

Hans-Günther Wagemann

Grenzen der Kausalität in der Halbleitertechnik

Da die Zahl beteiligter Atome oder Moleküle bei technischen Betrachtungen meist sehr groß ist, verwendet man zusammenfassende „Zustandsvariable“ der großen Zahl, wie Temperatur, Druck, Energie, Dichte, Geschwindigkeit und andere. Der Ablauf eines Vorganges wird in diesem Rahmen von Zustandsvariablen meist streng „kausal“ und oft recht einfach ablaufen. Das mikroskopische Bild eines Zustandes dagegen ist weitaus komplizierter: hätte es doch die Bewegungsgleichungen einer nahezu unendlich großen Zahl von beteiligten Individuen zu beschreiben, nämlich stets von mehr als 10^{23} Mitwirkenden! In unserer „natürlichen“ Umgebung ist die zusammenfassende und vereinfachende Zustandsbeschreibung meist angemessen. Die womöglich nicht vollständig kausale Beschreibung eines einzelnen mikroskopischen Einzelprozesses verschwindet in der mittelwertbildenden Aussage aller anderen über den wahrscheinlichen Ablauf. Allerdings ändern sich diese Verhältnisse dramatisch, wenn die große Zahl dahinschwindet. Machen wir uns das an einem einfachen Beispiel klar.

Studierende jeder Anfangsvorlesung werden leicht durch die Frage verblüfft „Wie kalt ist eigentlich der freie Weltraum außerhalb des Sonnensystems?“ (Abb. 1). Die meisten vermuten die Temperatur nahe dem absoluten Nullpunkt. Die Zusatzfrage, ob man das nachweisen kann, führt auf die richtige Antwort. Die wenigen schnellen Teilchen im freien Weltraum stehen untereinander nicht im energetischen Gleichgewicht. Wenn ein Thermometer beliebiger Art sich mit der geringen Strahlung aus der Umgebung ins Gleichgewicht zu setzen versucht, stellt sich ein für diese Messung charakteristischer, aber zeitlich nicht konstanter Wert der Temperatur ein. Auch für das Thermometer mit seiner Masse inmitten der wenigen Teilchen bleibt die Temperatur undefiniert. Eine Maxwell-Geschwindigkeitsverteilung der beteiligten, sich untereinander stoßenden und kinetische Energie austauschenden Teilchen beschreibt für die wenigen Teilchen des freien Weltraums nur sehr ungenau eine jeweils im „Ensemble“ gemeinsame Temperatur eines Gleichgewichtes.

Situationen wie die folgenden sind für die heutige Festkörpertechnik charakteristisch, wenn man bei einer geringen Zahl beteiligter Individuen in Form von Gitteratomen in dünnsten Schichten ankommt. Oder: wenn mit wenigen Elementarladungen in kleinsten



Abbildung 1
Eineinhalb Millionen Lichtjahre ist der Andromeda Nebel von uns entfernt.
Welche Temperatur herrscht im Weltraum auf halbem Wege?

Speichern ein Signal zu speichern ist. Schließlich: wenn von statistischen Störungen im Gitterverband eines Kristalls Materialkonstanten bestimmt werden. Wir beschäftigen uns vielfach in unseren technischen „Nano-Welten“ mit nur mühsam kausal ablaufenden Prozessen, die auf Grund der geringen Zahl beteiligter Partner nicht immer ihren abschätzbar „wahrscheinlichen“ Verlauf nehmen.

Beginnen wir mit den berühmten „soft-errors“ der frühen Speichertechnik der Mikroelektronik. „Speicher“ sind dabei Plattenkondensatoren auf dem Halbleiterchip. Ein Plattenkondensator besteht ursprünglich aus zwei sich gegenüberstehenden Metallplatten, zwischen denen ein Isolator wie eine dünne Glimmerschicht vorhanden ist. Hier sind es zwei Halbleiter-Platten, und zwischen ihnen liegt eine dünne Schicht Siliziumdioxid. Das gespeicherte Medium ist elektrische Ladung, die sich elektrisch erhöhen und verringern lässt (Abb. 2). Wenn zum Beispiel die Spannung U am Eingang des Kondensators unter Null absinkt, also $U < 0$ gilt, wird auf der Gegenseite positive Ladung $Q > 0$ gespeichert. Dieser positiven Ladung Q kann man dann beispielsweise eine logische „1“ zuordnen. Als die Abmessungen der Speicherzellen so klein wurden, dass die ordnungsgemäß gespeicherte Ladung, zum Beispiel für eine logische „1“, ungefähr der Ladung entsprach,

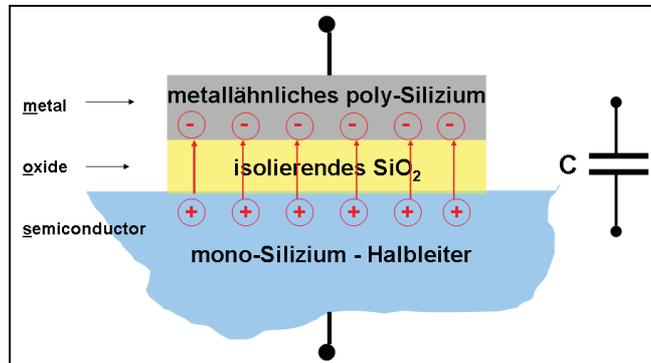


Abbildung 2

MOS-Plattenkondensator als integriertes Speicherelement (MOS ~ metal-oxide-semiconductor). Bei einer Fläche von $0,1\mu\text{m}^2$ und einer Dicke der SiO_2 -Schicht von 10nm lassen sich mit einer Spannung von 2V ca. 6000 Elementarladungen in dem Kondensator mit der Kapazität C speichern.

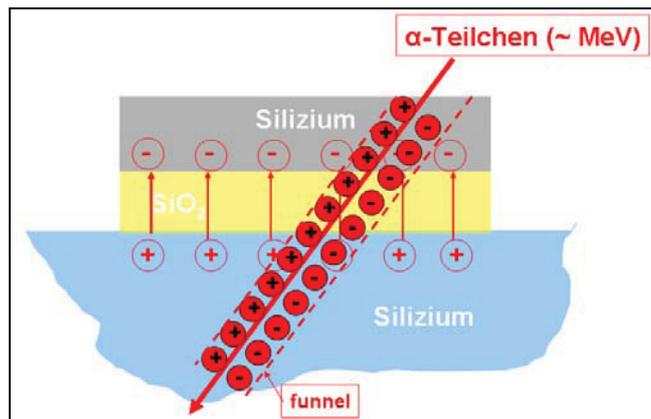


Abbildung 3

Soft error mit „ion funneling“ beim Einfall eines hochenergetischen α -Teilchens über alle Bereiche des Speicherelementes mit Veränderung in der Ladung Q

die ein auftreffendes hochenergetisches Teilchen als Ionenspur im System hinterlässt, geriet der Speicher hin und wieder in zunächst unerklärlicher Weise in Unordnung. Als Ursache erkannte man bald die Gamma-Quanten der kosmischen Höhenstrahlung, aber auch Alpha-Partikel aus dem meist keramischen Trägermaterial des Halbleiterchips. Diese „soft-errors“ aufzuklären und heutzutage weitgehend abzustellen, hat sehr viel Arbeit gekostet (Abb. 3).

Der Beginn der Unsicherheiten, nämlich ob „0“ oder „1“ in einer Speicherzelle steht, ist charakteristisch für die Auflösung kausaler Zuordnungen in miniaturisierten Systemen. Das typische heutige Speicherelement ist ein winziger Plattenkondensator mit SiO_2 -Dielektrikum und zwei Silizium-Elektroden bei einem Plattenabstand von 10 Nanometern und einer Fläche von 0,1 Quadratmikrometer. Man errechnet dafür eine Kapazität von $0,3 \cdot 10^{-15}$ Farad oder 0,3 Femtofarad, und bei einer Schaltspannung von -2 Volt speichert man weniger als 10^4 Elementarladungen als Information. Die vergleichbare Anzahl von Ladungen erzeugt ein Alpha-Teilchen auf seiner Abbremsspur von wenigen Mikrometern, die bei enger Packung der Speicherzellen durch eine oder mehrere Zellen führt und die Information zum Beispiel von „1“ auf „0“ umsetzen kann (Abb. 3).

Nicht nur Maßnahmen der Festkörpertechnologie, nämlich unter anderem die Wahl geeigneter synthetischer Substrat-Materialien, sondern auch Maßnahmen des dynamischen „refreshing“ der Speicher-Information, also des periodischen Wiedererzeugens der ansonsten verschwindenden gespeicherten Ladung, führten zur heutigen Beherrschung der „soft-errors“. Allerdings hört man immer wieder von unerwarteten Störungen der Halbleiterspeicher beim interkontinentalen Flug am Rande der Stratosphäre auf Grund der dort hohen Flüsse kosmischer Strahlung.

Ein weiterer Hinweis auf die sich auflösende Zuordnung kausaler Zusammenhänge in der Halbleitertechnik angesichts geringer Zahlen von zusammenwirkenden Partnern im atomaren Bereich ist die Beobachtung der sich über die Jahre hinweg stetig verringernden Eigenleitendichte n_i des Siliziums. Diese Größe beschreibt eine der wichtigsten Halbleitereigenschaften: das Gleichgewicht der freien Ladungsträger beider Arten (Elektronen und Löcher). Genau wie das Ionenprodukt des Wassers definiert ist und die Konzentrationen der H^+ - und OH^- -Ionen nach dem Massenwirkungsgesetz in einem multiplikativen Gleichgewicht zueinander stehen und den pH-Wert bestimmen (<7 : mehrheitlich H^+ -Ionen und saures Verhalten; >7 : mehrheitlich OH^- -Ionen und basisches Verhalten), so gilt eine entsprechende Aussage für die Ladungsdichten der Elektronen n und der Löcher p im Halbleiter: $n \cdot p = n_i^2$. Dem pH-Wert entspricht in diesem Sinne der Wert der Dichten n oder p bei Elektronen- oder Löcher-Leitung und deren jeweils im Gleichgewicht dominierender Dichte beim n-Halbleiter $n > n_i$ oder beim p-Halbleiter $p > n_i$. Seit 1965 hat sich nun der durch genaue Messungen an Reinst-Silizium bei $T=300\text{K}$ gewonnene Wert von $1,5 \cdot 10^{10} \text{ cm}^{-3}$ schrittweise auf $0,965 \cdot 10^{10} \text{ cm}^{-3}$ im Jahre 2003 verringert [1], also auf ca. 60 % des ursprünglichen Wertes, nicht unbedingt als Ergebnis verbesserter Messtechnik (Abb. 4).

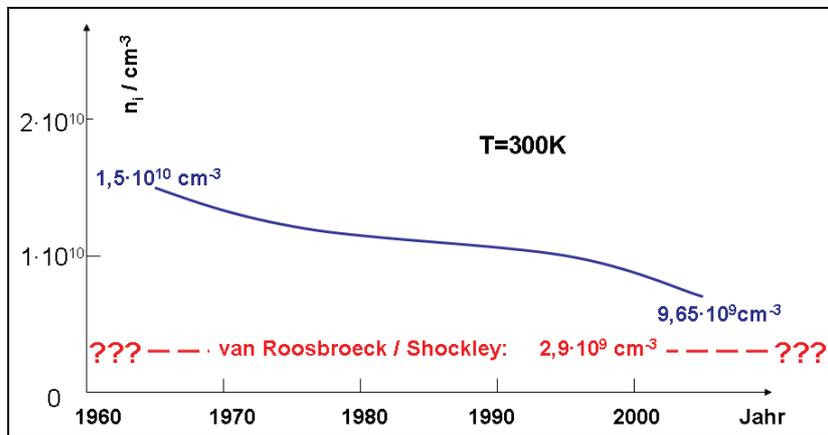


Abbildung 4
Verringerung der Eigenleistungsdichte n_i des Siliziums im Zeitraum 1965–2002

Woran mag diese für die Fachwelt unerwartete Verringerung liegen? Eine theoretische Abschätzung des Wertes n_i ist nach dem berühmten Prinzip von W. van Roosbroeck und W. Shockley [2] aus dem Jahre 1954 möglich. Die beiden Forscher betrachten die Emission von Strahlung als Umkehrung des Vorganges von Absorption, und ermitteln daraus die Emission aus dem Halbleitermaterial mit ihrer Eigenart. Die Methode war sehr erfolgreich bei der Ermittlung von technisch-nutzbaren Halbleitern für Lumineszenz-Dioden. Exakt gilt bei so genannten „direkten“ Halbleitern, dass beide gegenläufigen Vorgänge, die Absorption und die Emission von Photonen, sich quantitativ entsprechen. Bei Silizium als „indirektem“ Halbleiter ist das Prinzip nur eingeschränkt richtig: nicht-strahlende Übergänge mit Phononen-Mitwirkung stören die Bilanz. Für eine Abschätzung der Eigenleistungsdichte n_i jedoch kann man beim Silizium alle bekannten Übergänge miteinander verrechnen. Der damit gewonnene Wert liegt für Silizium bei $n_i(T=300\text{K}) = 2,9 \cdot 10^9 \text{ cm}^{-3}$. Ob er als „Konstante“ verlässlich ist, bleibt derzeit offen. Obwohl sich der experimentelle Wert tendenziell dem theoretischen nähert, bleibt die Spanne zwischen Experiment und Theorie immer noch beträchtlich.

Eine Erklärung für die beobachtete Verringerung des n_i -Wertes könnte sein, dass über die Jahre hinweg das einkristalline Reinst-Silizium immer perfekter hergestellt wird. Das bezieht sich nicht nur auf den Fremdstoffgehalt, den man beim Zonen-gereinigten Reinst-Silizium (FZ-Si) so perfekt beherrscht wie bei keinem anderen Werkstoff: es gelingt heute, Silizium herzustellen mit der Reinheit von $1:10^{-10}$, das heißt mit 1 Fremdatom auf 10

Milliarden Silizium-Atome! Hier geht es nun vor allem um die Dichte der nicht ordnungsgemäß eingebauten Gitteratome, vor allem um die so genannten Frenkel-Defekte, die als Punkt-Defekte jeweils ein Zwischengitteratom und eine Leerstelle umfassen, weniger um die Schottky-Defekte, bei denen Atome an die Oberfläche verlagert werden (Abb. 5). Jeder Temperatur entspricht nun eine Gleichgewichtsdichte der Frenkel-Defekte, deren theoretischer Wert sich aus Betrachtungen der Entropie als Maß für die thermodynamisch notwendige „Unordnung“ ermitteln lässt. Die thermischen Gitterschwingungen verlagern bei jeder Temperatur eine bestimmte Zahl von Gitteratomen auf Zwischengitter-Plätze. Experimentell lässt sich der Gleichgewichtswert der Frenkel-Defekte – von hohen Temperaturen kommend, wie es beim Kristall-Ziehen der Fall ist – bei tiefen Temperaturen nicht so leicht einstellen; denn es handelt sich um eine verringerte und zudem statistisch verteilte Anzahl von Defekten. Man wird sehr vorsichtig den fertigen Einkristall abkühlen müssen, um nicht eine Defektdichte höherer Temperatur „einzufrieren“. Dieses vorsichtige Abkühlen unter Schutzgas zieht sich unter Umständen über Wochen hin. Nur geringe Anteile der statistisch verteilten Frenkel-Defekte bei hoher Temperatur sollten bei Zimmertemperatur nach Rechnung verbleiben. Die Zwischengitteratome müssen dabei ihre regulären Gitterplätze wieder finden. Um dahin zu kommen, bedarf es nahezu alchemistischer Vorgehensweise mit Temperatursteuerung und katalysierender Gas-Umgebung, um für das thermodynamische Gleichgewicht die Zwischengitteratome wieder auf Leerstellen zurückzubringen.

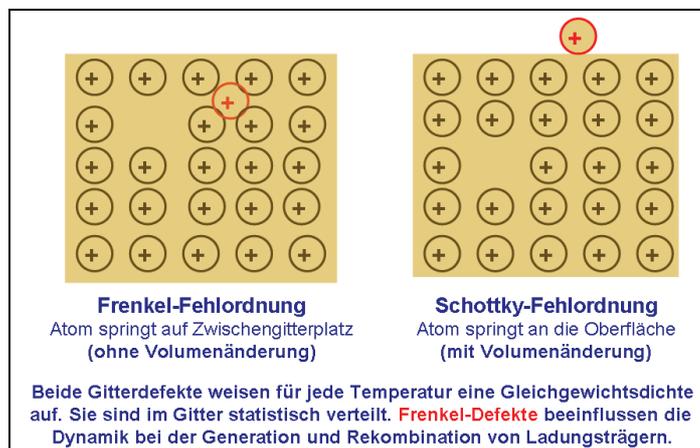


Abbildung 5
Frenkel- und Schottky-Defekte im Kristallgitter

Jede Messung der Eigenleitungsdichte ist mit Nichtgleichgewicht-Experimenten verbunden, bei denen zum Beispiel durch vorsichtige Lichteinstrahlung zeitweise mehr Ladungsträger beider Arten als im Gleichgewicht erzeugt werden und deren Abbau durch „Rekombination“ verfolgt wird. Rekombination bedeutet dabei die Wiedervereinigung der durch Strahlungsenergie freigesetzten Ladungsträger-Paare. Abweichungen vom perfekten Gitter erleichtern die Vereinigung von Ladungsträgern des Nicht-Gleichgewichtes. Die Rekombination erfolgt umso leichter, je mehr Frenkel-Defekte als Katalysatoren im Gitter vorhanden sind, was dann den n_i -Wert steigen lässt.

Messen wir über die Jahre hinweg in Form der stetig verringerten Konstanten n_i der Eigenleitungsdichte die sich stetig verbessernde Technik der Silizium-Abkühlung und die sie begleitende, immer perfektere Defekt-Ausheilung? Betrachten wir hier einen technisch bislang nicht völlig beherrschbaren Vorgang, ein kleine Störung der technologischen Kausalität, bei der ansonsten perfekt gehandhabten Technologie des Siliziums?

Schließlich möchte ich für die nicht kausal erfolgenden Ereignisse der Festkörperelektronik die „technologischen Einbrüche“ einer ganzen Bauelement-Fabrikation nennen, die von Zeit zu Zeit jedes Halbleiterlabor, auch die perfekttest organisierte Industrietechnologie, heimsuchen und für die zunächst einmal keine rationalen Gründe angeführt werden können. „Technologischer Einbruch“ steht für „... wir wissen nicht, warum ...“. Der unerwartete Niedergang des „yield“, der Ausbeute der Fertigung, der bei einer gut „eingefahrenen“ Technologie ohne weiteres über 95 % liegt, lässt Schlimmes ahnen. Im Sinne einer langen Kausalkette allmählich zurückzugehen und Ursachen aufzudecken, gelingt meist nicht und dauert auch zu lange. Zu stark verzweigt sind die Möglichkeiten, wie sich Unregelmäßigkeiten in den streng kontrollierten Betrieb einer Reinraum-Technologie eingeschlichen haben können, zu gering das Maß der Verunreinigungen, die eine ganze Produktion scheitern lassen. Es hat Fälle gegeben, bei denen man als Folge „technologischer Einbrüche“ eine gänzlich neue Technologie aufgebaut hat. Im Allgemeinen fängt man ganz vorn in der Kausalkette an und vertraut auf die Kraft des Neuanfanges.

In einem deutschen Halbleiterwerk stellte sich heraus, nachdem man längere Zeit mit einem nicht identifizierbaren „technologischen Einbruch“ bei ungefähr jedem zwanzigsten, dreißigstem Wafer gelebt hatte, dass eine besonders eifrige Mitarbeiterin regelmäßig Wafer vom geschützten Produktionsband nahm und sie genauester visueller Inspektion unterzog – fünf Zentimeter entfernt von ihrer Nase mit einer Schutzmaske aus Papier. Die Natrium-Ionen des Atems reichten trotz der Nasenmaske für den Totalausfall aller Bauelemente auf den besonders genau inspizierten Wafers (Abb. 6).

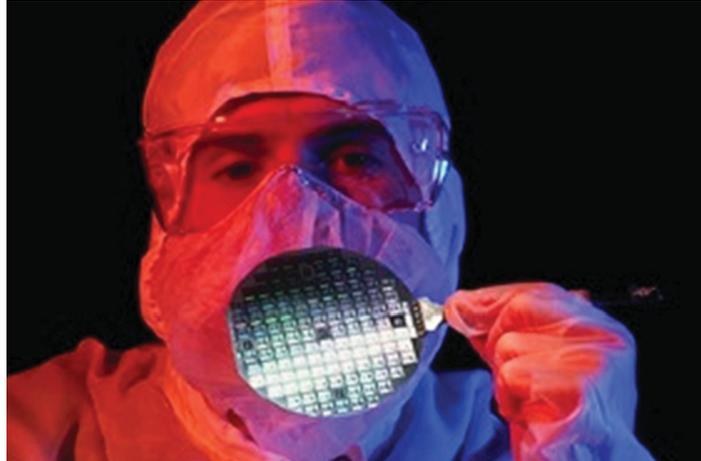


Abbildung 6

Bei der Chip-Produktion sollte man zu nahen Kontakt mit den ungeschützten Wafern vermeiden!

Auch hier existiert eine vom handelnden Menschen nur unzureichend verfolgbare Kausalkette. Zunächst gilt: je geringer die Menge bearbeiteten Werkstoffs, umso höher der Durchgriff der Mikrovorgänge mit ihren Wahrscheinlichkeitsverteilungen. Aber auch: je präziser über Mikrovorgänge Aussagen gefordert werden, umso mehr erkennt man deren Anfälligkeit durch unerwartete und unübersichtliche Einflüsse. Schließlich: je besser man eine technische Produktion als viel verzweigte Kausalkette beherrscht, umso folgenreicher ist ein unerwarteter, „akausaler“ Eingriff des Menschen, sozusagen von außerhalb der Grenzen des betrachteten Systems.

Literatur

- [1] Altermatt, P. P., Schenk, A., Geelhaar, F. & G. Heiser: J. Appl. Phys. 93 (2003), p.1598 u. f.
- [2] van Roosbroeck, W. & W. Shockley: Phys. Rev. 94 (1954), p.1558 u. f.