

SIEMENS



Mikroelektronik

Märkte, Megatechnik, Massenfertigung

Gigabit-Chips zur Jahrhundertwende?
Megabit-Multis fordern die Technologen heraus.
Das kleine Delta zwischen Gewinn und Verlust.

Sonderdruck: Beiträge aus TECHNISCHE RUNDSCHAU
Nr. 12 vom 22. März 1991, Nr. 23 vom 7. Juni 1991 und
Nr. 29 vom 19. Juli 1991

Autoren: Dr. F. S. Becker, Dr. H. Föll

Gigabit-Chips zur Jahrtausendwende?

Von Frank Stefan Becker und Helmut Föll

Die dynamischen Schreib-Lese-Speicher, die DRAMs, spielen in der Mikroelektronik die Rolle einer Technologielokomotive. Seit Anfang der siebziger Jahre der erste 1-kbit-Speicher auf den Markt kam, wurde etwa alle drei Jahre eine neue, um den Faktor 4 leistungsfähigere DRAM-Generation hervorgebracht. Am Ende dieser Entwicklung wird um die Jahrtausendwende wahrscheinlich der 1-Gbit-Chip stehen. Die möglichen neuen Anwendungen auf der Basis dieser Technologie sind heute, im Zeitalter des 1-Mbit-Chips, noch gar nicht vollständig zu übersehen.

Im Laufe seiner Evolutionsgeschichte war es stets das Bestreben des Menschen, seine Fähigkeiten durch den Einsatz von Hilfsmitteln zu steigern. Primitive Steinwerkzeuge, die Nutzung der Metalle und die Erfindung von Maschinen waren Schritte auf dem Weg zur «ersten industriellen Revolution», in der der Mensch seine Körperkräfte vervielfachte. Die Entdeckung der Elektrizität schuf die Möglichkeit der schnellen Wissensvermittlung und damit die Voraussetzung für die derzeit ablaufende «zweite industrielle Revolution». Übertragung, Speicherung und Weiterverarbeitung von Daten sind die zentralen Hilfsmittel, um jetzt auch die geistigen Fähigkeiten des Menschen zu potenzieren. Entscheidende Schritte waren hier der Bau des Zuse-Relaisrechners 1936, die Erfindung des Transistors 1948 und die Präsentation des ersten integrierten Germaniumschaltkreises 1958.

Im Jahre 1970 markierte die Realisierung eines 1-kbit-Speichers in Siliziumplanartechnik den Beginn einer neuen Phase, an deren Ende um die Jahrtausendwende voraussichtlich der 1-Gbit-Chip stehen dürfte. Exakt 1 073 741 824 Ja/Nein-Informationen, das entspricht gut 64 000 Schreibmaschinenseiten oder

dem Inhalt eines zwanzigbändigen Lexikons, gespeichert auf einem Siliziumplättchen, das nicht einmal die halbe Fläche einer Pro-Juventute-Briefmarke abdeckt – eine schwindelerregende Vorstellung. Die auf der Basis einer solchen Technologie möglichen neuen Anwendungen sind heute, im Zeitalter des 1-Mbit-Chips, noch nicht zu übersehen. Sie dürften aber Technik und Wirtschaft nicht weniger revolutionieren, als dies in den vergangenen 20 Jahren beim Weg von der Kilobit- zur Megabitgeneration der Fall war.

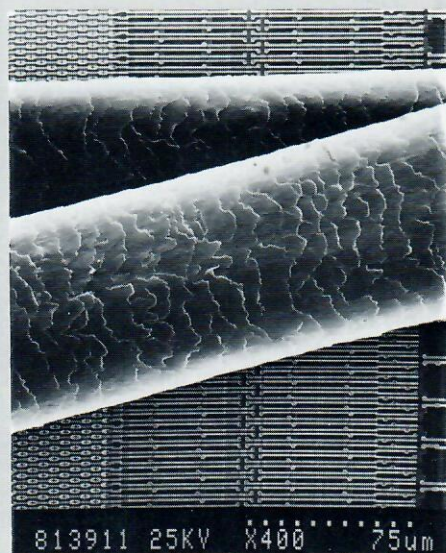
Die Speicher spielen eine Sonderrolle

Unter dem Schlagwort «Mikroelektronik» wird heute eine Vielzahl von Produkten am Markt angeboten. Die Spanne umfaßt zum Beispiel Leistungselektronik, Optoelektronik, Speicherchips und «Logikprodukte», wobei letzterer Begriff unterschiedlichste integrierte Schaltungen für Datenverarbeitung, Telekommunikation und Unterhaltungselektronik zusammenfaßt.

Speicherbausteine spielen traditionell eine Sonderrolle, da sie sehr regelmäßig aufgebaut sind und dank universeller Einsetzbarkeit in großen Stückzahlen produziert werden. Dies macht sie besonders geeignet, um neue Chipproduktionen mit ihren teilweise noch wenig erprobten Prozessen «einzufahren», das heißt, möglichst bald die für Halbleiterfabriken typische Endausbeute von rund 70 bis 90 % zu erreichen.

Modernste integrierte Schaltungen bestehen aus Millionen von Einzelelementen mit Strukturbreiten von einem Hundertstel eines Menschenhaars, wie Bild 1 illustriert. Deshalb können Defekte oder Prozeßprobleme fast nur noch über statistische Auswertemethoden lokalisiert werden, wozu Speicher wegen ihrer regelmäßigen Strukturen die beste Basis bilden. Einige Jahre später wird die so erprobte Technologie auch für die weniger regelmäßigen Logikprodukte mit ihren kleineren Fertigungsvolumina genutzt. Bild 2 zeigt diese Vorreiterrolle der Speicher anhand der Entwicklung der Integrationsdichte bei dynamischen Speichern

Bild 1. Menschliche Haare auf den Leiterbahnen eines 1-Megabit-Chips.



(DRAM, Dynamic Random Access Memory) und bei Mikroprozessoren. Halbleiterspeicher ersetzen magnetische Speichermedien nicht, sondern sie ergänzen sie. Der ideale Baustein müßte Daten dauerhaft halten, hohe Lese- und Schreibgeschwindigkeiten haben, häufig umprogrammierbar und kostengünstig sein, also nur wenig Schaltelemente für ein Bit benötigen. Da in der Praxis diese zum Teil widersprüchlichen Anforderungen nicht gleichzeitig realisierbar sind, wurden entsprechend den jeweiligen Einsatzgebieten verschiedene Prinzipien entwickelt. So gibt es die «flüchtigen» Speicher, die ihre Information bei Abschaltung der Spannungsversorgung verlieren, und die «nicht-flüchtigen», deren Umprogrammierung jedoch zumeist zeitaufwendig und limitiert ist. Tabelle 1 gibt einen Überblick über die wichtigsten Typen.

Sowohl technologisch als auch wirtschaftlich nehmen die DRAMs eine Sonderstellung ein. Wegen des geringen Platzbedarfs der nur aus Schalttransistor und Speicherkondensator bestehenden Elementarzelle lassen sich hohe Integrationsdichten und damit geringe und, wie Bild 3 zeigt, stetig fallende Bitpreise realisieren. Da DRAMs mit über 50 % Marktanteil den dominierenden Speichertyp darstellen, spielen sie traditionell die Rolle eines Motors für den Fortschritt in der Mikroelektronik, denn parallel zum Bitpreis sinkt auch der Preis für Rechenoperationen.

Seit zu Beginn der siebziger Jahre der erste 1-Kbit-Speicher auf den Markt kam, wurde etwa alle drei Jahre eine neue, um den Faktor 4 leistungsfähigere DRAM-Generation entwickelt. Diesem Rhythmus liegt keine physikalische Notwendigkeit zugrunde, sondern er stellt einen für Hersteller wie Anwender tragbaren Kompromiß zwischen zu vielen kleinen oder wenigen großen Schritten dar. Beide Seiten können so die technische Entwicklung und die Einführung neuer Produkte besser planen.

Hauptabnehmer für DRAMs sind heute Hersteller von Datenverarbeitungsanlagen, speziell Personal Computern. In Tabelle 2 sind typische Mittelwerte für derzeit beziehungsweise in Zukunft in verschiedenen Systemen installierte DRAM-Kapazitäten aufgeführt. Trotz des hohen Speicherbedarfs der Supercomputer ist jedoch der PC-Markt dank seiner hohen Stückzahlen der Hauptabnehmer für DRAMs, wie aus Bild 4 ersichtlich ist. Die derzeitige Absatzschwäche bei Personal Computern ist eine der wesentlichen Ursachen für den in den vergangenen zwei Jahren besonders ausgeprägten DRAM-Preisverfall, der vielen Speicherproduzenten tiefrote Zahlen bescherte.

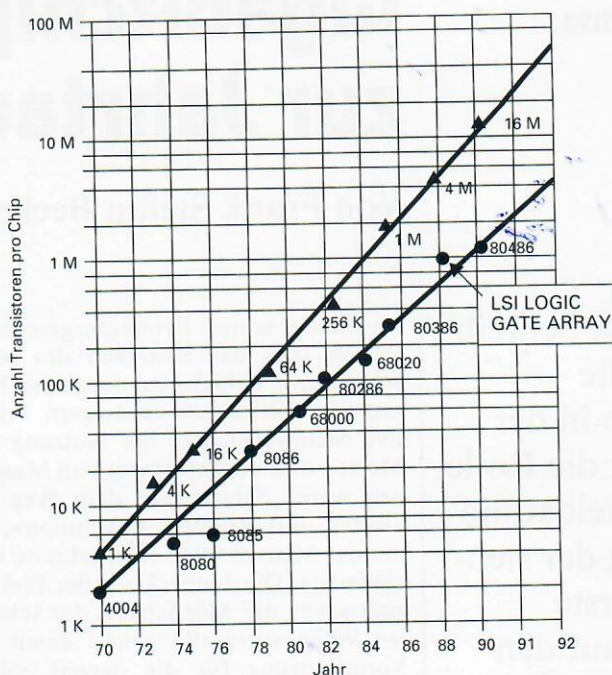


Bild 2. Entwicklung der Integrationsdichte, gemessen in Transistoren pro Chip, bei dynamischen Speichern und bei Mikroprozessoren. Die «Technologiekomotive Speicher» fährt schneller als der «Logikzug».

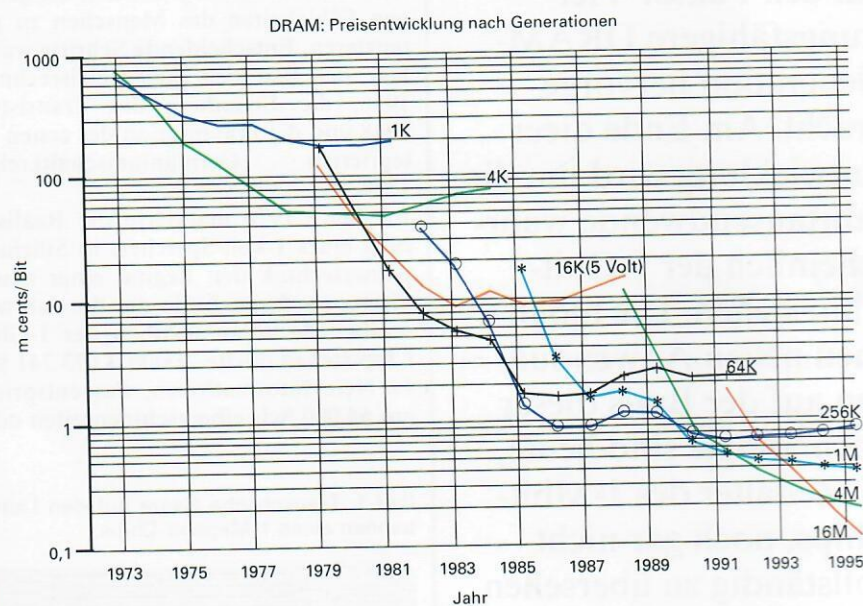


Bild 3. Entwicklung des Bitpreises für verschiedene DRAM-Generationen. Dieser Preisverfall für elektronische Funktionen ist ein treibender Faktor bei der Verbreitung der Mikroelektronik. (Quelle: Dataquest, Januar 1991)

Markt und Mächte

Im Jahre 1988 betrug der weltweite Markt für Mikroelektronik etwa 50 Mia \$, in der Bundesrepublik Deutschland waren es 2,8 Mia DM. Diese Zahlen geben jedoch die tatsächliche Bedeutung dieses Industriezweigs nicht wieder, da sehr viele moderne Produkte Chips benötigen. In der Bundesrepublik wurden zum Beispiel 1988 Güter aus den Berei-

chen Maschinenbau, Elektrotechnik, Feinmechanik/Optik/Uhren, Fahrzeugbau und Bürogeräte im Werte von über 600 Mia DM hergestellt. In dieser «Fünfergruppe», die etwa 30 % des Bruttosozialprodukts ausmacht, spielen die Chips wertmäßig zumeist nur eine untergeordnete Rolle (1988 rund 0,5 %). Ihr Einfluß auf die Leistungsfähigkeit des Produkts und damit auf seine Konkurrenzfähigkeit auf dem Weltmarkt ist

Tabelle 1. Vergleich verschiedener Halbleiterspeicher.

Speichertyp	Anwendungsbeispiele	verfügbare Kapazitäten (bit)	Datenerhalt ohne Speisung	Zahl der Umprogrammierungen	Umprogrammierungszeit	Zugriffszeit (ns)	Marktvolumen 1989	1994	lösbbare Grundeinheit
DRAM Dynamic Random Access Memory	Personal Computer bis Großrechner	64 K bis 16 M		unbegrenzt	100 bis 200 ns	50 bis 150	10 Mia \$	19 Mia \$	bitweise byteweise
SRAM Static Random Access Memory	Zwischenspeicher für Mikrocontroller, Großcomputer	16 K bis 1 M		unbegrenzt	20 bis 150 ns	10 bis 100	3,4 Mia \$	8,4 Mia \$	bitweise byteweise
EPROM Erasable Programmable Read Only Memory	Drucker, Terminals, Telefone	16 K bis 4 M	Jahre	10^2 bis 10^3	20 bis 30 min	70 bis 300	1,8 Mia \$	2,5 Mia \$	mit UV-Licht ganzer Chip
EEPROM Electrically Erasable Progr. Read Only Memory	Drucker, Terminals, Telefone	16 K bis 1 M	Jahre	10^5	5 bis 10 s	30 bis 200	300 Mio \$	700 Mio \$	blockweise oder ganz
Flash-EPROM	Speicherkarten für tragbare Kleincomputer	16 K bis 2 M	Jahre	10^4 bis 10^5	2 bis 5 s	90 bis 200	12 Mio \$	1,2 Mia \$	blockweise oder ganz
MOSROM	Sprachsynthese, elektronische Wörterbücher	16 K bis 16 M	unbegrenzt			90 bis 200	1,2 Mia \$	1,9 Mia \$	

jedoch häufig entscheidend. Für die Verhältnisse in der Schweiz gilt vereinfacht die Faustregel, daß 1 «Chipfranken» einen Industrieumsatz von etwa 200 Franken bestimmt. Unterhaltung in der Freizeit, Datenverarbeitung im Büro, schnelle und hochqualitative Datenübertragung, Automatisierung in der Fertigung, Energieeinsparung und Sicherheit beim Auto sowie Meßwertenerfassung im Umweltschutz sind nur einige Aspekte moderner Technik, die ohne Mikroelektronik nicht denkbar wären. Dementsprechend ist dieser Markt auch umkämpft wie kaum ein anderer, da er gleichzeitig enorme Chancen und Risiken bietet.

Bild 5 vergleicht den derzeitigen Stand und die voraussichtliche Entwicklung der Elektronik in Europa mit anderen Industriezweigen. Das erwartete Wachstum soll die Elektronik bis zum Jahre 2000 zum umsatzstärksten Zweig machen, wobei die Untermenge Mikroelektronik überproportional partizipieren dürfte. Für die Hersteller von Chips ist dies jedoch kein Anlaß zu ungetrübter Freude, da die Entwicklung, wie in Bild 6 dargestellt, alles andere als stetig verläuft. Trotz eines ungebrochenen Wachstums der Stückzahlen kann bei einem Zusammentreffen von Überangebot und schwacher Nachfrage der schon erwähnte Preisverfall (Bild 3) so stark sein, daß das Marktvolumen sogar schrumpft. Da ein Halbleiterproduzent jedoch wegen der enormen Vorleistungen für Entwicklung und Fertigung

Bits und Bytes

Grundlage jeder modernen Datenverarbeitung ist die binäre Digitaltechnik. Alle Informationen, seien es Bilder, Musik oder Texte, werden zuerst in Zahlen überführt und dann im Binärsystem dargestellt. Da hier nur die Zahlen 0 und 1 existieren, braucht die Schaltung nur zwischen definierten Zuständen wie «Spannung ein/aus» zu unterscheiden. Jede dieser Ja/Nein-Informationen läßt sich als ein «Bit» bezeichnen. Als Einheit für die Information wird bit klein und in Einzahl geschrieben (beispielsweise 64-Mbit-Chip).

Die Begriffe «Kilo», «Mega» und «Giga» stehen dabei für 2^{10} , 2^{20} und 2^{30} bit. Diese Einteilung wurde in Analogie zum vertrauten Dezimalsystem gewählt, entspricht diesem aber nicht genau. Die Großbuchstaben K, M und G stehen hier nämlich für die folgenden Werte: 1 K = 1024, 1 M = 1 048 576 und 1 G = 1 073 741 824. Der Kleinbuchstabe k steht für 1000.

Ein «Byte» ist die Einheit von 8 bit, die zur Darstellung eines Zeichens benötigt wird. Damit läßt sich eine achtstellige Zahl im Zweiersystem darstellen. Wegen der Existenz der 0 gibt es damit maximal $2^7 = 128$ verschiedene Möglichkeiten zur Definition von Buchstaben, Interpunktions- und anderen Zeichen. Die Darstellung einer Textseite mit 40 Zeilen zu 50 Zeichen benötigt im Binärsystem demnach 16 000 bit. Als Einheit wird Byte groß und in der Einzahl geschrieben (zum Beispiel 40-MByte-Speicherplatte).

kaum Möglichkeiten zu entsprechender Kostenreduktion hat, führte der Mitte der achtziger Jahre aufgetretene Einbruch bei einer Reihe von amerikanischen Herstellern zum Entschluß, zumindest das besonders risikoreiche DRAM-Geschäft aufzugeben.

Neben dieser kurzfristigen Entscheidung bewirkten aber auch längerfristige Trends, daß die USA als frühere Mikroelektroniknation Nr. 1 von Japan abgelöst wurde. Unter Führung des MITI baute Japan bereits in den siebziger Jahren systematisch und unbeirrt von Konjunkturschwankungen eine weit über den eigenen Bedarf hinausgehende Mikroelektronikindustrie auf. Ziel war es, eine dominierende Stellung in dieser Zukunftsindustrie zu erringen und sich damit einen oft entscheidenden Vorsprung im internationalen Wettbewerb zu sichern.

Den Erfolg dieser Politik demonstrieren die Bilder 7 und 8. Bild 7 zeigt die Entwicklung der prozentualen Anteile der USA, Westeuropas und Japans bei den Beiträgen zur «International Solid State Circuit Conference», einer jährlich in den USA stattfindenden wichtigen Halbleitertagung. Die noch vorhandene knappe Führerschaft der USA dürfte eher mit der Wahl des Tagungsorts als mit geringerer japanischer Kreativität zusammenhängen.

Bild 8 vergleicht die Umsätze der jeweils drei größten Halbleiterhersteller in dieser Triade Japan, USA und Europa. Noch deutlicher wird die japani-

sche Dominanz, wenn man berücksichtigt, daß selbst Matsushita als Nr. 6 in Japan noch vor Philips als der europäischen Nr. 1 liegt. Der leichte Rückgang des japanischen Weltmarktanteils im vergangenen Jahr sollte nicht mit einer Trendwende zugunsten der westlichen Triadenpartner verwechselt werden. Der Grund für die Verschiebung der Gewichte zugunsten von speziell Intel und Motorola liegt darin, daß diese Firmen auf dem Mikroprozessormarkt dominieren, wo die Preise nicht wie bei den von den Japanern bevorzugten DRAMs um zwei Drittel zurückgingen. Mit dem koreanischen Unternehmen Samsung ist darüber hinaus ein neuer, aggressiver Wettbewerber auf dem Speichermarkt erschienen, so daß sich die Konkurrenzsituation eher verschärfen dürfte.

IBM ist in Bild 8 nicht aufgeführt, da diese Firma integrierte Schaltungen bisher nur für den Eigenbedarf herstellt. Im Rahmen der Zusammenarbeit mit Siemens ist allerdings auch hier geplant, mit dem 64-Mbit-DRAM ein den internationalen Standards entsprechendes Produkt zu entwickeln.

Hintergrund dieser Entscheidung sind die für alle Firmen immer drückenderen Kosten, die die Teilnahme am Megachip-Technologie-Wettlauf verursacht. Bei der in Bild 9 dargestellten Entwicklung ist bemerkenswert, daß die Aufwendungen für die Fertigung gegenüber der Entwicklung mehr und mehr in den Vordergrund treten. Da es sich bei DRAMs um ein standardisiertes Massenprodukt handelt, liegt es nahe, hier durch Zusammenarbeit Geld und Zeit zu sparen.

Europa hat Nachholbedarf

Mit einem Weltproduktionsanteil von 10,5 % lag Europa nach der Dataquest-Statistik 1990 weit hinter den USA mit 36,5 % und Japan, das fast die Hälfte aller Chips produzierte. Eine entsprechende Reihenfolge ergibt sich, wenn man den Mikroelektronikverbrauch der Gerätehersteller auf die Bevölkerung normiert. Auch hier führte Japan 1989 mit knapp 250 Franken pro Kopf klar vor den USA mit etwa 95 Franken und Westeuropa mit 40 Franken. Obwohl also Europa von seiner gesamten Wirtschaftsleistung der der beiden anderen Blöcke durchaus vergleichbar ist, bestehen bei Produktion und Einsatz integrierter Schaltungen ebenso wie auf dem Forschungsgebiet (Bild 7) noch erhebliche Rückstände.

Woher kommt dieser Unterschied? Während die USA als Erfinder der Mikroelektronik einen Startvorteil hatten

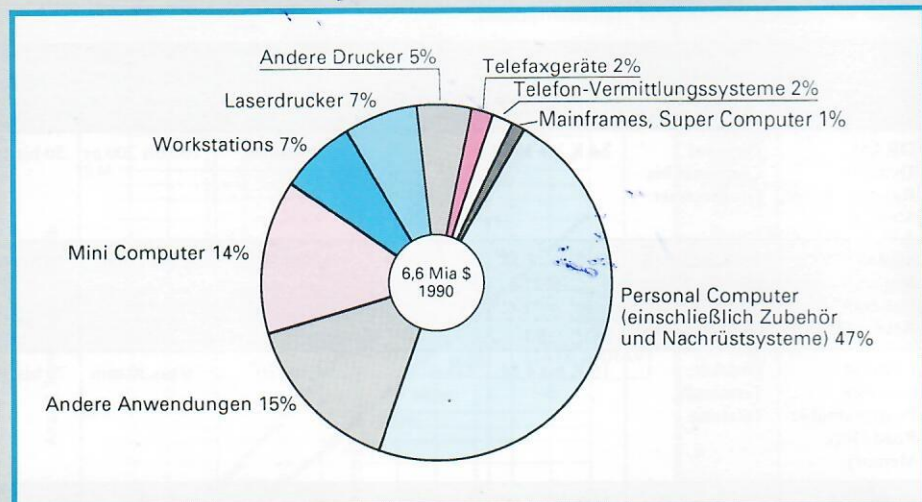


Bild 4. Aufteilung des Marktes für DRAMs nach Einsatzgebieten. (Quelle: Dataquest, August 1990)

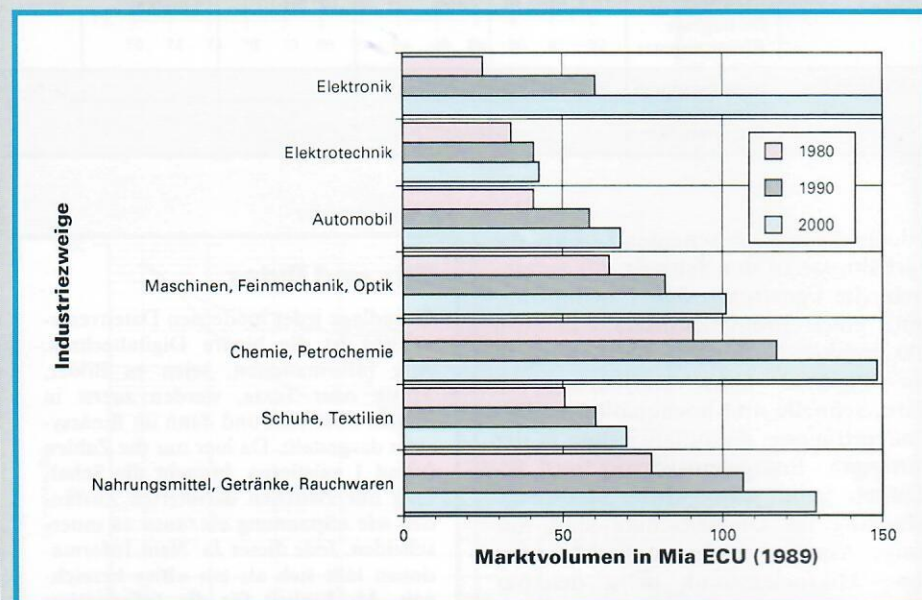


Bild 5. Entwicklung der Elektronikindustrie in Europa im Vergleich zu anderen Industriezweigen. (Quelle: JESSI 1989)

Tabelle 2. Mittelwerte in verschiedenen Systemen installierter DRAM-Kapazitäten.

Anwendung	mittlere Speicherkapazität 1990	mittlere Speicherkapazität 1995
Laserdrucker	8 Mbit	?
Personal Computer	16 Mbit	384 Mbit
Minicomputer	192 Mbit	736 Mbit
Technical Workstation	256 Mbit	1024 Mbit
Mainframe, Supercomputer	512 Mbit	1536 Mbit

und militärische Aufträge eine erhebliche Rolle spielten, setzte Japan unter staatlicher Koordination konsequent auf die zivile Nutzung der Chips für den Massenmarkt der Unterhaltungsindustrie. In Europa dagegen dominierte ein Produktspektrum, das es gestattete, mit ausgefeilter konventioneller Tech-

nik gutes Geld zu verdienen, so daß der Drang der Wirtschaft zu risikoreichen Experimenten gering war. Hinzu kam eine staatliche Politik, die aus Sorge um Arbeitsplätze lieber traditionelle Strukturen mit Milliardensummen stützte, als einen Bruchteil dieser Beträge für Neuerungen bereitzustellen. So erreichten

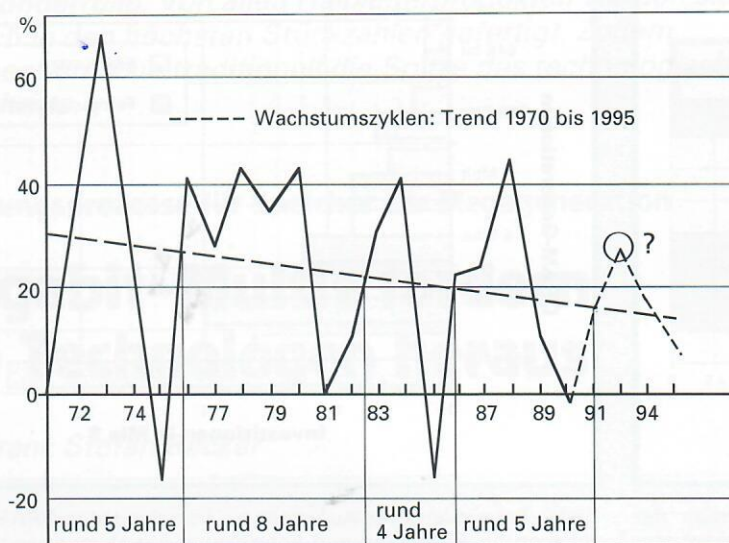


Bild 6. Wachstumsschwankungen des weltweiten Mikroelektronikmarkts. Die Prozentwerte beziehen sich auf Dollarumsätze; in Stückzahlen gerechnet, war immer eine Zunahme zu verzeichnen. (Quelle: Dataquest, INSTAT, Siemens)

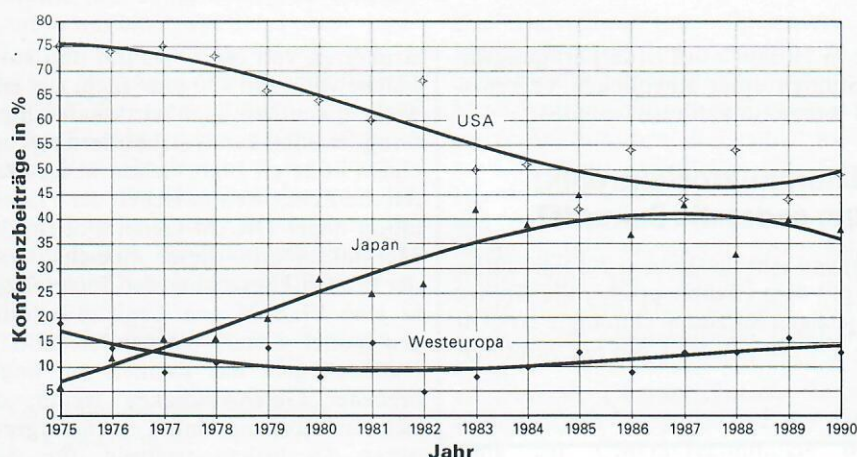


Bild 7. Prozentuale Verteilung der Beiträge auf der «International Solid-State Circuits Conference» auf die Triade USA, Westeuropa und Japan. Der sich im Bereich weniger Prozente bewegende Anteil der übrigen Welt wurde vernachlässigt. Erkennbar ist die Kräfteverschiebung zwischen den USA und Japan (nach einer Idee von Prof. W. Engl).

von JESSI ist dabei weder die Entwicklung eines ganz bestimmten Produkts noch die Errichtung einer gigantischen «JESSI-Fabrik», sondern die Schaffung der erforderlichen Technologiebasis und die Stärkung der europäischen Mikroelektronikinfrastruktur. Da bei der Komplexität moderner integrierter Schaltungen die Entwicklung neuer Fertigungsanlagen nur in enger Zusammenarbeit mit dem zukünftigen Kunden, das heißt dem Chiphersteller, erfolgreich durchgeführt werden kann, bedeutet die bisher geringe Präsenz Europas auf dem Equipmentmarkt ein zusätzliches Hindernis beim Aufbau eigener Halbleiterfertigungen.

Das in München angesiedelte JESSI-Büro übt in dem bis 1996 laufenden Projekt eine koordinierende Funktion aus, die Verantwortung für den Fortgang der Arbeiten liegt bei den jeweiligen Projektführern. Wenn ein Unternehmen oder ein Institut bereit ist, das von ihm vorgelegte Projekt zu mindestens 50 % selber zu finanzieren, wird das Vorhaben von JESSI geprüft und erhält eine positive Empfehlung, falls die Ziele mit den im «Grünbuch» definierten übereinstimmen. Mit diesem «JESSI-Stempel» kann der Antragsteller dann öffentliche Zuschüsse beantragen.

Die immer wieder genannte Summe von knapp 7 Mia Franken ist der Betrag, der als zur Realisierung der im «Grünbuch» aufgelisteten Projekte notwendig abgeschätzt wurde. Diese auf Preisen von 1988 basierende Summe ist also keineswegs in einem von JESSI verwalteten Topf vorrätig, sondern muß zu je 25 % von der Europäischen Gemeinschaft und den nationalen Regierungen im Rahmen der Projektförderung noch genehmigt werden. Die dafür gültigen Richtlinien sind allerdings nicht überall einheitlich, da in manchen Ländern die Unterstützung in engem Zusammenhang mit regionalen Strukturförderungsmaßnahmen gesehen wird.

Zusammen mit den von der EG verhängten Zöllen von 14 % auf importierte Chips hat nationale Strukturpolitik auch dazu geführt, daß eine Reihe von amerikanischen und vor allem japanischen Unternehmen in den letzten Jahren eigene Fertigungsstätten in Europa errichtet haben. Um die Klassifikation als ausländisches Produkt zu vermeiden, müssen die Chips dort «diffundiert» werden, das heißt einen bei etwa 1000 °C stattfindenden Prozeßschritt durchlaufen, der typischerweise in der Anfangsphase der Chipherstellung erfolgt. So wird verhindert, daß beispielsweise in Japan hergestellte integrierte Schaltungen nur noch zur Endmontage nach Europa kommen.

beispielsweise die vom deutschen Forschungsministerium 1985 bis 1989 an Siemens bewilligten Zuschüsse von rund 50 Mio DM pro Jahr für das Megaprojekt ungleich mehr Aufmerksamkeit als die jährlichen staatlichen Kohle-subsventionen von über 10 Mia DM. Ergänzt wurde die Problematik lange Zeit durch die Abneigung vieler Hochschulen, sich ihre akademische Freiheit durch zu nahe Anbindung an Industrieentwicklungen einengen zu lassen, sowie eine negative Einstellung in Teilen der Bevölkerung. «Chips als Jobkiller» und «Sonntagsarbeit» seien hier nur als Stichworte genannt.

Trotz grundsätzlich im Weltmaßstab unveränderten Kräfteverhältnissen trat jedoch in Europa seit Mitte der achtziger Jahre ein gewisser Bewußtseinswandel ein. Das als Technologiekooperation zwischen Siemens und Philips durchgeführte Megaprojekt schuf eine Know-how-Basis, ohne die beispielsweise ein Siemens-IBM-Bündnis nicht denkbar gewesen wäre. Aufbauend darauf, wurde im Rahmen von JESSI (Joint European Submicron Silicon) eine breite Zusammenarbeit zwischen Halbleiterindustrie, Forschungsinstituten, Herstellern von Produktionsanlagen und Anwendern konzipiert. Ziel

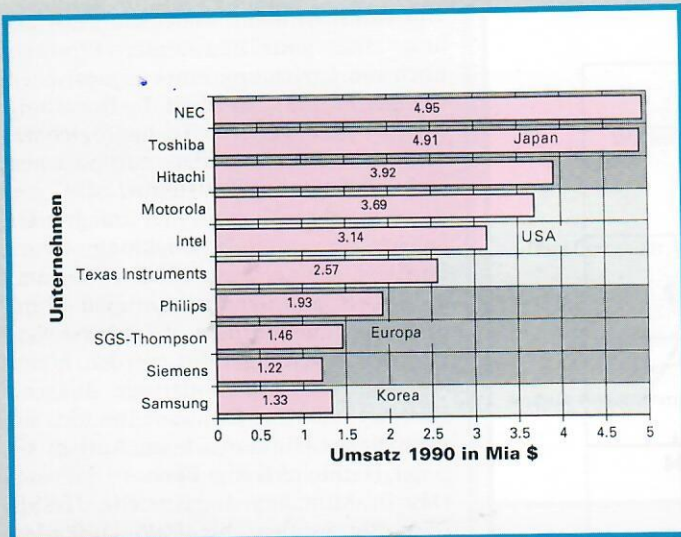


Bild 8. Umsätze der jeweils drei größten Halbleiterfirmen der Triade. Mit Korea hat sich inzwischen ein weiterer Konkurrent etabliert. (Quelle: Dataquest, Dezember 1990)

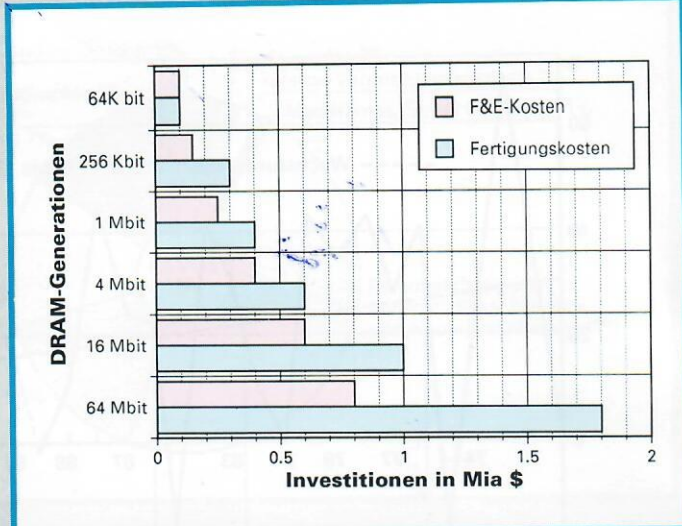


Bild 9. Entwicklung des Aufwands bei verschiedenen DRAM-Generationen für Forschung und Entwicklung beziehungsweise für die Errichtung einer Fabrik für die Produktion von 4 Mio Stück pro Monat. Die Kosten für die Fertigung steigen erheblich schneller als jene für die Entwicklung. (Quelle: Nikkei Microdevices 11/88; Prof. I. Ruge)

Beispiele für ausländische Investitionen in Europa sind die Fabriken von NEC in Livingston und von Motorola in East Kilbride (Schottland), die Prozeßlinie von Fujitsu in Durham (England), die Produktionsstätten von IBM in Sindelfingen (Bundesrepublik Deutschland) und in Corbeil (Frankreich), die neue Scheibenfertigung von Texas Instruments in Avezano (Italien) und die im Bau befindlichen Halbleiterfabriken von Mitsubishi in Alsdorf und von Hitachi in Landshut (beide Bundesrepublik Deutschland). Diese Präsenz ausländischer Unternehmen bedeutet zwar eine verschärfte Konkurrenz für die einheimischen Produzenten, verbreitert aber zugleich die europäische Mikroelektronikbasis.

Die von der früheren DDR-Regierung mit über 14 Mia Ostmark als Prestige-projekt geförderte Mikroelektronik in Ostdeutschland spielt in diesem Rahmen kaum eine Rolle. Zwar wurde unter anderem im Forschungszentrum Mikroelektronik Dresden oder im VEB Mikroelektronik Erfurt gute Arbeit – vor allem im Designbereich – geleistet, der Aufbau einer westlichen Maßstäben entsprechenden Fertigung gelang jedoch trotz allen Improvisationstalenten wegen des Cocom-Embargos nicht. So wurden zum Beispiel 1988 vom 64-Kbit-Speicher als dem Hauptprodukt bei einer Ausbeute von 60 % etwa 10 Mio Stück hergestellt, von der Nachfolgegeneration des 256-Kbit-Chips waren es 1989 ganze 60 000. Eine Massenproduktion von Standardbausteinen ist unter diesen Randbedingungen nicht konkurrenzfähig, andererseits leiden die Stand-

orte in Hinblick auf neue Fertigungsinvestitionen unter erheblichen Verkehrs- und Kommunikationsnachteilen.

Siliziummikroelektronik: Wann endet die Zukunft?

Heutigen Abschätzungen zufolge dürfte erst mit dem Gigabitspeicher die Grenze der jetzigen Siliziumtechnologie erreicht sein. Ob ein solcher Chip mit Minimal-

strukturen von etwa 0,15 µm und einer Fläche von rund 450 mm² nicht nur entwickelt, sondern auch wirtschaftlich gefertigt werden kann, steht allerdings auf einem anderen Blatt. Schon heute klaffen die Zeitachsen zwischen der Präsentation neuer DRAM-Generationen, der Marktakzeptanz dieser Speicher und der breiten Umsetzung der Technologie in eine Vielzahl von Logikanwendungen immer weiter auseinander (Bild 2). Dennoch geht das Rennen mit ungebremster Geschwindigkeit weiter, da sich nun Koreaner durch einen vorgezogenen Qualifikationstermin für den 16-Mbit-Speicher zu profilieren suchen. Der dadurch verursachte vorzeitige Preisverfall wird zu einer Welle von Firmenkooperationen führen, um die Kosten wenigstens etwas zu teilen. Die jetzt mit Mustern des 16-Mbit-Chips vorhandene Technologiebasis ist voraussichtlich bis zur Jahrtausendwende für Logikprodukte ausreichend. Falls bis dahin der Gigabitspeicher realisiert sein sollte, werden sicher mindestens weitere zehn Jahre bis zur Ausschöpfung der damit verbundenen Möglichkeiten vergehen. Doch auch danach dürfte Silizium wegen seiner einzigartigen Vorzüge das dominierende Halbleitermaterial bleiben. Es steht praktisch unbegrenzt zur Verfügung, hat gute Wärmeleitfähigkeit, bildet ein Oxid mit ausgezeichneten elektrischen und mechanischen Eigenschaften und verursacht als Abfall keine Umweltprobleme. Andere Materialien können es vermutlich ergänzen, aber wohl kaum verdrängen. Wir werden noch lange im «Siliziumzeitalter» leben.

Weiterführende Literatur

Zum technischen und wirtschaftlichen Umfeld:

WEINERTH H. (Hrsg.): Schlüsseltechnologie Mikroelektronik, Franzis-Verlag, München, 1990 (Elektronik).

FÖLL H., BECKER F. S.: Industrielle Entwicklung und Fertigung von Halbleiterbauelementen; Tagungsband «Festkörperforschung für die Informationstechnik» des 21. IFF, Seiten 16.1 bis 16.79, Institut für Festkörperforschung, KFA Jülich, Bundesrepublik Deutschland, März 1990.

Zur Technologie:

WIDMANN D., MADER H. und FRIEDRICH H.: Technologie hochintegrierter Schaltungen, Springer-Verlag, 1988.

BECKER F. S., FÖLL H. und SCHLÜTER K.: Die Mega-Generation, Microcomputer-Zeitung 12/90, Seite 60.

Zur Geschichte der Mikroelektronik:

ROGERS E. M. und LARSEN J. K.: Silicon Valley Fieber, Siedler Verlag, Berlin, 1985.

QUEISSER H.: Kristallene Krisen, Piper Verlag, München und Zürich, 1985.

Dynamische Speicherbausteine spielen in der Mikroelektronik eine Sonderrolle. Von allen Halbleiterprodukten werden sie nämlich in den höchsten Stückzahlen gefertigt. Zudem repräsentieren sie traditionell die Spitze des technologischen Fortschritts.

Fertigungsprozesse für Speicher der Megageneration

Megabit-Multis fordern die Technologen heraus

Von Frank Stefan Becker

Vor zwei Jahrzehnten machte der 1-Kbit-Speicher Furore. Seither wurde etwa alle drei Jahre eine neue Chipgeneration mit einer um den Faktor 4 gesteigerten Kapazität entwickelt. Bild 1 zeigt die wichtigen Schritte auf diesem Weg für die Generationen seit dem 16-Kbit-Speicher: erste Konferenzberichte, Verfügbarkeit von Kundenmustern und Produktionsaufnahme [1]. Dieser ungebrochene Fortschritt war nur durch stete Weiterentwicklung aller Aspekte der Halbleitertechnologie möglich.

Wesentliche Faktoren zur Steigerung der Leistungsfähigkeit der Chips sind:

- Vergrößerung der Chipfläche
- Verkleinerung der Strukturen
- Stapelung von Elementen übereinander
- Erhöhung der Prozeßkomplexität
- Optimierung der Schaltungsarchitektur

Bild 2 illustriert die ersten beiden Punkte, das heißt die Tendenz zu feineren Strukturen auf immer größeren Flächen [2]. Wie stark hochintegrierte Chips heute auch die dritte Dimension nutzen, wird aus Bild 3 ersichtlich, das einen Schnitt durch das Zellenfeld eines 16-Mbit-DRAMs wiedergibt. Diese «Dynamic Random Access Memories» sind der bei weitem wichtigste Speichertyp und stehen im Zentrum der folgenden Betrachtungen.

Prozeßkomplexität und Chiparchitektur

Zur Herstellung von Speichern der Megageneration sind Hunderte von Arbeitsschritten erforderlich, wobei jede Modifikation mühsam auf etwaige negative Auswirkungen untersucht werden muß. Daraus resultiert ein starker Druck von seiten der Fertigung, bewährte Verfahren so lange wie möglich beizubehalten. Da dies in prinzipiellem Widerspruch zur Dynamik der technischen Weiterentwicklung steht, werden Änderungen schubweise im Zusammenhang mit einer neuen Chipgeneration eingebracht. Charakteristische Innovationen waren beispielsweise der Übergang zur CMOS-Technologie beim 1-Mbit-DRAM, die Einführung einer

dreidimensionalen Speicherzelle beim 4-Mbit-DRAM und die Verwendung von zwei Metallisierungsebenen beim 16-Mbit-DRAM. Tabelle 1 faßt technologische Kerndaten dieser modernsten, als Muster bereits verfügbaren Speicherbausteine zusammen.

Seit dem 4-Mbit-DRAM sind dreidimensionale Speicherzellen Standard in allen Versionen der verschiedenen Hersteller [2]. Dies liegt im Zwang begründet, zur Erhöhung der Integrationsdichte den Platzbedarf der einzelnen Komponenten zu verringern. Besonders gilt dies natürlich für millionenfach vorkommende Elemente wie die jeweils ein Bit speichernden Elementarzellen, die typischerweise über die Hälfte der Chipfläche beanspruchen. Sie bestehen beim DRAM aus der Kombination eines Minikondensators, der die das Bit darstellende Ladung enthält, und dem über gekreuzte Wort- und Bitleitungen angesteuerten Schalttransistor [3]. Traditionell wurde mit jeder Speichergeneration die Kondensatorfläche verringert und parallel dazu auch die Dicke des Dielektrikums zwischen den Kondensatorplatten reduziert. Dies garantierte trotz verkleinerter Fläche die zur Speicherung der Information benötigte Minimalkapazität.

Mit dem 1-Mbit-DRAM gelangte jedoch diese Entwicklung an einen Endpunkt, da bei noch dünneren Dielektrika ein Ladungsverlust durch Tunnelströme befürchtet wurde. Um dennoch die Zelle weiter verkleinern zu können, bot sich als Ausweg an, durch Faltung des Kondensators auch die dritte Dimension zu nutzen. Bei der «Stapelzelle» (stacked) ist der Kondensator über dem Auswahltransistor angeordnet, beim «Grabenkondensator» befindet er sich in einem in das Siliziumsubstrat geätzten Loch (trench). Bild 4 zeigt einen vereinfachten «Stammbaum» [4], der bis zu exotischen Ansätzen wie dem mit mehreren «Flossen» versehenen Fujitsu-Konzept reicht.

Während beim 4-Mbit-DRAM Anfang 1989 noch der Grabenkondensator überwog [2], dominiert beim 16-Mbit-DRAM, wie Tabelle 1 zeigt, das Stapelprinzip. Dafür sind verschiedene Gründe maßgebend. Einerseits erwies sich zum Beispiel die beim Stapelkondensator erforderliche Erzeugung eines zuverlässigen Dielektrikums zwischen rauen Elektroden aus abgeschiedenem, polykristallinem Silizium als doch nicht so problematisch, während die Ätzung eines tiefen Grabens mit perfekter Form einen unerwarteten Aufwand erforderte. Weitere Probleme des Grabenkonzepts sind mögliche Rückstände im Graben nach der Reinigung sowie erhöhte Prozeßschrittzahl und geringere

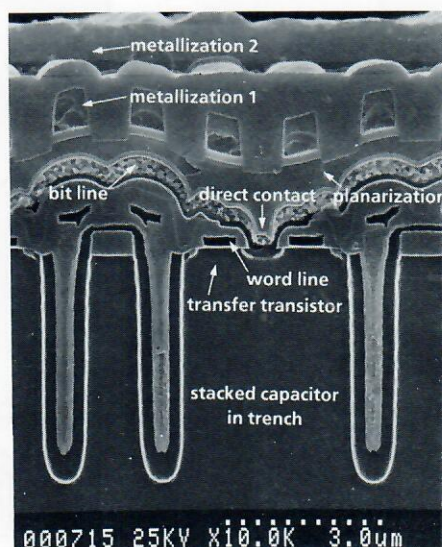


Bild 3. Schnitt durch das Zellenfeld des Siemens-16-Mbit-DRAM. Man erkennt die Dreidimensionalität der Strukturen. In jedem Graben wird ein Bit gespeichert.

Variationsmöglichkeiten im Vergleich zum Stapelprinzip. Dieses bedingt dafür erhebliche Höhenunterschiede auf dem Chip, die der Fototechnik Schwierigkeiten bereiten. Neueste Ansätze gehen inzwischen davon aus, durch bewußte Erzeugung von Körnigkeit auf der unteren Elektrode deren Oberfläche und damit die des Kondensators zu vergrößern [5] oder durch Verwendung anderer Isolatoren mit einer erhöhten Dielektrizitätskonstante die Kondensatorkapazität zu steigern [6].

Auch die Mehrlagenmetallisierung entspringt dem Wunsch, die lateralen Abmessungen zu verringern, ohne jedoch die Leiterbahnen auf den immer größeren Chips zu lang werden zu lassen. Längere Bahnen bedeuten größere Signallaufzeiten und erhöhte Eigenkapazitäten, die wiederum die Anforderung an die minimale Zellkapazität hochschrauben. Bei jetzt schon über 100 000 A/cm² Stromdichte ist auch eine Verringerung des Querschnitts der Aluminiumleitungen kaum mehr möglich, da sonst Phänomene wie Elektromigration zu deren Zerstörung führen können. Während mehrere Lagen aus leitendem polykristallinem Silizium, zum Teil in Kombination mit niederohmigen Siliziden, bisher schon Stand der Technik waren, bringt die Realisierung von zwei übereinandergelegten Netzen aus Aluminium eine Reihe von neuen Problemen mit sich. Besonders gravierend ist die durch dieses Metall gegebene Beschränkung auf Prozeßtemperaturen von unter 450 °C ab der ersten Lage, die viele Standardprozesse der Schichterzeugung und Planarisierung ausschließt [7]. Beim 16-Mbit-DRAM dürfte die erzielbare Endausbeute ganz entscheidend von der Beherrschung der mit der Mehrlagenmetallisierung verbundenen Schwierigkeiten abhängen. In den folgenden Abschnitten werden die wichtigsten Einzelprozesse diskutiert, die, periodisch wiederholt, zur Chipherstellung nötig sind.

Fototechnik

Durch die Tendenz zu immer kleineren Strukturen ist die zu deren Erzeugung benötigte Fotolithografie zur Schlüsseltechnologie geworden. Die erzielbare minimale Auflösung R wird bestimmt durch die Wellenlänge λ des verwendeten Lichts und die numerische Apertur NA des Abbildungssystems gemäß

$$R = 0,8 \frac{\lambda}{NA}$$

wobei der Faktor 0,8 ein aus der Praxis gewonnener Erfahrungswert ist. Sowohl

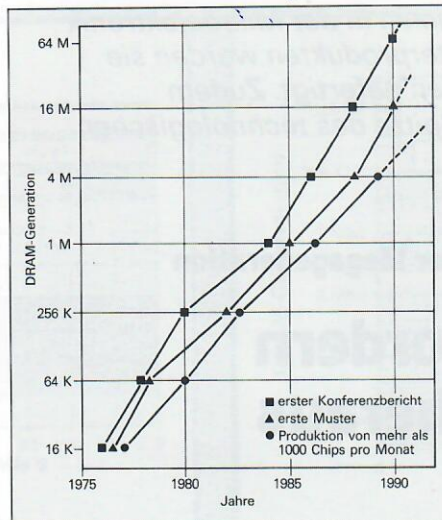


Bild 1. Zeitliche Abfolge wichtiger Schritte bei der Entwicklung neuer DRAM-Generationen. Der tendenziell größer werdende Zeitraum zwischen ersten Konferenzberichten und der Produktionsaufnahme ist unter anderem auf den verstärkten Einsatz von Elektronenstrahlolithografie in der Anfangsphase zurückzuführen (aktualisiert nach [1]).

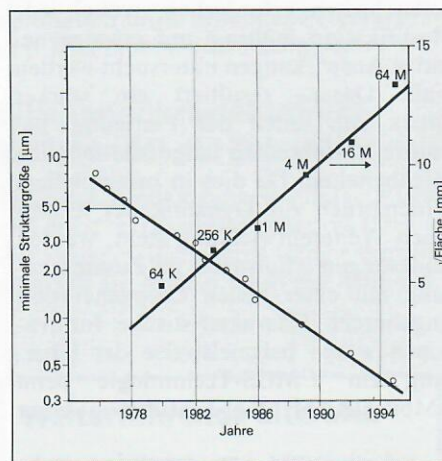


Bild 2. Entwicklung der minimalen Strukturgröße und der Chipfläche für verschiedene DRAM-Generationen. Immer größer werdende Chips tragen immer feinere Strukturen (Quelle: [2]).

durch Verkleinerung von λ als auch durch Vergrößerung von NA läßt sich die Auflösung verbessern. Da eine Änderung der Wellenlänge jedoch auch eine parallele Neuentwicklung bei den verwendeten lichtempfindlichen Fotolacken erfordert, zogen es bis zum 4-Mbit-Speicher die Halbleiterhersteller vor, die etablierte «g-Linie» der Quecksilber-Hochdrucklampe (436 nm) beizubehalten [2]. Dies bedingte allerdings Stepper mit immer größerer NA , das heißt mit aufwendigen Linsensystemen, die wiederum gemäß der Beziehung

$$F = \frac{\lambda}{(NA)^2}$$

nur einen verringerten Tiefenschärfebereich F aufweisen.

Als «Stepper» werden die modernen Belichtungssysteme bezeichnet, die im Gegensatz zu früher nicht mehr eine ganze Scheibe auf einmal belichten, sondern schrittweise jeweils nur wenige Chips in den Fotolack projizieren. Dieses Verfahren dauert zwar länger, mit dem alten System ließe sich jedoch bei den heutigen Scheibendurchmessern von 15 cm und Strukturbreiten unter 1 µm niemals die notwendige Exaktheit erreichen.

Nachdem beim 4-Mbit-DRAM die Grenzen der g-Linien-Lithografie erreicht wurden, bringt der 16-Mbit-Speicher den Übergang zu der mit 365 nm kürzeren i-Linie als Lichtquelle (Tabelle 1). Gemäß dem derzeitigen Diskussionsstand in der Fachwelt zeichnen sich sogar gute Chancen ab, auch die 64- und die 256-Mbit-Speicher noch mit Lichtoptik strukturieren zu können. Dabei hofft man, durch Einsatz von im tiefen UV-Bereich (200 bis 250 nm) emittierenden Excimer-Lasern und der Verwendung von «Phasenverschiebungsmasken» [8] noch Strukturen im 0,2-µm-Bereich herstellen zu können, ohne, wie früher prognostiziert, die optische Lithografie zugunsten neuer Techniken wie etwa der Röntgenlithografie verlassen zu müssen.

Phasenverschiebungsmasken beruhen auf dem Prinzip, durch das Aufbringen spezieller Schichten in Teilen der transparenten Maskenbereiche das dort durchgehende Licht um 180 °C in seiner Phase zu verschieben. Bei regelmäßigen Strukturen kann so über Interferenz mit benachbarten, nicht verschobenen Strahlen die Auslöschung störenden Streulichts erreicht werden. Diese Technik erfordert allerdings Negativfotolacke statt der bisher verwendeten Positivlacke sowie erheblich aufwendigere, in ihrer Beschichtung für die jeweiligen lokalen Strukturverhältnisse optimierte Masken. Unklar ist auch, inwiefern zukünftige Stepper die Anforderungen an Größe und gleichmäßige Ausleuchtung des Bildfeldes erfüllen können, die Chips mit Flächen von 300 bis 400 cm² stellen werden.

In engem Zusammenhang mit der Belichtungstechnik steht die Entwicklung der Lacktechnik. Hochempfindliche Fotolacke sind eine Grundvoraussetzung zur Nutzung der Auflösungskapazität moderner Stepper. Solche Lacke sind jedoch andererseits auch sehr anfällig für alle Arten von Sekundärlicht, das zum Beispiel von Stufen oder den verschiedenen optisch transparenten Schichten der Unterlage stammen kann. Die Folgen können Variationen in der Strukturbreite oder durch stehende Wellen ver-

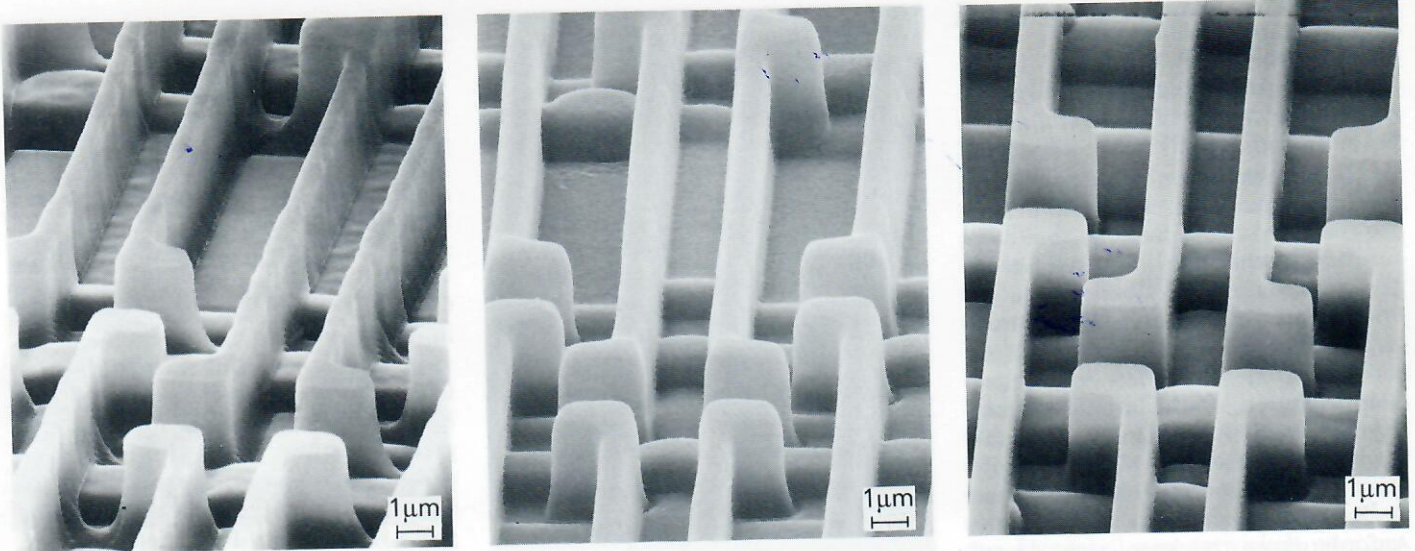
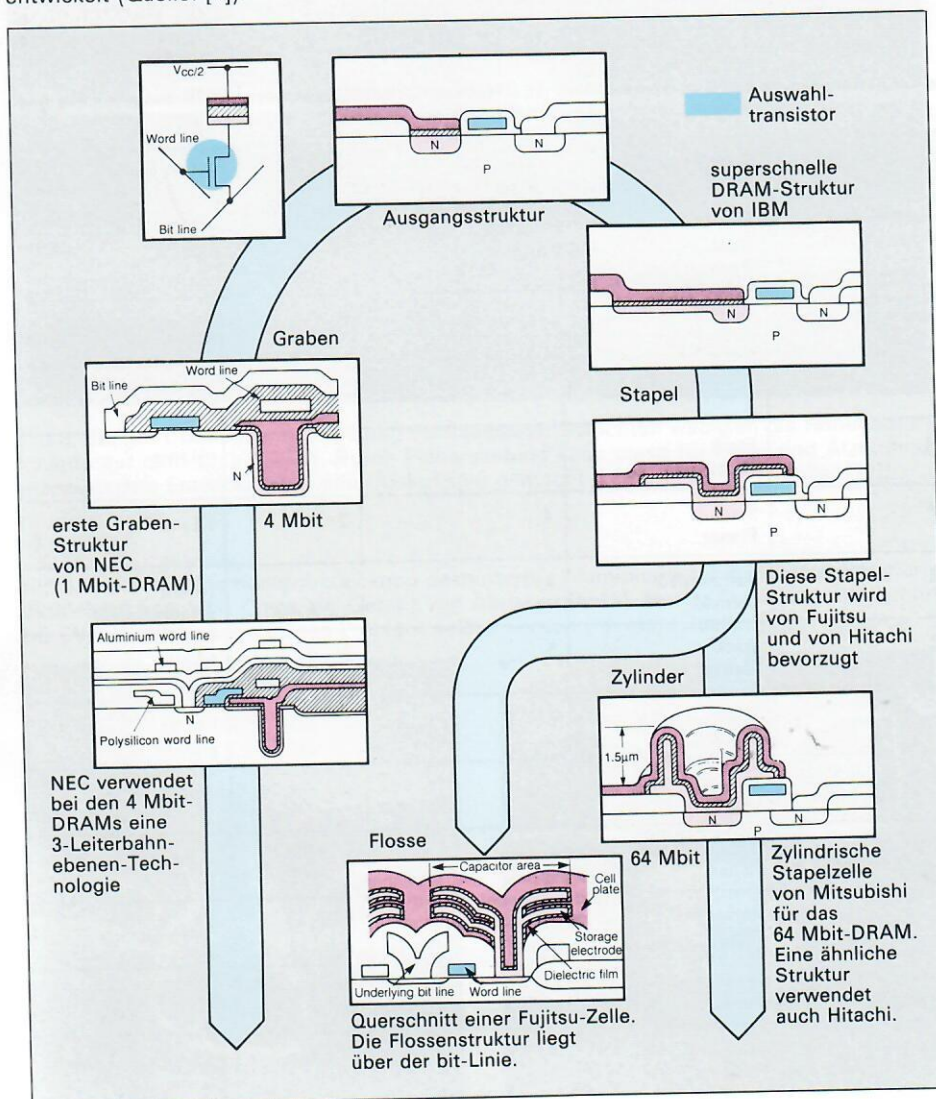


Bild 5. Bei hochempfindlichen Lacken kann Streulicht die Strukturtreue beeinträchtigen (links). Gefärbter Lack reduziert den Effekt (Mitte), das beste Ergebnis wird jedoch mit einer Antireflexschicht unter dem Fotolack erzielt (rechts) (Quelle: [2]).

Bild 4. «Stammbaum» der dreidimensionalen DRAM-Speicherzelle. Ausgehend vom ursprünglich planaren Ansatz, wurde sowohl das Grabenkonzept als auch die Stapelzelle entwickelt (Quelle: [4]).



ursachte Modulationen der Lackdicke sein. Manche Probleme sind durch den Einsatz gefärbter Lacke verringerbare, andererseits wird so die Auflösung im Bereich größerer Lackdicken herabgesetzt. Bessere Ergebnisse lassen sich mit Antireflexschichten unter dem Lack erreichen, wie der Vergleich in Bild 5 zeigt [2]. Der Preis ist allerdings eine erhöhte Prozeßkomplexität. Weitere im Rahmen der Weiterentwicklung der Fototechnik zu lösende Probleme sind beispielsweise die Toleranzen bei der Justierung von bis zu 20 übereinanderliegenden Maskebenen, die Abhängigkeit des Linsensystems von Druck und Temperatur sowie die routinemäßige Kontrolle der erzeugten Lackstrukturen.

Seit vielen Jahren schon wird die *Röntgenlithografie* als wahrscheinlichste Nachfolgetechnik der Lichtoptik gehandelt, der Einsatz in der Massenfertigung wurde bisher jedoch stets hinausgeschoben. Der Grund liegt in den fundamentalen Änderungen, die ihre Anwendung in der Chipherstellung bedingen würde. Hauptprobleme sind dabei Materialfragen und die hohen Kosten für die Strahlenquelle. Am besten läßt sich parallele, abstimmbare Röntgenstrahlung in einem Synchrotron erzeugen [9], dessen Preis allerdings in der Größenordnung von 45 Mio Franken liegt. Auch wenn damit dann rund zehn Stepper gleichzeitig betrieben werden können, so ist doch die Schwelle für den Einstieg in diese Technik für die meisten Anwender zu hoch. Von verschiedenen Partnern gemeinsam genutzte «Lithozentren» würden andererseits einen so häufigen Scheibentransport von und zu den je-

weiligen Prozeßlinien bedingen, daß eine Beeinträchtigung der Ausbeute kaum vermeidbar erscheint.

Einen zweiten Problemkreis bilden die Masken und speziell die Materialien [9]. Da bisher keine Optiken zur Bündelung der Röntgenstrahlen verfügbar sind, müßte von dem jetzigen Prinzip der um den Faktor 5 verkleinerten Abbildung der Maskenstrukturen auf den Chip abgewichen werden. Dies bedeutet in der Praxis, daß die Masken in ihrer Strukturgröße von derzeit minimal etwa 3 µm auf die Dimension der zukünftigen Chipstrukturen, das heißt unter 0,3 µm, reduziert werden müßten. Für die mit der Herstellung solcher Masken und ihrer hundertprozentigen Kontrolle auf absolute Fehlerfreiheit verbundenen Anforderungen zeichnen sich derzeit noch keine überzeugenden Lösungen ab.

Besondere Schwierigkeiten bereiten dabei die möglichen Maskenmaterialien. Während in der Lichtoptik Chrom auf Quarz eine Kombination von dünnem, undurchlässigem Absorber auf stabilem, transparentem Substrat ermöglicht, sieht die Situation für Röntgenstrahlen gänzlich anders aus. Mögliche Absorbermaterialien wie Gold oder Wolfram

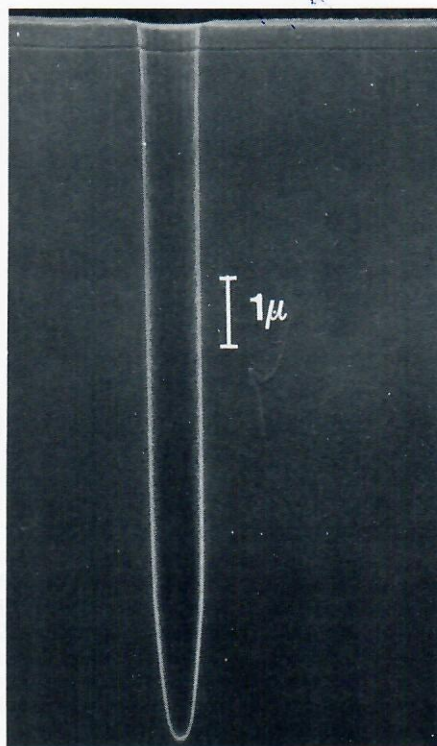


Bild 7. Beispiel für die extremen Aspektverhältnisse, die sich mit moderner Plasmaätzung erreichen lassen (Quelle: [2]).

müßten relativ dick, das heißt im Mikrometerbereich, auf Unterlagen wie Bornitrid, Siliziumkarbid oder Beryllium aufgetragen werden, die ihrerseits zur Erfüllung der Forderung nach Transparenz nicht wesentlich stärker sein dürfen. Daß solche empfindlichen Strukturen unter dem Bombardement mit Röntgenstrahlen leiden können, ist verständlich, aber nicht tolerabel, da bei Chipplängen von 2 cm bereits ein Verzug in der Größenordnung von einem tausendstel Prozent «tödlich» wäre.

Diesen Problemen steht andererseits eine Reihe von Vorzügen der Röntgenlithografie gegenüber. Dank des Fehlens eines Abbildungssystems existieren keine Begrenzungen bezüglich der Bildfeldgröße oder der Fokustiefe. Kleine, schwer zu eliminierende Partikel werden zumeist einfach durchstrahlt, und Streustrahlung ist nicht zu befürchten. Da allerdings auch diese Technik nur bis Strukturgrößen von rund 0,1 µm reicht, wird ihr Einsatz entscheidend davon abhängen, inwieweit sich ein 256-Mbit-Speicher noch lichtoptisch herstellen läßt und wie sich dann die Situation der Siliziummassenprodukte bezüglich Ausbeute und Marktchancen darstellt.

Kenndaten von 16-Mbit-DRAMs. Die Daten wurden aus Veröffentlichungen und Konferenzberichten (IEDM, ISSCC, SSDM) zusammengestellt und geben den Ende 1990 verfügbaren Stand wieder.

Hersteller	Chipgröße mm ²	Zellengröße µm ²	Kondensatortyp	Leiterbahnebenen		Lithografie
				Poly-Si, Silizid	Aluminium	
Siemens	142	4,8	Stack in Trench	3	2	g-Linie
IBM	141	4,13	Inverted Trench	1	2 W 1 Al	i-Linie
Texas Instruments	133	4,33	Trench	3	2	i-Linie g-Linie
Fujitsu	138	4,64	Stack Planar	4	2	i-Linie
Hitachi	127	4,35	Stack Planar	4	2	i-Linie
Matsushita	142	4,48	Surrounded Trench	4	2	i-Linie g-Linie
Mitsubishi	133	3,92	Stack Planar	5	2	i-Linie
NEC	~ 125	4,05	Stack Planar	4	2	i-Linie g-Linie
Oki	125	4,06	Stack in Trench	4	2	i-Linie
Samsung	135	4,39	Stack Planar	4	2	i-Linie
Sanyo	135	4,5	Stack Planar	4	2	
Sharp	128	4,5	Stack Planar	6	2	
Toshiba	134	4,62	Stack in Trench	3	2	i-Linie g-Linie

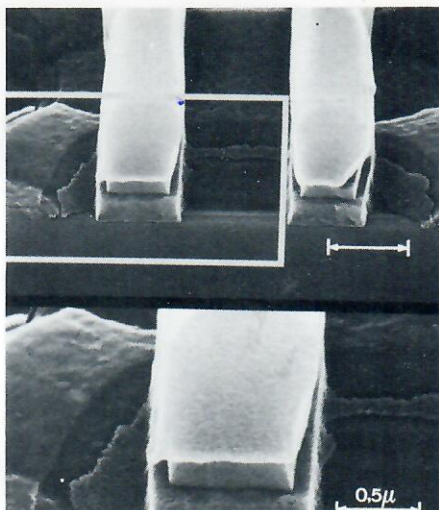


Bild 6. Bei stark anisotroper, das heißt gerichteter Plasmaätzung besteht die Gefahr, daß am Fuß von Kanten Reste zurückbleiben. Im oberen Bild entspricht die Balkenlänge 1 μm (Quelle: [2]).

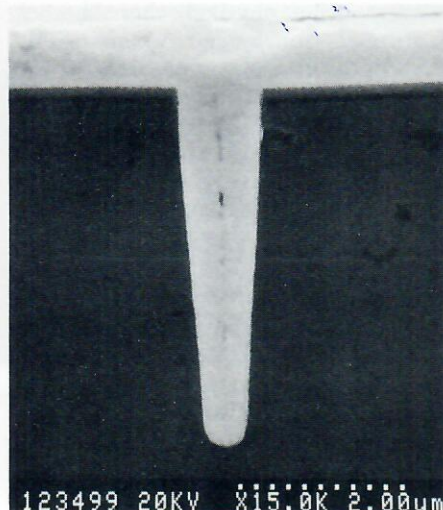


Bild 10. Bei einer optimierten Wolfram-CVD-Abscheidung lassen sich auch tiefe Testgräben mit guter Kantenbedeckung auffüllen (H. Körner, Siemens).

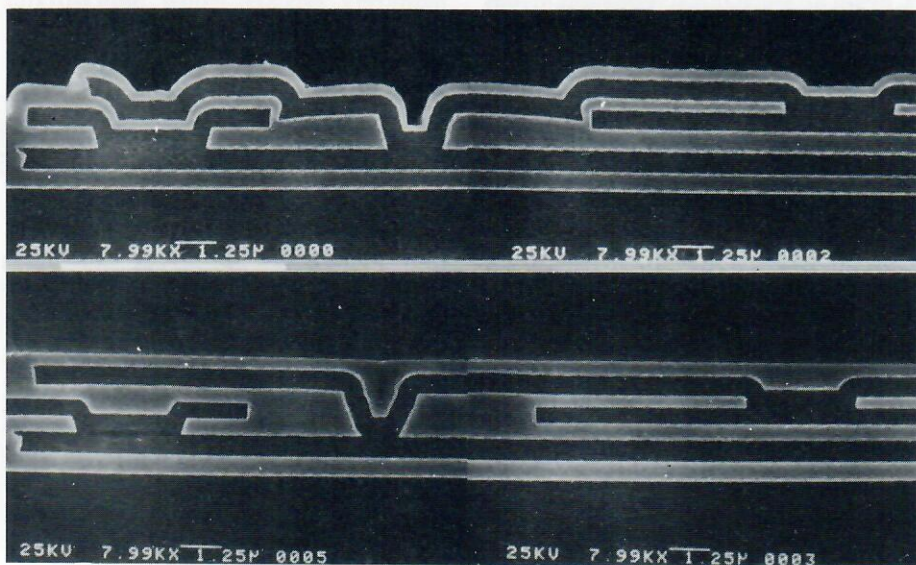
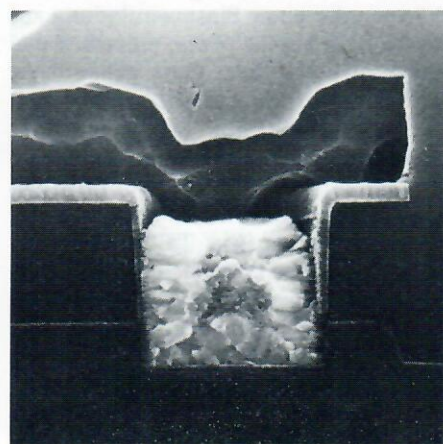
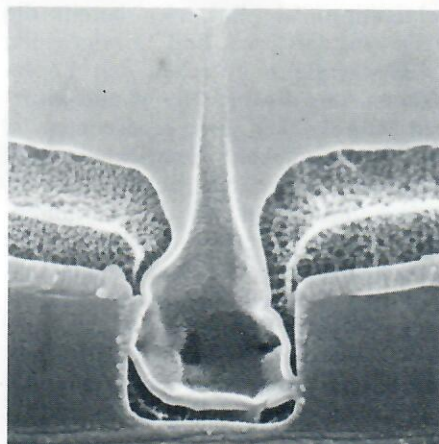


Bild 8. Bei der Übereinanderstapelung verschiedener Schichten wachsen die Höhenunterschiede auf dem Chip (oben). Durch Planarisierung kann diese für Foto- und Ätztechnik unerwünschte Entwicklung wieder rückgängig gemacht werden (unten) (Quelle: [11]).

Bild 9. Die schlechte Kantenbedeckung gesputterten Aluminiums führt in den engen Kontaktlöchern von Mbit-Chips zur Gefahr von Abrissen (links). Durch Auffüllung des Lochs mit CVD-Wolfram kann dieses Problem gelöst werden (rechts) (Quelle: [2]).



Ätztechnik

Die Ätztechnik überträgt die von der Fototechnik erzeugten Lackstrukturen in die darunterliegende Schicht. Dieser Schritt ist jedoch im Gegensatz zur Lackstrukturierung meist irreversibel und zusätzlich sehr komplex, da das Ergebnis von vielen Faktoren abhängt. Die zu ätzende Schichtfolge spielt dabei ebenso eine Rolle wie die Chiptopographie und die Art beziehungsweise die Vorbehandlung des Fotolacks. Der Idealprozeß sollte gleichzeitig schnell, reproduzierbar, selektiv, oberflächenschonend und anisotrop, das heißt gerichtet sein. Da in der Realität jedoch ein Konflikt zwischen diesen Anforderungen besteht, muß für jede Situation der Ätzprozeß neu entwickelt oder zumindest stark modifiziert werden.

Die beste Selektivität läßt sich naturgemäß mit chemisch kontrollierten Prozessen erreichen, wie sie die Naßätzung bietet. Besonders die in unterschiedlichen Konzentrationsgraden verwendete Flußsäure entfernt zum Beispiel das immer wieder vorkommende Siliziumdioxid rückstandsfrei und oberflächenschonend von Silizium.

Während früher Naßprozesse auch zur Strukturierung verwendet wurden, ist das heute nicht mehr möglich, da bei mit der Schichtdicke vergleichbaren Strukturbreiten ein isotroper seitlicher Abtrag nicht toleriert werden kann. Jedoch werden auch zur Herstellung hochintegrierter Schaltungen noch viele Naßprozesse eingesetzt, allerdings jetzt zur Entfernung nicht mehr benötigter Schichten beziehungsweise in der Reinigung. Das Standardverfahren zur Strukturübertragung ist heute die Plasmaätzung, bei der eine elektrische Entladung Ionen erzeugt, die dann auf die Siliziumscheibe beschleunigt werden und so einen gerichteten Abtrag bewirken. Wichtige Randbedingungen sind dabei gute Selektivität gegenüber der Ätzmaske und der Unterlage, hohe und gleichförmige Ätzrate über die Scheibe sowie die Vermeidung von Oberflächenschäden, Partikeln und Rückständen. Diese bleiben, wie Bild 6 zeigt, bei anisotroper Ätzung bevorzugt am Fuß von Kanten zurück, wo sie zum Beispiel im Fall eines leitenden Materials Kurzschlüsse verursachen können. Zu ihrer Vermeidung kann der Ätzprozeß so konzipiert werden, daß zum Schluß eine mehr isotrope Überätzphase Rückstände eliminiert. Dies ist durch Variation der Parameter der Plasmaentladung oder durch Verwendung anderer Ätzgase möglich. Generell sind zur Erzielung hoher Anisotropie stark physikalisch dominierte Prozesse mit chemisch etwas weniger reaktiven Brom-

verbindungen geeignet, während die aggressiveren Fluor- oder Chlorverbindungen einen mehr chemisch kontrollierten, das heißt materialelektiveren Abtrag bewirken. Daneben spielen auch bei der Ätzung entstehende Ablagerungen eine Rolle, die sich beispielsweise auf die Rauigkeit der Wände oder die Form eines Grabens auswirken können [3]. Bild 7 verdeutlicht die extremen Möglichkeiten moderner Plasmaätzung, eine Übersicht über verschiedene Verfahren findet sich beispielsweise in [10].

Oxidation

Zu Beginn der Planartechnik war die «lokale Oxidation» (LOCOS) der Standardprozeß zur Erzeugung des einzelnen aktiven Komponenten voneinander isolierenden dicken «Feldoxids». Das zweite traditionelle Einsatzgebiet sind dünne Oxide, wie sie für Transistor-Gates oder Kondensatordielektrika benötigt werden. Dank steter Prozeßverbesserung ist die Oxidation auch noch bei der 16-Mbit-Generation in der Lage, die gestellten Anforderungen zu erfüllen. Gegenüber allen Alternativen zeichnet sie sich durch eine überlegene Qualität sowohl der dünnen Schichten als auch der Grenzfläche Oxid-Silizium aus. Durchgeführt wird der Prozeß, indem man bei 900 bis 1000 °C Wasserdampf oder Sauerstoff über die sich in großen Quarzrohren befindenden Scheiben leitet, wobei zum Teil auch HCl zur Verbesserung der Oxidqualität eingesetzt wird. Wichtig ist dabei die Verhinderung der Rückdiffusion von Luft in den Ofen, die bei Beladung oder Entladung der Scheiben auftreten kann. Die Verwendung von Senkrecht- statt der traditionellen Waagrechtöfen ist hier ein vielversprechender Ansatz.

Planarisierung und Schichtherstellung

Da zur Erhöhung der Integrationsdichte immer mehr Lagen übereinandergestapelt werden, nimmt die Bedeutung der Schichttechnik zu. Typische Anforderungen sind hier gute Gleichmäßigkeit bei hoher Abscheiderate, geringe Partikeldichte, hohe Prozeßstabilität und eine möglichst weitgehende Bedeckung von Stufen. Da mit der Zahl der Schichten auch die Höhenunterschiede auf dem Chip zunehmen, muß wegen der beschränkten Fokustiefe der Fototechnik, der Gefahr von Ätzresten an Kanten und der schlechten Stufenbedek-

kung mancher physikalischer Abscheidungsprozesse zunehmend «planarisiert» werden. Dazu wurde eine Vielzahl von Verfahren entwickelt [7]. In Bild 3 befindet sich zum Beispiel unter der Bitlinie und der ersten Metallisierung jeweils eine Schicht von «Borphosphorsilikatglas». Dabei handelt es sich um ein «ternäres Glas», das aus einer genau definierten Mischung der Oxide von Silizium, Bor und Phosphor besteht. Diese Zusätze bewirken gegenüber reinem SiO_2 eine Herabsetzung des Schmelzpunktes, so daß ein nach der Abscheidung bei etwa 900 °C durchgeführter «Verfließschritt» die Konturen glättet. Bei anderen Verfahren wird die einebnende Wirkung von Fotolack genutzt, um in Kombination mit einer dafür optimierten Rückätzung eine ebene Oberfläche zu schaffen. Bild 8 gibt ein Beispiel für das Potential dieser Methode [11]. Ein anderer Ansatz ist das «Chemical-Mechanical Polishing», das durch eine Kombination von mechanischem und chemischem Abtrag die Chipoberfläche auch in fortgeschrittenem Prozeßstadium einebnet [12]. Unter den physikalischen Abscheidemethoden ist das «Aufstäuben», englisch «Sputtern», die wichtigste. Dabei wird ein aus dem Zielmaterial bestehendes «Target» gegenüber der Scheibe angebracht und so mit energiereichen Ionen beschossen, daß sich abgesputtertes Material auf der Scheibenoberfläche niederschlägt. Das Verfahren wird besonders für Elemente wie Aluminium eingesetzt, die keine einfach zu handhabenden Verbindungen bilden. Vorteile sind die Vermeidung aggressiver oder instabiler Gase, dafür sind Partikel und die dem Verfahren inhärente schlechte Stufenbedeckung gewichtige Nachteile. In Bild 9 links wird dies am Beispiel von Aluminium in einem Kontaktloch deutlich [2]. Zunehmend wird deshalb auch bei der Metallisierung versucht, mit «Chemical Vapour Deposition» (CVD) zu arbeiten. Dabei werden Gase in einem Reaktor über die Scheiben geleitet und durch Energie in Form von Wärme oder einer Plasmaentladung zur Reaktion gebracht. Zwar müssen dabei zum Teil gefährliche Gase gehandhabt werden, dafür können aber viele Scheiben gleichzeitig mit Schichten guter Gleichmäßigkeit und Stufenbedeckung überzogen werden. Aktuelle Entwicklungsgebiete sind die Abscheidung von neuen Schichten wie Ta_2O_5 als Kondensatordielektrikum [6] oder von Wolfram als Ergänzung der Aluminiummetallisierung [13]. Bild 9 rechts zeigt die mögliche Verbesserung, wenn das Kontaktloch vor dem Aluminiumsputtern mit CVD-Wolfram gefüllt wird. Wie aus Bild 10 ersichtlich, ist dieser Prozeß



Bild 11. Mikrokokken auf Leiterbahnen eines 4-Mbit-DRAM (H. Haudek, Siemens).

sogar in der Lage, tiefe Testgräben zu füllen. Nachteile des Verfahrens sind die Rauigkeit der Schichten und hohe Kosten.

Auf breiter Front wird die CVD-Technik zur Erzeugung der vielen auf einem Chip benötigten Hilfs- und Isolationschichten sowie für die Abscheidung des polykristallinen Siliziums eingesetzt. Bei den Isolatoren handelt es sich um Siliziumnitrid und vor allem um verschiedene Modifikationen von SiO_2 , wobei zunehmend organische Siliziumverbindungen als Ausgangsstoffe dienen [14]. Tabellarische Charakterisierungen wichtiger CVD-Verfahren finden sich zum Beispiel in [7] und in [10].

Analytik und Defekterkennung

Bei der Feinheit modernster Chipstrukturen können auch kleinste Luftverunreinigungen wie Bakterien bereits die Schaltung zerstören. Bild 11 zeigt einen Größenvergleich zwischen den Leiterbahnen eines 4-Mbit-DRAM und winzigen Mikrokokken. Eine Konsequenz ist die Notwendigkeit, Entwicklung und Fertigung integrierter Schaltungen in Reinräumen (Bild 12) durchzuführen. Der Hauptteil der Defekte stammt heute jedoch von der Scheibehandhabung und den Herstellprozessen selber. So können sich beispielsweise bei der Schichttechnik Partikel von den Reaktorwänden lösen, oder nach einer Ätzung können Rückstände bleiben (Bild 6). Eine immer aufwendigere prozeßbe-

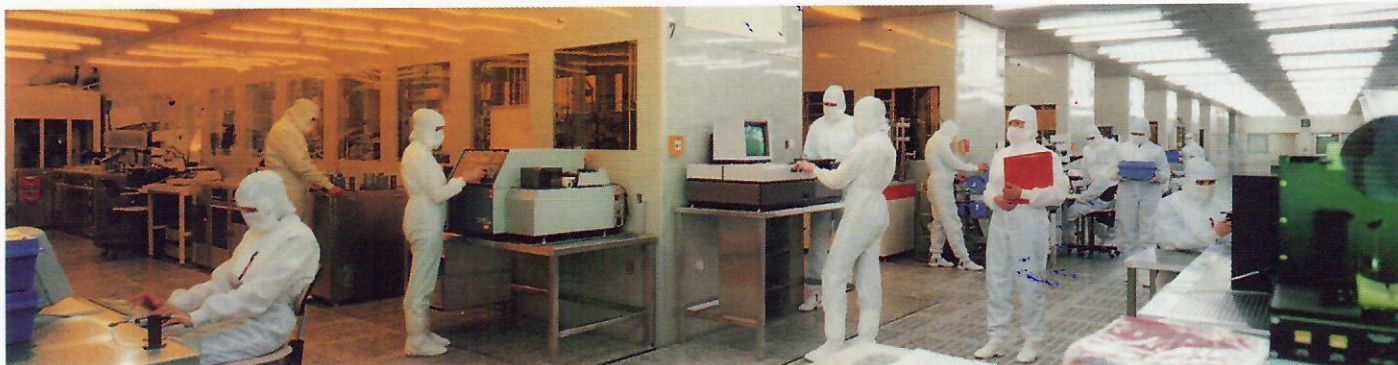


Bild 12. Blick in einen modernen Reinraum. Alle Mitarbeiter tragen Spezialkleidung zur Vermeidung von Partikelemissionen (Siemens).

gleitende Analytik ist deshalb essentiell als rasches «Frühwarnsystem». Dazu gehören Methoden wie Auger-Elektronenspektroskopie, Sekundärionen-Massenspektroskopie und Rasterelektronenmikroskopie, wobei letzteres Verfahren inzwischen auch für Routinekontrollen kritischer Dimensionen benötigt wird. Spezielle Präparationstechniken sind für die Untersuchung dreidimensionaler Strukturen wie der DRAM-Kondensatorgräben erforderlich [2].

Mit abnehmender Größe der für die Schaltung «tödlichen» Defekte wächst die Schwierigkeit, diese überhaupt noch zu finden. Dies gilt besonders für Partikel unterhalb eines halben Mikrometers, da dann der Streuquerschnitt des zur Detektierung üblicherweise benutzten Lichts stark abfällt. Bei heute etablierten Standardverfahren wird meist ein Laserstrahl über die Scheibe geführt und das Streulicht mit einem Fotodetektor erfaßt. Da das System nicht zwischen Defekten und normalen Strukturen unterscheidet, können so nur blanke

Testscheiben gemessen werden. Versuche, über polarisiertes Licht oder die Schaffung eines die regelmäßigen Strukturen ausblendenden Fourierfilters diese Einschränkung aufzuheben, reduzieren entweder die Empfindlichkeit oder treiben den Zeitaufwand in die Höhe [2]. Verfahren, wie zukünftige Multi-Megabit-Chips auf Defekte von $0,1 \mu\text{m}$ Durchmesser schnell und zerstörungsfrei kontrolliert werden, existieren bisher nicht. Sie werden aber bei der Frage, wie weit die Integrationsdichte erhöht werden kann, eine wichtige Rolle spielen und damit die wirtschaftlichen Grenzen der Siliziumtechnologie entscheidend mitbestimmen [15]. [72] ®

Literatur

- 1 Nikkei Microdevices 9/1988.
- 2 Becker F. S.: From 4-Mbit to 16-Mbit-DRAM Technology – Present Problems and Coming Challenges; Proceedings IV SB MICRO, Porto Alegre (Brasilien), S. 125, 1989.
- 3 Becker F. S., Föll H., Schlüter K.: Die Megageneration; Microcomputer-Zeitung 12, S. 60, 1990.
- 4 Electronic Engineering 7/1990.
- 5 Watanabe H. et al.: A New Stacked Capacitor Structure Using Hemispherical-Grain Poly-Silicon Electrodes; Ext. Abstr. 22nd SSDM Conference, Sendai/Japan, S. 873, 1990.
- 6 Shinriki H. et al.: Promising Storage Capacitor Structures with Thin Ta_2O_5 Film for Low-Power High-Density DRAMs; IEEE Trans. Electr. Dev. 37 (9), S. 1939, 1990.
- 7 Becker F. S.: Interlevel Dielectrics for Reduced Thermal Processing; in «Reduced Thermal Processing for ULSI», R. A. Levy (ed.), Plenum Publ. Corp., New York, S. 355, 1989.
- 8 Z. B. Jinbo H., Yamashita Y.: $0,2 \mu\text{m}$ or Less i-Line Lithography by Phase-Shifting-Mask Technology; IEDM Technol. Dig., S. 825, 1990.
- 9 Peters D. W., Frankel R. D.: X-Ray Lithography: The Promise of the Past and Reality of the Present; Sol. State Techn., Nr. 3, S. 77, 1989.
- 10 Föll H., Becker F. S.: Industrielle Entwicklung und Fertigung von Halbleiterbauelementen; in Tagungsband «Festkörperforschung für die Informationstechnik» des 21. IFF, Institut für Festkörperforschung, KFA Jülich, S. 16.1, 1990.
- 11 Mitsuhashi K. et al.: Etch-Back Planarization Technology for Interconnection of Stacked Structure; in «Proc. 1st Int. Symp. ULSI Science and Technology», Broydo S., Osburn C. M. (eds.), Electrochem. Soc. Proc., Vol. 87 (11), S. 557, 1987.
- 12 Davari B. et al.: A New Planarization Technique Using a Combination of RIE and Chemical-Mechanical Polish; IEDM Technol. Dig., S. 61, 1989.
- 13 Hieber K., Körner H., Treichel H.: Chemical Vapour Deposition of Oxide and Metal Films for VLSI-Applications; Thin Sol. Films 181, S. 75, 1989.
- 14 Becker F. S., Treichel H., Röhl S.: Low-Pressure Deposition of Doped SiO_2 by Pyrolysis of Tetraethylorthosilicate (TEOS), II. Arsenic Doped Films; J. Electrochem. Soc. 136, S. 3033, 1989.
- 15 Becker F. S., Föll H.: Gigabit-Chips zur Jahrtausendwende?; Technische Rundschau Nr. 12, S. 26, 1991.

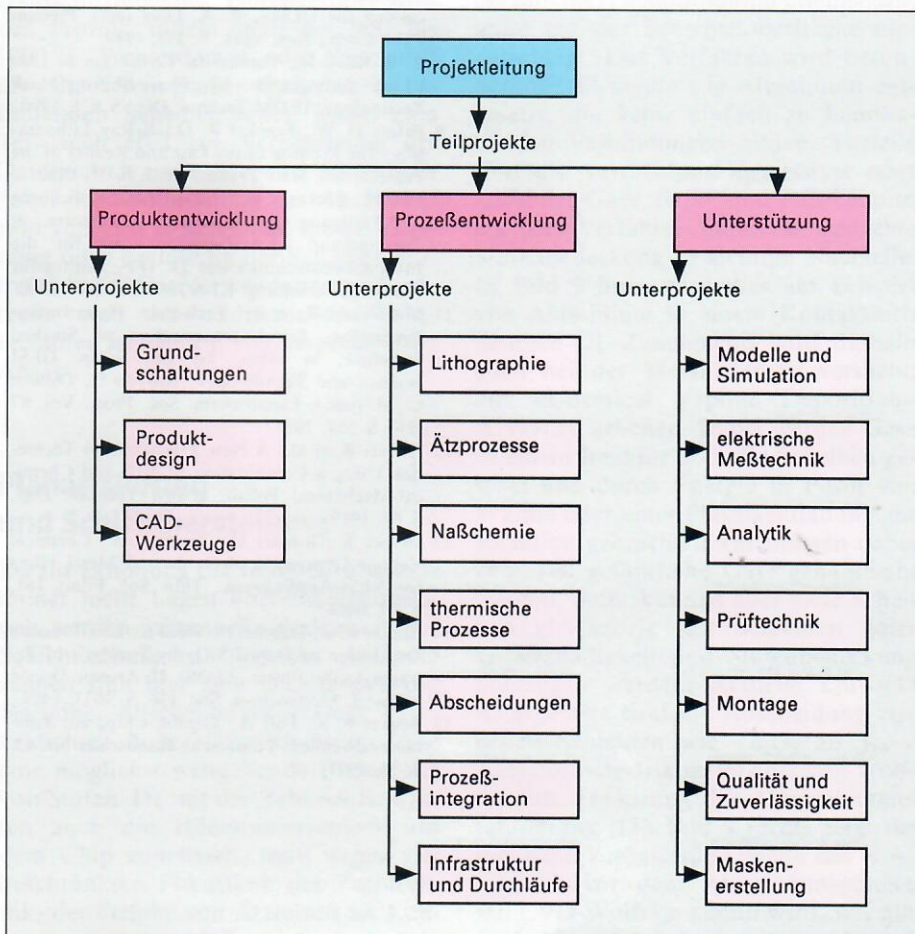
Beim Speicherwettbewerb sind stets mehrere Mannschaften gleichzeitig im Rennen: Während die Produktion der 1-Mbit-Speicher zurzeit weltweit auf Hochtouren läuft, werden jetzt die Prozeßlinien für die 4-Mbit-Chips hochgefahren, und die Wissenschaftler in den Labors bereiten parallel dazu bereits die nächste und die übernächste Speichergeneration vor. Die Investitionen sind enorm. Bevor mit dem 16-Mbit-Speicher der erste Franken Umsatz erzielt werden kann, müssen mehr als 1 Mia Franken in Entwicklung und Fertigung investiert werden. Und ob die Megabitmultis den europäischen Herstellern dereinst Gewinne einfahren, ist alles andere als gewiß, denn zwischen Gewinn und massiven Verlusten liegen lediglich einige Monate Abstand zu den ostasiatischen Marktleadern.

Megabitspeicher: vom Projekt zum Produkt und auf den Markt

Das kleine Delta zwischen Gewinn und Verlust

Von Helmut Föll und Frank Stefan Becker

Bild 1. Projektstrukturplan für die Entwicklung von Speichern in Anlehnung an das JESSI-Projekt «Joint Memory Project».



Während jetzt weltweit die Fertigung des 1-Mbit-DRAM-Speicherchips ihren Höhepunkt erreicht – bei Siemens zum Beispiel mit über 4 Mio Stück pro Monat –, wird gleichzeitig das 4-Mbit-DRAM-Produktionsvolumen hochgefahren. Parallel dazu arbeitet die Entwicklung natürlich bereits intensiv an den Nachfolgenerationen: dem 16-Mbit-DRAM und – zusammen mit der Forschung – am 64-Mbit-DRAM. Beim Speicherwettbewerb sind also mehrere Mannschaften gleichzeitig im Rennen. Insgesamt sind einige hundert hochqualifizierte Schaltungstechniker, Prozeßentwickler, Fertigungstechniker, Analytiker und Infrastrukturspezialisten mit Tausenden von Mitarbeitern in den Fertigungs- und Entwicklungslinien am Werk.

Allein das Großprojekt «Entwicklung der nächsten Speichergeneration» benötigt die direkte Mitarbeit von weit über 100 dieser Experten pro Jahr sowie Know-how oder Dienstleistungen eines großen Teils der restlichen Mannschaft. Um sicherzustellen, daß das Projekt optimal abläuft, wird deshalb häufig eine Projektorganisation aufgebaut, die das Projekt in überschaubare Teilbereiche gliedert, Aktivitäten verschiedener Bereiche koordiniert und den zeitlichen Ablauf strukturiert und kontrolliert. Bild 1 zeigt als Beispiel die im Rahmen von JESSI (Joint European Submicron Silicon) durchgeführte Strukturierung des «Joint Memory Projects» und verdeutlicht die Vielzahl der Aufgaben, die erst im Zusammenspiel zur nächsten Speichergeneration führt.

Das Pflichtenheft – oder das Diktat der kleinsten Strukturen

Am Beginn der Arbeit steht die Definition der wesentlichen Eckdaten des neuen Chips. Dabei sind eine Reihe von Randbedingungen zu beachten, deren Einhaltung unbedingt erforderlich ist. Den größten Einfluß hat die Festlegung des künftigen Gehäuses; in der Regel erfolgt dies drei bis vier Jahre vor Markteintritt. Das internationale Industriegremium JEDEC (Joint Electron Device Engineering Council) normt das Gehäuse, die elektrischen Funktionen der Anschlüsse (der «Beinchen» des Chips) sowie einige wesentliche elektrische Daten.

Dr. HELMUT FÖLL, Siemens AG, HL 16M P, Otto-Hahn-Ring 6, D-8000 München 83; Dr. FRANK STEFAN BECKER, Siemens AG, Pressereferat Halbleiter, Balanstraße 73, D-8000 München 80.

Aus den Gehäuseabmessungen läßt sich die Chipgröße ableiten und aus dieser dann die für eine einzelne Speicherzelle verfügbare Fläche. Eine Speicherzelle enthält einen Transistor, einen Kondensator, die entsprechenden Stücke der Bit- und Wortleitung sowie die Isolation der Elemente untereinander und zur Nachbarzelle. Als Aufgabe bleibt, diese Elemente so günstig wie möglich anzuordnen, um zu möglichst entspannten «Designregeln» zu kommen. Die Designregeln legen die minimalen Strukturgrößen und die zugehörigen Toleranzen fest; mit ihrer Definition ist gleichzeitig vorgegeben, was die Feinstrukturtechnik, insbesondere die Lithografie, leisten muß. Auch die Meßtechnik weiß damit, mit welcher Genauigkeit kontrolliert werden muß, denn nur Dimensionen, die auch routinemäßig gemessen werden können, lassen sich sinnvoll spezifizieren und später in einer Fertigung millionenfach herstellen. Bild 2 zeigt diese Wirkungskette am Beispiel des 16-Mbit-DRAM. Man kann den ganzen Ablauf aber auch «rückwärts» lesen – und der jeweilige Technologieführer tut das auch: Die kleinsten Dimensionen, die man sich zutraut, bestimmen die Designregeln, aus diesen folgt die Zellgröße usw.

Eine weitere wichtige Vorgabe ist die Höhe der Versorgungsspannung. Auch das 16-Mbit-DRAM muß noch bei 5 V arbeiten können. Möglicherweise ist aber auch eine 3,3-V-Version erforderlich – von der Schaltungstechnik und vom Design her gesehen, sind das zwei völlig verschiedene Produkte. Die durch Beibehaltung der Spannung bei gleichzeitiger Verkleinerung der Dimensionen entstehenden Probleme bei Parametern wie elektrische Feldstärke, Strom- oder Leistungsdichte wurden schon angesprochen; bei weiteren Generationen von Speichern wird eine Reduzierung der Versorgungsspannung unumgänglich sein. Auch beim 16-Mbit-DRAM wird chipintern teilweise schon mit reduzierter Spannung gearbeitet.

Weitere wichtige Produktziele, die schon frühzeitig feststehen müssen, sollen hier nur kurz gestreift werden: Zugriffszeiten, Organisation – soll der Speicher sich nach außen hin wie ein 16-Mbit-Speicher oder wie vier 4-Mbit-Speicher verhalten? –, Montagetechnik (beeinflusst die Lage der Anschlußfelder auf dem Chip und damit das «Layout»), Leistungsaufnahme sowie weitere produktspezifische Daten.

Aus all diesen Eckwerten lassen sich Anforderungen an die in Bild 1 gezeigten Teil- und Unterprojekte ableiten. Die dabei noch mögliche Vielfalt der Problemlösungen wird durch zwei weitere, überaus wichtige Vorgaben stark

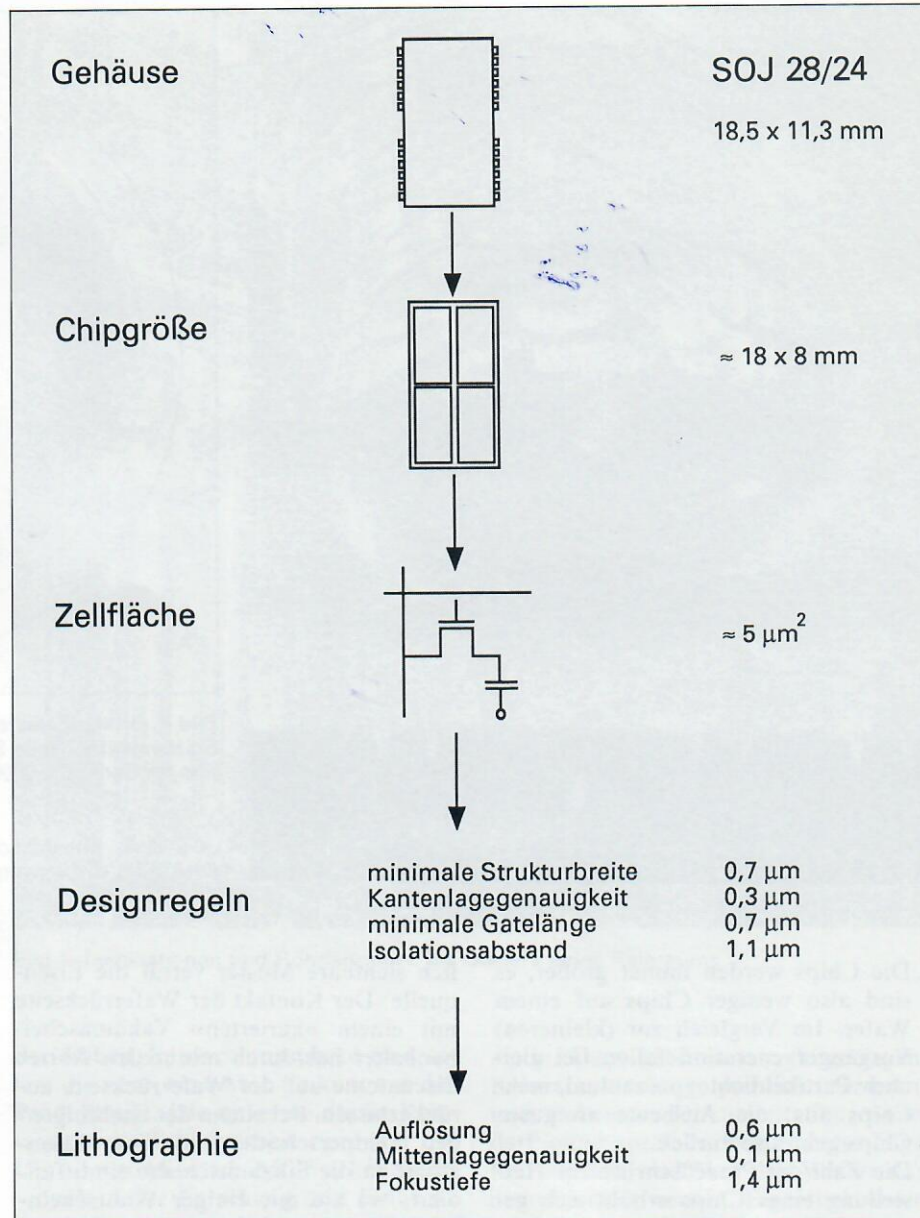


Bild 2. Der Weg von der Definition des 16-Mbit-DRAM-Gehäuses zu Designregeln und Lithografieanforderungen.

eingengt. Die erste heißt «Time to Market», das heißt eine möglichst kurze Zeitspanne bis zum Markteintritt; die zweite heißt fertigungsgerechter und kostengünstiger Prozeß. Denn ein Speicher, der zu spät den Markt erreicht oder aber nicht kostengünstig gefertigt werden kann, führt bei Megachips unweigerlich zu Millionenverlusten.

Mikropartikel bedrohen die Megachips

Auf der Querschnittfläche eines menschlichen Haars haben rund 50 Speicherzellen eines 4-Mbit-DRAM Platz (Bild 3). Eine einzige Schuppe des Haars in Bild 3 deckt ungefähr zwei Speicherzellen zu. Teilchen aller Arten oberhalb einer kritischen Größe sind also beim Herstellungsprozeß einer integrierten Schaltung unbedingt zu ver-

meiden. Aber auch Partikel mit Abmessungen weit unterhalb der Struktureinheit können zerstörend wirken; in den dünnen Dielektrikaschichten zum Beispiel führen manchmal schon Partikel im Tausendstelmikrometerbereich zu lokalen Stromdurchbrüchen. Selbst atomar verteilter «Dreck» in minimalster Konzentration kann für die Speicherzelle oder gar für den ganzen Chip tödlich sein. Insbesondere die Alkalimetalle (speziell Natrium) und Metalle wie Eisen, Kupfer oder Nickel sind schon bei Konzentrationen im ppt-Bereich (1 ppt = 0,000000001 %) schädlich und können zum Ausfall des Chips führen.

Die Vermeidung jeglicher «Kontamination» wird zunehmend das Schlüsselthema bei der Fertigung von Megachips. Dafür gibt es drei Gründe:

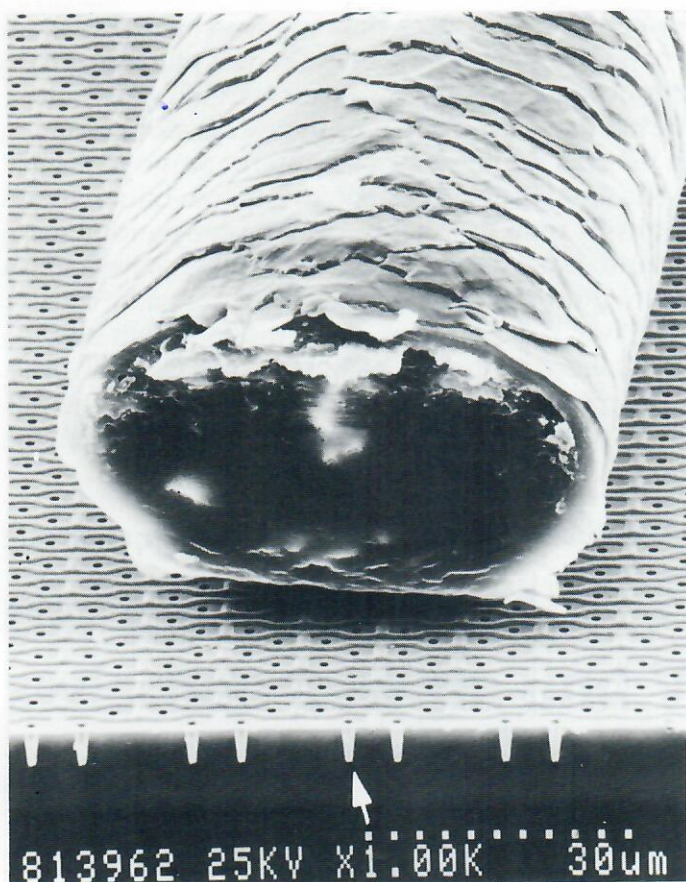


Bild 4. Ortsaufgelöste Darstellung der Eisenkonzentration in einem Siliziumwafer. Helle Stellen sind Gebiete «hoher» Eisenkonzentration im Bereich 0,0000003 %.

◀ Bild 3. Menschliches Haar und Speicherzellen des 4-Mbit-DRAM im Größenvergleich.

1. Die Chips werden immer größer, es sind also weniger Chips auf einem Wafer. Im Vergleich zur (kleineren) Vorgängergeneration fallen bei gleicher Partikeldichte prozentual mehr Chips aus; die Ausbeute an guten Chips geht also zurück.
2. Die Zahl der Prozessschritte zur Herstellung eines Chips erhöht sich gegenüber der Vorgängergeneration; die Wahrscheinlichkeit, daß ein Partikel auf die Scheibe fällt, erhöht sich entsprechend.
3. Die kritische Partikelgröße sinkt im Vergleich zur Vorgängergeneration, da die Strukturen kleiner geworden sind. Partikel, die bei der Vorgängergeneration noch nicht zum Ausfall führten, sind jetzt zerstörend.

Das bedeutet nichts anderes, als daß eine Fertigung, die beispielsweise 1-Mbit-DRAM mit einer Ausbeute von 70 % produziert, beim 4-Mbit-Chip kaum über 30 % Ausbeute erzielen wird, wenn sie ihr Kontaminationsniveau nicht drastisch reduziert.

Bild 4 illustriert, was es bedeutet, atomar gelöste Kontamination zu vermeiden. Dargestellt ist die Konzentration von Eisenatomen im Kristallgitter des Siliziums als «Bild» über die ganze Scheibe. Helle Stellen markieren Bereiche mit unzulässig hoher Eisenkonzentration von rund 0,0000003 %.

Das deut-lich sichtbare Muster verrät die Eisenquelle: Der Kontakt der Waferrückseite mit einem «karierten» Vakuumscheibenhalter hat durch minimalen Abrieb Eisenatome auf der Waferrückseite zurückgelassen. Bei einem der nachfolgenden Temperschritte sind diese Eisenatome in die Siliziumscheibe eindiffundiert, wo sie mit einiger Wahrscheinlichkeit zum Ausfall der Chips führen.

Da ist die Luft noch rein

Da ein Liter «saubere» Luft noch immer einige 10 000 Partikel enthält, muß die Chipherstellung in einem Reinraum erfolgen. Vorgefilterte, temperierte Luft wird in einem modernen Reinraum durch eine geschlossene Filterdecke gepreßt und fließt dann laminar mit relativ hoher Strömungsgeschwindigkeit (0,2 bis 0,4 m/s) durch einen gelochten Fußboden aus dem unmittelbaren Reinraumbereich heraus und durch große Ventilationschächte wieder zurück zu den Umwälzanlagen im Obergeschoß. Durch dieses Verfahren wird der eigentliche Reinraum weitgehend partikelfrei gehalten. Zum einem ist die Luft durch die Filterung selbst partikelfrei, zum anderen werden Staubeilchen, die im Reinraum entstehen, sehr schnell durch den Boden abgesaugt.

In der Praxis sieht das so aus, daß auf

ein Stockwerk des eigentlichen Reinraums ungefähr vier Stockwerke für die Luftumwälzung, Temperierung, Befeuchtung usw. fallen. Bei einem heutigen Klasse-10-Reinraum (weniger als zehn Partikel, die größer als 0,5 µm sind, in einem Kubikfuß Luft) mit 2000 m² Fläche werden stündlich ca. 250 000 m³ Luft umgewälzt. Die reinen Betriebskosten (Gas, Wasser, Strom) liegen bei über 10 Mio Franken pro Jahr. (Am Rande sei vermerkt, daß ein 2000-m²-Reinraum um einen Faktor 2 bis 3 zu klein ist für eine Speicherfabrik.)

Zum Reinraum gehört noch eine komplizierte Infrastruktur, insbesondere die Medienversorgung und -entsorgung. Eine Vielzahl von häufig gefährlichen Gasen und Flüssigkeiten muß aus zentralen Versorgungsstationen oder Gasflaschen zu der entsprechenden Anlage im Reinraum gebracht werden, und zwar *ohne* daß die Medien in den oft viele Meter langen Leitungen verunreinigt werden. Der «Keller» eines Reinraums sieht deshalb wesentlich interessanter aus als der eigentliche Arbeitsbereich (Bild 5).

Ein Reinraum für die Erforschung und die Entwicklung eines neuen Speicherchips (die «Pilotlinie») unterscheidet sich praktisch nur noch durch seine etwas geringeren räumlichen Abmessungen von einer Fabrik für die Massen-

produktion; Arbeitsabläufe, Infrastruktur und Logistik sind jedoch weitgehend identisch. Zunehmend wird auch in F+E-Reinräumen mit einem Teil der Kapazität für den Markt produziert, zum Teil, um die immensen Kosten etwas aufzufangen, vor allem aber auch, um professionelle Arbeitsabläufe sicherzustellen.

F + E rund um die Uhr

Betrachten wir den konkreten Fall der 16-Mbit-DRAM-Entwicklung. Der Prozeß baut zwangsläufig auf der 4-Mbit-DRAM-Produktion auf. Die Pilotlinie muß also zunächst die Herstellung der 4-Mbit-DRAM beherrschen. Das 16-Mbit-DRAM benötigt aber neue und bessere Prozesse. Diese müssen entwickelt werden. Zu allen innovativen Prozessen (zum Beispiel die beim 16-Mbit-DRAM erstmalig eingesetzte Wolframabscheidung) laufen Versuche; die Prozeßintegrationsexperten versuchen die einzelnen Prozeßschritte zu neuen Prozeßblöcken zusammenzusetzen. Einzelne Versuchswafer durchlaufen nur wenige Prozeßschritte, andere den kompletten 16-Mbit-Prozeß mit über 450 Einzelschritten.

Die meisten der am Projekt beteiligten Entwickler haben im Rahmen ihrer Versuchsreihen Wafers in der Pilotlinie. Sie wollen zunächst nur eines: ihre Scheiben möglichst rasch zur Auswertung zurück haben, um zu lernen, ob die Umsetzung ihrer Ideen erfolgreich war. Und danach wird gleich zum nächsten Lernzyklus gestartet.

Die Entwicklung und Optimierung des neuen 16-Mbit-DRAM erfordert einige vollständige Lernzyklen, also Komplettdurchläufe. Wie lange dauert so ein Lernzyklus? Eine Abschätzung dazu erhält man, wenn man die reine Bearbeitungszeit (die sogenannte physikalische Durchlaufzeit) mit dem «Flußfaktor» der Linie beaufschlagt. Der Flußfaktor gibt an, um wieviel länger ein Durchlauf, verglichen mit der physikalischen Durchlaufzeit, dauert.

Die physikalische Durchlaufzeit eines 16-Mbit-DRAM in der Entwicklungsphase liegt bei über 600 Stunden. Selbst bei einem (sehr guten) Flußfaktor von 2 dauert ein Durchlauf dann über 30 Wochen, falls die Linie «normal» arbeitet. Bei einer Arbeitszeit von knapp 40 Stunden pro Woche bedeutet das, daß die Entwickler längst «vergessen» haben, was sie eigentlich mit den Wafers wollten.

Die Marktlage (das 16-Mbit-DRAM muß 1992 in die Produktion) und die Mindestzahl an notwendigen Lernzyklen bestimmen zwingend eine Durchlaufzeit von maximal vier Monaten in

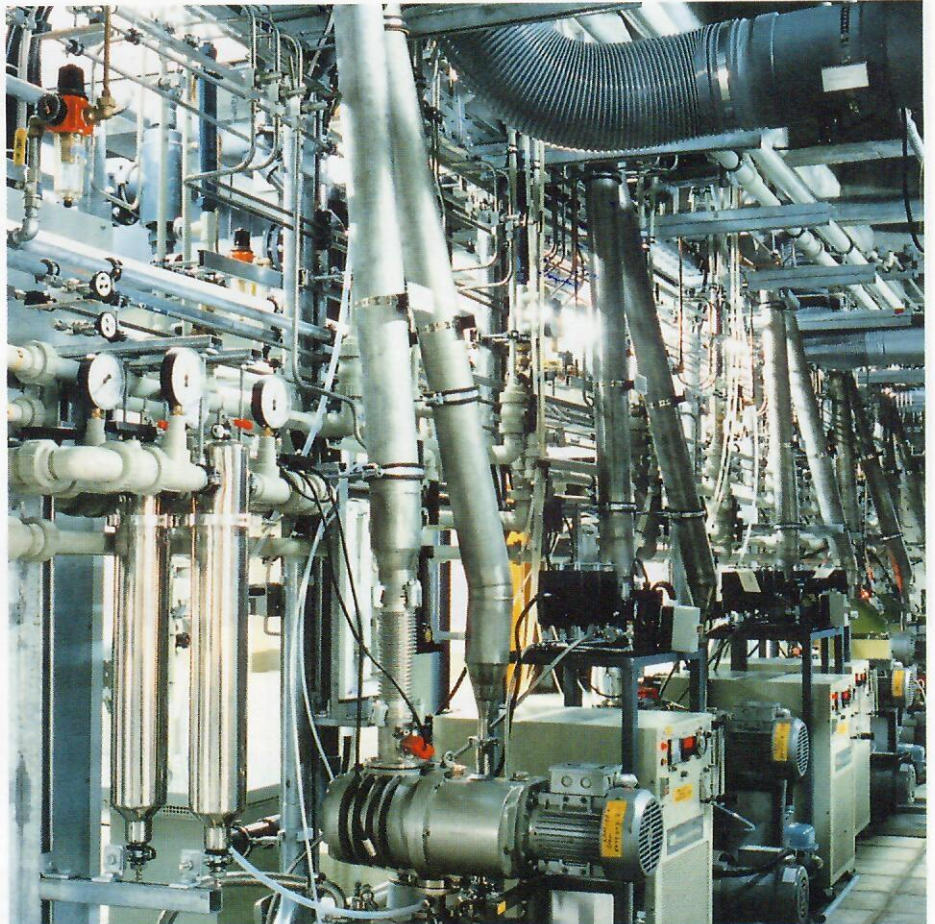


Bild 5. Installationen und Rohrleitungen im «Keller» eines Reinraums.

der frühen Entwicklungsphase und allenfalls noch zwei Monate kurz vor der Produktreife. Die unvermeidliche Konsequenz ist Schichtarbeit rund um die Uhr – einschließlich der Wochenenden – auch in der Forschung und Entwicklung!

Vermummungsgebot im Reinraum

Im Reinraum herrscht totales Vermummungsgebot; Bild 6 zeigt ein Beispiel der neuesten Mode. Menschen sind Partikelquellen erster Güte; ohne Schutzmaßnahmen umgibt uns eine Wolke von Hunderttausenden von Partikeln, die größer sind als $0,5\ \mu\text{m}$. Raucher exhalierten noch Stunden nach der letzten Zigarette Zigarettausende davon. Menschen sind außerdem Quellen des besonders gefährlichen Natriums. Es ist in Schweiß und Speichel reichlich vorhanden. Normaler Fingerschweiß ätzt Silizium (was den meisten Chemikalien nicht gelingt); deswegen lassen sich Fingerabdrücke nie wieder entfernen. Und Kosmetika enthalten für die Prozesse gefährliche Metalle in großen Mengen – von Schmuck ganz zu schweigen. Oberstes Gebot ist deshalb eiserne Dis-

ziplin: Dazu gehören korrekt getragene Reinraumkleidung, peinliche Ordnung und lückenlose Dokumentation; nichts darf ohne spezielle Handschuhe angefaßt werden, und über den Wafers darf nicht hantiert werden. Eigentlich unterscheiden sich die äußeren Rahmenbedingungen an diesen Arbeitsplätzen nur wenig von jenen in den Operationsräumen in den Kliniken.

Die Einführung der Samstag/Sonntag-Schicht in der Mikroelektronik gab in Deutschland Anlaß zu heftigen Diskussionen. Das gültige Gesetz aus dem späten 19. Jahrhundert erlaubt zwar die Wochenendarbeit als Ausnahme auch in der Industrie – allerdings nur unter einer einzigen Bedingung: Der Produktionsprozeß darf keine Unterbrechung zulassen. Die Auslegung dieses Gesetzes in den vergangenen 100 Jahren hat bei der Chipproduktion schließlich zu der Interpretation geführt, daß unter anderem ein Rückgang der Ausbeute oder ein Abfall der Qualität als Folge einer Unterbrechung des Herstellungsprozesses diese Bedingung erfüllt. In der Chipherstellung ist dies – wie die Experimente belegen – auch tatsächlich der Fall. Der Abfall in der Ausbeute ist jedoch nicht der einzige sachliche Grund

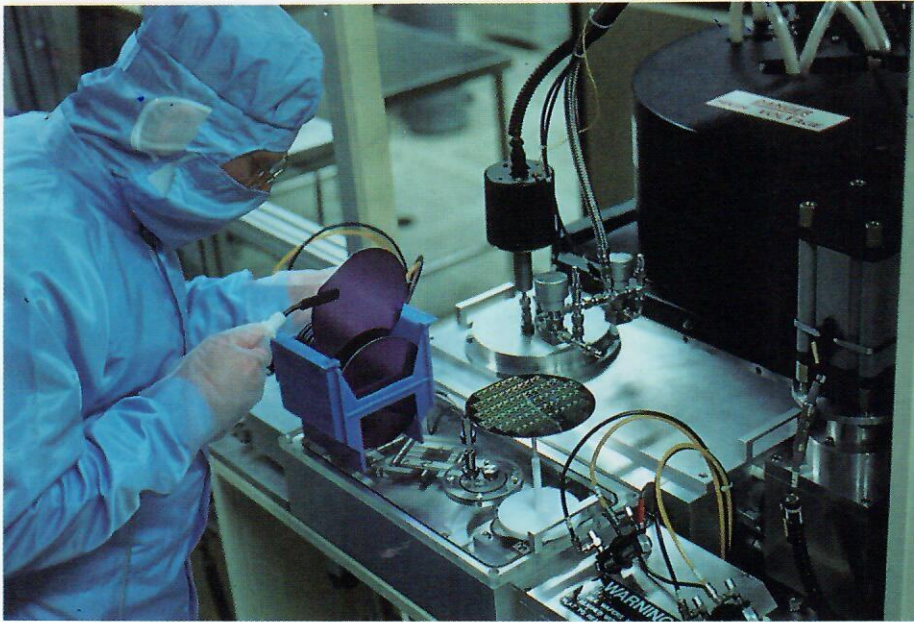


Bild 6. «Vermummter» Mitarbeiter im Reinraum.

für die Konti-Schicht (fünf Schichtgruppen) – lediglich der einzige rechtlich verwertbare. Bild 7 zeigt den Einbruch der Flächenausbeute beim nicht kontinuierlichen Produktionsbetrieb. Weitere wichtige Argumente sprechen gegen jeden unnötigen Prozeßunterbruch:

- zu lange Entwicklungszyklen
- schlechter beherrschbare Logistik (Wafers können nicht nach jedem Prozeßschritt für zwei Tage gestoppt werden).
- zu langsames Hochlaufen der Produktion auf die geforderte Ausbeute (abhängig von der Dauer der Lernzyklen)
- Verglichen mit den Kapitalkosten, sind die Personalkosten gering; ein möglichst hoher Nutzungsgrad der Investitionen ist zwingend.

Aus der Materialschlacht geht ein Chip hervor

Der Ablauf des Projekts «Nächste Speichergeneration» läßt sich am besten anhand der wichtigsten Meilensteine veranschaulichen (Bild 8). In der ersten Phase laufen vor allem Vorarbeiten bei Designern und Prozeßentwicklern; aus den technologischen Randbedingungen und den Möglichkeiten der Technologie folgt ein erster Satz von Designregeln. Einige Vorversuche zu prozeßtechnischen Fragen und die Simulation diverser schaltungstechnischer Ansätze kumulieren schließlich im «ersten Silizium» eines Testchips, also in ersten Si-Wafers mit den für den neuen Speicher relevanten Strukturen.

Mit den gewonnenen Erkenntnissen wird ein Technologie- und Designkonzept festgelegt, das durchaus noch Alternativen bei Prozeßkomplexen (zum Beispiel Zellkonzepten) und Einzelprozessen aufweisen kann. Zusammen mit einem ersten Produktdesign erhält man «erstes Silizium» des ersten neuen Speicherchips. Diese Speicher werden fast mit Sicherheit nicht voll funktionsfähig sein. Aus der Analyse der Design-, Technologie-, Montage- und Prüftechnikprobleme resultieren ein Neudesign und Lösungen der Technologieprobleme. Noch offene Alternativkonzepte werden entschieden; aus den entsprechenden Durchläufen resultieren jetzt «Labormuster», das heißt einige unter entspannten Bedingungen funktionierende Speicher. Entspannt heißt, daß der Speicher noch nicht unter allen spezifizierten Bedingungen (Temperatur, Spannung usw.) funktionieren muß. Erneut werden Prozesse und Design optimiert, Prüftechnik und Montage müssen jetzt stimmen. Prozesse werden festgeschrieben; Kontrollen und Qualitätssicherungsmaßnahmen verschärft. Ausbeuterelevante Messungen werden durchgeführt und Maßnahmen zur Ausbeutesicherung eingeführt. Die Zahl der eingeschleusten Wafers wächst. Schließlich erhält man kommerzielle Muster, die dem endgültigen Produkt schon sehr nahe kommen und auch zum Testen an gute Kunden ausgeliefert werden. Zunehmend beginnt nun eine Materialschlacht. Ziel ist, den Speicherbaustein auch im Grenzfeld der Spezifikationen funktionsfähig zu erhalten (Beispiel: $T = 85^\circ\text{C}$, Luftfeuchtigkeit = 100 %, $U = 4,5\text{ V}$), die Zuverlässigkeit für bei-

spielsweise zehn Jahre zu garantieren, Prozesse und Prozeßkomplexe auf Ausbeute und Kosten zu optimieren (mehr Ausbeute = mehr Kontrollen, mehr Kontrollen = mehr Kosten), Prüftechnik einschließlich Redundanz zu optimieren und die Ausbeute bei mindestens 10 % zu stabilisieren. Alles zu dem Zweck, die sogenannte Qualifikation zu bestehen, das heißt die Bestätigung der Qualitätsabteilung zu erhalten, daß der neue Chip alle Torturen im vorgeschriebenen Maße übersteht. Mit Bestehen der Qualifikation oder mit Vorserienfreigabe ist das Entwicklungsprojekt beendet, die Fertigung kann beginnen.

Hochlaufen auf der Lernkurve

Das Projekt wurde erfolgreich beendet, die Produktion wird aufgenommen. Was bedeutet das im einzelnen? Falls die Fertigungslinie nicht mit der Pilotlinie identisch ist, muß der Prozeß erstmals eingefahren und neu qualifiziert werden, da einige Anlagen und Prozesse sich von denen der Pilotlinie unterscheiden werden. Falls die Fertigungslinie neuer ist als die Pilotlinie, läuft sie möglicherweise schon mit größerem Waferdurchmesser. In diesem Fall müssen alle Prozesse neu optimiert werden. Die Fertigung verfolgt nun die Ziele:

- möglichst schnell die Ausbeute hochzubringen (die Ausbeute als Funktion der Zeit heißt «Lernkurve»)
- die Kosten der Bearbeitung eines Wafers zu minimieren
- die Kapazität der Prozeßlinie hochzufahren

Die beiden ersten Punkte bestimmen die Kosten pro verkaufsfähigen Chip; der letzte Punkt entscheidet darüber, wie viele Chips verkauft werden können. Der am Markt erzielbare Preis wird vom Weltmarkt festgesetzt; kleine Hersteller können ihn, im Gegensatz zu den dominierenden japanischen Häusern, nicht beeinflussen.

Betrachten wir die Lernkurve etwas genauer. Nehmen wir an, daß die Ausbeute bei Fertigungsbeginn bei 15 % liegt und daß der am Markt erzielbare Preis dann die Herstellkosten deckt und möglicherweise sogar noch einen Gewinn abwirft. Betrachtet man nun die typische Entwicklung des Speicherpreises als Funktion der Zeit, so läßt sich daraus die Lernkurve ableiten (Bild 9). Kurz gesagt: Jedesmal, wenn sich der Preis halbiert, muß sich die Ausbeute verdoppeln, damit die Kosten pro Chip konstant bleiben. Dabei wird zunächst einmal angenommen, daß vor allem die

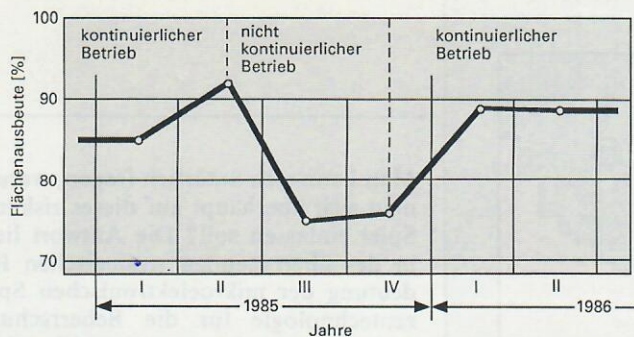


Bild 7. Einbruch der Flächenausbeute (Ausbeute an vollständig prozessierten Wafern) bei Unterbrechung der Kontinuierlichkeit in einer Speicherfabrik.

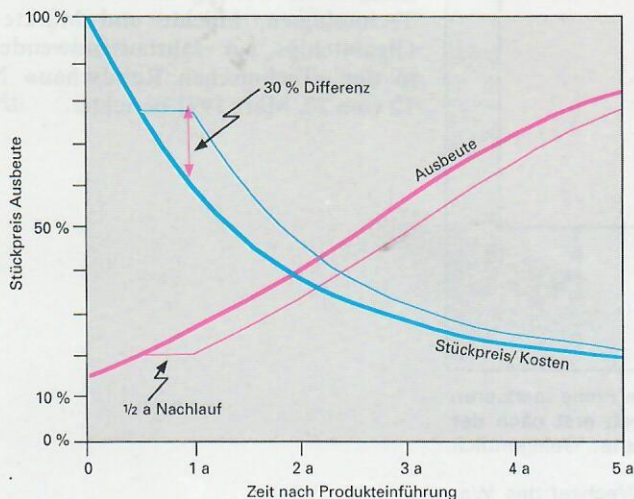


Bild 9. Preisverfall bei Speichern im Mittel der letzten Generationen und damit verknüpfter notwendiger Ausbeutehochlauf («Lernkurve»). Ein Nachlauf von nur einem halben Jahr in der Produkteinführungsphase führt zu 30 % höheren Kosten.

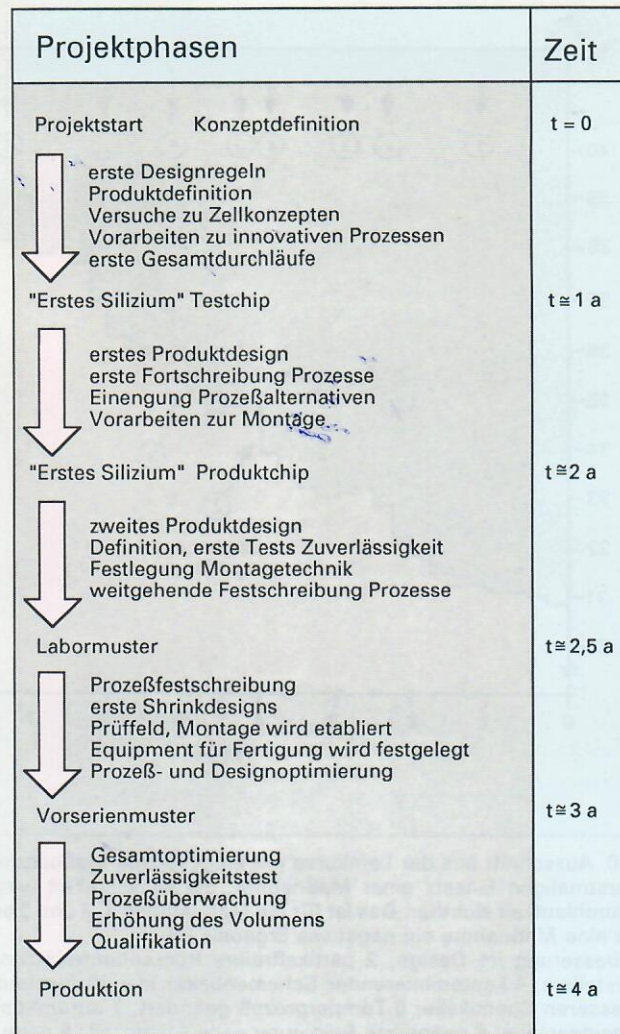


Bild 8. Wichtige Meilensteine eines Speicherentwicklungsprojekts mit den wesentlichen Arbeitsinhalten.

Herstellkosten für eine Scheibe konstant bleiben. Jedes Zurückbleiben hinter der Lernkurve führt dann automatisch zu Verlusten.

Das Hochlaufen auf der Lernkurve erfolgt keineswegs automatisch; jedes Prozent zusätzliche Ausbeute muß hart erkämpft werden. Eine Ausbeute von 80 % bedeutet, daß bei jedem der 400 Prozeßschritte die Prozeßausbeute 99,94 % erreichen muß und daß bei den rund 1000 Waferbewegungen (aus der Box in die Anlage, aus der Anlage in die Box, aus der Box zum Meßgerät usw.) nicht mehr als 10 bis 20 «Killerpartikel» auf die Scheibe fallen. Nur dauernde Analysen von Ausfallursachen und zähes Optimieren der Prozesse können zum Erfolg führen. Bild 10 zeigt an einem realistischen (wenn auch nicht realen) Beispiel, was Ausbeutesteigerung bedeutet. Die Hauptarbeit steckt im Erkennen der Ausfallursachen, die Behebung kann dann gelegentlich sehr einfach sein.

Die Anstrengungen, die gemacht werden, um auf der Lernkurve hochzulaufen, werden begleitet von den Anstrengungen, die Kosten pro Wafer zu sen-

ken. Diese sogenannten Rationalisierungsmaßnahmen können sein:

- Vereinfachung der Prozesse und der Prozeßführung
- Verwendung billiger Medien (sobald man weiß, worauf es ankommt)
- Beschaffung von leistungsfähigeren Anlagen
- besserer Fertigungsfluß (und damit höhere Kapazität) durch verfeinerte (computergesteuerte) Logistik
- Verkleinerung der Chipfläche durch Einführung eines sogenannten «Shrink»-Produkts

Die letzte Maßnahme ist zwar besonders wirksam, da man auf einem Wafer rund 20 % mehr Chips bekommt. Sie ist aber nicht ganz einfach zu realisieren. Sie bedingt in der Regel neues Lithografieequipment, da das vorhandene immer an der Grenze seiner Leistungsfähigkeit betrieben wird, darüber hinaus ein neues Design, eine teilweise geänderte Prozeßführung und in der Regel wiederum einige neue Einzelprozesse. Damit muß der «Shrink»-Chip einerseits neu qualifiziert werden (was viel Zeit und Geld kostet), und es muß auch der kleinere Chip erst wieder auf einer

Lernkurve hochlaufen. Mit der Rationalisierungsmaßnahme «Shrink» sind deshalb auch erhebliche Vorleistungen und Risiken verbunden.

Wenn alles gutgeht, liegt man mit der Ausbeute auf der Lernkurve, und die Rationalisierungsmaßnahmen greifen. Falls man auch noch einen Kunden findet, der die Chips kauft (das ist bei den wilden Fluktuationen des Weltchipmarkts alles andere als selbstverständlich), verdient die Fabrik Geld. Was passiert, wenn etwas schiefgeht?

Nehmen wir einmal an, daß die Lernkurve zum Beispiel plötzlich bei einer Ausbeute von 20 % für sechs Monate stagniert. Was in diesem Fall passiert, zeigt ebenfalls Bild 9. Die Kosten steigen gegenüber dem Weltmarktpreis um 30 %. Ist die Gewinnspanne kleiner (und das ist sie im Speichergeschäft mit Sicherheit), wird jetzt mit Verlust produziert. Eine simple Rechnung macht das deutlich: Der Weltmarktpreis beträgt 20 Franken. Die den Weltmarktpreis bestimmende Konkurrenz aus dem Fernen Osten produziert für 19 Franken; wir sind jedoch um 30 % teurer, wir fertigen für 25 Franken. Bei

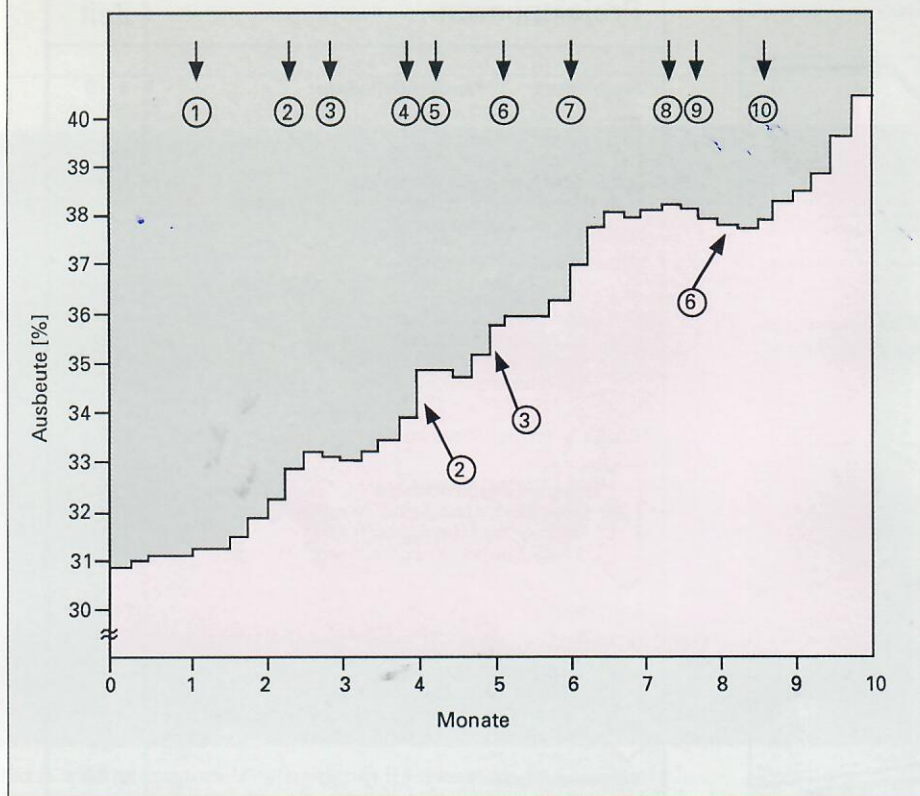


Bild 10. Ausschnitt aus der Lernkurve mit eingesetzten Maßnahmen. Die Pfeile markieren den erstmaligen Einsatz einer Maßnahme; die Wirksamkeit wird jedoch erst nach der Restdurchlaufzeit sichtbar. Das ist für die Maßnahmen 2, 3 und 6 angedeutet. Gelegentlich erzielt eine Maßnahme ein negatives Ergebnis (6).

1 Verbesserung im Design; 2 partikelfreiere Rückseitenfreiletzung; 3 Wechsel des Waferherstellers; 4 kontaminierender Scheibenhalter identifiziert und beseitigt; 5 Einsatz einer besseren Chemikalie; 6 Temperprozeß geändert; 7 automatisierte Beschickung einer Abscheideanlage; 8 geänderte Reinigung nach Ätzprozeß; 9 neue Anlage für Oxidätzung; 10 Temperprozeß wieder wie früher.

einem Umsatz von 500 Mio Franken macht der Marktleader also 25 Mio Franken Gewinn – und wir fahren einen Verlust von 125 Mio Franken ein. Da die in Bild 9 gezeigte Preiskurve nicht fiktiv ist, sondern ziemlich exakt der Realität entspricht, bedeutet dies in der Tat, daß der Unterschied zwischen Gewinn und massiven Verlusten in einigen Monaten zeitlichem Abstand zu den ostasiatischen Weltmarktführern liegt.

Unter dem Strich

Die Entwicklung der neuen Speichergeneration hat eine ganze Menge Geld gekostet – die Schätzung von etwa einer halben Milliarde Franken dürfte nicht ganz falsch sein. Diese Zahl erklärt sich leicht aus den Kosten eines Entwicklermannjahrs von rund und roh 1 Mio Franken pro Jahr. Denn ein Entwickler bezieht nicht nur ein Gehalt, sondern er braucht zum Arbeiten eine sehr teure Infrastruktur (Reinraum!) und verwendet sehr teure (da höchstreine) Materialien. Da während der Entwicklungsphase weit über 100 Experten pro Jahr am Werk sind, aber auch nach der Markteinführung noch namhafte Entwicklungsaufwände anfallen (beispielsweise für das «Shrink»-Produkt), kommt eine halbe Milliarde Franken schnell zusammen.

Wie schon gezeigt, ist es nicht allzu schwierig, in der Gesamtbilanz für eine Chipgeneration Verluste in Milliardenhöhe einzufahren; die Gewinnaussichten lassen sich dagegen auch im besten Fall eher mit Zehntelmilliarden beziffern. Ist man gezwungen, unter dem Herstellungspreis zu verkaufen, weil die Konkurrenz einige Monate voraus ist oder, wie schon geschehen, zu Dumpingpreisen anbietet, ist ein Fiasko unvermeidlich. Selbst für die großen internationalen Elektromultis erscheint dieses Risiko zunehmend als nicht mehr tragbar. Wir beobachten deshalb eine zunehmende Verflechtung der wesentlichen Firmen über Kooperationen, Joint-ventures und dergleichen.

Außer den Entwicklungskosten muß noch knapp 1 Mia Franken zum Bau einer Fabrik aufgebracht werden. Man hat also schon weit mehr als 1 Mia Franken ausgegeben, bevor der erste Franken Umsatz erzielt wird.

Man kann sich natürlich fragen, warum man sich überhaupt auf dieses riskante Spiel einlassen soll? Die Antwort liegt in der überragenden strategischen Bedeutung der mikroelektronischen Spitzentechnologie für die Beherrschung des Gesamtspektrums der Mikroelektronik und in der zentralen Rolle der Mikroelektronik für moderne Volkswirtschaften. Darüber haben wir zum Beispiel im Beitrag «Mikroelektronik: Technologien, Mächte und Märkte – Gigabit-Chips zur Jahrtausendwende?» in der «Technischen Rundschau» Nr. 12 vom 22. März 1991 berichtet. [12] ®

Weiterführende Literatur

Zum technischen und wirtschaftlichen Umfeld:

Weinerth H. (Hrsg.): Schlüsseltechnologie Mikroelektronik. Franzis-Verlag, München, 1990 (Elektronik).

Föll H., Becker F. S.: Industrielle Entwicklung und Fertigung von Halbleiterelementen. Tagungsband «Festkörperforschung für die Informationstechnik» des 21. IFF. Seiten 16.1 bis 16.79. Institut für Festkörperforschung, KFA Jülich, Bundesrepublik Deutschland, März 1990.

Zur Technologie:

Widmann D., Mader H., Friedrich H.: Technologie hochintegrierter Schaltungen. Springer-Verlag, 1988.

Becker F. S., Föll H., Schlüter K.: Die Mega-Generation. Microcomputer-Zeitung 12/90, Seite 60.

Ertl J., Mönkemeyer G., v. Otterloo H.: Verschiedene Fachartikel im GME-Fachbericht 5, «Halbleiterfertigung». VDE-Verlag, 1989.

Becker F. S.: Fertigungsprozesse für Speicher der Megageneration – Megabit-Multis fordern die Technologen heraus. Technische Rundschau Nr. 23, Hallwag AG, 1991.

Zur Geschichte der Mikroelektronik:

Rogers E. M., Larsen J. K.: Silicon-Valley-Fieber. Siedler Verlag, Berlin, 1985. Queisser H.: Kristallene Krisen. Piper Verlag, München, Zürich, 1985.

Die diesem Bericht zugrundeliegenden Arbeiten wurden mit Mitteln des Bundesministeriums für Forschung und Technologie und den Förderkennzeichen MT 2788 gefördert. Die Verantwortung für den Inhalt liegt jedoch allein bei den Autoren.