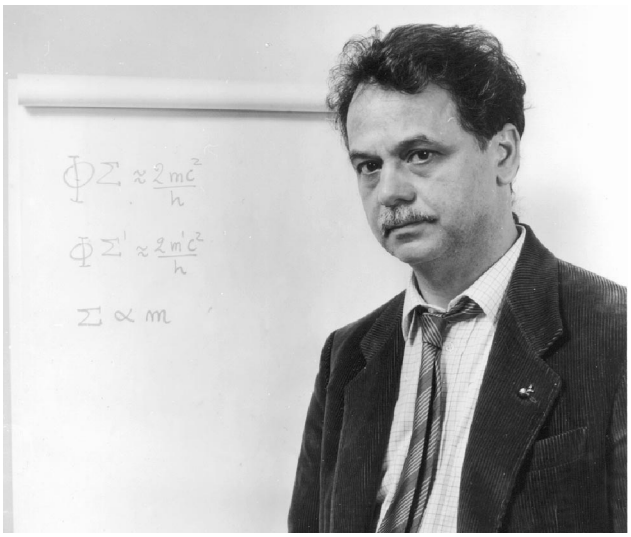

Figure 22. Geofranco Cerofolini.

Figure 23. (*left*) Cor Claeys and (*right*) Arnold Reisman.

MOS dynamic shift register leakage currents,[262] DRAM refresh time,[201,263] CCD imager performance by collecting minority-carriers generated by spurious radiation,[175] and CMOS circuit latch-up control by reducing the recombination lifetime in the substrate.[264] Significant improvements were observed in MOS,[263] bipolar,[265] and CPU ICs.[266] Louis Parrillo and colleagues described the integrated benefits of detailed attention to optimizing the response of the silicon wafer to the full suite of bipolar fabrication processes beyond just gettering, in order to achieve optimal IC performance.[265] The development of quantitative gettering processes and models were especially helpful in gaining insight into the microphysical processes operative, although continuous refinements ensure this will remain an active area of research.[260,267-272] Thomas Seidel, in particular, clearly enunciated at the 156th, Fall, 1979 meeting of The Electrochemical Society's Gettering symposium, that successful gettering required three essential steps. These steps were the release of metals from their deleterious location within/near the device, transport of the metals to the gettering arena (sufficiently removed from the spatial location or SCRs of the device), and capture at the gettering site (and retention during subsequent thermal processing.)

Gettering methodologies (nonoxygen techniques).—The concept of gettering to improve device performance was actually implemented before the point-defect dilemma had been appreciated. The first reported instance of gettering in silicon appears to have been the utilization of a phosphosilicate glass (PSG) film on the back surface of the silicon wafer by Goetzberger and Shockley to getter unwanted metallic impurities away from the device active regions.[115] It was initially believed that the metallic impurities were incorporated within the PSG liquidus. PSG gettering was also utilized in n-channel DRAMs as a result of the introduction of P via $POCl_3$ doping of the n-type source/drain. It subsequently became clear that gettering was due to metallic incorporation within the misfit dislocation array, formed as a result of the size difference between the tetrahedrally covalently bonded silicon (0.118 nm) and phosphorus (0.11 nm) atoms, when the diffused sheet resistance was less than about 4 ohms/sq.

A variety of back-surface external gettering (EG) mechanisms were subsequently examined,[1] especially mechanical damage,[137] ion-implant damage,[273,274] polysilicon gettering[275] and, to a much lesser extent, nitride gettering[276] and laser damage gettering.[277] The pre-eminent technique, however, was a variety of methodologies collectively referred to as preoxidation gettering procedures (POGO), as exemplified by George Rozgonyi (Fig. 21), ECS Fellow and Electronics Division 1981 Awardee and colleagues.[278] External gettering was found to be useful, although some procedures were often limited temporally due to the decrease in gettering efficiency with continued thermal processing. With the reduction of IC fabrication temperatures to below 1000°C, for example, PSG gettering becomes less effective and various combinations of the above methodologies were extensively examined.[1,279] The mechanical damage procedure, however, was prone to particulates flaking off the back-surface. Back-surface polysilicon gettering was a cleaner procedure and was extensively utilized. Theodore Kamins (Fig. 21), ECS Fellow and Electronics Division 1989 Awardee has discussed the properties and utilization of polysilicon, especially for the gate electrode for CMOS devices and the interpoly electrode for EEPROM devices.[280,281]

Interestingly, $POCl_3$ diffused samples at 1100°C and annealed at 800°C was noted by Geofranco Cerofolini (Fig. 22), co-author with Laura Meda of the textbook *Physical Chemistry of, in and on Silicon*, and colleagues to contain a complex network of dislocations and defects which were deduced to be effective gettering sites for reducing the leakage current in diodes.[282] Subsequent research reduced the $POCl_3$ deposition temperature to 920°C, followed by the 800°C anneal which resulted in a significant increase of the MOS capacitor storage time, of significance for DRAMs.[283] Concurrently and independently, Huff and co-workers used $POCl_3$ deposition at 1100°C or 1000°C, followed by a low-temperature N_2 anneal in the range of 700°C, for the fabrication of 16K DRAMs and also observed significantly improved MOS capacitor storage times.[284] The TEM micrographs of James Mayer and colleagues were a very useful visual representation of the bulk microgettering network,[285] consistent with the hypothesis current at the time.

Back-surface ion-implantation gettering was also successfully introduced, although as noted above, the healing of the ion-implant damage during subsequent thermal processing required multiple implants during IC fabrication. It was observed by Henry Geipel and Warren Tice that since the dislocation energy is proportional to the square of the Burger's vector, perfect dislocations were more effective than partial dislocations for gettering;[286] moreover, the heavier the ion, the more effective the gettering. Accordingly, argon damaged layers were particularly favored for this technique and found to be more effective for gettering Au than $POCl_3$ gettering at temperatures below 1000°C, whereas the two mechanisms were about equal from 1000°C to about 1150°C.[279] Experimental data by Yasuteru Ichida and colleagues indicated that a $POCl_3$ gettering process introduced at the end of device fabrication was more successful than its introduction at the initiation of the device fabrication process.[287]

HCl gettering, introduced by Rudolf Kreigler, DS&T Callinan 1979 Awardee, became an especially prevalent industrial technique.[288-290] Carlton Osburn (Fig. 18), ECS President in 2000-2001, ECS Fellow and recipient of the Electronics Division Award in 1991 and the DS&T Callinan Award in 1975, has studied the improvement of GOI via Cl methodologies[291] as part of an extensive

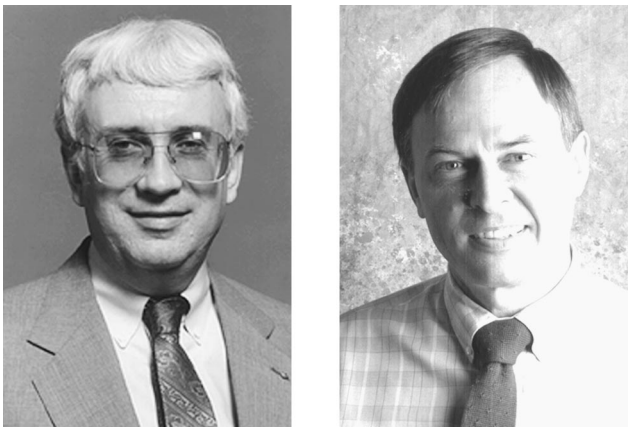

Figure 24. (*left*) Richard Fair and (*right*) Jim Plummer.

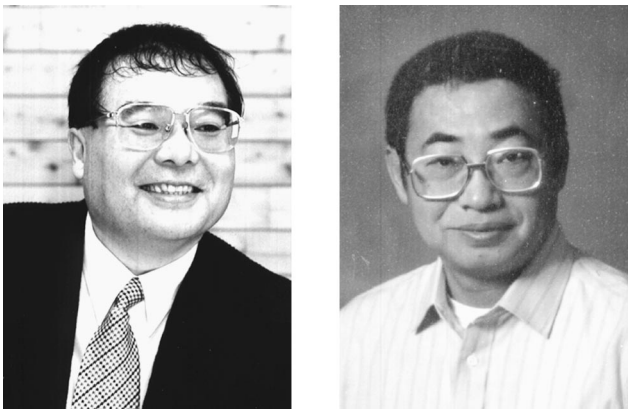

Figure 25. (*left*) Fumio Shimura and (*right*) Teh Tan.

series of analyses in the VLSI Laboratory of Arnold Reisman (Fig. 23), 2001 Solid-State Science and Technology Awardee, ECS Fellow and Electronics Division 1984 Awardee.[292] 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 OISF at sufficiently high temperature by Hiromitsu Shiraki, Cor Claeys (Fig. 23), coeditor of the *ECS High Purity Silicon*, *ULSI Process Integration*, and *Crystalline Defects and Contamination* proceedings and their colleagues,[293,294] thereby removing potential sites of metallic precipitation adjacent to the device. 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.[295] It was also suggested by Kreigler that Cl was most useful for cleaning the furnace quartz tube rather than actively incorporated into the oxide due to reliability concerns.[289,290]

A large solubility enhancement (about 10^3 times the intrinsic solubility) was observed by Meek and Seidel in heavily doped n-type silicon due to substitutional metal-substitutional donor pairing.[296] The pairing of interstitial metal donors with substitutional acceptors in heavily doped p-type silicon, however, resulted in an increased metal solubility only one to two orders of magnitude compared to the intrinsic solubility because of increased configurational and vibrational entropy effects associated with the interstitial metal-acceptor pair.[296] The description of this enhancement in solubility has also been described in terms of a Fermi energy effect by Aoki, Scott McHugo, and their colleagues.[297,298] Heavily doped isolation regions also gettered metals, due to both the diffusion induced damage as well as by the solubility enhancement of metals in heavily doped regions.

The beneficial effects of implantation gettering on both the generation and recombination lifetimes, over a limited annealing temperature range, was also observed. The efficacy of back-surface abrasion damage was often found to decrease with continued thermal processing, but the technique was useful in removing swirl defects and the monitoring of various gettering mechanisms led to an extensive literature. In particular, P introduced by $POCl_3$, and argon ion-implantation as well as oxygen precipitation were suggested to act as sources of silicon interstitials which mediated the transport of metals and facilitated gettering, as discussed by Jim Plummer (Fig. 24), ECS 1991 Solid State Science and Technology Medalist and colleagues in their extensive point-defect studies.[299] In that regard, Hu's research in the 1960s and 1970s was instrumental in establishing the dual mechanism of both vacancy- and interstitialcy-mediated diffusion of group III and V elements in silicon (although there is a dominant point-defect species, dependent on the size of the diffusing element)[300] and the interactions in sequential diffusion processes.[301] These point-defect studies of Hu have been summarized in the 1990s.[302,303] Gettering mechanisms which inject interstitials facilitated the transfer of 5d metals which predominately diffuse by an interstitial mechanism, from low mobility, high solubility substitutional (and electrically active) sites to high mobility, low solubility interstitial sites via the transport of metals to IG or EG captive sites away from the device active regions. The role of silicon interstitials in the generation of dislocation loop gettering complexes in (low oxygen content) magnetic CZ (MCZ) and FZ silicon was also noted as well as the role of interstitials in Si by Kimerling in the promotion of a host of solid-state defect and gettering reactions.[257,304] The chemical state of the transition metals was also noted by Kimerling to possibly be playing a role influencing their participation in various gettering schemes.[305] The role of vacancies, however, also continues to be acknowledged as playing some role in the overall diffusion and gettering schemes. The modeling of the emitter dip effect and the emitter retardation in relation to the vacancy charge states has been discussed by Richard Fair (Fig. 24) and his colleagues's ion-implantation and diffusion studies.[306,307] Scott Dunham has also explored the role of vacancies in transient enhanced diffusion[308] as have others.[1]

A number of other gettering techniques have been described although they have not yet been implemented into IC fabrication. These include porous gettering, cavity gettering, ultrasound gettering, electric field induced dissociation of Fe-B pairs to enhance gettering of Fe, gettering by enhanced wafer bending, and external gettering by an Al-SiO_2 film deposited on the back-surface and processed above the Al melting temperature.[1] There is also continued interest in replacing the highly doped substrate of an epitaxial wafer with an implanted ground plane (in a polished wafer) formed by an unmasked, high dose, MeV implant, which also has the potential for gettering.[309,310] Finally, it should be noted that gettering was not quite the panacea for every application.[1]

Gettering methodologies (internal oxygen gettering).—The predominant method of gettering became oxygen internal gettering (IG) with the implementation of a defect-free denuded zone (DZ) near the surface and precipitation dislocation complexes (PDC) in the bulk, as promulgated by Teh Tan (Fig. 25) and colleagues.[258] Both high performance and yield (minimal refresh loss) was achieved in MOS DRAMs by Huff and colleagues[263] by fabricating the DRAM in a sufficiently deep DZ with high generation lifetime material[1,121] in conjunction with a small bulk recombination lifetime. The DZ should be sufficiently small, however, to ensure effective gettering by the bulk gettering sites. In addition to the improved device characteristics described in the Gettering Section, the beneficial effect of IG was also reported to control epitaxial stacking faults,[311,312] reduce saucer pit formation,[313,314] retard slip formation[315] and improve MOS carrier generation lifetime.[263,316,317] Initial observations

also indicated both higher DRAM performance and yield from the lower sections of the CZ crystal.[318] A comprehensive discussion of oxygen in silicon was presented in *Oxygen in Silicon*,[149] edited by Fumio Shimura (Fig. 25) and in the host of studies collectively authored with his colleague Hideki Tsuya (Fig. 34).[319-322]

Successful IG required understanding the mechanism of oxygen incorporation and controlling its magnitude and spatial homogeneity during CZ crystal growth. Indeed, OISF often replicated the microdefect patterns in the grown crystal as shown by Abe and de Kock.[172,232] PDCs formed during IG can exhibit either a positive or a negative influence on IC electronic characteristics. An excessive number of PDCs, in conjunction with too shallow a DZ, significantly degrades the IC performance. If the PDCs are located sufficiently far away from the surface (typically $>30\ \mu m$ below the active device area), they can provide sites, due to the localized crystal disruption, for gettering metallic impurities inadvertently introduced during IC fabrication. Accordingly, process engineers aggressively pursued identification of the optimal oxygen concentration (and degree of super-saturation),[1,121,172] the role of dopant type and concentration as regards the influence of the point-defect density in promoting SiO_x precipitation,[322] the associated IC thermal processes (temperature, ambient and sequence), the usefulness and mechanisms of high-temperature annealing and out-diffusion (including the role of hydrogen, argon or nitrogen annealing),[229] and nucleation and growth procedures (including the role of ramp-up (and to a lesser extent cool down) techniques to process temperature) to make the gettering process more robust, as described by Seigo Kishino and colleagues.[1,267] In particular, nucleation and growth processes were mainly heterogeneous processes, including carbon[175] (and perhaps other impurities) as participants in providing thermodynamic pathways in addition to that provided by interstitials.[149,323] In that regard, the slower oxygen precipitation rate in n/n^+ material (occasionally attributed to the difficulty of incorporating oxygen in highly doped antimony crystals), (although doping species and related electrical effects have also been proposed) is especially enhanced by ramping procedures.

Quantitative modeling of the IG gettering efficiency with the density, morphology, and distribution of bulk microdefect structures (including their dependence with the concentration of oxygen and the role of the nuclei size distribution for condensed matter aggregation phenomena) was extensively studied by Fujimori.[324] It was also noted by William Patrick and colleagues that whereas devices processed from wafers near the seed end of a crystal eventually lost its gettering efficiency, devices from wafers from the tail end of the crystal (with a lower initial oxygen content) initially exhibited higher leakage currents, but with extended thermal processing, generated sufficient SiO_x precipitates to eventually exhibit lower leakage currents than wafers from the seed end of the crystal.[325] Internal gettering, however, does not perform as an infinite sink inasmuch as gettering will stop when the metal concentration in the DZ approaches the thermal equilibrium solubility,[248] consistent with the process kinetics, although this potential limitation continues to be explored.

The relationship of the various thermal processes associated with IC fabrication, in conjunction with the thermal donor anneal, historically performed at ~625°C for 30 min or so by the silicon wafer supplier (as well as, recently, via an RTA), led to an escalating number of publications postulating a host of solid-state reactions. A particular popular IC thermal process sequence became known as the hi-lo-mod hi process, in view of its compatibility with IG gettering. The initial hi temperature process facilitated the out-diffusion of oxygen and was characterized in terms of $(\sqrt{Dt}$, where D is the diffusion coefficient of oxygen and t is the time of the process). The low-temperature step, about 600-700°C, nucleated the bulk IG centers while the subsequent high-temperature anneal facilitated the growth of these bulk nuclei. At this point, the DZ could be readily distinguished by an appropriate etching or electrical technique, although the growth of these bulk nuclei continued to some extent with further IC thermal processing. The second high-temperature anneal was generally less than the initial high-temperature anneal so as not to dissolve those nuclei which were thermodynamically stable at the higher temperature. Of course, one may imagine many scenarios for the stabilization of the bulk gettering nuclei and IG gettering. John Borland has described an extensive number of macroscopic processes to enhance the IG methodology[326] while Edward Wijaranakula and John Matlock have described a number of the microscopic processes associated with the IG methodology.[327]

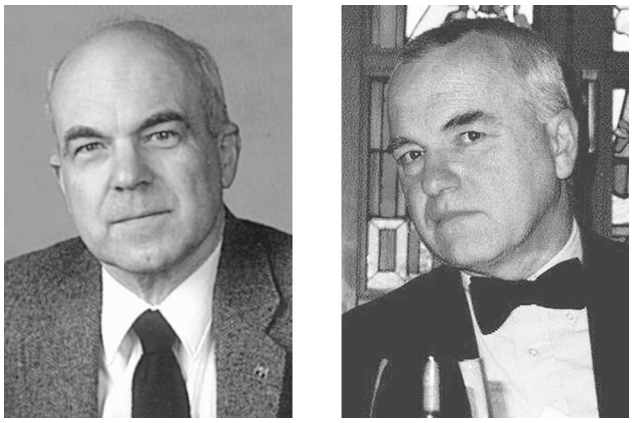

Figure 26. (*left*) W. Murray Bullis and (*right*) Laszlo Fabry.

The role of the ambient during the IG out-diffusion procedure was shown by Hu to be an especially important parameter.[328] For example, argon and nitrogen ambients facilitate the out-diffusion of oxygen from the bulk of the wafer; if the oxygen content of those ambients is sufficiently low (*i.e.*, low fugacity). In this case, the surface oxygen content is pinned to a value less than that obtained during an oxygen anneal. This methodology has also been helpful in improving the surface quality, reducing OISF and improving GOI. A hydrogen ambient has been noted to be useful in promoting the out-diffusion of oxygen as well as facilitating the surface phase reaction responsible for both perpendicular configurations of fingers often observed on epitaxial silicon surfaces along with improved surface quality.[1]

Epitaxial Material

Epitaxial material such as p/p^+ suppresses latch-up, and improves GOI and DRAM refresh performance compared to polished wafers[1,121,172,329-331] and, accordingly, has often been the material of choice for logic ICs such as ASICs, microprocessors, and EEPROMS. The onset of trench isolation, however, has reduced somewhat the concern over latch-up. The increased cost of epitaxial materials has been more than offset by the performance (and price) of these leading edge ICs. Although DRAMS were initially fabricated in polished wafers, epitaxial material became the material of choice for DRAMs during the first half of the 1980s, at least for U.S. manufacturers. The perceived benefits were the readily available latch-up control (U.S. manufacturers utilized the same design rules for both the n-MOS memory and the CMOS periphery circuitry) and the improved benefits of a heavily doped substrate in reducing the diffusion current (see Carrier Lifetime section). Japanese DRAM manufacturers, however, opted for polished wafers, due to the lesser performance requirements of DRAMs compared to logic ICs, and the attendant commodity pricing strategies. In this case, latch-up control was probably achieved by larger design rules for the periphery CMOS circuitry as well as, perhaps, by utilizing selectively diffused or implanted guard rings. U.S. DRAM manufacturers subsequently returned to polished wafers in the mid 1980s due to the severe DRAM pricing strategies. The beneficial reduction in the price ratio of epitaxial material compared to polished wafers, may be anticipated to continue its decline for 300 mm wafers, although

the utilization (and cost) of MCZ substrate wafers may be increasing faster than the decreasing cost of the epitaxial wafer deposition process, per se.

The improved GOI in epitaxial material, initially considered to result from its reduced surface microroughness specified over a defined spatial frequency range,[178,179] is now believed to result from, at least partially, the improved structural perfection of the epitaxial material, as compared to residual polishing damage in polished wafers. The utilization of p/p^- epitaxial material for DRAMs is also receiving attention; the benefit of improved surface quality and reduced system capacitance may offset the lack of Fermi-energy enhanced solubility gettering of Fe observed in the p^+ substrate of the conventional p/p^+ epitaxial system.[297,298] If significant IC benefits are perceived and implemented, the IC industry will successfully cope with the absence of solubility enhanced gettering. The oxygen content, however, may have to be re-assessed in the case of the p/p^- system since the influence of IG has been considered to be superior in p^+, compared to p^-, material.[332,333] Clarification of the enhanced oxygen precipitation in p^+, as compared to p^-, substrates has been related to a larger total quantity of precipitated structures in the former case,[332] whereas the strain around the precipitated complexes has been shown to be less in p^+, compared to p^-, material.[332]

The role of MeV implantation and associated annealing procedures in polished wafers as a replacement for epitaxial structures continues to receive attention, although the control of COPS and related defects will be aggravated in polished wafers. Indeed, the use of MeV implant/anneal procedures, in some cases in conjunction with epitaxial wafers, has been proposed to achieve enhanced device architectures,[309,310] although the implementation of this procedure has been somewhat slow to date.

Carrier Lifetime

Gold has been an especially critical element for controlling the carrier lifetime in silicon, as summarized by W. Murray Bullis (Fig. 26), ECS Treasurer, 1994-1997 and co-editor of the ECS *Defects in Silicon* proceedings.[334] Structural irregularities like OISF were also identified as significantly degrading carrier-lifetime.[335] Although this reference relates to GaAs, the result is generically applicable to Si, albeit with greater sensitivity to the degree of metallic decoration of the partial dislocations associated with the stacking fault, see Ref. 122. Alvin Goodman implemented a series of studies on the constant magnitude steady-state surface photovoltage (SPV) method for silicon,[336] determining the bulk recombination lifetime via the measurement of the minority carrier diffusion length and utilization of the Einstein relation. A variant of this method in which the surface photovoltage varies and the impinging photon flux is fixed has also been described.[337] The pulsed MOS capacitor generation lifetime technique, introduced by Manfred Zerbst,[338] is widely used because the MOS capacitor is utilized on most test structures. The relation between the recombination and generation lifetimes and related diagnostic techniques has been extensively studied by Dieter Schroder.[337,339,340] In a somewhat complementary fashion, Bullis and Huff have assembled the more common carrier lifetime techniques, their benefits and limitations on their interpretation as well as a proposed unified taxonomy.[341] Finally, the extensive studies of Claeys and colleagues of the influence of the perimeter to area ratio in determining the various lifetimes in scaled devices is of particular significance.[342]

Appropriate gettering procedures were implemented to both retain and improve the carrier lifetime. The carrier recombination lifetime is a particularly sensitive technique to assess the DZ quality. The recombination lifetime has also been extensively utilized to assess the interstitial Fe_i content by monitoring the Fe_i-B neutral pair species versus the significantly more electrically active Fe_i content by Gunther Zoth and Werner Bergholz[343] and extensively implemented by Jastrezbski and colleagues.[344] Both thermal and, more recently, optical techniques have been utilized to destroy the Fe_i-B pair.[343,344]

The carrier-lifetime for DRAMs requires consideration of two apparently contradictory requirements.[330] In one case, a high (generation) lifetime is desired to facilitate the retention of charge for an extended period of time (*i.e.*, high DRAM circuit refresh time). On the other hand, high (generation) lifetime material is especially susceptible to discharge of the DRAM storage node and, thereby, loss of information content due to transient electronic phenomena. A compromise is required for optimal DRAM performance. Internal gettering, in conjunction with the formation of the DZ, can reduce the influence of the transient phenomena while increasing DRAM performance and yield. This is accomplished by fabricating the DRAM and its voltage-dependent SCR in an appropriately deep DZ of sufficiently high (generation) lifetime material while, concurrently, maintaining the DZ sufficiently shallow to ensure effective gettering by the bulk gettering sites (and collection of carriers due to the low recombination lifetime). The lifetime in the DZ, called the generation lifetime, τ_g, and sampled within the voltage-dependent SCR of the IC, can be rather high due to the IG/EG in the silicon wafer. Concurrently, the recombination lifetime, τ_r, in the bulk of the wafer, sampled over the minority-carrier diffusion length, can be readily modified (decreased) via the nature of the gettering procedure utilized.

Because ICs typically operate in the temperature range of 75°C, the contribution of the diffusion current from the bulk becomes an important factor due to the exponential increase in n_i^2 with increasing temperature. A significantly reduced bulk recombination lifetime, τ_r, due to IG, for example, would increase the diffusion current; increased substrate doping, such as in a p^+ substrate, however, substantially decreases the diffusion current as discussed by Pallab Chatterjee and colleagues[329] as well as Hideki Iwai and colleagues.[330,331] The utilization of a lightly doped epitaxial layer on a heavily doped substrate is a particularly useful material configuration for both reducing the bulk diffusion current and achieving a high generation lifetime in the vicinity of the surface.[329-331,345] It has been noted that the benefits of IG as measured at room temperature, however, may be somewhat negated at device operating temperature due to the g-r current generated by the IG centers.[259]

Wafer Design

The incorporation of these strands into developing the optimal wafer was generically described as wafer design or defect engineering,[121] the latter exemplified by Rozgonyi's famous *wheel of misfortune*[346] and extended by Laszlo Fabry and colleagues (Fig. 26) as the *gear of challenge*.[347] The recognition of the silicon wafer as an early version of the finished IC product, rather than a raw commodity material, has become more fully appreciated throughout the IC industry. This insight is used to guide the design of a circuit-focused, multizoned silicon wafer using a modular systems approach to improve the consistency and performance of IC products.[121] By engineering the wafer specifications to achieve the variously configured multi-zone characteristics of silicon for the particular IC application, product performance, yield and reliability can be enhanced. While there may be no unique set of silicon wafer specifications applicable in all cases because of variations in both the IC fab processes and IC applications, an important cost-of-ownership (CoO) goal is to develop a sufficiently robust wafer and IC process in order to maintain cost control.

The multizone silicon wafer has been described and implemented by Howard Huff, ECS Fellow and Electronics Division 1996 Awardee in conjunction with Lawrence,[101,121] and independently by Shin Takasu, and Hans Richter and colleagues.[348-350] The structure consists of the circuit active zone (a), characterized by the requisite wafer surface flatness, metallics and localized light scatterers (LLS). This zone may be either the polished wafer surface or an epitaxial layer. The defect-free DZ region (b), is developed during the high-temperature thermal processes that accompanies the formation of the IG zone, region (c). Zone (c) getters the process-induced metallic contaminants (often responsible, in conjunction with structural de-

Figure 27. Harry Gatos (standing) and ''his orchestra.''

fects, for uncontrolled leakage currents) away from the circuit active zone during IC fabrication. The driving force for gettering is the reduction in the Gibbs free energy of the material system. The IG zone, however, generally exhibits an incubation time until it has sufficiently developed to form the getter complexes during IC fabrication. This incubation time is dependent on both the as-grown wafer characteristics and the detailed IC thermal process conditions, in conjunction with silicon interstitials, carbon and perhaps, dopants, resulting in the formation of IG centers. Recently, the formation of IG sites in the bulk gettering zone has been facilitated via engineering the vacancy concentration to enhance the clustering of oxygen, rather than focusing on the oxygen content per se (and the associated hi-lo-mod hi IG process) by Falster and colleagues.[351] The EG zone, region (d), is also a critical factor in the gettering process. This zone is generally active at the beginning of the IC fabrication process, whereas the conventionally formed IG zone, as noted above, requires a sequential set of process steps for its development. The back surface of the wafer, however, is increasingly being polished to reduce particle contamination, improve wafer flatness, and increase wafer strength. The polished back surface more readily exhibits microscopic contamination and wafer handling damage. As a result, back-surface cleanliness requirements may emerge and drive the need for more stringent robotic handler standards. Many EG techniques also degrade the quality of the polished back surface and the front-surface site flatness due to nonuniformities in the deposited back-surface gettering film and, consequently, may no longer be allowable.[22]

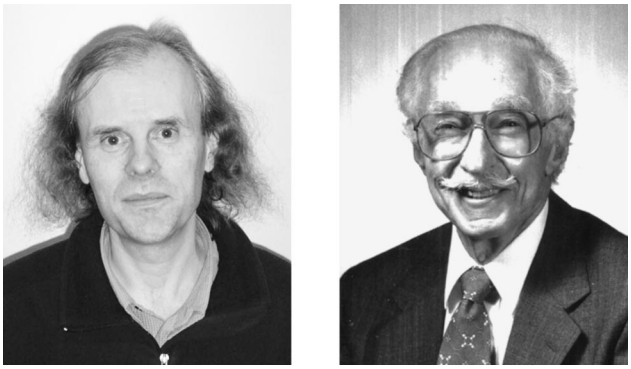

Figure 28. (*left*) Marc Heyns and (*right*) Bill Pliskin.

Figure 29. (*left*) Jerzy Ruzyllo and (*right*) Al Tasch.

Wafers with an epitaxial layer often incorporated an oxide seal zone (e), historically utilized in conjunction with external polysilicon (EG) gettering. The oxide back-seal has been utilized to prevent the transfer of silicon from the wafer to the susceptor in the case of front-surface radiant-heated epitaxial reactors and to minimize the autodoping phenomenon (substrate dopant transfer to the epitaxial layer of adjacent wafers by a complicated and variable time-dependent gaseous species transfer) as well as, presumably, the transfer of dopant to adjacent wafers during IC processing.[352] In some cases, an oxide-nitride seal zone or other composite films are preferred rather than the polysilicon-oxide configuration. The utilization of single wafer epitaxial reactors with both front- and appropriate back-surface heating lamps, however, has negated the silicon transfer effect while the autodoping effect during IC fabrication has been suggested, in retrospect, to not be that significant. In those cases where the back-surface oxide is still utilized during epitaxial deposition, the oxide is often subsequently removed per the IC customer's request. The oxide removal also significantly reduces the wafer bow. Accordingly, the back-surface oxide is no longer a recommended procedure for 300 mm p/p^+ epitaxial wafers; indeed CoO considerations has been noted by Goodall and Huff to become especially important with the transition to the 300 mm wafer diam.[22,353-355] Permutation of the multizone concept for variously configured zones or stacked layers leads to further specialized configurations.[101]

Surfaces and Wafer Cleaning

Although modern ICs are typically several cm^2 in spatial area and typically 0.75 mm thick (as fabricated), barely 1% by volume of the silicon material is required for the active device regions and interconnections. The surface structure and reactivity of silicon, including the importance of both wet and dry (plasma) cleaning methodologies, have therefore become of the utmost importance. Indeed, in recent years, the physical structure as well as the chemical nature of the wafer front surface has emerged as a critical concern. In the early days, preferential etchants were utilized for revealing defects such as dislocations in plastically deformed germanium as discussed by Gerald Pearson, 1981 ECS Solid State Science and Technology Awardee.[356] A wide variety of wafer cleaning recipes were also developed, as exemplified by William Pliskin (Fig. 27), ECS Fellow and Electronics Division 1972 Awardee,[357] Erhard Sirtl,[358] and summarized by Harry Gatos (Fig. 28) and Mary Lavine.[359] Cleaning the wafer surface was extended from these early studies to the current regime utilizing the SC1/SC2 methodology initiated by Kern[251] and subsequently expanded, taking into account both device performance and CoO implications. These studies included Marc Heyns (Fig. 27),[360-362] Tadahiro Ohmi,[363-365] and Jerzy Ruzyllo (Fig. 29) ECS Fellow,[366,367] and Richard Novak, co-editors of the ECS *Cleaning Technology in Semiconductor Device Manufacturing* proceedings and their colleagues. Takeshi Hattori has recently edited an

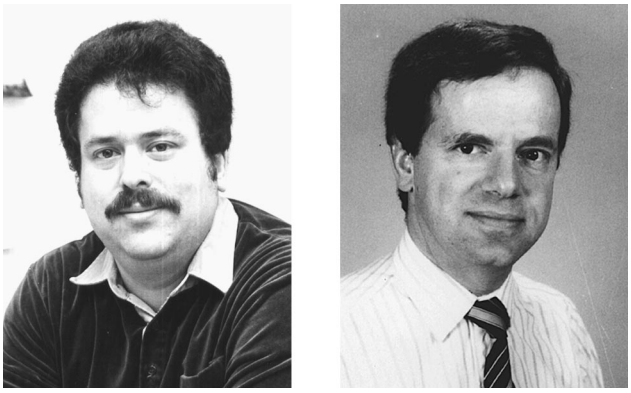

Figure 30. (*left*) Bernard Meyerson and (*right*) George Celler.

extensive monograph summarizing a number of cleaning technology and surface preparation processes and their relation to enhanced device performance.[368] The preparation of the silicon surface, cleaning, passivation, and surface morphology, has played a significant role in ensuring the development of improved device performance and yield.

Both polished and epitaxial wafers exhibit specific defects that must be controlled. Polished wafer defects include metal and organic surface chemical residues, particles, and grown-in microdefects, such as COPs. Epitaxial wafer defects include large structural defects ($>1\ \mu m$), (such as epitaxial stacking faults[192]) and small structural defects ($\leqslant 1\ \mu m$) as well as particles and residual chemical residues. Starting material requirements are expressed in terms of specific types of surface defects for different wafer types. The influence of the preoxidation clean on the SiO_2 growth rate, surface organics (apparently the carbon atom surface density) and its degrading effects on MOS GOI as well as the increased roughening of the silicon surface heated in a hydrogen ambient, compared to nearby surface regions covered with residual SiO_2 films, remain important arenas of research.[1] Indeed, the removal and prevention of surface defects is a current state-of-the-art challenge for silicon wafer technology with immense CoO implications. The development of laser scanning and other instrumentation to count, size, and determine the composition and morphology of these defects is a critical metrology challenge[369] and continues to be addressed by a variety of equipment and fab personnel.

IC Scaling

Gordon Moore's assessment of memory component growth in 1965, initially based on bipolar and then MOS memory density, was observed to quadruple every 2 years[20] (modified to ~3 years around the mid-later 1970s and currently taken as 3-4 years based on a 1995 assessment[21]), became enshrined as Moore's law. Indeed, Moore's law became the productivity criterion by which the IC industry grew at a 25% compound annual growth rate (CAGR). Device scaling has been the engine driving this revolution, with major contributions arising due to reductions in gate dielectric thickness, physical gate length and extension junction depth, as discussed by Robert Dennard and colleagues via constant electric-field scaling and, subsequently, constant voltage scaling.[370,371] These parameters were empirically related in 1980 by Simon Sze and colleagues;[372] Sze is the author of the influential textbook *Physics of Semiconductor Devices*.[373] For example, consider the scaling of the KDRAM from the early 1970s (4K DRAM) to today's leadership MPU part appropriate for the 100 nm technology generation in 2003.[22] The SiO_2 gate dielectric has proceeded from the range of 50-100 nm for the 4K DRAM to an anticipated value of about 1.2 nm oxide equivalent thickness (EOT) for the MPU part.[22] Likewise, the physical gate length has decreased from 7.5 μm for the 4K DRAM to a physical gate length of about 45 nm for the MPU part at the 100 nm technology generation.[22] Finally, the junction depth has decreased from several micrometers for the 4K DRAM to about 20 nm for the extension junction depth for the MPU part at the 100 nm technology generation.[22] In a related fashion, the critical figure of merit for transistor speed, CV/I, has become less than one psec for an NMOSFET and approaching one psec for a PMOSFET, as the physical gate length has decreased from 30 to 20 to 15 nm in leading edge research devices.[374]

Robert Dennard introduced the one-transistor memory cell in 1968,[375] thereby establishing the paradigm by which enhanced scaling has progressed. Scaling of the gate dielectric SiO_2 to the sub-2 nm regime, however, has exacerbated the occurrence of direct tunneling[376,377] as described by Yuan Taur and colleagues. An extensive global effort is in progress to identify an alternative, high-dielectric constant material to circumvent the gate dielectric direct tunneling leakage current in the case where the silicon oxynitride gate dielectric is less about 1.2 nm,[22,378-382] including the relevant diagnostic techniques as described by Diebold and colleagues.[19,383,384] Rajendra Singh, 1998 DS&T Callinan Awardee and co-editor of the ECS *Low and High Dielectric Constant Materials: Materials Science, Processing, and Reliability Issues* proceedings and the Materials Research Society's (MRS) March, 2002 issue of the MRS Bulletin, as well as other Symposia, continues the focus on alternative, high-k gate dielectric matierlas, gate electrodes (eventually requiring dual metal gates with differing work functions for CMOS optimization) and the issues of incorporating the gate stack into an integrated, conventional planar, initially poly electrode, IC process flow. Additionally, Al Tasch (Fig. 29), Electronics Division Awardee in 1997, has clarified the role of the quantum confinement effect in silicon in increasing the effective dielectric thickness of a MOSFET in inversion,[385] which cannot be avoided as compared to the poly-depletion effect in the polysilicon gate electrode, which can be negated by utilizing metal gate electrodes. In that regard, a dual metal system with differing work functions is under consideration as the gate electrodes for optimal CMOS performance.[22]

Concurrently, a host of studies are in progress to identify an ultrashallow junction fabrication methodology consonant with the sub-100 nm technology generations.[22,386] These studies may be grouped under the classification as classical CMOS structures. On the other hand, a plethora of nonclassical CMOS devices are under consideration wherein a unique combination of materials and/or structural configuration of the device may differ from the conventional or classical planar CMOS structure.[22] Of particular importance is the assessment of alternate channels for enhancement of the n- and p-channel mobility, ranging from strained silicon on unstrained Si-Ge on SOI, silicon-germanium on silicon and a host of alternative vertical transistor structural configurations[387,388] as well as the ballistic transistor.[389]

Silicon-Germanium

The limitations of silicon electronic devices have warranted the development of compound semiconductors such as gallium arsenide with a larger electron and hole mobility as compared to silicon for specialized, mainly high frequency, electronic applications. The limited device integration possible with compound semiconductors, however, and the complexity of processing compound semiconductors has driven the extension of silicon technology by forming alloys of silicon with germanium as exemplified by Bernard Meyerson (Fig. 30), Electronics Division 1993 Awardee[390-392] and John Bean,[393] and their colleagues as well as other research groups. The silicon energy gap ($E_g = 1.12$ eV) at the X symmetry point can be monotonically decreased with the addition of germanium break ($E_g = 0.66$ eV). The flexibility of tuning the energy gap in silicon-germanium alloys can improve the performance of specialized, high performance electronic devices.[103]

In some cases, the germanium content is changed abruptly at a certain distance from the surface of the semiconductor to form an abrupt heterojunction, with a discontinuity in the energy gap. Because silicon and germanium have similar electron affinities, the