

**The Analytic Forecast Method (AFM) -  
Entwicklung eines analytischen Ansatzes zur  
Materialanlieferungsprognose in der  
Halbleiterfertigung**

Marcin Mosinski

Vollständiger Abdruck der von der Fakultät für Informatik  
der Universität der Bundeswehr München zur Erlangung des akademischen Grades eines  
Doktor-Ingenieurs  
genehmigten Dissertation.

Gutachter:

1. Prof. Dr. rer. nat. Oliver Rose
2. PD Dr.-Ing. Gerald Weigert

Die Dissertation wurde am 02.07.2013 bei der Universität der Bundeswehr München  
eingereicht und durch die Fakultät für Informatik am 07.11.2013 angenommen.  
Die mündliche Prüfung fand am 27.11.2013 statt.



## Vorwort

Ich möchte an der Stelle allen meinen Freunden und Verwandten danken, die mich bei der Anfertigung dieser Arbeit unterstützt haben. Besonderer Dank gilt meiner Frau Anna für ihre Unterstützung, Verständnis, ständige Motivation und Übernahme größtenteils der Pflichten bei der Pflege unserer Tochter Emilia. Meinen Eltern danke ich für ihr Vertrauen in meine Fähigkeiten sowie für ihre Unterstützung und Motivation. Des Weiteren möchte ich meinen beiden Betreuern Herren Wolfgang Scholl und Herrn Prof. Dr. rer. nat. habil. Oliver Rose für die vielen Hinweise und Denkanstöße danken.

Bereits in der Grundschule hat mir meine Mathematik Lehrerin Frau A. Szelest eine Regel eingeprägt:

*„Es ist keine Herausforderung eine Lösung zu finden, die richtige Herausforderung ist die einfachste Lösung zu finden“*

Diese Aussage evaluierte und ergänzte Herr Prof. Dr. rer. nat. habil. Oliver Rose, der meint:

*„Um ein komplexes Problem zu vereinfachen und zu beschreiben, braucht man immer einen gewissen Grad an Komplexität.“*

Durch die Zusammensetzung von diesen zwei Gedanken entstand folgendes Motto für meine vorliegende Dissertation:

*„Versuche das gestellte Problem mit einem möglichst einfachen Ansatz zu lösen. Der Ansatz soll jedoch so komplexen sein, wie es nötig ist, das Problem möglichst treu nachzubilden.“*



# Inhaltsverzeichnis

---

<b>1</b>	<b>Einführung .....</b>	<b>1</b>
1.1	Zielsetzung.....	1
1.2	Motivation .....	2
1.3	Aufbau der Arbeit.....	3
<b>2</b>	<b>Problemstellung .....</b>	<b>5</b>
2.1	Funktionsprinzip der Halbleiterfabrik .....	5
2.2	Produktmixeinfluss auf die Fertigung .....	7
2.3	Die Problematik von Zwischenlagern .....	8
2.4	Stochastik in der Halbleiterfertigung.....	10
2.4.1	Prozess- und Rüstzeiten .....	11
2.4.2	Wartezeiten.....	12
2.4.3	Häufigkeit und Dauer von Ausfällen und vorbeugender Wartung .....	12
2.4.4	Hold-Zustand.....	13
2.4.5	Sampling.....	14
2.4.6	Lager-Ausschleusung .....	14
2.4.7	Übergang zwischen Fertigungs- und Testbereich .....	15
2.5	Anforderungen an das Vorhersagewerkzeug.....	15
2.5.1	Funktionale Anforderungen .....	15
2.5.2	Nichtfunktionale Anforderungen .....	16
2.5.3	Problembereichsanforderungen.....	17
<b>3</b>	<b>Analyse bestehender Prognoseansätze.....</b>	<b>19</b>
3.1	Theoretische Grundlagen von Prognosetechniken .....	19
3.1.1	Eigenschaften von Prognosetechniken.....	19
3.1.2	Prognosefehler und -bewertung .....	20
3.1.3	Entwicklung und Bedeutung von Prognosen .....	23
3.2	Simulation.....	24
3.2.1	Geschichte und Definition.....	25
3.2.2	Simulationsarten.....	27
3.2.3	Aufbau und Funktionsweise der diskreten ereignisorientierten Simulation .....	29
3.2.4	Anwendung in der Halbleiterfertigung .....	30

3.2.5	Bewertung als Prognosewerkzeug in der Halbleiterfertigung.....	34
3.3	Planzeitbasierte Prognose .....	35
3.3.1	Aufbau und Funktionsweise.....	35
3.3.2	Bewertung der planzeitbasierten Prognose in der Halbleiterfertigung .....	37
3.4	Warteschlangentheorie .....	38
3.4.1	Definition und Einsatzgebiete .....	39
3.4.2	Verwendung in der Halbleiterfertigung und Bewertung.....	40
3.5	Andere Vorhersageansätze .....	42
<b>4</b>	<b>Das Konzept der Analytischen Vorhersagemethode – AFM .....</b>	<b>45</b>
4.1	Problematik der Komplexität und Qualität der Basisdaten .....	46
4.2	Die Grundlagenfunktionen des Prognosealgorithmus .....	47
4.3	Bestimmung einer Zieloperationsmenge für die Prognose.....	50
4.4	Klassifizierung.....	51
4.4.1	Einfluss der Produktionssteuerung auf die Durchlaufzeiten .....	52
4.4.2	Auslastungsschwankungsbedingte Durchlaufzeitänderungen .....	60
4.4.3	Prozesszeitunterschiede bedingt durch Losgröße .....	62
4.4.4	Andere durchlaufzeitbeeinflussende Effekte .....	67
4.5	Statistik.....	71
4.5.1	Statistische Modellierung für eine Prognose.....	71
4.5.2	Loseinschleusung in die Fabrik.....	72
4.5.3	Durchlaufzeit .....	74
4.5.4	Sampling an der Zieloperation .....	78
4.5.5	Das Verhalten des Hold-Zustands.....	81
4.5.6	Wechsel der Produktionsbereiche .....	86
4.6	Vorhersageberechnung .....	90
4.6.1	Auslesen der Bestandsverteilung über die Fabrik und Kalkulation der Planeinschleusung .....	90
4.6.2	Bestimmung der Ankunftszeitpunkte an der Zieloperation .....	94
4.6.3	Ermittlung von Anlieferungsmengen an der Zieloperation.....	98
4.7	Vorhersageergebnisse und Anwendungsfälle.....	100
<b>5</b>	<b>Implementierung der Vorhersagesoftware .....</b>	<b>107</b>
5.1	Die Entwicklungsumgebung.....	107
5.2	Die Implementierung .....	109

5.2.1	Datenbeschaffung .....	111
5.2.2	Datenaufbereitung .....	113
5.2.3	Vorhersagebildung .....	116
5.2.4	Visualisierung der Resultate.....	118
5.3	Validierungsmodul .....	120
<b>6</b>	<b>Validierung.....</b>	<b>121</b>
6.1	Analyse der Zeitabweichung .....	121
6.2	Analyse des Anlieferungsverhaltens .....	126
6.3	Analyse der Prognose für ausgewählte Anwendungsfälle .....	132
6.4	Verifikation der Ansätze und Lösungswege durch Validierung der AFM für ausgewählte Problembereiche .....	135
6.4.1	Bewertung von unterschiedlichen Strategien in der Erstellung des operativen Plans für die Materialeinschleusung.....	135
6.4.2	Bewertung von unterschiedlichen Ansätzen für die Prognose des Los-Samplings .....	139
6.4.3	Analyse des Einflusses von Hold-Losen auf die Vorhersagequalität .....	143
6.4.4	Analyse des Los-Samplings hinsichtlich der Durchlaufzeit .....	147
<b>7</b>	<b>Vergleich unterschiedlicher Prognosemethoden in der Halbleiterfertigung .....</b>	<b>149</b>
<b>8</b>	<b>Zusammenfassung und Ausblick.....</b>	<b>157</b>
	<b>Abbildungsverzeichnis.....</b>	<b>160</b>
	<b>Tabellenverzeichnis.....</b>	<b>165</b>
	<b>Literaturverzeichnis.....</b>	<b>166</b>



# 1 Einführung

## 1.1 Zielsetzung

Die Herstellung elektronischer Bauteile, die auf Halbleitern basieren, ist einer der komplexesten Produktionsprozesse in der Industrie. Halbleiterfabriken umfassen mehrere Hundert Anlagen und arbeiten nach dem Job-Shop-Prinzip. Jedes Produkt wird innerhalb einer eindeutigen Route und im Verlauf von einigen hundert Schritten produziert. Im Fachjargon der Halbleiterbranche werden diese Schritte häufig Operation genannt. Im Regelfall bearbeiten dedizierte Maschinen die unterschiedlichen Produktionsschritte. Jedoch existieren einzelne Anlagen, die in der Lage sind, unterschiedliche Fertigungsprozesse auszuführen. Dies führt dazu, dass die Notwendigkeit besteht, jedem Herstellungsschritt eine Verordnung in Form eines Rezeptes mit Prozesskenndaten, wie beispielsweise Art und Dauer, zuzuordnen. Resultierend daraus werden innerhalb eines Fertigungsprozesses für ein Werkstück Anlagen wiederholt angefahren. Es ist möglich, mehrere Werkstücke zeitgleich zu bearbeiten, dabei kann jedoch der Prozessverlauf meistens nicht unterbrochen werden. Scheduling für Problemstellungen der Halbleiterfertigung, ähnlich dem Problem des Handlungsreisenden<sup>1</sup>, gehört in der Informatik zur Reihe der NP-schweren<sup>2</sup> Probleme [Pin08]. Aufgrund dieser Tatsache ist es auch bei einem Großteil der Prognosealgorithmen nicht möglich, diese trivial zu beschreiben oder aufzubauen.

In der Halbleiterindustrie wird hauptsächlich die Simulation zur Prognosebildung angewendet. Der Aufbau notwendiger Simulationsmodelle ist jedoch aufgrund der Fertigungsart, Anzahl von Maschinen und Vielzahl von Störfaktoren hoch komplex. Simulationsexperten benötigen deshalb sehr detaillierte Modelle und Informationen der einzelnen Anlagen bezüglich ihrer deterministischen und stochastischen Eigenschaften. Diese Eigenschaften zeigen sich beispielsweise in der Bearbeitungsgeschwindigkeit, Prozessdauer und Kapazität oder in der Verfügbarkeit, Ausfallrate und Reparaturdauer. Weiterhin sind Informationen über die Maschinenzuordnung für alle Produktionsschritte innerhalb der Route notwendig. Es entsteht eine Dedizierungsmatrix als Kreuzprodukt von Maschine, Operation, Prozessfreigabe und Fabrikat. Da die Werkstücke mehrmals zwischen verschiedenen Anlagen transportiert werden, ist ein detailliertes Transportsystemmodell erforderlich. In der Prozesskette der Halbleiterfertigung treten auch andere zufällige Störgrößen auf. Diese können beispielsweise Pro-

---

<sup>1</sup> Engl. Traveling Salesman Problem - TSP

<sup>2</sup> *NP-hard*, abgekürzt für *Non-deterministic Polynomial-time hard*

zessfehler und die damit verbundene Abarbeitung durch Fachkräfte oder Wiederbearbeitung<sup>3</sup> sein. Das Hauptproblem beim Modellaufbau, die voluminöse Menge an Basisdaten, kristallisiert sich hier heraus. Dabei spielt nicht nur die lange Zeit des Datenbankupdates eine Rolle, sondern auch Inkonsistenz, Verfügbarkeit und Format verfügbarer Datenquellen. Auch das Handling von fertigen Modellen ist nicht einfach, weil Überarbeitungen und anschließende Validierungen sehr zeitintensiv sind. Die Simulation von Halbleiterwerken hat jedoch auch mehrere Vorteile. Es sind zum Beispiel Aussagen über Kennzahlen auf Maschinenebene und für das gesamte Werk möglich. Diese detaillierten Informationen werden jedoch in vielen Fällen nicht benötigt. Vor allem im Bereich der Kurz- und Mittelfristsimulation wird nach Alternativen gesucht, die schneller, aber in der Genauigkeit vergleichbar arbeiten. Im Vergleich mit Simulationen entsteht kein detailliertes, beinahe vollständiges Fabrikmodell, sondern ein begrenztes mathematisches Modell und ein Algorithmus, der die Prognose von bestimmten Parametern bzw. Kennzahlen möglich macht.

In Kooperation mit unserem Projektpartner aus der Industrie wurden Anforderungen und Ziele des Projektes erarbeitet. Im Rahmen des Vorhabens wurde das zentrale Augenmerk auf eine Anlieferungsvorhersage unter Verwendung möglichst weniger Datenquellen für eine bestimmte Maschinengruppe in der Fabrik gelegt. Das fertige Produkt ermöglichte die Vorhersage in Echtzeit bzw. nur mit einer minimalen Verzögerung. Der Horizont der Prognose wurde auf eine Zeitspanne von 14 Tagen begrenzt. Ein Logikchip-fertigendes Halbleiterwerk mit breitem Produktspektrum wurde als Informationsquelle für die Experimente ausgewählt. Die Algorithmen müssen auch Lösungswege für typische Probleme in der Halbleiterfertigung, wie Hold-Zustand<sup>4</sup>, Sampling<sup>5</sup> oder auch Rework offerieren. Die Anlieferungsvorhersage wird ein Engpassfrühwarnsystem darstellen und sowohl für die Optimierungs- und Schedulingalgorithmen als auch Personal- und Wartungsplanung verwendet.

## 1.2 Motivation

Die aktuelle Marktsituation in der Halbleiterbranche wird durch ständig wachsenden Wettbewerb gekennzeichnet. Dabei spielen nicht nur die Entwicklung von neuen Technologien und Produktionskosten eine wichtige Rolle, sondern auch Durchsatz, Produktionsgeschwindigkeit und die Liefertermineinhaltung. Da sich der Konkurrenzkampf mit den Halbleiterherstellern

---

<sup>3</sup> Engl. Rework

<sup>4</sup> Ein Zustand der meistens nach einem Prozessfehler auftritt und Abarbeitung durch ausgebildete Fachkraft benötigt.

<sup>5</sup> Sampling ist ein Kontrollmechanismus in der Fabrik, der ermöglicht überspringen einer Messoperation wenn einer der vorher examinierten Losen der gleichen Spezifikation keine Anomalien aufgewiesen hat.

aus Schwellenländern bezüglich der Verkaufspreise extrem schwierig gestaltet, sind vor allem Optimierungsansätze für die Fertigung gefragt. Für sehr viele Bereiche der Logistik, Personalplanung, Wartungsorganisation und Produktionssteuerung existieren bereits erprobte und etablierte Lösungen für die Verbesserung von Fabrikperformance, Flexibilität oder Liefertreue. Scheduling- und Planungssoftware benötigen jedoch möglichst detaillierte Prognosen, die nicht von Haus aus in die Software integriert sind. Die automatisierte Vorhersage des Verhaltens komplexer Systeme, wie ein Halbleiterwerk, stellt eine sehr große Herausforderung dar. Da der Bedarf an valider Anlieferungsvorhersage beim Projektpartner gegeben war, wurden zuerst existierende Prognosealgorithmen untersucht. Die fehlende Akzeptanz bereits bestehender Vorhersagen von Produktions- und Planungsbereichen, ist auf die fehlende Treffsicherheit und Instabilität der Resultate zurückzuführen. Das Vorhaben des Autors der vorliegenden Arbeit, eine neue Prognosesoftware zu entwickeln, wurde deshalb sehr positiv aufgefasst.

### **1.3 Aufbau der Arbeit**

Die Arbeit wurde in insgesamt acht Kapitel unterteilt, von denen das erste eine Einführung in die Thematik darstellt. Im zweiten Teil wird die spezifische Problematik von Halbleiterwerken beschrieben. Es enthält die detaillierte Beschreibung des Halbleiterwerks mit allen typischen Problemen der Fertigung mit einem Fokus auf die Prognosebildung. Weiterhin wurden die notwendigen Anforderungen an die Prognosesoftware im zweiten Kapitel detailliert niedergeschrieben.

Kapitel 3 befasst sich mit der Analyse der bestehenden Prognosesoftware, wie auch mit den theoretischen Grundlagen der Prognoseerstellung. Im Besonderen wird die Simulation mit allen Vor- und Nachteilen bezüglich der Prognosebildung, als meistbenutztes Werkzeug in der Halbleiterbranche, ins Auge gefasst. Weiterhin werden die analytischen Methoden beschrieben, die beim Projektpartner eingesetzt werden.

Im vierten Kapitel wird detailliert das Konzept der entwickelten Methode zur Vorhersage der Materialanlieferung erläutert. Dabei steht nicht die Entwicklung der Software im Fokus, sondern die Beschreibung der Funktionsweise und Ideen, wie Losklassifizierung, Erzeugung des statistischen Hintergrunds wie auch Vorhersageberechnung und Darstellung der Ergebnisse.

Teil 5 der vorliegenden Arbeit setzt sich mit der Implementierung der Vorhersagemethode auseinander. Zusätzlich wird die Problematik des Umgangs mit enormen Datenmengen, wie der Notwendigkeit von der Verkettung inhomogener, historisch angewachsener Datensysteme

beschrieben. Anschließend erfolgt die Erläuterung des Aufbaus der Prognosesoftware mit Reporting und Validierungsmodulen.

Kapitel 6 und Kapitel 7 beinhalten die Beschreibung der Validierung der Prognose gegen die Realität und deren Bewertung. Dabei werden unterschiedliche Tests durchgeführt. Zum einen wird die Vorhersagbarkeit verschiedener Bereiche innerhalb eines festen Zeitrahmens getestet und zum anderen wird der Vergleich der Ergebnisse von Kurz- und Langfristsimulation sowie planzeitbasierten Prognosewerkzeugen durchgeführt.

Das abschließende Kapitel fasst die Ideen, deren Realisierung und die Ergebnisse der durchgeführten Tests zusammen und erstellt bezüglich der Weiterentwicklung und Erweiterung der vorgestellten Methode eine offene Konzeption.

## 2 Problemstellung

### 2.1 Funktionsprinzip der Halbleiterfabrik

Die Prognosebildung in einem derart komplexen System wie einer Halbleiterfabrik, die sich auf die Herstellung von integrierten Schaltkreisen spezialisiert hat, gestaltet sich sehr schwierig. Um das Konzept der vorliegenden Arbeit und die Funktionsweise des darin beschriebenen Vorhersagealgorithmus zu verstehen, muss zuerst der zugrunde liegende Produktionsprozess erläutert werden.

Die Fertigung von Halbleiterchips kann in zwei Segmente [Atherton95] unterteilt werden, wie in Abbildung 2-1 dargestellt. In der ersten Produktionsetappe werden Frontend-Prozesse, also die direkte Herstellung von integrierten Schaltkreisen wie Transistoren, Widerständen oder Kondensatoren direkt auf einer Siliziumscheibe, in der Halbleiterindustrie Wafer genannt, realisiert. Die Bearbeitung von Siliziumscheiben (Wafer) findet im Frontendbereich mittels fotolithografischem, thermischen und ähnlichen Verfahren statt. Für gewöhnlich wird die Bearbeitung neuer Wafer und die Durchführung von Tests und Inspektionen bereits fertiggestellter Wafer getrennt voneinander durchgeführt. Nach Bearbeitung und Inspektion der Siliziumscheibe wird diese in einem zweiten Produktionsschritt in Chips zersägt, die daraufhin weiter einzeln bearbeitet werden. Dieser Produktionsabschnitt wird Backend genannt und hat zum Ziel, die sensiblen Chips in die Gehäuse einzubauen. Um eine leichtere Wärmeabgabe zu ermöglichen, werden die Chips geschliffen. Innerhalb dieses Produktionsbereichs erfolgt auch das sogenannte Bonden, bei dem die geschliffenen Chips mit Schaltkreisen mit ICs, die den Kontakt mit Leiterplatten herstellen, versehen werden. Backend-Prozessketten sind im Vergleich mit Frontend-Prozessketten meist wesentlich weniger zeitintensiv.

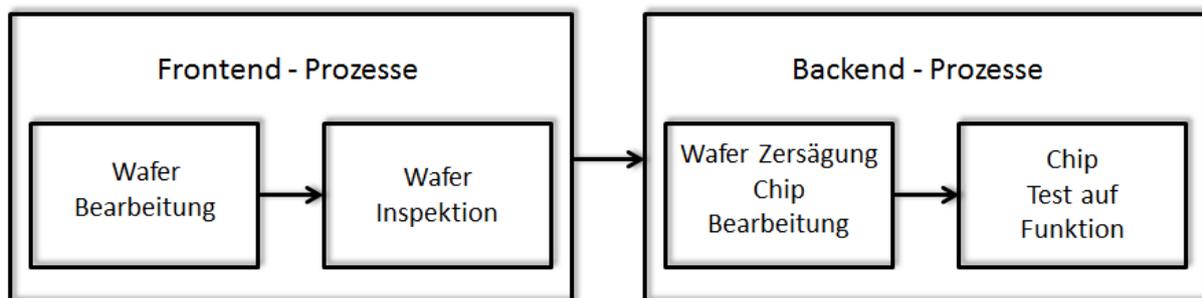


Abbildung 2-1: Produktionsetappen in der Halbleiterfertigung (Eigene Darstellung)

## **Problemstellung**

---

Die anschließende Prüfung, die ebenfalls vom Herstellungsprozess getrennt stattfinden kann, sichert die Funktionalität des Produktes, indem wichtige Eigenschaften überprüft werden und minderwertige Qualität aussortiert wird.

Der in der Arbeit vorgestellte Vorhersageansatz beschäftigt sich ausschließlich mit dem Frontend-Bereich. Aufgrund dessen wird an der Stelle auch der typische Ablauf der Schaltkreisherstellung erläutert. Der Prozessfluss in der Halbleiterfertigung ist in ca. 50 bis zu 1500 Schritte gegliedert [Atherton95]. Sie werden üblicherweise Operationen genannt und besitzen einen fest zugewiesenen Abarbeitungsprozess, wie z. B.:

- Oxidation – ermöglicht den Zuwachs der Siliziumoxidschicht auf dem Wafer, die als Isolator verwendet wird,
- Chemisch-mechanisches Polieren – vollständiges oder partielles Entfernen einer Strukturschicht,
- Abscheidung – Ablagern eines Dünnschichtfilms von Lack auf dem Wafer,
- Fotolithographie – Eintragen eines Musters auf einer lichtempfindlichen Schicht,
- Wafer-Reinigung – Entfernen von Verunreinigungen,
- Plasmaätzen – Wegätzen von einer Schicht in Abhängigkeit von der mit Fotolack markierten Struktur.

Die verschiedenen Prozesse unterscheiden sich deutlich voneinander. Problematisch ist die Abweichung bei den Parametern, die für die Vorhersage relevant sind. Das beste Beispiel dafür ist die Bearbeitungsdauer. Die längsten Abarbeitungszeiten treten meistens bei der thermischen Bearbeitung von Wafern in einem Ofen auf und können bis zu 72 Stunden dauern. Ein idealtypischer Abscheidungsprozess benötigt hingegen nur etwa eine Stunde. Sogar innerhalb eines Prozessstyps kann die zeitliche Dauer deutlich variieren. Beispielhaft dafür können die Parametermessungen nach Fertigstellung der ersten Metallisierungsebene angeführt werden, da dabei die Zeiten der unterschiedlichen Messungen die Spanne von ungefähr 10 Minuten bis über 24 Stunden betragen können. Problematisch für die Prognose sind auch die reihenfolgeabhängigen Rüstzeiten, die häufig mehr Zeit in Anspruch nehmen können als das eigentliche Prozessieren. Die genaue, vollständige Produktionsvorschrift wird in Form von Routen gespeichert, die für jede Operation ein Prozessrezept, Dedizierung an bestimmte Maschinen sowie weitere spezielle Parameter enthalten. Wie in Abbildung 2-2 dargestellt, werden ähnliche Prozessschritte mehrmals angefahren. Es werden neue Schichten im Chip erzeugt, um den notwendigen Platz für weitere Bauelemente gewährleisten zu können. Daraus resultiert eine schleifenförmige Bewegung der Werkstücke in der Halbleiterfabrik. Die vergleichbaren Prozesse werden mithilfe der gleichen Anlagen bzw. Anlagengruppen abgefertigt.

## Problemstellung

Für einen Arbeitsschritt werden parallele Maschinen zugelassen, die aber oft inhomogen sind, was ihre Leistung in der Abarbeitung betrifft. Die Beförderung des Materials erfolgt entweder manuell oder vollautomatisch über ein Transportsystem. Es werden spezielle Behälter<sup>6</sup> verwendet, die eine bestimmte Anzahl an Wafer enthalten. Die Werkstücke werden in der Halbleiterbranche Lose genannt. Die bereits beschriebenen Fertigungsmerkmale deuten auf eine Fertigung hin, die auf dem Job-Shop-Prinzip [Garey76] beruht.

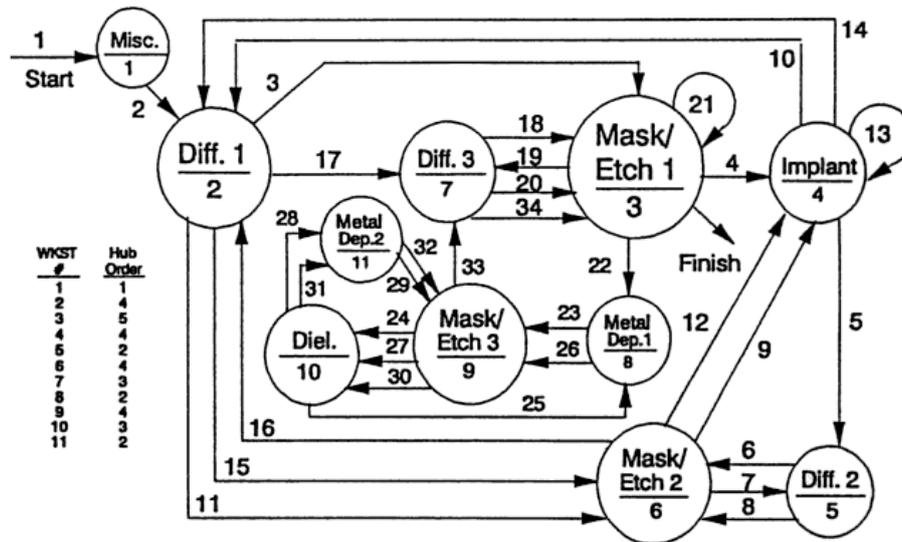


Abbildung 2-2: Beispiel einer Route – Quelle: [Atherton95]

## 2.2 Produktmixeinfluss auf die Fertigung

Die Halbleiterfertigung ist im direkten Vergleich mit der Fertigung anderer Güter, wie beispielsweise die Herstellung von Autos und Flugzeugen, durch einen stark schwankenden Produktmix gekennzeichnet [Atherton95]. Generell können Fabriken mit der Spezialisierung auf die Halbleiterfertigung, anhand der erzeugten Produktmenge unterschiedlicher Erzeugnisse und Produktionsformen, in eine von insgesamt 5 Gruppen eingeteilt werden:

- begrenzte Anzahl an Produkten, die in Serie produziert werden, wie Speicherfabriken,
- große Anzahl an Produkten, produziert meistens in Kleinserien, wie Logikfabriken,
- kleine Anzahl an Produkten, produziert in größeren Mengen, wie Solarmodulfabriken,
- sehr große Anzahl an Produkten und meistens Einzelfertigung, wie Halbleitermaskenfabriken,

<sup>6</sup> Im Halbleiterjargon auch „Horde“ genannt.

- nicht eindeutig zuzuordnender Produktmix und Produktionsform – zumeist die Fabriken, die keine eigene Produktpalette fertigen, sondern nur integrierte Schaltungen in Kundenauftrag produzieren.

Die Massenfertigung ist wesentlich effizienter als die Kleinserien- und Einzelfertigung, weil die Produktionsstätten sich nicht auf die Fertigung eines einzelnen Produktes, sondern auf einen Produktmix spezialisieren. Sowohl das Vermeiden von Umrüstzeiten als auch die hohe Wahrscheinlichkeit, Batch-Anlagen<sup>7</sup> im Full-Batch<sup>8</sup> Modus zu betreiben, führt zu einer effektiven Ausnutzung von Ressourcen. Je breiter der Produktmix ist, desto schwieriger wird es, die Fertigungsstätte optimal zu planen und zu betreiben. Das im Nanometerbereich arbeitende und komplexe Equipment ist zumeist sehr teuer, daher ist eine höchstmögliche Auslastung der Werke vonnöten. Der Betrieb der Fertigungsstätte im Grenzbereich der Auslastung hat durch häufige, unregelmäßige Ausfälle und der damit verbundenen Durchlaufzeitstreuung sehr großen Einfluss auf die Vorhersagegüte.

Die Häufigkeit der Änderung von Produktmischen ist ein wichtiges Merkmal für die Beschreibung der Fertigung, da Änderungen der Produktpalette sehr großen Einfluss auf das Fabrikationsverhalten nehmen. Es kommt dabei zur Verschiebung der Last auf andere Anlagengruppen und damit verbundene Durchlaufzeitveränderungen, die sogar bei den gleichen Produkten auftreten können. Vor allem häufiges Einfügen von vollständig neuen Produkten wie auch das Bearbeiten von Losen, die für Forschung und Entwicklung gedacht sind, führen zu Instabilitäten der Fertigungsstätte. Die erwähnten Probleme erschweren somit natürlich auch die Kennzahlenvoraussage.

### 2.3 Die Problematik von Zwischenlagern

Aufgrund der ähnlichen Grundstruktur und Funktion ist ein Teil unterschiedlicher, integrierter Schaltungen bis zu einer gewissen Ebene identisch. Aus Kostengründen wird Massenfertigung generell bevorzugt und möglichst viele Produkte werden deshalb bis zu einer gemeinsamen Schnittstelle im Produktionsprozess in der Fertigung vereinheitlicht. Anschließend werden die hergestellten Wafer mit den Routen der Endproduktion bearbeitet. Angesichts der derzeitigen Marktsituation spielt die Kundenorientierung in der Firmenpolitik der Halbleiterhersteller eine sehr wichtige Rolle. Daraus resultiert eine möglichst genaue Voraussage, wann

---

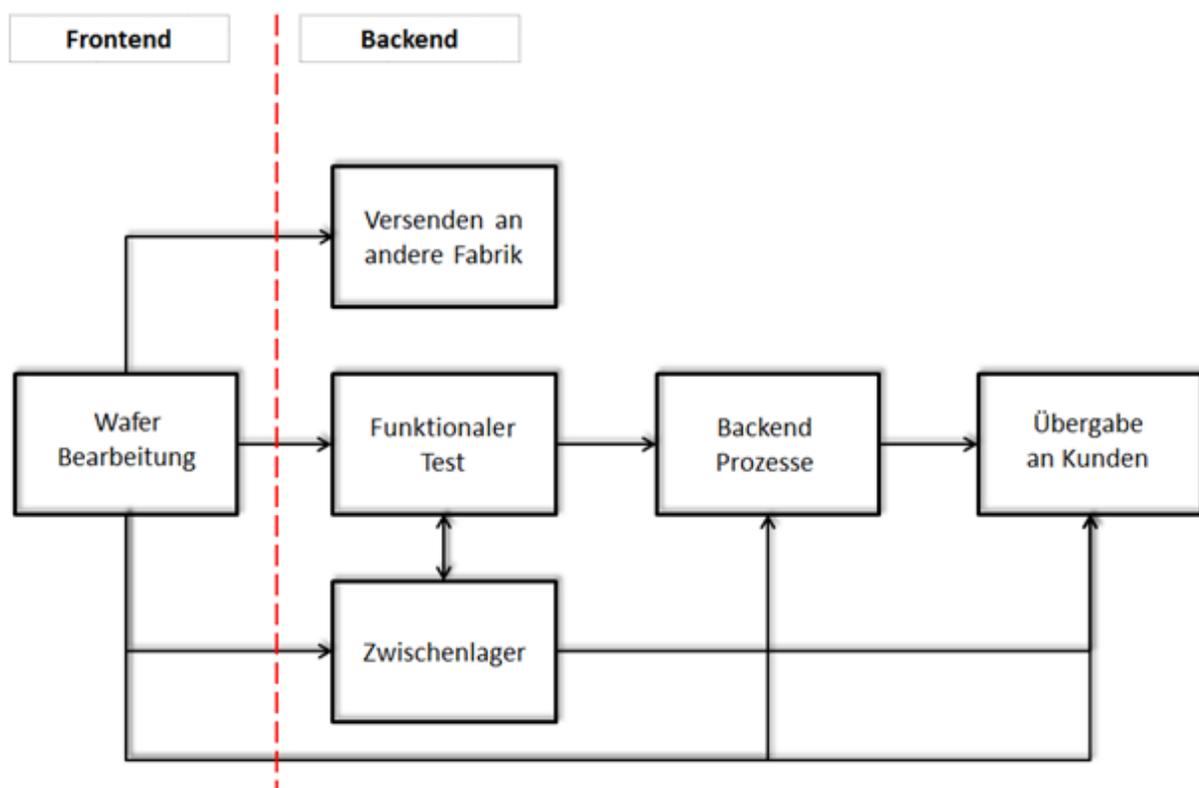
<sup>7</sup> Unter Batch-Anlage versteht man eine Anlage die Gleichzeitig entweder ein oder mehrere Werkstücke mit der gleichen Bearbeitungsvorschrift abarbeiten kann. Dabei ist es wichtig die minimale Anzahl von den Werkstücken zu bestimmen, damit die Abarbeitung rentabel ist.

<sup>8</sup> Full-Batch – betreiben der Batch-Anlage mit maximaler Anzahl an Werkstücken die für die Anlage möglich ist.

## Problemstellung

---

ein Produkt an die Kunden geliefert werden kann. Da in der Halbleiterfertigung die Fertigungszeiten über 3 Monate in Anspruch nehmen können, ist eine flexible Reaktion auf Kundennachfrage nicht einfach. Viele Firmen versuchen deshalb, die notwendige Produktionszeit durch das Einrichten von Zwischenlagern mit vorgefertigten Losen, die nur noch der Endbearbeitung nach Kundenwunsch bedürfen, zu verkürzen (siehe Abbildung 2-3). Die Bedarfe der Kunden an der produzierten Ware sind äußerst schwer abschätzbar und stellen für den Vorhersagenentwickler ein bedeutendes Problem dar. Leider lassen sich nicht alle Produkte teilweise vorproduzieren, sondern sind sogenannte Durchläufer, die von der Einschleusung bis zur Fertigstellung auf der entsprechenden Produktionsroute hergestellt werden.



**Abbildung 2-3: Transfer zwischen Fertigungsbereichen**

Da sehr viele Halbleiterhersteller die Teilung in Fertigungs- und Testbereich im Frontend bevorzugen, entsteht an dieser Stelle die Möglichkeit, ein weiteres Zwischenlager einzuführen. Hier werden bereits fertige Lose eingelagert, die nur noch einen Testdurchlauf benötigen. Bei Fabrikanten, die sich nicht auf Massenware spezialisiert haben, ist die Menge der eingelagerten Ware sehr gering, da die Bedarfe der Kunden nur sehr schwer vorhersagbar sind. Natürlich erfolgt auch im Testbereich die Abarbeitung entsprechend dem beschriebenen Konzept aus der Frontend-Fertigung, d.h. die Lose bewegen sich auf festgelegten Routen. Wie in Abbildung 2-3 dargestellt, werden unterschiedlichen Produkten zumeist abweichende Güteprüfungen zugewiesen. Im Regelfall ist die Testroute, die ein Produkt beim Verlassen des

Fertigungsbereiches und der Ankunft im Testbereich zugewiesen bekommt, festgelegt. Es besteht aber die Möglichkeit, dasselbe Produkt auf unterschiedliche Weise zu überprüfen. Art und Dauer dieser Prüfungen sind eng an die Bearbeitungsart, Stabilität des Prozesses im Backend und die Kundenwünsche gekoppelt. Die Flexibilität der Routensteuerung ist ein weiteres Kriterium, das bei der Vorhersagebildung eine wichtige Rolle spielt und möglichst genau abgebildet werden soll.

### 2.4 Stochastik in der Halbleiterfertigung

Das Wort Stochastik stammt von Altgriechischen *stochastikē technē*<sup>9</sup> und benennt: die Kunst des geschickten Vermutens. Der Begriff fasst die zwei Teilgebiete der Mathematik: Stochastik und Wahrscheinlichkeitstheorie zusammen [Dehling04]. Die mathematische Stochastik befasst sich mit der Untersuchung und Beschreibung von Zufallsexperimenten in Form von zufallsbeeinflussten Ereignissen und zeitlichen Entwicklungen bzw. räumlichen Strukturen. Die Experimente werden meistens mit Daten beschrieben. Die Statistik beschäftigt sich anschließend mit der Analyse des gesammelten Datenmaterials und versucht, mithilfe der Wahrscheinlichkeitstheorie ein Modell zu erzeugen, das das gegebene Problem möglichst genau widerspiegelt. Da das in der vorliegenden Arbeit vorgestellte Prognosewerkzeug auch auf der statistischen Datenanalyse und einem mithilfe von Wahrscheinlichkeitstheorie erstellten Modell basiert, sind die Begriffe immer wieder neu bedeutsam.

In der Definition von Stochastik spielt der Zufall eine zentrale Rolle. Allgemein betrachtet ist ein Wert bzw. Experimentergebnis zufällig, wenn auch bei genauer Beobachtung der Eingangsbedingungen die Resultate nicht richtig vorhersehbar sind [Krengel05]. Gründe dafür können sein, dass entweder keine Prognose möglich ist, die Anfangsbedingungen nicht genau genug geklärt wurden oder eine Ermittlung des Testergebnisses unmöglich ist. In jeder Prognosebildung ist der stochastische Einfluss in Form von Zufall die größte Störgröße der Treffsicherheit. Eine Antwort darauf zu finden, wo Zufälligkeiten in der Halbleiterfertigung auftreten können, ist nicht einfach, da das Auftreten von Zufälligkeiten im System völlig unvorhersehbar ist. Aus der Komplexität des Problems resultiert eine gewisse Pseudozufälligkeit, die auf Unmöglichkeit der vollständigen Beschreibung der Eingangsbedingungen bei Modellierung zurückzuführen ist.

---

<sup>9</sup> Griechisch: στοχαστική τέχνη

### 2.4.1 Prozess- und Rüstzeiten

Unter einer Prozesszeit wird nicht nur die reine Abarbeitungszeit in der Maschine verstanden, sondern auch die benötigte Zeit zur Beladung der Anlage durch einen Werker. Die Einbettung des Faktors Mensch in die Zeit der Bearbeitung führt zur Variation der Bearbeitungsdauer. Aber nicht nur die Unvollkommenheit menschlicher Arbeit führt zu Veränderungen innerhalb der Arbeitsperiode, sondern auch die individuellen Eigenschaften des Werkstückes, der eingesetzten Werkzeuge und des angestrebten Prozesses. Anlagengruppen, die gleiche Arbeitsprozesse ausführen, unterscheiden sich oftmals in ihrer Geschwindigkeit. Auch die Verwendung unterschiedlicher Modi an einer Anlage kann zu Unvorhersehbarkeiten innerhalb des Prozessablaufs führen. Dies trifft insbesondere auf Mehrkammer-Anlagen zu, bei denen der interne Wafer-Fluss variieren kann. Auch der generelle Zustand der Maschine sowie äußere Einflüsse können sich auf die Prozesszeit auswirken. Das Vorzeigebispiel dafür sind Anlagen, welche durch mechanisches Polieren unter Zuhilfenahme bestimmter chemischer Mittel, eine Schicht auf dem Wafer abbauen. Die Abtragsrate ist dabei von der Beschaffenheit des Belages der Schleifeinrichtung abhängig, die sich im Laufe der Zeit ungleichmäßig verändern kann. Um diese Abtragsrate zu bestimmen, werden vor der eigentlichen Produktion Vorläufer bearbeitet. Daraus resultierend ist es möglich, die notwendige Polierzeit für die erwünschte Zielschichtdicke zu errechnen. Der Zeitaufwand eines Arbeitsprozesses ist ebenfalls von der Wafer-Menge in einem Los abhängig. Die Siliziumscheiben werden zum Teil von Maschinen einzeln und nacheinander abgearbeitet. Ein typisches Beispiel eines Prozesses, dessen Bearbeitungsdauer abhängig ist von der Anzahl der Wafer, ist ein Messprozess. Mithilfe spezifizierter Parameter erfolgt die Messung an jedem bzw. an ausgewählten Wafern. Daraus resultiert eine bestimmte Streuung der Messzeit.

Natürlich ist ein Teil der Einflussfaktoren auf die Prozesszeit mit einem äußerst hohen Aufwand bestimmbar. Durch die Komplexität des Problems bleibt jedoch immer ein gewisser Grad an Stochastik in der Prozessdurchführung, der nur durch Approximation oder Heuristik dargestellt werden kann.

Unter der Rüstzeit der Anlage wird eine Zeitspanne verstanden, die notwendig ist, Maschinen für die Herstellung eines Zielproduktes vorzubereiten. Soll ein anderes Zielprodukt hergestellt werden, wird ein Umrüstvorgang an der Maschine notwendig. Die Umrüstung der Anlage erfolgt durch:

- Veränderung der physikalischen Eigenschaften, wie Abkühlung eines Ofens oder Druckänderung in der Implantationsanlage,

- Veränderung der mechanischen Eigenschaften, wie Ab- oder Einschalten einer Kammer in einem Clustertool,
- Veränderung der informationellen Eigenschaften, wie erneutes Einlesen eines Prüfprogramms von der Festplatte bei einem Messgerät.

Auch im Fall der Rüstzeiten ist ein genauer Nachbau von einer Matrix, die die notwendige Umrüstzeiten zwischen allen möglichen Produktkombinationen beinhaltet, ist nur bedingt möglich. An der Stelle werden bei der Modellierung unterschiedliche Approximationen verwendet.

### 2.4.2 Wartezeiten

Die meisten Halbleiterhersteller arbeiten mit Werkzeugen, welche die Warteschlange vor der Anlage bzw. vor Anlagengruppe mit entsprechenden Abfertigungsregeln<sup>10</sup> managen [Atherton95]. Der Zufall tritt an dieser Stelle vor allem in Form der Wartezeit bis zur Prozessierung auf. Angenommen die Spezifikation der Abfertigungsregel ist bekannt, so ist auch bekannt, an welcher Stelle das ankommende Los in die Warteliste eingefügt wird. Da aber die Prozesszeit der Wafer, die noch bearbeitet werden müssen, auch mit Zufall behaftet ist, überträgt sich das auf die wartenden Werkstücke. Weiterhin kommt auch in unvorhersehbaren Zeitpunkten neues Material hinzu, was die Länge der Warteschlange und somit auch die Wartezeiten deutlich vergrößern kann. Weitere typische Probleme, wie der Ausfall von Maschinen und Prozessfehler, können an dieser Stelle ebenfalls nicht außer Acht gelassen werden.

### 2.4.3 Häufigkeit und Dauer von Ausfällen und vorbeugender Wartung

Maschinenausfälle sind Vorzeigebeispiele für zufällig auftretende Probleme. Die Anlagen in der Halbleiterfertigung arbeiten im Nanometerbereich und sind sehr kompliziert aufgebaut. Aus diesem Grund sind sie sehr teuer in der Anschaffung und müssen entsprechend hoch ausgelastet werden. Die beschriebenen Ursachen führen in unvorhersehbaren Zeitabständen zu Ausfällen. Sowohl komplette als auch partielle Ausfälle sind möglich. Maschinen, die bestimmte Mängel haben, können nur noch für Teilprozesse in der Produktion freigegeben werden oder fertigen mit geringerer Geschwindigkeit. Im Fall des Ausfalls von Steuerungseinheiten wird jedoch nur die Abarbeitung, durch manuelle Ansteuerung des Tools, verlangsamt.

---

<sup>10</sup> Engl.: dispatch rule – Regel die zuständig ist für die Anordnung der Warteschlange. Das einfachste Verfahren heißt FiFo (First in, First out), also das Werkstück welches zuerst kommt, wird zuerst bearbeitet. Die Abarbeitungsregeln ermöglichen die Priorisierung und Steuerung anhand des Liefertermins. Ausführliche Erklärung in [Atherton95]

Anschließend erfolgt die Reparatur, deren Dauer ebenfalls schwer zu prognostizieren ist. Da keine gültigen Voraussagen über Ausfall bestimmter Bauteile gemacht werden können und Fachkräfte nicht immer verfügbar sind, kommt es zu zeitlichen Schwankungen in der Instandsetzungsdauer.

Die Häufigkeit der vorbeugenden Wartung der Anlagen kann durch zwei unterschiedliche Parameter bestimmt werden. Es kann nach zeitlichen Vorgaben oder nach Anzahl der produzierten Masse gewartet werden. Die Werte sind meistens konstant und Schwankungen erfolgen erst durch einen Wartungsspielraum. Der Wartungszeitpunkt befindet sich in einem Zeitfenster, somit ist der definitive Anfangszeitpunkt der Wartung variabel. Dies ist vor allem deshalb sinnvoll, da zu jedem Zeitpunkt unvorhergesehene Ereignisse stattfinden können, die eine standardisierte Maßnahme um einen unbestimmten Zeitabschnitt verlängern.

### **2.4.4 Hold-Zustand**

Der Begriff „Hold“ stammt aus dem Englischen und bedeutet „der Halt“. Ein Los kann sich in einem Hold-Zustand befinden, das heißt, wenn aus verschiedenen Gründen die Abarbeitung angehalten und ein Werkstück speziell markiert wurde. Die Benutzung solcher oder ähnlicher Markierungen ist für die Halbleiterfertigung üblich. Vor allem aus Gesichtspunkten der Prozessstabilität ist es sinnvoll, problembehaftete Werkstücke bereits beim Betreten eines Produktionsschritts zu markieren, um diese von einer Fachkraft bearbeiten zu lassen.

Es gibt mehrere Ursachen für den Hold-Zustand:

- Prozessfehler oder -abbruch,
- Zwang der Abarbeitung durch Fachkraft, meistens bei Entwicklungsprodukten,
- Problematische Abarbeitung bei vorherigen Operationen,
- Einführung von neuen Anlagen bzw. Prozessen,
- Unterschiedliche IT-Maßnahmen.

Wenn ein Los in den Hold-Zustand übergeht, treten die Bedingungen der Stochastik an zwei Stellen auf. Zum einen erfolgt die Auswahl des Werkstückes zufällig und zum anderen ist die Dauer der Situation willkürlich.

Die den Hold-Zustand betreffende Stochastik ist sehr schwierig zu beschreiben, da die Variation an Parametern hoch ist. Weiterhin fehlen zumeist Informationen vergangener Vorgänge, welche die Erstellung eines statistisch relevanten Modells überhaupt erst möglich machen würden.

### 2.4.5 Sampling

Das Wort „Sampling“ stammt von der englischen Sprache *sample* und bedeutet Stichprobe oder Auswahl. Möglich ist entweder das Los-Sampling, Auslassen einer Operation oder das Wafer-Sampling in Form der Abarbeitung einer bestimmten Anzahl an Siliziumscheiben. Die stichprobenartige Bearbeitung in der Halbleiterfertigung hat sich bei allen Arten von Messprozessen etabliert [Nurani96]. Sie erfolgen meistens in vorher festgelegten, produktspezifischen Intervallen:

- Zeitliches-Sampling – Prozessierung erfolgt immer am Anfang einer neuer Zeitperiode, danach überspringen alle im beobachteten Zeitabschnitt kommenden Lose die Operation, z. B.: Messung der Parameter alle 8 Stunden,
- Mengen-Sampling – Prozessierung erfolgt durch Erreichen einer bestimmten Stelle in einem Loszähler, es bedeutet immer nur der n-te Los in Folge wird bearbeitet,
- Misch-Sampling – Prozessierung erfolgt mittels Zähler wie bei Mengen-Sampling, jedoch werden durch das Einfügen einer Zeitperiode Prozessfehler ausgeschlossen bei Produkten die selten ankommen, z. B.: jeder zehnte Los jedoch nicht seltener als 8 Stunden
- Attribut-basiertes-Sampling – es werden nur die Lose prozessiert, die über ein spezielles Attribut verfügen.

Die Stochastik, die an dieser Stelle vorzufinden ist, hängt vor allem mit den Zufallswerten bei Abarbeitung von den vorherigen Operationen zusammen. Die Aussage über ein bestimmtes Los, ob die in der Zukunft liegende Prozessierung an der Sampling-Operation durchgeführt wird, ist zufallsbehaftet. Es ist vor allem wichtig bei der Anlieferungsprognose, wenn eine Vorhersagezieloperation nur stichprobenartig angeliefert wird.

### 2.4.6 Lager-Ausschleusung

Die Einführung von Zwischenlagern ermöglicht eine schnellere Reaktion auf Kundenanfragen. Der Bedarf an Material ist natürlich nicht konstant, sondern hängt von mehreren wirtschaftlichen Einflussgrößen ab (vgl. [Scholl08]). Die Ermittlung von Zeitpunkt und Menge der Bestellung ist außerdem durch ihre Zufälligkeit sehr kompliziert. Oft werden Trendprognosen als Nachfrageapproximation verwendet.

### 2.4.7 Übergang zwischen Fertigungs- und Testbereich

Der Übergang zwischen den einzelnen Fertigungsbereichen ist fabrikspezifisch. Viele Fertigungsstätten verfügen über keine Test- bzw. Backendbereiche, sondern produzieren nur Frontendschritte. Um die stochastischen Einflüsse zu veranschaulichen, wurde dazu ein Beispiel entworfen, das in Abbildung 2-3 dargestellt ist. Bereits mehrfach erwähnt wurde, dass die Produktion in der Halbleiterbranche außergewöhnlich nachfrageorientiert ist.

Ausgenommen sind meistens nur Massenprodukte, wie zum Beispiel Speicherbausteine. Um kundengerecht zu fertigen ist einerseits die Qualität und andererseits die Lieferzeit wichtig. Aus diesem Grund kann ein fertiger Wafer verschiedene Teststrecken absolvieren. Bei einer nicht ausreichenden Kapazität der Messbereiche wird lediglich ein bestimmter Teil der Lose eines Produktes vollständig getestet bzw. die Ware wird an ein anderes Werk verschickt, welches die notwendigen Aufgaben schneller erledigen kann. An dieser Stelle ist es angemessen, ein Zwischenlager einzuführen, das die Kapazitäts- und Nachfrageschwankungen verringert. Die Produkteinlagerung erfolgt entweder vor oder nach dem Funktionstest. Es ist möglich, ein Bereichsübergangsmodell aufzubauen. Dabei sind jedoch stochastische Einflüsse zu beachten, wie beispielsweise Anlagenausfälle im Testbereich, unvorhersehbare Kundennachfrage und Frontend-Ausschleusung. Das Systemverhalten wird also durch ein deutliches, von Zufall behaftetes Rauschen verfälscht.

## 2.5 Anforderungen an das Vorhersagewerkzeug

Bei der Beschreibung der notwendigen Voraussetzungen des Vorhersagewerkzeug werden die Grundkonzepte des modernen Softwareengineering verwendet [Sommerv01]. Dementsprechend können die Anforderungen in drei Kategorien geteilt werden:

- Funktionale Anforderungen,
- Nichtfunktionale Anforderungen,
- Problembereichsanforderungen.

### 2.5.1 Funktionale Anforderungen

Die funktionalen Anforderungen beziehen sich ausschließlich auf die Funktionalität des Systems – also auf den konkreten Bereich einer geplanten Problemlösung [Brugger05]. Sie beschreiben markante Interaktionen, welche die Software dem Anwender zur Verfügung stellen soll bzw. die Aktionen, die von einem System ausgeführt werden sollen.

Im Rahmen eines Projekts mit dem Industriepartner wurden die funktionalen Anforderungen detailliert erarbeitet. Die Hauptfunktion der Software ist die Erstellung einer in Wafer skalieren Anlieferungsvorhersage an bestimmten Schlüsselstellen der Fabrik. Die Prognose muss einen 7 bis 14-tägigen Zeitraum abdecken. Die Auswahl des Vorhersagezieles soll flexibel anpassbar sein und entweder über einen Punkt in der Route oder auch durch Zugehörigkeit zu einer Anlagengruppe spezifiziert werden. Die Funktionsweise des Werkzeugs muss auch ein Frühwarnsystem für drohende Kapazitätsengpässe umfassen. Die erstellte Software soll mindestens dreimal täglich eine aktualisierte Prognose liefern. Die Zeitspanne zwischen Fragestellung und entsprechender Antwort soll so gering wie möglich ausfallen. Weiterhin wurden auch die Angaben zur Visualisierung von Ergebnissen charakterisiert. Als Letztes ist zu erwähnen, dass die Prognoseresultate auch von nicht fachmännischem Personal benutzt werden. Sie sind daher in selbsterklärender Form zu präsentieren.

### **2.5.2 Nichtfunktionale Anforderungen**

Während die Bedeutung der funktionalen Anforderungen für die Kunden offensichtlich ist und akzeptiert wird, wird die Spezifikation von nichtfunktionalen oft vernachlässigt [Störrle05]. Nichtfunktionale Anforderungen beinhalten die Vorschriften zur Realisierung der Softwarefunktionen sowie die Voraussetzungen, welche an das System gestellt werden. Da sie ein wichtiger Bestandteil jedes Softwareprojektes sind ([Störrle05] spricht von ca. 80 %), werden sie in der vorliegenden Ausarbeitung näher beleuchtet.

Die nichtfunktionalen Anforderungen werden in drei Gruppen geteilt [Kotonya98] (siehe Abbildung 2-4). Die erste Gruppe beinhaltet die Entwicklungsprozessnormen. Die Wahl der verwendeten Programmiersprachen, Datenbanken und der Entwicklungsumgebung hat sich aus den industriellen Standards der Projektpartner ergeben. Erlaubt wurden also C++, C#, Oracle-Datenbanken sowie auch jede frei verfügbare Software. Die Entwicklungszeit wurde durch den gesetzten Projektrahmen begrenzt. Die zweite Gruppe der nichtfunktionalen Anforderungen beschrieb die Anforderungen an das Produkt selbst. Der wichtigste Gegenstand des Projektes, die Neuentwicklung einer Vorhersagemethode, wurde vollständig dem Entwicklungsteam überlassen. Da es sich bei dem Vorhaben hauptsächlich um die Entwicklung einer neuer Vorhersagemethode und anschließende Implementierung handelt, spielte das Thema Stabilität, Robustheit, Verfügbarkeit und Effizienz eine wichtige Rolle. Es wurden keine direkten Begrenzungen an die Größe der Software gestellt. Die Dimension, der zu entwickelten Software, sollte eine benutzerfreundliche Reaktionszeit des Systems bei der Ausführung auf einem Standard Desktoprechner gewährleisten. Die Daten- und Systemintegrität, also die Fä-

higkeit der Software, Daten und Systemfehler zu identifizieren, zu beheben bzw. den Administrator zu informieren, wurde als Hauptmerkmal für die einfache Wartung vorgeschrieben. Themen, die keinem der beiden vorhergehenden Punkte zugeordnet werden konnten, hauptsächlich die Kostenfrage, wurden in der letzten Gruppe der nichtfunktionalen Anforderungen zusammengefasst. Dem Entwicklerteam wurden sowohl gängige Entwicklungswerkzeuge als auch zusätzlich verfügbare Software kostenlos zur Verfügung gestellt. Jedoch wurden keine weiteren Investitionen in kommerzielle Software erlaubt.

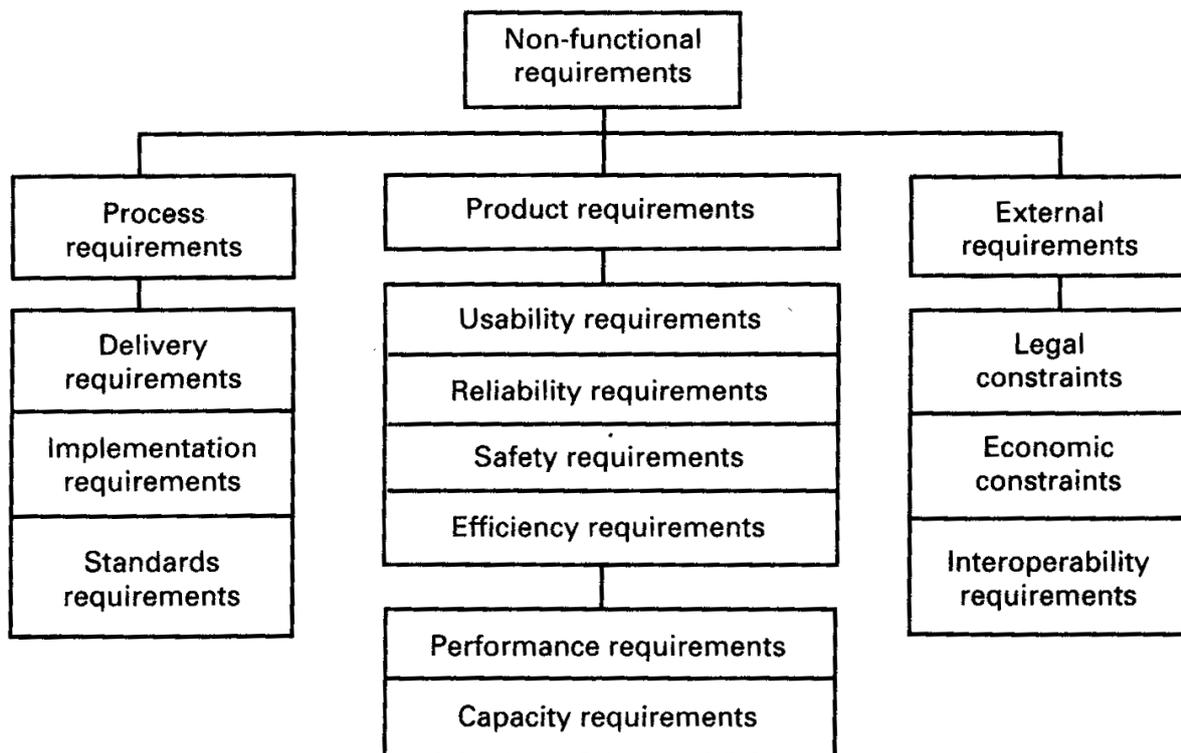


Abbildung 2-4: Nichtfunktionale Anforderungen – Quelle: [Kotonya98]

### 2.5.3 Problembereichsanforderungen

Die Problembereichsanforderungen nach [Uslar07] ergeben sich erst aus der Gesamtheit des Systems und nicht nur aus seinen Komponenten. Sie charakterisieren die Methodik und Implementierung mithilfe von sowohl funktionalen als auch nichtfunktionalen Anforderungen. Die Charakteristika und Einschränkungen des Systems basieren hauptsächlich auf Problematiken, die für die Halbleiterherstellung charakteristisch sind. Damit ist die enorme Komplexität des Systems gemeint, die keine direkte Abbildung in einem Modell ermöglicht. Die Anwendung von praktisch jeder Vorhersagemethode auf ein derartiges Problem zwingt den Entwickler, eine Reihe an Approximationsansätzen zu implementieren. Ein weiterer Nachteil bei der Modellierung einer dermaßen komplexen Fertigung ist das Übermaß, die Vielschichtig-

## **Problemstellung**

---

keit und die Fehleranfälligkeit der Basisdaten. Diese sollen also dementsprechend aufbereitet und korrigiert werden.

Nachdem in Kapitel 2 die Problemstellung erörtert wurde, und die Anforderungen an die technische Lösung definiert wurden, werden nun in Kapitel 3 die bestehenden Prognoseansätze analysiert. Der Fokus der Analyse liegt hauptsächlich bei der Eignung der Methoden für die Prognosen in der Halbleiterfertigung.

### 3 Analyse bestehender Prognoseansätze

Prognosen sind immer dann schwierig,  
wenn sie sich auf die Zukunft beziehen.

---

(Mark Twain)

Der Wunsch des Menschen, die Zukunft zu kennen, ist so alt wie die Menschheit selbst. Es gibt also einen guten Grund für viele Bereiche der Wissenschaft, sich mit der Thematik des Prognostizierens auseinanderzusetzen. Das Wort Prognose stammt aus der griechischen Sprache und bedeutet „das Vorwissen“ oder „die Voraus-Kenntnis“ – also die Aussage über zukünftige Ereignisse, Zustände und deren Entwicklung. Die Prognosebasis besteht aus Fakten, die mit formalisierten Methoden erhoben werden. Die Daten werden als Prädikatoren bezeichnet und dienen zur Erarbeitung einer Wissensbasis über ein System. Mit dem Erfahrungswissen und deren Extrapolation können anschließend – mit einer bestimmten Wahrscheinlichkeit – die Voraussagen getroffen werden.

#### 3.1 Theoretische Grundlagen von Prognosetechniken

##### 3.1.1 Eigenschaften von Prognosetechniken

Die einfachste bekannte Prognosetechnik ist die Naive Prognose, die sich in zwei Typen gliedert:

- Typ 1 (No-Change-Prognose) – verwendet den letzten gültigen Datenpunkt als Prognosewert,
- Typ 2 (Same-Change-Prognose) – bei der der ermittelte Trend aus dem letzten und vorletzten Wert einfach in die Zukunft fortgesetzt wird.

Die Naive Prognose wird als Qualitätsmaß anderer komplizierter Prognosealgorithmen verwendet. Die verschiedenen komplexeren Prognosetechniken unterscheiden sich bezüglich ihres zeitlichen Horizonts, der sich in kurz, mittel und langfristige Voraussagen einteilen lässt. Die Länge der Zeitperiode ist vom Prognoseproblem abhängig. Während die kurzfristige Wettervorhersage für Flugzeugpiloten nur die nächsten 1 bis 2 Stunden umfasst, erstreckt sich die kurzzeitige Voraussage der Zunahme von Biomasse eines Waldes über Wochen bis Monate. Das Prognoseergebnis beschreibt entweder eine Variable (Einzelprognose) oder ein System

mit mehreren Veränderlichen (Prognosesystem). Ein weiteres beschreibendes Kriterium ist die Art der Berechnung von Resultaten [Martens04]. Es gibt zwei Möglichkeiten hiervon:

- Top-Down – es wird eine gesamte Systemantwort berechnet und daraus resultieren die Antworten für die Teilsysteme. Die Anwendung derartiger Techniken ist nur im Falle von stabilem Verhalten des geteilten Systems sinnvoll.
- Bottom-Up – es werden zuerst die Teilsysteme untersucht, um die Systemantwort durch Aggregation zu ermitteln. Die Techniken sind durch die wesentlich größere Anzahl von Rechenschritten wesentlich komplizierter.

Die Prognosetechniken werden außerdem anhand des Berechnungsprozesses und dessen Ergebnis als qualitativ oder quantitativ bezeichnet [Martens04]. Die qualitativen Vorhersage-techniken basieren hauptsächlich auf dem Fachwissen von Experten, Meinungsabfragen und Lebenszyklusanalysen. Selten werden einfache Trends oder lineare Extrapolation in die Zukunft projiziert. Da die Ergebnisse derartiger Voraussagen wenig konkrete Zahlenergebnisse liefern, sind quantitative Techniken für die Industrie wesentlich interessanter. Derartige Prädiktionen wenden verschiedene, statistische Methoden an, um das Datenmaterial aufzubereiten. Anschließend wird mithilfe von Rechenverfahren ein zahlenmäßiges Resultat erstellt. Wenn die vorauszusagenden Größen von einstellbaren Variablen abhängen, wird die Prognose als Wirkungsprognose bezeichnet. Regressionsanalyse und GAP-Analyse werden als typische Beispiele derartiger Prädiktion angegeben. Anders sind die Entwicklungsprognosen, auch Informations- oder Trendprognosen genannt, welche die notwendigen Informationen aus der Umgebung gewinnen. Dabei spielt die eigentliche Unternehmung eine untergeordnete Rolle. Ein klassisches Beispiel dieser Art der Vorhersage ist für ein Unternehmen die Marktentwicklung der Verkaufsgüter. Die Voraussagegewinnung erfolgt zum Beispiel mittels Trend-, Saisonextrapolation, Glättungsverfahren oder autoregressiven Verfahren [Neusser06].

### 3.1.2 Prognosefehler und -bewertung

Trotz aller Versuche, die Vorhersagen technisch zu verbessern und zu korrigieren, entstehen immer gewisse Abweichungen zwischen der Vorhersage und den tatsächlich eingetretenen Ereignissen/Ergebnissen. Die Differenz zwischen Vorhersage und Ergebnis wird als Prognosefehler bezeichnet [Martens04].

Die Inkorrektheit der Vorhersage lässt sich erst nach dem Eintreten von untersuchten Vorkommnissen feststellen. Bei qualitativen Prognosetechniken sind Fehler zumeist sehr schwer quantifizierbar, deshalb ist die Prognosegenauigkeit auch nicht trivial zu ermitteln. Die Sys-

temexperten, die qualitativ prognostizieren, müssen sich deshalb auf typische Fehlerursachen konzentrieren [Sullivan77]. Die repräsentativen Fehlergründe liegen oft in falscher Problembeschreibung und deren Modellierung, wie zum Beispiel:

- Überbewertung von jüngeren Ereignissen, obwohl diese oft von natürlichem Verhalten abweichen,
- Erkennung und Gebrauch von Mustern, die jedoch in Wahrheit nicht existieren,
- Überbewertung von allgemein bekannten und populären Ereignissen, die in das Experiment einfließen,
- Problematik von Gedächtnis und Denkweise der Experten. Überbewertung von außergewöhnlichen Ereignissen gegenüber dem typischen Systemverhalten,
- großer Einfluss von Angst- und Wunschkonstruktionen.

Wesentlich interessanter sind quantitative Prognosetechniken, bei denen sich die Prognosegenauigkeit unkompliziert rechnerisch ermitteln lässt. Die Ursachen von Fehlerentstehung einer Prognose lassen sich oft in eine der vier Kategorien zuordnen:

- falscher Modellierungsansatz oder falsches Modell – ungenügende Kenntnisse über ein Problem bzw. Minderwertigkeit der Basisdaten,
- Strukturbrüche – inkonstanter Verlauf von Regressionsparametern über die gesamte statistischen Zeitreihe, mehr in [Zeileis03],
- Störvariablen – können sowohl als abhängige und unabhängige Variablen die Durchführung eines Experiments mit Rauschen beeinflussen [Schnell98],
- Schätzfehler – auch Stichprobenfehler oder Zufallsfehler – entstehen durch das Einbetten einer Zufallsvariable in die Berechnung eines Schätzwerts [Hartung95].

Bei der Modellierung sollten die beiden ersten Fehlerquellen möglichst vermieden werden. Aufgrund dessen ist eine detaillierte Analyse der Ausgangssituation und Systemzusammenhänge notwendig. Der Einfluss von Störquellen und Schätzfehlern auf das Vorhersageresultat ist unvermeidbar. Bei der Prognosebildung werden jedoch verschiedene Techniken angewendet, um die Störfaktoren zu kontrollieren [Hartung95], wie zum Beispiel:

- Randomisierung,
- Messwiederholung,
- Blockbildung,
- Eliminierung,

- Abschirmung,
- Konstanthaltung,
- Umwandlung von Störvariablen in unabhängige Variablen.

Für die quantitativen Prognosetechniken wurden auch mehrere Techniken zur Bewertung der Genauigkeit entworfen. Die gängigsten Verfahren und Formeln zur Messung des Prognosefehlers werden wie folgt aufgebaut [Buttler02] [Krengel05]:

- Mittlere absolute Abweichung (engl. mean absolute deviation – MAD) – wird aus der Summe der absoluten Fehler zwischen realem  $v_r(i)$  und prognostiziertem Wert  $v_f(i)$  geteilt durch die Anzahl der Stichproben  $n$  errechnet,

$$MAD = \frac{1}{n} \sum_{i=1}^n (|v_r(i) - v_f(i)|)$$

- Mittlerer absoluter prozentualer Fehler (engl. mean absolute percentage error – MAPE) – wird aus Summe der absoluten prozentualen Fehler zwischen realem  $v_r(i)$  und prognostiziertem Wert  $v_f(i)$ , geteilt durch Anzahl der Stichproben  $n$  errechnet,

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{v_r(i) - v_f(i)}{v_r(i)} * 100 \right|$$

- Mittlerer quadratischer Fehler (engl. mean square error MSE) – wird aus der Summe der Quadratpotenz der Subtraktion des prognostizierten Wertes  $v_f(i)$  von dem realen Wert  $v_r(i)$  geteilt durch Anzahl der Stichproben  $n$  errechnet,

$$MSE = \frac{1}{n} \sum_{i=1}^n (v_r(i) - v_f(i))^2$$

- Quadratwurzel des mittleren quadratischen Fehlers (engl. square root of mean square error – RMSE) – wird mittels Quadratwurzel aus der Summe der Quadratpotenz der Subtraktion des prognostizierten Wertes  $v_f(i)$  von dem realen Wert  $v_r(i)$  geteilt durch die Anzahl von Stichproben  $n$  errechnet,

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (v_r(i) - v_f(i))^2}$$

- Der mittlere prozentuale Fehler (engl. mean percentage error – MPE) – wird aus der Summe der prozentualen Fehler zwischen realem  $v_r(i)$  und prognostiziertem Wert  $v_f(i)$  geteilt durch Anzahl der Stichproben  $n$  errechnet.

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{v_r(i) - v_f(i)}{v_r(i)} * 100$$

Aufgrund der Häufigkeit von Störungen kommt es häufig in den untersuchten Zahlenreihen zu außergewöhnlichen Abweichungen im Verlauf eines Messwertes. Diese Datenpunkte werden Ausreißer genannt. Da derartige Werte die Mittelwertbildung zum Teil extrem beeinflussen können, wird auch oft die Median- oder  $p$ -Quantil-Funktion anstatt des gewöhnlichen Mittelwertes verwendet. Es existieren natürlich auch eine große Menge an weiteren, mehr oder weniger problemspezifischen Tests. Diese werden an dieser Stelle jedoch nicht näher erläutert. Die Problematik der Ermittlung des Prognosefehlers für eine Anlieferungsvorhersage in einem Halbleiterwerk (Fertigung auf Job-Shop-Prinzip) wird in den Kapitel 6 und 7 näher beschrieben und erläutert.

### 3.1.3 Entwicklung und Bedeutung von Prognosen

Die ersten, rechnergestützten Prognosen entstanden gleich nach Aufkommen der ersten Rechenanlagen. Die Anfänge waren jedoch nicht einfach. Charakteristisch für diese Zeit waren Redewendungen und Versprechungen wie: „*Die Computer haben vorhergesagt, dass im Jahr 2000...*“ [Martens04]. Die Genauigkeit der computergestützten Vorhersagen war aber bei Weitem nicht so zufriedenstellend wie die Resultate, die mit bekannten und natürlichen Methoden erzielt wurden. Trotz anfänglicher Schwierigkeiten wurden immer genauere Modelle entwickelt, die es ermöglichten, genauere Prognosen zu machen. Die Komplexität in der Methodik stieg proportional zum Fortschritt in der Computertechnik und zum Speicherplatzangebot.

Die ständige Erweiterung der Produktionssysteme und die Verfügbarkeit immer größerer Mengen an Basisdaten erschwert die Wartbarkeit und Aktualität von Prognosemodellen enorm. Martens [04] schreibt „*Nicht jedes Verfahren ist für jede reale Konstellation geeignet. Aus diesem Grunde erhalten leistungsfähige Prognosesysteme mittlerweile eine größere Anzahl alternativer Algorithmen. Allerdings erfordern Auswahl und Parametrierung meist mathematische Experten, was in der betrieblichen Praxis zu einem Akzeptanzproblem geführt hat.*“ Die Lösung für dieses Problem ist das Einfügen von einer Art Künstlicher Intelligenz in die Vorhersagealgorithmen. Dies erfolgt entweder mithilfe von lernfähigen Selbstvalidie-

rungsalgorithmen oder durch intelligente Filterung und Beobachtung von relevanten Parametern.

Die Anwendung von Vorhersagesystemen in der Industrie ist eng mit betrieblicher Planung gekoppelt und bringt sehr viele Vorteile mit sich. Die Aufzählung aller positiven Aspekte der Anwendung einer mehr oder weniger exakten Prognose, ist sehr mit dem problemspezifischen Prognoseziel und dem Anwendungsfeld verbunden. Infolgedessen wird nur ein kleiner Anteil des breiten Anwendungsspektrums, der für die Halbleiterindustrie relevant ist, in der vorliegenden Ausarbeitung beschrieben. In der heutigen Marktwirtschaft ist die Kostenreduktion ein wichtiger Schlüssel zum Erfolg. Die Minimierung von Ausgaben geschieht durch vorhergesagte Planung hinsichtlich:

- Personal,
- Ressourcen,
- Materialbeschaffung,
- Liefertermineinhaltung,
- momentaner und maximaler Kapazität für einen bestimmten Produktmix,
- vorbeugende Wartung usw.

Die Prognosesysteme ermöglichen außerdem die Erkennung von Markttrends und Anfrageschwankungen, was ein besseres Risikocontrolling ermöglicht.

### 3.2 Simulation

Simulationen eröffnen einen Zugang zu den Bereichen, die für herkömmliche Experimente zu klein oder zu groß, zu schnell oder zu langsam, zu gefährlich oder zu teuer sind

---

(Dr. Gabriele Gramelsberger [Grame08])

Der Begriff Simulation stammt aus dem lateinischen Wort „*simulare*“ und bedeutet „nachbilden“ oder „vortäuschen“. Eine Nachbildung wird immer dann benötigt, wenn das untersuchte Problem sich aus Kosten- und Komplexitätsgründen nicht direkt abbilden lässt bzw. wenn die theoretische oder formelmäßige Behandlung zu kompliziert ist [Domsch02]. Anfänglich wurde die Simulation in rein militärischen Bereichen genutzt. Sven Jensen schreibt [Jensen07] „*Es liegt in der Natur des Menschen, Dinge zu hinterfragen und nach Antworten zu suchen.*“ Aus diesem Grund entwickelte sich die Simulation zu einem interdisziplinären Instrument

von Wissenschaft und Praxis. Mittlerweile ist sie nicht mehr wegzudenken, um komplexe Fragestellungen der Physik, des Maschinenbaus, der Chemie oder der Industrie zu beantworten [Backes96].

### 3.2.1 Geschichte und Definition

Die Bezeichnung Simulation umfasst sowohl physikalische Experimente, die real nachgestellt werden, als auch eine rechnergestützte Systemabbildung. Als ein charakteristisches Beispiel für eine computerfreie Simulation kann ein Auto-Crashtest genannt werden. Aus Sicht der Informatik sind jedoch Computersimulationen wesentlich interessanter. Die Geschichte der Rechnersimulation fängt mit dem Manhattan-Projekt, das die Entwicklung einer Atombombe in den USA zum Ziel hatte, an [Jensen07]. Der nächste größere Schritt in Richtung heutiger Simulationswerkzeuge wurde bei IBM von Joffrey Gordon 1960 unternommen. Dort wurde das erste universelle Simulationssystem, genannt General Purpose Simulation System (GPSS), entwickelt [Gordon61]. Die weitere Entwicklung der Simulation erfolgt parallel zum Fortschritt in der Computertechnik. In den späten 70er Jahren war die Rechentechnik so ausgereift, dass erste Versuche der Visualisierung von Simulationen unternommen werden konnten. 1977 wurde, auf einer Tagung in München, einer der ersten grafisch-interaktiven Simulatoren namens SIMFLEX vorgestellt [Reinha03]. Mittlerweile waren mehrere Simulationssprachen und Simulationswerkzeuge entwickelt worden. Die Wahl der passenden Softwarelösung ist eng verbunden mit dem eigentlichen Anwendungsfall und dem erwarteten Resultat. Die Anwendung der Computersimulation in der Halbleiterindustrie lässt sich in drei Gruppen einteilen:

- Simulation von physikalischen Prozessen,
- Simulation von einzelnen Anlagen und Anlagengruppen,
- Simulation von den ganzen Halbleiterwerken.

Die Prozesssimulation fängt bereits in den 60er Jahren an, die eigentliche Evolution findet jedoch eine Dekade später statt [Cole90]. Die ersten einfachen Modelle, die nur analytische Gleichungen verwendeten, wurden durch komplexere, numerische Modelle ersetzt.

Der Erste 1-D Simulator SUPREM I wurde 1977 vorgestellt [Cole90]. Kurz darauf, in den früheren 80er Jahren, entstand auch die graphische 2-D Software, die bis in die heutige Zeit weiterentwickelt wird. Die Simulation von Anlagen und gesamten Fabriken ist miteinander eng verbunden. In komplexen Fertigungsmodellen werden somit also mehr oder weniger detaillierte Anlagenmodelle verwendet. Die Pioniere der Fabrik- und Anlagenmodellierung sind

größere Konzerne wie IBM oder Intel, die früh die Vorteile des Simulierens erkannt haben. Der Durchbruch kam aber erst mit dem Anlaufen der Massenproduktion. Simulation wurde zuerst für die Planung von Fabriklayouts mit Erfolg angewendet. Anschließend wurden die besten Möglichkeiten, Fertigungsstätten in Betrieb zu nehmen simuliert. Schließlich, nachdem die Produktion stabil erfolgte, unterstützte die Computersimulation mit Prognosen die Planungsabteilungen bei strategischen Entscheidungen (vgl. [Kempf96]). Aktuell verwenden aus Kostengründen die meisten Halbleiterhersteller speziell für diese Fertigungsart entwickelte kommerzielle Software. Für die Halbleiterindustrie spielt eher die Problematik der Modellentwicklung, Wartung und Validierung eine wesentlich wichtigere Rolle als die Implementierung der Simulationssoftware selbst.

Eine der einfachsten Definitionen von Simulation finden man in der Brockhaus Enzyklopädie [Brockh02]:

*„Simulation – die modellhafte Darstellung oder Nachbildung bestimmter Aspekte eines vorhandenen oder zu entwickelnden [...] Systems [...], insbesondere auch seines Zeitverhaltens“.*

Der Verein Deutscher Ingenieure definiert die Simulation in der VDI-Richtlinie 3633 wie folgt:

*„Simulation ist das Nachbilden eines Systems mit seinen dynamischen Prozessen in einem experimentierfähigen Modell, um zu Erkenntnissen zu gelangen, die auf die Wirklichkeit übertragbar sind.“* [VDI05].

Diese Definition erfuhr auch Kritik [Bauma96], da die Simulation nicht das System an sich nachbildet, sondern nur das Systemverhalten. Die Nachbildung eines Systems wird auch Modellierung genannt. Der Kern jeder erfolgreichen Simulation ist ein detailreiches Modell, was auch in der VDI-Richtlinie 3633 spezifiziert ist.

*„Ein Modell ist eine vereinfachte Nachbildung eines geplanten oder real existierenden Originasystems und -prozesses in einem anderen begrifflichen oder gegenständlichen System. Es unterscheidet sich hinsichtlich der untersuchungsrelevanten Eigenschaften nur innerhalb eines vom Untersuchungsziel abhängigen Toleranzrahmens vom Vorbild“* [VDI05].

Wichtig ist ebenso die Erläuterung des Begriffs „System“, der eine abgegrenzte Anordnung an Komponenten, die untereinander interagieren, bedeutet [Jensen07]. Die bisherigen Definitionen koppeln weder Simulation noch Modell an einen bestimmten Anwendungsbereich. Da auch der heutige Stand der Technik fast keine Grenzen in der Modellbildung setzt, lassen sich

die Anwendungsgebiete der Modellierung und Simulation nach Kosturiak [Kostur95] in fünf Gruppen unterteilen:

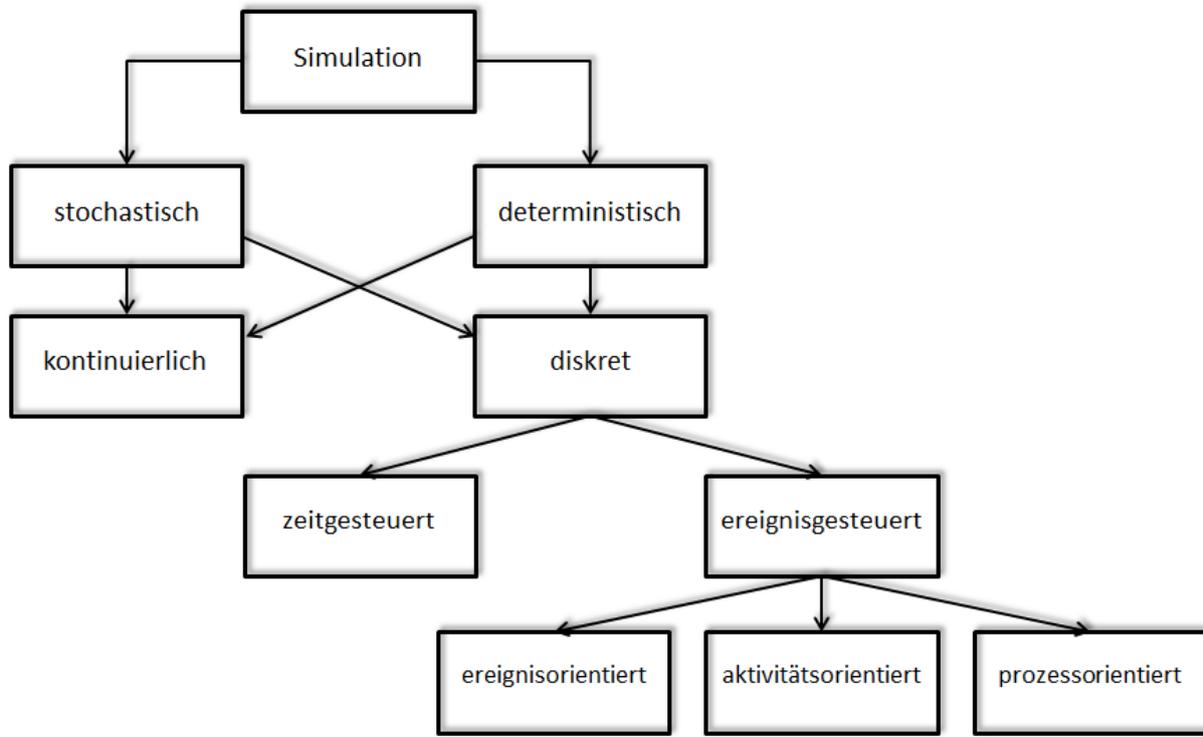
- Entscheidungsunterstützung – z. B. Erleichterung einer Entscheidung durch Prognose von deren Folgen – Auswirkung von Produktmix auf die Fertigung,
- Planung – z. B. Vereinfachung von Planung durch Vergleich von Auswirkungen einzelner, geplanter Szenarios – Fabrikerweiterung um neue Anlagen,
- Forschung und Entwicklung – z. B. Unterstützung der Forschung im Maschinenbau ohne kostspieligen Prototypenbau mittels Simulation der Materialermüdung,
- Organisationsgestaltung – z. B. Bewertung von Nebeneffekten nach Einführung neuer organisatorischer Maßnahmen,
- Ausbildung und Training – z. B. Unterstützung von Formel 1 Piloten durch Fahrt in Simulatoren an trainingsfreien Tagen.

In der Ausarbeitung wird insbesondere ein rechnerkonformes, detailliertes Simulationsmodell der Halbleiterfertigungsfabrik, mit allen Vor- und Nachteilen, charakterisiert. Auch die Resultate eines Simulationsverlaufs werden interpretiert und mit anderen Methoden verglichen.

### 3.2.2 Simulationsarten

Wie bereits in den vorherigen Kapiteln beschrieben, ist die Computersimulation an ein sehr breites Problem- und Anwenderspektrum adressiert. Infolgedessen existieren diverse Varianten der Modellierungs- und Simulationsdurchführung. Aufgrund der Thematik der vorliegenden Ausarbeitung wird der Anwendungsbereich auf die Produktion und teilweise auf die Logistik eingegrenzt. Die Abbildung 3-1 zeigt eine Klassifizierung der Simulation [Backes96] anhand der spezifischen Merkmale. Die erste Ordnung der Simulationsklassifizierung ergibt sich aus dem verwendeten Datenmaterial. Wenn die notwendigen quantitativen Informationen in der deterministischen Form vorliegen bzw. wenn es möglich ist, diese als solche zu behandeln, wird die Simulation als deterministisch eingruppiert. Sobald aber einer der notwendigen Werte zufallsabhängig ist, gilt die Computersimulation als stochastisch. Das zweite, quasi parallele Gruppierungskriterium resultiert aus der Zeitsteuerung der Simulationssoftware [Domsch02]. Bei der kontinuierlichen Simulation werden meistens mittels Differentialgleichungen die stetigen Prozesse abgebildet [Frauens90]. Für diese Arbeit interessanter und weiter verbreitet ist jedoch die diskrete Variante, in welcher der Simulator Zeit in Intervalle aufteilt und die Änderungen des Modellzustandes schrittweise erfolgen. Der Ansatz lässt keine Zustandsänderungen zwischen zwei Zeitpunkten zu [Page05]. Durch die zeitliche Abfolge

von allen Zustandsänderungen wird der Ablauf einer diskreten Simulation jedoch komplett beschrieben [Fishman01]. Die simulierte Zeitspanne kann keine Voraussage über den realen Zeitverlauf geben.



**Abbildung 3-1: Klassifizierung der Simulation im Bereich Produktion und Logistik**

Manche Simulatoren mit Visualisierungsmodus lassen jedoch die Ablaufgeschwindigkeit so einstellen, dass die Ereigniszeitpunkte von Simulation und Echtzeit nahezu identisch sind. Die diskrete Simulation wird, wie in der Abbildung 3-2 dargestellt, anhand der Einteilung der Ereignisabfolge einer der zwei Klassen zugeordnet. Wenn bei der zeitgesteuerten Variante der Simulationsuhr immer um ein konstantes Zeitintervall  $\Delta t$  inkrementiert wird, springt die Uhr bei dem ereignisgesteuerten Ablauf nach Initialisierung zum Zeitpunkt des nächstfolgenden Ereignisses in der Ereignisliste[Domsch02]. Die ereignisgesteuerte Simulation wird anhand der Modellierung von Ereignissen und Zustandsänderungen in vier weitere Untergruppen gegliedert [Hooper86]:

- ereignisorientiert,
- aktivitätsorientiert,
- prozessorientiert,
- transaktionsorientiert.

Da die meisten Simulatoren für das Abbilden einer Halbleiterproduktion auf dem Konzept der ereignisorientierten Zeitsteuerung basieren, steht dies im Fokus der vorliegenden Ausarbei-

tung. Die verbliebenen Ansätze werden ausführlich in der Literatur behandelt [Mehl94] [Liebl95] [Noche90].

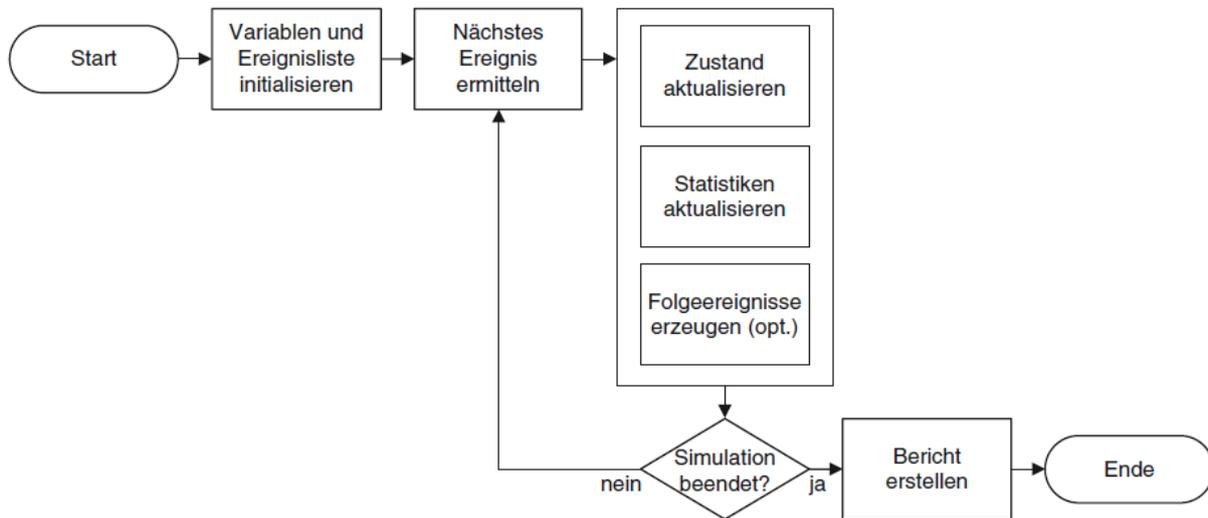
### 3.2.3 Aufbau und Funktionsweise der diskreten ereignisorientierten Simulation

Bei der diskreten ereignisorientierten Simulation (engl. Discrete Event Simulation – DES) wird das Systemverhalten mithilfe von Ereignisfolgen nachgebildet. Dem Eintreten eines Ereignisses folgt eine Zustandsänderung [Rose11]. Die für jede Fertigung charakteristischen Events sind zum Beispiel die Ankunft von Arbeitsmaterial an der Anlage, Prozessende oder Maschinenausfall. Der Kern von nach DES-Prinzip arbeitenden Simulatoren besteht aus [Law00]:

- Systemzustand – Menge an Parametern, die den Systemzustand zum gegebenen Zeitpunkt beschreiben,
- Simulationsuhr – Variable, welche die momentane Simulationszeit beinhaltet,
- Ereignisliste – nach Eintrittszeit sortierte Liste der Ereignisse,
- Statistische Zähler – Parameter zur Speicherung aller notwendigen, statistischen Daten über das Verhalten des Systems,
- Initialisierungsroutine – Routine, die bei Simulationsstart die Variablen initialisiert
- Zeitführungsroutine – Routine, die das nächste Ereignis in der Ereignisliste bestimmt und die Simulationsuhr überschreibt,
- Ereignisroutine – Routine, die für Systemzustandsänderungen nach Eintreten eines Ereignisses verantwortlich ist,
- Bibliotheksroutinen – implementierte Routinen, die für verschiedene Zwecke die wie Zufallszahlenerzeugung, Visualisierung u. a. dienen,
- Ergebnisroutine – Routine, die für das Reporting zuständig ist. Sie startet erst nach Abschluss der Simulation,
- Hauptprogramm – Routine, die für das Managen der anderen Routinen zuständig ist.

Ein Simulationslauf startet, wie in Abbildung 3-2 dargestellt, mit der Initialisierung des Systems. Den System-, Zustands- und Statistikvariablen, der Simulationsuhr und Ereignisliste werden die für die Simulation notwendigen Informationen zugewiesen. Der eigentliche Simulationszyklus beginnt, wenn die Zeitführungsroutine einen Eintrag aus der Ereignisliste holt. Das gefundene Ereignis muss zwingend den frühesten Eintrittszeitpunkt der gesamten Liste besitzen und dieser Zeitpunkt wird durch die Simulationsuhr festgehalten. Anschließend be-

ginnt die Ereignisprozedur nach Bedarf sowohl die Zustands- und Statistikvariablen als auch den Ereigniskalender zu aktualisieren. Nach Abschluss der Bearbeitung aller mit dem Ereignis verbundenen Variablen, ermittelt das Hauptprogramm, ob die Simulation durch Wiederholung des Zyklus weiter ausgeführt werden soll. Ein Simulationslauf wird entweder an einem



**Abbildung 3-2: Ablaufschema einer ereignisorientierten Simulation – Quelle: [Rose11]**

vor dem Start festgelegten Zeitpunkt abgebrochen oder wenn die Ereignisliste vollständig abgearbeitet wurde. Da das beschriebene Ablaufschema eines diskreten, ereignisorientierten Simulators prinzipiell einfach ist, liegt die Schwierigkeit der Simulation nicht in Durchführung eines Simulationsexperiments, sondern bei der Erstellung eines Simulationsmodells [Rose11]. Als problematisch könnte sich auch die Versuchsplanung erweisen, weil es oft zu einem Kompromiss zwischen Qualität der Ergebnisse und Effizienz bzw. Dauer des Experimentes kommen kann.

### 3.2.4 Anwendung in der Halbleiterfertigung

Die Komplexität der Fertigung, Produktvielfalt, sehr lange Prozessketten und andere, für die Industrie charakteristische Eigenschaften verhindern im gleichen Maße wie stochastische Einflüsse auf Produktionsprozesse die Benutzung von einfachen Vorhersagealgorithmen.

Mit ihrem umfangreichen Modellierungsansatz bietet die Simulation der Halbleiterindustrie ein umfangreiches und vielseitiges Werkzeug für die Erstellung von Prognosen. Dass sich Simulationssoftware als sehr nützlich erwiesen hat, wurde mittlerweile von vielen führenden Halbleiterherstellern (z. B.: IBM, Intel oder Infineon) in vielen Veröffentlichungen bestätigt (z.B.: [Bagchi08], [Scholl10], [Gan06], [Potoradi02]).

Da Simulationen, die auf detailreichen Modellen beruhen, über ein enormes Prognosepotential verfügen, werden in der vorliegenden Arbeit nur ausgesuchte Beispiele und Anwendungsfelder beschrieben. In der Halbleiterindustrie sind es vor allem [Bagchi08]:

- Bestimmung der Losposition und Bewegung im Werk mit entsprechenden Zeitangaben,
- simulationsbasierte Optimierung,
- Unterstützung bei strategischen Aufgaben der Produktionssteuerung,
- experimentelle Fabrikänderungen.

Die DES mittels Abarbeitung der Ereignisliste speichert implizit die Bewegung des Materials in der Fabrik. Die Simulation kann auch mit einem Zeitpunkt als Enddatum versehen werden. Dadurch gestaltet sich die Extraktion der Angaben für die Prognose als unkompliziert. Die Losanlieferungsvorhersage an bestimmten Schlüsselstellen in der Fabrik ermöglicht einen besseren Umgang mit typischen Problemen in der Halbleiterproduktion, wie zum Beispiel das Auftreten von Überbeständen und der Wartungsplanung. Oft ist es auch notwendig, die Fabrik so detailliert für die Simulation zu modellieren, dass die Bewegungsinformationen für einzelne Lose ermittelt werden können. Auf ähnliche Weise unterstützt die Simulation die Planung von strategischen Änderungen in der Produktionssteuerung. Der Planer, z.B. konfrontiert mit der Notwendigkeit von Modifikationen an Produktportfolio, bekommt durch Simulation eine Möglichkeit der Durchführung verschiedener Tests mit Fokus auf Fabrik Kennzahlen wie: Durchlaufzeit, Liefertermineinhaltung oder Fabrikauslastung. Die Anschaffung von im Nanometer-Bereich arbeitenden Anlagen für die Prozessierung der Halbleiter verursacht enorme Kosten. Mittels Simulation ist es auch möglich, entsprechende Szenarios zu entwerfen und zu simulieren, um die Wirtschaftlichkeit der Investition zu bewerten. Bei Durchführung derartiger Experimente betrachtet der Experte die Fabrik als eine Art Blackbox und verwendet nur den Unterschied zwischen Kapazität und Kalkulation an prognostizierten Gewinnen. Die Krönung aller Anwendungsfelder von DES ist die Prüfmöglichkeit von Optimierungsansätzen [Law00]. Die Möglichkeiten der Simulation in diesem Bereich sind vielfältig, was sich auch in der Literatur widerspiegelt. Die Anordnung der Lose in der Warteschlange erfolgt im Manufacturing Execution System<sup>11</sup> (MES) durch Anwenden von speziellen Abfer-

---

<sup>11</sup> Als Manufacturing Execution System (MES) wird eine prozessnah operierende Ebene eines mehrschichtigen Fertigungsmanagementsystems bezeichnet. Oft wird der deutsche Begriff *Produktionsleitsystem* synonym verwendet. Das MES zeichnet sich gegenüber ähnlich wirksamen Systemen zur Produktionsplanung, den sog. ERP-Systemen (Enterprise Resource Planning), durch die direkte Anbindung an die verteilten Systeme der Prozessautomatisierung aus und ermöglicht die Führung, Lenkung, Steuerung oder Kontrolle der Produktion in Echtzeit. Dazu gehören klassische Datenerfassungen und Aufbereitungen wie Betriebsdatenerfassung (BDE), Maschinen-

tigungsregeln. Durch Entwicklung und Einsatz neuer Abfertigungsregeln versuchen Experten, die Fertigungsstätten mit einem vorher bestimmten Ziel, zum Beispiel eine Verbesserung der Liefertreue, zu optimieren. Da es unmöglich ist, Neuentwicklungen sofort in realer Umgebung zu testen, wird auch hier eine Simulation verwendet [Akcali00]. Ein weiteres typisches Problem der Halbleiterfertigung ist die Beladestrategie der Batchanlagen und deren Optimierung. Die Anlagen können mehrere Lose gleichzeitig fertigen, aber nur mit einem Los, unter langen Prozesszeiten, arbeiten. Mithilfe der Simulation wird ein optimaler Schwellwert für die minimale Batchgröße gesucht [Rulkens98], [GanB06]. Weiterhin wird auch für die Transportsystemoptimierung [Gan06], Untersuchung der optimalen Losgröße [Potoradi99] und viele andere Optimierungsansätze eine Simulation als hilfreiches Werkzeug angewendet. Deren detaillierte Beschreibung wird aber ausgelassen, weil sie schnell den Rahmen der vorliegenden Ausarbeitung sprengen würde.

Der Prognosehorizont stellt einen der bedeutendsten Parameter dar, der die Simulation beschreibt. In der Halbleiterbranche werden meistens drei mögliche Bezeichnungen für die Vorhersagezeitspanne [Scholl10] gebraucht:

- Kurzfristvorhersage – engl. Short-term forecasting – mit der Simulationsperiode bis ca. 14 Tage, meistens wird jedoch wegen des Genauigkeitsgrads der Zeitraum auf 5 – 7 Tage verkürzt,
- Mittelfristvorhersage – engl. Medium-term forecasting – mit der Simulationsperiode bis ca. 3 Monate,
- Langzeitvorhersage – engl. Long-term forecasting – mit der Simulationsperiode über 3 Monate.

Die erwarteten Prognoseziele unterscheiden sich in Abhängigkeit des Horizonts deutlich voneinander. Am einfachsten lässt sich dies mit einem Beispiel veranschaulichen. Angenommen es werden für ein Schichtdickenmessgerät mit maximalem Durchsatz von 5.000 Losen pro Monat unterschiedliche Vorhersagen benötigt. Bei einer schichtfeinen Anlieferungsvorhersage des Materials an die Anlage innerhalb der nächsten drei Tage, spielen insofern sogar einzelne Lose eine sehr große Rolle. Wenn aber ein prognostizierter monatlicher Durchsatz derselben Maschine für die nächsten drei Monate ermittelt werden soll, ist eine Differenz einiger Werkstücke zulässig. Die Gesetzmäßigkeit, die sich aus diesem Beispiel ableiten lässt, wurde in Abbildung 3-3 dargestellt. Je kürzer der Zeithorizont des Experiments ist, desto detaillierter

---

datenerfassung (MDE) und Personaldatenerfassung, aber auch alle anderen Prozesse, die eine zeitnahe Auswirkung auf den Fertigungs-/Produktionsprozess haben. – Quelle: Wikipedia: Manufacturing Execution – Wikipedia die freie Enzyklopädie - [http://de.wikipedia.org/wiki/Manufacturing\\_Execution\\_System](http://de.wikipedia.org/wiki/Manufacturing_Execution_System) - [Online-stand 01.08.2011]

## Analyse bestehender Prognoseansätze

und genauer sollte das Modell sein [Reijers99]. Mittel- und Langzeitsimulationen benötigen aber auch deshalb wesentlich einfachere Modelle, weil die Abarbeitung eines detaillierten Ereigniskalenders viel Rechenzeit benötigt. Die Abbildung 3-4 veranschaulicht den Einfluss der Initialisierung auf die Simulationsresultate. Im Bereich der Kurzfristsimulation ist die exakte Abbildung von Bestandsverteilungen in der Fabrik, Maschinenzuständen, Position des Materials im Transportsystem u. a. überaus wichtig [Reijers99].

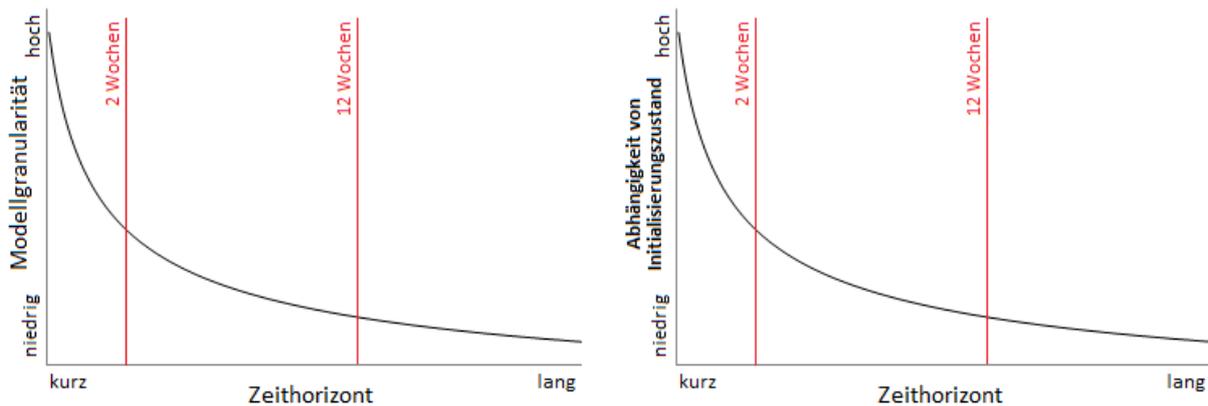


Abbildung 3-3: Abhängigkeit zwischen Prognosehorizont und Modellgranularität

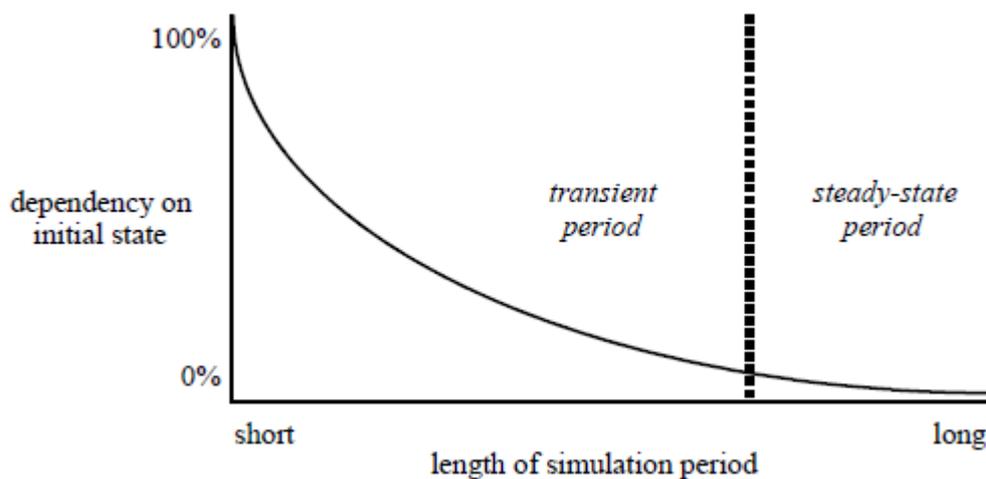


Abbildung 3-4: Verhältnis von Prognosehorizont und Abhängigkeit von exakter Initialisierung.  
Quelle: [Reijers99]

Aus der Charakteristik der Simulation in der Halbleiterfertigung kristallisiert sich eine komplizierte Aufgabe für Simulationsexperten heraus. Es wird immer ein Kompromiss zwischen Modellierungsaufwand, Güte der Vorhersage und benötigter Rechenzeit ausbalanciert werden müssen.

### 3.2.5 Bewertung als Prognosewerkzeug in der Halbleiterfertigung

Die detaillierte Auseinandersetzung mit der Simulation in den vorherigen Kapiteln diene der Veranschaulichung der Komplexität der Vorhersagemethode. Die Vielfältigkeit an Anwendungsmöglichkeiten und das breite Spektrum der Resultate begründen die Verwendung der Simulation als meistbenutztes Werkzeug zur Erstellung von Prognosen. In diesem Kapitel werden die Vor- und Nachteile dieses Ansatzes nur abstrakt beschrieben. Da bei unserem Projektpartner ein paralleles Projekt zur Entwicklung einer Kurzfristsimulation [Noack12] realisiert wurde, wurde auch die quantitative Bewertung und der Vergleich der Resultate von mehreren Prognosemethoden durchgeführt – wie in Kapitel 7 genauer beschrieben wird.

Der wesentliche Vorteil der Simulation wurde bereits mehrmals erwähnt. Es ist die Genauigkeit und die Vielfalt der Prognosen. Möglich sind Aussagen in sehr feiner Granularität, wie Anlieferungsvorhersagen, Maschinenverhaltensprognosen, Bottleneck-Früh-warnsysteme und ähnliches. Der Simulationsexperte, der an der gegebenen Aufgabenstellung arbeitet, erhält nach dem Simulationsabschluss meistens aber eine wesentlich komplexere Lösung, die auch für andere Fragestellungen von Nutzen sein könnte. Außer typischen Vorhersageaufgaben ermöglichen derartige Ansätze Untersuchungen bei der Entwicklung und der Feinplanung der jeweiligen Fertigungsstätte. Aus den beschriebenen Eigenschaften der Simulation konkludiert sich automatisch die vollständige Erfüllung der am Projektanfang zusammengestellten funktionalen Anforderungen (vgl. Kapitel 2.5.1).

Die Begründung für die Entwicklung neuer Prognosemethoden, statt der Anwendung der Simulation, liegt in den projektrelevanten, nichtfunktionalen Anforderungen (vgl. Kapitel 2.5.2). Ein Simulationslauf bei der Kurzfristsimulation beginnt mit einer detaillierten Initialisierungsphase. Anschließend folgt die eigentliche Bearbeitung der Daten durch einen Simulator. Zum Schluss werden die Daten für das Reporting entsprechend vorbereitet. Scholl [Scholl10] diskutiert die Simulationslaufzeiten für die 200 mm Produktionslinie von Infineon in Dresden. Simuliert wird mittels der kommerziellen Software AutoSched AP 8.0 mit 10 Konfidenzläufen. Die erforderliche Ausführungszeit beträgt:

- Mittelfristsimulation mit 3 Monaten Prognosehorizont – ca. 12 Stunden,
- Kurzfristsimulation mit 3 Tagen Prognosehorizont – weniger als 1 Stunde.

Deshalb wird die benötigte Zeit für die Simulation als entscheidender Nachteil eingestuft. Weiterhin beschreibt Scholl [Scholl10], dass auch andere Probleme bei der Erstellung einer Kurzfristsimulation auftreten können. Der Detaillierungsgrad des Modells erfordert enorme Mengen qualitativ sehr hochwertiger Basisdaten. Bei den über Jahre angewachsenen und zum Teil heterogenen Datenbanksystemen gestaltet sich die Aufgabe der Datenbereitstellung sehr

kompliziert. Die typischen Probleme solcher Systeme werden genauer in Kapitel 4.1 beschrieben. Da die Projekte – „Analytische Vorhersagemethode AFM“ – und „Kurzfristsimulation“ parallel gestartet wurden, war es auch möglich, den benötigten Arbeitsaufwand, der bei der Simulation ca. 5-mal höher war, zu vergleichen. Eine der größten Herausforderungen liegt auch in der Kalibrierung des Modells durch gezielte Validierung der Resultate. Dem Experten werden ein enormes Fachwissen und weitgehende Systemkenntnisse abverlangt. Nicht zu ignorieren sind auch die erforderlichen Kosten, die nach der Fertigstellung des Simulationsmodells auftreten. Einerseits sind es Hardwarekosten, Kosten für Softwarelizenzen und andererseits die Personalkosten. Notwendig ist der Einsatz von einem Simulationsexperten mit ausreichenden Informatik- sowie Fertigungskennnissen, dessen Hauptaufgabe es ist, das Modell aktuell zu halten und weiterzuentwickeln.

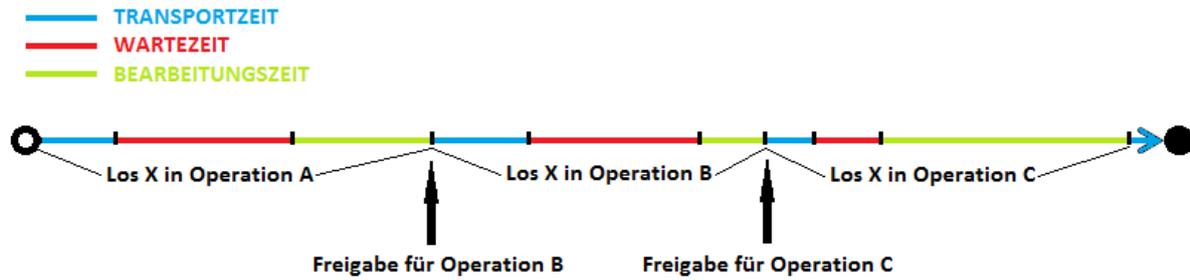
### 3.3 Planzeitbasierte Prognose

#### 3.3.1 Aufbau und Funktionsweise

Im Vergleich mit den in vorherigen Kapiteln beschriebenen Simulationen ist die Funktionsweise der Prognosealgorithmen, die auf den Planungsdaten basieren, unkompliziert. In Abbildung 3-5 wurde ein stark vereinfachtes Beispiel einer Route in der Halbleiterfertigung dargestellt. Ein Los  $X$  das sich auf der Route  $Rt$  bewegen soll, wurde in die Fabrik eingeschleust und durchläuft drei Operationen:  $\underline{A}$ ,  $\underline{B}$ ,  $\underline{C}$ . Einem Werkstück wird eine gegebene Operationsnummer zugewiesen, wenn es:

- sich im Transportsystem befindet und gerade zu der Warteschlange der zugewiesenen Anlage transportiert wird,
- sich in der Warteschlange befindet und auf die Bearbeitung an der zugewiesener Anlage wartet,
- prozessiert wird.

Für die komplette Fertigstellung eines Produktes, dessen Abarbeitungsfolge mittels Route  $Rt$  beschrieben ist, wird eine Zeit  $pt$  geplant. Die Planungsexperten schätzen auch den prozentualen Anteil der Gesamtabarbeitungszeit  $pt$  für Transport-, Warte- und Prozesszeit für jede Operation. Es entsteht eine Matrix, die als vereinfachtes Beispiel in Tabelle 3-1 abgebildet ist.



**Abbildung 3-5: Vereinfachtes Abbild einer Route in der Halbleiterfertigung**

Auf diese Weise müssen alle Routen für das gesamte Produktportfolio vorbereitet werden. Anschließend kann die erste Prognose kalkuliert werden. Es wird ein Datenschnappschuss der Materialverteilung in der Fabrik angefertigt. Vor allem ist eine Information über die Operationszugehörigkeit notwendig. Wichtig dabei sind aber auch Informationen, in welcher Phase des Produktionsschritts sich ein Los befindet, ob es transportiert, wartet oder bereits prozessiert wird. Angenommen für ein Los Y, das sich in Operation B befindet und gerade transportiert wird, wird eine Fertigstellungsprognose berechnet. Die noch notwendige Durchlaufzeit setzt sich wie folgt zusammen:

$$nct_Y = wt_B + pt_B + tt_C + wt_C + pt_C$$

Operation	Transportzeit <i>tt</i>	Wartezeit <i>wt</i>	Prozesszeit <i>pt</i>
A	8 % <i>pt</i>	17 % <i>pt</i>	12 % <i>pt</i>
B	8 % <i>pt</i>	13 % <i>pt</i>	7 % <i>pt</i>
C	4 % <i>pt</i>	7 % <i>pt</i>	24 % <i>pt</i>

**Tabelle 3-1: Vereinfachtes Abbild eines prozentualen Zeiteanteils an der Gesamtbearbeitungszeit einer Route für jede Operation.**

Die Berechnung des Ankunftszeitpunkts erfolgt durch Addieren der errechneten, notwendigen Zeit zum Zeitstempel des Fabrik-Snapshot. In einer realen Fabrik wird diese Prozedur für alle Lose durchgeführt. Dadurch, dass die realen Routen Hunderte von Operationen enthalten und damit wesentlich komplexer sind, wird auch die Berechnung zeitintensiver.

Um das Verfahren nachträglich anpassen zu können, arbeiten die Planer oft mit der mel:  $FF = CT/RPT$  (vgl. [Atherton95]), wobei  $FF$  – Flussfaktor<sup>12</sup>,  $CT$  – Durchlaufzeit<sup>13</sup> und

<sup>12</sup> Meistens wird englische Bezeichnung benutzt - flow factor

<sup>13</sup> Wie in 12 – cycle time

*RPT* – reine<sup>14</sup> Prozesszeit<sup>15</sup> bedeuten [Habeni10]. Für gewöhnlich wird für *RPT* die minimale physikalische Durchlaufzeit verwendet, also die Summe aller Prozess- und Transportzeiten abzüglich aller Wartezeiten. In der Halbleiterfertigung wird aber die Transportdauer oft vernachlässigt, weil deren Berechnung mit vielen Problemen verbunden ist. Außerdem nimmt das Prozessieren meistens entscheidend mehr Zeit in Anspruch als das Transportieren. Der Flussfaktor ist eine der wichtigsten Kennzahlen für die Beschreibung des Fabrikzustandes. In der optimalen Fertigung ist *FF* gleich 1, also die Fabrik arbeitet ohne jeglichen Wartezeiten – weder bei Abarbeitung noch Beförderung. In der Praxis erreicht der Flussfaktor meistens aber Werte zwischen 2 und 3, diese können aber auch höher werden [Kropik09]. Zu Prognosezwecken wird für jede Route die Summe der reinen Zeitangaben für die Prozessierung, also *RPT*, gespeichert. Nachträglich ermitteln die Anwender den Flussfaktor für die Fabrik, indem durch Multiplikation von *FF* und *RPT* die notwendige Durchlaufzeit errechnet und dieser Wert als Basis für den Vorhersagealgorithmus für Berechnungen eingesetzt wird. Der Flussfaktor als Steuerungsparameter nimmt Einfluss auf die Prognose und beinhaltet implizit den Zustand der Fertigungsstätte. Denkbar ist, dass speziell priorisierte Produkte andere Flussfaktoren als die normalen Produkte bekommen, wodurch die Qualität der Vorhersage verbessert würde.

### 3.3.2 Bewertung der planzeitbasierten Prognose in der Halbleiterfertigung

Das in Kapitel 3.3.1 beschriebene Prognoseverfahren wird sehr oft in der Industrie verwendet. Der größte Vorteil des Ansatzes ist der einfache Aufbau der Algorithmen. Dementsprechend ist auch die rechnergestützte Umsetzung nicht kompliziert und benötigt nur eine geringe Anzahl an Basisdaten. Nach der Implementierung kann mithilfe dieser Methode eine Vorhersage auch für komplexe Systeme wie eine Halbleiterfabrik schnell berechnet werden. Der Aufwand in der Wartung der Prognosesoftware ist sehr gering. Die Berechnung der Flussfaktoren und Aufbereitung von Informationen über Routen wird meistens von der Planungsabteilung vorbereitet. Deshalb müssen die erforderlichen Daten der Softwareentwickler nur von den entsprechenden Datenbanken abgerufen werden. Neben den vielen Vorteilen hat diese Art des Prognostizierens aber auch Nachteile, die zum Teil auch auf die Trivialität der Algorithmen zurückzuführen sind. Die typischen Probleme der Halbleiterfertigung wie Sampling, Hold, Rework und Ähnliches werden nicht unterstützt. Die Methode der Vorhersageerstellung lässt die Interaktion der Benutzer mit dem Programm in der Form von Änderung des Flussfaktors

---

<sup>14</sup> damit wird angenommen, dass Verzögerung durch Transport zwischen den Anlagen vernachlässigbar ist

<sup>15</sup> Wie in 12 – raw process time

als Steuergröße zu. Der fehlende Automatismus in der Steuerung kann zu extremen Vorhersageverfälschungen führen. Weiterhin ist auch die Verwendung von generellen Parametern für eine ganze Fabrik problematisch, weil es in der Praxis zu großen Flussfaktorunterschieden zwischen den einzelnen Routen kommen kann. Sinnvoller ist es, für jede Route spezifische Faktoren zu ergänzen. Bei hoch ausgelasteten Fabriken fehlt sogar oft die Eindeutigkeit des Flussfaktors innerhalb eines Abarbeitungsplans. Es ist so zu verstehen, dass für eine gegebene Route ein Produkt sich mit unterschiedlichen Flussfaktoren auf Teilstrecken bewegt. Im Vergleich mit der Simulationen fehlt bei dieser Methode die Unterstützung von Maschinendaten, wie Ausfallstatistiken, momentanen Zuständen, maximalen Durchsätzen, geplanten Wartungen etc. Bei instabilem Verhalten im Maschinenpark oder bei hoher Auslastung kommt es zusätzlich zu deutlichen Abweichungen zwischen Prognose und Realität. Die Prognosemethode ist nicht lernfähig, da sie statisch ist. Als Letztes ist zu erwähnen, dass der Vorhersageansatz keine Unterstützung von Zwischenlagern vorsieht.

Bei unserem Projektpartner aus der Industrie wurde bereits vor Jahren diese Methode implementiert und wird für die Vorhersagezwecke eingesetzt. Die Erfahrungen des Anwendergremiums sind geteilt. Während die Schnelligkeit der Berechnung bei der Verfügbarkeit der Ergebnisse als Vorteil dargestellt wird, wird die Vorhersagequalität meistens negativ bewertet. In Kapitel 7 wird auch dieser Ansatz quantitativ übergeprüft und sowohl mit der Analytischen Vorhersagemethode AFM als auch mit der Kurzfristsimulation verglichen.

### 3.4 Warteschlangentheorie

Im Jahr 1909 veröffentlichte der Däne A.K. Erlang eine Arbeit unter dem Titel: „*The Theory of Probabilities and Telephone Conversations*“, in der er sich mit mathematischen Verfahren zur Beschreibung von Telefongesprächen auseinandersetzte [Gross08]. Seit dieser Zeit wird er als der Vater der Warteschlangentheorie bezeichnet. In den dreißiger Jahren entwickelte Felix Pollaczek einen stark vereinfachten Berechnungsformalismus, die Pollaczek-Khintchine-Formel, welche die Verwendung der Theorie vereinfachte. Die heutige Notation der Warteschlangensysteme stammen aus dem Jahr 1951 von D. G. Kendall. Im Jahr 1957 wurden Lösungen für geschlossene und offene Warteschlangennetze präsentiert, die sogenannten Produktformlösungen<sup>16</sup>. In der Ära der Computertechnik wurde die Warteschlangentheorie ein leistungsfähiges Design- und Analyseinstrument.

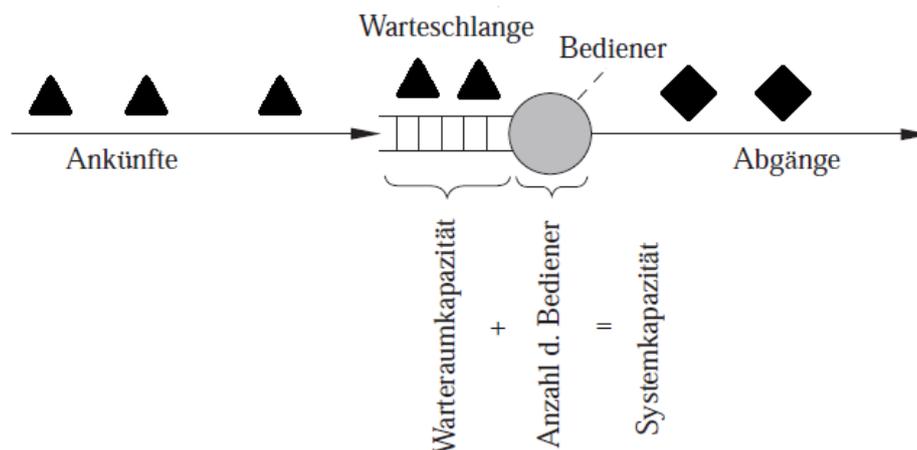
---

<sup>16</sup> Die Lösung eines Warteschlangennetzes lässt sich unter bestimmten Bedingungen aus Lösungen der einzelnen Warteschlangenknoten zusammensetzen (vgl. [Gross08].)

### 3.4.1 Definition und Einsatzgebiete

Die Warteschlangentheorie wurde von mehreren Autoren [Sauerw07], [Gross08] in der Literatur detailliert beschrieben, deshalb werden in der vorliegenden Ausarbeitung nur allgemeine Grundlagen zu diesem Thema behandelt. Ein Warteschlangenprozess ist in Abbildung 3-6 schematisch dargestellt und wird mit folgenden Angaben charakterisiert [Gross08]:

- Ankunftsprozess – wird mittels Angabe der Verteilung von Ankünften oder der Verteilung der Zwischenankunftszeit definiert,
- Bedienprozess – wird durch die Verteilung der Bedienzeiten (meistens durch mittlere Bedienrate) charakterisiert,
- Warteschlangenabfertigungsregel – enthält Informationen und Algorithmen für die Reihenfolge der Auswahl von Kunden der Warteschlange. Meistbekannte und einfachste Regel ist FIFO (engl. First In First Out), also wer zuerst kommt, wird zuerst bedient,
- Systemkapazität – gibt maximale Anzahl Aufträge oder Kunden im System vor.
- Anzahl der Bediener – parallele Bediener mit gemeinsamer oder autonomer Warteschlange,
- Zusammensetzung der Bedienstufen – Systembeschreibung mittels Warteschlangennetzen.



**Abbildung 3-6: Schematische Darstellung eines Warteschlangenprozess**

Mittels Warteschlangennetzen lassen sich sehr komplexe Systeme beschreiben. Das Ziel dieses Ansatzes ist die Berechnung unterschiedlicher Leistungskenngrößen wie zum Beispiel:

- Verkehrsintensität,
- Auslastung,

- Durchsatz,
- mittlere Anzahl der Kunden,
- mittlere Warteschlangenlänge,
- Verweilzeit.

Da die Verweilzeit angibt, wie lange ein Kunde im System verbracht hat, lässt diese Zeitspanne eine Prognose zu.

Die Warteschlangentheorie wird für strategische und operative Entscheidungen als analytische Lösungsalternative zu diskreten Simulationen angewendet. Mittels dieses Ansatzes lassen sich auch kompliziertere Probleme abbilden. Dies gilt auch für die Halbleiterindustrie, in der die Anfertigung von bestimmten Produkten von einer großen Anzahl an produktionsinternen und externen Faktoren abhängig ist [Mladen09]. Da sich fast jedes System als Warteschlangennetz darstellen und berechnen lässt, findet die Warteschlangentheorie Anwendung in Bereichen wie:

- der Prozessverwaltung [Held07],
- der Logistik [Held07],
- dem Straßenverkehr [Held07],
- der Medizin [Loeser06],
- der Fertigung [Mladen09],
- der Psychologie [Held07],
- bei Planung, Analyse und Vorhersage von diskreten Fertigungslinien [Zisgen08].

Mittlerweile wurden die Ansätze der Warteschlangentheorie in kommerzieller Software erfolgreich implementiert. Dank ausgereifter Computertechnik wurden Entwurf, Bedienung und Wartung von Warteschlangennetzen handlicher und zeitlich effizienter.

### **3.4.2 Verwendung in der Halbleiterfertigung und Bewertung**

Eine auf der Warteschlangentheorie basierende Software wurde bei unserem Projektpartner nicht benutzt bzw. nicht zur Verfügung gestellt. Aus diesem Grund wird die Bewertung der Methode nur anhand der Literaturrecherche erfolgen.

Dass die Methodik erfolgreich in der Halbleiterfertigung angewendet wird, beweist ein Werbeslogan einer der weltweit führenden Unternehmen für Soft- und Hardwarelösungen – IBM:

*„EPOS ist eine integrierte **Planungssoftware zur Analyse von Kapazitäten, Umlaufbeständen und Durchlaufzeiten sowie der Vorhersage von Fertigungsoutput für komplexe, diskrete Fertigungslinien.***

*EPOS nutzt von **IBM patentierte Methoden der Warteschlangentheorie und der mathematischen Optimierung zur Analyse der logistischen Leistungskennzahlen.** Dabei berücksichtigt EPOS auch zufällige Ereignisse in der Fertigung, wie Nacharbeit, Verschrottung und Maschinenausfälle durch die Verwendung stochastischer Verfahren.“<sup>17</sup>*

Horst Zisgen, ein Mitglied des EPOS Entwicklerteams berichtet über Versuche des Einsatzes der Software in einer 300 mm Halbleiterfabrik von IBM in East Fishkill [Zisgen08]. Die Softwarelösung basiert auf hoch entwickelten Algorithmen der Warteschlangentheorie und unterstützt die typischen Probleme in der Halbleiterfertigung, wie Batchprozesse, Stochastik von Prozess- und Ausfallzeiten, Sampling, Hold, Variation in der Losgröße, Rework und ähnliche. Entwickelt wird ein offenes Warteschlangennetz mit integrierter Unterstützung der flexiblen Verteilung von verschiedenen Produkten. Die wesentlichen Vorteile der Methode werden durch Vergleich mit diskreter, ereignisorientierter Simulation dargestellt. Wenn DES in Puncto Komplexität detailreicher ist, also wesentlich mehr Freiheiten bei Modellierung zulässt, punktet EPOS bei der Einfachheit der Modellbildung und in der Wartung. Anhand der Bottleneckanalyse wird die Treffsicherheit der auf der Warteschlangentheorie basierenden Methode besser eingestuft [Zisgen08]. Der Ansatz wurde von den Autoren für Mittel- und Langzeitvorhersagen empfohlen [Zisgen08] und unterstützt die Experten durch Vielseitigkeit an Analysemöglichkeiten im breitem Spektrum an Planungs- und Optimierungsaufgaben. Da aufgrund der Thematik der vorliegenden Dissertation der Fokus auf der Kurzfristprognose liegt, ist EPOS für derartige Prognosen ungeeignet. Zisgen empfiehlt für diesen Anwendungsfall Fluidmodelle zu verwenden [Zisgen08].

Ein weiterer Beweis für die Möglichkeit der Anwendung von Warteschlangensystemen wurde bereits 1988 in einem von Chen verfassten Artikel [Chen88] niedergeschrieben. Die Autoren beschreiben präzise den Weg der Entwicklung eines Warteschlangennetzes für Vorhersagen der Fabrikperformance. Weiterhin wurden auch die Ergebnisse von angefertigten Vorhersagen gegenüber der Realität validiert. Die Durchsatzvorhersage ist nur von sehr geringer Realitätsabweichung gekennzeichnet, die maximal 14 % beträgt.

---

<sup>17</sup> entstammt der IBM Internetpräsenz: <http://www-05.ibm.com/de/processindustry/solutions/epos.html> - Stand 23.08.11 12:32:00

Ansonsten weichen die Parameter im Mittel um ca. 11 % ab. Jedoch ist es nicht einfach, die Werte, welche durch Validierungsstudien ermittelt wurden, zu bewerten. Es fehlt die detaillierte Beschreibung der verwendeten Modelle und untersuchten Szenarios. Andere Quellen sind bei der Validität der Methode skeptischer. Lazowska beispielsweise schätzt die Methode nach einer ganzen Reihe von Experimenten schlechter ein [Lazow84]. Die Auslastung und der Durchsatz weichen im Mittel um 5 bis 10 % ab. Bei einer Anlieferungs- oder Durchlaufzeitvorhersage steigt der Fehlerquotient auf 10 bis 30 %.

Die Warteschlangentheorie wurde in der Vergangenheit für Prognosen in der Halbleiterfertigung oft angewendet. Mittlerweile hat sich jedoch die Simulation als die Prognosemethode schlechthin etabliert. Der Entwurf des Warteschlangenmodells ist für die gesamte moderne Halbleiterfabrik zum Teil komplexer geworden als eine Anfertigung des Simulationsmodells. Ähnlich kompliziert gestaltet sich die Wartung des Modells und das Design neuer Experimente. Es werden oft Warteschlangennetze verwendet, um Fertigungsausschnitte zu modellieren und zu analysieren. Diese Methodik wurde aber für die Erstellung von Anlieferungsprognosen in einer Logikfabrik mit mehreren Anlagen und zugehörigen Ankunftsverteilungen für zu kompliziert und eher ungeeignet befunden [Shanth07].

### **3.5 Andere Vorhersageansätze**

In der Literatur wurden bereits zahlreiche Prognoseansätze beschrieben. Die meisten davon wurden implementiert und anschließend in realen Systemen validiert. Die Anwendung der Vorhersagesysteme in der Fertigung, die mittels Job-Shop-Prinzip arbeiten, ist aber meistens nicht möglich. Außer einer Simulation, die sich auf die Plandaten basierender Systeme und Warteschlangennetze bezieht, existieren nur wenige Ansätze, die eine Prognose in der Halbleiterfertigung ermöglichen. Die Abbildung der Fabrik in einem Petri-Netz wurde von Nicholas G. Odrey [Odrey01] und Fei Qiao [Qiao04] vorgestellt. Der Ansatz basiert auf einem Modell der Fertigung in Form eines Prädikate/Transitions-Netzes, das nach dem Informatiker Carl Adam Petri benannt wurde. Die in der Literatur beschriebenen Modelle lassen nur Prognosen für einen sehr kurzfristigen Zeithorizont, also bis zu 8 Stunden, zu. Deshalb wird nur ein Fabrikausschnitt modelliert. Derartige Prognosen finden vor allem als Eingangsdatenstrom für die Optimierung Verwendung. Der Aufwand für den Aufbau und der Datenbedarf eines detaillierten Petri-Netzmodells, mit einem Prognosehorizont von 7 Tagen, ist mit dem Aufwand eines Simulationsmodells vergleichbar. Der netzbasierte Modellierungsansatz ist im Vergleich zur Simulation schwieriger in Handhabung und Validierung. Aus diesen Gründen wird der Ansatz für derart komplexe Fertigungssysteme nicht praktiziert. Ein wesentlich ein-

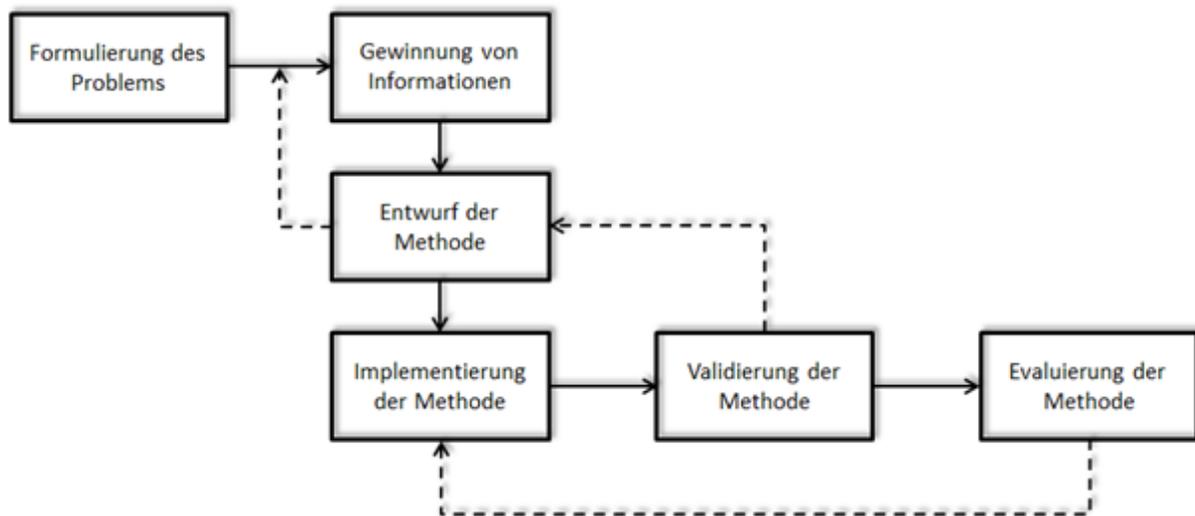
facheres Verfahren, eine Prognose zu bilden, basiert auf der Multiplen Regression, welche von Raddon und Grigsby in [Radd97] angewendet und beschrieben wurde. Diese Art der Regression berechnet die Systemantwort mittels einer mathematischen Formel, die mehrere Eingangsparameter beinhaltet. Die Formel inkl. Eingangsvariablen wird mithilfe der historischen Bewegungsabläufe bestimmt. Chen [Chen03] beschreibt die Methode als die einfachste und dadurch auch am weitesten verbreitete, jedoch ist die Qualität der Resultate eher mangelhaft. Chen verwendet deshalb bevorzugt Neuronale Netze anstatt der Regression [Chen03]. Dies wird in weiteren Veröffentlichungen bestätigt [Chang03] [Hsu04]. Die von den Autoren durchgeführten Vergleiche bezüglich der Güte, der erzeugten Vorhersageresultate beweisen, dass die Neuronale Netze deutlich validere Ergebnisse als die Multiple Regression liefern. Chang schreibt in [Chang03] von einem 40-prozentigen Anstieg der Qualität, gemessen mit RSME (Quadratwurzel des mittleren quadratischen Fehlers, s. Kapitel 3.1.2 ). Die Autoren betonen auch die Vorteile gegenüber der Simulation, wie ein geringerer Bedarf an Basisdaten und ein kleinerer Berechnungsaufwand. Der Nachteil der Methoden im Vergleich zu DES liegt jedoch in Qualität und Stabilität der Ergebnisse. Da die Untersuchungen nur wenige Produktionsschritte, wie auch einen Ausschnitt des Maschinenparks beinhalten, ist ihr Einsatz in einer Massenfertigung mit breitem Produktspektrum fraglich. Von Chang stammt auch ein Prognose-Verfahren, das auf dem Fallbasierten Schließen (engl. *case-based reasoning* mehr bei [Aamodt94]) basiert. Der in [Chang01] beschriebene Ansatz liefert genauere Prognosen als Neuronale Netze, birgt aber dieselben Probleme in Konfrontation mit der Gesamtheit der komplexen Systeme der Halbleiterfertigung. Weiterhin existiert eine Reihe von hybriden Prognosemethoden, in der sich zwei oder mehrere Prognoseansätze gegenseitig ergänzen. Ein Beispiel derartig komplexer Modellierungsansätze wurde von Kaplan und Unal in [Kaplan93] beschrieben. Die vorgestellte Methodik basiert auf einem Verbund der statistischen Analyse mit der Simulation. Liao und Wang stellen in [Liao04] einen Mix der neuronalen Netze mit analytischen Verfahren dar. Die hybriden Methodiken übernehmen von den Modellteilen neben den Vorteilen auch die Nachteile. Vor allem die Komplexität der Modellierung ist die Schwachstelle dieser Ansätze. Daraus resultiert, dass der von den Autoren beschriebene Komplexitätsgrad der Vorhersageprobleme deutlich geringer ist, als der einer gesamten Produktionslinie. Schon alleine der Aufwand bei Aufbau und Validierung der analytischen Komponente ist mit der Entwicklung der AFM vergleichbar. Vermutlich aus diesen Gründen wurden diese Methoden nur in einer Testumgebung erprobt und kontrolliert.

Die Methodik der Durchlaufzeitprognose, präsentiert von Backus, Janakiram, Mowzoon, Runger und Bhargava in [Backus06], besitzt ähnliche Eigenschaften wie die AFM, die in der

Ausarbeitung vorgestellt wird. Die Autoren gehen ebenfalls von der Annahme aus, dass Lose mit ähnlichen Charakteristika sich bezüglich der Durchlaufzeit nahezu gleich verhalten. Aus diesem Grund ist es möglich, zukünftige Ankunftszeitpunkte des Materials an der Zieloperation mithilfe der Auswertung historischer Daten zu prognostizieren. Weiterhin fasst die Methode der Durchlaufzeit mehrere Operationen zusammen. Es erfolgt aber nicht wie bei der AFM, zwischen Start und Zieloperation (mehr in Kapitel 4.5.3), sondern die Basis bildet der Bearbeitungszeit zwischen Flaschenhalsoperationen. Für die Berechnung der prognostizierten der Ankunft an der Zieloperation notwendige Durchlaufzeit setzt sich zusammen aus der Zeit zwischen Startoperation und ersten Bottleneckoperation, Summe aller Zeiten zwischen Bottleneckoperationen auf der Route zwischen Start und Ziel und zwischen der letzten Bottleneckoperation und der Zieloperationen. Ein ähnlicher Ansatz für den Umgang mit bekannten Kapazitätsengpässen erlaubt auch die Erweiterung der AFM, die in Kapitel 4.6.2 beschrieben wird. Der Hauptunterschied der Methoden liegt generell in der Durchlaufzeitbestimmung. Die Autoren beschreiben auch eine Art von Klassifizierung der Lose anhand von Produktzugehörigkeit und Prioritäten. Diese Klassifizierung ist deutlich einfacher als durch AFM. Die in [Backus06] beschriebene Methode, ausgehend von Little's Law, wendet Angaben über WIP zwischen Start- und Zieloperationen für Verfeinerung der Durchlaufzeitbestimmung. Genau dieser Punkt unterscheidet die Prognoseansätze. Wenn für die AFM die Position in der Warteschlange vor der Anlage bzgl. der Durchlaufzeit von Bedeutung ist (Deutlich verspäteten Werkstücke werden deutlich schneller die Produktionsschritte durchgehen, praktisch unabhängig von dem WIP, vgl. Kapitel 4.6), hat bei der Backus-Methode der Fabrikbestand Einfluss auf die Produktionszeit. Dieser Ansatz ist aber nur für nicht verspätete Lose interessant. Denkbar wäre daher die Verbindung der Vorteile beider Methoden, also die Integration der WIP-Abhängigkeit in die AFM.

## 4 Das Konzept der Analytischen Vorhersagemethode – AFM

Die Entwicklung einer neuen Prognosemethode erfolgt ähnlich zu einem typischen Softwareentwicklungsprojekt. In Abbildung 4-1 wurde ein Phasenmodell der Entwicklung einer rechnergestützten Vorhersagemethode dargestellt. Das Ziel des abgebildeten Vorgangs ist, eine funktionierende, valide Prognosesoftware zu erstellen.



**Abbildung 4-1: Phasen der Entwicklung einer rechnergestützten Prognosemethode**

Als Basis dient ein neu entworfener Prognosealgorithmus. Aus diesem Grund ist der Zyklus des Erstellungsprozesses deutlich anders als der Ablauf eines Projektes, das bereits bekannte Prognosevorschriften beinhaltet (vgl. [Armstr01]). Anfänglich werden die Ziele des Projektes definiert und das eigentliche Problem mit allen Rahmenbedingungen formuliert. Die Projektanforderungen wurden bereits in Kapitel 2.5 beschrieben. Anschließend wird versucht, alle notwendigen Informationen über das Problem zu erfassen. Es ist notwendig, allgemeine Angaben, wie in Kapitel 2.1 bis 2.4 beschrieben, zu erforschen und eine intensive, tiefgehende Analyse des Datenmaterials in qualitativer und quantitativer Form anzuschließen. Wenn dem Entwicklerteam ausreichende Kenntnisse des Systems zur Verfügung stehen, werden Entwürfe der Vorhersagemethoden angefertigt und anschließend prototypisch implementiert. Über Validierung mittels des Vergleichs von Prognose und Realität wird der beste Ansatz gesucht. Falls notwendig werden kleine Änderungen im methodischen Entwurf gemacht. Jetzt kann die Software implementiert werden (vgl. Kapitel 5). Die anschließende Validierung, (vgl. Kapitel 0) muss die richtige Funktion der Methode und der Software selbst bestätigen.

Als letzter Entwicklungsschritt wird die Evaluierung der Methode und der Software angesehen, die möglicherweise zu kleinen Verbesserungen in der Implementierung führen kann. Die Evaluierungsphase erfolgt durch Interaktion der Anwender mit der Software, die bereits in der Fertigstellungsphase das Werkzeug kennenlernen und die Möglichkeit haben, kleine Verbesserungsvorschläge bei den Entwicklern vorzuschlagen.

### 4.1 Problematik der Komplexität und Qualität der Basisdaten

Die enorme Komplexität von Prozessen in der Halbleiterfertigung, Produktionsart und dem Aufbau eines Werkes spiegelt sich in der Menge an unterschiedlichsten Daten wider. Die Entwicklung der Computertechnik und immer billiger werdender Speicherplatz ermöglichen die Archivierung nahezu aller Informationen über Infrastruktur, Prozesse, Maschinen etc. Die Kunst der prognoseerstellenden Experten liegt darin, interessante und valide Angaben auszuwählen. Dies ist aufgrund eines Übermaßes an Daten nicht immer einfach. Bei einer gerade

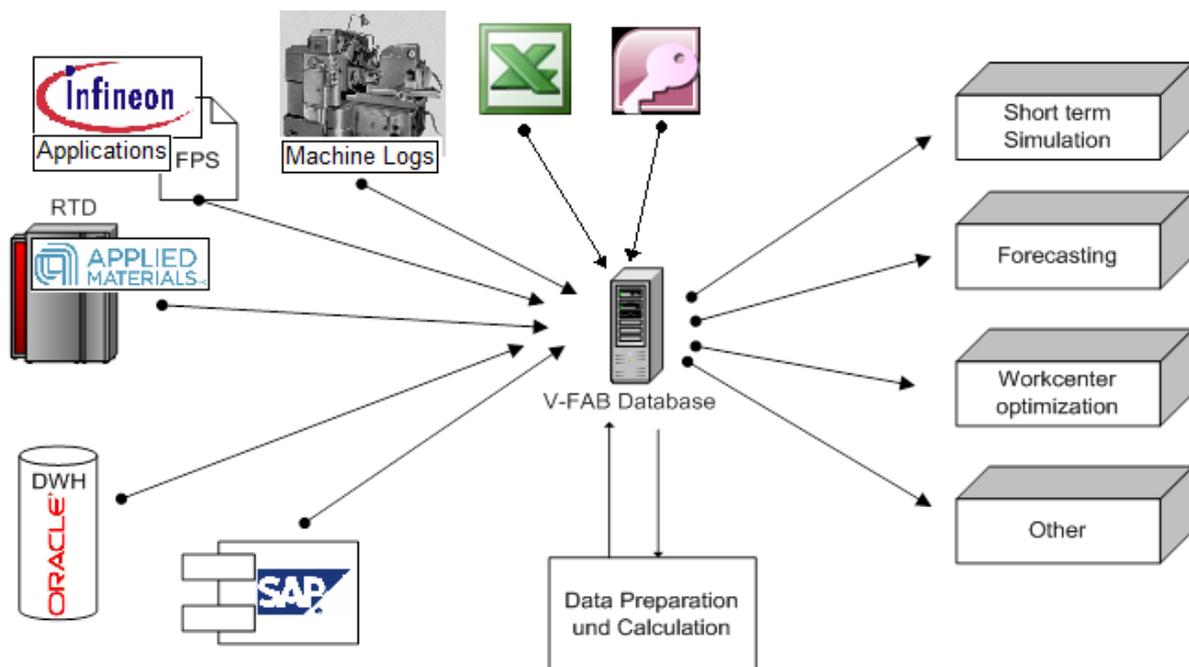


Abbildung 4-2: Schematische Darstellung der V-FAB Datenbank mit notwendigen Quellen

neu erbauten Fabrik, in der die Datenbankarchitektur noch nicht ausgereift ist, sind die notwendigen Angaben teilweise nicht zu finden. Bei bereits länger bestehenden Fertigungsstätten gestaltet sich das Problem anders. Datenbanken und unterschiedlichste Software wachsen unkontrolliert und unstrukturiert zu einem großen vernetzten System zusammen [Kohn08]. Heute übliche Standards sehen die Benutzung eines Data Ware House vor. Alte Strukturen auf neu einzurichten und die Gewohnheiten der Benutzer zu verändern, benötigt einen hohen

Arbeitsaufwand. Die Entwicklung von AFM wurde von derartigen Problemen geprägt. Im Rahmen des Projektes wurde der Versuch gestartet, eine gemeinsame Datenzwischenschicht für Kurzfristsimulation, AFM und Optimierung zu verwenden [Kohn08]. Bei unserem Projektpartner wird die Datenlandschaft durch Anwendung von diversen Softwarelösungen unterschiedlicher Hersteller gekennzeichnet. Abbildung 4-2 stellt einen Ausschnitt der notwendigen Interaktionen der verwendeten Datenbank mit funktionierenden Systemen des Industriepartners dar. Heterogenität und Alter der Softwareumgebung stellt einen Entwickler neuer Software vor typische Probleme beim Umgang mit verfügbaren Datenquellen (vgl. [Lee02] und [Gollas09]), wie:

- Verteilung der äquivalenten Daten – bspw. Pläne mit vorbeugender Wartung werden durch die jeweilige Abteilung betreut und in autonomen Datenquellen gespeichert,
- Dateninkonsistenz – bspw. mehrere Quellen für die Prozesszeit, also Datenbanken von Planungsabteilung, Prozessbetreuer oder Auszug aus historischen Bewegungsangaben,
- Fehlende Angaben – bspw. Samplingraten, die nur in zu grober Granularität erhältlich sind,
- Beschränkte Zugriffsmöglichkeiten – bspw. sicherheitsbegründete Zugriffsbegrenzung auf die produktionsrelevante Datenbanken, erlaubt wird meistens nur Zugriff auf vor-aggregierte Auszüge davon,
- Komplexität und Volumen von Datenbanken in der Halbleiterfertigung,
- Entwicklung von notwendigen Schnittstellen zwischen heterogenen Systemen.

Der Entwickler eines Prognosesystems muss Lösungen für derartige Probleme finden. Die Fehlerbehebung verlangt ausreichende Kenntnisse des Systemverhaltens und verbraucht dabei enorme Mengen an personalen Ressourcen. Die Methodik von AFM umgeht diese Hindernisse der Datenintegration, indem die Verwendung der Datenquellen auf eine valide Quelle minimiert wird. Es werden für die Vorhersage ausschließlich die historischen Losbewegungsdaten ausgewertet und mittels verschiedenen Algorithmen zu diversen Prozessstatistiken komprimiert. Weiterhin werden noch die Pläne der Materialeinschleusung verwendet. Durch die Datenquellenreduktion wird ein hohes Maß an Unabhängigkeit und Datenqualität erreicht.

### **4.2 Die Grundlagenfunktionen des Prognosealgorithmus**

Der Kern des Prognosealgorithmus basiert darauf, dass die Dauer von fertigungsbedingten Bewegungen der Lose auf einem Abschnitt einer vorherbestimmten Route bestimmt wird und

die gewonnenen Informationen in die Zukunft projiziert werden. Für diese Methode der Prognoseerstellung werden folgende Angaben verwendet:

- Zieloperation [*TO*],
- Startoperation [*SO*] – Position des Loses in der Fabrik zum Zeitpunkt der Ausführung einer Prognose,
- Route [*RT*],
- Produkt [*PD*],
- die Klasse der Bearbeitungsgeschwindigkeit [*SCI*],
- die Bearbeitungsdauer zwischen Start und Zieloperation für zugehörige Bearbeitungsgeschwindigkeitsklasse [*CtST*],
- Die Samplingrate an der Zieloperation [*SR*].

Die Prognosealgorithmen können anhand ihrer konkreten Aufgaben in eine von vier wesentlichen Funktionsgruppen eingeteilt werden:

- Erzeugung und Management von Statistiken,
- Vorhersagekalkulation,
- Reporting,
- Hilfsfunktionen.

Die Erzeugung einer Statistik wird mithilfe der Bestimmung eines Prognoseziels [*TO*] in Form einer Operation oder einer Menge an äquivalenten Operationen initialisiert. Die äquivalenten Produktionsschritte werden durch ein Äquivalenzmerkmal beschrieben, wie zum Beispiel eine zuständige Maschinengruppe, Prozessrezept oder -typ. Anschließend werden alle Routen [*RT*] spezifiziert, die in der Prozessreihenfolge die Zieloperation enthalten. Die Abbildung 4-3 stellt mögliche Konstellationen von Routen, Produkten und einen Zielschritt dar. Auf einer Route bewegen sich meistens mehrere unterschiedliche Produkte [*PD*], die zwar eines ähnlichen Typs sind, jedoch unterschiedliche, mittlere Durchlaufzeiten in der Fabrik aufweisen. Die Varianz lässt sich durch die Prozesszeitdifferenz, Priorisierung oder durch das Auslassen von alternativen Segmenten der Route begründen. Teile der Produkte in der Fabrik werden vorerst nur bis zur ersten Metallisierungsebene prozessiert und danach auf eine unbestimmte Zeit eingelagert.

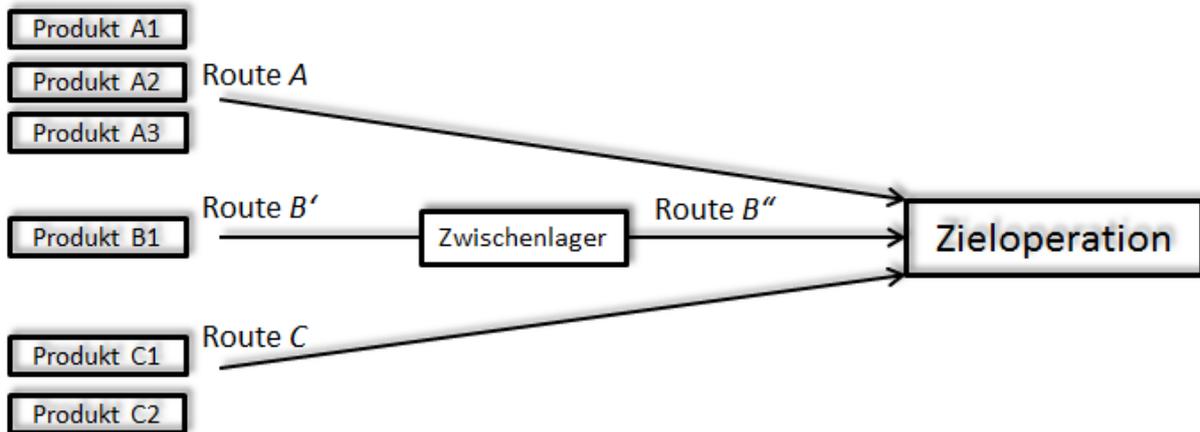


Abbildung 4-3: Beispiel einer Zieloperation und zugehörigen Routen sowie Produkte innerhalb eines Produktionsbereichs

Beim Verlassen des Vorratsraums bekommt ein Los eine neue Route zugewiesen. Deshalb ist eine Vorhersage für diese Lose unmöglich, die zwischen der derzeitigen Position und der Zieloperation ein Lager passieren müssen. Nach der Definition aller möglichen Tripel ( $[PD]$ ,  $[RT]$ ,  $[TO]$ ) werden die zugehörigen, statistischen Durchlaufzeiten von allen möglichen Startoperationen berechnet. Die, für die Vorhersage erforderlichen Statistiken für Sampling, Hold, Abarbeitungszeit usw. werden nach einmaliger Initialisierung periodisch upgedatet. Derartige Auffrischung statistischer Tabellen erfolgt unabhängig von der Prognosekalkulation. Als Basis für die Ermittlung der Informationen dienen historische Bewegungsdaten des Materials in der Produktionslinie. Kapitel 4.5 befasst sich detailliert mit Informationen zu diesem Thema. Wie in Abbildung 4-4 dargestellt, werden als mögliche Startoperationen  $[SO]$  alle Schritte der Route  $[RT]$  definiert, deren Ausführung vor dem Erreichen des Prognoseziels  $[TO]$  stattfinden. Die Definition von Quadrupel ( $[PD]$ ,  $[RT]$ ,  $[TO]$ ,  $[SO]$ ) erlaubt aber immer noch nicht das Auslesen oder die Manipulation von adäquater Durchlaufzeit in der AFM Durchlaufzeitstatistik.

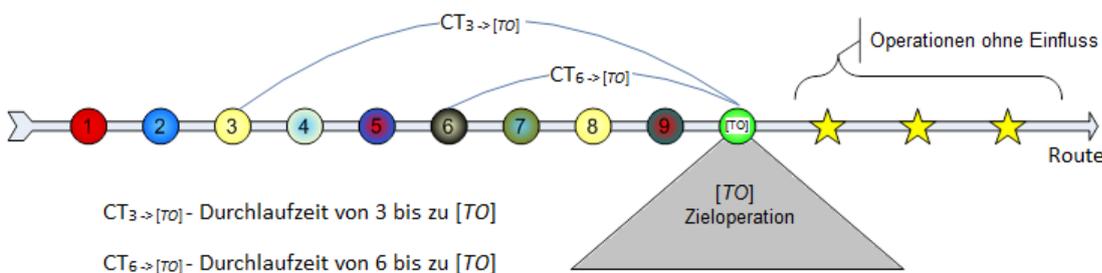


Abbildung 4-4: Schematische Darstellung einer Route mit Zieloperation und möglichen Startoperationen und dazugehöriger Durchlaufzeit

Aufgrund der Varianz der Fertigungszeiten von [SO] und [TO] wurde eine Einordnung anhand spezieller Loseigenschaften unternommen. Die genaue Beschreibung der Notwendigkeit und des Konzeptes einer Klassifizierung in [SCI] werden im folgenden Kapitel beschrieben.

Die Berechnung der Vorhersage erfordert genaue Angaben zur Lospositionen in der Fabrik. Die Initialen Operationen werden als Startoperationen [SO] gekennzeichnet. Für jedes Werkstück werden auch Angaben zu Produkt [PD] und Route [PD] ausgesucht. Erforderlich sind zusätzliche die Informationen für die Ermittlung der Klassen der Bearbeitungsgeschwindigkeit [SCI]. Anschließend werden möglichen Zieloperationen [TO] bestimmt. Ein Los kann mehrere Zieloperationen besitzen. Mithilfe der statistischen Durchlaufzeitangaben [CtST] werden die Zielankunftszeitpunkte berechnet, die dann zur Anlieferungsmenge aggregiert werden.

Mittels Reportingfunktionen werden die erarbeiteten Ergebnisse der Prognose anwendergerecht überarbeitet und dargestellt. Die Hilfsfunktionen dienen hauptsächlich der Kommunikation zwischen einzelnen Softwarepaketen und externen Bibliotheken. Die Funktionsweise und Implementierung von Reporting- und Hilfsalgorithmen werden im Rahmen der Beschreibung der Vorhersagesoftware erläutert.

### **4.3 Bestimmung einer Zieloperationsmenge für die Prognose**

Die AFM verwendet ein Kreuzprodukt von Routenname, Produktbezeichnung, Produktionsbereich und der Zieloperation zum Spezifizieren der Zielmenge für die Voraussage. Die Prognosemethode ermöglicht das Auslesen der möglichen Datensätze aus den Bewegungsdaten. Dadurch wird gewährleistet, dass die Zielangaben stets aktuell gehalten werden. Der Ausgang für die Spezifikation eines Parametersatzes als mögliches Vorhersageziel ist für den Anwender frei wählbar, wie zum Beispiel:

- ein Produktionsschritt, der typischerweise bei dem Verlassen eines Produktionsbereichs verwendet wird und die letzte Operation in einer Route darstellt,
- eine Anlagengruppe (Workcenter),
- eine Prozessrezept wird dann verwendet, wenn ähnliche Produktionsschritte an mehreren Stellen einer Route auftauchen und die Produktionspläne dies betreffend inhomogen sind.

Die Zielspezifikation wird von den meisten Statistikalgorithmen sowie bei der Prognoseerstellung als initialer Punkt angewendet. Der Nachteil von derartigen Verfahren ist die Wartezeit auf die Aktualisierung der Statistik bzw. die Notwendigkeit von manuellen Ausführungen der Auswertefunktionen bei der Einführung neuer Produkte.

### 4.4 Klassifizierung

Die Funktion einer Vorhersage ist es, zukünftige Ereignisse vorherzusehen. Die im industriellen Bereich verwendete Prognose muss ein großes Maß an Treffsicherheit aufweisen. Im Falle einer Materialanlieferungsprognose ist die Abweichung der prognostizierten zur realen Losmenge möglichst gering zu halten. Die Methodik der Berechnung von Anlieferungsmengen bei AFM basiert in einem ersten Schritt auf der Ermittlung der Ankunft einzelner Lose innerhalb einer Operation (auch Einzellosvorhersage). Die Zusammenfassung des Ankunftsstromes erfolgt in Schritt zwei durch Summieren der Werkstücke unter Anwendung von Samplingregeln. Demzufolge ist die Reduktion der Diskrepanz zwischen realem und prognostiziertem Eintreffen des Materials bei einer Operation die Hauptaufgabe. Die Untersuchungen der historischen Bewegungsdaten ergaben eine enorme Streuung der Durchlaufzeiten zwischen zwei bestimmten Operationen. Abbildung 4-5 stellt ein Histogramm der Häufigkeitsverteilung (vgl. [Scott79]) der Durchlaufzeitwerte zwischen Produktionsschritt A und B dar. Dabei wurden in einer längeren Zeitperiode alle Lose untersucht, die an den beiden Operationen bearbeitet wurden. Bereits in der Abbildung ist die enorme Streuung deutlich erkennbar und wird von quantitativen Ergebnissen bestätigt. Berechnungen ergaben, dass die Spannweite mehr als 200 % und die Standardabweichung mehr als 50 % des Mittelwertes betragen haben. Bei einer Entfernungsperiode von ca. 9 bis 10 Tagen zwischen A und B ist eine Fehlschätzung des Wertes um 5 Tage für eine Vorhersage nicht annehmbar. Aus der Notwendigkeit der Streuungsminimierung resultiert die erste Stufe der Klassifikation von Produktionsgegenständen. Es ist also eine separate Betrachtung einzelner Produkte erforderlich. Obwohl viele ähnliche Fabrikate gleiche Routen verwenden, liefert die Gruppierung mittels Bearbeitungswege deutlich schlechtere Ergebnisse in Bezug auf die Durchlaufzeitstreuung. Die Abbildung 4-6 stellt einen Vergleich der Histogramme der Häufigkeitsverteilung des prozentualen Verhältnisses von der Standardabweichung der Durchlaufzeit zu dem Mittelwert der Durchlaufzeit zwischen zwei Operationen dar. Klassifiziert wurde anhand der Produktbezeichnung oder Routencharakteristik. Die spezifischere Klassifizierung, im Diagramm blau, liefert mit Augenmerk auf die Minimierung der Streuung des Arbeitsablaufes zwischen zwei Operationen deutlich bessere Ergebnisse.

Über die Bildung von Produktgruppen konnte eine weitgehende Reduzierung der Durchlaufzeitstreuung nur teilweise erreicht werden. Abbildung 4-7 stellt einen Vergleich bezüglich der Häufigkeitsverteilung der notwendigen Bearbeitungsdauer zweier unterschiedlichen Produkte zwischen den Produktionsschritten A und B. Um Hindernisse auf dem Weg zu validen Prog-

nosen zu veranschaulichen, wurden explizit zwei zueinander inhomogene Fabrikate ausgesucht.

Einerseits differieren sie bezüglich der Durchlaufzeitdauer und andererseits ist die Durchlaufzeitstreuung des ersten Produkts wesentlich geringer als die des zweiten. Ähnliche Unregelmäßigkeiten bezüglich der Durchlaufzeitdispersion wurden auch bei fast der kompletten Produktpalette nachgewiesen. Die extrem breite Verteilung der Bearbeitungsperioden, wie zum Beispiel bei Produkt 2 in Abbildung 4-7, zwingt zu weiteren Versuchen, um die Streuung zu minimieren. Bei derartigen Nachforschungen sollte vorerst analysiert werden, welche Einflüsse auf die Durchlaufzeit existieren. Die charakteristischen Einflussgrößen in der Halbleiterfertigung, welche Durchlaufzeitschwankungen verursachen, wurden in der Literatur umfangreich diskutiert [Atherton95], [Lu94], [Rose99]. Die Variabilität in der Produktionsdauer haben ihren Ursprung vor allem in:

- Produktionssteuerung,
- Auslastungsschwankungen,
- Stochastik,
- Materialverteilung in der Fabrik,
- Losgröße,
- anderen Effekten.

### **4.4.1 Einfluss der Produktionssteuerung auf die Durchlaufzeiten**

Die Materialflusssteuerung der Halbleiterfabrik basiert hauptsächlich auf einer oder mehreren globalen bzw. lokalen Abfertigungsregeln. Die Regeln werden durch Simulation oder eine andere Art der Erprobung genau an die Produktionsstätte angepasst. Dabei spielen sowohl die ökonomischen Ziele (möglichst bester Durchsatz als Musterbeispiel) wie auch die Gewährleistung von kundenorientierten Qualitätsparametern (beispielsweise Liefertreue oder kurze Lieferzeiten ab Bestellung) eine beachtenswerte Rolle. Da die Anwendung und Funktionsweise verschiedener Dispatch-Regeln in der Literatur umfangreich beschrieben wurde [Atherton95], [Rose99] ist deren Einfluss auf die Durchlaufzeit bei der Problematik von Anlieferungsprognosen interessant. Bei unserem Projektpartner aus der Industrie wurde, wie auch bei den meisten führenden Halbleiterherstellern, global (mit mehreren lokalen Ausnahmen) eine erweiterte Version der Abfertigungsregel Operational Due Date (ODD) angewendet. Die Regel sortiert die Materialliste in der Warteschlange vor einer Operation mittels der Fälligkeitstermine der Fertigstellung des aktuellen Produktionsschritts. Es werden also die Lose bevorzugt, die zuerst an dieser Operation fertig sein sollen. Die ODD lässt als Fabriksteuerung zwar

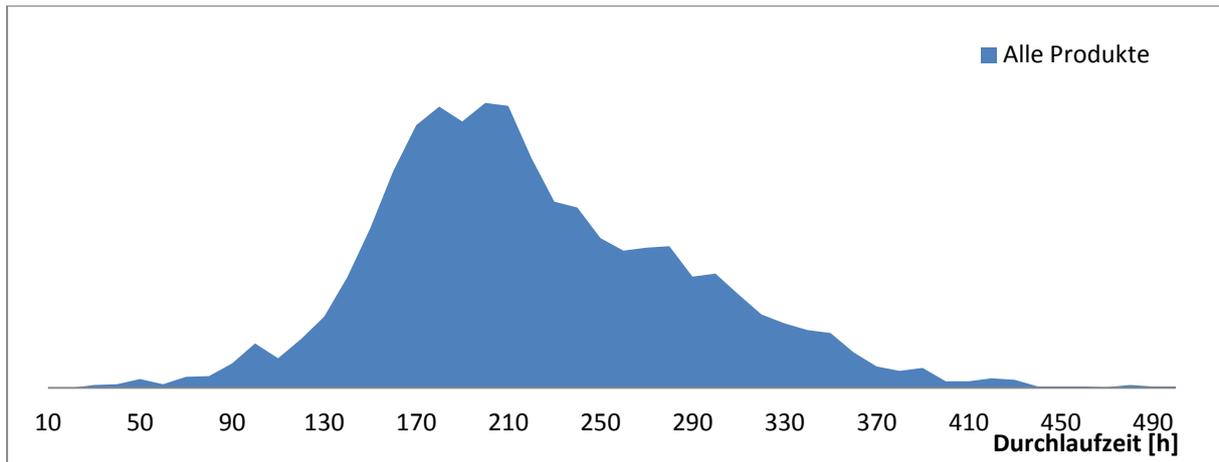


Abbildung 4-5: Häufigkeitsverteilung der Durchlaufzeitwerte zwischen Schritt A und B

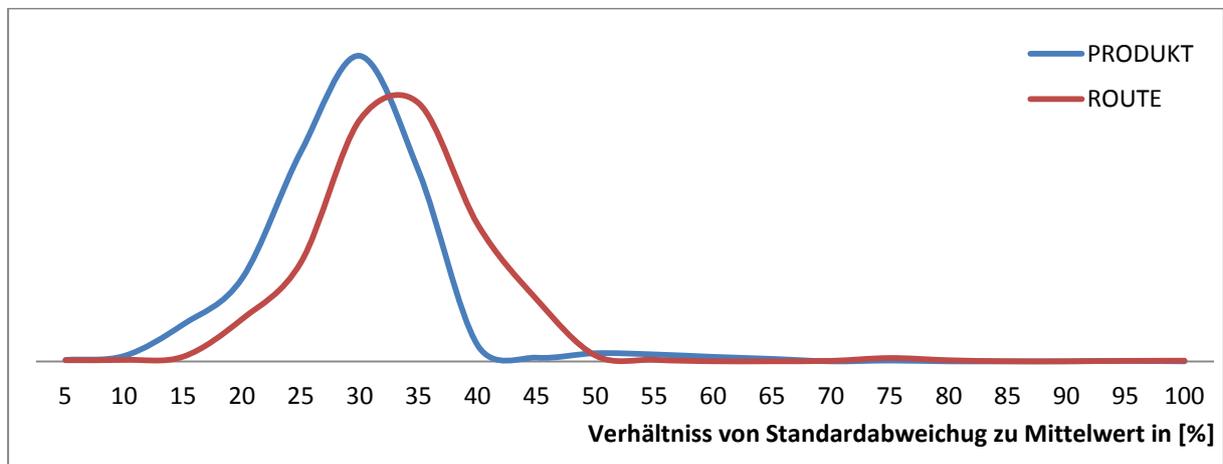


Abbildung 4-6: Häufigkeitsverteilung des prozentualen Verhältnisses der Standardabweichung zum Mittelwert der Durchlaufzeit zwischen zwei Operationen für mithilfe von Produkt und Route gebildeten Klassen

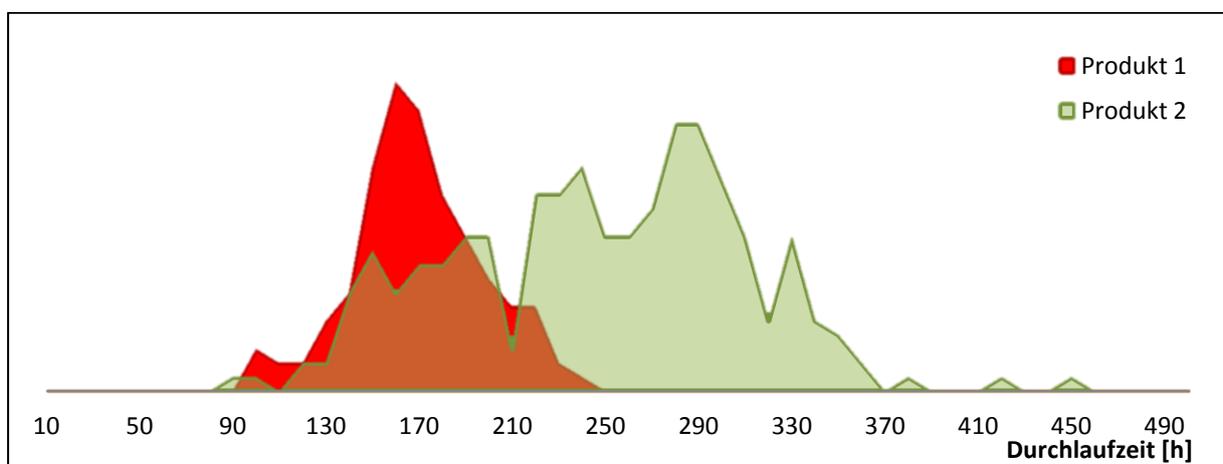


Abbildung 4-7: Häufigkeitsverteilung der Durchlaufzeitwerte zwischen Schritt A und B für unterschiedliche Produkte A und B

die nachträgliche Beschleunigung einzelner Werkstücke durch Änderung des Fertigstellungsdatums zu, dabei fehlt jedoch deren spezielle Kennzeichnung. Ein derartig begünstigtes Los taucht dann am Anfang der Abarbeitungsliste auf und bewegt sich schneller durch die Produktionslinie – dem Fertigungspersonal fehlt aber die Information über die Sonderbehandlung.

Die Handhabung einer derartigen Methode bei Sonderrechtvergabe ist bei mehreren tausend geplanten Fertigstellungsterminen sehr kompliziert. Aus diesem Grund wird eine zusätzliche Priorisierung der Fabrikate eingeführt. Als charakteristisch für die Halbleiterfertigung [Bonvik94], [Rose05] werden folgende Prioritätentypen genannt:

- Rocket (engl. Rakete) – meistens höchste Priorität, für das Los wird die Anlage beim nächsten Schritt freigehalten, damit die Fertigung ohne Verzögerung stattfinden kann. Der Transport erfolgt manuell ohne Verwendung des Transportsystems,
- Hot (engl. warm, gefragt) – zweithöchste Priorität, bevorzugt standardisierte Abarbeitung,
- Normal – keine Priorisierung,
- Anderes – es existieren weitere Priorisierungsarten, die jedoch fertigungsspezifisch sind. Das Material wird zwischen Hot und Normal priorisiert.

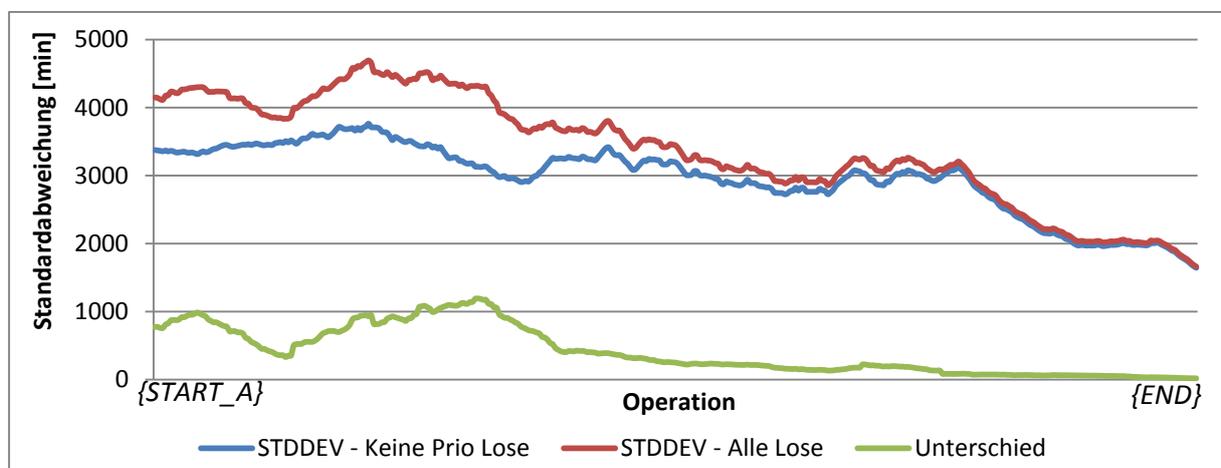
Die Untersuchung der Fabriksteuerung ergab mindestens zwei Parameter, die für eine Losklassifizierung und mögliche Reduzierung der Streuung der Durchlaufzeit aufschlussreich sein können. Der Einfluss der Prioritätensteuerung auf die Dauer der Abarbeitungsperiode und damit auch auf die separate Behandlung von Losen mit unterschiedlichen Prioritäten lässt sich trivial nachweisen. Die Berechnung beginnt mit einer Markierung der letzten bzw. ausgewählten Operation in der Produktion als Zieloperation  $\{END\}$ . Die Bewegungsdaten aus der Vergangenheit werden zuerst nach Produkten  $\{PROD\_*\}$  gruppiert und anschließend die Produktgruppen in weitere von der Startoperation  $\{START\_*\}$  abgängige Cluster geteilt. In jeder der Klassen erfolgt eine Berechnung der Standardabweichung  $STDDEV_{\{PROD\_*\}\{START\_*\}\{END\}}$  für die jeweiligen Zeitunterschiede zwischen End- und Startoperation. Dadurch ist die Erstellung einer Matrix, wie in Tabelle 4-1 abgebildet, möglich.

Die Kalkulation des Mittelwerts der Abweichung für alle Produkte  $\overline{STDDEV}_{\{PROD\_ALL\}\{START\_*\}\{END\}}$  ermöglicht eine Aussage über den kompletten Fertigstellungsprozess. Im Anschluss werden alle priorisierten Lose ausgefiltert und die Berechnungsprozedur wird wiederholt. Abbildung 4-8 stellt einen Vergleich zwischen dem Verlauf der Standardabweichungen, mit und ohne priorisierten Losen, dar. Die von gesamtem Losbestand

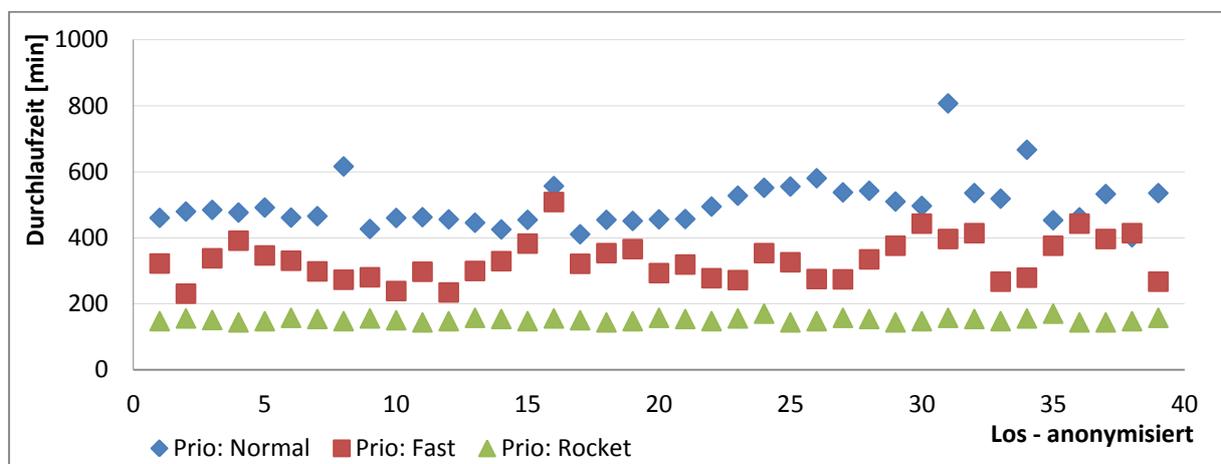
## Das Konzept der Analytischen Vorhersagemethode – AFM

Produkt	{Start_A}	{Start_B}	{Start_C}
{Produkt_1}	$STDDEV_{[PROD_1][START_A][END]}$	$STDDEV_{[PROD_1][START_B][END]}$	$STDDEV_{[PROD_1][START_B][END]}$
{Produkt_2}	$STDDEV_{[PROD_2][START_A][END]}$	$STDDEV_{[PROD_2][START_B][END]}$	$STDDEV_{[PROD_2][START_B][END]}$
{Produkt_3}	$STDDEV_{[PROD_3][START_A][END]}$	$STDDEV_{[PROD_3][START_B][END]}$	$STDDEV_{[PROD_3][START_B][END]}$
Mittelwert	$\overline{STDDEV}_{[PROD\_ALL][START_A][END]}$	$\overline{STDDEV}_{[PROD\_ALL][START_A][END]}$	$\overline{STDDEV}_{[PROD\_ALL][START_A][END]}$

**Tabelle 4-1: Beispiel einer Matrix von Standardabweichungen für die Klassen entstanden durch Kreuzprodukt von einem Produkt und zugehöriger Zeitperiode von Startoperation zu Zieloperation für alle Lose aus historischen Bewegungsdaten in der Fabrik**



**Abbildung 4-8: Der Vergleich von Standardabweichung der Durchlaufzeit in Klassen gebildet durch Start- und Endoperation mit Gesamtheit an Werten und nach Beseitigung durch Priorität beschleunigter Lose**



**Abbildung 4-9: Klassifikation der Lose mithilfe ihrer Priorität**

errechneten Streuungsmaße, rote Kurve, sind im gesamten Verlauf des Wertebereichs größer also schlechter als die, geschildert mit der blauer Kurve, bei der priorisierte Material aus der Berechnung ausgefiltert wurde. Aus diesen Gründen wurde das Priorität als weiteres Klassifizierungskriterium, mit 3 an möglichen Werten: Rocket, Hot incl. alle anderen Prioritäten und Normal, eingeführt.

Die Verspätung der Lose bezüglich festgelegter Fertigstellungstermine in aktuellen Operationen wurde als weiterer Klassifizierungsparameter untersucht. Die Auswahl des Klassifizierungsmerkmals wurde mit der Anwendung von ODD als globale Abfertigungsregel begründet. Den Beweis, dass die Durchlaufzeit abhängig ist von der Verspätung an der Startoperation liefert die negative Korrelation beider Zahlenreihen. Im Detail bedeutet dies, je verspäteter ein Werkstück ist, desto kürzer wird der Warteschlangenaufenthalt und dadurch auch das gesamte Pensum an Abarbeitungszeit bis zur Zieloperation. Als problematisch erweist sich die Auswahl der Begrenzung geeigneter Klassen, also die minimale und maximale Verspätung. Der Prognoseansatz erfordert die gleichen Merkmale für alle Produkte. Die Lösung dieses Problems liefert eine Clusteranalyse, welche die Strukturentdeckungen in großen Datenbeständen ermöglicht [Bacher10]. Die Gruppierung erfolgte durch die Anwendung des k-Means-Algorithmus, der eine Menge von Objekten in einer vorbestimmten Clusteranzahl  $k$  systematisiert [Kanun02]. Der rechnerisch sehr effektive Algorithmus startet mit einer zufälligen Verteilung der Klassifizierungsschwerpunkte, denen die Objekte nach ihren am nächsten liegenden Schwerpunkten zugeordnet werden. Für die Entfernungsberechnung wird eine geeignete Distanzfunktion verwendet. Anschließend ermittelt das Verfahren immer wieder das Zentrum eines Clusters und führt die Objektklassifizierung erneut durch. Terminiert wird erst dann, wenn sich die Clusterzentren nicht mehr bewegen. Um die Clusteranalyse anzuwenden, wurden die historischen Bewegungsdaten unter Verwendung von anderen, bereits bekannten Klassifizierungsmerkmalen zu Datensets zusammengefasst. Datensätze dieser Art wurden durch Produktzugehörigkeit, Prioritätsklasse, Start- und Zielschritt und andere Parameter begrenzt. Zu erforschen war die Gruppierungsmöglichkeit mithilfe der Verspätung der Startoperation und der Durchlaufzeit zwischen Start und Ziel. Zuerst wurde die Clusteranalyse mit den beiden Parametern als Eingangsvariablen und der euklidischen Norm als Distanzfunktion durchgeführt. Die Inhomogenität des Durchlaufzeitverhaltens in Abhängigkeit vom Zeitverzug, ein enormer Einfluss stochastischer Effekte wie auch die Funktionsweise des k-Means-Algorithmus selbst, führte zu Misserfolgen bei der Klassifizierung (ähnlich wie in [Backus06]). Abbildung 4-10 stellt ein Punktwolkendiagramm eines ausgewählten Datensatzes dar, der auf die beschriebene Weise gruppiert wurde. Problematisch ist die fehlende Ein-

deutigkeit der Verspätung zu der benötigten Durchlaufzeit. Cluster 2 besitzt im Vergleich zu Cluster 4 eine nahezu identische Verteilung der Fertigstellungszeit, jedoch verläuft die Verspätung in beiden Fällen von Grund auf unterschiedlich. Die Umkehrung des Problems wurde durch Cluster 3 und 4 verdeutlicht, da hier bei einem ähnlichen Verhalten bezüglich des Verzugs die Durchlaufzeitstreuung divergiert. Da die Prognoseerstellung nahezu in Echtzeit erfolgen und nicht rechenintensiv sein sollte, wurde eine möglichst einfache Methodik gesucht. Die Klassifizierungen gemäß der Verspätung wurden als fest, allgemeingültig und damit auch produktunabhängig spezifiziert. Da der bereits beschriebene Ansatz der Clusterbildung nicht erfolgreich war, wurde eine andere Art des Vorgehens gesucht, die mittels entsprechender Klassifizierung des Datenmaterials zur Reduktion der Durchlaufzeitstreuung innerhalb der Gruppen führen sollte. Die erwartete Dispersionsverringerung wird durch Gruppenbildung nach der Definition des k-Means-Clusteranalyseverfahrens erreicht, jedoch werden die Eingabedaten auf die Durchlaufzeit zwischen zwei Operationen für eine vorbestimmte Produktgruppe begrenzt. Die Auswahl  $k$  als Anzahl der Klassen ist frei wählbar und steht in sehr enger Beziehung mit dem Gegenstand der Untersuchungen. Die Einstellung aller Parameter lässt sich erst durch die Validierung der Vorhersageergebnisse kontrollieren. Daraus resultiert ein erheblicher Aufwand bei der Bestimmung der Gruppenzahl. Abbildung 4-11 stellt das Ergebnis der K-Means-Klassifikation mithilfe der Durchlaufzeit dar. Ein Vergleich der Diagramme von Abbildung 4-10 und Abbildung 4-11 verdeutlicht, dass die Probleme der Überlagerung bezüglich der in der Bearbeitungszeit ähnlichen Stichproben, über Anwendung dieser Verfahren behoben wurden. Die Problematik von Schnittmengenbildung von benachbarten Clustern in Bezug auf die Verspätung besteht weiterhin und wurde mittels Boxplot-Diagramm in Abbildung 4-12 veranschaulicht. Aus diesem Grund erweist sich eine Festlegung von Grenzen für jedes Cluster als schwierig. Im Rahmen des Projektes wurde ein Berechnungsverfahren entwickelt, das durch eine mehrfache Approximation die Ermittlung von allgemeingültigen, produktabhängigen Intervallen über die Verspätung ermöglicht. Angenommen werde, dass die historischen Bewegungsabläufe nach der Durchlaufzeit  $[ct]$  zwischen Start- und Zieloperation ausgewertet und anschließend angesichts ihrer Produktzugehörigkeit und Priorität zu eine der  $n$   $G_{SZPP}$  Gruppen zugeteilt werden. Die zeitliche Entfernung zwischen Start und Ziel sollte mithilfe des Vorhersagehorizonts begrenzt werden. Jeder derartiger Datensatz wird, wie bereits beschrieben, in  $k$  Cluster  $CL_k$  nach benötigter Abarbeitungszeit geteilt. Bei einer Gruppenanzahl von  $k$  entstehen  $k+1$  Gruppengrenzen  $G_i = \{G_0, G_1, \dots, G_k\}$ . Die äußeren Grenzen sind für jede  $G_{SZPPn}$  bekannt und betragen  $G_0 = -\infty$  und  $G_k = +\infty$  also bis unendlich zu früh bzw.

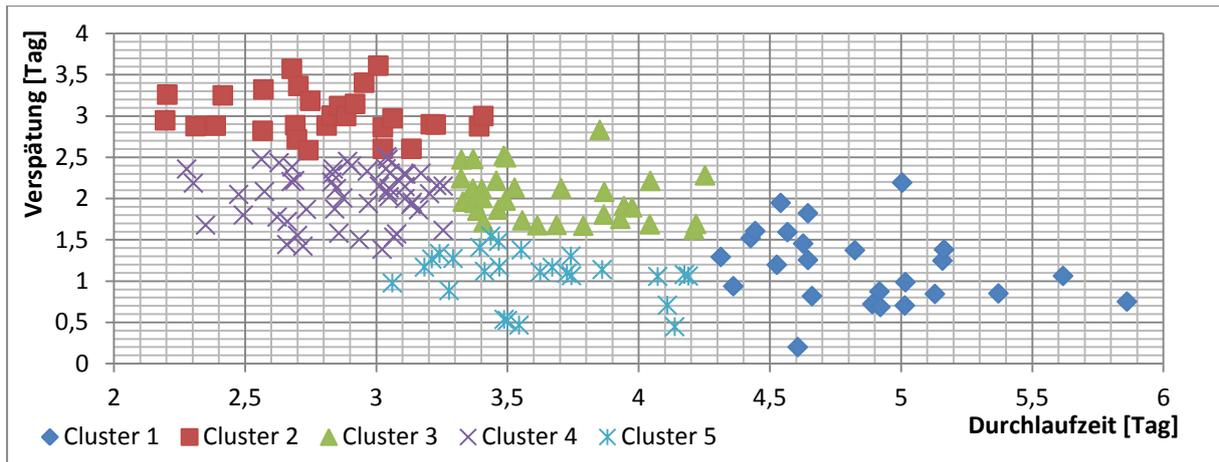


Abbildung 4-10: Clusteranalyse mittels K-Means-Algorithmus bei Verwendung von Verspätung und Durchlaufzeit als Ankunftsdatenstrom

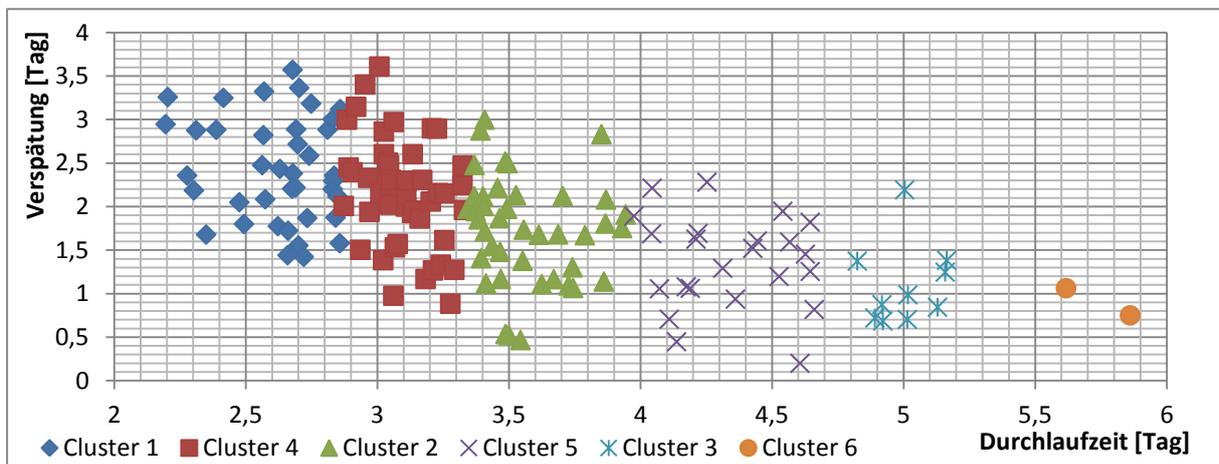


Abbildung 4-11: Clusteranalyse mittels K-Means-Algorithmus bei Verwendung von Durchlaufzeit als Ankunftsdatenstrom

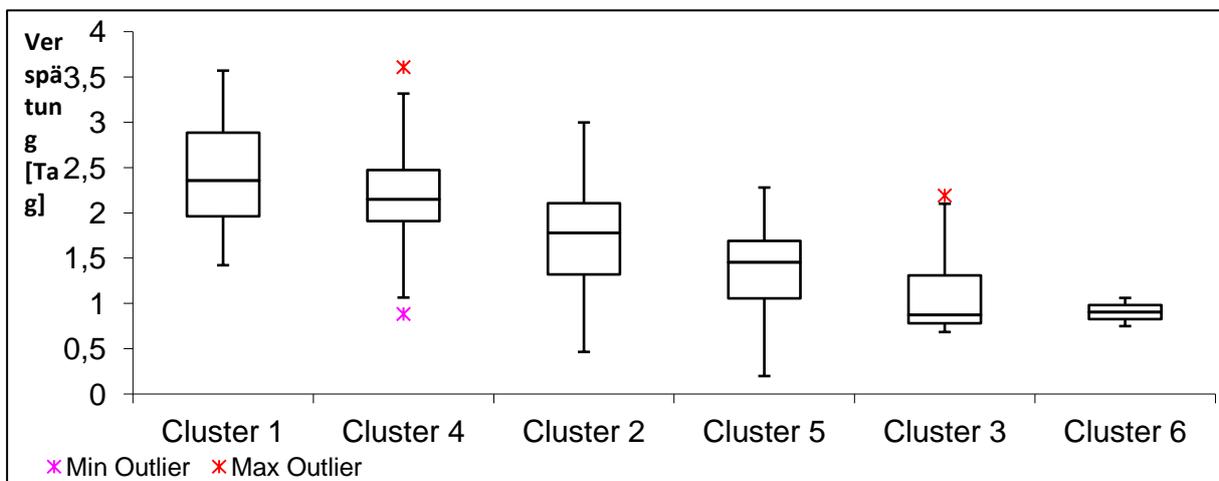


Abbildung 4-12: Boxplot der Verspätungsverteilung innerhalb der Cluster, die mittels K-Means-Verfahren bei Verwendung von Durchlaufzeit als Ankunftsdatenstrom erzeugt wurden

zu spät. Interessanter sind aber die inneren Clusterränder. Die Streuung der Verspätung innerhalb eines Clusters  $k$  kann zum Teil enorm sein, wenn die Grenze mittels einem 1.  $Q1$  und 3.  $Q3$  Quartil geschätzt wird. Nur 50 % der Werte weisen somit Gültigkeit auf. Die Ermittlung der Gruppengrenze  $G_i$  wird durch die Kalkulation der Anzahl aller Elemente

$$|A_{CL_{i-1}CL_i}| = \sum_{e=1}^n |A_{e_{CL_{i-1}}}| + |A_{e_{CL_i}}|$$

von zwei benachbarten Cluster  $CL_{i-1}$  und  $CL_i$  über alle  $n$  Gruppen  $G_{SZPP}$  initialisiert, wobei  $|A_{e_{CL_{i-1}}}|$  – die Anzahl von Elementen der durch Cluster  $CL_{i-1}$  begrenzten Menge für eine  $e$ -te Gruppe  $G_{SZPP}$  bedeutet.

Die Summe ist unerlässlich für die Gewichtung einzelner Cluster bezüglich deren Einfluss auf die Bestimmung der Gruppengrenzen. Die Mächtigkeit einer Menge begrenzt durch  $CL_k$  ist direkt proportional zur Bedeutsamkeit bei der Bestimmung der Grenzen. Dadurch wird gewährleistet, dass die Grenzen eher durch Massenprodukte entstehen, also diejenigen Produkte, deren Einwirkung auf die gesamten Resultate der zukünftigen Prognosen bedeutsamer ist. Die Gruppengrenze  $G_i$  wurde durch Ermittlung eines gewichteten Mittelwertes geschätzt:

$$G_i = \frac{1}{|A_{CL_{i-1}CL_i}|} \sum_{e=1}^n (|A_{e_{CL_{i-1}}}| * Q3_{e_{CL_{i-1}}} + |A_{e_{CL_i}}| * Q1_{e_{CL_i}})$$

Die Anwendung des unkomplizierten Verfahrens ermöglicht zwar das Ausrechnen der Grenzen, garantiert aber deren Gültigkeit nicht. Die Zahlenreihe  $G_1, G_2, \dots, G_{k-1}$  ist nach der Definition streng monoton steigend. Wenn aber durch die Bestimmung der Grenzen die Annahme der Monotonie verletzt wird, ist ein manueller Eingriff zwingend erforderlich. Einerseits besteht eine Möglichkeit, benachbarte Cluster zusammenzufassen, andererseits ist die erneute Kalkulation der Grenzen mit einer veränderten Clusteranzahl meist aufschlussreicher.

Der Erfolg derartiger Klassifizierungsansätze korreliert positiv mit der Stabilität der Fertigung und der Qualität der gesammelten Daten. Die Anwendung anderer komplizierterer Verfahren befindet sich zurzeit in der Untersuchungsphase. Es werden dennoch bei differenzierteren Fabriksteuerungen weitere bzw. abweichende Klassifikationsmerkmale verwendet. Ein gutes Beispiel dafür ist die SPTF<sup>18</sup> Regel [Rose01]. Diese Abfertigungsregel bevorzugt die Lose mit kürzeren Prozesszeiten in einer gegebenen Operation. Die Verspätung bezüglich ihres Fälligkeitsdatums spielt also in diesem Fall keine Rolle. Das Resümee der Untersuchungen zeigt die Notwendigkeit einer genauen Analyse der Fertigungsstätte mit Fokus auf die Abfer-

---

<sup>18</sup> SPF – engl. shortest processing time first

tigungsregeln und deren Einfluss auf die Durchlaufzeit. Im Falle der beschriebenen Fertigung wurden die Faktoren für die Losklassifizierung um die Faktoren Priorität und Verspätung erweitert.

### 4.4.2 Auslastungsschwankungsbedingte Durchlaufzeitänderungen

Die Zusammenhänge zwischen Auslastung und Durchlaufzeit, wie auch anderen Performanceparametern in Warteschlangensystemen, wurden bereits 1961 von John D. C. Little formuliert und bewiesen [Little61]. Die Auslastung  $\rho$  wird als ein Quotient der Ankunftsrate  $\lambda$  durch die Kapazität  $\mu$  definiert. Das Little-Theorem präzisiert den Begriff Umlaufbestand<sup>19</sup>  $WIP$  als Produkt der Ankunftsrate  $\lambda$  und der Durchlaufzeit  $t_c$ . Aus der Kombination beider Formeln, wie unten dargestellt, wird das umgekehrte proportionale Verhältnis der Durchlaufzeit

$$\rho = \frac{\lambda}{\mu} \text{ und } WIP = \lambda * t_c \Rightarrow t_c = \frac{WIP}{\rho * \mu}$$

und Auslastung deutlich. Wichtig dabei ist allerdings, dass ein so beschriebenes System nahezu ideal und störungsfrei arbeitet. Da in Halbleiterfertigungssystemen die Parameter auch von

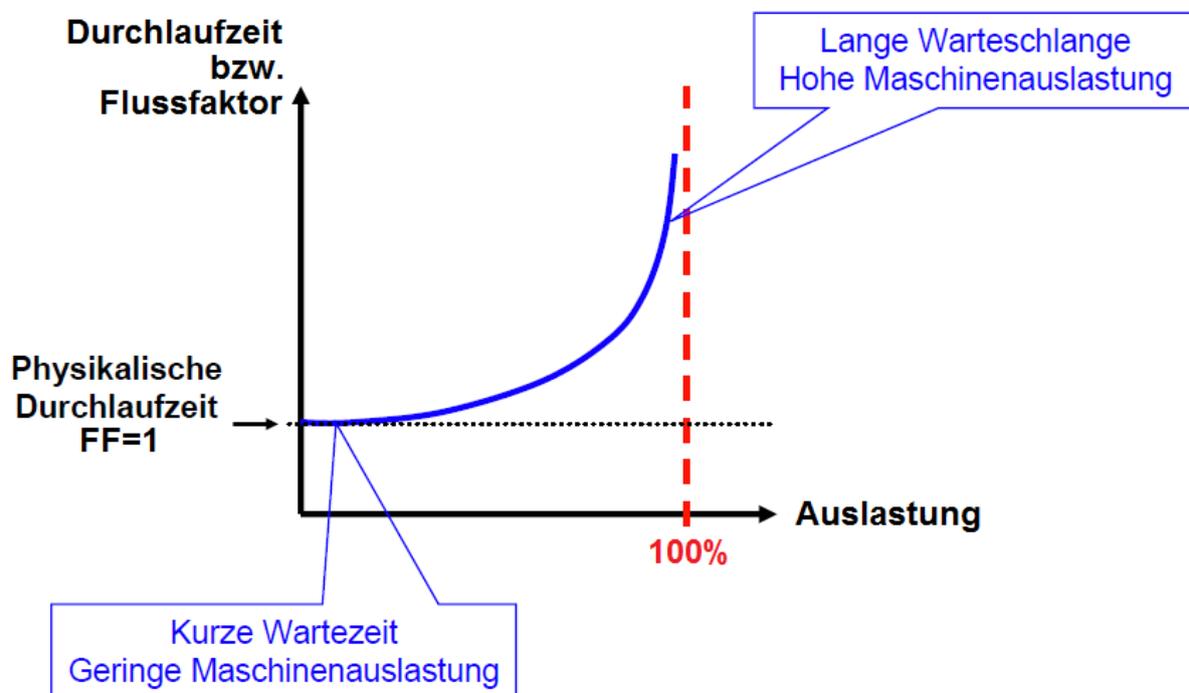
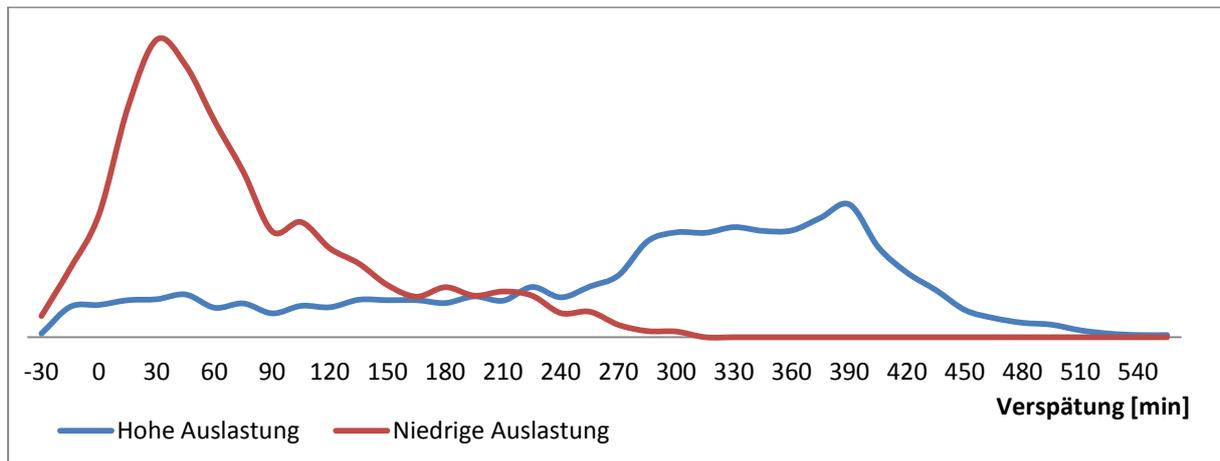


Abbildung 4-13: Beispiel einer Betriebskennlinie für ein Halbleiterwerk

<sup>19</sup> meistens wird das englische Begriff „Work in Process“ – WIP verwendet.

weiteren, zum Teil stochastischen Einflüssen abhängig sind, behelfen sich die Experten mit Erstellung einer Betriebskennlinie [Aurand97], [Fowler97]. Die in Abbildung 4-13 dargestellte charakteristische Kurve beschreibt jedoch, dass in einem Halbleiterwerk bei steigender Auslastung auch Durchlaufzeit und somit der Flussfaktor länger bzw. grösser werden. Obwohl die momentane Fabrikauslastung einer der wichtigsten durchlaufzeitbeeinflussenden Größen ist, hat sie nur einen bedingten Einfluss auf die Losklassifizierung und damit die Prognosebildung. Die Durchlaufzeit setzt sich aus Prozessierungs- und Wartezeit zusammen. Die steigende Fabrikauslastung hat keinen Einfluss auf die Prozessdauer (ausgenommen steigendes Ausfallrisiko) sondern nur auf die Länge des Aufenthalts in der Warteschlange. Daraus resultiert eine nahezu identische, von der Auslastung unabhängige Abarbeitungsdauer der priorisierten Lose. Der Einfluss der Auslastungsschwankungen macht sich aber in der Losverspätung bezüglich des Fälligkeitstermins, dem zweiten ausgesuchten Merkmal, bemerkbar.



**Abbildung 4-14: Häufigkeitsverteilung der Verspätung an einer gegebenen Operation bei hoch und niederausgelasteter Fabrik**

Das in Abbildung 4-14 dargestellte Histogramm der Verteilung der Verspätung für alle Produkte innerhalb einer Operation verdeutlicht das Problem. Die Streuung des Zeitverzugs und damit die Breite der Verteilung ist bei einer hoch ausgelasteten Fabrik deutlich höher als bei einer mit niedrigerer Last. Infolgedessen ist die Approximation der in der Statistik bekannten Verteilungsfunktionen erschwert. Die Problematik von wechselnder Fertigungsauslastung impliziert die Notwendigkeit des Monitorings dieser Kennzahl bezüglich ihrer Veränderungen. Die Reaktion auf die Auslastungsänderung ist von der Größe der Laständerung abhängig, wie zum Beispiel:

- Schwankungen bedingt durch ausgewogenen Produktionsmixwechsel – meistens geringe Änderungen der Auslastung, die sich in der statistischen Auswertung der Zeiten zwischen Start- und Zieloperation widerspiegeln. Die Lösung ist die höhere Gewich-

tung des Informationsgehalts von jüngeren Daten bei der Auswertung der Historie wie auch regelmäßige Updates der Statistik.

- Kurzzeitige Unter- oder Überladung – ähnlich wie bei ausgewogenem Produktionsmixwechsel sind keine Änderungen in der Klassifizierung notwendig.
- Gravierende Änderungen der Auslastung bedingt durch die Marktsituation (wie zum Beispiel die Krise im Jahr 2008 im Vergleich mit Aufschwung im 2011) – Notwendigkeit der erneuten Untersuchung der Verspätungsschwankungen und Überarbeitung der Klassengrenzen.

Die Güte der Materialflussprognose in den Fabriken mit der Auslastung im oberen Grenzbereich wird durch die enorme Streuung der Durchlaufzeit (vgl. Abbildung 4-14) beschränkt. Der negative Einfluss der Streubreite auf die Vorhersagequalität lässt sich zum Teil nur durch die Erweiterung der Klassenanzahl verringert. Resümierend lässt sich feststellen, dass die Einstellung aller Bestimmungsparameter der Prognosealgorithmen bezüglich des Fabrikzustands weitgehende Systemkenntnis, viel Feingefühl und eine umfassende Validierung mit unterschiedlichen Parameterwerten erfordert.

### **4.4.3 Prozesszeitunterschiede bedingt durch Losgröße**

Ein Los besitzt in der Halbleiterfertigung üblicherweise 25 Wafer. Die Zahl kann jedoch aus verschiedenen Gründen kleiner sein. Einerseits sind es Ausschusswafer, die nach einem Prozessfehler aus dem Los entfernt werden, andererseits werden beim Start die Horden nach Bedarf mit geringerer Anzahl an Siliziumscheiben bestückt. Eine weitere Variante entsteht durch Aufteilen des Loses in mehrere Teile, was in der englischen Sprache auch split bzw. splitten genannt wird. Ein derartig geteiltes Los bewegt sich getrennt über die Fertigung, zum Teil sogar mit abweichenden Prozessketten und dient häufig zu Entwicklungszwecken.

Die Abarbeitungsdauer zwischen Start und Zieloperation besteht aus der Summe von allen Wartezeiten und der gesamten Bearbeitungsdauer in den Anlagen. Im Werk des Projektpartners basiert die Steuerung, wie bereits im Kapitel 2.5.1 beschrieben, auf dem Fälligkeitstermin, der durch Priorisierungsregeln ergänzt ist. Das Steuerungskonzept sieht keinen Einfluss der Losgröße auf die Warteschlangenposition eines Loses vor. Die unterschiedliche Anzahl der Wafer korreliert jedoch bei den meisten Prozessstypen positiv mit der Länge der Prozesszeit. Das bedeutet also, je mehr Siliziumscheiben zu verarbeiten sind, desto länger wird auch das Prozessieren dauern. Die Maschinen in der Halbleiterfertigung lassen sich in folgende Gruppen bzgl. der Abarbeitungsmodi unterteilen:

- Einzel-Wafer Prozessierung [May06],
- Mini-Batch Prozessierung [Schmidt06] [Hosoe07],
- Batch Prozessierung [May06] [Mönch09].

Die Einzel-Wafer Anlagen werden nur mit einzelnen Wafer zur gleichen Zeit beladen. Anschließend erfolgt die Prozessierung. Erst nach erfolgreichem Abschluss des Entladevorgangs startet die Maschine die Bearbeitung der nächsten Siliziumscheibe. Bei derartiger Abarbeitung ist der Einfluss der Losgröße auf die Prozesszeit enorm. Die Analyse der notwendigen Bearbeitungszeit einiger Tools in der Fabrik hat die Hypothese bestätigt. Abbildung 4-15 stellt die Abhängigkeit der Prozesszeit von der Anzahl der Wafer in der Kassette bei Einzel-Wafer Prozessierung in einem Messgerät vor. Vereinfacht lässt sich der Zusammenhang der Prozesszeit und Losgröße als lineare Funktion beschreiben:

$$PT_{EWA}(w) = w * a + b$$

wobei  $w$  die Anzahl der Wafer und  $PT_{EWA}(w)$  die entsprechende Prozesszeit bedeuten. Die Parameter  $a$  und  $b$  sind konstant und prozessspezifisch. Die Hilfsgröße  $b$  ist von der Anzahl der Wafer unabhängig, wie zum Beispiel: die Dauer für das Finden und Laden eines notwendigen Abarbeitungsrezepts. Die Konstante  $a$  charakterisiert die erforderliche Bearbeitungszeit der einzelnen Wafer.

Als markantes Beispiel einer Mini-Batch-Prozessierungs Anlage wird eine chemisch-mechanische Poliermaschine Marke „Speedfam“ mit der Typbezeichnung „Auriga<sup>20</sup>“ näher dargestellt. Die Maschine entfernt die notwendige Dicke einer Schicht des Wafers mittels mechanischen Polierens mithilfe von chemischen Substanzen. Aus der Sicht des Einflusses der Losgröße auf die Prozesszeit ist die Arbeitsweise der Vorrichtung sehr interessant. Die Anlage holt 5 Wafer aus der Horde. Sie werden gleichzeitig bearbeitet und nach Fertigstellung werden sie in die Horde entladen. Die Maschine startet mit Abarbeitung von den nächsten 5 Wafers. Abbildung 4-16 stellt die Abhängigkeit der Prozesszeit von der Anzahl der Wafer in der Cassette bei Mini-Batch Prozessierung in der beschriebenen Anlage dar. Es entsteht eine monoton wachsende Treppenfunktion, die konstante Werte der Prozessdauer für einzelne Mini-Batches annimmt:

---

<sup>20</sup> Stammt aus der Entrepix Internetpräsenz: <http://www.entrepix.com/ipcc/speedfam/entrepix-auriga.php> - Stand 23.12.11 12:32:00

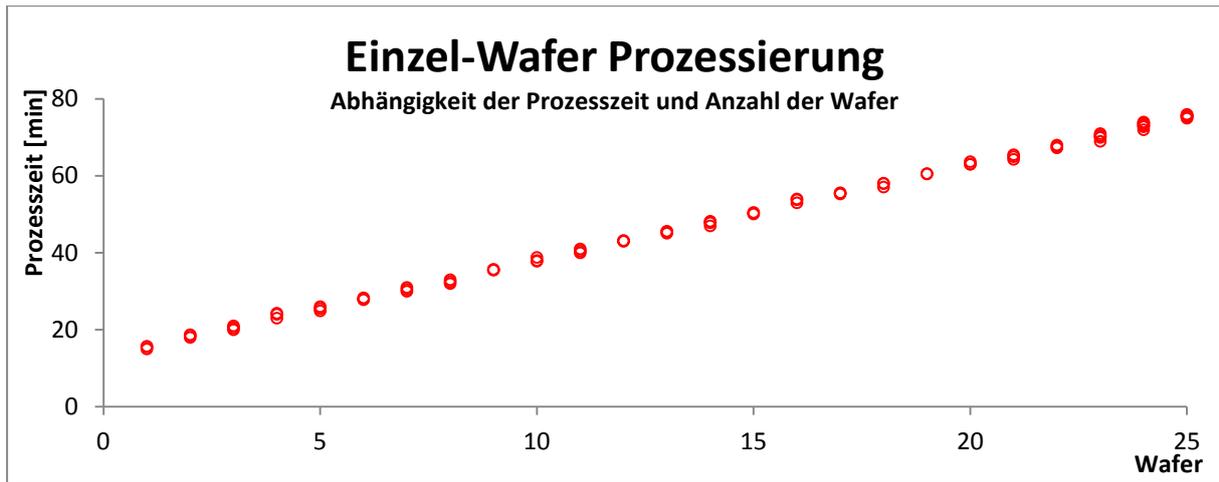


Abbildung 4-15: Die Abhängigkeit der Prozesszeit von der Anzahl der Wafer in der Cassette bei Einzel-Wafer Prozessierung

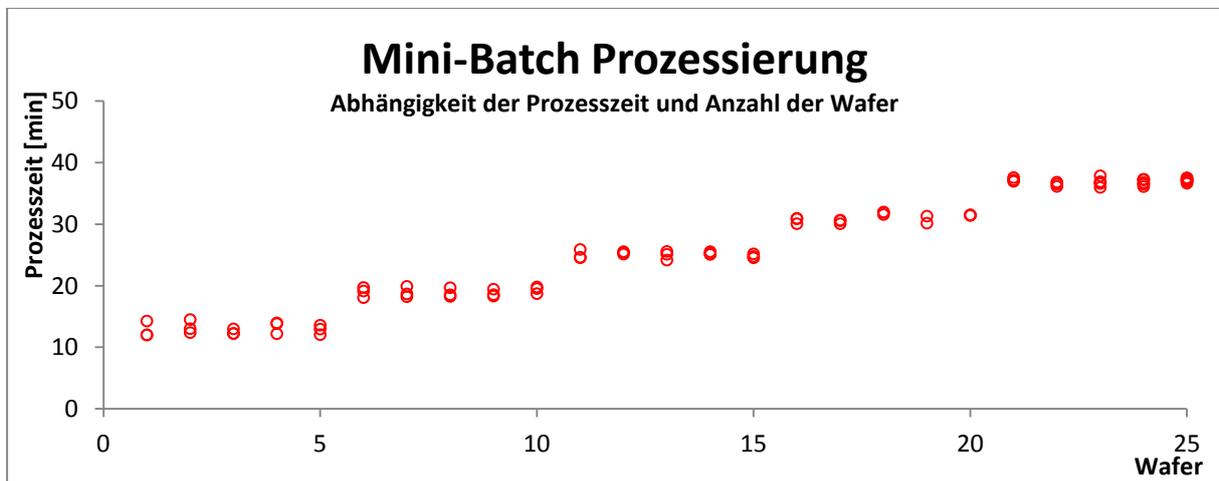


Abbildung 4-16: Die Abhängigkeit der Prozesszeit von der Anzahl der Wafer in der Cassette bei Mini-Batch Prozessierung

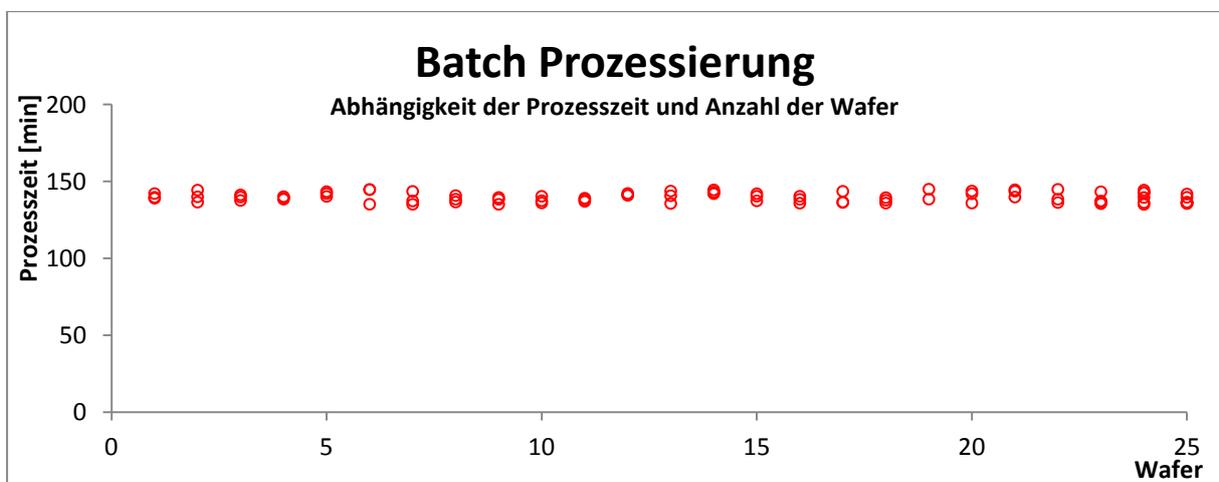


Abbildung 4-17: Die Abhängigkeit der Prozesszeit von der Anzahl der Wafer in der Cassette bei Batch Prozessierung

$$PT_{MBA}(w) = i * t + b \quad \text{für } w \in (w_{mbs*(i-1)}; w_{mbs*i}] \Leftrightarrow i \in N^+ \text{ and } mbs \in N^+$$

wobei  $i$  die Anzahl der Mini-Batches im bearbeiteten Los,  $mbs$  die Anzahl von Wafer in dem Mini-Batch und  $PT_{MBA}(w)$  die entsprechende Prozesszeit bedeuten. Die Parameter  $t$  und  $b$  sind auch in diesem Fall konstant und prozessspezifisch. Die Hilfsgrösse  $b$  ist von der Anzahl der Wafer unabhängig, wie zum Beispiel die Dauer des Wechsels von prozessnotwendigen, chemischen Substanzen (engl. Slurry). Die Konstante  $t$  charakterisiert die erforderliche Bearbeitungszeit der einzelnen Mini-Batches.

Die Batch-Prozessierung zeichnet sich durch gleichzeitige Bearbeitung der Lose in einer Anlage aus. Die Werkstücke werden aufgrund des gleichen oder kompatiblen Prozessschrittrepts ausgesucht und zu einem sogenannten Job zusammengefasst. Es ist nicht zwingend erforderlich, dass die Jobangehörigen zur gleichen Produktfamilie gehören oder eine gleiche Route abarbeiten. Ein Job kann eine unterschiedliche Anzahl an Wafer beinhalten, üblich ist aber die Definition einer minimalen Batchgröße. Die maximale Waferanzahl wird durch die Anlagenkapazität bestimmt. Da die Batch-Prozessierung praktisch nur bei zeitaufwendigen Prozessschritten angewendet wird, sind die Prozesszeitunterschiede bedingt durch die Beladedauer, die von der Losanzahl in dem gegebenen Job abhängig ist, vernachlässigbar. Die Abbildung 4-17 stellt einen typischen Verlauf der Prozessdauer in Abhängigkeit von der Waferanzahl in einem Los dar. Die Funktion für die Bestimmung der Prozesszeit bei einer gegebenen Anzahl an Siliziumscheiben ist trivial:

$$PT_{BA}(w) = c \Leftrightarrow c = const.$$

Ein markantes Beispiel einer Anlage, die mehrere Lose gleichzeitig und gemeinsam prozessiert, ist ein Oxidationsofen. Bei den meisten Ofenanlagen in der Halbleiterfertigung wird eine Batchgröße vom Hersteller vorgegeben. Bei einer geringen Zahl an Wafern, die zum Prozessstartzeitpunkt zur Verfügung stehen, wird die Anlage mit Dummy-Siliziumscheiben aufgefüllt.

Die Untersuchung des Einflusses der Waferanzahl auf die Fertigstellungszeit offenbarte ein bedeutendes Problem für Erstellung der Prognose. Ein wesentlicher Anteil der Lose mit geringerer Anzahl an Siliziumscheiben erreicht nie die Vorhersagezieloperation. Die Revision der Bewegung der unvollständigen Werkstücke über die Fabrik und deren Eigenschaften ergab folgende Aufstellung bezüglich der Anfangs- und Endoperation:

- Kategorie 1: Lose, die mit geringerem Siliziumscheibenvolumen eingeschleust wurden,

- Kategorie 2: Lose, deren Siliziumscheibenvolumen aufgrund von Prozessfehlern und Verwurf von Wafern verringert wurde,
- Kategorie 3: Engineering-Lose, die nur einen Teil der Route abarbeiten und zu Forschungszwecken dienen,
- Kategorie 4: Lose, die zu Forschungszwecken gesplittet wurden und auch gesplittet die Zieloperation erreichen,
- Kategorie 5: Lose, die prozessbedingt gesplittet wurden und die Zieloperation erst nach Vereinigung aller Teile erreichen.

Die Lose der Kategorie 1, 2 und 4 stellen für die Prognose keine Gefahr der Verfälschung der Ergebnisse dar. Selbstverständlich ist auch die Notwendigkeit der vollständigen Filterung der Engineering-Werkstücke aus der Vorhersage. Interessant sind aber die Lose der Kategorie 5, deren Anteil über den Bestand der Fabrik zu einem gegebenen, beliebigen Zeitpunkt ungefähr 2 bis 4 % beträgt. Da die Gefahr der Ergebnisverfälschung in dem Fall beträchtlich ist, entsteht die Notwendigkeit der Erstellung der Prognose erst nach Zusammenfassung aller Teile zu einem gemeinsamen Los. Der Mechanismus wird bei Erläuterung der Vorhersageberechnung in dem Kapitel 4.6.1 näher beschrieben.

Die Anlagen, die nach dem Einzel-Wafer oder Mini-Batch Prinzip arbeiten, machen in der untersuchten Fabrik über 85 % des Maschinenparks aus. Infolgedessen ist also aus der Sicht des Klassifizierungsansatzes in der AFM der Einfluss der Anzahl an Siliziumscheiben in einem Los auf die Durchlaufzeit von Bedeutung. Beim genaueren Betrachten der Verteilung der

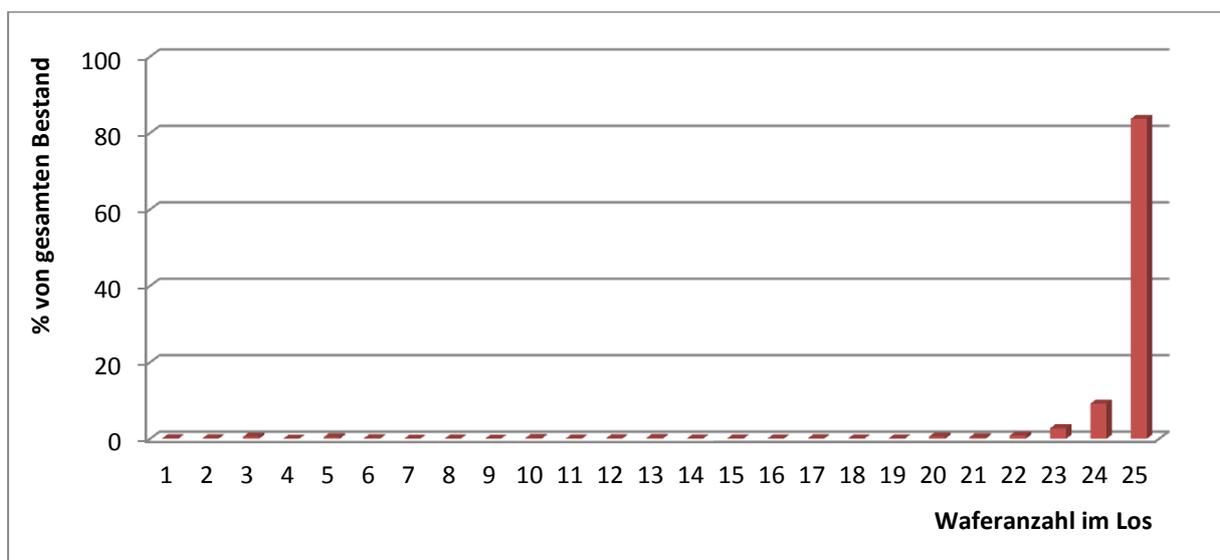


Abbildung 4-18: Häufigkeitsverteilung der Waferanzahl in einem Los über gesamten Bestand der Fabrik abzüglich der Split-Lose

Losgröße in der Fertigung, dargestellt in der Abbildung 4-18, wird nur der marginale Anteil der Lose mit Waferanzahl unter 23 festgestellt. Dementsprechend wurden bei der Modellierung der Fabrik lediglich zwei Losanzahl-Klassen eingeführt:

- $K_1 \Rightarrow [1,22]$  , also mit maximal 22 Wafer in einem Los, ca. 1,7 % WIP
- $K_2 \Rightarrow [23,25]$  , also mit 23, 24 oder 25 Wafer in einem Los, ca. 98,3 % WIP

Bei der Anwendung der Prognosemethode auf eine andere Fabrik ist die Eliminierung der Klassifizierung durch die Losgröße denkbar. Ein augenscheinliches Beispiel aus der Halbleiterfertigung wäre eine Massenfertigung in einem Speicherwerk.

### **4.4.4 Andere durchlaufzeitbeeinflussende Effekte**

In der Factory-Physic Literatur [Hopp08] wie auch in zahlreichen thematisch ähnlichen Publikationen [Ankenm08], [Na05] lassen sich auch weitere Faktoren ermitteln, welche die Durchlaufzeit einer Fertigung beeinflussen. Moacir Goodish Filho beschreibt [Filho10], dass er bereits bei scheinbar einfachen Problemen, wie einem Versuch der mathematischen Beschreibung der Abhängigkeit der Losgröße und Durchlaufzeit, mit dem enormen Einfluss zusätzlicher durchlaufzeitbeeinflussender Parameter konfrontiert wurde. Abbildung 4-19 verdeutlicht die Komplexität von Abhängigkeiten verschiedener Kennzahlen in der Halbleiterfertigung. Die Kennzahlenvorhersage erfordert aber eine genauere Untersuchung von weiteren Fertigungsspezifika zwecks der Verbesserung der Qualität der Ergebnisse. Bei der Erforschung eines neuen Prognosealgorithmus besteht die Gefahr, dass der immer höher werdende Bekanntheitsgrad des Systems zu immer grösseren Vektoren von Eingangs- bzw. Systemdaten bei nur minimalen Verbesserungen der Vorhersagetreffsicherheit führt. Aus diesem Grund spielt die Abwägung der Einflüsse eine bedeutende Rolle. Bei der Entwicklung von AFM wurden die ausgewählten durchlaufzeitbeeinflussenden Faktoren anhand deren Charakteristika im Verhalten untergliedert, in:

- stochastisch
  - Ausfallverhalten der Maschinen,
  - Verfügbarkeit des Personals,
  - Los-Sampling,
  - Eintrittswahrscheinlichkeit und Dauer des Hold-Zustandes,
  - Kapazitätsminderung der Anlage durch nicht produktive Bearbeitung von Entwicklungslosen und damit verbundene Effekte, wie Durchlaufzeitgewinn am Wochenenden,



Die Effekte, wie die durch die gehobene Auslastung erhöhte Wahrscheinlichkeit von einem erneuten Ausfall der reparierten bzw. nachfolgenden Maschinen werden nicht berücksichtigt. Bei der Betrachtung der Verfügbarkeit von Personal wurden nur minimale Abweichungen der Wartezeiten von den Losen in der Warteschlange vor der Prozessierung während der Ruhepausen festgestellt. Da die Verteilung der Wartezeiten über mehrere Tage aber annähernd konstant bleibt und die AFM an der ersten Stelle die prognostizierte Losanlieferung pro Tag ausrechnet, wurde die Personalverfügbar vollständig in die Durchlaufzeitstatistik einbezogen. Die ständige Entwicklung neuer Produkte ist in der Halbleiterbranche ein wichtiger Aspekt, um auf Dauer konkurrenzfähig zu bleiben. Die Forschung konsumiert einen Teil der verfügbaren Fabrikressourcen. In der Praxis schließen die Ingenieure die Maschinen aus der regulären Produktion für eine Zeitperiode aus. Daraus resultiert ein interessanter Effekt bezüglich der Durchlaufzeit, der in den typischen Urlaubszeiträumen oder an den Wochenenden kürzer ist. In der AFM wurde das Verhalten von Ingenieur Tätigkeiten an den Maschinen wie auch der Anlagenausfälle modelliert. Aufgrund des sehr schwer berechenbaren Beschleunigungsfaktors der Fertigungslinie wurde der Einfluss von entwicklungsfreien Phasen nicht in die Prognosealgorithmen eingebaut.

Der Begriff Hold-Zustand wurde bereits in Kapitel 2.4.4 beschrieben. Der Einfluss der Zeit, in dem die Lose angehalten wurden, auf die prognostizierte Durchlaufzeit variiert in Abhängigkeit vom Grund des Haltevorgangs. Die Untersuchung von Losbewegungen in der Fabrik ergaben eine negative Korrelation der Wahrscheinlichkeit des Auftretens eines Hold-Zustands zum Einführungszeitpunkt eines Produktes in die Fertigung. Aus der Relation lässt sich eine Gesetzmäßigkeit ableiten, die besagt, je neuer ein Artikel in der Fertigung desto höher ist die Aussicht auf das Anhalten eines Werkstückes. Die Produkte mit hochkomplexen, störungsempfindlichen Routen stellen jedoch eine typische Ausnahme von der Regel dar. Der verhältnismäßig erhebliche Einfluss des Holds auf die Durchlaufzeit erfordert eine Behandlung des Problems seitens der Prognosealgorithmen. Der in AFM implementierte Ansatz basiert in ähnlicher Weise auf den Standards, die auch in anderen Prognosewerkzeugen verwendet werden. Die historischen Bewegungsabläufe liefern die statistische Wahrscheinlichkeit des Eintreffens wie auch die Dauer des Holds an jeder Operation. Die Informationen werden für die Bestimmung der prognostizierten Durchlaufzeit bis zur Zieloperation angewendet. Die Gewinnung der Statistik wie auch die Berechnung der Prognose werden in den folgenden Kapiteln beschrieben. Der Einfluss des Holds auf die Vorhersage in der Fertigung wurde genauer untersucht und ist in Kapitel 6 detaillierter beschrieben.

Das Sampling eines Loses bedeutet die Verkürzung der Durchlaufzeit bedingt durch die Variation in der Notwendigkeit der Abarbeitung von gewissen Operationen in der Route. Aus der Sicht der Prognose ist das Verhältnis der gewonnenen Zeit zur gesamten Zeit interessant. In der untersuchten Fertigung betrifft das Problem des Überspringens gewisser Schritte ausschließlich die Vermessungsoperationen. Aufgrund der kurzen Prozesszeiten und dem vernachlässigbaren Zeitgewinn spielte das Sampling kaum eine Rolle bei der Durchlaufzeitbestimmung. Die Portierung der AFM-Software auf eine andere Fertigung, in welcher das Los-Sampling ein bedeutender Beschleunigungsfaktor ist, erfordert die Übernahme dieses Kriteriums in die Klassifizierung bzw. das Einfügen von Verzögerungsschritten. Im Rahmen des Projektes wurde die Variante mit der Durchlaufzeiterhöhung, ähnlich wie die Verzögerung durch den Hold-Zustand, ausgewertet. Die Ergebnisse der Validierung des Ansatzes wurden in Kapitel 7 dargestellt.

Bei der Entwicklung der Vorhersagealgorithmen wurde vorerst die Anlagenkapazität vernachlässigt. Bei der Untersuchung der Treffsicherheit der Prognose wurden jedoch gravierende Fehler entdeckt, die nur durch Einführung von Kapazitätsengpässen behoben werden können. Die genaue Funktionsweise des Algorithmus für die Engpassregelung ist in Kapitel 4.6 beschrieben. Ein weiterer der deterministischen Einflüsse auf die Durchlaufzeit in der Fertigung ist die Setupzeit beim Produktwechsel der Anlage. Der Setup bedeutet eine Anpassung der Anlagencharakteristik an einen anderen Prozess. Angenommen die Anlage ist ein Hochtemperaturofen und führt einen Oxidationsprozess durch. Bei einem Produkt beträgt die notwendige Temperatur 900 °C und bei einem anderen nur 750 °C. Bei einem Produktwechsel müssen die physikalischen Eigenschaften verändert werden, was eine gewisse Zeit dauert. Die Anpassungszeit wird in der Halbleiterfertigung Setupzeit genannt und lässt sich ohne größeren Aufwand ermitteln. Das Problem für AFM sind die unbekanntenen Losabarbeitungssequenzen an der Anlage. Somit ist die Verbesserung der Vorhersagequalität durch die Verwendung der Setupzeit konzeptionell unmöglich.

### 4.5 Statistik

Statistik ist eine Wanderkarte. Wenn man sie zu sehen bekommt, ist sie von der Realität schon etwas überholt. Dennoch gibt sie Orientierung. Man muss sie mit Verstand lesen können, sonst geht man in die Irre.

---

(Martin Kruse)

Die verbreitetste Definition der Statistik stammt von F. Vogel [Vogel97] und erläutert sie als „eine wissenschaftliche Disziplin, deren Gegenstand die Entwicklung und Anwendung formaler Methoden zur Gewinnung, Beschreibung und Analyse sowie zur Beurteilung quantitativer Beobachtungen (Daten) ist.“ Aus der Sicht einer Vorhersage ist eine Begriffserklärung der Statistik von Abraham Wald [Wald55] interessanter: „Eine Zusammenfassung von Methoden, die uns erlauben, vernünftige, optimale Entscheidungen im Falle von Ungewissheit zu treffen.“ Die AFM basiert auf einem Modell aus historischen Daten. Ein derartiges Modellierungsprinzip wird in der Literatur statistische Modellierung [Gelasch07] [Rinne02] genannt.

#### 4.5.1 Statistische Modellierung für eine Prognose

Die Prinzipien von statistischer Prognose beinhalten keine formalen Anforderungen, sondern sind eher durch ihre logische Bedeutung zu verstehen [Gelasch07]. Die Problemlösung einer derartigen Voraussage basiert auf dem Einsatz von Theorie und Methodologie der statistischen Wissenschaft, wie zum Beispiel: Clustering des Datenmaterials, verschiedene Mittelwert- und Medianfunktionen sowie Korrelations- und Regressionsanalyse usw. Der Kern der statistischen Prognose ist die Informationsbasis mit folgenden Inhalten (vgl. [Gelasch07]):

- Ziele, Aufgaben, Inhalt und Periodizität der Prognoserechnung,
- Möglichkeiten der Struktur- und Funktionenbeschreibung und Entwicklung des Prognosegegenstandes,
- retrospektive Entwicklung und gegenwärtiger Zustand des Systems und der beeinflussenden Faktoren,
- endogene und exogene Zusammenhänge des Prognosegegenstandes,
- Formalismen für die Darstellungsmöglichkeiten,
- notwendige Informationsquellen und die entsprechenden Informationsfilter.

Die Wissensbasis kann sowohl attributiven, quantitativen als auch qualitativen Charakter besitzen und ist in den meisten komplexeren Prognoseansätzen als Synthese der beiden zu erkennen. Die graphische Darstellung der Ergebnisse erleichtert die Feststellung der wesentlichen Entwicklungstendenzen oder Trends. Der Bewertungsfaktor ist, wie auch bei anderen Vorhersageansätzen, die Qualität der Ergebnisse.

Eines der größten Probleme beim Aufbau einer statistischen Prognose ist die Bereitstellung einer Informationsbasis. Die Problematik der Basisdaten wurde bereits in Kapitel 4.1 beschrieben und wird im nachfolgenden Kapitel vertieft.

### **4.5.2 Loseinschleusung in die Fabrik**

Die Materialeinschleusung ist in einer Produktionsstätte einer von vielen Planungsprozessen. Einen typischen Großkonzern charakterisieren drei Ebenen der Produktionsplanung [Ehrmann02]:

- strategische Planung – langfristige Taktik zur Sicherung der Wettbewerbsfähigkeit,
- taktische Planung – umfasst das mittelfristige Konzept für die Durchführung der Geschäftsprozesse und Vorgaben der Strategischen Planung,
- und operative Planung – ein Instrument für die Planung, Kontrolle und Steuerung der definierten Ziele in einer kurzfristigen Zeitperiode.

Die Definition, welche der drei Ebenen für die Loseinschleusung in der Fabrik verantwortlich ist, ist nicht möglich, da alle nacheinander sich mit der Problematik befassen und den Informationsfluss beeinflussen. Aus Sicht der Prognose sind drei Aspekte bezüglich der Produkteinschleusung von Bedeutung:

- Granularität der Information,
- Zeitpunkt der Verfügbarkeit der Information,
- Möglichkeit der Automatisierung von Informationsgewinnung,

Da die strategischen Planungsansätze durch groben und ungenauen Informationsinhalt gekennzeichnet sind und oft verändert werden, begrenzte sich die Suche nach einer geeigneten Datenquelle auf die taktische bzw. operative Planung. Die Bewertung der Qualifikation als Datenquelle für die kurzfristige Prognose wurde in Tabelle 4-2 dargestellt. Die taktische Planung gibt Vorgaben der notwendigen Waferstarts pro Produkt und Woche, die in dem operativen Bereich umgesetzt werden sollen. Die Daten liegen in einer maschinenauswertbaren Form vor, was die Automatisierung ermöglicht. Die Wissensbasis bezüglich Materialeinschleusung bei der operativen Planung ist wesentlich detaillierter und genauer. Die Angaben

bezüglich Startrate des Materials werden zumeist pro Schicht in speziellen Besprechungen erarbeitet und direkt dem zuständigen Personal mittgeteilt. Trotz der besseren Qualität der Information sind Auskünfte von der operativen Planung ungeeignet. Einerseits ist die maschinelle Auswertung unmöglich. Andererseits ist der Zeitraum begrenzt durch die Dauer einer Schicht, also zu kurz für eine Voraussage mit dem Prognosehorizont über 7 Tage. Die AFM

	Taktische Planung	Operative Planung
Granularität	Waferanzahl pro Produkt pro Woche	Waferanzahl pro Produkt pro Schicht
Verfügbarkeitszeitpunkt	Mehrere Tage vor der Einschleusung	Nach der Besprechung über Einschleusen in einer Schicht
Automatisierungsmöglichkeit	Gegeben	Nicht gegeben

**Tabelle 4-2: Vergleich der taktischen und operativen Planung bezüglich Eignung der Materialeinschleusungspläne für die AFM**

verwendet daher die Materialeinschleusungslisten der taktischen Planung. Die ungenügende Granularität der Angaben wird durch statistische Werte ergänzt und verfeinert. Die statistische Auswertung basiert auf den Bewegungsabläufen, die in der Vergangenheit erfolgten. Unabhängig von der Wahl einer Prognosemethode entsteht durch den falsch angenommenen Startzeitpunkt eines Loses noch vor der Kalkulation ein Fehler in der Voraussage. Um dieses Problem möglichst weit dämpfen zu können, wurde eine detaillierte statistische Analyse von dem Verhalten bei der Einschleusung durchgeführt. Sie umfasste unter anderem:

- Mitarbeiterverhalten
  - Merkmale abhängig von dem zuständigen Personal selbst – bevorzugter Arbeitsstil
  - Typische Merkmale für eine Schicht, wie zum Beispiel: unterschiedliche Pausenzeiten am Tag und in der Nacht oder Zeitpunkte für operative Meetings bzgl. Loseinschleusung
- Produktabhängigkeiten
  - Prozessbedingte Vorgaben für den gleichzeitigen Start von mehreren Losen.

Die Statistik umfasst produktfeine Informationen über:

- Einmalige Einschleusmengen – in eine Menge fallen die Lose, deren Startzeitpunkt nur gering voneinander abweicht (bestimmtes Zeitintervall)
- Typische Zeitpunkte für den Produktionsstart – falls vorhanden
- Die Zeitperioden zwischen den Einschleusungen – falls vorhanden

Für die Vollständigkeit des Verfahrens müssen auch einmalig die Pausenpläne aufbereitet und in der Datenbank gespeichert werden.

Die aufbereitete statistische Analyse mit Informationen über typische Einschleuszeitpunkte und Einschleusmengen ermöglicht die Kalkulation der prognostizierten Verteilung der Zeitabstände bei der Materialeinschleusung (mehr in Kapitel 4.6.1). Bei der Initialisierung der Prognose werden die Daten über Losstartmengen vorerst produktfein gruppiert. Anschließend werden die Losgruppen gleichmäßig über einen Zeitraum verteilt. Zusätzlich erfolgt eine Anpassung der Zeitpunkte des Produktionsstarts mithilfe des statistisch erfassten Schicht- und Mitarbeiterverhalten.

Die Aktualität der Statistik wird durch Anpassungen gewährleistet, die auf einem periodischen Vergleich mit den realen Daten basieren. Das gleiche Prinzip wird auch für die Loseinschleusung aus verschiedenen Lagerstätten angewendet.

### **4.5.3 Durchlaufzeit**

Den Kern jeder Prognose der Losbewegung in der Halbleiterfertigung bilden die Durchlaufzeiten durch die Fabrik. Unter dem Begriff Durchlaufzeit verbergen sich folgende Zeitangaben [Kaech04]:

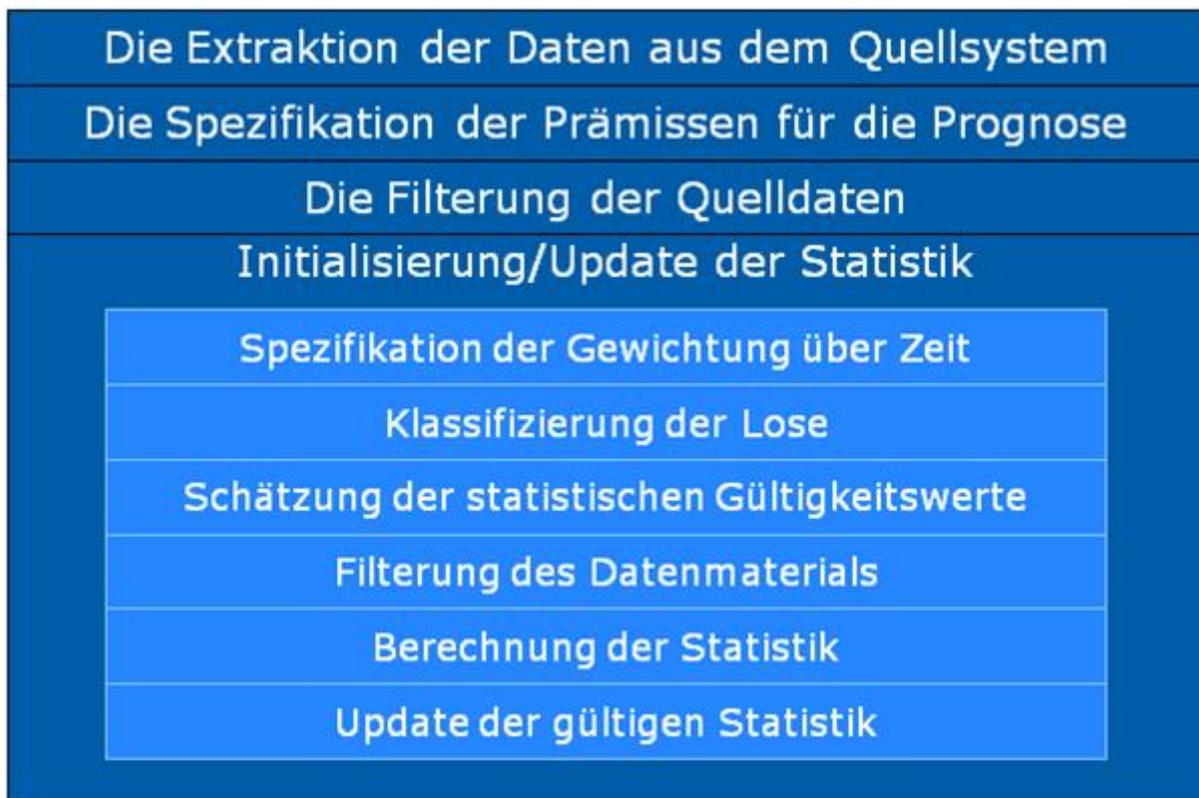
- die Bearbeitungszeit – die Zeit, die aus technischer Sicht für die Herstellung notwendig ist (vgl. Kapitel 2.4.1),
- die Liegezeit – auch Wartezeit, bedeutet die Zeit, in der ein Werkstück auf das Prozessieren wartet (vgl. Kapitel 2.4.2),
- Rüstzeit – ist die Zeit, die gebraucht wird, um die Ressourcen in den für die Bearbeitung eines Werkstücks notwendigen Zustand zu versetzen (vgl. Kapitel 2.4.1). Ein typisches Beispiel ist ein Oxidationsofen, der in mehreren Temperaturlevels abhängig vom Rezept fertigen kann. Die Rüstzeit entspricht der Dauer der Aufwärm- oder Abkühlphase.

Im Vergleich zu anderen bekannten Ansätzen wie Simulation, Warteschlangentheorie verwendet die AFM die aus Bearbeitungs-, Warte- und Rüstzeit zusammengefasste Zeit für die Prognoseerstellung. Dabei werden, wie bereits in Kapitel 4.4 detailliert beschrieben, auch die Zeitangaben für einzelne Operationen summiert.

Abbildung 4-20 stellt schematisch den AFM-Prozess für die Erstellung der Statistik für die Durchlaufzeitangaben dar. Nach dem Ansatz der in der Ausarbeitung beschriebenen Methode wird für die Erstellung der Durchlaufzeitstatistik nur eine Datenquelle extrahiert und ausgewertet. Sie beinhaltet die detaillierten sowohl zeitlichen Angaben über die Losbewegung in

der Vergangenheit als auch die Eigenschaften einzelner Lose selbst. Anschließend wird die Spezifikation der Vorbedingungen für die Prognose eingelesen. Die Prämissen beinhalten Informationen wie:

- die Klassifizierung,
- den Prognosehorizont,
- die Zieloperationen pro Produkt,
- und Ähnliches...



**Abbildung 4-20: Der Mechanismus für die Erstellung der Statistik für die Durchlaufzeitangaben in der AFM**

Die Änderungen in den Voraussageprämissen sind nur vor der Berechnung der Statistik möglich, ansonsten müssen viele Statistiken erneut erstellt werden. Im nächsten Schritt wird die Quelldatenmenge mittels Einsatz folgender Filter verringert:

- der Produktionsbereichsfilter – es werden alle Informationen als ungültig markiert, die andere Produktionsbereiche betreffen,
- der Produktfilter – es werden nur die Produkte weiterverfolgt, welche die Zieloperation überhaupt erreichen können,
- der Operationsfilter – es werden die Operationen pro Produkt bestimmt, von welchen die Möglichkeit besteht, die Zieloperation in einer bestimmten Zeit zu erreichen. Die

Periode wird von dem Prognosehorizont abgeleitet, sollte aber auf jeden Fall länger sein, was die Auswertung der im Grenzbereich liegenden Informationen für die Berechnung gewährleistet. Die Markierung der Gültigkeitsgrenze kann entweder mithilfe von Plan- oder Realwerten vorgenommen werden. Bei der Verwendung von Planzeiten sollte allerdings der Grenzbereich deutlich grösser sein.

Die Anwendung der Filter ist auch sofort bei der Extraktion der Daten aus den Quellsystemen denkbar. Da mit der AFM-Software unterschiedliche Experimente in der Entwicklungsphase ausgeführt wurden, wurden auch alle Losbewegungsinformationen in der ersten Phase in einer eigenen Datenbank zwischengespeichert. Damit wurde die Minimierung der Interaktion mit dem MES der Fabrik gewährleistet (vgl. Kapitel 5.2.1).

Die tatsächliche Kalkulation der statistisch relevanten Informationen für die Durchlaufzeiten greift nur auf die vorgefilterte Datenmenge zu. Wie bei der Erstellung anderer AFM-Statistiken wurde bei der Berechnung der Durchlaufzeitstatistik auch eine Gewichtung der Daten eingeführt. Sie korreliert mit dem Alter der Informationen. Die für AFM typische Auswerteperiode von 3 Monaten wurde in 3 Altersbereiche mit fallenden Gewichten wie folgt geteilt:

- 1. ... 14. Tag (größtes Gewicht  $GT_1$  (derzeit  $GT_1 = 1$  bei der AFM))
- 15. ... 30. Tag mittleres Gewicht  $GT_2$  (derzeit  $GT_2 = 0,7$  bei der AFM))
- 31. ... 90. Tag (minimales Gewicht  $GT_3$  (derzeit  $GT_{31} = 0,25$  bei der AFM))

Die Einstellung der Gewichtung ist ein mühsamer Prozess, der immer wieder eine erneute Berechnung der Statistik erfordert. Die Brauchbarkeit bzw. Güte der Informationen für die Prognose lässt sich nur durch Erstellung und Validierung einer umfangreichen Prognosemenge untersuchen. Die untersuchten Voraussagen sollten möglichst disjunkt bezüglich der Fabrikeigenschaften (wie Auslastung, Produktmix u. a.) bzw. Initialisierungszeit sein. Die Veränderung des Stellenwertes lässt sich ausschließlich vor der Erzeugung bzw. nur geringfügig vor dem Update der Statistik durchführen. Das Datenmaterial wird also vorerst in eine der Gewichtsgruppen eingeteilt. Anschließend erfolgt die Klassifizierung der einzelnen Lose im Sinne des AFM-Konzeptes (beschrieben in dem Kapitel 4.4). Es werden die Gültigkeitsbereiche  $G_b$  für die Durchlaufzeit für jede AFM-Klasse bestimmt, die für das Ausfiltern der Ausreißer angewendet werden. Gesucht werden also die Gültigkeitsgrenzen  $Gct_{min}_{CL1:S \rightarrow Z}$  und  $Gct_{max}_{CL1:S \rightarrow Z}$  für die Durchlaufzeit  $ct_{CL1:S \rightarrow Z}$  der Lose Klasse CL1 zwischen der Startoperation S und der Zieloperation Z. Die CL1 Gruppe wird durch die AFM-Klassifizierungsmerkmale wie: Produkt, Anzahl der Wafer, Priorität usw. beschrieben.

Die Information über die Durchlaufzeit jedes Loses der CL1 Gruppe werden nach der Extraktion aus der untersuchten, historischen Periode zu einem Mittelwert  $\overline{ct\_neu_{CL1:S \rightarrow Z}}$  zusammengefasst. Aus der bereits erzeugten Statistik wird der derzeitig geltende  $ct\_old_{CL1:S \rightarrow Z}$  ausgelesen. Falls die Statistik initialisiert wird, ersetzt die AFM den Wert von  $ct\_old_{CL1:S \rightarrow Z}$  durch einen aus den Plandaten  $pct_{CL1:S \rightarrow Z}$  extrahierten Wert. Der Gültigkeitsbereich wird mit der Formel bestimmt:

$$Gb = \begin{cases} Gct\_max_{CL1:S \rightarrow Z} = (1 + F) * \frac{WG1 * ct\_old_{CL1:S \rightarrow Z} + WG2 * \overline{ct\_neu_{CL1:S \rightarrow Z}}}{WG1 + WG2} \\ Gct\_min_{CL1:S \rightarrow Z} = (1 - F) * \frac{WG1 * ct\_old_{CL1:S \rightarrow Z} + WG2 * \overline{ct\_neu_{CL1:S \rightarrow Z}}}{WG1 + WG2} \end{cases}$$

Die Veränderung der Gewichte WG1 und WG2 skaliert den Einfluss des neuesten Durchlaufzeitverhaltens auf die Grenzen des Gültigkeitsbereiches. Der Faktor  $F$  ist im Bereich zwischen 0 und 1 frei wählbar und beschreibt die Breite des Gültigkeitsbereiches. In der analysierten Halbleiterfabrik liefert AFM die besten Ergebnisse, wenn der Faktor  $F = 0,35$  beträgt. Es werden also nur die Durchlaufzeitangaben erlaubt, die maximal 35 % von dem bestimmten Gültigkeitsbereichszentrum abweichen. Die restlichen Zeitinformationen werden als Ausreißer markiert und bei der Statistikbestimmung nicht betrachtet. Der Vorteil eines derartigen Ausreißerfilters im Vergleich zu der wesentlich einfacheren Median-Mittelwertrechnung besteht in der Reaktionsmöglichkeit auf ungewöhnliches Fabrikverhalten. Nach der Feststellung von Unregelmäßigkeiten in der Produktion, zum Beispiel durch: Linienstopp, Fabrikabschaltung oder Versorgungsproblemen, lässt sich vor dem Update der Durchlaufzeitstatistik der Faktor  $F$  einfach verringern, was die Verfälschung der Statistik minimiert. Das bereinigte Datenmaterial ermöglicht schließlich die Bestimmung der neuen, gültigen Durchlaufzeit  $ct\_neu_{CL1:S \rightarrow Z}$ . Die AFM wendet den mit  $GT_i$  gewichteten Mittelwert der Durchlaufzeitmittelwerten  $\overline{ct_{CL1:S \rightarrow Z}^i}$  einzelner Altersbereiche also die Formel:

$$ct\_neu_{CL1:S \rightarrow Z} = \frac{\sum_{i=1}^n GT_i * \overline{ct_{CL1:S \rightarrow Z}^i}}{\sum_{i=1}^n GT_i}$$

für die Kalkulation an. Dabei beschreibt der Parameter  $n$ , wie bereits beschrieben, die Anzahl der Altersklassen. Im letzten Schritt wird die bisher gültige Statistik mit den neuen Werten überschrieben.

**4.5.4 Sampling an der Zieloperation**

Die in Kapitel 2.4.5 beschriebene Stochastik in der Abarbeitung der Schritte, die mit Sampling behaftet sind, erfordert die Einführung entsprechender Statistiken. Da der Ansatz von AFM nur die Zeit zwischen Start- und Zieloperation aus der Historie verwendet, sind die Samplingraten ausschließlich bei der Zielankunft relevant. Die Prognose der Losanlieferung muss folglich die Information beinhalten, wie viel Material an der Operation die Bearbeitung in Anspruch nimmt. Die Lose, die den Zielschritt überspringen, sind für den Anwender uninteressant.

Die Basis für die Ermittlung der statistischen Samplingrate bildet, nach dem Ansatz von AFM, nur eine Datenquelle, nämlich die historischen Bewegungsdaten der Lose in der Fabrik. Die Samplingrate wird eindeutig für ein gegebenes Produkt und mögliche Zieloperationen gespeichert. Angenommen die prozentuale Rate der Notwendigkeit der Abarbeitung an der Operation A wird gesucht, dann ist es notwendig, die Lose herauszufinden, die zu einer analysierten Produktgruppe gehören und in einer gegebenen Beobachtungsperiode an einer Vor- und Folgeoperation bearbeitet wurden. Tabelle 4-3 veranschaulicht verschiedene Beispiele von Kombinationen der Gegebenheiten der Abarbeitung an Vor-, Ziel- und Folgeoperationen, die bei der Auswertung des Datenmaterials entstehen können. Grün wurden die Lose NR\_1 und NR\_2 gekennzeichnet, welche die Kriterien für die Samplingberechnung erfüllen,

Los Nummer	Abarbeitung an:		
	Voroperation	Zieloperation	Folgeoperation
LOS NR_1	Gegeben	Gegeben	Gegeben
LOS NR_2	Gegeben	Nicht Gegeben	Gegeben
LOS NR_3	Nicht Gegeben	Gegeben	Gegeben
LOS NR_4	Gegeben	Nicht Gegeben	Nicht Gegeben
LOS NR_5	Gegeben	Gegeben	Nicht Gegeben
LOS NR_6	Gegeben	Nicht Gegeben	Nicht Gegeben

**Tabelle 4-3: Fallunterscheidung bei der Auswertung von Sampling**

also mindestens einmal an der Operation mit dem kleineren und größeren Index bearbeitet wurden. Die rot markierten Lose werden aus der Berechnung ausgesondert, weil die Vollständigkeit der Daten in der Untersuchungsperiode nicht gewährleistet ist. Der vorletzte Fall stellt ein Los NR\_5 dar, das an der Voroperation und an der Zieloperation bearbeitet wurde, jedoch nicht an der Folgeoperation. Theoretisch könnten die Angaben für die Kalkulation von Sampling verwendet werden. Die Berechnung wurde dadurch aber fehlerhaft, da es unmöglich ist,

den dazu inversen Fall, wie bei Los NR\_6, zu ermitteln. Nach der Anwendung aller Filter wird die Samplingrate  $\sigma_{P/OP}$  für ein Produkt  $P$  an der Operation  $OP$  berechnet. Die Samplingrate ist ein Quotient der Summe der Lose, die an der Operation bearbeitet wurden, zu der gesamten Anzahl der relevanten Werkstücke der Route, also der Summe von bearbeiteten Lose und denen, die die Operation übersprungen haben. Da die ausreichende Anzahl an Stichproben gewährleistet werden muss, ist es notwendig, dementsprechend die Analyseperiode zu definieren. Die Auswahl eines zu großen Zeitraums der Untersuchung erhöht zwar die statistische Relevanz der Analyse, verschlechtert jedoch das Reaktionsvermögen des System auf die Änderungen von Samplingraten in der Fertigung.

Abbildung 4-21 stellt vereinfacht die Funktionsweise des Algorithmus dar, der im Rahmen AFM für die Berechnung der Samplingraten entwickelt wurde. Für die primäre Kalkulation (Initialisierung bei der Abbildung 4-21) des prozentualen Anteils der gesampelten Lose wird zuerst eine Untersuchungsperiode  $t_{period}$  festgelegt. Die Periode beschreibt den minimalen Zeitabschnitt der Vergangenheit, in dem die Beobachtung stattfindet. Obligatorisch ist auch die Bestimmung der minimalen Anzahl der Werkstücke eines Produktes an einem Produktionsschritt, welche die statistische Relevanz der Berechnung gewährleistet. Die gesamte Losanzahl ergibt sich aus Summe der Anzahl an der untersuchten Operation prozessiert Lose  $\sum_{i=1}^n Li$  und derjenigen, die den Prozessschritt übersprungen haben  $\sum_{i=1}^m Li$ . Bei der Festlegung der Untersuchungsperiode  $t_{period}$  wird davon ausgegangen, dass bei der Massenware die minimale Losquantität immer erreicht wird. Im Falle einer nicht ausreichender Anzahl der Vertreter wird die Periode schrittweise um einen Zeitabschnitt verlängert, und die Summe von  $\sum_{i=1}^n Li$  und  $\sum_{i=1}^m Li$  wird erneut berechnet. Die Wahl der Initialisierungsparameter hängt vor allem vom Produktportfolio einer Fabrik und der Häufigkeit des Samplings ab. Die Bestimmung der Parameter in AFM erfolgte unter der Verwendung von den realen Samplingraten als Vergleichsbasis zu den über die Experimente ermittelten Raten. An der Stelle ist es wichtig anzumerken, dass die enge Zusammenarbeit mit den Produktionsbereichen bei der Entwicklung der Prognose vorteilhaft oder sogar erforderlich ist. Nach der Festlegung der Konstanten und Extraktion der Daten aus der Losbewegungsdatei lässt sich die Samplingrate  $\sigma_{init}$  als Quotient der an der untersuchten Operation prozessierten Lose  $\sum_{i=1}^n Li$  zu der Summe von  $\sum_{i=1}^n Li$  und  $\sum_{i=1}^m Li$  berechnen. Die Samplingrate wird anschließend in der AFM-Datenbank gespeichert.

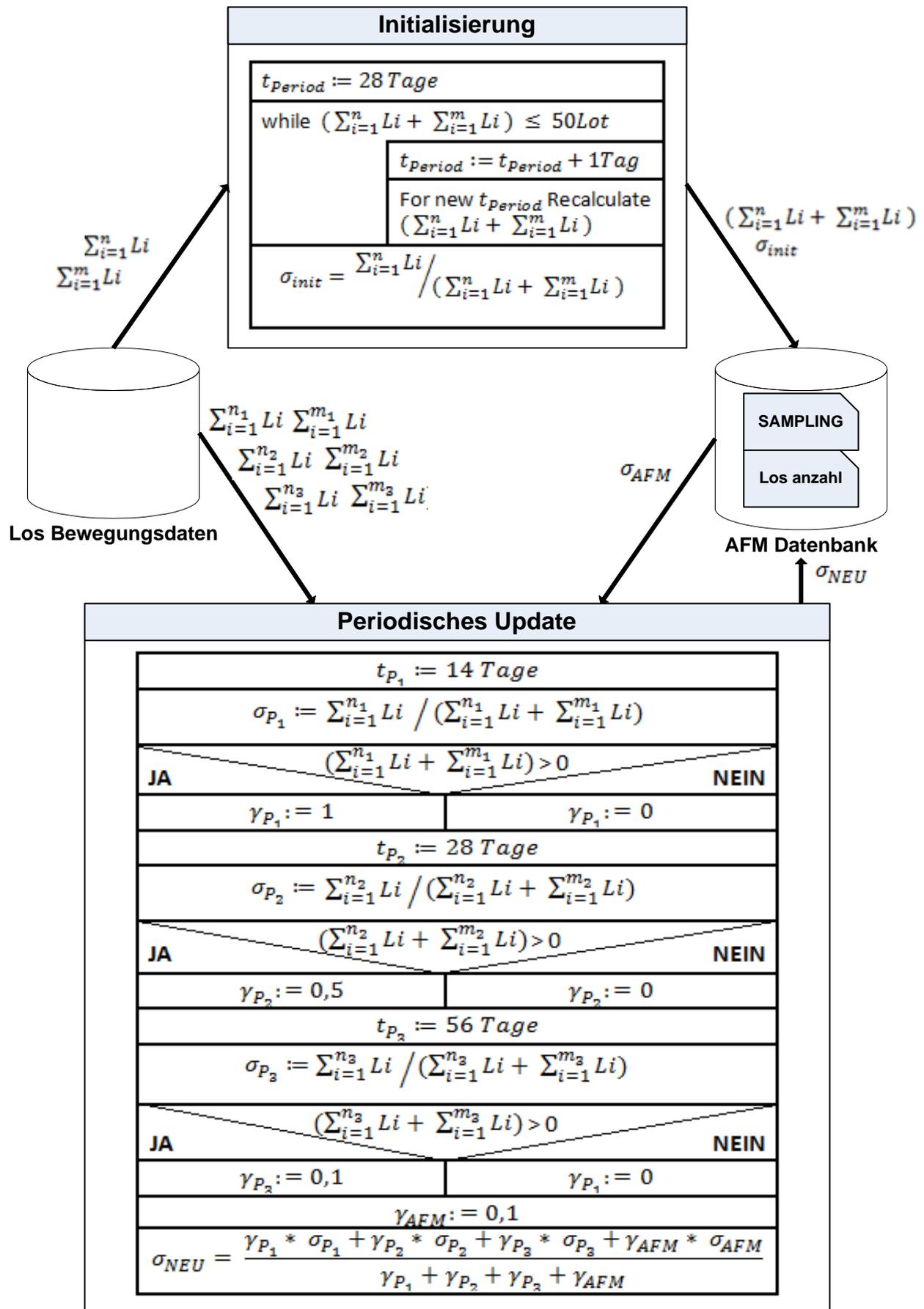


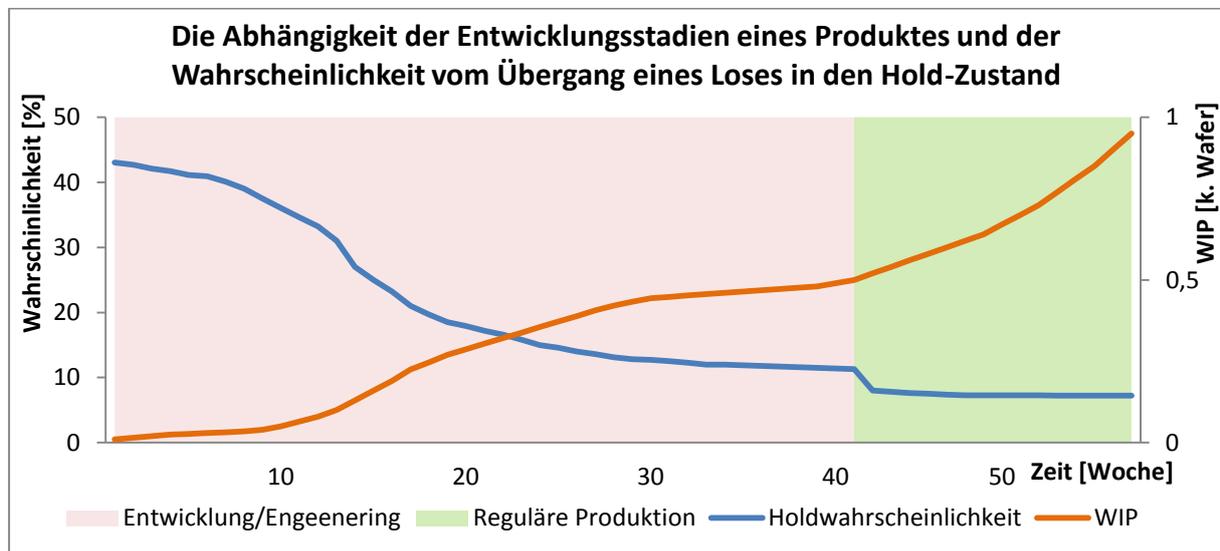
Abbildung 4-21: Initialisierung und Update der Samplingstatistik

In der realen Produktionsumgebung verändern sich die Samplingraten kontinuierlich. Einerseits sind es die neuen Produkte, die das Entwicklungsstadium abgeschlossen haben und in ein reguläres Produktportfolio aufgenommen werden. Die damit verbundene Intensität der Notwendigkeit von Kontrollen nimmt gleichzeitig ab. Andererseits werden die Inspektionen nach verschiedenen Problemen in der Produktionslinie oder bei den Veränderungen im Maschinenpark verschärft. Die Quintessenz der Überlegung ist die fortlaufende Beobachtung und ggf. das Update der Samplingraten. Die AFM-Algorithmen für die Auswertung der Raten ermöglichen die periodische Revision der Statistik (periodisches Update bei der Abbildung 4-21). Die historischen Bewegungsdaten der Lose werden in drei Zeitabschnitte  $t_{P_1}$ ,  $t_{P_2}$ ,  $t_{P_3}$  aufgeteilt. Für jeden dieser Abschnitte werden die zugehörigen Samplingraten  $\sigma_{P_1}$ ,  $\sigma_{P_2}$ ,  $\sigma_{P_3}$  berechnet. Wie in Abbildung 4-21 dargestellt, wird ähnlich wie bei der Initialisierung ein Quotient von den an der untersuchten Operation prozessierten Lose zu der gesamten Anzahl der Werkstücke, die über den Fokusproduktionsschritt bewegt wurden, errechnet. Die bei dem Update gültige Rate  $\sigma_{AFM}$  wird für die Berechnung des neuen Samplingwerts  $\sigma_{NEU}$  aus der AFM-Datenbank ausgelesen. Substanziell ist auch die entsprechende Gewichtung der Samplingraten  $\gamma_{P_1}$ ,  $\gamma_{P_2}$ ,  $\gamma_{P_3}$  wie auch  $\gamma_{AFM}$  bezüglich des Einflusses auf den  $\sigma_{NEU}$  Wert. Selbstverständlich ist dabei: je älter die Daten, desto geringer ist deren Rolle auf den momentanen Stand der Fabrik. Mit der Verlagerung der Gewichtung wird die Reaktion der AFM auf die Veränderungen im Sampling verlangsamt oder beschleunigt. Die Wahl der Zeitperioden sowie ein zugehöriger Bedeutungsgrad der Samplingraten ist von mehreren, spezifischen Faktoren abhängig. Die Justierung der Parameter für die AFM erfolgte über die Validierung der Samplingraten gegen Informationen der verantwortlichen Mitarbeiter. Wichtig dabei war ein längerer Analysezeitraum, damit die unterschiedlichen Veränderungen in der Samplingberechnung abgedeckt werden. Nach der Erforschung von notwendigen Faktoren lässt sich der Wert  $\sigma_{NEU}$  als ein gewichteter Mittelwert bestimmen (vgl. Abbildung 4-21). Bei der Kalkulation von  $\sigma_{NEU}$  ist auch der Fall denkbar, dass in einer Periode keine Vertreter existierten. Eine derartige Periode wird mit 0 gewichtet, um das Ergebnis nicht zu verfälschen. Bei gravierenden Problemen oder Veränderungen in der Fabrik lassen sich die Gewichte so verändern, dass die Ereignisse eine wesentliche Rolle bei der Kalkulation erhalten. Erlaubt ist auch ein jederzeitiger, manueller Start der Berechnung.

### 4.5.5 Das Verhalten des Hold-Zustands

Die Notwendigkeit der Einführung der Statistik über das Verhalten und die Dauer des Anhaltzustands resultiert aus den Anforderungen für das Prognosewerkzeug (vgl. Kapitel 2.4.4 und

2.5). Der Einfluss der Hold-Dauer auf die Gesamtdurchlaufzeit variiert in Abhängigkeit vom Entwicklungsstadium des gegebenen Fabrikates. Abbildung 4-22 stellt für fünf ausgewählte Produkte und Vermessungsoperationen die Abhängigkeit von deren Entwicklungsstadien und der Wahrscheinlichkeit vom Übergang eines Loses in den Hold-Zustand dar. In der Entwicklungsphase ist erwartungsgemäß die Chance deutlich höher, dass ein Los für die Sonderbearbeitung angehalten wird. Beim Übergang zur regulären Produktion verringert sich die Notwendigkeit einer gesonderten Behandlung der Werkstücke und der Erwartungswert für den Übergang in Anhaltezustand sinkt rapide. Den Beobachtungen zufolge korreliert die Wahrscheinlichkeit des Auftretens von Hold negativ mit der Anzahl der Wafer in der Fabrik. Im Fokus der AFM bezüglich Hold stehen also die entweder neu oder zu Testzwecken eingestarteten Fabrikate. Als Hauptproblem bei der Statistikbildung erweist sich die zu geringe Anzahl



**Abbildung 4-22: Die Abhängigkeit des Entwicklungsstadiums eines Produktes und die Wahrscheinlichkeit vom Übergang eines Loses in den Hold-Zustand**

von Stichproben bei derartigen Losen. Das Problem erfordert bei der Auswahl der Grunddaten die Untersuchung einer entsprechend langen Zeitperiode aus der Vergangenheit. Da aber eine Halbleiterfabrik zu den komplexesten Fertigungssystemen zählt, ist die Gefahr der Verfälschung der Vergangenheitsinformationen über ein derartig langes Zeitintervall erheblich. Als Fehlerquelle werden alle untypischen Störungen in der Produktionslinie erkannt, die zu vermehrtem Auftreten des Hold führen, wie zum Beispiel:

- Ein Produktionsstopp für ein oder ggf. mehrere Produkte aufgrund von Auffälligkeiten in der Bearbeitung. Die Lose werden angehalten, bis die Fehlerquelle gefunden und der Fehler behoben wurde. Ein typisches Beispiel ist das Auffinden von den Partikeln auf einem Wafer.

- Ein bedeutender Ausfall eines Medium/von Medien. Die Dauer des Hold-Zustands endet meistens nicht, sobald die Versorgung wieder gewährleistet ist, sondern erst, wenn alle notwendigen Untersuchungen abgeschlossen sind. Ein Exempel eines derartigen Produktionsausfalls ist eine Unterbrechung in Stromversorgung. Obwohl die Notversorgung augenblicklich startet, werden manche Anlagen oder die Pumpensysteme neu gestartet, was zu deutlichen Produktionseinschränkungen führen kann [Spinn11].
- Eine planmäßige Abschaltung der Fertigung, die in mehr oder weniger konstanten Perioden stattfindet und Softwareupdates, Neustart von Servern, Maschinen und Versorgungsnetzen ermöglicht. Die geplante Maßnahme führt zum Anhalten fast aller Lose in der Fabrik.

Außer dem vermehrten Eintreffen des Hold-Zustands wurde auch die untypisch lange Dauer des Zustandes als Fehlerquelle erkannt. Die AFM minimiert den Einfluss der beschriebenen Fehler bei der Statistikerzeugung durch mehrfache Filterung des Datenmaterials in Form von:

- quantitativem Tiefpassfilter der Stichproben für Hold-Dauer,
- quantitativem Hochpassfilter der Stichproben für Hold-Dauer,
- qualitativem Filtern der Stichproben über die Abweichung des gültigen Hold-Dauerwertes,
- qualitativem Filtern der Stichproben in Bezug auf die gültige Hold-Rate.

Unter dem Begriff „Tiefpassfilter für die Hold-Dauer“ verbirgt sich ein einfacher Mechanismus, der die Stichproben bezüglich der Dauer untersucht. Es wird eine Mindest-Hold-Dauer definiert und alle Hold-Einträge, deren Dauer unterhalb des Schwellwerts liegen, werden als ungültig markiert. Die Einführung eines derartigen Filteralgorithmus begrenzt den Berechnungsaufwand bei der Statistikerstellung praktisch ohne Verluste an der Vorhersagequalität (vgl. Kapitel 6.4.3.1). Interessanter ist jedoch der Hochpassfilter, der alle Stichproben herausfiltert, die eine definierte maximale Hold-Dauer überschritten haben. Die Bestimmung dieses Parameters basiert außer einer genauen Auswertung des Datenmaterials auch auf dem Erfahrungswissen der Experten aus der Produktion. Bei einer unerwarteten Störung in der Fertigung werden nahezu gleichzeitig mehrere Lose in einen Anhaltezustand versetzt. Nach der Behebung der Störung werden die Lose sequenziell von den Fachkräften für die Abarbeitung freigegeben, was sich sehr zeitintensiv gestaltet. Da im Normalfall (bei ca. 98,7 % der Stichproben) die Freigabe der Lose aus dem Hold innerhalb einer Schicht erfolgen soll, verwendet die AFM das Zweifache der Schichtdauer als oberen Schwellwert der Hold-Dauer. Die AFM-Software wurde als sehr flexibel in den Parametereinstellungen implementiert, was schnelle

Reaktionen auf die Störfälle in der Produktion ermöglicht. Qualitatives Filtern der Stichproben erfolgt durch eine Überprüfung, ob deren Hold-Dauer  $\varphi'$  zu dem vorbestimmten Gültigkeitsbereich  $G < g_u; g_o >$  gehört. Für jedes mögliche Tripel von Produkt, Route und Operation wird ein derartiges Gültigkeitsintervall für die Dauer des Hold-Zustandes bestimmt. Die Grenzen  $< g_u; g_o >$  von diesem Bereich  $G$  werden mithilfe von dem zur dieser Zeit gültigen Wert  $\varphi$  aus der Statistik festgelegt. Angenommen wird eine maximale, prozentuale Abweichung  $\tau$  [%]. Die Berechnung des Umfangs von  $G$  erfolgt mit den Formeln:

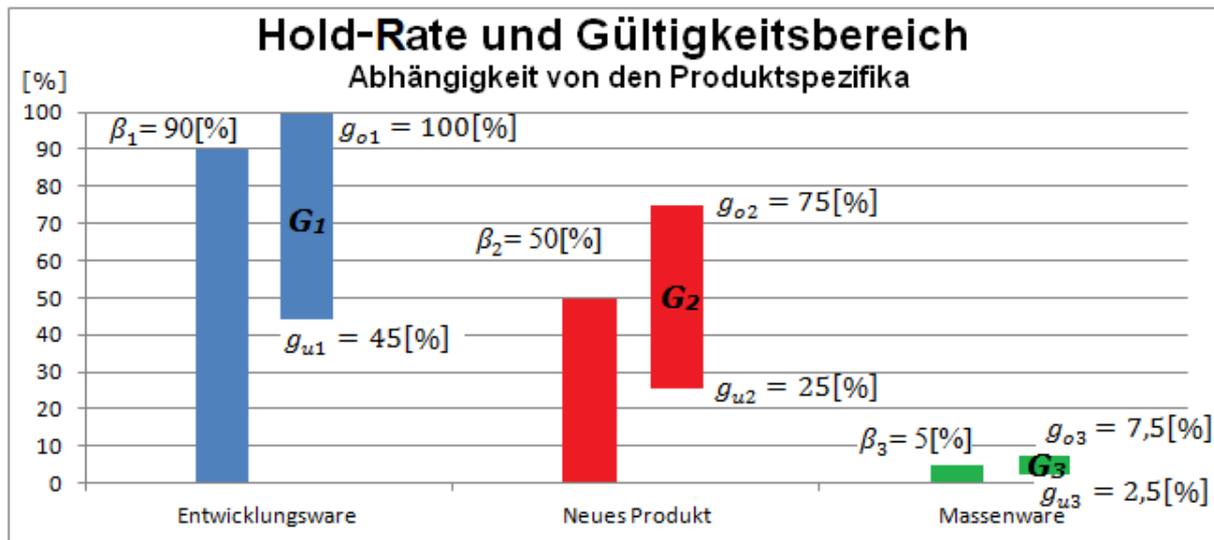
$$g_u = \begin{cases} \varphi - \tau * \varphi & \Leftrightarrow \varphi \geq \tau * \varphi \\ 0 & \Leftrightarrow \varphi < \tau * \varphi \end{cases} \text{ und } g_o = \varphi + \tau * \varphi$$

Im Halbleiterwerk, das als Basis für die Experimente und Entwicklung von AFM diente, wurden enorme Schwankungen in der Hold-Dauer festgestellt. Aus diesem Grund wurde auch die mögliche Abweichung  $\tau$  [%] großzügig gewählt. Dadurch wurde auch die Wirksamkeit dieser Filter begrenzt. Eine tiefere Analyse des Problems verdeutlichte, dass die Abweichung bei ausgewählten Produktionsschritten oder in ganzen Produktionsbereichen deutlich kleiner sein könnte als bei anderen. Die variable Einstellung der maximalen, prozentualen Abweichung in Abhängigkeit von der Operation führt jedoch nur zu geringfügigen Verbesserung der Prognoseresultate. Deutlich mächtiger in Bezug auf die Filterungsquote ist der zweite qualitative Filter, der die Hold-Raten bewertet. Ähnlich wie bei der Untersuchung der Dauer des Hold-Zustandes, wird auch eine maximale, prozentuale Abweichung  $\tau$  [%] angenommen und der Gültigkeitsbereich  $G < g_u; g_o >$  berechnet. Da es sich bei der zu untersuchenden Hold-Rate  $\beta'$  sowie bei der gültigen Hold-Rate  $\beta$  um eine Beschreibungsgröße eines Kontingents der Lose handelt, die in einen Hold-Zustand versetzt wurden, werden die Grenzen wie folgt berechnet:

$$g_u = \begin{cases} \beta - \tau * \beta & \Leftrightarrow \beta \geq \tau * \beta \\ 0 & \Leftrightarrow \beta < \tau * \beta \end{cases} \text{ und } g_o = \begin{cases} \beta + \tau * \beta & \Leftrightarrow \beta + \tau * \beta \leq 100 \text{ [%]} \\ 100 \text{ [%]} & \Leftrightarrow \beta + \tau * \beta > 100 \text{ [%]} \end{cases}$$

Die Funktionsweise des Filters und deren Vorteile für die Prognose lassen sich am einfachsten anhand eines Beispiels erläutern. Abbildung 4-23 stellt die Hold-Rate und den zugehörigen Gültigkeitsbereich für einen exemplarischen, hypothetischen Vertreter für die Entwicklungsware, für ein neues Produkt und für die Massenware dar. Die Hauptaufgabe des beschriebenen Filters ist das Erkennen der Abnormalität im Verhalten der Lose bezüglich des Eintretens eines Hold-Zustandes, bedingt durch untypische Störungen in der Fertigung oder ähnlichem. Im Rahmen der Reaktion auf die Irregularität sollen die neuen Hold-Raten als ungültig erklärt werden und die bis dato gültigen werden beibehalten. Wie die Abbildung 4-23 verdeutlicht, wirkt der Filter vor allem bei den bereits etablierten Produkten mit niedri-

ger Wahrscheinlichkeit des Eintreffens des Anhaltenzustands. Der Gültigkeitsbereich wird für derartige Fabrikate sehr eingengt. Die Einzelprodukte bzw. neuen Produkte sind von derartigen Problemen praktisch betroffen. Da sie sowieso mit einem höheren Erwartungswert für Hold gekennzeichnet sind, hat jegliche Abnormalität in der Fertigung kaum Einfluss auf dessen Hold-Verhalten. Andererseits erlaubt der Filter diesen Produkten, die Hold-Rate in der Statistik ebenso deutlich zu reduzieren. Dies stellt sich als notwendig bei der folgenden Übernahme der Ware in den regulären Produktmix heraus.



**Abbildung 4-23: Die Funktionsweise eines Filter der Hold-Raten am Beispiel von Entwicklungsware, neuem Produkt und Massenware**

Die Beschaffung der Stichproben erfolgt, wie in dem Anforderungskatalog beschrieben, von den Bewegungsdaten der Lose durch die Fertigung über einen vergangenen Zeitraum. Bei der Bestimmung der Zeitspanne wurde ähnlich wie bei der Wahl des Zeitintervalls der Samplingraten (vgl. Kapitel 4.5.4) vorgegangen. Es wurde ein vergangener Zeitraum von 100 Tagen in drei Perioden wie folgt geteilt:

- Tag 1 bis Tag 14 aus der Vergangenheit,
- Tag 15 bis Tag 43 aus der Vergangenheit,
- Tag 44 bis Tag 100 aus der Vergangenheit.

Die aus den Lot-Tracedaten gewonnenen Informationen wurden vorerst mit den quantitativen Filtern behandelt. Anschließend wurden für jeden dieser Zeitabschnitte für alle Möglichkeiten der Tripel Route, Produkt, Operation gebildet. Nachfolgend wurde die Wahrscheinlichkeit des Eintreffens des Hold-Zustands, sowie dessen voraussichtliche Dauer berechnet. Die Anwendung der qualitativen Filter bereinigt die Angaben. Im letzten Schritt werden die neuen Statistikangaben als ein gewichteter Medianwert ermittelt (vgl. Samplingraten – Kapitel 4.5.4),

wobei die neuesten Angaben deutlich mehr Einfluss auf das Ergebnis aufweisen. Die Bestimmung der Perioden für die Auswertung der Vergangenheit erfolgt durch das Abwägen der Dauer und der Anzahl der Stichproben. Der Übergang vom Entwicklungs- in das Produktionsstadium führt zu der größten Verfälschung der Ergebnisse und erfordert manchmal die Interaktion des Administrators. Aus diesem Grund werden bei einer deutlichen Diskrepanz im Erwartungswert der Hold-Dauer spezielle Mitteilungen generiert und dem Verantwortlichen zur Verfügung gestellt.

### 4.5.6 Wechsel der Produktionsbereiche

Unter dem Wechsel von Produktionsbereich verbirgt sich in der Ausarbeitung die Wahl einer Teststrecke im Testbereich für ein bereits fertiges Produkt (kommend aus dem Produktionsbereich). Abbildung 4-24 stellt die unterschiedlichen Varianten für den Übergang dar. Im Fall I ist für ein Produkt nur eine Teststrecke verfügbar, somit ist die Zuweisung der Route in dem Testbereich eindeutig. Aus der Erhebung der Statistik sind die Fälle II und III interessant.

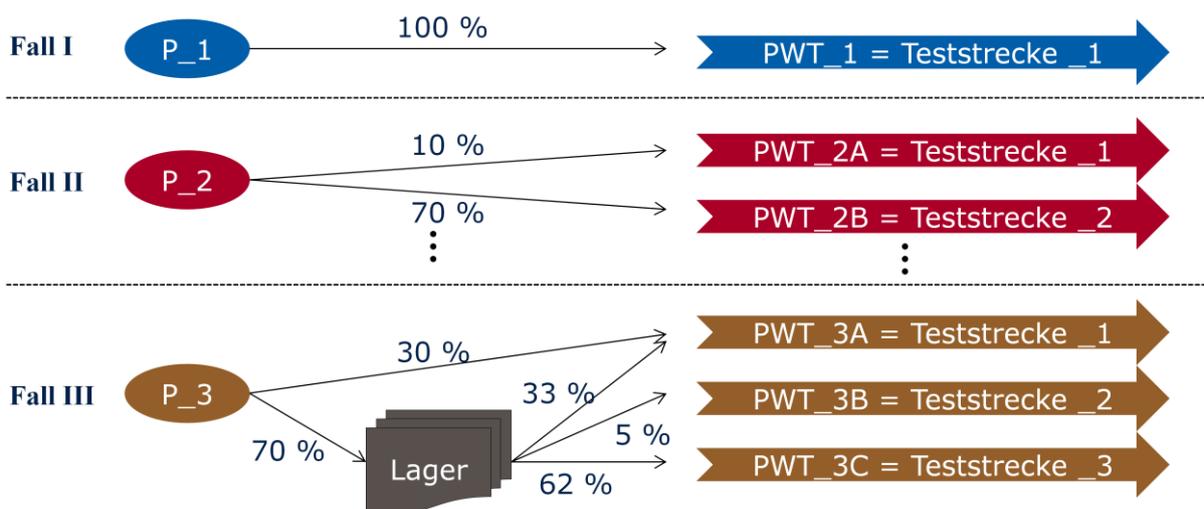
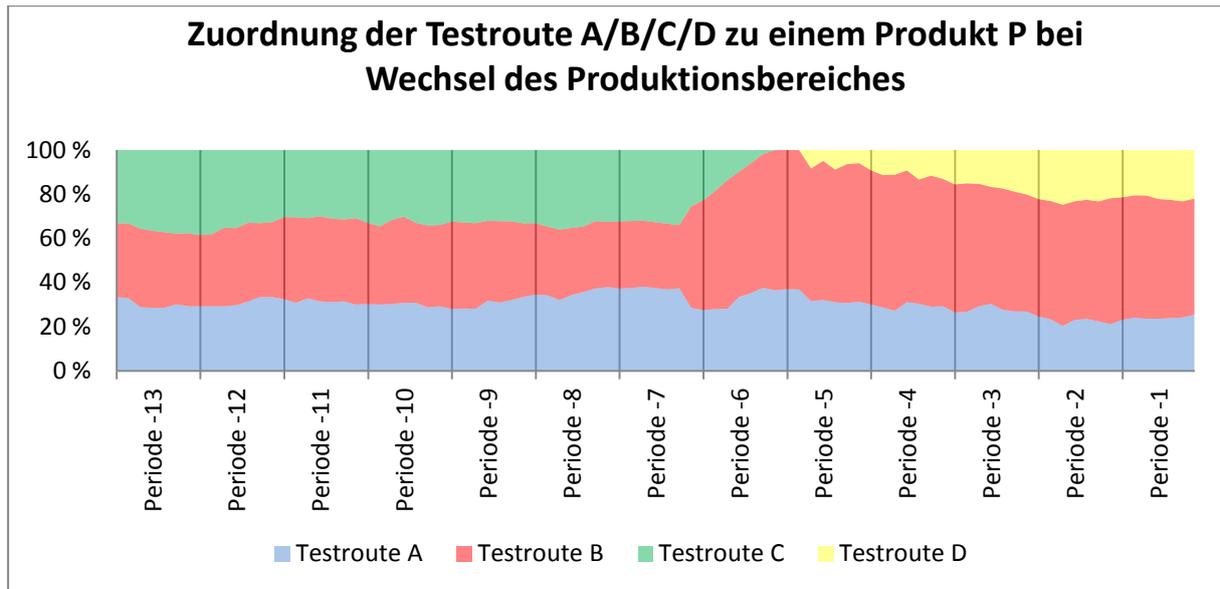


Abbildung 4-24: Die unterschiedlichen Varianten des Übergangs zwischen Fertigungsbereichen

Während in Fall 2 eine einfache Aufteilung des Losstromes auf mehrere Teststrecken vorgenommen wird, ist der Sachverhalt in Fall III deutlich komplizierter. Ein Teil der Lose wird vorerst in einem Lager zwischengelagert und erst dann erfolgt die Auswahl der Testroute. Aufgrund der sehr großen Streuung der Lagerungszeit wurde mit den Anwendern der AFM-Prognose entschieden, dass das Lager als ein Durchgangslager (Lagerungszeit = 0) modelliert wird, bei gleichzeitiger Kennzeichnung der Lose im Reporting. Dadurch liegt der Fokus bei der Voraussage des Übergangs zwischen Fertigungsbereichen ausschließlich auf dem Erwartungswert bei der Wahl einer Teststrecke, also wie groß der prozentuale Anteil der gegebenen



**Abbildung 4-25:** Die prozentuale Zuordnung für ein Ausgangsprodukt zu einer Testroute bei dem Wechsel von dem Produktionsbereich erstellt über längeres Zeitintervall mithilfe von realen Losbewegungsdaten (Beispiel)

Periode	P -13	P -12	P -11	P -10	P -9	P -8	P -7	P -6	P -5	P -4	P -3	P -2	P -1
<b>Testroute A</b>	30,15	30,77	31,44	30,02	30,36	35,14	35,83	32,13	32,89	29,36	27,73	22,62	23,98
<b>Testroute B</b>	33,81	34,06	37,83	37,04	37,02	30,77	32,72	57,05	62,12	59,37	55,20	54,41	54,21
<b>Testroute C</b>	36,05	35,18	30,73	32,94	32,62	34,08	31,45	10,82	0	0	0	0	0
<b>Testroute D</b>	0	0	0	0	0	0	0	0	4,99	11,27	17,07	22,97	21,81
<b>Summe</b>	100	100	100	100	100	100	100	100	100	100	100	100	100

**Tabelle 4-4:** Die Quantifizierung der prozentualen Zuordnung für ein Ausgangsprodukt zu einer Testroute bei dem Wechsel von dem Produktionsbereich erstellt über alle Zeitintervalle mithilfe von realen Losbewegungsdaten (Beispiel)

	Mittelwert	Mittelwert neu-este Periode (Trace)	Gewichteter Mittelwert	AFM Erwartungswert Vorstufe	AFM Erwartungswert Normiert
<b>Testroute A</b>	30,19	23,98	29,17	29,17	30,55
<b>Testroute B</b>	45,05	54,21	49,85	49,85	52,22
<b>Testroute C</b>	18,76	0	11,04	0	0
<b>Testroute D</b>	6,01	21,81	9,94	16,45	17,23
<b>Summe</b>	100	100	100	95,47	100
<b>Treffsicherheit bei mehreren Prognosen</b>	64,32[%]	73,14[%]	71,27[%]		84,61[%]

**Tabelle 4-5:** Der Erwartungswert der prozentualen Rate für die Testroutenzuordnung erstellt mit unterschiedlichen Techniken sowie ein Vergleich zu AFM bezüglich Treffsicherheit von den Prognosen

Lose, die genau diese Teststrecke absolvieren müssen, für ein Produkt ist. Die Grunddatenströme für die statistische Analyse bilden die historischen Bewegungsabläufe zwischen den Bereichen. In der Fertigung, die als Basis für die Entwicklung der Prognose diente, erfolgt eine Anpassung bzw. Überarbeitung der Auswahl der Teststrecken oft. Dies erfordert im Vergleich mit anderen AFM-Statistiken eine deutliche Erweiterung der Gewichtung der Daten nach dem Alter. Die Funktionsweise der AFM-Algorithmen für die Erzeugung von den für die Wahl der Testroute zuständigen Erwartungswerten lässt sich am einfachsten mit einem Beispiel erläutern. Abbildung 4-25 stellt graphisch die Eingangsdaten für ein ausgewähltes Produkt dar. Aus den historischen Bewegungsdaten wurde für einen Zeitraum von 13 Wochen die auf 100 % normierte, prozentuale Verteilung der 4 möglichen Teststrecken für jeden Tag extrahiert. Dies erfolgt über die Berechnung des Verhältnisses der Anzahl der Stichproben für eine Teststrecke zu der Gesamtanzahl der Elemente, die diese Testroute abarbeiten können. Die AFM-Algorithmen teilen den Analysezeitraum in mehrere Perioden, im Beispiel sind es 13 Perioden mit je 7 Tagen. Anschließend wird der Erwartungswert für jeden Volumenstrom für das Zeitintervall ermittelt und für das beschriebene Beispiel in der Tabelle 4-4 zusammengefasst. Der exemplarische Verlauf für Testroute C endet in der 6. Periode, wobei anschließend die Testroute D eingeführt wird. Bei Testroute B wird das Bewegungsvolumen vergrößert. Das Verhalten des Erwartungswertes der Auswahl der Strecke A ist nahezu konstant. Die AFM berechnet den Erwartungswert wie folgt:

- wenn alle Werte der zugehörigen Perioden  $> 0$  sind, wird ein nach dem Alter der Intervalle gewichteter Mittelwert errechnet,
- wenn die zwei neusten Perioden Null-Werte aufweisen, wird der Erfahrungswert auch gleich 0 gesetzt,
- ansonsten wird ein nach dem Alter der Intervalle gewichteter Mittelwert errechnet, jedoch werden die Zeitintervalle deren Erfahrungswert 0 sind, bei der Berechnung ignoriert.

Da die Summe der mit dem Verfahren errechneten Werte nicht immer 100 % betragen muss (vgl. Tabelle 4-5 – AFM-Erwartungswert Vorstufe), wurde die Normierung auf 100 % eingeführt. Die Validierung des AFM-Verfahrens sowie der Vergleich mit einfacheren Varianten wurde mithilfe von 100 Prognosen durchgeführt. Geprüft wurde dabei aber keine Einzellos-Prognose, also ob genau einem gegebenen Los eine richtige Teststrecke zugeordnet wurde, sondern die Mengenvorhersage. Angenommen von 100 Werkstücken, die an einem Tag den Bereich wechseln, wurden 60 der Teststrecke A und 40 der Teststrecke B zugeordnet, das

angewendete Verfahren berechnet aber eine 55 zu 45 Verteilung, beträgt die Treffsicherheit in dem Fall:

$$T = 100 \% - \left( \frac{|60 - 55|}{60} + \frac{|40 - 45|}{60} \right) * 100 \% = 100 \% - 20,83 \% = 79,17 \%$$

In dem Fall, dass die Abweichung der Prognose zur Realität größer als die reale Anlieferungsmenge ist, beträgt die Treffsicherheit 0 %. Bei der Zusammenfassung der Qualität der Prognosen zu allen Produkten wurde eine Gewichtung nach der Menge der Stichproben eingeführt. Damit wurde auch gewährleistet, dass der zum Teil sehr große Fehler von Sonderprodukten nur eine kleinere Rolle spielt. Tabelle 4-5 stellt den Vergleich der Vorhersagequalität dar. Interessanterweise ist sogar eine naive Prognose, also eine Projektion der Erwartungswerte der letzten Tage unmittelbar vor der Prognoseerstellung, genauer als ein Mittelwert über einen längeren Zeitraum oder ein einfach nach Zeit gewichteter Mittelwert. Durch das deutlich schnellere Abfangen der Änderungen in der Teststreckenwahl und bei gleichzeitiger Robustheit auf temporäre Änderungen ist die von AFM angewendete Methode für die Wahl erwarteter Teststrecke um über 10 % präziser als der nach Zeit gewichtete Mittelwert.

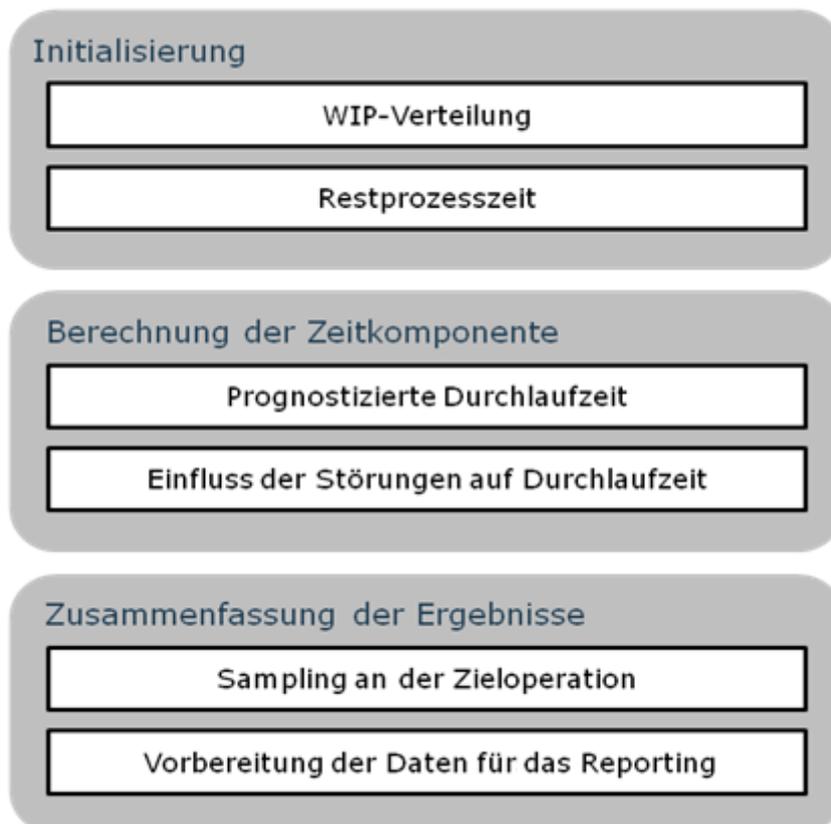


Abbildung 4-26: Der Prozess der Berechnung einer Prognose mittels AFM

## 4.6 Vorhersageberechnung

Die Berechnung einer Ankunftsprognose mit AFM erfolgt mittels der in der Abbildung 4-26 dargestellten Schritte, nämlich: Initialisierung, Berechnung der Zeitkomponente und Zusammenfassung der Ergebnisse. Mit der ersten Maßnahme wird die Verteilung der Lose in der Fabrik bestimmt. Interessant ist an dieser Stelle ebenfalls die Restprozesszeit, falls sich ein Los gerade in der Bearbeitung befindet. Folglich wird die Zeitkomponente der Prognose bestimmt, also die von den Losen benötigte Zeit, um das Ziel zu erreichen. Schließlich werden die Werkstücke von den AFM-Algorithmen bzw. dem Sampling an der Zieloperation untersucht. In einem letzten Schritt werden die Resultate einer Prognose für das Reporting aufbereitet.

### 4.6.1 Auslesen der Bestandsverteilung über die Fabrik und Kalkulation der Planeinschleusung

Wie bereits in den vorherigen Kapiteln beschrieben, bewegt sich das bearbeitete Material in den Halbleiterfabriken auf den vorgegebenen Routen [Atherton95]. Das Auslesen der Bestandsverteilung erfolgt durch das Einfrieren und Speichern der Losposition – also einer Operation im Ablaufplan zu einem gegebenen Zeitpunkt. Dieser Zeitpunkt wird zu einem Prognosestartzeitpunkt erklärt. Einem bereits eingeschleusten Los ist immer eine Operation zugewiesen. In Abbildung 4-27 wurde die Sequenz möglicher Zustände eines Loses in einem Produktionsschritt dargestellt. Aus Sicht der Prognose sind der Zustand und seine Dauer natürlich

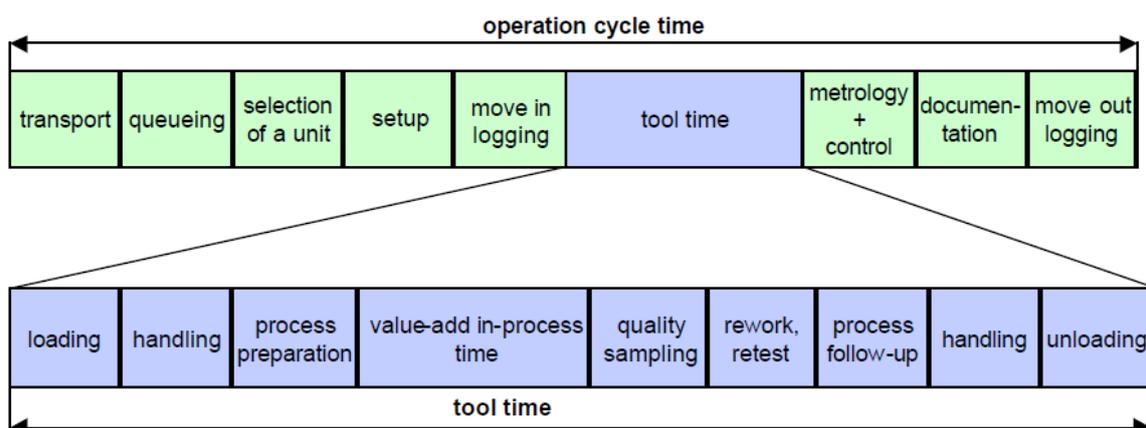


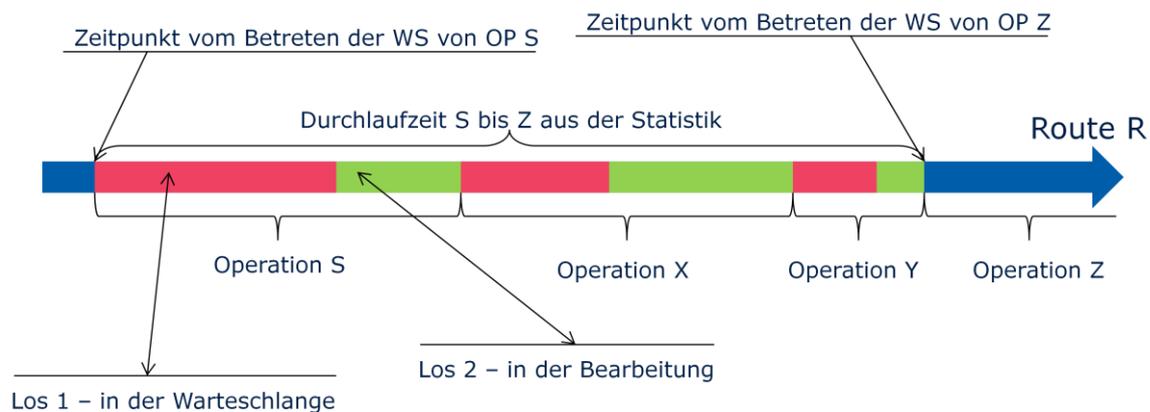
Abbildung 4-27: Die Sequenz möglicher Zustände eines Loses in einer Operation

von Bedeutung. Die möglichst genaue Initialisierung ist vor allem im Bereich der Ultrakurzfristvoraussage mit einem Horizont bis 12 Stunden erforderlich. Da die AFM ausschließlich auf der Beschreibung der historischen Losbewegungen basiert, stehen insofern detaillierte

Informationen über den Loszustand nicht zur Verfügung. Es lassen sich nur folgende Eintrittszeitpunkte von den Zuständen eines Werkstückes ermitteln:

- Eingang und Ausgang im Transportsystem,
- Betreten und Verlassen der Warteschlange,
- Start und Ende der Bearbeitung.

Da die Transportzeit eigentlich nur bei den priorisierten Losen eine Rolle bzgl. der Ankunftsprognose spielt und nur einen Bruchteil der gesamten Durchlaufzeit beim untersuchten Prognosehorizont beträgt, wurden bei der AFM die Transportzeit in die Wartezeit eingebettet. Das bedeutet, die Wartezeit fängt sofort beim Verlassen der Voroperation an. In der untersuchten Fertigung beeinflusst vor allem die Priorität und die Verspätung (bei dem sich an der Verspätung orientierenden Dispatching) die Dauer des Aufenthalts in der Warteschlange. Die AFM Durchlaufzeitstatistik ist so aufgebaut, dass bei der Startoperation S und der Zieleoperation Z, die Durchlaufzeit von S bis Z mit Betreten der Warteschlange in S anfängt und mit Betreten der Warteschlange in Z endet. Die Abbildung 4-28 stellt mit dem Los 1 ein Werkstück, das sich in der Warteschlange befindet, dar. Der Zeitpunkt des Betretens der Warteschlange wird gespeichert und dient als der Ausgangszeitpunkt bei der Vorhersageerstellung.



**Abbildung 4-28: Mögliche Losposition bei der Initialisierung der Prognose**

Der Vorteil des Ansatzes von AFM liegt vor allem in der Begrenzung der Datenmenge (ähnlich wie in [Backus06]). Aufgrund der Einbettung der Restwartezeit in der Gesamtdurchlaufzeit ist eine Wartezeitstatistik für jede Operation nicht erforderlich. Problematisch ist allerdings die Behandlung der Lose, die kurz vor dem Erreichen des Ziels eine ungewöhnlich lange Warteschlange betreten. Derartige Warteschlangen entstehen durch Maschinenausfälle oder andere Produktionsstörungen. Hypothetisch ist also der Fall möglich, in dem die gesamte Durchlaufzeit kleiner als die bereits absolvierte Wartezeit in der Startoperation ist. Daraus resultiert eine Prognose der Ankunft, die vor dem Initialisierungszeitpunkt liegt. Die AFM

weist allen derartigen Werkstücken den Zeitpunkt der Initialisierung der Prognose als den Ankunftszeitpunkt zu. Wenn für ein Los in dem Startschritt die Bearbeitung bereits angefangen wurde, sollte die prognostizierte Durchlaufzeit aus der Summe der Restprozesszeit und der gesamten Durchlaufzeit der nächsten Operation bis zum Ziel resultieren. Der Ansatz erfordert genaue Angaben zu Prozesszeiten. Da die AFM die Bearbeitungszeiten im benötigten Genauigkeitsgrad aus den Bewegungsdaten nicht extrahieren kann, müsste an dieser Stelle eine andere Datenquelle angebunden werden. Im Moment arbeitet AFM auch mit dem Betreten der Warteschlange als Startzeitpunkt – auch wenn ein Los sich bereits in der Anlage befindet. Das modulare Konzept der Prognosemethode erlaubt das Einbauen von Zusatzfunktionen und macht bei Bedarf die Erweiterung der Initialisierung denkbar.

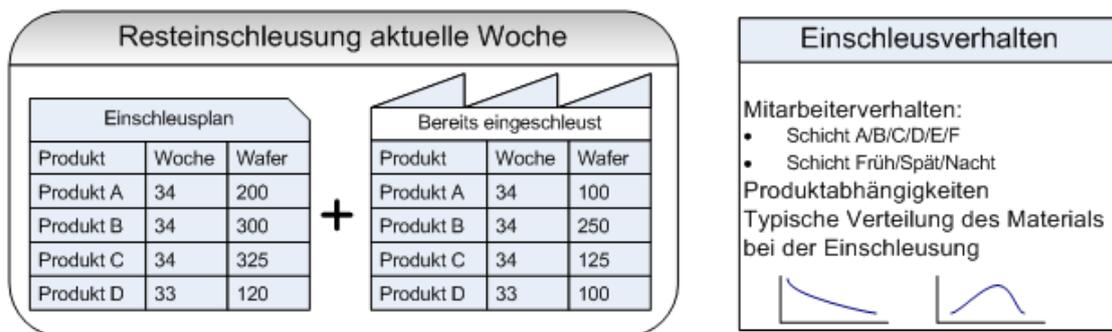
Die Behandlung der Lose, die aus verschiedenen Gründen geteilt wurden (vgl. Kapitel 4.4.3) stellt eine weitere Herausforderung bei der Erstellung der Prognose in der Halbleiterfertigung dar. In der Produktionslinie bewegen sich außer standardisierten auch Werkstücke mit einer kleinen Anzahl von Wafern in der Horde. Ein Teil der Lose wurde absichtlich in dieser Form eingeschleust und stellt kein Problem bei der Kalkulation der Prognose dar. Die Lose jedoch, welche während des Fertigungsprozesses gesplittet wurden, benötigen eine entsprechende Sonderbehandlung seitens der Vorhersagealgorithmen. Im Rahmen des Projektes wurden drei Ansätze implementiert und validiert:

- keine Behandlung – es wird für jedes Los eine Voraussage der Durchlaufzeit berechnet,
- die Voraussage nur für Mutter-Lose, die restlichen Lose werden ignoriert,
- die Zusammenfassung der gesplitteten Lose zu einem Mutterlos und eine entsprechende Voraussage der Durchlaufzeit.

Bei dem ersten Ansatz werden alle Split- und Mutterlose gleich behandelt, somit erhält jedes der Lose einen Ankunftstermin bei der Zieloperation. Der Vorteil der Methodik liegt in der Verwendung der Standardfunktionalität und erfordert keine Zusatzimplementierung. Die Nachteile disqualifizieren jedoch diesen Ansatz. Wenn die angelieferte Menge von Wafern kein Problem hervorruft, verfälscht die Anzahl der angelieferten Lose vor allem die Samplingkalkulation und das Bottleneck-Frühwarnsystem. Die Voraussage begrenzt auf die Mutter-Lose erfordert die Filterung der Lose bei der Initialisierung. Danach ist die Prognose mit den Standardalgorithmen möglich. Bei Verwendung der Methode wird allerdings die Anlieferungsmenge in der Waferanzahl verfälscht, zusätzlich werden die Werkstücke wegen einer geringeren Menge an Wafern falsch klassifiziert. Die dritte Methodik, die auch von AFM verwendet wird, versucht vorerst alle Mutterlose und von ihnen abgesplittete Tochterlo-

se zu finden. Anschließend werden die Lose durch Erhöhung der Anzahl der Siliziumscheiben in dem Ausgangslos künstlich zusammengefasst. Problematisch ist nur die Wahl der Startoperation für die Vorhersage für den Fall, dass sich die einzelnen Lose in unterschiedlichen Produktionsschritten befinden. Die AFM übernimmt an dieser Stelle die Angaben des Loses, welches am weitesten von der Zieloperation steht und platziert dort das künstlich zusammengefasste Werkstück. Die Validierung der drei Methoden in Bezug auf die Vorhersagequalität für das geteilte Material spricht deutlich für die Wahl der dritten Alternative. Allgemein ist jedoch der prozentuale Anteil derartiger Lose in der Fertigung marginal, deshalb hat die Wahl der angewendeten Methode auf die zusammengefassten Resultate kaum einen Einfluss.

Da der Prognosehorizont 7 bis 14 Tage beträgt, ist nicht nur die momentane Bestandsverteilung von Interesse, sondern auch die zukünftigen Losstarts. Abbildung 4-29 stellt alle Informationen dar, die notwendig sind, um einen detaillierten Plan für die Einschleusung des



**Abbildung 4-29: Notwendige Informationen für die Berechnung eines Planes für die Materialeinschleusung**

Materials zu erstellen. In Kapitel 4.5.2 wurde genauer erläutert, dass eine bei der taktischen Planung erstellte Aufstellung in eine operative Disposition übersetzt werden muss. Im ersten Schritt wird die Resteinschleusung für die aktuelle Woche als Quotient der geplanten Materialmenge und das bereits gestartete Loskontingent ermittelt. Die Zahl wird ebenso, falls vorhanden, um das verbliebene Restkontingent an Losen aus den vorherigen Wochen ergänzt. Die zeitliche Losverteilung erfolgt mittels Anwendung der erstellten Statistik. Die Resteinschleusung wird vorerst unter Benutzung von typischen Startmengen gleich verteilt auf alle zukünftigen Schichten gestreut. Anschließend wird schicht- und produktfein ein Schätzwert für einen Zeitpunkt der Einschleusung des ersten Loses ermittelt. Die Zeitpunkte für den gesamten folgenden Materialstrom werden unter den Ausschluss der Pausen gleichverteilt. Auf dieser Weise entsteht ein Entwurf eines operativen Plans für die Einschleusung. Das konzeptionelle Materialstartschema wird schließlich durch die AFM-Algorithmen mit dem Pausenplan verglichen und angepasst. Die Notwendigkeit der Anwendung eines relativ komplizier-

ten Berechnungsverfahren resultiert aus der wesentlichen Verbesserung der Vorhersagequalität im Vergleich zu anderen Ansätzen. Eine detaillierte Validierung durch einen Vergleich mit unterschiedlichen Planerstellungstrategien wurde durchgeführt und kann Kapitel 6.4.1 entnommen werden.

### 4.6.2 Bestimmung der Ankunftszeitpunkte an der Zieloperation

Der Ausgangspunkt der Bestimmung der Ankunftszeiten an den Zieloperationen bildet der Snapshot der Bestandsverteilung der Fabrik inkl. der zukünftigen Einschleusung von Material. Jedes der für die Prognose relevanten Lose verfügt über die Information hinsichtlich Startzeitpunkt  $T_S$ , Startoperation  $S$ , Zieloperation  $T$  sowie anderen für die Klassifizierung notwendigen Informationen. Die Berechnung des Ankunftszeitpunktes  $T_T$  erfolgt mit der Formel:

$$T_T = T_S + t_{HS \rightarrow T} + t_{CTS \rightarrow T} ,$$

wobei  $t_{HS \rightarrow T}$  die prognostizierte Dauer des Anhaltezustands und  $t_{CTS \rightarrow T}$  der Durchlaufzeit ohne Störungen bedeuten. Die Berechnung der Verzögerung durch Hold wurde näher in Kapitel 4.6.2.1 beschrieben. Die Kalkulation der Durchlaufzeit bestimmt zuerst die Einteilung den Losen in die AFM-Klassen. Dies erfolgt basierend auf dem gleichen Prinzip wie die Klassifizierung bei der Statistikerzeugung (vgl. Kapitel 4.4). Anschließend werden aus der Durchlaufzeitstatistik die entsprechenden Zeitinformationen extrahiert.

Der modulare Aufbau der AFM erlaubt eine einfache Erweiterung des Algorithmus für die Berechnung der Ankunftszeit um weitere Zeitkomponenten. Die Formel für die  $T_T$  Kalkulation wird einfach um neue Verzögerungsglieder  $t_{VxS \rightarrow T}$  ergänzt:

$$T_T = T_S + t_{HS \rightarrow T} + t_{CTS \rightarrow T} + t_{VxS \rightarrow T}$$

Die AFM verfügt konzeptionell über keine Kapazitätsmodelle der Anlagen. Es bedeutet für die Prognose, dass über jede Operation eine unbegrenzte Menge des Materials pro Zeiteinheit bewegt werden kann. Das führt in manchen Fällen zu einer zu optimistischen Prognose der Anlieferung. Derzeitig arbeitet unser Team an der Erstellung eines Verzögerungsgliedes für das Bremsen des Materials an den Bottleneck-Produktionsschritten. Abbildung 4-30 stellt schematisch das Konzept für die Einführung der Kapazitätsbegrenzung in das AFM Model dar.

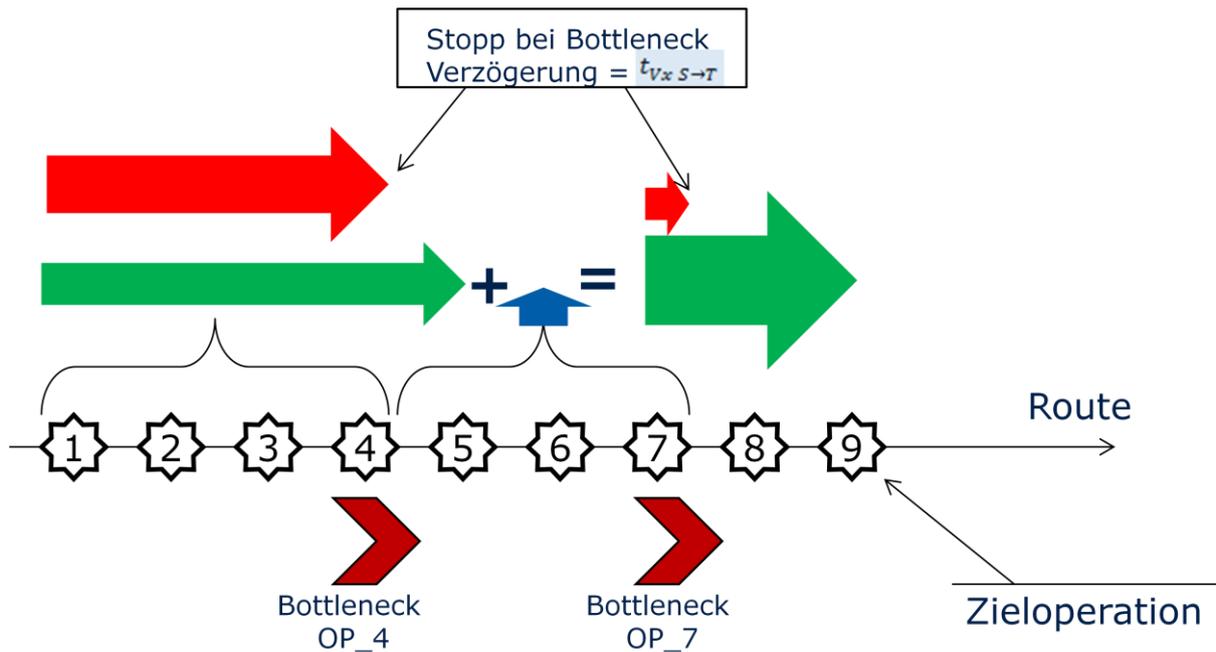


Abbildung 4-30: AFM-Konzept für den Umgang mit begrenzter Produktionskapazität an den Operationen zwischen Start- und Zieloperation

Im ersten Schritt werden die typischen Bottleneck-Operationen und deren Durchsatz eruiert. Die Engpässe werden als Zwischenziele deklariert. Anschließend wird eine AFM-Voraussage für die erste Zieloperation (Bottleneck OP\_4 in der Abbildung 4-30) erstellt. Die Lose, die aufgrund des beschränkten Durchsatzes nicht bearbeitet werden können, werden ausgebremst. Sie erhalten einen neuen Ankunftszeitpunkt durch Addition von der Verzögerungszeit  $t_{Vx S \rightarrow T}$  – deren Verspätung steigt dementsprechend. Diese Ankunfts voraussage dient im Anschluss als Ausgang für die weiteren Vorhersageschritte, bis auf dem ähnlichen Prinzip jedes Bottleneck abgearbeitet wurde. Die ersten durchgeführten Experimente beweisen die Verbesserung der Aussagequalität der Prognosen. Eine bisher unvollständige Umsetzung des Konzeptes verhindert jedoch deren umfangreiche Validierung.

#### 4.6.2.1 Die Verzögerung durch den Hold-Zustand

Die Ankunft eines Loses an der Zieloperation wird durch das mögliche Anhalten an den bestimmten Operationen der Route verzögert. Es erfolgt durch die Addition der prognostizierten Hold-Dauer  $t_H$  zu der regulären Durchlaufzeit. Angenommen ein Los befindet sich in einer Startoperation  $S$  und soll prognostiziert die Zieloperation  $T$  erreichen, wird die gesamte Dauer des Anhaltezustand  $t_{H S \rightarrow T}$  durch die AFM mittels der Formel ermittelt:

$$t_{H S \rightarrow T} = \sum_{i=1}^n t_{Hi} , \quad \text{wobei } t_{Hi} = \begin{cases} \varphi_i \Leftrightarrow \beta_i \geq \text{rand}[0,100] \\ 0 \Leftrightarrow \beta_i < \text{rand}[0,100] \end{cases}$$

Die produktspezifischen Angaben für die Hold-Dauer  $\varphi_i$  und die Hold-Rate  $\beta_i$  für eine bestimmte Operation wurden im Voraus statistisch erfasst (vgl. Kapitel 4.5.5). Die Anzahl der Operationen  $n$  zwischen Start und Ziel stammt auch aus der Statistik, da ausschließlich die Schritte interessant sind, bei denen die Hold-Rate größer als Null ist. Tabelle 4-6 stellt exemplarisch einen Statistikauszug für ein gegebenes Produkt dar. Die Startoperation  $S$  ist Schritt 5000 im Abarbeitungsplan und die Zieloperation Schritt 5064. Die Markierung der Operation als erste Operation  $i = 1$  bei der Berechnung der gesamten Hold-Dauer ist abhängig von der Losposition bei der Initialisierung der Prognose (Initiale Bestandsverteilung, vgl. Kapitel 4.6.1). Die AFM behandelt die Position des Materials für die Ermittlung des ersten Schrittes wie folgt:

- Das Los befindet sich in einer Operation  $X$  und die Abarbeitung wurde bereits gestartet, so wird ein nächsthöherer, mit Hold behafteter Schritt als  $i = 1$  markiert.
- Das Los befindet sich bei der Initialisierung im Hold. Es wird eine Restdauer des Holds ermittelt und zu  $t_{H\ S \rightarrow T}$  addiert. Ein nächsthöherer, mit Hold behafteter Schritt wird als  $i = 1$  markiert.
- Das Los befindet sich in der Warteschlange an einer Operation, für welche die Hold-Rate gleich Null ist. Somit wird ein nächsthöherer, mit Hold behafteter Schritt als  $i = 1$  markiert.
- Das Los befindet sich in der Warteschlange an einer Operation für welche, die Hold-Rate grösser Null ist, so wird genau dieser Schritt als  $i = 1$  markiert (so wie in Tabelle 4-6)

Operation $i$	$S = 5000$	5010	5020	5029	5030	5050	$T = 5064$
$i$	1	2	3	4	5	6(n)	
$\varphi_i$ [min]	12	3	145	122	34	22	15
$\beta_i$	25	5	5	4	15	3	25
$rand[0, 100]$	24	87	34	45	12	98	65
$t_{Hi}$	12	0	0	0	34	0	0

**Tabelle 4-6: Beispiel von der Verwendung der Hold-Statistik für die Berechnung der prognostizierten Verzögerung der Losankunft an der Zieloperation**

Nach der Ermittlung des Schrittes  $i = 1$  ( $S = 5000$  in der Tabelle 4-6) und dem Auslesen von weiteren, zugehörigen Hold-Raten  $\beta_i$  und Dauer  $\varphi_i$  aus der Statistik, wird für jede dieser Operationen eine Zufallszahl im Bereich von 0 bis 100 erzeugt. Für die Berechnung der gesamten Dauer des Anhaltezustandes  $t_{H\ S \rightarrow T}$  werden nur die Datensätze übernommen, bei denen der

entsprechende Zufallswert kleiner als die Hold-Rate ist. Im beschriebenen Beispiel werden dies dementsprechend nur die Schritte  $i = 1$  mit 12 min und  $i = 5$  mit 34 min sein. Der von AFM verwendete Algorithmus wird in derselben bzw. ähnlichen Form auch von sehr vielen gängigen Simulationswerkzeugen genutzt.

### 4.6.2.2 Die Berechnung der Anlieferungsprognose für den Testbereich der Fertigung

Die Berechnung der Anlieferungsprognose für den Wafer-Testbereich erfolgt bei der AFM in drei Phasen, die in Abbildung 4-31 dargestellt werden. In der ersten Phase werden die Zeitpunkte des prognostizierten Verlassens des Produktionsbereichs ermittelt. Die Methodik der Kalkulation wurde in Kapitel 4.6.2 beschrieben. In Phase II werden die Losinformationen erneut ausgewertet. Anhand der vorliegenden statistischen Informationen über den Übergang zwischen den Bereichen ordnen die AFM-Algorithmen jedem Los eine Teststrecke zu. Für die Teststreckenauswahl wird ein angepasstes Acceptance-Rejection-Verfahren [Knuth97]



**Abbildung 4-31: Das Phasenmodell für die Voraussage der Anlieferungszeitpunkte in dem Testbereich der Fertigung**

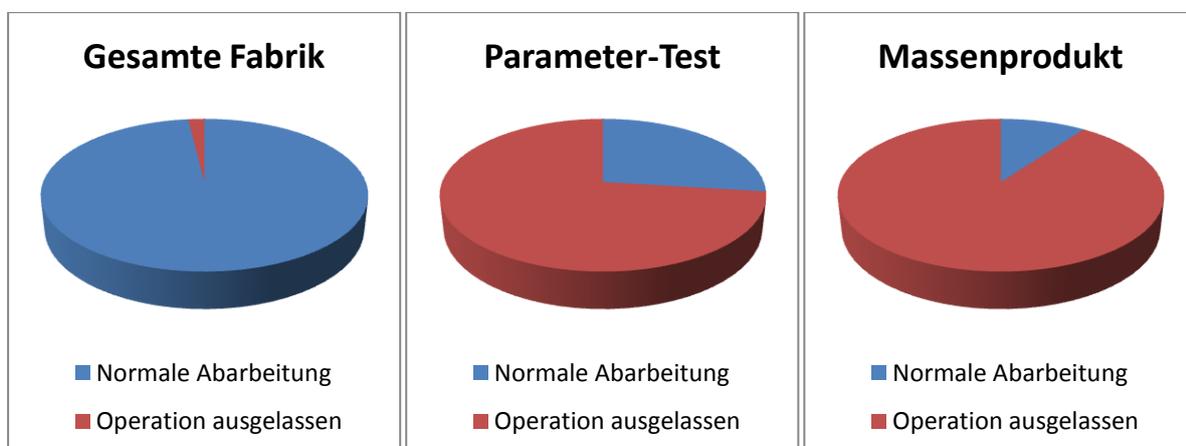
angewendet. Angenommen für ein Produkt stehen 3 Teststrecken zur Verfügung, stellen die entsprechenden Daten aus der Statistik eine Verteilung für die Wahl der Routen in dem Wafer-Testbereich wie folgt dar:  $R1 = 55\%$ ,  $R2=35\%$ ,  $R3=10\%$ . Daraus resultieren die Wertebereiche für  $R1$ :  $[1,55]$ ,  $R2$ :  $(55,90]$  und  $R3$ :  $(90,100]$ . Für jedes Los wird eine über einen Bereich von 1 bis 100 gespannte, normalverteilte Zufallszahl erzeugt. Anschließend erfolgt die Überprüfung, zu welchem der Wertebereiche aus der Statistik die Zahl gehört. In dem beschriebenen Beispiel wurde die Teststrecke  $R1$  angenommen, wenn der zufällige Wert kleiner gleich 55 wäre. Allen Werten die größer 55 und kleiner gleich 90 sind, würde  $R2$ , ansonsten  $R3$  zugeordnet. Nachdem die prognostizierte Teststrecke bestimmt wurde, werden die betroffenen Lose an deren Startoperation positioniert. Die Ermittlung des Ankunftszeitpunktes an der Zieloperation erfolgt wie im Produktionsbereich (vgl. Kapitel 4.6.2).

### 4.6.3 Ermittlung von Anlieferungsmengen an der Zieloperation

Die AFM fasst die prognostizierten Ankünfte des Materials an den Zieloperationen zu Ankunftsströmen zusammen. Dies erfolgt in zwei Schritten. Zuerst wird untersucht, ob die ankommenden Lose an den Zieloperationen überhaupt prozessiert werden oder die Operation überspringen. Die Kalkulation des Los-Sampling wird in dem Kapitel 4.6.3.1 detailliert beschrieben. In einem zweiten Schritt werden die als gültig markierten Ankünfte nach den Zieloperationen gruppiert und nach Ankunftszeit sortiert. In der letzten Phase werden für die an den Zweck angepassten Zeitintervalle (zum Beispiel: 24h - Tag, 8h - Schicht) die Materialmengen bestimmt. Die Ergebnisse der Prognose, sowie möglichen Anwendungsfälle, werden in Kapitel 4.7 beschrieben.

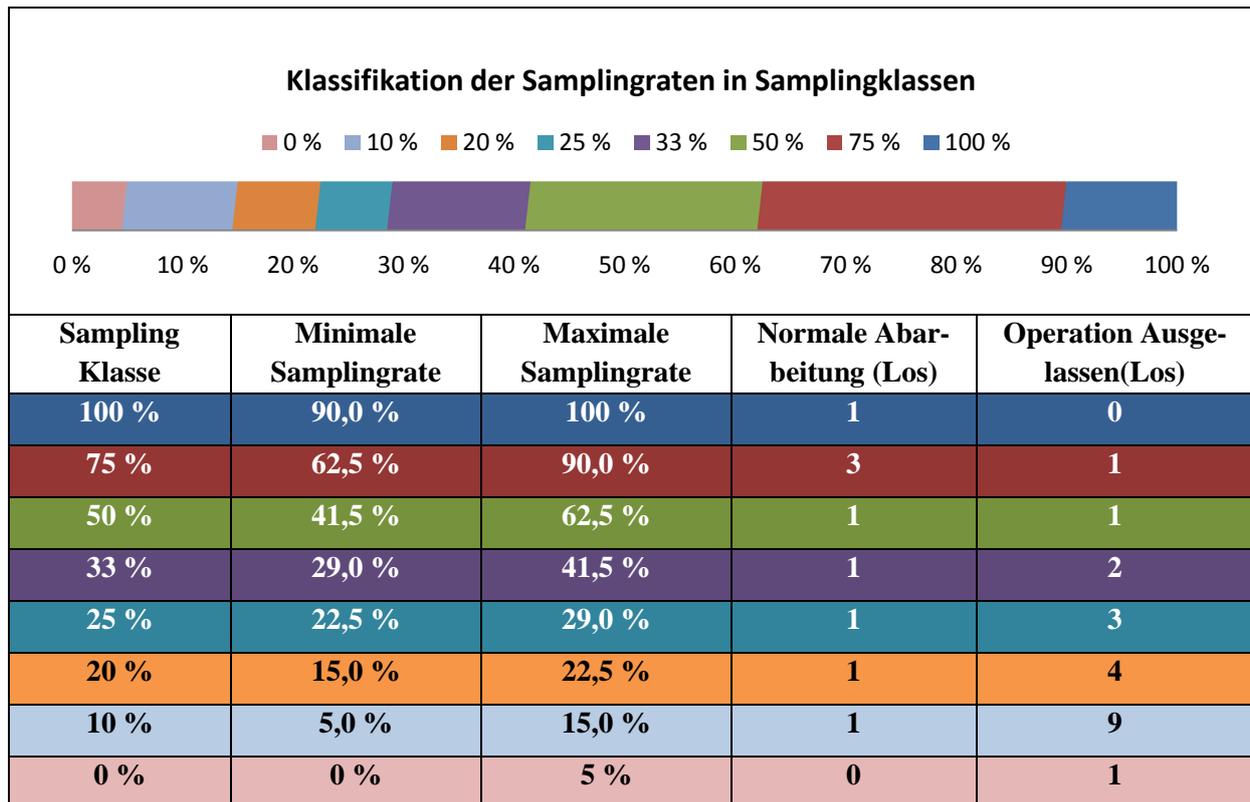
#### 4.6.3.1 Anpassung der Anlieferungsmengen aufgrund von Los-Sampling

Die Notwendigkeit der Anpassung der Anlieferungsvorhersage resultiert aus den Los-Sampling an verschiedenen Zieloperationen (mehr dazu im. Kapitel 2.4.5). Bei der Untersuchung der Gesamtheit der Bewegungen des Materials der Fabrik wird der Einfluss des Samplings deutlich, wie in Abbildung 4-32: Die Samplingrate für die gesamte Fabrik, Produktionsbereich Parameter-Test und ein Massenprodukt an einer gegebenen Testoperation dargestellt. Die Samplingrate für die gesamte Fabrik ist sehr gering. In der Abbildung sind allerdings auch zwei Gegenbeispiele ersichtlich. In allen Produktionsbereichen, die sich mit der Inspektion der Werkstücke befassen, zeigen die Samplingraten sehr große Auswirkungen auf die Materialanlieferung. Ganz deutlich ist es bezüglich der Massenprodukte, die an bestimmten Vermessungsoperationen eine konstante Samplingrate von 10 % besitzen, also nur jedes zehnte Los wird an der Operation abgearbeitet.



**Abbildung 4-32: Die Samplingrate für die gesamte Fabrik, Produktionsbereich Parameter-Test und ein Massenprodukt an einer gegebenen Testoperation**

Wie in Kapitel 4.5.4 wird die für die Funktionsweise von AFM notwendige Statistik über Samplingraten im Vorfeld berechnet und die Werte liegen zum Startzeitpunkt der Vorhersage vor. Der AFM-Ansatz für die Behandlung des Los-Sampling basiert auf festen Samplingklassen, die von den in der Realität angewendeten Algorithmen abgeleitet wurden. Vereinfacht ist die Klassifizierung in Tabelle 4-7 dargestellt, die für jede Klassenbezeichnung einen Zugehörigkeitsbereich und die Samplingregel beinhaltet. Angenommen die aus der



**Tabelle 4-7: Klassifikation der Samplingraten in Samplingklassen**

Statistik ausgelesene Rate für ein bestimmtes Produkt an einer Operation beträgt 27,4 %. In diesem Fall legt die Prognosemethode die Zugehörigkeit zur Klasse 25 % fest, für die gilt, dass nach der normalen Abarbeitung von einem Los die nächsten ankommenden drei Werkstücke die Operation überspringen sollen. Die Notwendigkeit für die Einführung von Clustern der Samplingrate ist bei AFM durch den Berechnungsansatz der Materialanlieferungsmengen bedingt. Die Funktionsweise der Samplingberechnung lässt sich am einfachsten an einem Beispiel erläutern. Angenommen ein Produkt an einer Operation wurde zu der Samplingklasse 33 % zugeordnet. Für die 33 % - Klasse folgen auf ein bearbeitetes Los zwei Lose die den Prozessschritt überspringen (vgl. Tabelle 4-7). Abbildung 4-33 stellt die theoretische Losankunftsfolge dar. Bei der Initialisierung der Prognoseberechnung ist die Information über die Anzahl der Lose obligatorisch, die nach dem letzten prozessierten Artikel, den analysierten

Schritt übersprungen haben. Im beschriebenen Beispiel wird ein Los um 12:05 prozessiert und bis zum Start der Prognose überspringt ein weiteres die Operation (vgl. Abbildung 4-33).

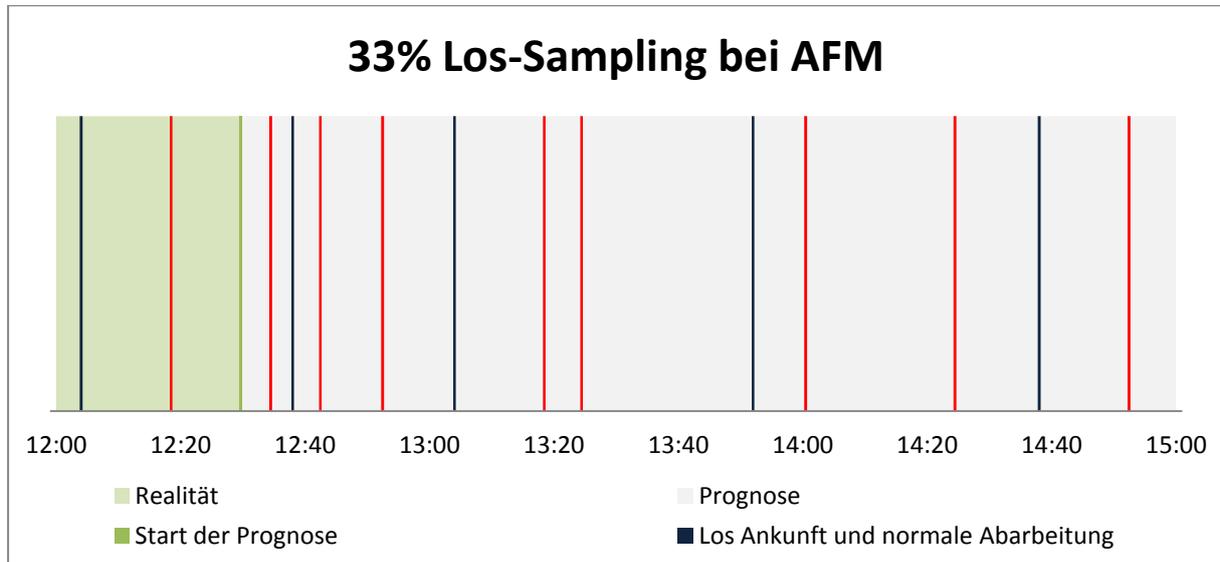


Abbildung 4-33: 33 % Los-Sampling bei AFM

Daraus resultiert, dass der nächste ankommende auch diesen Prozessschritt auslassen wird. Die AFM berechnet die Ankunftszeitpunkte des Materials an der Zieloperation und sortiert diese nach der zeitlichen Entfernung vom Zeitpunkt des Prognosestarts. Mithilfe der Initialisierungsinformationen wird das erste Los die Operation überspringen und die folgenden mit dem für die Samplingklasse spezifischen Regeln markiert (eins bearbeiten und zwei danach auslassen). Da in der Literatur auch weitere Methoden für die Prognose des Los-Samplings bekannt sind [Fishman01], wurden sie im Rahmen des Projektes validiert. Die Begründung für die Entscheidung, die obige Methode anzuwenden, wird detailliert in Kapitel 6.4.2 erläutert.

### 4.7 Vorhersageergebnisse und Anwendungsfälle

Die AFM liefert dem Anwender neben der in den funktionalen Anforderungen definierten Prognose von Anlieferungsströmen an der Zieloperation auch eine Voraussage von Ankunftszeitpunkten für einzelne Lose. Ausgehend von den beiden Informationsquellen wurden vielseitige Anwendungsmöglichkeiten der Prognose entworfen und realisiert. Da die AFM bereits implementiert wurde und auch von dem Projektpartner angewendet wird, werden an der Stelle nur wenige repräsentative Anwendungsfälle beschrieben, wie:

- Frühwarnsystem für Kapazitätsengpässe,
- Unterstützung bei der Planung von Anlagenreservierungen,

- Unterstützung bei der Priorisierung der Reparaturmaßnahmen,
- Assistenz bei der Optimierung der Personalplanung,
- Informationsboard zur Erfüllung der Produktionsziele,
- Liefervoraussagen für Sonderprodukte,
- Analyse der Liniensituation bzgl. Der Lieferzeit

Durch die Ergänzung der prognostizierten Mengen der Anlieferung um die Abarbeitungszeiten an der Zieloperation wurde ein Frühwarnsystem an den Engpässen der Produktionslinie entwickelt. Angenommen an der Flaschenhalsstelle werden  $n$  Anlagen mit der Verfügbarkeit  $\omega$  [%] für die Bearbeitung vorgesehen, dann lässt sich die maximal mögliche Belegungszeit  $MBT_{ges}$  [h] in der Zeitperiode  $tp$  [h] mit folgender Formel ausrechnen:

$$MBT_{ges} = \sum_{i=1}^n \omega_i * tp$$

Für die Approximation der prognostizierten gesamten Belegungszeit der Maschinen  $PBT_{ges}$  [h] verwendet die AFM die Formel:

$$PBT_{ges} [h] = m(tp) * \bar{t}_s + \sum_{i=1}^{m(tp)} (w_i * \bar{t}_h + \overline{t_p(\text{prod}_i)}),$$

wobei  $m(tp)$  die prognostizierte Anzahl der Lose in der Zeitperiode  $tp$ ,  $\bar{t}_s$  mittlere Setupzeit,  $w_i$  die Anzahl der prozessierten Wafer von Los  $i$ ,  $\bar{t}_h$  die mittlere Handlingzeit pro Wafer bedeuten. Die mittlere Prozesszeit über den  $n$  analysierten Maschinen  $\overline{t_p(\text{prod}_i)}$  wird mithilfe der Produktangehörigkeit des Loses  $i$  ermittelt. Bei der Annahme, dass die maximal mögliche Belegungszeit  $MBT_{ges}$  [h] in der Zeitperiode  $tp$  [h] 100 % beträgt, wird die prognostizierte Auslastung  $util(tp)$  [%] wie folgt berechnet:

$$util(tp) [\%] = \frac{PBT_{ges} [h]}{MBT_{ges} [h]} * 100 \%$$

Bei unserem Projektpartner wird die Prognose unter anderem im Parameter-Test-Bereich angewendet. Der Grund dafür ist vor allem die wellenhafte Anlieferung des Produktionsguts. Eine Zeitperiode  $tp$  [h] von 24 h ist ausreichend. Damit wird die approximierete tägliche Auslastung, wie in der Abbildung 4-34 dargestellt, prognostiziert. Das Frühwarnsystem erleichtert beispielsweise die Entscheidungsfindung über die Requalifikation von weiteren Maschinen in einem anderen Betriebsmodus. Aufgrund der früheren Reaktion auf die Bottleneck-Situation wird die Verlängerung der Durchlaufzeit minimiert. Die Auslastungsprognose hilft auch, das

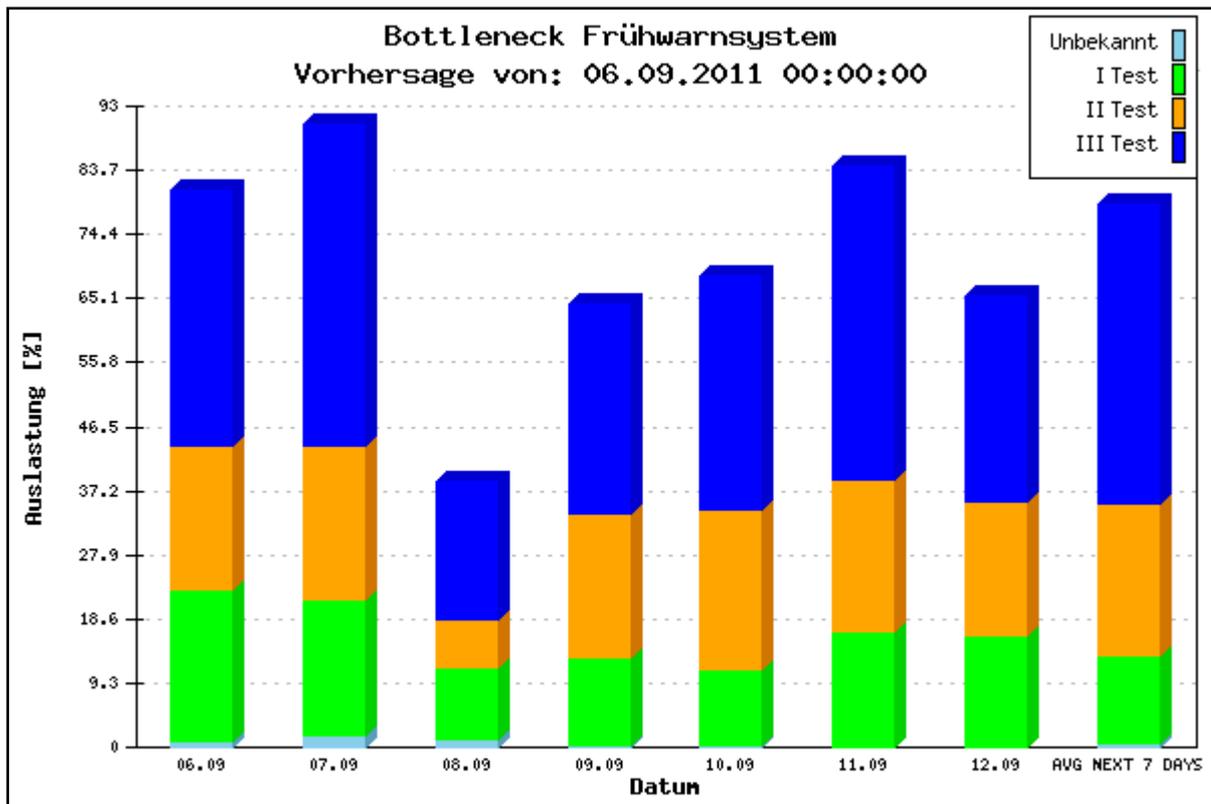


Abbildung 4-34: Das Bottleneck-Frühwarnsystem mittels AFM

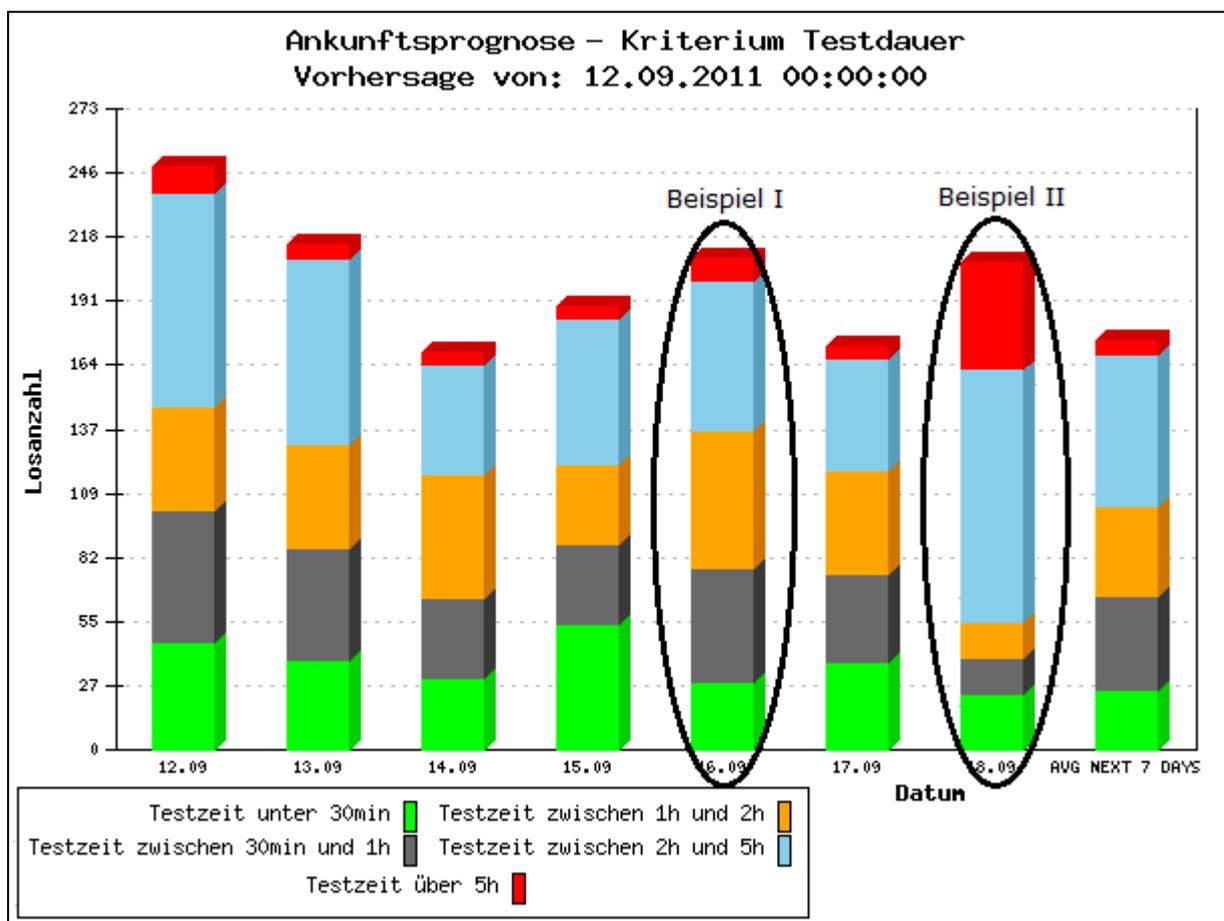


Abbildung 4-35: AFM Ankunftsprognose: Anzahl der Lose gruppiert nach der Prozesszeit

optimale Zeitfenster für die Anlagenbelegung zu finden, was bei der Planung der Anlagenreservierung für vorbeugende Wartungen oder Entwicklungstätigkeiten notwendig ist. Die Maschinenauslastung korreliert jedoch nicht ausreichend mit dem Personalbedarf. Bei der Verwendung der AFM-Prognose als eine der Prämissen der Personalplanung ist es notwendig, die Auslastungsprognose um die Information über die prognostizierte Anzahl der Lose zu ergänzen. Abbildung 4-35 stellt die Prognose der Anlieferung in einem der Testbereiche dar. Das Material wurde nach Länge der Prozesszeit (Test) in eine der 5 Gruppen eingeteilt. Das Personal ist in diesem Produktionsbereich hauptsächlich mit komplizierten Be- und Entladevorgängen beschäftigt. In der Abbildung wurden zwei Beispiele hervorgehoben:

- Beispiel I: Niedrige Auslastung aufgrund hoher Anzahl von angelieferten Produkten, deren Prozessdauer unter 2 Stunden liegt. Nur eine mittlere Anzahl an Be- und Entladevorgängen.
- Beispiel II: Auslastung ist sehr hoch, weil die Menge des Materials mit langer Abarbeitungsdauer sehr groß ist. Es ist eine typische Bottleneck-Situation bzgl. der Anlagenkapazität. Die Anzahl der Vorgänge, bei denen der Operator benötigt wird, ist aber geringfügig kleiner als in dem Beispiel I.

Obwohl sich also die Maschinenauslastung in beiden Beispielen deutlich voneinander unterscheidet, bleibt der Personalbedarf ungefähr gleich.

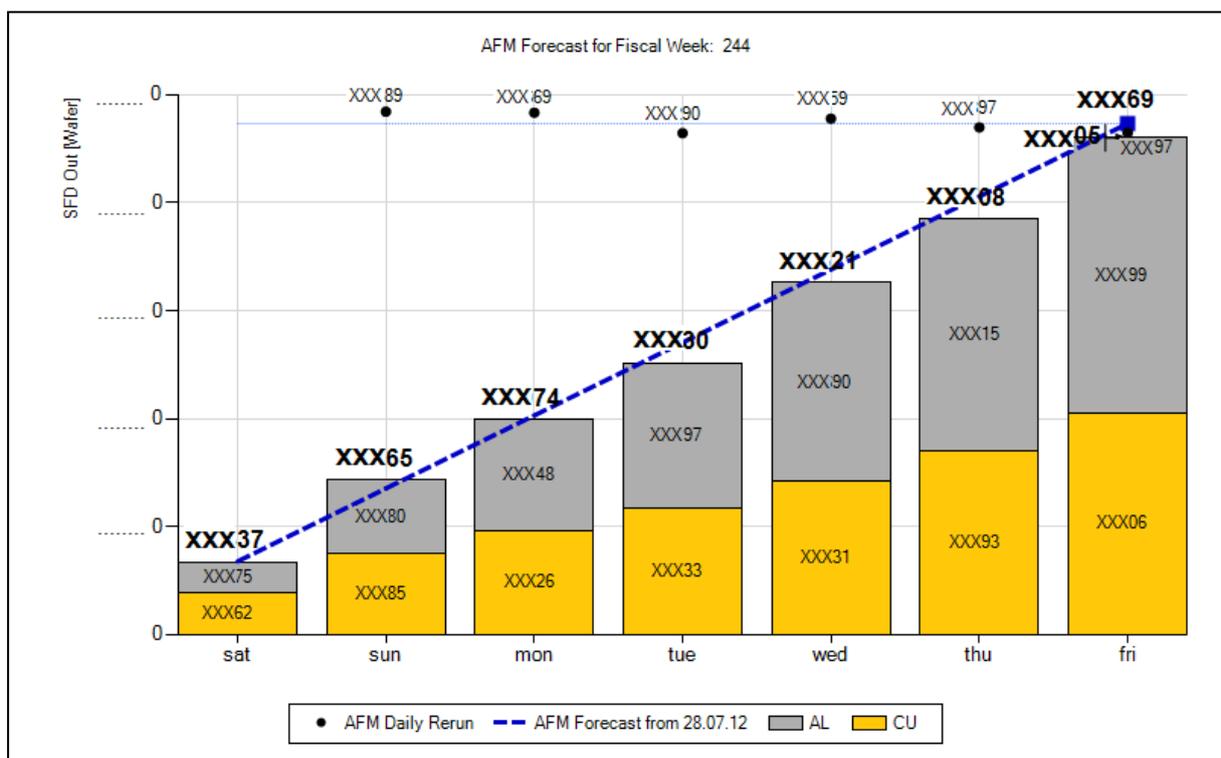


Abbildung 4-36: AFM-Prognose der Fabrikoutputs in gegebener Woche

Die AFM Voraussagen und Berichte werden in verschiedenen Meetings und Linienbriefings präsentiert. Aus diesem Grund wurde ein Informationsboard entwickelt, das die anwendungsorientierten Diagramme mit verschiedenen Informationen zur gesamten Produktionslinie oder einzelnen Produktionsbereichen beinhaltet. Ein Beispiel eines Musterdiagramms wurde in Abbildung 4-36 dargestellt. Der Graph stellt die Fabrikausschleusung (aufgrund des Betriebsgeheimnis anonymisiert) für die gegebene Woche dar. Vorerst wird eine Prognose für die Wochengesamtausschleusung erzeugt und als blaue Linie geplottet. Nach dem Ablauf von jedem Wochentag werden die Informationen über das bereits ausgeschleuste Material in der Form von Balken (AL/CU) abgebildet. Die AFM erzeugt jeden Tag eine Prognose. Dadurch ist ein tägliches Update der Wochenlieferung möglich, das sich aus der Summe der prognostizierten Restlieferungen und der bereits ausgeschleusten Materialmenge ergibt. Den Beobachtungen zufolge bietet der tägliche Wiederholungslauf eine Art Trend, der den Fabrikzustand im Backend of Line widerspiegelt. Bei gravierenden Problemen in der Fabrik treten deutlich fallende Trends auf. Daraus lässt sich ableiten, welche Auswirkungen auf die Liefermenge ein gegebenes Problem hatte. Da die Diagrammgestaltung, wie auch der Informationsbedarf von den Anwendern der Prognose stammt, befindet sich das AFM-Informationsboard immer noch im Entwicklungsstadium.

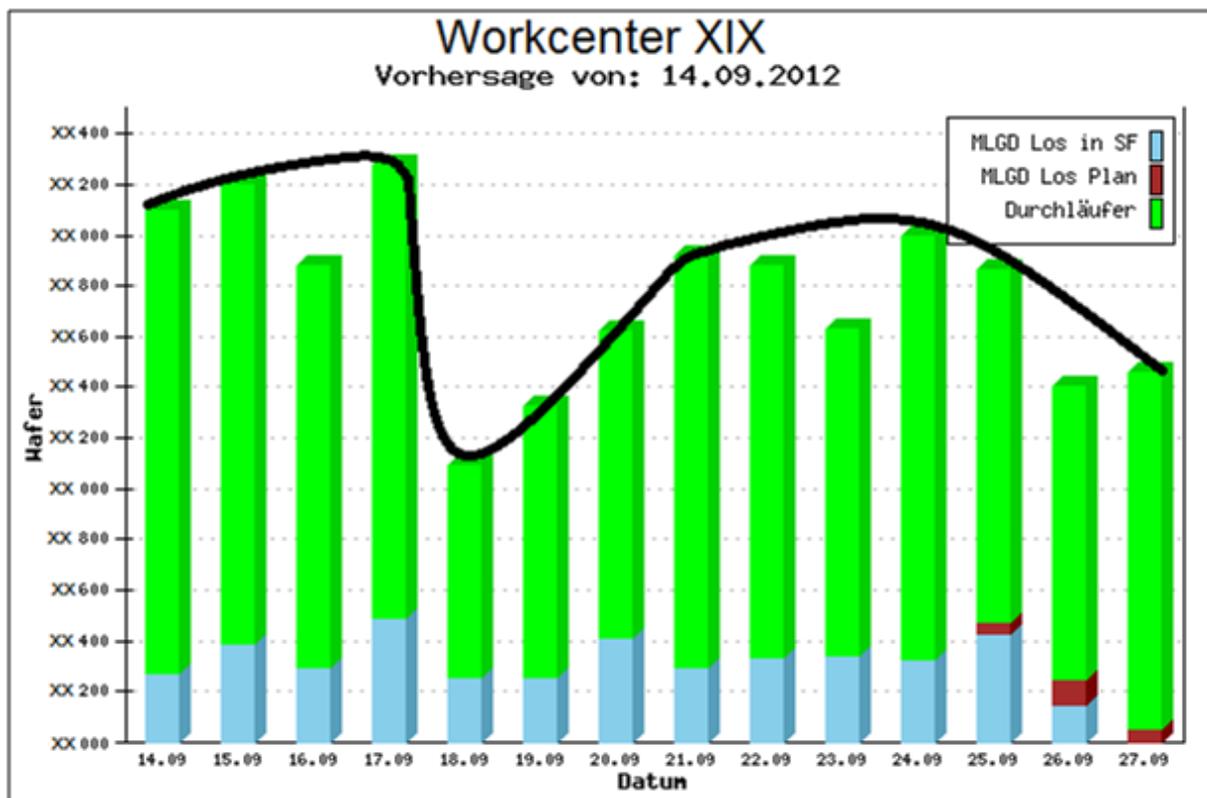
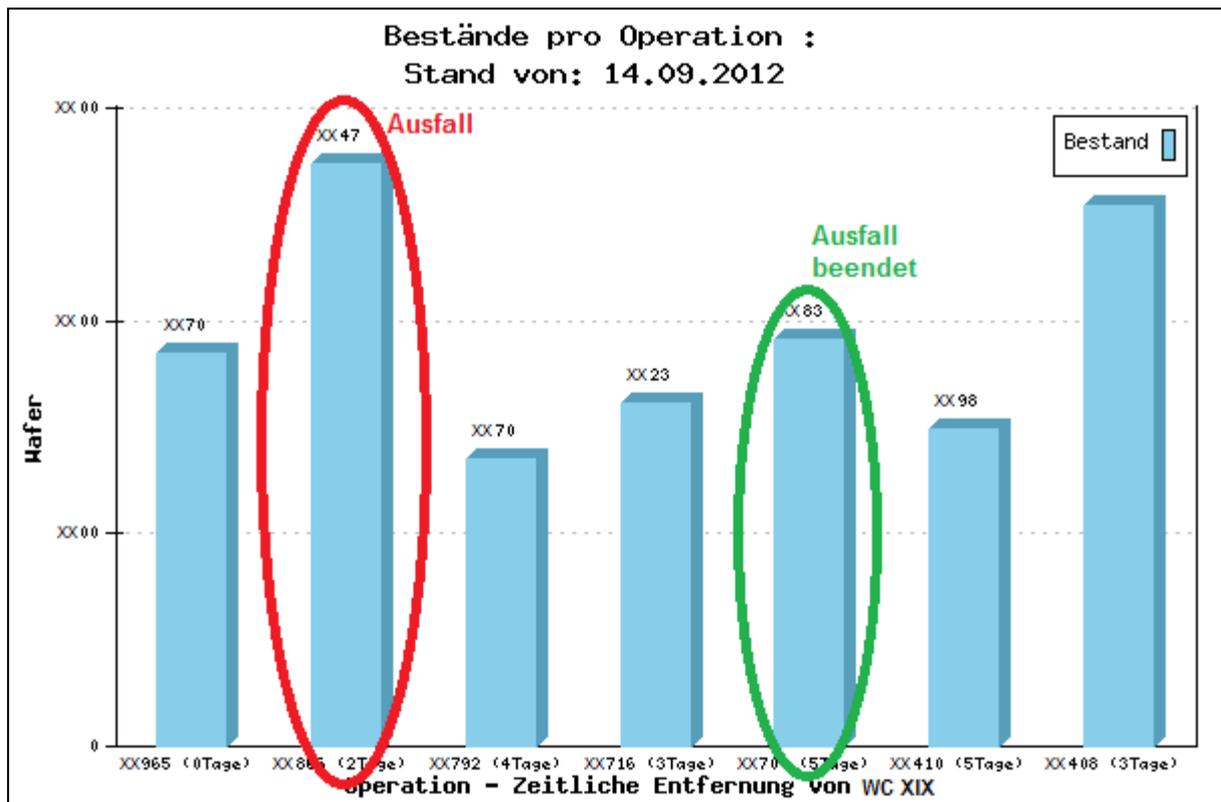


Abbildung 4-37: Report der AFM-Analyse der Abschnitte der Produktionslinie bezüglich der Ankunftsrate an einer Maschinenarbeitsgruppe



**Abbildung 4-38: Report der AFM-Analyse der Bestände pro Operation bei der Prognose der Anlieferung an einer Maschinenarbeitsgruppe**

Die AFM stellt außer dem Gesamtoutput der Fabrik auch detailliertere Prognosen für Liefermengen zur Verfügung. Die ausgeschleuste Materialmenge wird bei Bedarf in Gruppen geteilt. Die kleinste Granularität ist dabei das eigentliche Endprodukt, womit Lieferprognosen für wichtige Sonderprodukte ermöglicht werden.

Die AFM bietet außer der Vorhersagefunktionalität die Analysemöglichkeiten für Abschnitte der Produktionslinie bezüglich Lieferung oder Ankunftsrate an. Die Analyse ist vor allem interessant für die typischen Bottleneck-Anlagengruppen, Lagerstätten oder das Fertigungslinienende. Abbildung 4-37 stellt die 14-tägige Anlieferungsprognose des Materials an einer ausgewählten Maschinenarbeitsgruppe dar. Obwohl die Anlagengruppe im Mittel nur zu 75 % ausgelastet ist, verletzt sie öfter die Durchlaufzeitvorgaben. Ein typischer Grund dafür ist die wellenhafte Anlieferung, die als schwarze Kurve im Diagramm der Abbildung 4-37 abgebildet wurde. Als typisches Vorgehen für die Erkennung derartiger Situationen hat sich die Beobachtung des Linienprofils, also die kontinuierliche Betrachtung der Losposition, etabliert. Die vorgestellte Prognose macht die Kalkulation der Anlieferung einfacher und deutlich genauer. Dadurch lassen sich die Anlieferungswellen früher erkennen. Da die AFM die Kapazität der Anlagen sowie deren momentanes Ausfallverhalten nicht betrachtet, ist eine Aussage über ungewöhnlich hohe Bestände pro Operation zum Prognosestartzeitpunkt not-

wendig. Abbildung 4-38 stellt die Länge der Warteschlangen für einzelne Operationen dar. Der Anwender der Prognose wird mit der Liniensituation zum Initialisierungszeitpunkt konfrontiert und versucht, die einzelnen Abweichungen zu erforschen. Im Falle eines Anlagenausfalls wird in den meisten Fällen der Bestand steigen, solange der Ausfall nicht behoben wurde. Anschließend wird das Produktionsgut wellenhaft weitergeleitet. Da der approximierten Ankunftszeitpunkt sich mittels AFM (nach Bedarf wird die Prognose neu berechnet) voraussagen lässt, ist eine Vorbereitung auf diese Situation deutlich unkomplizierter.

## 5 Implementierung der Vorhersagesoftware

Die in Kapitel 2.5 beschriebenen Anforderungen an das Softwareprojekt wurden zusammen mit dem Projektpartner erarbeitet und bilden eine Grundlage für die Wahl der möglichen Softwarearchitektur, des Designs wie auch der Entwicklungsumgebung. Das Aufbau- und Struktur-Datenmodell wurde durch das Konzept und die Funktionsweise der Vorhersagealgorithmen geprägt. Zur Realisierung des vorgestellten Vorhabens wurde vor allem aus Performancegründen für jegliche mathematisch-statistische Berechnung die Oracle Datenbank selbst verwendet. Bei der Visualisierung von Prognoseergebnissen helfen gängige Webtechniken, die den globalen Datenaustausch und einen einfachen Zugriff auf die Vorhersageinformationen seitens der Anwender ermöglichen.

### 5.1 Die Entwicklungsumgebung

Die Wahl der Entwicklungsumgebung wurde den Projektpartner auf die in der Firma gängige Software eingegrenzt. Der Kern der Software ist die Datenbank. Zur Auswahl standen nur zwei Softwareprodukte, deren Unterstützung seitens der IT-Abteilung des Partnerunternehmens zugesichert wurde. Die Wahl zwischen MS Access und Oracle 10G Datenbanksystem fiel sofort auf das Produkt der Oracle Corporation. Die hauptsächlichen Gründe der Entscheidung waren die bessere Integration über das Firmennetz, deutlich performantere Arbeitsweise bei komplexen Problemen, wie auch ein weitgehendes Konzept der Programmierung mittels der gespeicherten Prozeduren. Als Entwicklungsumgebung der Datenbankprogrammierung dienten Oracles freiverfügbare SQL Developer (implementiert und empfohlen von der Oracle Corporation) wie auch der aus dem Hause Allround Automations stammende PLSQL Developer. Da beide in spezifischen Bereichen bestimmte Vor- und Nachteile aufweisen, wurden sie weitgehend parallel benutzt. Für die Entwicklung der Internetpräsenz mit Visualisierung der Prognoseergebnisse wurde Visual Studio von Microsoft eingesetzt. Die Validierung der Vorhersagen erfolgte zum Teil automatisch mit einem selbst entwickeltem Validierungstool. Bei besonderen, tiefgehenden, statistischen Untersuchungen, bei denen der Testautomatisierungsaufwand gegenüber dem Nutzungsgrad sehr hoch war, wurden auch Tabellenkalkulationsprogramme verwendet.

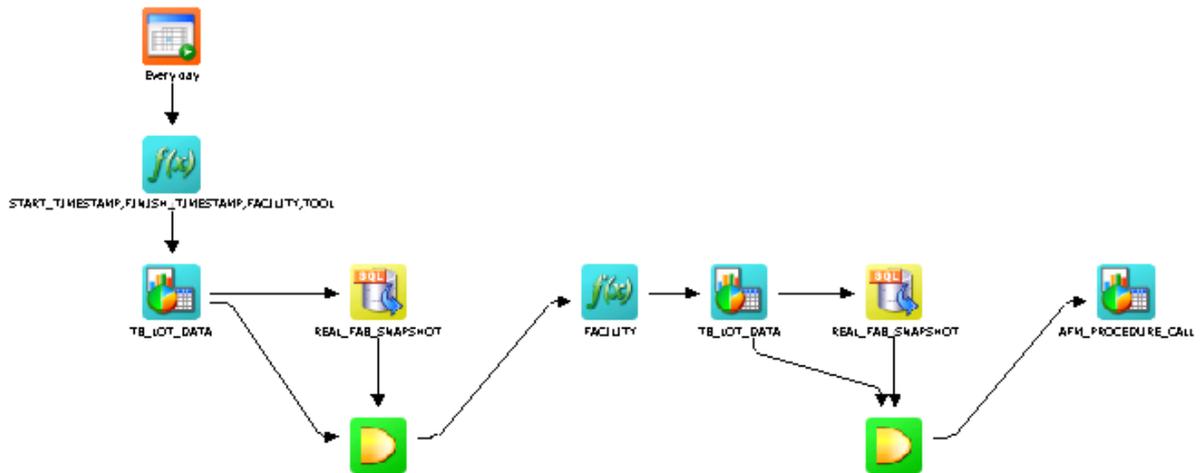


Abbildung 5-1: Täglicher Aufruf einer Abfrage über Bestandsverteilung in der Fabrik

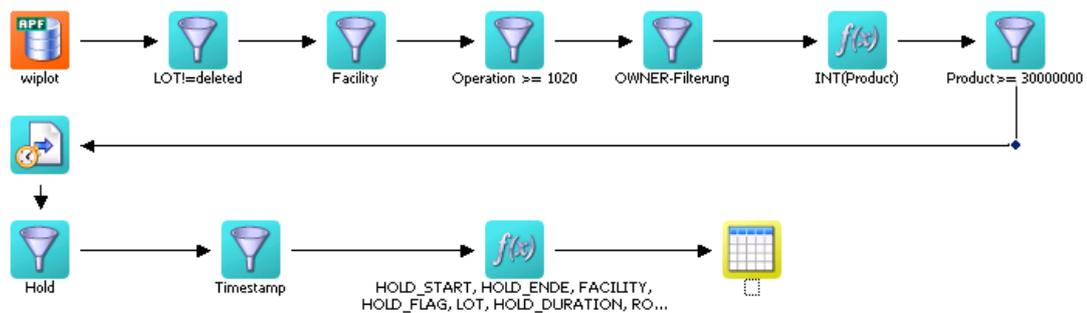


Abbildung 5-2: Eine einfache Abfrage einer Tabelle in RTD-Repositories

Für die Fabriksteuerung verwendet der Projektpartner den APF Real Time Dispatcher (RTD). Die von Applied Materials stammende Software wendet ein anderes Datenspeicherungsprinzip als gängige relationale Datenbanken an. Die Informationen werden in den meisten Fällen in sehr feiner Granularität in zeitlich begrenzt verfügbaren Repositorien abgelegt. Die RTD interne Datenbank lässt sich nicht direkt, sondern nur mithilfe einer speziellen Applied Materials Software, dem APF-Reporter, abfragen. Die Software verwendet eine eigentümliche, graphische Abfragesprache (Abbildung 5-1 und Abbildung 5-2), die im Konzept den Einstieg durch Visualisierung erleichtern soll. Die Einarbeitung in die spezifische Programmierumgebung ist für einen erfahrenen Datenbankprogrammierer zwar nicht schwer, jedoch stellt sich die Erstellung von umfangreichen, optimierten Abfragen sowie deren Wartung und Validierung sehr zeitintensiv dar. Aufgrund der kleinen Nachteile der Programmierumgebung und einer fehlenden Alternative wurde die Umgebung ausschließlich für die Zusammenfassung der feingranularen Daten und deren Portierung auf die von der AFM-Prognosesoftware unterstützte Oracle Datenbank verwendet. Auf diese Weise gelangen periodisch die historischen

Losbewegungsdaten und auch die momentane Bestandsverteilung in der Fabrik in die Datenlandschaft der AFM.

### 5.2 Die Implementierung

Die Implementierung der AFM-Algorithmen, wie in der Abbildung 5-3 dargestellt, beinhaltet vier Hauptkomponenten mit spezifiziertem Aufgabenfeld:

- Datenbeschaffung,
- Datenaufbereitung,
- Vorhersagebildung,
- Reporting.

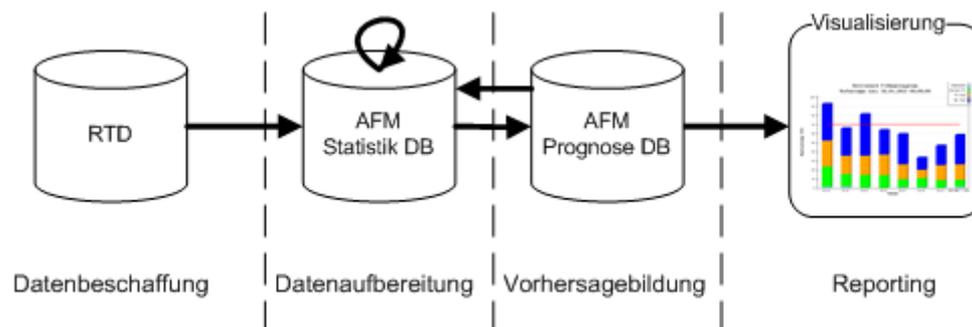


Abbildung 5-3: Komponenten der AFM-Software

Den Kern der Software bilden die Module für die Datenaufbereitung und Vorhersagebildung. Er wurde vollständig in der Oracle Datenbanksprache PL/SQL implementiert und beinhaltet folgende Pakete:

- AFM Conditions – ermöglicht die Konfiguration der Software durch den Anwender,
- AFM Functions – Satz der Funktionen, die in mehreren Paketen aufgerufen werden,
- AFM Initialisation – ein Paket, das für die Statistikinitialisierung zuständig ist,
- AFM Statistik – eine Sammlung von Algorithmen, die für die Prognose notwendige Statistiken erheben,
- AFM Main – ein Paket deren Prozeduren eine Prognose erstellen,
- AFM Validation – organisiert die automatisierte Validierung der Prognosen, es ist zwar kein notwendiges Modul für die Prognosefunktionalität, erleichtert aber die Auswertung enorm,
- AFM Clean – bereinigt alle Datentabellen von den veralteten Daten.

## Implementierung der Vorhersagesoftware

Die Interaktion der Pakete miteinander sowie mit den notwendigen Datenkollektionen wurde in Abbildung 5-4 dargestellt. Die im AFM-Kern entworfenen Tabellen lassen sich anhand ihrer Funktion in folgende Klassen einteilen:

- Konfigurationstabellen – z. B. AFM\_PROPERTIES – Sammlung an Konfigurationsparametern oder Angaben über Klassifikation,
- Zwischenspeichertabellen der Daten aus dem Manufacturing Execution System RTD – z. B. TEMP\_RTD\_EXPORT\_FOR\_TRACE,
- Tabellen für die Bewegungsabläufe – z. B. REAL\_HISTORY\_TRACE\_DATA,
- Statistiktabellen – z. B. STATISTIC\_HOLD – beinhalten die erhobenen statistischen Informationen,
- Einschleusung in die Fabrik – z. B. REAL\_MLGD\_RELEASE,
- Bestandsverteilungstabelle – REAL\_FAB\_SNAPSHOT,
- Tabellen mit den Vorhersageresultaten – z. B. AFM\_RESULTS\_SINGLE\_LOT,
- Validierungstabellen – z. B. AFM\_RESULT\_COMPARE\_DAILY,

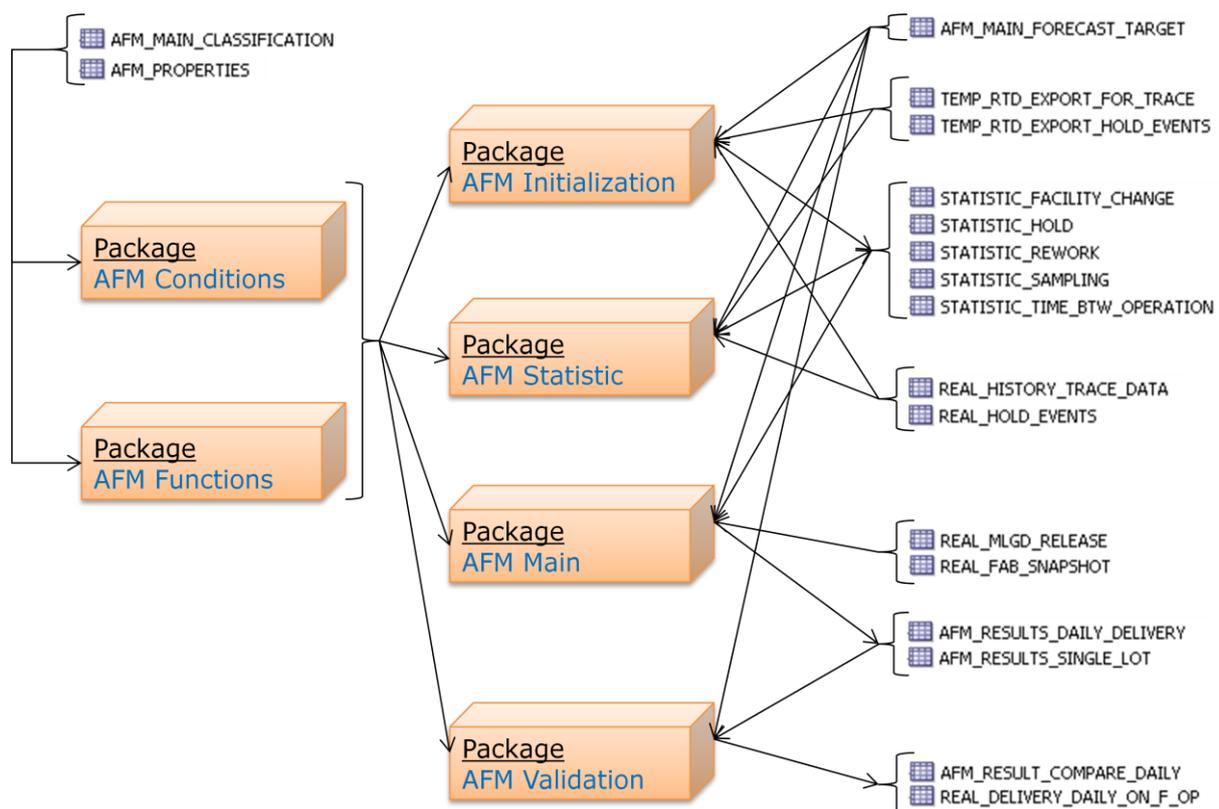


Abbildung 5-4: Der Aufbau des Kerns der AFM-Software in Paketen sowie notwendige Tabellen

Das Konzept des modularen Aufbaus der Software ermöglicht die Portierung des Softwarekerns auf andere, ähnliche Probleme (auch partiell) in einer abweichenden Softwareumgebung. Bei einem derartigen Softwareaufbau steht die klare Definition der Schnittstellen zwi-

schen einzelnen Komponenten im Fokus [Helmke01]. Durch das angewendete Programmierparadigma wurde die Verwendung des bereits implementierten Softwarekerns an anderen Standorten des Projektpartners gewährleistet.

### 5.2.1 Datenbeschaffung

Der Bedarf an Basisdaten über das zu beschreibende System (die Fabrik) wurde im Konzept von AFM auf das notwendige Minimum reduziert. Die Absicht der Reduktion von Datenquellen lässt sich mittels der in Kapitel 4.1 beschriebenen Probleme bei der Informationsgewinnung begründen. In Abbildung 5-5 wurden die erforderlichen Reports der Prognosemethode dargestellt. Der Hauptzweig der Datenbeschaffung ist das Auslesen der historischen Bewegung der Lose in der Fabrik aus dem Manufacturing Execution System. Es erfolgt automatisch alle zwei Wochen, kann aber auch jederzeit manuell initialisiert werden. Die Abfrage wurde entsprechend dem Real Time Dispatcher (MES des Projektpartners vgl. Kapitel 5.1) konzipiert. Der Kern der Abfrage beinhaltet zu jedem Los eine Abfolge der Zeitpunkte von Zustandsänderungen wie:

- Freigabe für die Folgeoperation / Betreten der Warteschlange,
- Freigabe für die Bearbeitung (Move In),
- Bearbeitung Start und Ende,
- Hold Start und Ende.

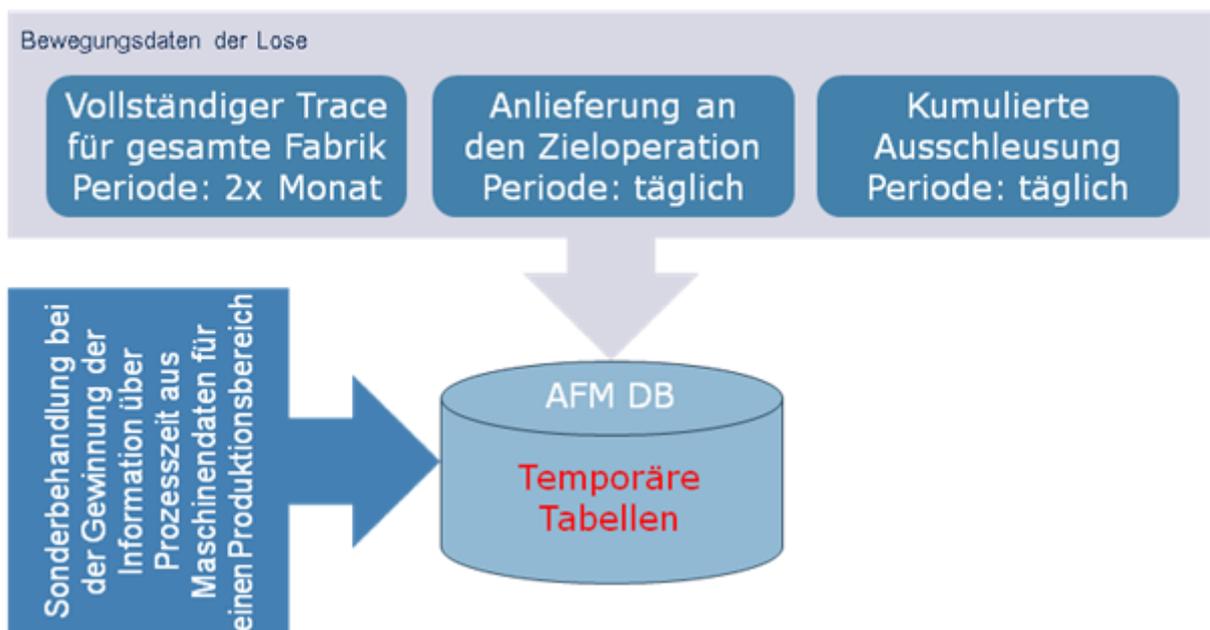
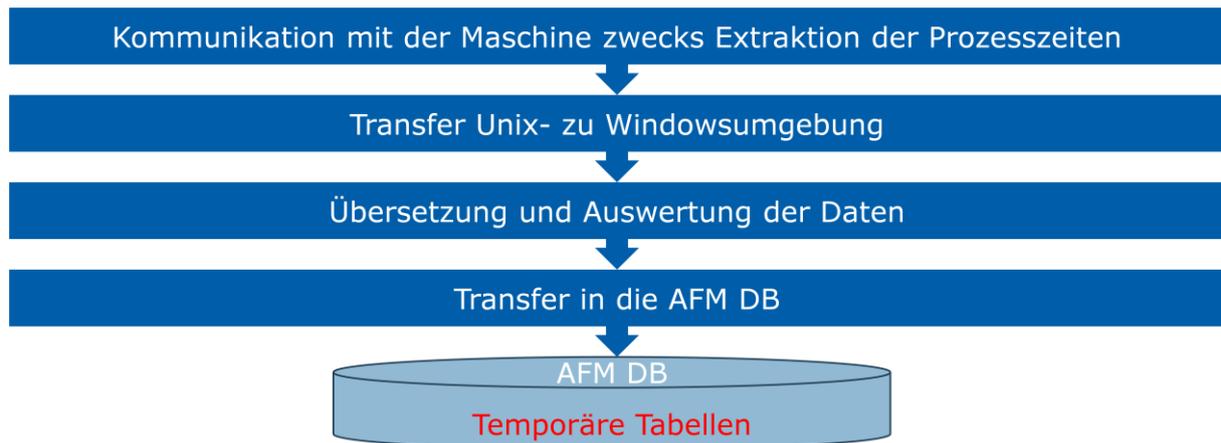


Abbildung 5-5: Die Datenquellen der AFM

Die AFM benötigt für die Statistikerzeugung auch weitere Angaben wie: Operationsnummer, Priorität, Verspätung bzgl. Fertigstellungsdatum, Route, Produkt, Anzahl der Wafer und weitere losspezifische Attribute, die im MES verfügbar sind. Da sich die zu extrahierende Datenmenge der Losbewegung über 2 Wochen als enorm herausgestellt hat, liegt die Vermutung nahe, dass die Periode falsch ausgewählt wurde. Folglich wäre also eine tägliche Abfrage günstiger. Die Interaktion mit dem MES der Fabrik ist zwar nur lesend, benötigt aber deren Ressourcen. Um den Produktionsablauf nicht zu gefährden, wurde ein Abfragefenster periodisch im 14 Tage Takt fest von den MES-Experten definiert. Weiterhin wurde die Segmentierung der Abfrage notwendig, sodass auch zwischen den einzelnen Abfragesegmenten die höher priorisierten produktionsrelevanten Prozesse ausgeführt werden können, ohne die AFM zu unterbrechen. Da beim Real Time Dispatcher die Operationen der Datumsarithmetik bemerkenswert schnell abgearbeitet werden, erfolgt die Segmentierung über die Zeit. Jeder vollständige Datenabschnitt wird anschließend in der AFM-Datenbank gespeichert. Für alle Etappen der Produktion (vgl. Abbildung 2-1) werden für die Datenextraktion die gleichen Algorithmen angewendet. Interessanterweise werden aber die Abfragen der Informationen, die relevant für die Validierung der Prognose sind, täglich abgefragt. Es ist damit zu begründen, dass durch ein eingebautes Filtersystem die Datenmenge so deutlich begrenzt wurde, dass die Ausführung keine Gefahr für das MES darstellt. Um die Sicherheit und Stabilität der Informationsqualität in der AFM-Datenbank zu gewährleisten, wurde an den Schnittstellen zwischen AFM und RTD eine Notifikation implementiert. Sobald eine Ausnahme an der Stelle auftritt, wird der Administrator darüber ausführlich via E-Mail benachrichtigt.

Die Realisierung des Anwendungsfalles als Bottleneck-Frühwarnsystem erfordert detaillierte und genaue Angaben über die Bearbeitungsdauer an der Zieloperation. In den meisten Fällen lassen sich die Informationen aus den Bewegungsdaten der Lose ermitteln. In Ausnahmefällen werden jedoch die Informationen vom Equipment nicht an das MES übermittelt. Ein Versuch, die Prozesszeiten aus den Move In und Move Out Zeitstempeln zu berechnen, scheiterte aufgrund der menschlichen Schwäche, nicht alles exakt zu protokollieren. Vor allem bei Überbestand wurden mehrere Lose vom Operator gleichzeitig oder in sehr kurzen Zeitabständen für die Bearbeitung freigegeben (Move In), obwohl Sie oft weiterhin in der Warteschlange warten mussten. In Kooperation mit dem Projektpartner wurde die Lösung gefunden, die Prozesszeiten direkt aus dem Maschinenlog zu extrahieren. Abbildung 5-6: Sonderbehandlung bei der Gewinnung der Information über die Prozesszeit aus Maschinendaten für ein Produktionsbereich stellt den Datengewinnungsprozess eines Parametermessgeräts beispielhaft dar. Die Anlage selbst bearbeitet die Informationen in einer Unix-Umgebung. Deshalb ist



**Abbildung 5-6: Sonderbehandlung bei der Gewinnung der Information über die Prozesszeit aus Maschinendaten für ein Produktionsbereich**

es vorerst notwendig, dementsprechend mit der Maschine zu kommunizieren, und anschließend die gewonnenen Informationen in die Windowsumgebung zu transferieren. Die Auswertung der Logdateien und der Informationstransfer an die AFM-Datenbank wurde mithilfe von einfachen Java-Programmen durchgeführt. Die Aufbereitung des Datenmaterials von AFM in eine brauchbare Statistik erfolgt wie in allen anderen Fällen direkt in der Datenbank. Der Sonderfall der Prozessdatengewinnung wurde an der Stelle provokativ erwähnt, um die Grenzen der Minimierung der Datenquellen zu veranschaulichen. Das AFM-Konzept der Minimierung von notwendigen Datenquellen wird also durch die Problematik der Basisdaten behindert. Die Prognose wird zwar ohne Probleme berechnet, die Realisierung von möglichen Anwendungsfällen wird jedoch dadurch eingeschränkt. Der Kraftaufwand für das Überwinden des Hindernisses brachte aber im Endergebnis einen positiven Aspekt für die Weiterentwicklung der AFM, nämlich die weitgehende Integration von Maschinendaten in das MES und das betriebliche Datawarehouse.

### 5.2.2 Datenaufbereitung

Die Aufbereitung der Daten mittels AFM wird durch einen Prozess beschrieben, der ausgehend von den aus dem MES importierten Daten eine Reihe von statistischen Informationen erstellt. Die Statistiken beschreiben das momentane Fabrikverhalten (vgl. Kapitel 4.5) in Bezug auf: Wechsel des Produktionsbereiches, Haltezustand, Sampling, Nacharbeit, sowie Durchlaufzeiten. Abbildung 5-7 stellt schematisch den AFM-Datenaufbereitungsprozess dar, der durch die periodische Ausführung die Aktualität der Statistiken gewährleistet. Den Ausgangspunkt bilden die aus dem MES extrahierten Bewegungsdaten (I) und detaillierte Informationen über den Hold-Zustand (II). Die Initialisierungsroutine gruppiert die Daten nach der Zugehörigkeit zu einer Route. Die Notwendigkeit einer derartigen Gruppierung resultiert aus-

schließlich aus den Berechnungsproblemen bedingt durch die sehr große Datenmenge. In jeder der Routengruppen wird eine aus dem AFM-Statistik-Package stammende Konsolidierungsroutine (III) durchgeführt. Das Konsolidierungsverfahren beinhalten folgende Abarbeitungsschritte:

- Plausibilitätstest der Eingangsdaten – war der Datentransfer erfolgreich?
- Fehlerkorrektur der fehlenden bzw. inkonsistenten Werte,
- Filterung des Datenmaterials – werden alle Daten für die Prognose gebraucht?
- Bereinigung der regulären Durchlaufzeit pro Operation durch den Abzug der Dauer des Hold-Zustands und ggf. der Lagerungszeit,
- Zusammenfassung der Informationen für einen Produktionsschritt.

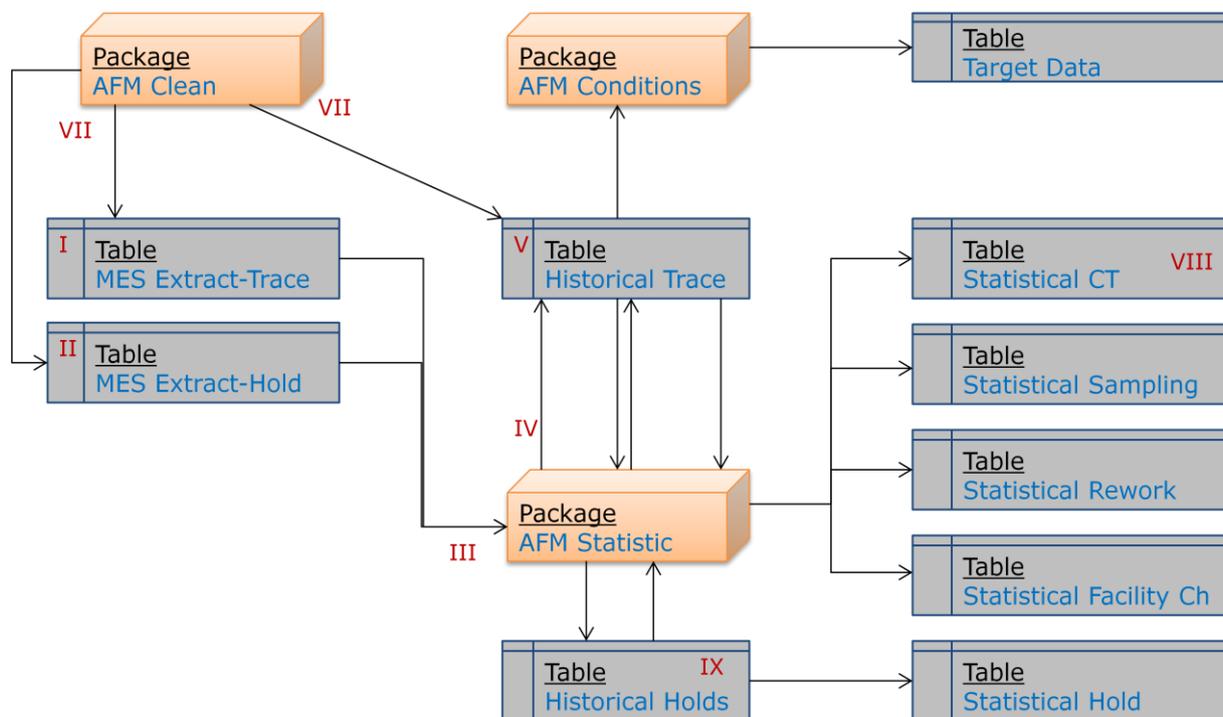
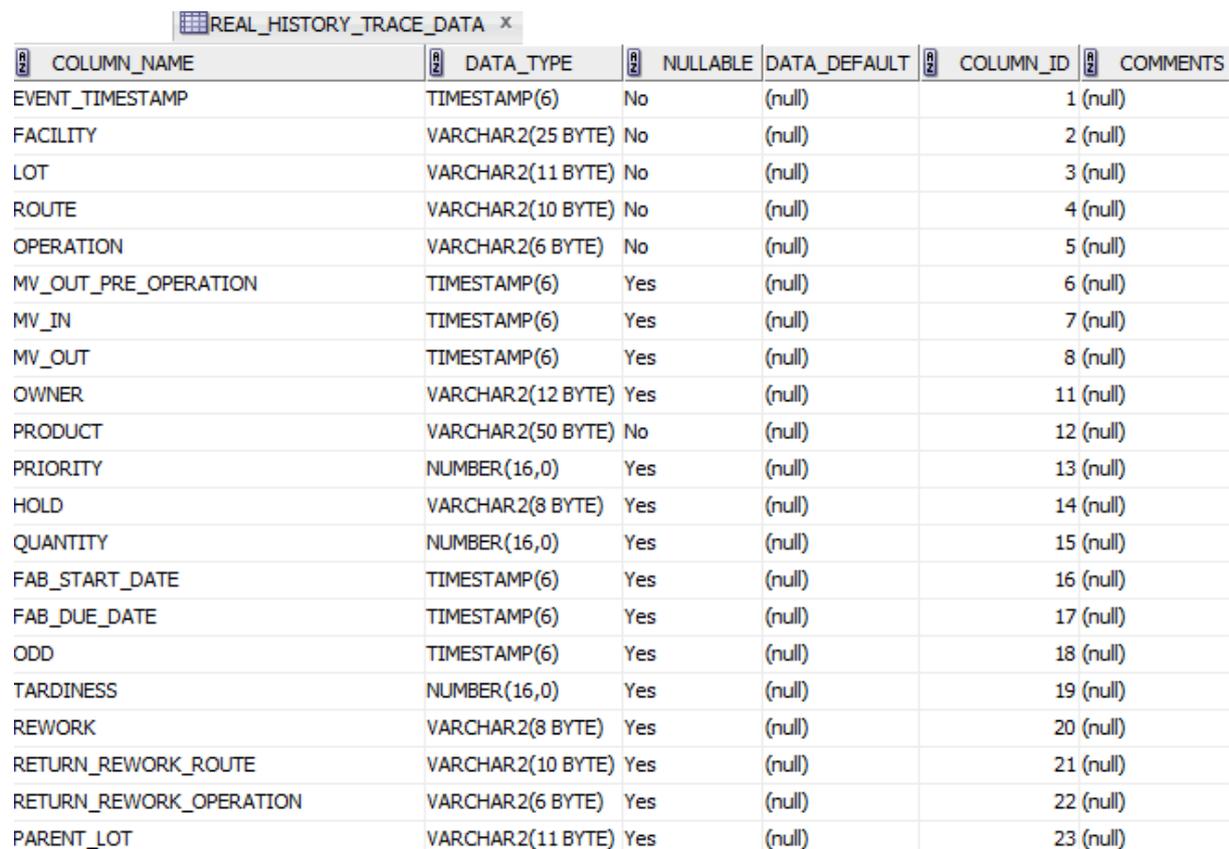


Abbildung 5-7: Die schematische Darstellung des AFM-Datenaufbereitungsprozesses

Die Menge der Daten für eine Operation nach der AFM-Spezifikation ist bedingt durch das Datenverwaltungsprinzip des Real Time Dispatcher. Das MES speichert separat jede einzelne Zustandsänderung für ein Los an der Operation. Die operationsfein aufbereiteten Informationen werden zur AFM-Tabelle mit historischen Bewegungsabläufen (V) hinzugefügt (IV). Dass diese Tabelle alle Angaben beinhaltet, die für die korrekte Erstellung einer Prognose mittels AFM notwendig sind, wurde in Abbildung 5-8 dargestellt. Die Tabelle bildet somit eine Schnittstelle für den Einsatz der AFM in anderen, ähnlichen Fertigungsstätten. Nach der Konsolidierung werden die neu eingefügten Daten in Bezug auf die bereits erhobenen Infor-

## Implementierung der Vorhersagesoftware

mationen wie: Anzahl Wafer, Zeitpunkte vom Verlassen oder Betreten der Operation, Verspätung und ähnliches verifiziert (VI). Abschließend werden alle veralteten oder nicht mehr gebrauchten Daten von den Bereinigungsroutinen (VII) gelöscht. Die eigentlichen AFM-Prozeduren, welche für die Erstellung der Statistiken zuständig sind, interagieren ausschließlich mit den konsolidierten und bereinigten Bewegungsdaten. Die Technik für die Erzeugung der Statistik wird anhand der Durchlaufzeitstatistik (VII) beschrieben. Die Routine ist modular aufgebaut, was das Ein- und Ausschalten deren einzelnen Bausteinen durch den Anwender ermöglicht. Dies erfolgt durch die Änderungen in der Konfigurationstabelle.



COLUMN_NAME	DATA_TYPE	NULLABLE	DATA_DEFAULT	COLUMN_ID	COMMENTS
EVENT_TIMESTAMP	TIMESTAMP(6)	No	(null)	1 (null)	
FACILITY	VARCHAR2(25 BYTE)	No	(null)	2 (null)	
LOT	VARCHAR2(11 BYTE)	No	(null)	3 (null)	
ROUTE	VARCHAR2(10 BYTE)	No	(null)	4 (null)	
OPERATION	VARCHAR2(6 BYTE)	No	(null)	5 (null)	
MV_OUT_PRE_OPERATION	TIMESTAMP(6)	Yes	(null)	6 (null)	
MV_IN	TIMESTAMP(6)	Yes	(null)	7 (null)	
MV_OUT	TIMESTAMP(6)	Yes	(null)	8 (null)	
OWNER	VARCHAR2(12 BYTE)	Yes	(null)	11 (null)	
PRODUCT	VARCHAR2(50 BYTE)	No	(null)	12 (null)	
PRIORITY	NUMBER(16,0)	Yes	(null)	13 (null)	
HOLD	VARCHAR2(8 BYTE)	Yes	(null)	14 (null)	
QUANTITY	NUMBER(16,0)	Yes	(null)	15 (null)	
FAB_START_DATE	TIMESTAMP(6)	Yes	(null)	16 (null)	
FAB_DUE_DATE	TIMESTAMP(6)	Yes	(null)	17 (null)	
ODD	TIMESTAMP(6)	Yes	(null)	18 (null)	
TARDINESS	NUMBER(16,0)	Yes	(null)	19 (null)	
REWORK	VARCHAR2(8 BYTE)	Yes	(null)	20 (null)	
RETURN_REWORK_ROUTE	VARCHAR2(10 BYTE)	Yes	(null)	21 (null)	
RETURN_REWORK_OPERATION	VARCHAR2(6 BYTE)	Yes	(null)	22 (null)	
PARENT_LOT	VARCHAR2(11 BYTE)	Yes	(null)	23 (null)	

Abbildung 5-8: Das Schema der AFM-Tabelle mit historischen Bewegungsabläufen

Für die Erstellung der Durchlaufzeitstatistik werden folgende Bausteine angewendet:

- Erstellung einer temporären Kopie der Bewegungsdatentabelle. Um die statistisch relevanten Aussagen bei möglichst geringer Datenmenge zu erreichen, ist es möglich die Länge der Periode der Bewegungsdaten über die Konfiguration einzustellen.
- Test auf Validität der Eingangsdaten – hier werden die typischen Ausreißer markiert. Die genaue Spezifikation des Tests wurde im Kapitel 4.5.3 beschrieben.

- Bereinigung der Eingangsdatenquelle – Ausreißer werden gelöscht. An der Stelle werden auch die Daten über bestimmte Routen entfernt, die einem Los keine Möglichkeit bieten, die als Prognoseziele bestimmte Operation zu erreichen.
- Update der Statistiktabelle,
- Bereinigung der Statistik Tabelle – an der Stelle werden die veralteten Statistikwerte gelöscht. Jeder Datensatz beinhaltet einen Zeitstempel des letzten Updates. Dadurch lässt sich das Alter der einzelnen Angaben herausfinden.
- Entfernen der temporären Tabellen.

Auf ähnliche Weise sind ebenso die Bausteine für die Erzeugung von anderen Statistiken aufgebaut. Die Erhebung der Hold-Statistik basiert jedoch zusätzlich auf einer weiteren Tabelle (IX), die ausschließlich die Hold-Dauer pro Los und Operation beschreibt. Die Tabelle bildet neben der Bewegungsdatentabelle (V) die zweite Schnittstelle der AFM.

Die AFM-Implementierung beinhaltet außer dem Standardstatistikpaket auch einen Baustein für die Initialisierung der Statistiken. Der Unterschied zwischen den Paketen liegt in der Datenquelle für die Referenz bei der Ausreißersuche sowie beim Umfang der Verifikation der Eingangsdaten. Da bei der Initialisierung keine statistischen Daten vorliegen, besteht die Möglichkeit, eine Referenzdatenquelle anzubinden – wie zum Beispiel die Planungsdaten o. ä.

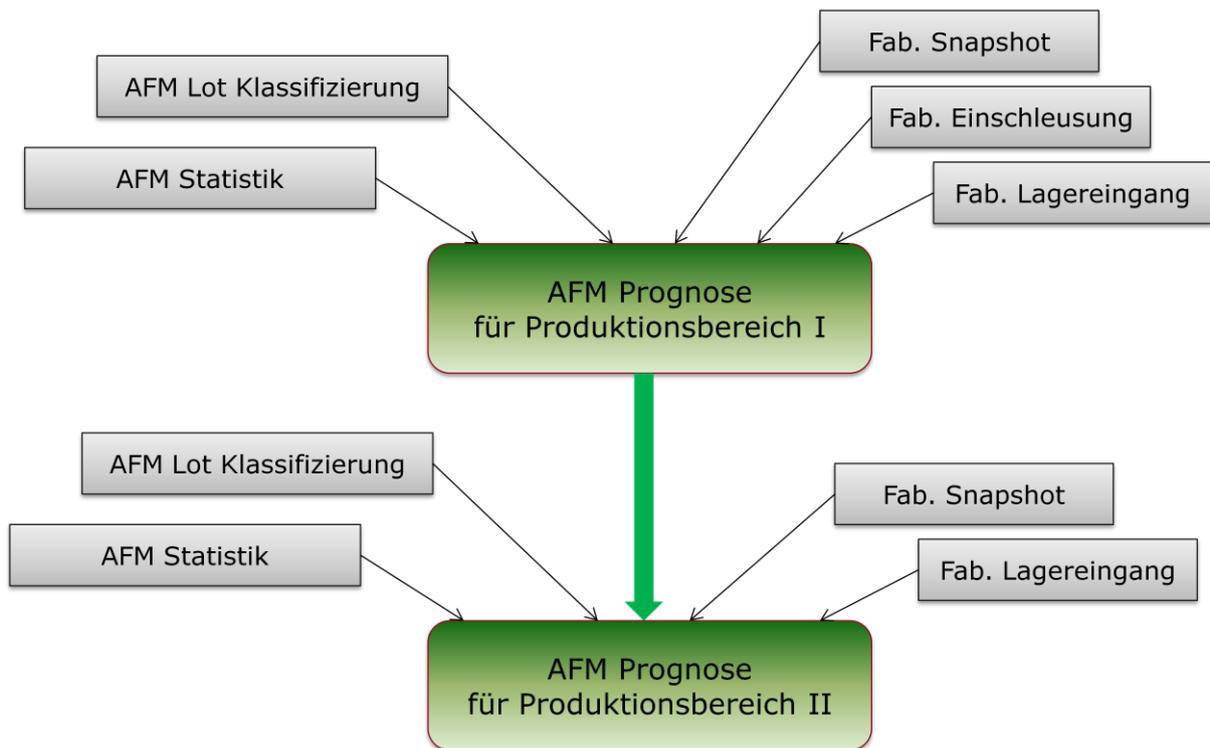
### 5.2.3 Vorhersagebildung

Abbildung 5-9 stellt schematisch den Datenfluss der AFM-Software für den Erstellungsprozess einer Vorhersage dar. Der Ausgangspunkt für die Berechnung der Prognose ist die Bestandsverteilung in der Fabrik. Dabei ist die Position einzelner Lose auf der Route von Interesse, sowie deren momentaner Zustand, wie: z. B., ob sich ein Los in der Warteschlange befindet oder gerade bearbeitet wird. Die Informationen bezieht die AFM von der gleichen Datenquelle wie die historischen Bewegungsabläufe, nämlich vom MES-Trace. Die dazu verwendete Abfrage gleicht der für die Statistikerzeugung – mit der Ausnahme, dass keine Historie abgefragt wird. Der Vorteil der Methodik liegt bei der Wiederverwendbarkeit und sie befolgt somit das AFM-Konzept der Minimierung der Datenquellen. Da der Prognosehorizont mehrere Tage umfasst, werden auch die Pläne für die Einschleusung des Materials in die Fabrik benötigt. Die Pläne sind im MES nicht vorhanden, was die Anbindung von zusätzlichen Informationsquellen erforderlich macht. Die Wissensbasis für die Prognose bildet die statistische Zusammenfassung der Parameter (Kapitel 4.5 und 5.2.2), welche die Fertigung beschreiben. Nach der Vervollständigung der notwendigen Informationen startet der Vorhersagebe-

## Implementierung der Vorhersagesoftware

rechnungsprozess des Softwarekernes. Der Prozess beinhaltet hauptsächlich folgende Unterprozeduren:

- Berechnung von den prognostizierten Ankunft an der Zieloperation
  - Bestimmung der Einschleusung (Eingang + Lager) für die gültige Periode
  - Klassifizierung der Lose
  - Zusammenfassung der anfallenden Zeitkomponenten (vgl. Kapitel 4.6)
- Markierung der Lose, die vom Sampling betroffen sind
- Markierung der Lose, die nach der Anwendung von AFM-Sampling-Verfahren, als gültig (kommen bei der Zieloperation an, siehe Kapitel 4.5.4) bezeichnet werden
- Zusammenfassung der gültigen Lose zu Anlieferungsmengen
- Ergänzung der einzelnen Anlieferungsströme um die Prozesszeit an der Zieloperation und um andere für das Reporting notwendige Informationen.



**Abbildung 5-9: Datenfluss bei der Erstellung einer Prognose mittels AFM**

Der modulare Aufbau der AFM erlaubt die Verkettung von unterschiedlichen Prozessen der Voraussageerzeugung. Das Konzept wird derzeit für den Wechsel des Produktionsbereichs angewendet. Die Antwort des Systems für die reguläre Fertigung dient als Materialeingangsstrom für den Testbereich. Die Ausweitung des Konzeptes mit mehreren einzelnen Modulen zwischen bestimmten Stellen in der Fertigung für deren Kapazitätsbegrenzung, wie in Kapitel 4.6.2 beschrieben, wird derzeit erprobt.

Der AFM-Kern für die Erstellung der Voraussage wurde ausschließlich in PL/SQL implementiert. Der Prozess wurde möglichst effizient entworfen, woraus eine weitestgehende Minimierung der Berechnungszeit resultiert. Die reine Kalkulationszeit einer Prognose beträgt ungefähr 10 Sekunden. Die Effizienz der Datenbanksprache beim Umgang mit größeren Datenmengen ergibt sich aus aufeinander abgestimmter Vergabe der Indices und deren effektiver Nutzung durch die gespeicherten Prozeduren.

### **5.2.4 Visualisierung der Resultate**

Die umfangreiche Auswertung der Ergebnisse einer Prognose ist ein wichtiger Punkt jedes Vorhersageprojektes. Da das Berichtswesen die Schnittstelle zwischen der Prognosesoftware und dem Anwender bildet, sollte es möglichst anwenderfreundlich gestaltet werden [Eichholz08], [Neuhe00]. Die Entwicklung der AFM orientierte sich an der Definition von Hans Jung hinsichtlich der Anforderungen an ein effizientes und modernes Reporting [Jung03]:

- Empfängerorientierung – Abstimmung der Berichte an den Anwenderbedürfnissen, also der erforderliche Verdichtungsgrad der Daten sowie zyklische Anfragen bezüglich des Informationsbedarfs,
- Wirtschaftlichkeit – Vermeidung der Darstellung von irrelevanten Informationen,
- Verständlichkeit und Übersichtigkeit – klare Struktur und Gliederung der Reports,
- Verlässlichkeit – Manipulation der Daten ausgeschlossen,
- Aktualität – schnelle und termingerechte Reporterstellung,
- Interpretation – die Reports müssen klar interpretierbar sein, z.B. durch Vergleiche mit bereits bekannten Kenngrößen aus der Vergangenheit,
- Reaktionsorientierung,
- Präsentation – das Zusammenspiel der Diagramme und Zahlenwerte in Tabellenform soll attraktiv präsentiert werden.

Um den Anforderungen gerecht zu werden, wurde das AFM-Reporting modular aufgebaut und stellt außer der Standardfunktionalität ein Framework zur Verfügung, das es erlaubt, einfach neue Reports zu entwickeln. Die Softwarelösung kombiniert die Vorteile Oracle Datenbank bezüglich der Performance beim Umgang mit größeren Datenmengen mit der Mächtigkeit der Visualisierung durch ein Web-Interface. Abbildung 5-10 stellt die Hauptbestandteile des AFM-Reporting und deren Interaktion miteinander dar. Das Softwaremodul für die Realisierung der Berichte ist in zwei Hauptkomponenten aufgeteilt:

## Implementierung der Vorhersagesoftware

- Datenbasis mit den Resultaten,
- Web Interface.

Nach der Erstellung einer Prognose mittels AFM werden die Resultate in einer Datenbanktabelle zwischengespeichert. Die Auswertung der Informationen erfolgt vorerst auf der Datenbankebene mittels im Vorfeld erstellter Datenbankprozeduren. Das Ziel der Prozeduren ist die Modifikation der Informationsform, um die Empfangsorientierung, Wirtschaftlichkeit und Übersichtlichkeit zu gewährleisten. Um eine bessere Akzeptanz der Prognose zu erzielen, werden zum Teil die Prognosen um die Vergangenheitsinformationen bzw. Ist- oder Plandaten ergänzt. Nach unseren Beobachtungen trauen die Anwender der Prognose eher, wenn Sie eine Möglichkeit bekommen, die gelieferten Resultate selbst zu validieren. Vor allem die Darstellung der näheren Vergangenheit in der Kombination mit der Prognose in einem Diagramm führte zu größer Akzeptanz. Die aufbereiteten Reportdaten werden in separaten Tabellen abgelegt. Der Vorteil dieser Lösung liegt vor allem in der Performance bei der Abfrage

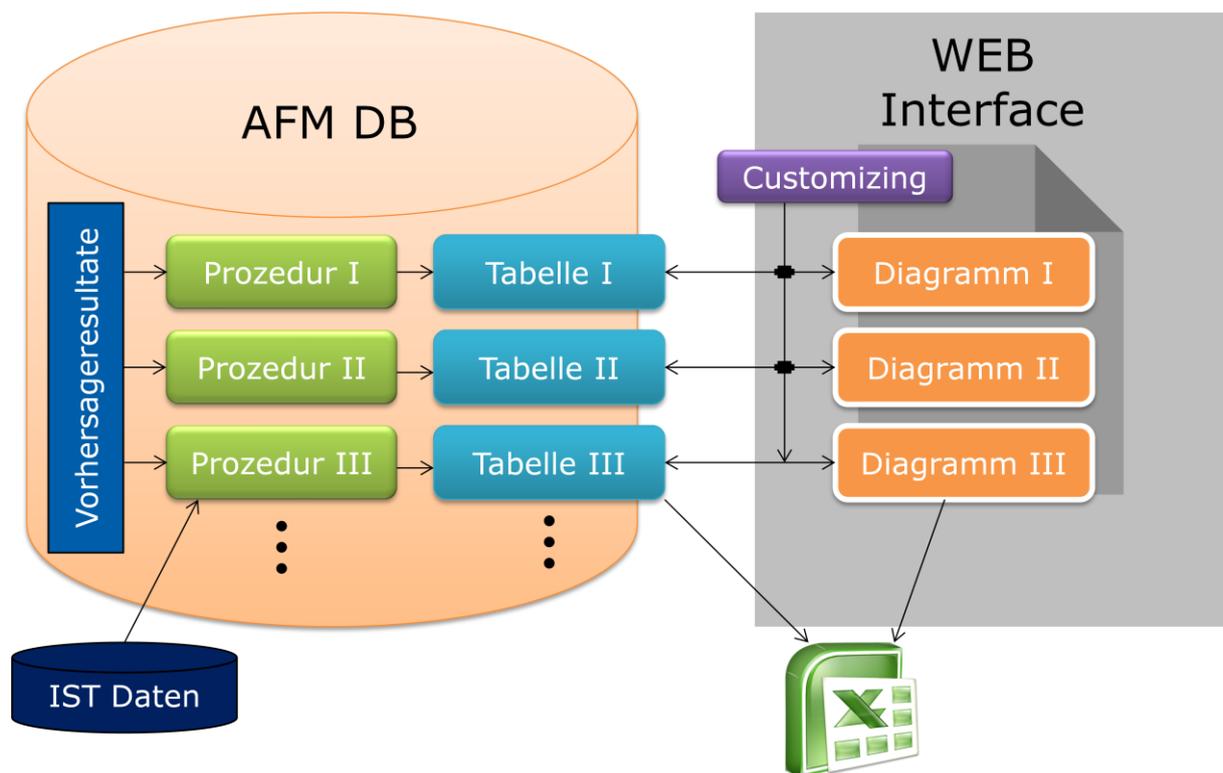


Abbildung 5-10: Der Aufbau des AFM-Reporting Moduls

der Informationen. Die Visualisierung der Prognose wurde mittels eines Web-Interface realisiert. Die einzelnen Diagramme holten die notwendigen Berichtsinformationen direkt aus den dedizierten Datenbanktabellen. Diese können vom Anwender nach Bedarf angepasst werden. Weiterhin ermöglicht die AFM die Abfrage der Daten in Zahlenform mittels des Downloads von Dateien in CSV-Format. Da die einzelnen Reports hauptsächlich auf den Anwendungsfäl-

len basieren, wurden die ausgewählten Berichtsdiagramme bereits in Kapitel 4.7 dargestellt und beschrieben.

### 5.3 Validierungsmodul

Das Validierungsmodul stellt keinen notwendigen Baustein für die Erstellung einer Vorhersage dar. Er ermöglicht lediglich die Bewertung der Ergebnisse einer Prognose. Die Vorteile der Implementierung liegen vor allem in:

- der Möglichkeit der Ausführung mehrerer Voraussagen als Gruppe – die einzelnen Szenarios werden in einer Testumgebung abgelegt,
- der Auswertung von mehreren Prognosen einzeln oder auch zusammengefasst,
- der standardisierten Visualisierung der Ergebnisse,
- der Unterstützung bei der Suche von Modellierungsfehlern,
- der Verkürzung der Bewertungszeit durch die Automatisierung.

Das Validierungsmodul erlaubt das Zwischenspeichern von jedem Vorhersageszenario zusammen mit den notwendigen Statistikdaten. Dabei werden die Statistikdaten mit einem Identifikationsschlüssel versehen, was zusätzlich die Nutzung einer Statistik für mehrere Prognosen erlaubt. Der Vergleich der Statistikdaten und die Vergabe der ID erfolgen automatisiert. Weiterhin wurde im Validierungsmodul eine Testumgebung geschaffen, in dem die AFM-Software installiert wurde. Die Validierung orientierte sich in erster Linie an der Bewertung der Prognosequalität beim Komponentenwechsel der AFM. Weiterhin wurden auch die Einstellungsparameter der Prognosesoftware (z. B.: Klassifikationsparameter, Statistikeinstellungen) und deren Einfluss auf die Vorhersagegüte überprüft. Die Durchführung sowie die Resultate ausgewählter Experimente werden in Kapitel 6 beschrieben.

## 6 Validierung

Gelaschwili [Gelasch07] konstatiert: „*Die statistische Prognose meint nicht die Vorausbestimmung aller Änderungen des gesamten Wandels irgendeines Ereignisses oder Prozesses (was meistens unmöglich ist), sondern die Feststellung der Grundtendenzen (der Grundtrends) der künftigen Änderungen. Dies ist besonders dann zu beachten, wenn man die Prognose auf Basis von Zeitreihen durchführt.*“ Die Validierung über den Vergleich der erzeugten Grundtendenzen mit realen Ereignissen ermöglicht also die Bewertung der Qualität einer Prognose.

### 6.1 Analyse der Zeitabweichung

Unter der Analyse der Zeitabweichung verbirgt sich die Untersuchung der zeitlichen Divergenz von den realen und den prognostizierten Ankunftszeitpunkten für ein Los an bestimmten Zieloperationen. Verglichen wurden also die Prognosen für die einzelne Lose mit der Realität. Für die Quantifizierung der Prognosequalität wurde, wie in Kapitel 3.1.2 genauer beschrieben, die mittlere absolute Abweichung MAD verwendet (vgl. [Buttler02], [Krengel05]). Die Basis für die Untersuchungen bildeten über 200 unabhängige Prognosen, die über ein gesamtes Produktionsjahr verteilt waren. Die Fabrik arbeitete in dem Zeithorizont mit einer einigermaßen hohen Auslastung. Der Untersuchungszeitraum lässt sich in Abschnitte teilen, die durch deutlich bemerkbare Anlagenausfälle sowie Änderungen im Produktionsmix gekennzeichnet waren. Die einzelnen Produktionsperioden sind deutlich voneinander divergent, was relativ hochwertige und statistisch relevante Ergebnisse garantiert.

Der MAD für die AFM mit einem Zeithorizont von 7 Tagen betrug ungefähr 19 Stunden. Die Untersuchung umfasste alle Zieloperationen und alle Produkte. Abbildung 6-1 stellt den Verlauf von dem MAD über den zu prognostizierten Zeitabschnitt dar. Der Verlauf der Kurve, die den Prognosefehler abbildet, ist nahezu linear. Der tägliche Anstieg des Fehlers beträgt näherungsweise 5 Stunden, wobei sich am ersten Tag der Prognose eine mittlere Abweichung von ungefähr 4 Stunden feststellen lässt. Sie steigt auf ca. 1,5 Tage beim siebten und letzten Tag. Auf den ersten Blick mag die Prognosequalität unzureichend erscheinen, jedoch ist sie typisch für die Einzellosvorhersage für Halbleiterfabriken mit einem breiten Produktspektrum (siehe auch Kapitel 7 oder [Noack12]). Abbildung 6-1 stellt ebenfalls den Unterschied in

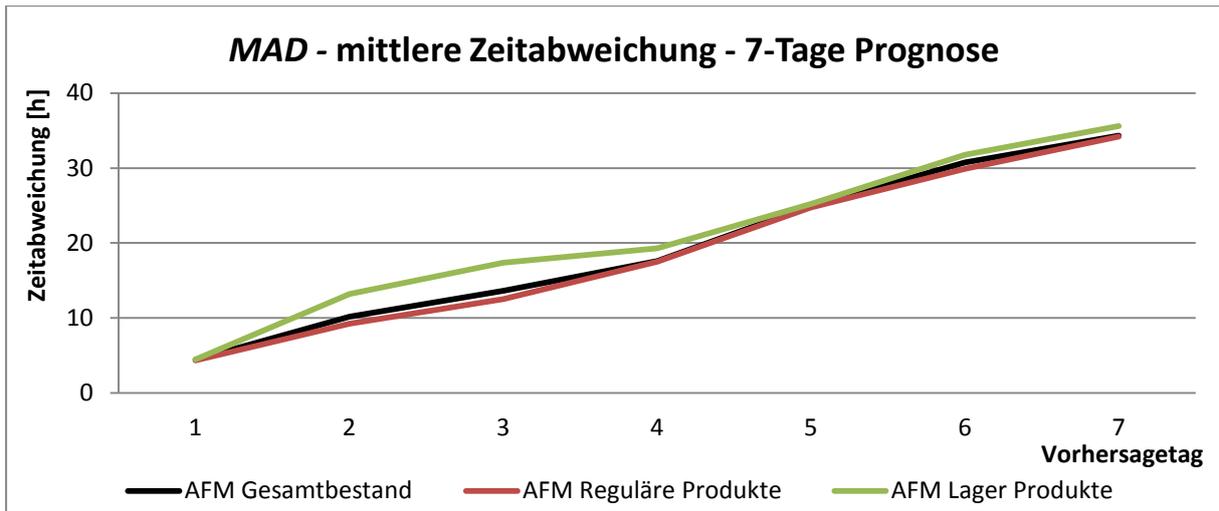


Abbildung 6-1: Der Verlauf der mittleren Zeitabweichung zwischen AFM-Prognose und Realität für den 7-tägigen Prognosehorizont

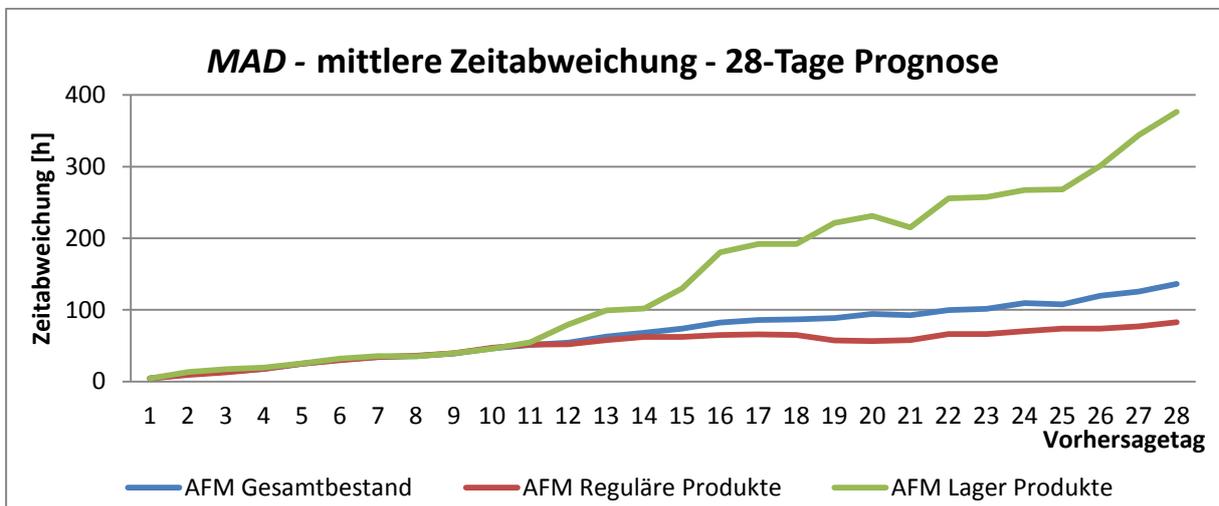


Abbildung 6-2: Der Verlauf der mittleren Zeitabweichung zwischen AFM-Prognose und Realität für den 28-tägigen Prognosehorizont

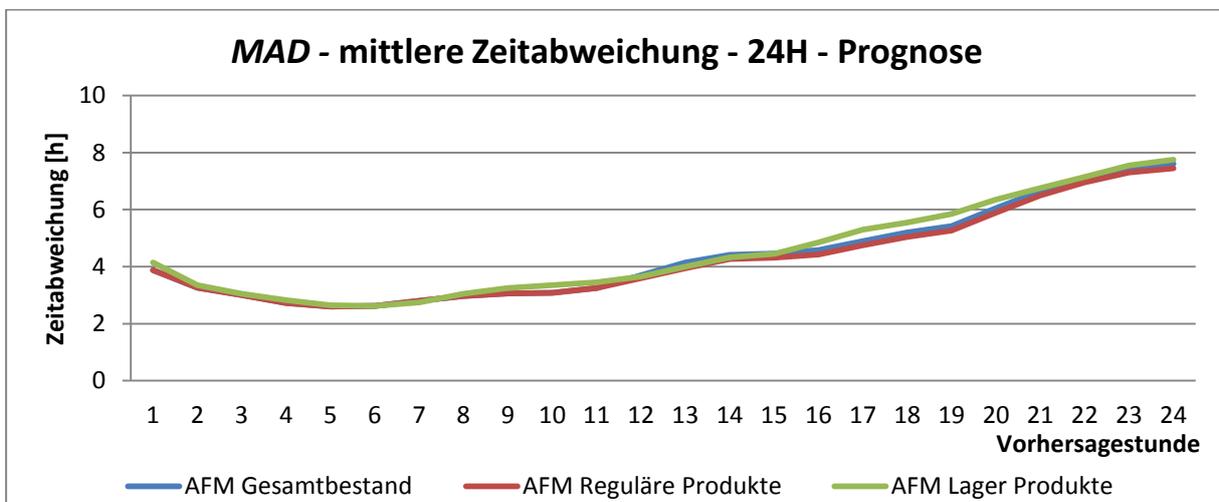


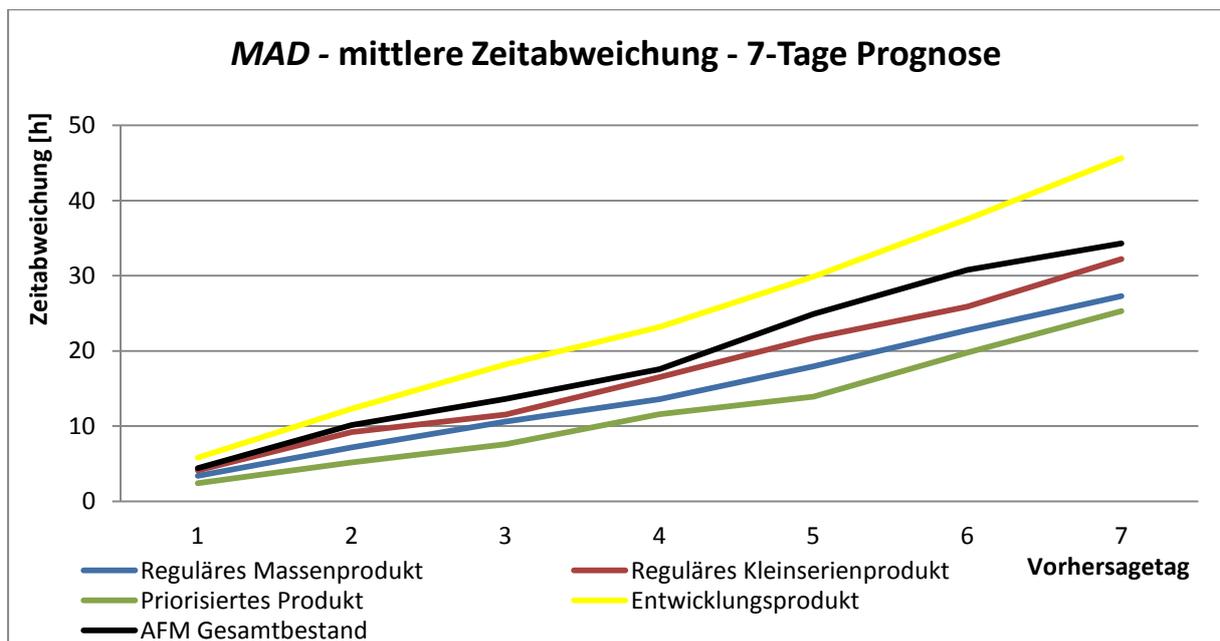
Abbildung 6-3: Der Verlauf der mittleren Zeitabweichung zwischen AFM-Prognose und Realität für den 24-stündigen Prognosehorizont

MAD zwischen den regulären Produkten und den Produkten, die aus dem Lager in die Fertigung eingespeist werden. Die Divergenz der beiden Verläufe zugunsten der konventionellen Produkte lässt sich vor allem in den ersten 4 Tagen feststellen. Danach wird der Fehler, bedingt durch die Prognoseunsicherheit der Einschleusung durch den obligaten Fehler der AFM nahezu überdeckt. Obwohl die funktionalen Anforderungen an die AFM den maximalen Vorhersagehorizont von 14 Tagen erfordern, wurde eine Prognose für die 4 Folgewochen untersucht. In Abbildung 6-2 wurde der Verlauf des MAD für eine 28-tägige Voraussage dargestellt. Der Anstieg der Prognosefehler innerhalb der Gesamtheit der Ergebnisse (blaue Kurve) bleibt auch in dem erweiterten Prognosebereich nahezu linear. Erklärungsbedarf besteht bei dem MAD-Verlauf für die Lagerprodukte. Der deutlich sichtbare Anstieg der MAD entsteht durch den Einfluss von fehlenden Informationen hinsichtlich der Materialeinschleusung. Die Lagereinschleusungsliste existiert in der untersuchten Fertigungsstätte nur im sehr kurzfristigen Bereich (maximal 7 Tage und minimal 1 Tag) und wird zusätzlich durch operationale manuelle Eingriffe verfälscht. Bei der Suche nach den fehlenden Informationen über die Einschleusung behilft sich die AFM mit der auf den Historienwerten basierenden Statistik für den Materialinput. Die künstlich erzeugten Zeitpunkte des Materialeingangs sind bereits von der Berechnung der Prognose mit einer Zeit- und Mengenabweichung behaftet und der Fehler wird beim Eintreffen an der Zieloperation fortgesetzt. Interessanterweise verringert sich der MAD-Anstieg bei der Ankunftsprognose der regulären Produkte für die Tage 15 bis 28 im Vergleich mit den zwei ersten Wochen des Prognosehorizonts. Die mittlere Abweichung der Ankunftszeitpunkte beim 28-ten Voraussagetag liegt unter 84 Stunden. Die Erklärung der erreichten Resultate liegt in der Entfernung von der Start- bis zur Zieloperation. Die Werkstücke, die durch einen Störeinfluss in der Abarbeitung verzögert werden, haben noch genug Zeit, um durch Verringerung der Wartezeit die verlorene Zeit wieder zu gewinnen. Sie erhalten automatisch eine Art von Priorisierung aufgrund der fälligkeitsgesteuerten Abfertigungsregeln. Weiterhin wurde eine Analyse der Prognosegüte der AFM im Ultrakurzfristbereich, also mit dem Prognosehorizont bis 24 Stunden, durchgeführt. Die Bewertung wurde durch eine große Zahl an möglichen Anwendungen beim Projektpartner bedingt. Die Ultrakurzfristvorhersage wird vor allem für die lokale Optimierung und Ablaufplanung anstelle des Dispatching als Inputgröße angewendet. Abbildung 6-3 stellt den Verlauf von MAD bei dem Voraussagehorizont von 24 Stunden dar. Der Prognosefehler beträgt bereits für die ersten Stunden der Vorhersage über 3 Stunden. Das resultiert aus der für dieses Anwendungsfeld unzureichenden Genauigkeit der Initialisierungsdaten (vgl. Kapitel 4.6.1). Resümierend wird die Verwendung der AFM im Ultrakurzfristbereich nicht empfohlen. Die entsprechende Pro-

gnosequalität für den in den funktionalen Anforderungen beschriebenen Horizont wurde erreicht. Weiterhin erlaubt die AFM-Vorhersagequalität ebenfalls die Trendprognosen von Materialankunftsmengen für die regulären Produkte für einen Horizont von 28 Tagen.

Interessant aus Sicht der Anwender ist die Bewertung der Prognosequalität mit Fokus auf die Ankunftsabweichung für einzelne Produkttypen. Abbildung 6-4 stellt den Vergleich von MAD für den Gesamtbestand in Kombination mit allen möglichen Zieloperationen (Schwarz) dar, mit:

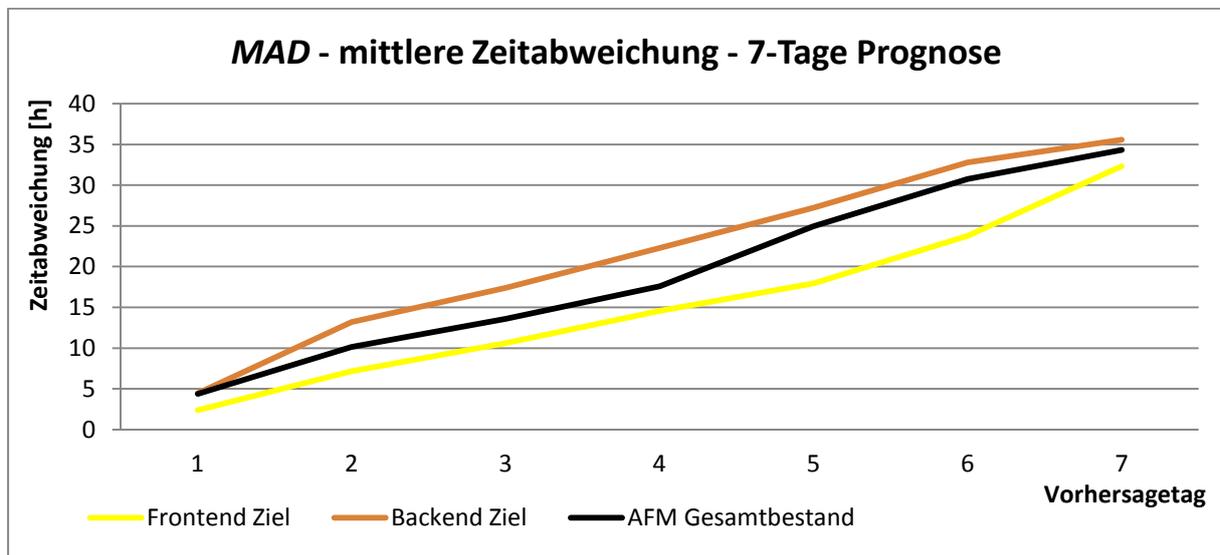
- ausgewähltem Massenprodukt (blau) – mit einem Bestand von ca. 2 % des Gesamtbestandes,
- priorisiertem Produkt (grün) – alle Werkstücke des Produktes sind durchgehend an allen Operationen priorisiert,
- einem etablierten Kleinserienprodukt – Produkt mit einem niedrigen Volumen, das bereits seit einem längerem Zeitraum produziert wird,
- Entwicklungsprodukt – ein typischer Vertreter von einem Produkt im Entwicklungsstadium oder einem ungewöhnlichen Fabrikat.



**Abbildung 6-4: Der Verlauf der mittleren Zeitabweichung zwischen AFM-Prognose und Realität für den 7-tägigen Prognosehorizont in Abhängigkeit von den Produkteigenschaften**

Das qualitativ beste Resultat liefert die AFM für die beschleunigten Produkte. Die mittlere absolute Abweichung beträgt für den 7-tägigen Prognosehorizont ungefähr 12 Stunden. Da derartige Produkte meistens lieferkritisch sind, ist eine genaue Prognose für die Ausschleusung sehr wichtig. Sie erlaubt eine frühzeitige Einstellung auf Folgeereignisse (wie z.B.: Mes-

sungen im Wafer-Test-Bereich oder Transport zu anderen Standorten oder Kunden) auf die prognostizierte Liefermenge und zusätzliche Beschleunigung des Materials falls die Lieferung nicht eingehalten werden kann. Geringfügig schlechter ist die Prognose für Massenprodukte. Die hohe Prognosequalität wird durch eine ausreichende Menge an Datenmaterial sowie durch typisches Verhalten dieser Produkte gewährleistet. Knapp unter dem Durchschnitt von 19 Stunden MAD für die gesamte AFM-Prognose liegen auch die etablierten Kleinserienprodukte. Anders jedoch gestaltet sich die Situation für die Entwicklungsprodukte, Einzelfabrikate oder neue, nichtetablierte Produktionsserien. Problematisch beim Prognostizieren von Ankünften bei derartigen Produkten ist die nicht ausreichende Menge an historischem Datenmaterial und die schwerwiegende Problematik der Statistikverfälschung durch Ausreißer. Der dargestellte MAD-Verlauf ist nur ein Beispiel, dessen Unterschied im Anstieg nur 15 % größer war als bei der Gesamtprognose. Beobachtet wurden aber auch Produkte, deren MAD am 7-ten Vorhersagetag 5 Tage betragen, also mehr als das Dreifache der Abweichung für die typische AFM-Voraussage. Trotz des sehr hohen Maßes an Qualität der Resultate der AFM-Prognose für etablierte oder priorisierte Produkte, die ca. 70 % des Bestandes ausmachen, liegt die Gesamtprognosequalität doch weit entfernt von den besten Ergebnissen. Die Erklärung dafür liegt bei dem sehr breiten Produktspektrum, dem ständig wachsenden Produktportfolio des Projektpartners sowie bei der erhöhten Auslastung der Fabrik im Analysezeitraum.



**Abbildung 6-5: Der Verlauf der mittleren Zeitabweichung zwischen AFM-Prognose und Realität für den 7-tägigen Prognosehorizont in Abhängigkeit von der Zielposition in der Fertigungslinie**

Die Analyse der MAD mit Fokus auf die Position der Prognoseziele in der Produktionslinie erfordert die Ausgliederung von zwei Bereichen aus der Fertigung:

- Frontend of Line – das erste Drittel der Produktionsschritte nach Einschleusung,

- Backend of Line – das letzte Drittel der Produktionsschritte vor Fertigstellung.

Abbildung 6-5 stellt den Vergleich der Vorhersagequalität mit der Annahme dar, dass sich das Ziel im Frontend (Gelb) oder Backend of Line (Orange) befindet. Die deutlich höhere Qualität der Vorhersageergebnisse im Frontend lässt sich durch einen geringeren Komplexitätsgrad der Produktionsprozesse und die deutlich kleinere Zahl von Operationen, die eine Interaktion Operator-Werkstück erfordern, wie z. B. die Bewertung von Messungen, erklären. Die Analysen der AFM im Bereich Backend of Line nach der Ausfilterung der Schwankungen der Losfreigabe durch die Fachkraft nach der Messung belegen, dass die Prognosequalität auch in diesem Produktionsbereich deutlich steigen würde.

## 6.2 Analyse des Anlieferungsverhaltens

Der Fokus der AFM liegt nicht auf dem Prognostizieren der Ankünfte von einzelnen Losen sondern auf Voraussagen der Anlieferungsströme des Materials. Aus diesem Grund wurde eine detaillierte Analyse und Auswertung der Prognosequalität nach dem Abschluss der Implementierungsphase durchgeführt. Die Quantifizierung der Treffsicherheit der Vorhersage erfolgt durch den über mehrere Stichproben errechneten mittleren absoluten prozentualen Prognosefehlers MAPE. Abbildung 6-6 stellt den Verlauf der Kurve, ergänzt durch das Intervall, durch die Standardabweichung einzelner Stichproben gebildet werden, dar. Die Basis für die Analyse bildeten ca. 200 unabhängige Prognosen in einem Zeitraum von einem Jahr für

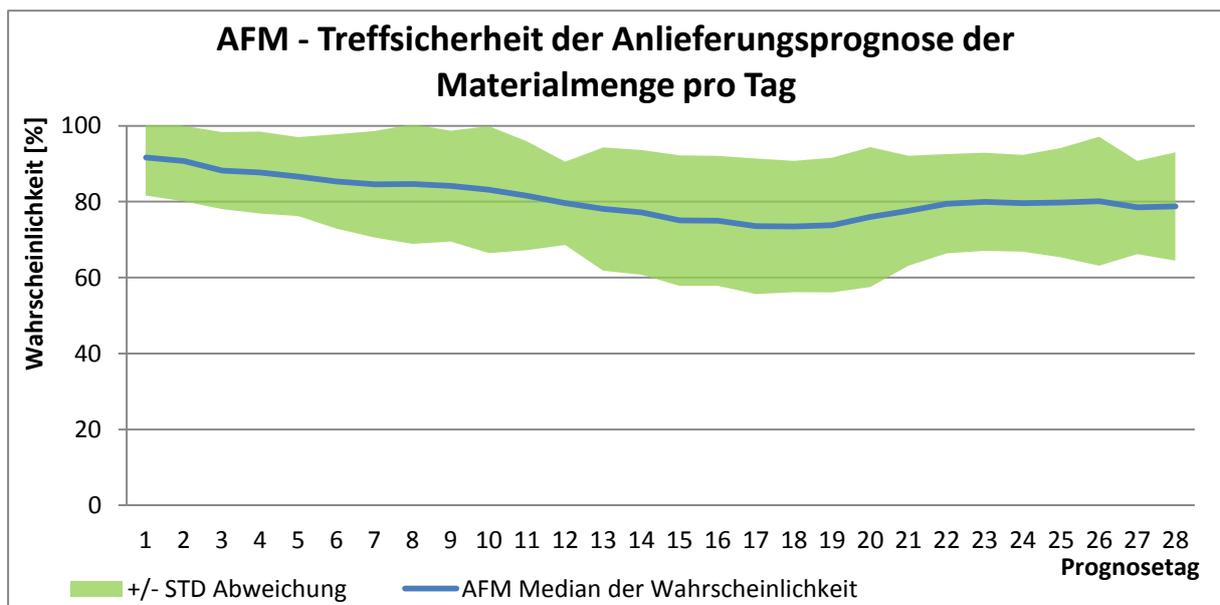


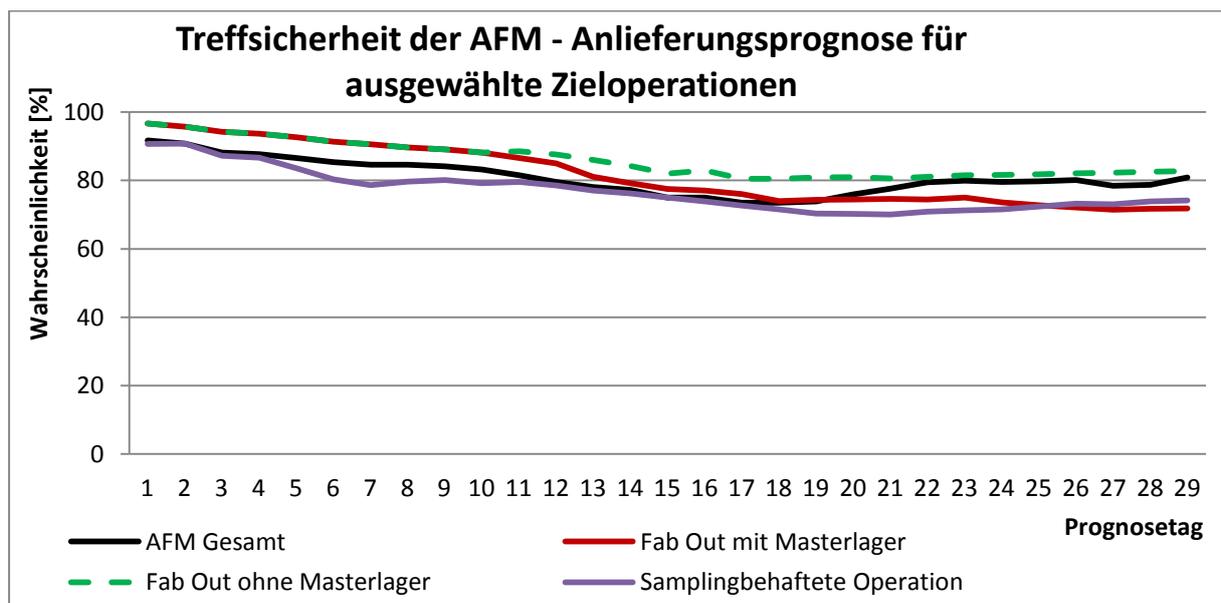
Abbildung 6-6: Treffsicherheit der Anlieferungsprognose der Materialmenge mittels AFM pro Tag

die Gesamtheit an möglichen Zieloperationen und Produkten. Vorerst wurde jede Vorhersage individuell bewertet. Anschließend erfolgte die Zusammenfassung mittels Medianbildung, der als blaue Kurve in Diagramm aus der Abbildung 6-6 abgebildet ist. Die X-Achse beinhaltet die Information über den Tag der Prognose. Die mittels AFM erreichte Treffsicherheit beträgt für:

- den ersten Tag: 91,7 % mit einer Standardabweichung von 10,1 %
- den zweiten Tag: 90,7 % mit einer Standardabweichung von 10,6 %
- den siebten Tag: 84,6 % mit einer Standardabweichung von 14,0 %
- den Mittelwert der ersten Woche: 87,8 % mit einer Standardabweichung von 11,2 %
- den Mittelwert der zweiten Woche: 81,2 % mit einer Standardabweichung von 15,0 %
- den Mittelwert für die 28 Tage: 80,9 % mit einer Standardabweichung von 14,3 %

Ab der zweiten Woche der Prognose ist die Qualität der Resultate unzureichend um tagfeine Aussagen zu treffen. Es wurde eine wochenfeine Granularität eingefügt und es wird nach den Anlieferungstrends gesucht.

Abbildung 6-7 stellt die Treffsicherheit in Abhängigkeit von den Eigenschaften der Zieloperation dar. Ausgesucht wurden mehrere Produktionsschritte, die mit einem Sampling an der Zieloperation behaftet sind. Die Untersuchungen wurden mit dem gesamten Produktionsmix durchgeführt. Beobachtet wurde nur die marginale Verringerung der Trefferquote für die Anlieferung im Vergleich zu der Gesamtaussage der AFM. Als ein typischer Vertreter für den Zielprozessschritt ohne den Einfluss von Sampling wurde die letzte Operation in dem Arbeitsplan ausgewählt. Sie gleicht der Ausschleusung aus der Fabrik. Abbildung 6-7 stellt zwei



**Abbildung 6-7: Treffsicherheit der Anlieferungsprognose der Materialmenge mittels AFM pro Tag an ausgewählten Zieloperationen**

Kurven für den Output der Produktionsstätte dar. Der Unterschied resultiert aus dem Abzug der neu eingeschleusten Lose vom Masterlager. Die Verfälschung der Anlieferungsprognose durch die Stochastik in der Auslagerungspolitik ist ab dem 11. Tag deutlich sichtbar. Aus diesem Grund repräsentiert also die grüne, gestrichelte Kurve den Treffsicherheitsverlauf für die nicht mit Sampling behafteten Operationen besser. Durch Vergleich der Zahlenwerte der Kurven lila und grün lässt sich ein Unterschied von:

- erster Tag: 6,6 %
- zweiter Tag: 5,5 %
- siebter Tag: 12,8 %
- Mittelwert erste Woche: 8,2 %
- Mittelwert zweite Woche: 8,8 %
- Mittelwert für 28 Tage: 8,9 %

feststellen, jeweils zugunsten der Zieloperation ohne Sampling. Die genauere Untersuchung der Treffsicherheit bei der Problematik des Überspringens der Zieloperation wurde im Kapitel 6.4.2 beschrieben.

Aus der Sicht der Anwender ist auch die Trefferquote in Abhängigkeit von Produkteigenschaften von Bedeutung. Die Analyse wurde im Rahmen der Bewertung der Prognose detailliert durchgeführt. Untersucht wurden verschiedene Produktspezifika, wie: Länge des Arbeitsplans, typischer Anzahl der Wafer im Los oder repräsentativer Verspätung etc. Da die vollständige Beschreibung der durchgeführten Tests den Rahmen dieser Ausarbeitung sprengen würde, werden nur die erzielten Resultate für ein Massenprodukt und ein Entwicklungsprodukt an dieser Stelle detailliert diskutiert. Die Abbildung 6-8 und Abbildung 6-9 stellen jeweils die Verläufe der Treffsicherheit der AFM für die ausgewählten Entwicklungsprodukte sowie die Massenware, ergänzt über das Standardabweichungsintervall dar. Die Datengrundlage für die Erstellung des Diagramms in Abbildung 6-8 umfasst im Vergleich zu den anderen Validierungsvorgängen deutlich weniger Stichproben. Es ist einerseits durch eine kleine Anzahl von Vertretern der Entwicklungsware in der Produktionslinie bedingt. Andererseits ist es schwer, den Stichprobenumfang durch die Verlängerung des Analysezeitraumes mit unabhängigen Prognosen zu vergrößern, weil die Produkte sich im Laufe der Zeit etablieren und das Entwicklungsstadium verlassen. Aus diesem Grund weist der Mittelwert der Treffsicherheit deutliche Schwankungen zwischen einzelnen Prognosetagen auf. Interessant ist, dass die Wahrscheinlichkeit von einer sehr guten Prognose für die ersten 48 Stunden recht hoch ist. Derartige Produkte werden meistens nicht priorisiert abgearbeitet, also bewegen sie sich nur

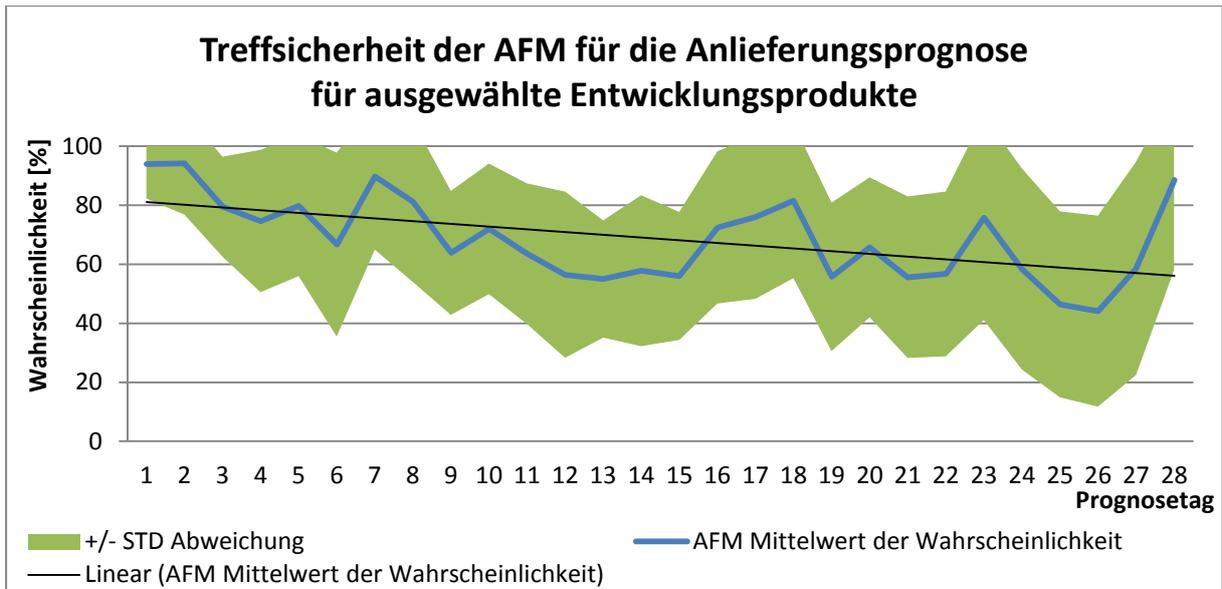


Abbildung 6-8: Treffsicherheit der Anlieferungsprognose der Materialmenge mittels AFM pro Tag für ausgewählte Entwicklungsprodukte

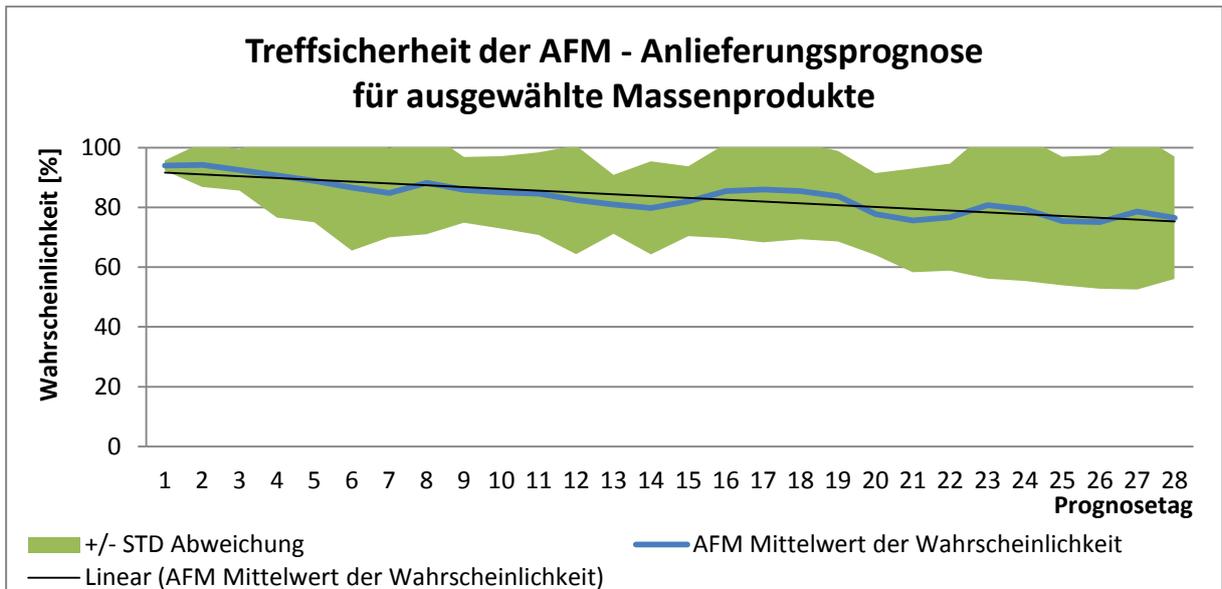


Abbildung 6-9: Treffsicherheit der Anlieferungsprognose der Materialmenge mittels AFM pro Tag für ausgewählte Massenprodukte

	Entwicklungsprod.	Entwicklungsprod. Lineare Approximation	Massenprod.	Massenprod. Lineare Approximation
1. Tag	94,0 %	81,1 %	94,2 %	91,6 %
2. Tag	94,2 %	80,1 %	94,0 %	91,0 %
7. Tag	89,8 %	75,5 %	84,7 %	88,0 %
Mittelwert 1. Woche	82,7 %	78,3 %	90,2 %	89,8 %
Mittelwert 2. Woche	64,3 %	71,8 %	83,8 %	85,6 %
Mittelwert 28 Tage	68,6 %	68,6 %	83,5 %	83,5 %

Tabelle 6-1: Der Vergleich der Treffsicherheit für die Entwicklungs- und Massenprodukte

über wenige Produktionsschritte pro Tag. Die enorme Schwankung der Hold-Rate sowie der Hold-Dauer führt als die hauptsächliche Störgröße zur Verfälschung der Ankunfts vorhersage. Die Wahrscheinlichkeit des Auftretens des ungewöhnlichen Hold-Verhaltens ist proportional zur Anzahl der abgearbeiteten Operationen. Für Massenprodukte ist die Stichprobenanzahl deutlich voluminöser. Aus diesem Grund fällt der sanftere Verlauf der Wahrscheinlichkeitskurve für die Massenprodukte in Verhältnis zu den Entwicklungsprodukten beim Vergleich der Diagramme auf. Zur Bewertung der Treffsicherheit beider Produktarten hilft die Approximation mittels linearer Trendlinien. Die tägliche Verringerung der Treffsicherheit liegt für die Entwicklungsprodukte bei 0,9 % bei der Treffwahrscheinlichkeit von ca. 81 % für den Vorhersagestarttag und für die Massenware entsprechend bei ca. 0,6 % und 92 %. Die Quantifizierung der erzielten Ergebnisse wurde in der Tabelle 6-1 dargestellt.

Außer der Treffsicherheit ist auch die Stabilität der Prognosegüte ein wichtiges Maß für die Qualität des Vorhersageansatzes. Die Beständigkeit der Brauchbarkeit der Resultate ist vor allem beim Anwendungsfall Informationsboard (vgl. Kapitel 4.7) aus Akzeptanzgründen von Interesse. Für die Analyse wurde die Wochenprognose der Ausschleusung der Fabrik gesucht, die immer am Anfang einer Woche ausgeführt wird und die Wochenausschleusung als Antwort liefert. Die Menge des ausgeschleusten Materials wird zusätzlich nach speziellen Produkteigenschaften in wenige Gruppen GE geteilt. Der sehr hohe Verdichtungsgrad der Resultate der Voraussage erfordert eine sehr hohe Treffwahrscheinlichkeit in Form der dargestellten Aussagen. Bei der AFM-Validierung wurde der annehmbare Bereich mit der Treffsicherheit über 90 % als ausreichend angenommen. Die Qualitätskennzahl, also Treffwahrscheinlichkeit  $P_{Ges}$  wurde mit der Formel:

$$P_{Ges} = \sum_i^n \frac{\Omega_i}{\Omega} * P_i \Big/ \sum_i^n \frac{\Omega_i}{\Omega} \Leftrightarrow P_i = \frac{|\omega_{r_i} - \omega_{p_i}|}{\omega_{r_i}} * 100 \%$$

berechnet, wobei  $n$  Anzahl der Gruppen GE,  $\Omega_i$  die prognostizierte Menge der Wafer der Gruppe  $i$ ,  $\Omega$  die gesamte prognostizierte Menge der Wafer und  $P_i$  die Trefferquote für die Gruppe  $i$  bedeutet. Der Parameter  $P_i$  wird für die  $i$ -te Gruppe GE durch die Division des absoluten Unterschieds zwischen der real ausgeschleusten  $\omega_{r_i}$  und prognostizierten Materialmenge  $\omega_{p_i}$  durch  $\omega_{r_i}$  ermittelt. Abbildung 6-10 stellt die Ergebnisse einer der oben beschriebenen Analysen der Stabilität der AFM Vorhersageresultate für die Wochenwerte graphisch dar. Die Analyseperiode erstreckte sich über ein Zeitraum von fast 18 Monaten und umfasste nur unabhängige Prognosen, also alle Vorhersagen des Fabrikoutputs für disjunkte Wochen. Der Fabrikoutput wurde aber nur in zwei Hauptgruppen geteilt. Nur zwei Stichproben aus diesem

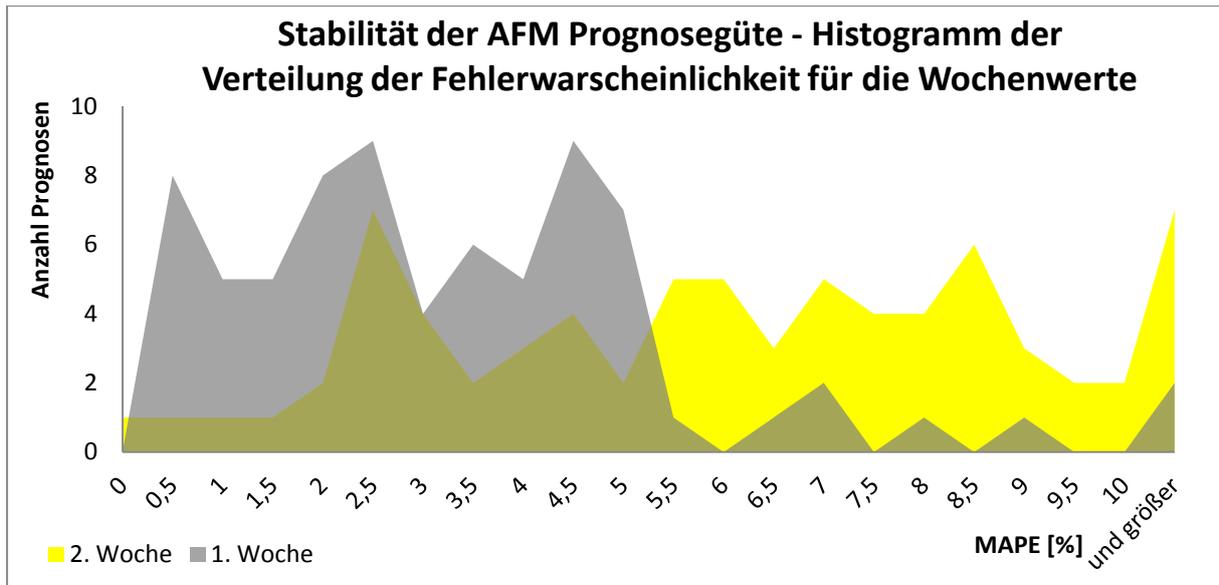


Abbildung 6-10: Histogramm der Verteilung der Fehlerwahrscheinlichkeit für die Wochenprognose

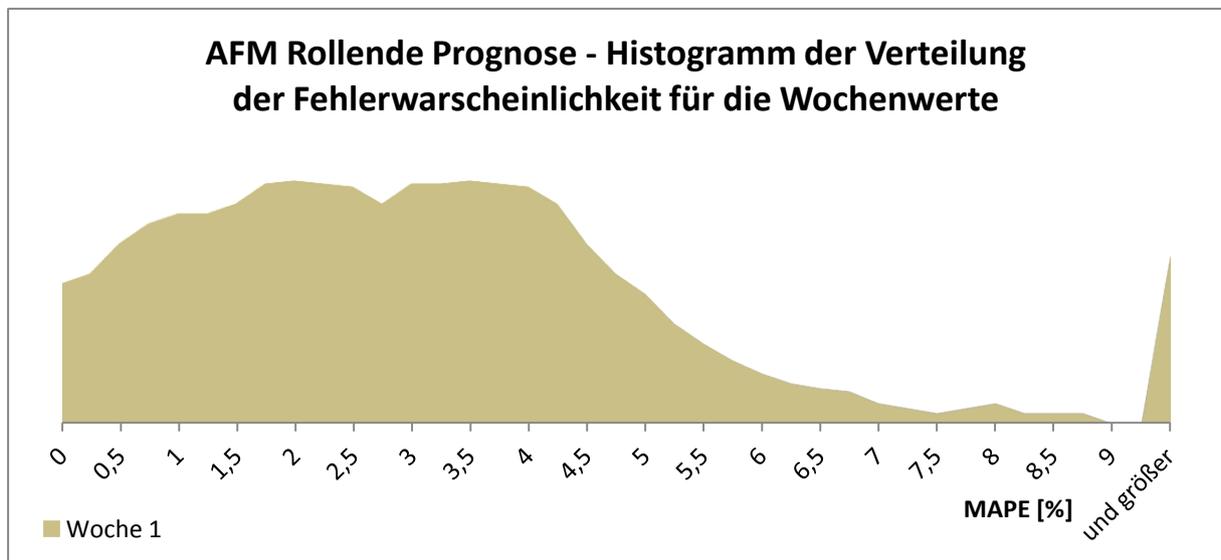


Abbildung 6-11: Histogramm der Verteilung der Fehlerwahrscheinlichkeit für die rollende Wochenprognose

Zeitraum wichen mehr als 10 % von der Realität in der ersten Vorhersageweche ab und bei ca. 90 % aller Versuche lag die Fehlergrenze bei maximal 5 %. Für die zweite Woche der Prognose streut der Prognosefehler von 2 % bis 10 % nahezu gleichverteilt und auch die Anzahl der Ausreißer wächst deutlich. Abbildung 6-11 stellt ein Histogramm der Verteilung von Prognosefehlern für die rollende Wochenprognose dar. Unter dieser Vorhersageart verbirgt sich nicht eine tägliche Prognose der nächsten 7 Tage, sondern ähnlich wie in Abbildung 4-36 veranschaulicht, die tägliche Verbesserung der Prognosequalität der laufender Woche durch eine Neuberechnung der Voraussage. Vor Wochenanfang wird eine Voraussage für gesamte 7-tägige Periode erstellt. Nach dem Abschluss des ersten Wochentages sucht die AFM die

reale Ausschleusung und summiert diese mit der Prognose für die nächsten 6 Tage. So werden fortlaufend für jede Woche 7 Prognosen erstellt, deren Qualität in den meisten Fällen mit der Menge an realen Daten positiv korreliert. Wie bei der vorherigen Untersuchung weisen auch bei dieser Analyse knapp 90 % aller Prognosen einen Fehler von unter 5 % auf.

### 6.3 Analyse der Prognose für ausgewählte Anwendungsfälle

Basierend auf der Anlieferungsanalyse wurden auch weitere Arten der Prognose, wie das Frühwarnsystem für eine kommende Flaschenhalssituation und das Ampelsystem für die Planung der Wartungsmaßnahmen (vgl. Kapitel 4.7) entwickelt. Da die Voraussagen außer der Materialanlieferung auch noch zusätzliche Informationen beinhalten, werden sie separat bewertet.

Die Treffsicherheit bei der Erkennung von Engpasszuständen  $P_{BtEWS}$  wird durch einen Vergleich mit der realen Auslastung  $util_{real}$  [%] mit der Formel:

$$P_{BtEWS} = 100 \% - abs|util_{real} - util_{AFM}|$$

bestimmt. Die Kalkulation von  $util_{AFM}$  wurde detailliert bereits im Kapitel 4.7 beschrieben.

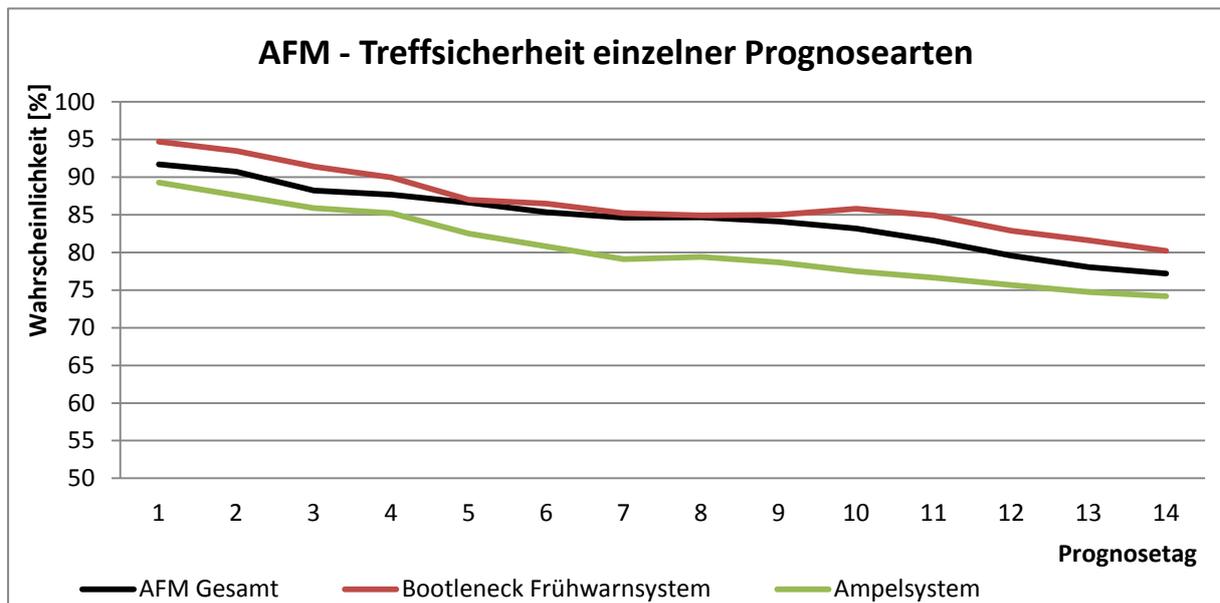


Abbildung 6-12: Die Treffsicherheit bezüglich der Eignung der AFM für die einzelnen Anwendungsfelder

Die Analyse der Prognosequalität wurde an mehreren Zieloperationen durchgeführt, die durch deutliche Schwankungen in der Prozesszeit gekennzeichnet sind und bei denen die Gefahr des Engpasses hoch ist. Die Ergebnisse der Studie wurden in Abbildung 6-12 dargestellt. Die durch AFM erreichte Wahrscheinlichkeit einer richtigen Prognose in dem beschriebenen Anwendungsfeld beträgt für:

## Validierung

- den ersten Tag: 94,7 %
- den zweiten Tag: 93,5 %
- den siebten Tag: 82,2 %
- den Mittelwert erste Woche: 89,8 %
- den Mittelwert zweite Woche: 83,6 %
- den Mittelwert für 14 Tage: 86,7 %

Die Ursache der höheren Qualität der Bottleneck-Prognose als der Anlieferungsprognose von AFM liegt vor allem an der Konvergenz von einzelnen Prozesszeiten für unterschiedliche Produkte, die trotz der deutlichen Prozesszeitschwankungen oft auftritt.

Die Visualisierung der Prognose mithilfe des Ampelsystems wurde gemeinsam mit den Entwicklern der Kurzfristsimulation (ein paralleles Projekt) und den Mitarbeitern des Projektpartners entworfen und entwickelt. Das Ziel einer derartigen Darstellungsform liegt in der Vereinfachung der Aussage für den Anwender. Für den Mitarbeiter, der zum Beispiel für die Planung der vorbeugenden Wartung der Anlagen zuständig ist, ist nur eine weitgehend verdichtete Information notwendig. Schließlich konzentriert sich das Interesse dieser Person bezüglich der Prognose auf die Frage, darf ich an diesem Tag die Wartungsmaßnahmen an der Anlage durchführen, oder wäre ein anderer Zeitpunkt dazu eher geeignet. Abbildung 6-13 stellt die Visualisierung der Prognose mithilfe von Ampeln dar. Im oberen Teil werden die durch die Prognose erzeugten Ampeln abgebildet, die unteren resultieren aus den realen Ankünften. Die so fortgeschrittene Verdichtung der Information erfordert eine gesonderte

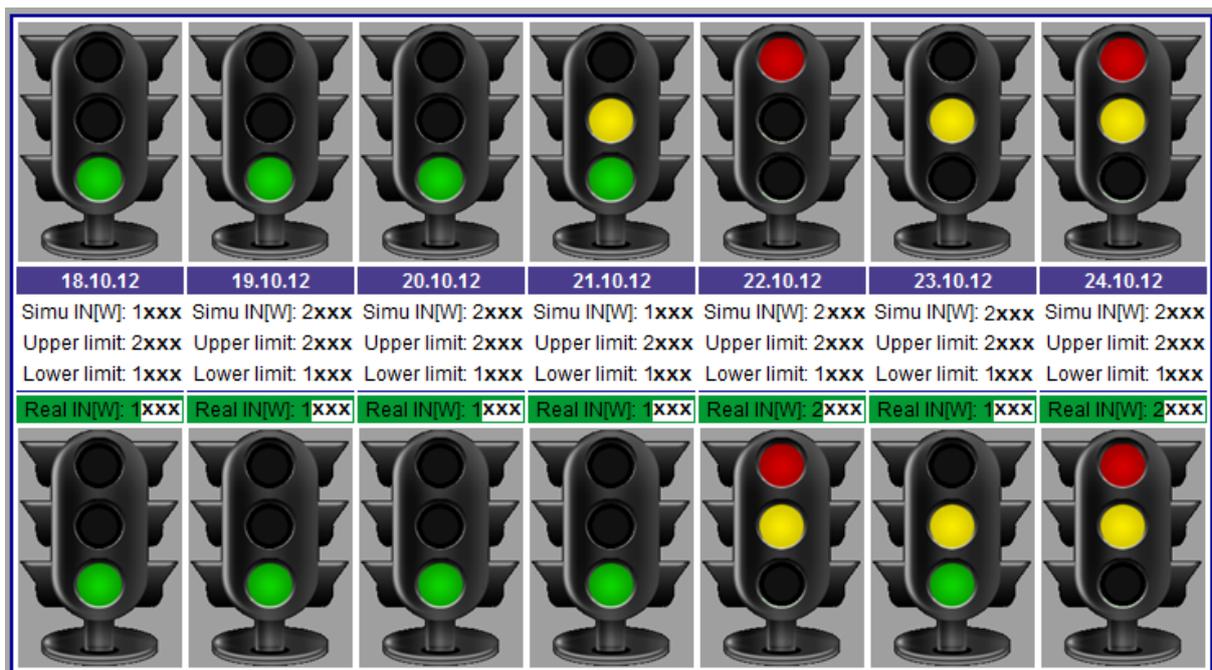


Abbildung 6-13: Die Visualisierung der Prognose und zugehörigen Realität mithilfe des Ampelsystems zwecks Validierung

Qualitätsbewertung. Die Prognosegüte  $Pq_{AFM\_Traffic}$  wird unter Anwendung der Formel:

$$Pq_{AFM\_Traffic} = \frac{\sum_{wc=1}^k m_{wc} * Pq_{wc}}{\sum_{wc=1}^k m_{wc}}, \text{ wobei } Pq_{wc} = \frac{\sum_{i=1}^n w_i * Pq_i}{\sum_{i=1}^n w_i}$$

berechnet, wobei sich unter  $Pq_{wc}$  die Vorhersagequalität einer einzelnen Maschinengruppe verbirgt. Die Quantifizierung von  $Pq_{wc}$  wird mithilfe von einem über die zeitliche Entfernung gewichteten Mittelwert der Qualität für einzelne Voraussageperioden  $Pq_i$  (zum Beispiel: 24 h) ermöglicht. Die Gewichtung  $w_i$  ist indirekt proportional zur Entfernung der Vorhersageperiode vom Initialisierungszeitpunkt. Die Werte für  $Pq_i$  für ein 5-Ampel-System wurden in Tabelle 6-2 dargestellt. Bei einem Ampel-System werden die Grenzen für die Anlieferungsmengen im Voraus in Gesprächen mit den Experten aus Produktionsbereichen bestimmt. Daraus resultiert eine 100 %-ige Treffsicherheit, wenn die gleiche Farbe in Realität und Prognose bestimmt wird. Angenommen Grüngelb wurde prognostiziert, aus der Realität resultiert aber rot, beträgt die Qualität der Prognose nur 25 %.

REAL	SIMU				
ROT	100 %	75 %	50 %	25 %	0 %
ROTGELB	75 %	100 %	75 %	25 %	0 %
GELB	50 %	75 %	100 %	75 %	50 %
GRÜNGELB	50 %	50 %	75 %	100 %	75 %
GRÜN	50 %	50 %	75 %	75 %	100 %

Tabelle 6-2: Die Quantifizierung der Vorhersagequalität für ein 5-Ampel-System

Die ungleichmäßige Verteilung der Qualitätswerte lässt sich mit der Anwendung begründen. Wenn die vorausgesagte niedrige Anlieferung in einer Periode zu einer Wartung führt, in der Realität jedoch eine sehr große Materialmenge ankommt, ergibt sich daraus ein kostenintensives Planungsproblem. Umgekehrt aber, wenn der Planer zwar ein rotes Licht sieht und keine Wartung einplant, allerdings in der Realität die Anlagen nicht ausgelastet sind, ist das Problem für den Fertigungsablauf deutlich geringer. Die Notwendigkeit der Einführung der Gewichtung  $w_i$  lässt sich mit der Flexibilität bei der Wartungsplanung begründen. Wenn eine Maßnahme für einen 7. Tag geplant wurde, aber der Tag sich doch nach ein paar Tagen als ungeeignet herausstellt, lässt sich die Wartung zumeist noch verschieben. Somit ist die Aussage für den 7. Tag deutlich weniger wichtig als für den Folgetag, an dem keine Verschiebung mehr möglich ist. Die Kalkulation der allgemeinen Qualität der Prognose – visualisiert durch ein Ampel-System  $Pq_{AFM\_Traffic}$  – umfasst mehrere Anlagengruppen, die nach dem mittleren Durchsatz  $m_{wc}$  gewichtet werden. Der Analysezeitraum umfasste über 100 Prognosen über ein Kalenderjahr verteilt. Die Ergebnisse der Studie wurden in Abbildung 6-12 abgebildet.

Die durch AFM erreichte mittlere Qualität einer Prognose in beschriebenen Anwendungsfeld beträgt für:

- den ersten Tag: 89,3 %
- den zweiten Tag: 87,6 %
- den siebten Tag: 79,1 %
- den Mittelwert erste Woche: 84,3 %
- den Mittelwert zweite Woche: 76,7 %
- den Mittelwert für 14 Tage: 80,5 %

wobei 100 % Qualität eine ideale Voraussage bedeutet. Die Werte gelten nur unter der Annahme der Qualitätsbewertung aus Tabelle 6-2.

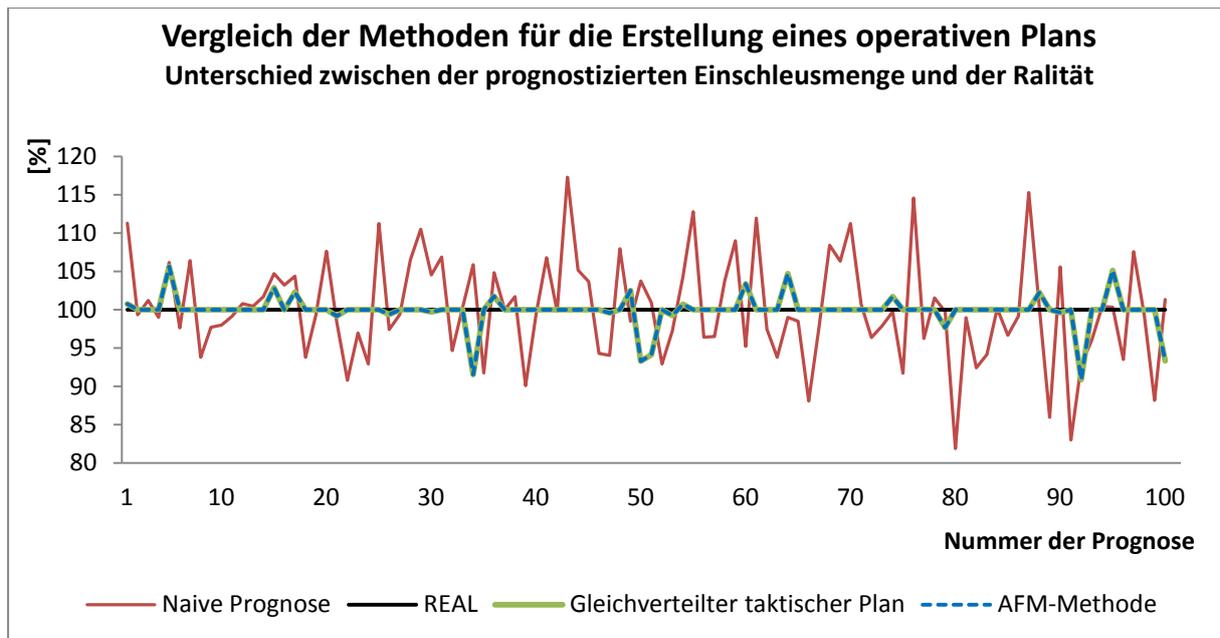
## **6.4 Verifikation der Ansätze und Lösungswege durch Validierung der AFM für ausgewählte Problembereiche**

### **6.4.1 Bewertung von unterschiedlichen Strategien in der Erstellung des operativen Plans für die Materialeinschleusung**

Der Nachweis der Notwendigkeit für einen Entwurf und die anschließende Implementierung einer intelligenten Aufbereitung des Datenmaterials, beschrieben in den Kapiteln 4.2 und 4.3, liefert die Gegenüberstellung von vorausberechneten und realen operativen Einschleusplänen sowie einen Vergleich der Ergebnisse der jeweiligen Prognosen. Es wurden drei Strategien im Rahmen des Projektes untersucht:

- eine Projektion der tatsächlichen Zeitpunkte für den Produktionsstart der Lose der letzten Woche,
- eine Gleichverteilung der Startzeitpunkte für das Material unter Einbeziehung der Losmengen aus dem taktischen Einschleusplan,
- auf der Statistik basierender, intelligenter Berechnungsansatz der AFM.

Die erste der drei Strategien ist ein charakteristisches Beispiel einer naiven Prognose [Kirchg05]. Der signifikante Vorteil einer derartigen Prognosemethode ist selbstverständlich die Einfachheit der Berechnung, bei der zu den Zeitpunkten der Loseinschleusung aus der Vergangenheit eine Zeitperiode von 7 oder 14 Tagen addiert wird. Der wesentliche Nachteil dieses Verfahrens ist die Ungenauigkeit. In Abbildung 6-14 wird ein Vergleich der untersuchten Techniken dargestellt. Angenommen die reale Einschleusung für die nächsten 7 Tage beträgt  $R_e = 100 \%$ , wird die Prognoseabweichung  $Q_p$  in [%] in Bezug auf  $R_e$  real mit der



**Abbildung 6-14: Die prozentualen Unterschiede in der Einschleusmenge zwischen der Realität und unterschiedlichen Methoden der Bestimmung des operativen Plans**

Formel  $Q_P [\%] = R_e * \frac{E_P}{E_R} = 100 \% * \frac{E_P}{E_R}$  berechnet, wobei  $E_p$  die prognostizierte und  $E_R$  die reale Menge der startenden Wafer bedeutet. Um die Verfahren quantitativ auswerten zu können, wird der Mittelwert über den absoluten Fehler ermittelt:  $\overline{E_P} = \frac{1}{n} \sum_{i=1}^n |100 \% - Q_P [\%]_i|$ , wobei  $n$  die Anzahl der untersuchten Prognosen bedeutet. Um die statistische Relevanz der Untersuchung zu gewährleisten, wurden als Basis 100 unabhängige Prognosen erstellt. Der mittlere Fehler für die untersuchten Verfahren beträgt:

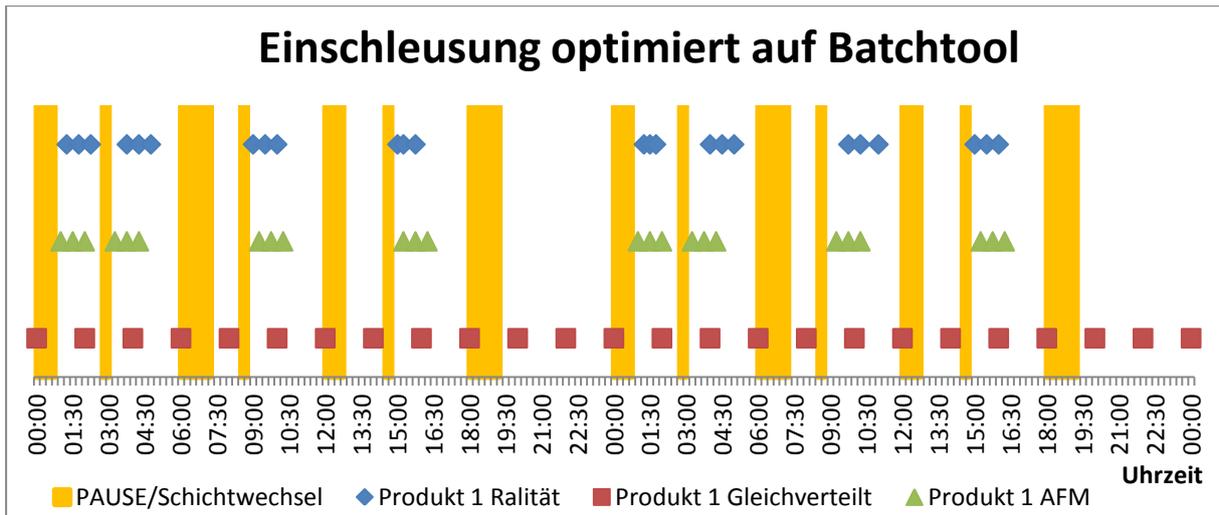
- $\overline{E_P} = 5,261 \%$  für die naive Methode,
- $\overline{E_P} = 0,497 \%$  für gleichverteilte Einschleusung,
- $\overline{E_P} = 0,497 \%$  für die Methode von AFM,

was ein Ausschlusskriterium für den naiven Ansatz bedeutet. Interessant ist die Abweichung von der Realität bei den zwei anderen Methoden. Die Untersuchung der Divergenz ergab die Unterschiede zwischen der geplanten und der realen Einschleusung. Da die beiden Methoden mit der gleichen Losmenge aus der Planung arbeiten, kann unmöglich festgestellt werden, welche davon nun besser ist. Erforderlich wurde folglich auch eine Revision der genauen Zeitpunkte für die Einschleusung. Ein Problem bei der Bewertung der Methoden stellt der Fakt dar, dass die eindeutigen Losnummern erst bei Fabrikationsstart vergeben werden. Aus diesem Grund wurde die Bildung von Zeitreihen erforderlich, die genaue aufsteigend sortierte Listen  $L_R$ ,  $L_G$  und  $L_{AFM}$  mit den Zeitangaben zur Materialeinschleusung beinhalten. Aus den Elementen mit der gleichen Position in der Liste resultieren Wertetripel in der Form  $\{TS_R;$

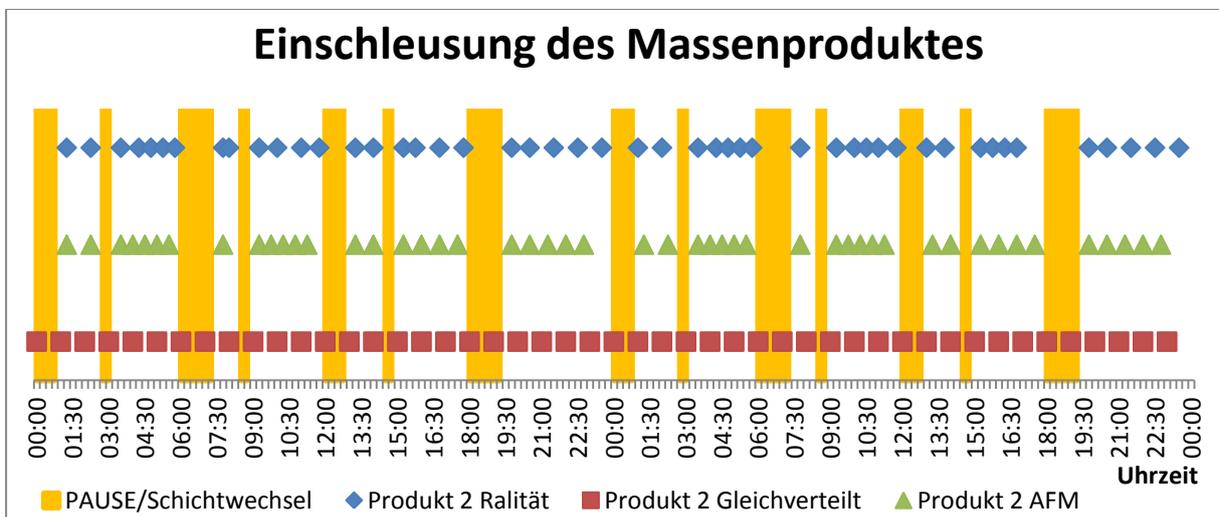
$TS_G; TS_{AFM}$ }, wobei die Werte  $TS_R, TS_G, TS_{AFM}$  die Zeitstempel der Einschleusung in der Realität bzw. bestimmt über die Gleichverteilung oder mithilfe des AFM-Ansatz bezeichnen. Anschließend wurde die mittlere Abweichung des Zeitunterschiedes  $\overline{TDev}$  für die beiden Ansätze mit den folgenden Formeln ermittelt:

$$\overline{TDev}_{R\_to\_G} = \frac{1}{n} \sum_{i=1}^n |TS_{R;i} - TS_{G;i}| \quad \text{und} \quad \overline{TDev}_{R\_to\_AFM} = \frac{1}{n} \sum_{i=1}^n |TS_{R;i} - TS_{AFM;i}| ,$$

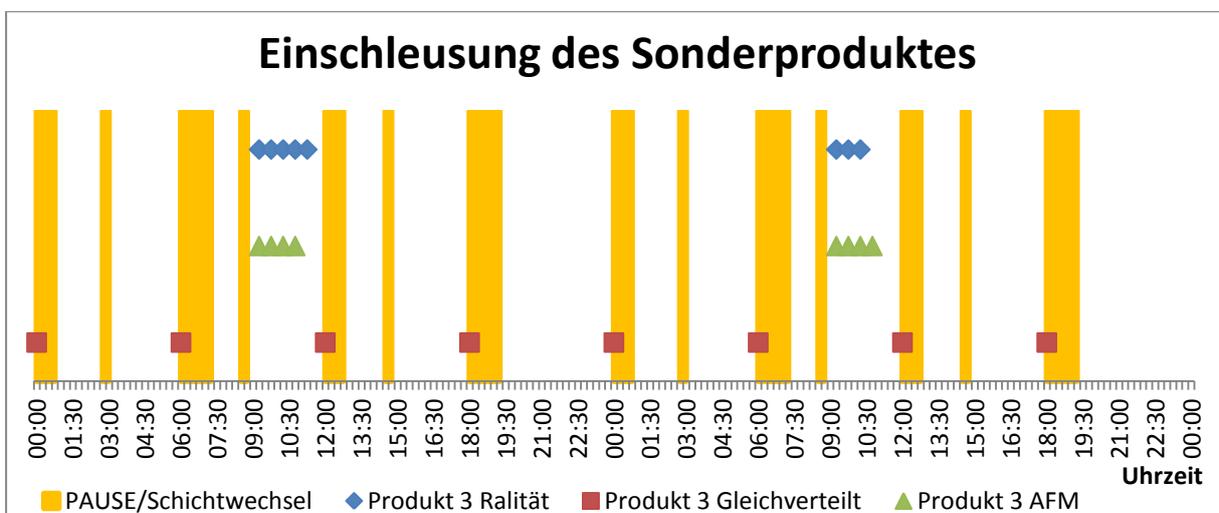
wobei  $n$  das Maximum gemeinsamer Elemente in der Listen  $L_R, L_G$  und  $L_{AFM}$  bedeutet. Die Materialmenge, die aus dem Nichteinhalten der Planwerte resultiert, werden in der Untersuchung nicht betrachtet. Interessant sind die Unterschiede in den Materialeinstartstrategien für verschiedene Produkte. Abbildung 6-15 stellt ein Beispiel einer speziellen Taktik für die Materialeinschleusung dar, bei der die Lose in Gruppen zusammengefasst werden und nahezu gemeinsam eingestartet werden. Diese Behandlungsweise wird vor allem dann bevorzugt, wenn sich in unmittelbarer Nähe des ersten Schritts auf der Route eines Produkts eine spezielle Operation befindet, die die Batchfertigung verlangt. Aus der Minimierung der Wartezeiten zur Batchbildung resultiert eine Verringerung der Durchlaufzeit. Zudem steigt die Auslastung der Batchanlage. Die Basis für die Untersuchung bildeten wie bei der Verifikation der Einschleusmengen 100 unabhängigen Prognosen. Die mittlere Abweichung der Zeitunterschiede für derartig gestartete Produkte beträgt für die Methode der Gleichverteilung  $\overline{TDev}_{R\_to\_G} = 202,53 [min]$ , was im Vergleich zum AFM-Ansatz mit  $\overline{TDev}_{R\_to\_AFM} = 23,01 [min]$  ca. 3 Stunden Prognosefehler bedeutet. Die Divergenz im Prognosefehler bei typischen Massenprodukten, wie in Abbildung 6-16 dargestellt, ist mit ca. 15 min bei  $\overline{TDev}_{R\_to\_G} = 33,75 [min]$  und  $\overline{TDev}_{R\_to\_AFM} = 16,81 [min]$  nicht mehr so gravierend. Die Erklärung des Sachverhaltes ist ein deutlich kürzeres Zeitintervall zwischen der Einschleusung einzelner Lose, sowie die breitere Verteilung über den gesamten Tag, die aus der Materialmenge resultiert. Schließlich wurden auch Sonder- und Entwicklungsprodukte auf dieselbe Weise untersucht. Abbildung 6-17 stellt ein typisches Beispiel für die Einschleustaktik derartiger Produkte vor. Die Fertigung startet also meistens einmal pro Tag in der Frühschicht. Die Gründe dafür müssen in der Organisation gesucht werden, weil über die Art und Anzahl der Werkstücke täglich entschieden wird, die Entscheidungsträger vor allem in der Normalschicht arbeiten und die Sondersitzungen normalerweise in den Morgenstunden stattfinden. Das zentrale Problem für beide Prognoseansätze stellt die ungleichmäßige und wechselhafte Verteilung der Lose über die Wochentage dar. Aus diesem Grund resultiert in beiden Fällen ein



**Abbildung 6-15: Vergleich der Methoden zur Abbildung des auf das Batchtool optimierten Einschleusverhalten**



**Abbildung 6-16: Vergleich der Methoden zur Abbildung des Einschleusverhaltens des Massenproduktes**



**Abbildung 6-17: Vergleich der Methoden zur Abbildung des Einschleusverhaltens des Sonderproduktes**

erheblicher Fehler in der Vorhersage mit  $\overline{TDev_{R\_to\_G}} = 359,05 [min]$  und  $\overline{TDev_{R\_to\_AFM}} = 208,31 [min]$ .

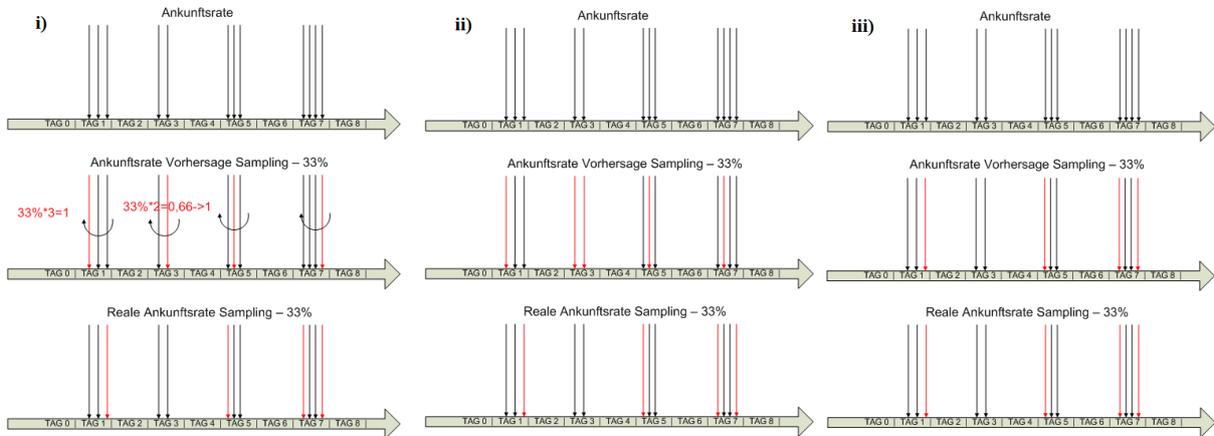
Zusammenfassend ist jedoch der Vergleich der mittleren Abweichung der Zeitunterschiede zwischen realen und ermittelten Zeitpunkten für alle Produkte relevant. Die für jedes Produkt ermittelten Daten wurden gewichtet zu einem Wert komprimiert. Das Gewicht für jedes Produkt entsprach dem Verhältnis der eingeschleusten Anzahl der Wafer zu der Gesamtwaferanzahl pro Untersuchungsperiode. Die Auswertung der Stichprobenmenge ergab für die AFM  $\overline{TDev_{R\_to\_AFM}} = 19,83 [min]$  und für die Gleichverteilung  $\overline{TDev_{R\_to\_G}} = 38,90 [min]$ . Dieser Wert ist signifikant vom Produktmix abhängig. Die Vorhersage eines operativen Einschleusplan mithilfe eines taktischen Plans für die Halbleiterfabriken, die hauptsächlich Massenware fertigen, wie zum Beispiel Speicher-Fabriken, könnte auch die Methode der Gleichverteilung anwenden. Bei den Fertigungsstätten mit breitem Produktspektrum ist die Anwendung des intelligenten Ansatz der AFM bezüglich der Qualität der Resultate von Vorteil.

### 6.4.2 Bewertung von unterschiedlichen Ansätzen für die Prognose des Los-Samplings

Die Bedeutung des Los-Samplings wurde bereits in den Kapiteln 2.4.5 und 4.5.4 beschrieben. In den bekannten Vorhersagemethoden für die Halbleiterfertigung werden beim Vorausberechnen der Materialanlieferung an bestimmten, mit Sampling behafteten Operationen unterschiedliche Ansätze angewendet. Im Rahmen des Projektes wurden verschiedeneartete Berechnungsalgorithmen miteinander verglichen sowie die mit deren Hilfe erstellten Prognosen gegenüber der Realität validiert.

Das einfachste Verfahren ist das meist verbreitete und basiert auf der Verwendung von Zufallszahlen für die Bestimmung der Art des Vorgehens bei der Zieloperation. Die Methodik wird u. a. auch in der diskreten ereignisbasierten Simulation angewendet [Fishman01]. Die Samplingrate wird in einen normierten Zahlenbereich zum Beispiel zwischen 0 und 1, umgewandelt (oft wird sie auch in dieser Form in den Basisdaten gespeichert). Anschließend wird eine Zufallszahl aus demselben Bereich erzeugt. Die prognostizierte Ausführung der Operation erfolgt nur, wenn die Zufallszahl kleiner als die normierte Samplingrate ist. Gewöhnlich bieten kommerzielle Simulatoren die Möglichkeit der Verwendung einer Verteilungsfunktion für die Berechnung. Ein zweiter Ansatz für die Samplingberechnung erfordert die quantitative Bestimmung des gesamten, prognostizierten Ankunftsstromes  $L_{Ges}$  des Materials an der Zieloperation. Angenommen, die Samplingrate beträgt  $\sigma_{All}$ , dann wird die tatsächliche An-

kunftsmenge  $L_{Ankunft}$  mit der Formel  $L_{Ankunft} = L_{Ges} - L_{Ges} * \sigma_{All}$  berechnet. Die AFM verwendet eine neuentwickelte Methode des Durchnummerierens der mithilfe der Ankunftszeit sortierten Ankunftsreihe (vgl. Kapitel 4.6.3.1).



**Abbildung 6-18: Beispiele für die Bestimmung von gesampelten Losen mit den Methoden: i) der Mengenkalkulation, ii) des Zufalls, iii) des Durchnummerierens (AFM).**

Der Vergleich der Methoden für die Bestimmung der gesampelten Lose ergibt vorerst die erwartete Differenz in der Komplexität und der Geschwindigkeit der Berechnung. Die Methode der Zufallszahlen konsumiert ein Minimum an Ressourcen. Im Gegensatz zu den beiden anderen Ansätzen erfordert sie auch keine vollständige Vorausberechnung des Prognoseresultates. Der wesentliche Nachteil bei der Mengenkalkulation resultiert aus der Behandlung der unvollständigen Lose. Bei einer hypothetischen Anlieferung von 2 Losen mit 25 Wafer und 4 Losen mit 12 Wafer bei einer Samplingrate von 33 % werden normalerweise nur zwei Werkstücke abgearbeitet, deren Quantität aber unbestimmt bleibt. Der Versuch, die Mengenberechnung nicht über die Anzahl der Lose sondern über Wafer zu gestalten, führt zu einer noch deutlicheren Verfälschung der Ergebnisse. Die Methoden unterscheiden sich auch bezüglich des Informationsbedarfs bei der Initialisierung. Während das auf den Zufallszahlen basierende und das mit Mengen arbeitende Verfahren keine zusätzlichen Daten für die Berechnung benötigen, ist die AFM-Variante auf detaillierte Initialisierungsinformationen angewiesen. Abgesehen von den methodischen Diskrepanzen stehen jedoch die quantitativen Unterschiede bezüglich der Vorhersagequalität im Fokus des Vergleichs. Die Basis für die Validierung der Methoden stellte eine aus der Realität extrahierte Abfolge der Materialbewegung über einen Zeitraum von 150 Tagen dar. Der Zeitraum wurde in mehrere 14-tägige Perioden unterteilt, für die eine Prognose berechnet wurde. Die Voraussagen umfassten allerdings nur ausgesuchte Produkte, die sich deutlich im Hinblick auf die Anzahl der Vertreter in der Produktionslinie sowie auf die Samplingrate voneinander unterschieden. Die Rate blieb je-

doch über den gesamten Analysezeitraum pro Produkt konstant. Durch die einmalige Erzeugung der Losankunftsströme, mit dem für AFM typischen Prognosefehler, wurde der Einfluss der Kalkulation der prognostizierten Ankünfte an der Zieloperation auf die Untersuchung ausgeschlossen. Jede Berechnungsmethode verwendete die gleichen Ankunftszeitpunkte. Die vorausgesagten Anlieferungsmengen an spezifizierten Zieloperationen, berechnet mit jedem der drei Algorithmen, wurden mit den realen Ankunftsströmen verglichen. Die Ermittlung der Vorhersagefehler fängt mit der Bestimmung der prognostizierten  $v_f$ , der realen Anlieferung  $v_r$  pro Tag, Produkt und Operation an. Für die Berechnungen der Prognosefehler bzw. der Treffsicherheit wurde der im Kapitel 3.1.2 vorgestellte mittlere absolute prozentuale Fehler *MAPE* modifiziert und wie folgt angewendet:

$$MAPE_{Sampl/day} = \frac{1}{n} \sum_{i=1}^n |PE(i)| \begin{cases} PE(i) = \frac{v_r(i) - v_f(i)}{v_r(i)} * 100 \Rightarrow |v_r(i) - v_f(i)| < v_r(i) \\ PE(i) = 0 \quad \Rightarrow |v_r(i) - v_f(i)| \geq v_r(i) \end{cases},$$

wobei mit  $n$  die Anzahl der möglichen Produkt/Operation-Kombinationen  $i$  bezeichnet wird. Die Modifikation ist notwendig für die Behandlung des Falls, weil wenn der absolute Fehler einen größeren Wert annimmt als die reale Anlieferung, die Vorhersage als unbrauchbar eingestuft wird. Abbildung 6-19 stellt tagfein die Treffsicherheit der Prognose, berechnet mit der Formel  $\vartheta = 100 - MAPE_{Sampl/day}$ , für das gesamte analysierte Produktspektrum dar. Bei der Methode des Durchnummerierens liegt die Trefferrate bei den ersten 7 Tagen über ca. 90 %. Die beiden anderen Ansätze sind schlechter als AFM-Ansatz bezüglich der Qualität der Vorhersageresultate. Bei den Prognosen für die zweite Woche beeinflussen die allgemeinen Prognosefehler die Qualität deutlicher. Aus diesem Grund fallen die Differenzen zwischen den Methoden wesentlich geringer aus. Interessanter sind aber die produktfeinen Vergleiche. Abbildung 6-20 stellt die Kurven der Prognosequalität für ein typisches Massenprodukt dar. Die Divergenz der einzelnen Verläufe ist marginal. Somit spielt die Wahl der Methode bei der Massenware eine untergeordnete Rolle. Unter den Kleinserienprodukten werden die Produkte eingegliedert, deren Anteil in der Produktion gering ist und deren Ankunft an der Zieloperation seltener stattfindet. Der MAPE bei derartigen Artikeln ist deutlich höher als bei Massenprodukten. Da das Verhältnis der Losanzahl pro Tag niedrig ist, wird die Prognose durch einzelne Lose deutlich verfälscht. Die in Abbildung 6-21 dargestellten Verläufe der Kurven mit der Prognosequalität verdeutlichen die allgemeine Verringerung der Prognosegüte, die mithilfe der Mengenermittlung- und Zufallszahlenmethode erstellt wurde. Das bei der AFM angewendete Verfahren schließt die Analyse auch für die Kleinserienprodukte am erfolgreichsten ab. Resümierend lässt sich an dieser Stelle feststellen, dass keine universale

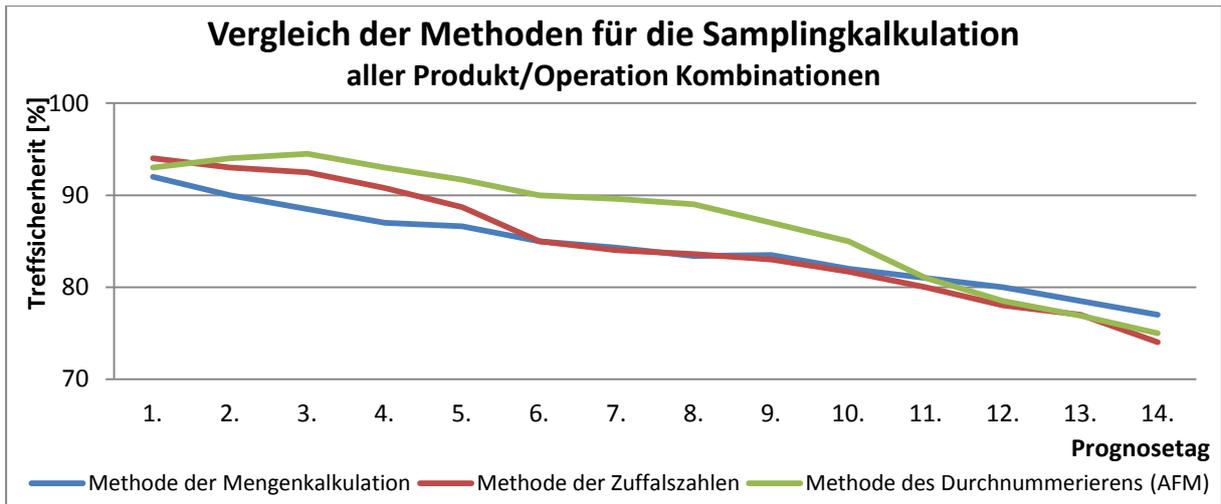


Abbildung 6-19: Trefferherheit der Prognose bei unterschiedlichen Algorithmen für die Samplingkalkulation für alle Produkte

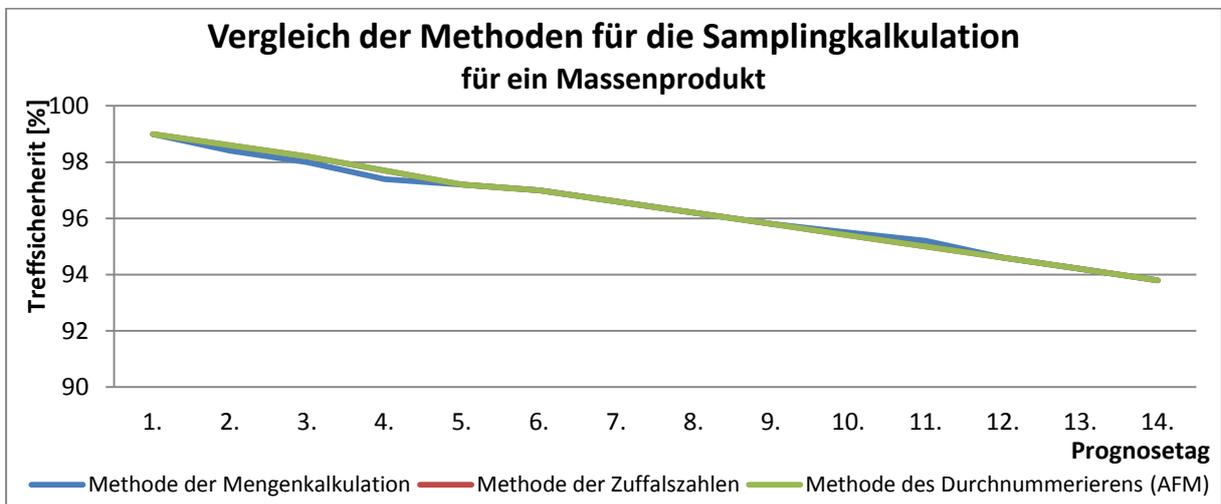


Abbildung 6-20: Trefferherheit der Prognose bei unterschiedlichen Algorithmen für die Samplingkalkulation für alle Produkte

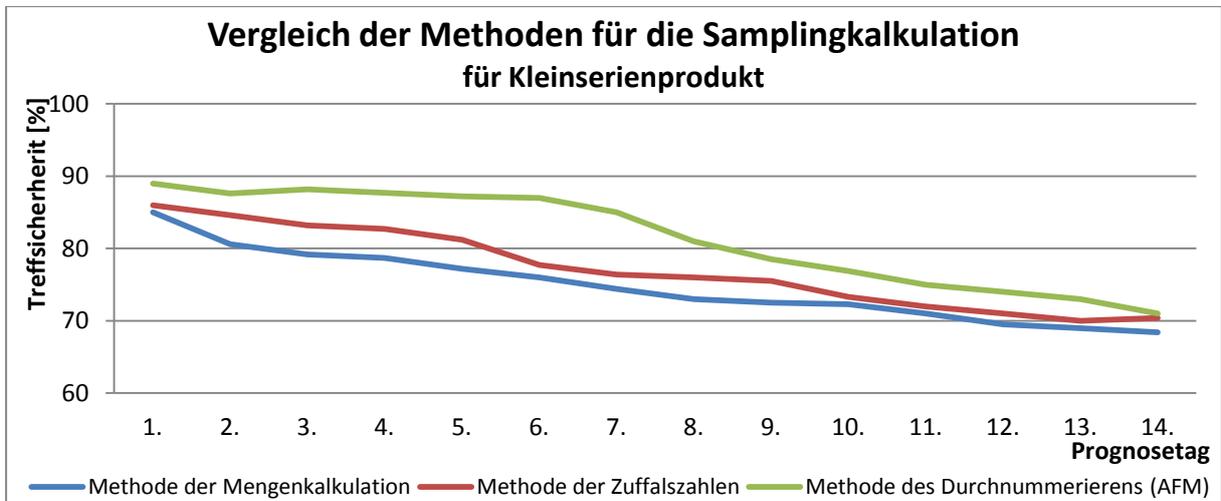


Abbildung 6-21: Trefferherheit der Prognose bei unterschiedlichen Algorithmen für die Samplingkalkulation für alle Produkte

Lösung für alle Halbleiterfertigungsstätten existiert. Die Wahl der Methode hängt eng mit dem Produktmix und der Art der Produktion zusammen. Die Anforderungen für die Prognosesoftware selbst wie auch Prognoseziele sollen bei der Entscheidung über das Samplingverfahren als richtungsweisend betrachtet werden.

### 6.4.3 Analyse des Einflusses von Hold-Losen auf die Vorhersagequalität

Die Dauer eines Anhaltezustands für jedes Los hängt von mehreren Faktoren, wie Produktzugehörigkeit, Zustand der Produktionsanlagen, Verfügbarkeit von geschultem Personal und vielen weiteren, ab. Aus der Sicht der Vorhersagequalität ist die Quantifizierung des Fehlers interessant, der durch vollständiges Vernachlässigen von Hold-Losen entstehen würde. Abbildung 6-22 stellt die Abhängigkeit zwischen der Hold-Dauer von in der Fabrik prozessierten Losen und deren regulären Durchlaufzeiten inkl. Hold dar. Um die statistische Relevanz zu gewährleisten, wurde eine längere Zeitperiode ausgesucht. Die gravierenden Zwischenfälle, die zu einem unnormalen Anstieg des Haltezustands geführt haben, wurden ausgefiltert. Mithilfe der roten Kurve wird nachgewiesen, dass für über 90 % der Lose der Hold-Anteil in der Durchlaufzeit nur marginal (unter 4 %) ist, und vernachlässigt werden kann. Bei über 99 % des Materials beträgt der Anteil des Anhaltezustands unter 10 % der gesamten Produktionszeit. Nach der Anwendung der in Kapitel 4.5.5 beschriebenen Filter und der Verbesserung der Datenqualität, sinkt der Einfluss des Hold auf die Prognosequalität noch deutlicher. Nach der näheren Analyse der Angaben liegt die Gedanke nahe, dass die Behandlung von Hold-Losen seitens des Prognosewerkzeuges nicht unbedingt notwendig ist.

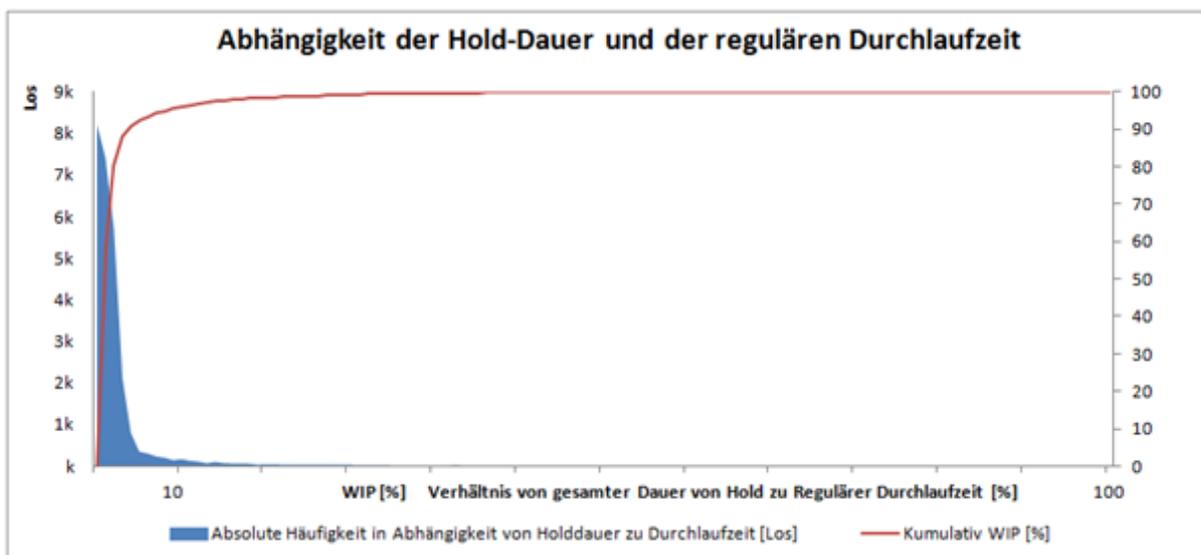
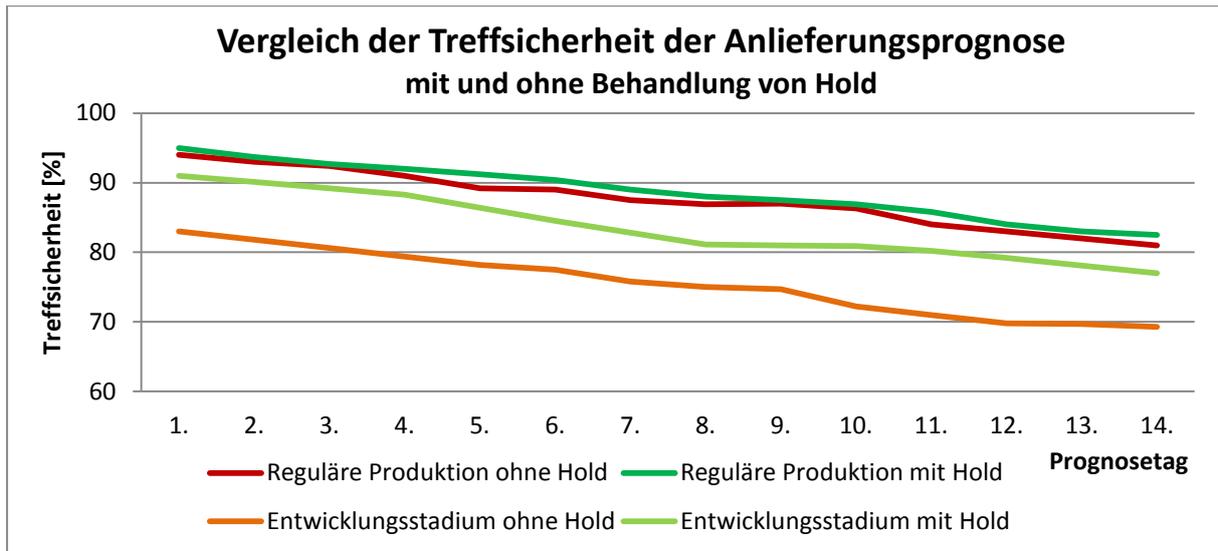


Abbildung 6-22: Die Abhängigkeit von der gesamten Hold-Dauer und der regulären Durchlaufzeit eines Loses über den vollständigen Bestand einer Fabrik in längerem Zeitabschnitt



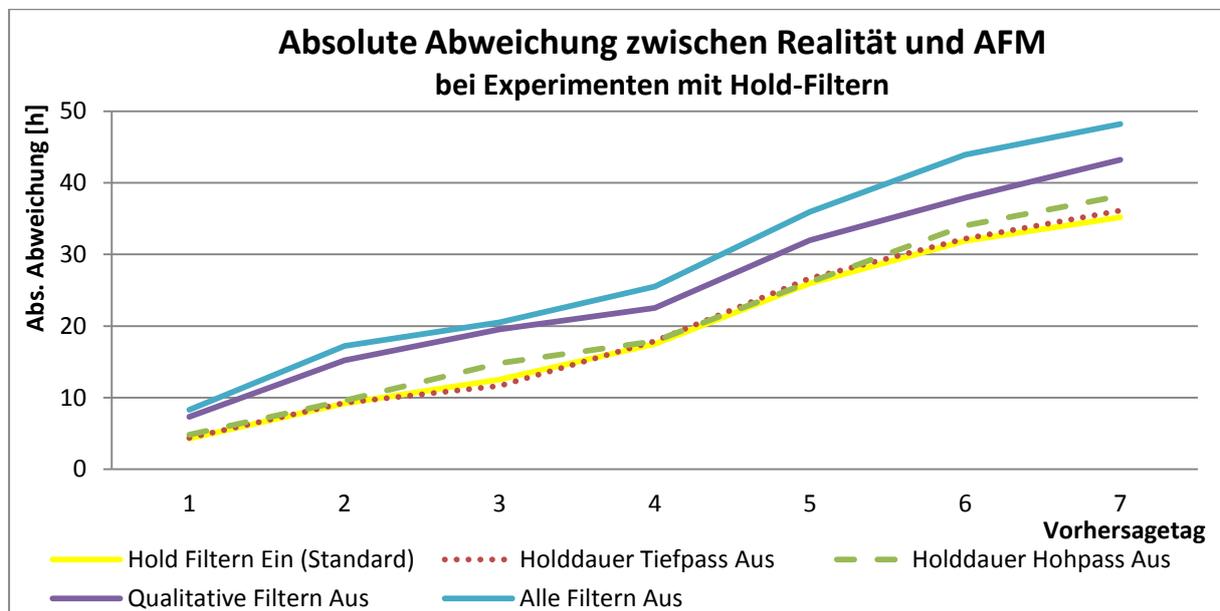
**Abbildung 6-23: Der Vergleich der Treffsicherheit der Anlieferungsprognose mit und ohne Behandlung des Hold-Zustands seitens AFM**

Um die Verschlechterung der Ergebnisse zu quantifizieren, wurden 100 unabhängige Prognosen mit und ohne das für die Hold-Behandlung zuständige Modul aufgestellt. Der Unterschied des mittleren absoluten prozentualen Fehlers MAPE für die Mengenanlieferung über alle Voraussagen beträgt 1,26 %. Im Schnitt sind also die AFM-Prognosen mit Hold-Prozeduren nur minimal treffsicherer. Die genauere Betrachtung der Problematik liefert den Nachweis der Brauchbarkeit und Notwendigkeit des Ansatzes für die Entwicklungsprodukte. Abbildung 6-23 stellt den Vergleich der Prognosequalität für die Anlieferungsprognose bei der Ausschleusungsoperation der Fabrik für ausgewählte Produkte aus der regulären Produktion sowie der Entwicklungsprodukte dar. Wie erwartet ist der Qualitätsunterschied der aufgestellten Prognosen für die etablierte Ware vernachlässigbar. Bei der Entwicklungsware beträgt der Unterschied im Qualitätsmaß MAPE ca. 6,32 % zugunsten der Hold-Behandlung. In der derzeitig implementierten AFM-Version wurde die Holdbehandlung nur für die Produkte eingeschaltet, deren Produktionsstart bei weniger als 9 Monaten liegt. Die Einstellung des Parameters ist frei wählbar im Bereich von 0 bis 5 Jahren, wobei bei 0 die Behandlung des Anhaltezustands gesperrt ist und bei 5 Jahren alle Produkte freigeschaltet sind. Da die Parametrisierung der AFM von vielen Eigenschaften der Fertigung abhängig ist, wird beim Einsatz der AFM in anderen Fertigungsstätten die Problematik des Einflusses von Hold auf die Durchlaufzeit detailliert untersucht werden müssen.

### 6.4.3.1 Analyse der Filteralgorithmen bei der Statistikberechnung für Hold

In Kapitel 4.5.5 wurde die Notwendigkeit der Einführung unterschiedlicher Filter für die Hold-Dauer und -rate angedeutet, die in einem umfangreichen zweiteiligen Validierungsexperiment nachgewiesen wurde.

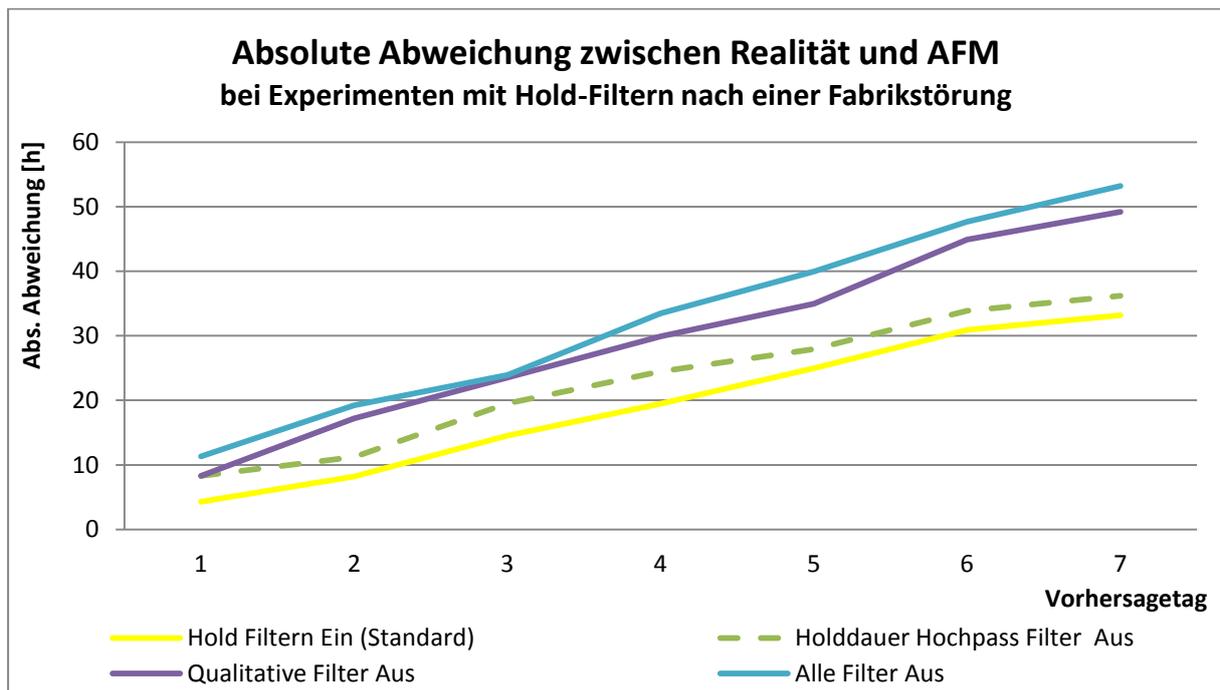
Im ersten Teil liegen, wie bei anderen Versuchen auch, 100 unabhängige Prognosen zugrunde, die jedoch über einen Zeitraum von über 4 Monaten ausgeführt wurden. Die entsprechenden Updates der Hold-Statistik wurden mithilfe des Basisdatenmaterial für je 14 Tage in der untersuchen Periode angefertigt. Dadurch wurde gewährleistet, dass jedes Voraussageexperiment über den zugehörigen statistischen Datensatz verfügt. Analysiert wurde fast ausschließlich die Abweichung der Ankunftszeit zwischen Realität und der AFM in verschiedener Konfiguration (Methodik aus Kapitel 6.1) der Prognosemethode bezüglich Hold Filters.



**Abbildung 6-24: Der Verlauf der absoluten Abweichung der Anlieferungszeit zwischen den AFM-Prognosen und der Realität bei Experimenten mit Hold-Filtern**

Abbildung 6-24 stellt die Resultate des Experimentes in der vereinfachten Form dar. Die Kurven charakterisieren den Anstieg der absoluten Abweichung zwischen den Ankunftszeitstempeln für die gleichen Lose aus der Prognose und der Realität für den jeweilig prognostizierten Tag. Der Vergleich der Kurven aus Abbildung 6-24 für die Vorhersagen mit eingeschalteten Filtern (gelb) sowie mit ausgeschalteten Hold-Dauer-Tiefpass-Filtern (rot gepunktet) ergibt nur einen vernachlässigbaren Unterschied zwischen den beiden. Aufgrund der minimalen Differenz wird eine derartige Filterung zur Verringerung des Datenvolumens verwendet. Das Ausschalten des Hochpass-Filters für die Hold-Dauer (grün gestrichelt) erzeugt die Verfälschung der statistisch erfassten Daten durch die Einzeleffekte. Da aber im Normalfall in der

Fertigung der ungewöhnlich lange andauernde Halt nur marginal häufig auftritt, ergibt auch der Vergleich der Abweichung der Ankunftszeitstempel nur eine minimale Divergenz zugunsten der Anwendung dieser Holdfilter. Eine wesentlich deutlichere Differenz entsteht durch das Ausschalten der qualitativen Filter (lila in Abbildung 6-24). Im normalen Betrieb der Fertigung entstehen von Zeit zu Zeit aufgrund des Auftretens unterschiedlicher Störungen temporäre Anhäufungen der Notwendigkeit zum Überführen mehrerer Lose in den Haltezustand. Derartige Situationen sind aber meistens nur kurz andauernd und stellen kein typisches Verhalten dar, verfälschen jedoch die statistische Auswertung enorm.



**Abbildung 6-25: Der Verlauf der absoluten Abweichung der Anlieferungszeit zwischen den AFM-Prognosen und der Realität bei Experimenten mit Hold-Filtern nach einer gravierenden Fabrikstörung**

Die AFM in Version 2.x arbeitete ohne die Hold-Filterung und lieferte dessen ungeachtet brauchbare Ergebnisse. Erst nach einem gravierenden Problem mit der Medienversorgung für die Produktionslinie und anschließendem Versetzen von über 30 % des Losbestandes in den Haltezustand wurde das Problem der Verfälschung der Statistik deutlich. Im zweiten Teil des Experiments werden auch 100 Prognosen über einen Zeitraum von 4 Monaten ausgewertet. Die untersuchte Periode beinhaltet eine auf der Realität basierende Fabrikstörung, die zu einem unnormalen Hold-Überschuss führt. Abbildung 6-25 stellt die Resultate des Experimentes für die beschriebene Problematik dar. Der Verlauf der Kurven zeigt eine absolute Abweichung zwischen den Ankunftszeitstempeln für das gleiche Los aus der Prognose und der Realität für den jeweilig prognostizierten Tag. Die durch die blaue Kurve dargestellte Abwei-

chung nach dem Abschalten jeglicher Art der Filterung ist nahezu 100 % größer als die durch die gelbe Kurve beschriebene Differenz, welche die Erstellung einer Prognose mit allen Hold-Filtern zeigt. Aus Sicht der Prognosequalität entsteht die größte Gefahr bei einem Auftreten der ungewöhnlichen Hold-Anhäufung auf einer Route kurz vor Ankunft der Lose an der Zieloperation. Nicht zu unterschätzen ist auch, dass, obwohl die Statistik in kurzen Perioden aktualisiert wird, die nicht gefilterten Daten die Prognosen über einen längeren Zeitraum verfälschen können (siehe Kapitel 4.5.5).

### 6.4.4 Analyse des Los-Samplings hinsichtlich der Durchlaufzeit

Die Möglichkeit für das Überspringen einer Operation wurde hauptsächlich aus wirtschaftlichen Gründen eingeführt. Die Maßnahme führt zur Verkürzung der Durchlaufzeit des Materials. Im Rahmen des AFM-Projektes wurde der Einfluss des Operations-Samplings auf die Prognosequalität untersucht. Angenommen zwischen der Startoperation  $S$  und der Zieloperation  $Z$  existieren  $n$  Produktionsschritte, deren Ausführung als optional gilt und  $m$  deren Abarbeitung zwingend erforderlich ist. Die Durchlaufzeit  $ct_{S \rightarrow Z}$  eines Loses ergibt sich aus der Summe einzelner Zeiten pro Operation, wobei an den übersprungenen Prozessschritten die Zeit 0 beträgt. Beim statistischen Erfassen des  $ct_{S \rightarrow Z}$  für die gleiche AFM-Klasse entsteht also eine durch das Sampling bedingte Durchlaufzeitschwankung. Analysiert wurden zwei Modellierungsansätze für dieses Problem:

- detailarm,
- detailreich.

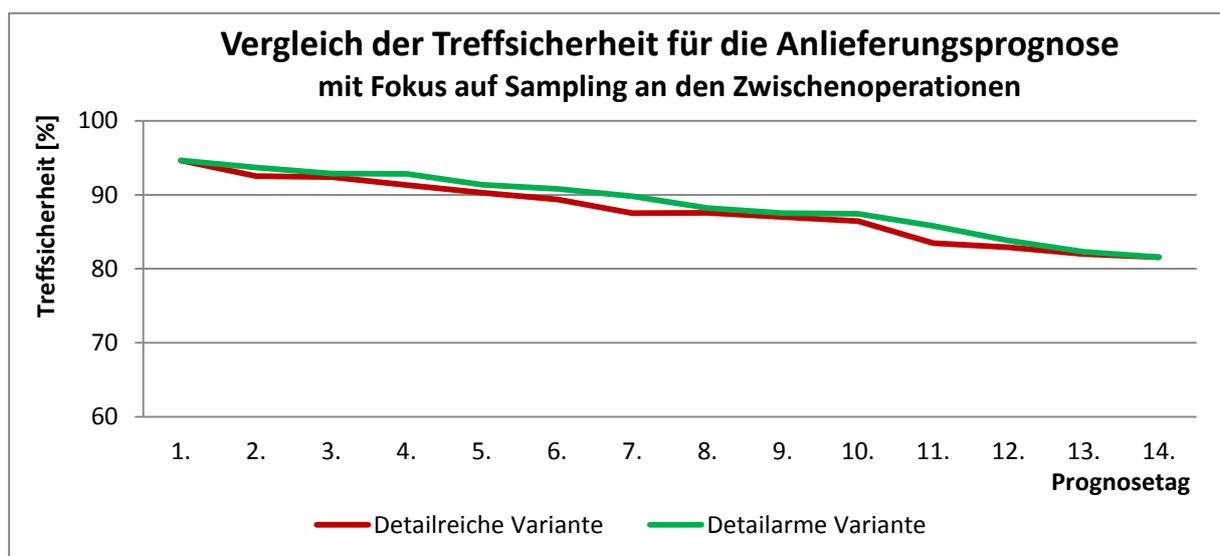


Abbildung 6-26: Der Vergleich der Treffsicherheit der Anlieferungsprognose mit dem detailreichen und -armen Ansatz für die Behandlung von Sampling an den Zwischenoperationen

Während bei der detailarmen Variante die Zeitunterschiede einzelner Werkstücke ohne Weiteres in die Durchlaufzeitstatistik eingebettet wurden, benötigt der detailreiche Ansatz zusätzliche statistische Informationen. Die Sampling-Statistik beinhaltet die prozentuale Abarbeitungsquote  $\beta_x$  pro Operation für jedes Produkt sowie die mittlere Operationsdurchlaufzeit  $ct_x$ . Bei der Berechnung der Durchlaufzeitstatistik werden die anfallenden Zeiten für den Aufenthalt in die mit Sampling behafteten Operationen von der Durchlaufzeit abgezogen. Dadurch wurden bereinigte Statistiken mit  $ct'_{s \rightarrow z}$  erzeugt. Die Berechnung der prognostizierten Ankunft erfolgt durch die Berechnung der gesamten Durchlaufzeit:

$$ct_{s \rightarrow z} = ct'_{s \rightarrow z} + \sum_{i=1}^n \begin{cases} RAND(0,1)_i > \beta_i \Rightarrow 0 \\ RAND(0,1)_i \leq \beta_i \Rightarrow ct_i \end{cases}$$

wobei  $RAND(0,1)_i$  eine Zufallszahl aus dem Bereich zwischen 0 und 1 mit den zugehörigen Samplingraten  $\beta_x$  verglichen wird. Abbildung 6-26 stellt das Ergebnis der Validierungsstudie mit Fokus auf das Sampling an den Zwischenoperationen dar. Das Beispiel zeigt deutlich, dass im Falle des Samplings die Steigerung des Detaillierungsgrades des Modells keine Verbesserung der Prognosegüte brachte. Da in der Qualität der Prognose ohne Sonderbehandlung des Sampling, also durch Einbettung der Störgröße in die Durchlaufzeitstatistik, bessere Ergebnisse erzielt wurden, wird in der derzeitigen Version die detailarme Variante angewendet.

## 7 Vergleich unterschiedlicher Prognosemethoden in der Halbleiterfertigung

Bei unserem Projektpartner aus der Industrie werden unterschiedliche Ansätze angewendet, um notwendige Prognosen für die Fertigung zu erstellen. Am meisten hat sich bereits die Lang- und Mittelfristsimulation etabliert, die zu einem unerlässlichen Werkzeug der Planung und des Krisenmanagements wurde. Die auf planerischen Abarbeitungszeiten basierenden Voraussagewerkzeuge wurden meistens trotz zum Teil enormer Fehlern akzeptiert, weil diese bereits mehrere Jahre im Einsatz sind und meistens effizient und schnell arbeiten. Die Kurzfristsimulation wurde parallel zu AFM im Rahmen desselben Projekts entwickelt. Die Entwicklung der auf der Simulation basierenden Prognose für den Nahzeitraum hat jedoch mindestens ein Fünffaches an Manpower als die AFM erfordert. Auch sonstige Projektkosten, wie: Hardware, Software, Expertenwissen etc. waren deutlich höher. Dank der Bereitstellung mehrerer Vorhersagewerkzeuge und weitgehender Unterstützung seitens des Halbleiterfertigungsunternehmens, wurden die folgenden Ansätze miteinander verglichen:

- AFM – Analytic Forecast Method,
- KFS – Kurzfristsimulation (Kapitel 3.2.4),
- LFS – Langfristsimulation (Kapitel 3.2.4),
- PBPA – Prognose basierend auf der planerischen Abarbeitungszeiten (Kapitel 3.3).

Die Ergebnisse des Vergleiches der unterschiedlichen Voraussagemethoden werden in Tabelle 7-1 dargestellt. Die Unterschiede im Bedarf an Basisdaten resultieren aus der Modellierung. Da das Modell der Kurzfristsimulation sehr detailliert ist, benötigt es wesentlich mehr Informationen als AFM oder PBPA. Die LFS verwendet für die Abbildung der Fabrik unterschiedliche Approximationsansätze, weil im Hauptfokus das Fabrikverhalten und nicht detaillierte Aussagen über einzelne Lastverbände (Maschinengruppen) oder Produktionsschritte stehen. Der Vorhersagehorizont der AFM und KFS sind nahezu gleich. Beide Instrumente liefern detaillierte Informationen für die ersten 7 Tage. Die Trendvoraussage für die zweite Woche wie in AFM ist wahrscheinlich auch mit KFS möglich. Die Ergebnisse der LFS liegen in Form von Wochenaussagen vor, jedoch erstreckt sich der Prognosehorizont über mehrere Wochen bis Jahre. Die PBPA liefert die Aussagen für einzelne Lose über einen sehr langen Zeitraum in der Zukunft. Als problematisch erweist sich jedoch die Qualität der Resultate.

## Vergleich unterschiedlicher Prognosemethoden in der Halbleiterfertigung

	<b>AFM</b>	<b>KFS</b>	<b>PBPA</b>	<b>LFS</b>
Basisdatenbedarf	Klein	Sehr hoch	Klein	Hoch
Vorhersagehorizont	< 7 Tage; Trends für 14 Tage	< 7 Tage	< 3 Monate	1 oder mehr Wochen
Zeitliche Granularität der Ergebnisse	Tageswerte	Tageswerte oder kleiner	Tageswerte	Wochenwerte
Berechnungszeit	< 2 Minuten	ca. 1 Stunde	Mehrere Minuten	12 Stunden für eine Jahres- prognose
Komplexität des Modells	Klein	Sehr Hoch	Klein	Hoch
Anwendungsfelder	Einige	Mehrere	Einige	Mehrere
Umfang der Resultate	Klein	Sehr Groß	Klein	Groß
Transferierbarkeit	Gegeben für ähnliche Fertigung	Abhängig von der Basisdatenlandschaft	Gegeben	Unmöglich
Wartungsaufwand	Minimal	Hoch	Mittel	Sehr Hoch

**Tabelle 7-1: Vergleich unterschiedlicher Ansätze für die Prognosen in der Halbleiterfertigung**

Approximierend korreliert die Erstellungszeit einer Prognose mit der Modellgröße multipliziert mit dem zeitlichen Horizont. Bei dem Vergleich gewinnt die AFM mit einer Rechenzeit unter zwei Minuten. Da im Normalfall die AFM aber nur einen Teil der Fertigung abdeckt, ist für die Vorhersage über die ganze Fabrik eine Berechnungszeit von ca. 30 Minuten nötig. Die KFS benötigt ca. 1 Stunde für die Bereitstellung der Ergebnisse. Bei der Langfristsimulation wird die Kalkulation mehrere Male gestartet, um die stochastischen Effekte darzustellen. Aus diesem Grund dauert die Berechnung über 12 Stunden. Die beim Industriepartner angewendete Software für PBPA verwendete keine modernen Programmierungstechniken und benötigt für die sehr einfache Rechnung ca. 6 Minuten für eine 28 Tage Prognose. Den Vergleich der analysierten Prognosemethoden bezüglich der Breite der Anwendungsfelder gewinnen eindeutig die Simulationsansätze. Deren Spektrum ist aber sehr verschieden, da die Kurzfristsimulation nur den Nahzeitraum abdeckt, jedoch detailreiche Informationen über Maschinengruppen oder Produktionsschritte liefert. Die Anwendungspalette von AFM und KFS ist ähnlich. Die KFS liefert allerdings qualitativ bessere Prognosen bei jedweder Art von Störungen in der Fabrik, die von der AFM nicht abgedeckt werden. Der Vorteil der AFM liegt im hohen Automatisierungsgrad und geringen Wartungsaufwand. Einmal installiert arbeitet die AFM störungsfrei ohne die Notwendigkeit der kontinuierlichen Überwachung. Bei der KFS wurden fast alle Prozesse automatisiert. Die Vielfalt an Datenquellen sowie die detailreiche Abbildung der Fertigung erfordern aber eine permanente Kontrolle des Modells. Da das LFS-Modell größtenteils manuell entwickelt wurde, ist der Wartungsauffand als sehr hoch einzu-

stufen. Der Automatisierungsgrad der Datenbeschaffung von PBPA ist zwar hoch, es existieren aber sehr viele notwendige Informationen, die von dem Anwender manuell überwacht werden müssen, wie zum Beispiel der momentane Flussfaktor der Fabrik. Das Reporting der Ergebnisse der LFS ist aufgrund des breiten Spektrums der zu beantwortenden Fragestellungen nur bedingt automatisierbar.

Der interessanteste Punkt des Vergleichs wurde jedoch nicht in der Tabelle erfasst. Es ist die durchgeführte Analyse der Prognosequalität der einzelnen Werkzeuge. Da sich nicht alle Resultate miteinander vergleichen lassen, werden in der Ausarbeitung nur ausgewählte Ergebnisse vorgestellt. Die Analyse der Zeitabweichung der Prognosemethoden zur Realität für die einzelnen Lose (vgl. Kapitel 6.1) mithilfe von MAD für AFM, PBPA und KFS wird in Abbildung 7-1 dargestellt. Der Vergleich mit LZS ist aufgrund der Granularität der Ergebnisse an der Stelle unmöglich. Der PBPA Fehler ist in der gesamten Untersuchungsperiode ca. 2-mal größer als bei der AFM. Der Unterschied der Realität zur Prognose betrug ca. 50 % der

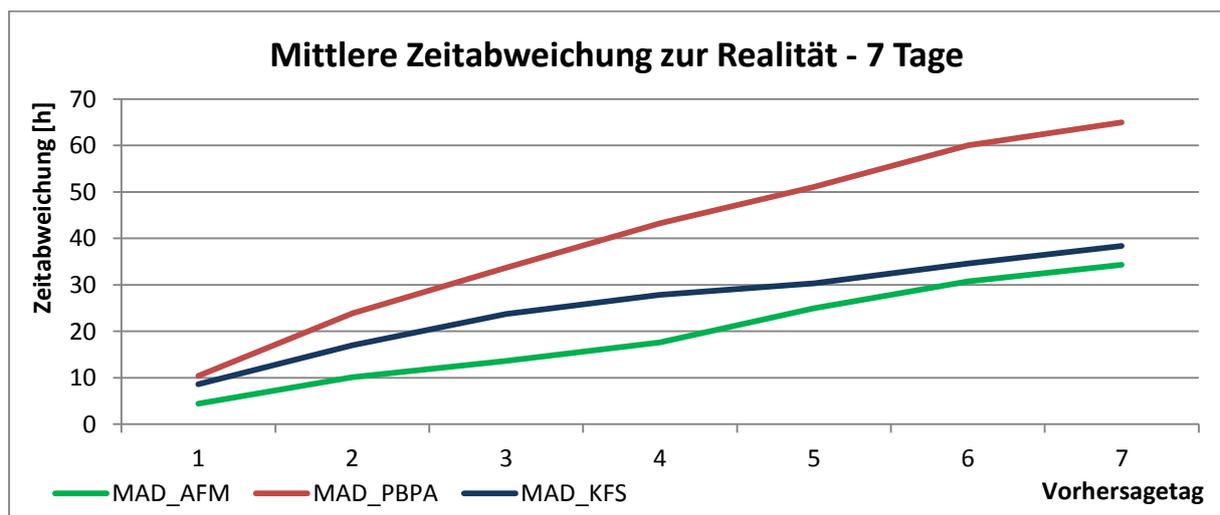


Abbildung 7-1: Der Vergleich der mittleren Zeitabweichung für die AFM, PBPA und KFS

Durchlaufzeit von der Start- bis zur Zieloperation. Die von AFM und KFS erzielten Ergebnisse sind vergleichbar, mit leicht besserer Qualität der nicht simulativen Methode. Die Simulation befand sich zur Zeit der Analyse aber in dem Validierungs- und Erweiterungsstadium. Das Entwicklerteam der KFS erhoffte sich eine deutliche Verbesserung der Prognosequalität in den ersten Tagen durch eine möglichst weitgehende Abkoppelung der stochastischen Einflüsse und der Verbesserung der Initialisierung. Die Abbildungen Abbildung 7-2, Abbildung 7-3 und Abbildung 7-4 stellen die Treffsicherheit der Anlieferungsprognosen für AFM, PBPA wie auch KFS dar. Die Wahrscheinlichkeit wird durch die Subtraktion des Median der MAPE einzelner Voraussagen von der Grundgesamtheit also von 100 % berechnet.

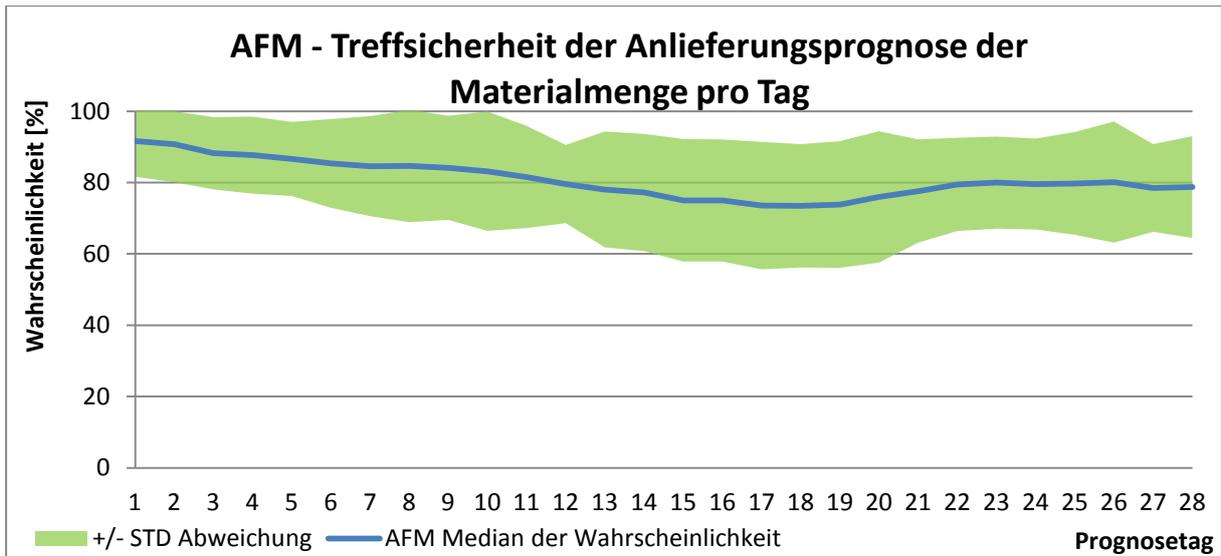


Abbildung 7-2: AFM - Treffsicherheit der Anlieferungsprognose der Materialmenge pro Tag

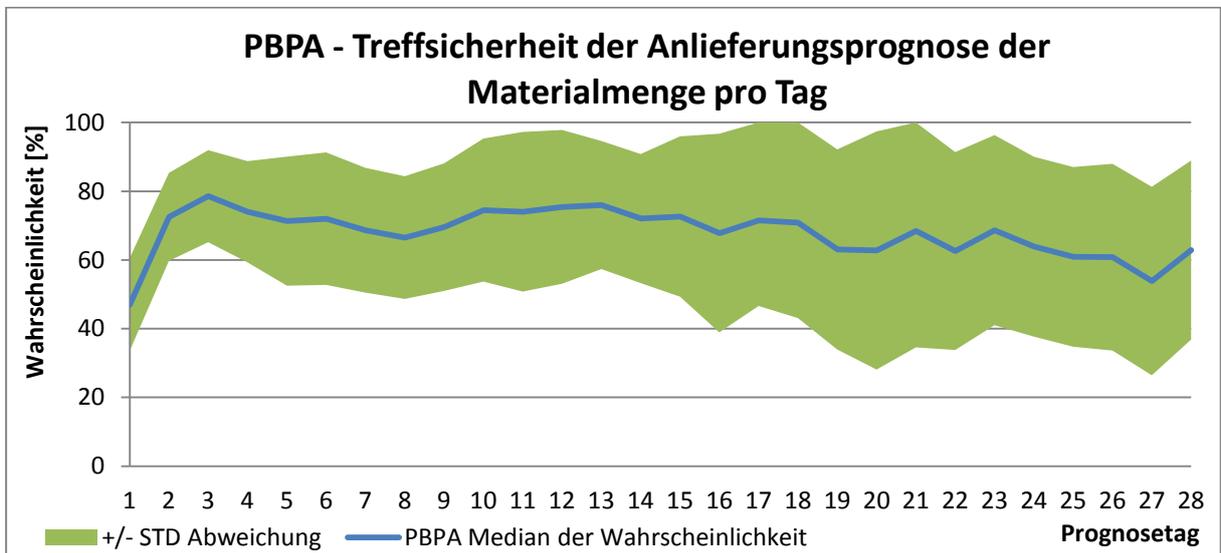


Abbildung 7-3: PBPA - Treffsicherheit der Anlieferungsprognose der Materialmenge pro Tag

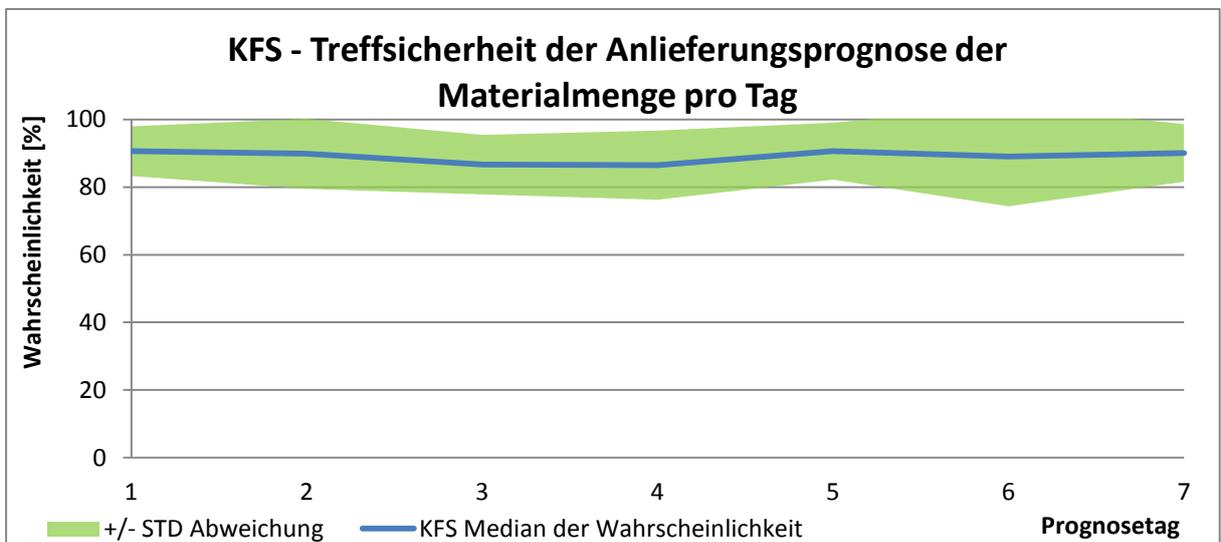


Abbildung 7-4: KFS - Treffsicherheit der Anlieferungsprognose der Materialmenge pro Tag

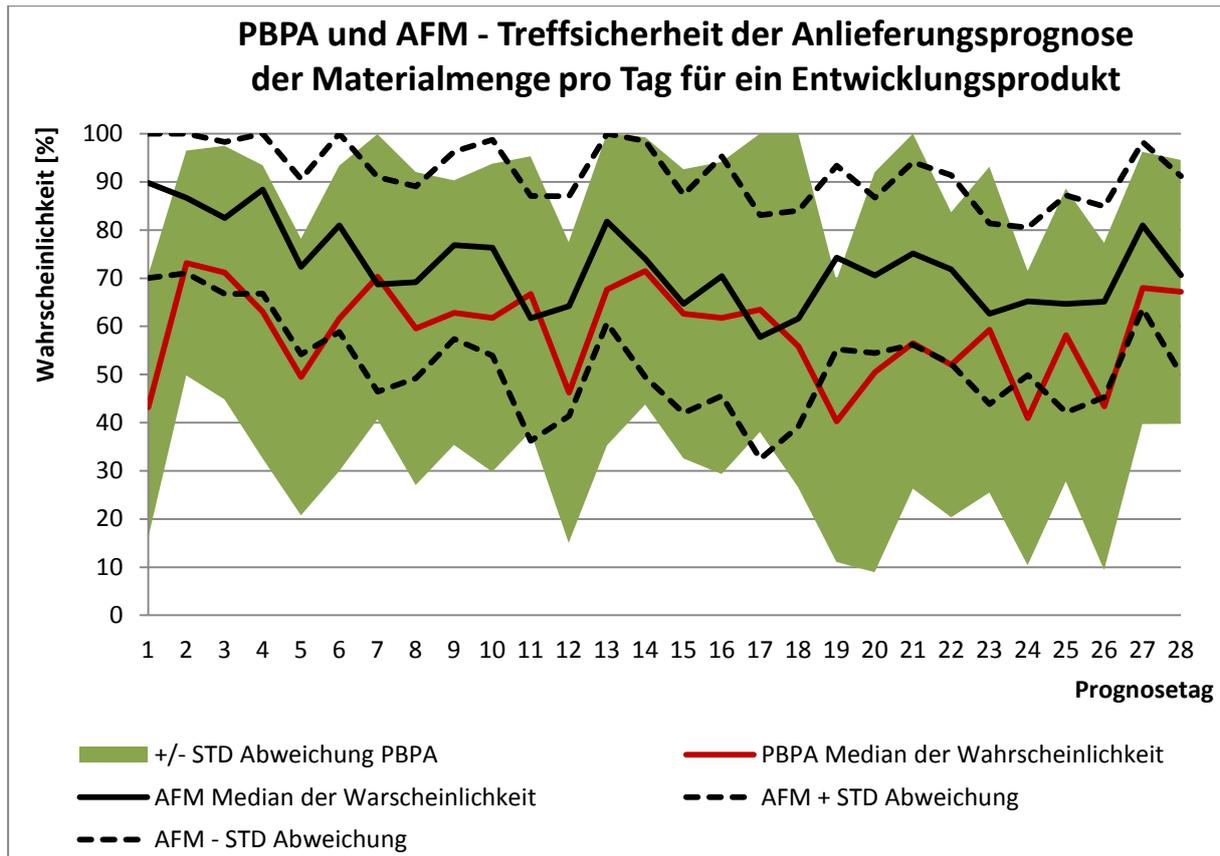


Abbildung 7-5: PBPA und AFM-Treffsicherheit der Anlieferungsprognose der Materialmenge pro Tag für ein Entwicklungsprodukt

Die in den Diagrammen abgebildete +/-Standardabweichung dient zur Veranschaulichung der möglichen Schwankungen in der Qualität der erzielten Resultate. In Abbildung 7-3 fällt die anormal niedrige Treffsicherheit bei den ersten drei Tagen auf. Der Grund dafür liegt in der Initialisierung der Kalkulation von PBPA. Weiterhin erzielt die Methode die niedrigste Treffsicherheit aller analysierten Verfahren. Ein zusätzliches Merkmal gegen die Verwendung der auf Planungsdaten basierenden Prognosemethode ist die Instabilität der Ergebnisse. Deutlich wurde dies bei der Analyse der Treffsicherheit für Entwicklungsprodukte, die in Abbildung 7-5 dargestellt ist. Da der Vorhersagehorizont der KFS nur 7 Tage beträgt, lässt sich nur dieser Zeitraum untersuchen. Die Wahrscheinlichkeit einer passenden Prognose liegt sowohl bei der AFM wie auch bei der KFS bei ca. 90 % für die erste Voraussageweche. Der Unterschied beider Methoden liegt in der Stabilität. Von Vorteil ist diesbezüglich die simulative Methode, die durch den Zugriff auf zusätzliche Datenquellen auch für die Prognosen bei ungewöhnlichem Fabrikverhalten ausgelegt ist.

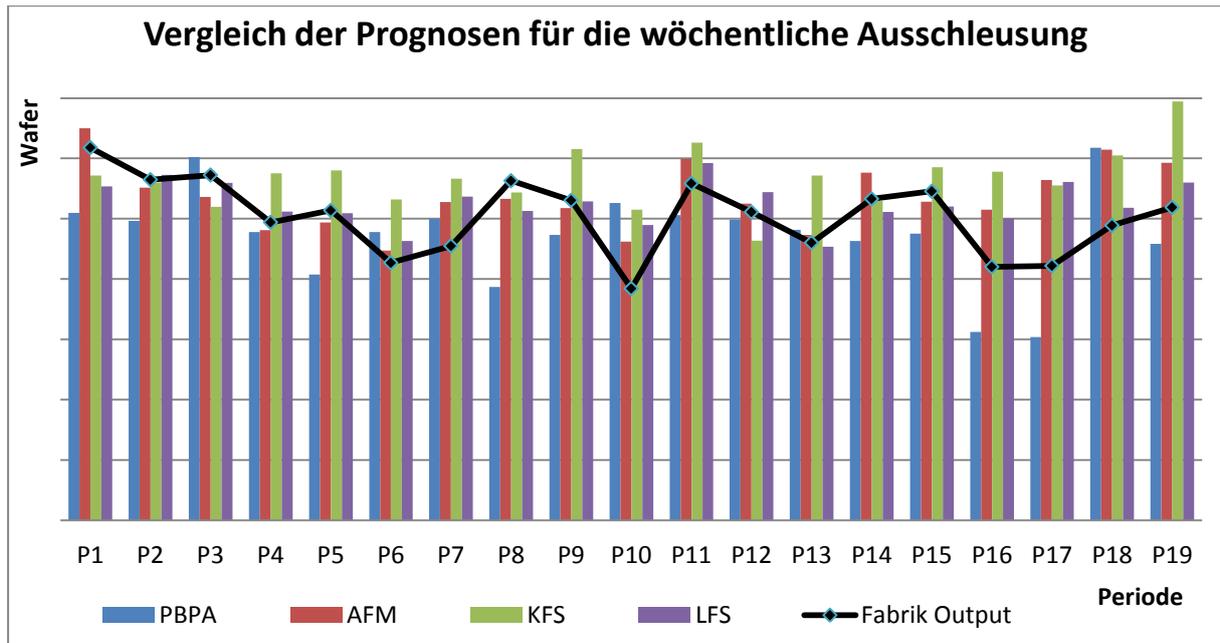


Abbildung 7-6: Vergleich der Prognosen für die wöchentliche Ausschleusung des Materials von der Fabrik

$f(P_{Met_{week}})$	AFM	KFS	PBPA	LFS
Mittelwert	89,67 %	83,35 %	85,40 %	91,70 %
Standardabweichung	8,70 %	11,31 %	9,14 %	8,49 %
Median	93,14 %	85,52 %	87,91 %	93,98 %

Tabelle 7-2: Die Quantifizierung des Vergleiches von Treffsicherheit der Prognosen für die wöchentliche Ausschleusung des Materials von der Fabrik

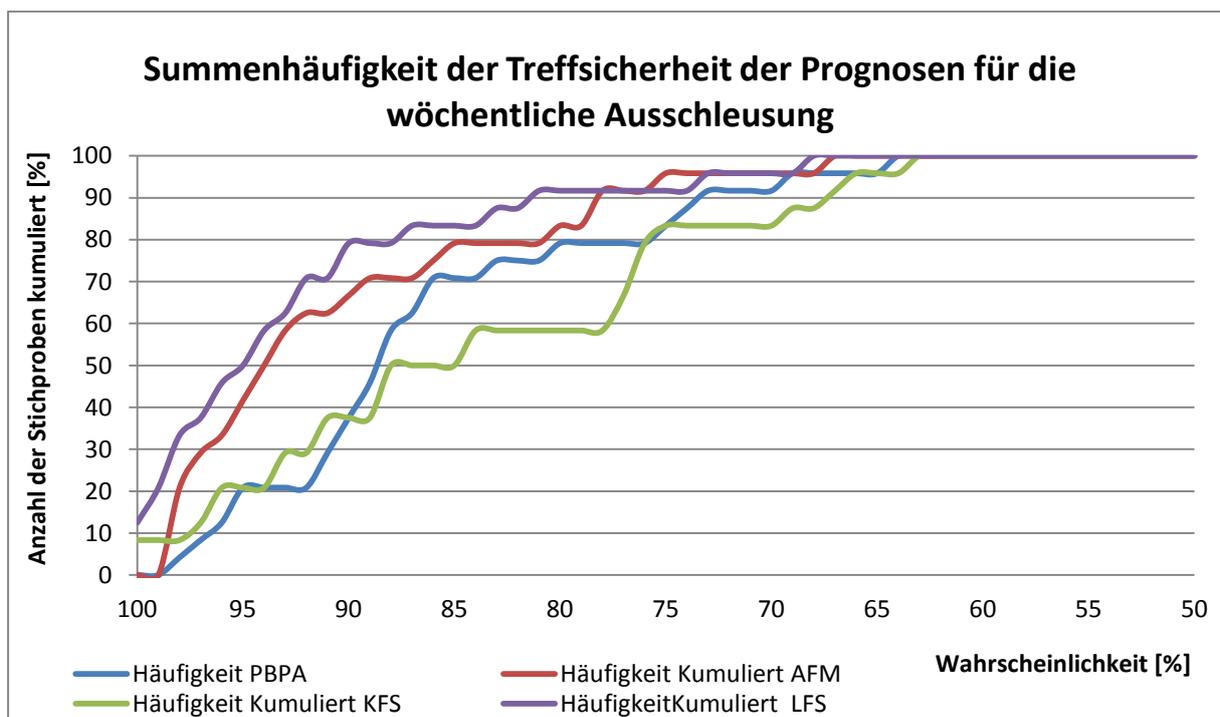


Abbildung 7-7: Summenhäufigkeit der Treffsicherheit der Prognosen für die wöchentliche Ausschleusung

## Vergleich unterschiedlicher Prognosemethoden in der Halbleiterfertigung

---

Der Vergleich aller Prognosemethoden erforderte die Entwicklung eines gesonderten Validierungsverfahrens. Da die Granularität der Aussagen der LFS keinen Tagesvergleich ermöglicht, wurden die Ergebnisse der sonstigen Verfahren zu einem Wochenwert zusammengefasst. Untersucht wurde die prognostizierte Ausschleusung der Fabrik für unabhängige Wochen. Das bedeutet, es wurden an einem bestimmten Wochentag die Voraussagen für die gesamte Woche erstellt und anschließend verglichen. Die Analyse umfasste einen fast 5 monatigen Zeitraum. Die Ergebnisse der Analyse wurden in Abbildung 7-6 dargestellt. Die Treffsicherheit  $P_{Met\_week}$  der einzelnen Vorhersagemethoden wurde für jede Wochenaussage mit folgender Formel quantifiziert:

$$P_{Met\_week} = \frac{Out_{real} - |Out_{real} - Out_{Met\_week}|}{Out_{real}} * 100 \% ,$$

wobei  $Out_{real}$  die reale und  $Out_{Met\_week}$  die prognostizierte dargestellt bedeuten. Die Resultate des Vergleichs aller Prognosen wurde in Tabelle 7-2 veranschaulicht. Wie erwartet liefert die LFS die treffsichersten Ergebnisse. Die Abstriche in der Qualität der Resultate seitens AFM resultieren aus dem detailärmeren Modell im Vergleich zu den simulativen Verfahren. Überraschend sind die sehr schlechten Prognosen der KFS. Die Begründung dafür liegt in der unzureichenden Validierung des Modells bezüglich dieses Anwendungsfeldes. Interessante Informationen über den Vergleich liefert das Diagramm über die kumulierte Häufigkeit der Treffsicherheit der Prognosen für die wöchentliche Ausschleusung, dargestellt in Abbildung 7-7. Wenn bei LFS ca. 70 % und bei AFM ca. 62 % der Prognosen mit einem maximalen Fehler von 10 % behaftet waren, waren es bei PBPA und KFS nur ca. 29 % und 37 % aller Stichproben. 90 % aller Stichproben sind mit einer Abweichung von maximal:

- 20 % für die LFS,
- 22 % für die AFM,
- 28 % für die PBPA,
- 33 % für die KFS

behaftet.

Resümierend lässt sich feststellen, dass keine der vorgestellten Prognosen alle Anwendungsfelder mit ausreichender Qualität abdeckt. Die Wahl der Prognosemethode ist also in erster Linie vom Anwendungsfeld abhängig. Die weiteren Vor- und Nachteile basieren hauptsächlich auf der wirtschaftlichen Bewertung der zukünftig eingesetzten Vorhersage. In der Prüfung werden die Kosten für Einkauf, Installation und Wartung gegenüber den Vorteilen der Benutzung des Verfahrens berechnet. Schließlich ist auch die Prognosequalität ein wichtiger

## **Vergleich unterschiedlicher Prognosemethoden in der Halbleiterfertigung**

---

Aspekt, weil nicht jede Anwendung einer Voraussage sehr detaillierte Aussagen verarbeiten kann und dadurch auch benötigt.

## 8 Zusammenfassung und Ausblick

Die Prognose der Materialanlieferung in einer Halbleiterfertigung findet sehr viele Anwendungsfelder in der Praxis. Die Erstellung einer validen Voraussagemethode ist durch die Spezifika der komplexen Fertigung geprägt und gestaltet sich als sehr kompliziert. Das meist verbreitete und mächtigste Werkzeug, das für diese Problematik eingesetzt wird, ist die Simulation. Oft jedoch übersteigen Aufwand und Kosten für die Modellierung der Fertigung und die anschließende, kontinuierliche Wartung des Simulationsmodells den praktischen Nutzen der Vorhersagen. An der Stelle – vor allem in den Bereichen kurz und mittelfristiger Prognosen – wird nach einer Alternative gesucht. Das in der Ausarbeitung präsentierte Werkzeug Analytic Forecast Method (AFM) wurde in diesem Kontext entworfen und entwickelt. Die Prognosemethode versucht, die Nachteile der Simulation zu umgehen. Einerseits erfolgt dies über die Anwendung eines neuen und zur Simulation verschiedenen Modellierungsansatz. Andererseits müssen aber auch kleine Einbußen in der Qualität der Resultate akzeptiert werden. Die AFM arbeitet praktisch kontinuierlich ohne Bedarf an Wartung. Dabei ist die Methode sehr robust im Umgang mit jeglichen Störungen, die im täglichen Fertigungsablauf passieren, sowie bei datentechnische Problemen.

Das Konzept der Prognosemethode, das in Kapitel 4 beschrieben wurde, basiert auf einer mehrstufigen Klassifizierung der Werkstücke mithilfe von Parametern, welche die Durchlaufzeit über die Fertigung beeinflussen. Dabei benötigt die AFM praktisch ausschließlich die historischen Angaben zu den Bewegungsabläufen des Materials in der Fertigung als Eingangsdatenstrom und extrahiert daraus alle für die Erstellung der Prognose notwendigen Informationen. Sie werden in Form von Statistiken in der Datenbank gespeichert und periodisch aufgefrischt. Dadurch wird die Aktualität der Wissensbasis gewährleistet. Die Verwendung einer Vielzahl an Filteralgorithmen verhindert die Grunddatenverfälschung durch unvorhersehbare Anomalien und Ereignisse in der Fertigung. Zusätzlich verwendet die AFM intelligente Konzepte für die Berechnung einer Prognose, wie das Durchnummerieren der mit Sampling behafteten Lose oder die Kalkulation der Materialeinschleusung aus der taktischen Planung. In Kapitel 5 wurde die Realisierung der Prognosemethode in einer Software-Anwendung beschrieben. Die Implementierung der AFM erfolgte auf einer Oracle Datenbank als Plattform. Zudem wurde das Implementierungskonzept um die Visualisierung mittels Webtechniken erweitert.

Die eindeutige Antwort auf die Frage nach der Qualität einer Prognose gestaltet sich sehr schwierig. Der Leiter der Kompetenzteam „Simulation“ bei einem der führenden Halbleiterhersteller, Wolfgang Scholl sagte bei einem Vortrag während Winter Simulation Conference 2010: *„Die Prognose ist erst dann gut und damit von ausreichender Qualität, wenn sich ein Anwendergremium findet, das mit der Prognose arbeitet.“* (Vom Autor frei übersetzt). Der Erfolg der AFM liegt also vor allem in der Weiterentwicklung der prototypischen Lösung, die für die Forschungszwecke erstellt wurde. Mittlerweile wurde eine stabile Version implementiert. Die Software wird bereits mit Erfolg und großer Akzeptanz in der Industrie angewendet. Die in den Kapiteln 0 und 7 durchgeführten Validierungsexperimente bestätigen das hohe Maß der Qualität der erzeugten Prognosen. Vor allem die hohe Treffsicherheit in den ersten 7 Tagen ermöglicht tagfeine Voraussagen der Materialanlieferung. Für die nächsten Zeitperioden ist jedoch die Verdichtung der Informationen zu Wochenwerten empfehlenswert. Aus dem Vergleich mit anderen Prognosemethoden, einschließlich der Simulation, resultiert keine eindeutige Empfehlung für die Wahl der Prognosemethode. Die Entscheidung sollte möglichst auf der Art der Anwendung und des zur Verfügung stehenden Budgets basieren. Wenn man jedoch das Verhältnis der Qualität der Resultate zu den anfallenden Installations- und Instandhaltungskosten betrachtet, zeigt sich die AFM als Preis-Leistungs-Sieger des Vergleiches.

Perspektivisch werden die unterschiedlichen Maßnahmen zw. Verbesserung der Qualität der AFM-Prognosen getestet und implementiert. Das Potenzial der Erweiterung der Initialisierung um Restprozesszeit oder Position in der Warteliste ist vor allem aus Sicht der Verwendung von AFM im Ultrakurzfristbereich von Interesse. Die Erweiterung der Klassifizierung bzw. produktfeine Klassifizierungsmerkmale sollten auch genauer untersucht werden. Es erfordert jedoch die weitgehende Automatisierung der Clusteralgorithmen für die Klassensuche. Einer der wichtigsten Aspekte, welcher die Verbesserung der Qualität der Prognosen bringen würde, ist die Abhängigkeit zwischen dem Materialbestand zwischen Start- und Zieloperation und der Durchlaufzeit. Ein Konzept eines Moduls für die Verlängerung der notwendigen Zeit aufgrund der Auslastung der Bottleneck-Anlagen wurde in Kapitel 4.6.2 beschrieben. Zur Zeit der Anfertigung der Ausarbeitung befand es sich noch in einem experimentellen Stadium und wurde nicht validiert. Der produktive Einsatz dieses oder eines ähnlichen Moduls wird aber in naher Zukunft erfolgen.

Die möglichst weitgehende Konsolidierung und Vereinheitlichung der Softwarelösungen zwischen einzelnen Standorten eines Herstellers in der Halbleiterindustrie ist zurzeit ein wichtiges Thema mit Fokus auf die Verringerung der anfallenden Kosten. Daraus resultiert die

## **Zusammenfassung und Ausblick**

---

Möglichkeit der Erprobung der AFM an weiteren Fertigungsstandorten. Wahrscheinlich besteht dabei die Notwendigkeit der Änderungen in der Klassifikation der Lose, also die Anpassung an zum Teil andere Abfertigungsregeln. Der erfolgreiche Einsatz der Methode in einer anders gesteuerten Fertigung mit einem anderen Produktmix würde die Universalität des beschriebenen Vorhersageansatzes in der Praxis bestätigen.

# Abbildungsverzeichnis

---

ABBILDUNG 2-1: PRODUKTIONSETAPPEN IN DER HALBLEITERFERTIGUNG (EIGENE DARSTELLUNG).....	5
ABBILDUNG 2-2: BEISPIEL EINER ROUTE – QUELLE: [ATHERTON95] .....	7
ABBILDUNG 2-3: TRANSFER ZWISCHEN FERTIGUNGSBEREICHEN .....	9
ABBILDUNG 2-4: NICHTFUNKTIONALE ANFORDERUNGEN – QUELLE: [KOTONYA98] .....	17
ABBILDUNG 3-1: KLASSIFIZIERUNG DER SIMULATION IM BEREICH PRODUKTION UND LOGISTIK.....	28
ABBILDUNG 3-2: ABLAUSCHHEMA EINER EREIGNISORIENTIERTEN SIMULATION – QUELLE: [ROSE11] .....	30
ABBILDUNG 3-3: ABHÄNGIGKEIT ZWISCHEN PROGNOSEHORIZONT UND MODELLGRANULARITÄT .....	33
ABBILDUNG 3-4: VERHÄLTNISS VON PROGNOSEHORIZONT UND ABHÄNGIGKEIT VON EXAKTER INITIALISIERUNG. QUELLE: [REIJERS99].....	33
ABBILDUNG 3-5: VEREINFACHTES ABBILD EINER ROUTE IN DER HALBLEITERFERTIGUNG .....	36
ABBILDUNG 3-6: SCHEMATISCHE DARSTELLUNG EINES WARTESCHLANGENPROZESS .....	39
ABBILDUNG 4-1: PHASEN DER ENTWICKLUNG EINER RECHNERGESTÜTZTEN PROGNOSEMETHODE .....	45
ABBILDUNG 4-2: SCHEMATISCHE DARSTELLUNG DER V-FAB DATENBANK MIT NOTWENDIGEN QUELLEN .....	46
ABBILDUNG 4-3: BEISPIEL EINER ZIELOPERATION UND ZUGEHÖRIGEN ROUTEN SOWIE PRODUKTE INNERHALB EINES PRODUKTIONSBEREICHS .....	49
ABBILDUNG 4-4: SCHEMATISCHE DARSTELLUNG EINER ROUTE MIT ZIELOPERATION UND MÖGLICHEN STARTOPERATIONEN UND DAZUGEHÖRIGER DURCHLAUFZEIT.....	49
ABBILDUNG 4-5: HÄUFIGKEITSVERTEILUNG DER DURCHLAUFZEITWERTE ZWISCHEN SCHRITT A UND B .....	53
ABBILDUNG 4-6: HÄUFIGKEITSVERTEILUNG DES PROZENTUALEN VERHÄLTNISS DER STANDARDABWEICHUNG ZUM MITTELWERT DER DURCHLAUFZEIT ZWISCHEN ZWEI OPERATIONEN FÜR MITHILFE VON PRODUKT UND ROUTE GEBILDETEN KLASSEN .....	53
ABBILDUNG 4-7: HÄUFIGKEITSVERTEILUNG DER DURCHLAUFZEITWERTE ZWISCHEN SCHRITT A UND B FÜR UNTERSCHIEDLICHE PRODUKTE A UND B .....	53
ABBILDUNG 4-8: DER VERGLEICH VON STANDARDABWEICHUNG DER DURCHLAUFZEIT IN KLASSEN GEBILDET DURCH START- UND ENDOPERATION MIT GESAMTHEIT AN WERTEN UND NACH BESEITIGUNG DURCH PRIORITÄT BESCHLEUNIGTER LOSE .....	55
ABBILDUNG 4-9: KLASSIFIKATION DER LOSE MITHILFE IHRER PRIORITÄT .....	55
ABBILDUNG 4-10: CLUSTERANALYSE MITTELS K-MEANS-ALGORITHMUS BEI VERWENDUNG VON VERSPÄTUNG UND DURCHLAUFZEIT ALS ANKUNFTSDATENSTROM .....	58
ABBILDUNG 4-11: CLUSTERANALYSE MITTELS K-MEANS-ALGORITHMUS BEI VERWENDUNG VON DURCHLAUFZEIT ALS ANKUNFTSDATENSTROM.....	58
ABBILDUNG 4-12: BOXPLOT DER VERSPÄTUNGSVERTEILUNG INNERHALB CLUSTER, DIE MITTELS K-MEANS- VERFAHREN BEI VERWENDUNG VON DURCHLAUFZEIT ALS ANKUNFTSDATENSTROM ERZEUGT WURDEN .....	58
ABBILDUNG 4-13: BEISPIEL EINER BETRIEBSKENNLINIE FÜR EIN HALBLEITERWERK .....	60

## Abbildungsverzeichnis

---

ABBILDUNG 4-14: HÄUFIGKEITSVERTEILUNG DER VERSPÄTUNG AN EINER GEGEBENEN OPERATION BEI HOCH UND NIEDERAUSGELASTETER FABRIK .....	61
ABBILDUNG 4-15: DIE ABHÄNGIGKEIT DER PROZESSZEIT VON DER ANZAHL DER WAFER IN DER CASSETTE BEI EINZEL-WAFER PROZESSIERUNG .....	64
ABBILDUNG 4-16: DIE ABHÄNGIGKEIT DER PROZESSZEIT VON DER ANZAHL DER WAFER IN DER CASSETTE BEI MINI-BATCH PROZESSIERUNG .....	64
ABBILDUNG 4-17: DIE ABHÄNGIGKEIT DER PROZESSZEIT VON DER ANZAHL DER WAFER IN DER CASSETTE BEI BATCH PROZESSIERUNG .....	64
ABBILDUNG 4-18: HÄUFIGKEITSVERTEILUNG DER WAFERANZAHL IN EINEM LOS ÜBER GESAMTEN BESTAND DER FABRIK ABZÜGLICH DER SPLIT-LOSE .....	66
ABBILDUNG 4-19: DIE KOMPLEXITÄT DER ABHÄNGIGKEITEN VERSCHIEDENER KENNZAHLEN IN DER HALBLEITERFERTIGUNG. QUELLE: [FILHO10] .....	68
ABBILDUNG 4-20: DER MECHANISMUS FÜR DIE ERSTELLUNG DER STATISTIK FÜR DIE DURCHLAUFZEITANGABEN IN DER AFM .....	75
ABBILDUNG 4-21: INITIALISIERUNG UND UPDATE DER SAMPLINGSTATISTIK .....	80
ABBILDUNG 4-22: DIE ABHÄNGIGKEIT DES ENTWICKLUNGSSTADIUMS EINES PRODUKTES UND DIE WAHRSCHEINLICHKEIT VOM ÜBERGANG EINES LOSES IN DEN HOLD-ZUSTAND.....	82
ABBILDUNG 4-23: DIE FUNKTIONSWEISE EINES FILTER DER HOLD-RATEN AM BEISPIEL VON ENTWICKLUNGWARE, NEUEM PRODUKT UND MASSENWARE.....	85
ABBILDUNG 4-24: DIE UNTERSCHIEDLICHEN VARIANTEN DES ÜBERGANGS ZWISCHEN FERTIGUNGSBEREICHEN .....	86
ABBILDUNG 4-25: DIE PROZENTUALE ZUORDNUNG FÜR EIN AUSGANGSPRODUKT ZU EINER TESTROUTE BEI DEM WECHSEL VON DEM PRODUKTIONSBEREICH ERSTELLT ÜBER LÄNGERES ZEITINTERVALL MITHILFE VON REALEN LOSBEWEGUNGSDATEN (BEISPIEL) .....	87
ABBILDUNG 4-26: DER PROZESS DER BERECHNUNG EINER PROGNOSE MITTELS AFM .....	89
ABBILDUNG 4-27: DIE SEQUENZ MÖGLICHER ZUSTÄNDE EINES LOSES IN EINER OPERATION .....	90
ABBILDUNG 4-28: MÖGLICHE LOSPOSITION BEI DER INITIALISIERUNG DER PROGNOSE.....	91
ABBILDUNG 4-29: NOTWENDIGE INFORMATIONEN FÜR DIE BERECHNUNG EINES PLANES FÜR DIE MATERIALEINSCHLEUSUNG.....	93
ABBILDUNG 4-30: AFM-KONZEPT FÜR DEN UMGANG MIT BEGRENZTER PRODUKTIONSKAPAZITÄT AN DEN OPERATIONEN ZWISCHEN START- UND ZIELOPERATION.....	95
ABBILDUNG 4-31: DAS PHASENMODELL FÜR DIE VORAUSSAGE DER ANLIEFERUNGSZEITPUNKTE IN DEM TESTBEREICH DER FERTIGUNG .....	97
ABBILDUNG 4-32: DIE SAMPLINGRATE FÜR DIE GESAMTE FABRIK, PRODUKTIONSBEREICH PARAMETER-TEST UND EIN MASSENPRODUKT AN EINER GEGEBENEN OPERATION .....	98
ABBILDUNG 4-33: 33 % LOS-SAMPLING BEI AFM .....	100
ABBILDUNG 4-34: DAS BOTTLENECK-FRÜHWARNSYSTEM MITTELS AFM .....	102
ABBILDUNG 4-35: AFM ANKUNFTSPROGNOSE: ANZAHL DER LOSE GRUPPIERT NACH DER PROZESSZEIT .....	102

## Abbildungsverzeichnis

---

ABBILDUNG 4-36: AFM-PROGNOSE DER FABRIKOUTPUTS IN GEGEBENER WOCHEN .....	103
ABBILDUNG 4-37: REPORT DER AFM-ANALYSE DER ABSCHNITTE DER PRODUKTIONSLINIE BEZÜGLICH DER ANKUNFTSRATE AN EINER MASCHINENARBEITSGRUPPE .....	104
ABBILDUNG 4-38: REPORT DER AFM-ANALYSE DER BESTÄNDE PRO OPERATION BEI DER PROGNOSE DER ANLIEFERUNG AN EINER MASCHINENARBEITSGRUPPE .....	105
ABBILDUNG 5-1: TÄGLICHER AUFRUF EINER ABFRAGE ÜBER BESTANDSVERTeilUNG IN DER FABRIK .....	108
ABBILDUNG 5-2: EINE EINFACHE ABFRAGE EINER TABELLE IN RTD-REPOSITORIIEN .....	108
ABBILDUNG 5-3: KOMPONENTEN DER AFM-SOFTWARE .....	109
ABBILDUNG 5-4: DER AUFBAU DES KERNS DER AFM-SOFTWARE IN PAKETEN SOWIE NOTWENDIGE TABELLEN .....	110
ABBILDUNG 5-5: DIE DATENQUELLEN DER AFM .....	111
ABBILDUNG 5-6: SONDERBEHANDLUNG BEI DER GEWINNUNG DER INFORMATION ÜBER DIE PROZESSZEIT AUS MASCHINENDATEN FÜR EIN PRODUKTIONSBEREICH .....	113
ABBILDUNG 5-7: DIE SCHEMATISCHE DARSTELLUNG DES AFM-DATENAUFBEREITUNGSPROZESSES .....	114
ABBILDUNG 5-8: DAS SCHEMA DER AFM-TABELLE MIT HISTORISCHEN BEWEGUNGSABLÄUFEN .....	115
ABBILDUNG 5-9: DATENFLUSS BEI DER ERSTELLUNG EINER PROGNOSE MITTELS AFM .....	117
ABBILDUNG 5-10: DER AUFBAU DES AFM-REPORTING MODULS .....	119
ABBILDUNG 6-1: DER VERLAUF DER MITTLEREN ZEITABWEICHUNG ZWISCHEN AFM-PROGNOSE UND REALITÄT FÜR DEN 7-TÄGIGEN PROGNOSEHORIZONT .....	122
ABBILDUNG 6-2: DER VERLAUF DER MITTLEREN ZEITABWEICHUNG ZWISCHEN AFM-PROGNOSE UND REALITÄT FÜR DEN 28-TÄGIGEN PROGNOSEHORIZONT .....	122
ABBILDUNG 6-3: DER VERLAUF DER MITTLEREN ZEITABWEICHUNG ZWISCHEN AFM-PROGNOSE UND REALITÄT FÜR DEN 24-STUNDIGEN PROGNOSEHORIZONT .....	122
ABBILDUNG 6-4: DER VERLAUF DER MITTLEREN ZEITABWEICHUNG ZWISCHEN AFM-PROGNOSE UND REALITÄT FÜR DEN 7-TÄGIGEN PROGNOSEHORIZONT IN ABHÄNGIGKEIT VON DEN PRODUKTEIGENSCHAFTEN ...	124
ABBILDUNG 6-5: DER VERLAUF DER MITTLEREN ZEITABWEICHUNG ZWISCHEN AFM-PROGNOSE UND REALITÄT FÜR DEN 7-TÄGIGEN PROGNOSEHORIZONT IN ABHÄNGIGKEIT VON DER ZIELPOSITION IN DER FERTIGUNGSLINIE .....	125
ABBILDUNG 6-6: TREFFSICHERHEIT DER ANLIEFERUNGSPROGNOSE DER MATERIALMENGE MITTELS AFM PRO TAG .....	126
ABBILDUNG 6-7: TREFFSICHERHEIT DER ANLIEFERUNGSPROGNOSE DER MATERIALMENGE MITTELS AFM PRO TAG AN AUSGEWÄHLTEN ZIELOPERATIONEN .....	127
ABBILDUNG 6-8: TREFFSICHERHEIT DER ANLIEFERUNGSPROGNOSE DER MATERIALMENGE MITTELS AFM PRO TAG FÜR AUSGEWÄHLTE ENTWICKLUNGSPRODUKTE .....	129
ABBILDUNG 6-9: TREFFSICHERHEIT DER ANLIEFERUNGSPROGNOSE DER MATERIALMENGE MITTELS AFM PRO TAG FÜR AUSGEWÄHLTE MASSENPRODUKTE .....	129
ABBILDUNG 6-10: HISTOGRAMM DER VERTEILUNG DER FEHLERWAHRSCHEINLICHKEIT FÜR DIE WOCHENPROGNOSE .....	131

## Abbildungsverzeichnis

---

ABBILDUNG 6-11: HISTOGRAMM DER VERTEILUNG DER FEHLERWAHRSCHEINLICHKEIT FÜR DIE ROLLENDE WOCHENPROGNOSE .....	131
ABBILDUNG 6-12: DIE TREFFSICHERHEIT BEZÜGLICH DER EIGNUNG DER AFM FÜR DIE EINZELNEN ANWENDUNGSFELDER .....	132
ABBILDUNG 6-13: DIE VISUALISIERUNG DER PROGNOSE UND ZUGEHÖRIGEN REALITÄT MITHILFE DES AMPELSYSTEMS ZWECKS VALIDIERUNG .....	133
ABBILDUNG 6-14: DIE PROZENTUALEN UNTERSCHIEDE IN DER EINSCHLEUSMENGE ZWISCHEN DER REALITÄT UND UNTERSCHIEDLICHEN METHODEN DER BESTIMMUNG DES OPERATIVEN PLANS .....	136
ABBILDUNG 6-15: VERGLEICH DER METHODEN ZUR ABBILDUNG DES AUF DAS BATCHTOOL OPTIMIERTEN EINSCHLEUSVERHALTEN .....	138
ABBILDUNG 6-16: VERGLEICH DER METHODEN ZUR ABBILDUNG DES EINSCHLEUSVERHALTENS DES MASSENPRODUKTES .....	138
ABBILDUNG 6-17: VERGLEICH DER METHODEN ZUR ABBILDUNG DES EINSCHLEUSVERHALTEN DES SONDERPRODUKTES.....	138
ABBILDUNG 6-18: BEISPIELE FÜR DIE BESTIMMUNG VON GESAMPELTEN LOSEN MIT DEN METHODEN: I) DER MENGENKALKULATION, II) DES ZUFALLS, III) DES DURCHNUMMERIERENS (AFM). .....	140
ABBILDUNG 6-19: TREFFSICHERHEIT DER PROGNOSE BEI UNTERSCHIEDLICHEN ALGORITHMEN FÜR DIE SAMPLINGKALKULATION FÜR ALLE PRODUKTE .....	142
ABBILDUNG 6-20: TREFFSICHERHEIT DER PROGNOSE BEI UNTERSCHIEDLICHEN ALGORITHMEN FÜR DIE SAMPLINGKALKULATION FÜR ALLE PRODUKTE .....	142
ABBILDUNG 6-21: TREFFSICHERHEIT DER PROGNOSE BEI UNTERSCHIEDLICHEN ALGORITHMEN FÜR DIE SAMPLINGKALKULATION FÜR ALLE PRODUKTE .....	142
ABBILDUNG 6-22: DIE ABHÄNGIGKEIT VON DER GESAMTEN HOLD-DAUER UND DER REGULÄREN DURCHLAUFZEIT EINES LOSES ÜBER DEN VOLLSTÄNDIGES BESTAND EINER FABRIK IN LÄNGEREM ZEITABSCHNITT .....	143
ABBILDUNG 6-23: DER VERGLEICH DER TREFFSICHERHEIT DER ANLIEFERUNGSPROGNOSE MIT UND OHNE BEHANDLUNG DES HOLD-ZUSTANDS SEITENS AFM.....	144
ABBILDUNG 6-24: DER VERLAUF DER ABSOLUTEN ABWEICHUNG DER ANLIEFERUNGSZEIT ZWISCHEN DEN AFM- PROGNOSEN UND DER REALITÄT BEI EXPERIMENTEN MIT HOLD-FILTERN .....	145
ABBILDUNG 6-25: DER VERLAUF DER ABSOLUTEN ABWEICHUNG DER ANLIEFERUNGSZEIT ZWISCHEN DEN AFM- PROGNOSEN UND DER REALITÄT BEI EXPERIMENTEN MIT HOLD-FILTERN NACH EINER GRAVIERENDEN FABRIKSTÖRUNG .....	146
ABBILDUNG 6-26: DER VERGLEICH DER TREFFSICHERHEIT DER ANLIEFERUNGSPROGNOSE MIT DEM DETAILREICHEN UND -ARMEN ANSATZ FÜR DIE BEHANDLUNG VON SAMPLING AN DEN ZWISCHENOPERATIONEN .....	147
ABBILDUNG 7-1: DER VERGLEICH DER MITTLEREN ZEITABWEICHUNG FÜR DIE AFM, PBPA UND KFS .....	151
ABBILDUNG 7-2: AFM - TREFFSICHERHEIT DER ANLIEFERUNGSPROGNOSE DER MATERIALMENGE PRO TAG..	152
ABBILDUNG 7-3: PBPA - TREFFSICHERHEIT DER ANLIEFERUNGSPROGNOSE DER MATERIALMENGE PRO TAG.	152

## Abbildungsverzeichnis

---

ABBILDUNG 7-4: KFS - TREFFSICHERHEIT DER ANLIEFERUNGSPROGNOSE DER MATERIALMENGE PRO TAG....	152
ABBILDUNG 7-5: PBPA UND AFM-TREFFSICHERHEIT DER ANLIEFERUNGSPROGNOSE DER MATERIALMENGE PRO TAG FÜR EIN ENTWICKLUNGSPRODUKT .....	153
ABBILDUNG 7-6: VERGLEICH DER PROGNOSEN FÜR DIE WÖCHENTLICHE AUSSCHLEUSUNG DES MATERIALS VON DER FABRIK.....	154
ABBILDUNG 7-7: SUMMENHÄUFIGKEIT DER TREFFSICHERHEIT DER PROGNOSEN FÜR DIE WÖCHENTLICHE AUSSCHLEUSUNG .....	154

# Tabellenverzeichnis

---

TABELLE 3-1: VEREINFACHTES ABBILD EINES PROZENTUALEN ZEITANTEILS AN DER GESAMTBEARBEITUNGSZEIT EINER ROUTE FÜR JEDE OPERATION. ....	36
TABELLE 4-1: BEISPIEL EINER MATRIX VON STANDARDABWEICHUNGEN FÜR DIE KLASSEN ENTSTANDEN DURCH KREUZPRODUKT VON EINEM PRODUKT UND ZUGEHÖRIGER ZEITPERIODE VON STARTOPERATION ZU ZIELOPERATION FÜR ALLE LOSE AUS HISTORISCHEN BEWEGUNGSDATEN IN DER FABRIK .....	55
TABELLE 4-2: VERGLEICH DER TAKTISCHEN UND OPERATIVEN PLANUNG BEZÜGLICH EIGNUNG DER MATERIALEINSCHLEUSUNGSPLÄNE FÜR DIE AFM .....	73
TABELLE 4-3: FALLUNTERSCHIEDUNG BEI DER AUSWERTUNG VON SAMPLING .....	78
TABELLE 4-4: DIE QUANTIFIZIERUNG DER PROZENTUALEN ZUORDNUNG FÜR EIN AUSGANGSPRODUKT ZU EINER TESTROUTE BEI DEM WECHSEL VON DEM PRODUKTIONSBEREICH ERSTELLT ÜBER ALLE ZEITINTERVALLE MITHILFE VON REALEN LOSBEWEGUNGSDATEN (BEISPIEL) .....	87
TABELLE 4-5: DER ERWARTUNGSWERT DER PROZENTUALEN RATE FÜR DIE TESTROUTENZUORDNUNG ERSTELLT MIT UNTERSCHIEDLICHEN TECHNIKEN SOWIE EIN VERGLEICH ZU AFM BEZÜGLICH TREFFSICHERHEIT VON DEN PROGNOSEN .....	87
TABELLE 4-6: BEISPIEL VON DER VERWENDUNG DER HOLD-STATISTIK FÜR DIE BERECHNUNG DER PROGNOSTIZIERTEN VERZÖGERUNG DER LOSANKUNFT AN DER ZIELOPERATION .....	96
TABELLE 4-7: KLASSIFIKATION DER SAMPLINGRATEN IN SAMPLINGKLASSEN .....	99
TABELLE 6-1: DER VERGLEICH DER TREFFSICHERHEIT FÜR DIE ENTWICKLUNGS- UND MASSENPRODUKTE .....	129
TABELLE 6-2: DIE QUANTIFIZIERUNG DER VORHERSAGEQUALITÄT FÜR EIN 5-AMPEL-SYSTEM .....	134
TABELLE 7-1: VERGLEICH UNTERSCHIEDLICHER ANSÄTZE FÜR DIE PROGNOSEN IN DER HALBLEITERFERTIGUNG .....	150
TABELLE 7-2: DIE QUANTIFIZIERUNG DES VERGLEICHES VON TREFFSICHERHEIT DER PROGNOSEN FÜR DIE WÖCHENTLICHE AUSSCHLEUSUNG DES MATERIALS VON DER FABRIK .....	154

## Literaturverzeichnis

---

- [Aamodt94] Aamodt Agnar, Plaza Enric. (1994). *Case-based reasoning: foundational issues, methodological variations, and system approaches*. AI Communications, v. 7 n. 1, p. 39-59
- [Akcali00] Akcali, E., Uzsoy, R., Hiscock, D. G., Moser, A. L., Teyner, T. J. (2000). *Alternative loading and dispatching policies for furnace operations in semiconductor manufacturing: a comparison by simulation*. In: Proceedings of the Winter Simulation Conference, pp. 1428–1435
- [Ankenm08] Ankenman, B., Nelson B. L., Tongarlak M., Fowler J., Mackulak G., and D. Pabst. (2008) *Cycle Time Prediction for Semiconductor Manufacturing via Simulation on Demand*. Working Paper, Department of Industrial Engineering & Management Sciences (2008), Northwestern University.
- [Armstr01] Armstrong. J. Scott. (Hrsg.). (2001). *Principles of Forecasting: A Handbook for Researchers and Practitioners*. Springer Science + Business Media Inc. Boston.
- [Atherton95] Atherton, L. F.; Atherton, R. W. (1995) *Wafer Fabrication: Factory Performance and Analysis*. Kluwer Academic Publishers, Boston, Dordrecht, London.
- [Aurand97] Aurand, S., and P. Miller. (1997). *The Operating Curve: A Method to Measure and Benchmark Manufacturing Line Productivity*. IEEE/SEMI Advanced Semiconductor Manufacturing Conference, 391-397.
- [Aydt08] Aydt, Heiko, Turner, Stephen John, Cai Wentong, Low, Malcolm Yoke Hean, Lendermann, Peter, Gan, Boon Ping. (2008). *Symbiotic Simulation Control in Semiconductor Manufacturing*. Proceeding ICCS '08 Proceedings of the 8<sup>th</sup> international conference on Computational Science, Part III. Springer-Verlag Berlin, Heidelberg.
- [Backes96] Backes, M. (1996). Simulationsunterstützung zur zielorientierten Produktionsprozessplanung und -regelung. Universität Mannheim, Dissertation.
- [Backus06] Backus, P., Janakiram, M., Mowzoon, S., Runger, G. C., Bhargava, A. (2006). *Factory cycle-time prediction with data-mining approach*. IEEE Trans. Semicond. Manuf. Vol. 19, no. 2. Pp. 252-258.
- [Bauma96] Baumann, Rüdiger. (1996). *Didaktik der Informatik*. Auflage 2, Klett Schulbuchverlag. Stuttgart.
- [Bagchi08] Bagchi, S., C.-H. Chen-Ritzo, S. T. Shikalgar, and M. Toner. (2008). "A full-factory simulator as a daily decision-support tool for 300mm wafer fabrication productivity." In Proceedings of the 2008 Winter Simulation Conference, eds. S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, J. W. Fowler, 2021–2029. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

## Literaturverzeichnis

---

- [Bacher10] Bacher, J., Pöge, A., Wenzig, K. (2010). *Clusteranalyse - Anwendungsorientierte Einführung in Klassifikationsverfahren*, 3. Auflage. München. Oldenbourg. ISBN 978-3-486-58457-8.
- [Becker81] Becker Fritz (1981). *Analyse und Prognose von wirtschaftlichen Zeitreihen der deutschen Schaden- und Unfallversicherung*. Hrsg.: Peter Albrecht, Egon Lorenz, Fritz Becker. Mannheimer Reihe - Veröffentlichungen des Instituts für Versicherungswissenschaft der Universität Mannheim, Band 19. ISBN 978-3-88487-025-9.
- [Bonvik94] Bonvik, Asbjorn M. (1994). *Estimating the Lead Time Distribution of Priority Lots in a Semiconductor Factory*. Massachusetts Institute of Technology, Operations Research Center. Operations Research Center Working Paper. P.: 293-94.
- [Box94] Box, G.E.P.; Jenkins, G.M. (1994). *Time series analysis, forecasting and control*. Prentice Hall. San Francisco.
- [Brockh02] Brockhaus: Der Brockhaus in fünfzehn Bänder. 2. F. A. Brockhaus. Mannheim.
- [Brugger05] Brugger, Ralph. (2005). *IT-Projekte strukturiert realisieren. Situationen analysieren, Lösungen konzipieren - Vorgehen systematisieren, Sachverhalte visualisieren - UML und EPKs nutzen*. Vieweg & Teubner, ISBN : 978-3-8348-0118-0.
- [Buttler02] Buttler Günter, Fickel Norman. (2002). *Einführung in die Statistik*. Rowohlt Tb. Verlag, ISBN 3-4995- 5645-6.
- [Chang01] Chang P-C, Hsieh J-C, Liao TW. (2001). *A case-based reasoning approach for due date assignment in a wafer fabrication factory*. Proceedings of the International Conference on Case-Based Reasoning (ICCBR 2001), Vancouver, British Columbia, Canada.
- [Chang03] Chang P-C, Hsieh J-C. (2003). *A neural networks approach for due-date assignment in a wafer fabrication factory*. Int J Ind Eng 10(1): 55–61.
- [Chen03] Chen T. (2003). *A fuzzy back propagation network for output time prediction in a wafer fab*. Appl Soft Comput 2(3): 211-222.
- [Chen88] Chen, H., J. M. Harrison, A. Mandelbaum, A. Van Ackere, L. M. Wein. (1988). *Empirical evaluation of a queueing network model for semiconductor wafer fabrication*. Operations Research, Vol. 36, No. 2, Operations Research in Manufacturing. S. 202–215.
- [Cole90] Cole, D. (1990). *The use of simulation in semiconductor technology development*. Solid-State Electronics, vol. 33, issue 6, pp. 591-623.
- [Dehling04] Dehling, H. und Haupt, B. (2004) *Einführung in die Wahrscheinlichkeitstheorie und Statistik*. Springer Verlag. Berlin.

## Literaturverzeichnis

---

- [Domsch02] Domschke W., Drexl A. (2002). *Einführung in Operational Research*. 5. Springer, Berlin 7, 10, 11, 71, 72.
- [Ehrmann02] Ehrmann, Harald. (2002). *Unternehmensplanung*. Kiehl, Ludwigshafen (Rhein). ISBN 3-470-46834-6 Harald.
- [Eichholz08] Eichholz, Rüdiger R. (2008). *Berichtswesen und Informationsmanagement*. Beck Juristischer Verlag. München. ISBN 978-3-406-57304-0.
- [Filho10] Filho, Moacir Goodish, Uzsoy, Reha. (2010). *The effect of shop floor continuous improvement programs on the lot size–cycle time relationship in a multi-product single-machine environment*. The International Journal of Advanced Manufacturing Technology Volume 52, Numbers 5-8, 669-681.
- [Fishman01] Fishman, G. S. (2011). *Discrete-event simulation – Modeling, programming, and analysis*. Springer. New York.
- [Fowler97] Fowler, J.W., S. Brown, H. Gold, and A. Schoemig. (1997). *Measurable Improvements in Cycle-Time-Constrained Capacity*. Proceedings of the Sixth International Symposium on Semiconductor Manufacturing (ISSM), San Francisco, U.S.A.
- [Frauens90] Frauenstein, T; Pape, U; Wagner, O. (1990). *Objektorientierte Sprachkonzepte und diskrete Simulation*. In: Möller, D. (Hrsg.); Schmidt, B. (Hrsg.): Fachberichte Simulation, Bd. 13. Springer. Berlin.
- [Gan06] Gan, B.P., Chan, L.P., Turner, S.J. (2006). *Interoperating simulations of automatic material handling systems and manufacturing processes*. In: Proceedings of the Winter Simulation Conference 2006, pp. 1129–1135.
- [GanB06] Gan, B.P., Lendermann, P., Quek, K.P.T., van der Heijden, B., Chin, C.C., Koh, C.Y. (2006). *Simulation analysis on the impact of furnace batch size increase in a deposition loop*. In: Perrone, L.F., Lawson, B., Liu, J., Wieland, F.P. (eds.) Proceedings of the Winter Simulation Conference, pp. 1821–1828.
- [Garey76] Garey, M.R. (1976). *The Complexity of Flowshop and Jobshop Scheduling*. *Mathematics of Operations Research* 1 (2): 117–129.
- [Gelasch07] Gelaschwili, Simon. (2007). *Einführung in die Statistische Modellierung und Prognose*. Statistische Diskussionsbeiträge. Band Nr. 26 / Lehrstuhl für Statistik und Ökonometrie, Wirtschafts- und Sozialwissenschaftliche Fakultät der Universität Potsdam.
- [Gollas09] Gollas Ferdinand. (2009). Mapping-gestützte Datensynchronisation für Heterogene relationale Datenbanken. Diplomarbeit. Der Ludwig-Maximilians-Universität München.
- [Gordon61] Gordon, G. (1961). *A general purpose systems simulation program*. In: Proceedings of the Eastern joint Computer Conference. Washington, D.C.

## Literaturverzeichnis

---

- [Grame08] Gramelsberger, Gabriele. (2008) *Computersimulationen - Neue Instrumente der Wissensproduktion* In: Mayntz, R./ Neidhardt, F. / Weingart, P./ Wengenroth, U. (Hrsg.): *Wissensproduktion und Wissenstransfer. Wissen im Spannungsfeld von Wissenschaft, Politik und Öffentlichkeit*, S 75-95.
- [Gross08] Gross Donald, Shortle John F., James M. Thompson James M., Harris Carl M. (2008). *Fundamentals of Queueing Theory, Solutions Manual (Wiley Series in Probability and Statistics)*. Wiley-Interscience; 4 edition. ISBN 0470077964.
- [Haben10] Habenicht Ilka. (2010). *Parallelisierbare Terminierungsverfahren als Baustein für ein Multiagentensystem*. Dissertation, Fernuniversität Hagen.
- [Hartung95] Hartung Joachim, Elpelt Bärbel Klösener Karl-Heinz. (1995) *Lehr- und Handbuch der angewandten Statistik*. Oldenbourg Wissenschaftsverlag. München, Wien.
- [Held07] Held Pascal. (2007). *Beschreibung von Warteschlangensystemen und deren psychologische Einflüsse auf den Menschen*. Otto-Von-Guericke Universität Magdeburg. Institut für Simulation und Graphik.
- [Helmke01] Helmke Hartmut, Isernhagen Rolf . 2001. *Softwaretechnik in C und C++ - Das Lehrbuch: Modulare, objektorientierte und generische Programmierung*. Hanser Fachbuch Verlag. ISBN-13: 978-3446216839.
- [Hosoe07] Hosoe H., Knanamori N., and Yoshida K.. (2007) *The methods of data collection and tool processing time estimation in lot processing*. in *Semiconductor Manufacturing. ISSM 2007. International Symposium on*, pp. 1–4.
- [Hooper86] Hooper, J. W. (1986). *Strategy-related characteristics of discrete-event languages and models*. In: *Simulation* 46, Nr. 4, S 153-159.
- [Hopp08] Hopp, W. (2008). *Factory Physics*. Mcgraw-Hill Higher Education; 3<sup>rd</sup> Revised edition. ISBN-13: 978-0071232463.
- [Hsu04] Hsu SY, Sha DY (2004). *Due date assignment in wafer fabrication using artificial neural network*. *Int J Adv Manuf Technol* 23(9–10):768–775.
- [Hüttner86] Hüttner Manfred. (1986). *Prognoseverfahren und ihre Anwendung*. Gruyter. Berlin. ISBN 3-11-010826-7.
- [Jensen07] Jensen, Sven. (2007). *Eine Methodik zur teilautomatisierten Generierung von Simulationsmodellen aus Produktionsdatensystemen am Beispiel einer Job Shop Fertigung*. Kassel University Press. Kassel. ISBN: 978-3-89958-289-5.
- [Jung03] Jung Hans. (2002). *Controlling*. Oldenbourg Wissenschaftsverlag. ISBN: 978-3486273151.
- [Kanun02] Kanungo T., Mount D. M., Netanyahu N. S., Piatko C. D., Silverman R., Wu A. Y. (2002). *An efficient k-means clustering algorithm: Analysis and implementation*. In: *IEEE Trans. Pattern Analysis and Machine Intelligence*. 24, S. 881–892.

## Literaturverzeichnis

---

- [Kaplan93] Kaplan, A. C., Unal, A. T. (1993). *A probabilistic cost-based due date assignment model for job shops*. Int. J. Production Res. Vol. 31, no. 12. Pp. 2817–2834.
- [Kaeche04] Kaechel Joachim, Teich Tobias. (2004). *Produktionswirtschaft*. 1. Aufl. GUC Verlag. ISBN 3-934235-19-0
- [Kempf96] Kempf, K.G. (1996). *Simulating semiconductor manufacturing systems: successes, failures, and deep questions*. In: Proceedings of the 1996 Winter Simulation Conference. ACM Press. New York.
- [Kirchg05] Kirchgässner, Gebhard, Wolters, Jürgen. 2005. *Einführung in die moderne Zeitreihenanalyse*. WiSo Kurzlehrbücher VII. Verlag Vahlen. ISBN 978-3-8006-3268-8
- [Knuth97] Knuth Donald E. (1997) *The Art of Computer Programming. Volume 2: Seminumerical Algorithms. 3. Auflage*. (Addison-Wesley Series in Computer Science and Information Processing). Addison-Wesley, Reading MA u. a. ISBN 0-201-89684-2. S. 120ff.
- [Kohn08] Kohn R., Noack D., Mosinski M., Zhou Z., and Rose O. 2009. "Evaluation of Modeling, Simulation and Optimization Approaches for Work Flow Management in Semiconductor Manufacturing" In *Proceedings of the 2009 Winter Simulation Conference*, eds. M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, 1592–1600. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- [Kostur95] Kosturiak, Jan, Milan, Gregor. (1995) *Simulation von Produktionssystemen*. Springer Verlag. Wien.
- [Kotonya98] Kotonya Gerald, Sommerville Ian. (1998). "*Requirements Engineering - Processes and Techniques*", John Wiley & Sons.
- [Krengel05] Krengel, U. (2005). *Einführung in die Wahrscheinlichkeitstheorie und Statistik*, 8. Auflage, Vieweg Verlag. Wiesbaden.
- [Kropik09] Kropik Markus. (2009). *Produktionsleitsysteme in der Automobilfertigung. (VDI-Buch)*. Springer Berlin Heidelberg; Auflage: 1. Berlin. ISBN: 3540889906
- [Law00] Law, Averill, Kelton, W. David. (2000). *Simulation modeling and Analysis*. McGraw-Hill Science/Engineering/Math. ISBN 0070592926
- [Lazow84] Lazowska E.D., Zahorjan j., Graham G.S., Cvencik K.C. (1984). *Quantitative System Performance*. Prentice-Hall, Engelwood Cliffs, N.J
- [Lee02] Lee, Y. W., D. M. Strong, B. K. Kahn, and R. Y. Wang. (2002). AIMQ: *A methodology for information quality assessment*. Information & Management 40 (2): 133 – 146.
- [Liao 04] Liao, D.Y. , Wang C.N. (2004). *Neural-network-based delivery time estimates for prioritized 300-mm automatic material handling operations*. IEEE Trans. Semiconduct. Manufact. Vol. 17, no. 3. Pp. 324–332.

## Literaturverzeichnis

---

- [Noack12] Noack, Daniel. (2012). *Online Simulation in Semiconductor Manufacturing*. Dissertation. <http://athene.bibl.unibw-muenchen.de:8081/doc/90413/90413.pdf>
- [Liebl95] Liebl, F. (1995). *Simulation: Problemorientierte Einführung*. 2. Oldenburg Verlag. München.
- [Little61] Little, J. D. C. (1961). *A Proof of the Queueing Formula  $L = \lambda W$* . Operations Research, 9, 383-387. <http://www.jstor.org/stable/167570> (von 25.09.11)
- [Loeser06] Loeser, Simon Christopher. (2006). *Modellierung und Prozessoptimierung der Organisationsstruktur in der orthopädischen Poliklinik*. Dissertation. Universität Marburg, Medizin. Deutsche National Bibliothek.
- [Lu94] Lu S. C. H., D. Ramaswamy D., Kumar and P. R. (1994). *Efficient scheduling policies to reduce mean and variance of cycle-time in semiconductor plants* IEEE Trans. Semiconduct. Manufact., vol. 7, pp. 374–388.
- [Martens04] Martens Peter, Rässler Susanne. (2004). *Prognoserechnung. Sechste, völlig neu bearbeitete und erweiterte Auflage*. Physica-Verlag. ISBN 3-7908-0216-6
- [May06] May G. S. and Spanos C. J.. (2006). *Fundamentals of semiconductor manufacturing and process control*. John Wiley & Sons, Inc.
- [Mehl94] Mehl, H. (1994). *Methoden verteilter Simulation*. Vieweg Verlag. Braunschweig.
- [Mladen09] Mladenova Teodora. (2009). *Einsatz der Warteschlangentheorie: Vergleich beider Entscheidungsinstrumente anhand eines Praxisbeispiels*. VDM – Verlag. ISBN 3639170229
- [Mosin11] Mosinski Marcin, Noack Daniel, Pappert Falk Stefan, Rose Oliver, Scholl Wolfgang. (2011). Cluster based analytical method for the lot delivery forecast in semiconductor fab with wide product range. In Proceedings of the 2011 Winter Simulation Conference S. Jain, R.R. Creasey, J. Himmelspach, K.P. White, and M. Fu, eds.
- [Mönch09] Mönch L., Fowler J. W., Dauzere-Peres S., Mason S. J., and Rose O. (2009). *Scheduling semiconductor manufacturing operations: Problems, solution techniques, and future challenges*. In MISTA Conference, 2009.
- [Na05] Na, L, Zhang, L., Zhang, M., Li Z. (2005). *Applied factory physics study on semiconductor assembly and test manufacturing*. Semiconductor Manufacturing, 2005. ISSM 2005, IEEE International Symposium.
- [Neuhe00] Neuheuser-Metternich, Sylvia. (2000) *Kommunikation und Berichtswesen*. Beck Juristischer Verlag. München. ISBN 978-3-406-45664-0.
- [Neusser06] Klaus Neusser. (2006). *Zeitreihenanalyse in den Wirtschaftswissenschaften*. Vieweg+Teubner. Wiesbaden. ISBN 383510117X.

## Literaturverzeichnis

---

- [Noack12] Noack Daniel. (2012). *Online Simulation in Semiconductor Manufacturing*. Universität der Bundeswehr München, Fakultät für Informatik.
- [Noche90] Noche, B. (1990) Simulation in Produktion und Materialfluss. TÜV Rheinland. Köln.
- [Nurani96] Nurani Raman K, Akella R, Strojwas A J. (1996) *In-line defect sampling methodology in yield management: an integrated framework*. IEEE Transactions on Semiconductor Manufacturing, Volume: 9, Issue: 4, Pages: 506-517
- [Odrey01] Odrey Nicholas G., Green Jonathan D., Appello Adrienne. (2001). *A generalized Petri net modeling approach for the control of re-entrant flow semiconductor wafer fabrication*. 10th International Conference on Flexible Automation and Intelligent, Robotics and Computer-Integrated Manufacturing, Volume 17, Issues 1–2, February 2001, Pages 5–11.
- [Page05] Page, Bernd; Kreutzer, Wolfgang. (2005). *The java Simulation Handbook*. Shaker Verlag. Aachen.
- [Pin08] Pinedo, M. (2008). *Scheduling: Theory, Algorithms and Systems*. Englewood Cliffs, New Jersey: Prentice-Hall
- [Potoradi02] Potoradi, J., Boon, O., Fowler, J., Pfund, M., Mason, S. (2002). *Using simulation-based scheduling to maximize demand fulfillment in a semiconductor assembly facility*. In: Proceedings of the Winter Simulation Conference, pp. 1857–1861
- [Potoradi99] Potoradi, J., Winz, G., Kam, L.W. (1999) Determining optimal lot-size for a semiconductor back-end factory. In: Proceedings of the Winter Simulation Conference, pp. 720-726
- [Rabe 07] Rabe, Spieckermann, Wenzel. (2007). *Verifikation und Validierung für die Simulation in Produktion und Logistik*. Springer Verlag.
- [Radd97] Raddon, A., Grigsby, B. (1997). *Throughput time forecasting model*. In Proc. IEEE/SEMI Advanced Semiconductor Manufacturing Conf. Pp. 430–433.
- [Reijers99] Reijers, H. A., W. M. P. van der Aalst. (1999). Short-term simulation: Bridging the gap between operational control and strategic decision making. *Proc. IASTED Internat. Conf. on Modeling and Simulation*. IASTED/Acta Press, Anaheim, CA 417–421.
- [Reinha03] Reinhardt, A. (2003). *Geschichten zur Simulation mit der Automobilindustrie*. In: Bayer, J. (Hrsg.); Collisi, T (Hrsg.); Wenzel, S (Hrsg.): Simulation in der Automobilproduktion. Berlin: Springer, S. 7-16
- [Rinne02] Prof. Dr. Rinne, Horst, Dr. Specht, Katja. (2002). *Zeitreihen. Statistische Modellierung, Schätzung und Prognose*. Verlag Vahlen. ISBN 978-3-8006-2877-3

- [Rose01] Rose O. (2001). *The Shortest Processing Time First (SPTF) dispatch rule and some variants in semiconductor manufacturing*. In Proceedings Winter Simulation Conference. vol. 2. pp. 1220–1224.
- [Rose05] Rose, O. (2005). *Transient Behavior During Single Product Start Peaks in a Semiconductor Fabrication Facility*. In Proceedings of the MASM 2005, pp. 147-154
- [Rose11] Rose, Oliver; März, Lothar; Krug Wilfried; Weigert, Gerald. (2011). *Simulation und Optimierung in Produktion und Logistik*. Springer Verlag. Berlin, Heidelberg.
- [Rose99] Rose, O. (1999). *Estimation of the Cycle Time Distribution of a Wafer Fab by a Simple Simulation Model*. In Proceedings of the 1999 International Conference on Semiconductor Manufacturing Operational Modeling and Simulation (SMOMS'99). San Francisco, CA, 133-138.
- [Rulkens98] Rulkens, H., van Campen, E., van Herk, J., Rooda, J. (1998). *Batch size optimization of a furnace and pre-clean area by using dynamic simulations*. In: Advanced Semiconductor Manufacturing Conference and Workshop (IEEE/SEMI), Boston, MA, USA, September 1998, pp. 439–444
- [Runkler10] Runkler Thomas A. (2010) *Data Mining: Methoden und Algorithmen intelligenter Datenanalyse*. 1. Auflage. Vieweg + Teubner.
- [Sauerw07] Sauerwein Fabian. (2007). *Mathematische Grundlagen der Warteschlangentheorie / Markov-Ketten*. Grin Verlag. ISBN: 3638774155
- [Schmidt06] Schmidt K., Weigang J., and Rose O. (2006). *Modeling semiconductor tools for small lotsize fab simulations*, in Simulation Conference, WSC 06. Proceedings of the Winter, pp. 1811 – 1816.
- [Schnell98] Schnell Reiner, Hill Paul B., Esser Elke. (1998). *Methoden der empirischen Sozialforschung*. Oldenbourg R. Verlag GmbH. Auflage: 5. ISBN 3486234897
- [Scholl08] Scholl, Wolfgang. (2008). *Coping with typical unpredictable incidents in a logic fab*. In Proceedings of the 2008 Winter Simulation Conference, ed. S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, and J. W. Fowler, 2030\_2034. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- [Scholl10] Scholl, W. B.P. Gan, D. Noack, P. Preuss, M. L. Peh, P. Lendermann, and O. Rose. 2010. *“Towards realization of a high-fidelity simulation model for short-term horizon forecasting in wafer fabrication facilities.”* In *Proceedings of the 2010 Winter Simulation Conference*, eds. B. Johansson, S. Jain, J. Montoya-Torres, J. Huan, and E. Yücesan, 2563–2574. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- [Scott79] Scott David W. (1979). *On optimal and data-based histogram*. In: *Biometrika*. 3, Nr. 66, S. 605–610

## Literaturverzeichnis

---

- [Shanth07] J. G. Shanthikumar, S. Ding, and M. T. Zhang. (2007) *“Queueing theory for semiconductor manufacturing systems: a survey and open problems”* IEEE Trans. Autom. Sci. Eng., vol. 4, no. 4, pp. 513–522.
- [Sommerv01] Sommerville Ian. (2001). *Software Engineering*. Pearson Studium, ISBN 3-8273-7001-9. Englische Originalausgabe: Addison-Wesley, 6th ed., ISBN 0-201-39815-X, 2001.
- [Spinn11] Spinnarke Sabine. (2011). *Japan: Stromausfälle setzen Halbleiterherstellern zu*. Internetpresänz; Abrufdatum: 12.01.2012:  
<http://www.produktion.de/automatisierung/Industrieelektronik/japan-stromausfaelle-setzen-halbleiterherstellern-zu/>
- [Störrle05] Störrle Harald. (2005). *UML 2 erfolgreich einsetzen. Einführung und Referenz*. Addison-Wesley, München. ISBN-10: 3827322685
- [Sullivan77] Sullivan Wiliam G., Claycombe Wiliam W. (1977). *Fundamentals of Forecasting*. Prentice Hall. ISBN 0879093005
- [Qiao04] Qiao Fei, Li Li, Wu Qidi, [2004]. Colored Petri net based hierarchical scheduling model for semiconductor production line. Intelligent Control and Automation, 2004. WCICA 2004. Fifth World Congress on 15-19 June 2004. Volume: 4 Page(s): 3014 - 3018 Vol.4
- [Uslar07] Mathias Uslar. (2007). *Referenzmodelle und Pattern in der Modellierung wissensintensiver Prozesse im Software Engineering*. GRIN Verlag, Auflage: 1. ISBN-13: 978-3638705059
- [VDI05] VDI: (2005) *VDI-Handbuch Materialfluss und Fördertechnik. Bd. 8: Materialfluss II (Organisation/Steuerung)*. VDI Verlag. Düsseldorf
- [Vogel97] Vogel, F. (1997). *„Beschreibende und Schließende Statistik“*. Oldenbourg: München, Wien, 6. Auflage.
- [Wald55] Wald, Abraham. (1955). *Selected Papers in Statistics and Probability*. McGraw-Hill. New York.
- [Weber91] Weber, Karl. (1991). *Prognosemethoden und -Software*. Schulz-Kirchner Verlag. ISBN 978-3-8248-0018-6
- [Zeileis03] Zeileis A, Kleiber C, Krämer W, Hornik K (2003). *Testing and dating of structural changes in practice*. Computational Statistics and Data Analysis 44: 109 – 123
- [Zisgen08] Zisgen Horst, Meents Ingo, Wheeler Benjamin R., Hanschke Thomas (2008). *A queueing network based system to model capacity and cycle time for semiconductor fabrication*. In Proceedings of the 2008 Winter Simulation Conference, ed. S. J. Mason, R. R. Hill, L. Mönch, O. Rose, T. Jefferson, and J. W. Fowler. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.