

Teil 2 Projektablauf und Fertigung

16.6 Aufbau, Organisation und Infrastruktur einer sub μ m-Prozeßlinie

16.6.1 Megachips und Mikropartikel

Auf der Querschnittsfläche eines menschlichen Haares haben ca. 50 Speicherzellen eines 4M DRAMs Platz (Fig. 21). Eine einzige Schuppe des Haares in Fig. 8.1 deckt ungefähr 2 Speicherzellen zu. Partikel jeder Art oberhalb einer kritischen Größe sind also unbedingt zu vermeiden, da sie direkt die geometrischen Strukturen zerstören können und damit zum Ausfall der Bauelemente führen. Aber auch Partikel mit Abmessungen weit unterhalb der Strukturfeinheit können zerstörend wirken. Beispielsweise können in den dünnen Dielektrikschichten (vgl. Kapitel 16.3 und 16.4.3) schon Partikel im nm-Bereich zu lokalen Stromdurchbrüchen führen. Aber auch kleinste Partikel, die Elemente enthalten, die in den kritischen Schichten oder im Silicium selbst die elektronischen Eigenschaften verändern, können tödlich sein. Dies gilt insbesondere für die Alkalimetalle (speziell Na) und die in Silicium schnell diffundierenden Metalle wie z.B. Cu, Ni, Fe, Au. Konzentrationen dieser Elemente im ppt-Bereich können zu elektrischen Ausfällen führen - bei völlig intakter Struktur der Schichten und dotierten Zonen.

Im folgenden werden mit "Partikel" die Fremdkörper bezeichnet, die auf der Si-Scheibe oder in den Schichten strukturzerstörend wirken - ein extremes Beispiel zeigt Fig. 22. Im Silicium oder in den Schichten enthaltene störende Fremdatome werden als "Kontamination" bezeichnet (im allgemeinen Sprachgebrauch umfaßt "contamination" beide Arten von Verunreinigungen).

Die Vermeidung von Partikeln und Kontamination wird zunehmend das Schlüsselthema bei der Fertigung von Megachips (also Chips mit sub μ m-Strukturen und Flächen in cm²-Bereich).

Dafür gibt es drei Gründe:

1. Die Chips werden immer größer, es sind also weniger Chips auf der Scheibe. Bei gleicher Partikeldichte wie bei der Vorgängergeneration verkleinert sich die Ausbeute. In einfacher Näherung gilt für die Ausbeute bei einem Prozeßschritt:

$$Y_{ps} = 1 - \frac{F}{F_0} \cdot z$$

mit Y_{ps} = Ausbeute des betrachteten Prozeßschritts,
 F = Fläche des Chips; F_0 = für Chips verfügbare Fläche der Scheibe; z = Zahl der "tödlichen" Partikel auf der Fläche F_0 beim betrachteten Prozeßschritt.

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. Pauschal gilt für die Ausbeute Y

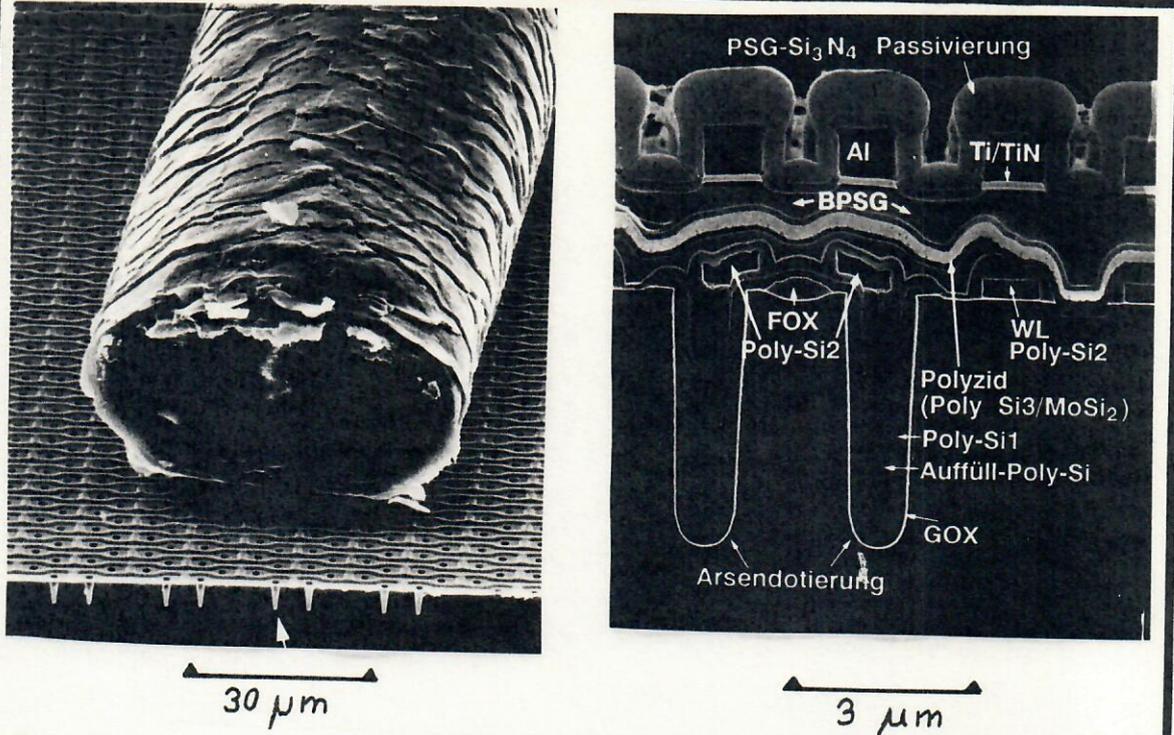


Fig. 21: a) Menschliches Haar auf 4M-DRAM Struktur. b) Querschnitt (nach vollständiger Prozessierung) durch die in a) mit Pfeil markierte Zelle.

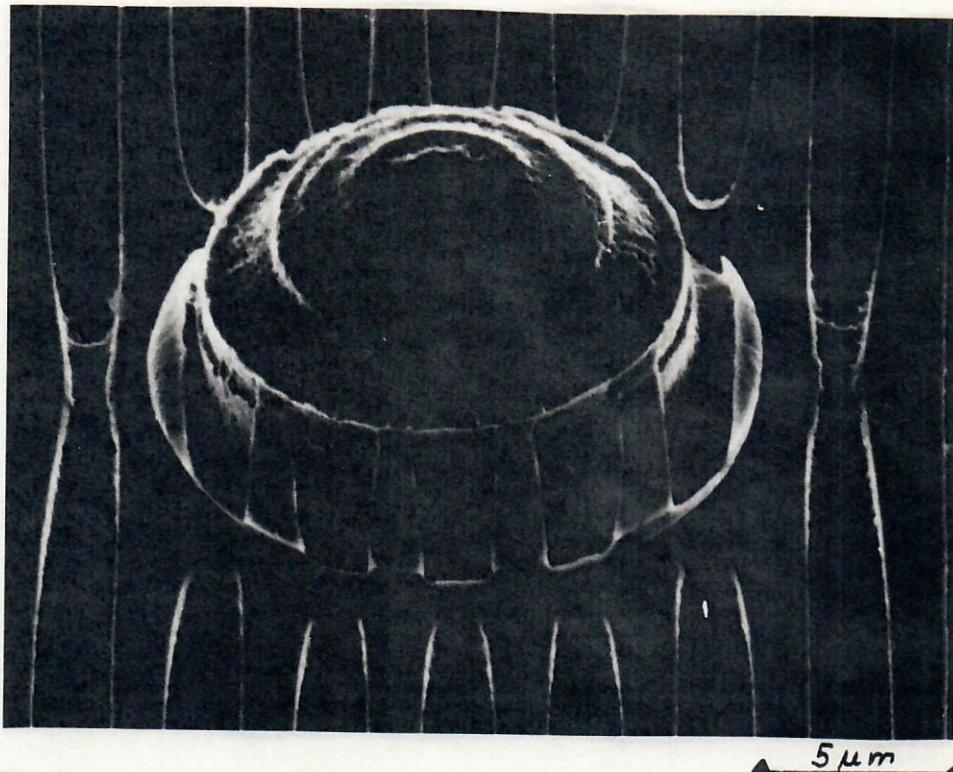


Fig. 22: Strukturzerstörendes, besonders großes (15 μm!) Partikel auf 4M-DRAM. Sechs Al-Bahnen sind kurzgeschlossen.

am Ende aller Prozessschritte

$$Y_{ps} = \left(1 - \frac{F}{F_0} \cdot z\right)^n$$

mit n = Zahl der kritischen Prozessschritte. Da manche Prozessschritte kritischer sind als andere, müßte eigentlich das Produkt aller Y_{ps} mit der für jeden Prozessschritt richtigen Partikelzahl ausgerechnet werden; wir vereinfachen hier grob, indem wir nur "kritische" Prozessschritte betrachten.

3. Die kritische Partikelgröße ist kleiner im Vergleich zur Vorgängergeneration, da auch die Strukturen kleiner sind. Partikel, die früher noch nicht zum Ausfall führten, sind jetzt tödlich - in anderen Worten: z wird größer. Zur Illustration folgendes, in den Absolutwerten relativ willkürliches, aber in der Relation nicht allzu falsches Beispiel. Für das 4M DRAM soll gelten:

$F_0 = 100 \text{ cm}^2$, $F = 1 \text{ cm}^2$, $z = 1$, $n = 50$. Daraus ergibt sich:

$$Y_{4M} = 61 \% .$$

Für das 16M DRAM könnte gelten: $F_0 = 100 \text{ cm}^2$; $F = 1,4 \text{ cm}^2$; $z = 1,2$ $n = 60$. Daraus ergäbe sich

$$Y_{16M} = 36 \% .$$

Die Lage ist weniger übersichtlich bei den noch tolerierbaren Kontaminationsniveaus, der Trend ist jedoch derselbe.

Die Konsequenz ist, daß das bei der Fertigung der Generation n noch akzeptierbare Partikel- und Kontaminationsniveau bei weitem nicht ausreichend für die Generation $n + 1$! Fig. 23 illustriert dies; siehe auch /33/

Das anzustrebende Ideal von Partikel- und kontaminationsfreien Prozessen hat zu eigenen Forschungsrichtungen und Industrien geführt, ein Blick in die Zeitschrift "microcontamination" illustriert dies. Dabei wird oft physikalisch/chemisches Neuland betreten. Mangels zuverlässiger und verstandener Nachweismethoden gibt es beispielsweise kaum Kenntnisse über Partikel, die kleiner sind als ca. $0.1 \mu\text{m}$. Wieviele solcher Partikel gibt es in der Luft/im Reinraum/im Vakuum einer Anlage (Vakuum heißt nicht Partikelfreiheit!)? Aus was bestehen sie? Wie entstehen sie? Wie fallen sie auf die Si-Scheibe? Wie fest haften sie dort (Schwerkraft spielt hier keine Rolle mehr, sondern nur Phänomene wie Elektrostatik, van-der-Waals Wechselwirkung, etc.)? Ähnlich die Situation bei Kontaminationen: Wie messe ich, ob irgendwo auf oder in der Scheibe z . B. die Fe-Kontamination kleiner ist als 10^{-12} cm^{-3} . Das sind 20 ppt oder $2 \cdot 10^{-9}\%$ oder 1/10 gr. Fe auf ein $50 \times 20 \times 2 \text{ m}^3$ Schwimmbad voll mit Silizium - und wahrscheinlich zuviel für einen 16M-DRAM Chip.

Fig. 24 zeigt zum Beispiel die Verteilung von Kontamination (hauptsächlich Fe) in einer Scheibe. Die Ursache der zu hohen

MAXIMAL ZULÄSSIGE TÖDLICHE DEFEKTE JE MASKENEbene	0,6cm ⁻²	0,3cm ⁻²	0,15cm ⁻²	0,08cm ⁻²	0,05cm ⁻²	
MAXIMALE GRÖSSE DIREKTER KILLER - PARTIKEL	1,0µm	0,6µm	0,4µm	0,25µm	0,15µm	0,1µm
MAXIMALE GRÖSSE INDIREKTER KILLER - PARTIKEL	0,4µm	0,25µm	0,15µm	0,1µm	0,07µm	0,05µm
MINIMALE DESIGN RULES FÜR KRITISCHE STRUKTUREN	4µm	2,5µm	1,5µm	1µm	0,7µm	0,5µm
MINIMALE GATEOXID-DICKE	0,1µm	0,05µm	0,025µm	0,0125µm	0,008µm	0,006µm
SCHEIBENDURCHMESS.	100mm	125mm	150mm	200mm		
PROZESSGENERATION						

Fig. 23: Graphische Darstellung von zulässigen Partikeldichten und -Größen bei den verschiedenen Speichergenerationen.

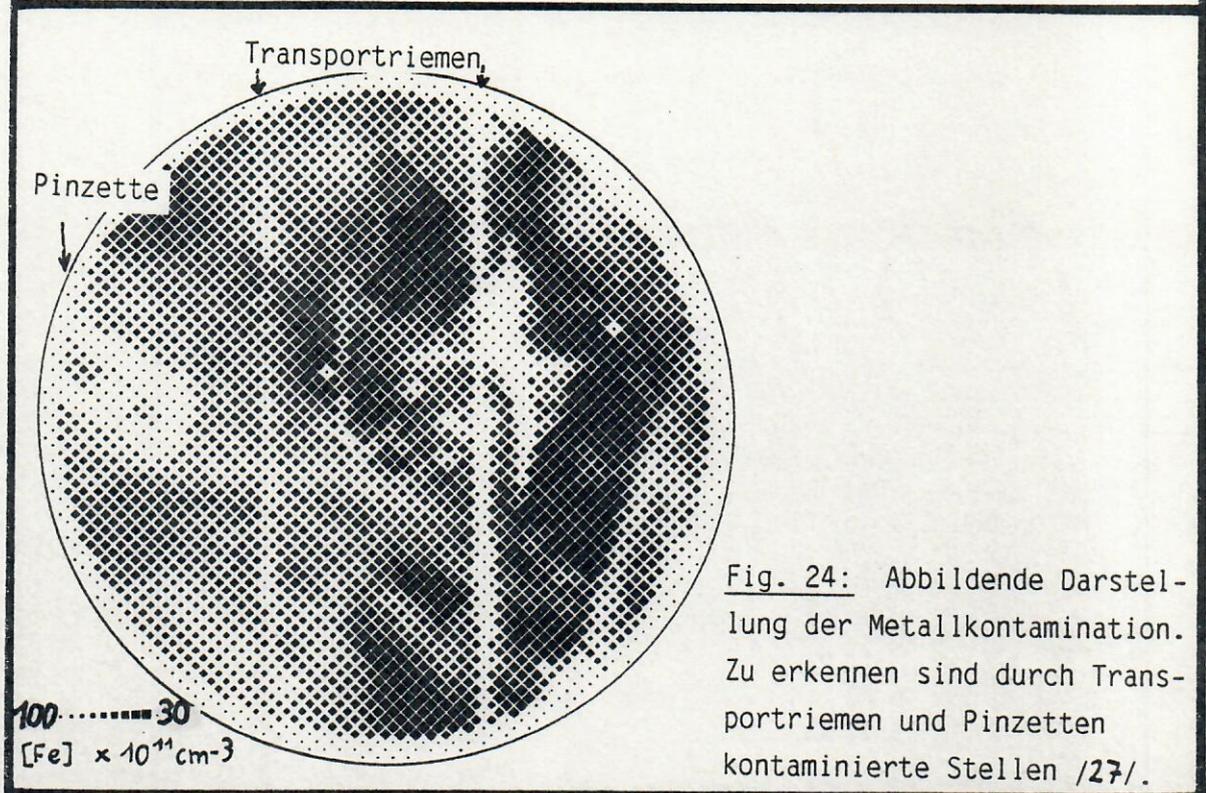


Fig. 24: Abbildende Darstellung der Metallkontamination. Zu erkennen sind durch Transportriemen und Pinzetten kontaminierte Stellen /27/.

Kontamination im ppb Bereich ist die einmalige Berührung der Scheibe mit einer "verschmutzten" Transporteinrichtung und Pinzette.

Wieviel Fe darf sich denn in all den Flüssigkeiten und Gasen befinden, damit weniger als 20 ppt in den Si-Kristall eingeschleppt wird? Wieso kann eigentlich so ein bißchen Fe (oder Cu, oder Ni, oder Na, ...) das Funktionieren des Chips verhindern?

Zu einigen dieser Fragen gibt es Antworten, zu den meisten nur Halbwahrheiten und Spekulationen. Wenn man ein bißchen weiter vorausdenkt (z.B. an das 256M-DRAM (ca. 1998?)) gibt es heute mit einiger Sicherheit noch nicht einmal die relevanten Fragen.

Pragmatisches Vorgehen erfordert, einerseits die Partikel- und Kontaminationsproblematik möglichst gut zu verstehen und zu erforschen, andererseits bekannte oder vermutete Partikel- und Kontaminationsquellen rigoros zu vermeiden - auch wenn im Einzelnen oft nicht klar ist, was eine bestimmte Maßnahme an Ausbeutesteigerung erbringt. Grenzen setzt dabei die Zumutbarkeit der Maßnahmen an die Mitarbeiter (Reinraumkleidung!) und die Bezahlbarkeit (ultrareine Materialien sind nicht billig).

16.6.2 Der Reinraum - eine komplexe Maschine

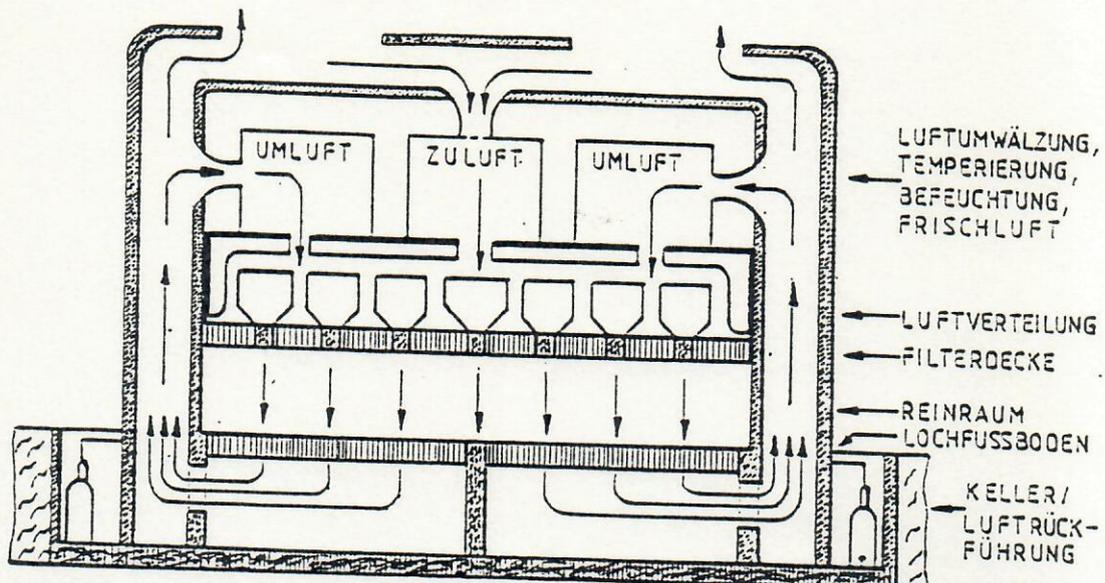
In modernen Reinräumen besteht die Decke aus einer geschlossenen Filterfläche. Von oben wird vorgefilterte, temperierte Luft mit eingestelltem Feuchtegehalt durch diese Filter gepreßt. Die Luft fließt laminar mit relativ hoher Strömungsgeschwindigkeit ((20-40)cm/sec) durch einen gelochten Fußboden aus dem unmittelbaren Reinraumbereich und durch große Ventilationsschächte wieder zurück zu den Umwälzanlagen im Obergeschoß. Durch dieses Verfahren wird der eigentliche Reinraum weitgehend partikelfrei gehalten. Zum einen ist die Luft durch die Filterung selbst partikelfrei, zum anderen werden Staubteilchen, die im Reinraum entstehen, sehr schnell durch den Boden abgesaugt.

In der Realität sieht das so aus, daß auf ein Stockwerk des eigentlichen Reinraums ungefähr vier Stockwerke für die Luftumwälzung, - Temperierung, etc. fallen. Bei einem heutigen "Klasse 1" Reinraum (weniger als 10 Partikel $> 0,5 \mu\text{m}$ im Kubikfuß Luft) mit 2000m^2 Fläche werden stündlich ca. 250000m^3 Luft umgewälzt. Die reinen Betriebskosten ("Gas, Wasser, Strom") liegen bei ca. 10^7DM/a . Fig. 25 zeigt einen Querschnitt durch einen Reinraum, Fig. 26 die Außenansicht einer der beschriebenen "Hallen". Am Rande sei vermerkt, daß ein 2000m^2 Reinraum um einen Faktor 2-3 zu klein ist für eine "richtige" Speicherfabrik.

Zum Reinraum gehört jedoch noch mehr; die wichtigsten Punkte sind:

1. Medienversorgung und -entsorgung:

KLASSE 10: weniger als 10 Partikel ($> 0.3 \mu\text{m}$)
in 15 ltr Luft



BEISPIEL:

	↑ GAS/CHEMIKALIEN VERSORGUNG
Reinraumfläche	2000 m ²
Strömungsgeschwindigkeit der Luft	
"Klasse 10"	0.45 m/sec
"Klasse 1000"	0.25 m ³ /sec
Umgewälztes Luftvolumen	ca. 2 500 000 m ³ /h
Baukosten	50 000 000 DM
Betriebskosten	
(Luft, Elektrizität, Medien)	7 000 000 DM/Jahr
Kosten der Anlagen	235 000 000 DM
Kosten/Arbeitsplatz	1 000 000 DM
(ca. 300 Personen)	

Fig. 25: Stark vereinfachter Querschnitt durch einen Reinraum der Klasse 10 - 1. In unkritischen Bereichen wird die Klasse auf 1000 reduziert um teure Luft zu sparen. Gasflaschen stehen außerhalb der Luftumwälzung.

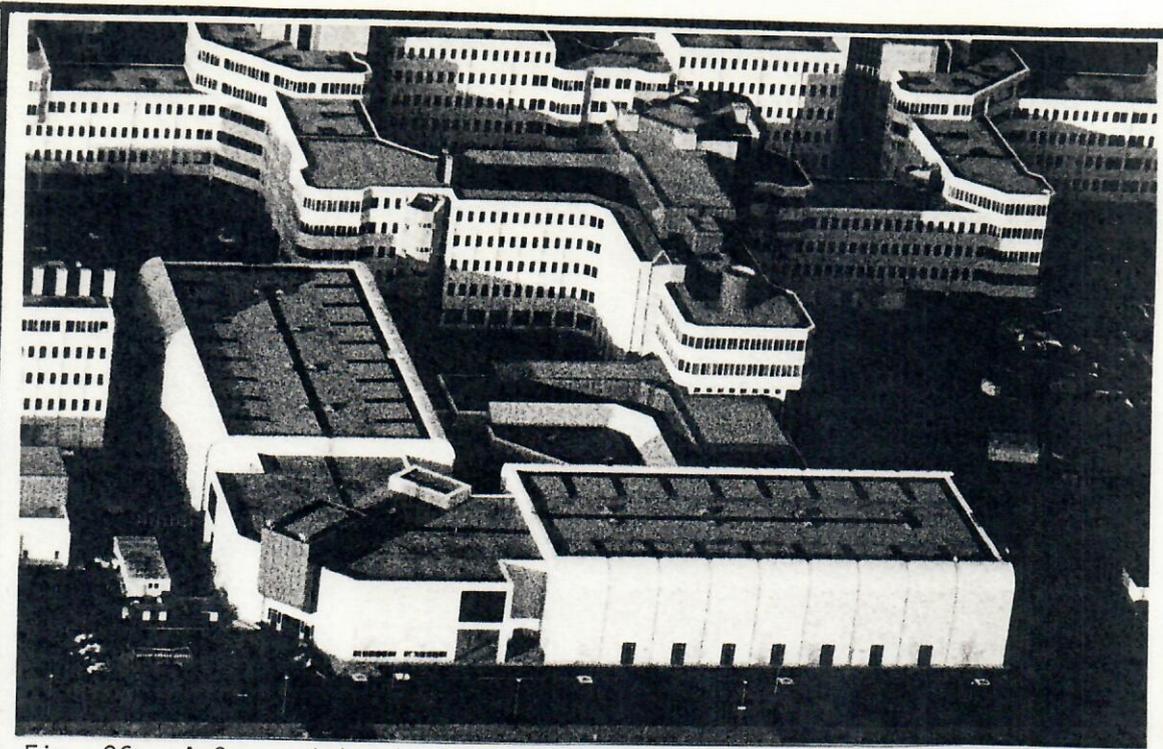


Fig. 26: Außenansicht der Reinräume ($2 \times 2000 \text{ m}^2$) von SIEMENS in München-Perlach. Der eigentliche Reinraum reicht vom Bodenniveau bis zur Höhe der (Notausgangs)türen. (Luftbildfreigabe Reg. Obb. G16/2051).

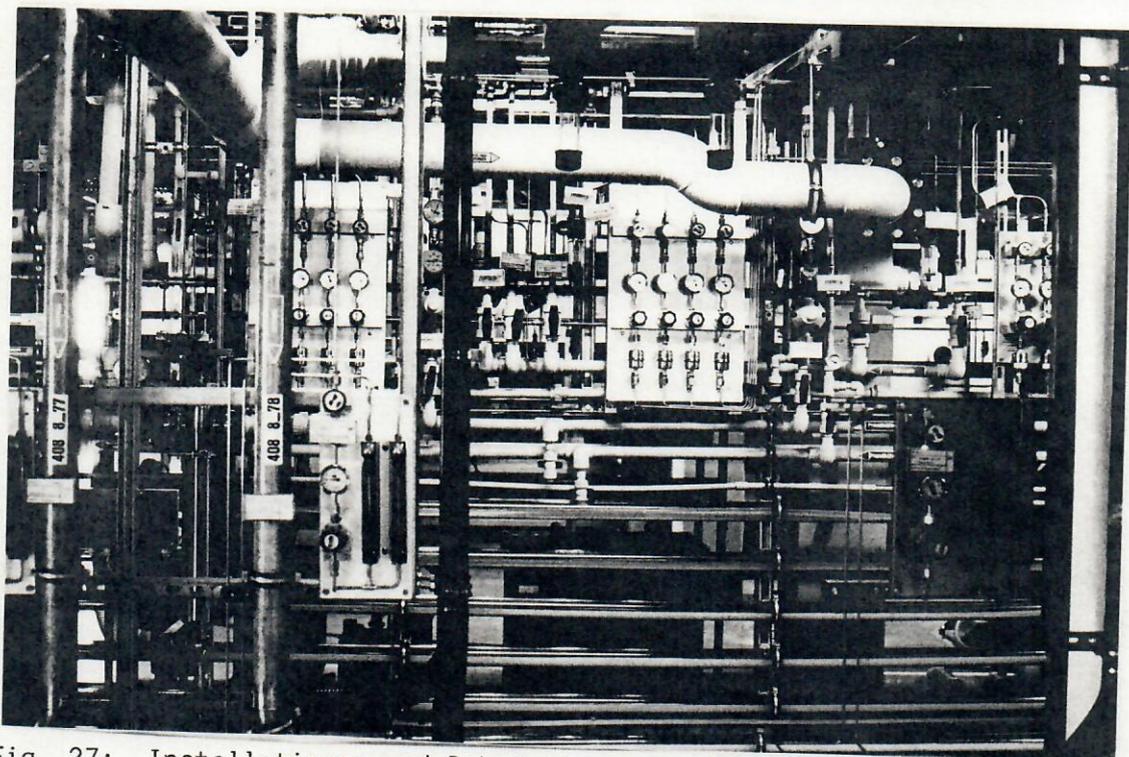


Fig. 27: Installationen und Rohrleitungen für Medien im "Keller". der Reinräume.

Die Herstellung von ultrareinem Wasser und die restlose Klärung des Abwassers bedingen eine beachtliche "Chemiefabrik" im Keller oder in unmittelbarer Nähe des Reinraumes. Die Zuführung der benötigten flüssigen Chemikalien erfolgt ebenfalls von zentralen Versorgungsstationen. Die Vielzahl der notwendigen Prozeßgase wird zu einem kleinen Teil von großen Zentralstationen aus zugeführt (N_2 , O_2 , Ar, H_2 , ...), die meisten Gase kommen jedoch aus Druckflaschen die in separaten "Bunkern" außerhalb der Umluft stehen. Alle Rohrleitungen, Ventile, Meßgeräte, etc., durch die Medien fließen, müssen speziell auf Partikelfreiheit hin konstruiert sein; für Gasleitungen werden z. B. innenpolierte Spezialstahlrohre verwendet. Die Gase müssen ebenso wie die Abwässer wieder entsorgt werden. Dazu werden Gaswäscher eingesetzt, die schädliche Stoffe auswaschen und im Abwasser binden. Alle Medien müssen darüberhinaus ständig auf Reinheit kontrolliert werden, sowohl am Einspeisepunkt als auch am Verbrauchspunkt.

Der "Keller" eines Reinraumes ist deshalb mit einer beachtlichen Vielfalt von Leitungen aller Art durchzogen (Fig. 27), die allein einen Wert von einigen zig Millionen darstellen.

2. Elektrizitätsversorgung

Die Prozeßanlagen sind häufig Stromfresser. Ein Epitaxiereaktor, beispielsweise, mit einer Standfläche von ca. $2m^2$ benötigt 150 kW nur für die Lampen mit denen "geheizt" wird; weitere 20 kW für die Kühlung der Lampen über Gebläse. Da ein Ausfall der Stromversorgung katastrophale Folgen hätte (in einer Fertigung befinden sich zu jedem Zeitpunkt ungefähr 50 000 Scheiben, davon 10 000 in Bearbeitung. Falls bei einem Prozeßausfall die Hälfte davon "abstürzt", kostet dies erheblich mehr als eine Million DM). Deshalb gibt es eine Notstromversorgung, die nach einigen Sekunden anspringt (die kurze Unterbrechung ist bei vielen Prozessen, z. B. Öfen, tolerierbar) und zusätzlich eine unterbrechungsfreie Stromversorgung mit geringerer Leistung, die die Steuercomputer etc. am Laufen hält.

3. Sicherheitstechnik

Vom Standpunkt der Sicherheit und des Umweltschutzes ist die Mikroelektronikfertigung eine Tätigkeit, die mit potentiell gefährlichen physikalischen Methoden chemische Prozesse durchführt; wobei einige der giftigsten oder korrosivsten Substanzen der anorganischen Chemie zum Einsatz kommen. Betrachten wir die physikalische Seite: Sehr hohe Spannungen (< 200 kV oder mehr bei Implantern); sehr hohe Frequenzen (GHz bei einigen Plasmaätzen); sehr hohe Leistungen und Leistungsdichten (100 kW in ca. $20cm^3$ bei Rapid Annealern); hohe Temperaturen ($1200^\circ C$ bei Ofenprozessen); Vakuum in Gefäßen mit m^3 -Inhalt; eventuell Hochdruck (10 bar in ca. $0,5m^3$ Volumen bei Hochdruckoxidation), automatisch ablaufende mechanische Bewegungen hoher Energie ("Fallbeil-Ventile"). Die Chemie bietet extrem toxische Gase wie Arsin (AsH_3), Phosphin (PH_3); explosive (in Kontakt mit Luft) oder leicht brennbare Gase wie Silan (SiH_4), Wasserstoff (H_2) sowie korrosive Gase wie Chlor

(Cl₂) oder Bortrifluorid (BF₃). Dazu kommen noch einige Organika wie Lösungsmittel oder Spezialgase.

Die mit der Entsorgung dieser Stoffe verbundenen Aufgaben des Umweltschutzes wurden schon angesprochen. Die überwiegend anorganische Natur der Chemikalien erleichtert hier die Aufgabe; eine moderne Linie wird heutzutage keine nennenswerte Emission bei Abwässern und Gasen mehr aufweisen.

Die angesprochenen sicherheitstechnischen Aspekte sind gravierend und genießen bei Bau, Einfahren und Routinebetrieb des Reinraums höchste Priorität. Alle kritischen Gasflaschen oder Chemikaliertanks sind beispielsweise in geschlossenen Kabinen untergebracht, die unter Unterdruck stehen (dauernde Absaugung). Entlang kritischer Medienpfade sind chemiespezifische Detektoren angebracht, die bei Lecks über ein zentrales (ausfallgeschütztes) Rechnersystem Alarm auslösen und alle Ventile schließen, falls Grenzwerte überschritten werden die noch deutlich unter den arbeitsrechtlich zulässigen MAK-Werten liegen (zum Vergleich: Der MAK-Wert (maximal zulässige Arbeitsplatzkonzentration) einiger anorganischer Gifte (z. B. Blausäure, HCN) ist kleiner als die Konzentration dieser Gifte im Rauch einer Zigarette).

Bei potentiellen Unfällen mit Gasen (den Autoren sind bis jetzt keine bekannt), helfen die ungeheuren zur Verfügung stehenden Luftmengen: Im betroffenen Sektor des Reinraums werden die zugehörigen Luftumwälzer blitzschnell hochgefahren, die Luft aber quantitativ nach außen geführt. Die riesigen, durch den Sektor strömenden Luftmengen können selbst größere Giftgasmengen in kürzester Zeit auf ungefährliche Konzentrationen verdünnen. Selbstverständlich werden die sicherheitstechnischen Maßnahmen (reiner Hardware-Wert ca. 20 Million DM) durch ständige Schulungen des Personals sowie sofort verfügbare Rettungsgeräte und Medikamente ergänzt.

16.6.3 Durchlaufzeit, Kapazität, Ausbeute

Im Reinraum werden Siliziumscheiben bearbeitet. In einer Fertigung, die (im Idealfall) nur ein Produkt herstellt, erfährt jede Scheibe exakt die gleiche Prozeßsequenz, auch jeder einzelne Prozeßablauf ist immer der gleiche. Jede Scheibe die "vorne" reingeht, kommt (hoffentlich) "hinten" wieder raus; die dazu benötigte Zeit heißt Durchlaufzeit (DLZ). Die Zeit, die benötigt würde, um die Scheibe unter Idealbedingungen zu prozessieren (d.h. keine Unterbrechungen im Prozeßablauf) heißt physikalische Durchlaufzeit (phys. DLZ).

Die Kapazität einer Linie wird in der Regel durch die Zahl der wöchentlich einlaufenden Scheiben ausgedrückt (WSPW = Wafer starts per week). Die Zahl der fertig prozessierten Scheiben, die "hinten" herauskommen, sollte idealerweise pro Woche (im Mittel) genauso groß sein, in der Realität ist sie kleiner da mal eine Scheibe zerbricht oder, bedingt durch einen Anlagenausfall während der Bearbeitung, die in der Anlage befindlichen Scheiben "abstürzen" und nicht weiterbearbeitet werden. Das Verhältnis zwischen Output und Input (in Prozent) heißt Scheibenausbeute Ys.

Die Zahl der benötigten Anlagen für eine bestimmte Einschleusrate folgt aus dem Durchsatz der Anlage (z.B. in Scheiben/hr) und ihrer Verfügbarkeit ("uptime"), dem prozentualen Anteil der Zeit in der die Anlage im Mittel genutzt werden kann. Die Anlagenverfügbarkeit ist niemals 100%; zu den diversen Arten der vorhersehbaren Wartung (festes Zeitraster, z.B. einmal pro Woche Reinigung; nutzungsbedingtes Raster, z.B. Pumpenölwechsel nach x Betriebsstunden; parameterbedingter Raster, z.B. durch überschrittenen Grenzwert) kommen die unvorhergesehenen Ausfälle, häufig ohne sofort ersichtliche Ursache.

Im Gegensatz zu einem Fließband, gibt es nicht für jeden Prozeßschritt eine (oder mehrere) eigene Anlagen. Beispielsweise mag ein Implanter für alle (ca. 10) Implantationen mit kleiner oder mittlerer Dosis ausreichen, während bestimmte Schichtabscheidungen drei oder mehr Anlagen erfordern. Eine Scheibe läuft also nicht linear durch die Linie, sondern kommt zu verschiedenen Prozessen immer wieder an einunddiesselbe Anlage, während sie für einen einzigen Prozeß an eine von mehreren (identischen) Anlagen gelangen kann.

Eine Fertigung arbeitet nur dann effektiv, wenn der Fluß der Scheiben durch die verschiedenen Prozesse einigermaßen gleichmäßig erfolgt und auch durch die unvorhersehbaren Anlagenausfälle nicht stark gestört wird. Aufgabe der Liniensteuerung ist es, durch Eingriffe bei der Scheibeneinschleusung, durch Umleiten von Scheiben, durch Prioritätsfestlegungen, etc. für den optimalen Scheibenfluß zu sorgen. Dabei kommt erschwerend hinzu, daß zwischen einzelnen Prozeßschritten oft Zwangskopplungen bestehen; z. B. muß ein Oxidationsprozeß relativ rasch nach der vorhergehenden Scheibenreinigung erfolgen.

In einer Forschungs- und Entwicklungslinie wird die Logistik erheblich unübersichtlicher. Betrachten wir den konkreten Fall der Entwicklung des 16M DRAMs. Es muß auf dem 4M DRAM aufbauen; unsere F+E-Linie muß also in der Lage sein den gesamten 4M DRAM Prozeß (ca. 400 Prozeßschritte) mit einer Mindestausbeute durchzuführen. Die einzige Möglichkeit, dies sicherzustellen, besteht darin, laufend 4M DRAMs zu machen. Erfahrungsgemäß genügt es nämlich nicht, jeden Prozeß für sich zu überwachen, da bei vielen Prozessen die für das Gelingen des 4M DRAM Gesamtprozesses wichtigen Kriterien nicht wirklich bekannt sind oder einer direkten Messung nicht zugänglich sind. Die F+E Linie fährt also eine sogenannte "base-load"; die dabei entstehenden Chips werden selbstverständlich verkauft - unsere F+E Linie ist zu einem Teil ihrer Kapazität eine Fertigung!

Nachdem das 16M DRAM neue oder bessere Prozesse benötigt, müssen diese in der F+E Linie entwickelt werden. Zu besonders neuartigen oder kritischen Prozessen wurden schon vorher in Labors Grundlagenuntersuchungen durchgeführt (z. B. zur Wolfram-Abscheidung); in die Linie geht ein neuer Prozeß erst wenn schon ein Minimum an Prozeß- und Geräte-Know-how vorliegt. Die Entwickler beginnen, gezielt 16M-DRAM Scheiben, einzuschleusen. Das Spektrum reicht von Scheiben, die mehr oder

weniger direkt nur zu einem neuen Prozeß gehen, und deshalb nur wenige Prozeßschritte sehen, bis zu Scheiben die den gesamten 16M-DRAM Prozeß in der bis dahin existierenden Form (ca. 450 Prozeßschritte) durchlaufen. Alle beteiligten Entwickler (Designer, Device-Spezialisten, Einzelprozeßentwickler und Gesamtprozeßentwickler) haben zu ihren speziellen Themen Scheiben in der Linie und wollen nur eines: Sie wollen diese Scheiben schnell und mit halbwegs funktionierenden oder vermeßbaren Strukturen.

Es ist klar, daß die Entwicklung und Optimierung eines 16M DRAM Prozesses mit vielen neuen Strukturen und Prozessen einige Lernzyklen, also Komplettdurchläufe, braucht. Wie lange dauert so ein Lernzyklus?

Eine Abschätzung dazu ergibt sich aus der physikalischen Durchlaufzeit des Gesamtprozesses, die mit guter Genauigkeit errechnet werden kann, und dem Flußfaktor (FF) der Linie. Der Flußfaktor ist der Quotient aus der tatsächlichen Durchlaufzeit und der physikalischen Durchlaufzeit; er gibt an um wieviel länger ein neuer Durchlauf im Vergleich zum idealen Durchlauf dauert.

Die physikalische Durchlaufzeit eines 16M DRAMs liegt bei ca. 600 hr. Bezogen auf eine 38,5 hr Woche sind dies 15,6 Wochen; bei 2-Schicht Betrieb (80 hr/Woche) 7,5 Wochen; bei Kontischicht (7 Tage die Woche, 24 hr pro Tag) 3,57 Wochen. Selbst bei auf Händen getragenen Scheiben, die oberste Priorität genießen, ist ein Flußfaktor von 2 ein sehr guter Wert, für normal laufende Entwicklungs-Chargen kann man einen Flußfaktor von 4 annehmen /46/. Die Durchlaufzeit einer Eil- oder Normalcharge beim 1-Schicht Betrieb der Linie liegt damit bei 31,2 bzw. 62,4 Wochen (in Wahrheit sogar noch höher, da ca. 1 hr der Arbeitszeit von 8 hr für das "Einschalten" oder "Ausschalten" verlorengelht) - der Entwickler hat, bis er Si bekommt, längst vergessen, was er eigentlich damit bezweckte.

Warum ist der Flußfaktor allenfalls 2? Neben einer Fülle von Trivialgründen (Anlage gerade besetzt, selbst bei Priorität muß Ende der laufenden Fahrt abgewartet werden, Bearbeiter krank/beim Essen, ...) die sich selbst bei optimaler Planung nicht immer vermeiden lassen, gibt es in der Entwicklungsphase zwei unvermeidliche Engpässe, die immer wieder auftreten: Neuartige, noch "wackelige" Prozesse, die nur von dem Entwickler selbst durchgeführt werden können, und unvorhergesehene Probleme die zu Abänderungen des vorgesehenen Ablaufs führen (und dazu einer Entscheidung oder gar einiger schneller Versuche bedürfen).

Aus der Marktlage (z.B. 16M DRAM muß 1992 fertig sein) und der Zahl der benötigten Lernzyklen ergibt sich zwingend eine maximale Durchlaufzeit von 4 Monaten in der früheren Entwicklungsphase und ca. 2,5 Monate in der späteren Entwicklungsphase. Die unvermeidliche Konsequenz ist Schichtarbeit (mindestens 3-Schicht Betrieb), auch in der F+E Linie.

Da in der Entwicklungsphase die Anwesenheit der hochqualifizierten Entwickler (in der Regel diplomierte und

promovierte Physiker/Ingenieure/Chemiker) bei vielen Prozeßschritten zwingend erforderlich ist; dies aber während der Nachtschicht praktisch nicht möglich ist, ist Samstag/Sonntag Arbeit auch in der F+E Linie abzusehen. Dies gilt für alle Wettbewerber, die zeitgleich mit den führenden Japanern am Markt vertreten sein wollen (wer das nicht will begeht betriebswirtschaftlichen Selbstmord), und ganz besonders für diejenigen, die sogar noch etwas schneller entwickeln müssen als die Japaner, weil sie einen Rückstand aufholen müssen. Es gilt insbesondere auch für rein universitäre Forschungsreinräume oder Mikroelektronikzentren die "state-of-the-art" Gesamtprozesse entwickeln wollen.

Die F+E Linie muß also auf Durchlaufzeit optimiert sein. Dies bedeutet, daß sie nur etwa 50% ihrer Kapazität nutzen kann (sonst kommt zwar mehr raus, aber es dauert länger) und damit erheblich höhere Kosten pro bearbeitete Scheibe, hat als eine vergleichbare Fertigung die auf maximalen Output fährt. Eine weitere spezielle Randbedingung ist die Forderung nach "Absturzfreiheit". Da jede Entwicklungscharge einzigartig ist (oft sind sogar die Scheiben innerhalb einer Charge bei bestimmten Prozessen noch "gesplittet, d.h. die Charge besteht aus Scheibenpaketen à 5 Scheiben,, die sich in einigen Prozessen noch unterscheiden), ist jeder "Absturz" eine Katastrophe. In einer Fertigung ist dies anders; falls eine Charge abstürzt, kommt sofort die nächste; Maßnahmen, die den Gesamtausstoß stärker erhöhen als die Absturzrate (z.B. die Einführung einer Nachtschicht in den ersten Monaten), sind für eine Fertigung betriebswirtschaftlich sinnvoll, nicht jedoch für eine F+E Linie.

16.6.4 Equipment und Medien

Für jeden der 400-500 Prozeßschritte, die zur Herstellung eines 4M- oder 16M-DRAMs benötigt werden, wird ein Gerät oder eine Maschine - kurz ein "equipment" - benötigt. Das Spektrum reicht dabei vom Hochstrom-Implanter als größtem Brocken (Flächenbedarf ca. 40m^2 ; Preis ca. $5 \times 10^6\text{DM}$) bis hin zu Meßgeräten (die als Anlagenzubehör zu sehen sind, beispielsweise Ellipsometer für Schichtdicken (Flächenbedarf ca. 1m^2 ; Preis ca. $5 \cdot 10^5\text{ DM}$)). Eine Linie die einen 16M-Speicher entwickeln will, braucht Equipment für mindestens $2 \cdot 10^8\text{DM}$.

Der größte Teil dieses Equipments wurde schon für die Vorgängergeneration verwendet, die schärferen Anforderungen der neu zu entwickelnden Generation bedingen dann bei vielen Prozeßschritten einen Betrieb außerhalb der vom Hersteller garantierten Leistungsfähigkeit (Beispiele: Schwankungen der Schichtdicke über die Scheibe, Zahl der im Gerät erzeugten Partikel, Meßgenauigkeit). Neues, speziell für die neu zu entwickelnde Generation hergestelltes Equipment, hat notgedrungen Prototyp-Charakter; d.h. hat seine Leistungsfähigkeit noch nicht bewiesen und wird im Betrieb öfter ausfallen.

Die mit dem Betrieb einer Anlage verbundenen Aufgaben: Routinewartung, Reparatur bei Ausfall, Einfahren nach

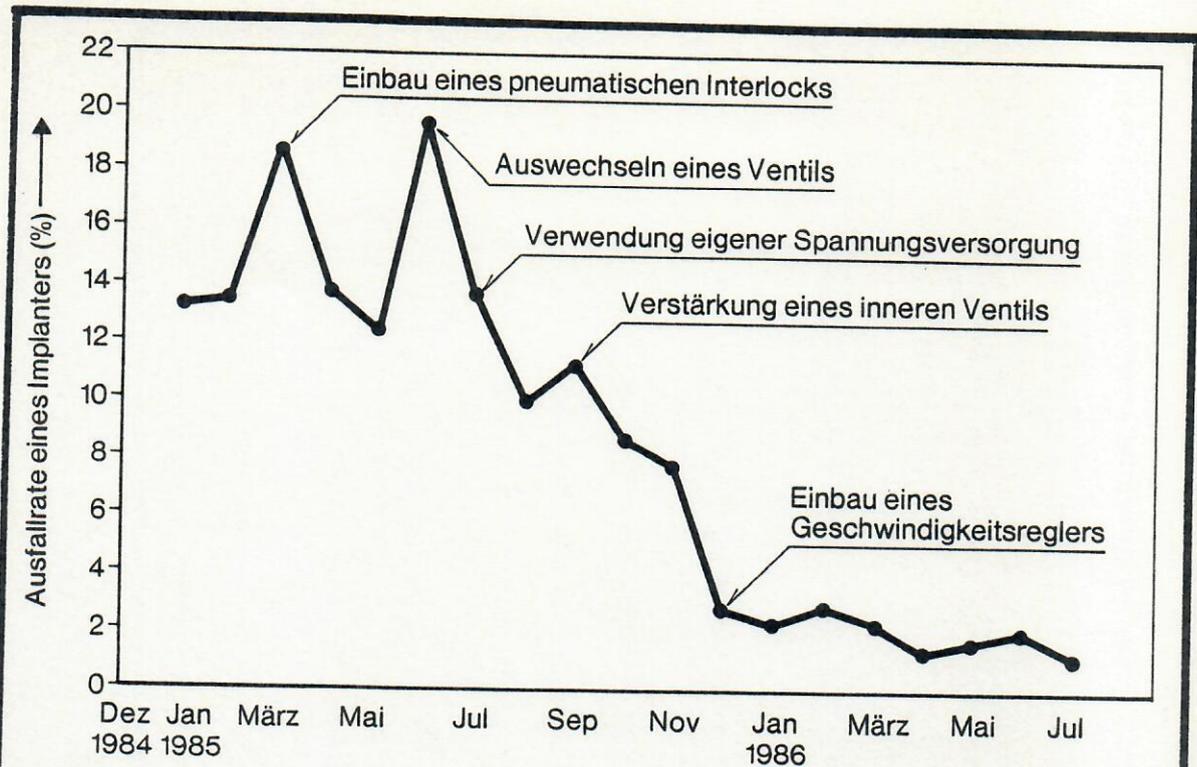


Fig. 28: Entwicklung der "Uptime" bzw. der Ausfallrate eines Implanters und vorgenommene Maßnahmen (Quelle /34/). Kleine Ursachen - große Wirkung.

Elektrischer Widerstand	18.2 MΩcm
TOC (Total Organic Content)	5 ppb
Max. Partikelzahl pro Liter (bestimmt durch Rasterelektronen- mikroskop)	0,1-0,2µm < 1000 0,2-0,3µm < 500 0,3-0,5µm < 10 >0,5µm < 1
Max. Partikelzahl pro Liter (bestimmt durch "on-line" Messung durch Lichtstreuung)	0,3-0,5µm < 10 >0,5µm < 1
Bakterien pro 100 ml	< 1
Gelöste Silikate	< 1 ppb
Bor	< 0,005 ppb
Ionen:	Na+ 0.025 ppb K+ 0,05 ppb Cl- 0,025 ppb Br- 0,05 ppb Max. Summe <0,2 ppb
Verdampfungsrückstand	<0,05 ppb

Tabelle 5: Zielspezifikation für H₂O für das 4M-DRAM (Quelle /35/).

Unterbrechung, Überwachung der Prozeßqualität, Fehlersuche bei Verletzung der Prozeßspezifikation ohne erkennbare Ursache, Verbesserungen durch gezielte Änderungen der Hard- und Software, muß die Linie zu einem beträchtlichen Teil selbst durchführen. Dazu wird eine hochspezialisierte Mannschaft benötigt, deren oberstes Ziel die Erreichung einer möglichst hohen "Up-time" ist. Die "Up-time" ist der Prozentsatz der Gesamtarbeitszeit, in der ein Equipment für den Prozeß uneingeschränkt verfügbar ist. Fig. 28 zeigt die zeitliche Entwicklung der Up-time und die für die "Down-time" verantwortlichen Ausfallgründe. Das "Up-time"-Spektrum des Equipments einer F+E-Linie reicht von traumhaften 96% bis (leider) weit unter 50%; wobei keine unmittelbare Korrelation zwischen Komplexität eines equipments und seiner up-time existiert. Nur am Rande sei vermerkt, daß die "Up-time" allein noch nicht ausreicht, um die Verfügbarkeit einer Anlage zu bewerten; theoretisch hätte auch eine Anlage die alle Stunde einmal für 5 min ausfällt noch eine (sehr gute) Up-time von 91%.

Eine Anlage gilt dann als ausgefallen und ist für den Prozeß gesperrt, wenn die Prozeßspezifikationen nicht mehr erreicht werden können. Dies kann sofort erkennbare Gründe haben (z.B. Quarzrohr gebrochen), anlagentechnische (z.B. Enddruck wird nicht erreicht), prozeßtechnische (z.B. Schichtdicke 5% über Sollwert) oder auch verfahrenstechnische (z.B. Stillstand > 24h). Nach Behebung der Störung muß in der Regel über Testfahrten nachgewiesen werden, daß Anlage und Prozeß wieder voll die Spezifikationen erfüllen.

Es versteht sich von selbst, daß das beste am Markt erhältliche Equipment gerade gut genug ist, um während der F+E Phase der neuen Generation auswertbare Chips zu erhalten. Manche Prozesse sind kritischer als andere, praktisch keiner ist völlig unkritisch. Um größtmögliche Reproduzierbarkeit der Prozesse bei minimaler Partikelbelastung und Kontamination sicherzustellen, sind nahezu alle Anlagen vollautomatisch; d. h. sie entnehmen die Scheiben aus der Box, transportieren sie zur Prozeßkammer, führen den Prozeß durch, machen (zum Teil) Messungen zur Prozeßkontrolle und bringen die Scheibe zurück in die Box. Schnittstelle zum Bearbeiter ist zunehmend nur noch der Computerbildschirm. Die Automatisierungs- und Überwachungsperipherie ist dabei häufig viel aufwendiger und teurer als die Prozeßkammer (ein "nackter" Lampenofen für "rapid thermal processing" für Handbetrieb im Labor kostet z.B. ca. $7 \cdot 10^4$ DM; derselbe Lampenofen kostet voll automatisiert ca. $7 \cdot 10^5$ DM).

Equipment für die Entwicklung neuer Chipgenerationen ist nicht nur komplex und teuer sondern kommt auch vorwiegend aus Übersee; nämlich aus den USA und aus Japan. Europäische Hersteller sind - von wenigen Ausnahmen abgesehen - in modernen subµm-Linien zunehmend weniger vertreten.

Für europäische Chiphersteller ergeben sich daraus einige Nachteile:

Transport-, Zoll-, Service- und Ersatzteilkosten sind naturgemäß höher als bei der überseeischen Konkurrenz. Darüber hinaus haben japanische Halbleiterhersteller ungefähr ein Jahr Vorlauf bei der Erprobung und Bewertung von japanischem Equipment, da japanische Firmen in der Regel ihre Anlagen erst dann exportieren, wenn sie sich im Inland schon bewährt haben.

Die überwiegende Mehrzahl der Prozesse verbraucht Material: Neben Unmengen von ultrareinem Wasser und H_2O_2 weitere flüssigen Medien; z.B. Säuren (HF , HCl , HNO_3 , H_3PO_4 , CH_3COH), Laugen (Cholin, NH_4). Lösungsmittel (Propanol, Aceton), Fotolacke, Spin-on-glas (Kolloidale Lösung aus SiO_x); TEOS ($(C_2H_5O)_4Si$), etc.. Gasförmige Medien sind z.B. Dotiergase (AsH_3 , PH_3 , BF_3 , B_2H_6 , $POCl$), Gase zur Schichtabscheidung (SiH_4 , SiH_2Cl_2 , NH_3 , WF_6), Ätze (CF_4 , O_2 , CHF_3 , $CClF_3$, $CBrF_3$, SF_6 , NF_3 , Cl_2 , CCl_4 , BCl_3) und die Inertgase He, Ar, N_2 sowie H_2 und O_2 . An festen Materialien werden benötigt: Si-Scheiben, Targets für die Sputterverfahren (Al (mit definierten Zusätzen), Mo, Ti, Ta, Si) und die diversen Materialien für die Montage (Keramikplättchen, Plastikgußmassen, etc.).

An alle Materialien werden extreme Reinheitsanforderungen gestellt; Tabelle 5 zeigt ein Beispiel für eine Materialspezifikation. Dies hat drei Konsequenzen: Erstens ist damit die Zahl der Lieferanten beschränkt (zunehmend auf japanische), zweitens wird Material enorm teuer, und drittens braucht man eine aufwendige und teure Eingangskontrolle. Denn jedes neue Materialgebilde, das in einem Prozeß eingesetzt wird und wegen unerkannter Kontamination oder sonstiger Abweichung von der Spezifikation (z.B. ist die Restfeuchte - der Gehalt an H_2O - in manchen Gasen sehr kritisch) Ausbeuten erniedrigt oder Lernzyklen in der F+E Phase beeinträchtigt, kann viele Millionen DM kosten.

16.6.5 Prozeßkontrollen und Ausbeuten

Bei 450 Prozeßschritten und einer erhofften Ausbeute von 80% muß die Ausbeute der einzelnen Prozeßschritte 99,95% betragen. Bei nur 99% Ausbeute der Einzelprozesse beträgt die Gesamtausbeute $0.99^{450} = 1\%$. Dies ist zwar sicherlich eine Milchpersonrechnung, zeigt aber doch deutlich das Problem: Ohne schärfste Überwachung von Anlage, verwendeten Medien und Prozeß an jeder einzelnen Stelle ist der Mißerfolg eines längeren Durchlaufs praktisch garantiert.

Ständige Überwachung aller relevanten Parameter und Zustandsgrößen der Anlagen und Prozesse und die Auswertung der erhaltenen Datenflut beschäftigt nicht nur einige Mitarbeiter; die sinnvolle Erstellung der Kontrollvorschriften gehört mit zu den Aufgaben des Prozeßentwicklers. Die damit verbundene Festlegung des "Prozeßfensters" nimmt häufig mehr Zeit in Anspruch als die Entwicklung des eigentlichen Prozesses. Wenn man sich unter Prozeßentwicklung die Festlegung von Material, Anlage und Parametersatz für einen bestimmten Prozeß vorstellt, heißt die Festlegung des Prozeßfensters die plus/minus Toleranz jedes Parameters zu definieren (siehe Beispiel in Tabelle 6). Bei der Vielzahl der Parameter läßt sich das mit endlichem

Beispiel: Dotierung des 4Mbit-Grabenkondensators /44/

Spezifikation: Arsenkonzentration von ca. $5 \times 10^{19} \text{ cm}^{-3}$ an der Grabenwand, Eindringtiefe ca. 150 nm nach Prozeßende.

Festzulegende und zu überwachende Parameter sind u.a:

Prozeßmaterialien: Reinheit der Prozeßflüssigkeiten Tetraäthylorthosilikat und Triäthylorthoarsenit sowie der Prozeßgase Sauerstoff und Stickstoff.

Abscheidung: Temperaturen der Verdampferquellen und der beheizten Leitungen, Menge des durchperlenden Trägergases, Temperatur und Druck der Abscheidung, Scheibenposition und -abstand, genaues Temperaturprofil (räumlich und zeitlich) incl. Ein- und Ausfahren.

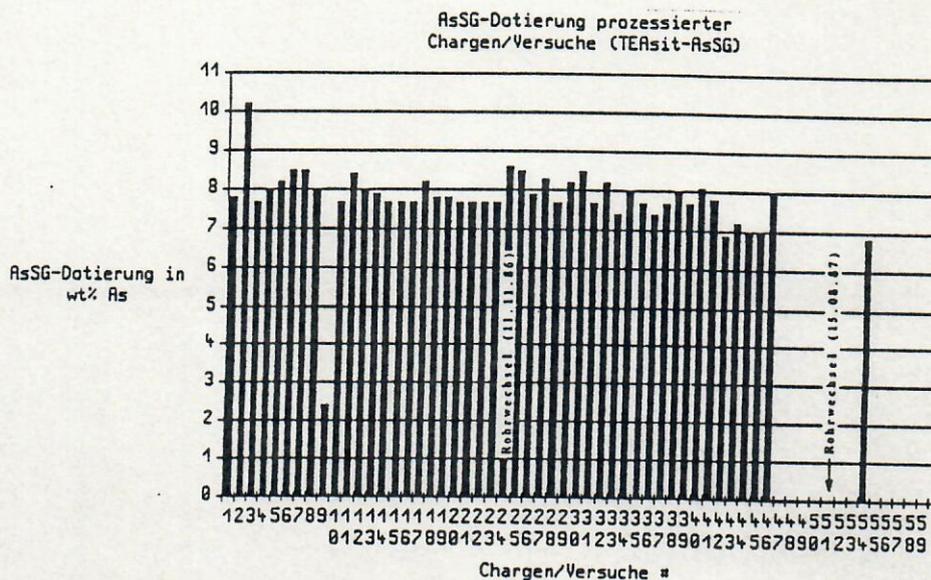
Scheiben: Grundmaterial, vorige Schritte (z.B. Reinigungen).

Diffusion: Atmosphäre, räumliches und zeitliches Temperaturprofil.

Abätzung: Ätzflüssigkeit zur Arsenglasentfernung, Temperatur, Zeitdauer.

Meßtechnik: Kontrolle des Arsengehalts im Glas und im Silizium (an speziellen Testscheiben), Entwicklung und Eichung solcher Verfahren, Gleichförmigkeit des Dotierresultats, Partikelfreiheit, Freiheit von Arsenglas-Rückständen, Qualität der Siliziumoberfläche am Prozeßende.

Zusätzlich müssen die anschließenden Prozeßschritte so abgestimmt werden, daß am Ende z.B. nicht durch zusätzliche Temperaturbelastung eine zu starke Ausdiffusion des Arsengebiets um den Graben auftritt!



Abbildungsbeispiel: Die Trendkurve der Messungen an den jeweiligen Testscheiben zeigt dem Prozeßingenieur, ob die Spezifikationen eingehalten werden, hier z.B. die Arsenkonzentration im Glas.

Tabelle 6: Beispiel für die bei einer Prozeßentwicklung zu berücksichtigende Parameter (mit +/- Toleranzen).

Zeit- und Kostenaufwand nicht mehr über die klassische Methode (einen Parameter variieren, alle anderen konstant halten), sondern nur über sogenannte statistische Versuchsplanung erreichen /36/. Dabei wird die Arbeit oft dadurch erschwert, daß die Spezifikation der Zielgröße nicht unbedingt klar ist. Ein Beispiel soll dies illustrieren: Der Entwickler eines Zwischenoxid-Prozesses kann zum Schluß kommen, daß eine Schichtdickenkonstanz von $\pm 8\%$ ausreicht um alle an den Prozeß direkt gestellten Anforderungen zu befriedigen. Erst später fällt dann auf, daß diese Schichtdickenschwankung, zusammen mit den Toleranzen der schon vorhandenen Schichten, die Scheibenoberfläche jetzt so "uneben" macht, daß der Schärfentiefenbereich der Lithographie nicht mehr ausreicht, um alle Chips richtig zu belichten.

Prozeßkontrollen sind notwendig, aber zeitraubend und teuer. Bei Entwicklungsdurchläufen (bei denen noch in voller Breite kontrolliert werden muß) machen sie einen erheblichen Prozentsatz der Durchlaufzeit aus. Die Kunst der Prozeßoptimierung besteht deshalb auch darin, die Kontrollen auf das notwendige Minimum zu reduzieren und möglichst schnelle und billige Methoden einzusetzen.

Im Idealfall ergeben sich aus der richtigen Abfolge von optimierten und kontrollierten Prozessen automatisch funktionierende Chips bei hoher Ausbeute. Im Realfall ist das leider nicht der Fall. Ein Grund dafür ist z.B. die ungenügende Spezifikation des Prozesses/Materials mangels Wissen. Beispielsweise gibt es von der Gesamtausbeute her gesehen nach wie vor wesentliche Unterschiede zwischen Chemikalien oder sogar Si-Scheiben verschiedener Hersteller, obwohl alle Materialien die Spezifikation erfüllen und sich in allen relevanten meßbaren Parametern nicht unterscheiden. Häufig versteht man erst Jahre später, durch welche speziellen Eigenschaften sich die Unterschiede in der Ausbeute ergaben.

Ein anderer Grund ist das häufig nicht vorhersehbare Zusammenwirken zweier oder mehrerer Prozeßschritte. Sind beispielsweise Schichtdicke und Brechungsindex eine Schicht zufällig so ausgefallen, daß diese Schicht für den optischen Sensor einer Anlage, die viele Prozeßschritte später den Wafer bearbeitet als Antireflexschicht wirkt (das Vorhandensein einer Scheibe innerhalb einer Anlage wird gelegentlich durch Lichtreflektion gemessen), erkennt die Anlage nicht, daß eine Scheibe vorhanden ist und macht dementsprechend garnichts oder etwas Falsches.

Alles in allem benötigt man mindestens 1-2 Jahre, bis man Einzelprozesse und Gesamtprozeß, aber auch Design und Montagetechnik sowie das Equipment und die Materialien soweit im Griff hat, daß man reproduzierbar eine Ausbeute im 10% Bereich erhält. Erst dann kann die eigentliche Fertigung beginnen; in der Fertigungsphase wird es wiederum einige Jahre dauern, bis ein Endausbeute von ca. 80% erreicht wird.

16.6.6 Arbeiten im Reinraum

Im Reinraum herrscht totales Vermummungsgebot (Fig. 29). Menschen sind Partikelquellen erster Güte; ohne Schutzmaßnahmen umgibt uns eine Wolke von mehreren hunderttausend Partikeln, die größer sind als $0,5 \mu\text{m}$. Raucher exhaliieren noch Stunden nach der letzten Zigarette zigtausende von Partikeln im Bereich $< 0,5 \mu\text{m}$. Menschen sind außerdem Quellen des besonders gefährlichen Natriums (reichlich vorhanden im Schweiß und Spucke; normaler Fingerschweiß ätzt Silicium (was den meisten Chemikalien nicht gelingt, deswegen gehen Fingerabdrücke nie wieder weg). Make-up ist voll mit gefährlichen Metallen, von Schmuck ganz zu schweigen.

Oberstes Gebot ist also eiserne Disziplin: Korrekt getragene Reinraumkleidung; nichts anfassen ohne spezielle Handschuhe; niemals über Scheiben hantieren; peinliche Ordnung und Dokumentation. Das mag inhuman klingen - aber sind die Arbeitsplätze, wo man so richtig schön im Dreck wühlen darf, besser (z.B. Bergwerk, Stahlindustrie, Bauindustrie, ...)? Vergleichbar ist durchaus der Operationsraum einer Klinik: Vermummung, äußerste Sauberkeit, eiserne Disziplin und Schichtarbeit incl. Samstag/Sonntag.

Alle, auch die Verfasser dieser Arbeit, sind sich einig, daß ein freies Wochenende einem Wochnende mit 2×8 hr Schichtarbeit bei weitem vorzuziehen ist. Auch ist die Regelarbeitszeit (5 Tage pro Woche, $35(?)$ hr pro Woche tagsüber) bedeutend angenehmer als jegliche Schichtarbeit. Noch besser wäre, nur dann zu arbeiten, wenn man gerade Lust dazu hat (bei vollem Lohnausgleich, selbstverständlich). Allerdings würde man dann, wenn man gerade nicht arbeitet, nur ungern auf die Dienstleistungen von anderen verzichten (die dann dummerweise arbeiten müssen); zumindest aber die technische Infrastruktur nützen, in der allerdings jede Menge Chips stecken (die man mit dem Geld das man hoffentlich noch erarbeitet, in Japan kaufen muß).

Diese kleine Polemik soll nur eines illustrieren: Die Frage der Arbeitszeiten ist zunächst keine ethisch-moralische Frage, sondern eine Frage der volkswirtschaftlichen Notwendigkeit. In der Herstellung von Chips hat man allmählich den Zustand erreicht wie bei Hochöfen, Restaurants, Verkehrssystemen, Vergnügungsparks, ärztlichen Notdiensten, etc.: Will man die damit verbundenen Produkte und Dienstleistungen haben, müssen Menschen samstags und sonntags arbeiten - es gibt keine Alternative. Das zuständige Gesetz aus dem späten 19. Jahrhundert erlaubt deshalb die Wochenendarbeit als Ausnahme auch in der Industrie, allerdings nur mit einer einzigen Begründung: Der Produktionsprozeß darf keine Unterbrechung zulassen. Die Auslegung dieses Gesetzes in den vergangenen 100 Jahren hat zu der Interpretation geführt, daß u.a. ein Abfall der Ausbeute oder der Qualität bei einer Unterbrechung des Herstellprozesses diese Bedingung erfüllt. In der Chipherstellung ist dies - experimentell ermittelt - tatsächlich der Fall (Fig. 30). Wie bereits beschrieben, ist dies aber nicht der einzige sachliche Grund für die



Fig. 29 "Vermummte" Mitarbeiter im Reinraum.

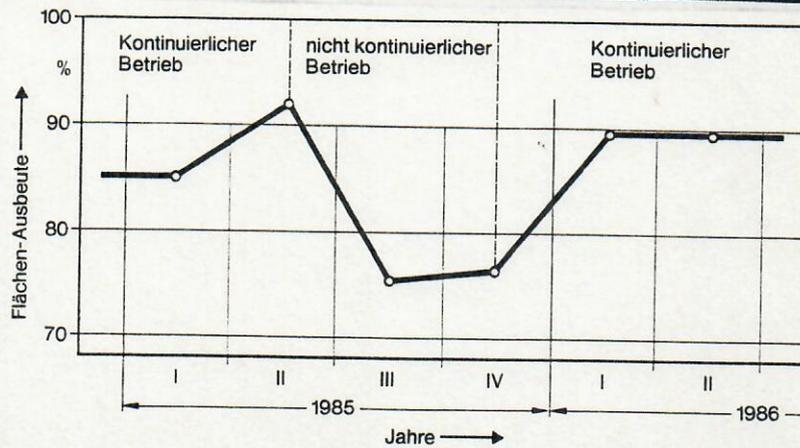


Fig. 30: Abfall der Ausbeute bei Unterbrechung der Kontinuität /37/.

3 shifts for 168 h/week
5 shift groups with average 37 h/week

Week	1							2							3							4								
	Mo	Tu	We	Th	Fr	Sa	So	Mo	Tu	We	Th	Fr	Sa	So	Mo	Tu	We	Th	Fr	Sa	So	Mo	Tu	We	Th	Fr				
Group A			L	L	N	N	N								L	L	N	N	N							L	L	N	N	
Group B	L	L					E	E	E	E	L	L							E	E	E	E	L	L						
Group C	E	E	E	E	L	L					E	E	E	E	L	L					E	E	E	E	L	L				
Group D	N	N	N	N						N	N	N	N	N						N	N	N	N	N						
Group E				E	E	L	L	L	L						E	E	L	L	L	L							E	E		

E = early shift L = late shift N = night shift

Fig. 31 Schichteinteilung beim 6/4 Modell

Kontischicht, sondern nur der einzige rechtlich verwertbare Grund. Dies bedeutet aber nicht, daß alle anderen Gründe moralisch verwerflich sind! Leider wird in der Öffentlichkeit die Problematik nicht in der gesamten Breite diskutiert, sondern immer verkürzt auf die legale Begründung oder auf eine moralisch-soziale Komponente. Welche Gründe sprechen noch für die Samstags-/Sonntagsarbeit?:

- Entwicklungszyklen sonst zu lang.
- Logistik schlechter beherrschbar (Scheiben können nicht nach jedem Prozeßschritt für 2 Tage gestoppt werden).
- Zu langsames Hochlaufen der Ausbeute in der Fertigung (hängt von Dauer der Lernzyklen ab).
- Die verglichen mit den Kapitalkosten geringen Personalkosten erfordern hohen Nutzungsgrad der Investitionen.

Wer die finanziellen Argumente nicht anerkennen will, muß bedenken, daß eine nicht konkurrenzfähige Linie mit 3-Schicht Betrieb sehr schnell gar keine Arbeitsplätze mehr hat, während eine in Kontischicht (= 5 Schichtgruppen) laufende Linie im direkten Vergleich 60% mehr Arbeitsplätze bietet, vom langfristigen Vergleich ganz zu schweigen.

Flexibilität der Arbeitszeitmodelle hilft, die zweifellos unangenehmen Folgen der Wochenendarbeit zu mildern. Die Palette der Möglichkeiten reicht vom 6/4-Modell (6 Tage arbeiten, 4 Tage frei, Fig. 31) bis zum reinen Wochenendschichtmodell (2 x 12hr Arbeit Samstag und Sonntag; die restlichen Tage frei), selbstverständlich muß auch der finanzielle Anreiz stimmen. Darüberhinaus sollte nicht vergessen werden: Wir befinden uns hier in einem klaren Wettkampf mit Fernost um die Dominanz der heutigen und zukünftigen Weltmärkte bei technischen Produkten. Wir werden in diesem Kampf nicht bestehen, wenn wir weniger und langsamer arbeiten als die Konkurrenz und dafür noch mehr verdienen wollen.

16.7 Struktur und Ablauf eines Entwicklungsprojektes für Speicher

16.7.1 Projektstruktur und Organisation

Ziel eines Entwicklungsprojektes für die nächste DRAM Generation mit vierfacher Speicherkapazität ist die Erarbeitung eines neuen Speicherchips, der:

- rechtzeitig in die Volumenproduktion gehen kann
- mit hoher Ausbeute fertigbar ist
- mit minimalen Kosten fertigbar ist
- von seinen technischen Daten her konkurrenzfähig ist
- vorhandene Infrastrukturen optimal nützt
- die Kosten der F+E-Phase minimiert.

Wird auch nur eines dieser Teilziele verfehlt, ist der wirtschaftliche Erfolg des Projektes unmittelbar gefährdet oder nicht mehr möglich. Weitere Ziele, die den wirtschaftlichen Erfolg des Projektes mittelbar beeinflussen sind:

- Schaffung einer Technologiebasis, die möglichst direkt auch für Logikprodukte verwendet werden kann.

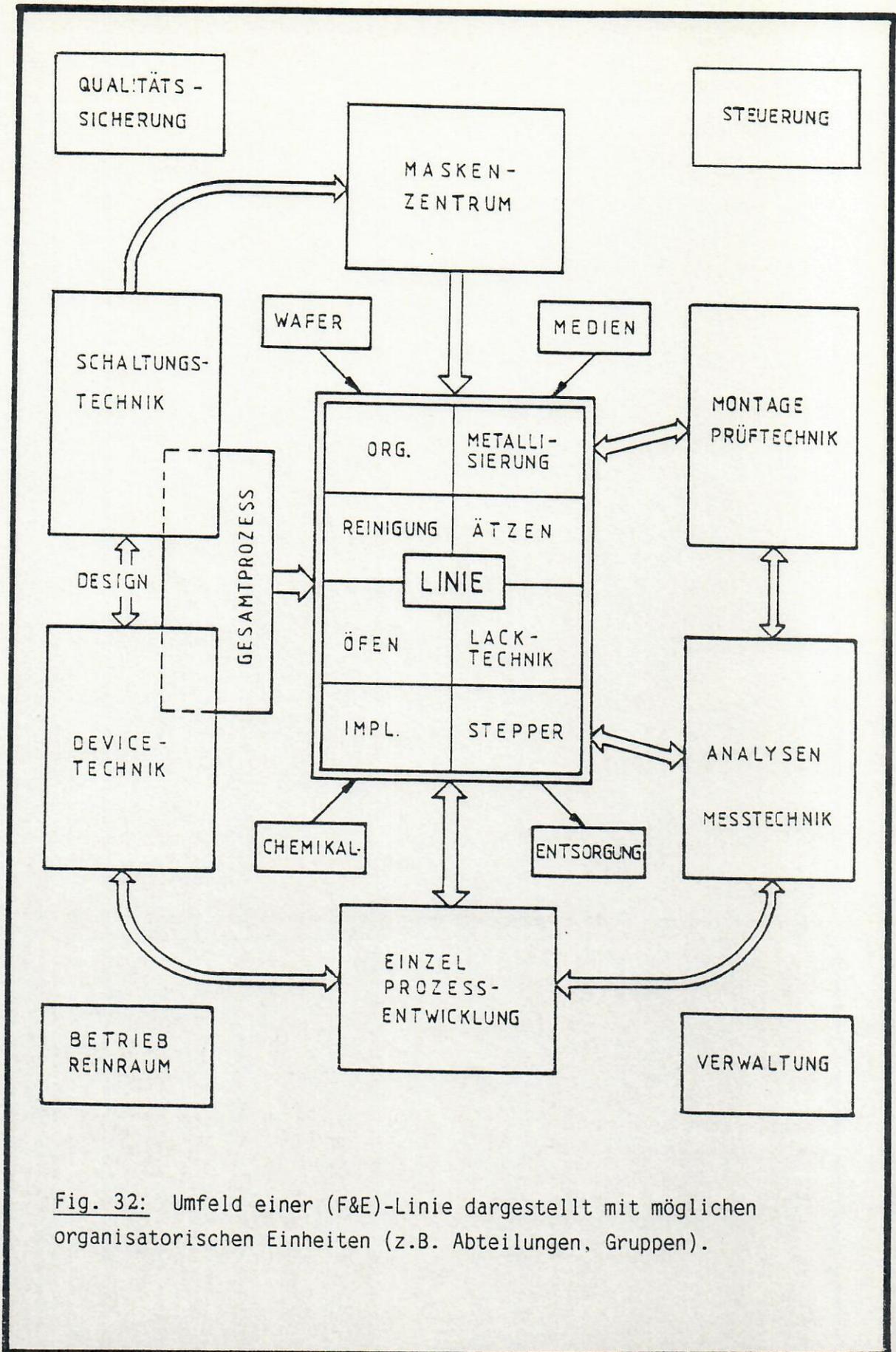


Fig. 32: Umfeld einer (F&E)-Linie dargestellt mit möglichen organisatorischen Einheiten (z.B. Abteilungen, Gruppen).

- Bereitstellung von frühen Informationen über notwendige Forschungsarbeiten für die übernächste Speichergeneration.
- "Ertüchtigung" der firmenexternen Infrastruktur (Forschungsinstitute, Equipment- und Materialhersteller, etc).

Verkürzt auf die technischen Notwendigkeiten bedeutet das, daß vor dem geplanten Anlauf der Volumenproduktion der neue Chip mit einer Mindestausbeute von ca. 10-20% aus der F+E-Pilotlinie herauskommen muß und dabei die technischen und wirtschaftlichen Kriterien erfüllt. Im Mittelpunkt des Projektes steht deshalb die Pilot-Linie, um sie herum gruppieren sich die in den vorhergehenden Kapiteln besprochenen Stellen. Fig. 32 zeigt dies schematisch, man erkennt die Instanzen Schaltungstechnik und Devicetechnik die den Chip "designen", das Maskenzentrum, das den Design umsetzt in die Masken ("Reticles") für die Lithographie, Einzel- und Gesamtprozeßentwicklung zur Schaffung der Prozeßgrundlagen; Meßtechnik, Analytik sowie Prüftechnik und Montage. Die Linie selbst hat ihre Feinstruktur mit Prozessen, Reinraumtechnik, Organisation etc. Alles in allem sind ca. 1500 Arbeitsplätze involviert. Diese ganze komplexe Struktur ist einerseits irgendwie in Bereiche, Abteilungen, Gruppen, etc. organisiert und hat andererseits erheblich mehr zu tun als "nur" einen neuen Speicher zu entwickeln.

Um sicherzustellen, daß das Großprojekt "neuer Speicher" optimal abläuft, wird deshalb häufig eine eigene Projektorganisation ausgewiesen, bestehend aus Projektleitung mit Teil- und Unterprojekten. Die Projektleitung strukturiert das Projekt in überschaubare Teilprojekte und Unterprojekte, definiert - in Zusammenarbeit mit den betroffenen Experten - die Arbeitsinhalte und ihre zeitliche Abläufe, plant die dazu notwendigen Ressourcen an "manpower", Linienkapazität, Investment, etc. und überwacht und steuert den Fortschritt des Projektes. Fig. 33 zeigt eine mögliche Projektorganisation mit Teil- und Unterprojekten. Die den "Kästchen" zugeordneten Teil- und Unterprojektleiter sind dem Projektleiter gegenüber für das rechtzeitige Erreichen der Ziele ihres Teil- und Unterprojektes gegenüber verantwortlich. Im Gegensatz zum Projektleiter sind sie gleichzeitig auch noch in eine der organisatorischen Einheiten der Linie oder des Linienumfeldes eingebunden und haben damit auch noch andere Aufgaben. Die bei großen Organisationen bei gleichzeitiger Abwicklung mehrerer Projekte unvermeidlichen Interessen-, Prioritäts- und Personalkonflikte konzentrieren sich dadurch auf die Teil- und Unterprojektleiter-Engpässe, Probleme und Konflikte sind dadurch faßbar und quantifizierbar - damit sind die Voraussetzungen für zügige Problemlösungen gegeben. Es ist insbesondere Aufgabe der Unterprojektleiter, mit ihrem Expertenwissen die elementaren Arbeitseinheiten des Projektes zu definieren und rechtzeitig auf Hindernisse hinzuweisen, die das Erreichen der Ziele gefährdet. Ein Beispiel für ein erfolgreiches Projekt dieser Art ist das erfolgreich abgeschlossene "MEGA-Projekt" /38/ der Firma Siemens und Philips.

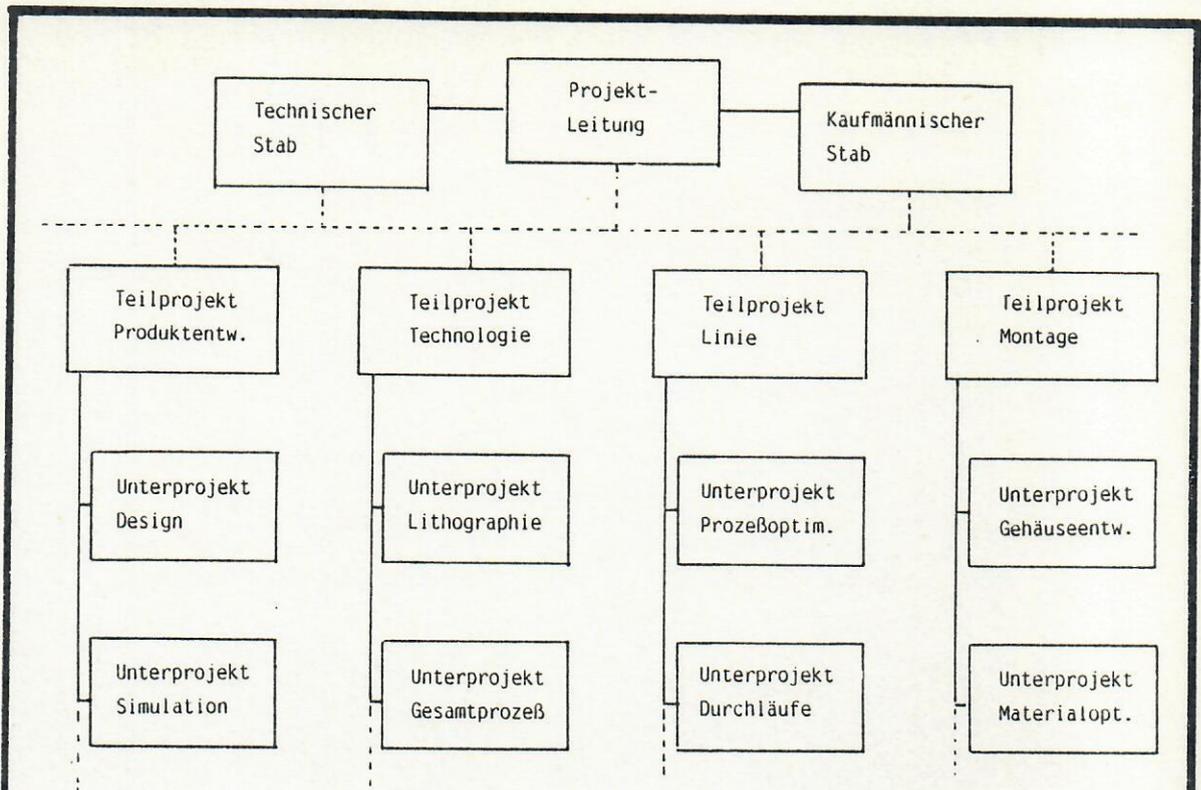


Fig. 33: Mögliche Projektstruktur. Nur der Projektleiter mit seinem Stab arbeitet 100% projektgebunden; die Teil- und Unterprojektleiter haben weitere Funktionen in der bestehenden Organisation.

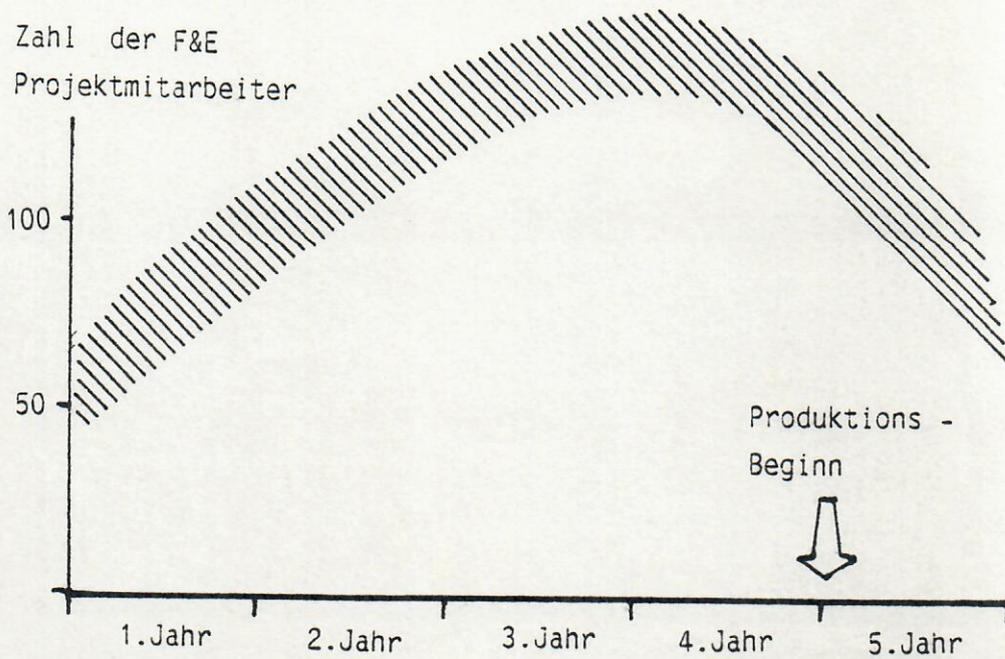


Fig. 34: Entwicklung der ungefähren Zahl der an einem Speicherprojekt mitarbeitenden Entwickler.

16.7.2 Projektplanung und -Kosten

Der besondere Nervenkitzel eines Speicherprojektes in der heutigen Zeit (also die Entwicklung von 16M-, 64M- und 256M-DRAMs in den 90-er Jahren) liegt darin, daß in der Gesamtbilanz für eine Generation es nicht allzu schwierig ist einen Gesamtverlust von mehreren 100 Millionen DM zu erleiden, aber es sehr schwierig ist mehr als ca. 100 Millionen DM Gewinn einzufahren.

Vor einigen Jahren waren obige Zahlen noch eine Null kürzer. Das war zwar auch noch viel Geld, aber für die großen Elektrotechnik-Firmen, die sich Halbleiteraktivitäten leisteten, noch im Rahmen des möglichen Risikos. Der heute beobachtete starke Trend zu Kooperationen, "Joint Ventures", Zusammenschlüssen, etc., auch und besonders unter den Großen der Branche, sowie der immer lauter werdende Ruf nach staatlicher Protektion und Förderung, kommt letztendlich daher, daß der Schritt zur nächsten Technologiegeneration ein auch für Großkonzerne nicht mehr allein tragbares finanzielles Risiko darstellt. Warum lassen sich die Firmen dann überhaupt auf ein derart riskantes und kaum profitträchtiges Geschäft ein? Die Antwort ist einfach: Wer den neuesten Speicher technologisch beherrscht, beherrscht auch die Technologie für andere Chips. Und von den knapp 4 Milliarden Umsatz der 1988 mit Halbleiterbauelementen in der BRD getätigt wurde, hängen ca. 550 Milliarden Umsatz der "Fünfergruppe" (Maschinenbau, Fahrzeugbau, Feinmechanik und Optik, Büro- und Datentechnik) /39/ ab - denn ohne die jeweils modernsten Chips läuft in diesen Sparten nicht mehr viel. Hat man diese Zahl im Auge, sind ein paar Milliarden kumulierter Verluste bei Speicherchips als kleiner Preis für die Beherrschung des von Chips abhängigen Marktes zu sehen - und in Fernost sieht man das auch so. Leider nützt diese Erkenntnis dem defizitären Halbleiterhersteller in Europa nicht viel, solange er isoliert wirtschaften muß und damit z. B. direkt in Konkurrenz steht zu der Firma NEC (IC-Umsatz 1987 ca. $5 \cdot 10^9$ DM), die der Sumitomo Gruppe angehört (Umsatz ca. $4 \cdot 10^{11}$ DM). Vergleichbar wäre im Deutschland ein Zusammenschluß von Allianz, Deutsche Bank, Hoechst, VW und Siemens /40/.

Wie kommen die Zahlen zustande? Schaut man auf die Projektstruktur und überlegt sich, wieviel Arbeit in den einzelnen Unterprojekten geleistet werden muß, kommt man zu Aussagen über die Zahl der benötigten F+E Mannjahre (oder Frauenjahre; kostenmäßig wird hier aber Gleichberechtigung praktiziert) pro Unterprojekt und Jahr. Es ergibt sich beispielsweise eine Kurve wie in Fig. 34. Am Projekt selbst arbeiten sehr viel mehr Menschen mit; wer aber nicht direkt Entwicklungsarbeit leistet wird unter "Kosten" und nicht als "Kopf" erfaßt. An der Kurve werden zwei Probleme sichtbar. Die zu Beginn des Projektes mit zunehmender Zahl benötigten Personen sind nicht irgendwelche Mitarbeiter, sondern zum überwiegenden Teil hochqualifizierte Experten, die per Definition einige Jahre Erfahrung haben müssen. Diese Experten wachsen nicht auf Bäumen, auch nicht (noch nicht?) in Universitäten und Fachhochschulen; sie müssen innerhalb der Firma herangezogen werden und dann zur Verfügung stehen. Auf

der anderen Seite des Mannjahrgebirges erhebt sich die Frage: Wo gehen die jetzt nicht mehr benötigten Mitarbeiter hin?

Ein Mitarbeiter verursacht Kosten. Neben seinem Gehalt, das eine vergleichbar untergeordnete Rolle spielt, braucht er Platz im Büro/Labor/Reinraum, er benützt eine Infrastruktur (von der Bibliothek über die Werkstatt zur Verwaltung), verbraucht Materialien (vom Bleistift bis zum DM 200.000.- Pt-Target für seine Sputteranlage), benötigt u. U. viel Zeit am Großrechner, verursacht Kapitalkosten über die Abschreibung des von ihm benötigten Equipments und "verbraucht" insbesondere viele Siliziumscheiben. Der letzte Posten hat es in sich: Wie bereits ausgeführt, könnte eine Linie, die Scheiben zur Entwicklung eines 16M-DRAMS prozessieren kann, ohne weiteres Scheiben mit 4M-DRAMS produzieren und die Chips verkaufen (in der Regel tut sie das auch). Eine 16M-Entwicklungsscheibe kostet deshalb konsequenterweise soviel, wie die dadurch blockierte Kapazität an 4M-Produktscheiben. Bei 140 4M Chips auf einer Scheibe, einer Ausbeute von 50% und einen Chippreis von 50,- DM (willkürliche, aber nicht ganz falsche Zahlen); ist eine 4M-Scheibe DM 3500,- wert, die 16M-Scheibe rund und roh das doppelte (eine 16M-Scheibe blockiert ca. zwei 4M Scheiben: Sie hat mehr Prozeßschritte, bindet mehr Spezialisten, läuft langsamer, da noch nicht routinemäßig). Die Gesamtzahl der für die F+E-Phase benötigten Scheiben liegt in der Größenordnung 30-60 Scheiben pro Mannjahr. Legt man alle Kosten auf das Mannjahr um, kostet ein "Kopf" pro Jahr im Mittel gut und gern DM 1 000 000,-; die Gesamtkosten der F+E-Phase sind demnach im Bereich einer halben Milliarde. Dabei ist noch nicht berücksichtigt, daß während der Entwicklungsphase noch die später benötigte Fabrik errichtet werden muß (Neubau einer 16M-DRAM Fabrik kostet ca. 10⁹DM).

Hat man Infrastruktur und Mitarbeiter in der benötigten Qualität und Quantität nicht zur Verfügung sondern muß die Ressourcen erst aufbauen, so kann man für die Kosten gut und gern das zwei- bis dreifache ansetzen. Dabei ist immer noch vorausgesetzt, daß keine größeren Pannen vorkommen und die Organisation wie geschmiert läuft. Einige Fallgruben, die Zeit und damit Geld kosten, sind beispielsweise:

- Verwendung von Technologien oder Equipment die, wie sich später herausstellt, nicht geeignet sind. Beispiele:
"Falsche" Silizidsorte, nicht geeignetes Lithographieequipment, Metallisierung erfüllt nicht die Zuverlässigkeitskriterien.
- Zu starkes Festhalten an alten Einzelprozessen.
Bis zur Einsicht, daß ein früher bewährter Prozeß nicht mehr gut genug ist, kann viel Zeit vergehen.
- Zu viele innovative Prozesse - die Optimierung all der neuen Prozesse und ihrer Wechselwirkung kann recht lange dauern.
- Zu viele Reibungsstellen beim Transfer des Know-hows.

Zum letzteren Punkt ist anzumerken, daß der- oder diejenigen, die zum guten Schluß einen bestimmten Prozeß in der Fabrik durchführen, von den Entwicklern, die in der frühen F+E-Phase zum erstenmal diesen Prozeß angingen, durch eine ganze Staffel von Personen getrennt sind. Im ungünstigen, aber nicht

unrealistischen Fall fließt das Know-how vom verantwortlichen Entwickler unter Einbindung seines Assistenzpersonals zu den Prozeß- und Fertigungsingenieuren der F+E- oder Pilotlinie wo es sich zu "festgeschriebenen" Prozessen in der Form von detaillierten Arbeitsanweisungen für das Linienpersonal verdichtet. Von der Pilotlinie erfolgt schließlich der letzte Know-how Transfer über die Ingenieure zu den Mitarbeitern der Volumenfertigung. Am besten funktioniert dieser Know-how Transfer, wenn mit den Know-how auch Menschen tranferiert werden und die Pilotlinie mit der Fertigung identisch ist, am schlechtesten, wenn nur große Papiermengen bewegt werden.

Für den Projektmitarbeiter kann dies im Extremfall heißen, daß er als "Forscher" anfängt, eine neuartige Kathode in einem Plasmaätzreaktor zu untersuchen, nach einem Jahr dann für die nächsten zwei Jahre einen neuen Ätzprozeß für Aluminium entwickelt, diesen Prozeß dann in der Pilotfertigung ein Jahr lang optimiert und anschließend in der Fertigung als "Fertigungsingenieur" eine Schicht der gesamten Metallätzung betreut.

Anhänger der "lebenslänglichen" reinen Forschung sollten mal darüber nachdenken, daß in einem Forscherarbeitsleben (ca. 35 Jahre) 12 neue Generationen (bei linearer Fortschreibung) der Mikroelektronik entwickelt werden. Wieviel Prozent der Jungforscher, die sich bei der F+E-Phase des 1M-, 4M-, 16M-DRAM noch mit Ruhm bedeckten, werden in der Lage sein, auch noch beim 16×10^{12} bit-DRAM (der 12. Generation nach dem 1M) an der vordersten F+E-Front mitzumarschieren? Selbstverständlich weiß niemand, ob es noch ein 1G-DRAM geben wird oder was danach kommt - die obige Frage ist aber auch losgelöst von Spekulationen über Speichergenerationen durchaus bedenkenswert.

16.7.3 Die Projektphasen

Der Ablauf des Projektes läßt sich am besten anhand der wichtigsten Meilensteine veranschaulichen (Fig. 35). In der ersten Phase laufen in erster Linie 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 Simulation diverser schaltungstechnischer Ansätze kumulieren schließlich im "First Silicon" eines Testchips; also in ersten Si-Scheiben mit für den neuen Speicher relevanten Strukturen.

Mit den gewonnenen Erkenntnissen wird ein Technologie- und Designkonzept festgelegt, das durchaus noch Alternativen bei Prozeßkomplexen (z.B. Zellkonzepten) und Einzelprozessen aufweisen kann. Zusammen mit einem ersten Produktdesign erhält man "First Silicon" des ersten neuen Speicherchips. Diese Speicher werden mit an Sicherheit grenzender Wahrscheinlichkeit nicht voll funktionsfähig sein. Aus der Analyse der Design-, Technologie; aber auch Montage- oder Prüftechnikprobleme resultieren ein Neudesign und Problemlösungen der Technologie. Noch offene Alternativkonzepte werden entschieden; aus den entsprechenden Durchläufen resultieren jetzt "Labormuster", d.h. einige unter entspannten Bedingungen funktionierende Speicher (Entspannt heißt, daß der Speicher nicht unter allen

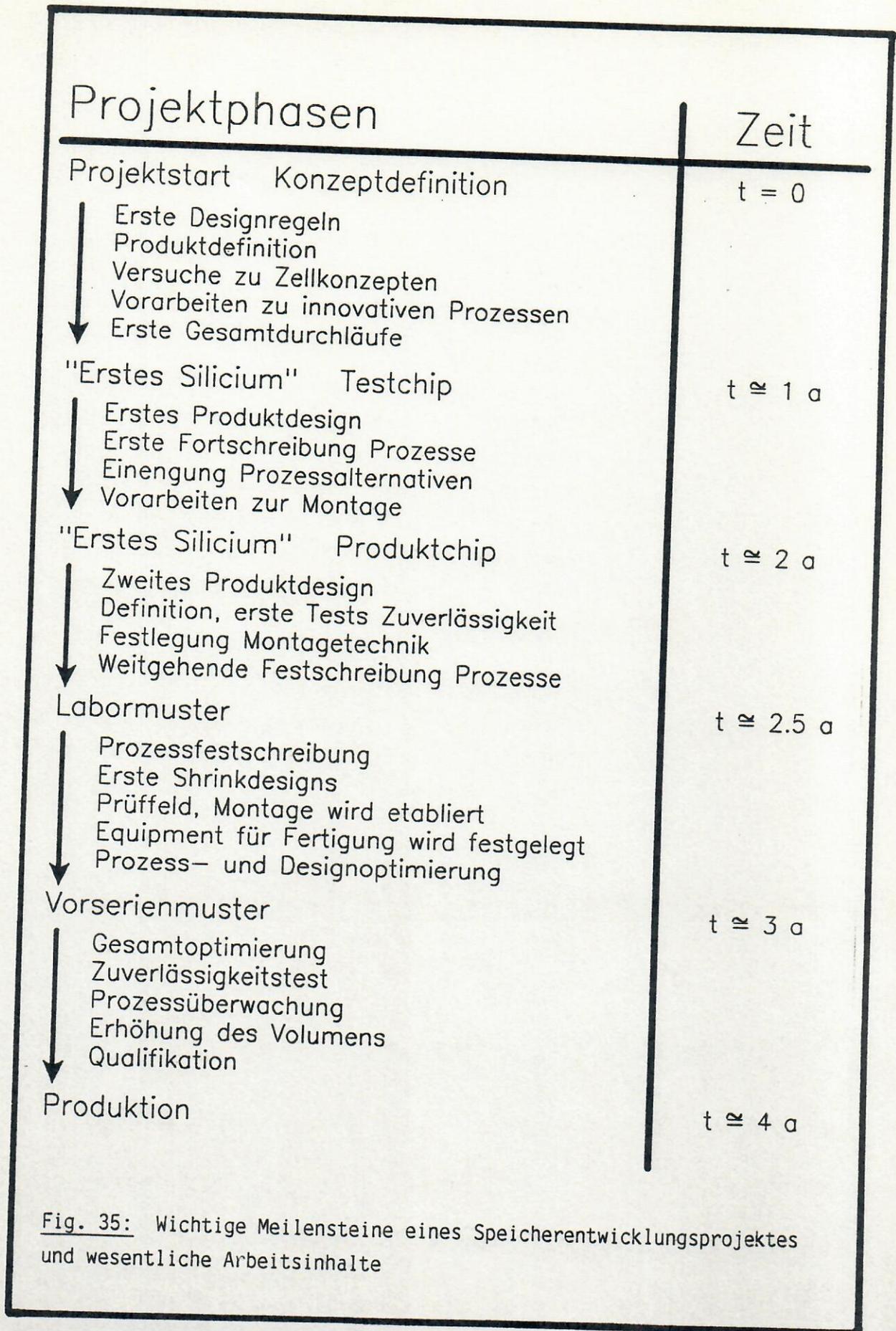


Fig. 35: Wichtige Meilensteine eines Speicherentwicklungsprojektes und wesentliche Arbeitsinhalte

spezifizierten Bedingungen (Temperatur, Spannung, ...) funktionieren muß).

Erneut werden Prozesse und Design optimiert, Prüftechnik und Montage müssen stimmen. Prozesse werden festgeschrieben; Kontrollen und Qualitätssicherungsmaßnahmen verschärft. Ausbeuterelevante Messungen werden durchgeführt, Maßnahmen zur Ausbeutesicherung eingeführt. Die Zahl der eingeschleusten Scheiben wächst - schließlich erhält man Vorserienmuster, die dem endgültigen Produkt schon sehr nahe kommen und auch zum Testen an gute Kunden weitergegeben werden.

Zunehmend beginnt nun eine Materialschlacht. Ziel ist, den Speicherbaustein auch im "worst-case" der Spezifikationen funktionsfähig zu erhalten (z.B. $T = 85^{\circ}\text{C}$, Luftfeuchtigkeit = 100%, $U = 4,5 \text{ V}$) die Zuverlässigkeit aller technischen Daten und damit der Betriebsfähigkeit für z. B. 10a zu garantieren, Prozesse und Prozeßkomplexe auf Ausbeute und Kosten optimieren (Mehr Ausbeute = mehr Kontrollen, mehr Kontrollen = mehr Kosten), Prüftechnik incl. Redundanz zu optimieren und die Ausbeute bei mindestens 10% zu stabilisieren. Alles zu dem Zweck, die sog. Qualifikation zu bestehen, d.h. die Bestätigung der Qualitätsabteilung zu erhalten, daß der neue Chip alle Torturen im vorgeschriebenen Maße überlebt. Mit Bestehen der Qualifikation ist das Projekt beendet, die Fertigung kann beginnen.

16.8 Die Fertigung

Das Projekt wurde erfolgreich beendet, die Produktion wird aufgenommen! Was versteckt sich hinter der einfachen Aussage "Die Produktion aufnehmen"? In Kapitel 16.1.1 haben wir genau diese Frage auch schon gestellt, die dortigen Antworten haben die Fertigungslinie aber noch als "black box" betrachtet. In diesem Kapitel soll schlaglichtartig beleuchtet werden, was sich in diesem schwarzen Kasten abspielt.

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 Scheibendurchmesser; 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 hochzufahren.

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

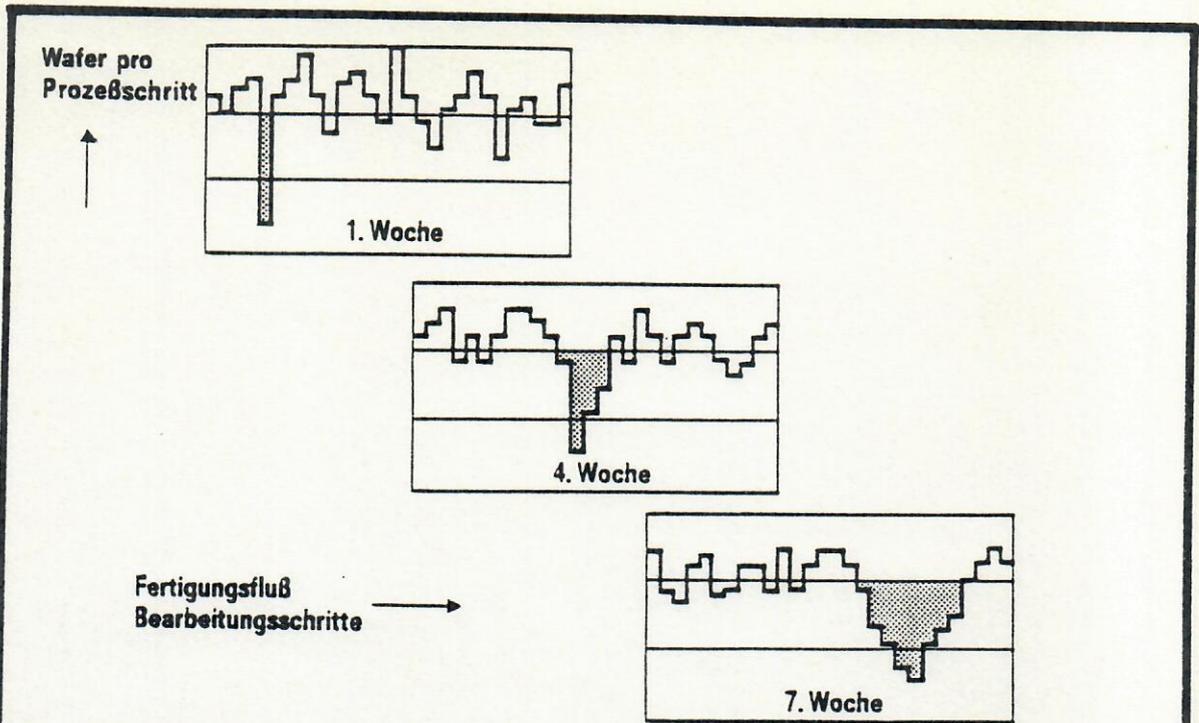


Fig. 36: Verteilung von Scheiben auf Prozeßschritte mit einem "Einbruch" in der ersten Woche. Bei einer voll ausgelasteten Fertigung wird dieser Einbruch nicht wieder aufgeholt /37/.

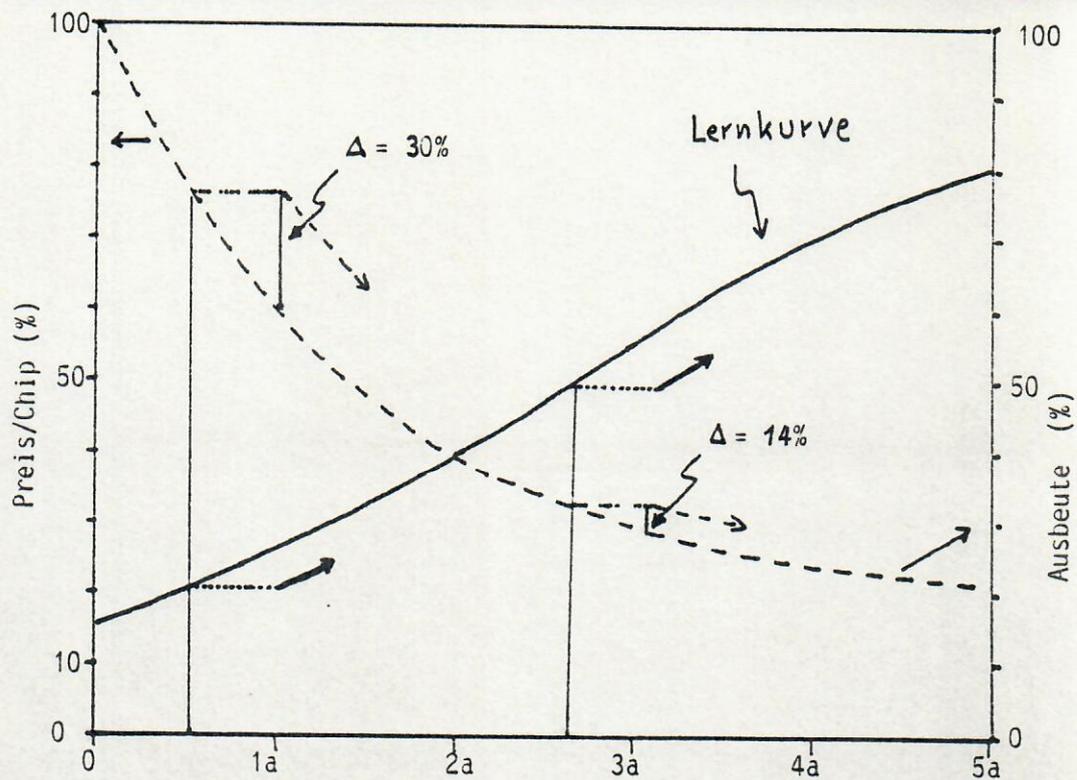


Fig. 37: Preisverfall bei Speichern und daraus resultierende Lernkurve für Kostenneutralität. Ein Stagnieren der Lernkurve für 6 Mon. führt zu Kostenerhöhungen von 30% bzw 14% in den eingezeichneten Beispielen.

Das Hochfahren der Kapazität ist nicht so einfach, wie es aussieht. Es genügt nicht sicherzustellen, daß ausreichend Gerätekapazität für jeden Prozeßschritt bereitsteht, sondern es muß auch der Fertigungsprozeß optimiert werden. Jeder einzelne Prozeß kann nur dann seine Tageskapazität erreichen, wenn der Scheibenanfall halbwegs kontinuierlich ist. Der Scheibenfluß wird optimiert durch geeignete Auswahl des Equipments (eine große oder drei kleine Anlagen - was ist besser?), Schaffung von Redundanz, Festlegung von Prioritäten bei Einschleusung und Weiterbearbeitung, Optimierung der Wartung des Equipments, etc.. Eine gute Logistik funktioniert nicht nur im störungsfreien Betrieb, sondern erlaubt auch ein möglichst rasches Abklingen von "Scheiben-Stoßwellen", die unvermeidlich entstehen wenn ein für einige Zeit ausgefallenes Equipment den entstandenen Scheibenstau schnell abbaut (Fig. 36).

Fällt das gesamte Equipment gleichzeitig aus - weil nämlich freitags dichtgemacht wird - müssen die Scheiben auf Parkpositionen, d.h. bei Prozeßschritten angehalten werden, die durch längere Wartezeiten nicht in ihrer Qualität beeinträchtigt werden. Ein 4M-DRAM Durchlauf, z.B., hat mehr als 50 Zeitkopplungen zwischen den Prozessen, d.h. Wartezeiten zwischen Prozeßschritt n und Prozeßschritt $n+1$ dürfen nicht länger sein als typischerweise 10 min. - 1 hr. In der Praxis sieht das so aus, daß freitags viele Scheiben schon lange vor Arbeitszeitende gestoppt werden müssen, da die verbleibende Arbeitszeit nicht mehr ausreicht, um die nächste mögliche Parkposition zu erreichen. Am Montag morgen stehen dann Scheibenberge vor bestimmten Anlagen, z.B. den Reinigungsmaschinen, an anderen Anlagen steht dafür nichts.

In Summe fehlt bei der Durchlaufzeit nicht nur die direkt weggefallene Arbeitszeit, sondern ein weiterer halber bis ganzer Tag pro Woche, da der Fertigungsfluß nicht mehr richtig ins Gleichgewicht kommt. Höhere Durchlaufzeiten aber bedingen immer kleinere Ausbeuten durch ein Zusammenwirken mehrerer Faktoren. Allein die statistische Wahrscheinlichkeit, daß ein Partikel auf eine Scheibe fällt, ist umso größer, je länger die Scheibe in der Linie ist. In Vakuumanlagen ist montags das Vakuum anders als Freitag abends - alles Faktoren, deren Auswirkungen auf die Ausbeute nur sehr schwer direkt zu erfassen sind, deren kumulierter Effekt sich jedoch mit Sicherheit in einem Ausbeuterückgang äußert.

Betrachten wir als nächstes die Lernkurve. Nehmen wir an, daß die Ausbeute bei Fertigungsbeginn bei 15% liegt und daß der am Markt erzielbare Preis zur Zeit $t=0$ die Herstellkosten deckt, möglicherweise sogar noch einen Gewinn abwirft. Betrachtet man nun die typische Entwicklung des Speicherpreises als Funktion der Zeit, so läßt sich daraus eine Lernkurve ableiten (Fig. 37), d.h. eine Forderung an die Ausbeutesteigerung als Funktion der Zeit, damit die erzielbaren Preise kostendeckend sind. Dabei wird zunächst einmal angenommen, daß die Herstellkosten für eine Scheibe konstant bleiben. Jedes Zurückbleiben hinter der Lernkurve führt automatisch zu Verlusten - es sei denn, man hat Glück und die Preise bleiben länger hoch als erwartet.

Das Hochlaufen auf der Lernkurve erfolgt keineswegs automatisch; jedes Prozent zusätzlicher Ausbeute muß erkämpft werden. Eine Ausbeute von 80% bedeutet, daß bei jedem der 400 Prozeßschritte die Prozeßausbeute 99,94% erreichen muß und bei den ca. 1000 Scheibenbewegungen (aus der Box in die Anlage; aus der Anlage in die Box, aus der Box zum Meßgerät, ...) nicht mehr als 10-20 "Killerpartikel" auf die Scheibe fallen. Nur dauernde Analysen von Ausfallursachen und zähes Optimieren der Prozesse wird zum Erfolg führen. Beispielsweise könnte sich ergeben, daß eine Senkung der Schichtdickenschwankung eines bestimmten Oxids von +/- 6% (über eine Scheibe, von Scheibe zu Scheibe, von Fahrt zu Fahrt, von Tag zu Tag, ...) auf +/- 5% die Ausbeute um 3% erhöht. Die Aufgabe, diese schärferen Toleranzen sicherzustellen (durch Optimierung der vorhandenen Prozesse und des vorhandenen Equipments - nicht durch neuere Geräte) ist eine Aufgabe für einen Fertigungsingenieur die erheblich anspruchsvoller sein kann, als es die eigentliche Entwicklung dieses Prozesses war.

Die Reduzierung der Partikeldichten ist einer der Schlüssel zum Ausbeutehochlauf. Auch hier ist nur durch harte Arbeit eine Verbesserung zu erreichen. Fig. 38 illustriert dies überdeutlich: Der Fortschritt wird nur im Mittel vieler Einzelergebnisse erreichbar, das Greifen von eingeführten Maßnahmen ist oft nicht unmittelbar nachweisbar.

Die Anstrengungen, die gemacht werden um auf der Lernkurve hochzulaufen, werden begleitet von Anstrengungen die Kosten pro Scheibe zu senken. Diese sogenannten Rationalisierungsmaßnahmen können sein:

- Vereinfachung der Prozesse und der Prozeßführung
- Verwendung billiger Medien (sobald man weiß, worauf es ankommt)
- Beschaffung von leistungsfähigem Equipment
- Erhöhung der Equipment "Uptime" durch geeignete Maßnahmen
- Besserer Fertigungsfluß (und damit höhere Kapazität) durch verfeinerte (computergesteuerte) Logistik
- Verkleinerung der Chipfläche durch "Shrink" Design; vgl. Fig. 3.

Die letzte Maßnahme ist zwar besonders wirksam, da man auf einer Scheibe ca. 20% mehr Chips bekommt, dafür aber nicht ganz einfach. Sie bedingt zumindest völlig neues Lithographieequipment, da das vorhandene immer an der Grenze seiner Leistungsfähigkeit betrieben wird und einfach hier keine feinere Strukturen herstellen kann. Darüberhinaus ein neues Design, eine teilweise geänderte Prozeßführung und in der Regel einige neue Einzelprozesse. Damit muß einerseits der "Shrink"-Chip neu qualifiziert werden (kostet viel Zeit und Geld), andererseits 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 gut geht, liegt man mit der Ausbeute auf der Lernkurve und die Rationalisierungsmaßnahmen greifen. Falls man noch einen Kunden findet, der die Chips auch kauft (bei den

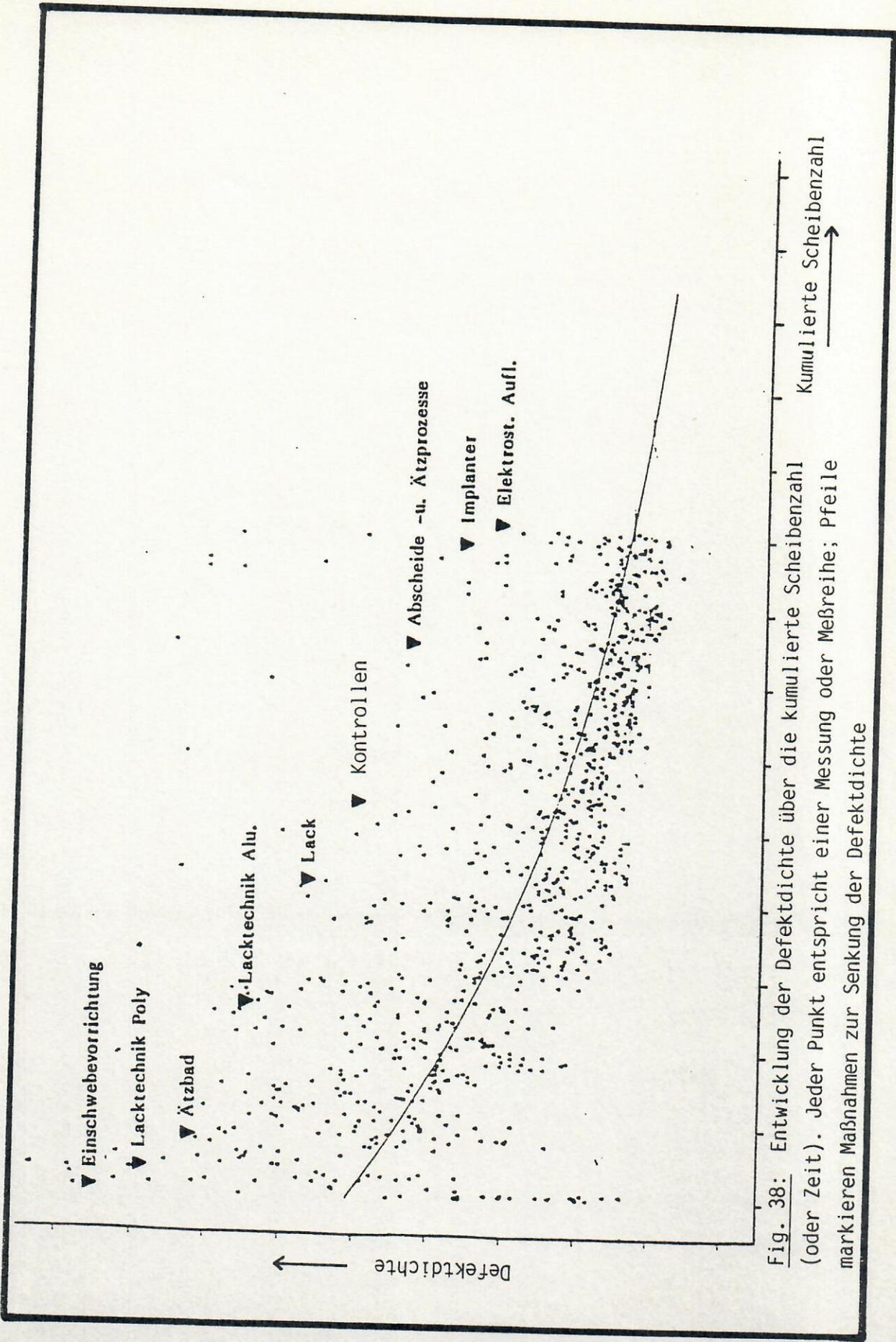


Fig. 38: Entwicklung der Defektdichte über die kumulierte Scheibenzahl (oder Zeit). Jeder Punkt entspricht einer Messung oder Meßreihe; Pfeile markieren Maßnahmen zur Senkung der Defektdichte

wilden Fluktuationen des Weltchipmarktes alles andere als selbstverständlich), verdient die Fabrik Geld. Was passiert wenn etwas schiefgeht?

Nehmen wir einmal an, daß die Lernkurve plötzlich für 6 Monate stagniert, z.B. bei einer Ausbeute von 20% oder 50%. Gründe dafür gibt es viele; zum Beispiel:

- die Leistungsfähigkeit eines Prozesses/Equipment ist (vorher nicht erkennbar) erreicht. Erkennen des Problems, Ausbau des "veralteten" und Einbau des besseren (welches?) Equipments erfordert 6 Monate (und einige zig Millionen DM).
- Ein bisher überhaupt nicht spezifizierter Parameter für ein Verbrauchsmaterial beeinträchtigt plötzlich die Ausbeute weil sich der Herstellprozeß oder der Lieferant geändert hat (Beispiel von früher: Kohlenstoffgehalt im Silizium geht hoch (von < 1 ppm auf > 1 ppm)). Ursachenfindung und Behebung dauern leicht 6 Monate.
- Streiks, neue Zollgesetze, Verzögerungen bei der Einführung der Kontischicht, etc. führen zu längeren Lernzyklen.

Was dann passiert, zeigt ebenfalls Fig. 37. Die Kosten steigen gegenüber dem Weltmarktpreis um 30 %, bzw. 14%. Ist die Gewinnspanne kleiner (und das ist sie im Speichergeschäft mit Sicherheit), wird jetzt mit Verlust produziert. Vereinfacht ausgedrückt: Ist der Weltmarktpreis bei 20 DM und produziert die den Weltmarktpreis bestimmende Konkurrenz für 18 DM; man selbst aber um 20% höher zu 22 DM, macht der eine bei 10⁹ DM Umsatz 100 Millionen Gewinn, man selbst 100 Millionen Verlust. Da die in Fig. 37 gezeigte Preiskurve nicht fiktiv ist, sondern ziemlich exakt der Realität entspricht, bedeutet dies in der Tat, daß der Unterschied zwischen massiven Gewinnen oder massiven Verlusten in einigen Monaten Abstand zu den ostasiatischen Weltmarktführern liegt. Geht man davon aus, daß der Weltmarktpreis für DRAMs ein Kampfpriß ist (wenn nicht gar direktes "dumping"), können sechs Monate Verzug zur Geschäftsaufgabe führen.

Vergleicht man diese Situation mit z.B. der Automobilindustrie, fällt auf, daß ein zu langsamer Modellwechsel (neue Generation) im Vergleich zur Konkurrenz zwar Gewinne und Marktanteile schrumpfen läßt (man beobachte z. B. die deutsche Oberklasse); aber nicht unbedingt massive Verluste bedingt. Die kritischen Zeiträume für die Einführung neuer Produkte wird darüberhinaus nicht in Monaten, sondern in Jahren gemessen.

Auch wenn nichts schiefgeht, ist die Fabrik kein Ort der Ruhe und Besinnung. Nehmen wir an, das erste Produkt der Fabrik ist das 1M-DRAM; die Fertigung läuft auf 150 mm-Scheiben (200 Chips/Scheibe), man liegt auf der Lernkurve und verdient Geld. Nach ca. 1,5 Jahren kommt die erste Shrink-Version (240 Chips/Scheibe) als Rationalisierungsmaßnahme, aber auch um ein besseres (schnelleres) Zukunftsprodukt anbieten zu können. Damit verbunden ist zwingend eine Beschaffung von besserem Equipment in Teilbereichen (insbesondere in der Lithographie) mit all den damit verbundenen Problemen. Bei der Lernkurve fängt man wieder unten an (wenn auch auf einer besseren Basis), das ganze Spielchen der Ausbeutesteigerung beginnt von vorne.

Nach drei Jahren kommt das 4M-DRAM. Möglicherweise kommt es auch nicht, weil zwischenzeitlich eine neue Fabrik gebaut wurde. Die neue Fabrik auf der grünen Wiese wird aber wahrscheinlich Teile des erfahrenen Personals abziehen, in der 1M-Fabrik läuft dann eine Einstellungs- und Ausbildungswelle (es gibt keine woanders ausgebildeten Mikroelektroniker am Personalmarkt). Kommt das 4M-DRAM auch in die "alte" Fabrik, bedingt dies massive Equipment- und Prozeßänderungen; es startet eine neue Lernkurve (die dritte in 3 Jahren). Nach weiteren 1,5 Jahren kommt der 4M-Shrink, nach 6 Jahren kommt der 16M-DRAM. Da jetzt aber nur noch ca. 80 Chips auf der Scheibe sind, geht das 16 DRAM aber eher in eine neue Fabrik mit 200 mm Scheiben (140 Chips per Scheibe). Möglicherweise kommt jetzt die auf der 1M- oder 4M-DRAM Technologie basierende Logik. Statt einigen wenigen Speicherprodukten macht die Fabrik jetzt viele, viele Logikprodukte (denn es gibt hier viele verschiedene Chips mit der gleichen Technologie) mit ganz anderen Anforderungen an Logistik, Prüftechnik und Montage. Nach 10-12 Jahren ist die Fabrik hoffnungslos veraltet; Teile landen im deutschen Museum (kein Witz). Zum Vergleich: In der Auto/Chemie/Stahlindustrie fängt man allmählich an, nicht mehr von der "neuen" Fabrik zu reden, sondern von "der" Fabrik.

Eine moderne Halbleiterfabrik ist also alles andere als ein langweiliger Platz. Die Herausforderungen an das wissenschaftlich-technische Können der Mitarbeiter steht hinter den Anforderungen von Forschungs- und Entwicklungslabors nicht mehr zurück.

Ziel dieses vom Umfang her überdimensionierten Beitrags war es, angehenden Naturwissenschaftlern und Ingenieuren die Probleme, aber auch die Faszination der Entwicklung und Fertigung moderner Mikroelektronik ungeschminkt und im Detail darzustellen.

Wer in den nächsten 30 Jahren sein jetzt gelerntes Wissen beruflich umsetzen will - sei es als "reiner" Forscher in einer Universität, als Produktioningenieur beim Kunden oder irgendwo dazwischen - wird mit hoher Wahrscheinlichkeit hautnah mit Mikroelektronik konfrontiert werden. Wünschen wir uns allen, daß die "technische" Intelligenz in unserem Land der damit verbundenen Herausforderung gewachsen sein wird.

Danksagung:

Besonderer Dank gebührt Fr. R. Pepper, die - wie immer in letzter Minute - aus einem großen Stapel Papier ein lesbares Manuskript gemacht hat. Unseren Kollegen L. Armbrecht, H. Fritzsche, J. Harter, H. Schuster-Woldan und S. Schwarzl danken wir für die Hilfestellung beim Verfassen und Redigieren des Manuskripts.

Fußnoten und Literaturhinweise

- Vorbemerkung:** Eine Erfassung der wichtigsten Arbeiten zu dem behandelten Themenkreis ist praktisch unmöglich und wurde nicht angestrebt. Die Literaturhinweise beschränken sich deshalb auf weiterführende, in der Regel leicht erhältliche Zeitschriften und Büchern sowie auf die Angabe der Quellen für Tabellen und Figuren.
- /1/ Abkürzungen und Mikroelektronik-Slang werden weitgehend im Text erläutert; ansonsten helfen die SIEMENS Broschüren: "Zum Thema Mikroelektronik" und "ABC der Mikroelektronik" (Bezug über Siemens, Infoservice Fürth), Postfach 2348, D-8510 Fürth
 - /2/ Anlässlich des Produktionsbeginns bei IBM/Sindelfingen im Juli 1989
 - /3/ Leider sind die oft scheußlichen, häufig falschen und immer amerikanischen Wortgebilde des Mikroelektronik-Slangs oft nicht durch griffige deutsche Bezeichnungen ersetzbar. Eine "Waferfab" ist keine Scheibenfabrik (auch keine Waffelfabrik), da sie eben nicht Wafer herstellt sondern Chips. "Chipfabrik" wäre zwar richtiger, aber auch nicht deutsch, und "Schnitzelfabrik" trifft es auch nicht ganz. Wir verwenden daher im Zweifelsfall den amerikanischen Slang.
 - /4/ H. Queisser; Kristallene Krisen; Diper (München, Zürich) 1985
 - /5/ W. Beinvogl und E. Hopf; The Development of a 4Mbit DRAM; Festkörperprobleme 28 (1988) 87
 - /6/ K. Horninger; Zellenkonzepte und Schaltungstechnik von dynamischen Megabit-Speichern; IGT-Fachberichte 102 (VDE-Verlag Berlin-Offenbach) 1988, S. 217
 - /7/ F. S. Becker; From 4M to 16M-DRAM Technology - Present Problems and Coming Challenges; Proc. IV SB MICRO Meeting, Porto Alegre, Brasil (1989) S. 125
 - /8/ D. Widmann, H. Mader, H. Friedrich; Technologie hochintegrierter Schaltungen; Halbleiter-Elektronik Bd. 19 (Springer Verlag) und einige andere Bände dieser Reihe
 - /9/ S. Wolf, R. N. Tauber; Silicon Processing for the VLSI Era; Lattice Press, 1988
 - /10/ F. S. Becker; Interlevel Dielectrics for Reduced Thermal Processing; in: Red. Therm. Proc. for VLSI; ed. R. A. Levi; (Plenum Publ. Corp., New York) 1989
 - /11/ W. Arden, K. H. Müller, Light vs. X-Rays: How Fine can we Get?; Semicond. Int., Sept. 1989, S. 128
 - /12/ A. Heuberger; X-ray Lithography, J. Vac. Sci. Technol. B6, p 107 (1988)

- 16.77
- /13/ S. Schwarzl; Proc. Symp. on Dry Proc.; Electrochem. Soc.; Proc. Vol. 88-7 (1988) S. 48
- /14/ Courtesy of S. Schwarzl; Siemens
- /16/ D. Roß: Cleaning and Drying Technology; Semicon Zürich, Techn. Proc., März 1989
- /17/ W. Reczek, W. Pribyl, B. Murphy, F. Bonner, J. Haster; Latch-up-Charakterisierung dynamischer Speicher; wie /6/, S. 292
- /18/ H. Reisinger, W. Hönlein; ONO-Technology; Appl. Surf. Sci. 39 (1989) S. 178
- /19/ Benzinger, Enders, Hasler, Higelin, Küsters, Mühlhoff, Müller, Röhl; 4Mbit DRAM Prozeß mit Trench-Zelle und FOBIC; wie /6/, S. 263
- /20/ A. Kohlhase, M. Mändl, W. Pamler; Performance and Failure Mechanisms of TiN in Submicron Devices; J. Appl. Phys. 65 (1989) S. 2464
- /21/ P. Thoren, J. W. Rangelow, R. Kassing, P. KÜcher; Micro-circuit Eng. 1988; Proc. 14th Int. Conf. Microlithography, Vienna, Sept. 1988, (North-Holland, Amsterdam) in print
- /22/ J. Berthold, C. Wieczorek; CVD-Tungsten Deposition and Dry-Etch Processes for Planarized Metallisations; Appl. Surf. Sci. 38 (198) S. 506
- /23/ L. Risch, W. Müller, P. Tielert; Probleme der Bauelemententwicklung im 4-Megabit Speicher; ntz Bd. 40 (1987) Heft 3, S. 192
- /24/ W. Pribyl, J. Harter, W. Müller; 4-Megabit DRAM in Submicron CMOS Technology with a FOBIC Trench Cell; Siemens Forsch.- und Entwicklungs-Ber. Bd 16 (1987) S. 253
- /25/ VLSI Electronics - Microstructure Science, Hrsg. N.G. Einspruch (Accademic Press, Inc.); insbesondere Band 9 und Band 12
- /26/ R. V. Criegern, H. Zeininger, S. Röhl; Proc. 6th Int. Conf. on Sec. Ion Mass Spectr.; "SIMS VI"; Versailles 1987 (eds. J. Wiley & Sons; 1988), S. 419
- /27/ H. Föll, V. Lehmann, G. Zoth, F. Gelsdorf, B. Göttinger, In-line Monitoring of Heavy Metal Contamination and Interface States by an Imaging Technique; J. Electrochem. Soc.; to be published Summer 1990
- /28/ V. Penka, W. Hub; Applications of TXRF in Semicond. Surface Analysis; Spectrochimica Acta 44b (1989) S. 483
- /29/ H. Oppolzer, H. Cerva, C. Fruth, V. Huber, S. Schild;

TEM Studies During Development of a 4M-DRAM; Inst. Phys. Conf. Ser. 87 (1987) S. 433

- /30/ S. Wurm, P. Alpern, D. Savignac, R. Kakoschke; Modulated Optical Reflectance Measurement on Amorphous Silicon Layers and Detection of Residual Defects; Appl. Phys. A, 47 (1988) S. 147
- /31/ B. Murphy, J. Stecker, H. Hemmert, J. Harter, W. Pribyl; Modellierung und Gesamtsimulation eines 4-Megabit DRAMs; wie /16/, S. 269
- /32/ H.-J. Hacke; Montage integrierter Schaltungen; Springer Verlag
- /33/ J. Ertl; Die Halbleiterproduktion auf dem Weg zum 64-Megabit-Chip; in: Halbleiterfertigung; Proc. Fachtagung Productronica München 1989; VDE-Verlag GmbH Berlin Offenbach
- /34/ Y. Funaki; Beyond Just in Time: Total Preventive Maintenance; Semicond. Int., Sept. 1989, S. 162
- /35/ K. Kearney; Ultrapure Wafer Requirements Squeeze into the Submicron Range; Semicond. Int., Jan. 1989, S. 80
- /36/ G. E. P. Box, W. G. Hunter, J. S. Hunter; Statistics for Experimenters; ed. Wiley New York, 1978
- /37/ G. Mönkemeyer; Gedanken zu einer wirtschaftlichen VLSI-Fertigung; wie /33/, S.29
- /38/ H. Friedrich; The MEGA Project; Siemens Review L II (1985) No. 2 S. 23
- /39/ Studie "Mikroelektronik 2000" in: micro-electronik, Heft 1, Sept./Okt. 1987, S. 30
- /40/ J. Knorr; Integrierte Schaltungen für ein integriertes Europa; mikro-elektronik, Heft 6, Nov/Dez. 1988, S.244
- /41/ ICE Status Report 1989 (Int. Circuit Eng. Corp.)
- /42/ H. E. Maes, G. Groeseneken, H. Lebon and J. Witters: Trends in semiconductor memories; Microel. Journal 20 (1-2), p. 9 (1989)
- /43/ F. S. Becker, D. Pawlik, H. Anzinger and A. Spitzer: Low-pressure deposition of high-quality SiO₂ by pyrolysis of tetraethylorthosilicate; J.-Vac. Sci Technol., B 5, p 1555 (1987)
- /44/ F. S. Becker, H. Treichel and S. Röhl: Low pressure deposition of doped SiO₂ by Pyrolysis of Tetraethylorthosilicate (TEOS) II. Arsenic Doped Films; J. Electrochem. Soc. 136, p. 3033 (1989)
- /45/ W. Kern: The evolution of silicon wafer cleaning technology; to be published in "Proc. 1 st. Int. Symp. on

"Cleaning", Electrochemical Society, Oct. 1989, Hollywood,
(Florida)

/46/ H. v. Otterloo: Experiences during the start-up of a wafer
fabrication unit; in Ref. 33, S. 9