

# Analytische Modellbildung von Fertigungssystemen

## – Eine Fallstudie

Ottmar Gühr  
IBM Deutschland  
Zentrum für Produktionstechnik  
Max-Eyth-Straße 6  
7032 Sindelfingen

Hermann Gold, Phuoc Tran-Gia  
Universität Würzburg  
Lehrstuhl für Informatik III  
Am Hubland  
8700 Würzburg

### Überblick

In diesem Artikel möchten wir die Möglichkeiten, Vorteile und Grenzen der analytischen Behandlung von logistischen Problemen in der Fertigungsindustrie aufzeigen. Für Modellbildung und Analyse werden die Methoden der Warteschlangentheorie angewendet. Die Vorteile einer analytischen Lösung liegen in der exakten und schnellen Berechnung der Resultate. Dies ist besonders wichtig bei extensiven Parameterstudien, die bei der Optimierung eines Systems oft notwendig sind. Die Grenzen der Analyse werden durch approximative Verfahren hinausgeschoben, sind jedoch durch die Komplexität oder die Funktionalität des Systems gegeben.

Das System der Gruppenbedienung wird exemplarisch behandelt. Die Analyse erfolgt mit der Methode der eingebetteten Markov-Kette.

## 1 Einleitung

Die Modellbildung und Leistungsanalyse von Fertigungssystemen gewinnt immer mehr an Bedeutung durch die steigende Komplexität der Systeme und den stärker werdenden internationalen Wettbewerb, der eine effiziente Betriebsweise erfordert. In diesem Beitrag möchten wir die Möglichkeiten und Vorteile der analytischen Modellbildung und Analyse aufzeigen und an dem Fallbeispiel der Dimensionierung von Gruppenbearbeitungsmaschinen demonstrieren.

Bei der analytischen Untersuchung von Fertigungssystemen wird aus dem Fertigungssystem ein stochastisches Modell abgeleitet, das mit Methoden der Warteschlangentheorie untersucht wird. Dabei ist man im allgemeinen an den Ergebnissen für das eingeschwungene (stationäre) System interessiert. Die Vorteile der analytischen Untersuchung liegen in der schnellen und exakten Berechnung der Resultate. Dies ist besonders wichtig für Parameterstudien oder Optimierung der Systemparameter.

Typische Methoden für die Analyse einfacher Systeme, meist bestehend aus Warteraum (Puffer) und Bedieneinheit, sind Markov-Prozeß, Markov-Kette und eingebettete Markov-Kette. Bei komplexeren Systemen, wie einer Fertigungslinie, werden die Methoden Markov-Prozeß oder Dekomposition zur Analyse angewendet. Bei der Methode des Markov-Prozesses sind einige Einschränkungen notwendig, die im allgemeinen in Fertigungssystemen nicht genügend gut erfüllt sind. Bei der Methode der Dekomposition wird das Gesamtmodell in die einzelnen Systeme zerlegt, die jeweils einzeln effizient untersucht werden können. Diese Dekomposition ist für einige Klassen von Modellen exakt, die auch mit der Methode des Markov-Prozesses untersucht werden können, aber im allgemeinen approximativ mit genügend guter Genauigkeit.

Die berechneten Leistungsparameter sind Durchsätze, Durchlaufzeiten ('Manufacturing Cycle Time'), Wartezeiten und Teilebestände ('Work in Progress'). Die Berechnung der Leistungsparameter ist im allgemeinen vernachlässigbar kurz.

Alternativ können die Systeme auch mit der Methode der Simulation untersucht werden. Die Simulation hat den Vorteil einer sehr genauen realitätsnahen Abbildung. Die Ausführungszeiten eines Simulationslaufs sind im allgemeinen sehr lang. Einige Stunden bis einige Tage sind keine Seltenheit. Daher müssen Simulationsläufe sehr genau geplant werden. Dies ist sehr lästig, wenn man einen großen Parameterraum abtesten will. Eine analytische Lösung ist im Gegensatz dazu sehr schnell, im Sekundenbereich, berechnet. Daher ist eine interaktive Arbeitsweise möglich, was dem Modellierer erlaubt während der Berechnung in seinem Denkprozeß zu bleiben.

Im nächsten Abschnitt werden wir die einzelnen Methoden detailliert einführen und ihre Vor- und Nachteile besprechen. Im dritten Abschnitt werden wir die Analyse von Gruppenbedienmaschinen ('Batch-Tools') mit der Methode der eingebetteten Markov-Kette darstellen. Durch die effiziente Berechnung der Leistungsgrößen lassen sich umfangreiche Parameterstudien durchführen, mit denen eine geeignete Dimensionierung durchgeführt werden kann.

Abschließend werden wir weitergehende Modelle diskutieren, wie sie bei komplexen Fertigungslinien auftreten. Dies umfaßt die 'KANBAN'-Steuerung für die Reduzierung der Teilebestände und die Steuerung nach dem Fälligkeitsdatum von Aufträgen.

## 2 Methoden der Analyse von Fertigungssystemen

In diesem Abschnitt werden wir die Methoden Markov-Prozeß, Markov-Kette, eingebettete Markov-Kette und Dekomposition für die Analyse von Fertigungssystemen darstellen. Es wird dabei besonderer Wert auf eine anschauliche Darstellung gelegt, und weniger auf eine mathematisch vollständige Darstellung.

### 2.1 Markov-Prozeß

Das charakteristische Kriterium eines Markov-Prozesses ist die Eigenschaft, daß die zukünftige Entwicklung des Prozesses nicht von der Vergangenheit abhängt. Dieser Sachverhalt wird auch oft als Gedächtnislosigkeit bezeichnet und gilt für jeden Zeitpunkt. Dies bedingt für die Verweilzeiten in den Zuständen eine negativ exponentielle Verteilung. Dies gilt ebenfalls für alle anderen Verteilungsfunktionen. Die stationären Zustandswahrscheinlichkeiten können über ein lineares Gleichungssystem bestimmt werden. Die Leistungsparameter sind aus den Zustandswahrscheinlichkeiten berechenbar. Falls der Zustandsraum unendlich ist, kann oft mit geeigneten Reihenentwicklungen eine geschlossene Lösung gefunden werden. Für einen endlichen, aber grossen Zustandsraum kann mit numerischen Methoden eine genügend genaue Lösung bestimmt werden.

Die negativ exponentielle Verteilungsfunktion bedeutet jedoch für die Anwendung auf Probleme in der Fertigungsindustrie eine sehr starke Einschränkung. So sind zum Beispiel Bearbeitungszeiten an Maschinen im allgemeinen als konstant und nicht als negativ exponentiell anzusehen. Mit dieser Methode können daher meist lediglich sehr grundsätzliche Fragestellungen angegangen werden. So wurde etwa in einer Arbeit von Gold und Hübner [1] der prinzipielle Unterschied zwischen „Push“- und „Pull“-Steuerungsmechanismen anhand eines  $M/M^{[X]}/N$  Gruppenbediensystems erörtert.

### 2.2 Markov-Kette

Werden die Zustandsübergänge des Systems ohne Berücksichtigung der Zeit ausgeführt, so spricht man von einer Markov-Kette. Eine Markov-Kette ist vollständig beschrieben durch die Zustandsübergangsmatrix. Für den eingeschwungenen Systemzustand (einer aperiodischen rekurrenten Markov-Kette) muß für die Lösung ebenfalls ein lineares Gleichungssystem gelöst werden. Die Markov-Kette findet vor allem Anwendung bei der zeitdiskreten Analyse von Systemen. Ein typisches Beispiel sind getaktete Systeme.

### 2.3 Eingebettete Markov-Kette

Die Methode der eingebetteten Markov-Kette wird angewendet, wenn der zu analysierende Prozeß kein Markov-Prozeß ist, aber Zeitpunkte in dem Prozeß gefunden werden können, an denen der Prozeß die Eigenschaft der Gedächtnislosigkeit besitzt. An diesen Punkten kann eine Markov-Kette in den zeitkontinuierlichen Prozeß eingebettet werden. Die Markov-Kette kann analytisch gelöst werden. Ausgehend von den Zustandswahrscheinlichkeiten an den eingebetteten Punkten können dann die Zustandswahrscheinlichkeiten zu jedem beliebigen Zeitpunkt berechnet werden. In dem

nächsten Hauptabschnitt werden wir diese Methode an dem Beispiel des Gruppenbediensystems demonstrieren.

## 2.4 Dekomposition

Die Methode der Dekomposition wird auf Fertigungslinein angewendet, die aus mehreren Maschinen bestehen. Das Prinzip dieser Methode ist eine Dekomposition der Fertigungsline in einzelne Maschinen, die getrennt analysiert werden. Als verbindendes Element dient der Teilefluß von Maschine zu Maschine. Um die stationären Teileflüsse zu ermitteln müssen ein lineares und ein nicht-lineares Gleichungssystem gelöst werden. Die Leistungsparameter sind bei bekannten Teileflüssen an jeder Maschine einfach mit expliziten Formeln berechenbar. Für diese Methode wurden Softwarepakete entwickelt, die die Anwendung ähnlich einfach machen wie bei der Simulation.

## 3 Fallstudie: Dimensionierung von Gruppenbediensystemen

### 3.1 Modellierung

Ein Gruppenbediensystem stellt eine Maschine dar, die bei einem Arbeitsgang eine grössere Anzahl von Teilen gleichzeitig bearbeiten kann. Ein typisches Beispiel aus der Halbleiterfertigung sind Öfen. Ein Ofen kann z.B. mit 100 Wafern gefüllt werden. Nachdem der Ofen gestartet wurde, kann er allerdings nicht mehr geöffnet werden, bis dieser Prozeßschritt, der einige Stunden dauern kann, abgeschlossen ist. Eine wichtige Fragestellung ist nun, wann der Ofen gestartet werden soll, falls die vorliegenden Teile nicht zu einer vollen Ofenladung ausreichen.

Wir betrachten nun ein solches Gruppenbediensystem, das auf zwei verschiedene Arten in einer Fertigungsumgebung betrieben werden kann. Dabei wird durch die Einführung geeigneter lokaler Steuerungsmechanismen der Weg zur Just-in-Time Produktion geebnet. Bei der ersten Betriebsweise sprechen wir von einem „Push“-System, weil der Teilevorrat den Motor des Prozeßflusses darstellt (siehe Bild 1).

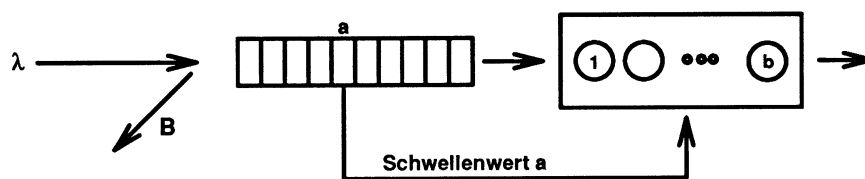


Bild 1: Das „Push“-Modell

Bei der zweiten Betriebsweise werden gemäß dem angezeigten Bedarf der nachfolgenden Maschine an vorverarbeiteten Teilen Rohteile in das Gruppenbediensystem zur Bearbeitung „hineingezogen“. Wir sprechen deshalb von einem „Pull“-System (siehe Bild 2). Beide Varianten des Gruppenbediensystems werden nun mit einer Steuerungsregel betrieben, mit deren Hilfe die Ziele der Just-in-Time Produktion – kurze Durchlaufzeiten, geringer WIP, Continuous Flow – realisiert werden sollen, bei gleichzeitiger wirtschaftlicher Nutzung der Ressourcen. Diese Regel lautet wie folgt: Wenn die Bedieneinheit frei ist und sich weniger als eine Anzahl  $a$  von Aufträgen in der Warteschlange befinden, so bleibt sie inaktiv bis sich  $a$  Aufträge in der Warteschlange angesammelt haben und beginnt sofort mit dem Eintreten dieses Ereignisses die Anzahl  $a$  von Aufträgen zu bearbeiten. Befindet sich nach Beendigung einer Bedienphase eine Anzahl von mehr als  $a$  Aufträgen in der Warteschlange, so wird sofort mit der Bedienung der wartenden Aufträge, jedoch nur mit einer Anzahl nicht größer als die maximale Kapazität  $b$  der Bedieneinheit, begonnen. Die Abfertigungsdisziplin ist dabei entsprechend der Reihenfolge bei der Ankunft (First Come First Served - FCFS).

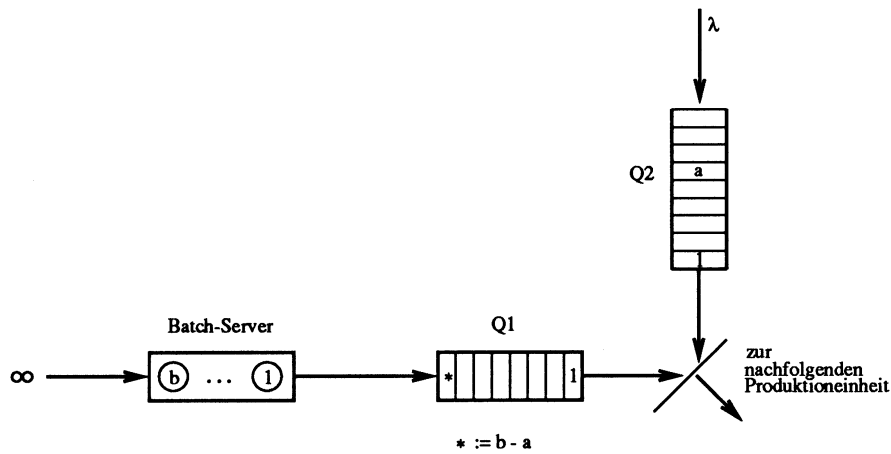


Bild 2: Das „Pull“-Modell

Im Falle des „Push“-Systems handelt es sich beim Warteraum um das Lager für vorrätige Teile, bei den Aufträgen um die Teile selbst. Für das „Pull“-System ist dagegen „Auftrag“ im eigentlichen Sinn zu verstehen, nämlich als Bedarfsanzeige für ein fertiges Teil, die sich in eine Warteschlange einreicht und gemäß FCFS bearbeitet wird. Vor dem „Pull“-System wird ein unendlicher Vorrat an rohen Teilen vorausgesetzt. Der Ankunftsstrom von Teilen im „Push“-Modell bzw. von Aufträgen im „Pull“-Modell habe die Markov-Eigenschaft (Gedächtnislosigkeit). Die Bedienzeit sei für beide Modelle allgemein verteilt. Aufträge bzw. Teile, die auf eine volle Warteschlange treffen, werden blockiert.

### 3.2 Analyse

Die Analyse der beiden vorgestellten Gruppenbedienmodelle, die wir gemäß der Kendall’schen Notation mit  $M/G^{[a,b]}/1-S$  (Push) bzw.  $M/G^{[a,b]}/1-S$  (Pull) bezeichnen, erfolgt mit Hilfe der Methode der eingebetteten Markov-Kette. Dazu betrachten wir je einen zweidimensionalen Zustandsprozess in der Zeit mit der Anzahl der wartenden Aufträge und der Restbedienzeit der Auftragsgruppe, die aktuell bedient wird. In diesen Prozess betten wir eine Markov-Kette ein, in dem wir die Betrachtung auf die Abgangszeitpunkte der Auftragsgruppen aus der Bedieneinheit, sogenannte Einbettungszeitpunkte, beschränken, zu denen der zweidimensionale Prozeß gedächtnislos ist. Eine Matrix der Übergangswahrscheinlichkeiten, die den Zusammenhang zwischen benachbarten Zuständen der eingebetteten Markov-Kette beschreibt, wird nun wiederholt auf den Initialzustand des jeweiligen Systems multiplikativ angewandt, bis wir die Zustandswahrscheinlichkeiten im stationären Zustand zu den Einbettungszeitpunkten erhalten. Diese führen uns zu den Zustandswahrscheinlichkeiten im stationären Zustand zu beliebigen Zeitpunkten, aus denen wir wiederum die charakteristischen Leistungsgrößen berechnen.

Sei nun die Wahrscheinlichkeit dafür, daß sich die Warteschlange für Teile im Falle des „Push“-Modells bzw. für Aufträge im Falle des „Pull“-Modells zu Einbettungszeitpunkten im Zustand „j Aufträge in der Warteschlange“ mit  $x(j)$ ,  $j = 1, \dots, S$  (Anzahl der Warteplätze) bezeichnet. Die entsprechenden Wahrscheinlichkeiten zu beliebigen Zeitpunkten seien mit  $x^*(j)$  bezeichnet.

Dann ergeben sich die charakteristischen Größen für das „Push“-Modell wie folgt.

Blockierungswahrscheinlichkeit  $B$ :

$$B = x^*(S) \tag{1}$$

mittlere Warteschlangenlänge:

$$EX^* = \sum_{i=0}^S i x^*(i) \tag{2}$$

mittlere Wartezeit eines Teiles bis zur Verarbeitung:

$$EW = \frac{EX^*}{\lambda(1 - B)}, \quad (3)$$

mittlere Anzahl von Teilen in der Bedieneinheit:

$$EY = \lambda(1 - B)EH. \quad (4)$$

mittlere Größe der Gruppen, in denen Aufträge gemeinsam bedient werden:

$$EY^{(A)} = a \sum_{i=0}^{a-1} x(i) + \sum_{i=a}^{b-1} i x(i) + b \sum_{i=b}^S x(i). \quad (5)$$

Die Formeln (1), (2), (3) und (4) gelten sinngemäß auch für das „Pull“-System, wobei die Größe  $EW$  gleichzeitig auch als die mittlere Systemzeit verstanden werden darf. Weiterhin werden für das „Pull“-System folgende charakteristische Größen betrachtet.

mittlere Anzahl von fertigen Teilen im Zwischenlager Q1:

$$EX_1^* = \sum_{i=a-b}^0 i x_1^*(i), \quad (6)$$

gewichtete Gesamtzahl von Teilen im Gruppenbediensystem (‘Work in Progress - WIP’):

$$WIP = \alpha_2 EX_1^* + \beta_2 EY^*. \quad (7)$$

Kosten:

$$COST = \alpha_1 EX_1^* + \beta_1 EX^*. \quad (8)$$

Für eine ausführliche Darstellung der Analyse sei auf die Bericht [2] und [3] verwiesen.

### 3.3 Numerische Ergebnisse

Wir präsentieren nun einige ausgewählte numerische Ergebnisse für ein Gruppenbediensystem mit folgenden Parametern:

Kapazität der Bedieneinheit  $b = 32$   
 Größe des Warteraums  $S = 64$   
 mittlere Bedienzeit  $EH = 1$

Im Bild 3 ist die mittlere Wartezeit im „Push“-System als Funktion der Verkehrsintensität dargestellt. Darin enthalten ist eine Schar von Kurven mit verschiedenen Schwellenwerten ( $a = 4$ ,  $a = 16$ ) und verschiedenen Variationskoeffizienten für die Bedienzeit. Für den Fall deterministischer Bedienzeiten erkennen wir, daß der Schwellenwert  $a = 4$  das bessere Wartezeitverhalten hervorbringt. Um stets die geringstmögliche Wartezeit zu garantieren, hätte man den Schwellenwert  $a = 1$  zu wählen, d.h. der Zwang zur Gruppenbildung wäre gänzlich aufzuheben. Bei schwankenden Bedienzeiten überschneiden sich die Wartezeitkurven für die Schwellenwerte  $a = 4$  und  $a = 16$ , da bei hoher Verkehrsintensität eine effiziente Nutzung der Bedieneinheit die Wahl einer hohen Schwelle erfordert. Bild 4 zeigt die Kosten gemäß Gleichung (8) in Abhängigkeit der Größe des Schwellenwertes  $a$  für das „Pull“-System. Dabei wurde  $\alpha_1 = 5$  und  $\beta_1 = 1$  gewählt.

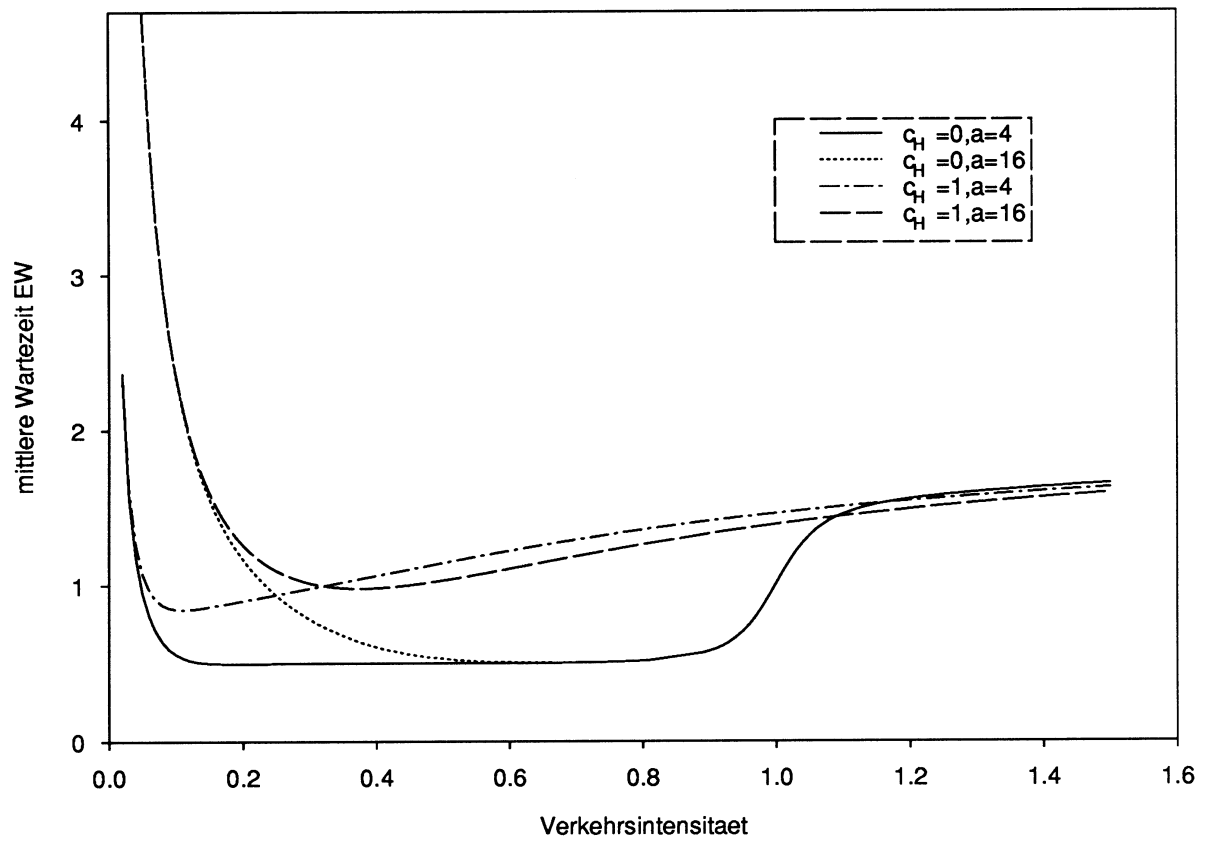


Bild 3: Wartezeitverhalten im „Push“-System

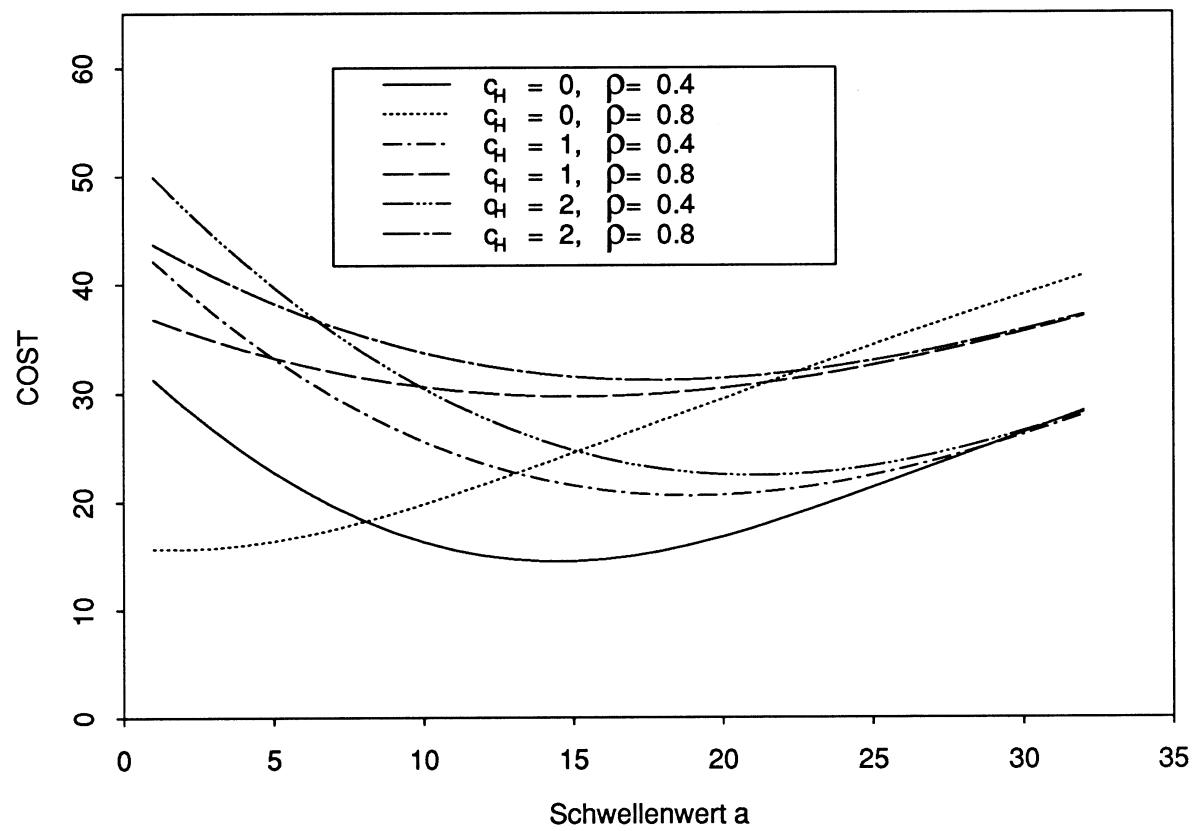


Bild 4: Kosten im „Pull“-System

## 4 Analytische Behandlung weiterführender Modelle

### 4.1 Due Date Scheduling

Das „Due Date Scheduling“ im Rahmen einer Due-Date-Managementstrategie ist vielleicht die konsequenteste Art und Weise, das zentrale Ziel der Just-in-Time Philosophie - ein Produkt eben gerade rechtzeitig fertigzustellen - zu verwirklichen. Ausgehend von einem Fälligkeitsdatum für das fertige Produkt, das mit realistischem Blick für das Machbare festgelegt wurde, werden entlang des Prozessflusses, jedoch in Gegenrichtung, relative „Due Dates“ an jeder Arbeitsstation bestimmt. Aufgabe des „Due Date Scheduling“ ist es nun, den Prozeßfluß so zu steuern, daß die relativen „Due Dates“ und insbesondere die Fälligkeitsdaten für die fertigen Produkte möglichst gut eingehalten werden. Als Konkretisierungen von „Due Date Schedules“ kommen beispielsweise die folgenden Abfertigungsdisziplinen in Betracht:

- Earliest Due Date First (EDD)
- Prioritätszuweisung für Aufträge mit Verspätung
- eine begrenzte Erlaubnis zur Überholung in Verspätungsfällen

Gegenwärtig findet eine analytische Untersuchung von Produktionssystemen, die nach der Due-Date-Managementstrategie und mit einer Abfertigungsstrategie nach Priorität für verspätete Aufträge betrieben werden, statt. Die „Due Dates“ werden dabei auf der Basis einer BCMP-Voranalyse (geeignete numerische Methode zur Analyse von Netzwerken aus Maschinen, die einen Markov-Prozeß bilden) der Produktionslinie bestimmt. Daraus resultiert ein Netz mit verschiedenen Auftragsklassen, Klassenwechsel und Prioritäten. Dieses wird rückverwandelt in ein Netz ohne Prioritäten durch Anpassung der Bediencharakteristiken für jede Auftragsklasse an jedem Prioritätsknoten. Eine nochmalige Analyse gibt Auskunft über die Fähigkeit, mit dem angewandten ‘Due Date Schedule’ und seinen speziellen Parametern Fälligkeitsdaten einzuhalten.

### 4.2 Kanban-Systeme

Die Implementierung der Just-in-Time-Philosophie durch ein Kanban-System verteilt die Kräfte zum „Ziehen“ von Materialien in die Produktionslinie hinein und schließlich von fertigen Produkten aus der Produktionslinie heraus auf geeignet gewählte Produktionssektoren. Diese bestehen aus einer Anzahl von Produktionseinheiten, innerhalb derer die Gesamtzahl von in Bearbeitung befindlichen Teilen (WIP) beschränkt ist. Dazu erhält jeder Sektor eine auf seine spezifische obere Schranke des ‘Work in Progress’ bemessene Anzahl von Kanbans (japanisches Wort für Karte), die gleichsam Berechtigungen zur Weiterverarbeitung von genau einem Teiles darstellen. Verfügbare Kanbans in einem Sektor üben so eine Zugwirkung auf den vorangehenden Sektor aus. Ein Mangel an Kanbans verhindert andererseits die Überschwemmung des Sektors mit vorverarbeiteten Teilen, die ohnehin gegenwärtig nicht weiterverarbeitet werden könnten und nur eine unnötige Erhöhung des ‘Work in Progress’ bewirkten.

Die Modellierung von Kanban-Systemen ist möglich, in dem man die einzelnen Sektoren als geschlossene Warteschlangennetze betrachtet, die zusätzlich sogenannte „Stop-and-Go“-Bedieneinheiten zur Reflektion der hemmenden Einflüsse benachbarter Sektoren enthalten. Dieses Modell läßt sich approximativ analysieren und kann zur Vorauswahl geeigneter Segmentierungen und Verteilungsmuster für Kanbans bezüglich einer gegebenen Produktionslinie zu Hilfe genommen werden.

## 5 Zusammenfassung

Wir haben in diesem Artikel die Methoden Markov-Prozeß, Markov-Kette, eingebettete Markov-Kette und Dekomposition zur mathematischen Analyse von Fertigungssystemen dargestellt. Grundsätz-

liche Ergebnisse können mit der Methode des Markov-Prozesses gewonnen werden. Die Abbildungstreue eines Fertigungssystems ist jedoch sehr beschränkt. Mit der Methode der eingebetteten Markov-Kette kann die Abbildungstreue erhöht werden. Die Methode der Dekomposition wird bei Fertigungslinien angewendet, die aus mehreren Maschinen bestehen.

Die Methode der eingebetteten Markov-Kette haben wir dann angewendet auf die Analyse von Gruppenbediensystemen. Abschliessend diskutierten wir die weitergehenden Modelle „Due-Date-Scheduling“ und KANBAN für die Steuerung von Produktionslinien.

## Literaturverzeichnis

- [1] H. Gold, F. Hübner, “Multi Server Batch Service Systems in Push and Pull Operating Mode - a Performance Comparison” Research Report Nr. 26, Institute of Computer Science, University of Würzburg, 1990.
- [2] H. Gold, P. Tran-Gia, “Performance Analysis of a Batch Service Queue arising out of Manufacturing System Modelling”, Research Report Nr. 16, Institute of Computer Science, University of Würzburg, 1990.
- [3] H. Gold, H. Grob, “Performance Analysis of a Batch Service System Operating in Pull Mode” Research Report Nr. 27, Institute of Computer Science, University of Würzburg, 1991.