

TECHNISCHE UNIVERSITÄT MÜNCHEN
LEHRSTUHL FÜR TECHNISCHE ELEKTRONIK
FACHGEBIET HALBLEITERPRODUKTIONSTECHNIK

Technisch-ökonomische Effizienzbetrachtungen für die Halbleiterfertigung

Walter Puffer

Vollständiger Abdruck der von der Fakultät für
Elektrotechnik und Informationstechnik der Technischen Universität München
zur Erlangung des akademischen Grades eines
Doktor-Ingenieurs
genehmigten Dissertation.

Vorsitzender: Univ.-Prof. Dr.-Ing. Hans-Georg Herzog
Prüfer der Dissertation: 1. Univ.-Prof. Dr.-Ing. Walter Hansch
2. Univ.-Prof. Dr. rer. nat. Ignaz Eisele,
Universität der Bundeswehr München

Die Dissertation wurde am 19.03.2007 bei der
Technischen Universität München eingereicht und durch die Fakultät für
Elektrotechnik und Informationstechnik am 25.10.2007 angenommen.

Inhaltsverzeichnis

1	Einführung.....	9
1.1	Zum Gegenstand der Arbeit.....	9
1.2	Bedeutung der Mikroelektronik und Dynamik des Halbleitermarktes.....	10
1.3	Darstellung der wichtigsten Produktgruppen im Halbleitermarkt.....	13
1.4	Ökonomische Rahmenbedingungen der Halbleiterindustrie.....	17
1.5	Entwicklung des Silizium-Grundmaterials.....	21
1.6	Entwicklung der wichtigsten Prozesstechnologien.....	23
1.7	Internationale Organisationen und Kooperationen verschiedener Hersteller.....	24
2	Aufbau und Herstellung des MOS-Transistors.....	25
2.1	Aufbau des MOS-Transistors.....	25
2.2	Technologieskalierung.....	26
2.3	Bauelementeeigenschaften.....	29
2.4	Typischer Prozessfluss zur Herstellung von integrierten Schaltungen.....	30
3	Fertigung von integrierten Schaltungen.....	35
3.1	Aufbau einer Halbleiterfabrik.....	35
3.1.1	Reinraum.....	35
3.1.2	Anordnung der Geräte in der Fabrik.....	36
3.2	Produktionsfluss in einer Halbleiterfabrik.....	39
3.2.1	Transportbehälter.....	39
3.2.2	Transportsysteme.....	40
3.3	Zukünftige Herausforderungen für die Fertigung.....	42
4	Wirtschaftliche Anforderungen an die zukünftige Höchstintegration von Schaltungen...	43
4.1	Die Rentabilität als unternehmerisches Oberziel.....	43
4.2	Kenngößen der Fertigung.....	47
4.2.1	Produktivität und Effizienz.....	47
4.2.2	Wirtschaftlichkeit.....	49
4.3	Leistung der IC-Fertigung.....	50
4.4	Kosten für die IC-Fertigung.....	52
4.4.1	Aufteilung der Kosten.....	52
4.4.2	Kostenaufwand für die Veränderung der einzelnen Einflussfaktoren.....	55
4.5	Berechnung der Effizienz der Fertigung.....	57
4.5.1	Erstellung eines Effizienz-Modells.....	57
4.5.2	Empfindlichkeit gegenüber Parameteränderungen.....	58
4.5.3	Verbesserungspotential der einzelnen Faktoren.....	60
4.6	Entwicklung beim Hochfahren der Fertigungskapazität.....	62
4.6.1	Zeitliche Entwicklung der Produktionsmenge und -ausbeute.....	62
4.6.2	Leistungs- und Kostenentwicklung.....	62
4.6.3	Entwicklung der Effizienz.....	64
4.7	Zusätzliche Einflussfaktoren auf die Rentabilität.....	66

4.7.1 Zeit.....	66
4.7.2 Qualität.....	69
4.7.3 Markt.....	71
5 Analyse des Einflusses der Produktionstechnik auf die Effizienz der Fertigung.....	75
5.1 Grundlagen und produktionstechnische Größen.....	75
5.1.1 Geräte für die IC-Fertigung.....	75
5.1.2 Gerätezustände.....	77
5.1.3 Beschränkung der Fertigungskapazität in einer Produktionslinie.....	78
5.1.4 Bestehende Größen zur Bewertung der Leistungsfähigkeit.....	80
5.2 Bewertungsgrößen der Fertigungslinie.....	81
5.2.1 Produktionsmenge.....	82
5.2.2 Durchlaufzeit.....	83
5.2.3 Zahl der halbfertigen Produkte in der Fertigungslinie.....	84
5.2.4 Kosten.....	85
5.2.5 Konkurrierende Ziele bei den Bewertungsgrößen.....	86
5.3 Auswirkungen auf die Effizienz der Fertigung.....	87
5.3.1 Verbesserungspotential der einzelnen Faktoren.....	87
5.3.2 Aufwand für die Veränderung der einzelnen Einflussfaktoren.....	90
6 Warteschlangentheorie zur Berechnung der Bewertungsgrößen.....	93
6.1 Warteschlangenmodelle für deterministische und variable Ankunfts- und Prozesszeit.....	93
6.1.1 Einzelgeräte.....	94
6.1.2 Parallele Geräte.....	99
6.1.3 Warteschlangennetzwerke.....	104
6.2 Betriebskennlinie.....	107
6.3 Einfluss der Variabilität auf die Betriebskennlinie.....	111
6.3.1 Schwankungen im Prozessfluss.....	111
6.3.2 Anwendbarkeit der einzelnen Warteschlangenmodelle auf die Prozesse der Halbleiterfertigung.....	112
6.4 Einfluss verschiedener Eingangsparameter auf die Betriebskennlinie.....	112
6.4.1 Gerätedurchsatz.....	112
6.4.2 Unterbrechungen im Prozessablauf.....	113
6.4.3 Losgröße.....	115
6.4.4 Blocking.....	117
6.4.5 Transportzeit.....	119
6.4.6 Variabilität.....	119
6.4.7 Position im Prozessfluss.....	120
6.5 Erstellung eines Warteschlangenmodells für einen Teil des Prozessflusses.....	121
6.5.1 Gerätedaten einer realen Fertigung.....	121
6.5.2 Auswertung der Daten des realen Prozessflusses.....	124
6.5.3 Warteschlangenmodell für einen Teil des Prozessflusses.....	131
6.5.4 Vergleich der Ergebnisse von Modell und Prozessfluss.....	134

7 Möglichkeiten zur Effizienzsteigerung bei der zukünftigen IC-Fertigung.....	139
7.1 Zusammenführung und Bewertung der ökonomischen und technologischen Aspekte der Fertigung.....	139
7.2 Steigerung der Effizienz der Fertigung.....	142
7.2.1 Anforderungen an die Effizienzsteigerung bedingt durch Preisverfall.....	142
7.2.2 Empfindlichkeit gegenüber Parameteränderungen.....	144
7.3 Berechnung der Auswirkungen auf die Effizienz mittels Warteschlangen- modell für Prozessfluss.....	147
7.3.1 Einfluss der Variabilität.....	148
7.3.2 Einfluss des Gerätedurchsatzes.....	151
7.3.3 Einfluss von Unterbrechungen im Prozessablauf.....	152
7.3.4 Einfluss der Position im Prozessablauf.....	156
7.3.5 Einfluss anderer produktionstechnischer Parameter.....	162
7.3.6 Auswirkungen der Maßnahmen auf die Kosten.....	166
7.3.7 Auswirkungen der Maßnahmen auf die Effizienz.....	168
7.3.8 Verbesserungspotential der einzelnen Faktoren.....	170
7.3.9 Maßnahmen zur Verringerung von Schwankungen im Prozessfluss.....	170
7.3.10 Gültigkeitsbereich der Betrachtungen.....	171
7.4 Verbesserungspotential beim Ramp Up.....	172
7.5 Auswirkungen von Veränderungen vor und während der Produktion.....	175
7.6 Zusammenfassung: Technische Faktoren mit dem größten Verbesserungs- potential.....	178
8 Ausblick.....	181
Literaturverzeichnis.....	183
Verzeichnis der verwendeten Abkürzungen und Symbole.....	193
Danksagung.....	197

Zusammenfassung

Die Mikroelektronik schreitet weiter voran. In den letzten 20 Jahren war auf dem Halbleitermarkt eine durchschnittliche Wachstumsrate von 16% pro Jahr zu verzeichnen. Starke technologische Verbesserungen, die zu Produkten mit immer höherer Leistungsfähigkeit und somit zu einer sehr hohen Innovationsgeschwindigkeit führen, sind charakteristisch für diesen Bereich.

Als problematisch stellen sich jedoch die Rahmenbedingungen in der Halbleiterindustrie dar. Bedingt durch die steigenden Anforderungen an die Fertigung war in den vergangenen Jahren ein starker Anstieg der Investitionskosten für den Aufbau und die Ausstattung einer Halbleiterfabrik zu beobachten. Die Kosten für moderne Fertigungslinien liegen deshalb inzwischen im Bereich von 1-2 Mrd. €. Ungeachtet der Leistungs- und Kostensteigerung ist auf dem Halbleitermarkt ein extremer Preisverfall zu verzeichnen. So ist beispielsweise im Verlauf der letzten 30 Jahre der Preis pro Bit bei Speicherchips um durchschnittlich 32% pro Jahr gesunken. Erschwerend kommt hinzu, dass der Halbleitermarkt starken zyklischen Schwankungen unterliegt, die durch einen Wechsel zwischen Überkapazitäten und Marktknappheit entstehen.

Diese Daten zeigen, dass es für die Unternehmen der Halbleiterindustrie ein Muss ist, die Kosten pro Funktion um ca. 30% pro Jahr zu senken, um die bisherige Wachstumsrate beibehalten zu können. Dies ist gleichbedeutend mit einer Halbierung der Kosten alle 1,5 Jahre. Da im gleichen Zeitraum eine Verdoppelung der Chipfunktionalität erzielt wird, bleiben die Preise pro Chip in etwa stabil.

Wegen dieser Herausforderungen und aufgrund der Tatsache, dass die Halbleiterfertigung inzwischen eine gewisse Reife erlangt hat, wird in Zukunft nicht mehr ausschließlich die technologische Machbarkeit im Vordergrund stehen, sondern es werden die ökonomischen Aspekte der Produktion weiter wachsende Bedeutung erlangen. Gordon Moore, Mitbegründer von Intel, formulierte diesen Sachverhalt vor einiger Zeit folgendermaßen: „I'm convinced that any limitation we run into in the near term will be more likely economic than technical.“ Ziel muss es also sein, durch Verbesserungsmaßnahmen in verschiedenen Bereichen eine hohe Produktivität bzw. Effizienz der Fertigung zu erzielen, um so die Voraussetzungen für Wettbewerbsfähigkeit und Rentabilität zu schaffen und zu vermeiden, dass möglicherweise ökonomische Begrenzungen der zukünftigen IC-Produktion auftreten. Eine große Rolle wird dabei in den nächsten Jahren auch der Bereich Produktionstechnik spielen, also Faktoren, die die Planung und den Ablauf der Fertigung betreffen.

Die oben dargestellten Überlegungen bilden den Rahmen der Arbeit, in der eine Kombination von ökonomischen und technischen Aspekten der Chipfertigung erfolgt. Ausgehend von realistischen Szenarien werden dabei durch umfangreiche Untersuchungen und Modellierung von Parametereinflüssen zukünftige Anforderungen an eine wirtschaftliche IC-Fertigung und der Beitrag verschiedener technischer Maßnahmen zur Verbesserung der Produktion unter besonderer Beachtung des Bereichs Produktionstechnik quantitativ angegeben.

Die Rahmenbedingungen und die allgemeine Situation in der Halbleiterindustrie werden in Kapitel 1 dargestellt.

Auf den Aufbau und die Herstellung des MOS-Transistors und den Fertigungsablauf in einer Fab wird in den Kapiteln 2 und 3 eingegangen.

In Kapitel 4 werden die wirtschaftlichen Ziele von Unternehmen dargestellt und geeignete

Kenngößen zur Bewertung der Leistungsfähigkeit der Produktion definiert. Als Effizienz der Fertigung wird dabei das Verhältnis aus der Zahl der produzierten und funktionsfähigen Chips zu den aufgewendeten Kosten festgelegt. Es wird ein analytisches Modell erstellt, mit dem diese Kenngröße in Abhängigkeit ihrer Eingangsparameter berechnet und die Empfindlichkeit gegenüber Änderungen dieser Parameter angegeben werden kann. Da die Rentabilität einer Fertigung nicht nur durch deren Effizienz bestimmt wird, werden zusätzlich allgemeine Einflussfaktoren und deren Einfluss auf die Gewinnerwartung des Unternehmens erläutert.

Kapitel 5 befasst sich mit dem Einfluss des Bereichs Produktionstechnik auf die Effizienz der Fertigung. In diesem Zusammenhang werden als geeignete Bewertungsgrößen für die Kombination mit den ökonomischen Überlegungen die Zahl der Waferstarts, die Durchlaufzeit, der WIP-Bestand der Produktionslinie und die Kosten eingeführt.

Um die Auswirkungen von Veränderungen mehrerer produktionstechnischer Einflussparameter auf diese Bewertungsgrößen berechnen zu können, wird in Kapitel 6 die Warteschlangentheorie benutzt. Die mathematischen Zusammenhänge dieser Theorie werden erläutert und für einen Teil des Prozessflusses wird ein Warteschlangenmodell, basierend auf Gerätedaten einer realen Fertigung, erstellt. Für die realen Daten wird ein Vergleich mit den mathematischen Verteilungen und dem Warteschlangenmodell vorgenommen. Spezielle Berücksichtigung findet dabei die Variabilität der Fertigung.

Eine Zusammenführung der ökonomischen und technischen Aspekte wird in Kapitel 7 vollzogen. Ausgehend von aktueller Technologie werden unter Berücksichtigung der Rahmenbedingungen in der Halbleiterindustrie mit den in Kapitel 4, 5 und 6 eingeführten Methoden die zukünftigen Anforderungen an die Steigerung der Effizienz berechnet und festgestellt, welchen Beitrag verschiedene Maßnahmen zur Erfüllung dieser Anforderungen leisten können und welches Verbesserungspotential für die einzelnen Faktoren zukünftig noch zur Verfügung steht. Vorteilhaft bei der Betrachtung der Bewertungsgröße „Effizienz der Fertigung“ ist dabei, dass sowohl die jeweilige Leistungssteigerung als auch der damit verbundene Aufwand berücksichtigt werden können. Darüber hinaus werden auf Basis der eingeführten Kenngrößen das Verbesserungspotential beim Ramp Up der Fertigung und die Auswirkungen von ungeplanten Veränderungen vor und während der Produktion berechnet.

Der Preisverfall in der Halbleiterbranche bewirkt, dass für die hergestellten Produkte nur noch ein immer geringerer Erlös erzielt werden kann. Aus diesem Grunde muss die Effizienz der Fertigung gesteigert werden, um so die Stückkosten zu senken und die negativen Effekte des Preisverfalls ausgleichen zu können. Bei einem angenommenen Rückgang des Preises pro Funktion um 32% pro Jahr ergibt sich, dass eine Steigerung der Effizienz um 47% pro Jahr notwendig wäre.

Als günstigste Maßnahme zur Steigerung der Effizienz erweist sich die Erhöhung des Waferdurchmessers von 200 mm auf 300 mm. Dadurch kann eine Effizienzsteigerung um rund 50% erzielt werden, wobei allerdings beachtet werden muss, dass eine Erhöhung des Waferdurchmessers nicht während der Betriebszeit einer Fab durchgeführt werden kann. Die wichtigste Maßnahme zur Steigerung der Effizienz während der Produktlebensdauer stellen Shrinkverfahren dar. Durch eine schnelle Verringerung der Strukturgröße können starke Verbesserungen erzielt werden. Eine Reduktion der Chipfläche um 30% resultiert beispielsweise in einer Effizienzsteigerung von ca. 49%. Weiterer großer Einfluss auf die Effizienz kann durch eine Verringerung der Zahl der Prozessschritte und durch eine Steigerung der Geräteleistungsfähigkeit erzielt werden. Eine um 20% niedrigere Zahl der Prozessschritte, wie sie beispielsweise durch eine optimierte Prozessfüh-

ung erzielt werden kann, führt beim dargestellten Szenario zu einer Effizienzsteigerung um ca. 10%, eine 20%ige Erhöhung des Gerätedurchsatzes zu einer um 7% höheren Effizienz. Durch eine Verringerung der Variabilität bei Prozessankunft und -durchführung um 20% ließe sich eine Effizienzsteigerung um 2-3% erzielen. Die Parameter Ausbeute, Anteil der Monitorwafer und Gerätezuverlässigkeit stellen ebenfalls wichtige Faktoren dar, sind aber bereits jetzt auf einem relativ hohen Niveau angelangt, so dass für die Zukunft kein großes Verbesserungspotential mehr besteht.

Das größte Verbesserungspotential bezüglich einer Verkürzung der Durchlaufzeit, die besonders bei flexiblen Mehrproduktfertigungen von Bedeutung ist, würde eine Verringerung der Losgröße von 25 Wafern pro Los auf Einzelwafertransport bieten. Allerdings wäre dafür eine aufwendige Änderung der Maschinen, Prozesse, Transportsysteme und der Fablogistik notwendig, die momentan nicht machbar und vorgesehen ist. Zusätzlich würde hieraus durch notwendige Setupzeiten eine mögliche Verringerung der Fertigungskapazität resultieren.

Insgesamt kann gesagt werden, dass sich zeitliche Verzögerungen bei der Produktion äußerst negativ auf die Situation des Unternehmens auswirken können, bzw. dass die Beschleunigung von Verbesserungsmaßnahmen ein sehr großes Potential beinhaltet. In der Arbeit wird berechnet, dass bei einer Verzögerung des Markteintritts (Time to Market) um 6 Monate und einer Produktlebensdauer von 5 Jahren bis zu 28% des Marktpotentials nicht genutzt werden können. In diesem Zusammenhang spielt auch die Geschwindigkeit des Ramp Up der Fertigung eine entscheidende Rolle. Bei einer Beschleunigung des Ramp Up von 8 auf 5 Monate kann beispielsweise eine um 50% höhere Zahl von Chips bezogen auf diese 8 Monate hergestellt werden. Einen sehr wichtigen Aspekt stellt dabei die konsequente und schnelle Ausnutzung von Lerneffekten dar.

1 Einführung

1.1 Zum Gegenstand der Arbeit

Die Unternehmen der Halbleiterindustrie stehen in einem extremen internationalen Wettbewerbsumfeld mit hoher Innovationsgeschwindigkeit, hoher Marktdynamik und ständig sinkenden Preisen.

Ziel der Unternehmen muss es sein, trotz steigender technologischer Anforderungen, enormer Investitionen und schwankender Rahmenbedingungen auch in Zukunft rentabel zu sein und die bisherige Innovationsgeschwindigkeit und Wachstumsrate beibehalten zu können. Ständige Verbesserungen in allen Bereichen sind dazu notwendig.

Die ökonomischen Aspekte der Fertigung, vor allem die Produktivität, stellen dabei einen zentralen und immer wichtigeren Punkt dar.

Diese Überlegung bildet den Rahmen der Arbeit. Betriebswirtschaftlich gesehen tragen viele Teilbereiche des Unternehmens, wie z.B. die Beschaffung, Logistik, Entwicklung, Marketing und Vertrieb zum Erfolg bei, was in Abbildung 1.1 [1] dargestellt wird. Hier soll für die ökonomischen und technischen Überlegungen aus der Wertschöpfungskette für die Halbleiterindustrie jedoch nur der Bereich Produktion betrachtet werden, d.h. dass der Einfluss der anderen Unternehmensbereiche und Aktivitäten auf die Effizienz nicht untersucht, bzw. als konstant angesehen wird. Im speziellen soll dabei nur auf die Silizium-Prozesskette der Fertigung vom unprozessierten bis zum vollständig prozessierten Wafer eingegangen werden. Nicht betrachtet werden hier abschließende Testvorgänge und die Häusung der Chips.

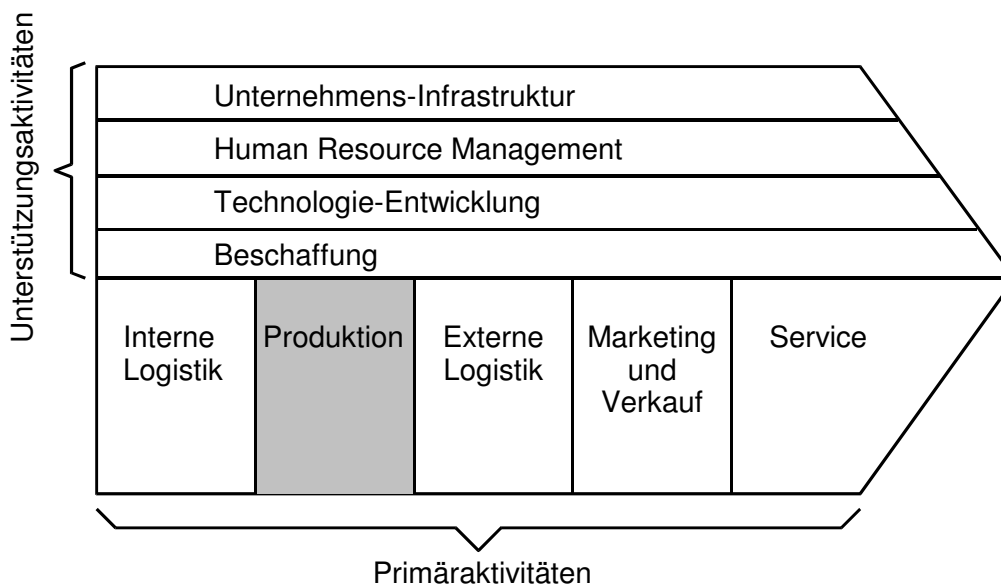


Abbildung 1.1: Wertschöpfungskette [1]

In der Arbeit werden zukünftige ökonomische und technologische Anforderungen an die Fertigung integrierter Schaltungen bei weiter sinkenden Strukturgrößen dargestellt.

Ziel ist es anzugeben, welche Maßnahmen am besten dazu beitragen, ökonomische Begrenzungen der zukünftigen IC-Fertigung zu vermeiden und welche Verbesserungen

durch diese Maßnahmen in der Fertigung erzielt werden können.

Ausgehend von einer grundlegenden Erläuterung von Rentabilität, Wirtschaftlichkeit, Produktivität und Effizienz, bezogen auf die speziellen Anforderungen und Rahmenbedingungen der Halbleiterfertigung, wird geklärt, welche Komponenten Einfluss auf die Wirtschaftlichkeit ausüben, welche gegenseitigen Abhängigkeiten bestehen und welches Verbesserungspotential in den einzelnen Bereichen besteht.

Speziell für den Bereich Produktionstechnik werden ausführliche Überlegungen angestellt, um so die technischen Auswirkungen von Veränderungen der Einflussgrößen in diesem Bereich und ihren Einfluss auf die Effizienz der Fertigung feststellen zu können. Aufbauend auf realen Fertigungsdaten wird ein Warteschlangenmodell erstellt, mit dem die Effekte von Parameteränderungen, insbesondere auch der Variabilität, berechnet werden können. Damit können Anforderungen an eine wirtschaftliche Fertigung und Maßnahmen zu deren Erfüllung unter besonderer Beachtung des Bereichs Produktionstechnik abgeschätzt werden.

1.2 Bedeutung der Mikroelektronik und Dynamik des Halbleitermarktes

Die Produkte der Mikroelektronik haben heutzutage ihren Platz in fast allen Bereichen unseres Lebens gefunden. In vielen Branchen, wie z.B. der Computerindustrie, Kommunikationstechnik oder der Automobilindustrie, besteht ein großer Bedarf an integrierten Schaltungen. Ein entscheidender Grund für die starke Verbreitung der Mikroelektronik in fast allen Märkten ist die Tatsache, dass es bis jetzt stetig gelungen ist, die Leistungsfähigkeit der in Siliziumtechnologie hergestellten Schaltungen zu erhöhen und gleichzeitig die Kosten pro Funktion zu senken, so dass die Produkte für die Kunden zu einem attraktiven Preis angeboten werden können.

Am besten wird die Evolution der Halbleitertechnologie beschrieben durch das Gesetz von Gordon Moore (Abbildung 1.2) [2], [5], das besagt, dass die Zahl der Komponenten pro Chip zeitlich exponentiell ansteigt, was z.B. zu einer Vervierfachung der Speicherkapazität von DRAM-Bausteinen alle drei Jahre führt.

Die zeitliche Steigerung des Integrationsgrades aufgrund der technologischen Fortschritte und die dafür üblicherweise verwendeten Bezeichnungen, seit der ersten Realisierung von integrierten Schaltungen in den 50er Jahren, sind in Tabelle 1.1 angegeben [3]. Die Industrie befindet sich zur Zeit im ULSI-Zeitalter mit 10^7 bis 10^9 machbaren Transistoren pro Chip. Eine weitere Steigerung der Integrationsdichte ins Giga-Zeitalter mit mehr als 10^9 Transistoren pro Chip wird in den nächsten Jahren erreicht.

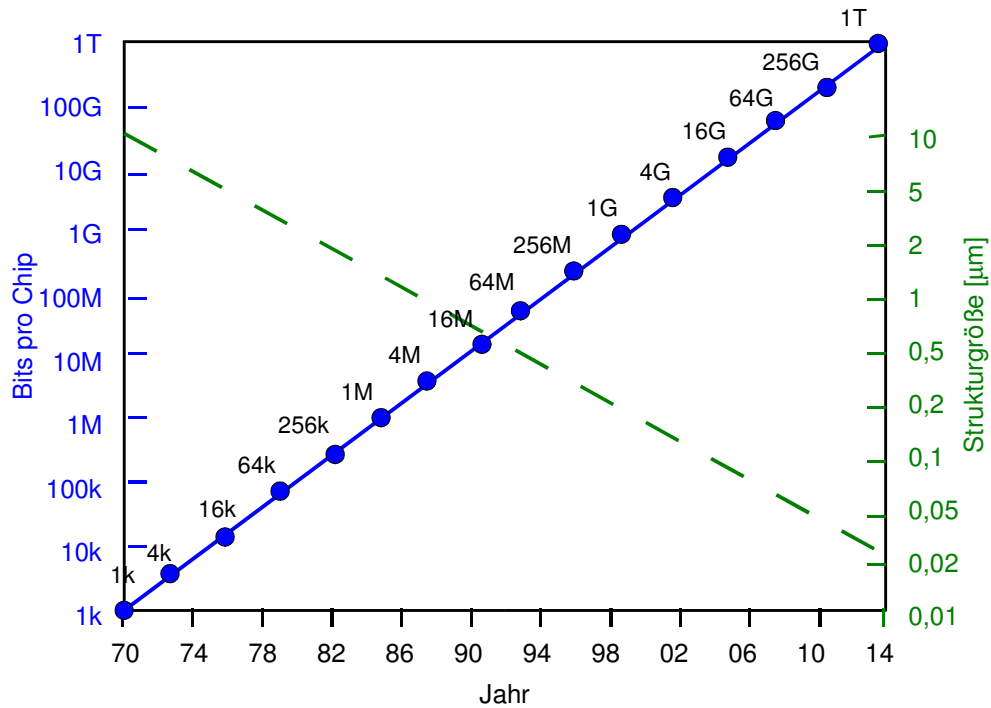


Abbildung 1.2: Mooresches Gesetz: Entwicklung der Speicherdichte und Strukturgröße [145]

Tabelle 1.1: Integrationstrends

Integrationsgrad	Jahr	Zahl der Transistoren	DRAM
SSI	50er	$< 10^2$	
MSI	60er	$10^2 - 10^3$	
LSI	70er	$10^3 - 10^5$	4k, 16k, 64k
VLSI	80er	$10^5 - 10^7$	256k, 1M, 4M
ULSI	90er	$10^7 - 10^9$	16M, 64M, 256M
SLSI	2000	$> 10^9$	1G, 4G, ...

Analog zur Gordon Moore Kurve existiert für Mikroprozessoren das sogenannte Gesetz von Joy, das die Entwicklung der Rechenleistung von Prozessoren vorhersagt. Man geht dabei von einer jährlichen Verdoppelung der Prozessorleistung in MIPS (Millionen Instruktionen pro Sekunde) seit dem Jahr 1984 aus, was durch die Formel

$$\text{MIPS} = 2^{\text{Jahr} - 1984}$$

beschrieben werden kann.

In Abbildung 1.3 wird die bisherige und prognostizierte Entwicklung der Rechenleistung dargestellt. Dabei ist anzumerken, dass die ‚Industrie-Kurve‘, also die tatsächliche Entwicklung der Prozessorleistung, in den letzten Jahren etwas hinter der im Gesetz von Joy vorhergesagten Entwicklung zurückbleibt.

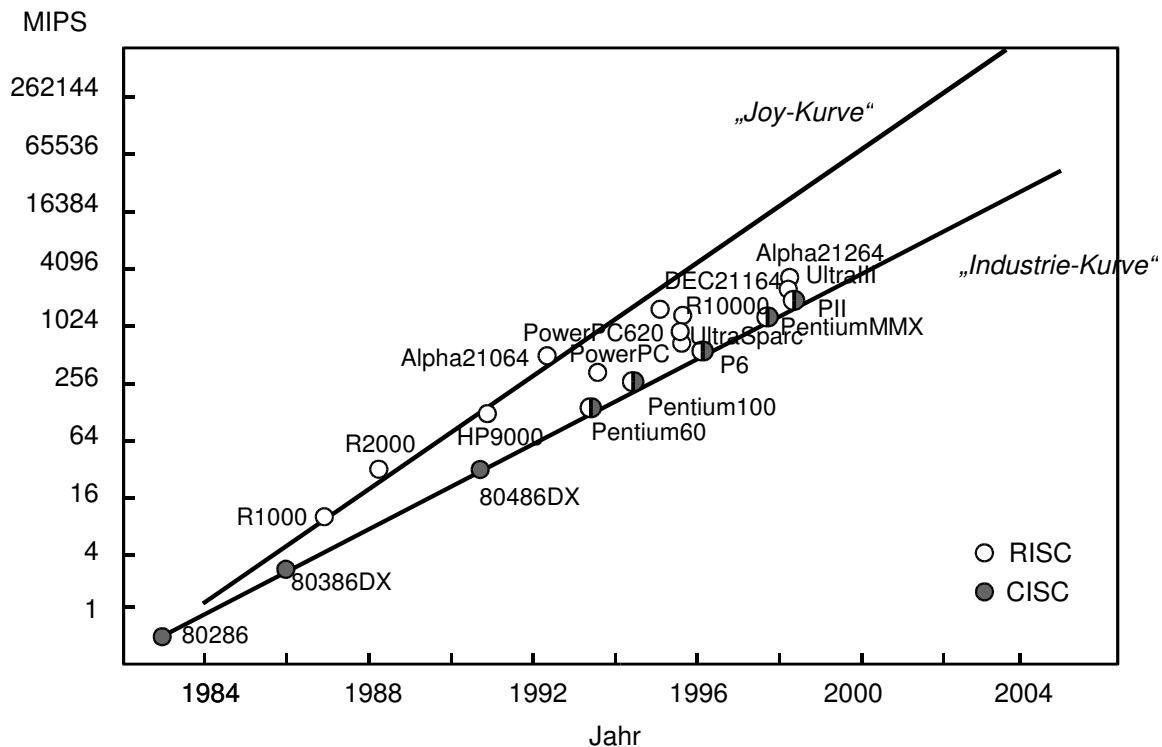


Abbildung 1.3: Entwicklung der Prozessorleistung [145]

Die prognostizierte Rangliste der größten Halbleiterhersteller im Jahr 2006 nach ihrem Umsatz und die Veränderung im Vergleich zu 2005 ist in Tabelle 1.2 angegeben [162]. Der Prozessorhersteller Intel ist die bei weitem größte Firma in diesem Bereich. Ihr Umsatz erreicht 2006 ungefähr 12% des weltweiten Halbleitermarktes und liegt damit weit höher als bei allen anderen Firmen.

Tabelle 1.2: Rangliste der Halbleiterhersteller 2006

Rang 2003	Firma	Umsatz 2005 [Mio. \$]	Umsatz 2006 [Mio. \$]	Veränderung 2005 - 2006
1	Intel	35395	31580	-11%
2	Samsung Electronics	17838	19475	9%
3	Texas Instruments	11300	13870	23%
4	Infineon + Qimonda	8297	10060	21%
5	Toshiba	9045	10030	11%
6	STMicroelectronics	8870	9930	12%
7	TSMC	8217	9715	18%
8	Renesas	8266	8170	-1%
9	Hynix	5599	7375	32%
10	NXP	5598	6365	14%

1.3 Darstellung der wichtigsten Produktgruppen im Halbleitermarkt

Grundsätzlich lassen sich die von der Halbleiterindustrie erzeugten integrierten Schaltungen in zwei Kategorien, nämlich in Speicher- und Logikbausteine einteilen:

Speicherbausteine:

Speicherbausteine dienen ausschließlich der Speicherung von Information und führen keine weitere Datenbearbeitung durch.

Die wichtigsten Vertreter dieser Kategorie stellen DRAMs, SRAMs und Flash-EPROMs dar.

Der historische Umsatz auf dem Speichermarkt ist in Abbildung 1.4 ([4] ergänzt durch [148], [161]) angegeben. Aufgrund des starken Preisverfalls waren immer wieder, trotz größerer verkaufter Stückzahlen, ein Rückgang des Marktes und starke Marktschwankungen bezogen auf den Umsatz festzustellen.

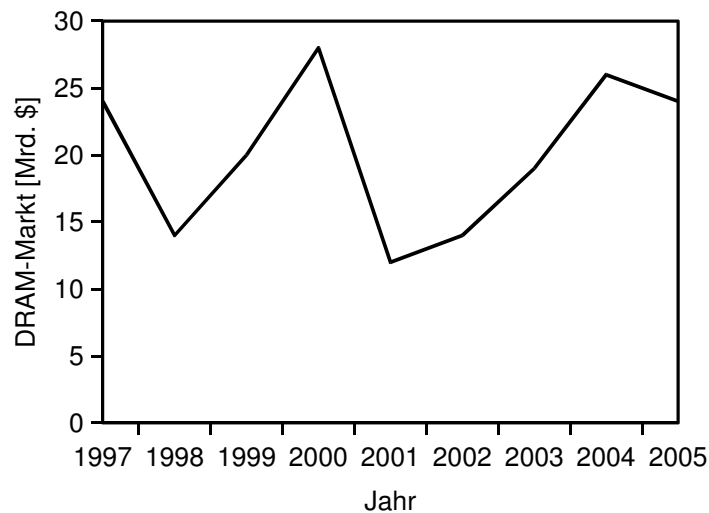


Abbildung 1.4: DRAM-Markt [4], [148], [161]

Den größten Teil des Speichermarktes nehmen DRAMs ein. Mit diesen Bausteinen wurde 2005 weltweit ein Umsatz von ungefähr 25 Mrd. US \$ erzielt.

Da DRAMs den größten Marktanteil im Speicherbereich besitzen und gleichzeitig die Bausteine mit der höchsten Speicherkapazität (bzw. Integrationsdichte) darstellen und deshalb auch sehr große Anforderungen an die Herstellungsprozesse stellen, konzentriert sich die Diskussion häufig auf diese Form der integrierten Schaltungen.

Speicherchips stellen Massenprodukte für den anonymen Markt dar und werden üblicherweise mit der modernsten Technologie hergestellt. Die Einführung neuer Produktgenerationen verläuft sehr rasch, wie im Mooreschen Gesetz dargestellt, es sind aber gleichzeitig mehrere Produkte unterschiedlichen Reifegrads auf dem Markt, was in Abbildung 1.5 gezeigt wird [163].

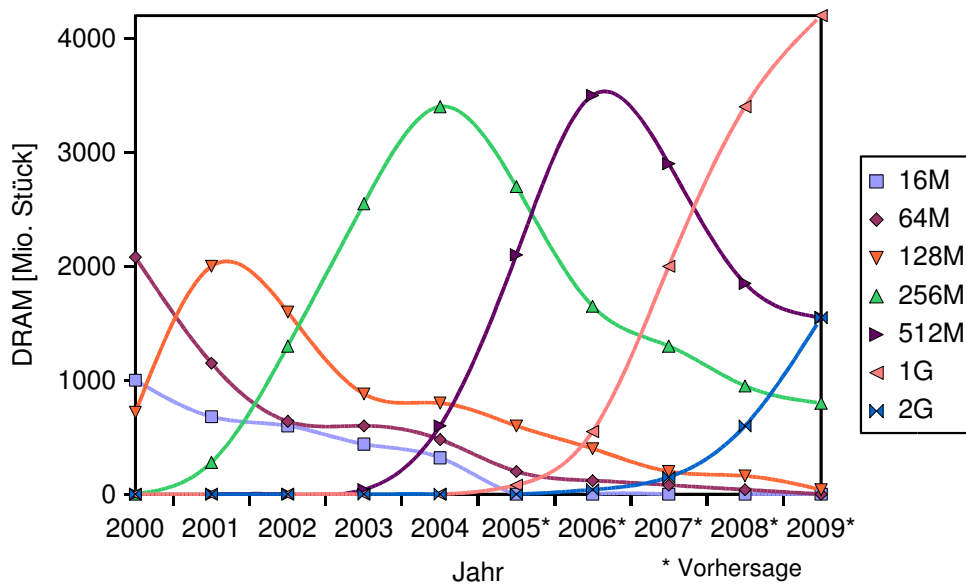


Abbildung 1.5: Verkaufte Zahl von DRAMs in Abhängigkeit von der Speicherkapazität [163]

Im Jahr 2003 war der 256 Mbit DRAM mit ungefähr 2,5 Mrd. Einheiten der meistverkaufte Speicherbaustein, während der 512 Mbit-Speicher erst am Anfang seiner Produktlebensdauer stand und, bei einer stetigen Steigerung der produzierten Menge, die maximale Stückzahl erst einige Jahre später erreichen wird.

Für eine neue Generation von Speichern muss üblicherweise die Speicherzelle neu entworfen und die Fertigungstechnologie neu entwickelt werden, was erhebliche Entwicklungskosten verursacht, die bei DRAMs teilweise durch Entwicklungskooperationen großer Firmen verringert werden.

Historisch gesehen stellten DRAMs die sogenannten Technologietreiber dar, also die Bausteine, bei denen die modernste Herstellungstechnologie zuerst angewendet und erprobt wird. Aufgrund der hohen produzierten Stückzahlen und der regulären Strukturen sind DRAMs für eine solche Aufgabe gut geeignet. Inzwischen ist die Rolle des Technologietreibers aber auf die Logikschaltungen übergegangen, die besonders im Bereich der Metallisierung deutlich aufwendiger als DRAMs zu fertigen sind.

Logikbausteine:

Logikbausteine nehmen, im Gegensatz zu Speichern, eine Bearbeitung der erhaltenen Daten vor und führen damit eine festgelegte oder programmierbare Funktion aus.

Wichtige Bestandteile des Bereichs der Logikchips stellen Mikrokomponenten und ASICs dar.

Unter der Bezeichnung Mikrokomponenten werden hier Mikroprozessoren, Mikrocontroller, DSPs und Microperipherals zusammengefasst.

Auch in diesem Marktbereich handelt es sich um Massenprodukte. Besonders bei Mikroprozessoren sind erhebliche Kosten und hoher Zeitaufwand für die Entwicklung einer neuen Generation notwendig. Während der Produktlebensdauer werden dann durch

technologische Verbesserungen, ohne großen zusätzlichen Entwicklungsaufwand, laufend neue Modelle mit höherer Taktfrequenz auf den Markt gebracht.

Wegen der höheren Verkaufspreise bei den Mikrokomponenten werden in diesem Bereich trotz der niedrigeren Verkaufszahlen höhere Umsätze erzielt als bei den Speicherchips. Die Marktveränderungen in diesem Bereich sind in Abbildung 1.6 dargestellt [4, S. 6-2].

ASICs (Anwendungsspezifische integrierte Schaltungen) stellen kundenspezifische Bausteine dar, die eine vom Kunden definierte Funktion erfüllen und die, je nach Bedarf, in kleineren oder größeren Stückzahlen produziert werden.

Die Fertigung von ASICs stellt eine auftragsorientierte Serienfertigung dar, für die eine festgelegte Fertigungstechnologie verwendet wird. Dabei müssen die Schaltungsentwicklung und Produktion nicht von der gleichen Firma durchgeführt werden. Die Herstellung kann also auch als Dienstleistung von darauf spezialisierten Firmen (Foundries) ausgeführt werden. Von dieser Vorgehensweise wird erwartet, dass sie in Zukunft stärker genutzt wird.

Der größte Anteil der ASICs wird in Form von CMOS-Standardzellen realisiert.

Der mit ASICs erzielte Umsatz lag 2003 bei ca. 17 Mrd. \$ (Abbildung 1.6) [148]. Wie auch bei den Mikrokomponenten ist hier festzustellen, dass aufgrund der Krise des Halbleitermarktes in den letzten Jahren die Gesamtumsätze rückläufig waren und sich erst jetzt wieder erholen.

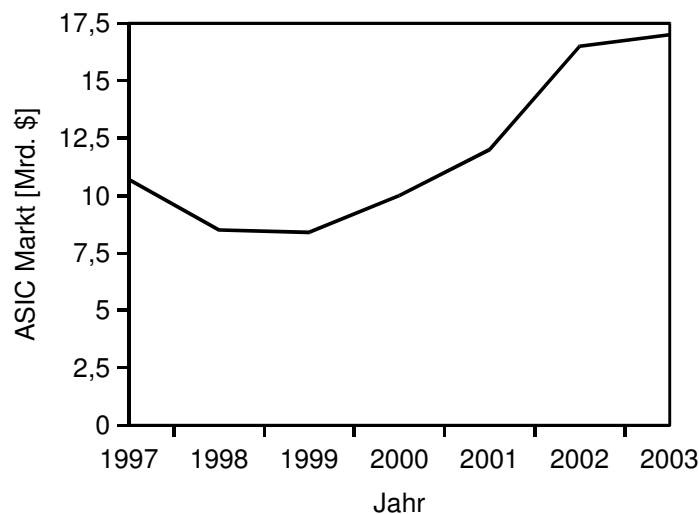


Abbildung 1.6: ASIC-Markt [148]

Die Entwicklung des gesamten Weltmarkts für Halbleiter ist in Abbildung 1.7 dargestellt ([4] ergänzt durch [160], [161]). Im Jahr 2005 betragen die Umsätze mit Halbleiterbauelementen ca. 228 Mrd. \$. Neben den bereits erwähnten Marktanteilen von Speicher- und Logikbausteinen ist im Weltmarkt für Halbleiter noch der wertmäßige Anteil von Bipolar-Technologie, Analogschaltungen und diskreten Bauelementen enthalten, auf die hier nicht näher eingegangen werden soll.

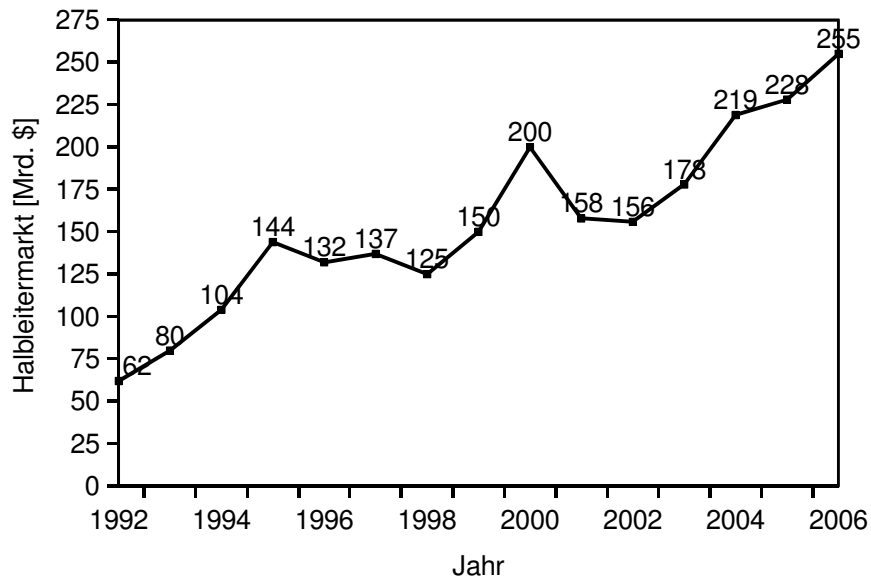


Abbildung 1.7: Weltmarkt für Halbleiter [4], [160], [161]

In Tabelle 1.3 ist die Aufteilung des Halbleitermarktes in die einzelnen Produktsegmente und deren Steigerungsrate von 2002 nach 2003 angegeben.

Tabelle 1.3: Aufteilung des Halbleitermarktes

Marktsegment	2002 [Mio. \$]	2003 [Mio. \$]	Veränderung [%]
Wireless	22435	28494	+27%
Compute	55485	64046	+15%
Consumer	26571	29567	+11%
Automotive	13068	14573	+12%
Indus/Med/Other	12127	13454	+11%
Wired	13109	14131	+8%
Storage	9090	9373	+3%
Mil/Aero	3744	3814	+2%
Gesamtmarkt	155629	177452	+14%

Im langjährigen Mittel ergibt sich eine Steigerung der Halbleitermarktes um durchschnittlich 11% pro Jahr. In der Abbildung ist jedoch auch erkennbar, dass in den letzten Jahren, bedingt durch den Preisverfall, ein Marktrückgang zu verzeichnen war, was zur Folge hat, dass der Halbleitermarkt in nächster Zeit bei weitem nicht die Höhe erreichen wird, die von vielen Experten noch vor kurzem vorhergesagt wurde.

1.4 Ökonomische Rahmenbedingungen der Halbleiterindustrie

Die Fertigung von integrierten Schaltungen ist einerseits gekennzeichnet durch eine extrem große Innovationsgeschwindigkeit und Steigerung der Leistungsfähigkeit ihrer Produkte, andererseits aber auch durch problematische Rahmenbedingungen wie sehr hohe Investitionskosten und ständig sinkende Preise.

Diese Rahmenbedingungen stellen ein ernstzunehmendes Problem für die Chipfertigung dar und müssen in die Betrachtungen mit einbezogen werden, wenn Aussagen über die Anforderungen und die ökonomischen Konsequenzen der zukünftigen IC-Fertigung gemacht werden sollen.

Bedingt durch die immer stärkeren Qualitätsanforderungen an die Fertigung war in den letzten Jahren ein stetiger starker Anstieg der Investitionskosten für den Aufbau einer Halbleiterfabrik zu beobachten (Abbildung 1.8) [4, S. 3-6], [8].

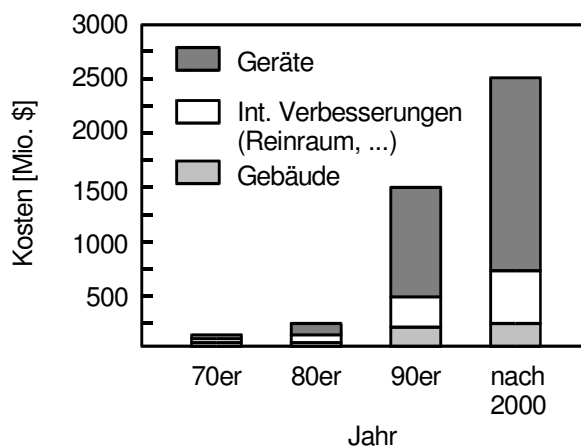


Abbildung 1.8: Kosten für den Aufbau einer Chipfabrik [4], [8]

Momentan liegen die Kosten für eine Halbleiterfabrik im Bereich von 1,5 Mrd. \$, wobei der Hauptanteil der Investitionen dabei auf die Anschaffung der Herstellungsgeräte entfällt (momentan ca. 70%). Dieser hohe Prozentsatz läßt sich damit erklären, dass es einerseits, durch die ständig sinkenden Strukturgrößen und den steigenden Waferdurchmesser schwieriger wird, die geforderten Prozesstoleranzen und die Prozesshomogenität auf den Wafern und von Wafer zu Wafer zu erfüllen, was zu einer Verteuerung der Geräte führt, bzw. auch eine Umstellung auf andere Prozesse notwendig machen kann. Andererseits wird in der Fabrik wegen der Vielzahl von Prozessschritten eine sehr große Anzahl der teuren Geräte benötigt, um die Massenproduktion durchführen zu können. Wegen der kontinuierlich sinkenden Strukturgröße der Bauelemente ist es zusätzlich notwendig während der Betriebsdauer einer Fabrik laufend alte Geräte durch neue zu ersetzen, was zu zusätzlichen Investitionen führt.

Hinzu kommt, dass aufgrund der hohen Anforderungen an die Qualität des Reinraums die Kosten auch in diesem Bereich in der Vergangenheit laufend gestiegen sind. Eine Erleichterung der Anforderungen an den Reinraum kann inzwischen durch die Einführung des sogenannten „SMIF-Konzepts“ mit einer lokalen Verbesserung der Reinraumqualität erreicht werden (siehe Kapitel 3).

Ungeachtet der Leistungs- und Kostensteigerung ist auf dem Halbleitermarkt ein extremer Preisverfall zu verzeichnen. Beispielsweise ist im Verlauf der letzten 30 Jahre der

Preis/Funktion (= Preis/Bit) für Speicherchips um durchschnittlich 32% pro Jahr gesunken (Abbildung 1.9) [6].

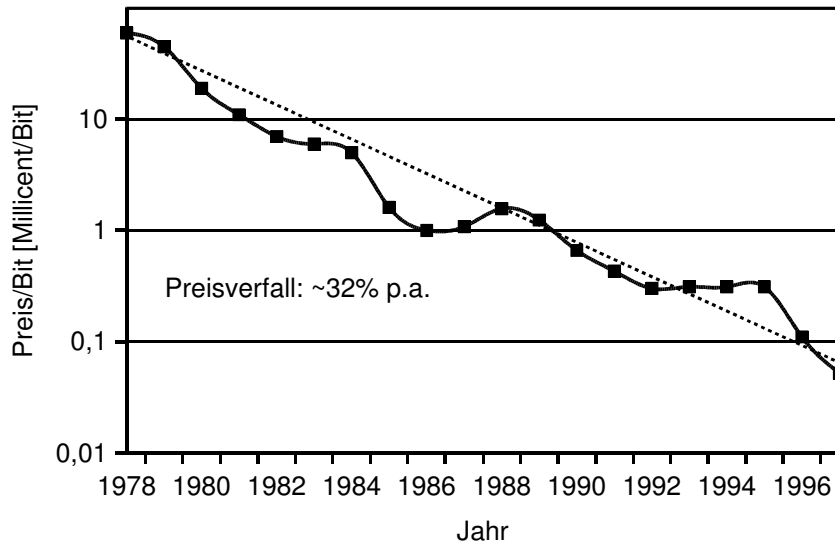


Abbildung 1.9: Preisentwicklung bei DRAM-Speicherchips [6]

Welche Schwankungen auch kurzfristig bei den DRAM-Preisen auftreten können, zeigt Abbildung 1.10 [9]. Während der Preis für ein 64 MBit DRAM im ersten Quartal 1999 bei 4 - 7 \$ lag, wurde, bedingt durch höhere Nachfrage und zusätzlich durch das Erdbeben in Taiwan, im Oktober ein Höhepunkt mit 27 \$ erreicht. Schon nach kurzer Zeit pendelte sich der Preis jedoch wieder bei ungefähr 10 \$ ein.

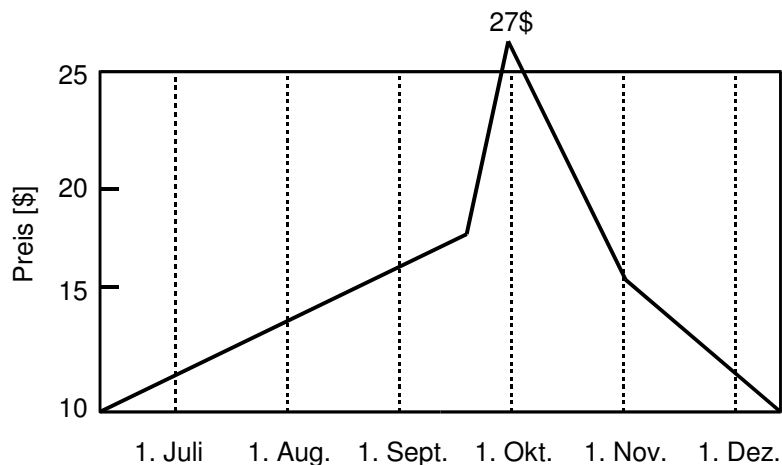


Abbildung 1.10: Preisentwicklung des 64 MBit DRAMs 1999 [9]

Ein ähnlich starker Preisverfall ist auch für den Mikroprozessormarkt typisch. Für ältere Prozessormodelle können nur erheblich geringere Preise erzielt werden als für die neueste, leistungsstärkste Generation. Allerdings liegen in diesem Marktsegment die Einstandspreise wesentlich höher als bei Speicherchips. Beispielsweise lag der Verkaufspreis für einen Intel Pentium 4 Prozessor mit 3 GHz Taktfrequenz Anfang des Jahres 2005 bei

rund 230 €, während dieses Modell im Herbst 2006 bereits für rd. 150 € zu kaufen war. [10].

Bei kundenspezifischen ASICs mit kleineren Stückzahlen dürfte die Preisentwicklung etwas entspannter verlaufen, wobei auch hier ein starker Wettbewerbsdruck besteht.

Der Halbleitermarkt weist zwar im Mittel eine hohe Wachstumsrate auf, diese Rate schwankt aber stark, was eine sichere Planung der Investitionen erschwert. In Abbildung 1.11 wird deutlich, dass das Marktwachstum über Jahre betrachtet zwischen ca. +50% und -30% pro Jahr schwankt ([4] ergänzt durch [160], [161]).

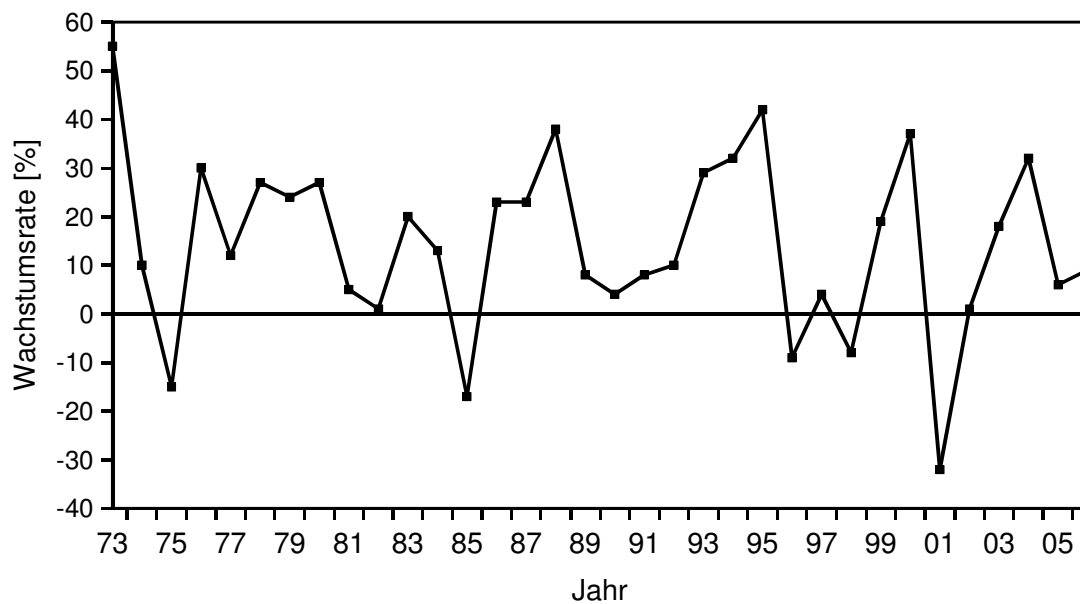


Abbildung 1.11: Wachstumsrate des Halbleitermarkts [4], [160], [161]

Um die Innovationsgeschwindigkeit beizubehalten und die modernen Prozesse einführen zu können, ist großer finanzieller Aufwand der Unternehmen notwendig. Die jährlichen Investitionen der Halbleiterindustrie liegen deswegen im langjährigen Durchschnitt bei 21% des Produktionswertes, wobei auch hier starke Schwankungen zu verzeichnen sind. Nach einem Höhepunkt der Investitionen im Jahr 1996 sanken z.B. 1997 und 1998 die Ausgaben aufgrund von Überkapazitäten (Abbildung 1.12) [4].

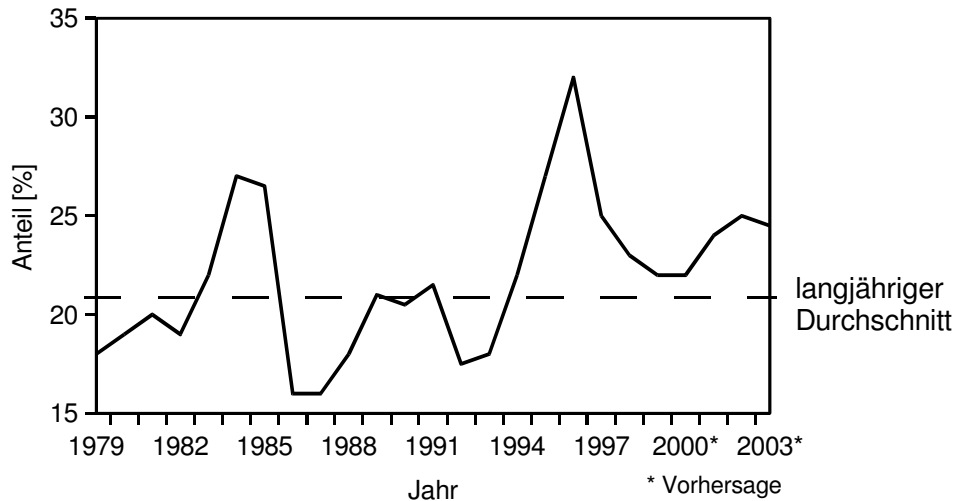


Abbildung 1.12: Prozentualer Anteil der Investitionen am Produktionswert der Halbleiterindustrie [4]

Ähnlich starke Schwankungen ergeben sich, wenn man nur die Ausgaben für Forschung und Entwicklung im Verhältnis zum Gesamtumsatz betrachtet. Im langjährigen Durchschnitt ergibt sich hier ein Anteil von knapp 11%, ein Wert, der weit über dem vieler anderer Industriebranchen liegt, wie Abbildung 1.13 zeigt und nur von der Software- und Pharmaindustrie übertroffen wird [4, S. 3-10].

Insgesamt zeigt sich, dass die schwierigen Rahmenbedingungen große Herausforderungen für die Chiphersteller darstellen, die ständige Verbesserungen der Technologie und Produktionstechnik erreichen müssen, um wettbewerbsfähig zu bleiben. Inzwischen haben ökonomische Faktoren fast die gleiche Bedeutung wie technologische Faktoren erlangt [11].

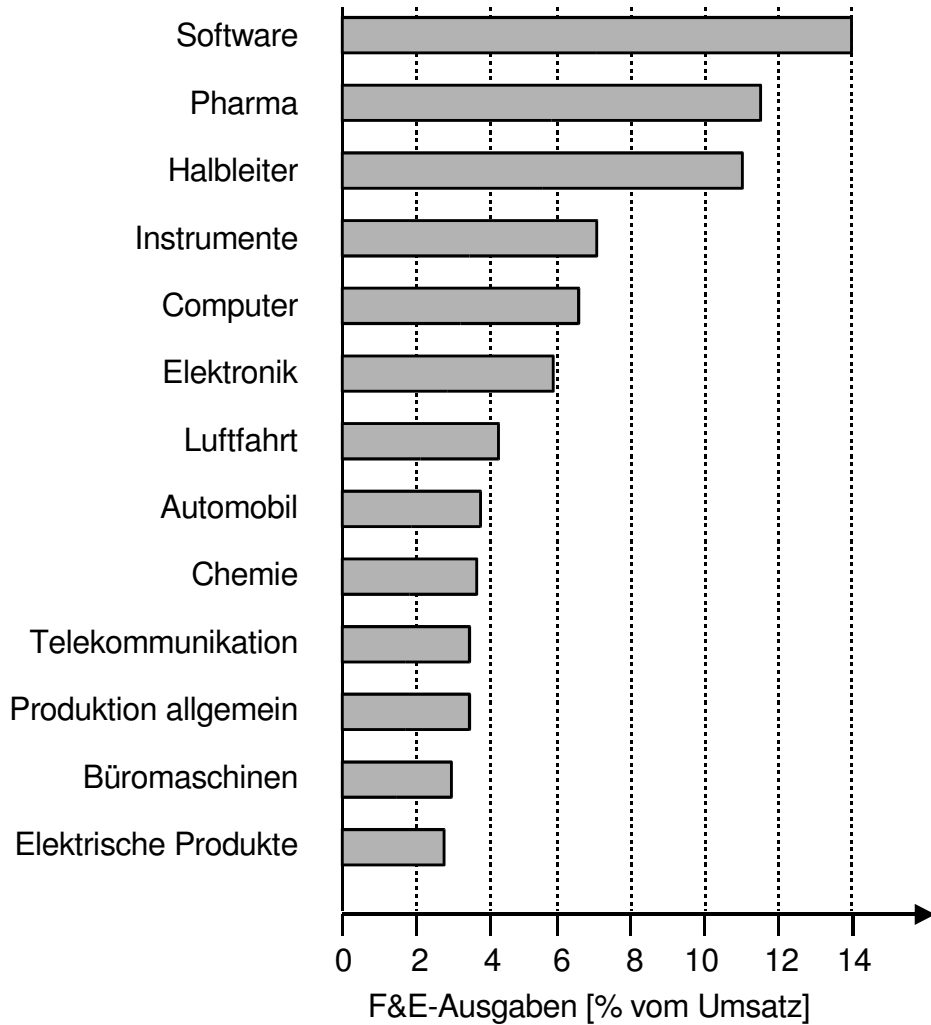


Abbildung 1.13: Prozentualer Anteil der F&E-Ausgaben vom Umsatz bei verschiedenen Industriebranchen 1999 [4]

Trotz all dieser Herausforderungen ist es der Halbleiterindustrie bis jetzt gelungen eine hohe Innovationsgeschwindigkeit beizubehalten und die Herstellungskosten laufend zu senken.

1.5 Entwicklung des Silizium-Grundmaterials

Das Halbleitermaterial mit der weitaus höchsten Bedeutung stellt Silizium (Si) dar. Rund 99% aller Bausteine werden zur Zeit in Si realisiert. Mit weitem Abstand folgen dahinter Gallium-Arsenid (GaAs) und Germanium (Ge), die nur für Spezialanwendungen, wie z.B. optische Schaltungen oder Hochfrequenzschaltungen, benutzt werden.

Grundsätzlich existieren verschiedene Varianten von Si-Wafern. Man unterscheidet zwischen herkömmlichen Wafern, bei denen eine einheitliche Si-Schicht über die gesamte Dicke des Wafers vorliegt, und Epi-Wafern, die eine dünne epitaktische Si-Schicht auf einem höher dotierten Si-Substrat aufweisen. Bei beiden Wafertypen kann entweder p- oder n-dotiertes Grundmaterial vorliegen. Für komplexe Logikschaltungen wurden bei kleinen Strukturgrößen zwischenzeitlich bevorzugt Epi-Wafer verwendet. Daneben existieren noch SOI-Wafer (Silicon on Insulator), bei denen durch eine zusätzliche isolierende Oxidschicht die Bauelemente vom Substrat getrennt werden.

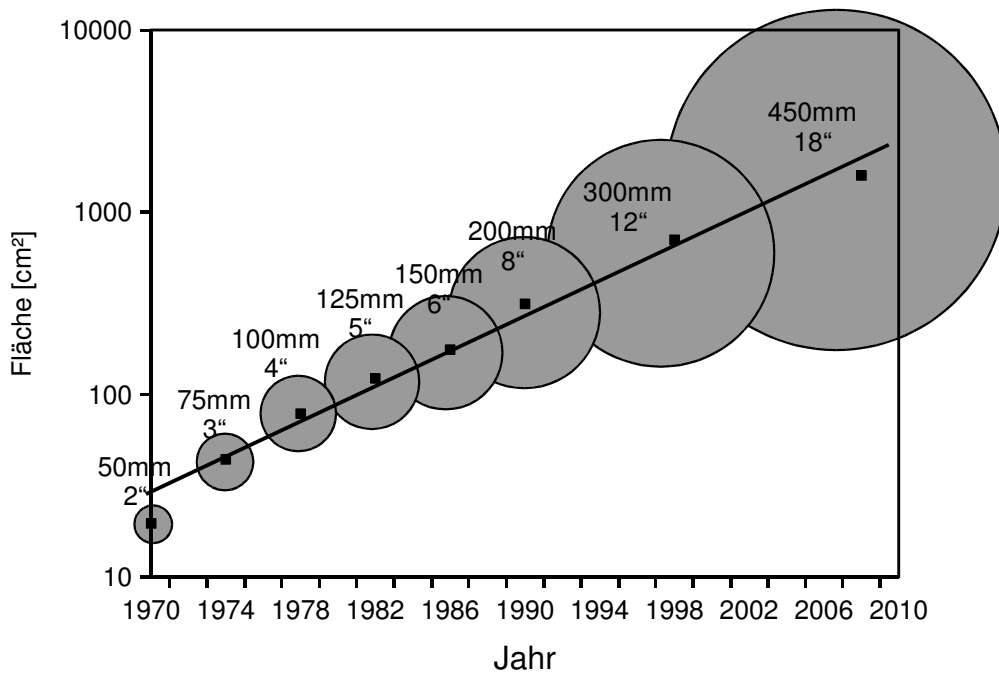


Abbildung 1.14: Entwicklung von Waferdurchmesser und -fläche [12], [13]

Im Laufe der Zeit ist es gelungen, den Durchmesser der Si-Wafer, die als Grundmaterial für die Chipfertigung benutzt werden, laufend zu steigern, was in Abbildung 1.14 dargestellt ist [12], [13].

Momentan stellt für moderne Fabriken 300 mm-Fertigung den Stand der Technik dar, wobei aber noch viele ältere Fabs mit geringerem Waferdurchmesser existieren.

Die Entwicklung einiger wichtiger Waferparameter in Abhängigkeit vom Waferdurchmesser ist in Tabelle 1.4 angegeben [15], [16].

Tabelle 1.4: Waferparameter

Parameter	Waferdurchmesser				
	100 mm	125 mm	150 mm	200 mm	300 mm
Fläche [cm ²]	79	123	177	314	707
Dicke [µm]	525	625	675	725	775
Masse [g]	10	19	28	53	128

Mit der Steigerung des Waferdurchmessers erhöht sich zwar die nutzbare Fläche, gleichzeitig erschwert sich aber auch die Herstellung der Wafer, da immer größere Mengen des einkristallinen Materials fehlerfrei aus der Schmelze gezogen werden müssen. (Eine Übersicht über die Herstellung von monokristallinen Si-Wafern ist z.B. in [17] und [18] enthalten.) Ebenfalls verkompliziert sich die Handhabung und der Transport der größeren und schwereren Si-Scheiben.

Wichtige Hersteller für Si-Grundmaterial stellen z.B. die Firmen Wacker Siltronic, Sumi-

tomo Sitix, Shin-Etsu oder MEMC dar.

1.6 Entwicklung der wichtigsten Prozesstechnologien

Die marktbeherrschende Stellung unter den Prozesstechnologien nimmt eindeutig CMOS ein. Aufgrund seiner günstigen Eigenschaften besonders in Bezug auf Leistungsverbrauch, Integrationsfähigkeit und durch die inzwischen angesammelte Erfahrung in Entwicklung und Fertigung hat CMOS eine Vormachtstellung erreicht, die auch in Zukunft ausgebaut werden wird.

Eine weitere Prozesstechnologie, die Steigerungsraten aufweist, ist BiCMOS. Es stellt eine Mischform dar und soll die Vorteile von CMOS und Bipolar vereinen, wobei jedoch höhere Komplexität und Kosten die Anwendungen beschränken. Wie in Tabelle 1.5 dargestellt ist, besitzen CMOS und BiCMOS gemeinsam ungefähr einen Marktanteil von 90% [19].

Bipolartechnologie wird hauptsächlich für analoge Schaltungen angewandt und hat wegen der Dominanz der digitalen Schaltungstechnik in den vergangenen Jahren an Marktbedeutung verloren.

Tabelle 1.5: Marktanteile der Prozesstechnologien

Prozesstechnologie	Marktanteil [%]			
	1970	1980	1994	1999
MOS				
PMOS	31	5	0	0
NMOS/HMOS	2	37	1	< 1
CMOS	2	10	78	85
BiCMOS	0	0	5	5
Bipolar				
ECL	3	3	1	<1
TTL	29	8	<1	0
S/LS TTL	7	13	2	<1
Linear	26	24	12	8
GaAs	0	0	<1	<1

Aufgrund der Vormachtstellung von CMOS werden im folgenden nur MOS-Transistoren und Verfahren zu ihrer Herstellung betrachtet.

1.7 Internationale Organisationen und Kooperationen verschiedener Hersteller

Das große Investitionsvolumen in der Halbleiterindustrie und die Komplexität der Herstellung führen dazu, dass sich immer wieder internationale Organisationen und Kooperationen zwischen verschiedenen Firmen bilden, um internationale Standards einzuführen und die hohen Kosten besser zu verteilen.

Wichtige internationale Organisationen stellen SEMI und SEMATECH dar.

SEMI (Semiconductor Equipment and Materials International) ist eine internationale Handelsorganisation mit 2300 Mitgliedsfirmen, die sich vor allem im Bereich der Standardisierungen engagiert, Messen und Konferenzen organisiert und Informationsmaterial sowie Marktstatistiken erstellt.

SEMATECH (Semiconductor Manufacturing Technology) stellt ein Forschungskonsortium dar, das sich ebenfalls, wie SEMI, mit der Weiterentwicklung der Halbleitertechnologie beschäftigt. Unter anderem wird die SIA Roadmap mit Prognosen über die Zukunft der Halbleiterindustrie von SEMATECH herausgegeben.

Speziell zur Erleichterung des Umstiegs auf 300 mm Wafer wurden das I300I-Konsortium (International 300 mm Initiative) und in Japan Selete (Semiconductor Leading Edge Technologies) gegründet. Die darin engagierten Mitgliedsfirmen bemühen sich z.B. durch Einigung auf allgemeine Standards die Risiken und Kosten des Umstiegs möglichst gering zu halten. Dabei arbeiten die beiden Konsortien in weiten Bereichen zusammen.

Weltweit existiert eine Vielzahl von Kooperationen zwischen Unternehmen der Halbleiterbranche. So hat beispielsweise das Joint Venture „Semiconductor300“ zwischen Infineon und Motorola in Dresden zum weltweit ersten Aufbau einer Produktionslinie für 300 mm Wafer geführt.

2 Aufbau und Herstellung des MOS-Transistors

2.1 Aufbau des MOS-Transistors

Der typische schichtweise Aufbau eines aktuellen nMOS- und pMOS-Transistors in einer CMOS-Schaltung, der sich auch bei den kommenden Bauelementegenerationen nicht wesentlich ändern wird, ist in Abbildung 2.1 dargestellt [20], [21].

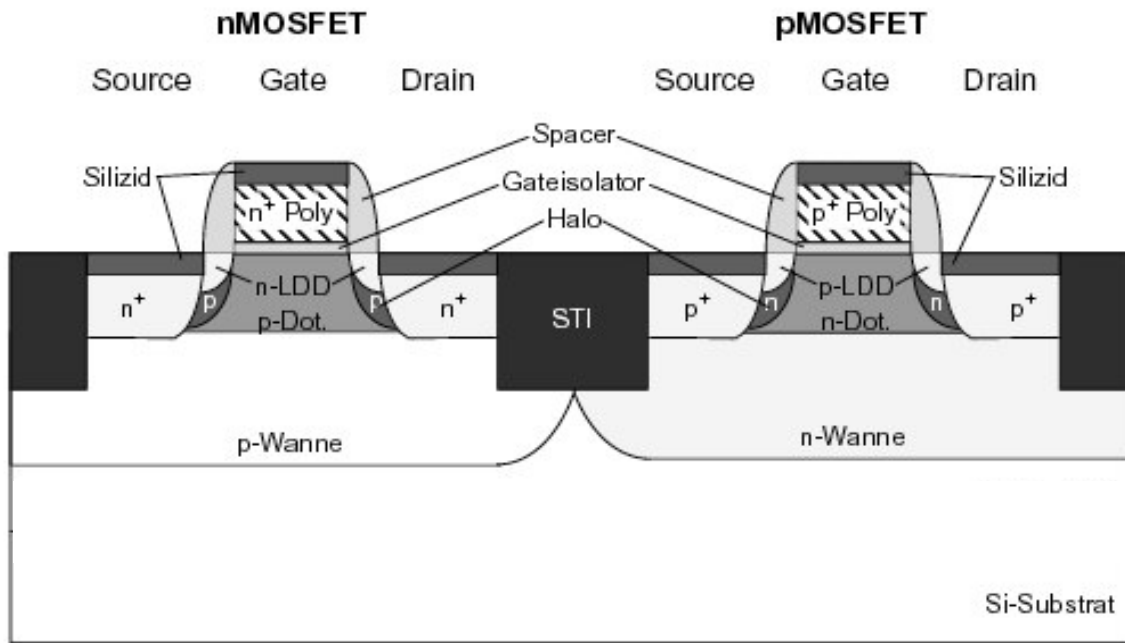


Abbildung 2.1: Prinzipieller Aufbau von pMOS- und nMOS-Transistor [20], [21]

Bei CMOS-Schaltungen müssen nMOS- und pMOS-Transistoren gleichzeitig auf einem Chip realisiert werden. Wie in Kapitel 2.4 dargestellt wird, werden diese Komponenten während des Herstellungsprozesses nacheinander gefertigt.

Wegen der weiter sinkenden Strukturgrößen und der dadurch möglicherweise auftretenden Kurzkanaleffekte kommt dabei, auch in Bezug auf die Fertigung und die Genauigkeit der Prozesse, der Kanalregion des Transistors, die in Abbildung 2.1 nur schematisch dargestellt ist, eine große Bedeutung zu. Üblicherweise wird die Einstellung der Schwellenspannung und die Unterdrückung des Punchthrough-Effekts durch getrennte Implantationen vorgenommen.

Die in Abbildung 2.1 gezeigte „Shallow-Trench“-Isolation (STI) zwischen den einzelnen Bauelementen, bei der die einzelnen Transistoren durch einen schmalen geätzten und mit Isolatormaterial aufgefüllten Graben voneinander getrennt werden, wird bei kleineren Strukturgrößen gegenüber der früher üblichen LOCOS-Isolation bevorzugt. Vorteilhaft wirkt sich dabei die geringere Breite der Isolation und deren bessere Planarität aus.

In Abbildung 2.1 nicht gezeigt ist die Verdrahtung der einzelnen Bauelemente untereinander. Zu diesem Zweck werden die Transistoren durch Metallleitungen, die auf mehreren Ebenen verlaufen können, miteinander verbunden. Die einzelnen Metallisierungsebenen werden durch IMD-Schichten (IMD = Inter-Metall-Dielektrikum) voneinander getrennt. Elektrische Verbindungen zwischen den Ebenen werden als Vias bezeichnet. Durch die Mehrlagenmetallisierung kann die Integrationsdichte erhöht werden und somit die Chip-

fläche klein gehalten werden, bzw. die Verdrahtbarkeit der Schaltung überhaupt erst gewährleistet werden. Während bei den sehr regelmäßig aufgebauten DRAMs die Zahl der Metallisierungsebenen bei 2 - 3 liegt, weisen moderne Logikschaltungen bis zu 6 oder 7 Metallisierungsebenen auf. Als Alternative zur herkömmlichen Aluminium-Metallisierung konnten dabei vor kurzem Kupfer-Leitungen technologisch realisiert werden. Vorteilhaft wirkt sich dabei der im Vergleich zu Aluminium um 50% höhere Leitwert aus, wodurch Schaltungen mit höherer Rechengeschwindigkeit bzw. geringerem Energieverbrauch realisiert werden können.

Ein Beispiel für den Aufbau einer Schaltung mit sieben Verdrahtungsebenen zeigt Abbildung 2.2 [4, S. 8-18].

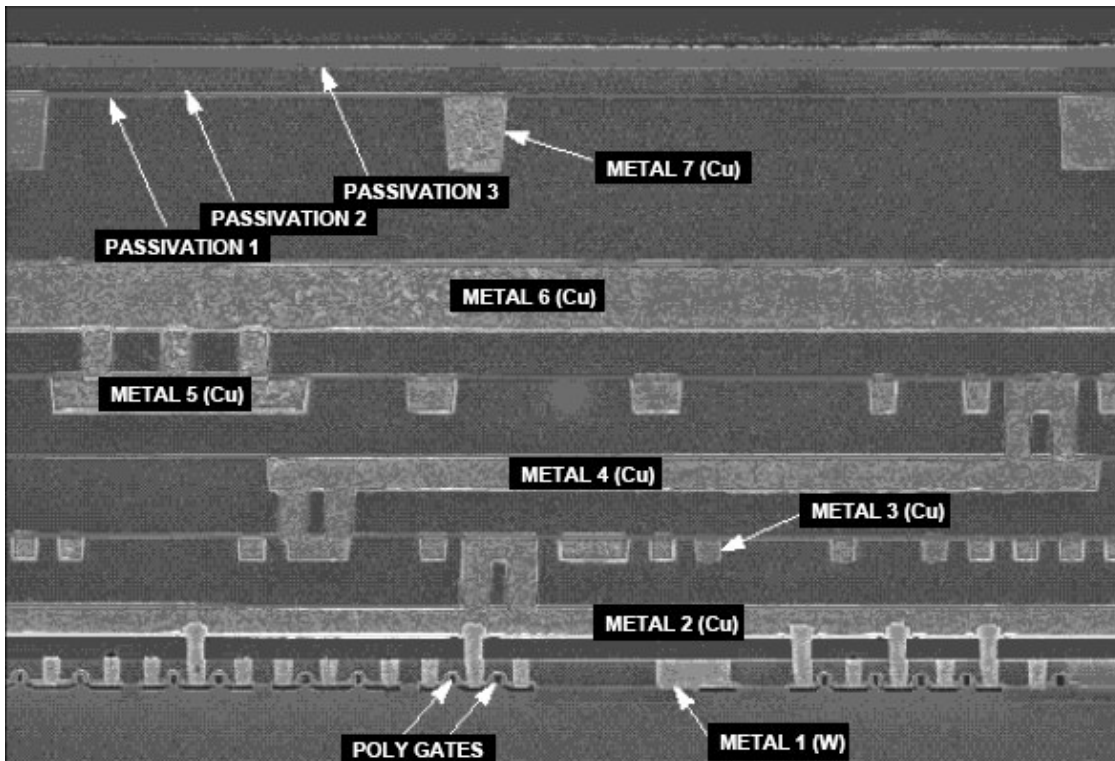


Abbildung 2.2: Mehrlagenverdrahtung [4]

Aufgrund der stetig steigenden Komplexität der Schaltungen wird die Bedeutung der Metallisierung und der Anteil der Prozessschritte für ihre Realisierung, im Verhältnis zur Gesamtzahl der Prozessschritte, in Zukunft weiter zunehmen.

2.2 Technologieskalierung

Ziel der IC-Entwicklung ist eine ständige Verbesserung der wichtigsten Bauelementeeigenschaften wie z.B. Integrationsdichte, Schaltgeschwindigkeit, Leistungsverbrauch und Kosten pro Funktion. Zum Erreichen dieser Ziele und zur Verhinderung negativer Auswirkungen der Strukturverkleinerung müssen Anpassungen der Bauelementeparameter vorgenommen werden. Dieses Vorgehen wird als Skalierung bezeichnet. Dabei werden üblicherweise die Abmessungen und Dotierungen des Bauelements sowie die auftretenden Spannungen angepasst. Die bekannten Skalierungsregeln sind in Tabelle 2.1 angegeben, ebenso die Auswirkungen auf die wichtigsten Bauelementeparameter (zusammengestellt aus [22], [23], [24]). Der Skalierungsfaktor α gibt an, wie sich der jeweilige Parameter verändert.

Als ideale Skalierung gilt dabei das sogenannte „constant field scaling“, bei dem die im Inneren des Transistors auftretenden Feldstärken konstant gehalten werden, womit mögliche, durch hohe Feldstärken auftretende, Zuverlässigkeitsprobleme vermieden werden. Aus Gründen der Kompatibilität mit anderen Schaltungsteilen und der längeren Verwendbarkeit der bestehenden Schaltungen wurde jedoch besonders in der Vergangenheit oft das sogenannte „constant voltage scaling“ angewendet, bei dem die Spannungen nicht verändert werden. Dieses Vorgehen führt dazu, dass zwar der resultierende Transistorstrom ansteigt und die Schaltgeschwindigkeit weiter verbessert wird, sich gleichzeitig aber auch die elektrischen Felder und der Leistungsverbrauch erhöhen.

Tabelle 2.1: Skalierungsregeln

Parameter	constant field scaling ($\alpha > 1$)	constant voltage scaling ($\alpha > 1$)
Veränderte Parameter		
Dimensionen ($W_{tr}, L_g, t_{gox}, X_j$)	$1/\alpha$	$1/\alpha$
Dotierungen (N_{Sub}, N_A, N_D)	α	α
Spannung (V_{DD}, V_{TN}, V_{TP})	$1/\alpha$	1
Beeinflusste Parameter		
Strom ($I_{DS} \propto \frac{W_{tr}}{L_g} \cdot \frac{\epsilon_{ox}}{t_{gox}} \cdot (V_{DD} - V_T)^2$)	$1/\alpha$	α
Transkonduktanz ($G_m = \frac{\Delta I}{\Delta V \cdot W_{tr}}$)	α	α^2
Gatekapazität ($C_g = \epsilon_{ox} \cdot \frac{W_{tr} \cdot L_g}{t_{gox}}$)	$1/\alpha$	$1/\alpha$
Transistor on-Widerstand ($R_{tr} \propto \frac{V_{DD}}{I_{DS}}$)	1	$1/\alpha$
Gate Delay ($\tau = \frac{C_g \cdot V}{I} = R_{tr} \cdot C_g$)	$1/\alpha$	$1/\alpha^2$
Leistungsverbrauch pro Gatter ($P_G = I \cdot V$)	$1/\alpha^2$	α
Power-Delay Produkt pro Gatter ($P_G \cdot \tau$)	$1/\alpha^3$	$1/\alpha$
Fläche ($A = W_{tr} \cdot L_g$)	$1/\alpha^2$	$1/\alpha^2$
Leistungsdichte ($P_d = \frac{P_G}{A}$)	1	α^3
Felder	1	α

Ziele für die künftige Entwicklung der charakteristischen Bauelemente- und Fabrikparameter für die nächsten Jahre werden in der „International Technology Roadmap for Semiconductors“ [12] angegeben. Diese Roadmap wird auf Initiative der Semiconductor Industry Association (SIA) gemeinsam von Halbleiterherstellern, Zulieferern, öffentlichen Organisationen und Universitäten erarbeitet und seit 1994 in regelmäßigen Abständen aktualisiert.

Die hier dargestellten Tabellen stammen von der letzten Überarbeitung im Jahr 2003. Bemerkenswert ist, dass sich die vorhergesagte Innovationsgeschwindigkeit in den letzten Ausgaben der SIA-Roadmap noch beschleunigt hat. Seit ihrem Erscheinen hat sich die Roadmap zu einer Art „selbsterfüllenden Prophezeiung“ entwickelt, da die Halbleiterfirmen versuchen, die Vorgaben zu erfüllen, bzw. sogar noch vor dem vorhergesagtem Zeitpunkt die entsprechende Technologie verfügbar zu haben.

Die erwarteten Entwicklungen der technologischen Prozesse und der erzeugten Transistorstrukturen, wie sie in Abbildung 2.1 und Abbildung 2.2 dargestellt werden, zeigt die folgende Tabelle 2.2 (Zusammenfassung der wichtigsten Parameter aus [12]). Grundlegende Änderungen des Transistoraufbaus oder der Herstellungsprozesse werden in den nächsten Jahren nicht erwartet.

Tabelle 2.2: Technologische Parameter

	2005 80 nm	2006 70 nm	2007 65 nm	2010 45 nm	2013 32 nm	2016 22 nm
Max. Waferdurchmesser (mm)	300	300	300	300	450	450
Min. Maskenzahl MPU	25	27	27	27	29	29
Min. Maskenzahl DRAM	24	24	24	26	26	26
Verdrahtungsebenen	9	9	9	10	10	10
Äquiv. Gateoxiddicke T_{ox} (nm)	1,1	1,0	0,9	0,7	0,6	0,5
Dicke Sidewall spacer (nm)	35,2	30,8	27,7	NA	NA	NA
Dicke Gateelektrode (nm)	32-64	30-60	25-50	18-36	13-26	9-18
Dicke Silizid (nm)	21	19	17	13	19	13
Struktur	Enhanced Planar Bulk CMOS			FDSOI, Elev. Cont.	FDSOI, Multi-Gate	
STI Tiefe (nm)						
Bulk-Si	367	359	353	331	314	198
SOI	26	22	20	14	10	7
LDD Tiefe X_j (nm)	17,6	15,4	13,8	7,2	10,4	7,2
LDD Schichtwiderstand PMOS (Ω/sq)	767	833	884	1875	514	549
LDD Schichtwiderstand NMOS (Ω/sq)	358	389	412	875	240	256
Unif. Kanaldotierung (cm^{-3}) für $V_T=0.4$	1,5- $2,5 \cdot 10^{18}$	2,0- $4,0 \cdot 10^{18}$	2,5- $5,0 \cdot 10^{18}$	NA	NA	NA

Wie von den Skalierungsregeln vorgesehen, werden die lateralen und vertikalen Abmessungen der Transistorstrukturen ständig kleiner, wogegen die Dotierungshöhen steigen. Anzumerken ist, dass ab dem Jahr 2010 der Übergang zu einer anderen Source/Drain-Struktur, der sogenannten „Elevated Source/Drain“ mit epitaktisch aufgebrachtene Source/Drain-Kontakten erwartet wird.

2.3 Bauelementeeigenschaften

In der folgenden Tabelle 2.3 (Zusammenfassung der wichtigsten Parameter aus [12]) werden die wichtigsten vorhergesagten zukünftigen Bauelementeeigenschaften, die sich durch Anpassung der technologischen Parameter ergeben, dargestellt. (Die Daten für „high performance“ bezeichnen Bausteine, die am oberen Ende der Leistungsfähigkeit angesiedelt sind, während mit „cost-performance“ preisgünstigere Bausteine mit niedrigerer Leistungsfähigkeit benannt werden.)

Wie erkennbar ist, wird bis zum Jahr 2016 ein weiteres uneingeschränktes Anwachsen der Leistungsfähigkeit vorhergesagt.

Tabelle 2.3: Charakteristische Bauelementeeigenschaften

Jahr	2005	2006	2007	2010	2013	2016
Strukturbreite (nm)	80	70	65	45	32	22
Zahl der Chip I/Os Chip-to-package (pads) ASIC	4000	4200	4400	4800	5400	6000
Zahl der Gehäusepins μ P/Controller ASIC	1760 2489	1936 2738	2140 3012	2782 4009	3616 5335	4702 7100
Min. V_{dd} (V)	0,9	0,9	0,7	0,6	0,5	0,4
Max. Leistungsverbrauch high-perf. mit Heat Sink (W) Batteriebetriebene Geräte (W)	170 3,2	180 3,5	190 3,5	218 3,0	251 3,0	288 3,0
Sub-threshold leakage $I_{sd leak}$ NMOS (25°C, μ A/ μ m)	0,05	0,05	0,07	0,1	0,3	0,5
$I_{d sat}$ NMOS (25°C, mA/mm)	1090	1170	1510	1900	2050	2400
On-chip lokaler Takt (high performance)	5173	5631	6739	11511	19348	28751
nom. Gate-Delay NAND Gate (ps)	21,72	18,92	16,23	9,88	6,52	3,74

Durch die verbesserten Bauelementeeigenschaften und die verbesserte Herstellungstechnologie steigt die Integrationsdichte der Schaltungen, d.h. es können mehr Transistoren pro Chip realisiert werden. Die Auswirkungen der technologischen Weiterentwicklungen auf die zukünftigen Integrationsdichten von DRAMs und Mikroprozessoren sind in Tabelle 2.4 dargestellt [12].

Tabelle 2.4: Integrationsdichte

	2005 80 nm	2006 70 nm	2007 65 nm	2010 45 nm	2013 32 nm	2016 22 nm
DRAM Bits/Chip bei Produkteinführung	8G	8G	16G	32G	64G	64G
DRAM Chipgröße bei Produkteinführung (mm ²)	454	359	568	563	373	186

	2005 80 nm	2006 70 nm	2007 65 nm	2010 45 nm	2013 32 nm	2016 22 nm
DRAM Bits/Chip bei Produktionsbeginn	2G	2G	4G	8G	32G	64G
DRAM Chipgröße bei Produktionsbeginn (mm ²)	147	116	183	181	239	238
Funktionsdichte (GBits/cm ²)	1,46	1,85	2,35	4,75	14,35	28,85
DRAM Zellfläche (µm ²)	0,039	0,031	0,024	0,012	0,004	0,002
MPU Transistoren/Chip (Mio.)						
high performance	697	878	1106	2212	4424	8848
cost-performance	487	614	773	1546	3092	6184
MPU Chipgröße (mm ²)	310	310	310	310	310	310
Funktionsdichte high performance (Mio. Transistoren/cm ²)	225	283	357	714	1427	2854
ASIC Chipgröße (mm ²)	572	572	572	572	572	572
ASIC Nutzbare Transistoren/cm ² (Auto Layout)	225	283	357	714	1427	2854

Wie in Tabelle 2.4 deutlich wird, steigt bei DRAMs die Zahl der gespeicherten Bits um den Faktor 4 pro Generation. Um diese Verbesserung zu erreichen, wird einerseits die Strukturgröße verringert (Faktor 2) und zusätzlich die Chipfläche im Vergleich zum Vorgängermodell um Faktor 1,5 erhöht. Der noch verbleibende Faktor 1,3 wird durch Modifizierung der Speicherzellen und der Chiparchitektur gewonnen.

Ebenfalls erwähnenswert ist die Steigerung der Integrationsdichte für DRAMs und Mikroprozessoren im Verlauf der Lebensdauer eines Produkts. Die laufende Verringerung der technologisch machbaren Strukturgröße wird dazu benutzt, die Chipgröße zu verringern, um dadurch mehr Chips pro Wafer fertigen zu können. Diese Vorgehensweise wird als Shrink bezeichnet.

2.4 Typischer Prozessfluss zur Herstellung von integrierten Schaltungen

Im folgenden soll ein Prozessfluss exemplarisch dargestellt werden, wie er typischerweise zur Herstellung von modernen integrierten Schaltungen bei Strukturgrößen im Bereich von 180 nm angewendet werden kann. Aufgrund des schematischen Charakters der Darstellung sind hier keine konkreten Prozessparameter, sondern nur die prinzipiellen Prozesse angegeben. Manche Prozessschritte, wie z.B. Lithographie oder Reinigung könnten dabei noch in mehrere Einzelschritte unterteilt werden.

Die Anzahl der Metallisierungsebenen des in Tabelle 2.5 dargestellten Prozessflusses kann variiert werden, Mess- oder Testvorgänge sind darin nicht enthalten (Prozessfluss zusammengestellt aus [22, S. 309 ff.], [25], [26], [27], [28], [29], [30], [31], [32]).

Tabelle 2.5: Exemplarischer Prozessfluss zur Herstellung von integrierten Schaltungen

Nr.	Bereich	Prozessschritt
1		Reinigung
2		Scribe
3	STI	Reinigung
4		Oxidation Pad-Oxid
5		Deposition Nitrid
6		Lithographie Aktives Gebiet
7		Ätzen Trench
8		Photolack entfernen
9		Reinigung
10		Oxidation Liner Oxid
11		Trench auffüllen mit Oxid
12		Therm. Oxid Dense
13		CMP Oxid
14		Reinigung
15		Ätzen Nitrid
16		Ätzen Pad-Oxid
17		Reinigung
18	N-Wanne	Oxidation Pad-Oxid
19		Reinigung
20		Lithographie N-Wanne
21		Implantation N-Wanne
22		Photolack entfernen
23		Reinigung
24	P-Wanne	Lithographie P-Wanne
25		Implantation P-Wanne
26		Photolack entfernen
27		Reinigung
28		Ätzen Pad-Oxid
29	Kanal	Oxidation Sac-Oxid
30		Schwellspannungsimplantation P-Vt
31		Lithographie Kanal
32		Schwellspannungsimplantation N-Vt
33		Photolack entfernen
34		Reinigung
35	Gate	Ätzen Sac-Oxid
36		Reinigung
37		Oxidation Gateoxid
38		Deposition Poly-Silizium

Nr.	Bereich	Prozessschritt
39		RTA Anneal
40		Lithographie Poly
41		Ätzen Poly-Silizium
42		Photolack entfernen
43		Reinigung
44		Oxidation Seitenwand-Oxid
45		Reinigung
46		Implantation Amorphisierung - Si
47	N-LDD	Lithographie N-LDD
48		Implantation N-LDD
49		Implantation Halo - P-Dotierung
50		Photolack entfernen
51		Reinigung
52	P-LDD	Lithographie P-LDD
53		Implantation P-LDD
54		Implantation Halo N-Dotierung
55		Photolack entfernen
56		Reinigung
57	Spacer	LPCVD TEOS Deposition Spacer
58		Ätzen Spacer
59		Reinigung
60	N-S/D	Lithographie N-S/D
61		Implantation N-S/D
62		Photolack entfernen
63		Reinigung
64	P-S/D	Lithographie P-S/D
65		Implantation P-S/D
66		Photolack entfernen
67		Reinigung
68		RTA S/D Anneal
69		Belackern Frontseite
70		Ätzen Rückseite
71		Photolack entfernen
72		Reinigung
73	Kontakt	PVD Deposition Cobalt
74		RTA Silizidierung
75		Ätzen Cobalt
76		CVD Deposition SiN
77		CVD Deposition IMD
78		Lithographie Via

Nr.	Bereich	Prozessschritt
79		Ätzen IMD
80		Photolack entfernen
81		Reinigung
82		Lithographie Metall
83		Ätzen IMD
84		Photolack entfernen
85		Reinigung
86		Ätzen SiN
87	Metall 1	PVD Deposition TaN
88		CVD Deposition Cu
89		CMP Cu
90		Reinigung
91	Via M1-M2	CVD Deposition SiN
92	Metall2	CVD Deposition IMD
93		Lithographie Via
94		Ätzen IMD
95		Photolack entfernen
96		Reinigung
97		Lithographie Metall
98		Ätzen IMD
99		Photolack entfernen
100		Reinigung
101		Ätzen SiN
102		PVD Deposition TaN
103		CVD Deposition Cu
104		CMP Cu
105		Reinigung
	weitere Metallisierungs ebenen :	
	Passivierung	Deposition Oxid
		Deposition Nitrid
		Lithographie Pads
		Ätzen Nitrid, Oxid
		Photolack entfernen

3 Fertigung von integrierten Schaltungen

3.1 Aufbau einer Halbleiterfabrik

3.1.1 Reinraum

Durch die ständig sinkenden Strukturgrößen bei der Chipfertigung ergeben sich immer höhere Anforderungen an die Herstellungsprozesse und an die Fertigungsumgebung. Der größte Teil der ausbeutemindernden Defekte wird durch Partikel unterschiedlicher Herkunft hervorgerufen. Die Prozessierung der Wafer muss deswegen unter Verwendung von Prozessmaterialien höchster Qualität und Sauberkeit in Reinräumen stattfinden, in denen definierte und konstante Umgebungsbedingungen z.B. bezüglich der Temperatur und Luftfeuchtigkeit herrschen und die Zahl der Partikel extrem reduziert ist. Die Qualität des Reinraums richtet sich nach der Zahl der Partikel, die in einem festgelegten Luftvolumen enthalten ist und wird in Klassen eingeteilt. In Tabelle 3.1 sind die Reinheitsanforderungen der verschiedenen Klassen dargestellt. Dabei muss beachtet werden, dass ein internationaler (ISO) und ein amerikanischer Standard (Fed. Std. 209E) existieren, die sich in ihren Festlegungen leicht unterscheiden [33].

Tabelle 3.1: Reinraumklassen

Klasse		Max. Konzentration pro m ³ für Partikel der angegebenen Größe					
Fed. Std. 209E	ISO	0,1 µm	0,2 µm	0,3 µm	0,5 µm	1 µm	5 µm
	1	10	2				
	2	100	24	10	4		
	3	1000	237	102	35	8	
1		1240	265	106	35,3		
	4	10000	2370	1020	352	83	
10		12400	2650	1060	353		
	5	100000	23700	10200	3520	832	29
100			26500	10600	3530		
	6	1000000	237000	102000	35200	8320	293
1000					35300		247
	7				352000	83200	2930
10000					353000		2470
	8				3520000	832000	29300
100000					3530000		24700
	9				35200000	8320000	293000

Der schematische Aufbau eines solchen Reinraums ist in Abbildung 3.1 gezeigt [34]. Die gereinigte Luft wird von oben in den Raum eingeblasen und durch den Fußboden wieder abgesaugt, wobei Wert auf einen möglichst ungestörten, geradlinigen Luftfluss gelegt wird

(laminar flow), durch den die Partikel zu Boden gedrückt und entfernt werden sollen.

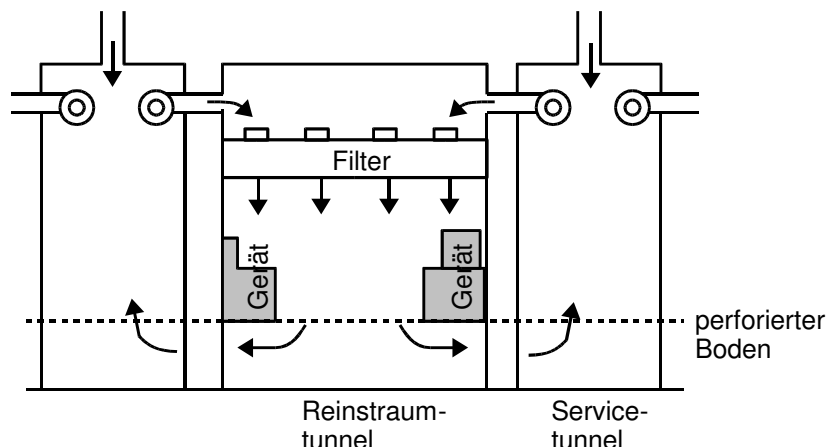


Abbildung 3.1: Schematischer Aufbau eines Reinraums [34]

Eine Halbleiterfabrik umfasst mehrere Stockwerke. Über der eigentlichen Produktionsebene muss durch eine geeignete Klimatechnik für eine Filterung der dem Reinraum zugeführten Luft gesorgt werden. Unter der Produktionsebene wird üblicherweise ein Stockwerk für die Versorgung mit Prozessmaterialien und -gasen, sowie Reinstwasser vorgesehen.

Die benötigte Reinraumfläche für eine Fabrik, die im Monat ca. 20000 Wafer bearbeitet, schwankt je nach angewendetem Reinraumkonzept. Ein Richtwert für die Größe beträgt ungefähr 5000 - 6000 m² [35]. Die dabei notwendige Reinraumklasse richtet sich ebenfalls nach dem jeweiligen Konzept. Bei Verwendung von SMIF-Behältern werden niedrigere Anforderungen an die Sauberkeit des umgebenden Reinraums gestellt als bei offenem Transport der Wafer. Übliche Werte für die Reinraumklasse bei modernen Fertigungen liegen deshalb im Bereich von 0,1 - 100 (US-Standard).

3.1.2 Anordnung der Geräte in der Fabrik

Die meisten 200 mm Fabriken für die IC-Fertigung sind nach dem Bay-and-Chase-Prinzip aufgebaut, wobei die Geräte in verschiedenen Bereichen der Fabrik (Bays) angeordnet sind [36, S. 119 ff.].

Der „Chase“-Bereich wird üblicherweise für Wartungs- oder Reparaturarbeiten an den Geräten verwendet. Um den Bedarf an teurer Reinraumfläche so gering wie möglich zu halten, sind nahezu alle Geräte so aufgebaut, dass nur auf ihrer Bedien- und Beladeseite absolut reine Bedingungen herrschen müssen. Die Rückseiten befinden sich im Installations- und Wartungsbereich und sind von dort aus zugänglich. Da sich in diesem Bereich keine Produktwafer befinden und dort nicht produziert wird, werden dort auch etwas geringere Anforderungen an die Reinraumklasse als in den Bays gestellt.

Am weitesten verbreitet ist die Anordnung nach dem sogenannten Farm-Layout. Dabei werden die Geräte getrennt nach ihrer Funktionalität in verschiedenen Bays der Fabrik aufgestellt. Das heißt, dass z.B. die Geräte für Lithographie, Implantation oder Diffusion räumlich voneinander getrennt in separaten Gängen angeordnet werden. Diese einzelnen Bays werden durch Korridore miteinander verbunden, auf denen der Zugang zu den Ge-

räten möglich ist und der Wafertransport durchgeführt wird. In Abbildung 3.2 ist prinzipiell ein solches Fabrik-Layout dargestellt [37], [38], [39].

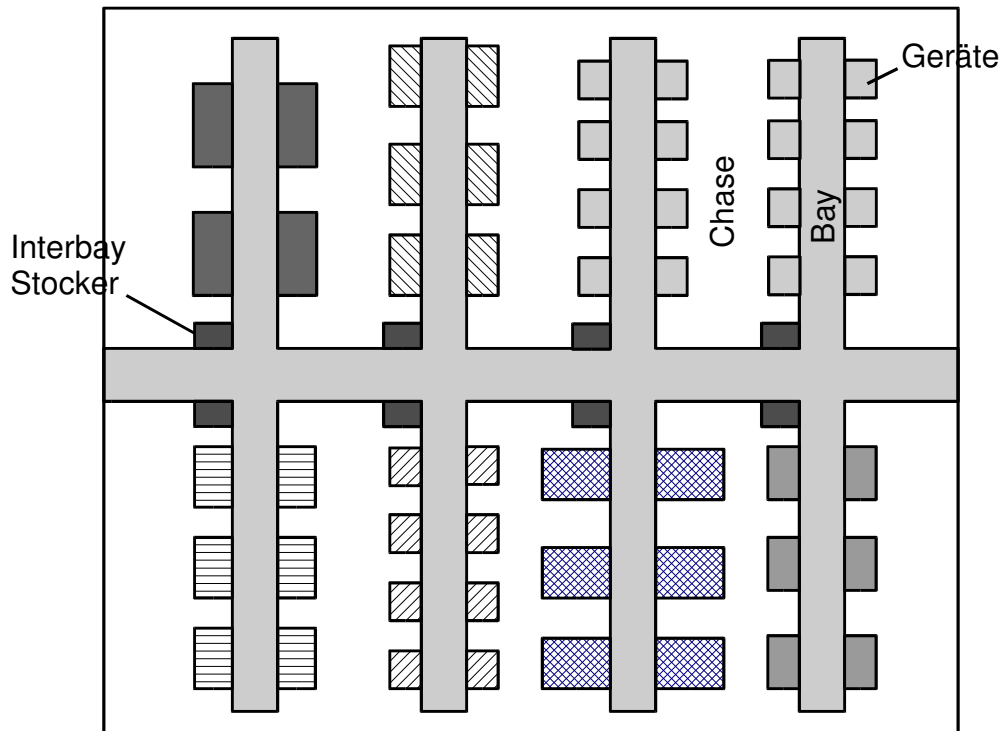


Abbildung 3.2: Farm-Layout einer Halbleiterfabrik [37]

Im allgemeinen wird eine solche funktionale Aufteilung der Geräte auch als Werkstattfertigung bezeichnet [40, S. 432 f.].

„Interbay Stockers“ dienen zur Aufbewahrung von Wafern, die dort auf ihre Bearbeitung oder auf ihren Weitertransport warten.

Das Farm-Layout führt zu einem relativ einfachen und kostengünstigen Design von Transport- und Versorgungssystemen, bringt aber Probleme mit sich, die die Reihenfolge des Prozessflusses und die Logistik betreffen, da die Wafer für den nächsten Prozessschritt jeweils in eine andere Bay transportiert werden müssen, was einen relativ großen Transportaufwand und längere Transportwege mit sich bringt.

Einen ersten Schritt diese logistischen Probleme zu überwinden, stellt das sogenannte „Cluster Layout“ dar, das auch als „Hybrid Layout“ bezeichnet wird und eine Gruppenfertigung darstellt. Dabei werden verschiedene Gerätetypen entsprechend der Reihenfolge im Produktionsablauf in unmittelbarer Nähe zueinander aufgestellt, um so den Transportaufwand zwischen den einzelnen Geräten zu verringern. Das bedeutet, dass z.B. Geräte zur Entfernung des Photolacks in die Bays für Implantation und Ätzen integriert werden, Implantationsgeräte gemeinsam mit RTP-Tools aufgestellt werden, oder Messgeräte über mehrere Bays verteilt werden [38]. Eine solche Verteilung der Geräte erweist sich jedoch als äußerst unflexibel gegenüber Prozessveränderungen. Fertigungsumstellungen wie z.B. die Einführung einer zusätzlichen Litho-Ebene sind hier nur sehr schwer umzusetzen. Abbildung 3.3 zeigt ein solches Konzept.

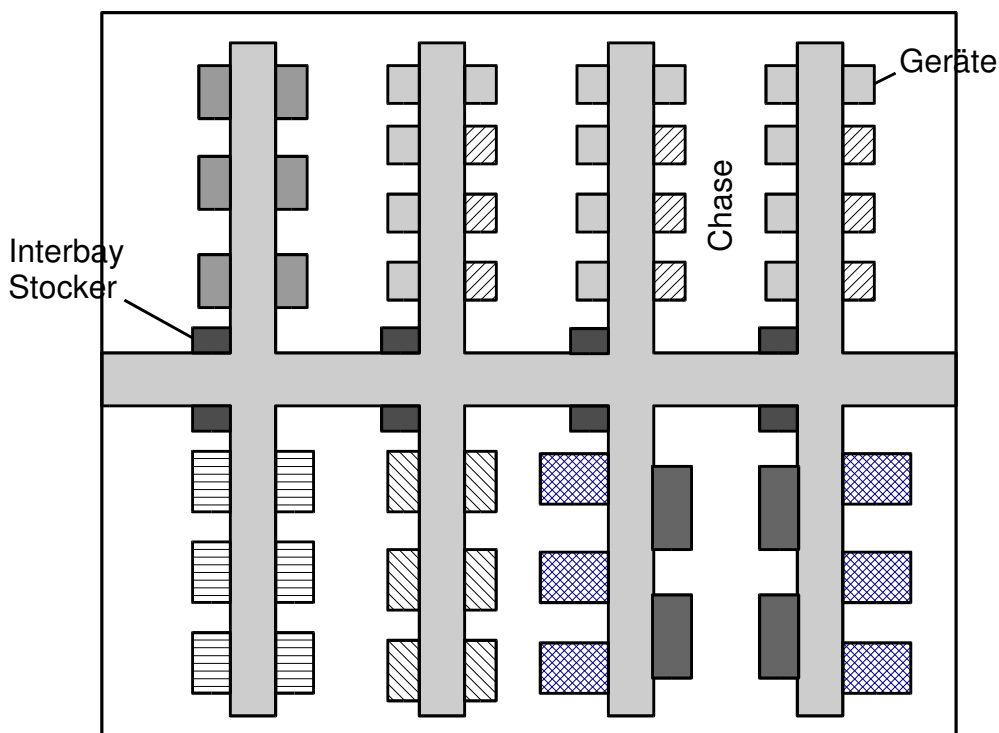


Abbildung 3.3: Cluster-Layout einer Halbleiterfabrik [38]

Abbildung 3.2 und Abbildung 3.3 stellen den prinzipiellen Aufbau der Konzepte dar. Leichte Abweichungen davon sind möglich. Tabelle 3.2 zeigt im Überblick die wichtigsten Eigenschaften und die Vor- und Nachteile der jeweiligen Konzepte (zusammengestellt aus [37], [38], [41], [42], [43]).

Tabelle 3.2: Vergleich von Fab-Layouts

	Farm-Layout	Cluster-Layout
Erweiterbarkeit	+	--
Flexibilität gegenüber Produktmix, Prozessveränderungen	+	--
Transportaufwand / Transportzeiten	-	+
Einfachheit des Versorgungssystems	+	-
Klarheit / Kontrollierbarkeit des Materialflusses	-	+
Planungsaufwand	+	-
Durchlaufzeit bei festgelegtem Prozessfluss	-	+

Neben den beiden oben dargestellten Fab-Layouts besteht auch noch die Möglichkeit, alle Geräte in einem einzigen, nicht unterteilten Reinraum aufzustellen, was als „Ballroom“-

Konzept bezeichnet wird. Der prinzipielle Aufbau eines solchen Reinraums ist in Abbildung 3.4 angegeben [42], [44].

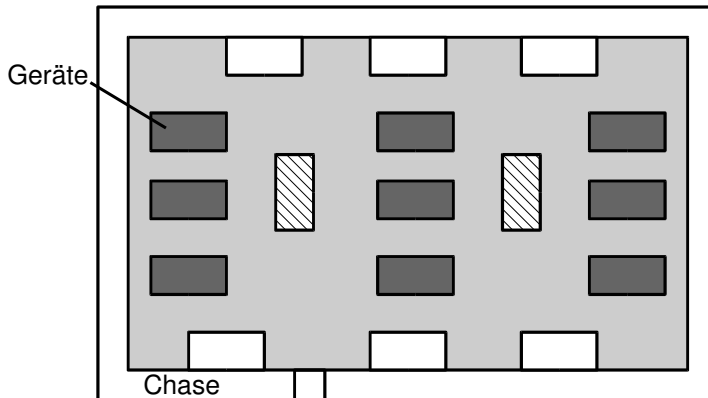


Abbildung 3.4: Ballroom-Layout [42], [44]

Ballrooms weisen einen einfachen Aufbau auf und sind flexibler bezüglich der Aufstellung der Geräte und der Automatisierung. Die Kontrolle und Verringerung der Partikelbelastung stellt in dem großen Raum jedoch eine größere Herausforderung dar. Zudem erweisen sich Geräte mit einem hohen Wartungsaufwand bei einem solchen Konzept als problematischer.

3.2 Produktionsfluss in einer Halbleiterfabrik

3.2.1 Transportbehälter

Wie in Abbildung 3.2 und Abbildung 3.3 deutlich wird, müssen die Wafer im Reinraum einerseits gelagert und andererseits innerhalb der einzelnen Bays (Intrabay Transport) und zwischen den Bays (Interbay Transport) transportiert werden. Zu diesem Zweck werden die Wafer in speziellen Transportgehäusen untergebracht, um eine Beschädigung oder Verunreinigung während der Transport- oder Wartezeiten zu verhindern. Am günstigsten für eine Fertigung mit sehr hohen Reinheitsanforderungen haben sich die sogenannten SMIF- (Standard Mechanical Interface) und FOUP- (Front Opening Unified Pod) Gehäuse erwiesen. Dabei handelt es sich um standardisierte Behälter, die im allgemeinen 25 Wafer enthalten. Die FOUP-Gehäuse, die in Abbildung 3.5 dargestellt werden [45], können 13 oder 25 Wafer enthalten und wurden speziell für die Anforderungen bei der zukünftigen Verwendung von 300 mm Wafern entwickelt, während SMIF-Gehäuse bei 200 mm Wafern Einsatz finden.

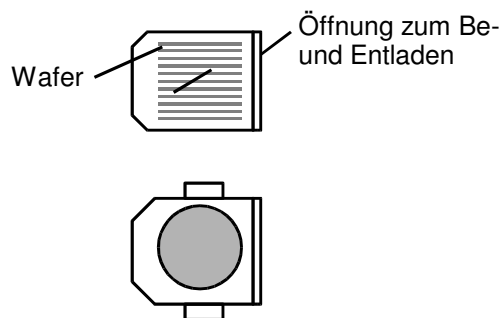


Abbildung 3.5: FOUP-Behälter [45]

Zusätzlich zum mechanischen Schutz der Wafer vor Beschädigungen bieten die abgeschlossenen SMIF- und FOUP-Behälter den Vorteil, dass in ihrem Inneren die Partikeldichte sehr gering gehalten werden kann, so dass dort bei richtiger Anwendung noch günstigere Reinraumbedingungen herrschen als außerhalb, was im Endeffekt zu etwas höheren Produktionsausbeuten führen kann [46]. Wie bereits erwähnt, lässt sich durch die Verwendung dieser Transportbehälter auch eine Verringerung der Anforderungen an den umgebenden Reinraum erreichen.

Durch die Standardisierung der Abmessungen und Funktionen bei den SMIF- und FOUP-Behältern wird das automatische Andocken an ein Gerät und das Be- und Entladen der Behälter ermöglicht. Außerdem wird dadurch der Einsatz von automatisierten Transportsystemen erleichtert, die auf die entsprechenden Abmessungen und Gewichte abgestimmt werden können.

Als übliche Losgröße für die Aufbewahrung und den Transport der Wafer haben sich 25 Wafer herausgestellt. Für 300 mm Wafer waren daneben auch noch 13 Wafer pro Los in der Diskussion. Mit dem INCAM-Behälter wurde auch bereits ein Gehäuse zur Aufnahme von einem einzelnen Wafer vorgeschlagen [47]. Bis jetzt ist aber der Einzelwafertransport innerhalb der Fab nicht üblich und bei Massenproduktion nicht machbar.

3.2.2 Transportsysteme

Für den Interbay-Transport werden bei fast allen der heutigen 200 mm Fabs AMHS Systeme (Automated Material Handling System) für den Wafertransport benutzt. Am weitesten verbreitet ist dabei der Einsatz von OHTs (Overhead Track), das sind an der Decke geführte Schienenbahnen. Im Bereich des Interbay-Transports werden auch beim Übergang zu 300 mm Wafern keine Änderungen erwartet.

Stärkere Veränderungen durch den Umstieg auf größere Waferdurchmesser werden jedoch für den Transport innerhalb der Bays erwartet. Bis jetzt werden diese Vorgänge im allgemeinen manuell vom Bedienpersonal durchgeführt. Ziel ist es jedoch, in allen Bereichen des Wafertransports eine vollständige Automatisierung zu erreichen [48]. Da teilprozessierte Wafer unter Umständen einen sehr hohen Wert darstellen, muss besonderer Wert auf eine sichere Handhabung gelegt werden. Wegen der Abmessungen und der Masse eines FOUP-Behälters mit 25 300 mm Wafern (Gesamtmasse ungefähr 8 kg), ist ein manueller Transport nicht sinnvoll und nur noch bedingt möglich. Einen weiteren Vorteil einer vollständigen Automatisierung stellt die Möglichkeit zur Einbindung in eine umfassende computergestützte Ablaufsteuerung dar, mit der die gesamten Prozesse erfasst und kontrolliert werden können.

Systeme, die für eine Intrabay-Automatisierung in Frage kommen, sind RGVs (Rail Guided Vehicles), AGVs (Automated Guided Vehicles) oder OHTs. RGVs sind spurgeführte

Fahrzeuge, die sich bei ihrer Bewegung an einer im Boden vorgegebenen Spur orientieren. Im Gegensatz dazu können sich AGVs direkt auf dem Reinraumboden bewegen und folgen dabei beispielsweise einer Induktionsschleife [49], [50].

Die verschiedenen Transportsysteme und ihre Anbindung an die Prozessgeräte sind in Abbildung 3.6 schematisch dargestellt [51]. Zur automatischen Be- und Entladung der Transportbehälter an den Load Ports der Geräte muss eine weitere Standardisierung vorgenommen werden [52].

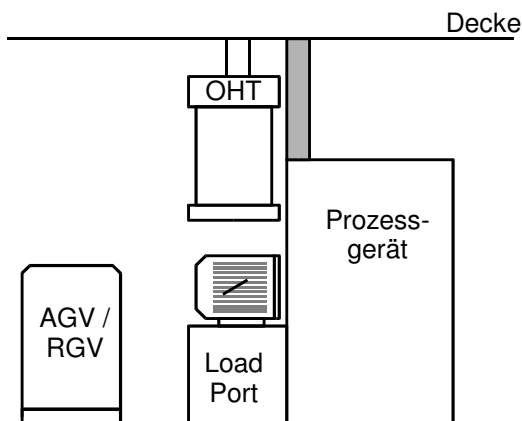


Abbildung 3.6: Verschiedene Transportsysteme [51]

Ein Vergleich der zukünftig denkbaren Intrabay-Transportsysteme wird in Tabelle 3.3 angestellt [53]. Man sieht, dass die Transportgeschwindigkeiten im Bereich von 30 - 50 m/min. liegen, wobei RGVs und AGVs gleichzeitig vier Behälter transportieren können. Die angegebenen Kosten beziehen sich auf ein Intrabay-System inklusive einem Transportfahrzeug.

Tabelle 3.3: Vergleich von Transportsystemen

	RGVs	AGVs	Overhead
Transportgeschwindigkeit [m/min]	50	30	40
Zahl der Lasten	4	4	1
Anzahl/Bay	2	2	10
Be-/Entladezeit [s]	30	90	30
Kosten [\$]	300000	250000	200000

Alle diese Systeme befinden sich noch in einem früheren Stadium ihrer Entwicklung.

Solange die vollständige Intrabay Automatisierung noch nicht möglich ist, wird der Einsatz von PGVs (Personal Guided Vehicles) erwartet, die von einem Operator bewegt werden. Im allgemeinen werden darunter kleine, reinraumtaugliche Wägen verstanden, die ein bis zwei Lose aufnehmen können und zwischen den Geräten hin- und hergeschoben und manuell be- und entladen werden [54].

3.3 Zukünftige Herausforderungen für die Fertigung

An den Aufbau und die Optimierung der Produktionslinie werden in Zukunft höhere Anforderungen gestellt werden. Als zukünftige „Schwierige Herausforderungen“ im Bereich der Fertigung von integrierten Schaltungen werden in der SIA-Roadmap folgende Punkte erwähnt [12]:

- steigende Investitions- und Betriebskosten
- steigende Produktivität der Fertigung gefordert
- sinkende Zeitspanne, um Investitionen zu amortisieren
- steigende Prozess- / Fabrikkomplexität

Allgemein ist erkennbar, dass zukünftig eine höhere Geschwindigkeit für die Optimierung der Parameter, ausgehend von einem höheren Anfangsniveau, bei einem höheren erreichten Endniveau angepeilt wird. So sollen beispielsweise die Werte für die Anfangs- und Endausbeute von momentan ungefähr 50% bzw. 85 - 90% im Lauf der Zeit weiter gesteigert werden. Ebenfalls wird eine verbesserte Gerätenutzung angestrebt, wobei typische OEE-Werte für Fertigungsgeräte heute im Bereich von 50% liegen. Auch bei der Zeitdauer für Planung und Errichtung einer Halbleiterfabrik und beim Hochfahren der Produktion ist eine weitere Beschleunigung geplant.

Zusätzlich zur Verbesserung der Fertigungsleistung muss es gleichzeitig gelingen, eine Verringerung der Kosten pro Funktion zu erreichen, die für alle Produkte der Mikroelektronik typisch ist. Die Kosten/Funktion müssen im Verlauf der Produktgenerationen und auch während der Lebensdauer eines Produkts sinken.

Wie an diesen Punkten erkennbar ist, hat sich die Halbleiterindustrie für die nächsten Jahre hohe Ziele gesteckt. Ziel der Unternehmen muss es sein, diese Vorgaben zu erreichen und trotz aller Herausforderungen auch in Zukunft rentabel zu sein, um die bisherige Innovationsgeschwindigkeit und Wachstumsrate beibehalten zu können. Bedeutende Anstrengungen müssen also unternommen werden, um eine wettbewerbsfähige Fertigung zu erreichen und zu erhalten. Welche Anforderungen an die zukünftige Chipfertigung gestellt werden und durch welche Maßnahmen diese erfüllt werden können, wird in den folgenden Kapiteln berechnet.

4 Wirtschaftliche Anforderungen an die zukünftige Höchstintegration von Schaltungen

4.1 Die Rentabilität als unternehmerisches Oberziel

Gewinn

In jedem Betrieb werden gewisse Zielvorstellungen vorgegeben, die durch die unternehmerischen Aktivitäten erreicht werden sollen. Dabei kann zwischen übergeordneten, speziellen Zielen, wie z.B. dem Streben nach Gewinn bei erwerbsorientierten Unternehmen oder nach Bedarfsdeckung bei Non-Profit-Organisationen, und allgemein gültigen Zielen, wie z.B. dem Streben nach Produktivität oder Wirtschaftlichkeit, die zur Erreichung der übergeordneten Ziele beitragen und wegen ihrer grundsätzlichen Bedeutung auch als Prinzipien bezeichnet werden, unterschieden werden [55].

Ein übergeordnetes Ziel der profitorientierten Unternehmen ist es, mit den erzeugten Produkten und Leistungen Gewinn zu erwirtschaften.

Der Erfolg eines erwerbswirtschaftlichen produzierenden Unternehmens kann durch die Differenz zwischen erzielten Erlösen und aufgewendeten Kosten ausgedrückt werden. Während der Verkauf von Gütern durch den dadurch erzielten Umsatz (= Erlös) zum Erfolg des Unternehmens beiträgt, ergibt sich durch den Aufwand für die Herstellung der Produkte (= Kosten) eine Erfolgsminderung. Für die Unternehmen bedeutet dies, dass die Differenz zwischen Erlösen und Kosten möglichst günstig gestaltet werden muss. Im Fall einer positiven Differenz spricht man von Gewinn, im Fall einer negativen Differenz von Verlust.

Üblicherweise werden solche Berechnungen auf eine gewisse Zeitdauer bezogen. Dies wird durch die Bezeichnungen Periodengewinn bzw. Periodenverlust ausgedrückt.

Allgemein kann der absolute Gewinn ausgedrückt werden durch die Gleichung:

$$\begin{array}{r} \text{Gesamterlöse} \\ - \text{Gesamtkosten} \\ \hline = \text{Gewinn} \end{array}$$

Abbildung 4.1: Gewinn

Diese Gleichung stellt eine übergeordnete Formel dar, wobei alle Leistungen des Unternehmens im Begriff Erlös und alle Aufwendungen unter Kosten zusammengefasst sind. Darin spiegeln sich alle beeinflussbaren und unbeeinflussbaren Größen wider, also auch die Rahmenbedingungen des Marktes, wie z.B. der Preisverfall, der sich über den Produktpreis direkt auf den Erlös auswirkt.

In erweiterter Form wird die Berechnung des Unternehmenserfolgs in der Gewinn- und Verlustrechnung dargestellt. Schematisch berechnet sich der Jahresüberschuss bzw. Jahresfehlbetrag wie folgt [56, S. 458], wobei auf einige der speziellen betriebswirtschaftlichen Größen im folgenden nicht mehr im Detail eingegangen wird.

	Umsatzerlöse
±	Bestandsveränderungen an Halb- und Fertigerzeugnissen
+	<u>aktivierte Eigenleistungen</u>
=	Gesamtleistung
-	<u>Materialaufwand</u>
=	Rohertrag
+	<u>sonstige Erträge</u>
=	erweiterter Rohertrag
-	Personalaufwendungen
-	Abschreibungen
-	Zinsaufwendungen
-	Steuern
-	<u>sonstige Aufwendungen</u>
=	Jahresüberschuß / Jahresfehlbetrag
±	Gewinnvortrag / Verlustvortrag
±	<u>Auflösung / Dotierung offener Rücklagen</u>
=	Bilanzgewinn / Bilanzverlust

Abbildung 4.2: Detaillierte Gewinnermittlung [56]

In der Darstellung wird klar, dass der Erfolg eines Unternehmens aus verschiedenen Bestandteilen resultiert. Bei einem produzierenden Betrieb sind zur konkreten Ermittlung des Unternehmensgewinns, neben den Erlösen und Aufwendungen, die direkt im Zuge der Produktion entstehen, auch zusätzliche Komponenten, wie z.B. Finanzerträge und -aufwendungen aus anderen Tätigkeiten (beispielsweise aus Beteiligungen) zu berücksichtigen.

Hier soll im folgenden nur auf den Beitrag eingegangen werden, der durch die Produktion geleistet werden kann.

Der Umsatz, der mit den produzierten und verkauften Produkten erzielt wird, berechnet sich aus der Menge der verkauften Güter und dem dafür erzielbaren Preis über die Formel (4.1) [57, S. 633].

$$U = \sum_i^n P_i \cdot M_i \quad (4.1)$$

U = Umsatz

P = Preis

M = Menge

i = Index der Produkte

n = Zahl der Produkte

Wie in Abbildung 4.2 dargestellt, setzen sich auch die Kosten aus mehreren Faktoren zusammen. Aufwendungen, die den Unternehmenserfolg verringern, stellen beispielsweise Material- und Personalkosten und Abschreibungen dar. In Bezug auf die Fertigung ist eine Gliederung dieser Aufwendungen in fixe und variable Kosten sinnvoll, wobei die fixen Kosten unabhängig von der Produktionsmenge sind, die variablen Kosten jedoch in Abhängigkeit von der produzierten Menge ansteigen [58, S. 411].

$$K = K_F + \sum_i K_{Vi} \cdot M_i \quad (4.2)$$

K = Kosten
K_F = fixe Kosten
K_V = variable Kosten

Eine genauere Erläuterung der Kosten für die Fertigung integrierter Schaltungen wird in Kapitel 4.4 vorgenommen.

Rentabilität

Während der Gewinn eine absolute Größe darstellt, handelt es sich bei der Rentabilität um eine relative Maßzahl, bei der der absolute Gewinn durch eine Bezugsgröße dividiert wird.

$$\text{Rentabilität} = \frac{\text{Gewinn}}{\text{Bezugsgröße}} \quad (4.3)$$

In diesem Zusammenhang sind, je nach Bezugsgröße, verschiedene Rentabilitätskennzahlen definierbar und in Benutzung [59, S. 670 ff.]. Beispielsweise ergibt sich der sogenannte Return on Investment (ROI) zu

$$\text{ROI} = \frac{\text{Jahresüberschuß vor Steuern}}{\text{Gesamtkapital}} \quad (4.4)$$

und die Umsatzrentabilität zu

$$\text{Netto-Umsatzrentabilität} = \frac{\text{Jahresüberschuß vor Steuern}}{\text{Umsatz}} \quad (4.5).$$

Mit Gleichung (4.4) und (4.5) ergibt sich folgender Zusammenhang zwischen dem Return on Investment und der Umsatzrentabilität:

$$\text{ROI} = \frac{\text{Jahresüberschuß}}{\text{Umsatz}} \cdot \frac{\text{Umsatz}}{\text{Gesamtkapital}} = \text{Umsatzrentabilität} \cdot \text{Kapitalumschlag} \quad (4.6)$$

Die Rentabilität ist also eine Kennzahl dafür, wie sich der Erfolg des Unternehmens im Vergleich zur eingesetzten Bezugsgröße verhält. Dadurch wird die Einordnung und Vergleichbarkeit der Werte vereinfacht. Beispielsweise kann sich bei zwei unterschiedlich großen Unternehmen die gleiche Umsatzrentabilität ergeben, obwohl bei beiden Firmen die Werte für den absoluten Gewinn stark unterschiedlich sein können.

Eine prinzipielle Zusammenstellung der oben eingeführten Begriffe wird noch einmal in Abbildung 4.3 vorgenommen.

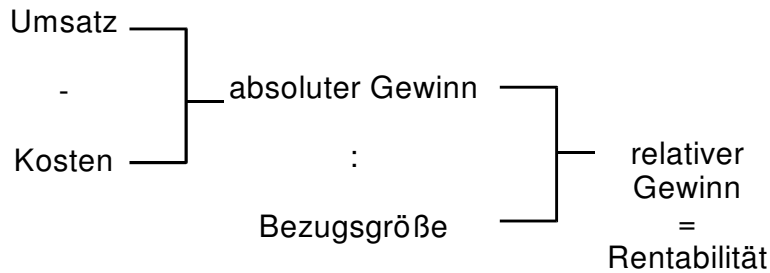


Abbildung 4.3: Bewertungsgrößen für Unternehmenserfolg

Ein Unternehmen muss zumindest mittel- oder langfristig rentabel arbeiten, da ansonsten bei ständigen Verlusten das vorhandene Eigenkapital der Firma aufgezehrt wird, bzw. keine Kapitalgeber mehr gefunden werden, was im Endeffekt zum Verlust der Liquidität (= Zahlungsfähigkeit) und damit zum Konkurs des Unternehmens führen kann.

Ziel der profitorientierten Unternehmen ist es, für die Investitionen, also für das eingesetzte Kapital, eine Rendite zu erwirtschaften. Ähnlich wie bei einer Geldanlage bei einer Bank, sollte für die aufgewendete Summe eine entsprechende Verzinsung erreicht werden. Diesem Ziel entspricht die Forderung nach einer angemessenen Rentabilität des Unternehmens. Besonders in der Halbleiterindustrie stellt dieser Punkt, wegen der hohen Investitionen für den Aufbau und Betrieb einer Fabrik, eine nicht einfach zu bewältigende Herausforderung dar.

Cash Flow

Zur Analyse der Finanzkraft eines Unternehmens wird häufig auch auf die Kenngröße „Cash Flow“ zurückgegriffen. Diese Größe drückt aus, in welcher Höhe dem Unternehmen während einer Periode finanzielle Mittel zugeflossen sind [60, S. 20 ff.]. Im allgemeinen ist der Cash Flow definiert durch:

$$\begin{array}{r} \text{Bilanzgewinn} \\ + \text{Zuführung zu den Rücklagen} \\ - \text{Entnahme aus den Rücklagen} \\ + \text{Abschreibungen} \\ \hline = \text{Cash Flow} \end{array}$$

Abbildung 4.4: Cash Flow [60]

Hauptunterschied bei der Betrachtung des Cash Flows im Vergleich zum Gewinn stellen also die Abschreibungen dar. Da die Abschreibungen eine rechnerische Wertverminderung der eingesetzten Produktionsmittel über ihre Nutzungsdauer darstellen (siehe Kapitel 4.4.1), werden sie bei der Ermittlung des zu versteuernden Gewinns abgezogen. Bei der Berechnung des Cash Flows werden sie jedoch addiert, da sie einen Zufluss an finanziellen Mitteln darstellen, dem in dieser Periode keine direkten Ausgaben gegenüberstehen. Diese Mittel spielen eine große Rolle in der Cash Flow Analyse und können für Reinvestitionen oder zwischenzeitlich auch anderweitig verwendet werden.

Der Cash Flow ist nicht als Gewinn zu interpretieren, da in ihm zu leistender Aufwand

enthalten ist. Die Abschreibungen stellen ja Aufwendungen für Betriebsmittel dar, die bereits erfolgt sind. Die Verwendung des Begriffs Cash Flow als Erfolgsindikator eines Unternehmens ist also problematisch, da auch ein Unternehmen mit positivem Cash Flow nicht automatisch Gewinn macht.

Ebenfalls muss angemerkt werden, dass sich ein positiver Cash Flow nicht positiv auf die Liquidität eines Unternehmens auswirken muss. Bei hohem Investitionsbedarf kann sich sogar die Liquidität verringern, da die finanziellen Aufwendungen für diese Investitionen sofort geleistet werden müssen und damit die verfügbaren Mittel des Unternehmens geringer werden. Steuerlich können diese Ausgaben in Form der Abschreibungen aber erst über einige Jahre verteilt geltend gemacht werden.

4.2 Kenngrößen der Fertigung

4.2.1 Produktivität und Effizienz

Einen wichtigen Erfolgsfaktor und damit ein Teilziel des Unternehmens im technischen Bereich stellt die Produktivität der Fertigung dar.

Im allgemeinen ist die Produktivität als mengenmäßige Verhältnisgröße von Output zu Input einer Fertigung definiert [40, S. 18]:

$$\text{Produktivität} = \frac{\text{Ertragsmenge}}{\text{Faktoreinsatzmenge}} \quad (4.7)$$

Eine Erhöhung der Produktivität wird also erreicht, indem die vorhandenen Produktionsfaktoren besser ausgenutzt werden und mehr produziert wird, oder indem bei konstanter Produktionsmenge die benötigte Faktoreinsatzmenge reduziert wird.

Die Produktivität stellt daher ein Maß zur Bewertung der Leistungen einer Fertigung dar. Da betriebliche Leistungen aus der Kombination mehrerer Größen entstehen, diese aber nur selten die gleiche Dimension besitzen und damit normalerweise nicht addierbar sind, behilft man sich mit der Bildung von Teilproduktivitäten, wie z.B. der

$$\text{Arbeitsproduktivität} = \frac{\text{Ertragsmenge}}{\text{Arbeitseinsatz}}$$

oder der

$$\text{Kapitalproduktivität} = \frac{\text{Ertragsmenge}}{\text{Kapitaleinsatz}},$$

die nur zusammenfassbare Größen enthalten, dafür aber auch nur einen Teil der Leistungserstellung beleuchten können. (Teil-)Kennzahlen dieser Art sind deshalb vorwiegend für interne Vergleichs- oder Kontrollzwecke geeignet.

Für den speziellen Fall der Halbleiterfertigung bietet sich die folgende Formel zur Berechnung der Kapitalproduktivität an:

$$\text{Kapitalproduktivität} = \frac{\text{Zahl der produzierten funktionsfähigen Chips}}{\text{Kapitaleinsatz}} \quad (4.8)$$

Als Ertragsmenge wird dabei die Zahl der produzierten und funktionsfähigen, also der absatzfähigen Chips benutzt.

Eine gewisse Lösung für das Problem der unterschiedlichen Dimensionen ergibt sich, wenn die Einsatzfaktoren über einen einheitlichen Maßstab vergleichbar gemacht werden.

Dazu bietet sich eine Verwendung der Kosten an.

Ausdruck (4.9) wird im folgenden als Effizienz der Fertigung bezeichnet. Da es sich bei den Kosten nicht um eine mengenmäßige Größe handelt, wird hier statt Produktivität die Bezeichnung Effizienz verwendet. Der Begriff „Effizienz“ ist im Gegensatz zu „Produktivität“ und „Wirtschaftlichkeit“ in der grundlegenden Literatur (z.B. [40], [56], [57]) nicht oder nicht eindeutig belegt, weshalb er hier in Form von Gleichung (4.9) definiert wird.

$$\text{Effizienz} = \frac{\text{Zahl der produzierten funktionsfähigen Chips}}{\text{Kosten}} \quad (4.9)$$

Ähnlich wie beim Produktivitätsbegriff können hier auch noch andere (Teil-)Effizienzen definiert werden. So kann zur Detailanalyse einer realen Produktion z.B. die Anzahl der gefertigten Ebenen bezogen auf die Kosten, den Personaleinsatz, den Kapitalaufwand oder den Materialeinsatz ausgewertet werden. Solche Kenngrößen werden auch teilweise in der Industrie angewendet.

Der wertmäßige Kostenbegriff interpretiert Kosten als den in Geldeinheiten bewerteten Verzehr von Produktionsfaktoren (Sachgüter und Arbeits-, Dienst- und öffentliche Leistungen zur Erstellung betrieblicher Leistungen). Er stellt also ein Maß für die Knappheit der Güter und Leistungen dar.

Der Unterschied zwischen Kosten und Kapitaleinsatz liegt darin, dass beim Kapitaleinsatz die Summe der investierten Mittel betrachtet wird, nicht aber z.B. Rohmaterial (Wafer). Insofern würde die Betrachtung des Kapitaleinsatzes alleine nicht zu einer vollständigen Abdeckung aller relevanten Punkte führen.

Als problematisch stellt sich dabei die genaue Erfassung der Kosten dar. Im allgemeinen wird es nicht möglich sein, alle Kostenfaktoren genau zahlenmäßig zu erfassen. Deshalb sollen im folgenden nur die Kostenveränderungen bewertet werden, die aufgrund von Änderungen der Produktion und der daraus resultierenden Auswirkungen auf die Produktivität entstehen können.

Für die Produktivität und Effizienz der Fertigung werden möglichst hohe Werte angestrebt. Eine gesteigerte Produktivität bzw. Effizienz zeigt eine bessere Beherrschung der Fertigung und wirkt sich unter sonst gleichen Bedingungen positiv auf den Gewinn aus.

Es ist jedoch zu beachten, dass sich durch eine isolierte Angabe der Produktivität keine Aussagen über die Rentabilität einer Firma erzielen lassen, d.h., dass auch Produktivitätssteigerungen keine Garantie für Gewinn darstellen. Um die Produktivitätsdaten einordnen und bewerten zu können, müssen zusätzlich auch noch andere Faktoren wie Zeitdauer, Qualität der Fertigung und die Rahmenbedingungen des Marktes beachtet werden. Ohne Betrachtung dieser Aspekte ist es nicht möglich, den Beitrag von Produktivitätsveränderungen zur Rentabilität eines Unternehmens festzustellen.

Ein Beispiel dafür, dass steigende Produktivität keine größeren Unternehmenserfolg bewirken muss, ist in Abbildung 4.5 angegeben.

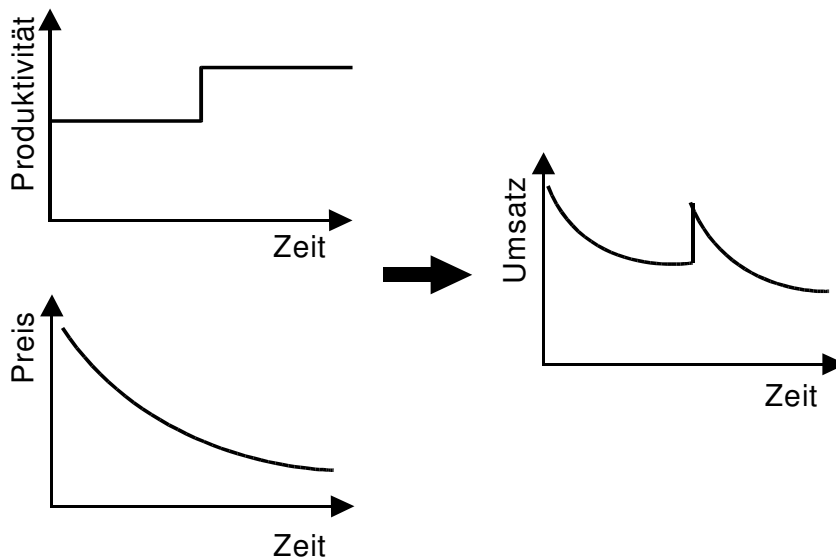


Abbildung 4.5: Umsatzverringering trotz steigender Produktivität

Unter der Voraussetzung einer konstanten Einsatzmenge kann durch eine Steigerung der Ertragsmenge, die zur dargestellten Produktivitätssteigerung führt, aufgrund des starken Preisverfalls, ein Absinken des Umsatzes und damit, bei angenommenen konstanten Kosten, eine Verringerung des Gewinns nicht vermieden werden. Um in diesem Fall einen steigenden Gewinn zu erzielen, müsste entweder die Produktivitätssteigerung noch höher ausfallen, oder es müssten zusätzliche Verbesserungen auf anderen Gebieten erfolgen.

Ein solcher Effekt kann auch darauf hindeuten, dass die Fertigung nicht so gut wie von anderen Wettbewerbern beherrscht wird, bzw. dass diese stärkere Verbesserungen erzielen konnten. Ziel muss es dann sein, die eigene Produktivität so stark zu steigern, dass wieder zu den Konkurrenzfirmen aufgeschlossen werden kann, bzw. diese sogar überholt werden können. Inzwischen betreiben viele Firmen Benchmarking, um ihre Stellung im Vergleich zu anderen Unternehmen analysieren zu können. Dadurch lassen sich auch notwendige Produktivitätssteigerungen und deren Auswirkungen noch besser beurteilen. Eine solche Wettbewerbsanalyse, die mit verschiedenen Methoden, wie z.B. Reverse Engineering, durchgeführt werden kann, stellt mittlerweile ein wichtiges Instrument dar. So können z.B. Konkurrenzprodukte in ihrem Aufbau untersucht werden, wodurch teilweise wichtige Informationen über die Prozessabfolge und damit sogar auch über ungefähre Prozessierungskosten ermittelt werden können.

4.2.2 Wirtschaftlichkeit

Durch eine monetäre Bewertung der Faktoren ergibt sich die sogenannte Wirtschaftlichkeit [56, S. 4]. Im Gegensatz zur Produktivität handelt es sich hier also nicht um mengenmäßige Größen, sondern um den marktmäßig bewerteten Ertrag und Aufwand.

$$\text{Wirtschaftlichkeit} = \frac{\text{Ertragsmenge} \cdot \text{Preis/Stück}}{\text{Faktoreinsatzmenge} \cdot \text{Preis/Einheit}} = \frac{\text{Ertrag}}{\text{Aufwand}} \quad (4.10)$$

Ausgehend von der Definition der Effizienz (4.9) kann damit die folgende Formel (4.11) für die Wirtschaftlichkeit aufgestellt werden. Als Ertrag wurde dabei die mit dem Marktpreis bewertete Zahl der verkaufsfähigen Chips, also der Umsatz, und als Aufwand die entstehenden Kosten für die Herstellung angesetzt.

$$\text{Wirtschaftlichkeit} = \frac{\text{Umsatz mit den produzierten funktionsfähigen Chips}}{\text{Kosten}} \quad (4.11)$$

Ziel muss es also sein, einen möglichst hohen Umsatz mit den produzierten Chips bezogen auf die eingesetzten Kosten zu erreichen [61].

Falls in Ausdruck (4.11) bei vorgegebenem Umsatz versucht wird, die Kosten möglichst niedrig zu halten, wird von Kostenwirtschaftlichkeit gesprochen. Andernfalls bei vorgegebenen Kosten und einem dabei optimierten Umsatz, wird dieser Ausdruck auch als Leistungswirtschaftlichkeit bezeichnet [62, S. 75].

Nach dem sogenannten ökonomischen Prinzip, lassen sich dabei drei verschiedene Zielsetzungen bei der Produktionsgestaltung unterscheiden [57, S. 287]:

- Ertragsmaximierung, bei der mit einem gegebenen Aufwand ein möglichst hoher Ertrag (= Nutzen) erzielt werden soll (Maximumprinzip).
- Aufwandsminimierung, bei der der nötige Aufwand, um einen bestimmten Ertrag zu erzielen, möglichst gering gehalten wird (Minimumprinzip).
- Erfolgsmaximierung, bei der die Differenz zwischen Ertrag und Aufwand maximiert wird (Optimumprinzip).

Die Erfüllung des ökonomischen Prinzips, und damit die Verbesserung der Wirtschaftlichkeit, trägt unter sonst gleichen Bedingungen zur Steigerung des Gewinns bei.

4.3 Leistung der IC-Fertigung

Als Leistung der Fertigung wird, wie in Ausdruck (4.11) beschrieben, die Anzahl der produzierten und funktionsfähigen Chips und der damit erzielte Umsatz verwendet.

Die Zahl der produzierten und funktionsfähigen Chips berechnet sich aus der Zahl der prozessierten Produktwafer, der Zahl der Chips pro Wafer und der Gesamt-Fertigungsausbeute.

$$C = \text{WSPM} \cdot (1 - m) \cdot \text{CPW} \cdot Y \quad (4.12)$$

C = Zahl der Chips pro Monat

WSPM = Wafer Starts per Month

m = prozentualer Anteil der Monitorwafer an den gesamten Waferstarts

CPW = Chips pro Wafer

Y = Ausbeute

Unter Benutzung der Formel von Hennessy/Patterson [63, S. 12] berechnet sich die Zahl der Chips pro Wafer bei gegebenem Waferdurchmesser und gegebener Chipgröße folgendermaßen:

$$\text{CPW} = \frac{\pi \cdot R_{\text{Wafer}}^2}{A} - \frac{\pi \cdot D_{\text{Wafer}}}{\sqrt{2 \cdot A}} \quad (4.13)$$

R_{Wafer} = Waferradius

D_{Wafer} = Waferdurchmesser

A = Chipfläche

Mit dem Modell von Poisson [64], [65] ergibt sich die Ausbeute abhängig von der Defektdichte und der Chipgröße, wobei die Defektdichte in dieser Formel keine rein physikali-

sche Größe, sondern eine produktbezogene, rechnerische Größe darstellt.

$$Y = \exp(-D \cdot A) \quad (4.14)$$

D = Defektdichte

In Gleichung (4.14) ist erkennbar, dass die Ausbeute steigt, wenn die Chipfläche und die Defektdichte reduziert werden.

Gleichung (4.13) und (4.14) in (4.12) eingesetzt, ergibt folgende Formel für die Zahl der produzierten und funktionsfähigen Chips:

$$C = WSPM \cdot (1 - m) \cdot \left(\frac{\pi \cdot R_{\text{Wafer}}^2}{A} - \frac{\pi \cdot D_{\text{Wafer}}}{\sqrt{2 \cdot A}} \right) \cdot \exp(-D \cdot A) \quad (4.15)$$

Durch Multiplikation mit dem Produktpreis ergibt sich der Umsatz, der mit den produzierten und funktionsfähigen Chips erzielt werden kann.

$$U = WSPM \cdot (1 - m) \cdot \left(\frac{\pi \cdot R_{\text{Wafer}}^2}{A} - \frac{\pi \cdot D_{\text{Wafer}}}{\sqrt{2 \cdot A}} \right) \cdot \exp(-D \cdot A) \cdot P \quad (4.16)$$

U = Umsatz

P = Produktpreis

Durch Einfügen von zusätzlichen Faktoren in Gleichung (4.16) kann die zeitliche Veränderung der Chipgröße und der Defektdichte berücksichtigt werden. Dabei gelten:

$$A(t) = A_0 \cdot s^t \quad (4.17)$$

A_0 = Anfangschipgröße

s = zeitlicher Veränderungsfaktor der Chipgröße

t = Zeit

$$D(t) = D_0 \cdot d^t \quad (4.18)$$

D_0 = Anfangsdefektdichte

d = zeitlicher Veränderungsfaktor der Defektdichte

Damit ergibt sich für die Zahl der produzierten Chips:

$$C(t) = WSPM \cdot (1 - m) \cdot \left(\frac{\pi \cdot R_{\text{Wafer}}^2}{A_0 \cdot s^t} - \frac{\pi \cdot D_{\text{Wafer}}}{\sqrt{2 \cdot A_0 \cdot s^t}} \right) \cdot \exp(-D_0 \cdot d^t \cdot A_0 \cdot s^t) \quad (4.19)$$

Zur Berechnung des erzielten Umsatzes muss die obige Gleichung (4.19) für die Zahl der produzierten Chips noch mit dem entsprechenden (ebenfalls zeitlich veränderlichen) Produktpreis $P(t) = P_0 \cdot p^t$ multipliziert werden.

$$U(t) = WSPM \cdot (1 - m) \cdot \left(\frac{\pi \cdot R_{\text{Wafer}}^2}{A_0 \cdot s^t} - \frac{\pi \cdot D_{\text{Wafer}}}{\sqrt{2 \cdot A_0 \cdot s^t}} \right) \cdot \exp(-D_0 \cdot d^t \cdot A_0 \cdot s^t) \cdot P_0 \cdot p^t \quad (4.20)$$

P_0 = Anfangspreis

p = zeitlicher Veränderungsfaktor des Preises

4.4 Kosten für die IC-Fertigung

4.4.1 Aufteilung der Kosten

Die Gesamtkosten für die Fertigung von integrierten Schaltungen setzen sich zusammen aus den fixen und den variablen Kosten.

Dabei stellen die fixen Kosten Aufwendungen dar, die von der erzielten Produktionsmenge nicht beeinflusst werden, also z.B. Geräte- oder Gebäudeabschreibungen. Diese Kosten sind grundsätzlich vorhanden, egal, ob in der Fabrik produziert wird oder nicht. Die variablen Kosten dagegen ändern sich in Abhängigkeit von der Produktionsmenge. Beispielsweise werden bei höherem Ausstoß mehr Roh- und Verbrauchsmaterialien benötigt, was in diesem Bereich zu einer Kostensteigerung führt.

$$K_G = K_V + K_F \quad (4.21)$$

K_G = Gesamtkosten

K_V = variable Kosten

K_F = fixe Kosten

Bei linearer Abhängigkeit der variablen Kosten von der Produktionsmenge ergibt sich also die Darstellung in Abbildung 4.6. Das Verhältnis zwischen fixen und variablen Kosten ist je nach Industriebranche und Unternehmen unterschiedlich. In der Halbleiterindustrie liegen aufgrund der hohen Investitionen die fixen Kosten sehr hoch.

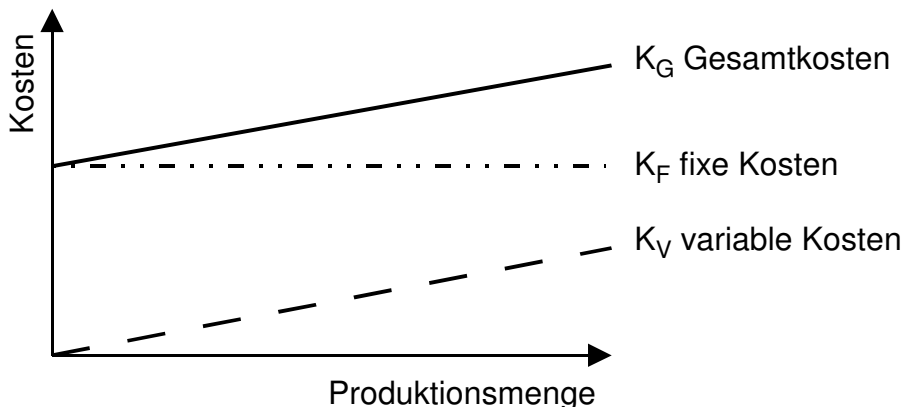


Abbildung 4.6: Kosten

Laut Motorola verteilen sich die Chipkosten für 200 mm und 300 mm Wafer bei einer Produktionsmenge von 20000 Wafern/Monat ungefähr wie folgt auf die einzelnen Kostenfaktoren [66].

Tabelle 4.1: Aufteilung der Chipkosten

	200 mm Wafer	300 mm Wafer
Geräteabschreibungen	44,5%	42,0%
Personal direkt	7,5%	5,5%
Personal indirekt	10,5%	7,0%
Gebäudeabschreibungen	9,5%	8,5%
<i>fixe Kosten</i>	<i>72,0%</i>	<i>63,0%</i>
Geräteunterhalt	12,0%	15,0%
Wafer	8,5%	12,5%
Verbrauchsmaterialien	6,5%	6,0%
Kontrollwafer	1,0%	3,5%
<i>variable Kosten</i>	<i>28,0%</i>	<i>37,0%</i>

Wie aus Tabelle 4.1 ersichtlich ist, nehmen die fixen Kosten, zu denen Abschreibungen für Geräte und Reinraum und Personalkosten gerechnet werden, bei der Chipfertigung einen sehr großen Anteil ein. Die variablen Kosten für Wafer, Verbrauchsmaterialien und Geräteunterhalt liegen dagegen nur bei rund 30% - 35% der Gesamtkosten. Erwähnenswert ist, dass ein gewisser Energieaufwand in der Halbleiterindustrie immer notwendig ist, um die Betriebsbereitschaft der Geräte und des Reinraums aufrechtzuerhalten. Das heißt, dass auch wenn nicht produziert wird in diesem Bereich Kosten entstehen, die ebenfalls als fix zu bezeichnen sind. Das immer vorhandene Grund-Kostenniveau in einer Halbleiterfab ist also sehr hoch und man muss sich bemühen, durch eine gute Auslastung der Fabrik, die eine hohe Produktionsmenge bewirkt, eine Verteilung der Kosten auf möglichst viele Chips und damit eine Senkung der Stückkosten zu erreichen.

Eine besondere Stellung kommt den Abschreibungen zu. Durch die Abschreibungen werden die betriebsbedingten Wertminderungen des Anlagevermögens (z.B. Gebäude, Produktionsanlagen) eines Unternehmens berücksichtigt. Der sich ergebende Wertverlust wird auf die Nutzungsdauer der Güter verteilt. Bei der linearen Abschreibung wird jedes Jahr der gleiche Betrag abgeschrieben. Bei einem Restwert von Null ergeben sich daraus folgende Abschreibungskosten pro Jahr [58, S. 78]:

$$\text{Abschreibung p.a.} = \frac{\text{Anschaffungspreis}}{\text{Abschreibungsdauer}} \quad (4.22)$$

Wegen der hohen erforderlichen Investitionen werden in der IC-Fertigung sehr hohe Geräteabschreibungen erreicht. Die bestehenden Abschreibungen wirken sich auf die Gesamtkosten aus und beeinflussen somit den Chippreis. Daraus folgt, dass die Herstellung von Chips in einer vollständig abgeschrieben Fab am kostengünstigsten wäre. Eine solche Vorgehensweise wäre jedoch nur für veraltete Produkte machbar, da für die Durchführung einer Produktion mit modernen Strukturgrößen, abhängig von der Nutzungsdauer der Geräte, immer wieder neue Investitionen erforderlich sind, so dass neue Abschreibungen entstehen. Bei Fabs, die ständig mit der modernsten Strukturgröße betrieben werden, ist es notwendig, pro Jahr 25 - 30% der Geräte neu anzuschaffen.

In steuerlicher Hinsicht und im Hinblick auf den Cash Flow wirken sich die hohen Abschreibungen allerdings günstig aus. Da die Abschreibungen keine direkten, sondern nur

steuerlich wirksame Ausgaben repräsentieren, stellt der Cash Flow die Menge an Geld dar, die dem Unternehmen im entsprechenden Zeitraum zufließt. Aus finanzieller Sicht stellen also die Abschreibungsmöglichkeiten einen wichtigen Aspekt für die Unternehmen dar. In den USA ist beispielsweise momentan eine minimale Abschreibungsdauer von fünf Jahren für Geräte der Chipfertigung gesetzlich erlaubt. Aufgrund der schnellen Veralterung dieser Herstellungsgeräte ist jedoch eine Reduktion auf minimal drei Jahre geplant [67], [68].

Für 200 mm Wafer und angenommene Gesamtinvestitionen von 1 Mrd. € ergibt sich aus den oben dargestellten Zusammenhängen bei Berücksichtigung einer fünfjährigen Abschreibungszeit für die Geräte und einer zehnjährigen Abschreibungszeit für das Gebäude die folgende Kostenaufteilung. Dabei wird davon ausgegangen, dass sich die Kosten pro Chip aus dem Verhältnis von Gesamtkosten zur Anzahl der produzierten Chips berechnen.

Geräteabschreibungen:	140 Mio. € p.a.
Personal direkt	23,5 Mio. € p.a.
Personal indirekt	33 Mio. € p.a.
<u>Gebäudeabschreibungen</u>	<u>30 Mio. € p.a.</u>
<i>fixe Kosten</i>	<i>226,5 Mio. € p.a.</i>
Geräteunterhalt	38 Mio. € p.a.
Wafer	27 Mio. € p.a.
Verbrauchsmaterialien	20,5 Mio. € p.a.
<u>Kontrollwafer</u>	<u>3 Mio. € p.a.</u>
<i>variable Kosten</i>	<i>88,5 Mio. € p.a.</i>
Gesamtkosten	315 Mio. € p.a.

Nicht enthalten in dieser Aufstellung ist eventuell anfallender Zinsaufwand, der sich bei der Kreditfinanzierung der Fabrik ergeben und sich in der Gewinn-und Verlustrechnung negativ niederschlagen würde.

Die gesamten entstehenden Kosten verteilen sich auf die Menge der produzierten und funktionsfähigen Chips, so dass die Stückkosten folgendermaßen berechnet werden können:

$$\text{Stückkosten.} = \frac{\text{Gesamtkosten}}{\text{Zahl der funktionsfähigen Chips}} \quad (4.23)$$

Eine isolierte Betrachtung der Gesamtkosten ist also nicht besonders aussagekräftig. Zusätzlich muss die mit diesem Aufwand erbrachte Produktionsleistung berücksichtigt werden. Eine Verringerung der Stückkosten kann durchaus mit einer Erhöhung der Gesamtkosten einhergehen, wenn dadurch eine entsprechend größere Menge an Produkten gefertigt werden kann.

Die Stückkosten des Produkts werden durch die hergestellte Menge und durch Erfahrungswerte beeinflusst. Bereits in den 60er Jahren wurde durch empirische Beobachtungen festgestellt, dass im Lauf der Zeit, bzw. mit steigender kumulierter Produktionsmenge, die Stückkosten sinken [69, S. 241 ff.]. Dieser Effekt kommt zum großen Teil durch die sogenannte Fixkostendegression zustande, die bewirkt, dass sich die fixen Kosten, die den konstanten Anteil der Gesamtkosten darstellen, auf eine größere Menge an Produk-

ten verteilen. Der Absolutwert der variablen Kosten steigt zwar dabei, die variablen Kosten pro Stück bleiben jedoch konstant, so dass sich insgesamt niedrigere Kosten pro Stück ergeben.

Zudem werden positive Effekte erzielt durch den technischen Fortschritt, also den Einsatz moderner technischer Verfahren, durch Betriebsgrößeneffekte (economies of scale), die davon ausgehen, dass sich mit zunehmender Betriebsgröße die Stückkosten senken lassen, und durch Lerneffekte, die sich im Lauf der Zeit ergeben [70]. Die Auswirkungen dieser Effekte sind in der Erfahrungskurve in Abbildung 4.7 dargestellt (nach McIvor [71, S. 2-21])). Durch große Produktionsmengen lassen sich also Kostenvorteile erzielen, ein Punkt, dem in der Halbleiterindustrie wegen des hohen Fixkostenanteils hohe Bedeutung zukommt.

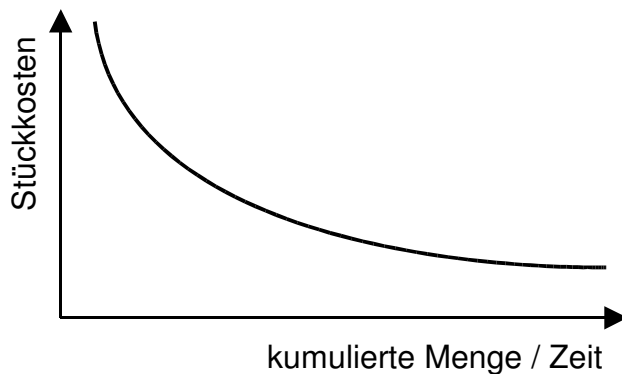


Abbildung 4.7: Lernkurve

4.4.2 Kostenaufwand für die Veränderung der einzelnen Einflussfaktoren

Von Interesse ist nun, welchen Einfluss die Parameter, die die Zahl der produzierten Chips, also die Leistung der Produktion, bestimmen, auf die Kosten ausüben. Um eine Aussage über die Machbarkeit von Verbesserungsmaßnahmen zu erhalten, muss die Empfindlichkeit gegenüber Parameteränderungen und das Verbesserungspotential mit dem damit verbundenen Aufwand, also mit den entstehenden Kosten, bewertet werden. Erst danach kann eine endgültige Einschätzung der Auswirkungen der Maßnahmen angegeben werden, da ein Parameter mit großem Einfluss mit einem sehr hohen Aufwand verbunden sein kann und sich so im Endeffekt eine Veränderung dieses Parameters als weniger „günstig“ als die eines anderen Parameters erweisen kann.

- Zahl der WSPM

Die Zahl der Waferstarts stellt einerseits eine strategische Planungsgröße vor Errichtung der Fabrik dar, durch die im wesentlichen die Fabrikkosten bestimmt werden. Eine größere Produktionsmenge erfordert im allgemeinen eine größere Anzahl von Geräten und eine größere Produktionsfläche und damit höhere Investitionen, sowie im allgemeinen mehr Personal. Damit erhöhen sich die fixen Produktionskosten und zusätzlich, durch die größere Menge an benötigten Wafern und Verbrauchsmaterialien, die Höhe der variablen Kosten.

Theoretisch kann die Produktionsmenge auch durch den parallelen Betrieb einer zweiten Produktionslinie gesteigert werden. Dabei ergeben sich dann zwar die doppelte Fertigungskapazität, aber auch die doppelten Kosten einer einzelnen Fertigungslinie.

Wirtschaftlicher kann es dagegen sein, eine bestehende Fab zu erweitern. Dabei können unter Umständen die bestehenden Geräte besser ausgenutzt werden, bzw. man erspart sich die Anschaffung einiger Geräte, da vorhandene ungenutzte Restkapazitäten bei einigen Anlagen dann verwendet werden können. Allerdings können beim Bau von Produktionsstätten auch andere Gründe eine Rolle spielen. Beispielsweise kann überlegt werden in einem wichtigen Absatzmarkt eine eigene Fertigungslinie aufzubauen.

Andererseits sind aber auch Steigerungsmöglichkeiten durch Verbesserung des Produktionsablaufs gegeben, d.h. dass durch eine verbesserte Nutzung der vorhandenen Geräte erhöhte Produktionsmengen ohne zusätzliche Investitionen möglich sind. Bei einer auf diese Weise gesteigerten Produktionsmenge ergeben sich dann nur höhere variable Kosten für Wafer und Verbrauchsmaterialien.

- Zahl der Monitorwafer

Monitorwafer (= Testwafer) werden eingesetzt, um die Qualität der jeweiligen Prozessschritte durch Messungen überprüfen zu können. Eine Nutzung dieser Wafer für die eigentliche Produktion scheidet also aus. Üblicherweise werden Testscheiben nur für jeweils einen einzelnen Prozess verwendet, d.h. dass sie an einem Gerät mitprozessiert und danach gemessen werden. Teilweise werden aber auch Entwicklungslose vollständig prozessiert oder in Prozesssequenzen bearbeitet.

Durch Verringerung der Zahl der Monitorwafer können mehr Produktwafer eingesetzt werden. Da Monitorwafer im Gegensatz zu Produktwafern teilweise mehrfach benutzt werden können, ergeben sich daraus etwas geringere Kosten im Vergleich zu Produktwafern [72].

Verbesserte Prozesskontrolle, und damit reduzierter Verbrauch von Monitorwafern, ist auch durch den Einsatz von in situ Messsystemen denkbar, der allerdings voraussetzt, dass derartige Sensoren in die entsprechenden Geräte integriert werden können und auch erhöhte Gerätekosten verursacht.

Andererseits ist aber auch durch die bessere Prozessbeherrschung eine Einsparung von Testvorgängen und damit von Testkosten denkbar.

- Waferdurchmesser

Die Steigerung des Waferdurchmessers bringt großes Verbesserungspotential mit sich, da z.B. beim Umstieg von 200 mm auf 300 mm Wafer die nutzbare Fläche mindestens um Faktor 2,25 ansteigt; sie erfordert aber auch erhebliche Investitionen, da mit den alten Geräten keine Wafer mit größerem Durchmesser prozessiert werden können und somit praktisch alle Geräte neu angeschafft werden müssen, wobei der Preis der Geräte für 300 mm Produktion um Faktor 1,3 - 1,8 über dem für 200 mm Geräte liegt [33], [73], [74].

Zusätzlich muss beachtet werden, dass der Preis für 300 mm Wafer höher ist als für 200 mm Wafer. Wie in der Vergangenheit wird aber auch hier erwartet, dass der Waferpreis im Lauf der Zeit sinkt.

Insgesamt wird erwartet, dass die Kosten für einen prozessierten 300 mm Wafer um 33 - 45% höher liegen als bei einem 200 mm Wafer. Wegen der größeren Fläche und damit der größeren Zahl produzierter Chips werden allerdings um 30 - 40% niedrigere Kosten pro Chip prognostiziert [42], [75].

Während der Lebensdauer einer Fabrik ist also der Umstieg auf einen höheren Waferdurchmesser nur sehr schwer möglich. Theoretisch ist auch ein zeitweiliger Mischbetrieb denkbar, der aber während der Umstellungsphase Probleme in der Fertigung erzeugt.

- Defektdichte

Eine niedrigere Defektdichte wird im allgemeinen durch bessere Prozesskontrolle und -beherrschung erreicht, so dass keine großen zusätzlichen Investitionen notwendig sind.

Es muss allerdings beachtet werden, dass im Lauf der Zeit bei immer weiter sinkenden Strukturgrößen stärkere Anforderungen an die Partikelvermeidung gestellt werden, da immer kleinere Partikel detektiert, ausgefiltert oder entfernt werden müssen, was auch dazu führen kann, dass manche Prozesse aufgrund ihrer Partikelbelastung durch andere ersetzt werden müssen. Insofern entstehen steigende Anforderungen an die Gerätehersteller und es ist auf lange Sicht ein steigender Aufwand anzunehmen [76].

- Strukturgröße, Chipgröße

Bei manchen Prozessen können nur kleine Verbesserungen der Strukturgröße mit den bestehenden Geräten erreicht werden. Um weitere Verringerungen herbeiführen zu können, ist der Kauf neuer Geräte, vor allem für die Lithographie, notwendig. Deswegen müssen pro Jahr 25 - 30% der Geräte neu angeschafft werden, wenn die Fabrik weiterhin mit modernsten Strukturgrößen produzieren soll. Dabei muss in Kauf genommen werden, dass der Preis für neue Geräte aufgrund der höheren Qualitätsanforderungen steigt. Da jedoch nur bei einigen kritischen Maskenschritten (z.B. Gatedefinition, first level contact) die minimale Strukturgröße erreicht werden muss, müssen nur diese Geräte neu gekauft werden und es ist nicht notwendig, alle Lithographiegeräte neu anzuschaffen (Mix and Match Strategie). Allerdings muss bei einer solchen Vorgehensweise beachtet werden, dass der Wartungsaufwand bei Verwendung unterschiedlicher Geräte ansteigt. Dies ist nicht nur bei Lithographiegeräten sondern allgemein, z.B. bei Plasmaätzern, Messgeräten oder Abscheideanlagen, der Fall.

Insgesamt kann gesagt werden, dass die Steigerung des Waferdurchmessers die größten Kosten verursacht. Jede Änderung, die den Zukauf neuer oder zusätzlicher Geräte notwendig macht, verursacht Investitionen in Millionenhöhe.

4.5 Berechnung der Effizienz der Fertigung

4.5.1 Erstellung eines Effizienz-Modells

Die wichtigsten Einflussfaktoren auf die Effizienz der Fertigung werden in einem Berechnungsmodell zusammengefasst. Ziel ist es, eine möglichst ausführliche und exakte Formel zu erstellen, die unter Beachtung der speziellen Situation in der Halbleiterfertigung folgende Kriterien erfüllt:

- Betrachtung der wichtigsten Einflussgrößen
- Berechnung der Empfindlichkeit gegenüber Veränderungen dieser Einflussgrößen
- Berücksichtigung der gegenseitigen Abhängigkeiten der Einflussgrößen
- Einbezug dynamischer Veränderungen der Parameter

Mit den Ergebnissen aus Kapitel 4.3 und Kapitel 4.4, Ausdruck (4.9) und Gleichung (4.15) ergibt sich für die Effizienz der Fertigung:

$$E = \frac{WSPM \cdot (1-m) \cdot \left(\frac{\pi \cdot R_{\text{Wafer}}^2}{A} - \frac{\pi \cdot D_{\text{Wafer}}}{\sqrt{2} \cdot A} \right) \cdot \exp(-D \cdot A)}{K_G} \quad (4.24)$$

E = Effizienz der Fertigung
 K_G = Gesamtkosten

Beim Einbezug der zeitlichen Veränderungen ergibt sich durch Division von Gleichung (4.19) durch die Kosten auch hier die Effizienz:

$$E(t) = \frac{WSPM \cdot (1-m) \cdot \left(\frac{\pi \cdot R_{\text{Wafer}}^2}{A_0 \cdot s^t} - \frac{\pi \cdot D_{\text{Wafer}}}{\sqrt{2} \cdot A_0 \cdot s^t} \right) \cdot \exp(-D_0 \cdot d^t \cdot A_0 \cdot s^t)}{K_G} \quad (4.25)$$

Die Wirtschaftlichkeit, die in Gleichung (4.11) definiert wurde, berechnet sich mit Gleichung (4.20):

$$W(t) = \frac{WSPM \cdot (1-m) \cdot \left(\frac{\pi \cdot R_{\text{Wafer}}^2}{A_0 \cdot s^t} - \frac{\pi \cdot D_{\text{Wafer}}}{\sqrt{2} \cdot A_0 \cdot s^t} \right) \cdot \exp(-D_0 \cdot d^t \cdot A_0 \cdot s^t) \cdot P_0 \cdot p^t}{K_G} \quad (4.26)$$

W = Wirtschaftlichkeit

Mit Hilfe der Formeln können zeitliche Veränderungen der Eingangsparameter beurteilt werden. Ziel der Fertigung ist es, eine möglichst große Steigerung der Effizienz zu erzielen. Da sich der Preis im Halbleiterbereich laufend ändert, ist der Vorteil, der aus einer hohen Effizienz gezogen werden kann, allerdings stark zeitabhängig. Vorteilhaft bei der Betrachtung der Wirtschaftlichkeit ist im Gegensatz dazu, dass der erzielbare Preis bereits in dieser Kenngröße enthalten ist.

Eine höhere Effizienz bedeutet ein günstigeres Verhältnis von Produktionsergebnis zu Kosten und damit eine besser beherrschte Fertigung. Produktivität bzw. Effizienz und Wirtschaftlichkeit stellen zwar wichtige Bewertungsgrößen dar, müssen aber, wie bereits erläutert, zusätzlich im Zusammenhang mit anderen Größen (z.B. Zeitfaktor, Rahmenbedingungen) betrachtet werden. Zusätzlich könnten theoretisch mit dieser Kennzahl verschiedene Fertigungslinien miteinander verglichen werden. Ein solcher Vergleich ist normalerweise jedoch schwierig, da die Kennzahlen auf dieselbe Weise ermittelt werden und die gleichen Bezugsgrößen enthalten müssten. Im allgemeinen ist also die Vergleichbarkeit solcher Kennziffern nicht immer gegeben.

4.5.2 Empfindlichkeit gegenüber Parameteränderungen

Zur Bewertung des bestehenden Verbesserungspotentials ist es wichtig die Empfindlichkeit des Ausgangsparameters (Zahl der funktionsfähigen Chips) gegenüber Variationen der Eingangsparameter festzustellen. Es soll also einerseits herausgefunden werden, welche Parameter das Ergebnis am stärksten beeinflussen und andererseits wieviel Raum

zur Verbesserung bei dem jeweiligen Parameter noch besteht.

Formel (4.15) zur Berechnung der Zahl der funktionsfähigen Chips hängt von 5 Parametern, nämlich von der Zahl der Waferstarts, dem Anteil der Monitorwafer, dem Waferdurchmesser, der Chipgröße und der Defektdichte ab. Zur Berechnung der Empfindlichkeit wird die partielle Ableitung nach allen Variablen gebildet.

$$\frac{\partial C}{\partial WSPM} = (1-m) \cdot \left(\frac{\pi \cdot D_W^2}{4 \cdot A} - \frac{\pi \cdot D_W}{\sqrt{2 \cdot A}} \right) \cdot \exp(-D \cdot A) \quad (4.27)$$

$$\frac{\partial C}{\partial m} = -WSPM \cdot \left(\frac{\pi \cdot D_W^2}{4 \cdot A} - \frac{\pi \cdot D_W}{\sqrt{2 \cdot A}} \right) \cdot \exp(-D \cdot A) \quad (4.28)$$

$$\frac{\partial C}{\partial D_W} = WSPM \cdot (1-m) \cdot \left(\frac{\pi \cdot D_W}{2 \cdot A} - \frac{\pi}{\sqrt{2 \cdot A}} \right) \cdot \exp(-D \cdot A) \quad (4.29)$$

$$\frac{\partial C}{\partial D} = WSPM \cdot (1-m) \cdot \left(\frac{\pi \cdot D_W^2}{4 \cdot A} - \frac{\pi \cdot D_W}{\sqrt{2 \cdot A}} \right) \cdot \exp(-D \cdot A) \cdot (-A) \quad (4.30)$$

$$\frac{\partial C}{\partial A} = WSPM \cdot (1-m) \cdot \left(\frac{-\pi \cdot D_W^2}{4 \cdot A^2} + \frac{\pi \cdot D_W}{2 \sqrt{2 \cdot A}^{3/2}} \right) \cdot \exp(-D \cdot A) \quad (4.31)$$

$$+ WSPM \cdot (1-m) \cdot \left(\frac{\pi \cdot D_W^2}{4 \cdot A} - \frac{\pi \cdot D_W}{\sqrt{2 \cdot A}} \right) \cdot \exp(-D \cdot A) \cdot (-D)$$

Da sich die Größenordnung der einzelnen Parameter (z.B. liegt die Zahl der Waferstarts bei 20000, während die Chipgröße im Bereich von einigen cm² liegt) und damit auch der Parametervariationen stark unterscheiden, ist es nicht möglich die absoluten Werte der partiellen Ableitungen direkt miteinander zu vergleichen. Um eine solche Vergleichbarkeit zu erhalten, wird die Änderung des Ergebnisses bei prozentual gleichen Abweichungen der Eingangsparameter um den Arbeitspunkt herum angegeben. Dies geschieht, indem die partielle Ableitung mit einer kleinen Änderung der Variablen multipliziert wird. Diese Vorgehensweise beruht auf der Tatsache, dass die Änderung eines Funktionswertes δy bei kleinen Veränderungen des Eingangsparameters δx durch die Formel

$$\delta y = \frac{\partial y}{\partial x} \cdot \delta x$$

abgeschätzt werden kann, was in Abbildung 4.8 verdeutlicht werden soll.

Um die Empfindlichkeit an verschiedenen Arbeitspunkten miteinander vergleichen zu können, wird das Ergebnis noch auf die Zahl der Chips im Arbeitspunkt C_{AP} normiert.

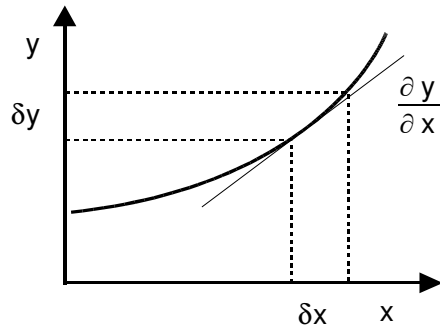


Abbildung 4.8: Linearisierung bei kleinen Parameteränderungen

Für die relative Empfindlichkeit ergibt sich damit:

$$\text{sensC} = \begin{bmatrix} \frac{\partial C}{\partial \text{WSPM}} \cdot \frac{\delta \text{WSPM}}{C_{\text{AP}}} \\ \frac{\partial C}{\partial m} \cdot \frac{\delta m}{C_{\text{AP}}} \\ \frac{\partial C}{\partial D_{\text{W}}} \cdot \frac{\delta D_{\text{W}}}{C_{\text{AP}}} \\ \frac{\partial C}{\partial D} \cdot \frac{\delta D}{C_{\text{AP}}} \\ \frac{\partial C}{\partial A} \cdot \frac{\delta A}{C_{\text{AP}}} \end{bmatrix} \quad (4.32)$$

Die resultierenden Werte sagen aus, um welchen Prozentsatz bezogen auf den Arbeitspunkt sich die Zahl der Chips ändert, wenn der jeweilige Eingangsparameter verändert wird.

Im allgemeinen handelt es sich hier um nichtlineare Zusammenhänge, d.h. die Empfindlichkeit gegenüber den einzelnen Parametern ist nicht konstant, sondern hängt vom jeweiligen Arbeitspunkt ab.

4.5.3 Verbesserungspotential der einzelnen Faktoren

Zusätzlich zur Empfindlichkeit gegenüber Parameteränderungen muss noch beachtet werden, welches Verbesserungspotential die einzelnen Parameter besitzen, d.h. welcher Spielraum für Änderungen, ausgehend vom heute erreichten Stand, noch ausgeschöpft werden kann. Ein Parameter, der bereits sehr gut beherrscht wird, kann zwar immer noch eine große Empfindlichkeit aufweisen, wird aber im großen und ganzen nicht mehr stark verbessert werden können. Deshalb sollen hier die bei modernen, gut kontrollierten Fertigungen momentan erreichten Daten und die zukünftigen Möglichkeiten abgeschätzt werden.

- WSPM

Typischer Wert: 20000 Wafer/Monat, bei sehr großen Chipfabriken auch bis zu 40000 - 50000 Wafer/Monat.

Durch geschicktes Ausnutzen der Geräte und durch Optimierung des Fertigungsablaufs

sind noch Verbesserungen denkbar. Da eine Steigerung um 10% in diesem Bereich eine Erhöhung der Produktionsmenge um ungefähr 2000 Wafer/Monat bedeuten würde, handelt es sich sicher um einen wichtigen Einflussparameter.

Aufgrund der Tatsache, dass die Fertigungsgeräte noch nicht ideal ausgenutzt werden (vgl. Kapitel 5), besteht noch ein großes Verbesserungspotential in diesem Bereich.

- Monitorwafer

Momentaner Anteil von Monitorwafern bezogen auf die Zahl der gestarteten Wafer: ca. 5% (Da Monitorwafer üblicherweise nur für jeweils einen einzelnen Prozessschritt verwendet werden, handelt es sich hier um einen Durchschnittswert.)

Spielraum für Verbesserungen besteht, da die Zahl der notwendigen Monitorwafer die Zahl der Produktwafer verringert und sich somit Veränderungen bei diesem Parameter direkt auf die nutzbare Produktionsmenge auswirken. Im Idealfall wäre die Fertigung so gut beherrscht, bzw. durch in situ Messmethoden so gut kontrolliert, dass vollständig auf Monitorwafer verzichtet werden kann [77], d.h., dass ausgehend vom heutigen Stand ungefähr 5% mehr Produktwafer prozessiert werden könnten.

- Waferdurchmesser

Momentaner Wert: 200 mm - 300 mm

Solange es den Silizium-Herstellern gelingt, Wafer mit immer größerem Durchmesser zu liefern, kann die Entwicklung in diesem Bereich weiterlaufen. Zusätzlich zur größeren Waferfläche (Faktor 2,25 beim Vergleich 300 mm / 200 mm) ergibt sich bei größerem Waferdurchmesser eine bessere Randausnutzung, so dass sich die Zahl der Chips, abhängig von der Chipgröße, bei 300 mm Wafern um Faktor 2,4 - 2,5 im Vergleich zu 200 mm Wafern steigern kann.

Es muss allerdings beachtet werden, dass der Umstieg auf größere Waferdurchmesser die Neuanschaffung fast aller Geräte erfordert, so dass ein solches Vorgehen zwar ein sehr großes Verbesserungspotential enthält, während der Lebensdauer einer Fabrik aber ohne große Investitionen nicht möglich ist.

- Defektdichte

Momentan erreichte Größenordnung: ca. 0,05 Partikel/cm²

Eine weitere Verringerung der Defektdichte ist zu erwarten und stellt eine anspruchsvolle Herausforderung dar. Da die davon beeinflusste Ausbeute aber schon heute im allgemeinen sehr hoch liegt [78], dürften die Auswirkungen dieser Verbesserungen geringer sein als in den anderen Bereichen. Hohe Bedeutung kommt der schnellen Verringerung der Defektdichte am Beginn der Produktion zu. Da bei der Einführung von neuen Prozessen in der Anfangsphase die Ausbeute relativ niedrig ist, muss es gelingen, durch schnelle Lernzyklen die Prozesse möglichst frühzeitig gut zu beherrschen [79].

- Chipfläche

Der Flächenbedarf eines Chips wird durch seine Funktion und die Strukturgröße festgelegt. Die sogenannten Shrinks sind dabei ein wichtiges Mittel in der Halbleiterindustrie, um die Produktivität der Fertigung zu steigern [80]. Dazu wird der Chip während seiner Produktlebensdauer ständig mit geringeren Strukturgrößen gefertigt, so dass die Fläche kontinuierlich sinkt, um so mehr Chips pro Wafer herstellen zu können. Ein schnelles und häufiges Durchführen von Shrinks ist also von enormer Bedeutung und beinhaltet ein sehr großes Verbesserungspotential. In [81] wird als Zielvorstellung ein Shrink um 30% alle 2 Jahre erwähnt.

4.6 Entwicklung beim Hochfahren der Fertigungskapazität

4.6.1 Zeitliche Entwicklung der Produktionsmenge und -ausbeute

Bei Inbetriebnahme einer neuen Fertigungslinie kann die Produktionskapazität und -ausbeute nur langsam im Lauf der Zeit gesteigert werden. Das Ausmaß, in welchem die vorhandene maximale Kapazität der Fertigung in Anspruch genommen wird, wird als Kapazitätsausnutzungsgrad oder Beschäftigungsgrad bezeichnet [58, S. 74].

$$\text{Kapazitätsausnutzungsgrad} = \frac{\text{Ist-Produktion}}{\text{Kann-Produktion}}$$

Da üblicherweise beim Aufbau einer neuen Fabrik neue Prozesse und Geräte verwendet werden, muss für deren Beherrschung eine gewisse Lernphase berücksichtigt werden. Für die in der Halbleitertechnik inzwischen extrem kleinen Strukturen sind sehr hohe Fertigungsgenauigkeiten notwendig. Deshalb muss in diesem Bereich mit einer besonderen Sorgfalt und Exaktheit vorgegangen werden. (Abbildung 4.9).

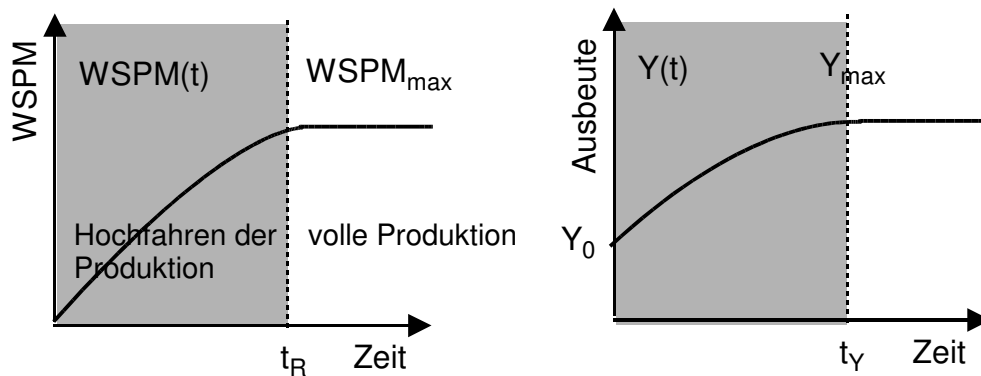


Abbildung 4.9: Ramp Up von Fertigungskapazität und -ausbeute

Ziel muss es sein, die Zeitspanne zu minimieren, die zwischen dem Produktionsbeginn und dem Erreichen der maximalen Produktionsleistung und Fertigungsausbeute liegt. Zusätzlich muss versucht werden, von einem möglichst hohen Anfangsniveau aus mit der Produktion zu beginnen. Durch schnelle Lernzyklen und ein Konzept, das alle Bereiche der Fertigung umfasst, kann ein erheblicher Wettbewerbsvorteil gegenüber anderen Firmen erreicht werden [82].

Für die aktuelle IC-Fertigung sind Zeiten von ungefähr 8 Monaten nach Beginn des Ramp Up bis zum Erreichen der vollen Fertigungskapazität und bis zum Erreichen der Endausbeute typisch [12]. Durch Prozessverbesserungen und Reduktion der Chipgröße steigt die Ausbeute allerdings auch nach dieser Zeit noch ständig leicht an.

4.6.2 Leistungs- und Kostenentwicklung

Bedingt durch die notwendigen Investitionen für die Planung, den Aufbau und die Ausstattung der Fabrik, besteht bereits bei Inbetriebnahme der Produktionslinie ein gewisses Kostenniveau. Da die ersten Kosten bereits bei Projektbeginn entstehen, kann grob gesagt werden, dass der erste Umsatz erst ungefähr 2 Jahre nach den ersten Kosten erzielt

werden kann. Zu diesen fixen Kosten, die unabhängig von der Produktionsmenge existieren, addieren sich bei Beginn der Produktion noch die variablen Kosten, die proportional zur Produktionsmenge ansteigen.

Die Leistung der Fertigung und damit der Umsatz, der mit den verkauften Chips erzielt werden kann, steigt ebenfalls langsam im Lauf der Zeit an. Bei den anfangs noch geringen Produktionsmengen kann auch nur ein geringer Umsatz erzielt werden, der zu Anfang unter den bestehenden Kosten liegen wird.

Wie in Abbildung 4.10 erkennbar ist, ergibt sich aus diesen Tatsachen die Folgerung, dass jede Fertigungslinie am Anfang, aufgrund der aufgelaufenen Vorlaufkosten, in der Verlustzone arbeitet.

Es muss also möglichst gelingen, die Anlaufverluste gering zu halten, bzw. den Umsatz so zu steigern, dass die entstandenen Anlaufverluste schnell kompensiert werden können. Besonders in der Halbleiterindustrie ist wegen des enormen Investitionsbedarfs für den Aufbau einer neuen Fertigungslinie, der im Bereich von 1-2 Mrd. \$ liegt, der damit entstehenden hohen fixen Kosten und der kurzen Lebensdauer der Produkte ein solches Vorgehen äußerst wichtig.

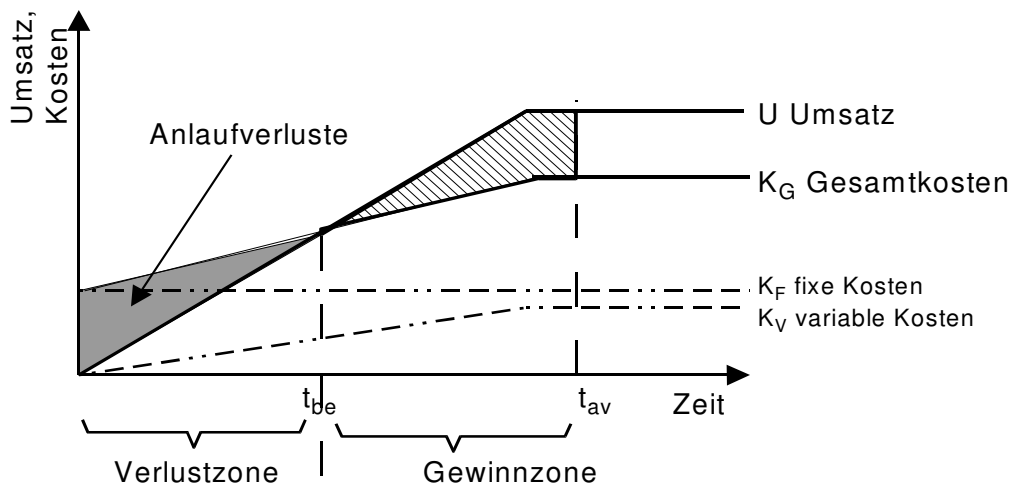


Abbildung 4.10: Zeitliche Entwicklung von Umsatz und Kosten beim Produktionsbeginn

Anzumerken ist, dass sich, aufgrund der hohen Investitionen und der damit verbundenen Abschreibungen, trotz der bestehenden Anlaufverluste ein positiver Cash Flow ergeben kann. Diese zufließenden Finanzmittel können beispielsweise für den weiteren Ausbau der Produktionskapazität durch den Kauf zusätzlicher Geräte genutzt werden (vergleichbar mit dem Lohmann-Ruchti-Effekt zur Kapazitätserweiterung [83]). Ergänzend sei an dieser Stelle allerdings noch einmal erwähnt, dass aufgrund des hohen Investitionsbedarfs in der Halbleiterindustrie ein positiver Cash Flow keine steigende Liquidität des Unternehmens bewirken muss.

Eine wichtige Marke bei der zeitlichen Entwicklung von Umsatz und Kosten stellt der Kostendeckungspunkt (Break-even-point) dar, bei dem der Umsatz den Kosten entspricht, wie in Abbildung 4.10 dargestellt [40, S. 1266].

$$\text{Umsatz}(t_{be}) = \text{Kosten}(t_{be}) \quad (4.33)$$

t_{be} = Kostendeckungspunkt (Break-even-point)

Zur Erwirtschaftung der Anlaufverluste müssen diese durch die nach dem Überschreiten des Break-even-points erreichten Gewinne im Lauf der Zeit kompensiert werden. Dieser Zeitpunkt ist erreicht, wenn gilt:

$$\int_0^{t_{av}} U(t) dt = \int_0^{t_{av}} K_G(t) dt \quad (4.34)$$

- $U(t)$ = zeitliche Entwicklung des Umsatzes
- $K_G(t)$ = zeitliche Entwicklung der Gesamtkosten
- t_{av} = Zeitpunkt, an dem die Anlaufverluste kompensiert werden

4.6.3 Entwicklung der Effizienz

Zur konkreten Berechnung der Zahl der produzierten Chips während des Hochfahrens der Produktion kann Formel (4.19) leicht abgewandelt werden. Dabei muss die zeitliche Entwicklung von Produktionsmenge und Ausbeute während der Ramp Up-Phase zusätzlich berücksichtigt werden.

Unter Voraussetzung einer linearen Steigerung der Zahl der Waferstarts ergibt sich folgende Abhängigkeit:

$$WSPM(t) = WSPM_0 + \frac{WSPM_{max} - WSPM_0}{t_R} \cdot t \quad (4.35)$$

- t = Zeit, mit $t \leq t_R$
- t_R = Zeit bis zum Erreichen von WSPMmax
- $WSPM_0$ = Zahl der Waferstarts zu Anfang
- $WSPM_{max}$ = maximale Zahl der Waferstarts

Entsprechend dem Modell von Poisson besteht für die Ausbeute eine exponentielle Abhängigkeit von der Defektdichte und Chipgröße.

$$Y(t) = \exp(-D(t) \cdot A(t)) \quad (4.36)$$

Die zeitlich veränderliche Chipgröße und Defektdichte lassen sich mit den gegebenen Größen wie folgt berechnen, wobei $t \leq t_y$ gelten soll.

$$A(t) = A_0 \cdot s^t = A_0 \cdot \left(\frac{A_{End}}{A_0} \right)^{t/t_y} \quad (4.37)$$

- t_y = Zeit bis zum Erreichen der Endausbeute
- A_0 = Anfangs-Chipgröße
- A_{End} = End-Chipgröße

$$D(t) = D_0 \cdot d^t = D_0 \cdot \left(\frac{D_{End}}{D_0} \right)^{t/t_y} \quad (4.38)$$

mit

$$D_0 = \frac{-\ln Y_0}{A_0} \quad (4.39)$$

und

$$D_{\text{End}} = \frac{-\ln Y_{\text{End}}}{A_{\text{End}}} \quad (4.40)$$

- Y_0 = Anfangs-Ausbeute
- Y_{End} = Endausbeute nach Ausbeute-Ramp Up
- D_0 = Anfangs-Defektdichte
- D_{End} = End-Defektdichte

Analog zur Situation im eingefahrenen Betrieb wurde hier ebenfalls das Ausbeutemodell von Poisson angewendet, was zu einer einheitlichen Darstellung beiträgt und auch mit den Feststellungen in [79] übereinstimmt. Damit ergibt sich der in Abbildung 4.11 dargestellte Anstieg der Ausbeute in Abhängigkeit von der Zeit.

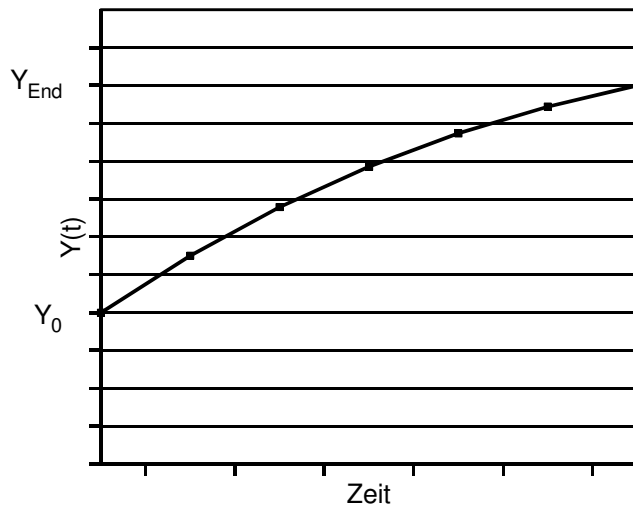


Abbildung 4.11: Ausbeuteentwicklung beim Ramp Up

Mit diesen Angaben ergibt sich ausgehend von Gleichung (4.19) für die Entwicklung der Zahl der produzierten funktionsfähigen Chips beim Hochfahren der Produktion:

$$C(t) = \left(WSPM_0 + \frac{WSPM_{\text{max}} - WSPM_0}{t_R} \cdot t \right) \cdot (1 - m) \cdot \left(\frac{\pi \cdot R_{\text{Wafer}}^2}{A_0 \cdot s^t} - \frac{\pi \cdot D_{\text{Wafer}}}{\sqrt{2 \cdot A_0 \cdot s^t}} \right) \cdot \exp(-D_0 \cdot d^t \cdot A_0 \cdot s^t) \quad (4.41)$$

Der erzielte Umsatz berechnet sich durch Multiplikation von Gleichung (4.41) mit dem zeitlich veränderlichen Preis zu:

$$U(t) = \left(WSPM_0 + \frac{WSPM_{\max} - WSPM_0}{t_R} \cdot t \right) \cdot (1-m) \cdot \left(\frac{\pi \cdot R_{\text{Wafer}}^2}{A_0 \cdot s^t} - \frac{\pi \cdot D_{\text{Wafer}}}{\sqrt{2 \cdot A_0 \cdot s^t}} \right) \cdot \exp(-D_0 \cdot d^t \cdot A_0 \cdot s^t) \cdot P_0 \cdot p^t \quad (4.42)$$

Wie in Abbildung 4.10 dargestellt, steigen die variablen Kosten in Abhängigkeit von der Produktionsmenge und damit auch die Gesamtkosten, ausgehend von den vorhandenen fixen Kosten.

$$K_G(t) = K_F + K_V(t) \quad (4.43)$$

Mit C(t) aus Gleichung (4.41) und mit (4.43) berechnet sich die Effizienz zu

$$E(t) = \left(WSPM_0 + \frac{WSPM_{\max} - WSPM_0}{t_R} \cdot t \right) \cdot (1-m) \cdot \left(\frac{\pi \cdot R_{\text{Wafer}}^2}{A_0 \cdot s^t} - \frac{\pi \cdot D_{\text{Wafer}}}{\sqrt{2 \cdot A_0 \cdot s^t}} \right) \cdot \exp(-D_0 \cdot d^t \cdot A_0 \cdot s^t) / (K_F + K_V(t)) \quad (4.44)$$

Aus (4.42) und (4.43) folgt damit für die Wirtschaftlichkeit beim Hochfahren der Produktionslinie:

$$W(t) = \frac{U(t)}{K_G(t)} \quad (4.45)$$

Auch in der Ramp Up-Phase können die Effizienz und die Wirtschaftlichkeit als Kenngröße für die Produktion benutzt werden. Ziel muss es sein, die Effizienz so schnell wie möglich zu steigern und die Phase der eingefahrenen und vollen Produktion zu erreichen.

4.7 Zusätzliche Einflussfaktoren auf die Rentabilität

Wie bereits erwähnt, wird die Rentabilität einer Fertigung nicht nur durch die Wirtschaftlichkeit und Produktivität allein bestimmt. Die im folgenden dargestellten Faktoren Zeit, Qualität und die Marktverhältnisse sind zum großen Teil mit ausschlaggebend für den ökonomischen Erfolg der Produktion.

4.7.1 Zeit

Geschwindigkeit von Verbesserungen

Im hart umkämpften Halbleitermarkt spielt vor allem der Faktor Zeit eine entscheidende Rolle. Der starke und schnelle Preisverfall im Halbleiterbereich erfordert von den Unternehmen erhebliche Anstrengungen. Es entsteht ein starker Innovationsdruck, d.h. die Unternehmen müssen neue Produkte mit verbesserten Eigenschaften (z.B. mit höherer Speicherdichte oder höherer Schaltgeschwindigkeit) möglichst früh einführen, um als einer der ersten Anbieter am Markt von den dann noch höheren Preisen profitieren zu können. Desweiteren müssen während der Lebensdauer eines Produktes starke Produktivitätssteigerungen erreicht werden, um den Preisverfall auszugleichen. Da sonst im Laufe der Zeit das Produkt nicht mehr wirtschaftlich hergestellt werden könnte, muss es gelingen, die Stückkosten laufend zu senken.

Es ist offensichtlich, dass der Vorteil einer hohen Produktivität, besonders im Bereich der Halbleiterfertigung, stark zeitabhängig ist. Je schneller die Produktivitätssteigerungen er-

reicht werden, desto größer ist der dadurch entstehende Vorteil, was in Abbildung 4.12 dargestellt ist.

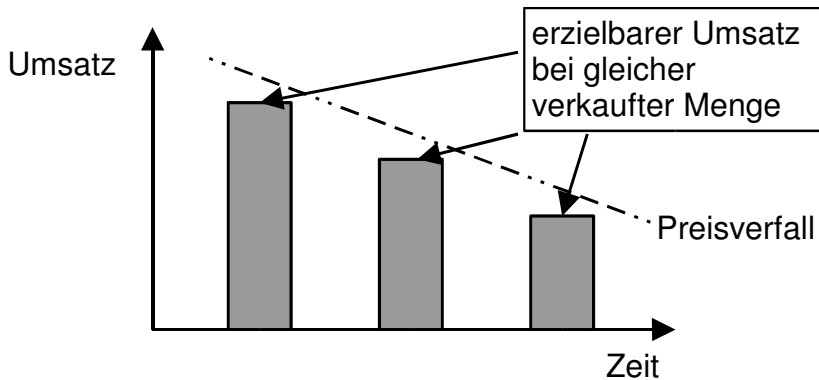


Abbildung 4.12: Höherer Umsatz durch frühzeitige Produktivitätssteigerung

Unter Berücksichtigung des zeitabhängigen Preisverfalls wird klar, dass durch eine frühzeitige Produktivitätssteigerung größere Umsatzvorteile erzielt werden können als durch die gleiche Steigerung zu einem späteren Zeitpunkt, bei dem dann nur noch ein geringerer Preis und damit für die gleiche Menge ein geringerer Umsatz erzielt werden kann.

Time to Market

Die Marktdurchdringung und der Erfolg eines Produkts unterliegen einer zeitlichen Entwicklung mit unterschiedlichen Phasen. Dieser typische klassische Lebenszyklus ist prinzipiell in Abbildung 4.13 dargestellt, wobei die Form der Kurve und die Zeitdauer der einzelnen Abschnitte je nach Produkt und Industriebranche variieren kann.

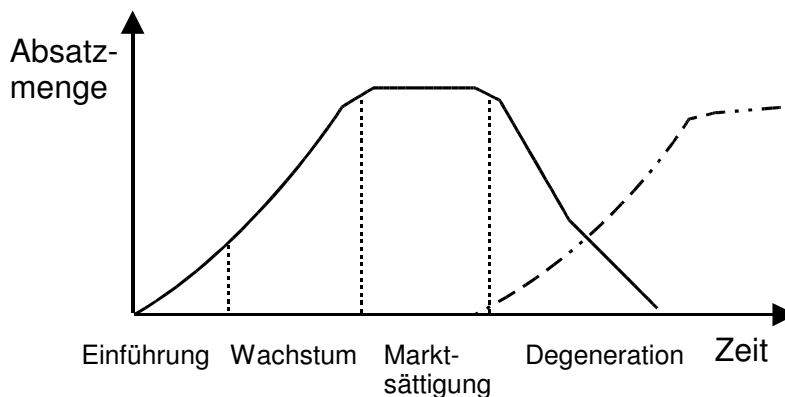


Abbildung 4.13: Lebenszyklus von Produkten

Üblicherweise wird dieser Zyklus in die Bereiche Einführung und Wachstum, mit steigender Marktdurchdringung, und Marktsättigung und Degradation, mit maximaler und danach sinkender Marktbedeutung des Produkts, eingeteilt. Im Laufe der Zeit erscheinen dann andere Produkte auf dem Markt, die das bestehende Produkt substituieren. Im sehr innovativen Bereich der Mikroelektronik ist die Produktlebensdauer, die im Bereich von 5 Jahren liegt, relativ kurz und damit die Kurve relativ schmal, da sehr schnell neue Produkte mit verbesserten Eigenschaften von mehreren Firmen auf den Markt gebracht werden.

Aus der Kurve ist offensichtlich, dass es für ein Unternehmen besonders wichtig ist, das entsprechende Produkt jeweils am Anfang seines Lebenszyklus auf den Markt zu bringen, da nur so die theoretisch erreichbare maximale Absatzmenge, also das maximale Markt-

potential, möglichst gut erreicht und ausgeschöpft werden kann. In Abbildung 4.14 ist vereinfacht dargestellt, welche Einbußen sich bei einem verspäteten Markteintritt im Vergleich zum maximal Erreichbaren (bzw. im Vergleich zu einem schnelleren Mitbewerber) ergeben können. Das Produkt kann einerseits nur für eine kürzere Zeitspanne verkauft werden und andererseits liegt der maximal erzielte Umsatz unter dem theoretisch erreichbaren. Zeitliche Verzögerungen, die sich durch Überschreitung der Entwicklungszeit oder durch auftretende technische Probleme ergeben, können also erhebliche Einbußen bedeuten.

Die Einbußen, die durch den verspäteten Markteintritt entstehen, lassen sich mit den Angaben aus Abbildung 4.14 folgendermaßen berechnen [84, S. 8], [85]:

$$L = R \cdot \frac{D \cdot (3 \cdot W - D)}{2 \cdot W^2} \quad (4.46)$$

- L = Mindereinnahmen über Produktlebensdauer
 R = Gesamt-Marktvolumen über Produktlebensdauer

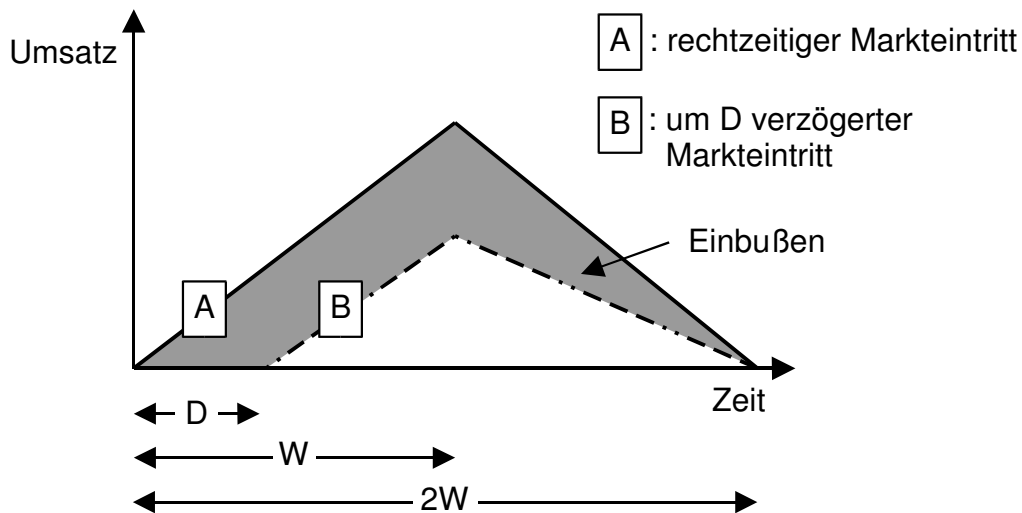


Abbildung 4.14: Einbußen durch verspäteten Markteintritt [85]

Unter Umständen können sich auch zeitlich begrenzte Monopolstellungen von einzelnen Unternehmen ergeben, wenn diese mit einem neuartigen Produkt vor allen anderen Konkurrenzunternehmen in den Markt eintreten. Eine solche Marktkonstellation hat zur Folge, dass für das entsprechende Unternehmen, neben einem Imagegewinn, eindeutige finanzielle Vorteile durch die stärkere Möglichkeit zur Preisbeeinflussung entstehen können. Da das betreffende Unternehmen die ersten Gewinne schon sehr früh erreicht und bereits Lerneffekte bei der Herstellung ausnutzen kann, besteht die Möglichkeit, bei Eintritt von Konkurrenzfirmen in den Markt, die Preise zu senken.

Ramp Up

Ein weiterer wichtiger Punkt, durch den starke Wettbewerbsvorteile erlangt werden können, ist die zeitliche Entwicklung von Produktionskapazität, Ausbeute und Kosten zu Beginn der Produktion (siehe Kapitel 4.6). Im günstigen Fall muss es gelingen, so früh wie möglich, ausgehend von einem möglichst hohen Anfangsniveau, innerhalb kurzer Zeit die maximale Produktionsmenge zu erreichen, wie dies in Abbildung 4.15 dargestellt ist, um so die entstandenen Vorlaufkosten schnell kompensieren zu können, und durch eine hohe Produktionsmenge bei den höheren Anfangspreisen größere Umsätze und höhere Marktanteile erreichen zu können. Zu diesem Zweck wird oft auch schon während des

Produktionsanlaufs eine Skalierung der Strukturgröße vorgenommen.

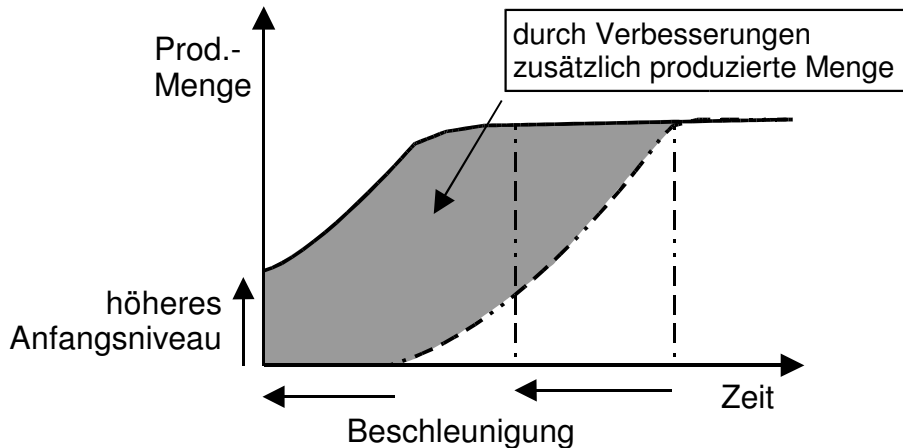


Abbildung 4.15: Verbesserungen beim Ramp Up

Zudem spielen im großen Maße Lerneffekte eine Rolle, die im Lauf der Zeit durch den Aufbau und den Betrieb mehrerer Fertigungseinrichtungen erworben werden. Der Einfluss von Lerneffekten auf die Produktionskosten wird in Kapitel 4.4 ausführlicher erläutert.

4.7.2 Qualität

Auswirkungen auf den Umsatz

Auch durch die Qualität der Fertigung kann entscheidender Einfluss auf den Erfolg des Unternehmens ausgeübt werden.

In Form der

$$\text{Fertigungsausbeute} = \frac{\text{Zahl der funktionsfähigen Teile}}{\text{Gesamtzahl der produzierten Teile}}$$

hat die Qualität direkte Auswirkungen auf die Zahl der funktionsfähigen und damit absetzbaren Produkte (siehe auch Kapitel 4.5.1). Je höher die Ausbeute, desto mehr Teile können verkauft werden, wodurch der Umsatz gesteigert wird. Im negativen Fall bei einer niedrigen Fertigungsausbeute ergibt sich ein hoher Anteil von Ausschuss. Das bedeutet, dass für die defekten Teile zwar Beschaffungs- und Herstellungskosten entstehen, aber dafür kein Erlös erzielt werden kann (Abbildung 4.16).

Bei einer Fabrik mit gleicher Fertigungskapazität aber höherer Ausbeute verteilen sich also die Kosten auf eine höhere absetzbare Stückzahl, was zu niedrigeren Stückkosten führt. Damit kann also bei gleichem Preis mehr Umsatz erzielt werden oder es können über eine attraktivere und flexiblere Preisgestaltung dem Kunden Vorteile geboten werden.

Dazu kommt, dass in einer Fabrik mit stabiler und hoher Ausbeute im gleichen Zeitraum mehr Chips hergestellt werden können, so dass eine vom Kunden gewünschte Anzahl von Teilen schneller produziert werden kann und auch in diesem Punkt schneller und damit flexibler auf den Kunden reagiert werden kann.

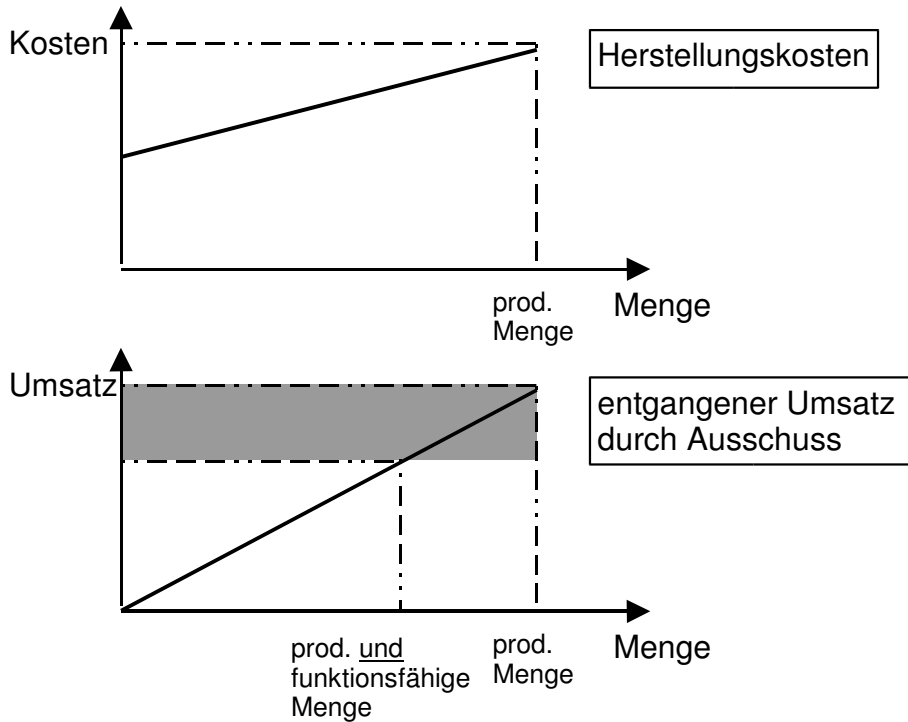


Abbildung 4.16: Einbußen durch verringerte Ausbeute

Änderungsvorgänge

Für den Fall, dass bei einem Produkt Fehler entdeckt werden, können Änderungsvorgänge notwendig werden, die sich auf die Bereiche Entwicklung und Fertigung auswirken. Solche Änderungen haben nicht nur erhöhte Kosten zur Folge, sondern auch erhebliche zeitliche Verzögerungen, was bedeuten kann, dass aufgrund des Preisverfalls mit dem modifizierten Produkt keine so hohen Umsätze mehr erzielt werden können.

Veränderungen am Schaltungsdesign, die sich auf die Fertigung auswirken, haben meistens zur Folge, dass Maskensätze für die Lithographie geändert werden müssen und Prozessschritte angepasst werden müssen. Dabei entstehen Kosten durch den zusätzlichen technischen Aufwand und den notwendigen Personalaufwand für die Fehlersuche und -behebung. Ergänzend kommt hinzu, dass bereits Kosten für die hergestellten fehlerhaften Teile entstanden sind. Unter Umständen müssen auch Aufwendungen für die Rücknahme von bereits an die Kunden ausgelieferten Teilen und sogar von ergänzenden Werbemaßnahmen berücksichtigt werden (siehe z.B. [86]).

Durch die Änderungsmaßnahmen werden ebenfalls zeitliche Verzögerungen verursacht. Die neue, geänderte Version des Chips ist erst verfügbar nach dem Zeitaufwand für Fehlersuche und -behebung, den Prozessänderungen und der Durchlaufzeit der Produktionslinie. Je nach Umfang und Komplexität der Veränderungen können auf diese Weise monatelange Verzögerungen die Folge sein.

Kundenzufriedenheit

Selbstverständlich müssen die erzeugten Produkte mit einwandfreier Funktion an die Kunden ausgeliefert werden, was bedeutet, dass Funktionstests des Herstellers an den Chips durchgeführt werden müssen. Bei schlechter Fertigungsqualität ist jedoch die Wahrscheinlichkeit größer, dass einige defekte Chips bei den Tests nicht erkannt werden und so an den Kunden geliefert werden, bzw. dass die Zuverlässigkeit der Bauelemente reduziert ist und deswegen mehr Ausfälle zu einem früheren Zeitpunkt während des Be-

triebs auftreten, was bedeutet, dass die vom Kunden geforderte maximale Ausfallrate während der Lebensdauer des Produkts eventuell nicht eingehalten werden kann.

Ebenfalls eine entscheidende Rolle spielt die Qualität im Zusammenhang mit der Kundenzufriedenheit. Falls durch Qualitätsmängel im Herstellungsprozess Lieferschwierigkeiten oder -verzögerungen auftreten, können diese für den Kunden, der zur Weiterverarbeitung auf die bestellten Bauteile angewiesen ist, zu erheblichen Schwierigkeiten führen. Im Extremfall können solche Probleme, abgesehen vom Imageverlust des Unternehmens, zu Konventionalstrafen oder zum Verlust des Kunden führen, der sich dann von einer Firma mit höherer Qualität und Zuverlässigkeit beliefern läßt.

Dass das Erkennen von Qualitätsproblemen in der Praxis problematisch ist und schwerwiegende Folgen haben kann, wird z.B. deutlich am sogenannten Pentium-Bug, einem Entwurfsfehler des Pentium II-Vorgängermodells, der erst nach Markteinführung des Prozessors entdeckt wurde und der Firma Intel neben Imageverlust einen geschätzten Verlust von 475 Millionen \$ einbrachte [87].

4.7.3 Markt

Angebot und Nachfrage

Beim Streben eines Unternehmens nach Rentabilität spielen die Markt- und Absatzchancen des jeweiligen Produkts eine wichtige Rolle. Die Halbleiterindustrie orientiert sich an weltweiten Absatzmärkten. Jedes große Unternehmen steht also im direkten Wettbewerb mit internationalen Mitbewerbern. Angebot und Nachfrage des Marktes und der Preis für ein Produkt sind nicht auf regionale oder nationale Grenzen beschränkt, sondern müssen weltweit betrachtet werden. Die Absatzchancen richten sich also auch nach der Zahl und Stärke der Konkurrenzunternehmen.

Überangebot am Markt oder zu geringe Nachfrage nach einem Produkt führen zu Preisrückgang und zu Absatzschwierigkeiten. Auch hohe Produktivität und hohe technische Kompetenz sind keine Garantie für Gewinne. Wenn ein Produkt am Markt eingeführt wird, das von den Kunden zu diesem Zeitpunkt nicht akzeptiert wird, wird sich der gewünschte unternehmerische Erfolg nicht einstellen. Von einer solchen mangelnden Nachfrage können auch hochinnovative Produkte mit sehr guter technischer Leistungsfähigkeit betroffen sein. Gründe für die schlechte Akzeptanz eines Produkts können z.B. sein, dass es zu früh oder zu spät auf den Markt gebracht wurde, oder, dass es am Trend vorbei entwickelt wurde. Als Beispiel für eine solche Entwicklung kann die Firma AMD gelten, die im Jahr 1996 500000 Prozessoren mit 100 MHz Taktfrequenz herstellte, davon aber, wegen des zu späten Markteintritts, 200000 Stück unter Herstellungskosten verkaufen und 300000 Stück verschrotten musste [88].

Ein anderer, die Absatzmenge und den Preis negativ beeinflussender Aspekt tritt auf, wenn die Zahl der auf dem Markt verfügbaren Chips die Nachfrage übersteigt. Besonders im Bereich der DRAMs ist immer wieder ein Schwanken des Marktes zwischen Überangebot und Knappheit zu verzeichnen. Anhand eines Ausschnitts aus der Preis pro Bit-Kurve für DRAMs wird in Abbildung 4.17 deutlich, welche Preisschwankungen sich durch Überkapazitäten oder durch Versorgungsengpässe ergeben können und in der Vergangenheit bereits aufgetreten sind [6, S. 7-66].

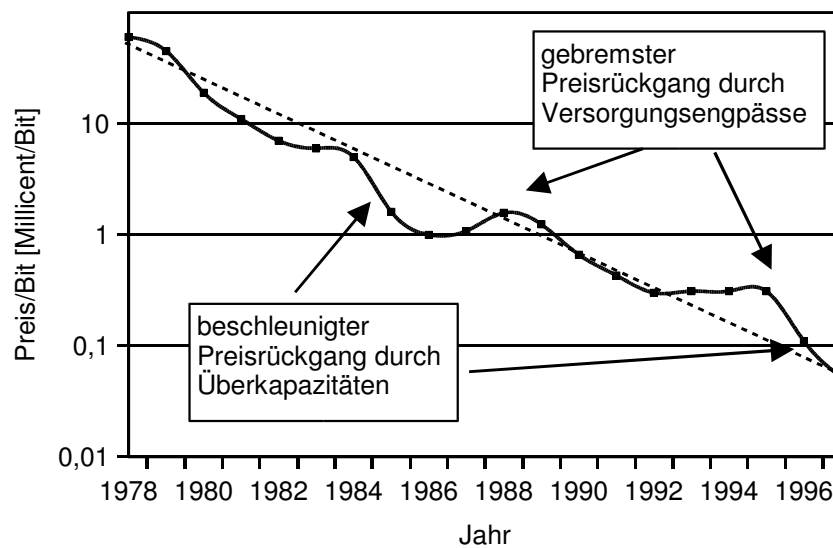


Abbildung 4.17: Einfluss des Marktes auf den Preis [6]

Image und Art des Produkts

Ebenso wichtig für die Absatzchancen und den erzielbaren Preis ist die Art des Produktes. Dabei spielt die Unterscheidbarkeit bzw. das Image eine wichtige Rolle. Während bei Speicherchips im allgemeinen kein starker Bezug zum Hersteller besteht und nur die Funktion des Chips entscheidet, ist bei Mikroprozessoren üblicherweise der Name des Herstellers für die Kaufentscheidung ausschlaggebend. Dies hat zur Folge, dass Firmen mit hohem Imagewert und bekanntem Namen, wie z.B. Intel, den Markt und den Preis zum großen Teil bestimmen, während weniger bekannte Firmen Probleme damit haben vom Markt akzeptiert zu werden. Die Bedeutung eines bekannten Markennamens unterstreicht auch der Werbeslogan „Intel Inside“, der dem Kunden signalisieren soll, dass es sich um ein bekanntes Qualitätsprodukt handelt.

In einem Bereich mit Standardprodukten, die sich nur in wenigen Eigenschaften unterscheiden, gibt es für die Unternehmen nur eingeschränkte Möglichkeiten, sich von den Konkurrenzfirmen abzuheben. Beispielsweise besteht im DRAM-Bereich praktisch nur die Unterscheidungsmöglichkeit bei Speicherkapazität und Zugriffszeit. Im Mikroprozessormarkt dagegen sind die Anforderungen an das Bauteil nicht derart festgelegt. Es ist für die Unternehmen möglich, sich in verschiedenen Leistungskriterien vom restlichen Markt zu unterscheiden. Deswegen spielt auch in diesem Marktsegment der Name und das Image des Chipherstellers eine wesentlich größere Rolle als im Speicherbereich. Die anderen bereits erwähnten Erfolgsfaktoren, wie Flexibilität oder Kundenzufriedenheit, bleiben davon jedoch unberührt.

Marktanteil

Der Marktanteil, also das Verhältnis von Absatzvolumen einer Firma zum Marktvolumen, spielt ebenfalls für die Rentabilität eine wichtige Rolle. Bei starkem Marktwachstum ist es für die Unternehmen auch bei geringen Veränderungen der Marktanteile möglich, hohe Steigerungen der Absatzmenge zu erzielen, bei gesättigten Märkten dagegen, sind größere Zuwachsraten praktisch nur noch durch Erhöhung des Marktanteils zu erzielen.

Ein weiterer Aspekt, der durch hohen Marktanteil zustande kommt, ist die Dominanz be-

stimmter Produkte oder Hersteller. Dies kann zur Folge haben, dass sich Kunden und auch andere Unternehmen an diesem Produkt orientieren und so eine Art „Quasi-Standard“ entstehen kann. Insofern führt hoher Marktanteil auch zu einem positiven Image des Herstellers. Für den Kunden liegt der Vorteil ein solches Produkt zu kaufen darin, dass er auf eine große Menge an kompatibelem Zubehör, aber auch Erfahrung zurückgreifen kann. Im Bereich der Mikroprozessoren stellt Intel das Maß der Dinge dar. Bisher sind Versuche anderer Unternehmen, nicht dazu kompatible Prozessoren auf dem Markt zu etablieren, die unterschiedliche Software benötigen, wie z.B. der Alpha-Prozessor von DEC, nicht sonderlich erfolgreich gewesen, während AMD mit Intel-kompatiblen Prozessoren in letzter Zeit Marktanteile gewinnen konnte.

5 Analyse des Einflusses der Produktionstechnik auf die Effizienz der Fertigung

5.1 Grundlagen und produktionstechnische Größen

5.1.1 Geräte für die IC-Fertigung

Mit dem Begriff Produktionstechnik sollen hier die Bereiche bezeichnet werden, die sich mit der Planung und dem Ablauf der Fertigung und deren Verbesserung befassen.

Ziel ist eine technische und ökonomische Bewertung der Auswirkungen von Änderungen ausgewählter Parameter im produktionstechnischen Bereich.

Die Fertigung von integrierten Schaltungen ist gekennzeichnet durch eine Vielzahl von Herstellungsschritten und damit auch einer großen Anzahl von Geräten (mehrere hundert pro Fabrik), die teilweise mehrfach im Verlauf des Herstellungsprozesses durchlaufen werden, was in Abbildung 5.1 beispielhaft für die Prozessschritte zur Erzeugung der Source/Drain-Struktur dargestellt ist.

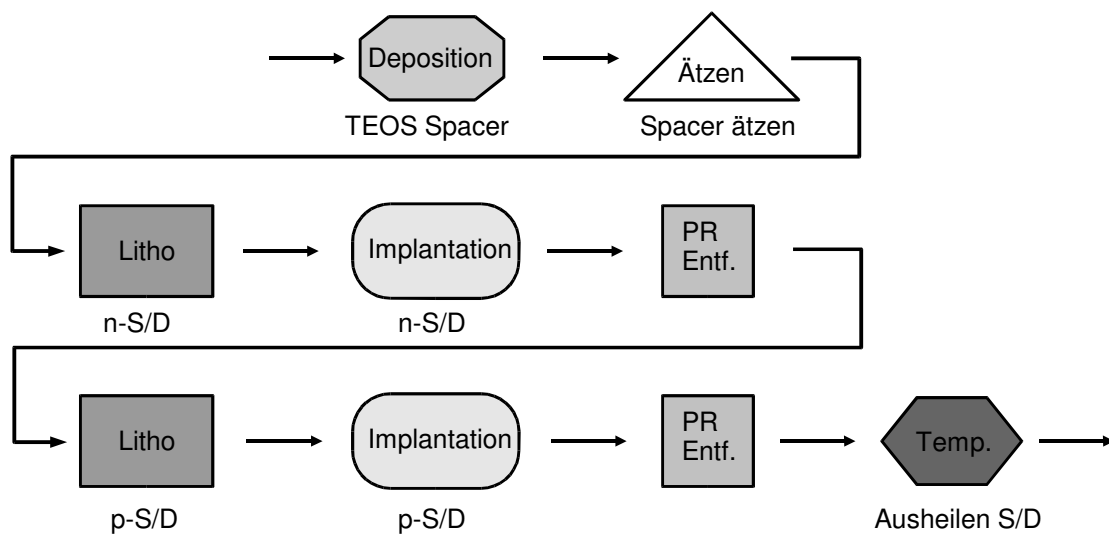


Abbildung 5.1: Beispiel für Prozesskette

Die Geräte weisen abhängig von der Art des Prozesses und der zu erzeugenden Struktur unterschiedlichen Durchsatz, d.h. unterschiedliche Produktionsgeschwindigkeit, auf. Ebenfalls kommt hinzu, dass es sich teilweise um Geräte handelt, die mehrere Wafer gleichzeitig prozessieren können (Batch-Prozess, z.B. Ofenprozesse) und teilweise um Geräte, die nur einen Wafer bearbeiten können (Single Wafer Prozess, z.B. Lithographie).

Da die Wafer üblicherweise in Losen mit 25 Wafern aufbewahrt und zwischen den Geräten transportiert werden, entstehen zwangsläufig unterschiedlich große Transport- und Wartezeiten, unter Umständen bilden sich Zwischenlager vor den Geräten. Wegen dieser Unwägbarkeiten kann eine statistische Ankunftsrate der Lose am Gerät, die um einen bestimmten Erwartungswert schwankt, angenommen werden. Die Durchlaufzeit einer solchen Produktionslinie ist, im Gegensatz zu reiner Fließfertigung, nicht 100%ig festgelegt und nicht für jedes Los konstant, sondern kann relativ stark um einen Mittelwert

schwanken.

Prinzipiell kann bei Fertigungen zwischen einem sogenannten Push- und Pull-Konzept unterschieden werden [71, S. 5-1 ff.], [109, S. 6 ff.]. Push-Systeme sind dadurch gekennzeichnet, dass die Produktion entsprechend einer Abschätzung des Bedarfs durch die eingeschleuste Menge bestimmt wird. Üblicherweise ergeben sich hier größere WIP-Bestände und Durchlaufzeiten. Im Gegensatz dazu wird bei Pull-Systemen die Produktion vom gewünschten Ergebnis ausgehend gesteuert, wodurch auf den momentanen Bedarf reagiert werden kann. Dies führt im allgemeinen zu niedrigeren WIP-Beständen und Durchlaufzeiten. In der Realität wird in der Halbleiterfertigung normalerweise kein reines Push- oder Pull-Prinzip angewendet, sondern es ergibt sich eine Mischung aus beiden Formen.

Wie in Kapitel 2.4 beschrieben, besteht die Herstellung von integrierten Schaltungen aus Hunderten von Prozessschritten, die auf mehreren hundert Geräten ausgeführt werden. Angestrebte und von SEMI prognostizierte Werte für theoretischen Durchsatz, Zuverlässigkeit und ungefähren Preis von verschiedenen Geräten bei Strukturgrößen von $0,18 \mu\text{m}$ sind in der folgenden Tabelle aufgelistet (zusammengestellt aus [25], [89], [90], [91]).

Tabelle 5.1: Gerätedaten für $0,18 \mu\text{m}$ Equipment

Gerät	theor. Durchsatz [W/h]	MTBF [h]	Preventive Maintenance [h/Woche]	Gerätekosten [\$M]
Litho Stepper	25	350	3	7,5
Litho Track	35	700	3	1,6
Asher	40	450	2	1
Plasmaätzen	20	350	7	3
Plasmaätzen Metall	25	350	7	3,1
CVD	20	275	6	3
CVD Metall	40	500	4	3
CMP	30	300	6	1,9
Ofen	20	1000	1,5	1,7
Ionenimplantation	60	200	10	2,9
RTA	40	1000	1	1,5
PVD	25	300	9	5
Reinigung	80	1000	1	2,5

MTBF = Mean Time Between Failure

Dabei ist anzumerken, dass auf der Litho-Track mehrere Prozessschritte teilweise mehrfach nacheinander ausgeführt werden. Es sind dies: Aufbringen von Photolack, thermisches Aushärten des Photolacks und Entwickeln des belichteten Photolacks. In den meisten modernen Fertigungsanlagen werden inzwischen sogenannte Photocluster eingesetzt, das sind Geräte, bei denen die Litho-Track fest mit dem Stepper verbunden ist. In diesem Fall entscheidet der langsamste Teil des Clusters über den Gesamtdurchsatz.

5.1.2 Gerätezustände

Produktionsgeräte verhalten sich im allgemeinen nicht ideal, d.h. sie können nicht 100% der Zeit für die Produktion genutzt werden, sondern benötigen beispielsweise Zeitaufwand für Wartung oder Reparaturen, wodurch die theoretisch zur Verfügung stehende Produktionszeit verringert wird.

Der Verlust an Produktionsleistung kann unterschiedlichen Gründen zugerechnet werden. Von SEMI wurden in diesem Zusammenhang sechs verschiedene Gerätezustände als Standard vorgeschlagen, die die Einordnung und Bewertung erleichtern sollen. (Abbildung 5.2) [90].

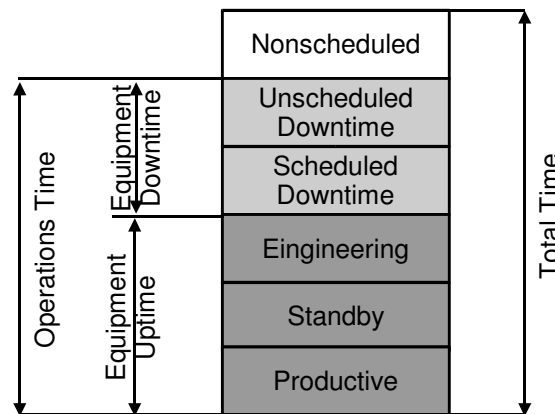


Abbildung 5.2: SEMI E10 Gerätezustände [90]

Dabei wird die Gesamtzeit (24 h/Tag) eingeteilt in „Equipment Uptime“, „Equipment Downtime“ und „Nonscheduled Time“.

Der Zustand „Nonscheduled“ kann für moderne Produktionsanlagen allerdings praktisch vernachlässigt werden, da üblicherweise die gesamte verfügbare Zeit für Produktion eingeplant ist. Die „Operations Time“, also die Zeit, die das Gerät in Betrieb ist, beträgt damit üblicherweise 24 h/Tag.

Als „Equipment Downtime“ wird die Zeit bezeichnet, in der das Gerät wegen Reparaturarbeiten, Wartung, Umrüstvorgängen, usw. nicht genutzt werden kann. Dabei wird unterteilt nach geplanter Downtime, während der z.B. regelmäßige Wartungsarbeiten durchgeführt werden, und ungeplanter Downtime, die entsteht, wenn am Gerät nicht vorhersehbare Probleme auftreten. Bekannte Größen, die die Zuverlässigkeit eines Gerätes angeben, sind die „Mean Time Between Failure“ (MTBF) und die „Mean Time Between Assist“ (MTBA), die die durchschnittliche Prozessdauer bis zum Auftreten einer Störung oder Unterbrechung angeben.

Während der „Equipment Uptime“ steht das Gerät bereit, kann aber auch dann nicht ständig produzieren, da z.B. auf Wafer gewartet werden muss (Standby). Der Anteil „Engineering“ kennzeichnet Zeiten, in denen das Gerät für Forschungs- oder Entwicklungsaufgaben genutzt wird und deswegen nicht für die normale Produktion eingesetzt werden kann. Tatsächlich für Produktion genutzt wird das Gerät nur im Zustand „Productive“. Ziel muss es sein, hier möglichst hohe Werte zu erreichen und so die theoretische Produktionskapazität des Geräts möglichst gut auszunutzen und eine möglichst hohe Produktionsmenge zu erreichen.

Für unterschiedliche Geräte können jedoch, aufgrund der verschiedenen ausgeführten Prozesse und der unterschiedlichen physikalischen Prozessabläufe, keine einheitlichen

Werte für die Nutzung erzielt werden.

Mit den Daten aus Tabelle 5.1 und [25] ergeben sich für die Geräte-Downtime und für den damit reduzierten und effektiv erreichbaren Durchsatz der Geräte die in Tabelle 5.2 dargestellten Werte.

Tabelle 5.2: Effektiver Gerätedurchsatz

Gerät	idealer Durchsatz [W/h]	Downtime [%]	effektiver Durchsatz [W/h]
Litho Stepper	25	8	23
Litho Track	35	10	32
Asher	40	10	36
Plasmaätzen	20	12	18
Plasmaätzen Metall	25	12	22
CVD	20	15	17
CVD Metall	40	15	34
CMP	30	10	27
Ofen	20	10	18
Ionenimplantation	60	15	51
RTA	40	10	36
PVD	25	15	22
Reinigung	80	12	71

Üblicherweise liegt die „Availability“, also die Zeit, in der das Gerät für die Produktion genutzt werden kann und die der „Uptime“ in Abbildung 5.2 entspricht, heutzutage bei 80-90% der Gesamtzeit. Unter Nutzungsgrad wird hier im folgenden die Zahl der tatsächlich prozessierten Wafer im Verhältnis zur theoretisch möglichen Produktionsmenge bezogen auf die Gesamtzeit bezeichnet.

5.1.3 Beschränkung der Fertigungskapazität in einer Produktionslinie

Die Fertigungskapazität der Produktionslinie wird durch die Leistungsfähigkeit der Geräte und ihre gegenseitige Abstimmung bestimmt.

Die oben erwähnten nicht-produktiven Gerätezustände führen zu einer Reduktion des Durchsatzes der Geräte und damit der gesamten Produktionslinie. Für die Festlegung der Fertigungskapazität darf nicht von dem theoretischen Durchsatzwert des Gerätes ausgegangen werden, sondern es müssen die Faktoren, die die Leistungsfähigkeit reduzieren, berücksichtigt werden. Bei den heutigen Produktionslinien wird auf diese Weise ein großer Anteil der Gerätekapazität verloren.

Großer Durchsatz von einzelnen Geräten allein garantiert jedoch noch keine große Produktionsmenge der Fertigung. Eine Produktionslinie im Halbleiterbereich besteht aus vie-

len hundert Geräten, die unterschiedliche Prozesse mit unterschiedlicher Geschwindigkeit nacheinander ausführen. Bei einer solchen Konfiguration wird sich gemäß der „Theory of Constraints“ (z.B. [92], [93]) immer ein sogenanntes „Bottleneck-Tool“ ergeben, d.h. also ein Gerät, dessen Produktionskapazität niedriger ist als die der anderen Geräte und das somit den „Flaschenhals“, also die beschränkende Komponente, der Produktion darstellt [94, S. 164 ff.]. In Abbildung 5.3 sollen diese Zusammenhänge prinzipiell verdeutlicht werden.

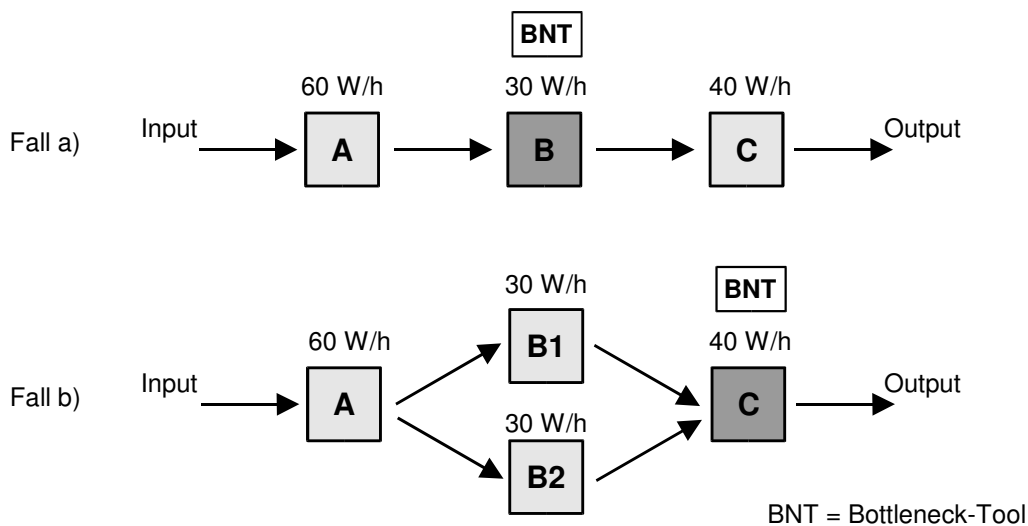


Abbildung 5.3: Beispiel für Bottleneck-Tool

In Fall a) stellt z.B. Gerät B mit 30 W/h das Bottleneck-Tool dar. Dieses Gerät limitiert also den maximal möglichen Gesamtdurchsatz der Produktionslinie auf 30 W/h. Die Input-Menge stellt auch die Output-Menge dar, da während des Produktionsablaufs keine Wafer entfernt oder hinzugefügt werden. Ein höherer Input und damit ein höherer Durchsatz ist bei dieser Gerätekonfiguration nicht möglich, da Gerät B, im Gegensatz zu den anderen Geräten, nicht mehr Teile bearbeiten kann und sich somit wiederum ein maximaler Output von 30 W/h ergeben würde, der Lagerbestand vor Gerät B bei mehr ankommenden Teilen sich jedoch immer weiter erhöhen würde. Ein solches unbegrenztes Anwachsen der Zwischenlager ist aus wirtschaftlichen und technischen Gründen nicht akzeptabel. Aus diesen Gründen begrenzt sich der Durchsatz einer Produktionslinie durch die Leistungsfähigkeit des Bottleneck-Tools.

Durch Zukauf und Parallelschalten eines weiteren Geräts vom Typ B läßt sich in Fall b) die Produktionskapazität der Fertigungslinie erhöhen. Die parallelen Geräte B1 und B2 besitzen dann insgesamt einen Durchsatz von 60 W/h, als neues Bottleneck-Tool ergibt sich dann jedoch Gerät C mit 40 W/h. Es ist damit also nur gelungen den Gesamtdurchsatz um 10 W/h zu steigern, ein Großteil der theoretischen Leistung des zusätzlichen Geräts bleibt ungenutzt.

Es zeigt sich, dass bei der Produktionsplanung eine exakte Strategie notwendig ist, um die Zahl der Geräte, die benötigte Produktionskapazität und die entstehenden Kosten aufeinander abzustimmen.

Wie an diesem Beispiel auch deutlich wird, ist das Bottleneck-Tool das Gerät mit dem höchsten resultierenden Nutzungsgrad. Da das Bottleneck-Tool die Fertigungskapazität beschränkt, weisen alle anderen Non-Bottleneck-Tools einen niedrigeren Nutzungsgrad auf. Der Nutzungsgrad der Geräte, also die Zahl der tatsächlich prozessierten Wafer be-

zogen auf ihren maximal möglichen Durchsatz, liegt heute üblicherweise durchschnittlich im Bereich von 50 - 60% [95]. Als Bewertungsgröße für die Leistungsfähigkeit der Fab ist der durchschnittliche Gerätenutzungsgrad aus den eben geschilderten Gründen nicht verwendbar [96].

Das Bottleneck-Tool kann entweder aufgrund von technischen Voraussetzungen entstehen, indem es tatsächlich das Gerät mit dem niedrigsten Durchsatz darstellt, oder es kann durch ungünstige Koordination und Organisation im Prozessablauf, z.B. durch mangelnde Versorgung mit Wafern, ein, technisch betrachtet, schnelleres Gerät so stark eingeschränkt werden, dass es zum Bottleneck-Tool wird [97]. Durch verschiedene Einflüsse kann das Bottleneck-Tool im Verlauf der Produktion wechseln, d.h., dass zu unterschiedlichen Zeiten verschiedene Geräte den beschränkenden Faktor der Produktionslinie darstellen können.

Die erläuterten Tatsachen zeigen, dass eine umfassende Beschreibung und Bewertung der Leistungsfähigkeit einer Produktionslinie kompliziert und aufwendig ist.

5.1.4 Bestehende Größen zur Bewertung der Leistungsfähigkeit

Einige Kenngrößen werden bereits jetzt in der HL-Industrie verwendet, um Bewertungsmöglichkeiten für die Qualität der Fertigung zu erreichen. Dabei ist zu beachten, dass die meisten dieser Größen nur für spezielle Zwecke und bei genau definierten Bedingungen einsetzbar sind.

CoO (Cost of Ownership)

Cost of Ownership (Einheit: \$/Wafer) stellt eine Größe dar, mit deren Hilfe der Käufer eines Gerätes Informationen über die Kosten erhält, die das jeweilige Gerät beim Kauf und während seiner Nutzung verursacht. Auf diese Art ist es möglich, die Geräte verschiedener Hersteller vor dem Kauf zu vergleichen. Die CoO-Daten werden üblicherweise vom Hersteller angegeben.

Im allgemeinen wird die Cost of Ownership eines Gerätes folgendermaßen berechnet [98].

$$COO = \frac{FC + VC + CY}{T \cdot Y \cdot U}$$

COO	= Cost of Ownership	T	= Durchsatz
FC	= fixe Kosten	Y	= Ausbeute
VC	= variable Kosten	U	= Nutzungsgrad
CY	= Kosten durch Ausbeuteverlust		

OEE (Overall Equipment Effectiveness)

Zur Bewertung der tatsächlichen Geräteleistung kann die Kenngröße OEE benutzt werden. Es handelt sich dabei um eine prozentuale Größe, die über die Angabe des einfachen Nutzungsgrades hinausgeht, und die angibt, welcher Teil der zur Verfügung stehenden Gesamtzeit tatsächlich zur Herstellung funktionsfähiger Teile aufgewendet wird. Der Schwerpunkt bei OEE liegt darauf, durch genaue Analyse des Gerätezustandes Verbesserungspotentiale aufzuzeigen, um so gezielt die Geräteeffizienz erhöhen zu können. Wie in Abbildung 5.4 dargestellt wird, wird dabei eine exakte Analyse und Einteilung der jeweiligen Gerätezeiten vorgenommen.

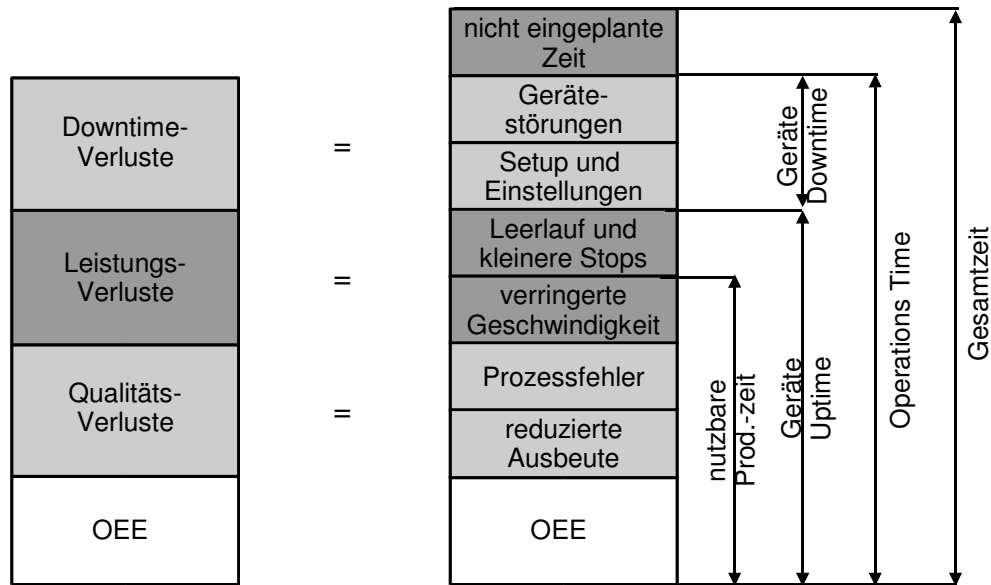


Abbildung 5.4: OEE [99]

Für die exakte Berechnung von OEE ergibt sich folgende Formel [99], [100]:

OEE = Verfügbarkeit x Leistungseffizienz x Qualitätsrate

$$\text{Verfügbarkeit} = \frac{\text{Geräte Uptime}}{\text{Gesamtzeit}}$$

$$\text{Leistungseffizienz} = \frac{\text{nutzbare Prod.-zeit}}{\text{Geräte Uptime}} \cdot \frac{\text{Prod.-menge}}{\text{theor. Prod.-menge i. nutz. Prod.-zeit}}$$

$$\text{Qualitätsrate} = \frac{\text{Prod.-menge} - \text{Ausschuss}}{\text{Prod.-menge}}$$

Vereinfacht kann OEE auch auf folgende Weise berechnet werden, dabei wird jedoch das Verbesserungspotential in den einzelnen Bereichen nicht deutlich.

$$\text{OEE} = \frac{\text{tatsächlich nutzbare Produktionsmenge}}{\text{theor. Produktionsmenge}}$$

OEE ist im allgemeinen nicht zum Vergleich verschiedener Geräte oder verschiedener Fabs geeignet, sondern gibt Aufschluss darüber, wo die momentanen Probleme des untersuchten Geräts liegen, und wie es besser ausgenutzt werden kann.

Insgesamt kann gesagt werden, dass der Vergleich unterschiedlicher Produktionen mit Hilfe der Kenngrößen OEE meist nicht möglich ist, da es sich dabei um Daten handelt, die nur im Zusammenhang mit dem entsprechenden Umfeld interpretiert werden können.

Eine Bildung von Durchschnittswerten über mehrere Geräte hinweg führt üblicherweise zu keinen nützlichen Aussagen.

5.2 Bewertungsgrößen der Fertigungslinie

Um eine Verbindung zu Kapitel 4 herstellen zu können und um den Einfluss auf die Effizienz der Fertigung beurteilen zu können, muss der Bereich Produktionstechnik nicht nur

auf seine technischen, sondern auch auf seine ökonomischen Auswirkungen hin betrachtet werden.

In diesem Zusammenhang können als wichtige Größen zur Beurteilung der Leistungsfähigkeit eines Produktionsablaufs die Parameter Produktionsmenge, Durchlaufzeit, Zahl der halbfertigen Wafer in der Produktionslinie und die entstehenden Kosten dienen. Dabei wirken sich die Anzahl der Waferstarts und die Kosten direkt auf die in Kapitel 4.5 bestimmte Wirtschaftlichkeit aus. Durchlaufzeit und WIP können durch eine ökonomische Bewertung in die Betrachtungen einbezogen werden.

5.2.1 Produktionsmenge

Die Produktionsmenge ist eine der offensichtlichsten Größen zur Bewertung einer Produktionslinie. Mit den vorhandenen Geräten sollte eine möglichst hohe Produktionsmenge erzielt werden. Durch das Erreichen dieses Ziels wird das in Kapitel 4 beschriebene ökonomische Prinzip umgesetzt.

Im allgemeinen ergeben sich bei steigender Ausstoßmenge Lerneffekte, Degression der Fixkosten und eventuell günstigere Einkaufsmöglichkeiten, was zur Folge hat, dass die Stückkosten des Produkts sinken. Bei Massenprodukten mit hohem Preisdruck ist es also vernünftig eine große Produktionsmenge anzustreben. Zur Zeit werden bei Hochvolumen-Produkten 20000 WSPM als die Produktionsmenge angesehen, ab der eine wirtschaftliche Fertigung möglich ist [101]. Kleinere Fabs werden für Massenproduktion inzwischen nicht mehr geplant.

Eine Änderung der Produktionsmenge hat unmittelbare Auswirkungen auf die Zahl der produzierten Chips und damit auch auf die Effizienz der Fertigung, die in Kapitel 4.5 beschrieben ist.

Die produzierte Menge richtet sich nach der Anzahl der pro Zeiteinheit in die Produktionslinie eingeschleusten Wafer (WSPM). Zur Steigerung der Produktionsmenge genügt es allerdings nicht, einfach die Zahl der Waferstarts zu erhöhen, da dies aufgrund der Bottleneck-Problematik meist nur zu erhöhtem Lagerbestand führen würde, sondern es müssen umfangreichere Verbesserungen vorgenommen werden.

Die maximale Produktionsmenge einer Fertigungslinie richtet sich, wie oben beschrieben, nach der Produktionskapazität und dem Nutzungsgrad des Bottleneck-Tools, wobei der Nutzungsgrad das Verhältnis von tatsächlich genutzter zu theoretisch möglicher Gerätekapazität darstellt.

$$WS = \text{cap}_{\text{BN}} \cdot u_{\text{BN}} \quad (5.1)$$

WS = Zahl der Waferstarts
 cap_{BN} = Fertigungskapazität Bottleneck-Tool
 u_{BN} = Nutzungsgrad Bottleneck-Tool

Durch Maßnahmen im Bereich der Produktionstechnik kann auf die entsprechenden Faktoren Einfluss genommen werden. Zur Ermittlung der tatsächlichen Produktionsmenge wird allerdings nicht nur die Kapazität des Bottleneck-Tools, sondern auch die resultierende Durchlaufzeit der Produktionslinie berücksichtigt, wie dies in der Betriebskennlinie in Kapitel 6.2 dargestellt wird.

5.2.2 Durchlaufzeit

Die Gesamt-Durchlaufzeit einer Fertigungslinie gibt die Zeit an, die ein Wafer benötigt, um die Produktionslinie von Anfang bis Ende zu durchlaufen und vollständig prozessiert zu werden. Die Durchlaufzeit berechnet sich aus der Summe der reinen Prozesszeiten des Wafers sowie aller Wartezeiten an den jeweiligen Geräten und Transportzeiten (inklusive Be- und Entladezeiten) zu den Geräten (Abbildung 5.5).

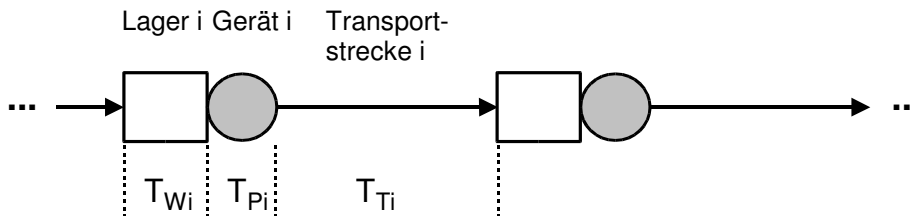


Abbildung 5.5: Bestandteile der Gesamt-Durchlaufzeit

$$P = \sum_{i=1}^n T_{Wi} + T_{Pi} + T_{Ti} \quad (5.2)$$

- TD_{ges} = Gesamtdurchlaufzeit
- T_{Wi} = Wartezeit am Gerät i
- T_{Pi} = Prozesszeit am Gerät i
- T_{Ti} = Transportzeit zum Gerät i
- n = Zahl der zu durchlaufenden Prozesse

Die tatsächlich erzielte Durchlaufzeit wird größer sein als die theoretisch erreichbare minimale Durchlaufzeit, die sich aus der Summe der reinen Prozesszeiten ergibt [102]. In diesem Zusammenhang wird auch oft das Verhältnis von tatsächlicher zu theoretischer Durchlaufzeit als Maß benutzt und als x-Factor [103] oder Cycle Time Multiplier [104], [105] bezeichnet. Kleine Werte dieser Größe bedeuten, dass in der Fertigung zusätzlich zu den reinen Prozesszeiten kein großer Zeitaufwand entsteht.

Eine geringe Durchlaufzeit der Produkte in der Fertigung ist grundsätzlich positiv zu bewerten. Vorteilhaft ist, dass Produkte schneller hergestellt werden können und damit schneller und flexibler auf Kundenwünsche reagiert werden kann, da dem Kunden die gewünschten Produkte innerhalb einer kürzeren „Reaktionsfrist“ zur Verfügung gestellt werden können. In diesem Sinne kann eine kurze Durchlaufzeit auch als Argument zur Kundenwerbung beurteilt werden.

Zudem können Änderungswünsche der Kunden bei den gewünschten Erzeugnissen schneller umgesetzt werden.

Außerdem ist es bei einer kurzen Durchlaufzeit möglich, die Effekte von Prozessänderungen und Prozessabweichungen schneller festzustellen, da eine nicht so lange Zeit abgewartet werden muss, bis der Einfluss dieser Veränderungen auf das Verhalten des fertigen Bauelements deutlich wird. Bei schnelleren Durchlaufzeiten können mehr Lernzyklen durchgeführt werden und es kann schneller korrigierend eingegriffen werden. In einigen Veröffentlichungen wird deshalb auch darauf hingewiesen, dass eine geringe Durchlaufzeit bei der IC-Fertigung positive Auswirkungen auf die Fertigungsausbeute hat. Als weiteres Argument dafür wird genannt, dass die Wafer sich kürzere Zeit im Produktionsablauf befinden und deswegen nicht so lange möglichen Partikel ausgesetzt sind, die die Ausbeute verringern können [106], [107]. Dazu kann allerdings angemerkt werden, dass aus-

beutevermindernde Effekte durch den Reinraum heute nicht mehr so stark wie früher ausgeprägt sind.

Bei einer kürzeren Produktionszeit können auch die Erlöse für die verkauften Produkte schneller erzielt werden und somit die Aufwendungen für Beschaffung und Herstellung der Teile schneller wieder eingebracht werden.

Insofern ist es im allgemeinen technisch und wirtschaftlich sinnvoll, eine kurze Durchlaufzeit anzustreben.

Die Durchlaufzeit einer Fertigung ist eng mit der Zahl der halbfertigen Produkte in der Fertigungslinie verknüpft. Ein großer WIP-Bestand führt aufgrund der entstehenden Wartezeiten gleichzeitig zu längeren Durchlaufzeiten. Beide Größen müssen deswegen gleichzeitig betrachtet werden.

5.2.3 Zahl der halbfertigen Produkte in der Fertigungslinie

Einen wichtigen Punkt stellt der WIP-Bestand (WIP = Work In Process), also die Anzahl der halbfertigen Wafer in der Fertigungslinie dar. Bei jeder Produktion existiert eine gewisse Menge an teilprozessierten Produkten, die sich in einem unterschiedlichen Grad der Fertigstellung befinden und noch mehrere Prozessschritte durchlaufen müssen. Aufgrund der hohen Anzahl von Einzelschritten bei der Herstellung von integrierten Schaltungen, wird sich hier ein hoher WIP-Bestand ergeben.

Da es sich bei WIP um gebundenes Kapital handelt und durch die Waferlose Lagerkapazität in der Fab belegt wird, wird es das Ziel jeder Fertigung sein, möglichst geringe WIP-Werte zu erreichen. Eine vollständige Vermeidung von WIP ist jedoch nicht möglich, da für den reibungslosen Ablauf der Produktion eine ausreichende Anzahl von Losen in der Produktionslinie vorhanden sein muss.

Ein Los mit 25 Wafern enthält z.B. bei 400 Chips pro Wafer insgesamt 10000 Chips, kann also je nach Produkt einen extrem hohen Wert darstellen. Dazu kommt, dass - bei einfachster Betrachtung - zum reibungslosen Ablauf der Fertigung an jedem Gerät ein Los in Bearbeitung sein sollte und zusätzlich mindestens ein Los zur Sicherheit im Lager zur Verfügung stehen sollte. Das bedeutet, dass sich bei einer großen Fertigungslinie mit vielen Geräten ein sehr hoher WIP-Bestand ergeben wird. Beispielsweise würden sich bei dieser einfachen Betrachtung bei einer Gesamtzahl von 300 Geräten insgesamt 600 Lose mit 15000 Wafern und 6 Mio. Chips ergeben. Besonders beim Übergang zu größeren Waferdurchmessern wird die Zahl der Chips, der Wert des Bestandes und damit die Bedeutung dieses Punktes immer größer.

Der mittlere Gesamtwert der Wafer, die sich in der Produktionslinie befinden, ergibt sich aus ihrer Anzahl und ihrem Durchschnittswert. Der Durchschnittswert liegt dabei, je nach Grad der Fertigstellung, zwischen den Einkaufskosten des Wafers, die von Anfang an vorhanden sind, und den kompletten Herstellungskosten des fertigprozessierten Wafers. Der Wert des einzelnen Wafers steigert sich also mit jedem Prozessschritt.

Für den mittleren Wert, den der WIP-Bestand darstellt, ergibt sich damit:

$$W_{\text{WIP}} = \text{WIP} \cdot W_{\text{Wafer}} \quad (5.3)$$

W_{WIP} = Wert des WIP-Bestandes

WIP = Zahl der Wafer im Fertigungsablauf

W_{Wafer} = Mittelwert des Wafers während der Produktion

mit dem Mittelwert des Wafers

$$W_{\text{Wafer}} = \frac{P_{\text{Wafer}} + W_{\text{Final}}}{2} \quad (5.4)$$

P_{Wafer} = Einkaufspreis des Wafers

W_{Final} = Wert des Wafers (Herstellungskosten) nach allen Herstellungsschritten

Die durch WIP und Durchlaufzeit entstehende finanzielle Belastung kann durch die Verzinsung des gebundenen Kapitals berechnet werden.

$$K_{\text{WIP}} = \text{WIP} \cdot W_{\text{Wafer}} \cdot T_{\text{D}} \cdot \frac{z}{360} \quad (5.5)$$

K_{WIP} = Kapitalverlust durch gebundenes Kapital

T_{D} = Durchlaufzeit [Tage]

z = Zinssatz p.a.

Der Grund, warum eine Verzinsung für WIP und Durchlaufzeit berechnet werden kann, liegt darin, dass für die Produktionsprozesse, bevor der Verkaufspreis erzielt werden kann, Rohstoffe und Verbrauchsmaterialien aufgewendet werden müssen, die unter Umständen zwischenfinanziert werden müssen, so dass sich abhängig von der Menge der vorhandenen halbfertigen Produkte unterschiedlich hohe finanzielle Belastungen ergeben.

Ebenfalls kann argumentiert werden, dass bei kürzerer Durchlaufzeit früher der Marktpreis erzielt werden kann, wobei bei längerer Durchlaufzeit dieser finanzielle Vorteil vorfinanziert werden müsste.

Durch die ökonomische Bewertung der Durchlaufzeit und der Zahl der halbfertigen Wafer in der Produktionslinie ist es möglich, ihren Einfluss auf die Effizienz der Fertigung zu beurteilen. Wegen der zusätzlichen finanziellen Aufwendungen, wirkt sich eine langsame Fertigung mit hohem WIP-Bestand negativ auf die Wirtschaftlichkeit aus.

Durch die Überlegungen zur Effizienz ist es auch möglich, die durch WIP und hohe Durchlaufzeit entstehenden finanziellen Nachteile im Zusammenhang mit der damit verbundenen Produktionsmenge und dem daraus erzielten Umsatz zu betrachten.

5.2.4 Kosten

Da das Ziel ist, eine möglichst hohe Wirtschaftlichkeit der Fertigung zu erreichen, muss sich insgesamt ein günstiges Verhältnis von Leistung zu Kosten für die entsprechende Auslegung der Produktionslinie ergeben. Eine Verbesserung der Bewertungsgrößen Produktionsmenge, Produktionszeit und WIP muss immer im Zusammenhang mit möglicherweise dabei entstehenden Kostenänderungen gesehen werden. Insofern stellen die Kosten eine Größe dar, die unbedingt mit betrachtet werden muss.

Kostenänderungen, die im Bereich Produktionstechnik entstehen, können sich auf die fixen oder variablen Kosten der Fertigung auswirken. Wie oben erläutert, stellen auch Durchlaufzeit und WIP einen Kostenaufwand dar.

5.2.5 Konkurrierende Ziele bei den Bewertungsgrößen

Ziel ist es, eine optimale Abstimmung zwischen den teilweise voneinander abhängenden Bewertungsgrößen zu finden.

Idealvorstellung wäre eine Produktion mit möglichst hoher Produktionsmenge, kurzer Durchlaufzeit, geringen Zwischenlagern und niedrigen Kosten, d.h. eine möglichst exakte Erfüllung der Pyramidenform in Abbildung 5.6 (Darstellung angelehnt an Atherton [94, S. 200]).

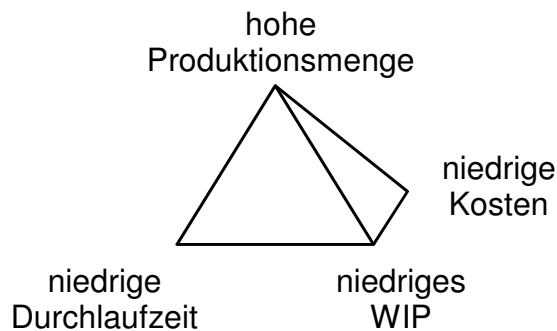


Abbildung 5.6: Abwägung zwischen den Bewertungsgrößen

Diese Ziele widersprechen sich jedoch in weiten Bereichen. Beispielsweise geht eine hohe Produktionsmenge üblicherweise mit einer höheren Durchlaufzeit und größerem Lagerbestand einher, was in Kapitel 6 erläutert wird. Im allgemeinen verursacht eine höhere Produktionsmenge auch höhere Kosten. Die Zielsetzung einer geringstmöglichen Durchlaufzeit bedeutet, dass meistens nur eine geringere Produktionsmenge erzielt werden kann.

An diesen Beispielen wird deutlich, dass für den jeweiligen Anwendungsfall ein möglichst guter Kompromiss zwischen den einzelnen Bewertungsgrößen gefunden werden muss. Die tatsächliche Lösung wird sich also im Innern der Pyramide aus Abbildung 5.6 befinden und im allgemeinen nicht die extremen Kanten erreichen.

Einflussgrößen

Die Produktion wird von einer Vielzahl von Parametern beeinflusst. Wichtige Einflussgrößen, die im weiteren betrachtet werden sollen, stellen dabei die folgenden Punkte dar, zwischen denen teilweise enge Zusammenhänge bestehen:

- WSPM
- Gerätedurchsatz
- Gerätezuverlässigkeit (MTBF, Reparaturzeit, Wartungszeit)
- Zahl der Prozessschritte
- Zahl der Geräte
- Fab-Layout
- Transportzeiten
- Losgröße

- Personalbestand
- Art des Produkts

Um den Einfluss des Bereichs Produktionstechnik auf die Effizienz der Fertigung feststellen zu können, muss es gelingen, die gegenseitigen Abhängigkeiten festzustellen und danach Veränderungen technisch und ökonomisch zu bewerten.

5.3 Auswirkungen auf die Effizienz der Fertigung

Variationen der Einflussparameter wirken sich durch die Veränderung der in Kapitel 5.2 angegebenen Bewertungsgrößen auf die Effizienz der Fertigung aus.

Im folgenden soll dargestellt werden, welche Werte für die einzelnen betrachteten Einflussfaktoren bereits erreicht wurden, welches Verbesserungspotential jeweils noch besteht und welcher Kostenaufwand bei den jeweiligen Änderungen entsteht.

5.3.1 Verbesserungspotential der einzelnen Faktoren

- Gerätedurchsatz

Aufgrund der unterschiedlichen physikalischen Prinzipien, die den Prozessen zugrunde liegen und der unterschiedlichen zu erzeugenden Strukturen, kann bei den verschiedenen Geräten kein einheitlicher Durchsatz erzielt werden. Die physikalischen Prozesshintergründe begrenzen im Endeffekt auch den maximal erreichbaren Durchsatz bei einem Gerät. Zusätzlich hängt der Durchsatz bei einigen Prozessen vom Waferdurchmesser ab, so dass für 300 mm-Wafer zu Beginn teilweise mit einem niedrigeren Gerätedurchsatz zu rechnen war als bei 200 mm-Wafern [125].

Die theoretischen, maximalen Durchsatzwerte für Halbleiterprozesse liegen im Bereich von ungefähr 30 W/h bis über 100 W/h [25]. Aufgrund der Bottleneck-Problematik und durch Stillstandszeiten der Geräte ist allerdings eine vollständige Ausnutzung der Produktionskapazität meistens nicht möglich. Ein erhöhter Gerätedurchsatz wirkt sich nur dann positiv aus, wenn davon auch das Bottleneck-Gerät, das die Leistung der Produktionslinie bestimmt, betroffen ist. Insofern werden sich die stärksten Bemühungen auf dieses Gerät konzentrieren. Da das Bottleneck-Gerät aber wechseln kann und von verschiedenen Einflüssen abhängig ist, ist es jedoch sinnvoll und notwendig, an allen Geräten nach Verbesserungsmöglichkeiten zu suchen.

Eine zweite Möglichkeit zur Steigerung des Gesamtdurchsatzes besteht in der Parallelschaltung von mehreren Geräten. Dadurch kann theoretisch auch ein beliebiger höherer Durchsatz erzielt werden, ohne die Notwendigkeit zur technischen Verbesserung der Geschwindigkeit einzelner Geräte.

- Gerätezuverlässigkeit

Die tatsächliche Geräteleistung wird, neben dem theoretischen Gerätedurchsatz, zu einem großen Teil von der Gerätezuverlässigkeit bestimmt. Durch geplante und ungeplante Downtime reduziert sich die für die Produktion nutzbare Zeit des Geräts. Eine Verbesserung in diesem Bereich hat also die gleichen positiven Auswirkungen wie eine Steigerung des Durchsatzes. Hauptaugenmerk der Entwicklung liegt derzeit im Bereich der MTBF, also der mittleren zu erwartenden Zeit bis zum Ausfall des Gerätes und im Bereich der „preventive maintenance“, das ist der Aufwand, der zur vorbeugenden Instandhaltung der Geräte notwendig ist.

Typische MTBF-Werte liegen momentan bei einigen hundert Stunden. Da sich die Ge-

samt-Stillstandszeit eines Gerätes aus der Ausfallshäufigkeit und der Reparaturdauer ergibt, bestehen zusätzlich zur Verbesserung der Gerätezuverlässigkeit auch Verbesserungspotentiale bei der Dauer von Reparatur- und Servicearbeiten.

Die regelmäßige Wartung und die ungeplante Downtime der Geräte nehmen zur Zeit durchschnittlich 10-15% der Gesamtzeit ein [89]. Obwohl das Idealziel eines fehler- und wartungsfreien Geräts nicht zu erzielen ist, bestehen in diesem Bereich noch große Verbesserungsmöglichkeiten.

- Zahl der Prozessschritte

Die Zahl der Prozessschritte liegt in der IC-Fertigung sehr hoch (300 - 400, teilweise auch bis zu 600 Einzelschritte). Durch besseres Prozessverständnis und durch Prozessänderungen ist es in den letzten Jahren gelungen, in einigen Bereichen Vereinfachungen zu erzielen und damit die Zahl der Prozessschritte nicht weiter anwachsen zu lassen und teilweise sogar zu verringern. Auch für die Zukunft besteht in diesem Bereich noch ein großes Verbesserungspotential, obwohl die steigende Anzahl von Metallisierungsebenen gegenläufige Auswirkungen mit sich bringt.

Je weniger Prozessschritte durchlaufen werden müssen, desto günstiger sind die Auswirkungen für die Durchlaufzeit und die Zahl der halbfertigen Wafer im Produktionsablauf, da neben den eigentlichen Prozesszeiten auch Warte- und Transportzeiten für die eingesparten Schritte wegfallen.

Ebenfalls positive Effekte sind für die Fertigungsausbeute zu erwarten, da im Normalfall bei keinem Einzelprozess eine ideale Ausbeute von 100% zu erreichen ist und durch eine Verringerung der Zahl der Prozessschritte insgesamt weniger Fehlerquellen bestehen.

Näherungsweise kann dieser Zusammenhang durch die Formel $Y = \prod_{i=1}^n Y_i$ [126, S. 14] angegeben werden, wobei Y die Gesamtausbeute, Y_i die Einzelschrittausbeute und n die Zahl der Prozessschritte darstellen. Daraus ergibt sich, dass bei gleicher Einzelschrittausbeute die Gesamtausbeute steigt, wenn die Zahl der Herstellungsschritte sinkt.

- Zahl der Geräte

Die Zahl der notwendigen Geräte ist eng mit der Zahl der Prozessschritte und dem Fab-Layout verknüpft. Durch Verbesserungen in diesen Bereichen ist eine Reduzierung der Zahl der benötigten Geräte möglich. Ebenfalls tragen ein gesteigerter Gerätedurchsatz und eine verbesserte Zuverlässigkeit dazu bei, dass die Anzahl der benötigten Geräte sinkt.

Der Gerätebestand in einer durchschnittlichen Fab liegt momentan im Bereich von 200 - 300 Stück, teilweise auch darüber [89].

- Fab-Layout

Durch das Layout der Fertigung, also durch die Anordnung der Geräte, kann ebenfalls Einfluss auf die Zahl der notwendigen Geräte und auf die Durchlaufzeit genommen werden. Durch eine andere Gruppierung der Geräte oder auch den Einsatz von Clustertools und Minienvironments sind weitere Verbesserungen möglich, wobei allerdings besonders bei Clustertools hoher Wert auf Zuverlässigkeit gelegt werden muss, da beispielsweise beim Ausfall einer Prozesskammer auch die anderen nicht genutzt werden können.

In [38] und [41] durchgeführte Simulationen zeigen Verbesserungsmöglichkeiten im Bereich von 10 - 14% bei der Durchlaufzeit für unterschiedliche Fab-Layouts.

- Transportzeiten

Bedingt durch die starke räumliche Verteilung der Geräte in der Fabrik und die daraus resultierenden unterschiedlich langen Transportwege, ergeben sich für die einzelnen Prozessschritte keine einheitlichen Transportzeiten. Zu dieser Tatsache trägt auch bei, dass noch keine einheitliche und vollständige Automatisierung innerhalb der Fab existiert, durch die ein sofortiger und gleichmäßig schneller Transport inklusive Be- und Entladung der Wafer möglich wäre, sondern viele dieser Tätigkeiten noch manuell vom Bedienpersonal vorgenommen werden, das für mehrere Geräte gleichzeitig zuständig ist. Durch eine durchgehende Automatisierung und verbesserte Logistikkonzepte würde es gelingen, die bestehenden Schwankungen zu reduzieren und die Transportzeiten im Mittel zu verringern. Eine vollständige Automatisierung von Wafertransport und -handling kann sich nach [127] auch positiv auf die Ausbeute auswirken.

In Anbetracht der Tatsache, dass in der Halbleiterfertigung sehr viele Transportschritte vorgenommen werden müssen, besteht in diesem Bereich noch großes Verbesserungspotential.

- Losgröße

Die momentan übliche Losgröße beträgt 25 Wafer. Im Zuge des Übergangs zu 300 mm-Wafern wurde auch die Benutzung von Losen mit 13 Wafern diskutiert. In der Praxis wird aber nun doch offensichtlich eine Losgröße von 25 Wafern beibehalten. Single-Wafer Transport ist momentan nicht vorgesehen, obwohl auch dafür bereits Vorschläge existieren.

Durch die Veränderung der Losgröße ergeben sich teilweise starke Änderungen bei Durchlaufzeit und WIP der Produktionslinie. Da die einzelnen Wafer bei Single Wafer Prozessierung warten müssen bis ein vollständiges Los prozessiert ist, führen kleinere Losgrößen zu geringeren Gesamtdurchlaufzeiten und damit auch zu geringerem WIP-Bestand [128], eine Tatsache, die mit steigendem Waferdurchmesser, steigendem Wert der Lose und zunehmender Verwendung von Single Wafer Prozessen immer größere Bedeutung erlangt. Je größer der Waferdurchmesser, desto mehr Chips pro Wafer können produziert werden und desto größer ist der Wert eines einzelnen Wafers. Das heißt, dass bei sehr großen Waferdurchmessern (in der SIA-Roadmap werden ab 2008 450 mm und in einigen anderen Veröffentlichungen für die Zeit danach sogar 675 mm vorhergesagt) die Beschädigung eines Loses bzw. eine fehlerhafte Prozessierung immer gravierendere Auswirkungen haben würde. Prozessfehler könnten bei größeren Losgrößen dann nur nach einer sehr großen Anzahl hergestellter Chips festgestellt werden. Besonders für sehr flexible ASIC-Fertigungen mit kleineren Stückzahlen erscheint eine Umstellung der Losgröße in der Zukunft damit am ehesten vorstellbar. Bei Massenprodukten, wie z.B. Speicherchips, kommt solchen Überlegungen momentan keine so hohe Priorität zu.

Momentan ist eine Umstellung auf Einzelwafertransport aus technischen Gründen jedoch noch nicht möglich. Um solche Änderungen umsetzen zu können, wären in weiten Bereichen neue Maschinen und Prozesse notwendig. Darüber hinaus können sich hohe Rüstzeiten an den Einzelscheibenanlagen ergeben. Auch aufgrund der notwendigen Standardisierung der Geräteschnittstellen und der Transportsysteme, ist eine beliebige Veränderung der Losgröße jedoch nicht möglich, ein solcher Schritt würde eine langfristige Planung und Koordination, auch mit den Geräteherstellern, erforderlich machen.

- Personalbestand

Der Personalbestand und damit die Personalkosten spielen in der Halbleiterindustrie keine so extreme Rolle wie in anderen sehr personalintensiven Branchen. Die Anzahl der

Mitarbeiter in einer herkömmlichen Fab liegt bei einer Schwankungsbreite von einigen hundert Mitarbeitern im Bereich von 800-1000 (z.B. [129], [130]). Trotzdem ist aufgrund von Sicherheits- und Reinheitsanforderungen im Reinraum mittelfristig mit einer weiter zunehmenden Automatisierung und damit auch einer Reduktion des Personals in der Produktionslinie zu rechnen. Ziel ist eine Steigerung der Personaleffizienz, also der Zahl der gefertigten Teil bezogen auf den Personalbestand.

Potential für Produktionsverbesserungen durch bessere Organisation (z.B. durch speziell aufgestellte Teams [131]) oder Schulung der Mitarbeiter besteht. Nicht unterschätzt werden darf dabei der Einfluss des Personals auf die Beschleunigung von Lerneffekten und damit auf die Flexibilität und die Geschwindigkeit mit der Neuerungen in der Fertigung umgesetzt werden können [132].

- Produkt

Die Art des Produkts bzw. der Produktmix hat ebenfalls starke Auswirkungen auf die Fertigungsgestaltung. Je mehr Produkte gleichzeitig hergestellt werden, desto aufwendiger ist die Koordination der Fertigung. Eine große Anzahl von verschiedenen Produkten ist vor allem bei der ASIC-Fertigung notwendig.

Die Art und Zahl der gefertigten Produkte stellt eine strategische Entscheidung dar, die während des Betriebes nicht beliebig verändert werden kann.

5.3.2 Aufwand für die Veränderung der einzelnen Einflussfaktoren

Zusätzlich zum Verbesserungspotential der einzelnen Einflussfaktoren muss beachtet werden, welche Kosten und welcher zusätzliche Aufwand bei den jeweiligen Änderungen entstehen.

- Gerätedurchsatz

Geräte mit höherem Durchsatz auf den Markt zu bringen ist die Aufgabe der jeweiligen Gerätehersteller, die teilweise gemeinsam mit den Chipproduzenten neue Prozesse und Geräte entwickeln. Im Zuge dieser Verbesserungen sind bei den meisten Geräten Preis-erhöhungen unumgänglich.

Bei Parallelschaltung von mehreren Geräten zur Steigerung des Gesamtdurchsatzes entsteht, entsprechend der Zahl der Geräte, ein Vielfaches der Kosten für ein einzelnes Gerät, wobei die Gerätepreise inzwischen im Bereich von einigen Millionen \$ liegen. Zusätzlich entsteht höherer Platzbedarf im Reinraum.

Positive Auswirkung einer Steigerung des Gerätedurchsatzes ist die Erhöhung der Produktionsgeschwindigkeit und damit auch der Produktionsmenge, so dass durch eine solche Maßnahme einerseits zwar höhere Kosten für Verbrauchsmaterialien entstehen, andererseits aber durch die höhere Produktionsmenge auch ein gesteigerter Umsatz erzielt werden kann.

Obwohl die größten Verbesserungen durch die Steigerung des Durchsatzes beim Bottleneck-Gerät erzielt werden können, ist es doch insgesamt sinnvoll, Aufwand für Verbesserungen an allen Geräten zu investieren, um die Fab als Ganzes voranzubringen.

- Gerätezuverlässigkeit

Gerätezuverlässigkeit und geringer zeitlicher Aufwand für die Gerätewartung muss ebenfalls vom Gerätehersteller gewährleistet werden. Für den IC-Hersteller verringert sich dadurch der Aufwand für Reparatur- und Servicearbeiten, eventuell auch der Personalaufwand.

Durch eine höhere Gerätezuverlässigkeit wird auch der effektive erreichbare Geräte-durchsatz erhöht, so dass sich eine bessere Ausnutzung des eingesetzten Kapitals ergibt.

- Zahl der Prozessschritte

Voraussetzung für eine Reduzierung der Zahl der Prozessschritte ist gutes Prozessverständnis und gute Prozessbeherrschung. Eine Verringerung der Zahl der Prozessschritte bedeutet, dass weniger Geräte benötigt werden, um das Produkt herzustellen, was geringere Investitionen nach sich zieht. Zum anderen verringern sich die variablen Kosten, also die Betriebskosten für die Geräte und die Aufwendungen für Verbrauchsmaterialien, wie z.B. Prozesschemikalien und Reinstwasser. Zusätzliche finanzielle Vorteile ergeben sich durch den Anstieg der Ausbeute, der bedeutet, dass mehr funktionsfähige Chips hergestellt und verkauft werden können.

Zusätzlich ergeben sich dadurch positive Auswirkungen für die Durchlaufzeit und den WIP-Bestand der Produktionslinie, was bedeutet, dass weniger Kapital in der Fertigung gebunden ist.

Ein Anstieg der Prozessschritte bedeutet im Umkehrschluss, dass erheblich höhere Kosten entstehen.

- Zahl der Geräte

Da Anschaffungskosten für Geräte inzwischen 75% der Gesamtinvestitionen einer Fabrik ausmachen, bringt eine Reduktion der Zahl der Geräte erhebliche Vorteile mit sich. Dadurch wird die Investitionssumme reduziert und damit die fixen Kosten während des Fab-betriebs. Zusätzlich ergeben sich Einsparungsmöglichkeiten beim Personalaufwand und bei den variablen Kosten für den Geräteunterhalt. Daneben muss beachtet werden, dass mehr Geräte auch mehr Stellfläche im teuren Reinraum mit sich bringen.

Zur Beseitigung eines Bottlenecks kann es jedoch unter Umständen sinnvoll sein, die Zahl der Geräte an bestimmten Stellen gezielt zu erhöhen. Dadurch ergeben sich zwar höhere Kosten, durch die resultierende Steigerung der Produktionsmenge wird aber auch ein höherer Umsatz erreicht.

- Fab-Layout

Durch das Fab-Layout können die Investitionen für den Aufbau und Betrieb einer Fab beeinflusst werden. Je nach Aufbau kann die Zahl der benötigten Geräte variieren. Ebenfalls können davon die Größe und Gestaltung des Reinraums und damit die Bau- und Installationskosten abhängen.

- Transportzeiten

Eine Verringerung der Transportzeiten setzt entweder mehr Personal zum schnelleren Abtransport der Lose oder eine stärkere Automatisierung voraus. Dadurch werden also einerseits die Personalkosten oder andererseits die Investitionen bei der Einrichtung der Fab für die Anschaffung der Transportsysteme beeinflusst. Als Voraussetzung für die Automatisierung ist dabei zusätzlich eine vollständige Standardisierung der Geräte und Geräteschnittstellen notwendig. Hinzu kommt die Notwendigkeit einer Modifikation der bestehenden Logistik.

- Losgröße

Die Verwendung einer anderen Losgröße macht im allgemeinen eine andere Logistik und unterschiedliche Geräteschnittstellen notwendig. Aufgrund der verwendeten Transportbehälter (SMIF oder FOUP) ist ein Umstieg auf größere Losgrößen nicht möglich, kleinere Losgrößen könnten theoretisch durch unvollständiges Auffüllen der Behälter erzielt wer-

den.

Zu beachten ist, dass sich bei unterschiedlichen Losgrößen auch die Zahl der insgesamt notwendigen Transportvorgänge ändert. Bei einer Losgröße von einem Wafer müssen 25-mal mehr Transportvorgänge als bei einer Losgröße von 25 Wafern abgewickelt werden. Daraus ergibt sich, dass eine andere Steuerung und auch eine andere Auslegung der Transportsysteme sowie eine modifizierte Logistik notwendig wäre [133]. Wegen der höheren Zahl an Transportvorgängen und Waferankünften an den Geräten würde sich auch ein erheblich höherer Datenverarbeitungsaufwand ergeben. Vorab wäre zusätzlich großer Aufwand notwendig, um alle Geräte und Prozesse auf Einzelscheibenfertigung umzustellen, bzw. neu zu entwickeln. Die Einführung einer Losgröße von einem Wafer würde also vollkommen neue Fertigungskonzepte erfordern und ist deswegen unter den momentanen Voraussetzungen für Massenfertigung noch nicht umsetzbar.

Durch den Einfluss der Losgröße auf Durchlaufzeit und WIP ergibt sich bei Verwendung größerer Lose eine größere Summe an gebundenem Kapital.

- Personalbestand

Veränderungen des Personalbestands wirken sich auf die fixen Kosten der Fertigung aus. Da die Personalkosten insgesamt bei 10-15% der Gesamtkosten liegen, kann durch eine Verringerung der Mitarbeiterzahl durchaus eine erhebliche Kosteneinsparung erreicht werden. Dabei muss jedoch beachtet werden, dass bei einer Reduktion des Personalbestandes aufgrund eines steigenden Automatisierungsgrads zwar die Personalkosten sinken, dafür aber zusätzliche Investitionen in die Automatisierungstechnik vorgenommen werden müssen und ein reibungsloser Ablauf der Fertigung sichergestellt sein muss.

- Produkt

Durch die Art und Zahl der hergestellten Produkte kann die Zahl der benötigten Geräte beeinflusst werden, was die bereits beschriebenen Konsequenzen für die entstehenden Kosten haben kann. Ebenfalls stark betroffen durch Produktveränderungen sind der Prozessablauf und die dazugehörige Ablaufsteuerung. Eine Umstellung der Produktion auf andere Produkte macht vorab hohen Aufwand notwendig.

6 Warteschlangentheorie zur Berechnung der Bewertungsgrößen

6.1 Warteschlangenmodelle für deterministische und variable Ankunfts- und Prozesszeit

Ziel ist es für realistische Gerätekonfigurationen die Bewertungsgrößen in Abhängigkeit von den Einflussparametern berechnen zu können.

Bei der Ankunft der Wafer am Gerät und ihrer nachfolgenden Bearbeitung handelt es sich um ein Warteschlangensystem, wie es in Abbildung 6.1 dargestellt ist. Waferlose kommen mit einer bestimmten Ankunftsrate am Gerät an, müssen in einem Zwischenlager vor dem Gerät warten, falls bereits andere Wafer bearbeitet werden, werden prozessiert und verlassen dann das System wieder.

Für einfache Fälle können daher mit Hilfe der Warteschlangentheorie durch Angabe von analytischen Formeln die Zusammenhänge zwischen den Kenngrößen Produktionsmenge (Gerätenutzung), Durchlaufzeit und WIP verdeutlicht werden.

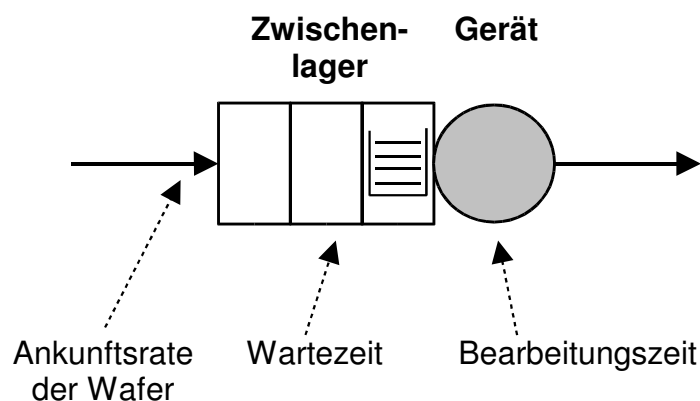


Abbildung 6.1: Prinzipieller Aufbau eines Warteschlangensystems

Anmerkung:

Da sich die Warteschlangentheorie nicht nur auf die Halbleitertechnik anwenden läßt, wird in der allgemeinen Nomenklatur von Kunden (in diesem Spezialfall: Waferlose), Bedienrate (hier: Gerätedurchsatz) und Bedienstellen (hier: Geräte) gesprochen.

Die Benennung von Warteschlangensystemen erfolgt nach ihren Eigenschaften. Dabei setzt sich der Name aus mehreren Teilen zusammen, die durch Schrägstriche voneinander abgetrennt werden.

Die allgemeinste Notation für ein Warteschlangensystem lautet:

$A/B/s/k/Y/Z$

Die einzelnen Symbole stehen dabei für verschiedene Eigenschaften und können unterschiedlich realisiert sein, was in Tabelle 6.1 aufgelistet ist [108], [109, S. 34].

Tabelle 6.1: Benennung von Warteschlangensystemen

Symbol	Eigenschaft	wichtige Realisierungsmöglichkeiten	Erklärung
A	Ankunftsrate	M D E_k H_k G	Exponentiell Deterministisch Erlang Hyperexponentiell Allgemein
B	Bedienrate	M D E_k H_k G	Exponentiell Deterministisch Erlang Hyperexponentiell Allgemein
s	Zahl der parallelen Bedienstellen	s	$s \in \mathbb{N}$
k	Maximal erlaubte Zahl der Kunden im System	k	$k \in \mathbb{N}$
Y	Maximale Warteschlangenlänge	Y	$Y \in \mathbb{N}$
Z	Warteschlangendisziplin	FCFS LCFS ROS	First come, first served Last come, first served Random order of service

Bei unbegrenzter erlaubter Zahl der Kunden im System und unbegrenzter Warteschlangenlänge, wovon im folgenden ausgegangen wird, werden die Symbole k und Y üblicherweise bei der Benennung der Warteschlange weggelassen.

Zusätzlich soll FIFO (First in first out) bzw. FCFS als Warteschlangendisziplin angenommen werden, d.h. das Los, das zuerst am jeweiligen Gerät ankommt, wird auch als erstes prozessiert.

6.1.1 Einzelgeräte

M/M/1-System:

Den analytisch am einfachsten zu betrachtenden Fall der Warteschlangen stellen Systeme mit einer exponentiellen Ankunfts- und Bedienrate dar. Die Annahme einer Exponentialverteilung ist unter der Voraussetzung zufälliger und voneinander unabhängiger Ankünfte und konstanter Ankunfts-wahrscheinlichkeit erlaubt [111, S. 665 ff.].

Der Nutzungsgrad u eines Systems wird definiert als das Verhältnis von tatsächlich bearbeiteten Wafern zu theoretischer Bearbeitungs-kapazität. Dabei ist zu beachten, dass immer $u < 1$ gelten muss, also die Ankunftsrate der Wafer kleiner als die Maschinenkapazität sein muss, da andernfalls die ankommende Menge der Wafer nicht mehr vollständig abgearbeitet werden könnte und das Lager vor dem entsprechenden Gerät unbeschränkt immer weiter anwachsen würde und somit das System nicht mehr stabil wäre, was bei der analytischen Lösung zu falschen Ergebnissen führen würde.

$$u = \frac{r_a}{r_e} \quad (6.1)$$

u = Nutzungsgrad
 r_a = mittlere Ankunftsrate [W/h]
 r_e = Maschinenkapazität [W/h]

Die Gesamtdurchlaufzeit des Systems läßt sich aus der Bearbeitungszeit im Gerät (= Prozesszeit) und der Wartezeit herleiten [110, S. 188 ff.], [111, S. 671].

$$TD = \frac{1}{r_e - r_a} = \frac{r_a}{r_e \cdot (r_e - r_a)} + \frac{1}{r_e} \quad (6.2)$$

TD = (Gesamt-)Durchlaufzeit

Der Anteil der Gesamtdurchlaufzeit, der auf die Wartezeit im System zurückzuführen ist, berechnet sich zu:

$$TD_q = \frac{r_a}{r_e \cdot (r_e - r_a)} \quad (6.3)$$

TD_q = Wartezeit

Die Bearbeitungszeit im Gerät ergibt sich aus der Bearbeitungsgeschwindigkeit:

$$TD_p = \frac{1}{r_e} \quad (6.4)$$

TD_p = Bearbeitungszeit

Die Gesamtdurchlaufzeit läßt sich aus der Wartezeit und der Bearbeitungszeit berechnen:

$$TD = TD_q + TD_p \quad (6.5)$$

Nach Little's Law [110, S. 17] berechnet sich die mittlere Anzahl der Teile im System als Produkt aus Waferankunftsrate und Gesamtdurchlaufzeit.

$$WIP = TD \cdot r_a \quad (6.6)$$

WIP = Anzahl der Teile im System

Little's Law kann auch auf die Zahl der Teile in der Warteschlange angewendet werden.

$$WIP_q = TD_q \cdot r_a \quad (6.7)$$

Für den mittleren WIP-Bestand eines M/M/1-Fertigungssystems ergibt sich damit aus Gleichung (6.2) und (6.6):

$$WIP = \frac{u}{1-u} = \frac{r_a}{r_e - r_a} \quad (6.8)$$

Grafisch dargestellt ergeben sich für die Durchlaufzeit TD und den WIP-Bestand eines M/M/1-Warteschlangensystems mit einer Bearbeitungsrate $r_e=25$ W/h in Abhängigkeit vom Nutzungsgrad die in Abbildung 6.2 dargestellten Kurven.

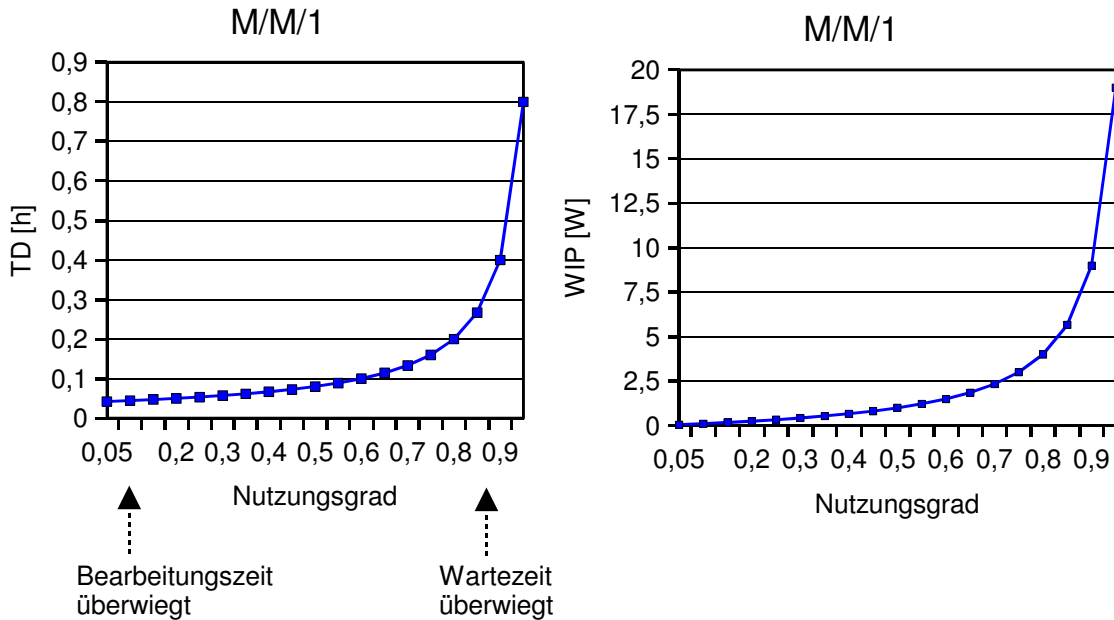


Abbildung 6.2: M/M/1-System

M/D/1-System:

Für ein Gerät mit exponentieller Ankunftsrate und festgelegter, nicht schwankender Bearbeitungsrate nimmt das Warteschlangensystem die Form von Abbildung 6.1 an und wird als M/D/1-System bezeichnet. Die relevanten Gleichungen zur Berechnung eines solchen Systems werden im folgenden erläutert [110, S. 188 ff.], [111, S. 688].

Die Gesamtdurchlaufzeit des Systems setzt sich wiederum zusammen aus der Bearbeitungszeit im Gerät (= Prozesszeit) und der Wartezeit.

$$TD = \frac{r_a}{2 \cdot r_e \cdot (r_e - r_a)} + \frac{1}{r_e} = \frac{u^2}{2 \cdot r_a \cdot (1 - u)} + \frac{u}{r_a} \quad (6.9)$$

TD = (Gesamt-)Durchlaufzeit

Wartezeit im System:

$$TD_q = \frac{r_a}{2 \cdot r_e \cdot (r_e - r_a)} \quad (6.10)$$

TD_q = Wartezeit

Bearbeitungszeit im Gerät:

$$TD_p = \frac{1}{r_e} \tag{6.11}$$

TD_p = Bearbeitungszeit

Nach Little's Law berechnet sich der mittlere WIP-Bestand in einem M/D/1-Fertigungssystem aus Gleichung (6.6) zu:

$$WIP = \frac{r_a^2}{2 \cdot r_e \cdot (r_e - r_a)} + \frac{r_a}{r_e} = \frac{u^2}{2 \cdot (1-u)} + u \tag{6.12}$$

Wie in den Gleichungen (6.3) und (6.10) erkennbar ist, ist bei gleicher Prozesszeit beim M/D/1-System die Wartezeit vor dem Gerät im Vergleich zum M/M/1-System um Faktor 1/2 geringer. Dies ist darauf zurückzuführen, dass Wartezeiten durch statistische Variationen verursacht werden und durch die festgelegte und konstante Bearbeitungszeit der Geräte beim M/D/1-System eine Quelle für Variationen ausgeschaltet wird, wogegen beim M/M/1-System jeweils statistische Verteilungen für Ankunfts- und Bearbeitungszeit angenommen werden.

Mit $r_e=25$ W/h verhalten sich TD und WIP wie in Abbildung 6.3 dargestellt. Wie oben erläutert, ist die Gesamtdurchlaufzeit und der WIP-Bestand geringer als beim M/M/1-System.

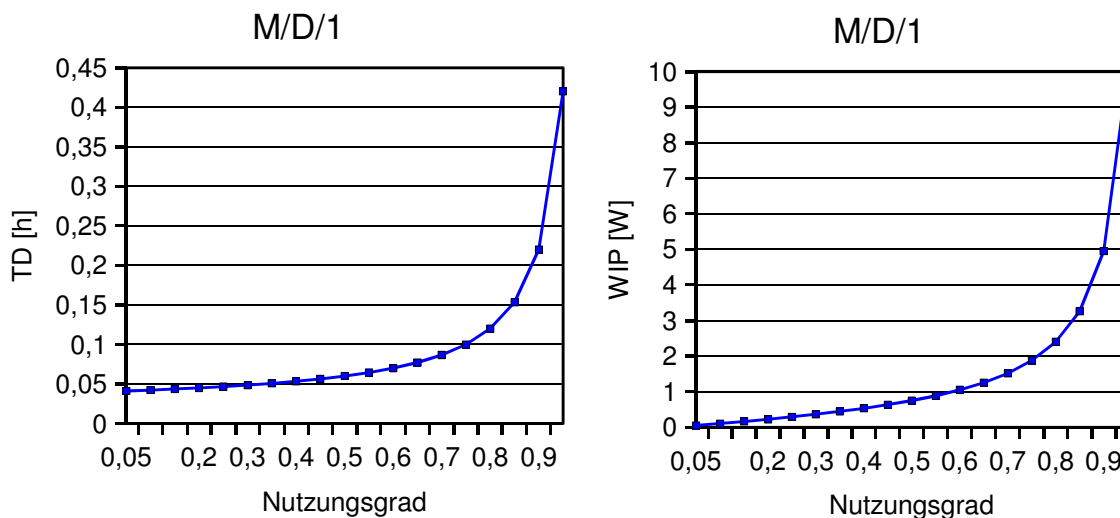


Abbildung 6.3: M/D/1-System

G/G/1-System:

Die meisten realen Systeme verhalten sich nicht exakt wie durch M/M/1- oder M/D/1-Warteschlangenmodelle beschrieben. Prozesszeiten und Ankunftszeiten in einem gesamten Fertigungssystem mit gegenseitigen Abhängigkeiten sind nicht notwendigerweise exponentiell verteilt oder gar deterministisch. Deshalb müssen G/G/1-Systeme mit allgemein verteilter Ankunfts- und Bearbeitungsrate betrachtet werden.

Eine Näherung zur Berechnung solcher Systeme wird in der sogenannten Kingman Formel angegeben [164, S. 270], [108]:

$$TD_q = \underbrace{\left(\frac{c_a^2 + c_e^2}{2}\right)}_V \cdot \underbrace{\left(\frac{u}{1-u}\right)}_U \cdot \underbrace{\frac{1}{r_e}}_T = VUT \quad (6.13)$$

c_a = Variationskoeffizient Ankunft
 c_e = Variationskoeffizient Bearbeitung

Wie in Gleichung (6.13) dargestellt, läßt sich die Kingman Formel in drei Terme unterteilen: einen dimensionslosen Term V, der die Variabilität beschreibt, einen Term U, der vom Nutzungsgrad abhängt und einen Term T, der die Bearbeitungsdauer im Gerät beschreibt.

$$TD_q = \frac{c_a^2 + c_e^2}{2} \cdot \frac{u}{1-u} \cdot \frac{1}{r_e} = \frac{c_a^2 + c_e^2}{2} \cdot \frac{r_a}{r_e \cdot (r_e - r_a)} \quad (6.14)$$

Die Gesamtdurchlaufzeit eines G/G/1-Systems berechnet sich wiederum aus der Wartezeit und der Bearbeitungszeit im Gerät.

$$TD = \frac{c_a^2 + c_e^2}{2} \cdot \frac{u}{1-u} \cdot \frac{1}{r_e} + \frac{1}{r_e} = \frac{c_a^2 + c_e^2}{2} \cdot \frac{r_a}{r_e \cdot (r_e - r_a)} + \frac{1}{r_e} \quad (6.15)$$

Nach Little's Law berechnet sich der mittlere WIP-Bestand eines G/G/1-Systems mit Gleichung (6.6) zu:

$$WIP = \frac{c_a^2 + c_e^2}{2} \cdot \frac{u^2}{1-u} + u = \frac{c_a^2 + c_e^2}{2} \cdot \frac{r_a^2}{r_e \cdot (r_e - r_a)} + \frac{r_a}{r_e} \quad (6.16)$$

Aus den Gleichungen lassen sich folgende Zusammenhänge ablesen:

G/G/1-System entspricht M/M/1-System für $c_a = c_e = 1$

G/G/1-System entspricht M/D/1-System für $c_a = 1, c_e = 0$

In Abbildung 6.4 und Abbildung 6.5 sind TD und WIP eines G/G/1-Systems für $r_e=25$ W/h und für $c_a=c_e=1,5$ bzw. $c_a=c_e=0,5$ dargestellt. Als Referenz sind zusätzlich die Kurven für das M/M/1- und M/D/1-System angegeben. Eindeutig ist der starke Einfluss der Variabilität auf die Durchlaufzeit und den WIP-Bestand zu erkennen.

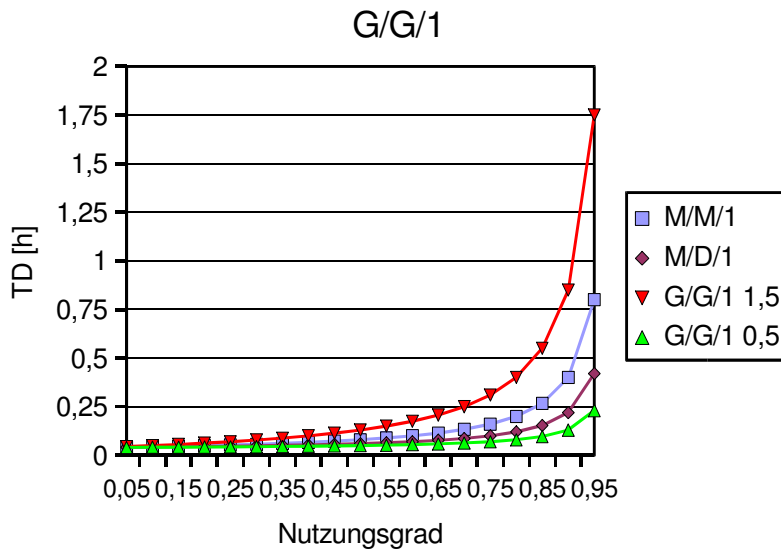


Abbildung 6.4: TD für G/G/1-System

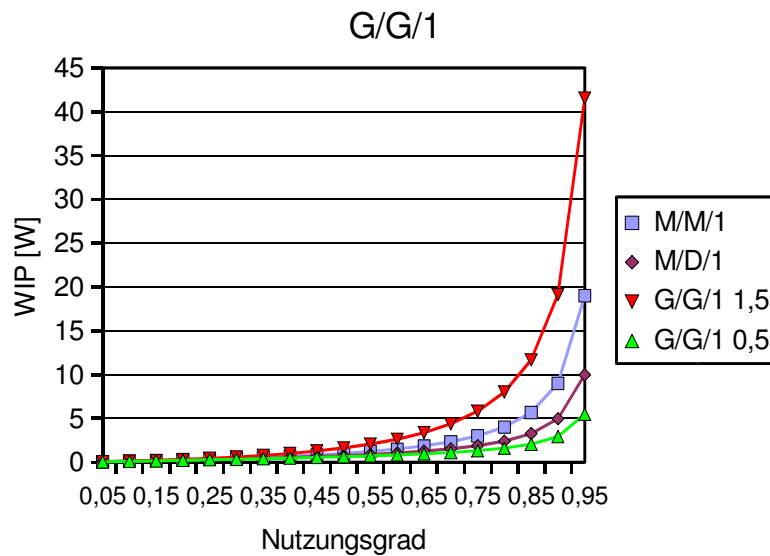


Abbildung 6.5: WIP für G/G/1-System

6.1.2 Parallele Geräte

M/M/s-System:

Für den Fall, dass mehrere Geräte zur Ausführung eines Prozessschritt parallel geschaltet werden (allgemein: s Geräte), zwischen denen die zu bearbeitenden Wafer verteilt werden können, ergibt sich ein Warteschlangensystem der Form, wie es prinzipiell in Abbildung 6.6 angedeutet ist. Die Arbeit verteilt sich dabei auf s Geräte, die alle die gleiche Maschinenkapazität r_e aufweisen sollen. Voraussetzung für die Parallelisierung der Prozesse ist, dass auf jedem Gerät der gleiche Prozess ausgeführt wird.

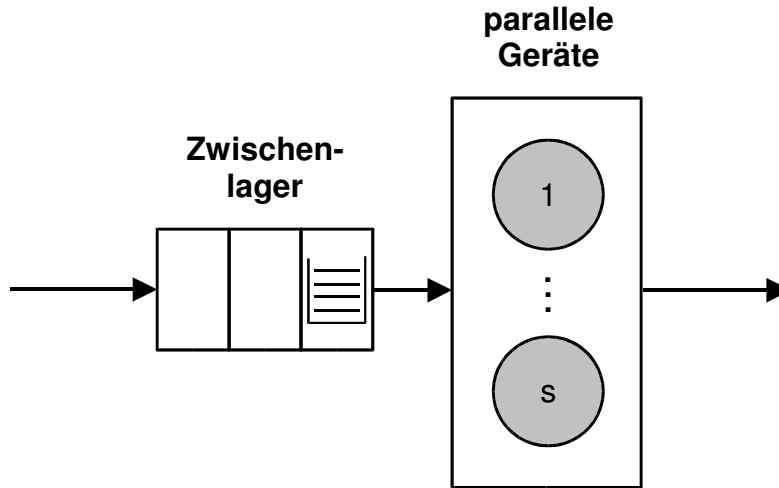


Abbildung 6.6: Parallelgeschaltete Geräte

Bei einem M/M/s-System mit exponentieller Ankunfts- und Bedienrate an allen s parallelen Geräten lassen sich folgende Formeln zur Berechnung von Durchlaufzeit und WIP angeben [108], [113]:

Nutzungsgrad des Gesamtsystems:

$$u = \frac{r_a}{s \cdot r_e} \quad (6.17)$$

s = Anzahl der parallelen Bedienstellen

WIP-Bestand im System:

$$P_0 = \frac{1}{P} \quad (6.18)$$

$$P = \left[\sum_{n=0}^{s-1} \frac{1}{n!} \cdot \left(\frac{r_a}{r_e} \right)^n \right] + \frac{1}{s!} \cdot \left(\frac{r_a}{r_e} \right)^s \cdot \frac{s \cdot r_e}{s \cdot r_e - r_a} = \frac{(s \cdot u)^s}{(1-u) \cdot s!} + \left[\sum_{n=0}^{s-1} \frac{(s \cdot u)^n}{n!} \right] \quad (6.19)$$

$$WIP = \left[\frac{r_a \cdot r_e \cdot \left(\frac{r_a}{r_e} \right)^s}{(s-1)! \cdot (s \cdot r_e - r_a)^2} \right] \cdot P_0 + \frac{r_a}{r_e} = \left[\frac{u \cdot \left(\frac{r_a}{r_e} \right)^s}{s! \cdot (1-u)^2} \right] \cdot P_0 + \frac{r_a}{r_e} \quad (6.20)$$

Für die Gesamtdurchlaufzeit des Systems folgt mit Little's Law analog zum Warteschlangensystem mit einem einzelnen Gerät:

$$TD = \frac{WIP}{r_a} \quad (6.21)$$

und damit:

$$TD = \left[\frac{r_a \cdot r_e \cdot \left(\frac{r_a}{r_e}\right)^s}{(s-1)! \cdot (s \cdot r_e - r_a)^2} \right] \cdot \frac{P_0}{r_a} + \frac{1}{r_e} = \left[\frac{u \cdot \left(\frac{r_a}{r_e}\right)^s}{s! \cdot (1-u)^2} \right] \cdot \frac{P_0}{r_a} + \frac{1}{r_e} \quad (6.22)$$

Von Sakasegawa wurde folgende Näherungsformel für die Berechnung der Durchlaufzeit des M/Ms-Systems angegeben [164, S. 272]:

$$TD = \frac{u^{\sqrt{2(s+1)}-1}}{s(1-u)} \cdot \frac{1}{r_e} + \frac{1}{r_e} \quad (6.23)$$

In Abbildung 6.7 und Abbildung 6.8 werden die Ergebnisse aus den Gleichungen (6.22) und (6.23) für 2 bzw. 3 parallel geschaltete Geräte mit $r_e=25W/h$ gegenübergestellt. Es zeigt sich praktisch kein Unterschied in den berechneten Ergebnissen.

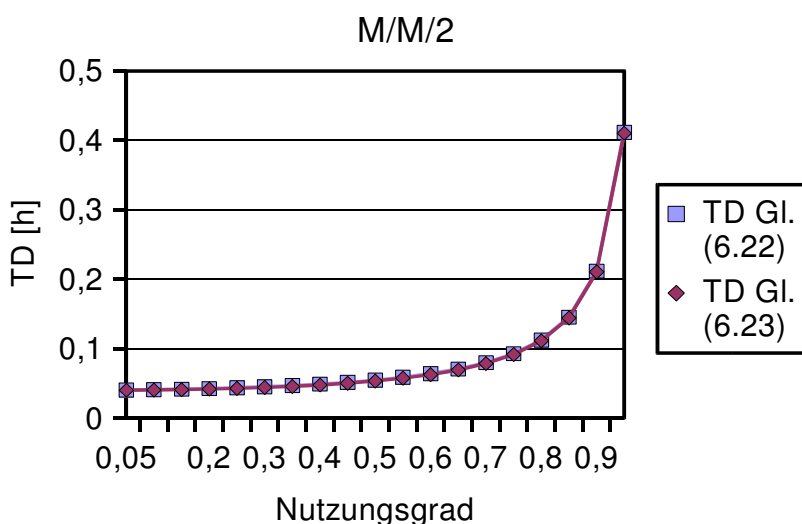


Abbildung 6.7: Vergleich von Gleichung (6.22) und (6.23) für s=2 Geräte

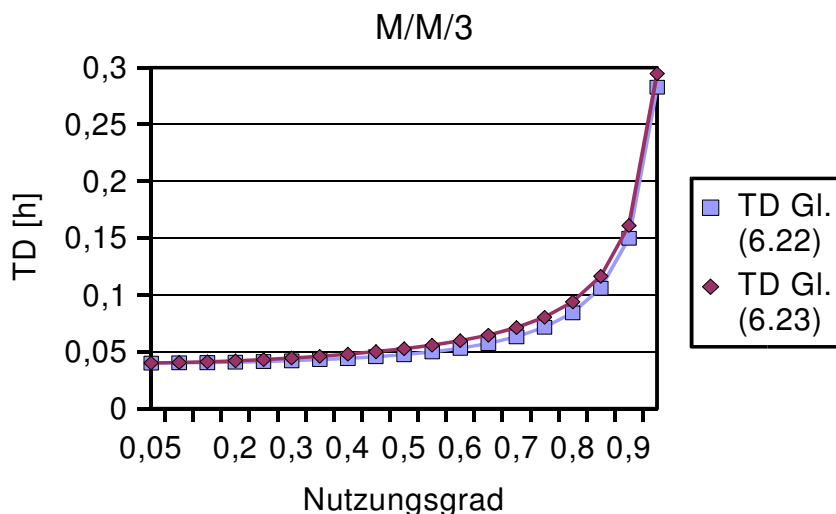


Abbildung 6.8: Vergleich von Gleichung (6.22) und (6.23) für s=3 Geräte

M/D/s-System:

Für M/D/s-Systeme, also für s parallel geschaltete Geräte (siehe Abbildung 6.6) mit exponentieller Ankunftsrate und konstanter Bedienrate, gilt folgende Näherung [113], [114]:

$$WIP_q \approx \frac{WIP_{qM/M/s}}{2} \quad (6.24)$$

Die Anzahl der Teile im Lager ist damit halb so groß wie beim M/M/s-System.

Für den WIP-Bestand beim M/D/s-System ergibt sich damit [113]:

$$P_0 = \frac{1}{P} \quad (6.25)$$

$$P = \left[\sum_{n=0}^{s-1} \frac{1}{n!} \left(\frac{r_a}{r_e} \right)^n \right] + \frac{1}{s!} \left(\frac{r_a}{r_e} \right)^s \cdot \frac{s \cdot r_e}{s \cdot r_e - r_a} = \frac{(s \cdot u)^s}{(1-u) \cdot s!} + \left[\sum_{n=0}^{s-1} \frac{(s \cdot u)^n}{n!} \right] \quad (6.26)$$

$$WIP = \left[\frac{r_a \cdot r_e \cdot \left(\frac{r_a}{r_e} \right)^s}{(s-1)! \cdot (s \cdot r_e - r_a)^2} \right] \cdot \frac{P_0}{2} + \frac{r_a}{r_e} = \left[\frac{u \cdot \left(\frac{r_a}{r_e} \right)^s}{s! \cdot (1-u)^2} \right] \cdot \frac{P_0}{2} + \frac{r_a}{r_e} \quad (6.27)$$

Für die Gesamtdurchlaufzeit des Systems folgt mit Little's Law:

$$TD = \left[\frac{r_a \cdot r_e \cdot \left(\frac{r_a}{r_e} \right)^s}{(s-1)! \cdot (s \cdot r_e - r_a)^2} \right] \cdot \frac{P_0}{2r_a} + \frac{1}{r_e} = \left[\frac{u \cdot \left(\frac{r_a}{r_e} \right)^s}{s! \cdot (1-u)^2} \right] \cdot \frac{P_0}{2r_a} + \frac{1}{r_e} \quad (6.28)$$

G/G/s-System:

Wie in den Gleichungen (6.3) und (6.14) ersichtlich, kann die Wartezeit in einem G/G/1-System auch dargestellt werden als

$$TD(G/G/1)_q = \left(\frac{c_a^2 + c_e^2}{2} \right) \cdot TD(M/M/1)_q \quad (6.29)$$

In [164, S. 273] wird deshalb vorgeschlagen, für die G/G/s Warteschlange analog zu verfahren. Damit ergibt sich mit Gleichung für die Durchlaufzeit des G/G/s-Systems mit s parallelen Geräten:

$$TD = \frac{c_a^2 + c_e^2}{2} \cdot \frac{u^{\sqrt{2(s+1)}-1}}{s(1-u)} \cdot \frac{1}{r_e} + \frac{1}{r_e} \quad (6.30)$$

Gleichung (6.30) wird benutzt, um für den allgemeinen Fall von G/G/s-Systemen die

Durchlaufzeit zu berechnen, wenn sich die ankommenden Lose zwischen den parallelen Geräten aufteilen.

In Abbildung 6.9 und Abbildung 6.10 werden die Ergebnisse aus den Gleichungen (6.23) und (6.30) miteinander verglichen. Dabei werden 2 parallel geschaltete Geräte mit jeweils einer Bearbeitungsrate von $r_e=25W/h$ angenommen. Wie bereits bei Einzelgeräten ergeben sich auch hier die längste Durchlaufzeit und der höchste WIP-Bestand für ein G/G/s-System mit hoher Variabilität.

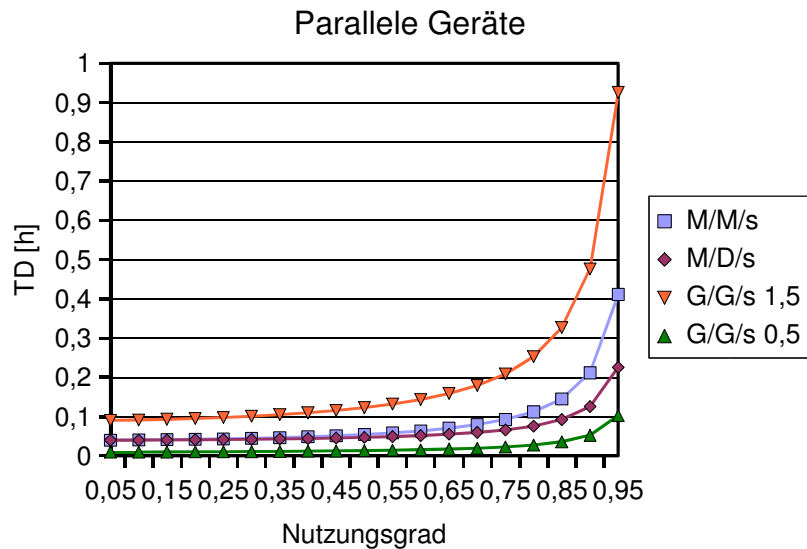


Abbildung 6.9: TD für 2 parallele Geräte

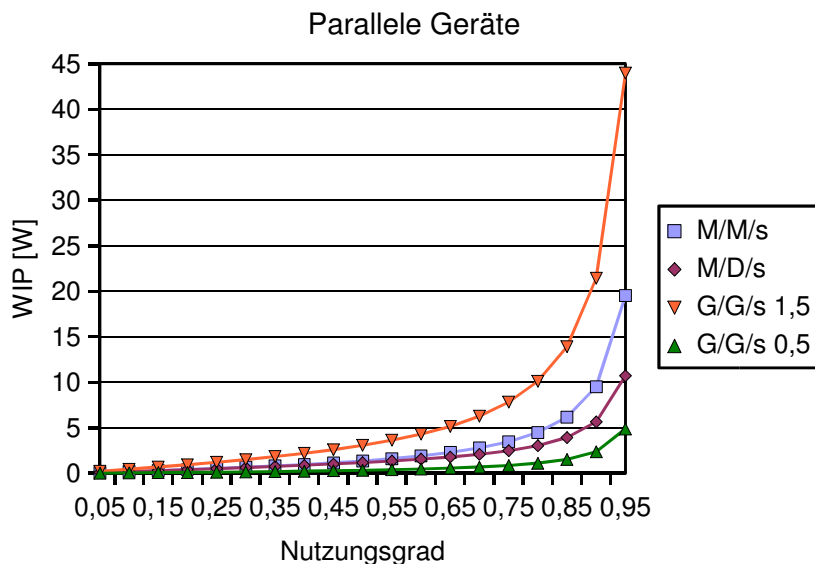


Abbildung 6.10: WIP 2 für parallele Geräte

6.1.3 Warteschlangennetzwerke

Die Formeln (6.1) bis (6.30) beziehen sich auf einzelne, isoliert arbeitende Geräte. In einer Produktionslinie müssen jedoch mehrere Prozessschritte hintereinander abgearbeitet werden, d.h., dass mehrere Geräte nacheinander durchlaufen werden müssen. Ein solches System wird in der Warteschlangentheorie oft als Tandem-Warteschlangennetz bezeichnet [110, S. 47 ff.], [116, S. 29].

Vereinfacht gesprochen ist ein Tandemsystem eine Hintereinanderschaltung endlich vieler elementarer Warteschlangensysteme bzw. Bedienstationen, wobei ein ankommender Kunde in die erste Bedienstation eintritt, nacheinander alle folgenden Stationen in festgelegter Reihenfolge durchläuft und schließlich nach Bearbeitung in der letzten Station das System verlässt. Im folgenden wird für alle beteiligten Geräte FIFO-Warteschlangendisziplin angenommen.

Der prinzipielle Aufbau eines solches Systems ist in Abbildung 6.11 dargestellt.

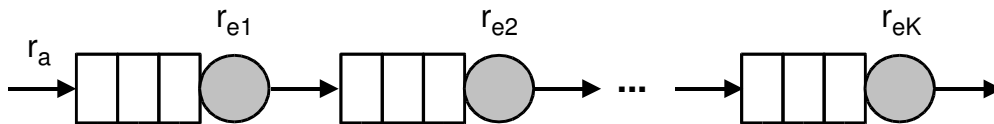


Abbildung 6.11: Warteschlangennetzwerk

Durch die Hintereinanderschaltung ergibt sich an allen Geräten die gleiche mittlere Ankunftsrate r_a .

Netzwerke von M/M/1-Systemen:

Bei einem M/M/1-Warteschlangensystemen mit K Geräten, die exponentielle Ankunfts- und Bedienrate aufweisen, stellt der Abgangsprozess vom Gerät ebenfalls einen Poissonprozess mit der gleichen Rate wie der Ankunftsprozess dar. Daraus ergibt sich, dass ein solches Netzwerk als Summe von K unabhängigen M/M/1-Warteschlangen aufgefasst werden kann [108], [116, S. 33 f.].

Der Gesamtwert für die Durchlaufzeit und die Wartezeit kann damit als Summe der Einzelwerte aus den Gleichungen (6.2) und (6.3) berechnet werden.

Gesamtdurchlaufzeit:

$$TD = \sum_{i=1}^K \frac{1}{r_{ei} - r_a} \quad (6.31)$$

Gesamtwartezeit:

$$TD_q = \sum_{i=1}^K \frac{r_a}{r_{ei} \cdot (r_{ei} - r_a)} \quad (6.32)$$

Für die Zahl der Teile im System gilt weiterhin Little's Law, d.h. mit der Formel

$$WIP = TD \cdot r_a \quad (6.33)$$

und Gleichung (6.32) kann der WIP-Bestand des Gesamtsystems berechnet werden.

Netzwerke von M/D/1-Systemen:

Für Netzwerke aus M/D/1-Systemen, die direkt hintereinandergeschaltet werden, ist die Verteilung des Abgangsprozesses unabhängig von der Anordnung der Bedienstationen, d.h. wenn man die Reihenfolge der Geräte ändert, bleibt der Abgangsprozess unverändert.

Als Beispiel können dafür Clustertools dienen, bei denen der Wafer ohne Unterbrechung direkt von einer Prozesskammer zur nächsten transportiert wird.

Weiterhin kann gezeigt werden, dass die Warteschlange der Station $i+1$ genau dann immer leer ist, wenn die Prozesszeit von Gerät $i+1$ kleiner ist als von Gerät i . Man sagt dann, Gerät i dominiert Gerät $i+1$.

Zur Analyse des gesamten Systems, ohne nähere Betrachtung der einzelnen Geräte, läßt sich das Netzwerk zu einem einfacheren System reduzieren, indem man die Geräte zunächst so anordnet, dass $r_{e1} \leq r_{e2} \leq \dots \leq r_{eK}$ gilt und dann die Geräte 2 bis K durch eine Station mit unendlich vielen Bedienern - also ohne Wartezeiten - und mit Bedienzeit

$$\frac{1}{r_{e2}} + \frac{1}{r_{e3}} + \dots + \frac{1}{r_{eK}}$$

ersetzt [116, S. 32 f.]. Die Gesamtwartezeit in einem solchen System ist also gleich der Wartezeit in einem elementaren M/D/1-System mit Bedienrate $r_e = \min\{r_{ei}\}$. Die Gesamtdurchlaufzeit ergibt sich aus der Summe dieser Wartezeit und der Bedienzeiten $1/r_{e1}, \dots, 1/r_{eK}$.

Dabei ist anzumerken, dass das System nur mathematisch so betrachtet wird, als ob das Lager nur am Bottleneck-Gerät existieren würde. In Wirklichkeit teilt sich dieses Lager auf mehrere Geräte auf.

Die Gesamtdurchlaufzeit des Systems ergibt sich mit diesen Überlegungen zu:

$$TD = \frac{r_a}{2 \cdot r_{e1} \cdot (r_{e1} - r_a)} + \sum_{i=1}^K \frac{1}{r_{ei}} \quad (6.34)$$

Die Gesamtwartezeit beträgt:

$$TD_q = \frac{r_a}{2 \cdot r_{e1} \cdot (r_{e1} - r_a)} \quad (6.35)$$

mit $r_{e1} = \min\{r_{ei}\}$.

Netzwerke von G/G/1-Systemen:

Anders als bei M/M/1- oder M/D/1-Systemen kann bei G/G/1-Warteschlangennetzwerken die Ankunftsrate am Gerät und die Bearbeitungsrate im Gerät statistisch beliebig verteilt sein. Bei mehreren aufeinander folgenden Geräten kann dann allerdings die Verteilung der Ankunftsraten nicht mehr als voneinander unabhängig angesehen werden.

Der Variationskoeffizient beim Verlassen des Gerätes berechnet sich zu [164, S. 263]:

$$c_{di}^2 = u_i^2 \cdot c_{ei}^2 + (1 - u_i^2) \cdot c_{ai}^2 \quad (6.36)$$

c_{di} = Variationskoeffizient beim Verlassen des Geräts i

Laut Gleichung (6.36) ist also die resultierende Variabilität am Ausgang des Geräts abhängig von den Variationskoeffizienten für Ankunft und Bearbeitung und vom Nutzungsgrad. Dieser Variationskoeffizient kann dann für den Eingang des nachfolgenden Geräts angenommen werden, so dass gilt $c_{ai+1}^2 = c_{di}^2$.

Für mehrere parallel geschaltete Geräte ergibt sich folgende Formel zur Berechnung von c_d [164, S. 263]:

$$c_d^2 = 1 + (1 - u^2) \cdot (c_a^2 - 1) + \frac{u^2}{\sqrt{m}} \cdot (c_e^2 - 1) \quad (6.37)$$

Für $m=1$ reduziert sich Gleichung (6.37) zu Gleichung (6.36).

Mit Gleichung (6.36) zeigt sich, dass für M/M/1-Systeme, bei denen $c_a=c_e=1$ gilt, $c_d=1$ ist. D.h., dass in diesem Fall der Variationskoeffizient an den einzelnen Geräten konstant bleibt.

Für M/D/1-Systeme ergibt sich mit Gleichung (6.36) $c_d^2=1-u^2$, d.h. dass hier der Variationskoeffizient abhängig vom Nutzungsgrad des Geräts zwischen 0 und 1 liegt.

Mehrfach benutzte Geräte:

Bei der Halbleiterfertigung werden manche Geräte mehr als einmal im Prozessablauf durchlaufen, d.h. es wird der gleiche Prozess mehrfach an unterschiedlichen Stellen des Herstellungsablaufs ausgeführt.

In Abbildung 6.12 ist schematisch die zweifache Benutzung eines Gerätes dargestellt.

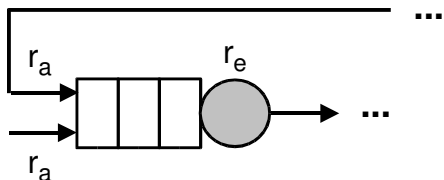


Abbildung 6.12: Mehrfach benutztes Gerät

Wie in Abbildung 6.12 deutlich wird, addieren sich bei einem solchen System die am Gerät ankommenden Ströme. Die Gesamt-Ankunftsrate berechnet sich dann zu [117], [118]

$$r_a' = V \cdot r_a \quad (6.38)$$

r_a' = Gesamt-Ankunftsrate

V = Anzahl der durchlaufenen Schleifen

Der Nutzungsgrad des Geräts ergibt sich analog zu Gleichung (6.1) mit

$$u = \frac{V \cdot r_a}{r_e} \quad (6.39)$$

Für parallelgeschaltete Geräte, die mehrfach durchlaufen werden, sind die jeweiligen Überlegungen analog anzuwenden. Als Nutzungsgrad ergibt sich dann:

$$u = \frac{V \cdot r_a}{s \cdot r_e} \tag{6.40}$$

s = Zahl der parallelen Geräte

Die Warteschlangentheorie eignet sich zur einfachen Berechnung und zur Abschätzung der Zusammenhänge. Sie stellt somit ein sehr nützliches Instrument dar, um grundsätzliche Erkenntnisse zu gewinnen. Für extreme Detailbetrachtungen sollte auf umfangreiche Verfahren zur Fab- oder Layoutsimulation zurückgegriffen werden, durch die eine exaktere Modellierung möglich ist. Durch Simulationssoftware wie z.B. AutoMod kann beispielsweise auch für jedes einzelne Los die Variation der Größen Durchlaufzeit und WIP festgestellt werden. Da die analytischen Formeln der Warteschlangentheorie nur statistische Mittelwerte ergeben, ist eine solche Betrachtung der Schwankungen von Los zu Los auf diesem Weg sonst nicht möglich.

6.2 Betriebskennlinie

Aus den in Kapitel 6.1 erläuterten Formeln der Warteschlangentheorie ergeben sich qualitativ folgende Abhängigkeiten für Durchlaufzeit und WIP vom Nutzungsgrad, die in Abbildung 6.13 dargestellt sind.

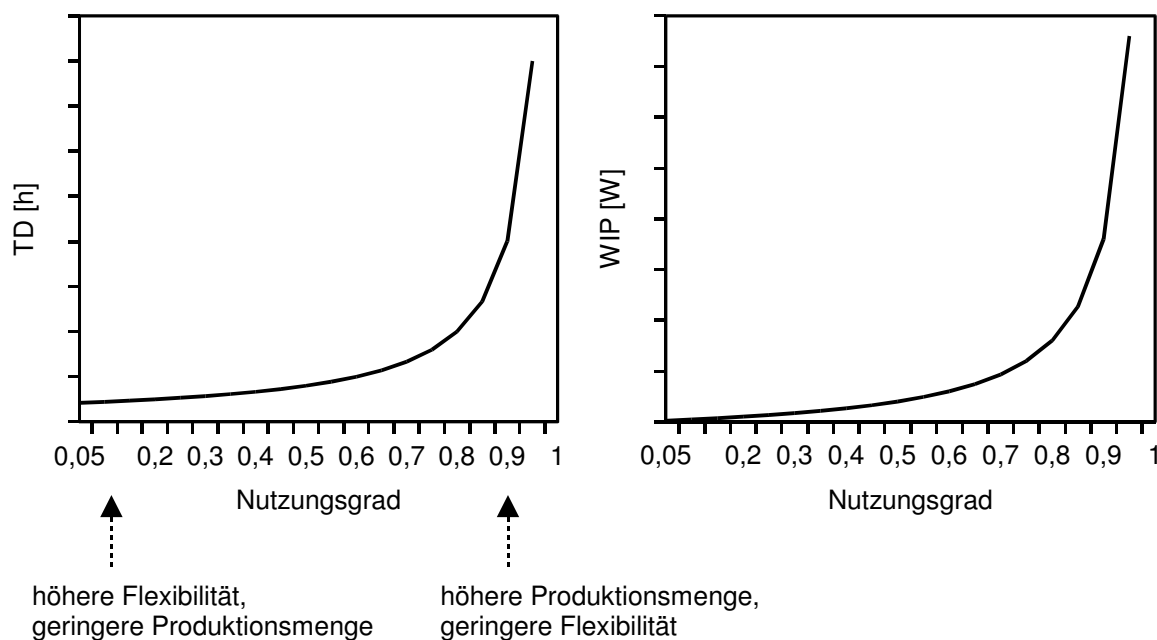


Abbildung 6.13: Betriebskennlinie: Zusammenhang zwischen Nutzungsgrad, Durchlaufzeit und WIP

Die Kurven aus Abbildung 6.13 werden auch als Betriebskennlinie oder Operating Curve bezeichnet. Das sogenannte "Operating Curve Management" ist seit kurzem in der Halbleiterindustrie wichtig geworden und betrachtet als Gesamtkonzept die gegenseitigen Abhängigkeiten von Mensch, Maschine, Material und Methode (4 M). Die Betriebskennlinie bildet die gesamte Fertigung ab und stellt ein sehr nützliches Instrument zur Analyse und

Planung dar, da auf einfache Weise festgestellt werden kann an welchem Punkt der Kennlinie man sich gerade befindet und welche Auswirkungen sich durch Veränderungen ergeben (s. z.B. [155], [169]). Diese Zusammenhänge werden im folgenden noch genauer erläutert.

Wie in Abbildung 6.13 erkennbar, ist hoher Nutzungsgrad eines Gerätes gleichbedeutend mit hoher Produktionsmenge, da bei besserer Ausnutzung des gleichen Gerätes laut Gleichung (6.1) mehr Wafer in der gleichen Zeit prozessiert werden können.

Bei höherem Nutzungsgrad steigen die Durchlaufzeit und die Zahl der Teile im System überproportional an, da für eine optimale Ausnutzung der Maschinenkapazität immer Wafer für die Produktion zur Verfügung stehen müssen und sich so zwangsläufig Wartezeiten vor dem Gerät ergeben. Das bedeutet, dass eine höhere Produktionsmenge des Gerätes mit längerer Produktionszeit und größerem Lagerbestand erkauft werden muss. In [173] wird beispielsweise das Verhältnis aus Durchsatz und Durchlaufzeit als Kenngröße für die Fab-Performance verwendet. Die Größen Durchlaufzeit und WIP sind dabei, wie bereits erläutert, formelmäßig über Little's Law (Gleichung (6.6)) miteinander verbunden, das besagt, dass bei längerer Durchlaufzeit auch die Zahl der Teile im System ansteigt.

Wie in Abbildung 6.13 angegeben, zeichnet sich die Produktionslinie, die mit niedrigem Nutzungsgrad betrieben wird, durch eine höhere Flexibilität aus. Diese Abhängigkeiten sind auch der Grund dafür, dass bei Massenprodukten (z.B. DRAMs) üblicherweise der Nutzungsgrad und damit die Produktionsmenge hoch ist, die Durchlaufzeit jedoch länger ist. Es ist zwar aus den in Kapitel 5.2.2 dargestellten Gründen auch hier technisch sinnvoll geringe Herstellungszeiten anzustreben, da hier aber sehr viele gleichartige und nicht unterscheidbare Produkte erzeugt werden, wirkt sich eine längere Durchlaufzeit für den Kunden nicht negativ aus. Im Gegensatz dazu kann bei Produkten mit kleinerer Stückzahl und mit hohem Termindruck (ASICs) die Gerätenutzung nicht beliebig gesteigert werden, da die Durchlaufzeit möglichst gering gehalten werden muss [119], [120].

Bei den Betrachtungen ist zu beachten, dass Durchlaufzeit und WIP-Bestand nicht konstant sind, sondern, aufgrund der statistischen Waferankunft, von Los zu Los teilweise großen Schwankungen unterliegen. Die mit Hilfe der Warteschlangentheorie berechneten Ergebnisse beziehen sich jeweils auf den statistischen Mittelwert. In der Realität schwankt die Durchlaufzeit (und damit auch der WIP-Bestand in der Fertigungslinie) von Los zu Los um einen Durchschnittswert herum, wobei der minimale Wert durch die theoretische Durchlaufzeit begrenzt wird, wie dies an einem Beispiel in Abbildung 6.14 angedeutet ist. Hier wurden mittels AutoMod Simulation die Daten für die Durchlaufzeit von 44 Losen mit einem Mittelwert von 0,2h ausgegeben. Falls es gelingen würde, eine Reduktion der Schwankungen zu erreichen, würden sich positive Auswirkungen auf die Durchlaufzeit und den WIP-Bestand ergeben [118], [122].

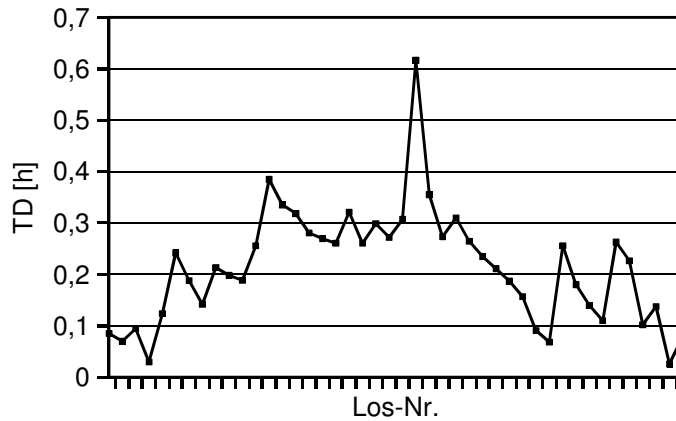


Abbildung 6.14: Beispiel für Schwankung der Durchlaufzeit

Da das (nicht erreichbare) Idealziel jeder Fertigung darin besteht, möglichst große Mengen in möglichst kurzer Zeit herzustellen, sollten Verbesserungsmaßnahmen dazu führen, dass die Betriebskennlinie aus Abbildung 6.13 möglichst flach verläuft, bzw. der steile Anstieg erst möglichst spät einsetzt.

Verbesserungen im Bereich der Produktionstechnik, durch die eine andere Kurvenform erreicht wird, können dann auf verschiedene Weise ausgenutzt werden. Wie in Abbildung 6.15 und in Abbildung 6.16 schematisch dargestellt ist, kann durch eine Abflachung der Kurve einerseits bei konstanter Produktionsmenge eine Reduktion der Durchlaufzeit erreicht werden oder andererseits die Produktionsmenge bei konstanter Durchlaufzeit gesteigert werden. Daneben ist auch Kombination beider Effekte, also eine Verkürzung der Durchlaufzeit bei gleichzeitiger Steigerung der Produktionsmenge möglich.

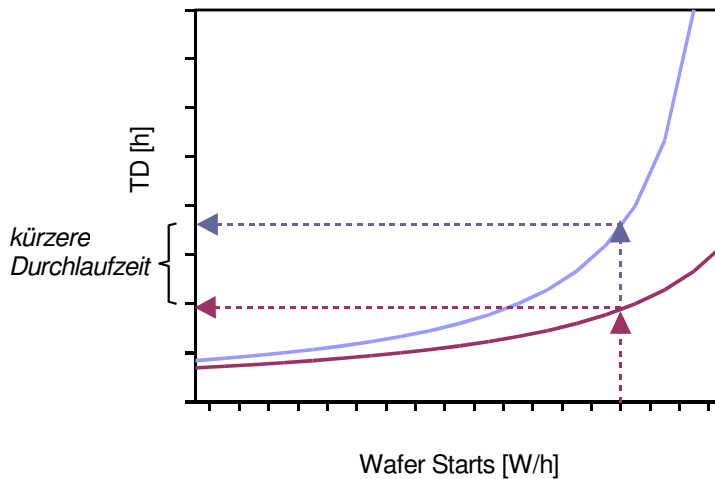


Abbildung 6.15: Verkürzung der Durchlaufzeit bei konstanter Produktionsmenge

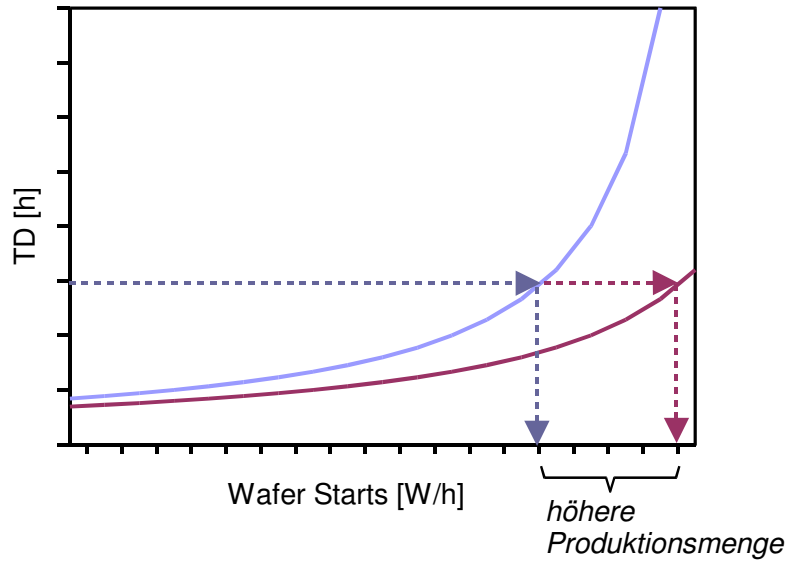


Abbildung 6.16: Steigerung der Produktionsmenge bei konstanter Durchlaufzeit

Wie bereits in Abbildung 6.15 und Abbildung 6.16 deutlich wird, ist die Empfindlichkeit gegenüber Parameteränderungen und damit die Möglichkeit für Veränderungen wegen des stark nichtlinearen Charakters der Kurven erheblich vom Arbeitspunkt abhängig. In Abbildung 6.17 ist dargestellt, welche Zahl an Waferstarts bei einer vorgegebenen maximalen Durchlaufzeit möglich ist und welcher WIP-Bestand sich in Abhängigkeit von der Durchlaufzeit ergibt. Man kann erkennen, dass beim Betrieb der Linie nahe an der Kapazitätsgrenze Steigerungen der Produktionsmenge nur durch Inkaufnahme erheblicher Steigerungen der Durchlaufzeit möglich sind. Bei niedrigeren Durchlaufzeiten ist aufgrund der größeren Kurvensteigung die Empfindlichkeit gegenüber Änderungen dagegen viel größer.

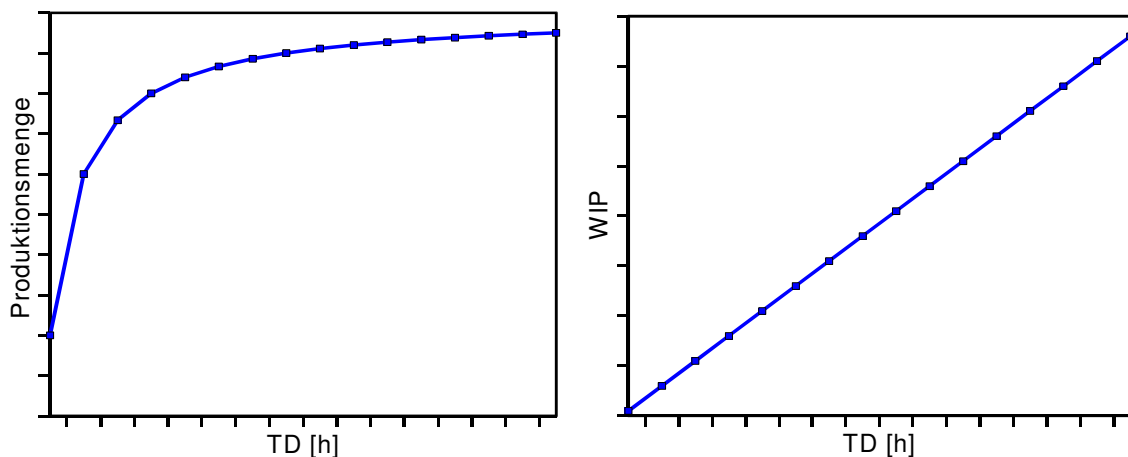


Abbildung 6.17: Abhängigkeit von Produktionsmenge und WIP von der Durchlaufzeit

Die gleichen Überlegungen gelten analog für den Zusammenhang zwischen Produktionsmenge und WIP-Bestand. Auch hier geht die Produktionsmenge bei sehr hohem WIP praktisch in eine Sättigung über, d.h., dass zusätzlich in die Fabrik eingeschleuste Wafer keine Steigerung der Produktionsmenge mehr bewirken. Die Durchlaufzeit steigt ausge-

hend von einem Grundniveau in Abhängigkeit des WIP-Bestands an, wie dies in Abbildung 6.18 dargestellt ist (siehe auch [123]).

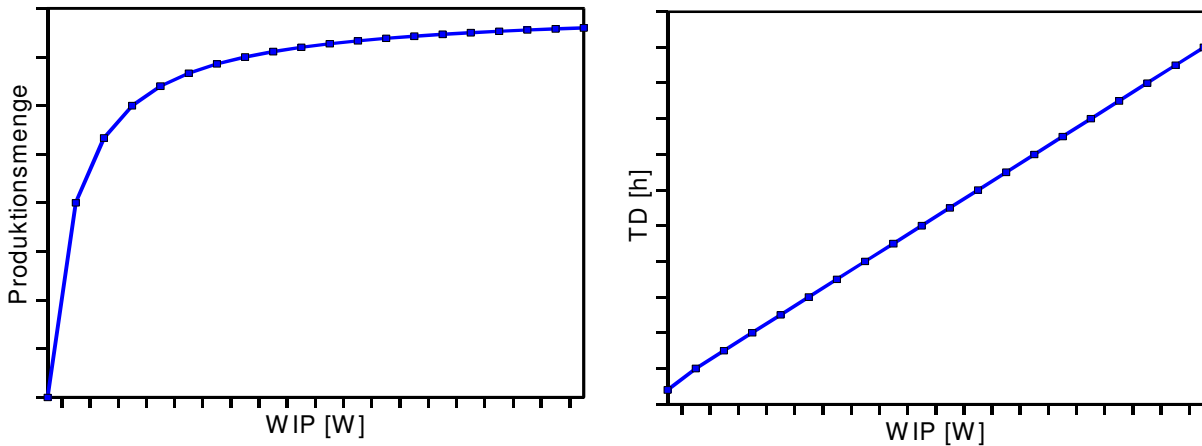


Abbildung 6.18: Abhängigkeit von Produktionsmenge und Durchlaufzeit vom WIP-Bestand

6.3 Einfluss der Variabilität auf die Betriebskennlinie

6.3.1 Schwankungen im Prozessfluss

Bei praktisch allen Fertigungsprozessen treten Schwankungen auf. Eine 100%ige Kontrolle aller Prozessparameter, wie z.B. Temperatur, Druck, Positioniergenauigkeit, ist im allgemeinen nicht möglich. Die Schwankungen eines Prozesses können durch ihren Mittelwert und ihre Standardabweichung beschrieben werden. Wie in Abbildung 6.19 dargestellt, ist es möglich, dass zwei Prozesse zwar den gleichen Mittelwert aufweisen, sich jedoch in ihrer Standardabweichung unterscheiden.

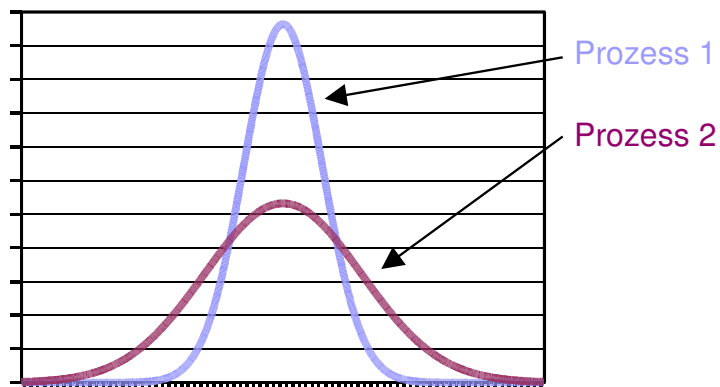


Abbildung 6.19: Prozesse mit unterschiedlicher Variabilität

Prozess 1 weist hier die geringere Standardabweichung auf verglichen mit Prozess 2. Man kann sagen, dass Prozess 1 besser kontrolliert ist, da bei ihm geringere Schwankungen

auftreten. Prozess 2 hat also eine höhere Variabilität und höhere Streuung als Prozess 1. Bezogen auf die Ankunftsrate an einem Gerät äußert sich Variabilität wie qualitativ in Abbildung 6.20 dargestellt. Bei geringer Variabilität ist die Ankunft der Lose am Gerät regelmäßig, bei hoher Variabilität unregelmäßig mit stark schwankenden Zeitabständen zwischen den einzelnen Ankünften.

Der für den Fall des allgemeinen Warteschlangensystems eingeführte Variationskoeffizient läßt sich aus der Standardabweichung und dem Mittelwert des zugehörigen Prozesses berechnen [164, S. 252], [110].

$$c = \frac{\sigma}{\bar{x}} \quad (6.41)$$

Ziel der Fertigung ist es im allgemeinen gut kontrollierte Prozesse mit geringer Standardabweichung zu erreichen.

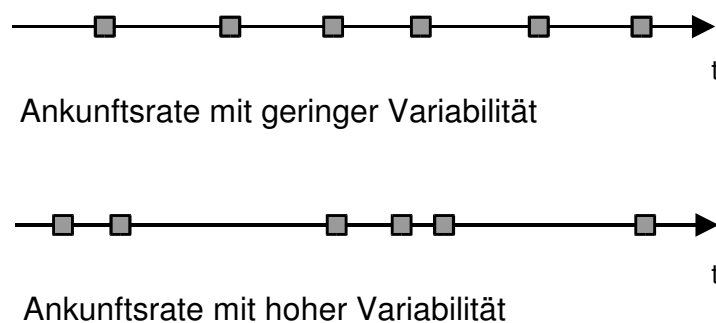


Abbildung 6.20: Variabilität bei der Ankunftsrate [164]

6.3.2 Anwendbarkeit der einzelnen Warteschlangenmodelle auf die Prozesse der Halbleiterfertigung

Wie in Kapitel 6.1 und 6.2 beschrieben, lassen sich Warteschlangenmodelle für Systeme mit unterschiedlicher Verteilung der Ankunfts- und Prozesszeiten aufstellen. Im allgemeinen Fall werden Variationskoeffizienten c_a und c_e für die Ankunfts- und Prozessrate angenommen. In einem realen Fertigungssystem wie der Halbleiterfertigung, bei dem Teile zwischen den einzelnen Geräten hin- und hertransportiert werden, unterschiedliche Produkte gefertigt werden und Eingriffe des Bedienpersonals an manchen Stellen notwendig sind, ist es sinnvoll die Modelle für den allgemeinen Fall unter Berücksichtigung der Prozessvariabilität zu betrachten.

Auch die funktionelle Aufteilung des Reinraums und die Anordnung der Geräte, die eher als beispielsweise eine Fließbandfertigung längere Transportwege und Zwischenlager erfordern, sprechen für eine Anwendung der Warteschlangenmodelle mit allgemeiner Verteilung der Ankunfts- und Bearbeitungsrate.

6.4 Einfluss verschiedener Eingangsparameter auf die Betriebskennlinie

6.4.1 Gerätedurchsatz

Wie in den vorangegangenen Kapiteln dargestellt, zeigt die Betriebskennlinie bei hohem Nutzungsgrad ein stark nichtlineares Verhalten. Je nach Produktionsgeschwindigkeit des vorhandenen Gerätes ergeben sich damit längere bzw. kürzere Durchlaufzeiten und WIP-

Bestand. In Abbildung 6.21 sind die Durchlaufzeit und der WIP-Bestand für ein G/G/1-System mit $c_a=c_e=1,5$ in Abhängigkeit von der Ankunftsrate dargestellt. Dabei wird der Gerätedurchsatz um 10% variiert.

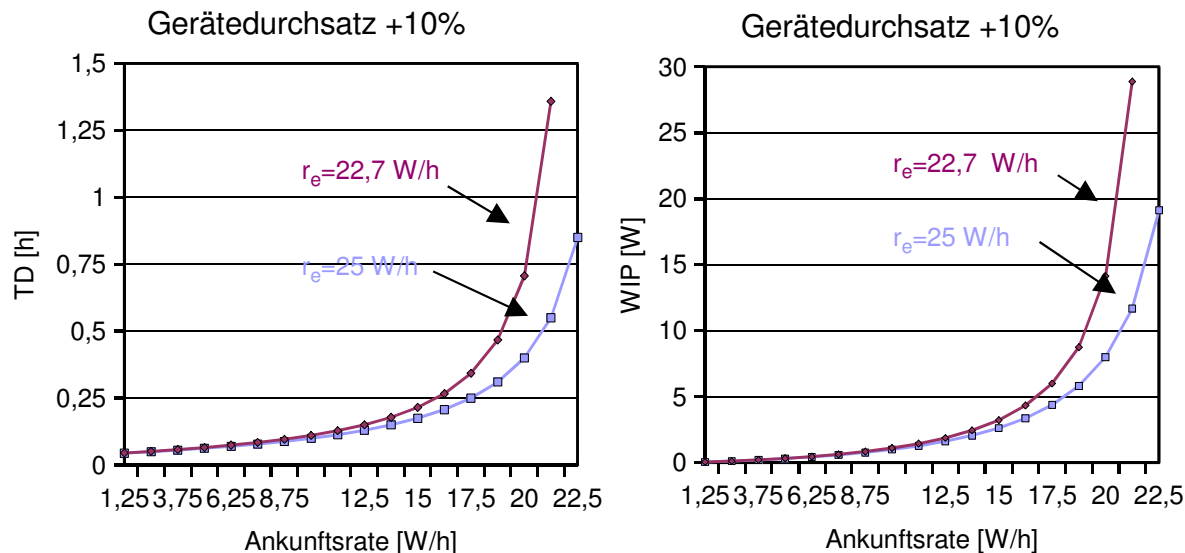


Abbildung 6.21: TD und WIP bei gesteigertem Gerätedurchsatz

Über den Nutzungsgrad aufgetragen würde sich hier bei beiden Geräten die identische Kurve für die Durchlaufzeit ergeben. Dabei ist allerdings zu berücksichtigen, dass, wie in Abbildung 6.21 dargestellt, das schnellere Gerät in absoluten Zahlen mehr Wafer bearbeitet.

6.4.2 Unterbrechungen im Prozessablauf

Wie in Kapitel 6.1.2 dargestellt, können reale Produktionsgeräte nicht 100% der Zeit für die Produktion genutzt werden. Stillstandszeiten können entstehen durch unvorhergesehene und unkontrollierbare Störungen wie z.B. Stromausfälle, technische Probleme oder durch geplante Maßnahmen wie z.B. Setupzeiten oder regelmäßige Wartungsarbeiten. Beide Formen der Produktionsunterbrechung wirken sich sowohl auf den erreichbaren Durchsatz des Gerätes wie auch auf die Variabilität des Prozesses aus.

Für unvorhergesehene Störungen läßt sich die Bearbeitungsrate des Gerätes folgendermaßen modifizieren [164, S. 256 ff.]:

$$r_{ebr} = r_e \cdot A = r_e \cdot \frac{m_f}{m_f + m_r} \tag{6.42}$$

r_{ebr} = Bearbeitungsrate unter Berücksichtigung von Breakdowns

m_f = MTTF = Mean Time To Failure

m_r = MTTR = Mean Time To Repair

A kann als Availability des Geräts bezeichnet werden.

Die Variabilität des Prozesses ergibt sich zu [164, S. 261]:

$$c_{ebr}^2 = c_e^2 + (1 + c_r^2) \cdot A \cdot (1 - A) \cdot m_r \cdot r_e \quad (6.43)$$

c_{ebr} = Variabilität unter Berücksichtigung von Breakdowns

c_r = Variabilität des Reparaturprozesses

Mit angenommenen Zahlen für MTTF von 300h, MTTR von 5h $r_e=25W/h$, $c_e=1,5$ und $c_r=1,0$ ergibt sich ein effektives $r_{ebr}=24,59W/h$ und $c_{ebr}= 2,51$. Die resultierenden Kurven sind in Abbildung 6.22 dargestellt.

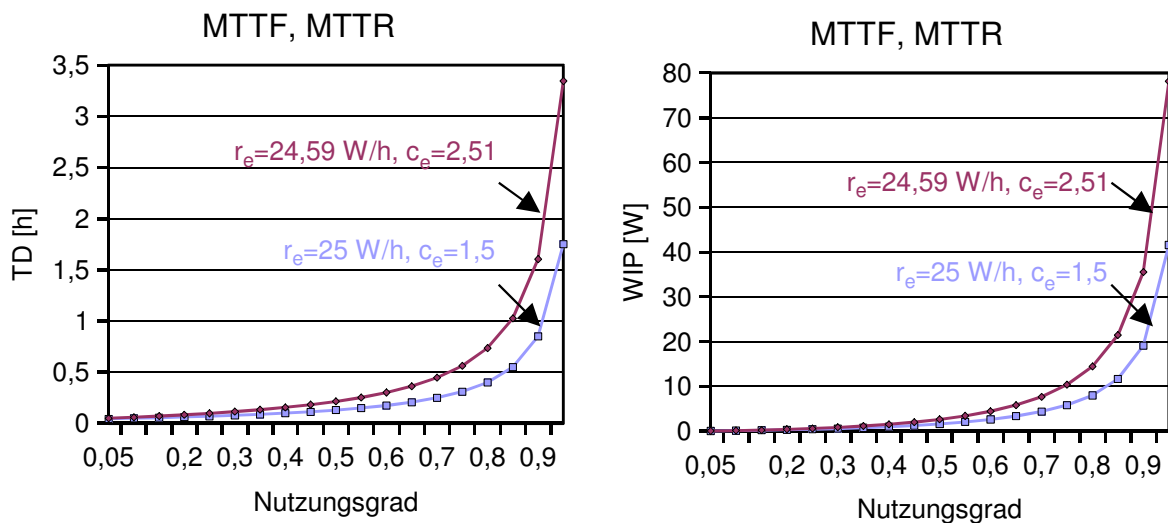


Abbildung 6.22: TD und WIP unter Berücksichtigung von Prozessstörungen

Die oben angegebenen Gleichungen lassen sich noch um den Einfluss von geplanten Setups, die vor der Produktion eines anderen Produkttyps durchgeführt werden müssen, erweitern. Damit ergibt sich für die Bearbeitungsrate [164, S. 259 ff.]:

$$t_{ebrs} = \frac{t_e}{A} + \frac{t_s}{N_s} = \frac{t_e}{\frac{m_f}{m_f + m_r}} + \frac{t_s}{N_s} \quad (6.44)$$

t_s = Zeitdauer für Setup

N_s = Zahl der prozessierten Teile zwischen Setups

Die Variabilität unter Berücksichtigung von Setups berechnet sich zu [164, S. 261]:

$$c_{ebrs}^2 = r_e^2 \cdot (t_0^2 c_0^2 + \frac{t_s^2 c_s^2}{N_s} + \frac{N_s - 1}{N_s^2} \cdot t_s^2) \quad (6.45)$$

c_s = Variabilität des Setupprozesses

Mit $t_s=0,1h$, $N_s=50$ und $c_s=1,5$ ergibt sich für das obige Beispiel unter Berücksichtigung von Störungen und Setups ein $r_{ebrs}=23,44W/h$ und ein $c_{ebrs}=2,46W/h$ (Abbildung 6.23).

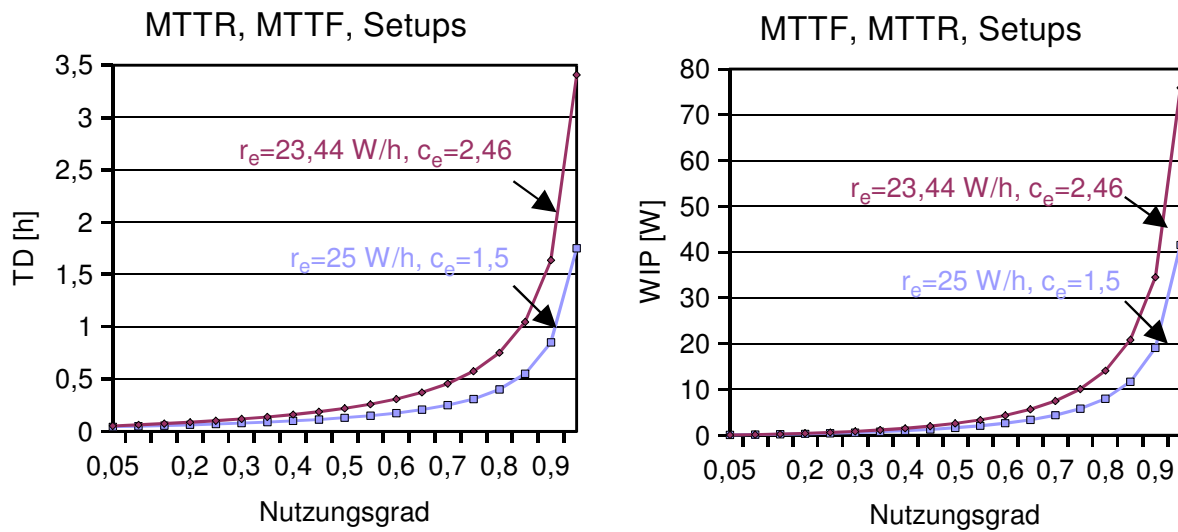


Abbildung 6.23: TD und WIP unter Berücksichtigung von Prozessstörungen und Setups

6.4.3 Losgröße

Abhängig von der Losgröße ändert sich die Variabilität eines Prozesses. Hier soll angenommen werden, dass alle Teile eines Loses gemeinsam transportiert werden und gemeinsam beim Gerät ankommen. Die Möglichkeit, dass Schwankungen auftreten und damit die Variabilität reduziert sich also mit der Losgröße. Die Variabilität für ein Los ergibt sich zu [164, S. 280]:

$$c_{\text{batch}} = \frac{c_0}{\sqrt{n}} \quad (6.46)$$

n = Zahl der Teile pro Los

Unter der Voraussetzung, dass alle Teile des Loses nacheinander gefertigt werden (single wafer processing), ergibt sich der Zeitaufwand um ein Los zu prozessieren aus der Bearbeitungsgeschwindigkeit des Geräts, der Zahl der Teile pro Los und der Setupzeit.

$$t_{\text{batch}} = \frac{n}{r_e} + t_s \quad (6.47)$$

Die Ankunftsrate am Gerät ausgedrückt in Losen ist:

$$r_{\text{abatch}} = \frac{r_a}{n} \quad (6.48)$$

Damit folgt für den Nutzungsgrad des Gerätes [164, S.309]:

$$u = r_a \cdot \left(\frac{1}{r_e} + \frac{t_s}{n} \right) \quad (6.49)$$

Mit der in Gleichung (6.14) angegebenen Wartezeit

$$TD_q = \left(\frac{c_a^2 + c_e^2}{2} \right) \cdot \left(\frac{u}{1-u} \right) \cdot t_{batch} \quad (6.50)$$

folgt für die gesamte Durchlaufzeit des Loses:

$$TD = TD_q + \frac{n}{r_e} + t_s = \left(\frac{c_a^2 + c_e^2}{2} \right) \cdot \left(\frac{u}{1-u} \right) \cdot t_{batch} + \frac{n}{r_e} + t_s \quad (6.51)$$

Mit Gleichung (6.51) ergeben sich in Abbildung 6.24 folgende Kurven für die Durchlaufzeit in Abhängigkeit von der Losgröße. Dabei wurde $r_e=25W/h$, $c_a=c_e=1,5$, $t_s= 0,1h$ angenommen. Angegeben wird hier die Durchlaufzeit in Abhängigkeit vom Nutzungsgrad, nicht von der absoluten Zahl der WSPM.

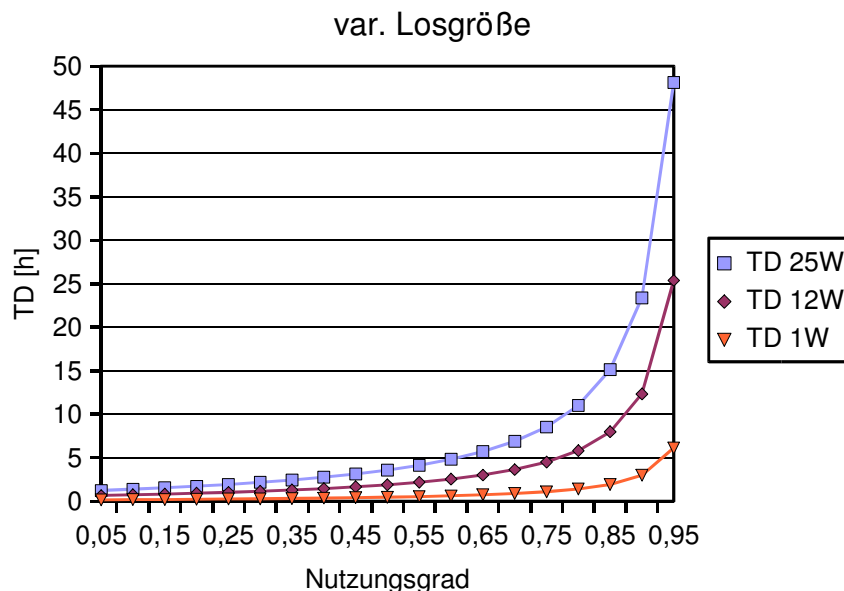


Abbildung 6.24: Durchlaufzeit in Abhängigkeit von der Losgröße

In Abhängigkeit von der Setupzeit ergibt sich ein unterschiedliches Verhältnis der Durchlaufzeiten bei variabler Losgröße. Je größer die Setupzeit, desto günstiger wirken sich größere Losgrößen aus.

Für Batch-Prozesse, bei denen mehrere Teile gleichzeitig prozessiert werden, können die Gleichungen folgendermaßen modifiziert werden, unter der Voraussetzung, dass Teile einzeln am Gerät ankommen [164]:

$$WTBT = \frac{B-1}{2r_a} \quad (6.52)$$

WTBT = Wartezeit bis zum Auffüllen des Batches
 B = Zahl der Teile im Prozess-Batch

$$u = \frac{r_a}{B \cdot r_e} \quad (6.53)$$

$$TD_q = \left(\frac{c_a^2/B + c_e^2}{2} \right) \cdot \left(\frac{u}{1-u} \right) \cdot \frac{1}{r_e} \quad (6.54)$$

Die Gesamtdurchlaufzeit setzt sich zusammen aus der Wartezeit bis alle Teile am Gerät angekommen sind, der Wartezeit vor dem Gerät und der Prozesszeit.

$$TD = WTBT + TD_q + 1/r_e \quad (6.55)$$

$$TD = \frac{B-1}{2r_a} + \left(\frac{c_a^2/B + c_e^2}{2} \right) \cdot \left(\frac{u}{1-u} \right) \cdot \frac{1}{r_e} + \frac{1}{r_e} \quad (6.56)$$

6.4.4 Blocking

In realen Systemen ist der zur Verfügung stehende Platz für die sich in der Warteschlange befindenden Teile meist beschränkt. Das bedeutet, dass die Warteschlange nicht bis zu einer unendlichen Größe anwachsen kann, sondern ab einer bestimmten Menge die weitere Anlieferung von Teilen an das Gerät gestoppt werden muss, d.h. die Ankunftsrate r_a am Gerät kann sich auf einen effektiven Durchsatz TH reduzieren. Dieser Effekt wird als Blocking bezeichnet und beeinflusst Durchsatz und Durchlaufzeit des Systems. Je kleiner die zur Verfügung stehende Lagerkapazität, desto stärker werden die Auswirkungen des Blocking-Effekts.

Ausgehend von Little's Law kann die Durchlaufzeit für den Blocking-Fall ausgedrückt werden durch:

$$TD_b = \frac{WIP_b}{TH_b} \quad (6.57)$$

wobei Index b den Fall mit Blocking bezeichnet.

TH = effektiver Durchsatz des Systems

Für ein M/M/1/b-System ergibt sich [164, S. 275 ff.]

$$WIP_b = \frac{u}{1-u} - \frac{(b+1) \cdot u^{b+1}}{1-u^{b+1}} \quad (6.58)$$

b = Größe des Zwischenlagers

$$TH = \frac{1-u^b}{1-u^{b+1}} r_a \quad (6.59)$$

Abbildung 6.25 zeigt den Einfluss der Größe des Zwischenlagers auf den effektiven Durchsatz des Systems. Bei sehr großem Zwischenlager ergeben sich die gleichen Ergebnisse wie für das System ohne Blocking.

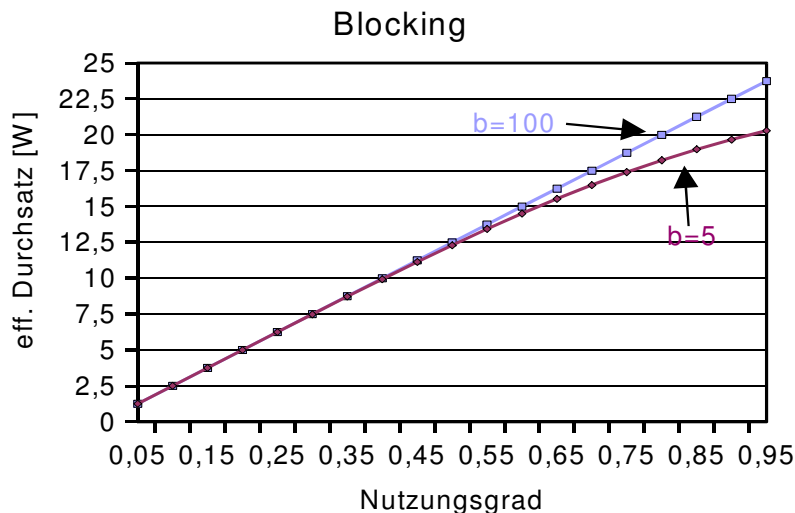


Abbildung 6.25: Effektiver Durchsatz bei unterschiedlicher Größe des Zwischenlagers im M/M/1/b-System

Für den allgemeinen Fall des G/G/1-Systems läßt sich ein korrigierter Nutzungsgrad ρ definieren [164, S. 278]

$$\rho = \frac{WIP_{nb} - u}{WIP_{nb}} \tag{6.60}$$

wobei Index $_{nb}$ den Fall ohne Blocking bezeichnet.

Mit Gleichung (6.59) ergibt sich

$$TH \approx \frac{1 - u \rho^{b-1}}{1 - u^2 \rho^{b-1}} r_a \tag{6.61}$$

Für WIP und Durchlaufzeit des G/G/1-Systems mit Blocking lassen sich die folgenden Abschätzungen angeben:

$$WIP < \min(WIP_{nb}, b - 1) \tag{6.62}$$

$$TD > \min \frac{(WIP_{nb}, b - 1)}{TH} \tag{6.63}$$

In Abbildung 6.26 wird deutlich, dass auch beim Blocking der Einfluss der Variabilität im G/G/1-System sehr ausgeprägt ist. Je kleiner die Variabilität ist, desto geringer wird die Reduktion des effektiven Durchsatzes bei beschränkter Lagerkapazität. In Abbildung 6.26 werden die entsprechenden Kurven für $b=5$ aufgetragen.

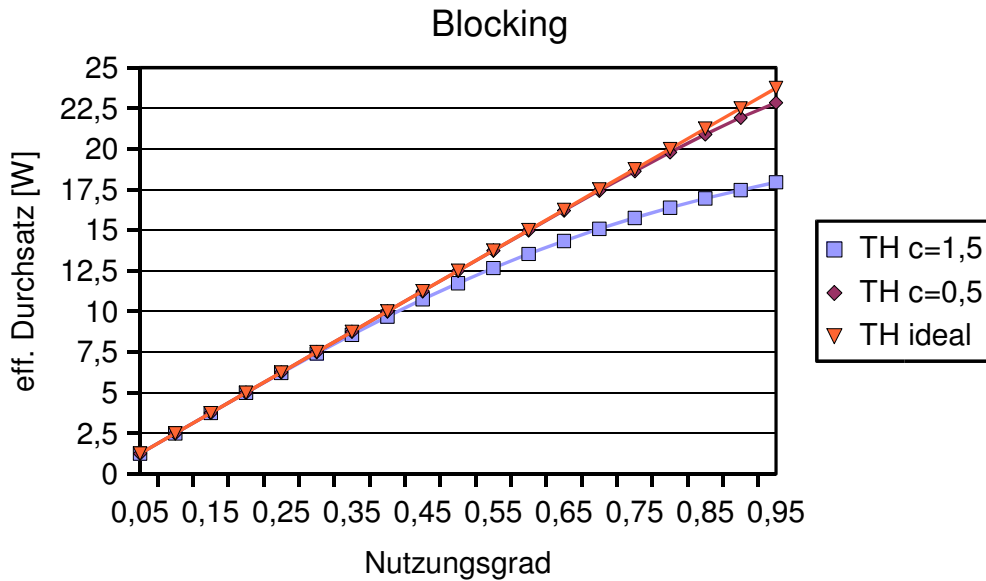


Abbildung 6.26: Effektiver Durchsatz bei unterschiedlichem Variationskoeffizienten für $b=5$ im G/G/1/b-System

6.4.5 Transportzeit

Für den Transport der Lose zwischen den Geräten muss eine gewisse Zeit aufgewendet werden, die sich zur Prozesszeit addiert. In Abbildung 6.27 wird für ein Gerät mit $r_e=25W/h$ und $c_a=c_e=1,5$ eine pauschale Transportzeit von 0,25h bzw. 0,1h angenommen.

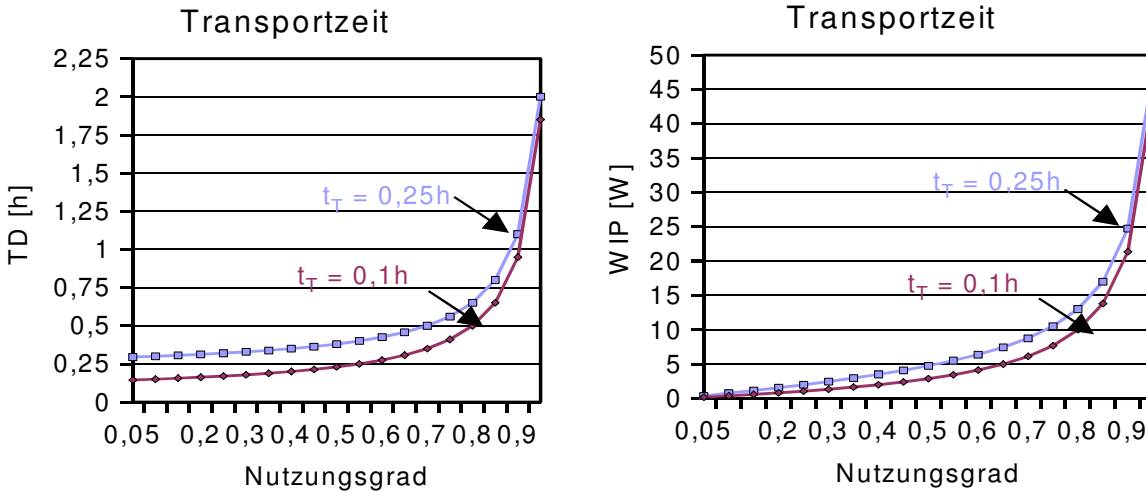


Abbildung 6.27: Auswirkungen der Transportzeit auf TD und WIP

6.4.6 Variabilität

Wie aus den grundlegenden Gleichungen ersichtlich, sind Durchlaufzeit und damit WIP-Bestand abhängig vom Wert der Variationskoeffizienten c_a und c_e . In Abbildung 6.28 sind die Auswirkungen einer Verringerung der Variabilität um 10% dargestellt.

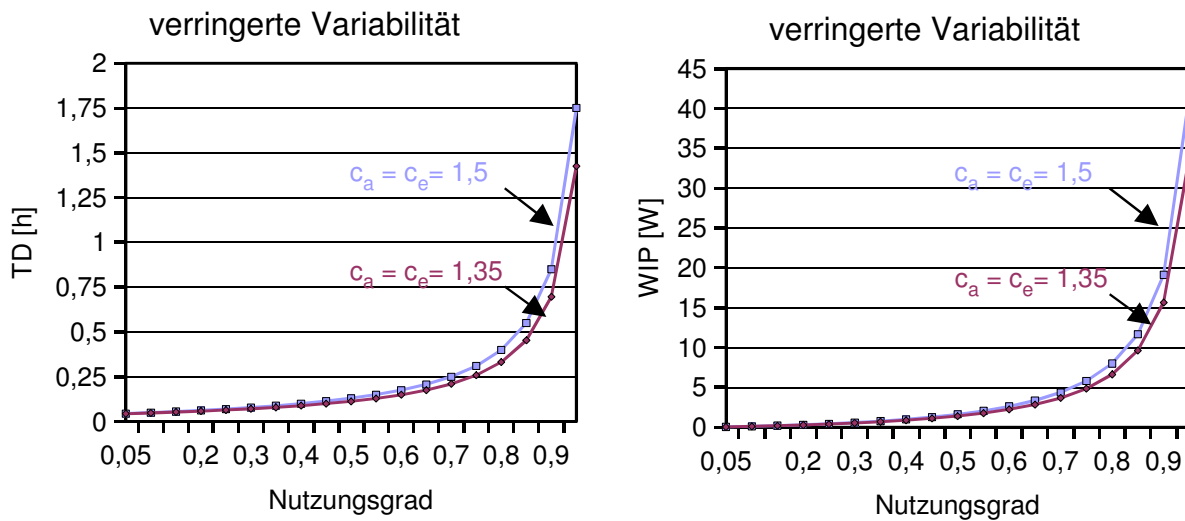


Abbildung 6.28: Auswirkungen einer verringerten Variabilität auf Durchlaufzeit und WIP

Wie bereits erwähnt, stellt die Verringerung der Variabilität ein sehr wirksames Mittel zur Verbesserung der Produktion dar.

6.4.7 Position im Prozessfluss

Die Auswirkungen von Variabilität auf die Durchlaufzeit sind nicht an jeder Stelle im Prozessfluss gleich. Abhängig vom Wert der Variationskoeffizienten und dem Durchsatz der Geräte an verschiedenen Positionen in der Prozesskette ergeben sich unterschiedliche Daten für die Gesamt-Durchlaufzeit.

In Abbildung 6.29 sind die Auswirkungen der Variabilität an verschiedenen Stellen im Prozessfluss am Beispiel von drei in Serie geschalteten Geräten mit jeweils $r_e=20$ W/h dargestellt. Als c_{a1} wurde 1 angenommen. Es zeigt sich, dass in diesem Beispiel hohe Variabilität am Anfang der Prozesskette ungünstigere Auswirkungen auf Durchlaufzeit und WIP hat, als am Ende des Prozessflusses.

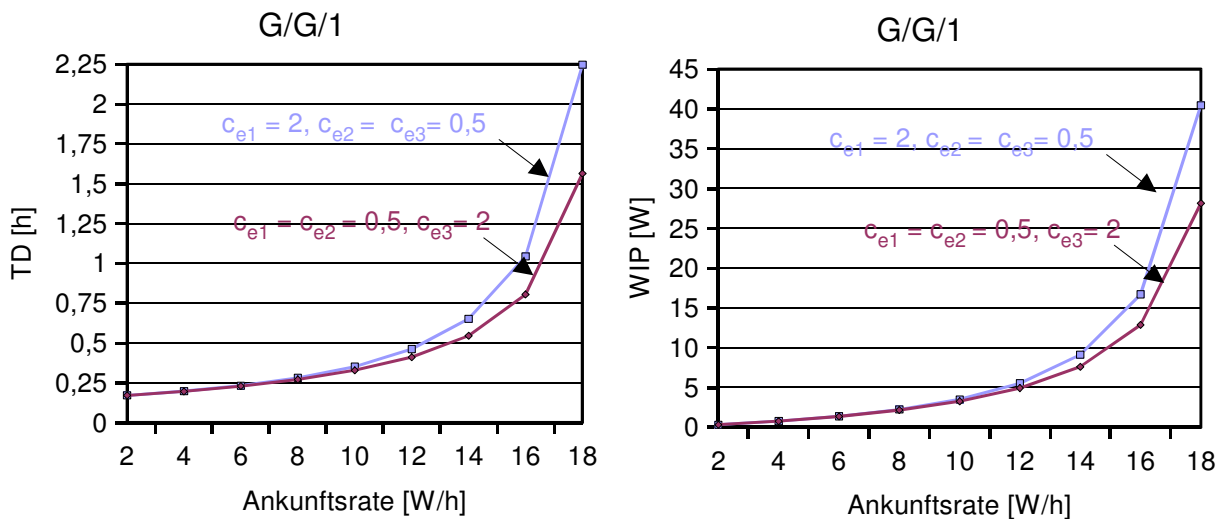


Abbildung 6.29: Auswirkungen der Variabilität an verschiedenen Stellen im Prozessfluss

6.5 Erstellung eines Warteschlangenmodells für einen Teil des Prozessflusses

6.5.1 Gerätedaten einer realen Fertigung

Ausgehend von den Überlegungen in den vorausgegangenen Abschnitten soll ein Warteschlangenmodell für einen Teil des Prozessflusses erstellt werden, mit dem die Einflüsse der verschiedenen Parameter auf die Bewertungsgrößen und damit auf die Effizienz berechnet werden können. Zur Überprüfung der Gültigkeit des Modells werden Daten aus einer realen Fertigung herangezogen.

Von Infineon wurden Daten aus einer kompletten Anlagengruppe zur Verfügung gestellt. Es handelt sich dabei um 14 Geräte gleicher Art, für die alle Vorgänge innerhalb eines Zeitraums von 8 Tagen aufgezeichnet wurden. Insgesamt fanden im Zeitraum 6.6.2005 bis 14.6.2005 in 195,5h ca. 7000 einzelnen Vorgänge statt, bei 72 verschiedenen Operations. Die Daten wurden im normalen Betrieb aufgenommen, es wurden keine nachträglichen Änderungen vorgenommen.

In der folgenden Tabelle 6.2 ist ein Ausschnitt aus den vorliegenden Daten dargestellt.

LOT	OWNER	ENTER_OPERATION _TIME_STAMP	MOVEIN_TIME_STAMP	TIME_STAMP_Moveout	OPERATION	QTY_OUT_ 1	EQUIPMENT
ID_1438	PROD	06.06.2005 13:16	06.06.2005 23:04	07.06.2005 00:28	7346	25	Tool_ID-06
ID_1402	PROD	06.06.2005 13:57	06.06.2005 23:04	07.06.2005 00:28	7693	25	Tool_ID-06
ID_2192	PROD	06.06.2005 14:43	06.06.2005 23:49	07.06.2005 01:03	7060	25	Tool_ID-06
ID_2009	PROD	06.06.2005 15:28	06.06.2005 23:49	07.06.2005 01:04	7060	25	Tool_ID-06
ID_2429	PROD	06.06.2005 18:43	06.06.2005 23:16	07.06.2005 00:37	7060	25	Tool_ID-03
ID_2008	PROD	06.06.2005 18:53	06.06.2005 23:16	07.06.2005 00:37	7060	25	Tool_ID-03
ID_2469	PROD	06.06.2005 19:33	07.06.2005 00:38	07.06.2005 02:13	5430	25	Tool_ID-03
ID_2469	PROD	06.06.2005 19:33	07.06.2005 00:38	07.06.2005 00:38	5430	25	Tool_ID-03
ID_3124	PROD	06.06.2005 19:38	06.06.2005 22:21	07.06.2005 01:19	4633	25	Tool_ID-10
ID_2453	PROD	06.06.2005 19:50	07.06.2005 00:38	07.06.2005 02:15	5430	25	Tool_ID-03
ID_2453	PROD	06.06.2005 19:50	07.06.2005 00:38	07.06.2005 00:38	5430	25	Tool_ID-03
ID_4134	PROD	06.06.2005 20:18	06.06.2005 22:21	07.06.2005 04:08	3267	25	Tool_ID-10
ID_2190	PROD	06.06.2005 20:54	07.06.2005 00:30	07.06.2005 01:53	7060	25	Tool_ID-06
ID_2190	PROD	06.06.2005 20:54	07.06.2005 00:30	07.06.2005 00:30	7060	25	Tool_ID-06
ID_2300	PROD	06.06.2005 21:26	06.06.2005 23:42	07.06.2005 00:46	7726	25	Tool_ID-11
ID_2011	PROD	06.06.2005 21:30	07.06.2005 00:30	07.06.2005 01:53	7060	25	Tool_ID-06
ID_2011	PROD	06.06.2005 21:30	07.06.2005 00:30	07.06.2005 00:30	7060	25	Tool_ID-06
ID_5433	PROD	06.06.2005 21:33	06.06.2005 23:32	07.06.2005 00:22	1490	25	Tool_ID-09
ID_6007	PROD	06.06.2005 21:38	06.06.2005 23:16	07.06.2005 00:08	1490	25	Tool_ID-11
ID_1473	PROD	06.06.2005 21:59	07.06.2005 00:58	07.06.2005 01:38	7669	25	Tool_ID-12
ID_1473	PROD	06.06.2005 21:59	07.06.2005 00:58	07.06.2005 00:58	7669	25	Tool_ID-12
ID_1173	PROD	06.06.2005 22:04	07.06.2005 00:58	07.06.2005 00:58	7726	25	Tool_ID-12
ID_1173	PROD	06.06.2005 22:04	07.06.2005 00:58	07.06.2005 01:38	7726	25	Tool_ID-12
ID_3040	PROD	06.06.2005 22:07	06.06.2005 23:19	07.06.2005 00:05	7560	25	Tool_ID-12
ID_3114	PROD	06.06.2005 22:13	06.06.2005 23:20	07.06.2005 00:04	3830	25	Tool_ID-01
ID_1245	PROD	06.06.2005 22:16	06.06.2005 23:46	07.06.2005 00:35	7726	25	Tool_ID-14
ID_4204	TE	06.06.2005 22:19	06.06.2005 23:33	07.06.2005 02:19	3727	25	Tool_ID-13
ID_5432	PROD	06.06.2005 22:23	06.06.2005 23:32	07.06.2005 00:22	1490	25	Tool_ID-09
ID_3088	PROD	06.06.2005 22:25	06.06.2005 23:08	07.06.2005 00:02	4685	25	Tool_ID-04
ID_2195	PROD	06.06.2005 22:30	06.06.2005 23:58	07.06.2005 01:21	7060	25	Tool_ID-03
ID_5210	LEX	06.06.2005 22:32	06.06.2005 23:21	07.06.2005 00:05	4229	25	Tool_ID-01
ID_1117	PROD	06.06.2005 22:33	06.06.2005 23:46	07.06.2005 00:35	7726	25	Tool_ID-14
ID_4133	PROD	06.06.2005 22:34	07.06.2005 00:37	07.06.2005 00:37	3267	25	Tool_ID-14
ID_4133	PROD	06.06.2005 22:34	07.06.2005 00:37	07.06.2005 01:10	3267	25	Tool_ID-14
ID_3014	PROD	06.06.2005 22:37	07.06.2005 00:47	07.06.2005 01:41	7560	25	Tool_ID-11
ID_3014	PROD	06.06.2005 22:37	07.06.2005 00:47	07.06.2005 00:47	7560	25	Tool_ID-11
ID_4169	PROD	06.06.2005 22:37	06.06.2005 23:09	07.06.2005 00:02	5044	24	Tool_ID-04
ID_0395	PROD	06.06.2005 22:37	06.06.2005 23:39	07.06.2005 00:34	7560	25	Tool_ID-04
ID_3117	PROD	06.06.2005 22:38	06.06.2005 23:53	07.06.2005 00:40	3830	25	Tool_ID-09
ID_2165	PROD	06.06.2005 22:39	06.06.2005 23:33	07.06.2005 00:13	7801	25	Tool_ID-02
ID_2045	PROD	06.06.2005 22:40	06.06.2005 23:50	07.06.2005 00:25	8150	25	Tool_ID-10
ID_0424	PROD	06.06.2005 22:44	06.06.2005 23:31	07.06.2005 00:20	9289	25	Tool_ID-05
ID_0421	PROD	06.06.2005 22:44	06.06.2005 23:39	07.06.2005 00:34	7560	25	Tool_ID-04
ID_6039	PROD	06.06.2005 22:45	06.06.2005 23:25	07.06.2005 00:13	1490	25	Tool_ID-14
ID_2324	PROD	06.06.2005 22:45	07.06.2005 00:09	07.06.2005 01:07	7726	25	Tool_ID-11
ID_2324	PROD	06.06.2005 22:45	07.06.2005 00:09	07.06.2005 00:09	7726	25	Tool_ID-11
ID_0229	PROD	06.06.2005 22:48	07.06.2005 00:19	07.06.2005 00:53	7642	25	Tool_ID-13

Tabelle 6.2: Gerätedaten der Anlagengruppe

Bei den einzelnen Spalten handelt es sich dabei um folgende Daten:

LOT: Losnummer

OWNER: Eigentümer / Zuständiger für das Los

ENTER_OPERATION_TIME_STAMP: Ankunft des Loses in der Warteschlange

MOVEIN_TIME_STAMP: Beginn des Prozesses

TIME_STAMP_MOVEOUT: Ende des Prozesses

OPERATION: charakteristische Nummer für den jeweiligen Prozessschritt

QTY_OUT_1: Zahl der prozessierten Wafer

EQUIPMENT: Gerätenummer

Die Zeit zwischen ENTER_OPERATION_TIME_STAMP und MOVEIN_TIME_STAMP

stellt die Wartezeit des Loses vor dem jeweiligen Gerät dar.

Die Prozesszeit für den jeweiligen Vorgang ergibt sich aus der Differenz zwischen TIME_STAMP_MOVEOUT und MOVEIN_TIME_STAMP.

Eine graphische Darstellung der Losankunft und Losbearbeitung wird in Abbildung 6.30 exemplarisch für einen kurzen Zeitraum an Tool 1 angegeben. Die Durchlaufzeit der einzelnen, mit verschiedenen Symbolen dargestellten, Lose sowie die Priorisierung einiger Lose, die trotz späterer Losankunft früher prozessiert werden, ist erkennbar.

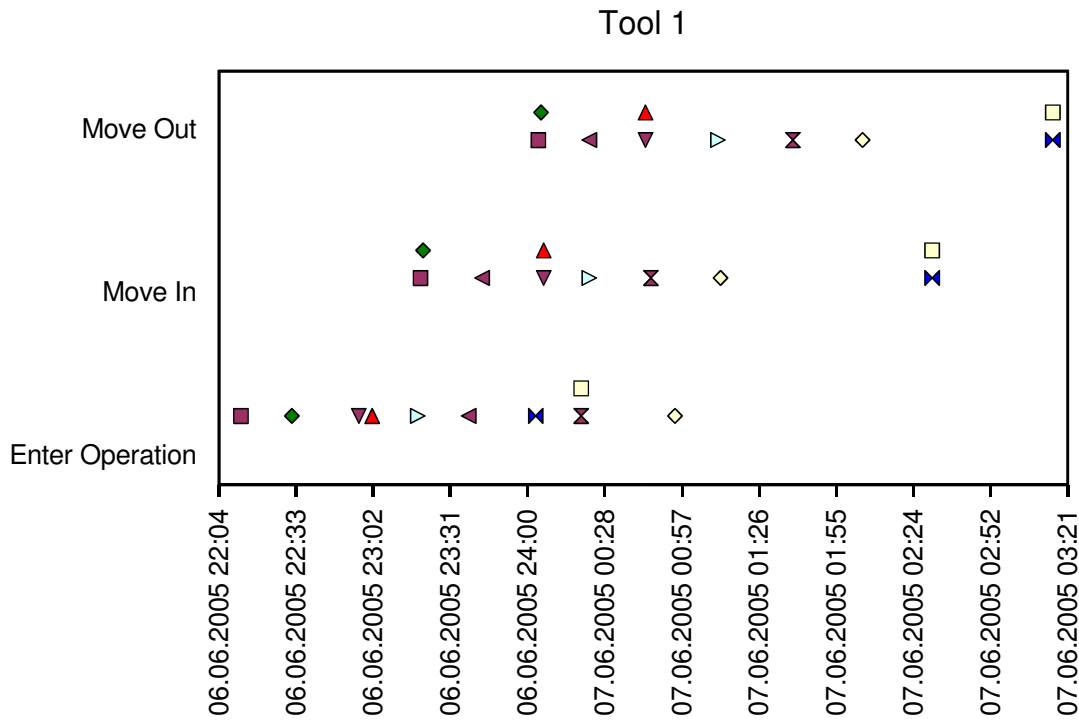


Abbildung 6.30: Grafische Darstellung einzelner Lose an Tool 1

Eine statistische Auswertung über alle bearbeiteten Lose der 14 Geräte ist in Tabelle 6.3 angegeben.

Tabelle 6.3: Maschinenbelegung der einzelnen Geräte

Tool	Prozessierte Lose	MW Prozesszeit [h]	MW Losankunft [h]	MW Wartezeit [h]
1	699	00:48:15	00:15:49	01:09:35
2	615	00:49:47	00:17:59	01:10:52
3	345	01:15:55	00:32:40	01:21:47
4	444	00:59:53	00:24:53	01:20:58
5	658	00:53:28	00:16:45	01:12:59
6	399	01:08:45	00:28:46	01:29:11
7	173	00:45:02	01:00:26	00:56:26
8	21	00:48:40	06:44:03	01:22:03
9	659	00:52:42	00:16:46	01:20:37
10	697	00:48:14	00:16:05	01:18:44

Tool	Prozessierte Lose	MW Prozesszeit [h]	MW Losankunft [h]	MW Wartezeit [h]
11	532	00:49:41	00:20:58	01:20:48
12	721	00:47:56	00:15:23	01:22:20
13	490	00:50:14	00:22:36	01:17:07
14	442	00:51:16	00:25:08	01:18:04

6.5.2 Auswertung der Daten des realen Prozessflusses

Bei den Geräten handelt es sich um eine Anlagengruppe gleichartiger Geräte, auf denen jeweils mehrere unterschiedliche Operations ausgeführt werden können. Bei der Analyse der Daten stellt sich heraus, dass einige Prozesse exklusiv von Tool 7 durchgeführt werden, die anderen Operations verteilen sich gleichmäßig auf die restlichen 13 Geräte. Zur Auswertung der Daten wird deshalb für Tool 7 ein Single Server Modell, für die restlichen Geräte 1-6 und 8-14 ein Multi Server Modell angenommen (Abbildung 6.31). Bei den Multi Server Tools wurden im betrachteten Zeitraum insgesamt 6722 Lose bearbeitet, beim Single Server Tool nur 173 Lose.

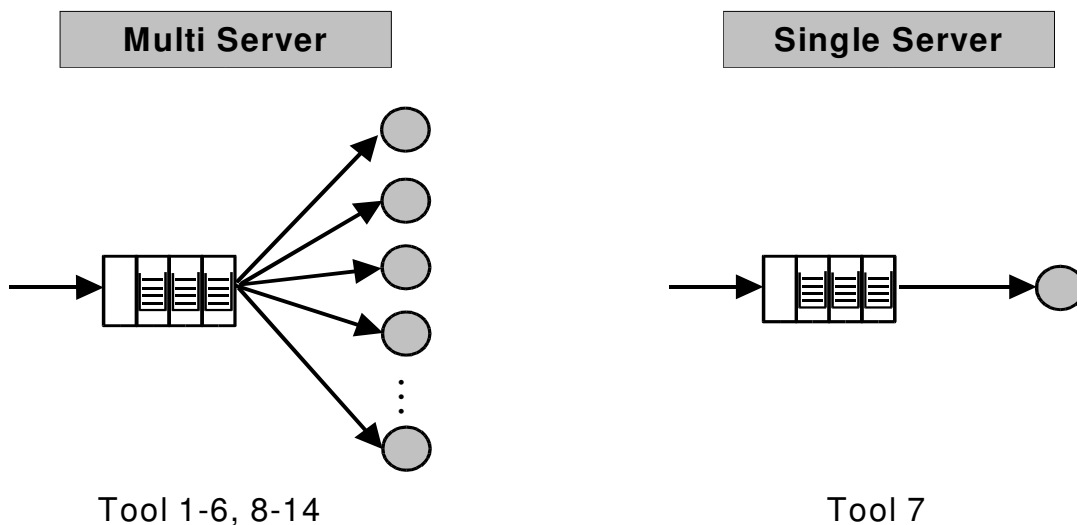


Abbildung 6.31: Auswertung der Gerätedaten, Unterteilung in Multi Server und Single Server Tool

Von besonderem Interesse ist, inwieweit sich die realen Daten von den theoretischen mathematischen Verteilungen, die in der Warteschlangentheorie betrachtet werden, unterscheiden. Um diese zu analysieren, wird für die jeweiligen Geräte die Zeitdauer zwischen den Losankünften ($\text{ENTER OPERATION TIME STAMP}_{n+1} - \text{ENTER OPERATION TIME STAMP}_n$) und die Prozessdauer ($\text{TIME STAMP MOVEOUT}_n - \text{MOVEIN TIME STAMP}_n$) ausgewertet.

Für die Geräte 1-6, 8-14 ergibt sich die in Abbildung 6.32 dargestellte Statistik für die Losankunft. Zum Vergleich wurden die Kurven einer idealen Exponential- ($\lambda=0,9$), Pareto- ($\alpha=1, \beta=1$) und Potenzverteilung ($\beta=1, d=1$) mit eingezeichnet. Folgende Funktionen für

die Verteilungen liegen den Kurven zugrunde:

Exponentialverteilung: $f(x) = \lambda e^{-\lambda x}$

Paretoverteilung: $f(x) = \frac{\alpha}{\beta} \left(\frac{\beta}{x}\right)^{\alpha+1}$

Potenzverteilung: $f(x) = \beta \frac{x^{\beta-1}}{d^\beta}$

Es ist erkennbar, dass die realen, gemessenen Daten relativ gut mit der Exponentialverteilung und Paretoverteilung übereinstimmen.

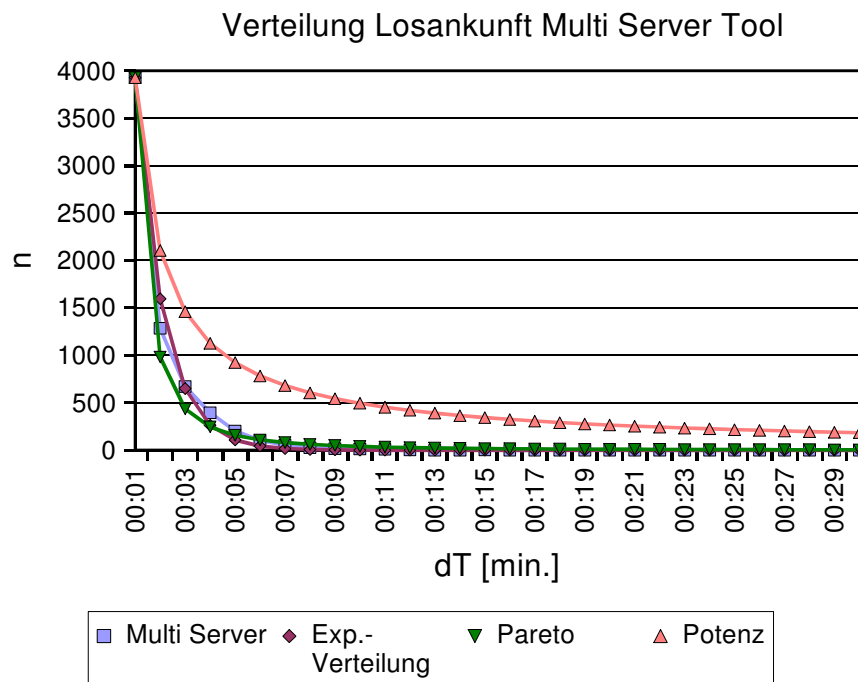


Abbildung 6.32: Losankunft am Multi Server Tool im Vergleich mit verschiedenen Verteilungen

Für Tool 7 wird in Abbildung 6.33 der Vergleich mit Exponential- ($\lambda=0,9$) und Paretoverteilung ($\alpha=1, \beta=1$) angegeben. Aufgrund der relativ geringer Anzahl von Losankünften kann die Statistik hier noch relativ stark von den idealen Verteilungen abweichen.

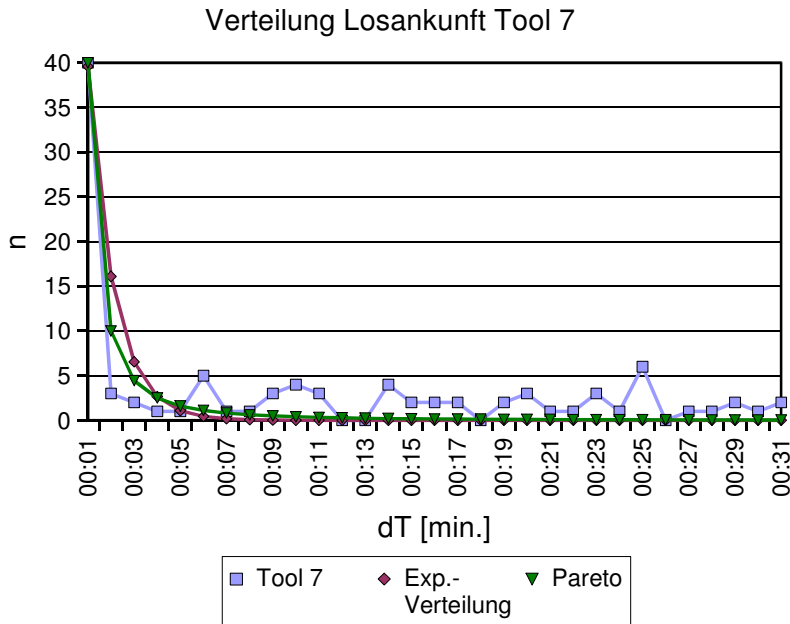


Abbildung 6.33: Losankunft am Single Server Tool im Vergleich mit verschiedenen Verteilungen

Zur weiteren Verdeutlichung wird die Verteilung der Ankunftszeit in Abbildung 6.34 gemeinsam mit der idealen Exponentialverteilung im logarithmischen Maßstab dargestellt. Für eine Exponentialverteilung ergibt sich bei dieser Darstellungsform die eingezeichnete Gerade. Besonders für den Multi Server Fall ergibt sich eine gute Übereinstimmung der Kurven, wie dies auch in der nicht-logarithmischen Darstellung in Abbildung 6.32 zu erkennen ist.

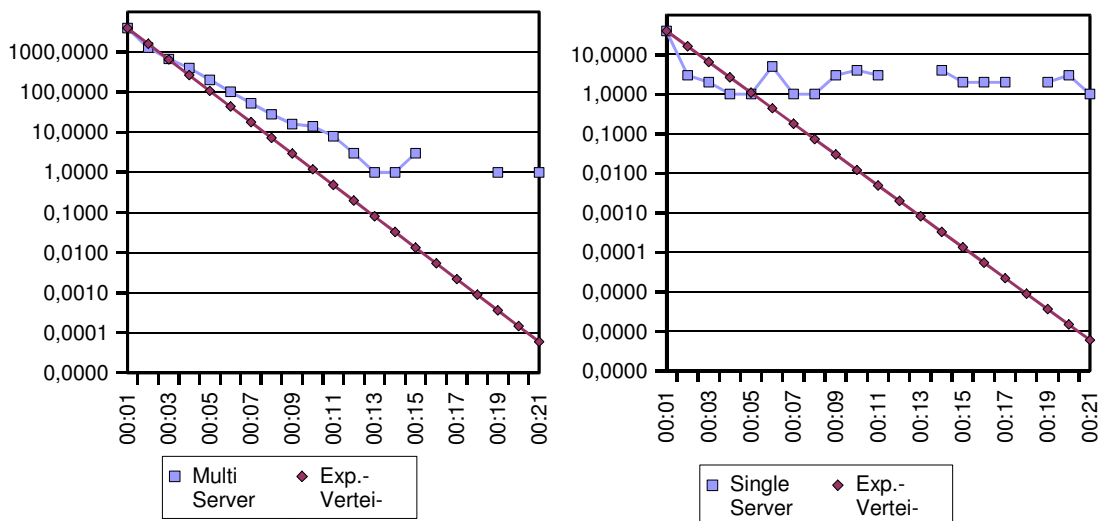


Abbildung 6.34: Darstellung der Losankunft im logarithmischen Maßstab

Die Prozessdauer der an den Geräten ankommenden Lose ist für die parallelen Geräte in Abbildung 6.35 zusammen mit den Kurven für eine ideale Standardverteilung ($\mu=50$, $\sigma=17$), Maxwell- ($\sigma=33$), Rayleigh- ($\Theta=60$), Weibull- ($\Theta=53$, $\beta=5$) und Lognormalverteilung ($\mu=4$, $\sigma=0,25$) dargestellt.

Folgende Funktionen entsprechen den beschriebenen Verteilungen:

Normalverteilung: $f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Maxwellverteilung: $f(x) = \sqrt{\frac{2}{\pi}} \frac{1}{\sigma^3} x^2 e^{-\frac{x^2}{2\sigma^2}}$

Rayleighverteilung: $f(x) = \frac{\beta}{\Theta} \left(\frac{x}{\Theta}\right)^{\beta-1} e^{-\left(\frac{x}{\Theta}\right)^\beta}$

Weibullverteilung: $f(x) = \frac{\beta}{\Theta} \left(\frac{x}{\Theta}\right)^{\beta-1} e^{-\left(\frac{x}{\Theta}\right)^\beta}$

Lognormalverteilung: $f(x) = \frac{1}{\sqrt{2\pi}\sigma x} e^{-\frac{(\ln x - \mu)^2}{2\sigma^2}}$

Für das Einzelgerät wird der Vergleich Standardverteilung ($\mu=40, \sigma=17$), Maxwell- ($\sigma=28$), Rayleigh- ($\Theta=55$), Weibull- ($\Theta=40, \beta=5$) und Lognormalverteilung ($\mu=3,8, \sigma=0,25$) in Abbildung 6.36 angegeben.

Es zeigt sich eine relativ gute Übereinstimmung der gemessenen Daten mit der Normalverteilung, Lognormal- und Weibullverteilung.

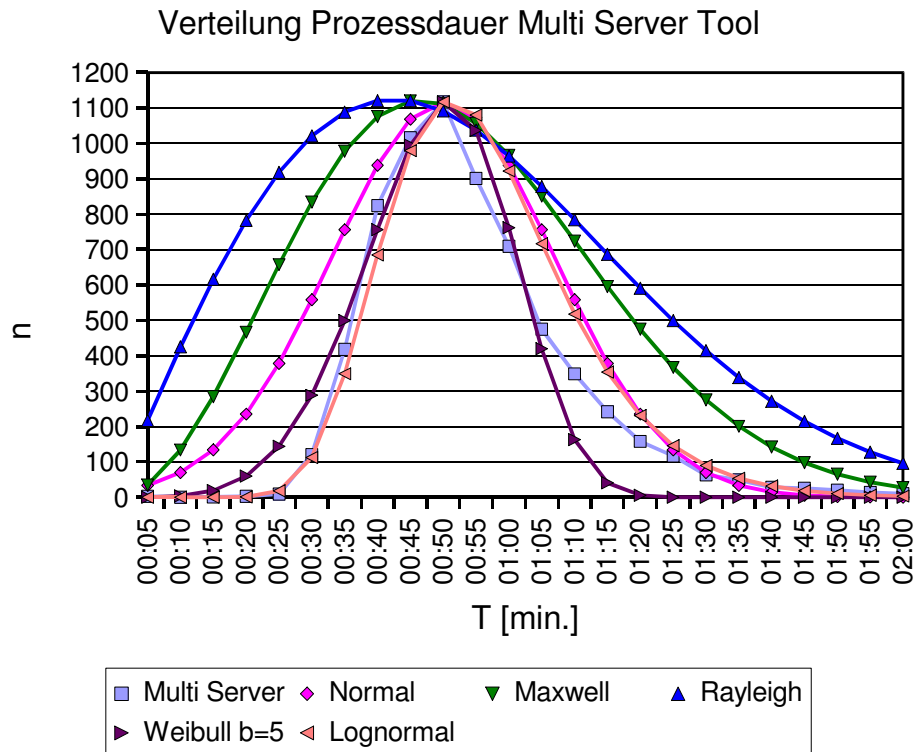


Abbildung 6.35: Prozessdauer am Multi Server Tool im Vergleich mit verschiedenen Verteilungen

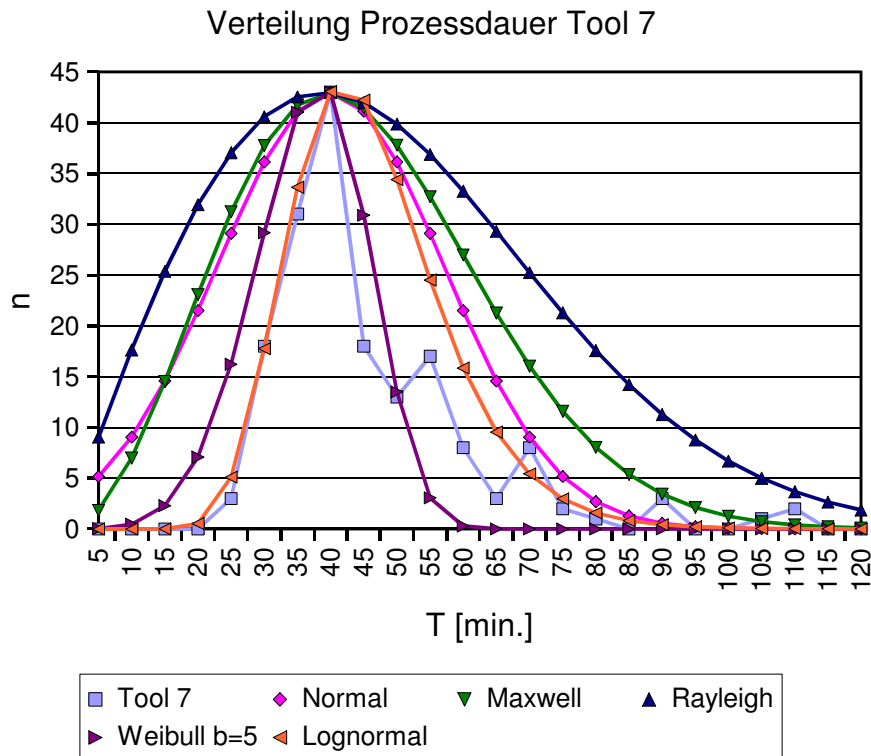


Abbildung 6.36: Prozessdauer am Single Server Tool im Vergleich mit verschiedenen Verteilungen

Um die Übereinstimmung der Gerätedaten mit der idealen Normalverteilung noch in anderer Weise grafisch darzustellen, sind in Abbildung 6.37 jeweils die reale und ideale Kurve für die parallelen Geräte und das Einzelgerät auf Normalpapier dargestellt. Dabei ergibt sich für die ideale Normalverteilung eine Gerade, je besser die realen Daten mit der Normalverteilung übereinstimmen, desto besser passen sie sich dieser Geraden an.

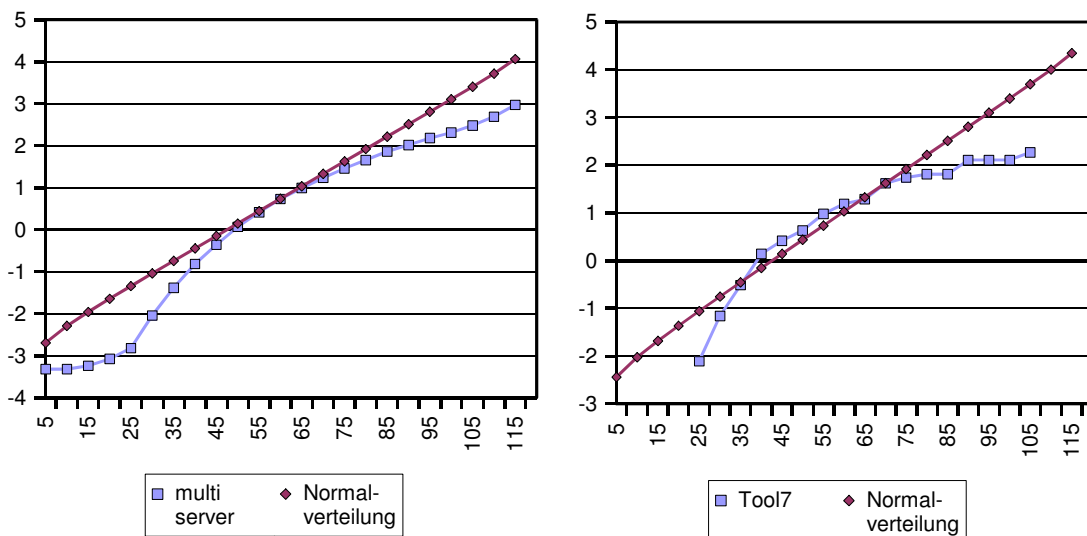


Abbildung 6.37: Darstellung der Prozessdauer auf Normalpapier

Besonders bei den Multi Server Daten zeigt sich eine relativ gute Übereinstimmung zwischen den realen Daten und der theoretisch angenommenen Verteilung. Besonders im Bereich des Mittelwerts stimmen die Kurven überein. Bei den Single Server Daten weicht auch hier die Statistik aufgrund der niedrigeren Anzahl prozessierter Lose noch relativ stark von den idealen Verteilungen ab.

An dieser Stelle soll noch erwähnt werden, dass sich entsprechende Ergebnisse auch zeigen, wenn statt der beschriebenen Gruppierung in Multi Server- und Single Server Tools einzelne Operations und Geräte aus den vorliegenden Daten betrachtet werden. Auch hier kann bei den statistischen Verteilungen eine gute Übereinstimmung mit der Exponential- bzw. Normalverteilung festgestellt werden. In Abbildung 6.38 und 6.39 wird beispielsweise für Tool 1 (699 Lose prozessiert) und Tool 2 (615 Lose prozessiert) die Statistik der ankommenden Lose zusammen mit der Exponentialverteilung im normalen und im logarithmischen Maßstab dargestellt. Es ergibt sich eine ähnliche Übereinstimmung wie in den Abbildungen 6.32, 6.33 und 6.34.

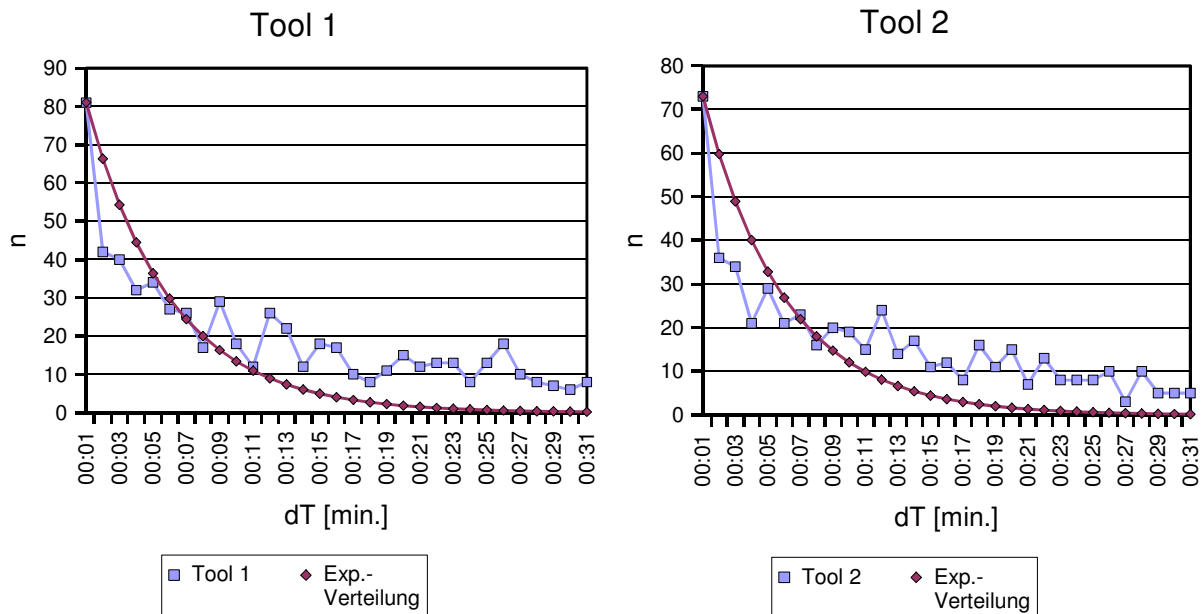


Abbildung 6.38: Losankunft an Tool 1 und 2 im Vergleich mit Exponentialverteilung

Für die Prozessdauer wird in den Abbildungen 6.40 und 6.41 ein Vergleich der gemessenen Daten mit der idealen Normalverteilung gezeigt. Für die relativ häufig vorkommenden Operations 4456 (147 Lose prozessiert, $\mu=54$, $\sigma=11$) und 5011 (126 Lose prozessiert, $\mu=45$, $\sigma=11$) wird wieder die normale Darstellung und die Darstellung auf Normalpapier angegeben. Auch hier stimmen die gemessenen und die theoretischen Kurven relativ gut überein.

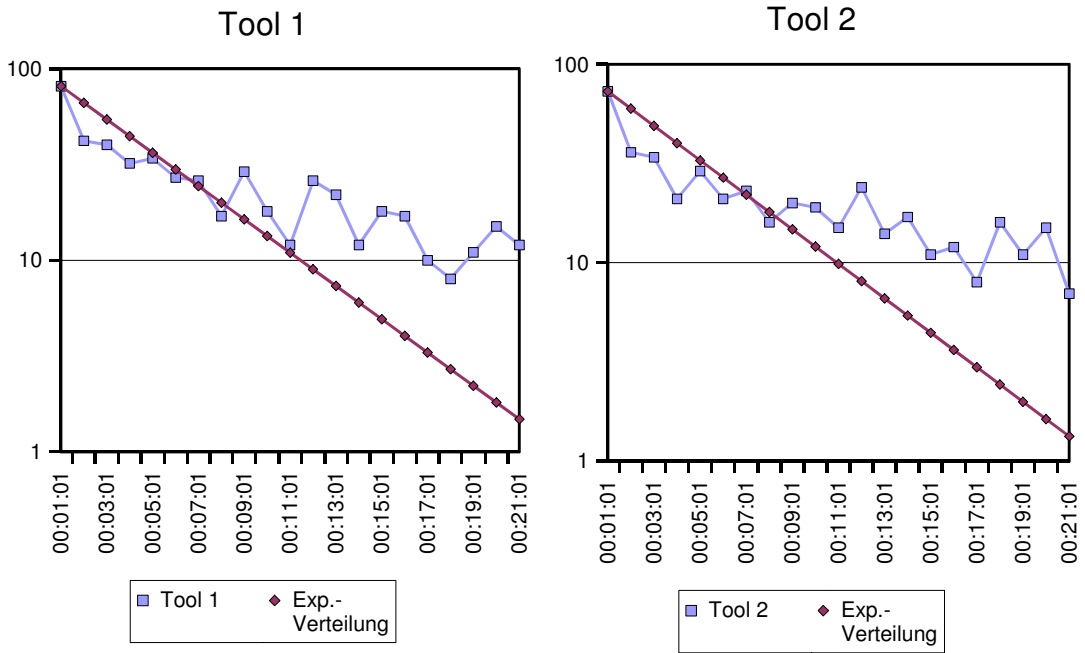


Abbildung 6.39: Losankunft an Tool 1 und 2 im logarithmischen Maßstab

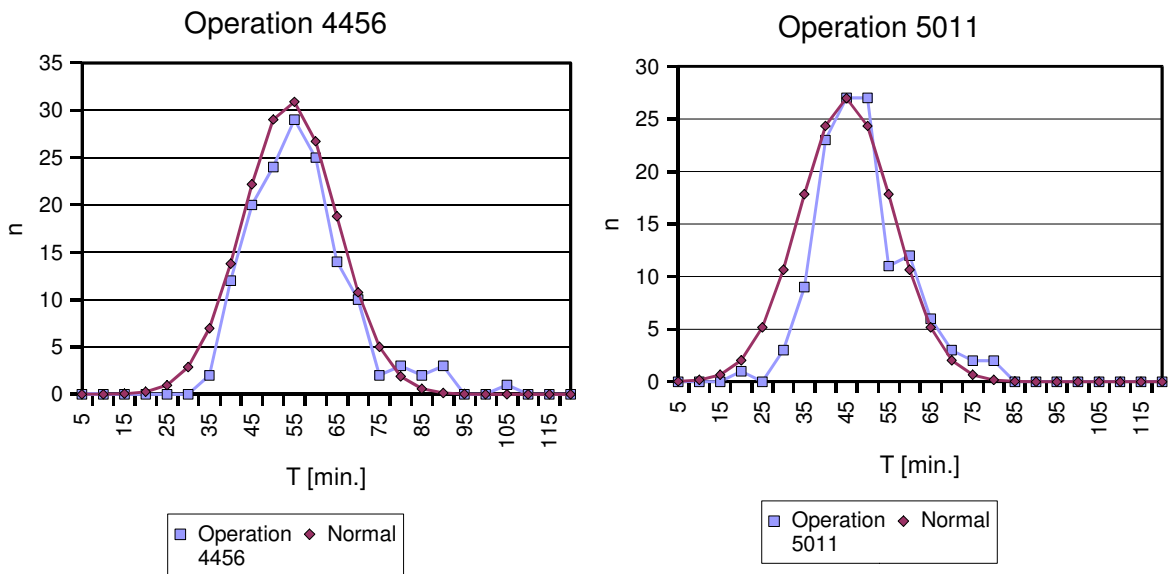


Abbildung 6.40: Prozessdauer bei Operation 4456 und 5011 im Vergleich mit Normalverteilung

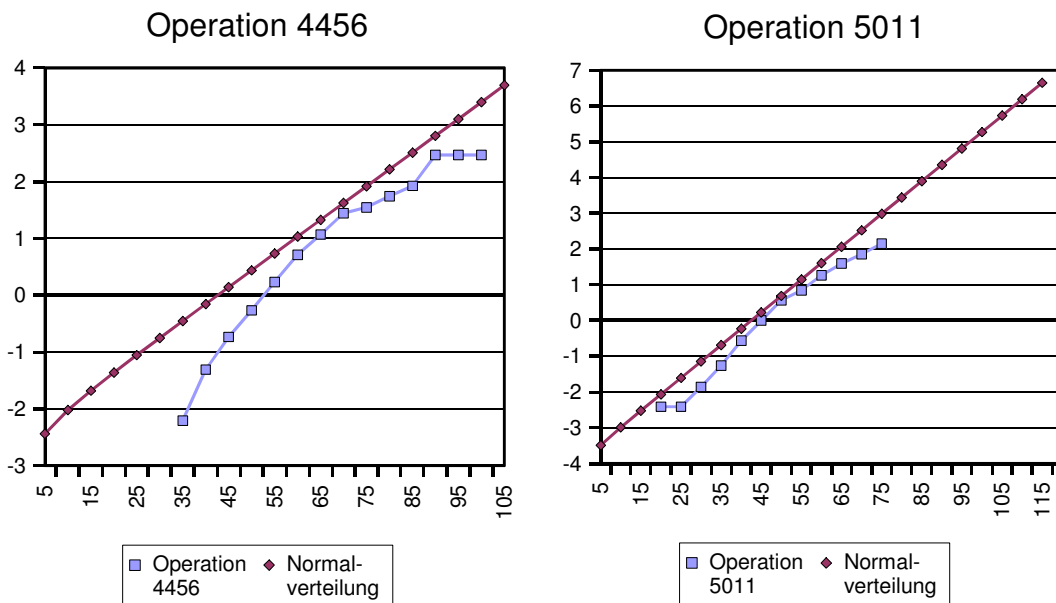


Abbildung 6.41: Prozessdauer bei Operation 4456 und 5011 auf Normalpapier

Insgesamt ergeben sich aus den Daten die in Tabelle 6.4 angegebenen Mittelwerte und die Werte für die jeweiligen Variationskoeffizienten.

Tabelle 6.4: Ermittelte Mittelwerte und Variationskoeffizienten

Multi Server		Tool 7	
Mittelwert Losankunft:	01:44 min.	Mittelwert Losankunft:	1:00:26 h
Mittelwert Wartezeit:	01:18 h	Mittelwert Wartezeit:	56:26 min.
Mittelwert Prozesszeit:	53:10 min.	Mittelwert Prozesszeit:	45:02 min.
Mittelwert Gesamtzeit:	02:11:10h	Mittelwert Gesamtzeit:	01:41:28h
Variationskoeffizient Losankunft c_a :	1,83	Variationskoeffizient Losankunft c_a :	1,23
Variationskoeffizient Wartezeit c_q :	0,74	Variationskoeffizient Wartezeit c_q :	1,93
Variationskoeffizient Prozesszeit c_e :	0,36	Variationskoeffizient Prozesszeit c_e :	0,39
Variationskoeffizient Gesamtzeit c_{ges} :	0,47	Variationskoeffizient Gesamtzeit c_{ges} :	1,08

Für den Fall, dass alle Geräte und Operations getrennt betrachtet werden, ergeben sich ähnliche Ergebnisse wie in Tabelle 6.4. Für den Variationskoeffizient der Losankunft c_a liegen die resultierenden Werte dann zwischen 1,0 und 3,6, für die Wartezeit beträgt c_q 0,6 - 2,0 und für die Prozesszeit beträgt c_e 0,2 - 0,55.

In [168] werden für die Variationskoeffizienten, die von den Autoren in der Philips MOS4YOU Fab ausgewertet wurden, ähnliche Werte angegeben. c_e liegt dort für fast alle Geräte unter 0,6, c_a liegt für die dort betrachteten Ofenprozesse im Bereich 1,4-2,5.

6.5.3 Warteschlangenmodell für einen Teil des Prozessflusses

Mit den Daten aus Kapitel 6.5.1 und 6.5.2 wird ein Modell aus maximal 10 hintereinander geschalteten Geräten aufgebaut, wie dies prinzipiell in Abbildung 6.42 dargestellt ist. Für den angenommenen Prozessfluss können Durchlaufzeit, WIP und damit die Operating

Curve in Abhängigkeit der Eingangsparameter r_a , r_e , c_a , c_e , t_{setup} , t_T und Losgröße berechnet werden.

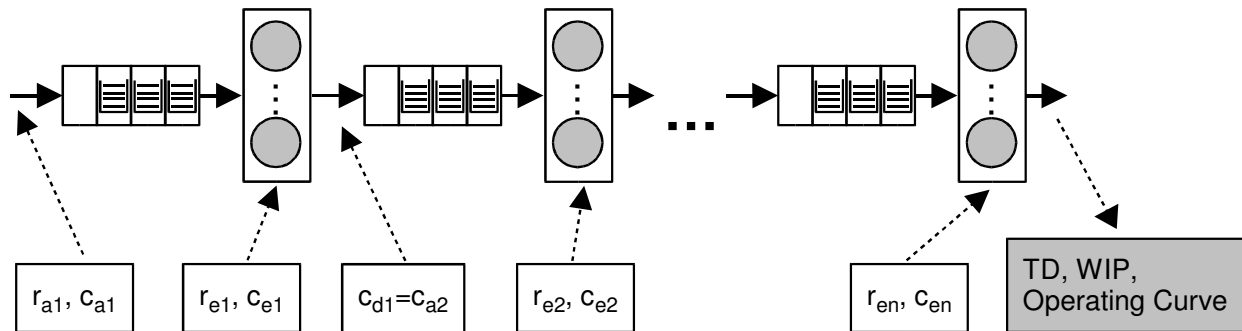


Abbildung 6.42: Prinzipieller Aufbau des Modells

Um den Einfluss der Variabilität betrachten zu können, wird zur Berechnung der Durchlaufzeit die Warteschlangenformel für ein G/G/s-System (Gleichung 6.30) verwendet.

$$TD = \frac{c_a^2 + c_e^2}{2} \cdot u \cdot \frac{\sqrt{2(s+1)} - 1}{s(1-u)} \cdot \frac{1}{r_e} + \frac{1}{r_e}$$

Zur Berücksichtigung von Unterbrechungen (MTBF, MTTR, Setups) werden für r_e die Gleichungen (6.42) und (6.47) eingesetzt.

$$r_{\text{ebr}} = r_e \cdot A = r_e \cdot \frac{m_f}{m_f + m_r}$$

$$t_{\text{batch}} = \frac{n}{r_e} + t_s$$

Laut Gleichung (6.36) und (6.37) ist die Variabilität am Eingang des nächsten Gerätes abhängig vom vorausgehenden Gerät. Dies wird im Modell berücksichtigt, indem $c_{di} = c_{ai+1}$ gesetzt wird.

$$c_d^2 = 1 + (1 - u^2) \cdot (c_a^2 - 1) + \frac{u^2}{\sqrt{m}} \cdot (c_e^2 - 1)$$

Die resultierende Durchlaufzeit TD_{ges} ergibt sich als $TD_{\text{ges}} = \sum_i TD_i$ für alle Geräte, WIP_{ges} berechnet sich mit Little's Law (6.7) zu $WIP_{\text{ges}} = TD_{\text{ges}} \cdot r_a$.

Für c_a und c_e können die in Kapitel 6.5.2 ermittelten und in Tabelle 6.4 angegebenen Daten aus dem realen Prozessfluss eingesetzt werden.

In Abbildung 6.43 sind die Eingangsparameter des Modells für die 10 Geräte anhand eines fiktiven Beispiels dargestellt. Es können für bis zu 10 Geräte die Maschinenkapazität r_e , MTBF m_f , MTTR m_r , Setupzeit t_{setup} , Transportzeit t_T und Losgröße, sowie die Variabilitätskoeffizienten und die Anzahl der parallelen Geräte angegeben werden.

Parallele Geräte in Serie geschaltet mit Transportzeit, Downtime und Losgröße

re1 [W/h]:	25	mf1 [h]:	300	mr1 [h]:	5	tsetup 1 [h]:	0,1
re2 [W/h]:	25	mf2 [h]:	300	mr2 [h]:	5	tsetup 2 [h]:	0,1
re3 [W/h]:	25	mf3 [h]:	300	mr3 [h]:	5	tsetup 3 [h]:	0,1
re4 [W/h]:	25	mf4 [h]:	300	mr4 [h]:	5	tsetup 4 [h]:	0,1
re5 [W/h]:	25	mf5 [h]:	300	mr5 [h]:	5	tsetup 5 [h]:	0,1
re6 [W/h]:	25	mf6 [h]:	300	mr6 [h]:	5	tsetup 6 [h]:	0,1
re7 [W/h]:	25	mf7 [h]:	300	mr7 [h]:	5	tsetup 7 [h]:	0,1
re8 [W/h]:	25	mf8 [h]:	300	mr8 [h]:	5	tsetup 8 [h]:	0,1
re9 [W/h]:	25	mf9 [h]:	300	mr9 [h]:	5	tsetup 9 [h]:	0,1
re10 [W/h]:	25	mf10 [h]:	300	mr10 [h]:	5	tsetup 10 [h]:	0,1
Transportzeit tT [h]:	0,1						
Losgröße [W]:	25						

Variabilitätskoeffizienten: Zahl der parallelen Geräte:

ca1:	1,5	m1:	1
ce1:	1,5	m2:	1
ce2:	1,5	m3:	1
ce3:	1,5	m4:	1
ce4:	1,5	m5:	1
ce5:	1,5	m6:	1
ce6:	1,5	m7:	1
ce7:	1,5	m8:	1
ce8:	1,5	m9:	1
ce9:	1,5	m10:	1
ce10:	1,5		

Abbildung 6.43: Eingabeparameter des Modells

Die Ausgabe des Modells mit den berechneten Ergebnissen zeigt Abbildung 6.44. Ausgegeben werden die Availability A, die reale Maschinenkapazität, die Prozesszeit für ein Los, sowie die Durchlaufzeit pro Gerät, die Gesamtdurchlaufzeit TD_{ges} und der WIP-Bestand jeweils in Tabellen- und in Kurvenform.

A =	0,98	reales re1 [W/h]:	24,59	te1 batch [h/Los]:	1,12
A =	0,98	reales re2 [W/h]:	24,59	te2 batch [h/Los]:	1,12
A =	0,98	reales re3 [W/h]:	24,59	te3 batch [h/Los]:	1,12
A =	0,98	reales re4 [W/h]:	24,59	te4 batch [h/Los]:	1,12
A =	0,98	reales re5 [W/h]:	24,59	te5 batch [h/Los]:	1,12
A =	0,98	reales re6 [W/h]:	24,59	te6 batch [h/Los]:	1,12
A =	0,98	reales re7 [W/h]:	24,59	te7 batch [h/Los]:	1,12
A =	0,98	reales re8 [W/h]:	24,59	te8 batch [h/Los]:	1,12
A =	0,98	reales re9 [W/h]:	24,59	te9 batch [h/Los]:	1,12
A =	0,98	reales re10 [W/h]:	24,59	te10 batch [h/Los]:	1,12

ra [W]	ra [Lose]	uBNT	u1	TD1	cd1=ca2	u2	TD2	...	cd9=ca10	u10	TD10	TD ges [h]	WIP
2	0,08	0,09	0,09	1,36	1,5	0,09	1,36	...	1,5	0,09	1,36	14,63	29,26
4	0,16	0,18	0,18	1,66	1,5	0,18	1,66	...	1,5	0,18	1,66	17,63	70,53
6	0,24	0,27	0,27	2,04	1,5	0,27	2,04	...	1,5	0,27	2,04	21,37	128,19
8	0,32	0,36	0,36	2,51	1,5	0,36	2,51	...	1,5	0,36	2,51	26,14	209,09
10	0,4	0,45	0,45	3,14	1,5	0,45	3,14	...	1,5	0,45	3,14	32,45	324,48
12	0,48	0,54	0,54	4,02	1,5	0,54	4,02	...	1,5	0,54	4,02	41,19	494,28
14	0,56	0,63	0,63	5,31	1,5	0,63	5,31	...	1,5	0,63	5,31	54,1	757,42
16	0,64	0,71	0,71	7,41	1,5	0,71	7,41	...	1,5	0,71	7,41	75,1	1201,55
18	0,72	0,8	0,8	11,42	1,5	0,8	11,42	...	1,5	0,8	11,42	115,23	2074,15
20	0,8	0,89	0,89	22,16	1,5	0,89	22,16	...	1,5	0,89	22,16	222,59	4451,77
22	0,88	0,98	0,98	143,56	1,5	0,98	143,56	...	1,5	0,98	143,56	1436,56	31604,34

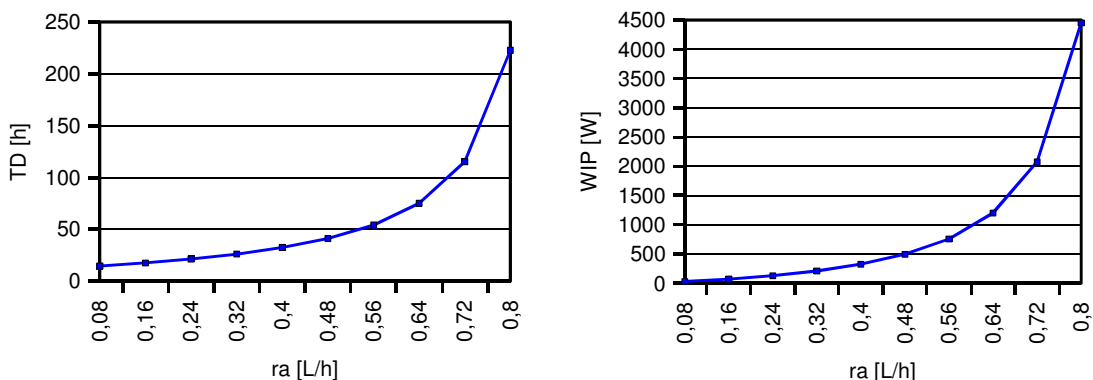


Abbildung 6.44: Ausgabe der Ergebnisse im Modell

6.5.4 Vergleich der Ergebnisse von Modell und Prozessfluss

Aus den Daten von Kapitel 6.5.1 und 6.5.2 ergibt sich für die Multi Server Geräte eine Durchlaufzeit von 2:11:10h unter Berücksichtigung der Parameter aus Tabelle 6.4. Für Tool 7 resultiert eine Durchlaufzeit von 1:41:28h, für Tool 1 von 1:57:50h und für Tool 2 von 2:00:39h.

Es soll nun überprüft werden, inwieweit diese realen Daten mit dem erstellten Modell nachgebildet werden können. Dazu werden die in Tabelle 6.4 angegebenen Werte als Eingangsparameter für das Modell verwendet. Es sollen die Ergebnisse jeweils für den Multi Server und Single Server Fall verglichen werden. Ebenso wird, wegen der umfangreicheren Statistik im Vergleich zu Tool 7, eine Kontrolle mit den Tools 1 und 2 durchgeführt.

Die Berechnung der Ankunftsrate r_a an den Geräten kann über die Statistik aus Tabelle 6.3 erfolgen. In den insgesamt 195,5h des Betrachtungszeitraumes kommen beispielsweise 699 Lose an Tool 1 an. Dies entspricht einer durchschnittlichen Ankunftsrate von ~89 W/h. Für Tool 2 beträgt analog die Ankunftsrate 78 W/h, für Tool 7 22 W/h. Für das

betrachtete Multi Server Modell bestehend aus den Geräten 1-6 und 8-14 beträgt $r_a = 838$ W/h.

Als problematisch stellt sich die Ermittlung der Bearbeitungsrate für die jeweiligen Geräte dar. Es stellt sich heraus, dass jeweils mehrere Lose gleichzeitig im Gerät bearbeitet werden, d.h. die reine Prozesszeit für ein Los, die im Bereich von 30 W/h liegt, kann für die Eingangsdaten des Modells nicht in dieser Form verwendet werden, sondern es muss von einem größeren r_e ausgegangen werden. In Abbildung 6.45 wird für das Multi Server Modell und für Tool 1 gezeigt, wieviele Lose sich gleichzeitig im Prozess befinden.

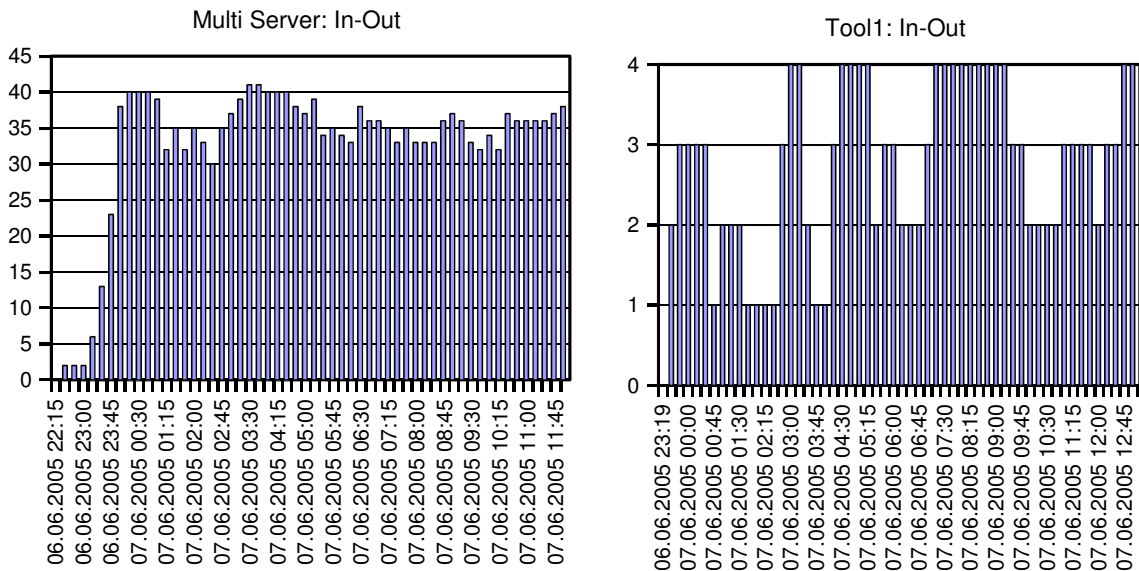


Abbildung 6.45: Zahl der gleichzeitig im Gerät befindlichen Lose

In Abbildung 6.46 werden die Ergebnisse des Modells für den Multi Server Fall mit 13 parallelen Geräten dargestellt. Mit den angegebenen Eingangsparametern lässt sich für eine Ankunftsrate von 850 W/h eine Gesamtdurchlaufzeit von 2,09 h berechnen.

Multi Server mit Transportzeit, Downtime und Losgröße

re1 [W/h]: 79 mf1 [h]: 300 mr1 [h]: 5 tsetup 1 [h]: 0,05

Transportzeit tT [h]: 0
 Losgröße [W]: 25

Variabilitätskoeffizienten:

Zahl der parallelen Geräte:

ca1: 1,9
 ce1: 0,4

m1: 13

A = 0,98 reales re1 [W/h]: 77,7 te1 batch [h/Los]: 0,37

Ankunftsrate [W]	ra [Lose]	u1	TD1 [h]	WIP [W]
50	2	0,06	0,37	18,59
100	4	0,11	0,37	37,17
150	6	0,17	0,37	55,76
200	8	0,23	0,37	74,37
250	10	0,29	0,37	93,02
300	12	0,34	0,37	111,77
350	14	0,4	0,37	130,72
400	16	0,46	0,38	150,08
450	18	0,51	0,38	170,17
500	20	0,57	0,38	191,59
550	22	0,63	0,39	215,39
600	24	0,69	0,41	243,53
650	26	0,74	0,43	279,89
700	28	0,8	0,48	333,11
750	30	0,86	0,57	426,06
800	32	0,92	0,81	644,04
825	33	0,94	1,12	921,55
850	34	0,97	2,09	1777,2

Abbildung 6.46: Modellberechnung für Multi Server Tool

In den Abbildungen 6.47, 6.48 und 6.49 werden die berechneten Ergebnisse des Modells jeweils für den Single Server Fall (Tool 7), für Tool 1 und für Tool 2 angegeben.

Single Server mit Transportzeit, Downtime und Losgröße

re1 [W/h]: 37 mf1 [h]: 300 mr1 [h]: 5 tsetup 1 [h]: 0,05

Transportzeit tT [h]: 0
Losgröße [W]: 25

Variabilitätskoeffizienten:

Zahl der parallelen Geräte:

ca1: 1,2
ce1: 0,4

m1: 1

A = 0,98 reales re1 [W/h]: 36,39 te1 batch [h/Los]: 0,74

Ankunftsrate [W]	ra [Lose]	u1	TD1 [h]	WIP [W]
1	0,04	0,03	0,75	0,75
2	0,08	0,06	0,77	1,55
4	0,16	0,12	0,82	3,26
6	0,24	0,18	0,86	5,18
8	0,32	0,24	0,92	7,35
10	0,4	0,29	0,98	9,83
12	0,48	0,35	1,06	12,72
14	0,56	0,41	1,15	16,12
16	0,64	0,47	1,26	20,21
18	0,72	0,53	1,4	25,26
20	0,8	0,59	1,58	31,67
22	0,88	0,65	1,82	40,14
24	0,96	0,71	2,16	51,9
26	1,04	0,77	2,67	69,45
28	1,12	0,83	3,52	98,65

Abbildung 6.47: Modellberechnung für Single Server Tool 7

Tool 1 mit Transportzeit, Downtime und Losgröße

re1 [W/h]: 126 mf1 [h]: 300 mr1 [h]: 5 tsetup 1 [h]: 0,05

Transportzeit tT [h]: 0
Losgröße [W]: 25

Variabilitätskoeffizienten:

Zahl der parallelen Geräte:

ca1: 1,1
ce1: 0,4

m1: 1

A = 0,98 reales re1 [W/h]: 123,93 te1 batch [h/Los]: 0,25

Ankunftsrate [W]	ra [Lose]	u1	TD1 [h]	WIP [W]
5	0,2	0,05	0,26	1,3
10	0,4	0,1	0,27	2,71
15	0,6	0,15	0,28	4,24
20	0,8	0,2	0,3	5,9
25	1	0,25	0,31	7,74
30	1,2	0,3	0,33	9,79
35	1,4	0,35	0,35	12,09
40	1,6	0,4	0,37	14,72
45	1,8	0,45	0,39	17,76
50	2	0,5	0,43	21,33
55	2,2	0,55	0,47	25,61
60	2,4	0,6	0,51	30,89
65	2,6	0,65	0,58	37,59
70	2,8	0,7	0,66	46,44
75	3	0,76	0,78	58,77
80	3,2	0,81	0,97	77,27
85	3,4	0,86	1,28	108,41
90	3,6	0,91	1,92	172,56
95	3,8	0,96	4,05	384,4

Abbildung 6.48: Modellberechnung für Tool 1

Tool 2 mit Transportzeit, Downtime und Losgröße

re1 [W/h]: 118 mf1 [h]: 300 mr1 [h]: 5 tsetup 1 [h]: 0,05

Transportzeit tT [h]: 0
 Losgröße [W]: 25

Variabilitätskoeffizienten:

Zahl der parallelen Geräte:

ca1: 1,6
 ce1: 0,4

m1: 1

A = 0,98 reales re1 [W/h]: 116,07 te1 batch [h/Los]: 0,27

Ankunftsrate [W]	ra [Lose]	u1	TD1 [h]	WIP [W]
1	0,04	0,01	0,27	0,27
5	0,2	0,05	0,29	1,43
10	0,4	0,11	0,31	3,08
15	0,6	0,16	0,33	5,01
20	0,8	0,21	0,36	7,25
25	1	0,27	0,4	9,89
30	1,2	0,32	0,43	13,02
35	1,4	0,37	0,48	16,76
40	1,6	0,42	0,53	21,27
45	1,8	0,48	0,6	26,8
50	2	0,53	0,67	33,69
55	2,2	0,58	0,77	42,45
60	2,4	0,64	0,9	53,92
65	2,6	0,69	1,07	69,48
70	2,8	0,74	1,31	91,66
75	3	0,8	1,68	125,65
76	3,04	0,81	1,77	134,72
77	3,08	0,82	1,88	144,86
78	3,12	0,83	2	156,26
79	3,16	0,84	2,14	169,17
80	3,2	0,85	2,3	183,92
85	3,4	0,9	3,6	306,04

Abbildung 6.49: Modellberechnung für Tool 2

Es zeigt sich, dass mit den vorliegenden Eingangsdaten eine vernünftige Übereinstimmung zwischen Modell und gemessenen Daten zu erreichen ist. Besonders für den Fall der einzelnen Geräte können mit dem Modell die realen Daten gut reproduziert werden. Allerdings muss angemerkt werden, dass aus den vorliegenden Fertigungsdaten nicht alle Eingangsparameter des Modells eindeutig abgeleitet werden können. Da sich teilweise mehrere Lose im Gerät befinden, kann z.B. r_e und der Nutzungsgrad für die jeweiligen Geräte nicht eindeutig ermittelt werden. Insofern sind also nicht alle Eingangsdaten klar definiert und können in einem bestimmten Umfang frei festgelegt werden.

7 Möglichkeiten zur Effizienzsteigerung bei der zukünftigen IC-Fertigung

7.1 Zusammenführung und Bewertung der ökonomischen und technologischen Aspekte der Fertigung

Da die Halbleiterfertigung inzwischen eine gewisse Reife erlangt hat und die technologische Machbarkeit nicht mehr ausschließlich im Vordergrund steht, erhalten Fragen nach Verbesserungen in anderen Bereichen eine höhere Bedeutung. Die Zusammenführung von ökonomischen und technologischen Aspekten in der Halbleiterfertigung wird in diesem Zusammenhang in Zukunft weiter steigende Bedeutung erlangen. Außerdem werden Verbesserungen auf dem Gebiet der Produktionstechnik, z.B. der möglichst optimierte Ablauf der Fertigung, einen großen Beitrag liefern. Gordon Moore, Mitbegründer der Firma Intel, sagte vor einiger Zeit: „I'm convinced that any limitation we run into in the near term will be more likely economic than technical.“ [134] Die Beachtung der ökonomischen Anforderungen an die Halbleiterfertigung wird also, auch unter Berücksichtigung der weiter steigenden Kosten für den Bau und die Ausstattung einer Halbleiterfabrik, immer wichtiger [135].

Ziel ist es, die Ergebnisse aus den Kapiteln 4 und 5 zusammen mit dem Modell aus Kapitel 6 für aktuelle und zukünftige Bauelemente anzuwenden, um damit Aussagen über die zukünftigen Anforderungen an die IC-Fertigung und über geeignete Maßnahmen zu deren Erfüllung treffen zu können. Wie bereits in Kapitel 1.1 erwähnt, wird hier nur der Bereich Produktion betrachtet. Der Einfluss anderer Bereiche, wie z.B. Entwicklung, Marketing oder Vertrieb, die selbstverständlich ebenfalls zur Steigerung des Unternehmenserfolgs beitragen können, wird nicht untersucht, bzw. als konstant angenommen.

Ausgehend von den Rahmenbedingungen und den unternehmerischen Vorgaben, ergeben sich Forderungen an die Entwicklung der Effizienz der Fertigung. Zum Erreichen dieser Forderungen werden mehrere Einflussparameter untersucht. Dabei werden auch die Auswirkungen der Durchlaufzeit und der Zahl der halbfertigen Produkte im Fertigungsablauf berücksichtigt, die sich auf die Gesamtkosten niederschlagen und zu den allgemeinen Produktionskosten addiert werden.

Mit den Formeln (4.24) und (5.5) ergibt sich, bei angenommener linearer Steigerung der variablen Kosten, für die Wirtschaftlichkeit der Fertigung:

$$W = \frac{WSPM \cdot (1 - m) \cdot \left(\frac{\pi \cdot R_{\text{Wafer}}^2}{A} - \frac{\pi \cdot D_{\text{Wafer}}}{\sqrt{2 \cdot A}} \right) \cdot \exp(-D \cdot A) \cdot P}{K_F + k_V \cdot WSPM + wip \cdot \bar{W}_{\text{Wafer}} \cdot T_D \cdot \frac{Z}{360}} \quad (7.1)$$

k_V = variable Kosten pro Waferstart

Zur Berechnung der Anforderungen an die Steigerung der Effizienz spielt der Preisverfall eine wichtige Rolle, da ohne Verbesserungsmaßnahmen in der Produktion der Preisverfall zu einer Verringerung des Umsatzes bei gleichbleibenden Kosten führen würde. Dies würde zu einem Absinken des Gewinns und damit auch der Rentabilität sowie zu einer verringerten Wirtschaftlichkeit der Fertigung führen. Deswegen muss durch geeignete Maßnahmen, wie z.B. die Steigerung der Produktionsmenge, die Effizienz der Produktion

so gesteigert werden, dass die negativen Effekte des Preisverfalls ausgeglichen werden. Da durch einen geringeren Stückpreis der Erlös pro Chip sinkt, ist es deshalb notwendig, die Stückkosten parallel zum Preisverfall zu senken, da sichergestellt werden muss, dass die Stückkosten unter dem Verkaufspreis liegen.

Die Kosten pro Chip ergeben sich aus dem Verhältnis der Gesamtkosten für die Herstellung und der damit erzeugten Zahl funktionsfähiger Chips.

$$k_{St} = \frac{K_G}{C} \quad (7.2)$$

k_{St} = Stückkosten

K_G = Gesamtkosten

C = Zahl der funktionsfähigen Chips

Um die Stückkosten entsprechend dem Preisverfall zu senken, muss Gleichung (7.3) gelten:

$$\frac{K_1}{C_1} = \frac{P_1}{P_0} \cdot \frac{K_0}{C_0} \quad (7.3)$$

P = Stückpreis

Index $_1$ = Situation nach Preisverfall

Index $_0$ = Situation vor Preisverfall

Bei Erfüllung von Gleichung (7.3) ist sichergestellt, dass der prozentuale Gewinn pro Stück konstant bleibt. Der Gewinn pro Stück berechnet sich als Differenz zwischen dem Stückpreis und den Stückkosten.

$$g_{St} = P - k_{St} \quad (7.4)$$

g_{St} = Gewinn pro Stück

Unter Verwendung von (7.3) und (7.4) ergibt sich ein gleichbleibender Gewinn pro Stück bezogen auf den Verkaufspreis vor und nach Preisverfall.

$$\frac{P_1 - k_{St1}}{P_1} = \frac{P_1 - \frac{P_1}{P_0} \cdot k_{St0}}{P_1} = \frac{P_0 - k_{St0}}{P_0} \quad (7.5)$$

Wie durch leichte Umformung aus Gleichung (7.3) abzuleiten ist, ist die Forderung nach einer angemessenen Senkung der Stückkosten gleichbedeutend mit einer konstanten Wirtschaftlichkeit der Fertigung, wie sie in Gleichung (4.11) definiert ist.

$$\frac{C_1}{K_1} \cdot P_1 = \frac{C_0}{K_0} \cdot P_0 \quad (7.6)$$

$$\Rightarrow W_1 = W_0$$

Die Forderung nach einer gleichbleibenden Wirtschaftlichkeit entspricht ebenfalls der Forderung nach gleichbleibender Umsatzrentabilität. Da die Umsatzrentabilität ohne Beachtung sonstiger Einflussfaktoren auf den Gewinn gegeben ist durch

$$R_U = \frac{G}{U} = \frac{C \cdot P - K}{C \cdot P} = 1 - \frac{1}{W} \quad (7.7),$$

R_U = Umsatzrentabilität
 G = Gewinn
 U = Umsatz

folgt, dass konstante bzw. steigende Wirtschaftlichkeit gleichzeitig zu konstanter bzw. steigender Umsatzrentabilität führt.

Bei unterschiedlichen Kombinationen von Gewinn und Umsatz können sich gleiche Werte für die Umsatzrentabilität ergeben. So kann beispielsweise bei kleinem Gewinn und kleinem Umsatz das gleiche Verhältnis wie bei großem Gewinn und großem Umsatz resultieren. Bei einem insgesamt wachsenden Markt wie dem Halbleitermarkt würde eine Reduktion des Umsatzes jedoch bedeuten, dass der Marktanteil des Unternehmens sinkt. Um den Marktanteil zu sichern oder zu steigern, ist es also notwendig, durch Steigerung der Produktionsmenge den Umsatz trotz Preisverfall möglichst beizubehalten oder sogar zu erhöhen. Insofern wird hier im folgenden der Schwerpunkt auf eine Steigerung der Produktionsleistung gelegt.

Der Return on Investment berechnet sich nach Gleichung (4.4) zu:

$$ROI = \frac{G}{Kap} = \frac{C \cdot P - K}{Kap} \quad (7.8)$$

Kap = Gesamtkapital

Bei konstanter Wirtschaftlichkeit bleibt auch der Return on Investment konstant, wenn es gelingt bei gleichbleibenden Kosten die Produktionsmenge so zu steigern, dass sich ein konstanter Umsatz ergibt. Im allgemeinen Fall, bei dem sich die Produktionskosten in Abhängigkeit von der Produktionsmenge steigern, bedeutet die Forderung nach konstanter Wirtschaftlichkeit ein leichtes Ansteigen des ROI, da die konstante Wirtschaftlichkeit ein strengeres Kriterium darstellt. Damit folgt, dass bei gleichbleibender Wirtschaftlichkeit auch bei Preisverfall gilt:

$$ROI_1 \geq ROI_0$$

Bei Einhaltung einer konstanten Wirtschaftlichkeit wird also sichergestellt, dass die Stückkosten entsprechend dem Preisverfall sinken, bzw. die Effizienz der Fertigung entsprechend steigt und dass Umsatzrentabilität und ROI unter sonst gleichen Voraussetzungen mindestens konstant bleiben.

Um die Effizienz der Fertigung zu steigern, bzw. die Stückkosten zu senken, besteht grundsätzlich die Möglichkeit die Leistung (= Produktionsmenge) zu steigern oder die Produktionskosten zu senken, mit dem Ziel das Verhältnis dieser beiden Größen zu maximieren. Aus Gleichung (7.3) und (7.6) folgt für das Verhältnis C_1/K_1 :

$$\frac{C_1}{K_1} = \frac{P_0}{P_1} \cdot \frac{C_0}{K_0} \quad (7.9)$$

Dieser Zusammenhang ist in Abbildung 7.1 dargestellt. Je größer der Preisverfall wird, desto stärker müssen die Stückkosten sinken, bzw. desto größer muss auch das Verhältnis C/K werden.

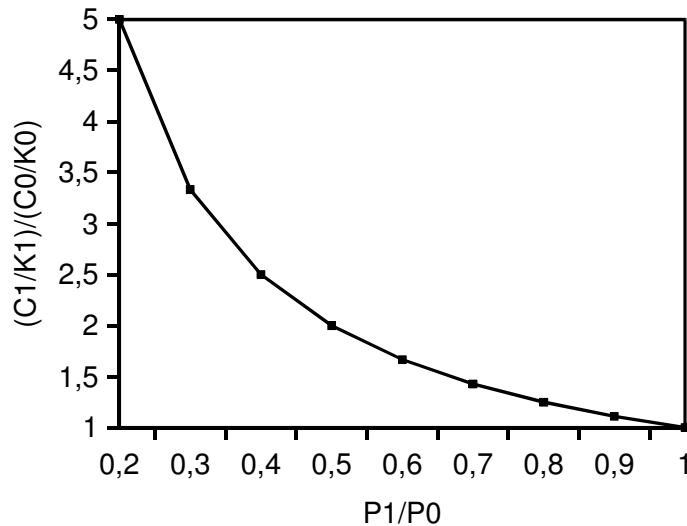


Abbildung 7.1: Notwendige Steigerung des Verhältnisses von Produktionsmenge zu Kosten bei Preisverfall

7.2 Steigerung der Effizienz der Fertigung

7.2.1 Anforderungen an die Effizienzsteigerung bedingt durch Preisverfall

Wie in Kapitel 4 erläutert, spielen die Rahmenbedingungen für die Rentabilität der IC-Fertigung eine große Rolle. Im Bereich der DRAMs hat bis jetzt ein ständiger starker Preisverfall stattgefunden, mit einer durchschnittlichen jährlichen Verringerung des Preises pro Bit um 32%, wie dies in Kapitel 1.4 dargestellt ist. Es ist also auch in Zukunft von einer kontinuierlichen, massiven Verringerung des erzielbaren Preises auszugehen.

Die Strukturgröße wird im Laufe der Produktlebensdauer durch Shrinkverfahren ständig verringert. Üblicherweise ist bei einer Verringerung der Strukturgröße um Faktor 0,7 die Rede von einer neuen Technologiegeneration oder einem neuen Technologieknoten. In den vergangenen Jahren war eine stetige Verringerung der Strukturgröße um Faktor 0,89 pro Jahr zu beobachten, was im Verlauf von 3 Jahren zu einer Reduktion um Faktor 0,7 führt. Unter Annahme der idealen Skalierbarkeit einer Schaltung kann davon ausgegangen werden, dass sich die Chipfläche entsprechend dem Quadrat des Verkleinerungsfaktors verringert. Eine Reduktion der Strukturgröße auf das 0,7-fache führt damit also ungefähr zu einer Halbierung der Chipfläche.

Bei einem jährlichen Preisverfall von 32%, wie er im langjährigen Durchschnitt auf dem Speichermarkt zu beobachten ist, ergibt sich nach Gleichung (7.9) für das Verhältnis C/K die Forderung:

$$\frac{C_1}{K_1} = 1,47 \cdot \frac{C_0}{K_0} \quad (7.10)$$

Das heißt, die Effizienz der Fertigung muss bei einem Preisverfall von 32% jährlich um 47% steigen. Bei einem für die Unternehmen günstigeren angenommenen Preisrückgang von 25% pro Jahr würde sich beispielsweise die notwendige Effizienzsteigerung auf 33% reduzieren.

In Abbildung 7.2 ist dargestellt, welche Auswirkungen sich durch konstanten mehrjährigen Preisverfall ergeben. Bei einem durchschnittlichen Preisverfall von 32% p.a., wie er für den DRAM-Markt typisch ist, liegt der Stückpreis des gleichen Produktes nach 6 Jahren nur noch bei ungefähr 10% des Ausgangspreises.

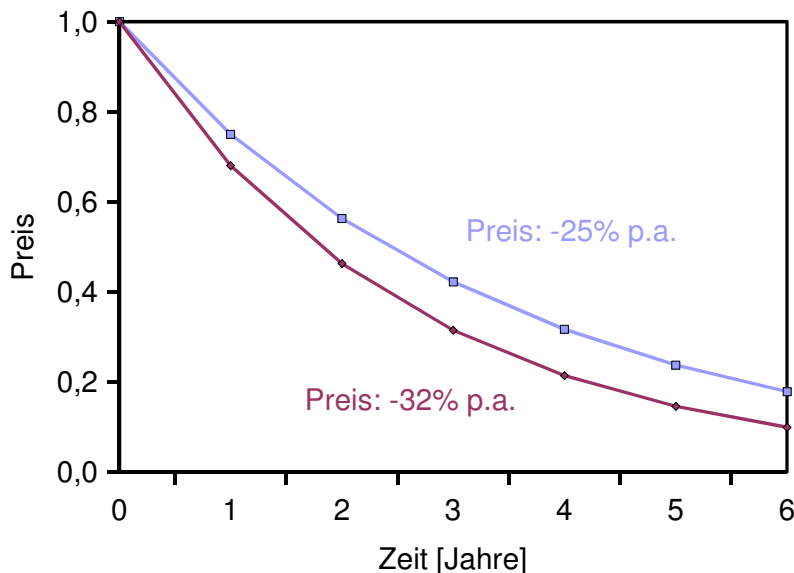


Abbildung 7.2: Entwicklung des Stückpreises bei kontinuierlichem Preisverfall über mehrere Jahre

Das bedeutet, dass innerhalb dieser Zeit die Effizienz der Fertigung insgesamt ungefähr um Faktor 10 gesteigert werden muss, was in Abbildung 7.3 dargestellt ist. Bei einem weniger starken Preisrückgang würden sich diese Anforderungen erheblich reduzieren. Ein durchschnittlicher Preisverfall von 25% pro Jahr würde beispielsweise bedeuten, dass innerhalb von 6 Jahren die Effizienz lediglich um Faktor 5,5 gesteigert werden müsste. Das zeigt, welchen enormen Einfluss der Preisverfall im Bereich der Chipfertigung besitzt und welchen großen Druck auf die Unternehmen dadurch entsteht.

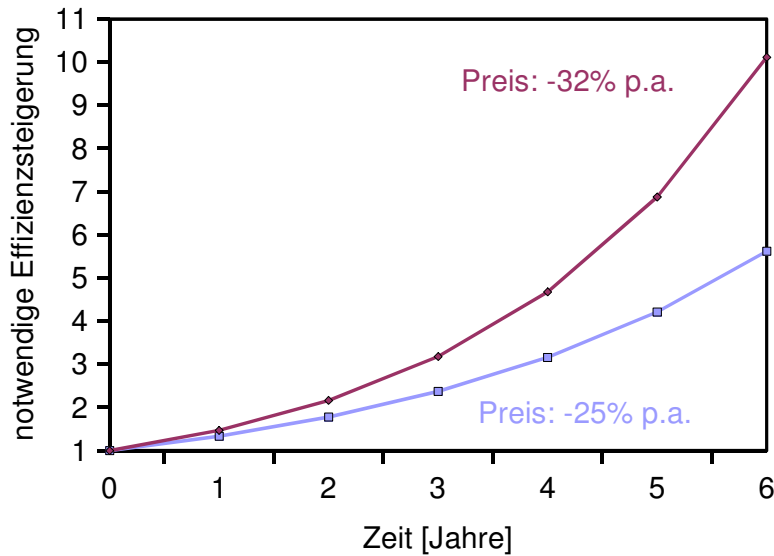


Abbildung 7.3: Notwendige Effizienzsteigerung bei konstantem Preisverfall über mehrere Jahre

7.2.2 Empfindlichkeit gegenüber Parameteränderungen

Nachdem festgestellt worden ist, welche Anforderungen an die Entwicklung der Effizienz der Fertigung gestellt werden müssen, um den Preisverfall zu kompensieren, muss untersucht werden, welchen Beitrag die einzelnen in Kapitel 5 und 6 beschriebenen Einflussparameter dazu leisten können.

Durch eine Berechnung der Empfindlichkeiten gegenüber Parameteränderungen kann festgestellt werden, welcher Parameter den stärksten Einfluss ausübt. Durch zusätzliche Angabe des Aufwands, der bei den jeweiligen Änderungen entsteht, läßt sich ermitteln, welche Maßnahmen unter ökonomischen und technischen Gesichtspunkten am ehesten geeignet sind.

Für angenommene 20000 WSPM, einen Anteil der Monitorwafer von 5%, einen Waferdurchmesser von 200mm, eine Chipfläche von 100mm² und eine Defektdichte von 0,001mm⁻² (Ausbeute = 90%) ergeben sich unter Benutzung von Gleichung (4.27) bis (4.32) die folgenden Werte für die relative Empfindlichkeit der Zahl der produzierten und funktionsfähigen Chips gegenüber Änderungen der Eingangsparameter.

$$\text{sensC} = \begin{bmatrix} \frac{\partial C}{\partial \text{WSPM}} \cdot \frac{\delta \text{WSPM}}{C_{AP}} \\ \frac{\partial C}{\partial m} \cdot \frac{\delta m}{C_{AP}} \\ \frac{\partial C}{\partial D_W} \cdot \frac{\delta D_W}{C_{AP}} \\ \frac{\partial C}{\partial D} \cdot \frac{\delta D}{C_{AP}} \\ \frac{\partial C}{\partial A} \cdot \frac{\delta A}{C_{AP}} \end{bmatrix} = \begin{bmatrix} 0,05 \\ -0,0026 \\ 0,1082 \\ -0,0050 \\ -0,0591 \end{bmatrix} \quad (7.11)$$

Das negative Vorzeichen bei einigen Zahlen bedeutet, dass sich die Zahl der produzierten Chips erhöht, wenn der Parameterwert verringert wird. Beispielsweise ändert sich die Zahl der Chips um 5,9%, wenn die Chipgröße um 5% sinkt. Die berechneten Werte sind im allgemeinen vom Arbeitspunkt abhängig und nicht konstant.

Mit (7.11) ergibt sich folgende Reihenfolge für die Empfindlichkeit gegenüber Parameteränderungen:

$$\text{sens}C_{DW} > \text{sens}C_A > \text{sens}C_{WSPM} > \text{sens}C_D > \text{sens}C_m$$

Es zeigt sich, dass die Änderung des Waferdurchmessers den größten Einfluss ausübt, gefolgt von der Chipfläche und der Zahl der gestarteten Wafer. Den geringsten Einfluss haben die Defektdichte und die Zahl der Monitorwafer, was auch in Abbildung 7.4 dargestellt ist, wobei für die Wafergröße nicht 5% Änderung, sondern der realistische Umstieg von 200 mm auf 300 mm angegeben ist.

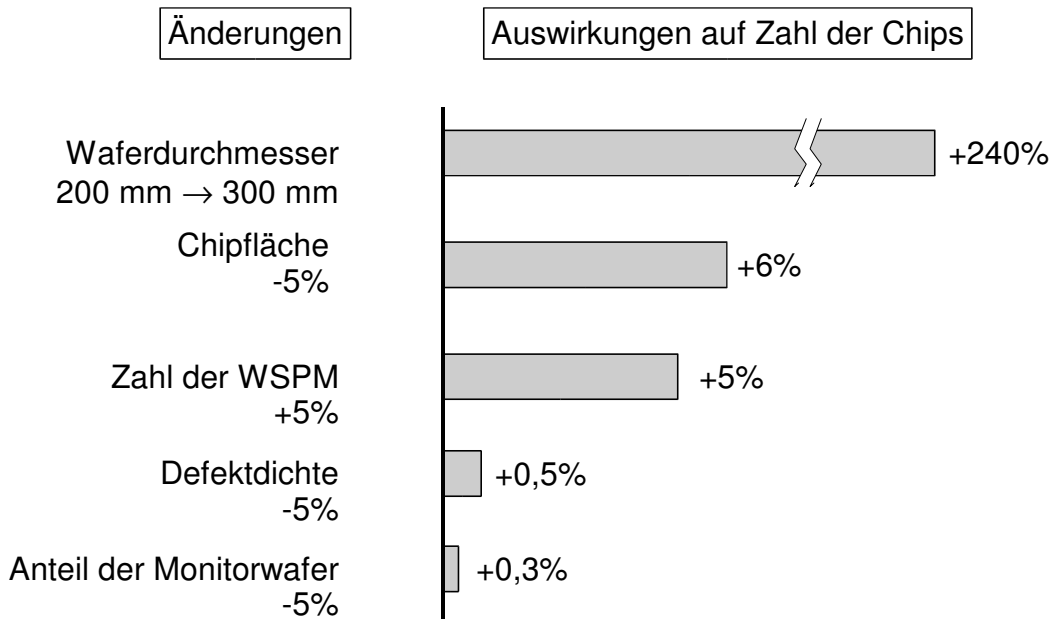


Abbildung 7.4: Änderung der Zahl der Chips bei Verbesserungen

In Abbildung 7.5 wird für die einzelnen Parameter noch einmal der Einfluss auf die Zahl der produzierten Chips angegeben.

Ein Umstieg von 200 mm- auf 300 mm-Wafer ist aber während des Fabbetriebs wegen des damit verbundenen hohen Aufwands praktisch ausgeschlossen. Als erste und wichtigste Maßnahme zur Steigerung der Effizienz bietet sich also die Verringerung der Chipgröße durch Shrinkverfahren an.

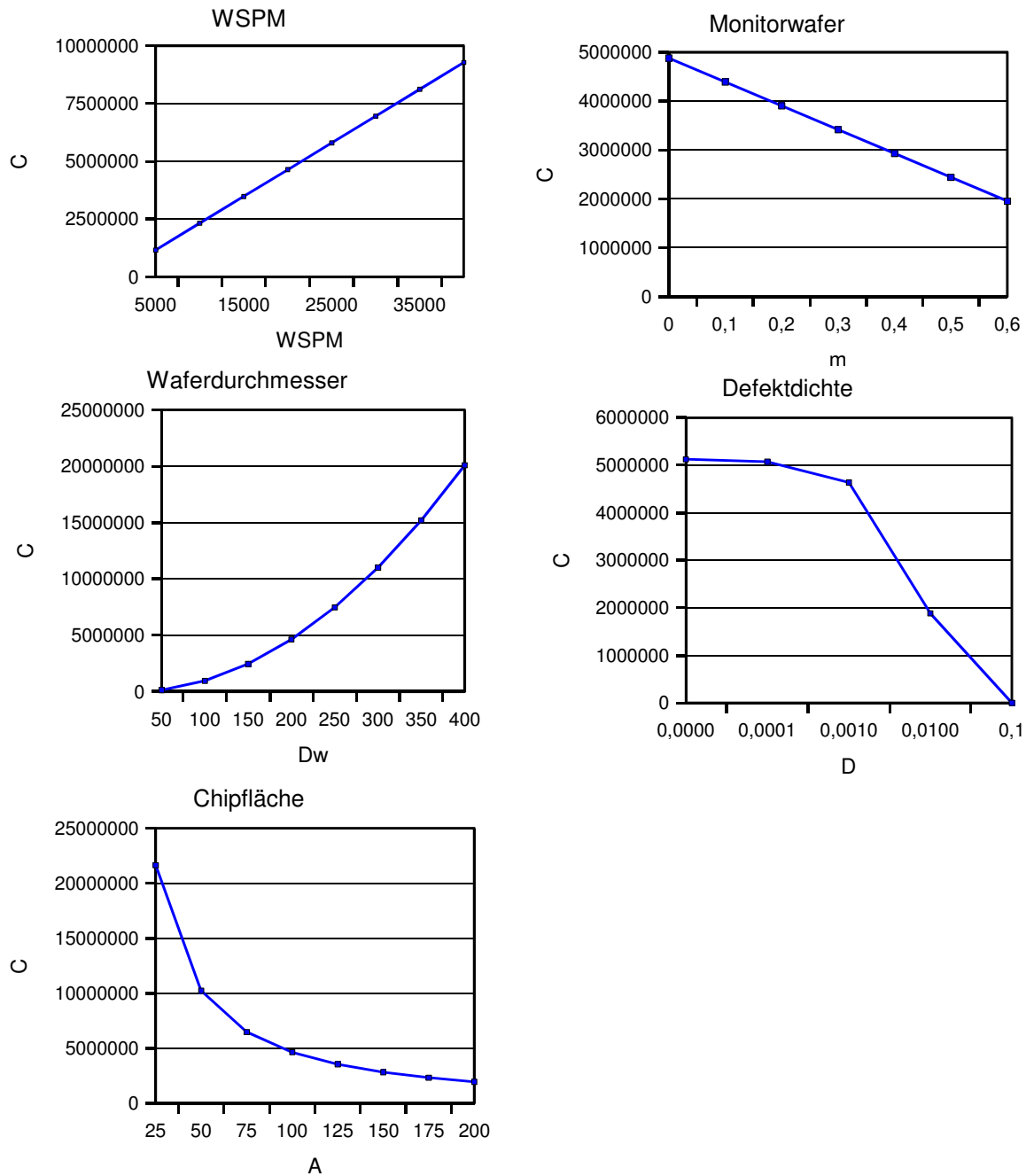


Abbildung 7.5: Empfindlichkeit gegenüber Parameteränderungen

Bei einem angenommenen Preisverfall von 32% pro Jahr muss die Effizienz der Fertigung um 47% p.a. steigen, um die negativen Aspekte des Preisverfalls ausgleichen zu können. Ohne Berücksichtigung von möglichen Kostensteigerungen ergibt sich mit Gleichung (4.19), dass die Chipfläche pro Jahr um 27% fallen müsste, um durch diese Maßnahme allein den Preisverfall auszugleichen. Das bedeutet, dass eine Verringerung der Strukturgröße um Faktor 0,7 in ungefähr 2,25 Jahren durchgeführt werden muss. Eine Reduzierung der Strukturgröße um Faktor 0,7 alle 3 Jahre, wie sie früher meistens üblich war, würde erst ab einem Preisverfall, der günstiger als 25% p.a. ausfällt, zu einem gleichbleibenden Ergebnis führen.

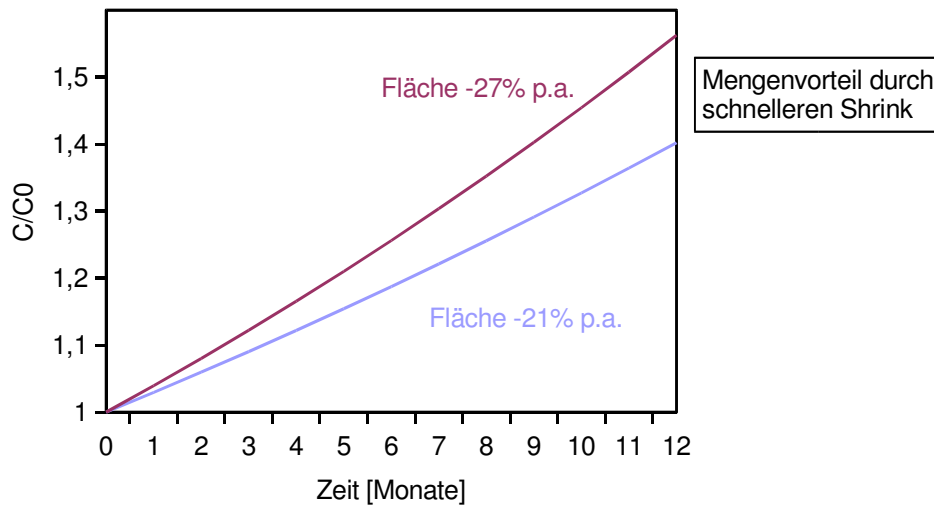


Abbildung 7.6: Vorteile bzgl. der Produktionsmenge bei schnellerem Shrink

In Abbildung 7.6 werden die beiden Shrinksszenarien miteinander verglichen. Während bei dem schnellen Shrink die Zahl der Chips nach einem Jahr um Faktor 1,47 gestiegen ist, wird beim langsameren Shrink nur eine Steigerung um Faktor 1,33 erzielt.

Wie in Gleichung (7.11) deutlich wird, kann die Produktionsleistung zusätzlich durch Verringerung der Defektdichte positiv beeinflusst werden. Falls dadurch die Ausbeute um 5% p.a. gesteigert werden kann, ergibt dies eine geringfügige Erleichterung der Anforderungen an die Strukturgrößenverringern. Bei einem Preisverfall von 32% p.a. würde dann eine Verringerung der Chipfläche um 26% p.a. ausreichen, um eine konstante Wirtschaftlichkeit zu erreichen, bzw. könnte bei einer Reduzierung der Strukturgröße um Faktor 0,7 alle 3 Jahre ein Preisverfall von 27% p.a. ausgeglichen werden.

Wenn zusätzlich berücksichtigt wird, dass durch die Shinks Kostensteigerungen verursacht werden, erhöhen sich die Anforderungen an die notwendige Leistungssteigerung. Bei einer angenommenen 5%-igen Kostensteigerung pro Jahr und einem Preisverfall von 32% muss eine Steigerung der Chipmenge um 54% p.a. erzielt werden. Bei einer berücksichtigten Ausbeutesteigerung von 5% pro Jahr bedeutet dies, dass die Chipfläche um 30% pro Jahr sinken muss, was einer Verringerung der Strukturgröße um Faktor 0,7 alle 2 Jahre entspricht. Eine etwas langsamere Strukturgrößenverkleinerung um Faktor 0,7 in 3 Jahren würde dann zu einer konstanten Wirtschaftlichkeit führen, wenn der Preis nur um höchstens 23% p.a. sinkt.

7.3 Berechnung der Auswirkungen auf die Effizienz mittels Warteschlangenmodell für Prozessfluss

Im folgenden soll mit Hilfe des in Kapitel 6 erstellten Warteschlangenmodells dargestellt werden, welche Auswirkungen sich durch Veränderung der Eingangsparameter ergeben. Als Ausgangspunkt für die Betrachtungen werden die Parameter aus den realen Gerätedaten gewählt. Mit den Eingangsparametern $r_e=30W/h$, $m_f=300h$, $m_r=5h$, $t_{\text{setup}}=0,1h$, $t_T=0,1h$, Losgröße=25W, $c_a=2$, $c_e=0,5$ ergeben sich die in Tabelle 7.1 und Abbildung 7.7 dargestellten Ergebnisse für die Durchlaufzeit des Modells bestehend aus 10 aufeinander folgenden Geräten.

ra [W/h]	ra [Lose]	TD ges. [h]	WIP ges. [W]
2	0,08	12,09	24,17
4	0,16	13,76	55,04
6	0,24	15,33	91,99
8	0,32	16,7	133,62
10	0,4	17,86	178,61
12	0,48	18,9	226,83
14	0,56	20,02	280,22
16	0,64	21,48	343,71
18	0,72	23,72	427,04
20	0,8	27,55	551,09
22	0,88	35,13	772,92
24	0,96	55,76	1338,34
26	1,04	288,24	7494,29

Tabelle 7.1: TD und WIP für Startparameter des Modells

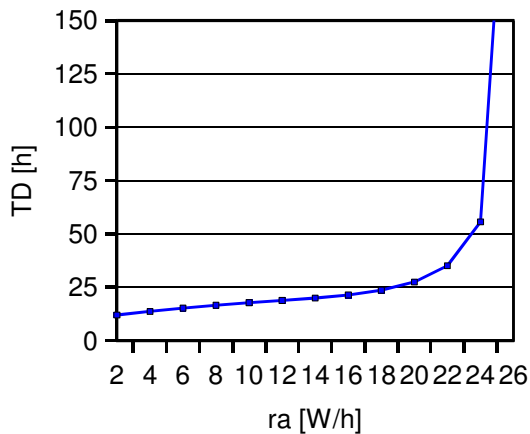


Abbildung 7.7: Durchlaufzeit des Modells

Auf Basis der Daten aus Tabelle 7.1 sollen nun die Verbesserungsmöglichkeiten durch Änderung der Eingangparameter beurteilt werden. Zur besseren Vergleichbarkeit werden dazu die Parameter jeweils um 10% bzw. 20% verändert und die Auswirkungen auf die Durchlaufzeit berechnet.

7.3.1 Einfluss der Variabilität

Ausgehend von den Startwerten ($c_a=2$, $c_e=0,5$) wird die Variabilität des Ankunftsprozesses c_a im Prozessfluss um 10% bzw. 20% auf $c_a=1,8$ bzw. $c_a=1,6$ verringert. Die sich daraus ergebenden Änderungen der Durchlaufzeit sind in Tabelle 7.2 und in grafischer Form in Abbildung 7.8 dargestellt. Wie zu erkennen ist, ist die prozentuale Verbesserung größer bei hohem Nutzungsgrad der Geräte, da sich im steilen Bereich der Betriebskennlinie Änderungen stärker auswirken als im flachen Bereich bei niedrigem Nutzungsgrad.

ra [W/h]	ca=2,0	ca-10%		ca-20%	
	TD [h]	TD [h]	Änderung [%]	TD [h]	Änderung [%]
2	12,09	11,80	-2,38	11,54	-4,51
4	13,76	13,18	-4,22	12,66	-7,99
6	15,33	14,49	-5,50	13,73	-10,43
8	16,70	15,65	-6,31	14,71	-11,96
10	17,86	16,66	-6,75	15,58	-12,78
12	18,90	17,59	-6,92	16,42	-13,12
14	20,02	18,62	-6,95	17,38	-13,18
16	21,48	19,99	-6,95	18,65	-13,16
18	23,72	22,07	-6,98	20,59	-13,23
20	27,55	25,59	-7,12	23,84	-13,48
22	35,13	32,54	-7,38	30,22	-13,99
24	55,76	51,40	-7,83	47,49	-14,83
26	288,24	263,70	-8,51	241,74	-16,13

Tabelle 7.2: Durchlaufzeit bei Verringerung von c_a

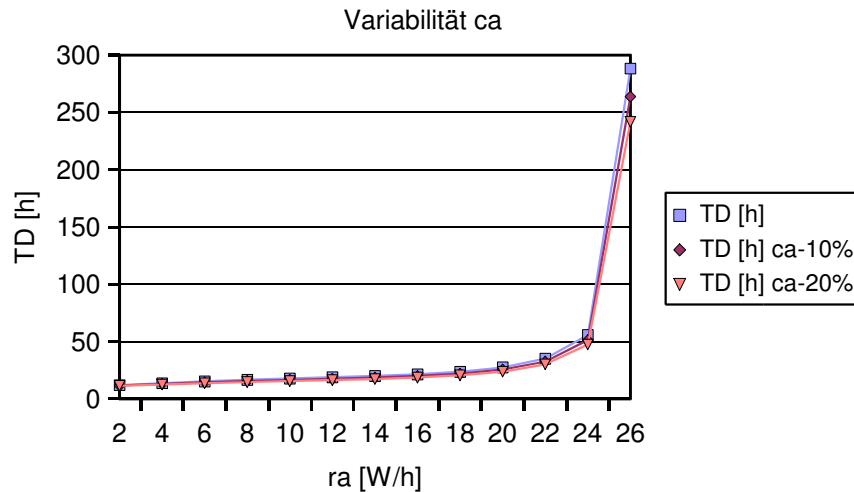


Abbildung 7.8: Durchlaufzeit bei Verringerung von c_a

Im weiteren wird die Variabilität des Bearbeitungsprozesses c_e bei gleichbleibendem c_a um 10% bzw. 20% auf $c_e=0,45$ bzw. $c_e=0,4$ gesenkt. Die Ergebnisse dieser Änderung sind in Tabelle 7.3 und Abbildung 7.9 angegeben.

ra [W/h]	ce=0,5	ce-10%		ce-20%	
	TD [h]	TD [h]	Änderung [%]	TD [h]	Änderung [%]
2	12,09	12,07	-0,16	12,05	-0,30
4	13,76	13,72	-0,32	13,68	-0,61
6	15,33	15,25	-0,52	15,18	-0,98
8	16,70	16,57	-0,78	16,46	-1,47
10	17,86	17,66	-1,12	17,48	-2,11
12	18,90	18,61	-1,55	18,35	-2,94
14	20,02	19,59	-2,10	19,22	-3,99
16	21,48	20,88	-2,79	20,35	-5,29
18	23,72	22,86	-3,63	22,09	-6,88
20	27,55	26,27	-4,66	25,12	-8,84
22	35,13	33,04	-5,95	31,17	-11,28
24	55,76	51,52	-7,60	47,73	-14,41
26	288,24	260,01	-9,80	234,74	-18,56

Tabelle 7.3: Durchlaufzeit bei Verringerung von c_e

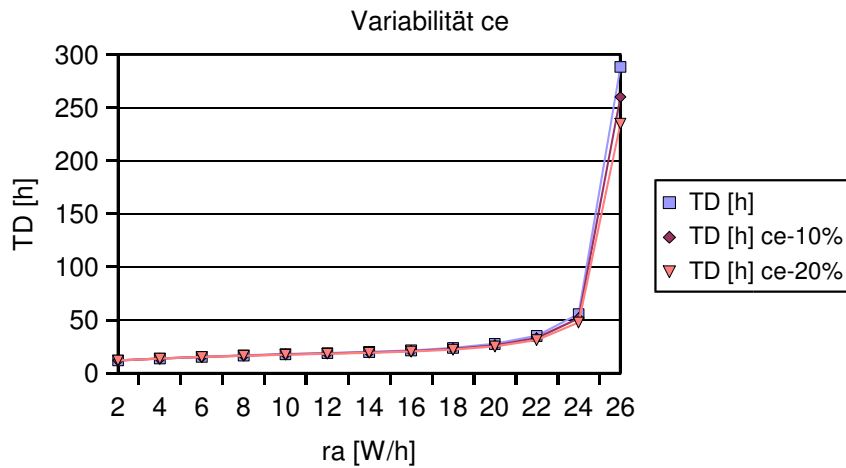


Abbildung 7.9: Durchlaufzeit bei Verringerung von c_e

Schließlich soll noch untersucht werden, welche Verbesserungen der Durchlaufzeit sich durch eine gemeinsame Verringerung von c_a und c_e erzielen lassen. Die Ergebnisse für $c_a=1,8$, $c_e=0,45$ und $c_a=1,6$, $c_e=0,4$ sind in Tabelle 7.4 und Abbildung 7.10 dargestellt.

ra [W/h]	ca=2,0, ce=0,5	ca, ce-10%		ca, ce-20%	
	TD [h]	TD [h]	Änderung [%]	TD [h]	Änderung [%]
2	12,09	11,78	-2,54	11,50	-4,81
4	13,76	13,13	-4,54	12,58	-8,60
6	15,33	14,41	-6,02	13,58	-11,41
8	16,70	15,52	-7,09	14,46	-13,43
10	17,86	16,46	-7,86	15,20	-14,89
12	18,90	17,30	-8,47	15,87	-16,06
14	20,02	18,20	-9,06	16,58	-17,17
16	21,48	19,39	-9,74	17,52	-18,45
18	23,72	21,21	-10,61	18,95	-20,11
20	27,55	24,31	-11,78	21,40	-22,32
22	35,13	30,45	-13,34	26,25	-25,27
24	55,76	47,16	-15,43	39,46	-29,24
26	288,24	235,47	-18,31	188,24	-34,69

Tabelle 7.4: Durchlaufzeit bei Verringerung von c_a und c_e

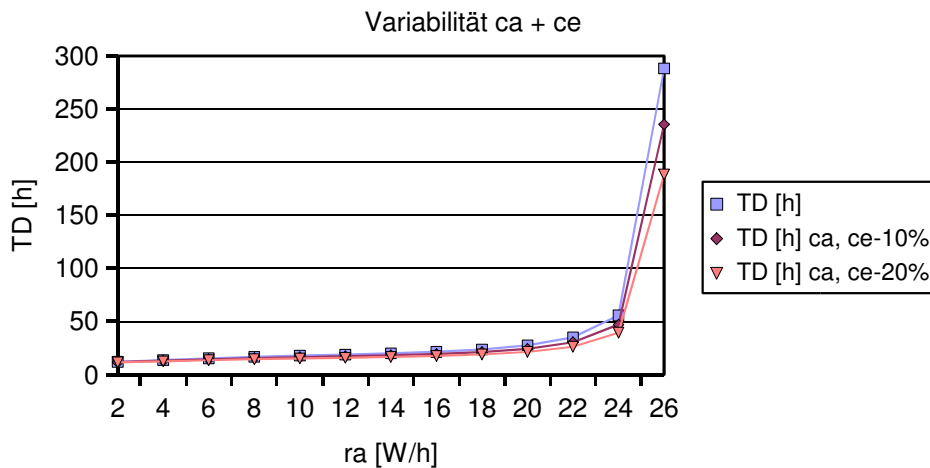


Abbildung 7.10: Durchlaufzeit bei Verringerung von c_a und c_e

Es zeigt sich, dass, ausgehend von den Startwerten, die Verringerung von c_a und c_e jeweils ungefähr gleichen Einfluss auf die Durchlaufzeit hat. Durch die Verkleinerung von c_a und c_e gemeinsam kann in etwa eine Verdoppelung des Effekts erreicht werden.

7.3.2 Einfluss des Gerätedurchsatzes

In Tabelle 7.5 und Abbildung 7.11 werden die Auswirkungen einer Änderung des Gerätedurchsatzes um 10% bzw. 20% angegeben. Es zeigen sich besonders bei hohem Nutzungsgrad sehr starke Auswirkungen, da bei einer Erhöhung des Gerätedurchsatzes insgesamt eine höhere Zahl von Waferstarts möglich ist und damit die Betriebskennlinie viel flacher verläuft.

ra [W/h]	re=30 W/h	re+10%		re+20%	
	TD [h]	TD [h]	Änderung [%]	TD [h]	Änderung [%]
2	12,09	11,06	-8,49	10,22	-15,42
4	13,76	12,47	-9,34	11,44	-16,89
6	15,33	13,83	-9,80	12,61	-17,72
8	16,70	15,04	-9,96	13,69	-18,04
10	17,86	16,08	-9,99	14,63	-18,10
12	18,90	16,98	-10,15	15,44	-18,30
14	20,02	17,87	-10,74	16,20	-19,09
16	21,48	18,89	-12,08	16,98	-20,94
18	23,72	20,28	-14,53	17,95	-24,34
20	27,55	22,39	-18,74	19,29	-30,01
22	35,13	25,91	-26,25	21,30	-39,37
24	55,76	32,61	-41,52	24,59	-55,91
26	288,24	49,41	-82,86	30,65	-89,37
28	-	159,24	-	44,98	-
30	-	-	-	116,19	-

Tabelle 7.5: Durchlaufzeit bei Erhöhung des Gerätedurchsatzes

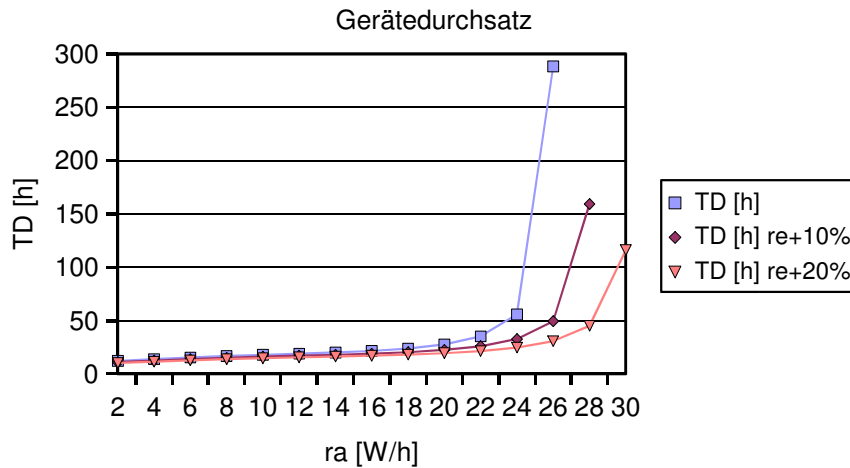


Abbildung 7.11: Durchlaufzeit bei Erhöhung des Gerätedurchsatzes

7.3.3 Einfluss von Unterbrechungen im Prozessablauf

An dieser Stelle soll untersucht werden, wie sich die Gerätezuverlässigkeit (MTBF, MTTR) auf die Durchlaufzeit auswirken kann. Tabelle 7.6 und Abbildung 7.12 zeigen die Gesamtdurchlaufzeit bei einer Änderung von MTBF und MTTR von 10% und 20%. Die Änderungen sind, außer bei sehr hohem Nutzungsgrad, relativ gering, da bei diesen Werten schon ein relativ hohes Niveau erreicht ist und somit das Verbesserungspotential eher niedrig ausfällt.

ra [W/h]	MTBF=300h, MTTR=5h	MTBF+10% MTTR-10%		MTBF+20% MTTR-20%	
	TD [h]	TD [h]	Änderung [%]	TD [h]	Änderung [%]
2	12,09	12,05	-0,28	12,02	-0,52
4	13,76	13,72	-0,31	13,68	-0,57
6	15,33	15,28	-0,33	15,24	-0,60
8	16,70	16,65	-0,33	16,60	-0,61
10	17,86	17,80	-0,33	17,75	-0,61
12	18,90	18,84	-0,34	18,78	-0,62
14	20,02	19,94	-0,37	19,88	-0,67
16	21,48	21,39	-0,43	21,31	-0,78
18	23,72	23,60	-0,54	23,49	-0,99
20	27,55	27,35	-0,75	27,18	-1,36
22	35,13	34,72	-1,17	34,38	-2,13
24	55,76	54,44	-2,38	53,38	-4,27
26	288,24	245,86	-14,70	219,17	-23,96

Tabelle 7.6: Durchlaufzeit bei Erhöhung der Gerätezuverlässigkeit

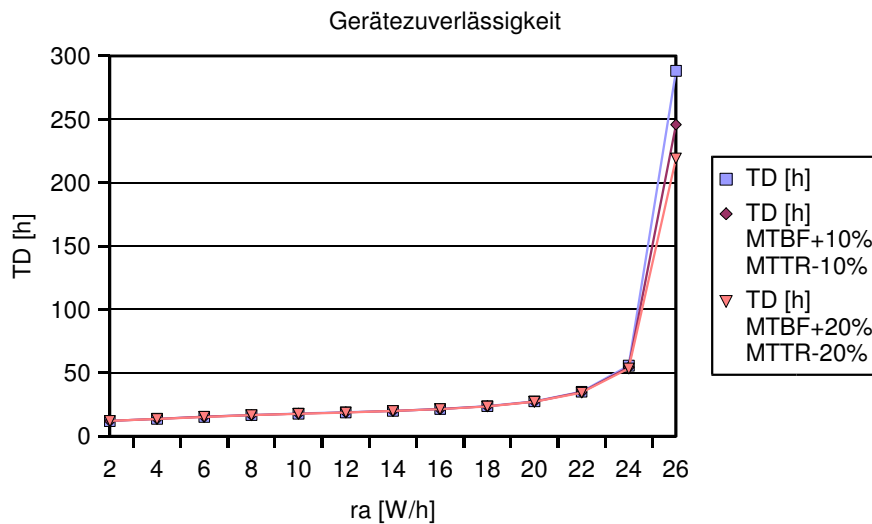


Abbildung 7.12: Durchlaufzeit bei Erhöhung der Gerätezuverlässigkeit

Zusätzlich werden die Auswirkungen von Verkürzungen der Setupzeit berechnet. Diese Daten sind in Tabelle 7.7 und Abbildung 7.13 angegeben. Bei hoher Waferankunftsrate wirkt sich eine Verringerung der Setupzeit, die vor jedem Los anfällt, höher aus als bei niedrigem Nutzungsgrad.

ra [W/h]	tsetup=0,1h	tsetup=10%		tsetup=20%	
	TD [h]	TD [h]	Änderung [%]	TD [h]	Änderung [%]
2	12,09	11,95	-1,11	11,82	-2,22
4	13,76	13,59	-1,23	13,42	-2,45
6	15,33	15,13	-1,29	14,94	-2,58
8	16,70	16,48	-1,31	16,27	-2,61
10	17,86	17,63	-1,31	17,39	-2,62
12	18,90	18,65	-1,34	18,40	-2,68
14	20,02	19,73	-1,45	19,44	-2,88
16	21,48	21,12	-1,68	20,77	-3,32
18	23,72	23,22	-2,11	22,74	-4,15
20	27,55	26,76	-2,89	26,00	-5,63
22	35,13	33,56	-4,48	32,13	-8,56
24	55,76	50,88	-8,76	46,83	-16,02
26	288,24	171,97	-40,34	123,42	-57,18

Tabelle 7.7: Durchlaufzeit bei Verringerung der Setupzeit

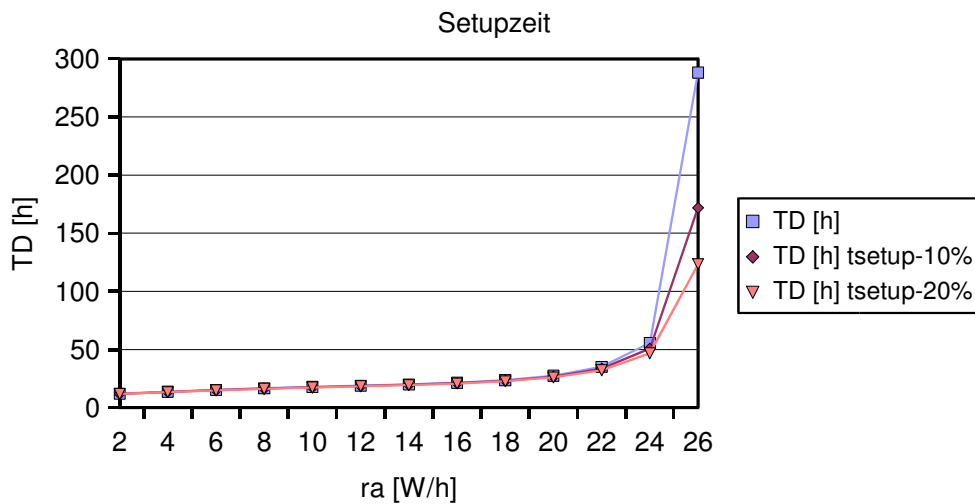


Abbildung 7.13: Durchlaufzeit bei Verringerung der Setupzeit

In Abbildung 7.14 wird der Einfluss der verschiedenen Faktoren auf die Durchlaufzeit noch einmal zusammenfassend dargestellt. Dabei werden die Verbesserungen bei einem Nutzungsgrad von 90% (entspricht $r_a=24W/h$) und 75% (entspricht $r_a=20W/h$) angegeben. Wegen der Steilheit der Betriebskennlinie fallen die Verbesserungen bei hohem Nutzungsgrad größer aus. Qualitativ ergibt sich aber in beiden Fällen die gleiche Reihenfolge der Maßnahmen.

Die größten Verbesserungen können durch eine Erhöhung des Gerätedurchsatzes erzielt werden, gefolgt von einer Verringerung der Variabilität c_a+c_e . Etwas geringere Verbesserungen ergeben sich durch die Verringerung der Setupzeit und der Variabilität c_a oder c_e . Die geringsten Effekte lassen sich durch eine weitere Erhöhung der Gerätezuverlässigkeit erzielen.

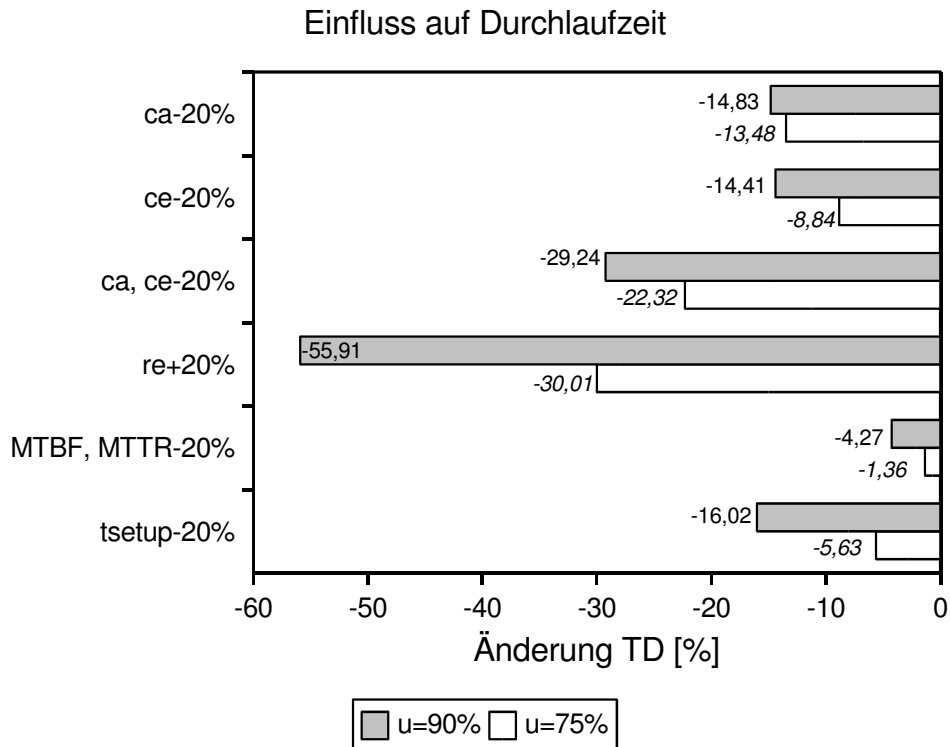


Abbildung 7.14: Einfluss auf Durchlaufzeit - Zusammenfassung

Eine Verbesserung im Produktionsablauf und damit eine Änderung der Betriebskennlinie kann einerseits zur Verringerung der Durchlaufzeit bei gleichbleibender Produktionsmenge und andererseits zur Steigerung der Produktionsmenge bei konstanter Durchlaufzeit genutzt werden. In Abbildung 7.15 ist angegeben, welche Steigerung von r_a sich durch die angegebenen Maßnahmen ergeben würde, ausgehend von einer konstanten Durchlaufzeit von 55,76h (entspricht $u=90\%$) bzw. 27,55h (entspricht $u=75\%$). Hier ist auffällig, dass die Änderung bei niedrigerem Nutzungsgrad stärker ausfällt als bei hohem Nutzungsgrad. Hintergrund dafür ist, dass man sich bei niedrigem Nutzungsgrad in einem relativ flachen Bereich der Betriebskennlinie befindet und sich damit Änderungen bei der Zahl der Wafer Starts noch vergleichsweise schwach auf die Durchlaufzeit auswirken.

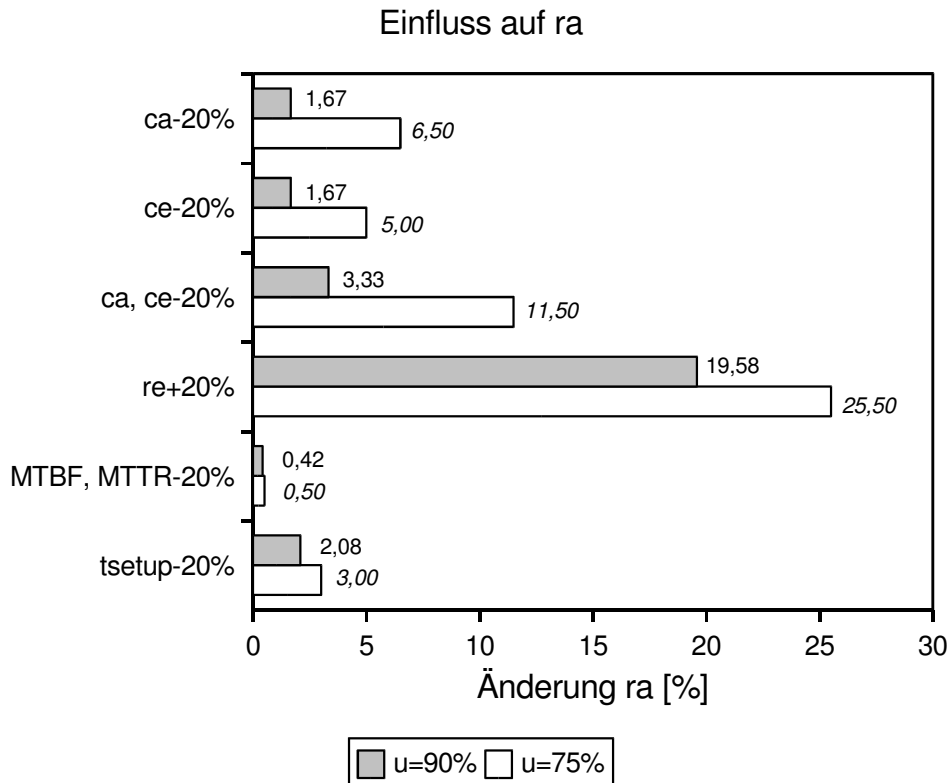


Abbildung 7.15: Einfluss der Maßnahmen auf Produktionsmenge

7.3.4 Einfluss der Position im Prozessablauf

Im folgenden soll untersucht werden, an welcher Stelle im Prozessablauf sich Veränderungen am stärksten auswirken und an welcher Stelle es damit am interessantesten wäre Verbesserungen durchzuführen.

Diese Untersuchungen werden hier für die Variabilität, den Gerätedurchsatz und die Setupzeit durchgeführt. Dazu wird im Modell der jeweilige Parameter beim ersten Gerät des Prozessflusses, beim fünften Gerät und beim letzten Gerät um 20% verändert und die Änderung der Durchlaufzeit berechnet.

In Tabelle 7.8 und Abbildung 7.16 sind die Auswirkungen einer Änderung der Variabilität c_e an unterschiedlichen Stellen im Prozessfluss dargestellt. Der geringste Einfluss ergibt sich, wenn die Variabilität beim letzten Prozessschritt reduziert wird. Am Anfang und in der Mitte des Prozessablaufs ergeben sich ungefähr gleich große Veränderungen.

ra [W/h]	ce=0,5 TD [h]	Anfang		Mitte		Ende	
		TD [h]	Änderung [%]	TD [h]	Änderung [%]	TD [h]	Änderung [%]
2	12,09	12,08	-0,03	12,08	-0,03	12,08	-0,03
4	13,76	13,75	-0,07	13,75	-0,06	13,75	-0,06
6	15,33	15,32	-0,11	15,32	-0,10	15,32	-0,08
8	16,70	16,67	-0,18	16,68	-0,15	16,68	-0,11
10	17,86	17,82	-0,26	17,82	-0,22	17,84	-0,15
12	18,90	18,84	-0,35	18,84	-0,32	18,87	-0,19
14	20,02	19,92	-0,47	19,93	-0,44	19,97	-0,24
16	21,48	21,35	-0,61	21,36	-0,58	21,42	-0,31
18	23,72	23,54	-0,77	23,55	-0,75	23,63	-0,39
20	27,55	27,29	-0,97	27,29	-0,96	27,42	-0,48
22	35,13	34,71	-1,22	34,71	-1,21	34,92	-0,61
24	55,76	54,91	-1,53	54,91	-1,53	55,34	-0,77
26	288,24	282,60	-1,96	282,60	-1,96	285,42	-0,98

Tabelle 7.8: Einfluss der Variabilität c_e an verschiedenen Stellen im Prozessfluss

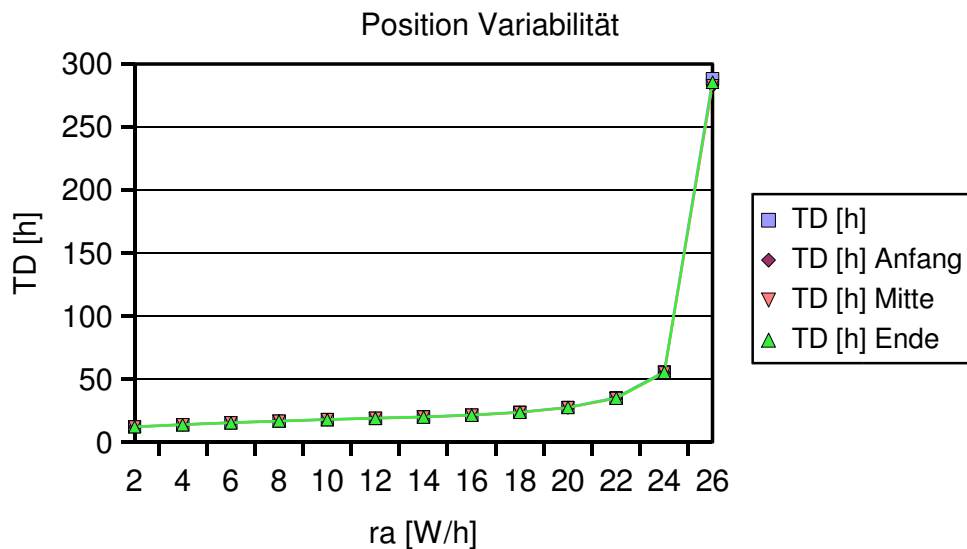


Abbildung 7.16: Einfluss der Variabilität c_e an verschiedenen Stellen im Prozessfluss

Tabelle 7.9 und Abbildung 7.17 zeigen den Einfluss eines veränderten Gerätedurchsatzes an unterschiedlichen Stellen im Prozessfluss.

ra [W/h]	re=30W/h TD [h]	Anfang		Mitte		Ende	
		TD [h]	Änderung [%]	TD [h]	Änderung [%]	TD [h]	Änderung [%]
2	12,09	11,90	-1,54	11,90	-1,54	11,90	-1,54
4	13,76	13,53	-1,68	13,53	-1,67	13,53	-1,67
6	15,33	15,06	-1,75	15,07	-1,73	15,07	-1,71
8	16,70	16,41	-1,78	16,42	-1,72	16,42	-1,66
10	17,86	17,53	-1,84	17,56	-1,67	17,58	-1,57
12	18,90	18,52	-2,01	18,59	-1,64	18,63	-1,47
14	20,02	19,54	-2,40	19,69	-1,63	19,73	-1,41
16	21,48	20,81	-3,13	21,12	-1,69	21,17	-1,44
18	23,72	22,68	-4,39	23,30	-1,81	23,35	-1,59
20	27,55	25,76	-6,52	27,00	-2,03	27,04	-1,88
22	35,13	31,54	-10,21	34,27	-2,45	34,30	-2,38
24	55,76	46,21	-17,14	53,93	-3,29	53,94	-3,28
26	288,24	196,22	-31,92	273,48	-5,12	273,48	-5,12

Tabelle 7.9: Einfluss des Durchsatzes an verschiedenen Stellen im Prozessfluss

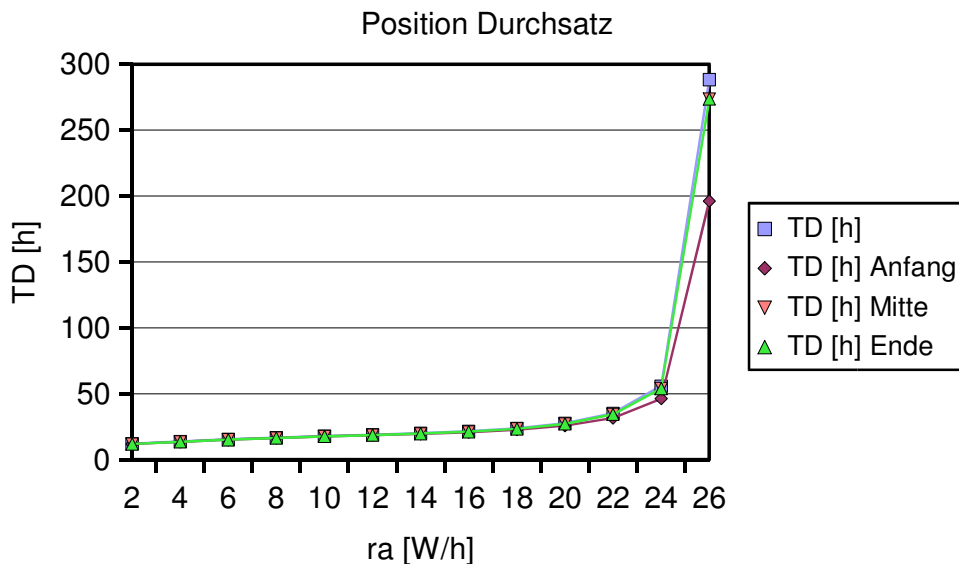


Abbildung 7.17: Einfluss des Durchsatzes an verschiedenen Stellen im Prozessfluss

Beim ersten Prozessschritt ergeben sich eindeutig stärkere Auswirkungen als bei einer Änderung in der Mitte oder am Ende des Prozessablaufs.

Der Einfluss von Unterbrechungen an verschiedenen Stellen im Prozessablauf durch MTBF und MTRR ergibt sich aus Tabelle 7.10 und Abbildung 7.18. Auch hier ergibt sich der stärkste Effekt beim ersten Prozessschritt.

ra [W/h]	MTBF=300h MTTR=5h		Anfang		Mitte		Ende	
	TD [h]		TD [h]	Änderung [%]	TD [h]	Änderung [%]	TD [h]	Änderung [%]
2	12,09		12,08	-0,05	12,08	-0,05	12,08	-0,05
4	13,76		13,75	-0,06	13,75	-0,06	13,75	-0,06
6	15,33		15,32	-0,06	15,32	-0,06	15,32	-0,06
8	16,70		16,69	-0,06	16,69	-0,06	16,69	-0,06
10	17,86		17,85	-0,07	17,85	-0,06	17,85	-0,06
12	18,90		18,89	-0,08	18,89	-0,06	18,89	-0,05
14	20,02		20,00	-0,10	20,00	-0,06	20,01	-0,05
16	21,48		21,45	-0,15	21,47	-0,07	21,47	-0,05
18	23,72		23,67	-0,23	23,71	-0,08	23,71	-0,06
20	27,55		27,45	-0,38	27,53	-0,09	27,53	-0,08
22	35,13		34,88	-0,72	35,09	-0,13	35,09	-0,13
24	55,76		54,80	-1,73	55,63	-0,25	55,63	-0,25
26	288,24		255,86	-11,23	284,29	-1,37	284,29	-1,37

Tabelle 7.10: Einfluss von MTBF, MTTR an verschiedenen Stellen im Prozessfluss

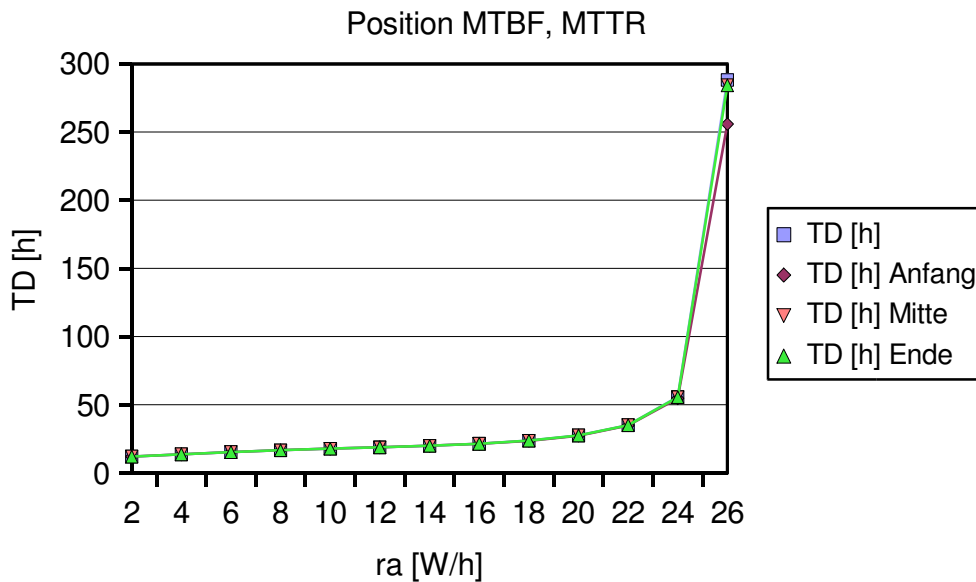


Abbildung 7.18: Einfluss von MTBF, MTTR an verschiedenen Stellen im Prozessfluss

Für die Setupzeit werden die Auswirkungen in Tabelle 7.11 und Abbildung 7.19 dargestellt.

ra [W/h]	tsetup=0,1h	Anfang		Mitte		Ende	
	TD [h]	TD [h]	Änderung [%]	TD [h]	Änderung [%]	TD [h]	Änderung [%]
2	12,09	12,06	-0,22	12,06	-0,22	12,06	-0,22
4	13,76	13,73	-0,25	13,73	-0,25	13,73	-0,24
6	15,33	15,29	-0,26	15,29	-0,26	15,29	-0,25
8	16,70	16,66	-0,27	16,66	-0,26	16,66	-0,25
10	17,86	17,81	-0,29	17,82	-0,26	17,82	-0,24
12	18,90	18,84	-0,34	18,85	-0,26	18,86	-0,22
14	20,02	19,93	-0,43	19,96	-0,27	19,97	-0,22
16	21,48	21,35	-0,61	21,42	-0,29	21,43	-0,23
18	23,72	23,50	-0,93	23,65	-0,32	23,66	-0,27
20	27,55	27,13	-1,54	27,45	-0,38	27,46	-0,35
22	35,13	34,14	-2,83	34,95	-0,53	34,95	-0,51
24	55,76	52,26	-6,28	55,24	-0,93	55,25	-0,93
26	288,24	213,19	-26,04	278,82	-3,27	278,82	-3,27

Tabelle 7.11: Einfluss der Setupzeit an verschiedenen Stellen im Prozessfluss

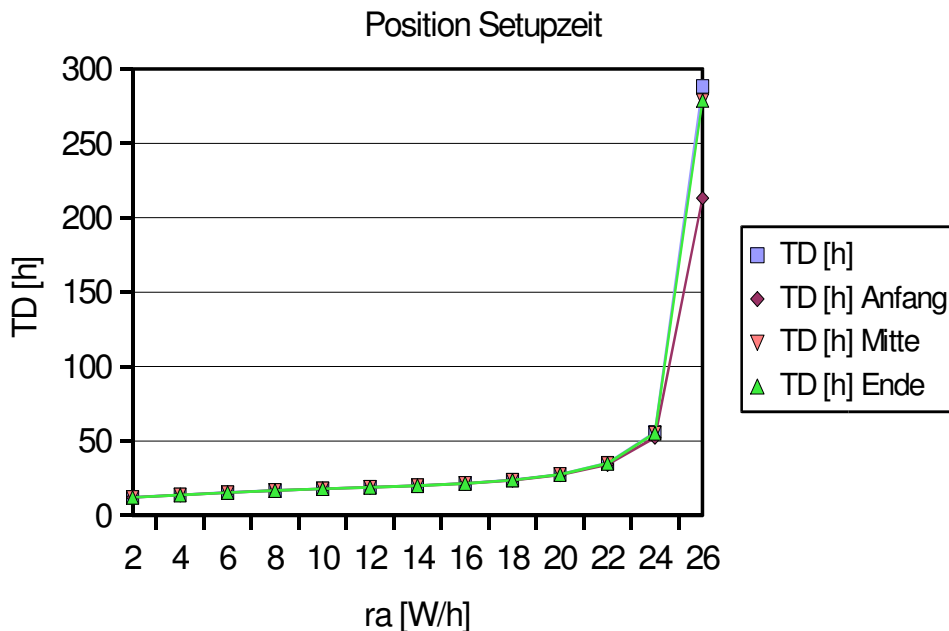


Abbildung 7.19: Einfluss der Setupzeit an verschiedenen Stellen im Prozessfluss

Auch für die Unterbrechungen im Prozessablauf zeigt sich eindeutig, dass der erste Prozessschritt als der kritischere angesehen werden muss. Der Einfluss von Änderungen in der Mitte oder am Ende des Prozessflusses erweist sich als klar geringer.

Abbildung 7.20 zeigt in der Zusammenfassung den Einfluss der einzelnen Faktoren an unterschiedlicher Stelle im Prozessablauf auf die Durchlaufzeit, ausgehend von $r_a=24\text{W/h}$ ($u=90\%$) und $r_a=20\text{W/h}$ ($u=75\%$). Die stärksten Auswirkungen zeigen sich jeweils beim ersten Schritt im Prozessablauf. Die Effekte in der Mitte und am Ende des Prozesses sind jeweils ungefähr gleich. Damit ergibt sich, dass Änderungen und Verbesserungen im Prozessfluss am effektivsten sind, wenn sie am Anfang durchgeführt werden.

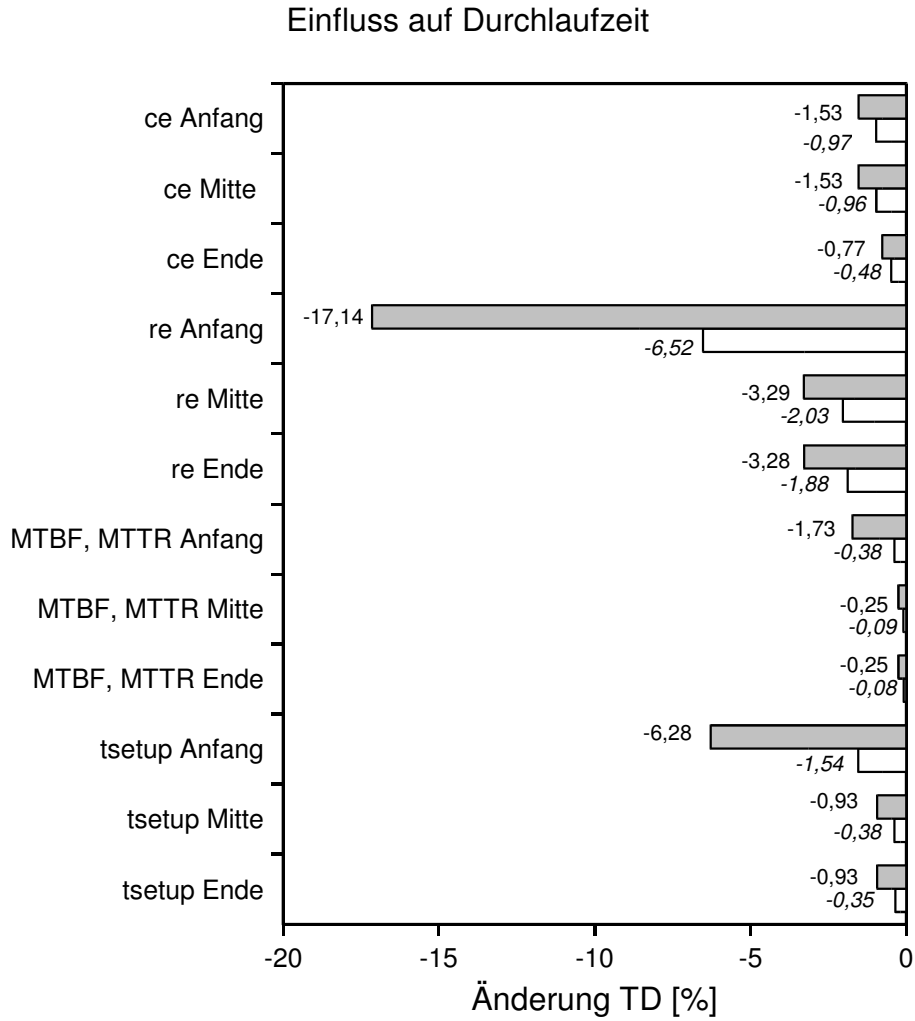


Abbildung 7.20: Einfluss auf Durchlaufzeit an unterschiedlichen Stellen im Prozessfluss - Zusammenfassung

Wie in Kapitel 7.3.3 soll hier auch noch dargestellt werden, welcher Einfluss sich auf die Zahl der Wafer Starts durch die Maßnahmen an unterschiedlichen Stellen im Prozessfluss ergibt (Abbildung 7.21). Analog zu Abbildung 7.20 ergeben sich natürlich auch hier die größten Auswirkungen beim ersten Prozessschritt.

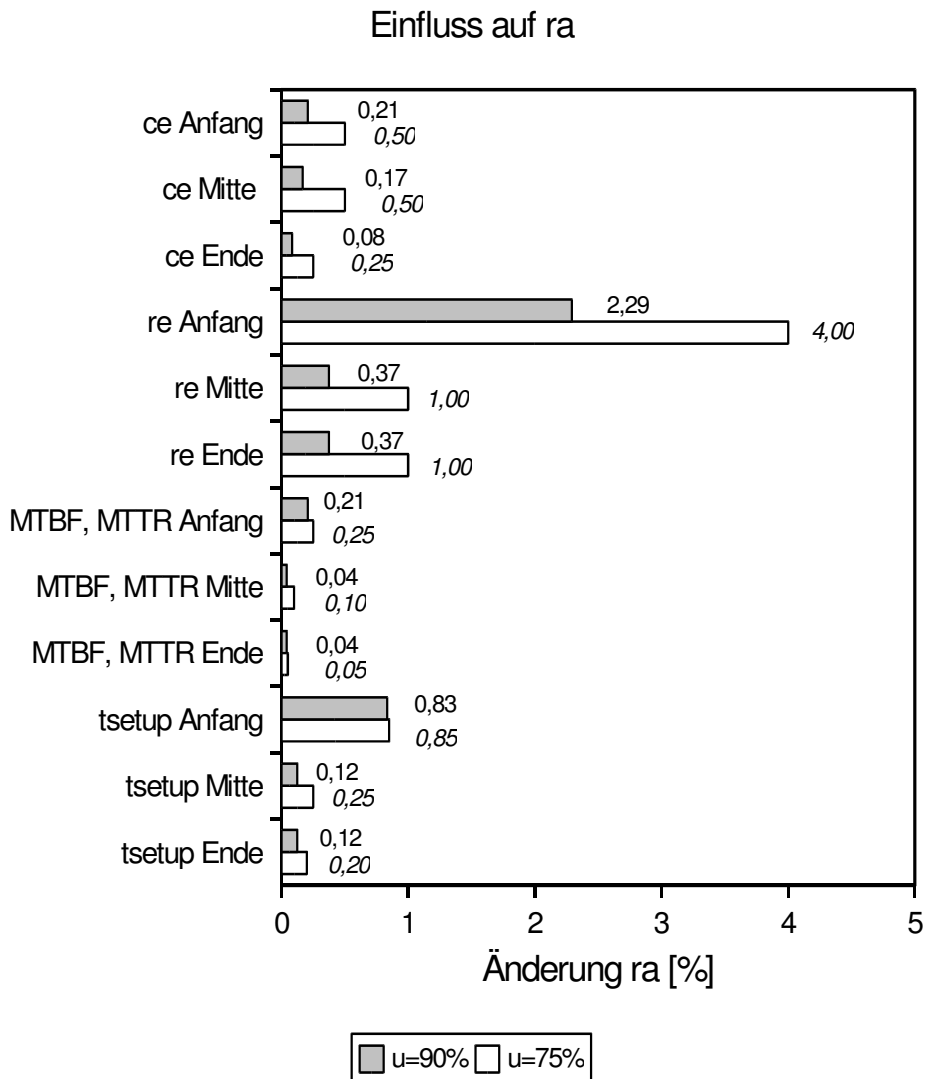


Abbildung 7.21: Einfluss der Maßnahmen an unterschiedlichen Stellen im Prozessfluss auf Produktionsmenge

7.3.5 Einfluss anderer produktionstechnischer Parameter

Zusätzlich zu den bereits angegebenen Parametern soll hier noch der Einfluss der Losgröße betrachtet werden.

Die Durchlaufzeit ist für eine Losgröße von 25 Wafern, 12 Wafern und 1 Wafer in Tabelle 7.12 und Abbildung 7.22 dargestellt.

ra [W/h]	L=25W		L=12W		L=1W	
	TD [h]	TD [h]	Änderung [%]	Änderung [%]	TD [h]	Änderung [%]
2	12,09	7,03	-41,81		3,13	-74,07
4	13,76	8,02	-41,68		3,70	-73,11
6	15,33	8,93	-41,75		5,30	-65,46
8	16,70	9,70	-41,95		-	-
10	17,86	10,34	-42,11		-	-
12	18,90	10,97	-41,95		-	-
14	20,02	11,77	-41,17		-	-
16	21,48	13,03	-39,35		-	-
18	23,72	15,29	-35,57		-	-
20	27,55	20,17	-26,78		-	-
22	35,13	36,99	5,30		-	-
24	55,76	-	-		-	-
26	288,24	-	-		-	-

Tabelle 7.12: Einfluss der Losgröße auf die Durchlaufzeit

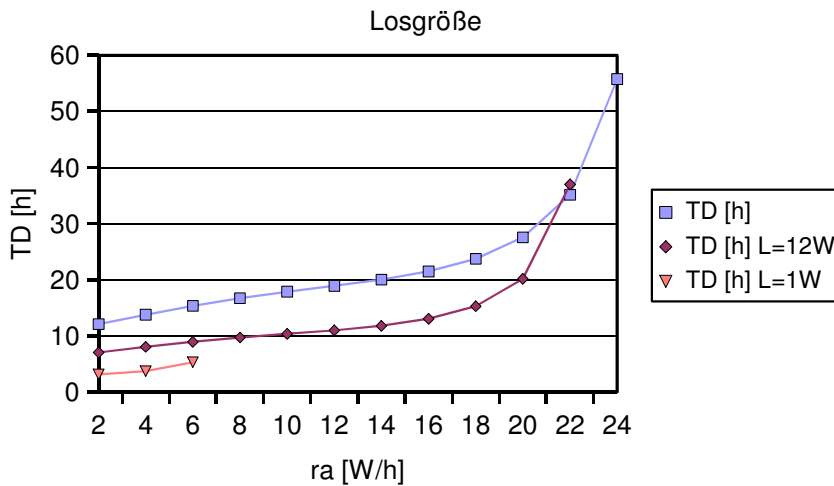


Abbildung 7.22: Einfluss der Losgröße auf die Durchlaufzeit

Entscheidenden Einfluss auf die Durchlaufzeit hat an dieser Stelle die Setupzeit, die bei jedem Los anfällt. Dies ist auch der Grund, warum bei der in diesem Beispiel angenommenen $t_{\text{setup}}=0,1\text{h}$ die Durchlaufzeit bei geringerer Losgröße zwar eindeutig kleiner wird, aber die Fertigungskapazität insgesamt sinkt. Wenn man im Modell eine niedrigere Setupzeit ansetzen würde, würde zwar die Fertigungskapazität in allen Fällen ansteigen, wäre aber qualitativ unverändert bei einer Losgröße von 12W und 1W niedriger als bei 25W.

Änderungen im Bereich der Losgröße würden zwar in gewissen Bereichen zu einer Verringerung von Durchlaufzeit und WIP führen, würden aber die gesamte Fab-Organisation und -Logistik betreffen, so dass sie nur mit großem Aufwand umgesetzt werden könnten. Aufgrund dieser technischen Probleme ist eine Verringerung der Losgröße in naher Zukunft nicht vorgesehen.

Zusätzlich soll noch untersucht werden, welche Veränderungen sich durch eine Reduktion der Zahl der Prozessschritte ergeben würden. In Tabelle 7.13 und Abbildung 7.23 wird die resultierende Durchlaufzeit bei 10, 9 und 8 Prozessschritten dargestellt. Wie zu erwarten,

ergibt sich bei einer Verringerung der Zahl der Prozesse eine eindeutige Verkürzung der Durchlaufzeit. Im allgemeinen werden sich bei einer eingefahrenen Fertigung jedoch keine Veränderungen bei den Prozessen ergeben, solche Veränderungen sind meist nur denkbar bei neuen Produkten oder Produktstrukturen.

ra [W/h]	10 Prozesse		9 Prozesse		8 Prozesse	
	TD [h]	TD [h]	Änderung [%]	TD [h]	Änderung [%]	
2	12,09	10,88	-9,97	9,67	-19,95	
4	13,76	12,41	-9,79	11,06	-19,62	
6	15,33	13,89	-9,40	12,43	-18,91	
8	16,70	15,23	-8,83	13,72	-17,85	
10	17,86	16,40	-8,18	14,90	-16,60	
12	18,90	17,47	-7,56	16,00	-15,37	
14	20,02	18,60	-7,08	17,14	-14,36	
16	21,48	20,03	-6,78	18,54	-13,68	
18	23,72	22,16	-6,61	20,58	-13,27	
20	27,55	25,76	-6,50	23,97	-13,01	
22	35,13	32,90	-6,36	30,67	-12,72	
24	55,76	52,34	-6,14	48,92	-12,27	
26	288,24	271,53	-5,80	254,81	-11,60	

Tabelle 7.13: Einfluss der Zahl der Prozessschritte auf die Durchlaufzeit

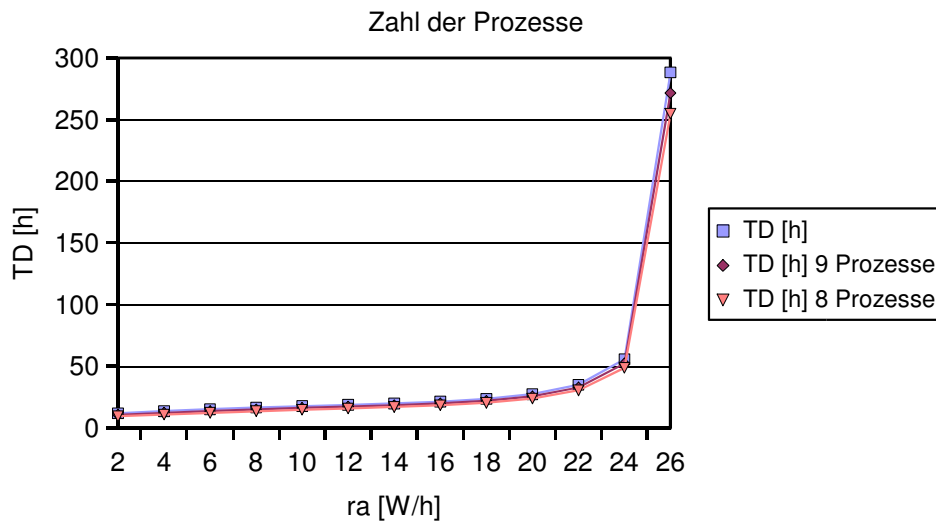


Abbildung 7.23: Einfluss der Zahl der Prozessschritte auf die Durchlaufzeit

Im Folgenden soll betrachtet werden, welche Auswirkungen durch Phasen mit hoher bzw. niedriger Beschäftigung in der Fertigung entstehen. Dazu wird angenommen, dass in Zeiten mit niedriger Beschäftigung eine Waferankunftsrate von 10W/h, entsprechend einem Nutzungsgrad von ca. 40%, und in Phasen mit hoher Beschäftigung eine Waferankunftsrate von 24W/h, entsprechend einem Nutzungsgrad von 90% erreicht wird.

ra [W/h]	Nutzungsgrad	TD [h]
1	0,04	11,26
2	0,08	12,09
3	0,11	12,93
4	0,15	13,76
5	0,19	14,57
6	0,23	15,33
7	0,27	16,05
8	0,3	16,7
9	0,34	17,31
10	0,38	17,86
11	0,42	18,39
12	0,45	18,9
13	0,49	19,44
14	0,53	20,02
15	0,57	20,68
16	0,61	21,48
17	0,64	22,47
18	0,68	23,72
19	0,72	25,36
20	0,76	27,55
21	0,8	30,62
22	0,83	35,13
23	0,87	42,38
24	0,91	55,76
25	0,95	88,5
26	0,99	288,24

Tabelle 7.14: TD und Nutzungsgrad für unterschiedliche Waferankunftsrate

Man kann in Tabelle 7.14 erkennen, dass die Durchlaufzeit bei einem r_a von 10W/h mit 17,86h relativ niedrig liegt, bei einem r_a von 24W/h dann aber bei 55,76h liegt. Für den Fall, dass die beiden Phasen mit hoher und niedriger Beschäftigung gleichmäßig verteilt sind, wäre in diese Fall eine ausgeglichene Beschäftigung mit $r_a=17W/h=const.$ und einer daraus resultierenden Durchlaufzeit von 22,47h günstiger, was die mittlere Durchlaufzeit betrifft (Abbildung 7.24).

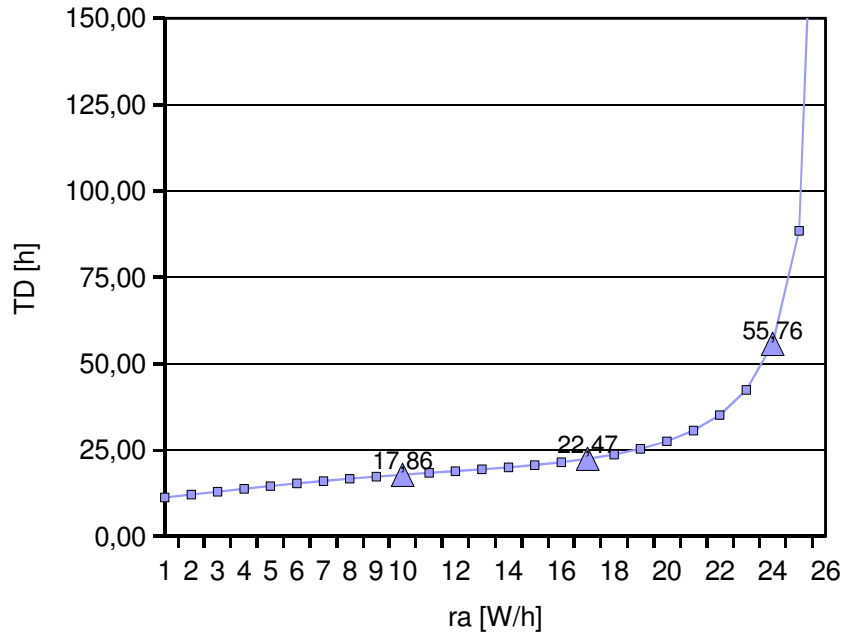


Abbildung 7.24: TD bei unterschiedlichem Beschäftigungsgrad

Zusammenfassend kann gesagt werden, dass durch Veränderungen der produktionstechnischen Faktoren noch große Verbesserungen erzielt werden können. Die Steigerung des Gerätedurchsatzes und die Verringerung der Variabilität stellen zwei Punkte dar, mit denen mit weniger Zusatzaufwand große Verbesserungen erzielt werden können.

Wie in Kapitel 6.2 bereits erläutert wurde, können die Verbesserungen prinzipiell dazu benutzt werden, um bei gleichbleibender Durchlaufzeit die Produktionsmenge zu erhöhen, oder bei gleichbleibender Produktionsmenge die Durchlaufzeit zu senken. Während sich die erste Variante besonders bei Massenprodukten anbietet, bei denen eine möglichst hohe Ausstoßmenge erzielt werden soll, liegt bei der ASIC-Fertigung ein Schwerpunkt auf der Verringerung der Durchlaufzeit.

Aufgrund der Nichtlinearität der Zusammenhänge, die in der Betriebskennlinie deutlich wird, sind die jeweiligen Verbesserungen stark vom Arbeitspunkt abhängig, so dass die Steigerungsmöglichkeiten nicht allgemein angegeben werden können.

7.3.6 Auswirkungen der Maßnahmen auf die Kosten

Gemäß Tabelle 4.1 ergeben sich für die Fertigung mit 200 mm-Wafern bei 20000 WSPM unter Annahme einer Investitionssumme von 1 Mrd. € und linearer Abschreibung der Geräte in 5 Jahren und des Gebäudes in 10 Jahren die in Tabelle 7.15 dargestellten Gesamtkosten.

Tabelle 7.15: Kosten für 200 mm-Fertigung

Geräteabschreibungen:	140,0 Mio. € p.a.
Personal direkt	23,5 Mio. € p.a.
Personal indirekt	33,0 Mio. € p.a.
Gebäudeabschreibungen	30,0 Mio. € p.a.
<i>fixe Kosten</i>	<i>226,5 Mio. € p.a.</i>

Geräteunterhalt	38,0 Mio. € p.a.
Wafer	27,0 Mio. € p.a.
Verbrauchsmaterialien	20,5 Mio. € p.a.
Kontrollwafer	3,0 Mio. € p.a.
<i>variable Kosten</i>	<i>88,5 Mio. € p.a.</i>
<i>Gesamtkosten</i>	<i>315,0 Mio. € p.a.</i>

Bei Annahme einer linearen Abhängigkeit der variablen Kosten von der Produktionsmenge berechnen sich die monatlichen Gesamtkosten zu:

$$K_G = 18,875 \text{ Mio.€} + 369 \text{ €/WSPM} \quad (7.13)$$

Bei Maßnahmen zur Steigerung der Effizienz der Fertigung muss beachtet werden, dass bei Erhöhung der Produktionsmenge durch Steigerung der Zahl der Waferstarts auch die variablen Kosten der Fertigung und damit die Gesamtkosten ansteigen. Bei Shrinkverfahren steigen die variablen Kosten dagegen nicht an, wenn sich die Zahl der Waferstarts nicht ändert.

Um die notwendige Reduzierung der Strukturgröße durchführen zu können, müssen laufend alte Geräte für die Chipfertigung gegen neue ausgetauscht werden. Die Kosten für den Aufbau und die Ausstattung einer Halbleiterfertigung steigen weiter an. Aufgrund der momentanen Situation kann von einem Anstieg um ca. 14% pro Jahr ausgegangen werden [137]. Dabei ist anzumerken, dass sich die Anstiegsrate der Kosten, die zwischenzeitlich bei 25% p.a. lag, in den letzten Jahren verlangsamt hat [135].

Die steigenden Kosten werden zum großen Teil durch den Anstieg der Gerätekosten bestimmt. Da während der Betriebsdauer einer Fabrik laufend ein Teil der Geräte durch neue ersetzt wird, wirkt sich dieser Kostenanstieg über die steigenden Geräteabschreibungen, die ungefähr 40% der Gesamtkosten ausmachen, auch auf die Betriebskosten aus. Wenn man annimmt, dass die übrigen Kostenfaktoren aus Tabelle 7.14 nur geringfügig ansteigen, bzw. die Kosten pro cm² Si-Wafer historisch gesehen pro Jahr leicht sinken [138], würde sich damit ein Anstieg der Gesamtkosten um ca. 5 - 6% pro Jahr ergeben.

Für die kommenden Jahre wird bei den großen IC-Herstellern der weitere Umstieg von 200 mm auf 300 mm Wafer in der Produktion geplant. Die notwendigen Investitionen für den Bau einer 300 mm Fab liegen höher als bei 200 mm Fabs. Momentan werden Kosten von 1,5 bis 2 Mrd. \$ für die ersten Produktionslinien angegeben [14]. Gemäß Tabelle 4.1 ergeben sich für die Fertigung mit 300 mm Wafern bei 20000 WSPM unter Annahme einer Investitionssumme von 1,5 Mrd. € und linearer Abschreibung der Geräte in 5 Jahren und des Gebäudes in 10 Jahren die in Tabelle 7.16 dargestellten Gesamtkosten.

Tabelle 7.16: Kosten für 300 mm-Fertigung

Geräteabschreibungen	210,0 Mio. € p.a.
Personal direkt	27,5 Mio. € p.a.
Personal indirekt	35,0 Mio. € p.a.
Gebäudeabschreibungen	42,5 Mio. e p.a.
<i>fixe Kosten</i>	<i>315,0 Mio. € p.a.</i>

Geräteunterhalt	75,0 Mio. € p.a.
Wafer	62,5 Mio. € p.a.
Verbrauchsmaterialien	30,0 Mio. € p.a.
Kontrollwafer	17,5 Mio. € p.a.
<i>variable Kosten</i>	<i>185,0 Mio. e p.a.</i>
<i>Gesamtkosten</i>	<i>500,0 Mio. € p.a.</i>

Die Investitionen und die daraus resultierenden Gesamtkosten liegen also um einiges höher als bei Fabs der 200 mm Wafergeneration. Den steigenden Kosten gegenüber steht jedoch ein starker Anstieg der Leistung der Fertigung. Aufgrund des größeren Waferdurchmessers können ungefähr um Faktor 2,3 - 2,4 mehr Chips pro Si-Scheibe produziert werden, wobei Faktor 2,25 durch den reinen Flächengewinn zustande kommt und sich ein weiterer geringerer Vorteil durch die geringere Randkrümmung des größeren Wafers ergibt. Die hohe Empfindlichkeit der Zahl der Chips gegenüber Änderungen des Waferdurchmessers wird auch in Ausdruck (7.11) deutlich.

7.3.7 Auswirkungen der Maßnahmen auf die Effizienz

Für die in Kapitel 7.3.3 untersuchten Maßnahmen Variabilität, Gerätezuverlässigkeit und Setupzeit wird angenommen, dass sich keine zusätzlichen Investitionen ergeben. Gemäß der Steigerung der Produktionsmenge lt. Abbildung 7.15 steigen dann nur die variablen Kosten in Abhängigkeit der mehr produzierten Menge und die Änderung der Effizienz berechnet sich nach Gleichung 4.24 und 4.21. Für die Steigerung des Gerätedurchsatzes wird angenommen, dass der Preis für die neuen Geräte um 15% ansteigt.

In Abbildung 7.25 wird der Einfluss auf die Effizienz jeweils bei einer Änderung der Parameter um 20% dargestellt. Die größten Auswirkungen ergeben sich durch eine Steigerung des Gerätedurchsatzes gefolgt von einer Verbesserung der Variabilität bei der Prozessankunft und -durchführung.

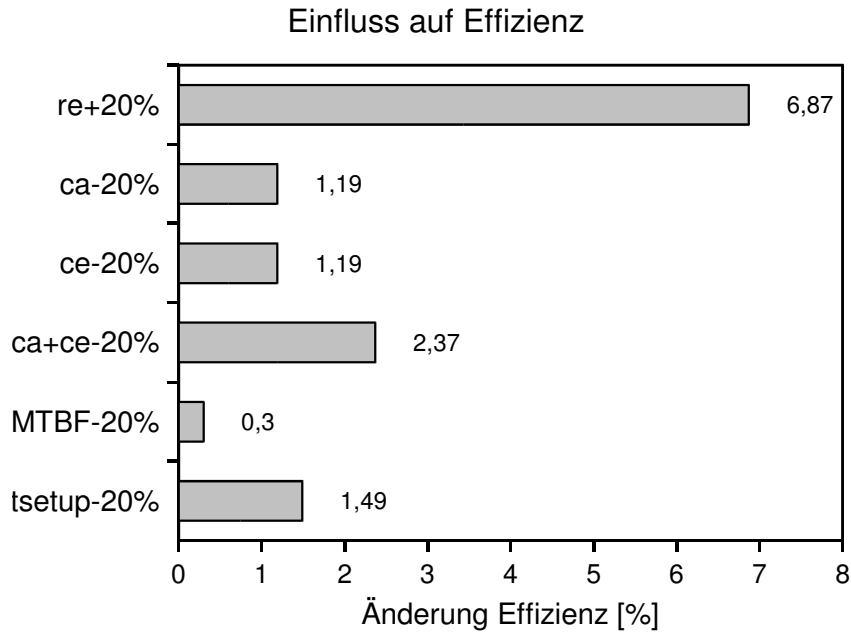


Abbildung 7.25: Einfluss der Maßnahmen auf die Effizienz

Der Einfluss der weiteren Faktoren auf die Effizienz wird in Abbildung 7.26 zusammenfassend dargestellt. Dabei wurden für den Umstieg auf 300mm Wafer die zusätzlichen Kosten aus Tabelle 7.16 angenommen. Für den Shrink wurde vorausgesetzt, dass pro Jahr 25% der Geräte mit jeweils 15% höherem Preis neu angeschafft werden müssen. Bei der verringerten Anzahl der Prozessschritte wurde fiktiv angenommen, dass sich ebenfalls die Zahl der Geräte um 10 bzw. 20% reduziert und damit die entsprechenden Investitionen.

Es zeigt sich, dass die größte Effizienzsteigerung mit ca. 50% durch die Einführung der 300mm Wafer entsteht, gefolgt vom Shrink der Chipfläche um 30%.

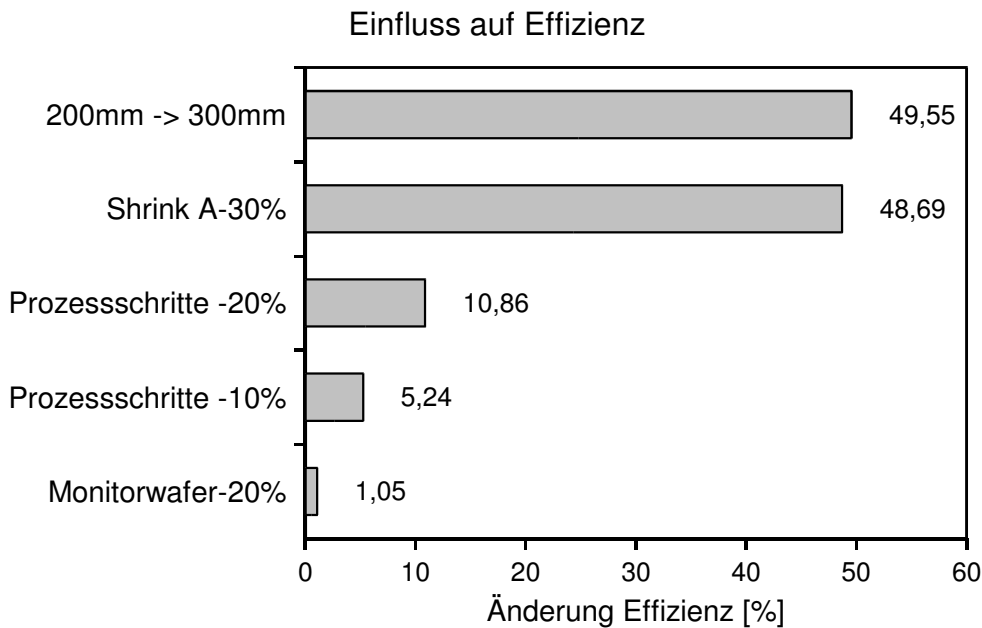


Abbildung 7.26: Einfluss zusätzlicher Maßnahmen auf die Effizienz

Unter Zugrundelegung der Daten aus Tabelle 7.15 und Tabelle 7.16 ergibt sich eine Steigerung der Effizienz bei 300 mm Produktion um ungefähr 50% gegenüber 200 mm Fertigung bzw. eine Verringerung der Kosten pro Stück um ungefähr 35%, was mit anderen Veröffentlichungen zu diesem Thema übereinstimmt (z.B. [66], [139]).

Die Steigerung des Waferdurchmessers von 200 mm auf 300 mm ist eine Methode, die jetzt und in den kommenden Jahren zur Steigerung der Effizienz angewandt wird. Aufgrund des hohen Aufwands, der dabei entsteht, wird der Umstieg auf die nächste Wafergeneration, dann vermutlich mit 450 mm Durchmesser, frühestens erst ab dem Jahr 2012 erwartet [12].

7.3.8 Verbesserungspotential der einzelnen Faktoren

Das Verbesserungspotential im Bereich der Ausbeute und der Zahl der Monitorwafer wird aufgrund der hohen bereits jetzt erreichten Werte bei zukünftigen Generationen sinken, d.h., dass die Effizienzsteigerungen, die bisher in diesen Bereichen erzielt worden sind, in Zukunft zu einem gewissen Teil mit anderen Maßnahmen erreicht werden müssen. Shrinkverfahren bilden weiterhin, solange Moore's Law bestehen bleibt, die wichtigste Maßnahme zur Effizienzsteigerung während der Produktlebensdauer.

Bei der Gerätezuverlässigkeit sind zwar schon relativ gute Werte erreicht, jedoch wirkt sich jede Unterbrechung und besonders ungeplante und überraschende Breakdowns gerade bei hohen Produktionsmengen und hohem Nutzungsgrad stark auf die Fab-Performance aus.

Bezüglich der Reduktion von Setupzeiten besteht noch Verbesserungspotential. Dieser Bereich muss auch zusammen mit den Themen Operator-Verfügbarkeit und Automatisierungsgrad der Fab betrachtet werden.

Eine Reduktion der Zahl der Prozessschritte bietet tatsächlich großes Verbesserungspotential, ist aber abhängig von neuen Prozessen oder neuen Produktstrukturen. Bei einer eingefahrenen Produktion stellt sich eine Veränderung bzw. Reduzierung der Prozessschritte als eher unrealistisch dar.

Im Hinblick auf die steigenden Waferdurchmesser und den damit steigenden Wert eines Wafers und eines Transportloses könnte auch der Verringerung der Losgröße höhere Bedeutung zukommen. Besonders für den Bereich der ASIC-Fertigung würde sich der Vorteil kürzerer Durchlaufzeiten verbunden mit einer Reduktion des gebundenen Kapitals und einem geringeren finanziellen Verlust bei Beschädigung oder fehlerhafter Prozessierung eines Loses vorteilhaft auswirken. Voraussetzung wäre allerdings eine entsprechende Handhabung mit niedrigen Setupzeiten.

7.3.9 Maßnahmen zur Verringerung von Schwankungen im Prozessfluss

Wie in den vorhergehenden Kapiteln beschrieben, kann sich erhöhte Variabilität im Prozessablauf vor allem auf die Durchlaufzeit der Fab auswirken, bzw. eine verringerte Kapazität und ggf. erhöhten Gerätebestand zur Folge haben. Ebenfalls können Schwankungen in der Prozessdurchführung zu verringerter Ausbeute führen. Es ist also offensichtlich, dass im Sinne einer guten Fab-Performance die Variabilität an allen Stellen möglichst reduziert werden sollte.

Vor einiger Zeit wurden im Projekt „Measurement and Improvement of Manufacturing Capacity (MIMAC)“ die folgenden 5 Faktoren als Haupteinflussgrößen auf die Performance beschrieben [170], [171], [172]:

- Equipment Downtime
- Setup
- Dispatch
- Yield Loss
- Batching

Als weitere Faktoren wurden angegeben:

- Alternative Tools
- End of Shift Effect
- Hot Lots
- Operator Skills
- Order Release / WIP Limits
- Redundant Tools
- Rework
- Blocking
- Factory Shutdown
- Lot Sizes
- Operator Availability
- Product Mix
- Re-Entrant Flow
- Time Bound Sequences

Praktisch alle dieser Faktoren sind am Entstehen von Variabilität beteiligt oder können, wie z.B. Ausbeuteverluste, durch Variabilität verursacht werden.

In [165] werden ebenfalls Produktmix und Nutzungsgrad als Einflussfaktoren auf die Variabilität beschrieben.

Die Verringerung der Variabilität ist also ein Thema, das sowohl bei der Fabplanung und -organisation als auch bei jedem einzelnen Prozessschritt betrachtet werden muss. Ein vollkommenes Vermeiden von Variabilität ist zwar grundsätzlich nicht möglich, durch einen regelmäßigen Ablauf und genaue Kontrolle der Fertigung sollte es aber gelingen, größere Schwankungen zu vermeiden.

7.3.10 Gültigkeitsbereich der Betrachtungen

Für die Betrachtungen wurden Daten aus einer realen Gerätegruppe zugrunde gelegt und als Eingabeparameter im Modell verwendet. Insgesamt sind solche Daten wegen der komplizierten Prozessflüsse in der Halbleiterindustrie und beispielsweise wegen mehrfach benutzten Geräten sehr schwer aufzunehmen und auszuwerten. Wie in Kapitel 6 dargestellt wurde, besteht auch eine gewisse Freiheit bei der Auswahl einiger Modellparameter, da nicht alle Daten eindeutig bestimmt sind.

Ziel war es nicht, die Prozesse im Detail zu modellieren, sondern diese sollten von einer höheren Ebene betrachtet werden. Insofern enthält das Warteschlangenmodell nicht detaillierte Prozess-Parameter, sondern bezieht sich auf die von außen zugänglichen Daten wie Gerätedurchsatz oder Variabilität der Prozesse für einen Ausschnitt des Gesamtprozesses.

Der Vergleich mit verschiedenen mathematischen Verteilungen in Kapitel 6 zeigt, dass die Modelle der Warteschlangentheorie für die Ankunft und Bearbeitung der Lose gut mit der Wirklichkeit der erhaltenen Gerätedaten übereinstimmen. Ebenfalls stimmen die erhaltenen Daten mit anderen Veröffentlichungen zu diesem Thema überein.

Grundsätzlich muss gesagt werden, dass sich eine Effizienz-Optimierung auf alle Ebenen des Unternehmens bezieht und nicht nur eine reine Prozess-Optimierung darstellt. Wie in Kapitel 1.1 abgegrenzt, werden hier nur Parameter aus Teilbereichen betrachtet.

7.4 Verbesserungspotential beim Ramp Up

Die Entwicklung der Zahl der produzierten und funktionsfähigen Chips beim Ramp Up ist in Abbildung 7.27 dargestellt. Hauptsächlich von Bedeutung beim Ramp Up sind das Anfangsniveau und die Zeitdauer bis zum Erreichen des angestrebten Endniveaus bei der Zahl der Waferstarts und der Ausbeute. Die (nicht erreichbare) Idealvorstellung wäre, die Produktion sofort mit der gewünschten Anzahl von Waferstarts und der Endausbeute beginnen zu können. Durch die langsame Steigerung der Produktionsmenge ergibt sich jedoch am Anfang eine geringere Produktionsmenge als bei idealem Ramp Up.

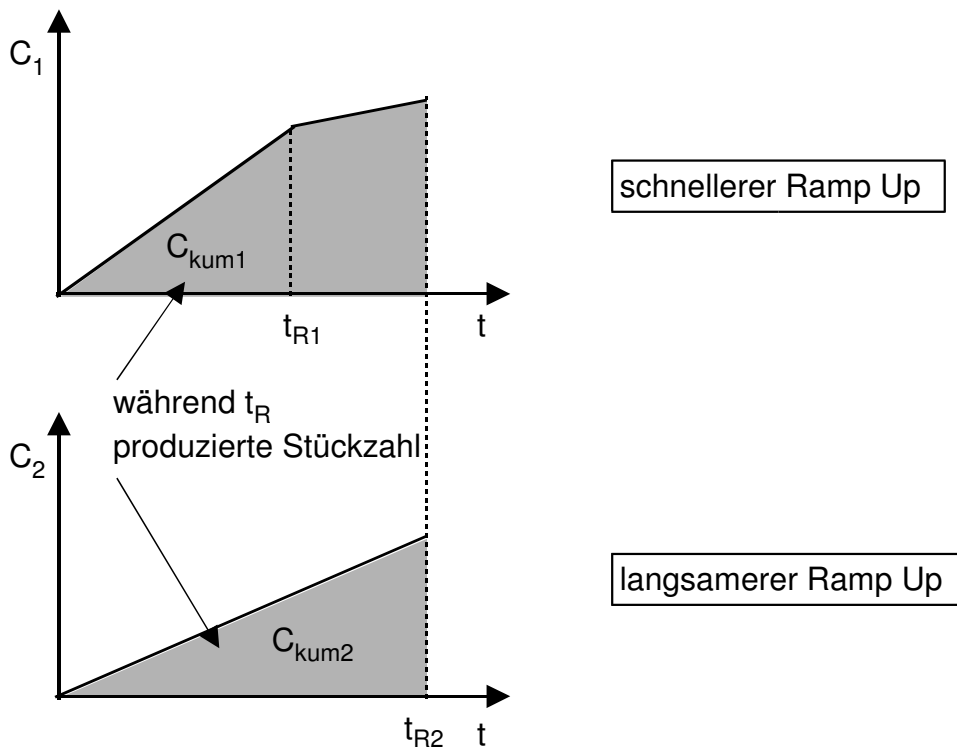


Abbildung 7.27: Entwicklung der produzierten Stückzahl beim Ramp Up

Im folgenden soll festgestellt werden, welche Verbesserungsmöglichkeiten sich bei Verkürzung der Zeitdauer des Ramp Up ergeben. Zu diesem Zweck werden jeweils zwei Szenarien miteinander verglichen und die prozentuale Veränderung gegenüber dem Ausgangswert angegeben.

Während des Ramp Up wird folgende Menge an funktionsfähigen Chips insgesamt produziert:

$$C_{\text{kum}} = \int_0^{t_R} C(t) dt \quad (7.14)$$

C_{kum} = kumulierte Anzahl der produzierten Chips während Ramp Up

t_R = Zeitdauer des Ramp Up

Für eine bessere Vergleichbarkeit der Daten werden die Ergebnisse C_{kum1} und C_{kum2} jeweils auf die Zeitdauer des längeren Ramp Up t_{R2} bezogen, wie dies in Abbildung 7.27 dargestellt ist. Das bedeutet, dass im schnelleren Fall der Ramp Up bereits abgeschlos-

sen ist und sich die Fabrik im eingefahrenen Zustand befindet, bei dem nur noch leichte Steigerungen im Lauf der Zeit möglich sind.

Der relative Ausfall, der durch die längere Zeitdauer des Ramp Up entsteht, läßt sich berechnen durch:

$$A_C = \frac{\int_0^{t_{R2}} C_1(t) dt - \int_0^{t_{R2}} C_2(t) dt}{\int_0^{t_{R2}} C_1(t) dt} = \frac{C_{\text{kum1}} - C_{\text{kum2}}}{C_{\text{kum1}}} \quad (7.15)$$

A_C = prozentualer Ausfall an produzierten Chips

Dabei muss allerdings beachtet werden, dass aufgrund des Preisverfalls für die Chips am Anfang der Produktlebensdauer ein höherer Preis erzielt werden kann als am Ende. Das bedeutet, dass der Umsatzausfall höher liegt als der reine Ausfall an Produktionsmenge.

Der kumulierte Umsatz berechnet sich durch Integration des Produkts von Produktionsmenge und zeitabhängigem Preis

$$U_{\text{kum}} = \int_0^{t_R} C(t) \cdot P_0 \cdot p^t dt \quad (7.16)$$

U_{kum} = kumulierter Umsatz während Ramp Up

Zur Berechnung des relativen kumulierten Umsatzausfalls werden wieder die Werte U_{kum1} und U_{kum2} miteinander verglichen.

$$A_U = \frac{\int_0^{t_{R2}} C_1(t) \cdot P_0 \cdot p^t dt - \int_0^{t_{R2}} C_2(t) \cdot P_0 \cdot p^t dt}{\int_0^{t_{R2}} C_1(t) \cdot P_0 \cdot p^t dt} = \frac{U_{\text{kum1}} - U_{\text{kum2}}}{U_{\text{kum1}}} \quad (7.17)$$

A_U = prozentualer Umsatzausfall

Wie in (7.17) deutlich wird, hat der Anfangspreis keinen Einfluss auf A_U , sondern nur die zeitliche Veränderung des Preises. Dagegen ist der kumulierte relative Umsatzausfall nicht gleich dem kumulierten relativen Produktionsausfall.

Für das in Kapitel 7.2.2 dargestellte Szenario ergibt sich, dass während einer 8-monatigen Ramp Up-Phase insgesamt ca. 14 Mio. Chips hergestellt werden. In Tabelle 7.16 wird angegeben, welche relativen Veränderungen bezüglich Produktionsmenge und Umsatz sich ergeben, wenn die Zeitdauer des Ramp Up unter sonst gleichen Verhältnissen auf 5, 6 oder 7 Monate verkürzt wird. Bei der Umsatzveränderung wird ein Preisverfall von 32% pro Jahr angenommen.

Tabelle 7.17: Veränderungen bei Verkürzung des Ramp Up

Zeitdauer des Ramp Up [Monate]	Veränderung der Zahl der Chips [%]	Veränderung des Umsatzes [%]
8	Referenzwert	Referenzwert
7	+15,9	+16
6	+32,5	+33,1
5	+50,3	+51,6

Wie in Tabelle 7.17 deutlich wird, sind aufgrund des stetigen Preisverfalls die Auswirkungen auf den Umsatz etwas größer als auf die Produktionsmenge. Bei einer Verkürzung der Zeitdauer für Volumen- und Ausbeute-Ramp Up auf 7 Monate könnte während der ersten 8 Monate der Produktion ein um 16% höherer Umsatz erzielt werden. Bei einer Reduktion um 3 Monate würde sich die relative Steigerung bereits auf über 50% belaufen. In Abbildung 7.28 wird der relative Produktionsmenge bezogen auf 8 Monate graphisch dargestellt.

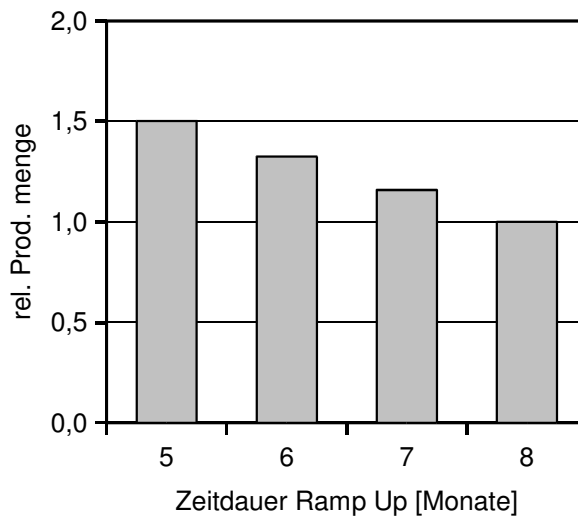


Abbildung 7.28: Relative Änderung der Produktionsmenge bei Verkürzung des Ramp Up

Bei einer Veränderung des Anfangsniveaus der Ausbeute ergeben sich, unter sonst gleichen Umständen und bei einer Ramp Up-Dauer von 8 Monaten, die in Tabelle 7.18 angegebenen Veränderungen bezogen auf eine Anfangsausbeute von 50%.

Tabelle 7.18 Veränderungen bei Änderung des Anfangsausbeute

Anfangsausbeute [%]	Veränderung der Zahl der Chips [%]	Veränderung des Umsatzes [%]
30	-6,1	-6,5
40	-3	-3,1
50	Referenzwert	Referenzwert
60	+3	+3,1
70	+6	+6,4

In Abbildung 7.29 sind die Daten aus Tabelle 7.18 graphisch gegenübergestellt. Die Empfindlichkeit gegenüber einer Änderung des Anfangsniveaus ist deutlich geringer als gegenüber einer Änderung der Zeitdauer. Das bedeutet, dass beim Ramp Up der Fertigung möglichst großer Wert auf Schnelligkeit und kurze Lernzyklen gelegt werden muss (siehe auch [167]).

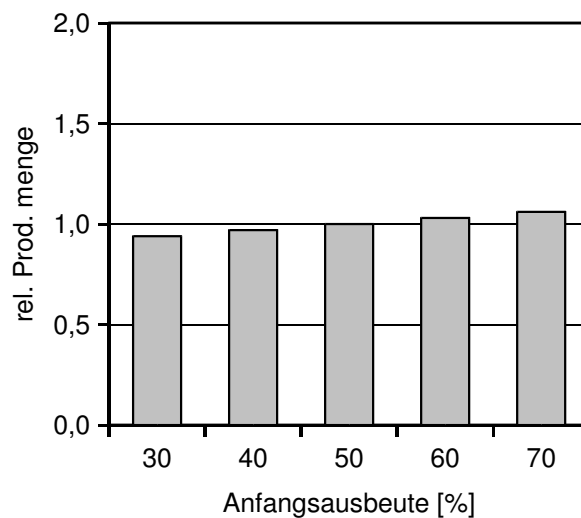


Abbildung 7.29: Relative Änderung der Produktionsmenge bei Änderung der Anfangsausbeute

7.5 Auswirkungen von Veränderungen vor und während der Produktion

Im folgenden soll ermittelt werden, welche Auswirkungen sich durch Veränderungen oder Abweichungen vom Sollwert vor bzw. während des Produktionsablaufs ergeben können.

Für die folgenden Betrachtungen werden für die entsprechenden Parameter die nachfolgenden Werte angenommen:

Investitionen: 1 Mrd. €

Entwicklungskosten: 0,5 Mrd. €

produzierte Chips pro Jahr: 50 Mio. Stück

Produktlebensdauer: 5 Jahre

Investitionen um 10% höher:

Eine 10%ige Erhöhung der Investitionssumme entspricht einer Steigerung um 100 Mio. €. Bei einer Abschreibungsdauer von 5 Jahren bedeutet dies einen zusätzlichen finanziellen Aufwand von 20 Mio. € p.a., was umgerechnet auf 50 Mio. Chips pro Jahr eine Erhöhung der Stückkosten um 0,4 € bedeutet. Da die Abschreibungen ungefähr 50% der Gesamtkosten betragen, würde eine Steigerung um 10% zu einer Erhöhung der Gesamtkosten und damit auch der Stückkosten um ca. 5% führen.

Bei einer Abschreibungsdauer von 3 Jahren würden sich sich beim dargestellten Szenario zusätzliche Kosten in Höhe von 33,5 Mio. € p.a. ergeben, was einer Steigerung der Stückkosten um 0,67 € entspricht.

Entwicklungskosten um 10% höher:

Bei angenommenen Entwicklungskosten von insgesamt 0,5 Mrd. € bedeutet eine Veränderung um 10% insgesamt 50 Mio. € mehr Entwicklungsaufwand, der auf die Produktlebensdauer umgelegt werden muss. Bei 5 Jahren Laufzeit ergeben sich damit 10 Mio. € p.a. zusätzliche Kosten, was bei 50 Mio. Chips p.a. 0,2 €/Chip ausmacht. Bei Investitionskosten von 1 Mrd. € und Entwicklungskosten von 0,5 Mrd. € werden dadurch die Stückkosten um ca. 2,5% erhöht.

Falls die Entwicklungskosten, wie oft im Speicherbereich üblich, auf mehrere Kooperationspartner aufgeteilt werden, verringert sich die entstehende Mehrbelastung. beispielsweise 3 Partnern reduzieren sich die zusätzlichen Stückkosten dann auf 0,07 €/Chip. Die Kosten pro Stück steigen ungefähr um 0,7%.

Ausbeute um 10% zu niedrig:

Bei einer um 10% verringerten Ausbeute verringert sich die Zahl der produzierten und funktionsfähigen Chips pro Jahr von 50 Mio. auf 45 Mio. Stück, wobei die Herstellungskosten allerdings konstant bleiben. Das bedeutet, dass sich die Gesamtkosten nur noch auf 90% der Chips verteilen und damit die Kosten pro Stück um 11% steigen.

Absatzprobleme, 10% der Chips können nicht verkauft werden:

Falls aufgrund von Absatzproblemen ein Teil der Produktion nicht verkauft werden kann, verringert sich zwar die Zahl der verkauften Chips, nicht jedoch die Herstellungskosten, da natürlich auch für die nicht verkäuflichen Chips Produktionskosten anfallen.

Für den Fall, dass 10% der Jahresproduktion nicht abgesetzt werden können, verringert sich in diesem Beispiel die Zahl der verkauften Chips von 50 Mio. auf 45 Mio.. Das bedeutet, dass sich pro verkauftem Chip die Stückkosten um 11% steigern. Falls der Preis pro Chip nicht erhöht werden kann, ergibt sich eine Verringerung des Umsatzes um 10%.

1 Monat nicht produzieren:

Bei einem Produktionsausfall von einem Monat verringert sich die jährliche Produktionsmenge um 1/12. Dabei muss allerdings beachtet werden, dass die variablen Kosten für ein Monat ebenfalls wegfallen. Bei einem Anteil der variablen Kosten von 30% an den Gesamtkosten bedeutet dies, dass sich die jährlichen Gesamtkosten um 2,5% reduzieren. Damit ergibt sich insgesamt eine Steigerung der Kosten pro Stück um ungefähr 6%.

Nicht enthalten in dieser Betrachtung ist allerdings der möglicherweise weitaus schlimmere Effekt, der sich ergibt, wenn Kunden aufgrund des Produktionsausfalls nicht bedient werden können und zur Konkurrenz abwandern.

In Abbildung 7.30 sind die Auswirkungen der oben erwähnten Abweichungen auf die Stückkosten graphisch dargestellt.

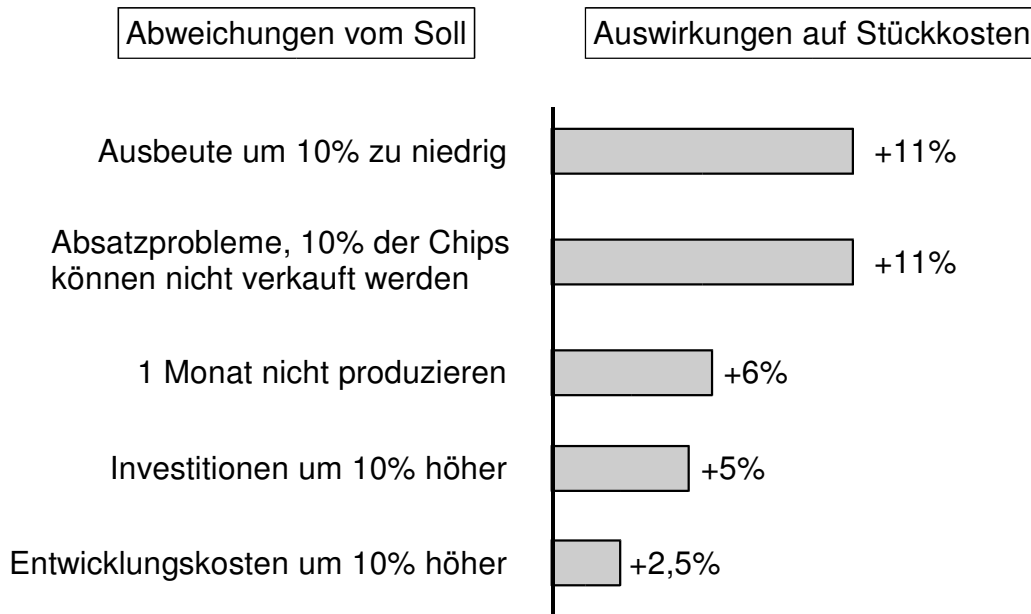


Abbildung 7.30: Auswirkungen von Abweichungen auf die Stückkosten

1 Monat zu spät am Markt:

Bei einem verspäteten Markteintritt kann bei einem angenommenen kontinuierlichen Preisverfall, der von den Firmen nicht beeinflusst werden kann, nur noch ein geringerer Erlös für die gleiche Zahl von Chips erzielt werden. Bei sonst gleichen Voraussetzungen ergeben sich keine Auswirkungen für die Stückkosten, der kumulierte Umsatz sinkt jedoch wegen der zeitlichen Verzögerung. Wenn man annimmt, dass die Firma zwar später in den Markt eintritt, dann das Produkt aber doch 5 Jahre lang produziert, ergibt sich bei einem angenommenen Preisverfall von 32% pro Jahr ein Umsatzausfall von ungefähr 3%, wenn die Produktion um einen Monat verspätet begonnen wird. Bei einer Verzögerung um 6 Monate, die bei einer Produktlebensdauer von 5 Jahren einer 10%igen Verzögerung entspricht, würde sich der Umsatzausfall auf rund 17,5% erhöhen.

Wenn man andernfalls annimmt, dass ein Produkt nur für eine begrenzte Zeit am Markt absetzbar ist, bedeutet dies, dass ein Unternehmen, das zu spät in den Markt eintritt, das Produkt nur für eine insgesamt kürzere Zeitdauer produzieren kann, wie dies in Abbildung 4.14 dargestellt ist. Nach Gleichung (4.46) würde ein solches Szenario bedeuten, dass bei einer zeitlichen Verzögerung von 1 Monat 5% des Marktpotentials nicht genutzt werden können. Bei einer Verzögerung um 6 Monate würde sich dieser Ausfall auf 28% belaufen (Abbildung 7.31).

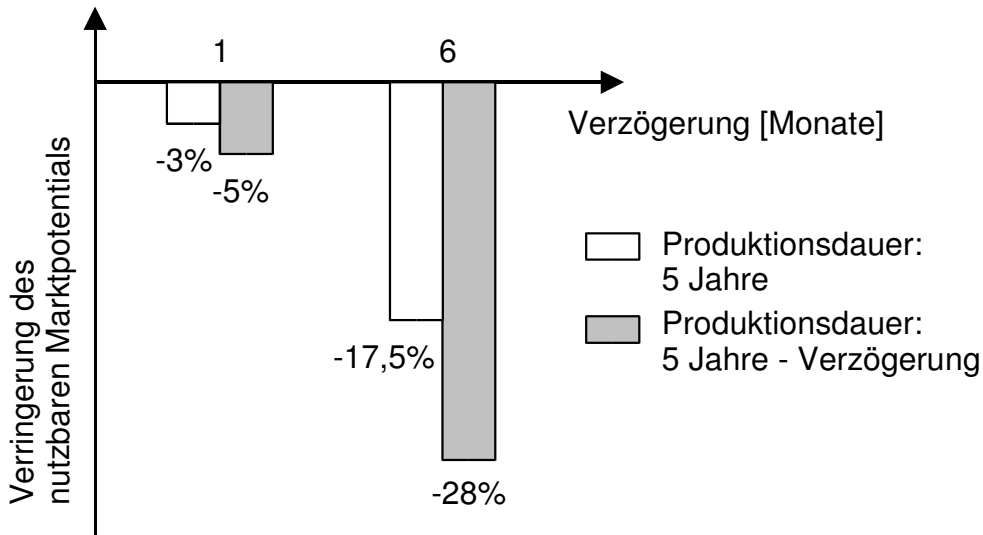


Abbildung 7.31: Nicht nutzbares Marktpotential durch Verzögerungen

7.6 Zusammenfassung: Technische Faktoren mit dem größten Verbesserungspotential

Die problematischen Rahmenbedingungen in der Halbleiterfertigung erfordern von den Unternehmen erhebliche Anstrengungen. Besonders der starke Preisverfall erzwingt ständige starke Verbesserungen. Ein Preisverfall von 32% p.a. erfordert eine ebenso große Verringerung der Stückkosten, bzw. eine Steigerung der Effizienz der Fertigung um 47% p.a.. Durch eine leichte Abflachung der Preiskurve auf beispielsweise 25% p.a. würden sich die Anforderungen an die Effizienzsteigerung auf 33% p.a. reduzieren. Dabei wird eine Absetzbarkeit der Produkte auf dem Markt vorausgesetzt. Für den Fall, dass mit einem Produkt keine Gewinne erzielt werden, sind noch stärkere Verbesserungen notwendig, um in die Gewinnzone zu gelangen.

Unter Berücksichtigung der Ergebnisse aus den oben dargestellten Kapiteln ergibt sich, dass die technischen Einflussgrößen aus Abbildung 7.26 und Abbildung 7.25 am stärksten zur Verbesserung der Produktions-Effizienz beitragen.

Dabei muss allerdings beachtet werden, dass die Erhöhung des Waferdurchmessers im Gegensatz zu den anderen Maßnahmen nicht während des Betriebs einer Fab durchgeführt werden kann.

Shrinkverfahren stellen die wichtigste Maßnahme zur Steigerung der Effizienz während der Produktlebensdauer dar. Je schneller die Strukturgröße verringert wird, desto größer sind die für das Unternehmen resultierenden Verbesserungen. Durch aggressive Shrinkstrategien können also Wettbewerbsvorteile gegenüber der Konkurrenz erreicht werden.

Ebenso bestehen noch große Verbesserungsmöglichkeiten bei der Steigerung der Gerätekapazitäten bzw. der besseren Ausnutzung der bestehenden Gerätekapazitäten und beim Prozessverständnis.

Großer Einfluss auf die Produktion entsteht durch Schwankungen im Fertigungsablauf. Wie sich gezeigt hat, besteht besonders bei der Variabilität der Ankunftsprozesse noch eindeutig Verbesserungspotential. Eine Verringerung der Variabilität würde insgesamt zu einer Verbesserung des Fabablaufs beitragen.

Die Parameter Defektdichte und Anteil der Monitorwafer stellen zwar wichtige Faktoren

dar, die zur Steigerung der Effizienz beitragen und an denen die Qualität einer Fertigung deutlich wird, sind aber bei gut beherrschten Fertigungslinien heute bereits auf einem Niveau angelangt, von dem aus kein großes Verbesserungspotential mehr besteht. Desweiteren bestehen bezüglich der Geräte-Downtime und der Transportzeiten noch Verbesserungsmöglichkeiten, die jedoch ebenfalls im Lauf der Zeit immer geringer werden.

Das größte Verbesserungspotential bezüglich einer Verkürzung der Durchlaufzeit bietet prinzipiell eine Verringerung der Losgröße, die jedoch, wie in Kapitel 7.3.5 dargestellt, große Umstellungen erfordern würde und beim gegenwärtigen Stand der Produktion nicht möglich ist.

Allgemein kann gesagt werden, dass durch eine allgemein optimierte Ablaufplanung und Logistik, die zu einer regelmäßigen Waferankunft an den Geräten und damit zu einer Vermeidung von Variabilität führt, die Wartezeiten vor den Geräten und damit die Gesamtdurchlaufzeit positiv beeinflusst werden können.

Insgesamt zeigt sich, dass sich zeitliche Verzögerungen äußerst negativ auf die Umsatz- und Gewinnerwartungen auswirken. Durch einen schnellen Markteintritt oder durch eine Beschleunigung des Ramp Up (siehe Kapitel 7.4) lassen sich sehr große Verbesserungen erzielen. Bei einer Verzögerung des Markteintritts (Time to Market) um 6 Monate und einer Produktlebensdauer von 5 Jahren können beispielsweise bis zu 28% des Marktpotentials nicht genutzt werden. Durch eine Beschleunigung des Ramp Up von 8 auf 5 Monate kann eine um 50% höhere Zahl der Chips bezogen auf diese 8 Monate erreicht werden. In diesem Zusammenhang spielen auch Lerneffekte eine entscheidende Rolle, die z.B. auch für eine Verringerung der Zahl der Prozessschritte genutzt werden können.

8 Ausblick

Im Laufe der Zeit haben sich die Herausforderungen und Schwerpunkte bei der Fertigung integrierter Schaltungen verändert. Während am Anfang der Schwerpunkt auf der technischen Machbarkeit und der Prozessoptimierung lag, spielt inzwischen auch die Optimierung der Fabrikabläufe und der Softwaresteuerung eine immer größere Rolle, wie dies in der unten angeführten Übersicht dargestellt ist (nach [99]).

- 60er Jahre: Schwerpunkt auf technischer Machbarkeit, manuelle Kontrolle, lange Servicezeiten für Geräte
- 70er Jahre: steigende technologische Anforderungen, zunehmende Automatisierung, steigende Kosten
- 80er Jahre: automatisierte Geräte werden Standard, Gerätepreis > 1 Mio. \$, steigende Bedeutung der Zuverlässigkeit, Einführung der statistischen Prozesskontrolle
- 90er Jahre: 200 mm Wafer, Gerätepreis 3-5 Mio. \$, beginnende Produktivitätsanalyse auf Fabrikebene
- 2000: 300 mm Wafer, hoher Automatisierungsgrad, bessere Prozesskontrolle

Aufgrund der Komplexität der Prozesse und der üblicherweise sehr konservativen Haltung der Chiphersteller, dass funktionierende Prozesse möglichst nicht geändert, bzw. möglichst unverändert übernommen werden sollten, wurden Änderungen früher nur relativ langsam in der Fertigung umgesetzt, so dass die Fertigungsabläufe noch nicht so optimiert sind wie in manchen anderen Industriebranchen [140].

Für die Zukunft geplant ist eine vollständige Automatisierung der Fab und die Steuerung und Integration der Fertigungsabläufe durch CIM-Konzepte auf allen Ebenen. Eine große Rolle werden dabei die MES-Systeme (Manufacturing Execution System) spielen, deren Aufgabe die Geräteüberwachung, Datenerfassung und -auswertung und zunehmend die Ablaufplanung und -steuerung umfasst (siehe z.B. [141], [142], [143]) und die in ein umfangreiches Softwarekonzept eingebunden sind, das von der Materialflußplanung bis zur Fabsteuerung mit ERP-Systemen (Enterprise Resource Planning) reicht. Dabei wird zunehmend die Fabrik als Ganzes gesehen und das Zusammenwirken der einzelnen Teilbereiche berücksichtigt.

Ebenso wird der Einsatz von Simulationswerkzeugen zur Fabplanung und die Integration verschiedener Simulationsebenen weiter intensiviert werden, um bereits in der Planungsphase verschiedene Szenarien durchspielen zu können und den Fabaufbau und -ablauf optimieren zu können.

Diese Maßnahmen sollten auch dazu beitragen, dass eine weitere Steigerung der Flexibilität und eine Reduzierung der Variabilität in der Mehrproduktfertigung erreicht wird. Besonders in der ASIC-Fertigung ist eine optimierte Ablaufplanung von großer Bedeutung, da viele Produkte gleichzeitig unter Einhaltung eines Datums für die Fertigstellung hergestellt werden müssen.

Der Weltmarkt für Halbleiter wird auch in Zukunft im Mittel ein sehr hohes Wachstum aufweisen.

Neue Produkte und steigender Elektronikanteil in bestehenden Systemen werden in den nächsten Jahren den Markt weiter vorantreiben [144]. Der Halbleitermarkt wird jedoch auch weiterhin starken zyklischen Schwankungen unterliegen, die durch einen Wechsel zwischen Überkapazität und ungesättigter Nachfrage entstehen und für die Unternehmen eine extreme Herausforderung darstellen. Wegen der hohen Kosten für Entwicklung und

Fertigung und wegen der starken Marktschwankungen, besonders im DRAM-Bereich, ist in den nächsten Jahren in der Halbleiterbranche mit Fusionen bzw. mit weiter intensivierten Kooperationen einiger Unternehmen zu rechnen.

Auch aufgrund der hohen Kosten für eine eigene Halbleiterfertigung sind momentan viele Firmen bestrebt, einen Teil ihrer Produktion auszulagern und von Foundries fertigen zu lassen. Besonders kritische Produkte und Prozesse, die die Kernkompetenz eines Unternehmens darstellen, können allerdings nur schwer nach außen gegeben werden.

Von großer Wichtigkeit für die Halbleiterindustrie wird es sein, in Zukunft die richtige Balance zwischen Investitionsbedarf, Innovationsgeschwindigkeit und Preisverfall zu finden. Die ständig steigenden Investitionen, vor allem für Geräte, machen es immer schwieriger, den nötigen Return on Investment zu erwirtschaften. Zudem machen Investitionskosten in Milliardenhöhe einen Einstieg für kleinere Unternehmen praktisch unmöglich. Eine weitere Abflachung des Kostenanstiegs würde dazu führen, dass die Anforderungen an die Effizienzsteigerungen gesenkt werden können. Eine hohe Innovationsgeschwindigkeit bedingt eine kurze Produktlebensdauer, da in kurzen Abständen neue Produkte mit besseren Eigenschaften auf den Markt kommen. Je kürzer die Produktlebensdauer, desto schwieriger ist es jedoch die Investitionen verzinst wieder zu erwirtschaften. Der Preisverfall stellt einen entscheidenden Faktor in der Halbleiterfertigung dar. Je größer der Preisverfall ist, desto attraktiver sind zwar die Produkte der Mikroelektronik für die Kunden, desto schwieriger ist es aber für die Unternehmen profitabel zu sein und umso mehr und schneller muss die Effizienz der Fertigung im Lauf der Zeit gesteigert werden.

Falls es in Zukunft nicht gelingen sollte, die Effizienz der Fertigung in der gleichen Weise wie bisher zu steigern, ist es denkbar, dass sich eine Verringerung der Innovationsgeschwindigkeit oder ein Abflachen der Preiskurve ergibt. Solche Konsequenzen wären übrigens auch denkbar, falls sich langfristig betrachtet ein Monopol einer Firma am Markt ergeben sollte. Momentan ist eine solche Entwicklung jedoch nicht in Sicht. Die Vorhersagen der „International Technology Roadmap for Semiconductors“ reichen inzwischen bis zum Jahr 2018. Bis dahin werden keine prinzipiellen Änderungen beim Aufbau des Transistors und bei den Herstellungsprozessen erwartet. Ebenso wird keine Änderung der Innovationsgeschwindigkeit vorhergesagt. Bis zum Jahr 2016 soll eine minimale Strukturweite von 22 nm erreicht werden, was einem 64 GBit DRAM-Speicher entsprechen soll.

Literaturverzeichnis

- [1] F.X. Bea, J. Haas, *Strategisches Management*, Verlag Fischer, 1995, Stuttgart
- [2] G. Moore, *Proceedings International Electron Devices Meeting*, S. 11-13, 1975
- [3] W. Kuo, T. Kim, „An Overview of Manufacturing Yield and Reliability Modeling for Semiconductor Products“, *Proceedings of the IEEE*, Vol. 87, No. 8, S. 1329 -1344, August 1999
- [4] *ICE Status 1999*, Integrated Circuit Engineering Corporation
- [5] G. Moore, „Cramming more components onto integrated circuits“, *Electronics*, Vol. 38, No. 8, April 19, 1965
- [6] *ICE Status 1998*, Integrated Circuit Engineering Corporation
- [7] A. Chatterjee, D. Gudmundsson, R.K. Nurani, S. Seshadri, J.G. Shanthikumar, „Fabless-Foundry Partnership: Models and Analysis of Coordination Issues“, *IEEE Transactions on Semiconductor Manufacturing*, Vol. 12, No. 1, S. 44-52, February 1999
- [8] K. Mautz, R. Bloom, „Plasma Etch Tool Selection Criteria for Minimizing Cost of Ownership“, *Proceedings Advanced Semiconductor Manufacturing Conference*, S. 29, 1995
- [9] „Der DRAM-Zyklus wird optimiert“, *Markt&Technik*, Nr. 51/52, 17.12.1999, S. 18
- [10] Heise Zeitschriften Verlag, www.heise.de
- [11] R. De Jule, „CMOS: A Hard Act to Follow“, *Semiconductor International*, December 1999
- [12] *International Technology Roadmap for Semiconductors*, Semiconductor Industry Association (SIA), 2003
- [13] P.O.Hahn, H.G. Fusstetter, „300mm Si-Wafer: Chance and Challenge“, *Proceedings Semiconductor Equipment and Materials beyond JESSI - 300mm and Single Wafer Processing*, S. 43-55, Productronica München, 1995
- [14] B. Arnold, „300 mm Standoff“, *Semiconductor Magazine*, Vol. 1, No. 2, February 2000
- [15] H.M. Grimes, „Economics and Materials Science of Large-Diameter Silicon Wafers“, *TI Technical Journal*, S. 5-12, September-October 1995
- [16] W. Murray Bullis, William C. O'Mara, „Large diameter silicon wafer trends“, *Solid State Technology*, S. 59-65, April 1993
- [17] I. Ruge, H. Mader, *Halbleiter-Technologie*, Springer-Verlag, 1991
- [18] G. Schumicki, P. Seegebrecht, *Prozeßtechnologie*, Springer-Verlag, 1991
- [19] A. Plettner, *Leiterbahnkonzept für Hochfrequente Integrierte Schaltungen*, Dissertation TU München, 1998
- [20] Y. Taur, „The incredible shrinking transistor“, *IEEE Spectrum*, S. 25 - 29, July 1999
- [21] L. Peters, „Who Will Gamble at 0.13 mm?“, *Semiconductor International*, January 2000
- [22] D. Widmann, H. Mader, H. Friedrich, *Technologie hochintegrierter Schaltungen*, Springer, 1996
- [23] H.B. Bakoglu, *Circuits, Interconnections, and Packaging for VLSI*, Addison-Wesley,

1990

- [24] *Giga-Scale CMOS Technology*, Short Course, IEDM 1997
- [25] *180 nm Equipment Performance Metrics*, International 300 mm Initiative, Revision 1, August 1998
- [26] *300 mm Factory Layout and Material Handling Modeling: Phase II Report*, International SEMATECH, November 1999
- [27] C.M. Pichler, R. Plasun, R. Strasser, S. Selberherr, „Simulation of Complete VLSI Fabrication Processes with Heterogeneous Simulation Tools“, *IEEE Transactions on Semiconductor Manufacturing*, Vol. 12, No. 1, S. 76-86, February 1999
- [28] L. Peters, „Choices and Challenges for Shallow Trench Isolation“, *Semiconductor International*, April 1999
- [29] M. Nandakumar, A. Chatterjee, S. Sridhar, K. Joyner, M. Rodder, I.-C.Chen, „Shallow Trench Isolation for advanced ULSI CMOS Technologies“, *Proceedings International Electron Devices Meeting*, S. 133-136, 1998
- [30] J. G. Ryan, R. M. Geffken, N. R. Poulin, J. R. Paraszczak, „The evolution of interconnection technology at IBM“, *IBM Journal of Research & Development*, Vol. 39, No. 4, S. 371-382, 1995
- [31] M. Igarashi, A. Harada, H. Amishiro, H. Kawashima, N. Morimoto, Y. Kusumi, T. Saito, A. Ohsaki, T. Mori, T. Fukada, Y. Toyoda, K. Higashitani, H. Arima, „The Best Combination of Aluminium and Copper Interconnects for a High Performance 0.18mm CMOS Logic Device“, *Proceedings International Electron Devices Meeting*, S. 829-832, 1998
- [32] A. Brand, A. Haranahalli, N. Hsieh, Y.C. Lin, G. Sery, N. Stenton, B.J. Woo, S. Ahmed, M. Bohr, S. Thompson, S. Yang, „Intel's 0.25 Micron, 2.0 Volts Logic Process Technology“, *Intel Technology Journal*, 3rd Quarter `98
- [33] *I300I Factory Guidelines*, Version 4.1, International 300 mm Initiative, 1999
- [34] P. van Zandt, *Microchip Fabrication, A practical guide to semiconductor processing*, 1. Aufl, McGraw-Hill, New York, 1986
- [35] *High -Tech in europäischer Dimension, Halbleiterwerk Dresden*, Informationsschrift zur Eröffnung des Halbleiterwerks Dresden, Siemens AG, 1995
- [36] R. Zorich, *Handbook of Quality Integrated Circuit Manufacturing*, Academic Press, 1991
- [37] P. Csatory, D. Nolan, P. Wolf, A. Honold, „300mm Fab Layout and Automation Concepts“, *Future Fab 5*, Technology Publishing, London, S. 37-41, 1998
- [38] P. Csatory, D. Nolan, P. Wolf, A. Honold, „Evaluation of 300mm Fab Layout Concepts“, *Proceedings of the 7th International Symposium on Semiconductor Manufacturing, Ultra Clean Society*, Tokyo, S. 283-286, 1998
- [39] C.D. Geiger, R. Hase, C.G. Takoudis, R. Uzsoy, „Alternative Facility Layouts for Semiconductor Wafer Fabrication Facilities“, *IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part C: Manufacturing*, Vol. 20, No. 2, S. 152-163, April 1997
- [40] E. Heinen, *Industriebetriebslehre, Entscheidungen im Industriebetrieb*, 9. Auflage, Verlag Gabler, 1991
- [41] R. Wright, C. Cunningham, K. Benhayoune, E. Campbell, V. Swaminathan, R. White, „300mm factory layout and automated materials handling“, *Solid State*

Technology, Vol. 42, Issue 12, December 1999

- [42] D. Meyersdorf, A. Tazhizadeh, „Fab Layout Design Methodology: Case of the 300 mm Fab“, *Semiconductor International*, July 1998
- [43] A.G. Hughes, S.J. Udoka, „A Simulation Based Analysis of Semiconductor Wafer Fabrication with Layout Considerations“, *Proceedings International Conference on Semiconductor Manufacturing Operational Modeling and Simulation*, S. 176, 1999
- [44] M. Weiss, „New Twists on 300 mm Fab Design and Layout“, *Semiconductor International*, July 1999
- [45] P.H. Ballentine, W. Dorl, S. Hoppe, „The 300mm Equipment Front End Modules“, *Proceedings of the SEMI Technology Symposium, SEMICON Kansai, Japan*, 1998
- [46] D. Calhoun, „Facing the challenges of intratool automation in the 300-mm era“, *Micro*, S. 91-101, March 1998
- [47] A. Honold, W. Scheler, „Minienvironment System for 300mm Wafer Manufacturing“, *Semiconductor Fabtech*, Vol. 5, ICG Publishing, London, S. 65-69, 1996
- [48] T.R. Behrens, „Using strategic planning to prepare for 300-mm fab construction“, *Micro*, S. 103-111, October 1997
- [49] J.J. Plata, „300mm Fab Design - A Total Factory Perspective“, *Future Fab Issue 4*, Technology Publishing, London, 1998
- [50] J. Zimmermann, „*Untersuchung der Möglichkeiten berührungsloser Handhabung von Wafern mittels akustischer Levitation*“, Diplomarbeit am Lehrstuhl für Integrierte Schaltungen, TU München, 1999
- [51] *300 mm Integrated Vision for Semiconductor Factories*, Version 3.0, I300I and J300E, November 1999
- [52] A. Honold, P. Ballentine, F. Panchevre, P. Csatory, D. Nolan, P. Wolf, „The Impact of Front Tool Automation on 300mm Fab Productivity“, *Proceedings of the SEMI Technology Symposium 98*, SEMICON Japan, SEMI, Tokyo 1998, S. 2-46-2-51
- [53] M. Weiss, „Evaluating 300-mm fab automation technology options and selection criteria“, *Micro*, S. 65-78, June 1997
- [54] J. Pflueger, P. Ballentine, „Operational Considerations for PGVs in 300mm Wafer Fabs“, SEMICON Korea, 1997
- [55] D. Witt, „Rationalisierung in der öffentlichen Verwaltung“, *Der Gemeindehaushalt*, 5/77, Verlag W. Kohlhammer, Deutscher Gemeindeverlag, 78. Jahrgang
- [56] H. Schierenbeck, *Grundzüge der Betriebswirtschaftslehre*, 8. Auflage, Oldenburg Verlag, 1986, München
- [57] *Vahlens Kompendium der Betriebswirtschaftslehre*, Band 1, 4. Auflage, Verlag Franz Vahlen, 1998
- [58] R. Schwinn, *Betriebswirtschaftslehre*, 2. Auflage, Oldenburg Verlag, 1996, München
- [59] W. Hopfenbeck, *Allgemeine Betriebswirtschafts- und Managementlehre*, 11. Auflage, Verlag Moderne Industrie, 1997, Landsberg/Lech
- [60] C.S. Park, G.P. Sharp-Bette, *Advanced Engineering Economics*, John Wiley & Sons, Inc., 1990, New York
- [61] S. McIntosh, „Doubling Capital Effectiveness“, *Proceedings Fab Managers Forum on Productivity, The Challenge of Managing Fabs in Europe*, Semicon Europa 97

- [62] E. Zahn, U. Schmid, *Produktionswirtschaft I: Grundlagen und operatives Produktionsmanagement*, Verlag Lucius & Lucius, 1996, Stuttgart
- [63] D.A. Patterson, J.L. Hennessy, *Computer Architecture - A Quantitative Approach*, Second Edition, Morgan Kaufmann Publishers, San Francisco, 1996
- [64] T. Chen, M.-J.J. Wang, „A Fuzzy Set Approach for Yield Learning Modeling in Wafer Manufacturing“, *IEEE Transactions on Semiconductor Manufacturing*, Vol. 12, No. 2, S. 252-258, May 1999
- [65] L.S. Milor, „Yield Modeling Based on In-Line Scanner Defect Sizing and a Circuit's Critical Area“, *IEEE Transactions on Semiconductor Manufacturing*, Vol. 12, No. 1, S. 26-35, February 1999
- [66] B. Johnson, S. Mastroianni, T. Stanley, D. Tull, „300 mm Fab Architecture“, *Future Fab International Volume 1 Issue 3*, Technology Publishing London, S. 51-58, 1997
- [67] „Semiconductor producers and suppliers welcome tax depreciation legislation; measure would accurately reflect life of semiconductor equipment“, *SIA Pressemitteilung*, 16.3.1999
- [68] B. Ojo, „U.S. semiconductor sector celebrates restoration of R&D tax credits“, *Electronic Buyers` News*, 23.11.1999
- [69] B. Lutz, M. Hartmann, H. Hirsch-Kreinsen (Hg.), *Produzieren im 21. Jahrhundert, Herausforderungen für die deutsche Industrie, Ergebnisse des Expertenkreises „Zukunftsstrategien“* Band I, Campus-Verlag, 1996, Frankfurt/Main
- [70] W.J. Abernathy, K. Wayne, „Limits of the Learning Curve“, *Harvard Business Review*, S. 109-119, Sept.-Oct. 1974
- [71] R. McIvor, *Managing for Profit in the Semiconductor Industry*, Prentice Hall, 1989
- [72] G.A. Leonovich, A.P. Franchino, W.J. Miller, U.E. Tsou: „Integrated cost and productivity learning in CMOS semiconductor manufacturing“, *IBM Journal Research & Development*, Vol. 39, No. 1/2, S. 201-213, January/March 1995
- [73] B. Johnson, „The Conversion to 300mm Wafer Processing“, *Proceedings Semiconductor Equipment and Materials beyond JESSI - 300mm and Single Wafer Processing*, Productronica München, 1995
- [74] „Größere Scheiben für die Massenproduktion“, *VDI nachrichten*, Nr. 13, 31.März 2000
- [75] „Rosige Zukunft für Europas Halbleiterindustrie“, *VDI nachrichten*, Nr. 13, 31.März 2000
- [76] A. Honold, S. Dürr, „Advanced Waferfab Technology“, *Semiconductor Fabtech*, Vol. 5, S. 1-3, 1996
- [77] David W. Reed, Steven D. Leeke, „Industry and MMST Program Overview“, *TI Technical Journal*, S. 4-19, September-October 1992
- [78] D. Anderson, „Stoking the productivity engine with new materials and larger wafers“, *Solid State Technology*, S. 57, March 1997
- [79] C. Weber, D. Jensen, E.D. Hirlleman, „What drives defect detection technology?“, *Micro*, S. 51-75, June 1998
- [80] O. Ruelle, „Continuous Flow Manufacturing: The Ultimate Theory of Constraints“, *Proceedings Advanced Semiconductor Manufacturing Conference*, S. 216, 1997
- [81] R. Helms, „Fab Automation: Fab-wide automation is critical to microelectronics` fu-

- ture", *Solid State Technology*, January 2000
- [82] J. Harter, C. Esser, K.H. Küsters, G. Rauter, J. Rüstig, „Fast Feedback as Key to Successful Fab Ramp-Up", *Proceedings Fab Managers Forum on Productivity, The Challenge of Managing Fabs in Europe*, Semicon Europa 97
- [83] *Gabler Wirtschaftslexikon*, 12. Auflage, Betriebswirtschaftlicher Verlag Dr. Th. Gabler GmbH, 1988, Wiesbaden
- [84] *The Medea Design Automation Roadmap*, 1st release, 1999
- [85] M. Penn, „Cycle times squeeze", *European Semiconductor*, S. 10, July/August 1995
- [86] U. Lindemann, R. Reichwald, *Integriertes Änderungsmanagement*, Springer, 1998, Berlin
- [87] Dr. R. Drechsler, Vortrag am Lehrstuhl für Integrierte Schaltungen, TU München, 1999
- [88] I. Ruge, K. Habegger, Reisebericht über den Besuch von Halbleiterfirmen in USA, 5. - 21. September 1996
- [89] *300 mm Factory Layout and Material Handling Modeling: Phase I Report*, International SEMATECH, December 1999
- [90] *Equipment Performance Metrics - Revision 3*, International 300 mm Initiative, 1997
- [91] *Lithography CoO Analysis - Revision 3*, International SEMATECH, June 1999
- [92] J.D. Witte, M.L. Ashline, „Determining Product Value Based on Consumption of Factory Constraints", *Proceedings Advanced Semiconductor Manufacturing Conference*, S. 141, 1997
- [93] R. Villforth, „Applying Constraint Management Theory in a Wafer Fab", *Proceedings Advanced Semiconductor Manufacturing Conference*, S. 175, 1994
- [94] L.F. Atherton, R.W. Atherton, *Wafer Fabrication: Factory Performance and Analysis*, Kluwer Academic Publishers, 1995, Boston/Dordrecht/London
- [95] S. McIntosh, „Is the Two Billion Dollar Fab Economic?", *Future Fab 5*, Technology Publishing, London, 1998
- [96] L. Sattler, R. Schlueter, „Semiconductor Metrics: Conflicting Goals or Increasing Opportunities?", *Proceedings Advanced Semiconductor Manufacturing Conference*, S. 55, 1998
- [97] R.C. Leachman, „Closed-Loop Measurement of Equipment Efficiency and Equipment Capacity", *IEEE Transactions on Semiconductor Manufacturing*, Vol. 10, No. 1, S. 84-97, February 1997
- [98] D.L. Dance, T. DiFloria, D.W. Jimenez, „Modeling the Cost of Ownership of Assembly and Inspection", *IEEE Transactions on Components, Packaging, and Manufacturing Technology, Part C: Manufacturing*, Vol. 19, No. 1, S. 57-60, January 1996
- [99] Course: Overall Equipment Effectiveness (OEE) - Improving Equipment Productivity, SEMICON Europa `99
- [100] *Semiconductor Manufacturing Productivity, Overall Equipment Effectiveness (OEE) Guidebook*, Revision 1.0, SEMATECH, 1995
- [101] D.L. Benavides, J.R. Duley, B.E. Johnson, „As Good As It Gets: Optimal Fab Design and Deployment", *IEEE Transactions on Semiconductor Manufacturing*, Vol. 12, No. 3, S. 281-287, August 1999

- [102] R. Randall, M.J. Harry, „Six Sigma - Reaching Our Goal", *TI Technical Journal*, S. 78-87, September-October 1995
- [103] S.S. Aurand, P.J. Miller, „The Operating Curve: A Method to Measure and Benchmark Manufacturing Line Productivity", *Proceedings Advanced Semiconductor Manufacturing Conference*, S. 391, 1997
- [104] N.G. Pierce, A. Yost, „Cycle Time Metrics for R&D Semiconductor Wafer Fabrication", *Proceedings Advanced Semiconductor Manufacturing Conference*, S. 105, 1995
- [105] D. Meyersdorf, T. Yang, „Cycle Time Reduction For Semiconductor Wafer Fabrication Facilities", *Proceedings Advanced Semiconductor Manufacturing Conference*, S. 418, 1997
- [106] L. M. Wein, „On the Relationship Between Yield and Cycle Time in Semiconductor Wafer Fabrication", *IEEE Transactions on Semiconductor Manufacturing*, Vol. 5, No. 2, S. 156-158, 1992
- [107] S. P. Cunningham, J. G. Shanthikumar, „Empirical Results on the Relationship Between Die Yield and Cycle Time in Semiconductor Wafer Fabrication", *IEEE Transactions on Semiconductor Manufacturing*, Vol. 9, No. 2, S. 273-277, May 1996
- [108] R.C. Larson, A.R. Odoni, *Urban Operations research*, Prentice-Hall, 1981, NJ
- [109] A. Schömiß, *Performance Analysis of WIP-controlled Manufacturing Systems*, Würzburger Beiträge zur Leistungsbewertung Verteilter Systeme, Bericht 01/97, Bayerische Julius-Maximilians-Universität Würzburg
- [110] L. Kleinrock, *Queueing Systems, Volume I: Theory*, John Wiley & Sons, 1975, New York
- [111] K. Neumann, M. Morlock, *Operations Research*, Hanser Verlag, 1993, München
- [112] L. Kleinrock, *Queueing Systems, Volume II: Computer Applications*, John Wiley & Sons, 1975, New York
- [113] T.T. Allen, ISE 513 Lecture 16, M/M/c/K/K Queue, Networks, and Review, Department of Industrial, Welding & Systems Engineering, Ohio State University
- [114] C. Wang, R.W. Wolff, „The M/G/c queue in light traffic", *Queueing Systems: Theory and Applications*, 29, S. 17-34, 1998
- [115] L. Sattler, „Using Queueing Curve Approximations in a Fab to Determine Productivity Improvements", *Proceedings Advanced Semiconductor Manufacturing Conference*, S. 140, 1996
- [116] W. Sandmann, „Wartezeitverteilungen in Tandem-Warteschlangennetzen mit stückweise approximierenden Dichten maximaler Entropie", Diplomarbeit am Institut für Informatik, Friedrich-Wilhelms-Universität Bonn, Januar 1998
- [117] P.R. Kumar, „Re-Entrant Lines", *Queueing Systems: Theory and Applications, Special Issue on Queueing Networks*, Vol. 13, Nos. 1-3, S. 87-110, May 1993
- [118] S. C. H. Lu, D. Ramaswamy, P. R. Kumar, „Efficient scheduling policies to reduce mean and variance of cycle-time in semiconductor manufacturing plants", *IEEE Transactions on Semiconductor Manufacturing*, Vol. 7, No. 3, S. 374-385, August 1994
- [119] O. Ruelle, „Operations Management System for Continuous Flow Manufacturing", *Proceedings Advanced Semiconductor Manufacturing Conference*, S. 36, 1996
- [120] B. Sherlock-Lynn, „Comparing Lithography Systems Using Cost of Ownership",

- Vortrag bei der SEMI Europe Konferenz ,Cost Effective IC Manufacturing", Brüssel, 28.10.1998
- [121] R. Doering, „The Outlook for Single-Wafer and Large Wafer Processing", *Proceedings Semiconductor Equipment and Materials beyond JESSI - 300mm and Single Wafer Processing*, Productronica München, 1995
- [122] D.W. Collins, „Inventory modeling yields significant cycle-time improvements", *Solid State Technology*, S. 44-48, January 1999
- [123] H. Koriyama, Y. Yazaki, I. Honboti, Y.Kato, T. Kusakibaru, „Virtual Manufacturing System", *International Symposium on Semiconductor Manufacturing*, S. 5, 1995
- [124] D. Gscheidlen, Applied Materials GmbH, persönliche Mitteilung, Januar 2000
- [125] C. Cunningham, „Structured Approach to Modeling Advance Semiconductor Trends", *Proceedings International Conference on Semiconductor Manufacturing Operational Modeling and Simulation*, S. 167, 1999
- [126] W. Moore, W. Maly, A. Strojwas (Hrsg.), *Yield Modelling and Defect Tolerance in VLSI*, IOP Publishing, Bristol, 1987
- [127] R.L. Guldi, D.E. Paradis, M.T. Whitfield, F.D. Poag, D.P. Jensen, „Strategy and Metrics for Wafer Handling Automation in Legacy Semiconductor Fab", *IEEE Transactions on Semiconductor Manufacturing*, Vol. 12, No. 1, S. 102-108, February 1999
- [128] S.C. Wood, „Cost and Cycle Time Performance of Fabs Based on Integrated Single-Wafer Processing", *IEEE Transactions on Semiconductor Manufacturing*, Vol. 10, No. 1, S. 98-111, February 1997
- [129] P. Patrino, M. Montier, „300mm Fab for CMOS Logic Process and Cost of Ownership Driven", *Proceedings Semiconductor Equipment and Materials beyond JESSI - 300mm and Single Wafer Processing*, S. 116-124, Productronica München, 1995
- [130] Firmeninformation 1st Silicon, Malaysia, www.1stsilicon.com
- [131] R. Cholewa, „Productivity Improvement by Implementing Self-Directed Working Teams", *Proceedings Fab Managers Forum on Productivity, The Challenge of Managing Fabs in Europe*, Semicon Europa 97
- [132] B. J. Cabibi, „The Human Aspects of Organizational Change in Semiconductor Manufacturing", *Proceedings Advanced Semiconductor Manufacturing Conference*, S. 328, 1997
- [133] G. Horn, W.A. Podgorski, „What Gain from Small Batch Manufacturing?", *Semiconductor Fabtech*, 8th Edition, S. 35-37, 1998
- [134] „Intel`s Moore sees economics as new ,law`", *Electronic Engineering Times*, S. 28, 21. August 1995
- [135] J.R. Lineback, „How real is the \$10 billion fab?" *Semiconductor Business News*, July 1999
- [136] I.T. Hadar, J. Nubnan, K. Achiwa, O. Turbahn, „Economic Challenges and Opportunities in the 300 mm Transition", *Semiconductor International*, October 1998
- [137] „Industry Outlook", *European Semiconductor*, S. 14, September 1999
- [138] M.E. McGraw, B. Rowe, „Lessons from Fab Woebegone", *Semiconductor International*, October 1998
- [139] D. Seligson, „Planning for the 300mm Transition", *Intel Technology Journal*, 4th

Quarter `98

- [140] D. Leonard, „Time to Start Mastering Change“, *Semiconductor Magazine*, S. 70-73, March 2000
- [141] T. Kaufmann, „The Paradigm Shift for Manufacturing Execution Systems in European Projects and SEMI Activities“, *Semiconductor Fabtech*, 8th Edition, S. 17-25, 1998
- [142] D. Kozian, „Der Datenfluß wird rund“, *Computer&Automation*, S. 23-25, 4/99
- [143] J. Baliga, „MES and CIM: At the Center of Productivity“, *Semiconductor International*, July 1998
- [144] J.-P. Dauvin, „Chip Market`s Long-Term Growth Unaffected by Short Term Cycles“, *Electronics Journal*, S. 9-12, August 1998
- [145] I. Ruge, *Integrierte Schaltungen 1*, Skript zur Vorlesung, Technische Universität München, 1999
- [146] Halbleiterforum: „Auf zu höheren Sphären“, *Markt&Technik*, Nr. 50, S. 28-33, 10.12.99
- [147] J.A. Davis, R. Venkatesan, A. Kaloyeros, M. Beylansky, S.J. Souri, K. Banerjee, K.C. Saraswat, A. Rahman, R. Reif, J.D. Meindl, „Interconnect Limits on Gigascale Integration (GSI) in the 21st Century“, *Proceedings of the IEEE*, Vol. 89, No. 3, March 2001
- [148] J. Donovan, „Gartner Dataquest Announces Final 2003 Semiconductor Market Share“, 23 March 2004
- [149] P. Singer, „Dual Damascene Challenges Dielectric-Etch“, *Semiconductor International*, August 1999
- [150] D. Yost, „Semiconductor Equipment Industry: Migration From Equipment to Entire Process Module Solutions“, Keynote Address, *IEEE International Integrated Reliability Workshop Final Report*, 1998
- [151] L. Peters, „Is the 0.18 mm Node Just a Roadside Attraction?“, *Semiconductor International*, January 1999
- [152] H.-S.P. Wong, D.J. Frank, P.M. Solomon, C.H.J. Wann, J.J. Welser, „Nanoscale CMOS“, *Proceedings of the IEEE*, Vol. 87, N0. 4, S. 537-570, April 1999
- [153] S. Thompson, P. Packan, M. Bohr, „MOS Scaling: Transistor Challenges for the 21st Century“, *Intel Technology Journal*, 3rd Quarter `98
- [154] D.J. Frank, R.H. Dennard, E. Nowak, P.M. Solomon, Y. Taur, H.-S.P. Wong, „Device Scaling Limits of Si MOSFETs and Their Application Dependencies“, *Proceedings of the IEEE*, Vol. 89, No. 3, March 2001
- [155] A. Schoemig, „On the Corrupting Influence of Variability in Semiconductor Manufacturing“, *Proceedings of the 1999 Winter Simulation Conference*
- [156] Y. Taur, „CMOS design near the limit of scaling“, *IBM Journal of Research and Development*, Vol. 46, No. 2/3, S. 213-222, March/May 2002
- [157] E. J. Nowak, „Maintaining the benefits of CMOS scaling when scaling bogs down“, *BM Journal of Research and Development*, Vol. 46, No. 2/3, S. 169-180, March/May 2002
- [158] S. Natarajan, M. Bost, D. Fisher, D. Krick, C. Kenyon, C. Kardas, C. Parker, R. Gasser, „Process Development and Manufacturing of High-Performance Micropro-

- cessors on 300mm Wafers“, *Intel Technology Journal*, Vol. 6, Issue 2, S. 14-22, May 16, 2002
- [159] S. Thompson, M. Alavi, M. Hussein, P. Jacob, C. Kenyon, P. Moon, M. Prince, S. Sivakumar, S. Tyagi, M. Bohr, „130nm Logic Technology Featuring 60nm Transistors, Low-K Dielectrics, and Cu Interconnects“, *Intel Technology Journal*, Vol. 6, Issue 2, S. 5-13, May 16, 2002
- [160] T. Yancey, IC Insights, Semiconductor Market Mid-Year Forecast, GBC Summer Conference 2004, June 21, 2004
- [161] Gartner Dataquest 2006
- [162] IC Insights, 2006
- [163] Tang Kai Choon, Infineon Technologies, „World DRAM Market Overview“, VIA Technology Forum 2004
- [164] W.J. Hopp, M.L. Spearman, *Factory Physics*, Chicago, Irwin, 1996
- [165] Kan Wu, „An Examination of Variability and Its Basic Properties for a Factory“, *IEEE Transactions on Semiconductor Manufacturing*, Vol. 18, No. 1, S. 214-221, February 2005
- [166] T. Wakabayashi, S. Watanabe, Y. Kobayashi, T. Okabe, A. Koike, „High-Speed AMHS and Its Operation Method for 300-mm QTAT Fab“, *IEEE Transactions on Semiconductor Manufacturing*, Vol. 17, No. 3, S. 317-323, August 2004
- [167] C. Weber, „Yield Learning and the Sources of Profitability in Semiconductor Manufacturing and Process Development“, *IEEE Transactions on Semiconductor Manufacturing*, Vol. 17, No. 4, S. 590-596, November 2004
- [168] J. H. Jacobs, P. P. Van Bakel, L. F. P. Etman, J. E. Rooda, „Quantifying Variability of Batching Equipment Using Effective Process Times“, *IEEE Transactions on Semiconductor Manufacturing*, Vol. 19, No. 2, S. 269-275, May 2006
- [169] S. Aurand, P. Miller, „The Operating Curve: A Method to Measure and Benchmark Manufacturing Line Productivity“, *IEEE/SEMI Advanced Semiconductor Manufacturing Conference 1997*, S. 391-397
- [170] M. Mittler, N. Gerlich, A. Schömig, „On Cycle Times and Interdeparture Times in Semiconductor Manufacturing“, University of Würzburg, Institute of Computer Science, Research Report Series
- [171] J.W. Fowler, J.K. Robinson, „Measurement and Improvement of Manufacturing Capacity (MIMAC) Project Final Report“, *SEMATECH Technology Transfer #959062861A-TR*, 1995
- [172] J. Robinson, J. Fowler, E. Neacy, „Capacity Loss Factors in Semiconductor Manufacturing“, Working Paper, 2003
- [173] A.J. de Ron, J.E. Rooda, „Fab Performance“, *IEEE Transactions on Semiconductor Manufacturing*, Vol. 18, No. 3, S. 399-405, August 2005

Verzeichnis der verwendeten Abkürzungen und Symbole

Abkürzungen:

AGV	= Automated Guided Vehicle
AMHS	= Automated Material Handling System
ASIC	= Application Specific Integrated Circuit
CIM	= Computer Integrated Manufacturing
CMOS	= Complementary Metal Oxide Semiconductor
CMP	= Chemical Mechanical Polishing
CoO	= Cost of Ownership
CVD	= Chemical Vapor Deposition
DRAM	= Dynamic Random Access Memory
DSP	= Digitaler Signalprozessor
ERP	= Enterprise Resource Planning
FOUP	= Front Opening Unified Pod
I300I	= International 300 mm Initiative
IC	= Integrated Circuit
IMD	= Inter-Metall-Dielektrikum
ISO	= International Standards Organization
LDD	= Lightly Doped Drain
LOCOS	= Local Oxidation Of Silicon
LSI	= Large Scale Integration
MES	= Manufacturing Execution System
MMST	= Microelectronics Manufacturing Science and Technology
MOS	= Metal Oxide Semiconductor
MSI	= Medium Scale Integration
MTBA	= Mean Time Between Assist
MTBF	= Mean Time Between Failure
MTTR	= Mean Time To Repair
OEE	= Overall Equipment Effectiveness
OHT	= Overhead Track
PGV	= Personal Guided Vehicle
PMD	= Pre-Metal Dielectric
PVD	= Physical Vapor Deposition
RGV	= Rail Guided Vehicle
ROI	= Return On Investment
RTA	= Rapid Thermal Anneal
RTP	= Rapid Thermal Processing
Selete	= Semiconductor Leading Edge Technologies
SEMATECH	= Semiconductor Manufacturing Technology

SEMI	= Semiconductor Equipment and Materials International
SIA	= Semiconductor Industry Association
SLSI	= Super Large Scale Integration
SMIF	= Standard Mechanical Interface
SRAM	= Static Random Access Memory
SSI	= Small Scale Integration
STI	= Shallow Trench Isolation
ULSI	= Ultra Large Scale Integration
VLSI	= Very Large Scale Integration
WIP	= Work In Process
WSPM	= Wafer Starts per Month
WSPW	= Wafer Starts per Week

Symbole:

A	= Chipfläche
A_{RC}	= Produktionsausfall durch Ramp Up
A_{RU}	= Umsatzausfall durch Ramp Up
C	= Zahl der Chips
c_a	= Variationskoeffizient Ankunft
c_e	= Variationskoeffizient Bearbeitung
CPW	= Zahl der Chips pro Wafer
C_g	= Gatekapazität
D	= Defektdichte
DS	= Gerätedurchsatz
d	= Veränderungsfaktor der Defektdichte
D_{Wafer}	= Waferdurchmesser
E	= Effizienz der Fertigung
G	= Gewinn
G_m	= Transkonduktanz
g_{St}	= Gewinn pro Stück
I_{DS}	= Transistorstrom
K	= Kosten
K_F	= fixe Kosten
K_V	= variable Kosten
K_{wip}	= Kapitalverlust durch gebundenes Kapital
k_{St}	= Stückkosten
k_V	= variable Kosten pro Stück
Kap	= Gesamtkapital
L	= Anzahl der Teile im System (Warteschlangentheorie)
L_g	= Gatelänge
m	= Anteil der Monitorwafer

m_f	= Mean Time To Failure (MTTR)
m_r	= Mean Time To Repair (MTTR)
N_A	= Akzeptordotierung
N_D	= Donatordotierung
N_{Sub}	= Substratdotierung
n	= Zahl der Geräte
P	= Stückpreis
P_G	= Leistungsverbrauch
p	= Veränderungsfaktor des Preises
R	= Rentabilität
r_a	= mittlere Ankunftsrate
r_e	= Maschinenkapazität
R_{tr}	= Transistor on-Widerstand
R_U	= Umsatzrentabilität
R_{Wafer}	= Waferradius
s	= Veränderungsfaktor der Chipgröße
s	= Zahl der parallelen Bedienstellen (Warteschlangentheorie)
TD	= Durchlaufzeit
TD_p	= Bearbeitungszeit im Gerät
TD_q	= Wartezeit vor Gerät
t	= Zeit
t_{gox}	= Gateoxiddicke
TH	= effektiver Durchsatz des Systems
t_L	= Produktlebensdauer
t_R	= Dauer des Ramp Up
t_{setup}	= Setupdauer
t_T	= Transportzeit
t	= Gate-Delay
U	= Umsatz
u	= Nutzungsgrad
V_{DD}	= Versorgungsspannung
V_T	= Schwellenspannung
W	= Wirtschaftlichkeit
W	= Gesamtdurchlaufzeit (Warteschlangentheorie)
W_{tr}	= Transistorweite
W_{Wafer}	= Wert des Wafers nach allen Herstellungsschritten
W_{Wafer}	= Mittelwert des Wafers
W_{wip}	= Wert des WIP-Bestandes
wip	= Zahl der Wafer im Fertigungsablauf
X_j	= Tiefe des pn-Übergangs
Y	= Ausbeute
z	= Zinssatz

Danksagung

An dieser Stelle möchte ich allen, die zum Gelingen dieser Arbeit beigetragen haben, meinen herzlichen Dank aussprechen.

An erster Stelle möchte ich meinem Doktorvater Herrn Prof. Dr.-Ing. Walter Hansch für seine Betreuung während der Bearbeitung danken. Sein Interesse an der Thematik und seine Unterstützung haben diese Arbeit erst möglich gemacht.

Für die Übernahme des Zweitreferats danke ich Herrn Prof. Dr. rer. nat. Ignaz Eisele von der Universität der Bundeswehr München.

Herrn Prof. Dr.-Ing. Hans-Georg Herzog bin ich für die Übernahme des Prüfungsvorsitzes zu Dank verpflichtet.

Bei Herrn Prof. Dieter Witt möchte ich mich für seine Unterstützung betreffend den betriebswirtschaftlichen Teil der Arbeit bedanken. Zahlreiche Diskussionen und Hinweise waren eine große Hilfe für mich.

Herrn Dr. Alexander Schömig sei für die Bereitstellung von Gerätedaten der Firma Infineon gedankt.