Betriebswirtschaftliche Forschungsergebnisse

Band 132

„Losüberlappung“

Verfahren zur Effektivitätssteigerung in der operativen Produktionsplanung

Von

Martin Feldmann

Duncker & Humblot · Berlin

MARTIN FELDMANN

„Losüberlappung“

Betriebswirtschaftliche Forschungsergebnisse

Begründet von

Professor Dr. Dr. h. c. mult. Erich Kosiol (1899 – 1990)

Fortgeführt von dessen Schülerkreis

Herausgegeben von

Professor Dr. Ernst Troßmann
Universität Hohenheim

in Gemeinschaft mit

Professor Dr. Oskar Grün
Wirtschaftsuniversität Wien

Professor Dr. Wilfried Krüger
Justus-Liebig-Universität Gießen

Professor Dr. Hans-Ulrich Küpper
Ludwig-Maximilians-Universität München

Professor Dr. Gerhard Schewe
Westfälische Wilhelms-Universität Münster

Professor Dr. Axel von Werder
Technische Universität Berlin

Band 132

„Losüberlappung“

Verfahren zur Effektivitätssteigerung in der operativen Produktionsplanung

Von

Martin Feldmann

Duncker & Humblot · Berlin

Die Fakultät für Wirtschaftswissenschaften der Universität Bielefeld hat diese Arbeit im Jahre 2004 als Habilitationsschrift angenommen.

Bibliografische Information Der Deutschen Bibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über <http://dnb.ddb.de> abrufbar.

Fotoprint: Berliner Buchdruckerei Union GmbH, Berlin
Printed in Germany

ISSN 0523-1027
ISBN 3-428-11857-X

Gedruckt auf alterungsbeständigem (säurefreiem) Papier
entsprechend ISO 9706 ♾

Internet: http://www.duncker-humblot.de

Vorwort

Die vorliegende Monographie entstand als Habilitationsschrift während meiner Tätigkeit als wissenschaftlicher Assistent am Lehrstuhl für Betriebswirtschaftslehre und Unternehmensforschung an der Fakultät für Wirtschaftswissenschaften der Universität Bielefeld.

Zum Gelingen dieser Arbeit haben viele Kollegen und Freunde beigetragen, denen ich an dieser Stelle ganz herzlich danken möchte. Mein sehr verehrter akademischer Lehrer Herr Professor Dr. Klaus-Peter Kistner hat an seinem Lehrstuhl stets für ein produktives und angenehmes Arbeitsklima gesorgt und mir in vielen Punkten mit sehr guten Ratschlägen weitergeholfen. Meine Kollegin PD Dr. Marlies Rogalski fand immer Zeit für ein fachliches Feedback, leider hat sie die Fertigstellung dieser Arbeit nicht mehr erleben dürfen. Umso häufiger hat mich meine Kollegin Dipl.-Kff. Stephanie Müller mit ihren kritischen Hinweisen unterstützt, wenn es darum ging, diverse Modelle und Argumentationen zu diskutieren. Frau Dorothea Becker hat unzählige Seiten Korrektur gelesen und die Zahl meiner Tippfehler signifikant reduziert.

Mit meinem ehemaligen Kollegen, guten Freund und mittlerweile vielfachem Koautor Dr. Dirk Biskup konnte ich immer rechnen und habe diese Zusammenarbeit stets als äußerst produktiv, fair und zielgerichtet erlebt.

Den Herren Professoren Dr. Hermann Jahnke und Dr. Christian Bierwirth sei ganz herzlich für die zügige Begutachtung und ihre Hinweise gedankt. Dem Herausgeberkreis und insbesondere Herrn Professor Dr. Ernst Troßmann danke ich für die Aufnahme in die Reihe „Betriebswirtschaftliche Forschungsergebnisse“ sowie dem Team der Duncker & Humblot GmbH für die gute Betreuung.

Mit ihrer moralischen Unterstützung und mit ihrem Verständnis für dieses zeitintensive Projekt hat mir meine Familie immer sehr geholfen. Meiner Frau Christiane möchte ich dieses Buch widmen.

Bielefeld, im Juli 2005 *Martin Feldmann*

Inhaltsverzeichnis

Tabellenverzeichnis

Abbildungsverzeichnis

Symbolverzeichnis

j	Index der Freigabelose (j = 1,..., J)
m	Index der Stufen (m = 1,..., M)
i	Index der Transferlose bei Stückverfügbarkeit (i =1 ,....,I)
e, g, h, k, u	Hilfsindices
a_m	antizipativer Rüstvorgang auf der Stufe m
c_g	Prüfgröße für die kritische Stufe g
C	Durchlaufzeit bei konsistenten Losen
C_{jm}	Zeitpunkt der Fertigstellung des j-ten Freigabloses auf der Stufe m
C_{min}	minimale Durchlaufzeit
C^{D}_{min}	minimale Durchlaufzeit bei diskreten Losen
C^{K}_{min}	minimale Durchlaufzeit bei kontinuierlichen Losen
d_j	Liefertermin für Los j
E_j	Verfrühung des j-ten Loses
$h(i,j)$	das letzte Freigabelos i das Teile enthält, die in das Freigabelos j eingehen
I	Zahl der Transferlose bei Stückverfügbarkeit
(j, m)	Knoten des j-ten Loses auf der m-ten Stufe im Netzwerk
J	das letzte Los, die Loszahl
(J, M)	Senke des Netzwerks mit J Losen und M Stufen
k'	das k-te Los im modifizierten Plan
K_o	Kapazität des Ofens
KiV	prozentuale Verschlechterung konsistente versus variable Losen
$KoKi$	prozentuale Verschlechterung konstante versus konsistente Lose
KoV	prozentuale Verschlechterung konstante versus variable Lose
m^*	Stufe mit der höchsten Rüstzeit
M	die letzte Stufe (Stufenzahl)

o	Ofenstufe
O	Stufenzahl im LINGO Kode
O(..)	Laufzeitkomplexität
p_m	Ausführungszeit je Stück auf Stufe m
p_o	Ausführungszeit des Ofens
$\underline{p}$	Vektor der Ausführungszeiten je Stück
$P(e,g)$	Summe der Ausführungszeiten von der Stufe e bis zur Stufe g
P(0)	Losüberlappungsproblem unter Losverfügbarkeit
P(1)	Mehrstufen-Mehrlosüberlappungsproblem bei konsistenten Losen
P(2)	Losüberlappungsproblem bei variablen Losen und Losverfügbarkeit
q	Quotient der Ausführungszeiten von zwei aufeinanderfolgenden Stufen: $q = p_2/p_1$
q_m	Quotient der Ausführungszeiten zweier aufeinanderfolgender Stufen im Mehrstufenfall
Q	Vorgegebene Losgröße
R	Sehr große Zahl (100.000)
S	Loszahl im LINGO Kode
t_m	Transportzeiten von der Stufe m zu der Stufe m+1
T_j	Verspätung des j-ten Loses
v_m	verbundener Rüstvorgang auf der Stufe m
x_{jm}	Größe des Loses j auf Stufe m
x_j	Größe des j-ten konsistenten Freigabeloses
$\underline{x}$	Vektor der Losgrößen bei konsistenten Freigabelosen
x	Größe des ersten Loses im Zweilosfall
$\tilde{x}_g$	obere Schranke für die Losgröße x, wenn g die kritische Stufe ist
$\underset{\sim}{x}_g$	untere Schranke für die Losgröße x, wenn g die kritische Stufe ist
X_j	Kumulierte Stückzahl über die ersten j konsistenten Lose
X_{jm}	Kumulierte Stückzahl über die ersten j Lose auf der Stufe m
y_{im}	Größe des i-ten Transferloses mit (i = 1,..., I) auf der Stufe m bei Stückverfügbarkeit
α_j	Verfrühungsstrafe je ZE

β_j	Verspätungsstrafe je ZE
δ_{jmk}	Binärvariable zur Abbildung der Verfügbarkeit.
$\delta_{jmk} = 1$	wenn das *j*-te Freigabelos auf der Stufe *m* nicht gestartet wird, bevor das *k*-te Freigabelos auf der Stufe *m*–1 beendet ist,
$\delta_{jmk} = 0$	in den übrigen Fällen
π	durchlaufzeitminimaler Plan
π'	modifizierter Plan
ρ	Verhältnis der Losgrößen im Zweilosfall
ω_{jm}	Binärvariable, nimmt den Wert 1 an, wenn $x_{jm} > 0$ ist.
@, #, |	Steuerzeichen des LINGO Kodes
Δ ZE	geringe Verlängerung in den Ausführungszeiten einer Stufe
Δ	Grenzeffekt eines zusätzlichen Loses
$\rightarrow$	Zuweisung

1. Einleitung

Bei der mehrstufigen Fertigung sind neben der Konkurrenz um die Maschinen einer Produktionsstufe auch Abhängigkeiten zwischen vor- und nachgelagerten Stufen zu berücksichtigen. Diese können als sachlich-horizontale und als zeitlich-vertikale Beziehungen auftreten (vgl. Steven, 1994). Hinzu kommen die Interdependenzen auf aggregiertem Niveau, die zwischen den Planungsaufgaben bestehen. Für die Bestimmung der Losgrößen müssen beispielsweise die Kapazitäten und insbesondere die benötigten Rüstzeiten bekannt sein. Diese sind aber erst mit der Festlegung der Reihenfolge gegeben. Umgekehrt kann über die Reihenfolge der Einlastung nur entschieden werden, wenn der Umfang der einzuplanenden Arbeitsgänge bekannt ist. Dieser liegt aber erst mit der Entscheidung über die Losgrößen fest.

Zur Berücksichtigung aller Interdependenzen bietet sich prinzipiell eine simultane Planung in einem (Total-)Modell an. Dieses Vorhaben muss allerdings scheitern, da bereits Komponenten des Problems schwer im Sinne der Komplexitätstheorie sind (vgl. Garey/Johnson, 1979; Bachem, 1980; Brüggemann, 1995). Im Fall der mehrstufigen Fertigung unter Kapazitätsrestriktionen kann bereits das Auffinden einer zulässigen Lösung zu einem schweren Problem führen (vgl. Maes et al., 1991). Um in dieser Situation die Planung durch geeignete Modelle unterstützen zu können, muss die Komplexität der Fragestellung reduziert werden.

Ein häufig angewandtes Vorgehen besteht darin, nur Teilaspekte in (Partial-) Modellen zu erfassen und diese dann sukzessiv zu behandeln. Dennoch bleiben die zu lösenden Fragen häufig so schwer, dass statt optimierender Verfahren Heuristiken eingesetzt werden müssen. Leisten spricht hier von einer Verringerung des Anspruchsniveaus (Leisten, 1996, S. 19). Im Kern werden somit drei gängige Konzepte zur Reduktion der Komplexität verwendet:

- Aufteilung des Planungsproblems,
- sukzessives Lösen,
- Einsatz von heuristischen Verfahren.

Ein Problem bei der Anwendung dieser Konzepte liegt in der breiten Akzeptanz bestimmter, einschränkender Denkmuster. So setzt eine Problemvereinfachung durch die Aufteilung des Planungsproblems die Identifikation der Aspekte voraus, an denen man sich orientieren will. Bereits bei dieser Identifikati-

on spielt aber die herrschende Lehrmeinung eine wesentliche Rolle, da Alternativen häufig unbeachtet bleiben, wenn sie nicht der aktuellen Auffassung, dem Mainstream folgen. Weiter sind für eine Vereinfachung durch Aufteilung Annahmen notwendig, mit denen jeder Teilaspekt des Problems von den anderen abgegrenzt wird. Diese Annahmen füllen die Lücken, die durch das Fallenlassen von Details und Zusammenhängen entstehen. Trotz ihrer hohen Bedeutung für die Modellentwicklung wird die Frage, ob eine Annahme gerechtfertigt ist, häufig nur am Rande behandelt. Stattdessen wird das Postulieren bestimmter Eigenschaften einer Problemstellung zum Standard. Aus der breiten Akzeptanz bestimmter Annahmen resultieren eingeschränkte Sichten auf das zu lösende Problem, die sich zur Lehrmeinung verdichten, wenn sie über längere Zeit vorherrschen. Auch den Bereich der Produktionsplanung dominieren bestimmte Annahmen beziehungsweise Kombinationen von Vereinfachungen. Zum Beispiel wird die Fragestellung:

„Wann sind die Teile für die Bearbeitung auf der nächsten Stufe verfügbar?“

überwiegend so modelliert, dass die geschlossene Weitergabe der Lose zwischen den Produktionsstufen vorausgesetzt wird oder dass jedes einzelne Stück unmittelbar nach seiner Produktion der nächsten Stufe zur Verfügung steht (vgl. Hackman/Leachman, 1989). Zwischenformen, wie z.B. die teilweise und gemeinsame Weitergabe, werden nur selten betrachtet (Hancock, 1991).

Als zweites Beispiel für ein gängiges, einschränkendes Denkmuster in der Produktionsplanung kann die Verteilung der Planungsaufgaben an die einzelnen Entscheidungsebenen genannt werden. Dieses Beispiel ist eng verknüpft mit dem ersten, da üblicherweise zunächst die Losgrößenplanung und nachgelagert – unter der Annahme der geschlossenen Weitergabe – die Reihenfolgeplanung vorgenommen wird. Die meisten Modelle der Reihenfolgeplanung setzen als Standardannahme voraus, dass eine Unterbrechung der Bearbeitung oder eine Aufteilung von gegebenen Aufträgen (den Jobs) nicht zulässig ist (French, 1982). Zunächst verhindert diese Annahme, dass die festgelegten Arbeitsinhalte, die Lose, in einer anderen Größe bearbeitet werden als in der Losgrößenplanung vorgesehen. Erweisen sich allerdings, z.B. auf Grund vernachlässigter Kapazitätsbeschränkungen, die ermittelten Losgrößen als nicht realisierbar, werden Anpassungsmaßnahmen eingeleitet. Diese verschieben beispielsweise genau die Menge zeitlich, die zur Unzulässigkeit führt, teilen die Lose zur parallelen Bearbeitung auf, passen die Kapazität der Anlagen kurzfristig an etc. (vgl. Zäpfel, 1996, S. 179 und 192). Dabei existiert mit der Losüberlappung eine Maßnahme, die einerseits wirkungsvoll ist und andererseits hilft, die Entscheidungen hinsichtlich der Losgrößen beizubehalten (vgl. Kurbel, 2003, S. 149 ff.). Sie soll hier wie folgt verstanden werden:

Bei der *Losüberlappung* (lot streaming) werden gegebene Lose stufenüberlappend bearbeitet: Teile des Loses können bereits auf der folgenden Stufe eine

Bearbeitung erfahren, während die Fertigstellung des restlichen Loses noch andauert.

Zur Verdeutlichung der Idee dient die folgende dreistufige Reihenfertigung zur Herstellung eines Produkts. Auf Stufe 1 werden zur Produktion je Stück 0,7 Zeiteinheiten (ZE), auf der zweiten Stufe 1,0 ZE und auf Stufe 3 0,6 ZE benötigt. Zudem wird davon ausgegangen, dass die Dauer der jeweiligen Transporte vernachlässigbar klein ist. Für ein geschlossen weitergegebenes Los mit 30 Teilen beträgt die Durchlaufzeit 69 ZE, da es auf der ersten Stufe 21 ZE, auf der zweiten Stufe 30 ZE und auf der dritten Stufe 18 ZE lang bearbeitet wird.

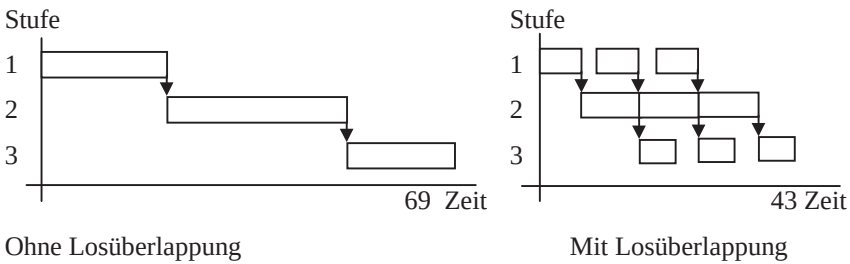


Abbildung 1: Situation ohne Losüberlappung versus mit Losüberlappung

Wird das Los nicht geschlossen in einem Transfer weitergegeben, sondern überlappend in drei Losen mit je 10 Teilen, kann die Durchlaufzeit von 69 ZE um fast 40% auf 43 ZE gesenkt werden. Obwohl diese Technik erheblich zur Reduktion der Durchlaufzeit beitragen kann, wird sie selbst in einschlägigen Veröffentlichungen nur selten erwähnt. In der betrieblichen Praxis hingegen zählt die Losüberlappung zu den regelmäßig verwendeten Techniken und auch die siebte OPT-Regel (vgl. Fry et al., 1992) empfiehlt, dass die Transport- und die Fertigungslosgrößen voneinander abweichen dürfen. Dennoch bieten viele der gängigen PPS-Systeme keine nennenswerte Entscheidungsunterstützung für die Fertigung mit überlappenden Losen. Insbesondere eine Integration der wissenschaftlich fundierten Ergebnisse und Verfahren zur Losüberlappung (vgl. z.B. Trietsch/Baker, 1993; Baker/Jia, 1993) in die bestehenden PPS-Systeme ist nicht erfolgt. Entsprechend stellen Studien zum Stand der PPS-Systeme fest, dass in vielen Systemen keine planerische Unterstützung bei der Disposition der – aus praktischer Sicht naheliegenden – Losüberlappung angeboten wird (vgl. Gronau/Ibelings, 2002; François/Wolf, 2001; Fandel/François, 2000).

Für die vorliegenden Untersuchung ergeben sich damit zwei Zielsetzungen, die zugleich die Struktur der Arbeit bestimmen:

In den Kapiteln zwei und drei wird den Ursachen für die geringe Resonanz nachgegangen, die die Losüberlappung bislang gefunden hat. Es wird untersucht, worin die Gründe dafür liegen können, dass ein in der Praxis bewährtes Vorgehen keine besondere Beachtung im Mainstream der Produktionsplanung

findet. Dies geschieht mit der Absicht, ein stärkeres Problembewusstsein zu wecken und einen Beitrag dazu zu leisten, dass das Forschungsgebiet der Losüberlappung einen höheren Stellenwert erhält.

In den Kapiteln vier und fünf werden ausgewählte Verfahren zur Losüberlappung vorgestellt, die belegen, dass eine Reihe von theoretisch fundierten Ansätzen existieren, die unmittelbar umgesetzt werden können und deren ökonomische Konsequenzen gut analysierbar sind. Die Auswahl erfolgt unter Berücksichtigung der drei wesentlichen Aspekte: Praxisrelevanz, Rechenbarkeit und Erkenntnisbeitrag. In Kapitel vier werden zunächst die Rahmenbedingungen und Einsatzmöglichkeiten der Losüberlappung aufgearbeitet und wichtige Begriffe eingeführt. Das fünfte Kapitel gliedert sich in drei Schwerpunkte und dient sowohl der Darstellung als auch der Erweiterung einzelner Verfahren. Im ersten Schwerpunkt werden die Ansätze vorgestellt, die für den Zwei- und den Dreistufenfall entwickelt wurden und die notfalls auch manuell umsetzbar sind. Im zweiten Schwerpunkt werden die Verfahren behandelt, die mit linearer Programmierung arbeiten und eine Aufteilung der Losgröße in konsistente Teillosen vornehmen. Es wird gezeigt, wie die Interpretation der Dualvariablen und die Sensitivitätsanalyse sinnvoll und praxisrelevant bei der Lösung von Losüberlappungsproblemen eingesetzt werden können. Diese Form der ökonomischen Analyse geht wesentlich über die bislang in der Literatur vorgeschlagene Vorgehensweise hinaus und ermöglicht neben der Beurteilung a priori auch eine angemessene Form der Reaktion auf Störungen a posteriori. Die bei der Sensitivitätsanalyse gefundenen Intervalle erlauben, diverse Anpassungsmaßnahmen hinsichtlich ihrer Konsequenzen zu beurteilen, so dass dem Disponenten eine What-If-Analyse ermöglicht wird. Im dritten Schwerpunkt wird ein Modell zur Festlegung variabler Lose unter Losverfügbarkeit bei kontinuierlicher Zeitführung vorgestellt, das eine seit Jahren bestehende Forschungslücke schließt. Das Potential dieser Losart wird untersucht und gezeigt, dass sich variable Lose sehr gut zur Reduktion der Durchlaufzeit eignen, wenn z.B. hohe Rüstzeiten vorliegen oder Terminabweichungen zu berücksichtigen sind. Das fünfte Kapitel endet mit einer Charakterisierung derjenigen Verfahren, die den oben genannten drei Aspekten – zumindest bislang – nicht genügen können.

Im Folgenden wird den Ursachen für die weitgehend fehlende Berücksichtigung der Techniken zur Losüberlappung nachgegangen. Da diese Ursachen schon auf der Ebene der Modellbildung und der damit verbundenen Wahrnehmung der Problemstellung ansetzen, ist mit der Argumentation auch dort zu beginnen.

2. Losüberlappung, Planung und Modellbildung

Eine Untersuchung der überlappenden Fertigung ist aus mehreren Gründen relevant: Neben dem Effekt den die überlappende Weitergabe auf die Durchlaufzeiten und damit einhergehend auf die Lagerbestände und die Kapitalbindung nimmt, betrifft sie nicht nur den Prozess der betrieblichen Planung, sondern verändert auch die betriebliche Struktur. Dieser Zusammenhang soll mit Rückgriff auf Chandler (1962) verdeutlicht werden. Wie Chandler mit seiner These „structure follows strategy" betont, stellt sich die Struktur des Unternehmens in Abhängigkeit von der Strategie ein. Eine auf Kostenführerschaft ausgerichtete Strategie wird zu anderen Strukturen im Unternehmen – etwa hinsichtlich der Ausstattung mit Maschinen – führen, als eine Strategie, in der eine schnelle Reaktion auf Nachfrageänderungen oder die Entwicklung von innovativen Produkten dominiert. Kistner/Steven (1991) weisen darauf hin, dass diese Wirkungskette um das eingesetzte Planungskonzept zu erweitern ist. Um Entscheidungen treffen zu können, die in die Richtung der gewünschten Strategie führen, ist ein Konzept festzulegen, nach dem geplant wird. Diesem kommt indirekt ein erheblicher Einfluss auf die Struktur des Unternehmens zu. Zusammenfassend lässt sich dies so formulieren: „structure follows planning concept." Je nach Wahl des Planungskonzepts sind die im Unternehmen aufzubauenden Strukturen anzupassen. Die Wahl des Konzepts impliziert zudem eine Entscheidung für eine Reihe von Modellen, die dem jeweiligen Konzept zugrunde liegen, beziehungsweise die erlauben, dass es umgesetzt wird. Diese Modelle bestimmen indirekt die Bahnen, in denen gedacht und nach Lösungen gesucht wird.

Als Beispiel kann die Entscheidung für ein bestimmtes PPS-System betrachtet werden, das dem Pull- oder dem Push-Prinzip folgt. Die Kapazitäten der Lager und der Transportmittel, die maschinelle Ausstattung etc. stellen sich in Abhängigkeit des PPS-Systems ein und sind damit abhängig von der Logik der Modelle, die ihm zugrunde liegen. Würde man beim Entwurf neuer PPS-Systeme von vornherein zulassen, dass Lose auch überlappend weitergegeben werden, dann resultieren daraus veränderte betriebliche Strukturen, die für diese Form der Fertigung günstige Bedingungen schaffen. Greift man hingegen auf Modelle zurück, die dem Paradigma der geschlossenen Fertigung folgen, dann ist zunächst die Sicht verstellt, dieses Potential wahrzunehmen. Werden die entsprechenden betrieblichen Strukturen und Regelungen für die geschlossene Weitergabe aufgebaut, ist ein späterer Wechsel zur überlappenden Fertigung nur durch eine umfassende Restrukturierung möglich.

Als zweites und sehr prominentes Beispiel für ein Paradigma, dem ein erheblicher Einfluss auf die betrieblichen Strukturen zukommt, kann die Arbeit von Wagner/Whitin (1958) genannt werden. Aus der Analyse ihres Modells folgt für optimale Lösungen eine Complementary Slackness Bedingung für den Lagerbestand und die Produktionsmenge in einer Periode. Obwohl nur wenige Jahre später Haehling von Lanzenauer (1970) zeigt, dass diese Bedingung nur bei unbeschränkter Kapazität Gültigkeit besitzt und dass es kostenminimal sein kann, über die Stufen variierende Losgrößen zu verwenden, hat es sehr lange gedauert, bis der Mainstream diese wichtigen Einschränkungen wahrgenommen hat. Da sich insbesondere aus der Complementary Slackness Bedingung erhebliche Vereinfachungen für den Planungsprozess ergeben, entstanden über viele Jahre – wie z.B. bei Moily (1986) angesprochen – Verfahren, die die Eigenschaft der Complementary Slacknesss übertragen und sie zur Vereinfachung nutzen, selbst wenn die Voraussetzungen nicht gegeben sind. Die Konsequenzen, die man aus den Eigenschaften optimaler Pläne nur im nichtkapazitierten Fall ziehen kann, sind außerdem in die Logik der PPS-Systeme eingeflossen.

Beide Beispiele verdeutlichen, dass die Entscheidung für ein Konzept, nach dem man planen will, respektive für den Einsatz eines PSS-Systems auch eine Entscheidung für die zugrundeliegenden Modelle und deren Annahmen ist. Diese Entscheidung hat Konsequenzen, die bis in die Struktur des Unternehmens reichen. Um den Ursachen für die geringe Resonanz der Losüberlappung nachzugehen, muss somit beim Prozess der Planung angesetzt werden.

Den Ausgangspunkt jeder Planung bildet nach Schneeweiß (1991) ein Gestaltungswunsch. Um diesen realisieren zu können, muss zunächst in einer Systemgrobanalyse festgestellt werden, über welchen Objektbereich sich die Planung erstrecken und welche Zielsetzung verfolgt werden soll. Dieses gedankliche Durchdringen geschieht, indem ein Modell entwickelt wird, das den zu untersuchenden Sachverhalt und die dort vorliegenden Zusammenhänge verdeutlicht. Dabei besteht eine unbestrittene Aufgabe darin, Ähnlichkeiten mit bereits untersuchten Phänomen zu erkennen und die dort erarbeiteten Ergebnisse zu nutzen. Insofern ist der Rückgriff auf bewährte Annahmen und Ergebnisse ein notwendiger Bestandteil bei der Modellbildung und dem darauf aufsetzenden Entwurf von Planungssystemen, wie z.B. den PPS-Systemen. Allerdings birgt der sinnvolle Rückgriff auf Bekanntes immer die Gefahr, dass ein reales Problem bereits in einer gewissen Erwartungshaltung betrachtet wird. Werden Ähnlichkeiten zu anderen, bereits untersuchten Fragestellungen deutlich, liegt es nahe, wieder in die bewährte Richtung zu denken. Es besteht die Gefahr, dass Modelle entwickelt werden, ohne zu überprüfen, ob die Annahmen tatsächlich gegeben sind.

Das Vorgehen des modellgestützten Planens wird hier in Anlehnung an Schneeweiß (1999, S. 260 sowie 1992, S. 4) verdeutlicht. Den Ausgangspunkt

bildet das Realproblem, z.B. die Suche nach einer geeigneten Vorgehensweise für die Produktionsplanung einer mehrstufigen, seriellen Fertigung. Aus dem Realmodell geht durch Abstraktion ein hinreichend genaues Abbild des zu bewältigenden Problems hervor. Dabei wird von bestimmten Aspekten, etwa von der Dynamik, abstrahiert. Schneeweiß bezeichnet diese Zwischenstufe als Masterproblem. Darauf aufbauend werden Zusammenhänge relaxiert, so dass ein „formal-mathematisches" Modell, das Formalmodell entsteht, welches der Entscheidungsfindung dient. Ein solches Formalmodell kann z.B. darin bestehen, dass das Problem aufgeteilt wird, wobei zunächst die Losgrößen geplant werden und anschließend die Lossequenzplanung vorgenommen wird.

Dabei hat der Schritt vom Master- zum Formal-Modell die Qualität einer Relaxation. Im Gegensatz zur Abstraktion werden bei der Relaxation Bedingungen oder Eigenschaften nur vorübergehend fallen gelassen (vgl. Chen/Thizy, 1990; Fisher, 1981).

Weiterhin sind rückkoppelnde Schritte zwischen den beiden Modellen vorgesehen. Die gefundenen Entscheidungen werden im Rahmen des Mastermodells überprüft, wobei festzustellen ist, ob sie auch für die nichtrelaxierte Situation akzeptabel sind. Man spricht hier von einer „Entscheidungs-Validierung", da die Prüfung möglicher Formalmodelle über die Betrachtung der gefundenen Entscheidungen erfolgt. Für kleine Ausschnitte des Mastermodells ist zudem eine empirische Validierung vorzusehen, mit der sichergestellt werden soll, dass das Mastermodell die Realität angemessen wiedergibt. Diese grobe Struktur gibt Abbildung 2 wieder (vgl. Schneeweiß, 1992, S. 4):

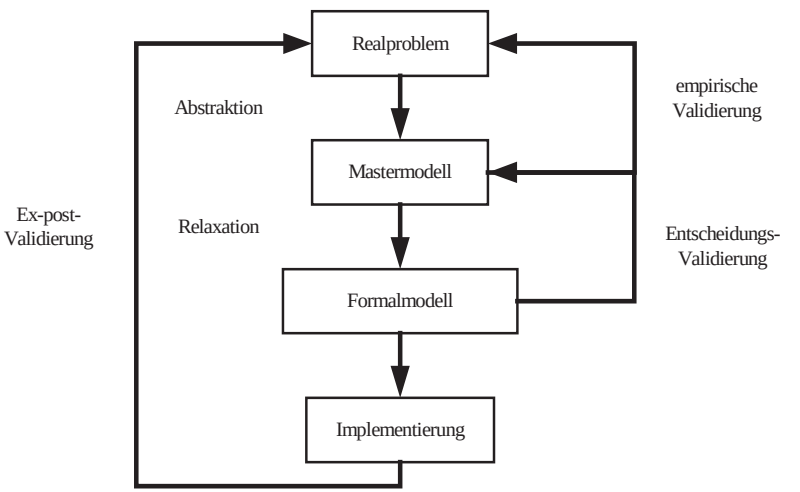


Abbildung 2: Die Grobstruktur des Planungsprozesses

Schneeweiß stellt fest, dass durch die Entkoppelung zwischen Master- und Formalmodell ein erheblicher Freiraum bei der Konstruktion der Formalmodelle entsteht. Es ist nicht wesentlich, dass die Formalmodelle die Realität bis ins Detail richtig beschreiben, sondern dass sie zu Entscheidungen führen, die für das Mastermodell akzeptabel sind. (vgl. Schneeweiß, 1992, S. 5). Diese Überprüfung soll hier als eine Anforderung an ein Modell und den Planungsprozess verstanden sein. Sie wird im Weiteren als die „Validierung nach Schneeweiß“ bezeichnet.

Die Darstellung des Planungsprozesses nach Schneeweiß verdeutlicht, dass wiederholt auf Modelle zurückzugreifen ist. Ursachen für die geringe Resonanz der Losüberlappung könnten somit schon auf der Ebene dieser Modelle liegen.

Ein Modell stellt, wie Malinvaud (1980) betont, die formale Repräsentation der Kenntnisse dar, die über ein Phänomen vorliegen. Jedes Modell ist damit abhängig vom Stand der wissenschaftlichen Erkenntnis im Moment seiner Formulierung. Dieser Aspekt ist für die vorliegende Untersuchung relevant, da z.B. das Konzept der PPS-Systeme nach MRP-II Logik letztlich den Stand des Wissens umsetzt, der den Modellen entspricht, die in den siebziger Jahren entwickelt wurden. Hält man an dieser Logik fest, akzeptiert man implizit die damaligen Paradigmen. Wird weiterhin bereits auf der Ebene des Mastermodells eine Formulierung gewählt, die nur die geschlossene Weitergabe abbilden kann, liegt es nahe, im nächsten Schritt des Planungsprozesses jene Formalmodelle zu wählen, die wieder dieses Paradigma teilen.

Mit der Abstraktion und der Relaxation kommt es beim Übergang zwischen den Modellen immer zu notwendigen Vereinfachungen. Auf diese weisen z.B. Bamberg/Coenenberg (2000) hin. Sie verstehen unter einem Modell ein vereinfachendes Abbild realer Tatbestände. Dabei soll man sich auf die Elemente und Relationen konzentrieren, die für die gegebene Problemstellung wesentlich sind (Zweckorientierung) und trotz der Vereinfachung im Modell zumindest eine strukturähnliche Abbildung des Realsystems erreichen. Wie Bretzke (1980, S. 28 ff.) ausführt, setzt allerdings das Problem schon bei der notwendigen „Abbildung“ ein. Man muss z.B. über klar definierte Begriffe, wie den des Loses verfügen, um ein zu modellierendes Phänomen richtig erfassen zu können. Dann aber sind wesentliche Schritte durch die Begriffsfindung bereits implizit geleistet. Nimmt man den Begriff des Loses, ist z.B. zu fragen, ob ein Los immer nur aus Teilen des gleichen Produkts bestehen darf, ob seine Produktion unterbrochen oder seine Teile vorzeitig an die nächste Stufe weitergegeben werden dürfen etc. Die vorherrschenden Begriffe und ihre gängige Interpretation spielen für die Modellentwicklung somit eine sehr große Rolle. Solange schon im Begriff des Loses eine Identität von produzierter und transportierter Menge angelegt ist, muss eine Annahme wie „Lose werden nur geschlossen weitergegeben“ naheliegend erscheinen. Solange sie als Standard akzeptiert

bleibt, werden auch Konzepte und Modelle gewählt, die zu dieser Begrifflichkeit passen. Diese bedingen wiederum betriebliche Strukturen, in denen die Einschränkung der Sichtweise auf die geschlossene Weitergabe nicht weiter auffällt. Während dem bisher angesprochenen Verständnis eines Modells eine gewisse Starrheit zu eigen ist, schlägt Bretzke (1980, S. 245) vor, einen konstruktivistischen Modellbegriff zu verwenden. Ein Modell soll als Ergebnis von Versuchen angesehen werden, womit der modellgestützte Entscheidungsprozess zu einer konstruktiven gedanklichen Leistung wird, die gerade „ohne natürlichen Abschluß“ ist. Jedes Modell hat damit den Charakter einer vorläufigen und noch weiter zu verbessernden Abbildung. Daraus folgt unmittelbar, dass auch häufig bewährte Annahmen regelmäßig zu hinterfragen sind.

Für eine tiefere Betrachtung der Erkenntnislehre soll hier auf Popper (1959) verwiesen werden. Eine umfassende Darstellung weiterer Definitionen und der wissenschaftlichen Diskussion zur Modellierung fasst Zschocke (1995) zusammen.

Halten wir diese Punkte in Form von Thesen fest:

- Die Struktur des Unternehmens hängt neben der Strategie auch von der Wahl des Konzepts ab: structure follows planning concepts.
- Die Diskussion der Losüberlappung ist relevant, da sie erheblich zur Reduktion der Durchlaufzeit beitragen kann und sich ihre Berücksichtigung bis in die betrieblichen Strukturen auswirken wird.
- Solange Annahmen wie „Lose werden nur geschlossen weitergegeben“ als Standard akzeptiert bleiben, werden Konzepte gewählt, die betriebliche Strukturen bedingen, in denen dies nicht als Einschränkung auffällt.
- Der Planungsprozess setzt die Formulierung von Modellen voraus. Diese basieren auf einer bestimmten Sicht der Realität und folgen bestimmten Paradigmen. Sie sind damit abhängig vom Stand der Erkenntnis im Zeitpunkt ihrer Formulierung.
- Modelle erfassen formal die Realität in einer zweckorientierten Abbildung. Sie setzen auf Begriffen, Definitionen und Annahmen auf, wobei ihre Formulierung konstruktivistisch aufzufassen ist.
- Nach Schneeweiß (1992) müssen Modelle im Planungsprozess einer Entscheidungs- und einer ex-post-Validierung genügen.

2.1 Anforderungen an Modelle

Um weitere Anforderungen an die Modelle formulieren zu können, werden diese klassifiziert, indem die Berücksichtigung der Interdependenzen und der Zweck der Modellbildung als Kriterium herangezogen werden.

Das Ausmaß, in dem Interdependenzen berücksichtigt werden, bezeichnet man nach Zäpfel (1982) als den Integrationsgrad des Planungsmodells. Nur im (theoretischen) Extremfall umfasst ein sogenanntes Totalmodell alle relevanten Interdependenzen. Da eine sinnvolle Modellierung aber definitionsgemäß Interdependenzen vernachlässigen muss, können Modelle immer nur Teilaspekte berücksichtigen. Ein sinnvolles Modell ist damit immer ein Partialmodell (vgl. Bretzke, 1980, S. 125 ff.). In diese Richtung argumentieren auch Kistner/Steven (2001, S. 191), die anstatt der herkömmlichen, irreführenden Bezeichnung Totalmodell, den Begriff des monolithischen Modells vorziehen.

Hinsichtlich ihres Zwecks sind Beschreibungs-, Erklärungs- und Entscheidungsmodelle zu unterscheiden. Um letztere geht es in dieser Arbeit. Sie entsprechen den Formalmodellen in Abbildung 2 und dienen unmittelbar zur Ableitung von Entscheidungsempfehlungen. Nach Dinkelbach (1982, S. 30) müssen diese Modelle „die formale Darstellung eines Entscheidungsproblems“ wiedergeben und dabei „wenigstens eine Alternativenmenge und wenigstens eine auf dieser definierte Zielfunktion“ umfassen. Weiterhin sollte ein Entscheidungsmodell den folgenden sechs Anforderungen genügen, die zwar schon vor über dreißig Jahren von Little im „decision calculus“ zusammengefasst wurden, aber nach wie vor aktuelle Kriterien zur Beurteilung von Modellen darstellen (vgl. Little, 1970, S. B-470):

Einfachheit: Das Modell soll möglichst einfach sein, da dies das Verständnis fördert. Nur die wichtigen Phänomene sollen modelliert werden. Verfeinerungen im Detail sind erst dann sinnvoll, wenn der Modellnutzer bereit ist, sie zu erfassen.

Mit diesem ersten Punkt greift Little die schon seit dem 13. Jahrhundert bekannte grundsätzliche Anforderung an einem Modell nach William Of Ockham (1285-1349) auf. Diese als „Ockham’s razor“, als „law of ecomony“ und als „law of parsimony“ bekannte Forderung lautet: „pluralitas non est ponenda sine neccesitate“ oder anders formuliert: „non sunt multiplicanda entia praeter necessitatem“ (vgl. Leff, 1975, S. 35). Sie lauten übersetzt: „Wesenheiten soll man nicht über Gebühr vermehren“ und „Außer dem Notwendigen soll Nichts hinzugefügt werden“. Nach Ockham’s razor ist somit alles aus einem Modell oder aus einer Aussage zu entfernen, was nicht notwendig ist. Eine Spezifizierung ist nur dann vorzunehmen, wenn die zusätzlich abgebildeten Details für die Entscheidungsfindung beziehungsweise für die Erklärung der Zusammenhänge erforderlich sind.

Robustheit: Das Modell soll auch durch unsachgemäße Nutzung nicht zu einem inadäquaten Ergebnis gebracht werden können. Es soll z.B. sicherstellt sein, dass Variable nur Werte innerhalb eines sinnvollen Wertebereiches annehmen können. Zudem sollte das Modell unempfindlich gegenüber kleinen Datenänderungen und Schätzfehlern sein.

Kontrollfähigkeit: Der Zusammenhang zwischen den Variationen der eingegebenen Daten und den daraus resultierenden Veränderungen der vom Modell empfohlenen Entscheidung soll für den Benutzer plausibel sein.

Dem Benutzer sind für bestimmte Datenkonstellationen die Entscheidungen vorzuschlagen, die er erwartet oder zumindest nachvollziehen kann. Reagiert ein Modell nicht nachvollziehbar sensitiv auf Veränderungen der Daten, wird der Entscheidungsträger dem Modell nicht vertrauen. Dies wird er nur dann, wenn er die Konsequenzen, die eine Datenänderung bewirkt, a priori in ihrer Tendenz abschätzen kann.

Anpassungsfähigkeit: Das Modell soll an veränderte Situationen angepasst werden können, indem es in Modulen mit wohldefinierten Schnittstellen formuliert wird.

Vollständigkeit: Das Modell soll die wesentlichen Aspekte vollständig erfassen. Diese Anforderung steht in gewissem Widerspruch zum ersten Kriterium und ist nur im Rückgriff auf die bei Bamberg/Coenenberg (2000) betonte Zweckorientierung der Modelle sinnvoll.

Nur wenn der Zweck des Modells feststeht, kann über den angemessenen Grad an Vollständigkeit entschieden werden. Dieser sollte dann mit einer möglichst einfachen Formulierung erreicht werden. Vollständigkeit ist damit als Restriktion zu verstehen, die aus dem Zweck des Modells resultiert, während Einfachheit eine Zielsetzung darstellt.

Kommunikationsfähigkeit: Der Output des Modells soll derart aufbereitet sein, dass Missverständnisse bei der Interpretation vermieden werden. Außerdem muss die Geschwindigkeit, mit der die Lösungen gefunden werden, so hoch sein, dass die Entscheidung noch relevant ist, wenn die Lösung vorliegt. Es muss genügend Zeit bleiben, um unterschiedliche Konstellationen prüfen zu können.

Auch unter Berücksichtigung des „decicion calculus“ wird die Konstruktion eines Entscheidungsmodells nicht zu einem eindeutigen Vorgang. Immerhin liefert er aber Kriterien, mit denen Entscheidungsmodelle und die darauf aufsetzenden Verfahren beurteilt werden können.

2.2 Elementare Problemvereinfachungsstrategien

Die drei Konzepte der Komplexitätsreduktion: Aufteilen, sukzessives Lösen und Einsatz von Heuristiken, werden bei der Formulierung von Modellen durch elementare Vereinfachungsstrategien ergänzt. Zu diesen zählen:

- Abstraktion, mit den Unterformen Idealisierung und Generalisierung,
- Aggregation und Disaggregation, sowie
- Dekomposition, die verbunden mit Hierarchisierung und Koordination erfolgt.

Abstraktion bezeichnet das bewusste Vernachlässigen bestimmter Details eines vorliegenden Problems. Abstrahiert wird von den Zusammenhängen und Eigenschaften, deren Fallenlassen eine wesentliche Vereinfachung erlaubt. Zugleich ist aber die Struktur des Problems im Modell zu erhalten, damit die ermittelten Lösungsvorschläge für das Realproblem zulässig sein können.

Ein weit verbreitetes Beispiel ist der Rückgriff auf deterministische Daten. In vielen Fällen kann ein Zusammenhang nur dann sinnvoll modelliert und einer Lösung zugeführt werden, wenn von der Unsicherheit abgesehen wird. Betrifft dies z.B. die Schwankungen in den Ausführungszeiten auf den Maschinen, resultieren – bei entsprechend großen Produktionsmengen – keine negativen Konsequenzen, da sich die Abweichungen ausgleichen. Ein weiteres Beispiel ist die Entwicklung statischer statt dynamischer Modelle. Diese abstrahieren von wesentlichen zeitlichen Interdependenzen. Als drittes Beispiel werden bestimmte Veränderungen, die einen stetigen Verlauf nehmen, wie z.B. das Fortschreiten in der Produktion oder die Abnutzung einer Maschine, in ein Raster gezwängt; sie werden diskretisiert. Das damit verbundene Abstrahieren von Details der betrieblichen Abläufe durch das Festlegen eines Zeitrasters ist eine wichtige, oft aber nicht mehr wahrgenommene Problemvereinfachung (vgl. Ovacik/Uzsoy, 1997, S. 34 ff.).

Betrachtet man zur Verdeutlichung des letzten Punkts die Produktionsmengenmodelle, sind hinsichtlich der Modellierung der Perioden diejenigen mit sogenannten „large bucket“, den „Makroperioden“, von denen mit „small bucket“, den „Mikroperioden“, zu unterschieden (vgl. Eppen/Martin 1987, Domschke et al. 1997, S. 150). In letzteren gilt implizit die „all-or-nothing“ Bedingung: Entweder wird während der gesamten Periode genau ein Produkt produziert oder die Anlage wird umgerüstet oder die Maschine steht leer. Hingegen können im large bucket-Fall mehrere Produkte in einer Periode gefertigt werden. Von einer expliziten Modellierung der Reihenfolgeplanung wird dann aber abstrahiert (vgl. Drexl/Kimms 1997, S. 225). Wie Schneeweiß (1999, S. 224) betont, kommt es insbesondere durch das Erfassen von auflagefixen Kosten durch Binärvariable und die Verwendung von Mikroperioden dazu, dass auf der realen

Ebene eine Diskretisierung der Produktion erzwungen wird, d.h. nur zu bestimmten Zeiten, zu Beginn der Perioden, können Lose aufgelegt werden, die dann oft auch noch eine bestimmte Größe (oder deren Vielfaches) haben müssen. Beispielsweise muss im Modell von Haehling von Lanzenauer (1970) die Periodenlänge auf das kleinste gemeinsame Vielfache der Ausführungszeiten der Stufen gesetzt werden, damit das Modell numerisch beherrschbar bleibt. Damit können die Losgrößen auf den Stufen und die weitergegebenen Mengen aber nur noch als das Vielfache dieser Größe auftreten. Die zunächst unproblematisch erscheinende Einschränkung hat allerdings weitreichende Folgen. Zu überlegen ist insbesondere, ob nicht die oft unterstellte Annahme einer geschlossenen Fertigung, d.h. die Weitergabe von ganzen Losen zwischen den Stufen, gerade durch das Diskretisieren naheliegend wird. Aus der Akzeptanz einer sinnvollen Annahme: Die Fertigung schreitet in diskreten Schritten fort, drängt sich eine andere – nicht unbedingt sinnvolle – Annahme auf: Auch die Weitergabe erfolgt diskret, nämlich in ganzen Losen und nicht in Teilen der Lose.

In einer ähnlichen Richtung argumentiert Jahnke (1998). Wird zur Berücksichtigung der Fertigungskapazität der Zeit- und nicht nur der reine Mengenaspekt abgebildet, hat zwangsläufig auch die Lösung des Problems eine Zeitstruktur. Häufig kann aber weder für die Wahl der Periodenlängen noch für die des Planungshorizonts auf „natürliche Werte“ zurückgegriffen werden, die unmittelbar aus dem Produktionsprozess oder der Nachfragestruktur abzuleiten sind. Jahnke/Biskup (1999) charakterisieren ihre Festlegung als „kritisch“, da sie für die Güte des gesamten weiteren Planungsverfahrens eine sehr hohe Bedeutung haben (vgl. Jahnke, 1995; Jahnke/Biskup, 1999, S. 57ff.).

Die Abstraktion mündet dann in eine *Idealisierung*, wenn Eigenschaften außer Acht gelassen werden, damit eine idealisierte Situation, z.B. die vollständige Information, der vollkommene Markt oder der homo oeconomicus modelliert werden kann. Ihr wesentliches Kennzeichen ist, dass Situationen unterstellt werden, die so in der Realität nicht zu beobachten sind und die nur in der Modellwelt, d.h. im Mastermodell vorliegen. Beispielsweise existieren eine Reihe von Modellen, in denen angenommen wird, dass Produzieren und Transportieren keine Zeit benötigt, beziehungsweise dass die entsprechenden Kapazitäten unbeschränkt verfügbar sind.

Eine weitere Unterform der Abstraktion ist die *Generalisierung*. Dabei wird unterstellt, dass ein tatsächlich beobachtbarer Zusammenhang beziehungsweise ein Zustand generelle Gültigkeit hat:

- Es werden risiko-averse Entscheidungsträger modelliert, alle Aufträge und Maschinen stehen im Zeitpunkt Null zur Verfügung, die Nachfrage der Kunden tritt jeweils gebündelt am Periodenanfang auf, etc.

- Für ein Intervall wird ein proportionaler Zusammenhang angenommen, obwohl er nur für bestimmte kleine Ausschnitte beobachtbar ist. Bei Erhöhung der Intensität steigt der Energieverbrauch einer Maschine z.B. zunächst proportional, dann aber überproportional an. Modelliert wird oft nur der proportionale Zusammenhang.
- Ein konstanter Prozentsatz der in einer Periode unbefriedigten Nachfrage – in der Regel wird sogar 100% postuliert – führt zu einer Bestellung für die folgende Periode (back-order) oder entfällt ganz (lost-sales).

Als zweite elementare Vereinfachungsstrategie kann die *Aggregation* angesehen werden. Indem einander ähnliche Daten und Entscheidungsvariablen sinnvoll unter einem gemeinsamen Oberbegriff zusammengefasst sind, werden entscheidungsirrelevante Details fallengelassen (vgl. Hax/Meal, 1975; Switalski, 1988; Leisten, 1996). Ähnliche Produkte werden zu Familien aggregiert, was die Datenbeschaffung und die Modellierung vereinfacht. Die zu einem Los zusammengefassten Teile bilden z.B. ein derartiges Aggregat. Eine Zielsetzung der Aggregation ist dabei die Entlastung der Entscheidungsträger von überflüssigen Details. Aggregierte Größen sind zudem verlässlicher zu prognostizieren als detaillierte Größen. Weiterhin dient die Betrachtung weniger Aggregate der Verständnisförderung (vgl. Steven, 1994). Wie stark aber beim Aggregieren abstrahiert wurde, zeigt sich, wenn die *Disaggregation* durchlaufen wird. Sie hat die auf Basis der aggregierten Daten ermittelten Ergebnisse auf die bei der Aggregation beteiligten Einzeldaten aufzuschlüsseln (vgl. Zäpfel, 1996; Leisten 1996). Neben der sachlichen, z.B. am Rüstaufwand orientierten Zusammenfassung einzelner Produkte zu einer Erzeugnisgruppe kann auch eine zeitliche Aggregation und Disaggregation notwendig werden, indem z.B. Produktionsmengen pro Monat geplant und anschließend auf einzelne Tage oder Schichten disaggregiert werden. Eine an der Zeit vorgenommene Aggregation legt aber eine diskrete Zeitführung nahe und liefert damit wieder einen Anstoß, Produktionsmengenmodelle zu nutzen.

Als dritte elementare Vereinfachungsstrategie bezeichnet die *Dekomposition* die Zerlegung eines Problems in einfacher handhabbare Teilprobleme. Auch hier kommt es zu einer Abstraktion. Bei der Aufteilung des Problems wird versucht, möglichst wenige Interdependenzen zu verletzen. Dabei wird die Dekomposition sehr häufig im Sinne einer zeitlichen Zerlegung des Problems mit anschließender sukzessiver Lösung eingesetzt: Das Problem wird in Partialmodelle mit unterschiedlicher zeitlicher Reichweite und Genauigkeit des Zeitrasters zerlegt. Es können sowohl sachlich horizontale als auch zeitliche vertikale Abhängigkeiten fallen gelassen werden. Die Art der Abstimmung zwischen den Partialmodellen wird dann wesentlich durch die Reihenfolge bestimmt, in der die Partialmodelle gelöst werden, wenn die Lösungen des einen Partialmodells den Input des nächsten bilden. Für die Festlegung dieser Reihenfolge spielt die

Hierachisierung eine wichtige Rolle (zur Hierarchisierung vgl. Stadtler, 1988; Kistner/Switalski, 1989; Steven, 1994; Zäpfel, 1996; Schneeweiß, 1999b). Im Idealfall führt sie dazu, dass die im Unternehmen gegebene organisatorische Aufteilung in Entscheidungsebenen erhalten bleibt. Dabei sollten die einzelnen Teilprobleme so modelliert werden, dass bevorzugt die Zusammenhänge betrachtet werden, für die die Daten bereitgestellt werden können und die Verantwortlichkeit an die jeweilige Ebene delegierbar ist. Die Struktur des Planungsprozesses kann damit der organisatorischen Struktur des Unternehmens angepasst werden und umgekehrt, womit wieder die „structure follows planning concept"-These angesprochen ist. Bei der Abstimmung der dekomponierten Probleme spielt die Art der *Koordination* der Partialmodelle eine wesentliche Rolle. In Anlehnung an Kistner (1992) und Steven (1994) ist die Koordination nach dem Kriterium der Richtung, in der die Kopplung erfolgt, als einseitig, wechselseitig und einseitig mit Rückkopplung möglich. Weiterhin unterscheidet man nach dem Rhythmus, in dem sie erfolgt, in die periodische oder die fallweise Abstimmung. Wird eine periodische Abstimmung gewählt, bedingt die Form der Koordination eine diskrete Zeitführung und legt damit den Einsatz eines Produktionsmengenmodells nahe.

Im Folgenden rückt die Zeitführung im Modell und die Form der Ausgestaltung der Planung als Reaktion auf die betriebliche Unsicherheit in den Vordergrund.

2.3 Unsicherheit und Ausgestaltung der Planung

Um den realen Gegebenheiten möglichst nahe zu kommen, müsste neben einem nichtbeschränkten Planungshorizont auch die Modellierung der Vorgänge in kontinuierlicher Zeit vorgenommen werden. Beide Ausprägungen finden allerdings nur selten in den betriebswirtschaftlichen Modellen Verwendung. Wird ein unendlicher Planungshorizont unterstellt, geschieht das in der Regel nicht, um zeitliche Interdependenzen zu erfassen, sondern weil bei unbeschränktem Horizont Lösungen in geschlossener Form vorliegen. Der Rückgriff auf einen unendlichen Planungshorizont hat damit den Charakter einer vereinfachenden Annahme. Folgt man Jahnke/Biskup (1999, S. 53), führt die Modellierung von kontinuierlich fortschreitenden Prozessen oft dazu, dass Standardmethoden des Operations Research nicht zum Einsatz kommen, da diese eine kontinuierliche Zeitführung nur schwer verarbeiten können. Auch aus diesem Grund dominieren Ansätze mit diskontinuierlicher Zeitführung und beschränktem Planungshorizont. Außerdem ist, wie Jahnke ausführt, das Diskretisieren der Zeit schon in der traditionellen Kostenrechnung üblich und notwendig, da die Zusammenhänge anders als durch ein regelmäßiges, vergangenheitsbezogenes Schätzen nicht fassbar sind (vgl. dazu Jahnke, 1995, S. 135). Hinzu treten die im externen

Rechnungswesen üblichen und ebenfalls periodenweise durchgeführten Rechnungen sowie die in der Regel auf diskrete Zeitpunkte bezogenen Kalküle aus der Investitions- und Finanzierungsrechnung. Die Systeme, die die Produktionsplanung umgeben, nutzen damit eine diskrete Zeitführung. Aus dem Umfeld der Produktionsplanung ergeben sich dadurch Vorgaben für den Planungshorizont und die Planungsperioden, die bezogen auf die Produktionsplanung willkürlich sind. Diese Vorgaben eines zeitlichen Rasters schränken die Alternativenmenge der Produktionsplanung ein. Lose können nur zu bestimmten Zeitpunkten eingelastet werden und auch die Abbildung einer überlappenden Fertigung wird erschwert. Eine vorzeitige Weitergabe von Losen wäre nur innerhalb des (willkürlichen) Zeitrasters erfassbar und damit a priori beschränkt. Prinzipiell kann zwar das zeitliche Raster so eng wie benötigt gewählt werden, doch führt eine detailliertere zeitliche Erfassung zu einer Vervielfachung der Entscheidungsvariablen, zu erhöhtem Datenbeschaffungs- und Pflegeaufwand und zu numerisch aufwändiger zu lösenden Modellen.

Der Verwendung vieler kleiner Perioden steht zudem die betriebliche Unsicherheit entgegen: Die Prognosequalität geht bei zunehmendem zeitlichen Abstand zurück während die zeitlich stärker aggregierten Größen verlässlicher prognostiziert werden können. Als Komponenten der betrieblichen Unsicherheit unterscheidet Zäpfel (1996, S. 51f.) die Beschaffungs-, die Nachfrage- sowie die Prozessunsicherheit und erfasst damit einen wesentlichen Aspekt nur teilweise. Dieser wird bei Morey (1985) hervorgehoben, der ergänzend die Informationsunsicherheit berücksichtigt. Sie betrifft die Unsicherheit der Informationen über den Systemzustand in zwei Dimensionen: die Unsicherheit hinsichtlich der Menge (z.B. den Produktionsfortschritt bis zu einem beliebigen Zeitpunkt) und die Unsicherheit bezüglich des Zeitpunkts (z.B. Fertigstellung eines Loses in einem vom Erfassungszeitraster abweichenden Zeitpunkt). Eine Ursache für fehlerhafte Informationen über den Systemzustand sieht Tempelmeier (1999, S. 360 ff.) beispielsweise darin begründet, dass Lagerbestände nur periodisch fortgeschrieben werden und damit innerhalb der Überwachungsintervalle die Veränderungen nicht erfasst sind. In der Realität geschieht etwas, über das keine Daten beschafft werden, da die Diskretisierung, die zur Vereinfachung bei der Modellbildung genutzt wird, eine zwischenzeitliche Erfassung nicht nahe legt. Werden die Informationen aber nur periodisch fortgeschrieben, sinkt die Wahrscheinlichkeit, dass das überlappende Produzieren als Planungsalternative erkannt wird. Soll dennoch aus einem Los vorzeitig eine Menge weitergegeben werden, liegen keine verlässlichen Daten vor und die Entscheidung wird erschwert.

Bei Tempelmeier umfasst die Informationsunsicherheit auch eventuell auftretende Planungs- und Kommunikationsfehler, die durch unsachgemäße Verfahren oder durch mangelhafte Abstimmung auftreten (vgl. Tempelmeier, 1999, S. 361). Derartige Mängel können z.B. mit dem Einsatz eines PPS-Systems ein-

hergehen. Tempelmeier betont somit einen Aspekt, der bei Zäpfel (1996) mit der Prozessunsicherheit zwar angelegt, nicht aber in dieser Deutlichkeit ausgeführt wird. Aus der Wahl des Planungskonzepts, das zur Steuerung und Kontrolle der Vorgänge eingesetzt wird, resultieren Probleme, die letztlich nicht problemimmanent, sondern verfahrensimmanent sind. Das bedeutet, dass durch einen Wechsel des Planungsverfahrens Teile der Problemstellung entfallen könnten oder sich zumindest anders darstellen würden.

Für eine angemessene Reaktion auf diese vielfältigen Formen der Unsicherheit kommt die rollierende Planung (Kistner/Steven, 2001, S. 214, Tempelmeier, 1999) in Frage. Sie ist als eine wichtige Form der Reaktion auf sich ändernde Daten und als eine weit verbreitete Form des Umgangs mit den verschiedenen Formen der Unsicherheit anzusehen. Zwar werden bis zum Planungshorizont Entscheidungen festgelegt, tatsächlich realisiert werden allerdings nur die für die ersten Periode(n). Die Entscheidungen für die restlichen Perioden haben den Charakter eines vorläufigen Plans, so dass jede Periode mehrfach vorläufig und nur einmal verbindlich geplant wird.

Wird das Konzept der rollierenden Planung – wie in weiten Teilen der Literatur angenommen – mit einer Periodisierung und einem endlichen Planungshorizont verbunden eingesetzt, zeigt sich, dass auch aus dem Umgang mit der Unsicherheit eine Tendenz zur Periodisierung erwächst. Kombiniert man rollierende Planung mit einer Zunahme des Aggregationsgrads in Richtung des Planungshorizonts, kann der Aufwand erheblich verringert werden. Beispielsweise kann die Planung für den ersten Monat auf den Tag genau erfolgen, für den zweiten Monat könnte die Periodenlänge auf zwei Tage erhöht und im dritten Monat wochenweise geplant werden. Führt man außerdem eine Dekomposition des Planungsproblems durch, nennt Zäpfel (1996, S. 52) als Vorteil, dass sich mit der Aufteilung der Planungsaufgabe in mehrere Teilplanungen mit abnehmendem Horizont und zunehmendem Detailliertheitsgrad der Daten in Verbindung mit der rollierenden Planung eine stufenweise Unsicherheitsbeseitigung realisieren lässt: Die übergeordneten Ebenen planen mit längerem (zumindest gleich langem) Horizont als die untergeordneten Ebenen; dafür ist aber der Detaillierungsgrad der Daten geringer. Die untergeordneten Ebenen erhalten neben den bindenden Vorgaben der höheren Ebenen für den gegenwärtig fest einzuplanenden Zeitraum vorläufige Entscheidungen, so dass frühzeitig auf den nachgelagerten Ebenen die demnächst zu erwartenden Vorgaben bekannt sind. Die Koppelung in bestimmte Zeitpunkten macht es aber schwer, Modelle mit diskreter und Modelle mit kontinuierlicher Zeitführung vermischt einzusetzen. Ein sinnvoller Ausweg besteht darin, auf den stärker aggregierten Stufen mit Produktionsmengenmodellen zu arbeiten und auf den weniger aggregierten Stufen Modelle mit kontinuierlicher Zeitführung einzusetzen. Da in beiden Modellwelten die gängigen Standardannahmen die überlappende Weitergabe ausschließen, lassen sich aus der Bewältigung der Unsicherheit mittels rollierender

Planung und dem Einsatz hierarchisch dekomponierter Modelle Gründe für die eher geringe Resonanz der Losüberlappung ableiten.

Als Zusammenfassung der Ergebnisse kann festgehalten werden:

- Little's „decision calculus" (1970) folgend, ist bei der Entwicklung von Entscheidungsmodellen insbesondere auf eine angemessene Balance zwischen Vollständigkeit und Einfachheit zu achten.
- Als elementare Problemvereinfachungsstrategien werden die Abstraktion, mit ihren Unterformen Idealisierung und Generalisierung, sowie die Aggregation und die Dekomposition eingesetzt.
- Diskrete Zeitführung ist ebenso notwendig wie willkürlich. Sie verdeckt den kontinuierlichen Verlauf und verleitet dazu, nur noch im diskreten Raster zu denken.
- Informationsunsicherheit als verfahrensimmanentes Problem kann die Wahl einer überlappenden Fertigung erschweren.
- Dekomposition und Hierarchisierung stellen insbesondere in der Verbindung mit einer rollierend ausgeführten Planung eine wichtige Form des Umgangs mit Unsicherheit dar.

Die Hierarchisierung und zeitliche Dekomposition legen den Einsatz von Modellen nahe, deren Standardannahmen die überlappende Weitergabe ausschließen und ihre geringe Berücksichtigung im Schrifttum erklären können.

3. Produktionsplanung und Problemvereinfachung

Ziel dieses Kapitels ist die Darstellung von zwei Konzepten, die der Problemvereinfachung und der Modellbildung in der Produktionsplanung häufig zu Grunde liegen. Zunächst wird mit der Produkt-Prozess-Matrix nach Hayes/Wheelwright (1979) und ihrer Weiterentwicklung durch Spencer/Cox (1995) eine gängige Vereinfachung hinsichtlich der Wahl der geeigneten Produkt-Prozess-Kombination vorgestellt. Diese Wahl betrifft die Aufbauorganisation und legt die Strukturen fest, in denen ein Produktionsplaner typischer Weise denkt und für die er Entscheidungsmodelle entwickelt. Danach wird auf Anthony (1965) zurückgegriffen, der einen in der Betriebswirtschaftslehre weitgehend akzeptierten Ausgangspunkt zur Komplexitätsreduktion durch Dekomposition vorgeschlagen hat. Sein Konzept sieht eine bestimmte Anordnung von Planungsebenen sowie eine Verteilung und Zuordnung der Aufgaben vor und umfasst damit wesentliche Aspekte der Ablauforganisation.

3.1 Produktionsstrukturen

Um Zusammenhänge zwischen den strukturellen Eigenschaften der Produkte und den zur Produktion genutzten Prozessen herauszuarbeiten, stellen Hayes/Wheelwright bereits 1979 die sogenannte Produkt-Prozess-Matrix vor. Sowohl für die Produkte als auch für die Prozesse bilden sie je vier Klassen und identifizieren innerhalb der so aufgespannten Matrix die Felder, die sie als typische Kombinationen mit einigen Beispielen belegen. Die Prozesse charakterisieren sie hinsichtlich des Materialflusses und identifizieren die zugehörigen Organisationsformen der Produktion. Die Produkte unterscheiden Hayes/Wheelwright (1979) hinsichtlich der Stückzahl und ihrem Heterogenitätsgrad. Es ergeben sich vier als typisch herausgearbeitete Kombinationen:

- Vermischte, sich überkreuzende Materialflüsse, die bei Werkstattfertigung (*job shop*) zur Produktion kleiner, nicht standarisierter Volumen dienen.
- Nichtverbundene, serielle und separierbare Materialflüsse, die bei Fließfertigung (*flow shop*) zur Produktion geringer Stückzahlen in zahlreichen Varianten, den Serien, eingesetzt werden.
- Verbundene, serielle und synchronisierte Flüsse, die für hohe Stückzahlen in wenigen Ausprägungen (den Sorten) typisch sind.

- Kontinuierliche, serielle Flüsse bei stetiger Weitergabe, die bei der Massenfertigung auf Fließbändern angetroffen werden.

Für die seriellen und separierbaren Flüssen stellen Hayes/Wheelwright (1979) die Losbildung als typisches Merkmal heraus. Auf der Produkt-Prozess-Matrix von Hayes/Wheelwright setzen Spencer/Cox (1995) auf und prüfen sie an Hand einer empirischen Studie. Sie unterscheiden die Einzelfertigung (vermischte Materialflüsse), die losweise Fertigung (nicht verbundene Materialflüsse), die Fertigung auf einem Fließband (verbundener, synchronisierter Materialfluss) sowie die natürliche Fließfertigung, die z.B. bei der Herstellung von Chemikalien auftritt. Weiter ist hervorzuheben, dass Spencer/Cox einen kontinuierlichen Übergang auf den Achsen annehmen und damit eine größere Vielfalt betrieblicher Strukturen zulassen, als Hayes/Wheelwright (1979). Hinsichtlich einer empirischen Überprüfung der erweiterten Matrix kommen Spencer/Cox (1995, S. 1289 f.) unter anderem zu folgenden Ergebnissen:

- Die Prozess-Struktur ist oft, aber nicht immer, durch die Art des Produkts gegeben.
- Innerhalb einer Fertigung können alle vier Prozess-Typen gemeinsam auftreten. Oftmals finden sich in einem Endprodukt mehrere, den unterschiedlichen Produkt-Prozess-Kombinationen zuzuordnende Komponenten wieder.
- Produkte beziehungsweise Komponenten können im Laufe der Zeit, z.B. dem Produktlebenszyklus folgend, in so stark veränderten Quantitäten benötigt werden, dass sie zwischen verschiedenen Produkt-Prozess-Kombinationen wechseln.
- Aus einer vorgegebenen Produkt-Prozess-Kombination folgt nicht eindeutig, welches Konzept zur Produktionsplanung und -steuerung das geeignetste ist.

Zum ersten Punkt haben bereits Finch/Cox (1988, S. 124) festgestellt, dass in der industriellen Fertigung für ein und dieselbe Produktart sehr unterschiedliche Produkt-Prozess-Kombinationen gewählt wurden. Ein Grund kann darin gesehen werden, dass die Nutzungsdauer einer Maschine i.d.R. die Dauer des Produktlebenszyklus um ein vielfaches übertrifft, z.B. erreichen in der Metallverarbeitung Stanzmaschinen eine Nutzungsdauer von mehreren Jahrzehnten. Sie stehen somit noch zur Verfügung, wenn die Produkte, für die sie ursprünglich gedacht waren, schon lange nicht mehr produziert werden. Neben dem Kriterium, eine der Produktart möglichst angemessene Prozess-Struktur zu wählen, ist auch der wirtschaftliche Einsatz der bereits verfügbaren Kapazität zu berücksichtigen. Eine eindeutige Zuordnung zwischen Produkten mit speziellen Charakteristika und bestimmten Prozessen stellen Finch/Cox (1988) in der Praxis als nicht gegeben fest.

Ergänzend zu den Ergebnissen von Spencer/Cox (1995) lassen die Beobachtungen von Finch/Cox (1988) eine weitere mögliche Antwort auf die Frage zu, warum die wissenschaftliche Diskussion der Losüberlappung in erheblicher Diskrepanz zum praktischen Einsatz dieser Techniken steht. Zieht man ein beliebiges Lehrbuch zur Produktionsplanung heran, kann man fast sicher sein, eine Kategorisierung der typischen Produktionsstrukturen ähnlich der Matrix von Hayes/Wheelwright (1979) oder Spencer/Cox (1995) zu finden. Diese sinnvolle Beschränkung auf typische Kombinationen verleitet aber auch dazu, vornehmlich Modelle und Verfahren zu entwickeln, die diese idealtypischen Strukturen voraussetzen. In der dabei unterstellten Modellwelt werden nachträgliche Anpassungsmaßnahmen leicht als Ausdruck dafür gesehen, dass es nicht gelungen ist, die mit den Modellen ermittelten Vorgaben richtig umzusetzen. Es liegt dann nicht an dem Modell, wenn sich herausstellt, dass der Plan und die Realität voneinander abweichen. Folglich sind auch die Techniken, die eine Anpassung der Ergebnisse außerhalb des Modells leisten, nicht vorrangig zu diskutieren. Wenn aber, wie z.B. die Untersuchungen von Finch/Cox (1988) und Spencer/Cox (1995) zeigen, in der Praxis eher die nicht dem Ideal entsprechenden Situationen typisch sind, dann stimmen die Modelle im Allgemeinen nicht mit den betrieblichen Gegebenheiten überein. In der Konsequenz stellt sich für die betriebliche Praxis die Anpassung von ermittelten Ergebnissen an reale Situationen als eine dauernd zu bewerkstelligende Aufgabe dar. Diese Überlegung deckt sich mit der Empfehlung von Lee et al. (1997), die explizit fordern, klassische Algorithmen und Modelle dahingehend zu erweitern, dass sie sich stärker den tatsächlichen Gegebenheiten annähern.

Ausgehend von den oben genannten Charakteristika kann das Untersuchungsobjekt weiter festgelegt werden. Die vorliegende Arbeit betrachtet in erster Linie die Serien- und Sorten-Fertigung. Beide geben das Umfeld einer mehrstufigen, oft seriellen Produktion vor, die einerseits in der industriellen Fertigung häufig anzutreffen ist und für die andererseits Losgrößenprobleme typisch sind (Domschke et al., 1997).

3.2 Planungsebenen nach Anthony

Während die von Hayes/Wheelwright (1979) vorgestellte Matrix letztlich eine Vereinfachung durch Klassifikation vornimmt, setzt ein zweites wesentliches Konzept zur Problemvereinfachung bei der Verteilung der Inhalte an: Es ist die Einteilung in die Planungsebenen nach Anthony (1965). Er unterscheidet drei hierarchisch angeordnete und vom Zeit- und Entscheidungsumfang gestaffelte Ebenen der Planung: die strategische, die taktische und die operative Planung.

Die breite Akzeptanz, die diese Aufteilung gefunden hat, kommt unter anderem in der klassischen Trilogie Zäpfels zum Produktions-Management (1982:

operatives, 1989 und 2000: taktisches, sowie 1989b und 2000b: strategisches Produktions-Management) zum Ausdruck und findet sich in vielen Lehrbüchern z.B. bei Tempelmeier/Günther (2003, S. 25), Kistner/Steven (2001, S. 12) und Jahnke/Biskup (1999, S.13) wieder. Sie bedingt, dass auch die Entwicklung von PPS-Systemen in diesem gedanklichen Rahmen erfolgt(e). Er beeinflusst sowohl die Forschung als auch die Umsetzung der Ergebnisse in der Praxis. Eine Diskussion der Konsequenzen, die sich aus dem Konzept von Anthony ergeben, hat somit aus heutiger Sicht eine hohe Relevanz.

3.2.1 Strategische Produktionsplanung

Unter dem Begriff der strategischen Planung wird die langfristig ausgerichtete, Potential aufbauende und Rahmenbedingungen setzende beziehungsweise erhaltende Planung verstanden. Nach Anthony (1965, S. 15 ff.) wird dabei über die Ziele des Unternehmens entschieden und festgelegt, welche Ressourcen wie zu beschaffen und zu verwenden sind, damit diese Ziele erreichbar werden. Auf der strategischen Ebene stehen somit qualitative, aufbauorganisatorische Aspekte im Vordergrund. Als Ergebnis der strategischen Planung erhält man sogenannte Strategien, die den Rahmen für das weitere Vorgehen bei der Beschaffung und bei der Verwendung der Ressourcen festlegen.

Zwar entsprechen sich langfristige und strategische Entscheidungen oftmals, doch folgt, wie Anthony betont, allein aus der Langfristigkeit einer Entscheidung nicht zwangsläufig, dass sie auch strategischen Charakter hat. Neben der zeitlichen Reichweite der Konsequenzen dient als Kriterium deren Qualität, d.h. ihre Bedeutung für die Zukunft des Unternehmens. Um eine Entscheidung als strategische Entscheidung zu klassifizieren, muss ihre Tragweite a priori richtig eingeschätzt werden. Zur Identifikation der strategischen Entscheidungen bietet es sich an, ihre typischen Entscheidungsfelder zu benennen. Nach Domschke et al. (1997, S. 8) umfasst die strategische Produktionsplanung die langfristigen, qualitativen Strukturentscheidungen hinsichtlich der vier Kriterien: *Sortiment*, *Verfahren*, *Standort* und *Produktionsfaktoren*.

Die Wahl des Konzepts, nach dem auf den nachgelagerten Ebenen die Produktion geplant und gesteuert werden soll, fassen Domschke et al. (1997, S. 8 f.) nicht explizit als strategische Entscheidung auf. Berücksichtigt man aber die ursprüngliche Definition von Anthony (1965), beinhaltet die strategische Planung auch die Festlegung von Grundsätzen zur Nutzung der Ressourcen und darunter fällt auch die Auswahl des Konzepts zur Planung und Steuerung der Produktion. Da sich die im Unternehmen vorliegenden Strukturen zum Teil auch aus der Wahl des Planungskonzepts ergeben, muss dessen Festlegung als strategische Entscheidung verstanden werden.

Andererseits sind die grundsätzlichen Entscheidungen hinsichtlich der zu fertigenden Produktart, der Auswahl der Standorte und der Technologie eher auf der Ebene der unternehmensweiten Strategiefestlegung zu behandeln, d.h. sie sollten nicht exklusiv der strategischen Produktionsplanung zugerechnet werden. Die strategische Produktionsplanung behandelt daher qualitative Entscheidungen hinsichtlich der folgenden fünf Kriterien:

- Sortiment: Was soll produziert werden?
- Standort: Wo soll produziert werden?
- Verfahren: Welche Technologien sollen zur Produktion genutzt werden?
- Produktionsfaktoren: Womit soll produziert werden?
- Konzept: Wie soll die Produktion geplant und gesteuert werden?

Für den weiteren Verlauf der Untersuchung wird angenommen, dass die Grundsatzentscheidungen hinsichtlich der Wahl der Produktart und des Sortiments, der Auswahl der Standorte sowie der eingesetzten Technologie bereits vorliegen, so dass die übrigen Entscheidungen in den Vordergrund treten. Sie sind näher zu betrachten, da sie die Struktur des zu lösenden Problems definieren oder Maßnahmen auslösen, die Veränderungen der Problemstruktur nach sich ziehen. Mit dem Setzen dieser Rahmenbedingungen werden aber zugleich Möglichkeiten für den Einsatz der Losüberlappung geschaffen oder aufgegeben. Von besonderer Relevanz sind dabei:

- Bereitstellung und Ausgestaltung des maschinellen Potentials
  - Zahl und Art der Maschinen
  - Segmentierung (wird mitunter als Modularisierung bezeichnet).
- Gestaltung der Produkte
  - Standardisierung und Modularisierung
  - Verlagerung des Entkopplungspunkts.
- Wahl des Konzepts zur Planung und Steuerung der Produktion
  - Prognosegetrieben: Push-Konzept
  - Kundenauftragsgetrieben: Pull-Konzept
  - Mischformen.

Unter den genannten Möglichkeiten hat die Wahl des Konzepts besonders weit reichende Konsequenzen, doch setzt sie ergänzende Maßnahmen bei der maschinellen Ausstattung und der Gestaltung der Produkte voraus. Aus diesem Grund ist zunächst auf diese beiden Formen der Problemgestaltung einzugehen.

3.2.1.1 Bereitstellung des maschinellen Potentials

Die Entscheidung hinsichtlich der maschinellen Ausstattung ist nicht nur einmal, sondern regelmäßig zu betrachten. Das vorhandene und alternde Potential muss kontinuierlich überwacht und gegebenenfalls ergänzt werden. Dabei können drei wesentliche Ansatzpunkte unterschieden werden: die Zahl der Maschinen, die Art der Maschinen und die Gruppierung der Maschinen.

a) Die Zahl der verfügbaren Maschinen kann durch Installation weiterer identischer Maschinen erhöht oder durch Deinstallation verringert werden. Mit Hilfe dieser Maßnahmen wird versucht, eine Ausstattung vorzuhalten, die der zu erwartenden Belastung angemessen ist. Die Belastung – respektive die Verteilung der Belastung über die Zeit – hängt aber auch davon ab, wie Lose dimensioniert und weitergegeben werden. Eine geschlossene Weitergabe führt in der Tendenz zu höheren Belastungsspitzen, die durch zwischenzeitliche Leerstände unterbrochen werden. Eine überlappende Weitergabe hingegen harmonisiert den Material- und Güterfluss, er wird verstetigt. Dies beeinflusst indirekt wieder die Entscheidung über die bereitzuhaltende Kapazität.

b) Die Einsatzmöglichkeiten der Maschinen können verändert werden. Neben der Installation von Maschinen mit verändertem Nutzungsspektrum kann die Flexibilität der bereits bestehenden Ausstattung erhöht werden. Rüstvorgänge können automatisiert, die Spektren für die intensitätsmäßige Anpassung erweitert oder Mehrzweck-Maschinen installiert werden. Wird der Aufwand für das Rüsten z.B. durch einen automatischen Werkzeugwechsel verringert, erhöht sich die zur Produktion tatsächlich nutzbare Zeitspanne. Das Rüsten kann – geeignete Produktionsprozesse vorausgesetzt – an Bedeutung für die Festlegung der Losgrößen verlieren (vgl. Wiendahl, 1995, S. 24). Damit werden kleinere Lose wirtschaftlich und das Gewicht zwischen der Losgrößen- und der Reihenfolgeplanung verlagert sich. Kleinere Lose bewirken eine höhere Flexibilität, reduzieren die Zwischenlagerbestände sowie die Durchlaufzeiten und stärken damit die operative wie auch die strategische Positionierung des Unternehmens (vgl. Knolmayer, 1987, S. 56). Außerdem führt eine Reduktion der Rüstzeiten und Rüstkosten dazu, dass auch eine nachträgliche Veränderung der Losgröße – beispielsweise in Form der weiteren Aufteilung oder der Überlappung – ökonomisch sinnvoll wird. Die Auswahl der Maschinen beeinflusst somit erheblich die Struktur des Problems für die übrigen Entscheidungsebenen. Aggregate werden installiert und Strukturen geschaffen, die es erlauben, Entscheidungsspielräume ökonomisch sinnvoll zu nutzen.

Damit im Zuge der strategischen Planung fundierte Entscheidungen getroffen werden, sind die Konsequenzen für die nachfolgenden Ebenen zu antizipieren. Wird die nachträgliche Aufteilung gegebener Lose als Normalfall und nicht als Ausnahme gesehen, ist dieser Aspekt schon bei der Wahl der Maschinen, des Transportsystems und bei der Dimensionierung der Lager zu berücksichtigen.

c) Als weitere strategische Maßnahme können Segmente definiert werden. Durch eine Entflechtung der Materialflüsse und eine entsprechende An- beziehungsweise Umordnung der Maschinen wird die Voraussetzung für eine lokale, dezentrale Steuerung geschaffen. Kleinere und oft serielle Ausschnitte können separat betrachtet, modelliert und geplant werden. Gerade innerhalb der Segmente mit seriellem Materialfluss können die Durchlaufzeiten durch überlappende Fertigung erheblich gesenkt werden.

Die Definition von Segmenten zielt in die Richtung der Beobachtungen von Spencer/Cox (1995) beziehungsweise von Finch/Cox (1988). Das Netzwerk der sich beliefernden einzelnen Produktionsstufen kann in vielen Fällen auf eine Kombination serieller Strukturen reduziert werden, indem man einige Maschinen mehrfach installiert und den Materialfluss entflechtet. Die Planung fokussiert auf die wesentlichen Komponenten eines Produkts, die übrigen Komponenten sind auf dieses Ergebnis hin anzupassen. Eine Empfehlung, auf solch umfassende Änderungen des Systems zurückzugreifen, ist auch bei Knolmayer (1987) angelegt, der argumentiert, dass in der wissenschaftlichen Analyse und in der Praxis das folgende, einseitige Denkmuster vorherrscht: Ein Problem und seine Daten sind gegeben, die Planung ist daraufhin anzupassen. Er hingegen fordert, dass das System anzupassen ist, wenn es im Analysezeitpunkt die nötigen Eigenschaften für ein als richtig erkanntes Prinzip nicht besitzt. Es soll von einem gestaltbaren statt von einem gegebenen Objekt-System ausgegangen werden (vgl. Knolmayer, 1987, S. 61).

Im Kontext der Diskussion zur Losüberlappung würde diese Segmentierung kleinere Einheiten mit seriellem Materialfluss sicherstellen. Wie Smunt et al. (1996) und Kher et al. (2000) zeigen, sind es gerade die seriellen Strukturen, für die sich ein Einsatz der Losüberlappung besonders empfiehlt. Den Ergebnissen ihrer Simulationsstudien folgend, profitiert ein Flow Shop stärker vom Einsatz der Losüberlappung als eine Struktur, die einem Job Shop mit verzweigten Materialflüssen entspricht (vgl. Smunt et al., 1996; Kher et al., 2000). Soll das Prinzip der Losüberlappung zum Einsatz kommen, bedeutet die Forderung Knolmayers, dass eine Umstrukturierung in kleine serielle Einheiten in Erwägung gezogen wird. Erneut ist eine Form der Anpassung des Systems zu beobachten, die mit der These „structure follows planning concept“ zu umschreiben ist.

Jede der drei Maßnahmen zur Umstrukturierung setzt allerdings voraus, dass erkannt wurde, wo sich eine Änderung beziehungsweise eine Ergänzung des Systems anbietet. Dem Bottleneck-Prinzip folgend, wird man die Engpässe suchen und dort ansetzen. Typisch für diese Orientierung ist ihr Charakter des Reagierens. Es besteht die Gefahr, dass Strukturen entstehen, die nur noch historisch aus einer Folge von „symptombekämpfenden“ Maßnahmen erklärbar sind.

Aus den folgenden Gründen wird man dieses Manko nie vollständig ausräumen können:

- Selbst eine detaillierte Problemanalyse kann an der Komplexität des Problems scheitern. In der Konsequenz werden Entscheidungen getroffen, die nicht an den Ursachen einer Fehlentwicklung sondern an ihren Symptomen ansetzen.
- Es wird immer Entscheidungen geben, die mit so kurzer Reaktionszeit zu treffen sind, dass sie nur oberflächlich z.B. vor dem Hintergrund der bisherigen Erfahrungen und unter dem bewussten Vernachlässigung wesentlicher Interdependenzen behandelt werden können.
- Einmal erschaffene Strukturen und Regelungen zeichnen sich allgemein durch ihr hohes Beharrungsvermögen aus. Häufig bleiben gerade die als provisorisch angesehenen Maßnahmen über lange Zeit bestehen. Dies gilt insbesondere, wenn spätere Entscheidungen auf dem Provisorium aufsetzen.

Neben den Faktoren, die unmittelbar einzelne Ressourcen betreffen und zum Auftreten eines Engpasses führen können, wie z.B. die geringe Kapazität einer Maschine oder ihre hohe Störanfälligkeit, bestimmt das gegenwärtige Planungsverfahren (z.B. über die Größe der Lose) und die in diesem Kontext benötigten Annahmen (wie z.B. ihre Form der Weitergabe) die Lage der Engpässe mit. Engpässe sind daher immer nur als bedingte Engpässe zu verstehen, für deren Beseitigung neben der Anpassung des Systems auch der Einsatz von nachträglich greifenden Anpassungsmaßnahmen in Frage kommen kann. Eine solche Maßnahme kann z.B. darin liegen, die Intensität oder die Dauer der Nutzung der Maschinen zu erhöhen. Geschieht dies aber dauerhaft, werden die bei Normal-Intensität beziehungsweise Normal-Arbeitsdauer auftretenden Engpässe auch dauerhaft verdeckt. Es besteht hier eine Parallele zur negativen Sicht der Lager, die im Zuge der Diskussion um die Just in Time Production aufkam (vgl. Silver, 1992). Dort wird argumentiert, dass die Existenz von Lagern das Erkennen von Planungsfehlern verhindert. Der vorgehaltene Bestand hilft, kurzfristige Schwankungen, beispielsweise in der Ausschussrate, auszugleichen. Einhergehend fallen für das Vorhalten dieses Bestands Kosten an. Da sie laufend und in etwa gleicher Höhe anfallen, besteht die Gefahr, sie zu übersehen. Zudem verdeckt die gelagerte Menge ihre eigentliche Verursachung (die schwankende Ausschussrate), da keine unmittelbar negativen Konsequenzen (Nachlieferungen) auffallen. Folglich besteht auch kein Anreiz, nach einer Verbesserung, z.B. nach Möglichkeiten zur Verringerung der Ausschussrate, zu suchen. In ähnlicher Form kann sich eine intensitätsmäßige Anpassung „verselbständigen“. Anpassungsmaßnahmen können somit Engpässe verdecken; Sie können aber auch zur Reaktion auf Engpässe notwendig sein. Zieht z.B. das PPS-System in Reaktion auf eine erwartete Auftragsspitze die Produktion von Vor- und Zwischen-

produkten vor, erscheinen Ressourcen als Engpass. Ursächlich für das Auftreten dieses Engpasses ist die (vom Entwickler vorhergedachte) Reaktion des Systems auf eine antizipierte Entwicklung. Es ist nicht ausgeschlossen, dass letztlich nur das Planungsverfahren selber (auf Grund seiner Logik und seiner problemvereinfachenden Annahmen) den limitierenden Faktor bildet und das bestehende System aus Maschinen, Lagern, Transporteinrichtungen etc. zu einer höheren Leistung im Stande wäre. Der Empfehlung Knolmayers (1987) folgend, ist dann der Ersatz des Planungssystem zu prüfen. Stellt sich dabei heraus, dass dieses Verfahren (zumindest mittelfristig) nicht durch ein besser abgestimmtes System ersetzt werden kann, sind nachträglich greifende, dezentral ansetzende Anpassungsmaßnahmen notwendig, um geeignet mit den Schwächen des PPS-Systems umgehen zu können. Kehrt man die Blickrichtung um und sieht die Losüberlappung zur Senkung der Durchlaufzeiten und zum Einhalten von Terminen als eine derartige Maßnahme an, dann erklärt sich ihre weite Verbreitung in der Praxis aus der Notwendigkeit, auf die Schwächen der PPS-Systeme zu reagieren.

3.2.1.2 Gestaltung der Produkte

Bei der Gestaltung des Produkts spielen die Standardisierung, die Modulbildung und die Verlagerung des Kundenauftragsentkopplungspunkts eine große Rolle. Bereits bei der Konstruktion sollte darauf geachtet werden, standardisierte Prozesse vorzusehen und viele Gleichteile beziehungsweise gleiche Baugruppen (Module) zu verwenden. Mit diesen Maßnahmen wird oft erst die Voraussetzung geschaffen, einzelne Segmente für die Produktion der häufig verwendeten Teile vorsehen zu können. Überspitzt formuliert kann man hier fordern: Das Produkt sollte derart konstruiert sein, dass sich die Steuerung und Planung seiner Produktion leicht bewerkstelligen lässt.

Wird außerdem der Punkt in der Produktion, in der das Produkt kundenspezifisch ausgestaltet wird, so spät wie möglich vorgenommen, vereinfacht sich die zu behandelnde Fragestellung im Detail erheblich (Zäpfel, 1989b). Die Standardisierung der Teile, das Verwenden von Modulen und das Verlagern des Entkopplungspunkts führt zu einer Reduktion der zu bewältigenden Vielfalt. Mit seiner Festlegung können zwei miteinander gekoppelte Regelkreise geschaffen werden (vgl. Missbauer, 1998, S. 14). Der erste steuert die kundenauftragsneutrale Fertigung vor dem Entkopplungspunkt, der zweite Regelkreis die kundenauftragsbezogene Fertigung, die im Entkopplungspunkt beginnt.

Die genannten Maßnahmen tragen dazu bei, eine Situation zu schaffen, in der – insbesondere bei kundenauftragsneutraler Fertigung – größere Mengen der Komponenten in seriellen Strukturen auf Basis von prognostizierten Bedarfsmengen losweise gefertigt werden. Schon bei der Produktentwicklung

werden damit Anforderungen an die Kapazität der Maschinen, die Dimensionierung des Transportsystems etc. definiert. Diese Maßnahmen in Kombination mit den entflochtenen Materialflüssen bedingen, dass bestimmte Maschinen in bestimmten Konstellationen zu installieren sind. Um die Planung zu vereinfachen, kommt somit wiederum eine Veränderung der maschinellen Ausstattung in Frage. Außerdem bestimmt die Gestaltung der Produkte und dabei insbesondere die Festlegung des Entkopplungspunkts die Konzeptwahl im wesentlichen mit.

3.2.1.3 Konzeptwahl

Eine Verlagerung des Entkopplungspunkts läuft letztlich auf eine Verlagerung der Bevorratungsebene als Schnittstelle zwischen der prognose- und auftragsgetriebenen Wertschöpfung hinaus. Eine wichtige Möglichkeit der strukturellen Problemgestaltung besteht darin, das Konzept zu wählen oder zu verändern, nach dem die Steuerung und Planung der Produktion vor oder nach dem Entkopplungspunkt erfolgen soll. Als grundsätzliche Alternativen stehen das Push-Konzept und das Pull-Konzept zur Verfügung. Durch den oder die Entkopplungspunkt(e) kann neben dem reinen Einsatz eines der beiden Konzepte auch eine Kombination aus beiden in Frage kommen.

Wird das *Push-Konzept* gewählt, dann sollen sich die Aufträge derart durch das System schieben, dass sie als Endprodukte in der gewünschten Spezifikation spätestens dann zur Verfügung stehen, wenn die Nachfrage zu befriedigen ist. Von der vorgelagerten zu der nachgelagerten Stelle liegt dabei Bringpflicht vor (vgl. Wildemann, 1988). Als Auslöser für die Weitergabe von Teilen dient deren Fertigstellung beziehungsweise die Vollendung des zugehörigen Loses bei geschlossener Fertigung. Da es im Allgemeinen unwirtschaftlich beziehungsweise wegen der kundenspezifischen Ausgestaltung unmöglich ist, die Nachfrage vollständig aus dem Endproduktlager zu befriedigen und zudem für die Herstellung der Produkte in der Regel mehr Zeit benötigt wird, als der Kunde zu warten bereit ist, muss von Prognosen über die benötigten Endproduktmengen ausgegangen werden, die die Produktion auslösen. Aus den prognostizierten Mengen sind mit Hilfe einer Stücklistenauflösung und der Vorlaufverschiebung die Bedarfsmengen und Zeitpunkte für die vorgelagerten Stufen zu ermitteln. Das Push-Konzept ist somit als ein zentral steuernder Ansatz charakterisierbar, bei dem die Disposition programmorientiert vorgenommen wird. Auf Grund von Rüstvorgängen und begrenzter Kapazität ist davon auszugehen, dass eine bedarfssynchrone Produktion nicht sinnvoll zu realisieren ist. Die einzelnen Bedarfsmengen müssen vielmehr zu Losen zusammengefasst werden (vgl. Missbauer, 1998, S. 24). Mit diesem Vorgehen ist aber bereits die Notwendigkeit der sequentiellen Problemlösung vorgezeichnet, da die Komplexität der Fragestellung nur diese zulässt. Werden in diesem Kontext viele kleine Lose

statt einiger großer Lose verwendet, gestaltet sich der Materialfluss gleichförmiger (vgl. dazu auch die Diskussion hinsichtlich der „Losgröße 1“). Zu dem Paradigmenwechsel, der in den späten achtziger Jahren von der Materialbestandsoptimierung hin einer Materialflussorientierung führte, merkt Knolmayer (1987) an, dass große Lose mit dem Öffnen von Schleusen an einem Staudamm verglichen werden können. „Die Existenz des Dammes und das zeitweilige Öffnen seiner Schleusen führt zu ungleichförmigen Systemzuständen, die einen kontinuierlichen Fluss hemmen.“ (vgl. Knolmayer, 1987, S. 65). Als Beispiel führt er Simulationsstudien auf Basis der belastungsorientierten Auftragsfreigabe an, die zeigen, dass sich bei einer Losteilung trotz zusätzlicher Rüstzeiten, die Leistung des Produktionssystems erhöhen kann, da große Lose in Push-Systemen dazu führen, „dass sich Engpasswirkungen über das Produktionssystem in unvorhersehbarer Weise verteilen“ (vgl. Knolmayer, 1987, S. 57). Neben den kleinen Losen kann aber auch die überlappende Fertigung eine sinnvolle Alternative zur Verstetigung des Materialflusses bieten.

Wird hingegen das *Pull-Konzept* gewählt, löst das Eintreffen der Nachfrage die Produktion aus, indem die Information über die benötigte Nachlieferung von der letzten Stufe ausgeht und die Produktion auf der oder den vorgelagerten Stufen initiiert. Die einzelnen Stellen haben eine Holpflicht: Erst mit dem Abholen beziehungsweise dem Bestellen durch die nachgelagerte Stufe wird das (Nach-)Produzieren auf der vorgelagerten Stufe ausgelöst (vgl. Kistner/Steven, 2001, S. 281, Zäpfel, 1996, S. 260). Als Vorteil des Pull-Konzepts ist nach Schneeweiß zu nennen, dass man auf den unteren Stufen nicht auf die unter Umständen fehlerhaft prognostizierten Bedarfe zeitlich weit entfernter höherer Stufen festgelegt ist. Diese Flexibilität wird aber erkauft durch hohe Sicherheitsbestände beziehungsweise hohe Fertigungskapazitäten, die es ermöglichen, auf die vergleichsweise späte Bedarfsmeldung schnell zu reagieren (vgl. Schneeweiß, 1999, S. 232). Das Pull-Konzept führt somit zu einer dezentralen Steuerung, bei der die Disposition verbrauchsgesteuert vorgenommen wird und die Informationen myopisch weitergegeben werden. Für ein Pull-Konzept sind folglich die gleichen Rahmenbedingungen wie für eine fließartige Fertigung zu schaffen: Die Stellen müssen in der Lage sein, flexibel auf die Änderungen des maschinellen und des personellen Kapazitätsbedarfs zu reagieren. Eine Konkurrenz von mehreren nachfolgenden Stufen um die Lieferung einer vorgelagerten gemeinsamen Stufe muss vermieden werden, da sonst im Konfliktfall eine der Stufen leer ausgeht und das rechtzeitige Nachproduzieren gefährdet wird. Die Mehrfachinstallation von Maschinen ist daher vorauszusetzen. Die Struktur des Systems muss im Extremfall so einfach ausgestaltet werden, dass auf den nachgelagerten Ebenen keine Planung mehr erforderlich ist. Dieser Zusammenhang kann als ein Substitutionsverhältnis charakterisiert werden (vgl. Wildemann, 1988, Kistner, 1994). Die bei einer Steuerung nach dem Push-Konzept häufig beobachtbaren hohen Bestände an Zwischenprodukten können vermieden wer-

den, wenn der Fertigungsbereich organisatorisch angepasst beziehungsweise reorganisiert wird. Das Kapital wird nicht mehr im Lager (im Umlaufvermögen) sondern in den zusätzlich zu beschaffenden Maschinen (im Anlagevermögen) gebunden. Insofern werden mit der Entscheidung für den Einsatz einer Pull-Steuerung langfristig Strukturen auf maschineller Ebene geschaffen, die ein flexibles Anpassen behindern. Zugleich wird die Notwendigkeit zur Planung auf den nachgelagerten Ebenen durch die Planung auf der strategischen Ebene substituiert. Im Extremfall entstehen vermaschte, sich selber steuernde Regelkreise, die keine laufende Planung erfordern. Einhergehend verliert das Unternehmen auch einen Teil seiner Flexibilität. Setzt man z.B. das Pull Prinzip mit einer Kanban-Steuerung um, fällt zudem auf, dass wieder nur ganze Container (ganze Lose) zur nächsten Stufe transportiert werden. Eine überlappende Fertigung, bezogen auf den Inhalt eines Containers ist in einer Kanban-Steuerung ausgeschlossen.

Im Zuge der strukturellen Problemgestaltung kann es sinnvoll sein, nur bestimmte Teile oder Produkte Pull- beziehungsweise Push- gesteuert zu fertigen (vgl. Zäpfel, 1996, S. 265 ff.) und entsprechende Segmente einzurichten.

Insgesamt ist festzuhalten:

- Für die praktische Umsetzung sind Mischformen und Provisorien typischer als ideale Strukturen, für die Modelle und Verfahren üblicherweise entwickelt werden. Maßnahmen zur Anpassung haben daher eine hohen Stellenwert.
- Standardisierung, Modulbildung und die Verlagerung des Kundenauftragsentkopplungspunkts spielen eine große Rolle zur strukturellen Problemgestaltung und bedingen sich in hohem Maß gegenseitig.
- Die auf den nachfolgenden Ebenen getroffenen Entscheidungen bestimmen, wie sich aus Sicht der strategischen Planung die Situation darstellt. Dies gilt insbesondere für das Erkennen von Engpässen.
- Anpassungsmaßnahmen können Engpässe verdecken, sie können aber auch dauerhaft notwendig sein, um den Schwächen eines Planungssystems zu begegnen. Daher ist ihr Einsatz schon a priori ins Kalkül zu ziehen.
- Nachgelagerte Ebenen müssen von ihren Gestaltungsmöglichkeiten Gebrauch machen. Das schnell als Manko interpretierte Abweichen von Vorgaben ist gerade nicht zu verhindern und sollte daher auch nicht als negativ angesehen werden.

3.2.2 Taktische Produktionsplanung

Die zweite Planungsebene bezeichnet Anthony (1965, S. 15ff.) als taktische Planung. Sie ist mittelfristig ausgerichtet und versucht, eine effektive Nutzung des bereitgestellten Potentials mit dem vorgegebenen Konzept innerhalb der strategischen Vorgaben sicherzustellen. Während für das gesamte Unternehmen als Ziel der taktischen Ebene die Gewinnmaximierung oder der Ausbau des Marktanteils vereinbart sein kann, überwiegt in der taktischen Produktionsplanung die Kostenminimierung.

Nach Missbauer (1998, S. 55), ist bereits die Annahme dieses einheitlichen Ziels eine wesentliche Idealisierung. Prinzipiell ist die Planung und Steuerung der Produktion immer unter mehreren, teilweise konfliktären Zielen zu betrachten. Neben der Kostenminimierung kommen zeit- und bestandsorientierte Ziele in Frage. Die Vielfalt der zu regelnden Größen macht zudem eine entsprechende Vielfalt im Regelungsmechanismus erforderlich. Zäpfel verweist in diesem Kontext auf den von Ashby (law of requisite variety) formulierten Zusammenhang: „only variety destroys variety“ (vgl. Ashby, 1956; Zäpfel, 1996b, S. 862). Ist die Vielfalt im Regelungsmechanismus nicht zu gewährleisten, müssen entweder die zu planenden Prozesse in ihrer Komplexität reduziert werden, oder es muss von Ausprägungen abstrahiert werden. Da die Maßnahmen, die die Struktur verändern, vornehmlich auf der strategischen Ebene ansetzen, bleiben der taktischen Planung die Abstraktion und die Dekomposition als problemvereinfachende Schritte. Häufig gesetzte Annahmen sind z.B., dass die Kapazität unbeschränkt verfügbar ist, oder dass von Abhängigkeiten zwischen den Produkten abstrahiert und eine separate Einproduktplanung auch im Mehrproduktfall vorgenommen wird. (vgl. z.B. Helber, 1994; Petersen, 1998, S. 65; Tempelmeier, 1999, S. 283).

Als Vorgaben dienen der taktischen Produktionsplanung die mit anderen Partialmodellen (z.B. der Absatzplanung) ermittelten Mengen. In der Regel handelt es sich dabei um aggregierte Größen, die aus Absatzschätzungen, aus den vorliegenden Aufträgen, oder dem gewünschten Bestand resultieren. Die dabei zu bewältigenden Aufgaben sind (Kistner/Steven, 2001, S. 4):

„*Wieviel wird produziert: Ausbringungsmengen und Losgrößen*“ und

„*Wieviel wird eingesetzt: Bereitstellung der Ressourcen*.“

Die taktische Planung sollte so gestaltet sein, dass sie effektive Entscheidungsalternativen für die beiden Fragen aufzeigt. Sie ist für das gesamte Unternehmen wesentlich, da hier repetitiv Entscheidungen zu treffen sind, die die Kosten und die erzielbaren Erlöse maßgeblich determinieren. Bei der Festlegung der Produktionsmengen ist ein Ausgleich zwischen gegenläufigen, kostenverursachenden Einflüssen zu finden. Entscheidungsvariable sind die Losgrößen, wobei ein Los die Menge genau eines Guts bezeichnet, die gemeinsam be-

schafft oder im Produktionsprozess ohne Unterbrechungen und Umrüstungen auf einer Anlage hergestellt wird (vgl. Kistner/Steven, 2002, S. 242). Während von den Rüstkosten eine Tendenz zur Zusammenfassung von vielen Teilen zu großen Losen ausgeht, resultiert aus der Berücksichtigung der variablen Kosten (Lagerhaltungskosten) eine entgegengerichtete Tendenz.

Bei der Entscheidung müssen neben den sachlichen (Konkurrenz um Ressourcen) auch zeitliche Interdependenzen berücksichtigt werden. Da Produzieren Zeit benötigt, ergeben sich gerade bei mehrstufigen Erzeugnisstrukturen vielfältige Abhängigkeiten zwischen den Perioden, die dann gravierende Folgen für die Lösbarkeit des Problems haben, wenn die Ressourcen kapazitiert sind und Rüstvorgänge erfasst werden sollen (vgl. Maes et al., 1991). Die Fragestellung der taktischen Produktionsplanung ist somit durch folgende Aspekte charakterisiert (Schneeweiß, 1999, S. 222):

- Mehrperiodizität,
- Mehrstufigkeit,
- Kapazitätsbeschränkungen.

Unter Einsatz geeigneter Modelle besteht die wesentliche Aufgabe darin (vgl. Domschke et al., 1997, S. 8):

detailliert art- und mengenmäßig das Produktionsprogramms hinsichtlich einzelner Produkte festzulegen,

mittelfristig Kapazitätsanpassungsmaßnahmen zu wählen, die Maschinen, das Personal und den Lagerbestand betreffen können).

Mit dieser üblichen Aufteilung der gesamten Problemstellung wird die Frage, in welcher Reihenfolge die ermittelten Lose einzulasten sind, zum Inhalt der nachgelagerten, operativen Planung. Es findet somit eine Dekomposition statt.

3.2.2.1 Losgrößenplanung

Zu den Fragestellungen der ein- oder mehrstufigen, statischen oder dynamischen, kapazitierten oder nicht-kapazitierten, Rüstkosten bzw. -zeiten berücksichtigenden Losgrößenmodelle und den zu ihrer Lösung empfohlenen Verfahren, existiert mittlerweile eine praktisch unüberschaubare Fülle an Arbeiten (z.B. Billington et al., 1983; Bahl et al., 1987; Tersine, 1994; Nahmias, 1997; Petersen, 1998; Drexl/Kimms, 1998; Silver et al., 1998; Tempelmeier, 1999). Dabei entspricht die Festlegung von Losen immer einer Problemvereinfachung durch Aggregation, da die Entscheidungen für alle Güter des Aggregats getroffen werden. Beispielsweise erfahren alle Teile eines Loses die gleiche Bearbeitung. In Anlehnung an Potts/van Wassenhove (1992) kann die Losbildung aber

auch derart verstanden werden, dass eine vorgegebene Menge, z.B. die während eines Monats insgesamt auftretende Nachfrage, in Lose aufzuteilen ist. Potts/van Wassenhove kehren damit die Blickrichtung um. Nicht das Zusammenfassen sondern das Aufteilen einer gegebenen Quantität in kleinere Einheiten sehen sie als Losbildung an.

Während im deutschen Sprachraum die Bezeichnung Losgrößenplanung üblich ist, wird in der englischsprachigen Literatur (z.B. Salomon, 1991; Drexl/Kimms, 1997; Potts/Kovalyov, 2000) neben dem „lotsizing" auch das so genannte „batching" behandelt. Auch wenn einige Autoren die beiden Begriffe als Synonyme verstehen: „Batching is also called lotsizing." (vgl. Kuik et al., 1994, S. 243), ist in Anlehnung an Woodruff/Spearman (1992) und Kimms (1997) explizit zu unterscheiden: Ein batch (wörtlich übersetzt: ein „Stapel") fasst unterschiedliche Güter, z.B. vorliegende Kundenaufträge, als Einheit zusammen, die wie ein Los ohne Unterbrechungen abgearbeitet werden. Dabei können zwischen den Gütern eines batch unwesentliche Rüstvorgänge, so genannte „minor setups" anfallen (vgl. Potts/van Wassenhove, 1992, S. 398). Im Rahmen dieser Arbeit soll auf die Besonderheit des batching nicht näher eingegangen werden, die z.B. von Jordan (1996) behandelt wird.

Sind in der betrieblichen Praxis Losgrößen festzulegen, überwiegt der Fall der mehrstufigen, mehrere Produkte und bzw. Bauteile umfassenden Problemstellung unter Berücksichtigung von Kapazitätsrestriktionen. Typischer Weise ist das Problem außerdem dynamisch und je nach vorliegender Produktionsstruktur für eine serielle, konvergierende, divergierende oder eine generelle Erzeugnisstruktur zu formulieren (Helber, 1994; Derstroff, 1995; Drexl/Kimms, 1997). Vorgeschlagen werden überwiegend Produktionsmengenmodelle, in denen für ein vorgegebenes zeitliches Raster, z.B. für jede Woche, jeden Tag oder jede Schicht zu entscheiden ist, ob und wie viel produziert wird. Im Fall der generellen Produktionsstruktur läuft die Modellierung auf das so genannte MLCLSP (*Mutli-Level Capacitated Lot Sizing Problem*) hinaus bzw. auf eines der damit nah verwandten Modelle (Billington et al., 1983; Helber, 1994; Tempelmeier, 1999; Stadtler, 1997). Zu deren grundsätzlichen Schwachstellen führt Petersen (1998, S. 78) aus:

> „Es ist mit den betreffenden Modellen nicht möglich, Lose stark unterschiedlichen Umfangs (die [...] durch die Divergenz der Kostenstrukturen verschiedener Produktarten, sondern auch durch Bedarfs- und Kapazitätsschwankungen im Zeitverlauf auftreten können) adäquat abzubilden."

Durch ein feines Periodenraster müsste eine bedarfsnahe Bereitstellung zugleich aber auch eine Einplanung umfangreicher Belegungen, die mehrere Perioden überdauern, ermöglicht werden. Zu diesem Zweck ist dann aber der Rüstzustand im Modell fortzuschreiben und genau zu erfassen, in welchem Zustand sich die Ressource am Anfang bzw. am Ende einer Periode befindet, oder alternativ ganz auszuschließen, dass eine Ressource durch mehr als einen Ar-

beitsgang pro Periode beansprucht wird. In derartigen „small bucket“-Modellen werden „Mikroperioden“ genutzt (vgl. Eppen/Martin, 1987; Domschke et al., 1997, S. 150), in denen implizit die „all-or-nothing“ Bedingung gilt: Entweder wird während der gesamten Periode genau ein Produkt produziert oder die Maschine steht leer. Der Lösung wird damit eine Struktur aufgepresst, da die gefundenen Mengen nur ganzzahlige Vielfache der in Stück gemessenen Kapazität sein können. Wird andererseits ein Modell mit so genannten „large bucket“, den „Makroperioden“ genutzt, wird überwiegend von der expliziten Modellierung der Reihenfolgeplanung abstrahiert (vgl. Drexl/Kimms, 1997, S. 225). Eine Variante bietet die Modellierung Dinkelbachs (1964). Dort wird zugelassen, dass Perioden nur teilweise genutzt werden, allerdings steht dann die restliche Periode über die Maschine leer, was wiederum wenig realitätsnah erscheint. Wird in die Modelle eine Modellierung der Rüstzustände explizit eingeführt, bzw. auf Mikroperioden zurückgegriffen, so sieht Petersen (1998) dies als eine (zumindest teilweise) Integration der Ablaufplanung in die Auftragsplanung (Losgrößenplanung) an. Eine Erweiterung, auf die nur in vergleichsweise wenigen Veröffentlichungen für den mehrstufigen Fall bisher eingegangen wurde. Für diesen Untersuchungsgegenstand hat sich im englischen Sprachraum die Bezeichnung: „lotsizing and scheduling“ etabliert (vgl. z.B. Haase, 1994; Drexl/Kimms, 1997; Grünert, 1998; Meyr, 2000). Die folgenden Gründe scheinen allerdings einer stärkeren Beachtung dieser Forschungsrichtung entgegenzustehen: Grundsätzlich läuft die Bestimmung einer optimalen Reihenfolge auf ein im Sinne der Komplexitätstheorie schweres Problem hinaus. Zudem sind auch Losgrößenmodelle nicht rechenbar, wenn realitätsnahe Annahmen gelten sollen (Maes et al., 1991). Wenn aber schon jedes Problem für sich genommen schwer ist, dann kann das kombinierte Problem schnell zu einer nicht sinnvoll handhabbaren Aufgabe werden.

Mit der klassischen Aufteilung in die taktische (Losgrößen festlegen) und die operative Ebene (Reihenfolgen bestimmen) bleiben zwar Interdependenzen unberücksichtigt, dennoch wird neben der deutlichen Reduktion des Schwierigkeitsgrads erreicht, dass jeweils die Aspekte behandelt werden, für die man sinnvoll die notwendige Datengrundlage bereitstellen, pflegen und verarbeiten kann. Es widerspricht der vorwiegend anzutreffenden Organisation des Planungsablaufs, den verfügbaren Daten und einer sinnvollen Verteilung der Aufgaben, wenn man versucht, Losgrößen und Sequenzen simultan zu behandeln.

3.2.2.2 Anpassung des Kapazitätsangebots

Unmittelbar mit der Entscheidung über die Größe der Lose ist auf der taktischen Ebene zu prüfen, ob die benötigte maschinelle Kapazität bereitsteht oder wie diese kostenminimal anzupassen ist. Prinzipiell wird die benötigte Kapazi-

tät erst nach der Entscheidung über die Reihenfolgen festliegen. Sind aber z.B. im Mehrproduktfall zur Vereinfachung die Produkte zunächst separat geplant worden, kann sich beim Zusammenführen der Pläne schon auf der taktischen Ebene zeigen, dass die vorgeschlagenen Losgrößen mit der verfügbaren Normalkapazität nicht realisierbar sind. Es kommen dann die folgenden Anpassungsmaßnahmen in Frage (vgl. z.B. Fogarty et al., 1991, S. 422 ff.; Jahnke/Biskup, 1999, 178 ff.; Schneeweiß, 1999, S. 243 ff.; Zäpfel, 2001, S. 192; Günther/Tempelmeier, 2003, S.210 ff.):

- Die Inbetriebnahme von Reservemaschinen (quantitative Anpassung),
- die zeitliche Anpassung durch Überstunden oder Zusatzschichten,
- die intensitätsmäßige Anpassung, indem z.B. Akkordarbeit vereinbart oder die Geschwindigkeit der maschinellen Bearbeitung erhöht wird,
- das innerbetriebliche Verlagern von Arbeit oder Arbeitnehmern zwischen Abteilungen,
- die Bereitstellung von zusätzlicher, extern beschaffter Kapazität z.B. durch die Beschäftigung von Leiharbeitern,
- die Verminderung oder Erhöhung der Fehlzeiten, indem z.B. den Mitarbeitern in auftragsschwachen Monaten der Ausgleich ihrer Überstunden empfohlen wird oder in auftragsstarken Monaten Urlaubstage nur eingeschränkt gewährt werden,
- das Vorziehen von Wartungsarbeiten, um in den auftragsstarken Perioden eine größere Zeitspanne produktiv nutzen zu können.

Während diese Maßnahmen in erster Linie das Kapazitätsangebot betreffen und den Umfang der Belastung, d.h. die Losgröße nicht verändern, kann im Zuge der mittelfristigen Anpassung auch eine Rückkopplung in die Losgrößenplanung nötig werden:

- Die Reduktion der zu leistenden Menge durch Fremdvergabe,
- die zeitliche Verschiebung von Produktionsmengen (Produktionsglättung) in spätere Zeiträume, womit sich die Losgrößen ändern, die an die operative Ebene weitergemeldet werden.

Durch diese Anpassungsmaßnahmen können offensichtliche Unzulässigkeiten auf der taktischen Ebene gemindert werden, allerdings wird man erst nach der Festlegung der Reihenfolgen tatsächlich wissen, welche Kapazität benötigt wird. Die dann einsetzenden Maßnahmen, wie z.B. weitere und kurzfristig vereinbarte Überstunden sind in der Regel sehr kostenintensiv. Auch aus diesem Blickwinkel ist die Diskussion von Maßnahmen geboten, die nachträglich greifen und dabei Vorgaben weitgehend erhalten.

Als wesentliche Charakteristika der taktischen Produktionsplanung sind damit festzuhalten:

- Da die notwendige Vielfalt im Regelungsmechanismus nicht zu gewährleisten ist, muss zur Komplexitätsreduktion von Interdependenzen und mehrfacher Zielsetzung abstrahiert werden.
- Die kombinierte Problemstellung aus Sequenz und Losgrößenplanung kann nicht sinnvoll realisiert werden. Die Aufteilung der Problemstellung mit sequentieller Lösung ist unumgänglich.
- Bei der Losgrößenplanung sind Mehrperiodizität, Mehrstufigkeit und Kapazitätsbeschränkungen zu berücksichtigen, allerdings haben die dafür entwickelten Modelle diverse Schwächen.

3.2.3 Operative Produktionsplanung

Zur konkreten Umsetzung der Entscheidungen dient die dritte Planungsebene nach Anthony, die als operative Planung bezeichnet wird. Sie ist kurzfristig ausgerichtet und versucht die Entscheidungen der taktischen Ebene effizient auszufüllen. Im Vordergrund stehen die zeitlichen Aspekte der Produktionsplanung. Nach Kistner/Steven (2001, S. 4) ist hier zu fragen:

„*Wann wird produziert: Terminplanung*“ und

„*In welcher Reihenfolge wird produziert: Reihenfolgeplanung*.“

Die operative Ebene legt somit die Reihenfolgen und Einlastungstermine fest. Sie steuert und überwacht den Ablauf und greift bei Unzulässigkeiten ein (vgl. Fogarty et al., 1991, S. 844, der diese Ebene als „shop floor control“ bezeichnet). Auf dieser können nur Zeitaspekte berücksichtigt werden, da eine Zuordnung von Kosten nicht praktikabel ist. Eine Reihenfolgeplanung unter Berücksichtigung von Kosten wäre nur dann möglich, wenn adäquate Knappheitspreise gegeben sind. Diese stehen aber erst nach Abschluss der Reihenfolgeplanung fest, da wesentliche Teile der Kapazität als ablaufbedingte Leerzeiten auf den Maschinen ungenutzt verstreichen. Es muss demnach von Schätzungen oder von Erfahrungswerten für den Grad der Nutzung und damit für die Knappheitspreise der Ressourcen ausgegangen werden. Wie Tempelmeier (1999, S. 193) darstellt, „überfordert“ aber die Aufgabe, Rüstkosten zu quantifizieren, das betriebliche Rechungswesen.

Als Zielsetzung der operativen Planung dienen daher Zeitkriterien, wie die Minimierung der Durchlaufzeiten. Diese haben den Vorteil, indirekt eine Kostenreduktion zu erwirken, da mit einer Senkung der Durchlaufzeiten im allgemeinen auch eine Senkung der Bestände und damit der Kapitalbindung einhergeht. Die operative Ebene versucht nun unter Berücksichtigung der Zeitziele,

Vorgaben der taktischen Planung umzusetzen. Ergänzend muss diese Planung im Stande sein, schnell auf Störungen zu reagieren und Eilaufträge in den Ablauf einzugliedern. Die wesentliche Herausforderung für die operative Ebene resultiert aber daraus, dass sich häufig die Vorgaben der taktischen Ebene als unzulässig erweisen: Unterschätzt man bei der taktischen Planung die ablaufbedingten Leerzeiten auf den Maschinen, wird die vorgenommene Anpassung der Kapazität auf der taktischen Ebene nicht zum Erfolg führen. Wird im Mehrproduktfall eine separate Planung für einzelne Produkte durchgeführt und darauf verzichtet, gegenseitige Abhängigkeiten zu berücksichtigen oder werden zur Vereinfachung Rüstzeiten ganz ausgeblendet, erhält man regelmäßig nicht realisierbare Vorgaben. Dies zeigt sich aber erst bei der Umsetzung des Plans. Da die Berücksichtigung zusätzlicher Restriktionen im Allgemeinen eine Einschränkung des Handlungsspielraums bedeutet, ist es praktisch unmöglich, dass sich ein aggregierter, unzulässiger Plan auf operativer Ebene als zulässig erweist (vgl. Tempelmeier, 1999).

Es ist daher auf der operativen Ebene notwendig, die Vorgaben anzupassen, damit die Planung der taktischen Ebene annähernd umgesetzt werden kann. Dazu bieten sich an:

- Eine erneute Anpassung der Kapazität, die aber i.d.R. sehr kostenintensiv ausfällt.
- Die Veränderung der Vorgaben. Zum Beispiel können Losgrößen modifiziert werden, indem die Lose aufgeteilt und abweichend von der Vorgabe Mengen zeitlich in andere Perioden verlagert werden.
- Durch eine überlappendende Weitergabe zwischen den Maschinen oder durch Splittung der Lose auf parallele Maschinen kann versucht werden, die gegebenen Losgrößen beizubehalten und dennoch termingerecht die Aufträge fertig zu stellen. Dadurch fallen zwar zusätzliche Transporte an bzw. es werden beim Splitten zusätzliche Rüstvorgänge notwendig, die Vorgabe der taktischen Ebene (die Losgröße) bleibt aber erhalten.

Insofern verfügt die operative Ebene über Maßnahmen, die zur Durchsetzung taktischer Vorgaben geeignet sind. Können die Vorgaben der taktischen Planung im wesentlichen beibehalten werden, entfallen umfangreiche Rückkopplungen zur taktischen Ebene. Es existiert damit ein Handlungsspielraum, den es zunächst wahrzunehmen gilt. Insbesondere die Möglichkeit, überlappend zu fertigen, sollte schon bei der taktischen Planung als dezentrale Variante der Anpassung bekannt sein. Sollen aber derartige Maßnahmen greifen, müssen sie im Kontext des gegenwärtigen Umfelds, also den in der Praxis der Produktionsplanung und -steuerung vorherrschenden Strukturen beurteilt werden. Deren Mängel sind im folgenden Abschnitt kurz zusammengefasst.

3.3 Mängel der PPS-Systeme

Die vorherrschenden Verfahren zur Produktionsplanung und -steuerung setzen in ihrer Struktur nach wie vor auf dem konventionellen Manufacturing Ressource Planning (MRP-II) auf und unterliegen damit dem phasenbezogenen Sukzessivplanungskonzept (vgl. Vollmann et al., 1992; Tempelmeier, 1999). Sie bestehen aus Modulen zur Primärbedarfsplanung, Bedarfsplanung für die Vorprodukte, Durchlaufterminierung, Kapazitätsplanung sowie Feinplanung bzw. Produktionssteuerung (vgl. Fleischmann, 1988). Sie gewährleisten eine terminierte Nettobedarfsermittlung mit entsprechender Vorlaufverschiebung. In diese ist allerdings die Losbildung inkludiert, da die Bedarfsermittlung für die Positionen einer Stufe die Lose der übergeordneten Stufen zum Ausgangspunkt hat (vgl. Missbauer, 1998, S. 11 ff.). Um dieses Problem lösen zu können, folgt die notwendige Problemaufteilung einem festen Schema, das dem hierarchischen Dekomponieren im Sinne Anthonys (1965) entspricht. Es wird sukzessiv erst über die Größe der Lose unter Kostenaspekten und unter Vorgabe von Plandurchlaufzeiten entschieden. Anschließend wird in Abhängigkeit der verfügbaren Kapazität angepasst bzw. rückgekoppelt und schließlich werden die Reihenfolge der Einlastung festgelegt und auf weitere Unzulässigkeiten reagiert.

Um numerisch beherrschbar zu bleiben, können die entstehenden Teilprobleme, selbst wenn man auf den Einsatz exakter, optimierender Verfahren verzichtet, nur sukzessiv abgearbeitet werden. Das Gesamtproblem wird in Phasen zerlegt, die aufeinanderfolgend mit zunehmendem Detailliertheitsgrad und abnehmendem Planungshorizont betrachtet werden. Innerhalb der Phasen wird außerdem erzeugnisweise sukzessiv vorgegangen. Daraus resultieren zahlreiche Schwächen, die hier kurz zusammengefasst werden (vgl. Fleischmann, 1988; Drexl et al., 1994; Missbauer, 1998, S. 23 ff.; Kistner/Steven, 2001, S. 256 ff.).

In der Primärbedarfsplanung wird dem *Synchronisationsprinzip* folgend, der Absatzplan dem Produktionsplan gleichgesetzt und damit impliziert, dass dieser realisierbar ist. Eine Annahme, die gerade bei knappen Kapazitäten regelmäßig nicht gegeben ist (vgl. Fleischmann, 1988).

Die Mengenplanung erfolgt als eine *erzeugnisbezogene Sukzessivplanung*, die weder die Konkurrenz der Erzeugnisse um knappe Ressourcen noch die aus der Mehrstufigkeit der Erzeugnisstruktur resultierenden kostenmäßigen Interdependenzen berücksichtigt (vgl. Drexl et al., 1994). Insbesondere wird der Mehrproduktfall so behandelt, als ob voneinander unabhängige Probleme vorliegen. In der Konsequenz ergeben sich bei der Zusammenführung der Lösungen unzulässige Vorschläge.

Die Terminierung von Aufträgen und die Kapazitätsplanung setzen auf Schätzungen der Durchlaufzeiten auf (vgl. Kistner/Steven, 2001, S. 279). Die in der Vergangenheit ermittelten Werte werden um einen Sicherheitszuschlag er-

höht und der Berechnung der Freigabezeitpunkte zugrunde gelegt. Tendenziell werden damit viele der Aufträge zu früh für die Einlastung freigegeben, was einerseits die Durchlaufzeit noch weiter erhöht und andererseits das zu lösende kombinatorische Problem erschwert. Das *Durchlaufzeit-Syndrom* tritt auf, wenn die realisierten Durchlaufzeiten wieder als Basis für den nächsten Planungsdurchlauf genutzt werden. Die auf Grundlage gegebener Plan-Durchlaufzeiten ermittelten Freigabezeitpunkte müssten das Ergebnis der Planung und nicht ihr Datum sein. Eine Berücksichtigung der Kapazität findet bei der Planung auf der taktischen Ebene nur implizit statt, indem die Plandurchlaufzeiten höher (bei knappen Ressourcen) oder niedriger (bei nicht knappen Ressourcen) gesetzt werden (vgl. Schneeweiß, 1999).

Zeigt sich bei der Planung der Einlastung, dass die Vorgaben nicht realisierbar sind, wird der Disponent eingreifen. Achtet dieser bei der Einlastung der Aufträge darauf, dass die Kapazitäten ausgeschöpft werden, wird er mit dem *Dilemma der Ablaufplanung* konfrontiert. Eine Einlastung weiterer Aufträge reduziert zwar Leerzeiten, führt aber zugleich zu einer Erhöhung der Wartezeit der Aufträge auf freie Maschinen in ihrer Maschinenfolge. Wird hier myopisch, z.B. mittels Prioritätsregeln entschieden, stauen sich die Aufträge vor den Engpässen (vgl. Kistner/Steven, 2001, S. 279).

Wird die Losgröße unter der Annahme unbeschränkter Kapazität bestimmt, kommt es zum *Tannenbaum-Effekt*. Die Lose können in Richtung der vorgelagerten Stufe nicht kleiner, sondern nur größer werden. Wird eine Teilmenge eines großen Loses zu einem frühen Zeitpunkt für die Produktion auf der nächsten Stufe benötigt, muss bei geschlossener Weitergabe das gesamte Los zu diesem Zeitpunkt fertiggestellt sein. Wesentliche Teile des Loses warten dann aber auf ihre Weiterbearbeitung, so dass hohe Bestände auftreten. Außerdem führen große Lose bei geschlossener Weitergabe zu ablaufbedingt hohen Leerzeiten. (vgl. Kistner/Steven, 2001, S. 280).

Da einerseits die Überprüfung der Umsetzbarkeit von Entscheidungen für die nachfolgenden Stufen in den gängigen PPS-Systemen a priori nicht erfolgt, sondern lediglich auf Unzulässigkeit durch Neuplanung oder durch myopisches Anpassung reagiert wird und sich andererseits die Systeme mit laufend aktualisierten Daten und manuellen Eingriffen konfrontiert sehen, zeigen sie eine hohe *Nervosität*. In der Folge akzeptieren die Mitarbeiter die vom PPS-System vorgeschlagenen Entscheidungen nicht, Planung und Realisation entfernen sich stark voneinander. Auf den Aspekt der Nervosität in PPS-Systemen und auf die Ursachen von Stabilitätsverlusten, insbesondere durch interne Faktoren wie der Wahl der Dispositionsregeln und deren Parametern, gehen z.B. Inderfurth/Jensen (1997) näher ein.

Wie Drexl et al. (1994, S. 1026) pointiert zusammenfassen, ist die Vorgehensweise des MRP-II als eine Aneinanderreihung heuristischer Improvisatio-

nen bei durchgängiger Vernachlässigung von kapazitierten Ressourcen zu beschreiben, deren geringe Qualität nicht überraschen kann.

3.4 Aufgaben der Untersuchung

Angesichts der beiden aufgezeigten Problemkreise, d.h. dem Denken und Modellieren innerhalb bestimmter Paradigmen und den in der Praxis vorherrschenden Mängeln der PPS-Systeme bieten sich nun zwei Alternativen für den Aufbau der weiteren Untersuchung an:

- Einerseits kann betrachtet werden, wie bei einer idealisierten Problemstellung vorzugehen ist. Als Ergebnis erhält man zwar theoretisch fundierte Handlungsempfehlung, allerdings betreffen sie Problemstellungen, die oft praxisfern sind.
- Andererseits können die Vorgehensweisen betrachtet werden, die sich in der Praxis bereits bewähren aber bislang nicht im Fokus des wissenschaftlichen Interesses stehen. Forschungsgegenstand sind damit die in der Empirie bewährten Maßnahmen; untersucht wird, unter welchen Umständen diese vorteilhaft sind.

In den meisten Veröffentlichungen wird die erste Alternative gewählt. Wie schon die Diskussion der Produkt-Prozess-Matrix (vgl. Spencer/Cox, 1995) zeigt, sind häufig nicht ideale Strukturen sondern Mischformen in den Unternehmen anzutreffen. Insofern kann die Modellwelt erheblich von der Realität abweichen. So kritisiert z.B. Petersen (1998) zunächst die geringe Praxisnähe vieler Modelle und postuliert dann aber für sein Modell der kapazitätsorientierten Auftragsplanung bei Serienfertigung (vgl. Petersen, 1998, S. 103, Postulate 10 und 11):

> „Jedes Fertigungslos belegt hinsichtlich seines Arbeitsganges genau einen Arbeitsplatz, d.h. es findet kein Lossplitting statt."

> „Jedes Fertigungslos wird im ganzen von einem Arbeitsplatz zum nächsten bzw. zum Bereitstellungslager weitergegeben, d.h. es liegt eine geschlossene Produktionsweise vor."

Eine größere Praxisnähe, die z.B. auch darin bestehen könnte, dass eine überlappende Weitergabe der Lose oder eine Lossplittung zugelassen wird, schließen diese Postulate aus. Würde man versuchen, diese beiden Punkte zusätzlich in das Modell aufzunehmen, steigt die Kompliziertheit so stark und schnell an, dass eine praktische Anwendung nicht möglich erscheint. Es würden Modelle entstehen, die den Anforderungen des im zweiten Kapitel dargestellten „decision calculus" (vgl. Little, 1970, S. B-470) nicht genügen können. Insofern kann es nicht sinnvoll sein, diese Aspekte zusätzlich in den bestehenden Mo-

dellen berücksichtigen zu wollen. Das Ergebnis wäre eine weitere Blackbox, deren betriebliche Akzeptanz von vornherein fraglich ist.

Außerdem kann an dieser Stelle auf die Beobachtungen des Wissenschaftstheoretikers Kuhn zurückgegriffen werden (vgl. Kuhn, 1970). Er stellt fest, dass die Entwicklung einer Wissenschaft in Schüben erfolgt, die durch den Wechsel von Paradigmen gekennzeichnet sind. Nach der Etablierung eines Paradigmas folgt nach Kuhn die Normalphase der wissenschaftlichen Entwicklung, die durch kleine Fortschritte geprägt ist. Sieht man die geschlossene Weitergabe als ein Paradigma an, stellt sich die weite Verbreitung dieser Annahme in den Modellen als ein gut begründbares und verbreitetes Phänomen des wissenschaftlichen Arbeitens dar.

Insgesamt dürfen diese Beobachtung allerdings nicht so verstanden werden, dass hier der vorherrschenden Vorgehensweise ihre Berechtigung streitig gemacht wird – vielmehr besteht das Anliegen dieser Arbeit darin, eine Perspektive aufzuzeigen.

Aus diesen Überlegungen wird die zweite Alternative gewählt und mit der Losüberlappung eine Vorgehensweise betrachtet, die sich in der Praxis bereits bewährt. Es zählt dort zu den gängigen Praktiken, die unter Kostengesichtspunkten gebildeten Lose weiter aufzuteilen. Es kommt zur Losteilung, zur Lossplittung und zur Losüberlappung (vgl. Kurbel, 2003, S. 149 ff.). Als Arbeitsdefinitionen können diese Begriffe wie folgt verstanden werden:

- Bei der *Losteilung* werden gegebene Lose in kleinere Quantitäten aufgeteilt und diese zeitlich getrennt eingelastet. D.h. aus dem ursprünglichen Los werden kleinere Lose gebildet und separat aufgelegt, während das ursprüngliche Los untergeht.
- Bei der *Lossplittung* (lot splitting) werden gegebene Lose aufgeteilt. Die entstehenden Teillose werden gleichzeitig auf parallelen Maschinen bearbeitet.
- Bei der *Losüberlappung* (lot streaming) werden gegebene Lose stufenüberlappend bearbeitet. Teile des Loses können bereits auf der folgenden Stufe eine Bearbeitung erfahren, während die Fertigstellung des restlichen Loses noch andauert.

Die Begriffe für die nachträgliche Aufteilung von Losen haben sich allerdings im Schrifttum noch nicht klar etabliert, beispielsweise wird „lot streaming“ mit „Lossplitting“ bei Domschke et al. (1997, S. 364) übersetzt. Eine ähnliche Verwendung der Begriffe findet sich bei Bogaschewsky et al. (1999) und bei Buscher (2000). Baker/Pyke (1990, S. 575) sehen das „lot splitting“ sogar als den Oberbegriff an, der den Spezialfall des „lot streaming“ einschließt.

Während die Lossplittung und die Losüberlappung nur einen temporären Charakter haben, da das gegebene Los als solches bestehen bleibt, geht bei der Losteilung die Zugehörigkeit der Teile zu dem gegebenen Aggregat „Los“ verloren, es entstehen neue Lose. Die Losteilung verändert damit die Vorgaben der taktischen Ebene, wohingegen die anderen beiden Maßnahmen eine Reaktion ermöglichen, die vorgegebene „Eckdaten“ beibehält.

Auf Basis einer empirischen Studie führen François/Wolf (2001) aus, dass unter den von Ihnen untersuchten 210 Software-Systemen zur Produktionsplanung und Steuerung immerhin 161 Systeme die Lossplittung dispositiv unterstützen. 173 Software-Systeme bieten weiter die Möglichkeit, überlappend gefertigte Lose zu verwalten, allerdings können lediglich 115 der 210 Systeme dem Disponenten auch eine planerische Unterstützung bei der Anpassung der Losgrößen anbieten. François/Wolf sehen darin eine erhebliche Systemschwäche bestehender PPS-Systeme. Insbesondere die Überlappung von Losen wird häufig nur verwaltet, eine tatsächliche Planung mit Hilfe von Algorithmen findet hier nicht statt (vgl. auch Gronau/Ibelings, 2002; Fandel/François, 2000). Insofern kann die Diskussion der Techniken zur Losüberlappung helfen, eine gegenwärtige bestehende Lücke zu schließen.

Diese Lücke kann ansatzweise auch damit erklärt werden, dass bislang in der wissenschaftlichen Diskussion die Thematik der nachträglichen Anpassung von Losgrößen nur wenig Resonanz gefunden hat. So wird der Aspekt z.B. bei Tempelmeier (1999, S. 267) in einer Fußnote kommentiert:

> „Zwar ist in vielen PPS-Systemen die Möglichkeit vorgesehen, die unter Kostengesichtspunkten gebildeten Lose u.U. wieder in kleinere Einheiten zu zerlegen und damit die Ergebnisse der Mengenplanung zu modifizieren. Diese sogenannte Losteilung ist ein weit verbreitetes Hilfsmittel zur Verkürzung der geplanten Durchlaufzeit der Aufträge. Sie entbehrt aber jeglicher methodischer Grundlage und zeigt nur, dass die Produktionsplaner dem Ergebnis der Losgrößenplanung nur geringe Bedeutung zumessen.“

Andere Autoren sehen das Forschungsgebiet der nachträglichen Modifikation gegebener Lose weniger kritisch (vgl. z.B. Zäpfel (2001), S. 179-182). Eine vergleichsweise ausführliche Darstellung findet der Leser bei Glaser/Geiger/Rohde (1992, S. 161-171). Sie sprechen alle drei Möglichkeiten der Modifikation an und stellen die Losüberlappung am Beispiel des Programmpakets „MIACS-TD“ vor. Glaser et al. (1992) enden mit dem Ergebnis, dass in der dort vorgesehene Überlappung schwerwiegende logische Fehler und Ungereimtheiten auftreten und einem potenziellen Benutzer die Verwendung nur abgeraten werden kann. Insgesamt kommen die Autoren zu dem Ergebnis, dass zur Umsetzung der nachträglichen Modifikation von Losen keine operationalen Verfahren Anwendung finden, die in der Lage sind, die Auswirkungen dieser Techniken richtig zu erfassen. Dabei hat sich, initiiert durch den Beitrag von Jacobs/Bragg (1988) und Potts/Baker (1989), eine mittlerweile breite und theo-

retisch fundierte Diskussion ergeben. Die folgende Tabelle zeigt eine Auswahl der Arbeiten:

Tabelle 1: **Wesentliche Veröffentlichungen zum Lot Streaming von 1988-1995**

Autor(en)	Thema	Zeitschrift
Jacobs/Bragg (1988)	Repetitive lots: flow-time reductions through sequencing and dynamic batch sizing.	Decision Sciences, Vol. 19, 281-294
Potts/Baker (1989)	Flow shop scheduling with lot streaming.	Operational Research Letters, Vol. 8, 297-303
Baker/Pyke (1990)	Solution procedures for the lot-streaming problem.	Decision Sciences, Vol. 21, 475-491
Trietsch/Baker (1993)	Basic techniques for lot streaming.	Operations Research, Vol. 41, 1065-1076
Baker/Jia (1993)	A comparative study of lot streaming procedures.	Omega, Vol. 21, 561-566
Glass/Gupta/Potts (1994)	Lot streaming in three-stage production processes.	European Journal of Operational Research, Vol. 75, 378-394
Vickson (1995)	Optimal lot streaming for multiple products in a two-machine flow shop	European Journal of Operational Research, Vol. 85, 556-575
Baker (1995)	Lot streaming in the two-machine flow shop with setup times.	Annals of Operations Research, Vol. 57, 1-11

Zieht man zur Einordnung der Thematik die deutschsprachigen Standardlehrbücher der Produktionsplanung hinzu, ist festzustellen, dass die meisten Autoren auf die Darstellung von Algorithmen zur Losüberlappung ganz verzichten. Offensichtlich messen sie dieser Technik bzw. den dafür entwickelten Algorithmen keine besondere Bedeutung zu. Schlägt man zur Einordnung der Thematik bei Silver et al. (1998, S. 654) nach, wird dort im Rückgriff auf Baker/Pyke (1990) die Technik der Losüberlappung zwar vorgestellt, behandelt wird sie aber im Kontext „Just-in-time“, „Optimized production technology (OPT)“ und „Conwip“. Beim Leser entsteht der Eindruck, dass Silver et al. (1998) bewusst eine Zuordnung zur Losgrößenplanung oder zur Reihenfolgeplanung vermeiden. Letztere würde der Einschätzung von Lee et al. (1997) entsprechen, die die Losüberlappung zu den „current trends in deterministic scheduling“ zählen; eine Zuordnung, die sich auch bei Baker (1998, S. 9.1) wiederfindet. Hingegen verzichtet Brucker (2001) ganz darauf, die Verfahren zu diskutieren, und auch bei Domschke et al. (1997, S. 365) wird der Sachverhalt nur sehr kurz angesprochen, dafür aber auf die weiterführende Literatur verwiesen.

Zur Begründung der Schwierigkeit, die Verfahren einzuordnen, können die folgenden Punkte angeführt werden:

Auf den ersten Blick erscheint es eher als Herausforderung, nach solchen Vorgehensweisen zu suchen, die ein nachträgliches Modifizieren unnötig machen, als die Techniken zu betrachten, die den Charakter von Reparaturmaßnahmen haben. Beispielsweise bezeichnet Petersen (1998, S. 105) das Lossplitting als eine „Notmaßnahme“ zur Abwendung eines möglichen Verzugs. Weiterhin diskutiert er die Berücksichtigung einer Überlappung von Arbeitsgängen im Kontext der Modellierung von Transportzeiten und argumentiert, dass eine Überlappung durch das Vorgeben negativer Mindestübergangszeiten, „problemlos“ integrierbar sei (vgl. Petersen, 1998, S. 105 f.). Diese Einschätzung ist allerdings nicht zutreffend, da es bei der überlappenden Fertigung wesentlich auf die Menge der vorzeitig weitergegebenen Güter ankommt. Die von Petersen vorgeschlagene und von der Größe des Loses unabhängige negative Übergangszeit kann nur den Fall abbilden, dass immer eine bestimmte Quantität überlappt. Dann aber ist das Ausmaß der Überlappung keine Entscheidungsvariable, sondern ein Datum, so dass potenzieller Entscheidungsspielraum verloren geht.

Die Losüberlappung entspricht zudem nicht der üblichen Aufteilung der Aufgabenstellungen gemäß der Ebenen nach Anthony. Tatsächlich ist die Frage, ob die Losüberlappung eher der Losgrößenplanung oder eher der Ablaufplanung zuzuordnen ist, nicht eindeutig beantwortbar. Einerseits werden Entscheidungen getroffen, die die zu realisierenden Losgrößen betreffen bzw. diese verändern und andererseits werden Aspekte der Reihenfolgeplanung berücksichtigt. Insofern wirft hier die betriebliche Realität eine Diskussion auf, die insbesondere für die Reihenfolgeplanung eine Herausforderung darstellen kann. Der Job, als die kleinste Einheit, wird zu einer disponierbaren Größe. Gegen eine Zuordnung in den Bereich der Losgrößenplanung spricht hingegen ein Argument, das Glaser et al. (1992, S. 170 f.) anführen. Sie heben die Schwierigkeit hervor, dass a priori die Kostenwirkungen nicht erfassbar sind, die von einer nachträglichen Modifikation der Lose ausgehen werden. In der hier vorliegenden Untersuchung wird die Losüberlappung somit als Teil der Ablaufplanung aufgefasst.

Ingesamt ist ohnehin fraglich, wie das Zusammenwirken von kapazitierten deterministischen Losgrößenmodelle und den betrieblichen PPS-Systemen zu beurteilen ist. Wie Jahnke (1998, S. 109) ausführt, können „gute Lösungen für ein bestimmtes Teilproblem als Bestandteil einer umfassenden Gesamtlösung zu beliebig schlechten Gesamtlösungen führen“. Es sollte daher zunächst in umfangreichen Simulationsstudien gezeigt werden, dass die Approximationsgüte derartiger Modelle tatsächlich besser ausfallen kann als die der klassischen PPS-Systeme. Es fehlt sonst die von Schneeweiß (1992 S. 4 ff.) geforderte Validierung, die im ersten Abschnitt angesprochen wurde. Jahnke kommt zu dem

Ergebnis, „daß in dieser Richtung nicht mit einem größeren Einfluß der Losgrößentheorie auf die Planungspraxis zu rechnen ist.“ (vgl. Jahnke, 1998, S. 110). Er liefert damit ein weiteres Argument, die Losüberlappung nicht in Losgrößenmodelle einzubeziehen, sondern sie im Kontext der Ablaufplanung zu behandeln.

Wie bei der Diskussion zur strategischen Produktionsplanung ausgeführt, sollte die vorherrschende Sichtweise und die mit ihr einhergehende geringe Resonanz der Losüberlappung überdacht werden. Ihr Einsatz ist nicht als ein Manko oder eine Notmaßnahme zu verstehen, sondern als Ausdruck dafür, dass hier ein notwendiger Gestaltungsspielraum genutzt wird. Da sich zudem die Problemstellung für die vorgelagerte Ebene stets in Abhängigkeit von den Maßnahmen darstellt, die von den nachgelagerten Ebenen getroffen werden, liegen Rückkopplungen vor. Sind nachgelagert erhebliche Reaktionsmöglichkeiten verfügbar, stellt sich aus Sicht der vorgelagerten Ebene das Problem anders dar. Die Alternativenmenge der vorgelagerten Ebene wird größer und eine andere Verteilung der Aufgaben möglich.

Eine Entscheidung für die zweite der beiden Möglichkeiten zum weiteren Aufbau der Arbeit lässt sich abschließend aus der bisherigen Umsetzung von anspruchsvollen, theoretischen Konzepten zur Planung und Steuerung der Produktion in der Praxis rechtfertigen. Stellt man die zahlreichen Ansätze, die zur Lösung der Aufgaben der taktischen Produktionsplanung entwickelt wurden (vgl. zur Übersicht z.B. Domschke et al., 1997, S. 69-180; Petersen, 1998, S. 41-96 oder Tempelmeier, 1999, S. 105-348), den tatsächlich in der Praxis eingesetzten Verfahren gegenüber (Glaser et al., 1992; Fandel et al., 1997; Silver et al., 1998; François/Wolf, 2001), zeigt sich eine große Diskrepanz. Die wenigsten Methoden und Modelle konnten bislang ihren Weg in die praktische Umsetzung finden.

Gleichzeitig wird aber seit Jahren in einschlägigen Publikationen der Eindruck erweckt, dass die breite Verwendung auch sehr anspruchsvoller Verfahren und Modelle z.B. aus der Lagerhaltungstheorie unmittelbar bevorsteht. Nur in wenigen Arbeiten wird diese Einschätzung nicht geteilt, beispielsweise schreibt Tempelmeier (1999) im Vorwort zur vierten Auflage der Material-Logistik: „Dabei wird besonders auf den immer noch sehr unbefriedigenden – und von Praktikern auch so empfundenen – Entwicklungsstand der dem konventionellen MRP-Sukzessivplanungskonzept folgenden EDV-gestützten Systeme zur Produktionsplanung und -steuerung eingegangen. Es wird gezeigt, daß die dort vorhandenen Planungsdefizite teilweise durch jetzt verfügbare Modellierungskonzepte und Planungsverfahren beseitigt werden können.“ Insbesondere der letzte Satz verdeutlicht, dass es auch mit den gegenwärtig leistungsfähigsten Verfahren nur in Teilen möglich ist, die bestehenden Defizite auszuräumen. Eine ähnlich zurückhaltende Einschätzung findet sich in Gün-

ther/Tempelmeier (2000). Dort wird ausgeführt, dass zwar mit dem Vorschlag von Drexl et al. (1994) zum Aufbau und zur Ausgestaltung eines kapazitätsorientierten PPS-Systems gezeigt ist, wie prinzipiell vorgegangen werden müsste: Es sollten mehrstufige Mehrprodukt-Losgrößenmodelle bei beschränkter Kapazität eingesetzt werden und diese sollten mit Hilfe eines hierarchischen Planungssystems aufeinander abgestimmt werden. Hinsichtlich der Umsetzung kommen die Autoren dann allerdings zu folgendem, ernüchternden Ergebnis (Günther/Tempelmeier (2000, S. 322):

> „Die Entwickler von PPS-Systemen haben diesen Ansatz eines kapazitätsorientierten PPS-Systems noch nicht aufgegriffen. Es mehren sich allerdings die Anzeichen dafür, daß das Problem der Vernachlässigung der Ressourcen und die unerwünschten Folgen für die Leistungsfähigkeit [...] nicht nur von den Software-Anbietern, sondern vor allem auch von ihren Kunden, den Anwendern von PPS-Systemen, erkannt worden sind."

Nimmt man die drei Jahre später erschienene fünfte Auflage zur Hand (Günther/Tempelmeier, 2003), stellt man fest, dass an Stelle einer Weiterentwicklung der PPS-Systeme zunehmend ein anderer Schwerpunkt in den Vordergrund tritt und sich ein erneuter Paradigmenwechsel andeutet. Statt das auf ein Unternehmen beschränkte Planungsproblem lösen zu wollen, wendet sich die Diskussion dem weitgesteckten Ziel zu, ganze Supply Chains bzw. Supply Networks zu betrachten. Die dafür entworfenen APS (Advanced Planning Systems) versuchen, „die verbesserte Informationslage für die Optimierung der Prozesse über die gesamte Wertschöpfungskette hinweg – auch unternehmensübergreifend – zu nutzen." (Günther/Tempelmeier, 2003, S 326). Allerdings schließen die Autoren mit der deutlichen Feststellung, dass der Sprung von einem PPS-System nach Sukzessivplanungskonzept, „in dem (fast) gar nicht geplant wird zu einem System, in dem „advanced" geplant wird, doch etwas hoch erscheint" (Günther/Tempelmeier, 2003, S. 338). Zudem klaffen die Versprechungen der APS-Anbieter (hinsichtlich der Qualität der von ihren Systemen „optimierten" Entscheidungen) und der Ist-Zustand der gebotenen Entscheidungsunterstützung sehr stark auseinander (vgl. dazu Knolmayer, 2001).

Trotz aller Defizite und der Erweiterung auf die Supply Chain ist eine Abkehr von den überwiegend MRP-II-basierten PPS-Systemen weder zu verzeichnen noch in Kürze zu erwarten. Daher hat die Diskussion von Maßnahmen zur Reaktion auf ihre Schwächen einen hohen Stellenwert. Ein Ersatz der gegenwärtigen PPS-Systeme scheitert, da nach wie vor wirkliche Alternativen fehlen. Weiterhin zeigt sich, wie bei der Diskussion der strategischen Maßnahmen bereits ausgeführt, dass es i.d.R. schwierig ist, von einmal vereinbarten Regelungen abzuweichen, da diese eine Tendenz zur Verfestigung haben. Aus der Verwendung eines MRP-II-basierten PPS-Systems resultieren zudem Strukturen im Unternehmen, die einen schnellen Wechsel erschweren. Folglich kommt den Maßnahmen, die im Rahmen dieser PPS-Systeme als losaufteilende Techniken

notwendigerweise praktiziert werden, noch für lange Zeit hohe Relevanz zu. Gleichzeitig muss aber auch deutlich gesagt sein, dass die Losüberlappung nur einen kleinen Beitrag liefern kann und keinesfalls alle Probleme der PPS-Systeme ausräumen wird.

In dieser Arbeit wird somit die zweite der eingangs genannten Alternativen gewählt. Das Vorgehen hat den Vorteil, dass es dem Akzeptanzproblem begegnet, da hier vornehmlich Techniken untersucht werden, die sich schon lange bewährt haben, bzw. deren Einsatz angesichts der erheblichen Mängel der gegenwärtigen PPS-Systeme nicht vermeidbar ist. Werden den Disponenten Regeln zur Überlappung von Losen angeboten, kann besonders groben Fehlern begegnet werden. Als Zielsetzung soll eine theoretisch fundierte Basis für eine Unterstützung der Disponenten bei der Planung überlappender Lose erreicht werden.

4. Rahmenbedingungen der Losüberlappung

In diesem Kapitel werden die Rahmenbedingungen der Losüberlappung vorgestellt, wobei zunächst auf die technischen und organisatorischen Einschränkungen einzugehen ist. Anschließend wird zur detaillierten Beschreibung der Überlappung eine Unterscheidung in Produktions-, Freigabe- und Transferlose vorgenommen. Schließlich ist die Losüberlappung von den übrigen Techniken abzugrenzen, die zur nachträglichen Anpassung der Losgrößen einsetzbar sind.

4.1 Technische und organisatorische Einschränkungen

Die in der Literatur als offene und geschlossene Fertigung charakterisierten Situationen werden in diesem Abschnitt als Ausgangspunkt gewählt und ihre Eigenschaften im Hinblick auf die Losüberlappung diskutiert. Dabei ist auf die Fertigung in Öfen gesondert einzugehen, da sie die Möglichkeiten, überlappend zu fertigen, stark einschränkt. Zur genaueren Beschreibung der Prozesse werden drei Formen unterschieden, in denen die Stücke nach ihrer Bearbeitung zum Transfer an die nächste Produktionsstufe oder für die Auslieferung an den Kunden verfügbar werden. Einschränkungen hinsichtlich der Einlastung der Stücke auf den Maschinen sind anschließend vorzustellen. Diese können sich aus dem Rüstprozess, einer eventuellen Unterbrechungs- oder Leerzeitfreiheit auf den Maschinen, sowie den Transport- und Lagerrestriktionen ergeben. Die folgende Abbildung 3 verdeutlicht die unterschiedlichen Perspektiven der Restriktionen für die Losüberlappung.

	Restriktionen hinsichtlich der:
Maschine A:	Weitergabe (Verfügbarkeit)
Ausgangslager A:	Lagerkapazität und -bedingungen
Transport:	Transportkapazität
Eingangslager B:	Lagerkapazität und -bedingungen
Maschine B:	Einlastung (Rüstprozess)

Abbildung 3: Einschränkungen bei der Anpassung von Losgrößen

Wiedergegeben ist nur der Ausschnitt von der teilweisen Weitergabe der Maschine A bis zur teilweisen Einlastung auf der Maschine B. Danach wiederholt sich das Schema. Neben den maschinenspezifischen Restriktionen sind unter Umständen lager- und transportspezifische Einschränkungen zu berücksichtigen. Ergänzend können bereichs- beziehungsweise unternehmensexterne Restriktionen auf der ersten Stufe den Bezug des Materials und auf der letzten Stufe die Auslieferung an die Kunden betreffen und zu weiteren Einschränkungen führen.

4.1.1 Offene und geschlossene Produktion

Wird jedes einzelne Teil nach seiner Fertigstellung direkt an die nachfolgende Stufe weitergeleitet, liegt die sogenannte offene Produktion vor, in der ein gleichmäßiger Produktionsausstoß zur Weiterbearbeitung erfolgt. Wird hingegen abgewartet, bis alle Teile des Loses bearbeitet sind, bezeichnet man dies als geschlossene Produktion, in der ein zeitpunktgeballter Produktionsausstoß beziehungsweise ein einmaliger Lagerzugang erfolgt (vgl. Kilger, 1973, S. 385 f.). Bei der überlappenden Fertigung liegt nun eine Situation vor, die weder als offene noch als geschlossene Fertigung zutreffend charakterisierbar ist:

- Gegen die Einordnung als offene Fertigung spricht, dass nicht jedes Teil einzeln, sondern Teile des Loses gemeinsam weitergegeben werden.
- Gegen die Einordnung als geschlossene Fertigung spricht, dass das Los nicht zu Ende gefertigt wird, bevor Teile daraus weitergegeben werden.

Insofern nimmt die überlappende Fertigung eine Zwischenstellung ein. Die produzierten Teile können „offen“, d.h. Stück für Stück verfügbar werden, der Transport zwischen den Maschinen findet dennoch „geschlossen“ statt, d.h. Teile werden angesammelt und gemeinsam weitertransportiert. Neben rein technischen Gründen kann aus organisatorischen Überlegungen oder aus wirtschaftlichem Kalkül die Verfügbarkeit der Teile beziehungsweise die Art der Produktweitergabe eingeschränkt und eine Losaufteilung verhindert werden (vgl. Kilger, 1973, S. 386; Domschke et al., 1997, S. 73).

Wird die geschlossene Fertigung beziehungsweise Weitergabe als generelle Regel zur Vereinfachung der Planung vorgegeben, ist zu prüfen, ob die Konsequenzen der Vereinfachung in einem angemessenen Verhältnis zu dem Potential stehen, das wegen dieser Vorgabe nicht genutzt werden kann:

- Die geschlossene Fertigung vereinfacht die Überwachung und Steuerung der Produktion, da zumindest die Bereitstellung der Teile sichergestellt ist, bevor das Los eingelastet wird.

- Schwankungen in den Produktionsgeschwindigkeiten einer Stufe können sich innerhalb eines Loses ausgleichen. Das Los schützt die Stufe somit vor dem Aushungern (*Starving*). Erfolgt hingegen die Weitergabe Stück für Stück, entfällt dieser Puffer.
- Die je Transport zu bewältigenden Mengen sind bei geschlossener Fertigung größer als bei offener oder bei überlappender Fertigung.
- Für eine geschlossene Fertigung muss der Raum ausreichend groß dimensioniert sein, in dem das Los auf seine Bearbeitung oder auf die Weitergabe wartet.
- Geschlossene Fertigung führt zu hohen Durchlaufzeiten.

Als Sonderfall, der eine geschlossene Produktion aus technischer Hinsicht erforderlich macht, ist die Fertigung in Öfen zu nennen. Hier ist sowohl die teilweise Einlastung als auch die vorzeitige Weitergabe problematisch, da auf der Maschine simultan mehrere Teile bearbeitet werden, wofür auch die Bezeichnungen „batch processor“ oder „batching machine“ (vgl. Jordan, 1996, S. 11, Brucker et al. 1998) geläufig sind. Neben den Öfen, in denen z.B. Porzellan gebrannt, Metall gehärtet oder Kunststoff durch Erhitzen in seiner Struktur verändert wird, finden sich derartige Prozesse bei der Veredelung von Oberflächen, wie beim Färben oder Lackieren in Tauchbädern. Beispielsweise werden Brillengläser zur Hartbeschichtung oder zur Entspiegelung unter Reinraumbedingungen in Lack getaucht und anschließend für mehrere Stunden in einem Ofen erhitzt, damit sich Lack und Glas dauerhaft verbinden. Weiterhin ist die gemeinsame Bearbeitung im Kontext der Chargenfertigung, z.B. in der chemischen oder pharmazeutischen Industrie typisch. Dort laufen chemische Reaktionen z.B. in Kesseln oder auf Reaktor-Anlagen ab. Charakteristisch für diese Art der Fertigung sind die folgenden Eigenschaften:

- Mehrere Teile erfahren simultan eine Bearbeitung, d.h. die Güter, deren Bearbeitung in einem Zeitpunkt gemeinsam begonnen wird, werden auch zu einem gemeinsamen Zeitpunkt fertig. Die vorzeitige Weitergabe ist nicht möglich.
- Die Bearbeitungsdauer ist in der Regel unabhängig von der Zahl der Teile.
- Die Kapazität (das Fassungsvermögen) ist beschränkt.
- Die Unterbrechung einer bereits laufenden Bearbeitung zur Eingliederung weiterer Güter (Lose) kann nicht realisierbar sein. Folglich wird die sukzessive Einlastung unzulässig.

Nimmt man als Beispiel einen Ofen, in dem Metall getempert wird, sollten weder Teile vor Beendigung der Bearbeitung entnommen, noch nach Beginn der Bearbeitung hinzugefügt werden, da sukzessiv eine Folge von Temperatur-

änderungen vorgenommen wird, die das Metall härten. Es muss somit die gesamte Menge bereitstehen, bevor die Bearbeitung beginnt. Bei einer Ofenfertigung kann es folglich aus rein technischen Gründen notwenig sein, sowohl die Entnahme als auch die Einlastung der Lose geschlossen vorzunehmen. Weitere Eigenschaften dieser Problemstellung und eine Abgrenzung zu nah verwandten Fragestellungen behandeln Lee et al. (1992); Webster/Baker (1995); Baker (1998); Brucker et al. (1998). Neben rein technischen Gegebenheiten können wirtschaftliche Gründe die Unterbrechung der Bearbeitung zur teilweisen Weitergabe oder Nachbestückung ausschließen. Ist z.B. der Energieverbrauch nach einem zwischenzeitlichen Öffnen eines Ofens zu hoch oder wird der bereits laufende Prozess, etwa in der Halbleiterfertigung, durch Störungen beim Nachbestücken gefährdet, ist von diesen Alternativen abzusehen, selbst wenn sie technisch machbar wären. Weiterhin kann eine notwendige Unterscheidung der Teile zu aufwendig sein, wenn sie beispielsweise in ein gemeinsames Säurebad zum Anlaugen der Oberfläche gegeben werden, darin aber nur eine bestimmte Zeitspanne liegen sollen.

Als Alternative zu einem geschlossenen Ofen können Durchlauf- oder Durchstoßöfen installiert werden. Bei diesen zieht ein Gitternetzband die Produkte langsam durch den Ofen oder beispielsweise durch ein Farbbad. Die Güter können einzeln in den Prozess eingegliedert, entnommen und weitergegeben werden. Allerdings kommt die mit hohen Investitionen verbundene Installation dieser Aggregate nur in Betracht, wenn eine Nutzung gewährleistet ist, die dauerhaft und kontinuierlich stattfindet. Eine überlappende Weitergabe auf den Vorstufen des Durchlaufofens kann helfen, diese Bedingungen zu erfüllen.

Lässt die Technik des Produktionsprozesses und das wirtschaftliche Kalkül hingegen eine überlappende Fertigung zu, muss die Produktion nicht erst in dem Moment beginnen, da alle Teile eines Loses an der Maschine vorliegen. Die Produktion kann einsetzen, wenn die Bearbeitung auf der Vorstufe noch nicht für alle Teile des Loses zu Ende gebracht ist. Damit steigt die Gefahr, dass auch kleine Störungen nachhaltige Verzögerungen hervorrufen. Unter Umständen kommt es sogar zu einer Unterbrechung der Produktion. Das eigentlich festgelegte Los, d.h. die Menge, für die eine Fertigung ohne Unterbrechungen vorgesehen war, ist nicht realisierbar, da sie z.B. nicht rechtzeitig angeliefert wurde. Je nach Art des Produktionsprozesses muss dann die Maschine z.B. gereinigt oder neu gerüstet werden. Bei der Entscheidung für oder gegen eine überlappende Fertigung sind daher die folgenden Aspekte zu berücksichtigen:

- Die Vorteile der schnelleren Weiterbearbeitung bei Überlappung.
- Die verringerten beziehungsweise verlagerten Bestände.
- Das frühere Erkennen von fehlerhaften Teilen.

- Die Konsequenzen für die Dimensionierung der Lagerflächen und der Kapazität der Transporteinrichtungen.
- Der höhere organisatorische und planerische Aufwand, den eine überlappende Fertigung gegenüber einer geschlossenen Fertigung mit sich bringt.
- Die bei Überlappung mit größerer Wahrscheinlichkeit auftretenden Unterbrechungen.

Der dritte Aspekt erklärt sich daraus, dass Fehler häufig erst bei der nachfolgenden Bearbeitung bemerkt werden. Da bei Losüberlappung Teile früher weitergegeben werden, fallen Fehler auch früher auf. Allerdings kann der Prozess auch vergleichsweise schnell aushungern. Die Argumente für und gegen die überlappende Weitergabe weisen damit eine strukturelle Ähnlichkeit zu der Diskussion der Vor- und Nachteile der Produktion auf Abruf (Kistner/Steven, 2001) auf. Die Maschinen müssen erhebliche Anforderungen hinsichtlich ihrer Störungsfreiheit erfüllen, will man sicherstellen, das der Prozess auch bei teilweise weitergegebenen Losen ohne zu Stocken abläuft.

4.1.2 Formen der Verfügbarkeit

Eine Aufteilung der Lose kann innerhalb des Produktionsprozesses als teilweise Weitergabe oder nach dem Produktionsprozess als Teillieferung an den Kunden vorgenommen werden. In beiden Fällen können drei Formen vorliegen, in denen die Teile verfügbar werden (vgl. Dobson et al., 1987; Potts/van Wassenhove, 1992; Şen et al., 1998):

Definition: *Stückverfügbarkeit*

> Stückverfügbarkeit (item availability; item completion; item flow)ist gegeben, wenn jedes einzelne Stück sofort nach seiner Fertigstellung zur nächsten Stufe transportiert oder an den Kunden ausgeliefert werden kann.

Definition: *Losverfügbarkeit*

> Losverfügbarkeit (batch availability; sublot completion; batch flow) liegt vor, wenn die Stücke für den Transport zur nächsten Stufe oder die Lieferung an den Kunden erst dann verfügbar sind, wenn die Bearbeitung des Teillloses abgeschlossen ist, aus dem sie hervorgehen.

Definition: *Auftragsverfügbarkeit*

> Auftragsverfügbarkeit (job availability; job completion) ist gegeben, wenn keines der Lose kleiner als ein Auftrag gewählt werden darf beziehungsweise die Aufträge erst dann an den Kunden auszuliefern sind, wenn sie komplett vorliegen.

Bei der bereits vorgestellten offenen Produktion wird die Stückverfügbarkeit in vollem Umfang ausgenutzt, womit die offene Produktion den ersten von drei Extremfällen darstellt:

Jedes einzelne Stück steht unmittelbar zur Bearbeitung an der nächsten Stufe bereit, sobald seine Bearbeitung auf der Vorstufe abgeschlossen ist. Obwohl es die Bezeichnung „stückverfügbar“ nahe legt, ist dabei nicht gefordert, dass die verfügbare Menge ganzzahlig sein muss. Gemeint ist lediglich, dass separat für jede beliebige gefertigte Quantität über die Weitergabe entschieden werden darf und dies unabhängig von der Fertigstellung anderer Einheiten erfolgt. Je nach Prozess kann es zulässig sein, einen beliebigen, nicht ganzzahligen Anteil der Lose weiterzugeben, die dann z.B. in Gewichts-, in Längen- oder Volumeneinheiten bemessen sind. Ebenso sind Fertigungsprozesse denkbar, in denen nur ganze Stücke weitergegeben werden dürfen. Die Stückverfügbarkeit kann folglich sowohl für eine Losüberlappung mit diskreten als auch mit kontinuierlichen Losen vorgegeben sein.

Im zweiten Extremfall wird die Anzahl der Teillose auf eins und damit die Größe des (Teil-)Loses auf die Auftragsgröße gesetzt, so dass sich der Fall der Auftragsverfügbarkeit ergibt. Verlangt man diese sowohl während des Produktionsprozesses als auch für die Auslieferung an den Kunden, liegt der klassische, in der Maschinenbelegungsplanung fast durchgängig unterstellte Fall vor (vgl. z.B. French, 1982, S. 8 f.). Jeder Auftrag (traditionell als „Job“ bezeichnet) ist eine nicht weiter unteilbare Einheit. Wird ein Job einmal auf einer Maschine eingelastet, ist seine Produktion auch zu Ende zu führen. Jede überlappende Fertigung – und damit der Untersuchungsgegenstand dieser Arbeit – wird durch Standardannahmen ausgeschlossen. Die Diskussion der Losüberlappung geht somit über den traditionellen Fokus des Scheduling und Sequencing hinaus. Da Losüberlappung durch Auftragsverfügbarkeit im Produktionsprozess ausgeschlossen wäre, wird diese Form der Verfügbarkeit hier nicht weiter behandelt.

Der dritte Extremfall ist die Ofenfertigung. Hier ist die Situation dadurch charakterisierbar, dass sich die Losverfügbarkeit durch die Technik des Prozesses ergibt. An dieser Stelle wird deutlich, dass die Verfügbarkeit nur eine stufenbezogene Eigenschaft darstellt. Auf einen Ofen mit Losverfügbarkeit können z.B. Stufen folgen, auf denen Stückverfügbarkeit gilt. An die Ergebnisse von Spencer/Cox (1995) anknüpfend, sind derartige Mischformen möglich und realistisch. Auch hier ist an die Empfehlung von Lee et al. (1997) zu erinnern, die eine größere Realitätsnähe für die Modelle des Scheduling fordern.

Als Annahme bei der Modellierung der Losüberlappung wird die Losverfügbarkeit am häufigsten gesetzt. Der Einsatz der Güter auf der nächsten Produktionsstufe kann hier frühestens für den Zeitpunkt eingeplant werden, in dem das letzte Los fertiggestellt ist, das Teile enthält, die auf der nächsten Stufe gemein-

sam einzulasten sind. Auf den ersten Blick ergibt sich hier ein Widerspruch. Zieht man nämlich die Definition der Losüberlappung heran, wird dort gefordert, dass die Lose stufenüberlappend bearbeitet werden. Der Widerspruch löst sich, wenn man berücksichtigt, dass sich die in der Definition verlangte Überlappung auf die ursprünglich vorgegebene Losgröße bezieht und nicht unbedingt auch die zu bildenden Teillose betreffen muss. Weiterhin bedeutet Losverfügbarkeit nicht, dass die Lose über die Stufen hinweg höchstens größer werden können. Eine Aufteilung in kleinere Lose ist möglich, indem aus einem größeren Los – allerdings erst nach seiner Fertigstellung– kleinere Lose für die Folgestufe gebildet werden. Ebenso kann sich aus dem Zusammenziehen kleinerer Lose wieder ein größeres für die Folgestufe ergeben (das so genannte „mating“ tritt auf). Als Einwand gegen die Unterscheidung der Stück- und der Losverfügbarkeit kann nun argumentiert werden, dass sich die Weitergabe bei Stückverfügbarkeit ex post als eine Situation unter Losverfügbarkeit darstellen lässt: Es werden von vornherein kleinere Lose aufgelegt und diese nach ihrer Fertigstellung (also unter Losverfügbarkeit) komplett weitergegeben. Die folgende Abbildung 4 verdeutlicht den Zusammenhang. Auf ihrer linken Seite ist die Situation unter Stückverfügbarkeit gezeigt, während rechts die Losverfügbarkeit unterstellt wird. Vergleicht man zunächst die Auslieferung an den Kunden, die sich an die Fertigung auf der Stufe 2 anschließt, werden in beiden Fällen die gleichen Mengen in übereinstimmenden Zeitpunkten ausgeliefert.

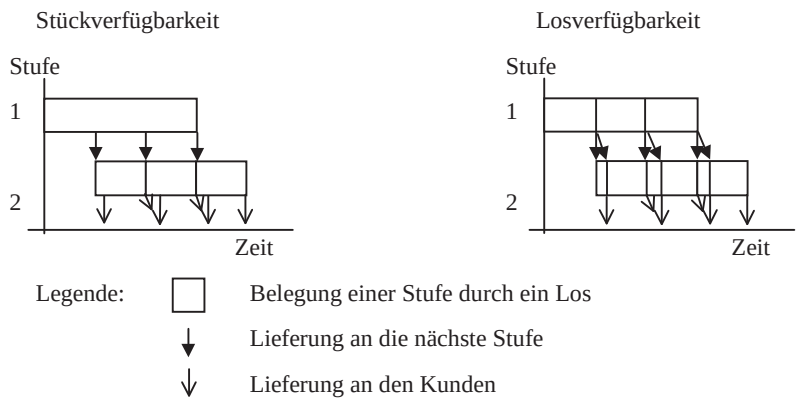


Abbildung 4: Stück- versus Losverfügbarkeit

In der Situation unter Stückverfügbarkeit wird auf Stufe 1 ein Los aufgelegt. Dieser Sachverhalt ist durch den ununterbrochenen Balken auf der ersten Stufe dargestellt. Noch während die Bearbeitung andauert, können wegen der Stückverfügbarkeit Teile an die Stufe 2 weitergegeben werden. Dort erfolgen aus der laufenden Bearbeitung der 3 Lose 4 Lieferungen an die Kunden, teilweise warten aber auch Teile auf ihre Weitergabe, was die diagonal verlaufenden Pfeile andeuten. Auf der rechten Seite der Abbildung 4 ist hingegen der Fall der Los-

verfügbarkeit dargestellt, der zu dem gleichen Ergebnis (den gleichen auslieferten Mengen in den gleichen Zeitpunkten) aus Sicht des Kunden führt. Hier müssen wesentlich mehr Lose aufgelegt werden, um den gleichen Sachverhalt abzubilden. Insofern kann die Annahme der Stückverfügbarkeit auch zu einer Vereinfachung der Darstellung führen. Dies ist insbesondere dann der Fall, wenn Rüstvorgänge abzubilden sind. Nehmen wir an, dass mit dem Beginn der Bearbeitung eines Teilloses auf der Stufe 2 jeweils ein Rüstvorgang nötig wird, dann ist bei Losverfügbarkeit 6 mal zu rüsten, bei Stückverfügbarkeit nur 4 mal. Diese Gegenüberstellung macht deutlich, dass zwar die Stückverfügbarkeit ex post wie eine Losverfügbarkeit darstellbar ist, dabei aber wichtige Aspekte, wie z.B. das Rüsten nicht richtig erfassbar sind. Insofern haben beide Formen ihre Berechtigung. Generell kann sich die Losverfügbarkeit aus den folgenden Gründen ergeben:

- Der Produktionsprozess lässt rein technisch gesehen keine zwischenzeitliche Entnahme aus einem bereits laufenden Los zu. Dies ist z.B. bei einer Ofenfertigung der Fall.
- Das Los ist an ein Transportmittel (Palette oder Container) beziehungsweise an einen Werkstückträger gebunden und muss auf diesem Träger bis zum Ende der Bearbeitung des letzten Teils des Loses verbleiben. Im Unterschied zum Ofen findet hier keine simultane Bearbeitung statt, d.h. die Dauer der Bearbeitung variiert je nach Größe des Loses. Bei einem Ofen ist dies gerade nicht der Fall, selbst wenn er nur zur Hälfte bestückt ist, ändert sich die Dauer z.B. eines Härteprozesses nicht.
- Die Vereinfachung der organisatorischen Abläufe während der Produktion kann ein Argument für die Losverfügbarkeit darstellen.
- In Folge der Qualitätssicherung kann es erforderlich sein, dass die Bearbeitung der einmal zusammengefassten Stücke immer abzuschließen und zu dokumentieren ist, bevor Teile an die nächste Stufe weitergegeben werden.

Der letzte Aspekt berührt die bislang nur wenig diskutierte Thematik des sogenannten Lot Tracing (vgl. Petroff/Hill, 1991; Steele, 1995). Dort wird argumentiert, dass es auf Grund der zunehmend umfassenderen Garantievorschriften für jedes einzelne Stück problemlos möglich sein muss, festzustellen, welche anderen Teile die gleiche – z.B. fehlerhafte – Bearbeitung erfahren haben. Um ein effizientes Lot Tracing zu ermöglichen, ist es zweckmäßig, die Losverfügbarkeit als eine generelle Regel vorzugeben.

4.1.3 Rüsten, Warten, Transportieren und Lagern

Während die Formen der Verfügbarkeit die Art der Weitergabe von einer Maschine charakterisieren, berühren sie nicht den Aspekt, dass auch hinsichtlich der Einlastung Restriktionen bestehen können. Wichtig sind in diesem Kontext die Rüstvorgänge, die Warte- und Leerzeitfreiheit, die Transporte, sowie die Kapazität der Lager, die den Anwendungsbereich der teilweisen Einlastung und damit der Losüberlappung einschränken können.

Hinsichtlich des Rüstens sind die maschinellen Anlagen dahingehend zu unterscheiden, inwiefern sie den Rüstzustand nach Beendigung der Produktion eines Loses beibehalten.

- Im ersten Fall bleibt der Rüstzustand (z.B. die eingegebene Folge vorzunehmender Bearbeitungen auf einer NC-Maschine) nach Beendigung der Bearbeitung erhalten. Schließt sich – auch durch einen zwischenzeitlichen Leerstand unterbrochen – die gleiche Art der Bearbeitung an, ist kein erneutes Rüsten erforderlich.
- Im zweiten Fall bleibt der Rüstzustand nur dann erhalten, wenn sich die Produktion des nächsten Loses ohne Leerzeit anschließt. Jede (längere) Unterbrechung der Nutzung der Maschine bedingt einen erneuten Rüstvorgang.
- Im dritten Fall verliert die Maschine mit dem Ende der Bearbeitung ihren Rüstzustand. Dieser Fall ist z.B. dann gegeben, wenn immer nach der Beendigung der Bearbeitung eines Loses Reinigungsarbeiten erforderlich werden, die nicht unter Beibehaltung des Rüstzustandes möglich sind.
- Weiter kann in das antizipative, getrennte Rüsten (detached setup) und das verbundenen Rüsten (attached setup) (vgl. Chen/Steiner, 1998; Skriskandarajah/Wagneur, 1999) unterschieden werden.

Beim verbundenen Rüsten werden ein oder mehrere Stücke unmittelbar für den Rüstvorgang benötigt. Beispielsweise ist es bei spanabhebenden Prozessen i.d.R. erforderlich, dass mindestens eines der weiter zu bearbeitenden Teile auch tatsächlich vorliegt, um versuchsweise die Bearbeitung durchzuführen und dabei die Maschine richtig einzustellen. Unter Umständen müssen für den Rüstvorgang sogar alle Teile des Loses an der Maschine vorliegen (vgl. Potts/Kovalyov, 2000). Dieser Fall ist beispielsweise dann gegeben, wenn die Teile in einen gemeinsamen Werkstückträger einzulegen sind. Weiterhin kann es im Zuge der Qualitätssicherung nötig sein, dass Probeteile gefertigt und nachgemessen werden, bevor die Produktion auf der Maschine freigegeben wird. Fasst man diesen Vorgang als Rüsten auf, ist er nur möglich, wenn einige Teile zur Bearbeitung an der Maschine verfügbar sind, aus denen dann zufällig

für die probeweise Bearbeitung und Qualitätskontrolle ausgewählt wird. Offensichtlich trägt eine überlappende Fertigung dazu bei, dass das verbundene Rüsten früher einsetzen kann. Ist zudem der Rüstprozess sehr zeitintensiv, kann es sinnvoll sein, bereits das erste gefertigte Stück von der Vorstufe an die zu rüstende Maschine weiterzugeben. Dieser Zusammenhang wurde bereits von Kropp/Smunt (1990) bei der Entwicklung ihrer „Flag-Heuristik“ genutzt. Sie behandeln einen Einprodukt Flow Shop und schlagen im Fall von umfangreichen Rüstzeiten vor, dass durch das Vorausschicken einer sehr kleinen Quantität sichergestellt wird, dass sich die Maschinen im richtigen Rüstzustand befinden, bevor das eigentliche Los eintrifft.

Im Unterschied zum verbundenen Rüsten kann das antizipative Rüsten (vgl. Chen/Steiner, 1997) ohne Vorliegen der zu bearbeitenden Teile vorgenommen werden. In der Regel kann z.B. ein Werkzeugwechsel an einer Maschine unabhängig von der Verfügbarkeit zu bearbeitender Stücke erfolgen.

In der betrieblichen Praxis werden allerdings nicht nur diese beiden idealtypischen Situationen vorliegen, sondern es können alle genannten Formen auch vermischt – zwischen den Maschinen und an einzelnen Maschinen – auftreten. Teile des Rüstvorgangs, wie die Bestückung mit Werkzeugen können unabhängig erfolgen, zum Justieren und zur Qualitätssicherung müssen hingegen einige oder alle zu bearbeitenden Teile vorliegen. Die in der einschlägigen Literatur vorgenommene Unterscheidung, in der ausschließlich einer der beiden Fälle unterstellt wird, bedeutet wieder eine Generalisierung; die Zwischenstufen bleiben bislang unberücksichtigt.

Eine weitere Einschränkung bei der Anpassung von Losgrößen ergibt sich, wenn auf Grund des Produktionsprozesses Leerzeit- oder Wartezeitfreiheit erreicht werden muss.

Im ersten Fall, dem „no-idling“, sind für die Maschinen Unterbrechungen durch Stillstände unzulässig (vgl. Trietsch/Baker, 1993; Baker/Jia, 1993). In einem Maschinen-Gantt-Diagramm betrifft diese Eigenschaft somit horizontal die einzelne Stufe. Sie kann unmittelbar aus dem Produktionsprozess folgen, wenn z.B. Farbe oder Klebstoff in der Maschine einzutrocknen droht und dann ein erneutes Rüsten erforderlich wird. Andererseits kann diese Einschränkung auch als generelle Regel vorgegeben werden. Beispielsweise können die Energiekosten bei einem zwischenzeitlichen Herunterfahren der Aggregate zu hoch ausfallen. Ebenso kann die Maschine bei häufiger Unterbrechung der Produktion überproportional verschleißen etc. In Folge der Leerzeitfreiheit sind die Lose auf der Stufe ohne Unterbrechung nacheinander einzulasten. Innerhalb des Planungshorizonts darf lediglich führende Leerzeit vor der ersten Bearbeitung auftreten. Weder zwischen den Losen noch innerhalb eines Loses sind Stillstände zulässig.

Im zweiten Fall wird die wartezeitfreie Fertigung („no-wait“) für die Teile gefordert. Sie gilt damit (im Maschinen-Gantt-Diagramm vertikal) für ein Los über die Stufen seiner Bearbeitung und schließt dabei aus, dass ein Stück zwischen dem Ende einer und vor dem Beginn der nächsten Bearbeitung warten (vgl. Hall/Sriskandarajah, 1996). Im Kontext der Ablaufplanung ist die Berücksichtigung derartiger Restriktionen gebräuchlich und sinnvoll, da die dort betrachteten Aufträge, die Jobs, kleinste Einheiten darstellen, die nicht weiter unterteilt werden können. Wird diese Restriktion auch bei der Betrachtung einer losweisen Fertigung unterstellt (vgl. z.B. Sriskandarajah/Wagneur, 1999; Kumar et al., 2000; Hall et al., 2003), scheint auf den ersten Blick ein Widerspruch vorzuliegen. Üblicherweise besteht ein Los doch aus mehreren Stücken und wenn diese nicht in einer Ofenfertigung hergestellt werden, folgt aus der Forderung der unterbrechungsfreien Bearbeitung, dass jedes einzelne Teil auch sofort weitergegeben und unmittelbar weiter bearbeitet werden muss. Die triviale Lösung ist, die Losgröße auf ein Stück zu setzen. Wird der Begriff der Unterbrechungsfreiheit auf Lose angewendet, die größer als ein Teil sind, warten zwangsläufig die zuerst fertiggestellten Teile des Loses so lange auf ihre Bearbeitung auf der nächsten Maschine, wie die Fertigung der übrigen Teile des Loses noch andauert. Dieser Widerspruch löst sich, wenn man annimmt, dass die Weiterbearbeitung nicht unmittelbar einsetzen muss, sondern innerhalb eines Zeitfensters nach der Fertigstellung zu beginnen hat. Wenn beispielsweise die Bearbeitung eines Stücks auf einer Maschine 3 Minuten dauert und das Werkstück dabei auf 250° Celsius erhitzt wird, die Weiterbearbeitung auf der nächsten Maschine 5 Minuten in Anspruch nimmt und erfolgen muss, bevor das Stück auf unter 100° ausgekühlt ist, ergibt sich z.B. bei einer Auskühlung um 3° pro Minute, dass spätestens nach 50 Minuten die Weiterbearbeitung zu erfolgen hat. Die weiterzugebende Losgröße ist durch die Produktionsgeschwindigkeit der zweiten Maschine und durch die Auskühlrate auf maximal 10 Stück beschränkt, da sonst die Stücke des Loses, die zuletzt auf der ersten Maschine gefertigt wurden, mehr als 50 Minuten warten müssten. Dieses Beispiel zeigt auch, dass es im Fall der unterbrechungsfreien Bearbeitung notwendig sein kann, die Teile innerhalb eines (Teil-)Loses immer in genau der gleichen Reihenfolge zu bearbeiten

Weiterhin können Transportrestriktionen vorliegen, die den Transfer einzelner Stücke aus organisatorischen oder aus wirtschaftlichen Gründen ausschließen. Ebenso können aus den verfügbaren Transporteinrichtungen Bedingungen für die mindestens beziehungsweise höchstens gemeinsam weiterzugebende Menge folgen, womit sich der Entscheidungsspielraum für die Anpassung der Losgrößen entsprechend eingrenzt. Selbst wenn die „liefernde“ Maschine eine offene Weitergabe ermöglicht, sind Transportrestriktionen denkbar, die dies einschränken. Auf den Extremfall, dass allein das Transportsystem die geschlossene Weitergabe trotz offener Fertigung erzwingt, weist bereits Adam

(1969, S. 100) hin. Zu denken sind weiterhin an (vgl. Trietsch/Baker, 1993; Kalir, 1999):

- Vorgegebene Höchst- oder Mindestmengen je Transport.
- Einschränkungen hinsichtlich der Zahl der Transporte.
- Vorgegebene Mengeneinheiten (z.B. nur ganze Paletten).
- Vorgaben hinsichtlich der Relation der Transportmengen untereinander (z.B. ein gleichmäßiges Aufteilung eines Loses auf mehrere Transporte).
- Abhängigkeit oder Unabhängigkeit der Transportzeit von der transportierten Menge.
- Ressourcenverbund von Transport- und Produktionskapazität.

Mit der letztgenannten Variante wird der Fall abgedeckt, dass entweder die liefernde Stufe oder die empfangende Stufe den Transport durchführt und während dieser Zeitspanne die Produktion unterbrochen werden muss. Dieser Fall wird z.B. dann auftreten wenn bei einer Fertigung mit hohem manuellem Anteil der Mitarbeiter auch für den Transfer zur nächsten Stufe zuständig ist (Bring- resp. Hohlpflicht).

Schließlich können sich auch aus der Gestaltung und Dimension des Lagers Höchstmengen für die teilweise Anlieferung ergeben. Beispielsweise kann der an der Maschine vorhandene Platz so knapp bemessen sein, dass nicht alle Teile eines Loses dort auf ihre Bearbeitung warten können, sondern diese zunächst in ein Zwischenlager zu bringen sind.

4.2 Produktions-, Freigabe- und Transferlose

Versteht man als das wesentliche Charakteristikum eines Loses die nicht weiter unterteilte beziehungsweise unterbrochene Bearbeitung, hat die gerade geführte Diskussion einen Aspekt verdeutlicht: Die Frage, welche Größe das tatsächlich realisierte Los hat, ist nicht unmittelbar beantwortet, wenn mehrere Lose des gleichen Produkts auf einer Maschine so nacheinander eingelastet werden, dass keine zwischenzeitliche Leerzeit und kein Umrüsten erforderlich wird. A priori liegen mehrere kleine Lose vor, a posteriori hat sich aus den kleineren Losen eine größere Einheit gebildet, die ebenso als ein Los angesehen werden kann. Insbesondere wegen der Möglichkeit einer überlappenden Weitergabe ist es sinnvoll, in Anlehnung an Jacobs/Bragg (1988) und Lockwood et al. (2000), die folgenden drei Arten der Aggregatbildung zu unterscheiden, die im Weiteren als Lostypen bezeichnet werden. Alle drei Lostypen werden in der Zahl der Stücke gemessen, die sie umfassen. Alle Losgrößen sind auf Endprodukteinheiten umgerechnet.

Definition: *Lostypen*

Transferlos: Die Menge an Gütern, die gemeinsam zwischen den Maschinen beziehungsweise zwischen den Lagern und den Maschinen transportiert oder auf der letzten Stufe an den Kunden geliefert wird.

Freigabelos: Die Menge an Gütern die in einem Zeitpunkt gemeinsam für die Bearbeitung auf einer Maschine freigegeben wird.

Produktionslos: Die Menge an Gütern, die ohne Unterbrechung und ohne (beziehungsweise nur mit vernachlässigbarem) Umrüsten gefertigt wird.

Generell ist es möglich, dass sich ein Produktionslos aus mehreren Freigabelosen zusammensetzt. Dies geschieht, wenn die Freigabe zur Produktion derart vorgenommen wird, dass es zu keiner Unterbrechung und zu keinen zwischenzeitlichen Rüstvorgängen kommt. Ebenso kann ein Freigabelos unmittelbar einem Transferlos entsprechen, es kann sich aber auch aus mehreren Transferlosen zusammensetzen oder nur Teile eines Transferloses umfassen. In den letzten beiden Fällen kommt es im Zwischen- beziehungsweise im Eingangslager der Maschine zu einer veränderten Zusammenfassung beziehungsweise Aufspaltung, da die bisher als Einheit aufgetretenen Mengen neu zusammengestellt werden. Dieser Vorgang ist z.B. von Bedeutung, wenn im Mehrproduktfall stufenweise Familien zur Reduktion des Rüstaufwands gebildet werden. Als dritte Größe wird das Freigabelos eingeführt, da die bei einer Maschine bereitstehenden Güter oft nicht sofort und auch nicht in ihrer Gänze eingelastet werden. Die folgende Abbildung verdeutlicht den Zusammenhang

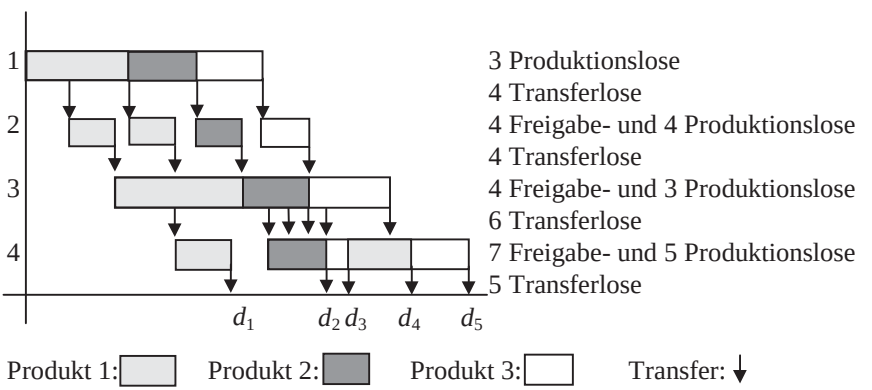


Abbildung 5: Unterscheidung in Transfer-, Freigabe- und Produktionslose

In Abbildung 5 wird die Situation eines vierstufigen Flow Shops unterstellt, auf dem drei Produkte gefertigt werden. Es wird angenommen, dass für die Zeitpunkte d_2, ..., d_5 Bestellungen durch Kunden vorliegen und eine Einlastung gesucht wird, die dieser Anforderung genügt. Von Rüstvorgängen wird abstra-

hiert und angenommen, dass der Transfer zwischen den Stufen keine Zeit in Anspruch nimmt, dass die Lager nicht kapazitiert sind und die Bereitstellung vor der ersten Stufe immer gewährleistet ist.

Auf Stufe 1 wird in drei Produktionslosen gefertigt. Die Weitergabe an Stufe 2, auf der mit größerer Produktionsgeschwindigkeit gearbeitet wird, muss in vier Transferlosen erfolgen, damit auf Stufe 3 die Fertigung derart freigegeben werden kann, dass nach der Bearbeitung auf Stufe 4 die vorgegebenen Termine eingehalten werden. Auf Stufe 1 und 2 genügt zur Umsetzung des Plans Losverfügbarkeit, auf Stufe 3 wird bei der Weitergabe des Loses des Produkts 3 Stückverfügbarkeit notwendig.

Um auf die Unterschiede zwischen den drei Arten von Losen einzugehen, ist zunächst auf die Aufteilung des ersten Produktionsloses für Produkt 1 auf Stufe 1 in zwei Transferlose hinzuweisen. Auf Stufe 3 ergibt sich durch die lückenlose Einlastung der beiden Freigabelose des Produkts 1 ein Produktionslos für Produkt 1 und auf Stufe 4 vereinigen sich drei Freigabelose für Produkt 2 zu einem Produktionslos des Produkts 2. Im zweiten, dritten oder vierten Transferlos zur Stufe 4 wird sowohl Produkt 1 als auch Produkt 2 befördert. Auf Stufe 4 werden die Produkte 1 und 3 zweimal aufgelegt, um den Lieferterminen zu genügen. Damit liegt für das Produktionslos des Produkts 3 im Übergang zur Stufe 4 eine Losteilung vor, da Teile dieses Loses auf Stufe 4 erst nach der Fertigstellung des Produktionsloses für Produkt 1 bearbeitet werden. Durch die Unterscheidung in die drei Losarten wird somit auch der Fall beschreibbar, dass ein Freigabelos, z.B. durch eine maschinelle Störung, nicht ununterbrochen bearbeitet wird, sondern sich daraus zwei Produktionslose ergeben. Dabei ist zu beachten, dass in ein Freigabelos immer nur die Teile einfließen können, die in dem Moment der Freigabe tatsächlich verfügbar sind.

Mit Hilfe des Beispiels ist deutlich geworden, dass das Produktionslos lediglich die Losgröße bezeichnet, die sich tatsächlich realisiert; es wird somit zu einer residualen Größe. Insbesondere muss keine Identität zwischen Transferlosen und Freigabelosen eines Produktionsloses bestehen. Entscheidungsvariable sind die Größen der Transferlose und die der Freigabelose mit den zugehörigen Zeitpunkten.

4.3 Andere Techniken zur Anpassung gegebener Losgrößen

Die drei Techniken zur Anpassung gegebener Losgrößen in der Ablaufplanung: Lossplittung, Losteilung und Losüberlappung haben gemeinsam, dass sie auf der zuvor geplanten Losgröße aufsetzen und kurzfristige Reaktionen ermöglichen. Sie agieren ‚ex post' bezüglich der Entscheidung, die die vorgelagerte Planungsebene unter Kostengesichtspunkten hinsichtlich der Losgrößen

getroffen hat (Bukchin et al., 2002). Die Techniken unterscheiden sich hinsichtlich des Loszusammenhalts, der so genannten Losintegrität. Mit ihr wird untersucht, ob sich die Veränderung einer gegebenen Losgröße nur als vorübergehend oder als endgültig erweist (Steele, 1995; Kher et al., 2000).

Die *Lossplittung* (z.B. Cheng/Sin, 1990; Fogarty et al., 1991; Baker, 1998) ist als Aufteilung eines gegebenen Loses zur gleichzeitigen Bearbeitung auf mehreren parallelen Maschinen – mit u.U. entsprechend hohem Rüstaufwand – zu verstehen. Die Aufteilung hat nur einen temporären Charakter. Insbesondere wird die Information über die Loszugehörigkeit der Teile nicht aufgegeben, da eine anschließende erneute Zusammenfassung zu einer Einheit – trotz der unterschiedlichen Produktionslose – intendiert ist. Der Vorgang der Lossplittung wird in der weiteren Untersuchung aus folgenden Gründen nicht betrachtet:

- Viele PPS-Systeme unterstützen die Lossplittung bereits dispositiv (vgl. François/Wolf, 2001), so dass hier keine drängende Forschungslücke wie bei der Losüberlappung besteht.
- Oft werden funktionsgleiche beziehungsweise einander unmittelbar ergänzende Maschinen zu so genannten „Production Units“ aggregiert (Missbauer, 1998). Ein Los wird bezogen auf diese Einheiten vorgegeben und abgearbeitet. Innerhalb einer Production Unit kann es gesplittet werden. Dabei ist hinsichtlich der parallel installierten Maschinen nicht davon auszugehen, dass sie identisch sind. Häufig sind zwar funktionsgleiche, dafür aber unterschiedlich alte Maschinen installiert. Der jeweilige technische Entwicklungsstand führt zu abweichender Abnutzung, zu Unterschieden hinsichtlich der Ausschussraten und Rüstzeiten, zu unterschiedlicher Produktionsgeschwindigkeit etc. Es ist daher sinnvoll, die Aufteilung auf dieser maschinennahen Entscheidungsebene den Meistern zu übertragen. Die Unit meldet ihre Kapazität – bei deren Bestimmung die Erfahrungswerte über den Effekt des internen Splittens eingehen – an die losgrößen-planende Ebene. Diese kann entscheiden, ob zwischen den Production Units überlappend gefertigt wird oder ob z.B. vorgegebene Lose geteilt und separat abgearbeitet werden sollen. Um hervorzuheben, dass es hier nicht um die Zuordnung von Aufträgen zu einzelnen Maschinen sondern um die stufenüberlappende Weitergabe geht, wird im Weiteren von Stufen statt von Maschinen gesprochen.

Im Fall der *Losteilung* geht der ursprüngliche Loszusammenhalt verloren, da das gegebene Los in neue Lose geteilt wird. Die entstandenen Lose können und sollen separat bearbeitet und weitergegeben werden, ohne dass eine spätere erneute Zusammenfassung zu der alten Einheit a priori vorgesehen ist. Die Losteilung wird oft im Rahmen der mittelfristigen Planung zur Produktionsglättung (Kistner/Steven, 2001, S. 94 ff.) eingesetzt. Ausgangspunkt ist dabei die stark variierende Belastung der Kapazität über die Perioden. Unter Umstän-

den reicht die vorgesehene Kapazität in einigen Perioden nicht aus, um die eingeplanten Mengen zu fertigen, während in anderen Perioden eine Unterauslastung festzustellen ist. Es bietet sich an, die bestehenden Lose aufzuteilen und die Mengen so umzulegen, dass eine gleichmäßige und zulässige Auslastung erreicht wird (Kurbel, 2003, S. 155 ff.). Der ursprünglich geplante Loszusammenhalt geht bei Anwendung dieser Maßnahme vollständig verloren. Auch diese Variante wird hier nicht weiter betrachtet, da die mit der Losteilung neu entstehenden Lose wiederum den Ausgangpunkt für die kurzfristig auf der operativen Ebene anwendbare Losüberlappung bilden können. Außerdem führt die Losteilung dazu, dass sich die Lose in anderer Größe realisieren, als in der Losgrößenplanung vorgesehen. Diese größeren Abweichungen von der Vorgabe führen tendenziell zu umfassenderen Rückkopplungen, wohingegen die Losüberlappung dazu beitragen kann, dass Vorgaben wie geplant umzusetzen sind.

Mit der Unterscheidung in Freigabe-, Produktions- und Transferlose kann zudem das *Preemption* (vgl. z.B. Brucker et al., 2003) abgebildet und in die Diskussion der übrigen Techniken subsumiert werden. Unter Preemption versteht man in der Ablaufplanung die Unterbrechung der Bearbeitung eines Auftrags, um einen anderen Auftrag vorzuziehen. Der teilweise fertiggestellte Auftrag wird von der Maschine genommen und ein anderer Auftrag (Job) wird diesem vorgezogen, wobei der unterbrochene Auftrag an der Maschine wartet. Eine teilweise Weitergabe an die nächste Maschine findet nicht statt, so dass das Preemption nicht der Überlappung entspricht. Die Bearbeitung des unterbrochenen Auftrags wird nach der Fertigstellung des eingeschobenen Auftrags fortgesetzt, d.h. der ursprüngliche Auftrag bleibt als solcher erhalten. Das Preemption kann damit als eine nachträgliche, degenerierte Losteilung aufgefasst werden. Darunter soll verstanden werden, dass ein freigegebenes Los, dessen Bearbeitung bereits begonnen hat, durch eine weitere Freigabe eines Loses eines anderen Produkts unterbrochen wird. Damit realisiert sich ein kleineres Produktionslos, als zunächst erwartet wurde. Die bereits fertiggestellten Stücke verbleiben aber an der Maschine und werden erst dann gemeinsam weitergegeben, wenn auch das übrige Los bearbeitet wurde, damit der für das Preemption charakteristische Zusammenhalt als Einheit nicht verloren geht. Die Frage, ob diese Unterbrechung ex ante vorgegeben ist oder sich erst während der Bearbeitung ergibt, ist von nachgelagertem Interesse, so dass hier eine separate Behandlung des Preemption ausgeklammert werden kann. Für die weitere Untersuchung wird zudem angenommen, dass alle Teile eines Loses entweder identisch sind oder nur vernachlässigbar geringe Unterschiede z.B. hinsichtlich des Rüstaufwandes auftreten. Eine Diskussion bezüglich der Planung und Aufteilung von „*batches*“ (Potts/Van Wassenhove, 1992) wird somit ebenfalls nicht vorgenommen.

5. Verfahren der Losüberlappung

5.1 Begriffe und Beziehungen

Zur Verdeutlichung der Vorgehensweise bei der Losüberlappung werden vier verschiedene Losarten vorgestellt und typische Eigenschaften von Lösungen an Hand kleiner Beispiele exemplarisch gezeigt. Dies geschieht, um die zwischen den Lösungen bestehende hierarchische Beziehung herzuleiten, die bei der Entwicklung beziehungsweise der Auswahl von Verfahren nützlich ist. Den Ausgangspunkt bildet das Beispiel der bereits im ersten Kapitel vorgestellten dreistufigen Reihenfertigung eines Produktes. Auf Stufe 1 werden zur Produktion je Stück 0,7 ZE, auf der zweiten Stufe 1,0 ZE und auf Stufe 3 0,6 ZE benötigt. Weiterhin wird davon ausgegangen, dass die Dauer der jeweiligen Transporte vernachlässigbar klein ist. Für ein geschlossen weitergegebenes Los mit 30 Teilen beträgt die Durchlaufzeit 69 ZE, da es auf der ersten Stufe 21 ZE, auf der zweiten Stufe 30 ZE und auf der dritten Stufe 18 ZE lang bearbeitet wird.

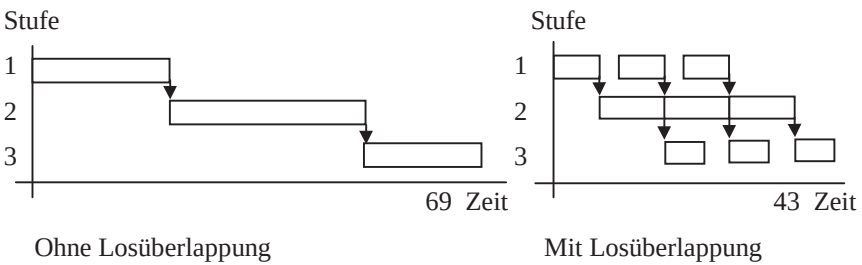


Abbildung 6: Geschlossene Weitergabe versus Weitergabe in gleichen Losen

Wird das Los nicht geschlossen in einem Transfer weitergegeben, sondern in drei gleichgroßen Losen mit je 10 Teilen, kann die Durchlaufzeit um fast 40% auf 43 ZE gesenkt werden. Die Entscheidung, welche Losart verwendet werden soll, kann sich prinzipiell auf alle drei Lostypen (Transfer-, Freigabe- und Produktionslose) beziehen. Bislang wird dieser Aspekt allerdings nicht berücksichtigt. Entweder tritt die Größe der Freigabelose (sublots) in den Vordergrund (z.B. bei Potts/Baker, 1989; Trietsch/Baker, 1993; Baker/Jia, 1993; Glass et al., 1994; Baker, 1998; Kumar et al., 2000), oder es wird auf die Transferlose (transfer lots) Bezug genommen (z.B. bei Çetınkaya, 1994; Topaloğlu et al., 1994; Benli, 1997).

5.1.1 Losarten

In Erweiterung der bislang üblichen Dreiteilung werden hier neben den variablen, konstanten und konsistenten Losen (vgl. Baker, 1998) die stufenweise konstanten Lose eingeführt und somit vier Arten von Losen unterschieden.

Definition: *Konstante Lose*

Konstante Lose liegen vor, wenn nur gleich große Lose vorgesehen sind.

Bei der Verwendung konstanter Lose bleibt die Losintegrität, also die Zugehörigkeit eines Teils zu einem Los, über alle Stufen erhalten. Außerdem ist die organisatorische Regelung zum Erreichen eines Plans mit konstanten Losen besonders einfach umzusetzen. Sollen nur konstante Lose verwendet werden, reicht Losverfügbarkeit aus. Allerdings bedeutet die Vorgabe, nur nach Lösungen mit konstanten Losen zu suchen, eine erhebliche Vereinfachung. Betrachtet man zudem die in Abbildung 6 rechts angegebene Situation, fallen einige Besonderheiten auf:

- Statt eines Produktionsloses sind auf der ersten und dritten Stufe drei Produktionslose vorgesehen, wobei für jedes der drei Lose Wartezeitfreiheit realisiert ist.
- Die wartezeitfreie Bearbeitung der Lose kann auf Grund der von Stufe zu Stufe unterschiedlich hohen Fertigungsgeschwindigkeit nur erreicht werden, wenn auf den Stufen 1 und 3 Leerzeit auftritt. Die gezeigte Lösung bietet sich an, wenn kein oder nur ein vernachlässigbarer Rüstaufwand zwischen den drei Produktionslosen auf Stufe 1 und auf Stufe 3 anfällt.
- Der zweite und der dritte Transport zur Stufe 2 liegt zeitlich parallel mit dem ersten und zweiten Transport zur dritten Stufe. Nimmt man an, dass z.B. nur ein Gabelstapler als Transportmittel verfügbar ist, kann diese Lösung nicht umgesetzt werden.

Lässt man unter der Annahme der drei gleichgroßen Lose die wartezeitfreie Bearbeitung fallen, kann die Problemstellung wie in Abbildung 7 gelöst werden.

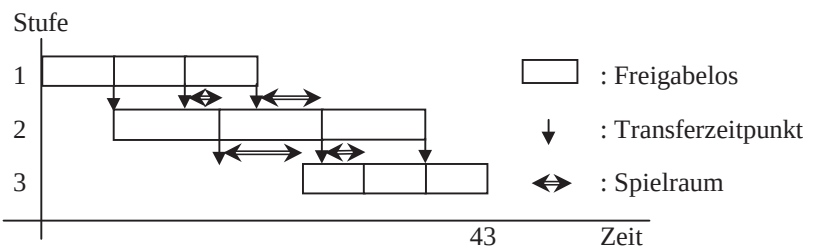


Abbildung 7: Konstante Lose ohne Leerzeiten

Dort liegt pro Stufe nur ein Produktionslos vor, das sich jeweils aus 3 Freigabelosen zusammensetzt, die wiederum den 3 gleich großen Transferlosen entsprechen. Teile des zweiten und dritten Freigabeloses gehen bei Stufe 2 ein, bevor sie dort benötigt werden. Der gleiche Sachverhalt zeigt sich auf Stufe 3. Alternativ kann z.B. das zweiten Transferloses von der Stufe 1 zurückgehalten werden, bis die Teile auf Stufe 2 benötigt werden.

Es entstehen damit Entscheidungsspielräume, die erstens den Zeitpunkt des Transfers und zweitens die Menge der transferierten Stücke und damit drittens den Ort der Zwischenlagerung betreffen. Um diese in der Abbildung 7 aufzuzeigen, wurden die horizontalen Doppelpfeile eingefügt. Jeder vertikal verlaufende Pfeil gibt den frühesten Zeitpunkt an, in dem ein Transport zur nächsten Stufe durchgeführt werden kann. Der horizontale Doppelpfeil endet in dem Zeitpunkt, in dem der Transport spätestens durchzuführen ist. Wird in der gezeigten Lösung Stückverfügbarkeit zugelassen, d.h. auch Teile eines noch nicht zu Ende bearbeiteten Freigabeloses dürfen transferiert werden, können die Größen der Transfer- und der Freigabelose voneinander abweichen. Dieser Fall tritt z.B. dann auf, wenn der zeitliche Spielraum zum Transfer des zweiten Loses von der ersten Stufe zur zweiten genutzt wird und in diesen Transfer die mittlerweile fertiggestellten Stückes des dritten Freigabeloses der ersten Stufe eingehen. In diesem Fall werden zwar konstante Freigabelose genutzt, die Transferlosgrößen weichen aber von diesen ab. Durch derartige Verlagerungen können z.B. beschränkte Lagerkapazitäten oder die Restriktionen des Transportsystems berücksichtigt werden.

Werden mehr als die gezeigten drei Transporte zwischen den Stufen eingesetzt, reduziert sich die maximale Durchlaufzeit. Im Extremfall wird die Zahl der Transporte extrem erhöht und die Transferlosgröße jeweils auf nur ein Stück gesetzt. Mit dieser Vorgabe kann immer die kleinstmögliche Durchlaufzeit – unter der Voraussetzung ganzzahliger Losgrößen – realisiert werden. Im vorliegenden Beispiel beträgt sie 31,3 ZE und ergibt sich wie folgt: Das erste Stück benötigt auf der ersten Stufe 0,7 ZE und wird danach sofort an die zweite Stufe weitergegeben. Die zweite und langsamste Stufe benötigt 1 ZE pro Stück. Die zweite Stufe wird, da sie offensichtlich den Engpass darstellt, ohne Unterbrechungen genutzt und ist somit 30 ZE lang, also bis zum Zeitpunkt 30,7 belegt. Die auf Stufe 2 gefertigten Stücke können direkt zur weiteren Bearbeitung an Stufe 3 weitergegeben werden. Da diese schneller arbeitet als die zweite Stufe (sie benötigt nur 0,6 ZE statt 1,0 ZE), ist sie immer frei, wenn wieder ein Stück von der Stufe 2 angeliefert wird. Das Schema setzt sich fort, bis das letzte Stück die Stufe 2 verlässt, was im Zeitpunkt 30,7 geschieht. Da die Bearbeitung des letzten Stückes auf Stufe 3 0,6 ZE dauert, ergibt sich als minimale Durchlaufzeit für die Produktion der 30 Stück die bereits genannten 31,3 ZE. Wie schon in dem zuvor betrachteten Beispiel kann durch Vorverlagern beziehungsweise durch spätere Freigabe der Bearbeitung die Leerzeitfreiheit pro Stu-

fe realisiert werden. Da sich mit der Verwendung von 30 konstanten Losen die Zahl der insgesamt notwendigen Transporte auf 60 erhöht, bietet es sich an, Lose zusammenzufassen und Transporte zu sparen. Dabei können stufenweise konstante Lose entstehen.

Definition: *Stufenweise konstante Lose*

> Stufenweise konstante Lose liegen vor, wenn die Lose auf jeder Stufe gleich groß sind.

Im Gegensatz zu den konstanten Losen ist es bei stufenweise konstanten Losen zulässig, dass die Zahl der Lose über die Stufen hinweg variiert, womit die Losintegrität verloren geht. Allerdings bleibt auf jeder Stufe die Größe der Freigabelose gleich. Um Missverständnissen vorzubeugen, werden im Folgenden die Transferlose immer der Stufe zugerechnet, auf der sie gebildet werden. Die nachfolgende Abbildung 8 zeigt auf der linken Seite einen Plan, bei dem mit stufenweise konstanten Transferlosen gearbeitet wird. Zudem wird dort angenommen, dass Stückverfügbarkeit auf der Stufe 2 gegeben ist, so dass aus den laufenden vier Freigabelosen der Größe 7,5 drei Transferlose der Größe 10,0 gebildet werden können.

Vergleicht man die Lösung mit der, die man für drei konstante Lose (Abbildung 8 rechte Seite) erhält, wird deutlich, wie mit einer Erhöhung der Zahl der insgesamt benötigten Transporte von 6 auf 7 die Durchlaufzeit um fast 4% von 43 ZE auf 41,25 ZE gesenkt werden kann. Es bedarf dazu nur einer geringfügigen Ergänzung der organisatorischen Regel: Anstatt über alle Stufen hinweg immer nach 10 Teilen ein Los zu bilden, wird hier der zweiten Stufe vorgegeben, jeweils nach der Fertigstellung von 7,5 Einheiten ein Los an die nächste Stufe weiterzugeben. Eine Lösung mit stufenweise konstanten Losen zeichnet sich dadurch aus, dass es pro Stufe nur eine über die Zeit hinweg unveränderte Arbeits- beziehungsweise Aufteilungsregel gibt, die aber den Vorgaben der anderen Stufen nicht entsprechen muss. Die in Abbildung 8 gezeigte Durchlaufzeit ergibt sich aus: $0{,}7 \cdot 7{,}5 + 1 \cdot 30 + 0{,}6 \cdot 10 = 41{,}25$ ZE und wurde durch Schraffierung hervorgehoben.

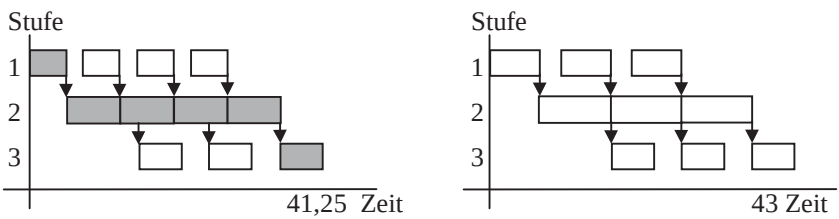


Abbildung 8: Stufenweise konstante versus konstante Lose

Aus Sicht der unmittelbar in den Produktionsprozess eingebundenen Mitarbeiter hat die Verwendung von stufenweise konstanten Losen den Vorteil, dass

bestimmte Einstellungen an den Maschinen unverändert bleiben können und dennoch eine deutliche Reduktion der Durchlaufzeit mit einer vergleichsweise geringen Zahl zusätzlicher Transporte realisiert wird. Stufenweise konstante Lose können daher in Reaktion auf knappe Transportmittel gewählt werden. Ausgehend von einer Lösung mit konstanten Losen sollten die zusätzlichen Transporte so eingeplant werden, dass die Produktion auf dem Engpass möglichst früh einsetzen kann.

Während zur Charakterisierung der beiden bisher vorgestellten Losarten in erster Linie horizontal die Größe der Loses auf einer Stufe oder im Vergleich von ganzen Stufen untereinander herangezogen wurde, wird jetzt die Blickrichtung geändert und jedes Los vertikal über die Stufen hinweg begleitet. Diese Perspektive führt zu der Unterscheidung in die konsistenten und die variablen Lose.

Definition: *Konsistente Lose*

> Konsistente Lose liegen vor, wenn kein Los über die Stufen hinweg seine Größe ändert.

Konsistente Lose haben die wichtige Eigenschaft, dass die Zugehörigkeit eines Stückes zu seinem Los erhalten bleibt. Auf den Erhalt der Losintegrität zielt dementsprechend auch das in der Literatur wiederholt vorgebrachte Argument ab, eine Verwendung konsistenter Lose sei leicht zu implementieren und vereinfache die Koordination (vgl. Potts/Baker, 1989; Smunt et al., 1996; Chen/Steiner, 1997). Aus Sicht der übergeordneten Produktionsplanung bedeutet die Einschränkung auf konsistente Lose unbestritten eine Vereinfachung, da Losverfügbarkeit, nicht aber Stückverfügbarkeit verlangt und im Entscheidungsmodell erfasst werden muss. Die ermittelten Lose verändern über die Stufen ihre Größe nicht, was die Koordination und das Lot Tracing vereinfachen. Allerdings muss der Mitarbeiter auf einer Stufe u.U. eine Abfolge unterschiedlich großer Lose zur Produktion freigeben. Treffen konsistente Freigabelose und Losverfügbarkeit aufeinander, dann gestaltet sich der Ablauf insofern einfach, als dass ein zu der Stufe transferiertes Los auch unverändert zur Produktion freizugeben ist.

Wird in dem Beispiel mit 3 konsistenten und zudem ganzzahligen Losen die Durchlaufzeit minimiert, ergibt sich die in der folgenden Abbildung wiedergegebene Lösung. Es werden 9 Stück im ersten Transferlos, 13 Stück im zweiten und 8 Stück im dritten Transferlos weitergegeben. Die damit realisierbare Durchlaufzeit beträgt 41,1 ZE. Lässt man die Ganzzahligkeit fallen, haben die 3 Lose (auf 2 Nachkommastellen gerundet) die Größe 9,13; 13,04 und 7,83. Mit dieser Aufteilung sinkt die Durchlaufzeit auf 41,08 ZE = 9,13·0,7 ZE + 30·1 ZE + 7,83·0,6 ZE.

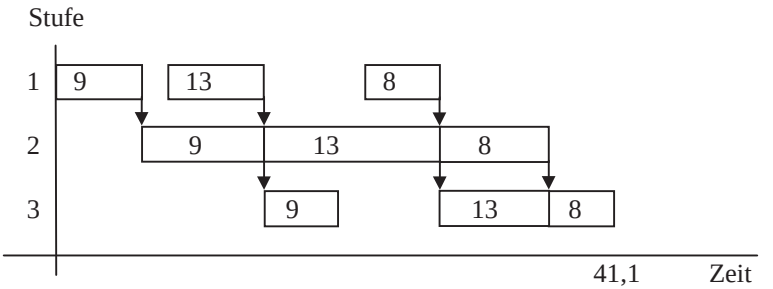


Abbildung 9: Ganzzahlige konsistente Lose mit Leerzeiten

Da konstante Lose auch konsistent sind, bilden die Pläne, die für konstante Losgrößen ermittelt werden, eine Untermenge der Pläne, für die konsistente Lose verlangt werden. Wie im Fall der konstanten Lose kann sich bei Losverfügbarkeit die Zahl der Transporte zwischen den Stufen nicht ändern. Im Vergleich der konsistenten zu den stufenweise konstanten Losen fällt auf, dass letztere keine Untermenge der Pläne mit konsistenten Losen darstellen. Lässt man schließlich alle Beschränkungen hinsichtlich der Verhältnisse der Losgrößen untereinander fallen, liegen variable Lose vor.

Definition: *Variable Lose*

> Variable Lose liegen vor, wenn keinerlei Restriktionen hinsichtlich der Größe der Lose gegeben sind.

Derartige Pläne zeichnen sich dadurch aus, dass es weder über die Zeit hinweg unveränderte Arbeits- beziehungsweise Aufteilungsregel pro Stufe noch Entsprechungen der Regeln zwischen den Stufen geben muss. Insbesondere ist es für einzelne Lose zulässig, dass ihre Größe auf Null gesetzt wird und sie damit ganz entfallen. Die Zahl der Transfers variiert in diesem Fall von Stufe zu Stufe. Wird nach einer Lösung unter der Verwendung von variablen Losen gesucht, ist die Vorgabe einer festen Zahl von Transporten zwischen den Stufen nur als Obergrenze sinnvoll. Wird hingegen unter der Zielsetzung der Minimierung der Durchlaufzeit ein Plan mit konstanten oder stufenweise konstanten Losen gesucht, kann sich für keines der Lose die Größe Null ergeben, da in diesem Fall nicht produziert wird. Während bei konstanten und konsistenten Losen die Losintegrität erhalten bleibt, ist dies bei den anderen beiden Losarten, den stufenweise konstanten und den variablen Losen, nicht mehr der Fall.

Eine Lösung mit variablen Losen unter Stückverfügbarkeit, die zum Erreichen der minimalen Durchlaufzeit von 31,3 ZE führt, benötigt aber nicht unbedingt 30 Transferlose mit der Größe 1 und damit 60 Transporte. Die folgenden Tabellen zeigen, dass insgesamt nur 16 Transporte in dafür aber unterschiedlich großen ganzzahligen Transferlosen notwendig sind, um die Durchlaufzeit auf 31,3 ZE zu senken.

Tabelle 2: **Transferlosgrößen zwischen der Stufe 1 und der Stufe 2**

Tranferlos-Nr.	1	2	3	4	5	6	7	8	9
Größe	1	1	1	2	3	4	6	8	4
frühester Transferzeitpunkt	0,7	1,4	2,1	3,5	5,6	8,4	12,6	18,2	21
spätester Transferzeitpunkt	0,7	1,7	2,7	3,7	5,7	8,7	12,7	18,7	26,7

Tabelle 3: **Transferlosgrößen zwischen der Stufe 2 und der Stufe 3**

Tranferlos-Nr.	1	2	3	4	5	6	7
Größe	9	9	5	3	2	1	1
frühester Transferzeitpunkt	9,7	18,7	23,7	26,7	28,7	29,7	30,7
spätester Transferzeitpunkt	13,3	18,7	24,1	27,1	28,9	30,1	30,7

Sowohl die Zahl der Lose als auch die Größe der Lose auf den Stufen muss für diese Lösung variieren. Sind es zwischen Stufe 1 und Stufe 2 insgesamt neun Transporte, werden zwischen Stufe 2 und 3 nur sieben benötigt. Entsprechend müssen die Stücke in unterschiedlicher Konstellation zu Transferlosen und später zu Freigabelose zusammengefasst werden. Beispielsweise setzt sich das erste Transferlos zwischen der zweiten und der dritten Stufe nicht nur aus den ersten fünf Transferlosen zwischen Stufe 1 und 2 zusammen, sondern es umfasst zusätzlich noch das erste Stück aus dem sechsten Transferlos zwischen diesen Stufen. Die gezeigte Lösung erfordert somit, dass Stückverfügbarkeit gegeben ist, da aus den noch laufenden Losen weiterzugeben ist (z.B. beim zweiten Transferlos der Stufe 2).

Setzt man statt der variablen Lose die 16 Transporte in Kombination mit ganzzahligen konsistenten Losen (also für 8 Transfer- respektive Freigabelose) ein, erhält man die folgende optimale Sequenz von Losgrößen: 2, 3, 4, 6, 7, 4, 2, 2, mit denen eine Lösung in 32,7 ZE möglich wird. Wählt man hingegen 8 konstante Lose, erhält man als Durchlaufzeit 34,875 ZE. Die Differenz von 32,7 ZE (mit konsistenten Lose) beziehungsweise 34,875 ZE (mit konstanten Lose) zu 31,3 ZE (mit variablen Lose) beträgt 1,4 ZE beziehungsweise 3,75 ZE. Da die Zahl der Transporte in jeder Lösung identisch ist, muss die Reduktion der Durchlaufzeit dem erhöhten Planungs- und Koordinationsaufwand sowie der (für die Lösung mit variablen Losen notwendigen) Sicherstellung der Stückverfügbarkeit gegenübergestellt werden.

Deutlich wurde mit diesem Beispiel, dass der resultierende Zuwachs an organisatorischem Aufwand erheblich ist, wenn variable Lose eingesetzt werden.

Als Nebeneffekt wurde erneut die Zweckmäßigkeit der hier genutzten Unterscheidung hervorgehoben: Transfer-, Freigabe- und Produktionslose fallen teilweise auseinander, teilweise entsprechen sie sich.

Der Zusammenhang zwischen den zusätzlich eingesetzten Transporten, der Zunahme an Koordinations- und Kontrollaufwand und der damit erreichten Reduktion der Durchlaufzeit ist hier bislang nur exemplarisch gezeigt worden. Dennoch wurde bereits deutlich, dass offensichtlich vielfache Substitutionsbeziehungen im Kontext der Losüberlappung bestehen. Dabei sind ertragsgesetzliche Verläufe z.B. hinsichtlich der Erhöhung der Zahl an Transporten und der damit bewirkten Reduktion der Durchlaufzeit zu erwarten. Im Folgenden werden die Dominanzbeziehungen zwischen den Problemstellungen der Losüberlappung behandelt, da sie den Untersuchungsgegenstand sinnvoll strukturieren.

5.1.2 Dominanzbeziehungen

Die Restriktion, variable Lose zu verwenden, schränkt den Lösungsraum am wenigsten ein. Daher bilden alle Pläne mit konstanten, konsistenten und stufenweise konstanten Losen eine Teilmenge der Pläne, die mit der Vorgabe variabler Lose ermittelt werden können. Eine ähnlich gelagerte Überlegung gilt hinsichtlich der Forderungen nach Leerzeitfreiheit und Ganzzahligkeit der Lose. Die Problemstellungen der Losüberlappung und damit die für sie entwickelten Verfahren stehen in folgenden Dominanzbeziehungen zueinander:

(1) Existiert ein Verfahren, das Pläne generiert, die Leerzeiten auf den Stufen aufweisen dürfen, so kann das Verfahren auch Lösungen finden, in denen die Stufen ohne Unterbrechungen genutzt werden.

(2) Wird ein Verfahren entwickelt, das Pläne mit variablen Losen generiert, können sich auch Pläne ergeben, in denen stufenweise konstante, konstante oder konsistente Lose auftreten.

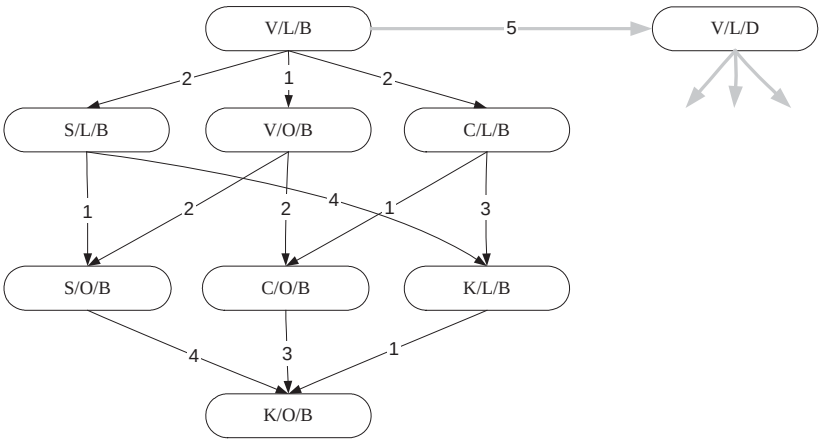
(3) Existiert ein Verfahren, das Pläne mit konsistenten Losen generiert, können sich auch Pläne ergeben, in denen konstante Lose auftreten.

(4) Wird ein Verfahren gefunden, das Pläne mit stufenweise konstanten Losen generiert, können sich auch Pläne ergeben, in denen konstante Lose auftreten.

(5) Existiert ein Verfahren, das Pläne mit beliebigen, kontinuierlichen Losgrößen generiert, können die optimalen Lösungen trotzdem ganzzahlig sein.

Die folgende Abbildung lehnt sich an die Darstellung bei Trietsch/Baker (1993, S. 1069) an und ergänzt diese um die stufenweise konstanten Lose. Die oben genannten fünf Dominanzbeziehungen werden genutzt, um die Relationen

zwischen den einzelnen Problemen der Losüberlappung besser nachvollziehen zu können.



Notation:

V / C / S / K	variabel / konsistent / stufenweise konstant / konstant
L / O	Leerzeit zulässig / Leerzeiten unzulässig
D / B	diskrete / beliebige Lose

Abbildung 10: Dominanzbeziehungen der Losüberlappung

In der Abbildung der hierarchischen Beziehungen wird jeweils die oben genannte Nummer der Dominanzbeziehung auf dem zugehörigen Pfeil vermerkt. Die oben in der Hierarchie angegebene Problemstellung V/L/B stellt die allgemeinste Fragestellung dar. Deshalb gehen die Pfeile in Abbildung 10 auch von ihr aus und münden nicht dort, wie es z.B. bei Trietsch/Baker (1993) der Fall ist. Wenn für einen Plan Leerzeiten zwischen den Bearbeitungen zulässig sind, schließt diese Gruppe von Plänen diejenigen ein, in denen Leerzeiten mit der Länge Null auftreten, d.h. die unterbrechungsfreien Pläne (O). Weiterhin können sich unter der Vorgabe von variablen Losen alle übrigen Formen der Aufteilung ergeben, also stufenweise konstante, konstante oder konsistente Lose. Da die Pläne mit konstanten Losen eine Untermenge der Pläne mit konsistenten Losen bilden, erklärt sich die übrige hierarchische Beziehung. Nicht in die Abbildung aufgenommen wurde der Ast der Losaufteilungen mit diskreten Wertebereichen, da dieser identisch zu dem für beliebige Losgrößen verläuft. Es ist aber festzuhalten, dass entsprechend der grau hervorgehobenen Beziehung zwischen der Problemstellung V/L/B und V/L/D auch für die übrigen Relationen zwischen den Lösungen mit beliebigen Losgrößen und denen mit diskreten Losgrößen bestehen. Der Lösungsraum für eine Fragestellung mit beliebigen Losen umfasst immer auch die Lösungen, in denen nur ganzzahlige Lose auf-

treten. Im Extremfall ist es möglich, dass durch die Forderung nach ganzzahligen Losen das gegebene Problem trivial wird. Soll z.B. eine Produktionsmenge von 100 Stück mit konstanten Losen realisiert werden, können als Losgrößen lediglich {100, 50, 25, 20, 10, 5, 4, 2 sowie 1} auftreten und das Problem kann enumerativ gelöst werden. Dass eine Einschränkung des Lösungsraums auf ganzzahlige Lösungen eine derartige Vereinfachung mit sich bringt, ist allerdings nicht typisch. Üblicherweise sind Fragestellungen unter Ganzzahligkeitsbedingungen erheblich schwerer zu lösen als ihre kontinuierlichen Entsprechungen (vgl. Kistner, 2003, S. 237). Wird bei der Modellentwicklung nur als vereinfachende Annahme zugelassen, dass beliebige Wertebereiche für die Lose ermittelt werden, sind diese Pläne regelmäßig nicht direkt umsetzbar, was im Rückgriff auf die von Little (1970) formulierten Anforderungen des *decicion calculus* kritisch zu beurteilen ist. Sieht man von Flüssigkeiten oder Schüttgütern ab, ist es in vielen Fällen notwendig, ganzzahlige Losgrößen zu bilden. Werden dazu Losüberlappungspläne mit beliebigen Losgrößen einfach auf die nächstliegenden ganzzahligen Werte abgerundet, müssen auch alle Freigabezeitpunkte angepasst und die durch Runden verlorenen Stücke an anderer Stelle hinzugefügt werden. Werden die Losgrößen hingegen aufgerundet, muss die Veränderung konsistent erfolgen, da sich sonst Unzulässigkeiten ergeben. Für die durch Runden entstandenen Pläne können außerdem nur in Ausnahmefällen Angaben über die Planungsgüte und über die Eigenschaften der Einlastungen (wie z.B. Leerzeitfreiheit oder Wartezeitfreiheit) hergeleitet werden (vgl. Chen/Steiner, 1997b). Daher hat auch die Entwicklung und Darstellung von Verfahren zur Ermittlung von ganzzahligen Losen eine hohe Bedeutung.

5.1.3 Weiterer Gang der Untersuchung

Die gesuchte Lösung wird im Folgenden Losüberlappungsplan genannt und ein konkreter Plan mit π bezeichnet. Dieser ist dann vollständig gegeben, wenn er die Größe der Lose (die der Transferlose und bei Abweichungen zwischen Transfer- und Freigabenlosen auch deren Größe) festschreibt und (falls erforderlich) die Zeitpunkte angibt, in denen die Lose transportiert oder zur Produktion freigegeben werden sollen. Um einen Plan π aufzustellen, müssen zunächst die Ausprägungen der Problemstellung bekannt sein, die aus der verwendeten Technologie oder aus generellen organisatorischen Vorgaben resultieren und die die vorzeitige Weitergabe oder die teilweise Einlastung einschränken können. Benötigt werden insbesondere Angaben hinsichtlich der

- Anzahl der Stufen,
- Produktionsgeschwindigkeit je Stufe,
- vorgegebene Losgröße,

- Wartezeitfreiheit,
- Leerzeitfreiheit,
- Verfügbarkeit der Teile (stück- oder losweise),
- Wertebereiche der Lose (ganzzahlig oder beliebig).

Außerdem müssen von den Entscheidungsträgern die folgenden Größen beziehungsweise Eigenschaften des gesuchten Plans vorgegeben werden:

- Zielkriterium (sofern kein anderes Kriterium benannt ist, wird im Weiteren von der Minimierung der maximalen Durchlaufzeit ausgegangen),
- Art der Lose: konstant, stufenweise konstant, konsistent oder variabel,
- Zahl der Transfer- beziehungsweise der Freigabelose,
- Anspruchsniveau: optimale oder heuristische Lösung.

Die folgenden Ausführungen sind in drei Schwerpunkte gegliedert. Es werden jeweils ausgewählte Fragestellungen betrachtet, wobei als wesentliche Auswahlkriterien die Aspekte: Praxisrelevanz, Rechenbarkeit und Erkenntnisbeitrag herangezogen wurden.

Im ersten Schwerpunkt (Kapitel 5.2) wendet sich die Untersuchung der Losüberlappung im Zwei- und Dreistufenfall zu. Die dort vorgestellten Verfahren für die verschiedenen Losarten sind vergleichsweise einfach umzusetzen, da sie mittels geschlossener Formeln oder einfacher Algorithmen darstellbar sind. Ihre Herleitung setzt aber voraus, dass entweder die Zahl der Stufen oder die Zahl der Lose restringiert ist. Bei der Konstruktion von Heuristiken sowie von exakten Verfahren für die Mehrstufenprobleme dienen diese Erkenntnisse als wertvolle Bausteine: Sie erlauben, Aussagen über generelle Eigenschaften optimaler Pläne zu formulieren und eröffnen damit wichtige Einsichten in die grundsätzliche Struktur optimaler Lösungen von Losüberlappungsproblemen.

Im zweiten Schwerpunkt (Kapitel 5.3) wird die Losüberlappung mit konsistenten Losen im Mehrstufenfall betrachtet. Die hohe Betonung, die den konsistenten Losen damit zukommt, ist aus mehreren Gründen geboten. Konsistente Lose stellen in vielerlei Hinsicht sinnvolle Kompromisse dar: Die Modelle sind vergleichsweise leicht lösbar, ihre Umsetzung ist organisatorisch wenig aufwändig und dennoch können die wesentlichen, mit Losüberlappung realisierbaren Effekte mit Hilfe der konsistenten Lose genutzt werden. Neben einem iterativen Verfahren für die Festlegung von zwei konsistenten Losen im Mehrstufenfall müssen bei größeren Problemen Ansätze der linearen Programmierung zum Einsatz kommen. Ihre Lösung – und wegen der Umfangs der Modelle auch ihre Formulierung – erfordert den Einsatz spezieller Software. Hier wird auf LINDO 6.1 und LINGO 7.0 zurückgegriffen. Mit dieser Software ist es möglich, optimale Pläne für konsistente Lose bei praxisrelevanter Problemgröße so schnell

zu ermitteln, dass What-If-Analysen möglich werden. Dabei bietet insbesondere die postoptimale Analyse einen leicht zugänglichen Ansatz, der weit über die bislang in der Literatur aufgezeigten Möglichkeiten zur Interpretation hinausreicht. Ergänzend wird untersucht, wie Problemdetails in die Modelle zu integrieren sind und ein Rescheduling im Kontext der Losüberlappung unterstützt werden kann.

Im dritten Schwerpunkt (Kapitel 5.4) werden ausgewählte Fragestellungen betrachtet, die entstehen, wenn variable Lose zugelassen werden. Vorgeschlagen wird ein lineares, binäres Modell, mit dem es erstmals möglich ist, Losüberlappungsprobleme mit variablen Losen unter Losverfügbarkeit bei kontinuierlicher Zeitführung optimal zu lösen. Analog zum zweiten Schwerpunkt wird die Modellierung mittels LINGO 7.0 vorgestellt. An Hand einiger Beispiele kann der Effekt der variablen Lose gegenüber den konsistenten Losen aufgezeigt werden. Dieser Ansatz schließt eine seit den ersten Arbeiten zum Lot Streaming (Potts/Baker, 1989) bestehende Forschungslücke. Er hat zudem den Vorteil, dass Erweiterungen, wie z.B. die Ganzzahligkeit der Lose, Rüst- und Transportzeiten, sowie andere Zielfunktionen leicht in das Modell eingebracht werden können.

Das fünfte Kapitel wird mit einem Abschnitt beendet (5.5) in dem ein Überblick über die restlichen Ansätze der Losüberlappung vorgenommen wird. Er umfasst Vorläufer der Losüberlappung und die Ansätze, die hinsichtlich der oben genannten Kriterien: Praxisrelevanz, Rechenbarkeit und Erkenntnisbeitrag nicht für eine unmittelbare Integration in die betriebliche Praxis geeignet erscheinen.

5.2 Losüberlappungsprobleme im Zwei- und Dreistufenfall

Die formale Analyse beginnt sinnvoller Weise mit einer Darstellung des Problems im Einproduktfall unter der Annahme der Losverfügbarkeit (Glass et al., 1994; Chen/Steiner, 1997). Dazu wird die folgende Notation benötigt:

Index des Freigabeloses:	j ($j = 1,..., J$)
Index der Stufen:	m ($m = 1,..., M$)
Vorgegebene Losgröße:	Q
Ausführungszeit je Stück auf Stufe m:	p_m
Größe des Loses j auf Stufe m:	x_{jm} für $1 \le m \le M$; $1 \le j \le J$
Fertigstellung des Loses j auf Stufe m:	C_{jm} für $1 \le m \le M$; $1 \le j \le J$
Letztes Freigabelos i das Teile enthält, die in Freigabelos j eingehen:	$h(i,j)$

Gesucht ist eine Aufteilung der Losgröße Q in J Freigabelose. Festzulegen ist deren Größe x_{jm} und deren Fertigstellungszeitpunkt C_{jm}. Dies geschieht bis auf weiteres unter der Zielsetzung, den Fertigstellungszeitpunkt des letzten Freigabeloses auf der letzten Stufe C_{JM} und damit die Durchlaufzeit der vorgegebenen Menge Q zu minimieren. Aus diesen Daten lässt sich unmittelbar der gesuchte Plan π aufstellen. Formal stellt sich das Problem P(0) wie folgt dar, wobei sich die Formulierung an die Ausführungen bei Potts/Baker (1989), Glass et al. (1994) und Chen/Steiner (1997) anlehnt:

P(0): Das Losüberlappungsproblem unter Losverfügbarkeit

$$\text{Min } C_{JM} \tag{1}$$

$$C_{jm} \geq C_{j-1,m} + p_m x_{jm} \qquad (j = 2,...,J; m = 1,...,M) \tag{2}$$

$$C_{jm} \geq C_{h(i,j),m-1} + p_m x_{jm} \qquad (i = 1,...,J;\ j = 1,...,J; m = 2,...,M) \tag{3}$$

$$C_{11} \geq p_1 x_{11} \tag{4}$$

$$\sum_{j=1}^{J} x_{jm} = Q \qquad (m = 1,...,M) \tag{5}$$

$$x_{jm} \geq 0 \qquad (j = 1,...,J; m = 1,...,M) \tag{6}$$

Die Zielfunktion (1) minimiert den Fertigstellungszeitpunkt des letzten Loses auf der letzten Stufe. Die Restriktion (2) stellt sicher, dass die Stufen jeweils verfügbar sind, d.h. die Bearbeitung des vorangehenden Loses abgeschlossen ist. Die Restriktion (3) gewährleistet, dass jedes Loses auf der vorhergehenden Stufe fertiggestellt ist, bevor Teile dieses Loses weiterbearbeitet werden. Die im Index der Ungleichung (3) verwendete Funktion h liefert für das Argument (i,j) das letzte Freigabelos i, das noch Teile umfasst, die in das Freigabelos x_j eingehen. Diese Beziehung wird jeweils zwischen den Stufen m und m–1 betrachtet. Durch die Verwendung der Funktion h ist das Problem P(0) als ein Losüberlappungsproblem unter Losverfügbarkeit formuliert.

Die Restriktion (4) ist notwendig, da für die Bearbeitung des ersten Transferloses keine Vorgängerstufe zu berücksichtigen ist. Die Restriktion (5) gewährleistet, dass das gesamte Los auf allen Stufen abgearbeitet wird. Für alle Losgrößen gilt die Nicht-Negativitätsbedingung.

Die vorgegebene Losgröße Q in Restriktion (5) explizit zu verwenden, bedeutet, dass die Freigabelose in Stücken gemessen werden. Das ist z.B. dann vorteilhaft, wenn Ganzzahligkeit hinsichtlich der zu bildenden Lose verlangt wird. Können die Lose beliebige reelle Werte annehmen, bietet es sich an, die vorgegebene Losgröße Q auf 1 zu normieren. Die gesuchten Freigabelose geben dann die jeweiligen Anteile wieder, und (5a) ersetzt die Restriktion (5).

$$\sum_{j=1}^{J} x_{jm} = 1 \qquad (m = 1,....,M) \qquad (5a)$$

Betrachtet man die optimalen Losgrößen der Freigabelose aus Sicht der Reihenfolgeplanung, dann ist die inhaltliche Nähe zum Flow Shop offensichtlich. Jedes Freigabelos *j* kann als ein Job aufgefasst werden, der in Position *j* auf jeder der *m* Stufen eines *M*-stufigen Flow Shops zu bearbeiten ist. Die Ausführungszeit des Jobs entspricht $p_m x_{jm}$. Für Flow Shop Probleme unter regulärer Zielsetzung existiert eine Äquivalenzbeziehung zwischen einem gegebenen Problem und dem umgekehrten, dem inversen Scheduling Problem. Diese Äquivalenz gilt auch für die Losüberlappungsprobleme unter der Zielsetzung der Durchlaufzeitminimierung (vgl. Trietsch/Baker, 1993, S 1067).

Das zu einem Losüberlappungsproblem inverse Problem erhält man, indem $p_i = p_{M-i+1}$ für $i = 1,..., M$ gesetzt wird. Inhaltlich bedeutet es lediglich, dass die Maschinenfolge der Produkte umgekehrt durchlaufen wird. Die von einem Produkt bisher als letztes zu durchlaufende Stufe *M* ist im inversen Problem die erste. Die bisher erste Stufe wird im inversen Problem zur letzten Stufe etc. Hat man für das inverse Problem eine durchlaufzeitminimale Lösung ermittelt, so gilt diese Lösung – zurücktransferiert mit $x_{ji} = x_{j,M-i+1}$ für $i = 1,..., M$ und $j = 1,..., J$ – auch für das Ausgangsproblem. Man kann somit statt der gegebenen Problemstellung auch das inverse Problem lösen und die Ergebnisse übertragen. Die optimalen Lösungen sind in gewisser Hinsicht symmetrisch (vgl. Potts/Baker, 1989; Baker, 1998). Diese Eigenschaft wird im Weiteren als *Invers-Äquivalenz* bezeichnet.

Lemma 1: *Invers-Äquivalenz*

> Unter der Zielsetzung der Durchlaufzeitminimierung sind das Losüberlappungsproblem und das dazu inverse Problem äquivalent.

Aus dieser Eigenschaft kann für die optimalen Lösungen von P(0) eine wesentliche Besonderheit abgeleitet werden, die hier als Randkonsistenz bezeichnet wird. Die Beweisführung ist bei Potts/Baker (1989, S. 298 f.) oder leicht modifiziert bei Chen/Steiner (1997) angegeben:

Theorem 1: *Randkonsistenz*

> Es existiert für P(0) ein optimaler Plan π mit konsistenten Freigabelosen auf den ersten beiden und auf den letzten beiden Stufen.

Um das Theorem 1 zu beweisen, wird ein beliebiger, durchlaufzeitminimaler Plan π gewählt, dessen Losgrößen mit x_{jm} bezeichnet sind. Die zugehörige Durchlaufzeit des Plans beträgt C_{JM}. Es wird nun angenommen, dass ein alternativer Plan π' mit x'_{jm} existiert, der bis auf die Losgrößen der vorletzten und der letzten Stufe mit dem Plan π identisch ist. Auf der vorletzten und der letzten Stufe werden für π' konsistente Lose verwendet, während hinsichtlich der Lose

in π keine Einschränkungen vorgegeben sind. Zu untersuchen ist, ob die Veränderung von π zu π' zu einer Erhöhung der Durchlaufzeit führen kann.

Dazu wird in dem Plan π' ein beliebiges Los *k'* gewählt, das weder auf der letzten Stufe noch auf der vorletzten Stufe die Größe Null hat und somit tatsächlich existiert. Der Fertigstellungszeitpunkt des Loses *k'* auf der Stufe *M–1* ist mit $C'_{k',M-1}$ bezeichnet. Für den modifizierten Plan folgt aus (1) und (2):

$$C'_{JM} \geq C'_{k',M-1} + \sum_{j=k'}^{J} p_M x'_{jM} \tag{7}$$

Für die Fertigstellung des letzten Loses *J* im Plan π' gilt, dass sich dieser Zeitpunkt aus dem Fertigstellungszeitpunkt des Loses *k'* auf der vorletzten Stufe und der leerzeitfreien Folge der übrigen Lose auf der letzten Stufe *M* ergibt. Um eine untere Schranke für die maximale Durchlaufzeit im Plan π angeben zu können, wird im Plan π ein Los *k* ausgewählt, das das erste Los auf der Stufe *M* ist, welches Teile umfasst, die im Plan π' auf der Stufe *M*–1 im Los *k'* enthalten sind. Zur Vereinfachung bezeichnet

$$X_{jm} = \sum_{k=1}^{j} x_{km}$$

die Summe der Stücke in den ersten *j* Freigabelosen im ursprünglich Plan π und analog X'_{jm}, diese Summe im Plan π'. Für das Los *k* gilt in Bezug zu dem Los *k'*:

$$X_{k-1,M} \leq X_{k'-1,M-1} < X_{k,M} \tag{8}$$

Die beiden äußeren Terme stehen in der strengen Ungleichheitsbeziehung zueinander, da auf der Stufe *M* die Bearbeitung des Loses *k* nach der des Loses *k*–1 erfolgt und angenommen wurde, dass das Los *k'* nicht die Größe Null hat. Daraus folgt, dass ein Los, das Teile aus *k'* umfassen soll, ebenfalls nicht die Größe Null haben kann. Die Beziehungen zum mittleren Term folgen aus der Forderung, dass *k* das erste Los im Plan π auf der Stufe *M* ist, das Teile umfasst, die im Plan π' auf der Stufe *M*–1 im Los *k'* enthalten sind. Ist es nicht das erste Los, kann die Ungleichheit zum linken Term nicht erfüllt sein. Umfasst es die geforderten Teile nicht, ist die strenge Ungleichheit zum rechten Term nicht gegeben. Sieht man sich nun den mittleren Term in (8) näher an, ergibt sich:

$$X_{k'-1,M-1} = X'_{k'-1,M-1} = X'_{k'-1,M} \tag{9}$$

Da das Los *k* das erste Los im Plan π ist, das Teile umfasst, die im Plan π' auf der Stufe *M*–1 zum Los k' gehören, ist die Summe der Teile in den zuvor aufgelegten *k'*–1 Losen im Plan π und im Plan π' gleich und es folgt die linke Entsprechung in (9). Da im Plan π' mit konsistenten Losen auf den letzten beiden Stufen gearbeitet wird, folgt das rechte Gleichheitszeichen in (9). Wird auf die Terme (2) und (3) zurückgegriffen, zeigt sich, dass die folgende Ungleichheit gilt:

$$C_{JM} \geq C_{h(k,J),M-1} + \sum_{j=k}^{J} p_M x_{jM}$$

Entsprechend den Annahmen der Beweisführung ergibt sich daraus:

$$C_{JM} \geq C_{k',M-1} + \sum_{j=k}^{J} p_M x_{jM} \tag{10}$$

Da die Lose auf den ersten M–1 Stufen für beide Pläne identisch sind, kann die Differenz zwischen dem Plan π und dem Plan π' im Rückgriff auf (7) wie folgt bestimmt werden:

$$C_{JM} - C'_{JM} \geq p_M \left(\sum_{j=k}^{J} x_{jM} - \sum_{j=k'}^{J} x'_{jM} \right)$$

Daraus erhält man:

$$C_{JM} - C'_{JM} \geq p_M \left[\left(Q - X_{k-1,M}\right) - \left(Q - X'_{k'-1,M}\right) \right]$$

$$C_{JM} - C'_{JM} \geq p_M \left[X'_{k'-1,M} - X_{k-1,M} \right] \tag{11}$$

Aus (8) und (9) folgt, dass die Differenz in der Klammer der Gleichung (11) nicht negativ werden kann und als Konsequenz die Ungleichung (12) gelten muss:

$$C_{JM} \geq C'_{JM} \tag{12}$$

Es wurde gezeigt, dass ein Plan mit konsistenten Losen auf den letzten beiden Stufen keine höhere Durchlaufzeit verursachen kann, als ein optimaler Plan mit beliebiger Losgröße. In Kombination mit der Invers-Äquivalenz (Lemma 1) gilt das Ergebnis auch für die ersten beiden Stufen. Außerdem weisen Potts/Baker (1989) darauf hin, dass zur Beweisführung nicht verlangt wurde, dass nur ein Job vorliegt. Das Ergebnis hat deshalb auch für den Mehrproduktfall Gültigkeit. Als wesentliche ökonomische Konsequenz ist festzuhalten, dass es unter der Zielsetzung der Durchlaufzeitminimierung nicht sinnvoll sein kann, wenn ein Plan generiert wird, bei dem die Lose auf den äußeren beiden Stufen ihre Größe ändern, da es einen mindestens ebenso guten Plan gibt, der diese Änderungen der Losgrößen nicht erfordert und folglich leichter umsetzbar ist. Diese Regel hilft, viele dominierte oder gleichgute Alternativen mit einem einfachen Kriterium von der weiteren Betrachtung auszuschließen.

In Kombination mit der *Invers-Äquivalenz* folgt für die Lot Streaming Probleme mit geringer Stufenzahl aus der Randkonsistenz eine noch weitergehende Aussage.

Folgerung 1:

Für den Zwei- und den Dreistufenfall existiert für P(0) ein optimaler Plan π mit konsistenten Freigabelosen.

Die Folgerung 1 ergibt sich unmittelbar aus der Randkonsistenz in Verbindung mit der Invers-Äquivalenz. In einem Flow Shop mit zwei Stufen ist die vorletzte Stufe die erste Stufe und daher folgt aus der Randkonsistenz unmittelbar, dass eine Beschränkung auf konsistente Pläne die optimale Lösung von P(0) nicht ausschließt. In einem Flow Shop mit drei Stufen gilt in Folge der Randkonsistenz für die zweite und die dritte Stufe, dass ein optimaler Plan existiert, der konsistente Lose auf diesen beiden Stufen vorsieht. Wegen der Invers-Äquivalenz kann statt P(0) auch die zu P(0) inverse Problemstellung gelöst werden. Da die Lösungen äquivalent sind, bedeutet eine Einschränkung auf konsistente Lose auf den Stufen 1 und 2 (respektive 2 und 3) nicht, dass die durchlaufzeitminimale Lösung für P(0) ausgeschlossen wird. Damit müssen die Lose auf Stufe 2 aber gleichzeitig konsistent hinsichtlich der Lose der Stufe 1 und hinsichtlich der Lose der Stufe 3 sein. Dieser Forderung kann nur genügt werden, wenn die Lose über alle drei Stufen konsistent sind.

Folgerung 2:

Für P(0) mit $M \leq 3$ kann die Analyse auf konsistente Pläne beschränkt werden, ohne deshalb die optimale Lösung für P(0) auszuschließen.

Folgerung 3:

Für P(0) mit $M > 3$ existiert ein optimaler Plan π, bei dem die Losgrößen der ersten mit denen der zweiten Stufe übereinstimmen.

Folgerung 4:

Für P(0) mit $M > 3$ existiert ein optimaler Plan π, bei dem die Losgrößen der letzten mit denen der vorletzten Stufe übereinstimmen.

Folgerung 5:

Folgerungen 2 – 4 sind auf den Mehrproduktfall übertragbar.

5.2.1 Losüberlappung im Zweistufenfall

Aus Folgerung 1 ist deutlich geworden, dass eine Einschränkung auf konsistente Freigabelose für $M \leq 3$ die durchlaufzeitminimale Lösung nicht ausschließt und somit auch die Vorgabe der Losverfügbarkeit keine Einschränkung bedeutet. Nicht geklärt ist allerdings die Frage, wie diese Losgrößen zu ermitteln sind.

Im Zweistufenfall kann dazu eine einfache Formel angegeben werden. Ihre Herleitung deckt einige strukturelle Eigenschaften der Problemstellung auf und erlaubt den Begriff des kritischen Loses einzuführen (vgl. Potts/Baker, 1989; Baker/Pyke, 1990; Baker, 1998). Zu diesem Zweck wird zunächst die Problemstellung in Form eines Netzwerks präsentiert (vgl. Monma/Rinnooy Kan, 1983) und im Weiteren der Darstellung bei Glass et al. (1994) gefolgt. Die Knoten (j, m) repräsentieren in diesem Netz die Bearbeitung eines Loses auf einer Stufe. Die horizontal verlaufenden Kanten stellen sicher, dass die Stufe frei ist, bevor das nächste Los eingelastet wird und sind analog zur Restriktion (2) zu interpretieren. Die vertikal verlaufenden Kanten geben die Beziehungen analog zur Restriktion (3) wieder: Die Bearbeitung des Loses ist zu beenden, bevor es auf der nächsten Stufe begonnen werden kann (Losverfügbarkeit). Das folgende Beispiel zeigt ein Netzwerk für 5 Lose auf 2 Stufen. Da die Losgrößen konsistent sind, kann der Stufenindex entfallen und $\underline{x} = (x_1, x_2, x_3, x_4, x_5)$ vereinfachend geschrieben werden. Hinsichtlich der Bearbeitungszeiten sei $\underline{p} = (5, 4)$.

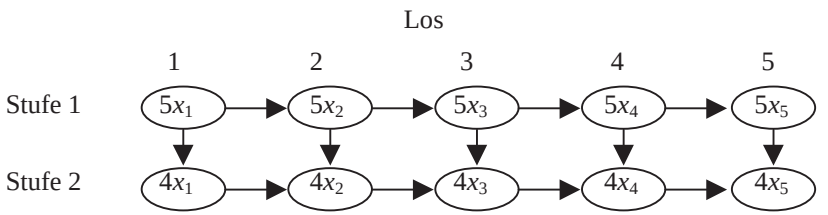


Abbildung 11: Losüberlappung als Netzwerk

Die Quelle des Netzwerks ist die Bearbeitung des ersten Loses auf der ersten Stufe, d.h. der Knoten (1,1). Ausgehend von diesem Knoten kann jeder andere Knoten auf mindestens einem Weg erreicht werden. Die Bearbeitung des Loses ist im Knoten (J, M) abgeschlossen und damit die Senke erreicht. Im Beispiel ist dies der Knoten (5,2). Die Durchlaufzeit ergibt sich als Länge des längsten Pfads vom Knoten (1,1) bis zum Knoten (J, M) durch das Netz. Dieser längste Pfad wird auch als der kritische Pfad bezeichnet. Als Gewichtung der Knoten dienen die mit $\underline{x}$ multiplizierten Bearbeitungszeiten $\underline{p}$. Insofern unterscheidet sich dieses Netz von den klassischen Netzen, da die Gewichtung nicht vorgegeben ist, sondern als Entscheidungsvariable gesucht wird: Der Vektor $\underline{x}$ ist so zu wählen, dass die Länge des kritischen Pfads minimiert wird. Soll z.B. ein Los der Größe 440 bearbeitet werden und sind die Losgrößen (50, 80, 90, 100, 120) festgelegt worden, ergibt sich das folgende Netz:

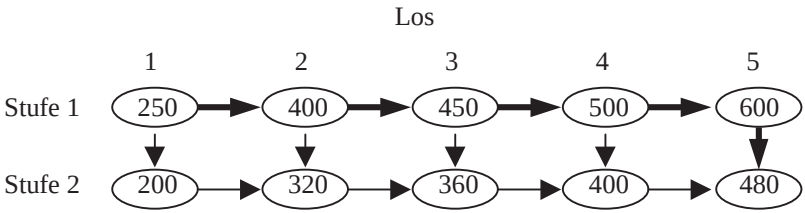


Abbildung 12: Eine beliebige Lösung im Netzwerk

In diesem Netzwerk beträgt die maximale Durchlaufzeit 2680 ZE. Der kritische Pfad führt über die Knoten: (1,1) – (2,1) – (2,1) – (4,1) – (5,1) – (5,2) und wurde in der Abbildung 12 fett hervorgehoben. Wird der Vektor der Losgrößen auf (130,89; 104,71; 83,76; 67,01; 53,61) geändert, sinkt die Länge des kritischen Pfads auf 2414,45 ZE und es ergibt sich das folgende Netz, dessen Werte gerundet angegeben sind:

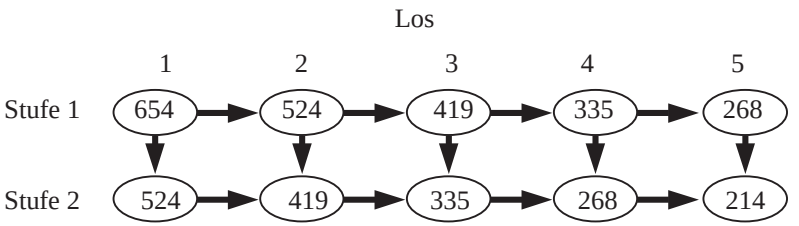


Abbildung 13: Eine verbesserte Lösung im Netzwerk

In diesem Fall ergeben sich 5 kritische Pfade: (1,1) – (1,2) ... – (5,1) – (5,2); (1,1) – (1,2) ... – (1,4) – (4,2) – (5,2); ... ; (1,1) – (1,2) ... – (5,2). Es stellt sich allerdings die Frage, ob die gezeigte Lösung bereits optimal ist, beziehungsweise wie sich eine optimale Lösung in der Netzwerkdarstellung auszeichnet. Zur ersten Annäherung an die gesuchte Antwort dient die folgende Überlegung: Die Fertigstellung des letzten Loses auf der letzten Stufe bestimmt die Durchlaufzeit. Treten auf dieser Stufe Leerzeiten auf, ist es möglich, die Lose durch zeitliches Verlagern (im Sinne einer späteren Freigabe) lückenlos aufeinander folgen zu lassen, ohne die Durchlaufzeit zu verlängern. Die Forderung, auf der letzten Stufe ohne Leerzeiten zu arbeiten, verlängert somit die minimale Durchlaufzeit nicht. Es existiert – für beliebig große M – immer ein optimaler Plan, auf dessen letzter Stufe keine Leerzeit zwischen der Bearbeitung von zwei Losen auftritt. Infolge der Invers-Äquivalenz gilt dieses Argument auch für die erste Stufe.

Folgerung 6:

Wird Leerzeitfreiheit verlangt, verlängert sich die Durchlaufzeit nicht, sofern diese Forderung die erste oder die letzte Stufe betrifft.

Hat P(0) nur zwei Stufen, folgt unmittelbar, dass die Forderung nach Leerzeitfreiheit auf allen Stufen keine Einschränkung bedeutet (vgl. Baker, 1998).

Folgerung 7:

Für Losüberlappungsprobleme mit zwei Stufen existiert ein leerzeitfreier, optimaler Plan.

Um einen leerzeitfreien Plan auf den Stufen zu erreichen, muss die Stufe 2 immer genau in dem Moment frei werden, in dem die Anlieferung des nächsten Loses von der Stufe 1 erfolgt. Wenn man nun Abbildung 13 näher betrachtet, fällt auf, dass dort die Aufteilung in die Losgrößen so gewählt wurde, dass sich diese lückenlose Abfolge ergibt, so entsprechen sich z.B. die Bewertungen in den Knoten (2,1) = 524 = (1,2). Allerdings ist damit noch nicht gezeigt, dass die Werte in Abbildung 13 optimal gewählt sind. Um diesen Zusammenhang aufzudecken, wird das kritische Los benötigt.

Definition: *Kritisches Los (critical sublot)*

Ein Los k heißt kritisch, wenn es nicht möglich ist, den Umfang des Loses k zu verändern (und den Umfang der übrigen Lose entsprechend anzupassen), ohne dass die Durchlaufzeit zunimmt.

Um formal die strukturellen Zusammenhänge herauszuarbeiten, kann die Darstellung der Losüberlappung für den Zweistufenfall vereinfacht werden. Ohne Einschränkung der Allgemeinheit wird zunächst vorausgesetzt, dass $p_1 \neq p_2$ ist. Die Ausführungszeit wird so normiert, dass $p_1 = 1$ gilt. Auf der zweiten Stufe beträgt sie damit: $q = p_2/p_1$. Die Losgröße Q wird ebenfalls auf 1 normiert. Da konsistente Lose verwendet werden, kann der Stufenindex bei den Losgrößen wegfallen. Betrachtet man ein beliebiges Los k, so ergibt sich die Durchlaufzeit C aus:

$$C \geq \sum_{j=1}^{k} x_j + q \sum_{j=k}^{J} x_j \tag{13}$$

Ist k das letzte der Lose ($k = J$), dann gilt: $C \geq 1 + q \cdot x_J > 1$; ist k hingegen das erste Los, so ist $C \geq q + x_1 > q$. Zusammengefasst erhält man:

$$C > \max\{1, q\} \tag{14}$$

Da sich die maximale Durchlaufzeit für mindestens ein Los k in Term (13) als eine Gleichheitsbeziehung ergeben muss, wird dieses Los im Weiteren betrachtet. Es ist das Los, über das der kritische Pfad von der ersten zur zweiten Stufe wechselt. Der kritische Pfad führt damit zweimal über das Los k. Daher

wird in Formel (13) auch zweimal über das Los *k* summiert: Einmal auf der ersten Stufe und einmal auf der zweiten Stufe. Das kritische Los ist somit alternativ als das Los *k* definiert, für das sich Ausdruck (13) maximiert.

In der Netzwerkdarstellung zeichnet sich ein kritisches Los *k* dadurch aus, dass ein kritischer Pfad existiert, der vertikal von einem Knoten (*k*,*m*) zu einem Knoten (*k*,*m*+1) führt. Zur Verdeutlichung wird im Folgenden die Lösung aus Abbildung 12 aufgegriffen. In der Gegenüberstellung von Gantt-Diagramm und Netzwerkdarstellung wird die Bedeutung des kritischen Loses offensichtlich.

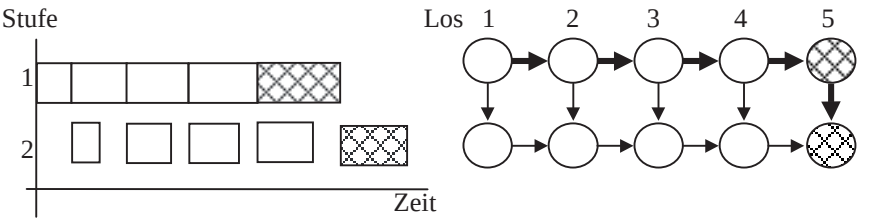


Abbildung 14: Kritisches Los im Gantt-Diagramm und der Netzwerkdarstellung

Der kritische Pfad durchläuft das Los *k*, das in Abbildung 14 schraffiert hervorgehoben ist, zweimal. Weiterhin fällt auf, dass in dieser Lösung nur ein kritisches Los auftritt. Wie Potts/Baker (1989) aber zeigen, gilt für eine optimale Lösung von P(0) mit $M = 2$, dass alle Lose kritisch sein müssen. Die Beweisführung erfolgt durch einen Widerspruch. Angenommen, es existiert ein Plan π, der zu einer optimalen Lösung von P(0) führt und in dem ein Los *h* nicht kritisch ist. Dann muss es möglich sein, zu diesem Plan π einen modifizierten Plan π' zu erstellen, in dem das Los *h* geringfügig vergrößert und die anderen entsprechend verkleinert werden, ohne dass sich deshalb die Durchlaufzeit im Plan π' gegenüber der Durchlaufzeit im Plan π erhöht. Durch die Verkleinerung der übrigen Lose muss aber das kritische Los in π betroffen sein und damit die aus (13) ermittelte Durchlaufzeit sinken, was einen Widerspruch zu der anfänglich getroffenen Annahme bedeutet. Gezeigt ist damit, dass ein Plan π nicht optimale Lösung des Problems P(0) sein kann, wenn eines der Lose nicht kritisch ist. Damit ist die Aussage in einer Richtung bewiesen. Gibt man umgekehrt vor, dass in einer Lösung alle Lose kritisch sind, dann folgt direkt aus der Definition des kritischen Loses, dass die Durchlaufzeit nicht verringert werden kann und somit die Lösung bereits optimal sein muss. Damit ist auch die Gegenrichtung der Aussage bewiesen.

Details zur Beweisführung sind bei Potts/Baker (1989) oder in Baker (1998, Abschnitt 9.5) dargestellt. Eine Verallgemeinerung der Aussage auf $M \geq 3$ gelang Glass et al. (1994), die ebenfalls mit einem Widerspruchsbeweis arbeiten.

Folgerung 8:

Eine Lösung von P(0) ist genau dann optimal, wenn alle Lose kritisch sind.

5.2.1.1 Kontinuierliche Losgrößen im Zweistufenfall

Ausgehend von den Folgerungen 7 und 8 zeigt sich, dass die in Abbildung 13 gezeigte Lösung genau die typische Struktur einer optimalen Lösung für den Zweistufenfall aufweist. Die Größe der Lose ist so festgelegt, dass über jedes Los mindestens ein kritischer Pfad führt. Diese Eigenschaft ist aber nur zu erreichen, wenn die Bearbeitung des Loses *j* auf der Stufe 2 mit der Bearbeitung des Loses *j*+1 auf der Stufe 1 übereinstimmt. Die folgende Abbildung 15 zeigt das Gantt-Diagramm und die korrespondierende Netzwerkdarstellung für die optimale Lösung des Beispiels aus Abbildung 13, in der die Entsprechung der Bearbeitungszeiten deutlich zu sehen ist.

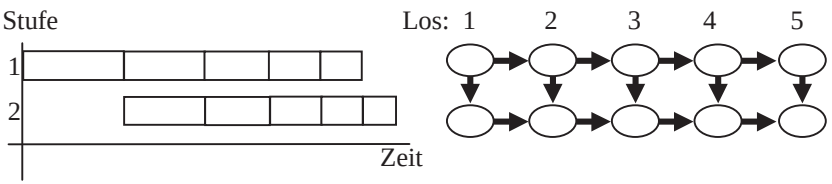


Abbildung 15: Eine optimale Lösung im Gantt-Diagramm und im Netzwerk

Formal gilt somit $p_1 x_{j+1} = p_2 x_j$. Durch iteratives Einsetzen ergibt sich:

$$x_j = x_1 q^{j-1} \tag{15}$$

Da sich die Summe der Losgrößen zu Q addieren muss, gilt die folgende Entsprechung:

$$Q = \sum_{j=1}^{J} x_j = x_1 \sum_{j=1}^{J} q^{j-1} \tag{16}$$

Aus dieser Beziehung kann die optimale Größe der Lose bestimmt werden (vgl. Baker, 1998):

$$x_j = Q \cdot q^{j-1} \Big/ \sum_{j=1}^{J} q^{j-1} \tag{17}$$

Da der Quotient zweier aufeinanderfolgender Losgrößen konstant ist, bilden die Losgrößen eine geometrische Folge: $x_j = q\, x_{j-1}$.

Folgerung 9:

Für P(0) mit $M = 2$ bildet die Menge der optimalen Losgrößen x_j eine geometrische Folge mit $x_j = Q \cdot q^{j-1} \Big/ \sum_{j=1}^{J} q^{j-1}$ mit $q = p_2/p_1$.

Betrachten wir hierzu nochmals das Beispiel aus Abbildung 15, in dem $Q = 440$, $J = 5$ und $p_1 = 5$ sowie $p_2 = 4$ gesetzt ist, ergibt sich $q = p_2/p_1 = 0{,}8$.

Für x_1: $(440 \cdot 1) \quad / (1 + 0{,}8 + 0{,}8^2 + 0{,}8^3 + 0{,}8^4) = 440 / 3{,}3616 = 130{,}890$
Für x_2: $(440 \cdot 0{,}8) \ / (1 + 0{,}8 + 0{,}8^2 + 0{,}8^3 + 0{,}8^4) = 352 / 3{,}3616 = 104{,}714$
Für x_3: $(440 \cdot 0{,}8^2) / (1 + 0{,}8 + 0{,}8^2 + 0{,}8^3 + 0{,}8^4) = 281{,}6 / 3{,}3616 = 83{,}769$
Für x_4: $(440 \cdot 0{,}8^3) / (1 + 0{,}8 + 0{,}8^2 + 0{,}8^3 + 0{,}8^4) = 225{,}28 / 3{,}3616 = 67{,}016$
Für x_5: $(440 \cdot 0{,}8^4) / (1 + 0{,}8 + 0{,}8^2 + 0{,}8^3 + 0{,}8^4) = 180{,}22 / 3{,}3616 = 53{,}611$

440,000

Aus Folgerung 6 ist bekannt, dass die optimale Lösung nicht ausgeschlossen wird, wenn auf den Randstufen Leerzeitfreiheit gefordert ist. In der Konsequenz gilt, dass sich in der optimalen Lösung die minimale Durchlaufzeit als Summe der Ausführungszeit des ersten Loses auf der ersten zuzüglich der Bearbeitung der gesamten Menge auf Stufe 2 ergibt:

$$C_{\min} = p_1 x_1 + Q\, p_2 \tag{18}$$

Im Beispiel erhält man: $C_{\min} = 130{,}89 \cdot 5 + 440 \cdot 4 = 2414{,}45$ ZE.

Sollte für P(0) im Zweistufenfall $p_1 = p_2$ gelten, vereinfacht sich die Formel (17). In diesem Fall ist es optimal, mit J gleich großen Losen zu arbeiten:

$$x_j = Q/J \tag{19}$$

Folgerung 10:

Für P(0) mit $M = 2$ und $p_1 = p_2$ sind die optimalen Losgrößen konstant.

Während sich die bisherigen Ausführungen auf die Festlegung von kontinuierlichen Losgrößen beschränken, wird im folgenden Abschnitt als zusätzliche Restriktion eingeführt, dass die Losgrößen ganzzahlig sein müssen.

5.2.1.2 Diskrete Losgrößen im Zweistufenfall

Mit der Frage der Losüberlappung im Zweistufenfall unter Einsatz ganzzahliger Losgrößen beschäftigte sich zunächst Truscott (1985 und 1986). Allerdings ging es ihm in diesen Untersuchungen nur um die Ermittlung von zulässigen ganzzahligen Lösungen, die er bei der Betrachtung restringierter Transportkapazität nutzen wollte. Er behandelte nicht die Optimierung der Losgrößen bei

Vorgabe der Anzahl an Transporten. Dieser Fragestellung ging Trietsch nach, dessen Ergebnisse in Trietsch/Baker (1993) zusammengefasst sind. Später wurden sie auch bei Baker (1998) aufgegriffen. Zum Beweis, dass das im Folgenden vorzustellende Verfahren eine optimale Lösung liefert, wird in der Literatur auf das mittlerweile nicht mehr erhältliche Arbeitspapier von Trietsch (1989) verwiesen.[1]

Um zunächst die Problemstellung zu verdeutlichen, wird die für das Beispiel gefundene optimale Lösung mit nicht-ganzzahligen Werten: $\underline{x}$ = (130,890; 104,714; 83,769; 67,016; 53,611) aufgegriffen und es werden durch Auf- beziehungsweise Abrunden ganzzahlige Losgrößen erreicht. Man erhält: $\underline{x}$ = (131; 105; 84; 67; 54) und stellt fest, dass nicht mehr Q = 440 sondern 441 Stück eingeplant wurden. Insofern kann das Runden zu unzulässigen Lösungen führen. Weiter ist in dieser Situation nicht zu erkennen, welche der Losgrößen auf- und welche abzurunden sind. Beispielsweise könnte $\underline{x}$ = (130; 105; 84; 67; 54) oder ebenso gut $\underline{x}$ = (131; 105; 84; 67; 53) zu einer durchlaufzeitminimalen Lösung führen. Aus diesem Grund wird im Folgenden ein Verfahren zur Ermittlung der optimalen Losgrößen vorgestellt, das bei Baker (1998) beschrieben ist. Den Ausgangpunkt bildet die Überlegung, dass sich durch die zusätzliche Berücksichtigung der Ganzzahligkeit $C_{\min}$ nicht reduzieren wird. Insofern kann die für kontinuierliche Losgrößen ermittelte Durchlaufzeit als Schranke genutzt werden. Sie wird im Weiteren mit $C^{K}_{\min}$ bezeichnet und für die Durchlaufzeit bei ganzzahligen (diskreten) Losgrößen die Bezeichnung $C^{D}_{\min}$ genutzt. X_j bezeichnet wieder die über die ersten j Lose kumulierte Stückzahl. Hinsichtlich des Startzeitpunkts des j-ten Loses auf der zweiten Stufe (er ist auf der linken Seite der Formel (20) durch $C_{j2} - p_2\, x_j$ gegeben), muss die folgende Beziehung gelten:

$$C_{j2} - p_2\, x_j \le C^{D}_{\min} - p_2\,(Q - X_{j-1}) \tag{20}$$

Die in der Klammer angegebene Menge $Q - X_{j-1}$ entspricht den Stücken, die nach dem Los j –1 auf der Stufe 2 zu fertigen sind, d.h. die Lose x_j, x_{j+1},.., x_J. Wird das Los j auf der Stufe 2 in dem spätest zulässigen Zeitpunkt gestartet, gilt in (20) die Gleichheit. Hinsichtlich der Bereitstellung des j-ten Loses gilt, dass die Produktion der Teile auf der ersten Stufe abgeschlossen sein muss:

$p_1 X_j \le C_{j2} - p_2\, x_j = C^{D}_{\min} - p_2\,(Q - X_{j-1})$ und damit:

$$X_j \le [C^{D}_{\min} - p_2\,(Q - X_{j-1})] \,/\, p_1 \tag{21}$$

[1] Trietsch, D.: Polynomial Transfer Lot Sizing Techniques for Batch Processing on Consecutive Machines. Technical Report NPS-54-89-011. Naval Postgraduate School, Monterey, California, 1989.

Aus (21) ergibt sich ein rekursiver Zusammenhang zwischen X_j und X_{j-1}. Dabei ist X_j so groß zu wählen, wie die größte ganze Zahl, die (21) erfüllt, d.h. die mit (21) ermittelten Größen werden abgerundet. Alternativ können direkt x_1, x_2, ..., x_J bestimmt werden, wenn mit $x_0 = 0$ begonnen wird. Als Startwert für C^D_{min} bietet es sich an, C^K_{min} zu nutzen. Sollten alle Ausführungszeiten ganzzahlig sein, kann C^D_{min} auf den nächsten ganzzahligen Wert aufgerundet werden. Die sukzessive Berechnung der kumulierten Losgröße ist dann abzubrechen, sobald $X_J = Q$ oder $X_J < Q$ wird.

Im ersten Fall ($X_J = Q$) ist die optimale Lösung gefunden. Die aus den X_j ermittelbaren Losgrößen x_j stellen optimale Lösungen des Problems dar; die Durchlaufzeit entspricht dem vorgegebenen Wert C^D_{min}.

Im zweiten Fall ($X_J < Q$) wurde C^D_{min} zu klein gewählt. Bei rein ganzzahligen Ausführungszeiten wird $C^D_{min} = C^D_{min} + 1$ gesetzt und das Verfahren erneut durchgeführt. Sollten die Ausführungszeiten nicht ganzzahlig sein, empfiehlt Baker (1998) C^D_{min} um den kleinsten Bruchteil zu erhöhen, der zum Aufrunden der mittels Formel (21) festgelegten X_j auf die nächste ganze Zahl erforderlich wäre.

Die Vorgehensweise wird nun an Hand des oben eingeführten Beispiels demonstriert. Für dieses sind die folgenden Daten gegeben: $p_1 = 5$; $p_2 = 4$; $Q = 440$. Für $J = 5$ Lose wurde $C^K_{min} = 2414{,}45$ ermittelt.

Aus $C^K_{min} = 2414{,}45$ und den beiden ganzzahligen Ausführungszeiten $\underline{p} = (5,4)$ folgt $C^D_{min} = 2415$. Es wird nun sukzessiv Formel (21) verwendet:

$X_1 \leq [2415 - 4\,(440-0)] / 5 = 131 \Rightarrow X_1 = 131$

$X_2 \leq [2415 - 4\,(440-131)] / 5 = 235{,}8 \Rightarrow$ Abrunden liefert: $X_2 = 235$

$X_3 \leq [2415 - 4\,(440-235)] / 5 = 319 \Rightarrow X_3 = 319$

$X_4 \leq [2415 - 4\,(440-319)] / 5 = 386{,}2 \Rightarrow$ Abrunden liefert: $X_4 = 386$

$X_5 \leq [2415 - 4\,(440-386)] / 5 = 439{,}8$

Da $439{,}8 < 440$ ist der Startwert $C^D_{min} = 2415$ zu klein gewählt. Als nächster Wert wird $C^D_{min} = 2415 + 1 = 2416$ vorgegeben. Man erhält die folgende Berechnung:

$X_1 \leq [2416 - 4\,(440-0)] / 5 = 131{,}2 \Rightarrow$ Abrunden liefert: $X_1 = 131$

$X_2 \leq [2416 - 4\,(440-131)] / 5 = 236 \Rightarrow X_2 = 236$

$X_3 \leq [2416 - 4\,(440-236)] / 5 = 320 \Rightarrow X_3 = 320$

$X_4 \leq [2416 - 4\,(440-320)] / 5 = 387{,}2 \Rightarrow$ Abrunden liefert: $X_4 = 387$

$X_5 \leq [2416 - 4\,(440-387)] / 5 = 440$

Aus den gegebenen kumulierten Losgrößen lassen sich die optimalen ganzzahligen Losgrößen bestimmen: $\underline{x}$ = (131; 105; 84; 67; 53). Stellt man diese Lösung in Form eines Gantt-Diagramms dar, wird deutlich, dass die Lösung nicht mehr zugleich leerzeit- und wartezeitfrei sein kann, da z.B. zwischen dem ersten und dem zweiten Los auf der zweiten Stufe 1 ZE Leerzeit auftritt und die Stufe 2 für die Bearbeitung des fünften Loses erst im Zeitpunkt 2204 bereitsteht, das Los aber auf der Stufe 1 bereits im Zeitpunkt 2200 fertig bearbeitet ist.

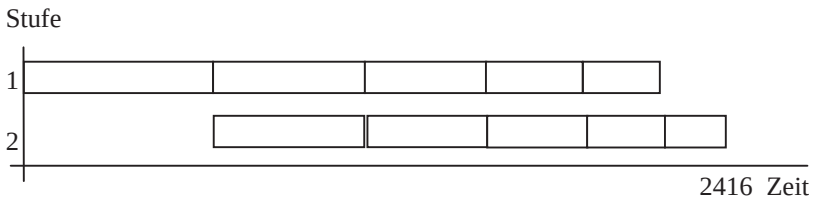


Abbildung 16: Die optimale ganzzahlige Lösung im Gantt-Diagramm

Wie Trietsch/Baker (1993) anmerken, ist der mit diesem Verfahren konstruierte Plan nicht unbedingt eindeutig, d.h. es kann andere Aufteilungen geben, die zu der gleichen Durchlaufzeit führen. Ein Möglichkeit diese zu finden, besteht darin, das inverse Problem (p_1 = 4, p_2 = 5) zu lösen. Für das vorliegenden Beispiel ergibt sich (zurücktransformiert) die Sequenz $\underline{x}$ = (130; 105; 84; 67; 54), die ebenfalls zu der Durchlaufzeit von 2416 ZE führt.

Zur Beurteilung des Verfahrens ist festzuhalten, dass hier mit einer einfachen iterativ anwendbaren und leicht mit Hilfe einer Tabellenkalkulation umsetzbaren Formel die optimale, ganzzahlige Lösung gefunden werden kann.

Wie Trietsch (1989) laut Baker (1998) gezeigt hat, ist der Rechenaufwand des Algorithmus zudem polynomial begrenzt. Dies ist insofern ein wichtiges Ergebnis, da die Alternative in der Lösung eines ganzzahligen Optimierungsproblems z.B. mittels eines Branch & Bound Verfahrens liegt.

Halten wir auch hier die wesentlichen Ergebnisse fest:

Folgerung 11:

Für P(0) mit M = 2 können die optimalen diskreten Losgrößen mit Hilfe eines polynomial begrenzten Suchalgorithmus gefunden werden. Die optimalen kumulierten Losgröße X_j erfüllen für den kleinstmöglichen Wert C^{D}_{min} die Relation: $X_j \leq [C^{D}_{min} - p_2 (Q - X_{j-1})] / p_1$

Folgerung 12:

Für P(0) mit M = 2 und diskreten Losgrößen muss die optimale Lösung nicht eindeutig sein.

Folgerung 13:

> Für P(0) mit $M = 2$ und diskreten Losgrößen kann für einen optimalen Plan π nicht zugleich Wartezeit- und Leerzeitfreiheit garantiert werden.

Der Algorithmus ist zudem von Relevanz, da er als Baustein die Konstruktion von Lösungsverfahren für Losüberlappungsprobleme mit mehr als zwei Stufen ermöglicht. Diese Erweiterung wird im Abschnitt 5.2.2.3 aufgegriffen. Zuvor sind allerdings die generellen Eigenschaften der Losüberlappung im Dreistufenfall zu behandeln.

5.2.2 Losüberlappung im Dreistufenfall

Erweitert man die Betrachtung der Eigenschaften und Verfahren der Losüberlappung vom Zweistufenfall auf den Dreistufenfall, rücken die Wartezeit- und die Leerzeitfreiheit sowie die Art der verwendeten Lose in den Vordergrund. Sie bestimmen das Vorgehen bei der Festlegung der optimalen Losgrößen und die Eigenschaften der Pläne wesentlich mit. Dieser Zusammenhang wird z.B. unmittelbar aus der Folgerung 6 erkennbar. Dort war festgehalten worden, dass die Forderung nach Leerzeitfreiheit auf der ersten und auf der letzten Stufe einen Plan mit minimaler Durchlaufzeit nicht ausschließt. Sind allerdings mehr als zwei Stufen zu berücksichtigen, deckt Folgerung 6 nicht mehr alle Stufen ab. Entsprechend ist zu untersuchen, welche Änderungen sich für die optimalen Pläne bei dem Übergang von zwei auf drei Stufen ergeben. Wie die vorangehenden Ausführungen zum Zweistufenfall dienen auch die Überlegungen zum Dreistufenfall in erster Linie dem Erkennen von Strukturen und zur Konstruktion heuristischer Verfahren. So kann es für größere Probleme sinnvoll sein, von den Stufen, die vergleichsweise schnell arbeiten, vorübergehend zu abstrahieren. Der Plan wird auf die Engpässe ausgerichtet festgelegt und von den Einflüssen der übrigen Maschinen auf die Losgrößen abstrahiert. Hinsichtlich des Erkennens von Engpässen gelten dabei allerdings die im Abschnitt 3.2.1.1 genannten Einschränkungen.

5.2.2.1 Wartezeit- und Leerzeitfreiheit im Dreistufenfall

Wie in Folgerung 7 festgehalten und in Abbildung 15 verdeutlicht, erfolgt in der optimalen Lösung für den Zweistufenfall die Bearbeitung auf den Maschinen ohne Unterbrechung, d.h. leerzeitfrei. Wird genau diese Eigenschaft verlangt, liegt die Idee nahe, auch für Probleme mit drei oder mehr Stufen die Formel (16) iterativ für jeweils zwei aufeinanderfolgende Stufen anzuwenden. Ist Stückverfügbarkeit gegeben, liefert dieses Vorgehen die optimale Lösung (Trietsch/Baker, 1993). Das folgende Beispiel zeigt die Vorgehensweise. Es sei:

$Q = 60$; $M = 3$; $p_1 = 3$; $p_2 = 5$; $p_3 = 7$. Gesucht ist eine Lösung mit $J = 4$ Freigabelosen. Da die Freigabelose x_{jm} nicht unbedingt mit den Transferlosen übereinstimmen müssen, werden diese mit y_{im} ($i = 1,..., I$) bezeichnet. Dabei gilt, dass die Freigabelose der ersten Stufe immer den Transferlosen der ersten Stufe und diese auch den Freigabelosen der zweiten Stufe entsprechend müssen, da eine abweichende Setzung nicht vorteilhaft sein kann (sofern nicht Restriktionen hinsichtlich der Bereitstellung auf der ersten Stufe dem entgegenstehen).

$x_{11} = y_{11} = x_{12} = 5{,}9959$; $x_{21} = y_{21} = x_{22} = 9{,}926$;

$x_{31} = y_{31} = x_{32} = 16{,}544$; $x_{41} = y_{41} = x_{42} = 27{,}573$.

Für die Transferlose von der zweiten Stufe (y_{i2}) gilt diese Entsprechung hinsichtlich der Freigabelose der dritten Stufe (x_{j3}).

$y_{12} = x_{13} = 8{,}445$; $y_{22} = x_{23} = 11{,}824$; $y_{32} = x_{33} = 16{,}544$; $y_{42} = x_{43} = 23{,}175$.

Das erste Transferlos von Stufe 2 zur Stufe 3 erfolgt nach der Fertigstellung von 8,445 Einheiten. Es umfasst damit das erste Freigabelos der Stufe 2 komplett und aus dem zweiten Freigabelos die ersten 2,486 Einheiten. Die Abbildung 17 zeigt die teilweise Weitergabe aus einem laufenden Los indem auf der zweiten Stufe die Rechtecke horizontal geteilt sind. Die Durchlaufzeit beträgt 480,09 ZE.

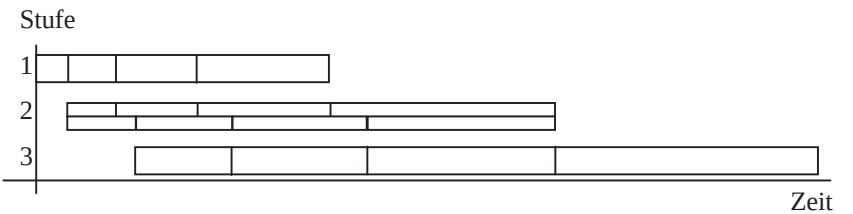


Abbildung 17: Leerzeitfreie, optimale Lösung mit variablen Losen, Stückverfügbarkeit und kontinuierlicher Losgröße

Inhaltlich kann man sich die in Abbildung 17 gezeigte Struktur so erklären, dass sich jedes mehrstufige System in eine Folge von überlappenden Zweistufenproblemen (Stufe 1 zu Stufe 2; Stufe 2 zu Stufe 3, Stufe 3 zu Stufe 4 etc.) zerlegen lässt. Da Stückverfügbarkeit vorausgesetzt und Leerzeitfreiheit gefordert wird, lassen sich je zwei Stufen separieren und unabhängig von den übrigen betrachten. Wegen der Stückverfügbarkeit beeinflusst die Aufteilung und Weitergabe, die zwischen zwei Stufen festgelegt wird, die Planung für die jeweils folgenden Stufen nicht. Diese Überlegung gilt für beliebig große Probleme, in denen diese beiden Anforderungen aufeinandertreffen (vgl. Baker/Jia, 1993). Optimal gelöst ist damit – unter der Annahme der Stückverfügbarkeit – der Fall V/O/B, der in Abbildung 10 hinsichtlich seiner Beziehungen zu den anderen Problemstellungen auf der zweithöchsten Ebene eingeordnet wurde. Von ökonomischer Bedeutung ist dieses Ergebnis, da es den Produktionsplaner in die

Lage versetzt, mit einer vergleichsweise einfachen Formel eine optimale Lösung für eine beliebige Zahl an Stufen und Losen ermitteln zu können.

Die so gefundene minimale Durchlaufzeit kann zudem als eine Schranke verwendet werden, um für stärker restringierte Probleme (z.B. solche mit ganzzahligen Losen) das Potential abzuschätzen, das man durch die Verbesserung eines Plans höchstens erreichen kann.

Folgerung 14:

Bei Stückverfügbarkeit und Leerzeitfreiheit existiert ein durchlaufzeitminimaler Plan, wenn für die Transferlosgrößen y_{im} (i = 1,..., I) gilt: $y_{im} = Q \cdot q_m^{\,i-1} \Big/ \sum_{i=1}^{I} q_m^{\,i-1}$ mit $q_m = p_{m+1}/p_m$ für m = 1,..., M-1.

Um die minimale Durchlaufzeit direkt angeben zu können, betrachten wir wieder den Dreistufenfall und nutzen die Eigenschaft, dass die Leerzeit vor dem Start der Bearbeitung auf Stufe 3 leicht zu ermitteln ist. Sie entspricht der Summe der Bearbeitungszeit des ersten Transferloses der Stufe 1 zuzüglich des ersten Transferloses der Stufe 2. Die Größe des jeweils ersten Loses pro Stufe wird in der folgenden Formel mit den beiden hinteren Summanden in der eckigen Klammer bestimmt (vgl. Baker/Jia, 1993).

$$C_{\min} = Q\left[p_3 + \frac{p_1 - p_2}{1-(q_1)^J} + \frac{p_2 - p_3}{1-(q_2)^J}\right] \tag{22}$$

Setzt man die in Abbildung 17 verwendeten Daten: Q = 60; M = 3; p_1 = 3; p_2 = 5; p_3 = 7 und J = 4 in Formel (22) ein, bestätigt sich das vorne angegebene Ergebnis:

$$C_{\min} = 60\left[7 + \frac{-2}{-6{,}716} + \frac{-2}{-2{,}8416}\right] = 480{,}1\ ZE$$

Im Dreistufenfall gilt für durchlaufzeitminimale Lösungen bei Stückverfügbarkeit und variablen Losen weiterhin, dass es einen optimalen Plan gibt, der eine wartezeitfreie Weiterbearbeitung (*no-wait*) der Lose ermöglicht (vgl. Trietsch/Baker, 1993). Zum Beweis durch Widerspruch wird angenommen, dass in einem optimalen Plan π ein Los k auf seine Weiterbearbeitung warten muss. Ein Los k muss dann warten, wenn die nächste Stufe in dem Moment noch nicht frei ist, in dem mit der Bearbeitung des Loses k begonnen werden kann. Da es nicht optimal sein kann, auf einer Stufen das erste Los auf seine Weiterbearbeitung warten zu lassen, folgt unmittelbar, dass $k > 1$ sein muss. Dann aber ist es möglich, Los k geringfügig zu vergrößern und dafür ein davor gefertigtes Los $g < k$ entsprechend zu verkleinern, so dass die Wartezeit in der Weiterbearbeitung des Loses k kleiner wird oder ganz entfällt. Das verkleinerte Los g wird früher weitergegeben, die Wartezeit für Los k sinkt und damit sinkt auch die Durchlaufzeit. Da vorausgesetzt wurde, dass der Plan π bereits durch-

laufzeitminimal ist, liegt ein Widerspruch vor. Gezeigt ist, dass bei variablen Losen ein Plan, bei dem ein Los auf seine Bearbeitung warten muss, nicht optimal sein kann. Aus Sicht eines Produktionsplaners ist diese Eigenschaft interessant, da sie ein einfaches Kriterium liefert, um unter alternativen Plänen die auszuschließen, die nicht optimal sein können.

Folgerung 15:

> Bei Stückverfügbarkeit und variablen Losen existiert ein optimaler Plan, der eine wartezeitfreie Bearbeitung für alle Lose ermöglicht.

Betrachtet man unter diesem Aspekt die in Abbildung 17 gezeigte Lösung, fällt auf, dass sie nicht nur wartezeitfrei, sondern zugleich leerzeitfrei ist. Ob eine optimale Lösung existiert, die diese beiden Eigenschaften zugleich erfüllt, hängt von dem Verhältnis der Ausführungszeiten auf den Stufen ab (vgl. Baker, 1998; Glass et al., 1994). Um diesen Zusammenhang zu erkennen, wird zunächst einmal angenommen, dass nur auf den Stufen 1 und 3 unterbrechungsfrei gefertigt wird, d.h. auf der Stufe 2 treten zwischen den Losen Leerzeiten auf. Wie in Folgerung 6 bereits festgehalten, zeichnen sich optimale Lösungen durch Leerzeitfreiheit auf der ersten und der letzten Stufe aus, so dass diese Annahme keine Einschränkung bedeutet. Damit die Lose eine wartezeitfreie Bearbeitung über die drei Stufen erfahren, muss die folgende Entsprechung gelten:

$$x_{j+1}\,(p_1 + p_2) = x_j\,(p_2 + p_3) \quad \text{für } j = 1,\ldots, J\text{-}1 \tag{23}$$

Sie ist wie folgt zu interpretieren: Die kumulierte Bearbeitungszeit des j-ten Loses auf den Stufen 2 und 3 muss der kumulierten Bearbeitungszeit des j+1-ten Loses auf den Stufen 1 und 2 entsprechen. Warum dieser Zusammenhang eine Bedingung für die wartezeitfreie Bearbeitung der beiden Lose darstellt, verdeutlicht die folgende Grafik an Hand eines Ausschnitts eines Plans mit wartezeitfreier Bearbeitung.

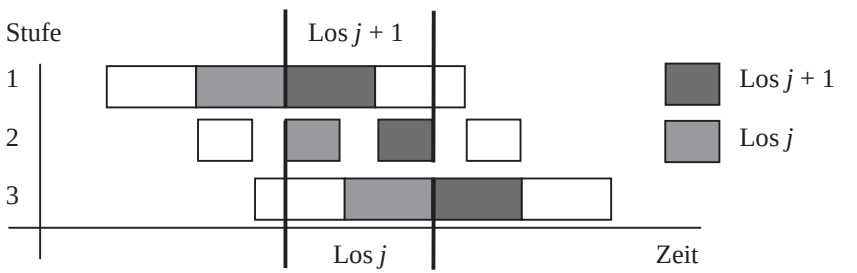


Abbildung 18: Entsprechung der Bearbeitungszeiten im wartezeitfreien Plan

Gemäß Folgerung 6 (Leerzeitfreiheit auf den Randstufen) beginnt die Bearbeitung des j+1-ten Loses auf Stufe 1 und auf Stufe 3 unmittelbar nach der Fertigstellung des vorangehenden Loses j. Damit sich eine wartezeitfreie Bearbeitung ergeben kann, muss Stufe 3 frei sein, wenn ein Los auf Stufe 2 fertigge-

stellt ist. Für Los j muss die Zeitspanne zwischen dem Ende der Bearbeitung auf Stufe 1 und dem Ende der Bearbeitung auf Stufe 3 genau so lang sein wie für die Bearbeitung des Loses j+1 auf den Stufen 1 und 2 benötigt wird. Diese Zeitspanne ist in Abbildung 18 durch die fett eingezeichneten Geraden angedeutet. Nur wenn beide Zeitspannen übereinstimmen, ist sichergestellt, dass die Stufe 3 exakt in dem Moment frei wird, wenn das Los j+1 von Stufe 2 eintrifft. Wird die Stufe 3 früher frei, liegt ein Widerspruch gegen die Leerzeitfreiheit auf den Randstufen vor, wird sie später frei, ist die wartezeitfreie Bearbeitung nicht gegeben.

Formel (23) kann so umgestellt werden, dass erneut die geometrische Folge der Losgrößen in der warte- und leerzeitfreien Lösung offensichtlich wird. Dazu bezeichnet $q = (p_2 + p_3) / (p_1 + p_2)$ und analog zur Herleitung von Formel (15) erhält man wieder den Zusammenhang $x_j = x_1 q^{j-1}$.

Betrachtet man statt zwei aufeinanderfolgender Lose eine Sequenz von Losen, wird sie zu einem wartezeitfreien Plan führen, wenn folgender Zusammenhang sichergestellt ist:

$$p_1 (x_1 + x_2 + ... + x_{j+1}) \geq p_1 x_1 + p_2 (x_1 + x_2 + ... + x_j) \quad \text{für } j = 1,..., J\text{-}1 \tag{24}$$

Die Bearbeitungszeitsumme der ersten j + 1 Lose darf nicht kleiner sein als die Summe der Bearbeitungszeiten des ersten Loses auf der ersten Stufe, zuzüglich der Summe der Bearbeitungszeiten der ersten j Lose auf Stufe 2. Zur Verdeutlichung zeigt in Abbildung 19 ein Pfeil auf dieses erste Los auf Stufe 1. Nur wenn (24) erfüllt ist, kann Stufe 2 frei sein, wenn das Los j+1 von der ersten zur zweiten Stufe transferiert wird. Dies ist erforderlich, damit das von Stufe 1 angelieferte Los ohne Wartezeit weiterbearbeitet werden kann. Dieser Zusammenhang wird mit Hilfe der Abbildung 19 verdeutlicht:

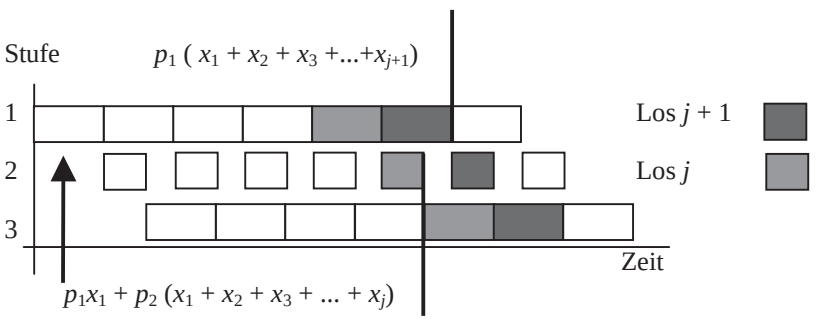


Abbildung 19: Erläuterung zur Wartezeitfreiheit

Wesentlich ist dabei, dass der untere Term $p_1 x_1 + p_2 (x_1 + x_2 + ... + x_j)$ nicht größer werden darf als der obere Term $p_1 (x_1 + x_2 + ... + x_{j+1})$, da sonst die Stufe 2 nicht frei ist, um das j +1-te Los ohne Wartezeit weiter zu bearbeiten. Aus-

gehend von Formel (24) kann vereinfacht und wegen des Zusammenhangs $x_j = x_1 q^{j-1}$ (Formel 15) derart umgestellt werden, dass nur x_1 und q auftreten:

$$p_1(x_2 + x_3 + ... + x_{j+1}) \geq p_2(x_1 + x_2 + x_3 + ... + x_j)$$

$$p_1 x_1(q + q^2 + ... + q^j) \geq p_2 x_1(1 + q + q^2 + ... + q^{j-1}) \quad \text{für } j = 1,...,J \tag{25}$$

Um den Klammerausdruck zu vereinheitlichen, wird q auf der linken Seite in (25) ausgeklammert und wieder durch $(p_2 + p_3) / (p_1 + p_2)$ ersetzt:

$$qp_1 x_1(1 + q + q^2 + ... + q^{j-1}) \geq p_2 x_1(1 + q + q^2 + ... + q^{j-1})$$

$$qp_1 \geq p_2$$

$$p_1 p \geq p_2^2 \tag{26}$$

Wie die Herleitung zeigt, korrespondiert die Relation in Formel (26) mit den oben getroffenen Annahmen, dass alle Lose wartezeitfrei bearbeitet werden und zugleich auf der Stufe 2 Leerzeiten zugelassen sind. Wären auf der Stufe 2 keine Leerzeiten zulässig, dann müsste in (24) statt „≥“ ein „=“ gelten. Wie Baker/Jia (1993) ausführen, ist Formel (26) inhaltlich so zu interpretieren, dass dort geprüft wird, ob Stufe 2 dominiert wird oder nicht. Wenn $p_1 p_3 \geq p_2^2$ gilt, wird die Stufe 2 von den Stufen 1 und 3 dominiert. Gilt hingegen $p_1 p_3 < p_2^2$, dann dominiert die Stufe 2 die Stufen 1 und 3. Es ist damit durch einen einfachen Vergleich der Ausführungszeiten – unabhängig von der Zahl und der Art der Lose – möglich, festzulegen, welche Maschine(n) dominiert werden. Mit Hilfe dieser einfachen Ungleichung können drei Fälle betrachtet werden:

i) Die zweite Stufe wird dominiert, wobei $p_1 p_3 > p_2^2$ gilt

Wird die Stufe 2 dominiert und gilt in Formel (26) die strikte „>“-Relation, dann sind die mit $x_j = x_1 q^{j-1}$ und $q = (p_2 + p_3) / (p_1 + p_2)$ ermittelten Lose konsistent und bilden eine geometrische Folge (vgl. Trietsch/Baker, 1993). Der Plan ist auf der zweiten Stufe wartezeitfrei, erlaubt aber keine leerzeitfreie Nutzung dieser Stufe. Wird beispielsweise $p_1 = 8$; $p_2 = 4$ und $p_3 = 3$ gesetzt und sollen $Q = 40$ Stück mit $J = 3$ Losen bearbeitet werden, ergibt sich folgende Lösung: $x_1 = 20{,}794$; $x_2 = 12{,}129$; $x_3 = 7{,}0762$ mit einer Durchlaufzeit von 369,53 ZE. Das folgende Gantt-Diagramm zeigt, dass tatsächlich die Stufe 2 nicht leerzeitfrei genutzt werden kann, wenn wartezeitfreie Bearbeitung gefordert wird.

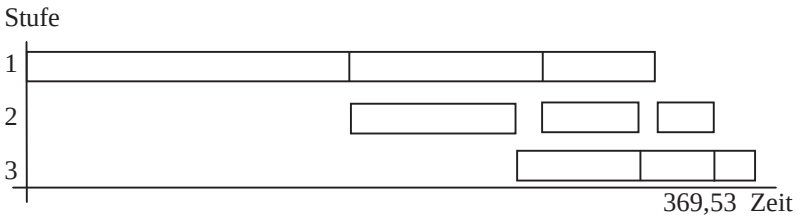


Abbildung 20: Optimale Lösung mit Leerzeit auf Stufe 2

Für die optimale Losgröße des ersten Loses ergibt sich durch Umstellen von Formel (16) der folgende und für den praktischen Einsatz leicht anwendbare Zusammenhang:

$$x_1 = \frac{Q(1-q)}{(1-q^J)} \tag{27}$$

Im Beispiel erhält man: $x_1 = 40\ (1 - 0{,}5833) / (1 - 0{,}5833^3) = 20{,}794$. Aus dem optimalen Wert für x_1 lassen sich unmittelbar die übrigen Losgrößen ableiten, da für sie nach Formel (15): $x_j = x_1 q^{j-1}$ gelten muss. Neben den optimalen Losgrößen kann auch die Durchlaufzeit mittels einer einfachen Formel bestimmt werden. Zu ihrer Herleitung wird die Weitergabe des ersten Teilloses über die Stufen betrachtet und die unterbrechungsfreie Bearbeitung auf der dritten Stufe (Folgerung 6) herangezogen. In einer optimalen Lösung ist ausgeschlossen, dass das erste Los auf einer der Stufen auf seine Weiterbearbeitung wartet. Nach Folgerung 6 ist bekannt, dass die optimale Lösung nicht ausgeschlossen wird, wenn auf den Randstufen Leerzeitfreiheit gefordert ist. In der Konsequenz gilt, dass sich in der optimalen Lösung die minimale Durchlaufzeit als Summe der Ausführungszeit des ersten Loses auf der ersten und der zweiten Stufe, zuzüglich der Bearbeitung der gesamten Menge auf Stufe 3 ergibt:

$$C_{\min} = (p_1 + p_2)\, x_1 + Q\, p_3 \tag{28}$$

Dieser Zusammenhang kann gut an Hand der Abbildung 20 nachvollzogen werden. Setzt man die Zahlen des Beispiels in diese Formel ein, bestätigt sich die oben sukzessiv ermittelte Durchlaufzeit, so dass der Produktionsplaner schnell die Auswirkungen dieser Aufteilung auf die Fertigstellung des Loses ermitteln kann:

$C_{\min} = (4 + 8)\ 20{,}794 + 40 \cdot 3 = 369{,}53$ ZE

ii) Die zweite Stufe wird dominiert, wobei gilt: $p_1 p_3 = p_2^2$

Die mit $x_j = x_1 q^{j-1}$ und $q = (p_2 + p_3) / (p_1 + p_2)$ ermittelten Lose sind konsistent und der Plan ist sowohl leerzeit- als auch wartezeitfrei. Allerdings kann die Relation $p_1 p_3 = p_2^2$ auf zwei Arten gegeben sein:

- sie kann trivial erfüllt sein, wenn alle Stufen die identische Bearbeitungszeit aufweisen ($p_1 = p_2 = p_3 = p$) und
- sie kann bei unterschiedlichen Bearbeitungszeiten gelten.

Im trivialen Fall ist $q = 1$ und die Lösung mit konstanten Losen der Größe Q/J ist die optimale Lösung. Der Ausdruck zur Bestimmung der minimalen Durchlaufzeit kann weiter vereinfacht werden, aus Formel (28) erhält man:

$$C_{\min} = \frac{Qp(2+J)}{J} \tag{29}$$

Angenommen, in einer dreistufigen Produktion mit identischer Ausführungszeit p = 7 ZE ist eine Losgröße von 400 Stück zu bearbeiten. Weiterhin hat der Produktionsplaner für diese Menge einen Liefertermin einzuhalten, der eine Durchlaufzeit von maximal 3150 ZE verlangt. Für diese Setzung kann er die Anzahl der gleichgroßen Lose direkt aus Formel (30) ermitteln, indem er $C_{\min} = 3150$ setzt und J gegebenenfalls aufrundet.

$$J = \frac{2Qp}{C_{\min} - Qp} \tag{30}$$

Es zeigt sich, dass der Liefertermin mit 16 konstanten Losen der Größe 25 zu realisieren ist.

In dem nicht trivialen Fall gilt $p_1 p_3 = p_2^2$, ohne dass die Ausführungszeiten auf den Stufen gleich sein müssen. Beispielsweise erfüllen $p_1 = 2$; $p_2 = 6$ und $p_3 = 18$ die Formel (24) mit „=“. Für diese Setzung ergibt sich $q = 3$. Angenommen, es sollen $Q = 40$ Stück mit 3 Losen bearbeitet werden, haben die optimalen Lose die Größen: $x_1 = 3.0769$; $x_2 = 9{,}3408$ und $x_3 = 27{,}6923$.

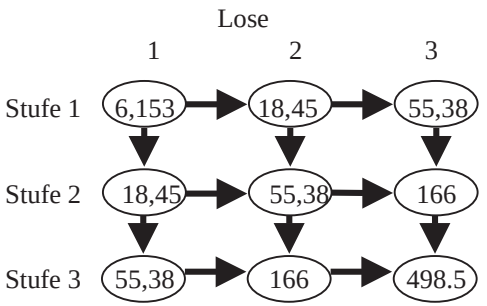


Abbildung 21: Eine warte- und leerzeitfreie Lösung mit konsistenten Losen

Die Darstellung als Netzwerk zeigt, dass diese Lösung tatsächlich leer- und wartezeitfrei ist, da jeder Weg durch das Netz 744 ZE lang ist. Zur Bestimmung der einzelnen Losgrößen und der minimalen Durchlaufzeit kann auf die Formeln (27) und (28) zurückgegriffen werden.

iii) Die Stufe 2 dominiert

Ergibt die Relation hingegen, dass $p_1 p_3 < p_2^2$ gilt, sind nicht nur die Stufen 1 und 3 sondern es ist auch die Stufe 2 leerzeit- und wartezeitfrei zu nutzen. Die optimalen Losgrößen lassen sich durch das stufenweise Anwenden der Formel (16) ermitteln. Die optimalen Losgrößen variieren dann allerdings über die Stufen. Ein Beispiel für diesen letzten Fall wurde am Anfang dieses Kapitels in Abbildung 17 gezeigt. Setzt man zur Kontrolle die im Beispiel verwendeten Bearbeitungszeiten in Formel (26) ein, gilt $p_1 p_3 < p_2^2$, da $\underline{p} = (3; 5; 7)$. Es zeigt sich, dass in diesem Beispiel tatsächlich der Fall *iii)* eintritt, womit die Struktur

der Lösung in Abbildung 17 bestätigt ist. Zur Bestimmung der optimalen Losgrößen und der Durchlaufzeit ist Folgerung 14 und Formel (22) anzuwenden.

Wie die Fallunterscheidung zeigt, ergibt sich für den Produktionsplaner im Fall einer dominierten zweiten Stufe keine Einschränkung, wenn er konsistente Lose fordert, da es immer eine optimale Lösung gibt, die diese Struktur hat. Ist hingegen die Stufe 2 dominant, dann führt die Verwendung von variablen Losen zu kürzeren Durchlaufzeiten, als wenn konsistente Lose verlangt werden (Baker/Jia, 1993).

Folgerung 16:

Gilt $p_1p_3 > p_2^2$, existiert eine optimale wartezeitfreie Lösung mit konsistenten Losen: $x_j = x_1 q^{j-1}$ mit $q = (p_2+p_3) / (p_1+p_2)$. Auf der Stufe 2 treten Leerzeiten auf.

Folgerung 17:

Gilt $p_1p_3 = p_2^2$, existiert eine optimale leer- und wartezeitfreie Lösung mit konsistenten Losen: $x_j = x_1 q^{j-1}$ mit $q = (p_2+p_3) / (p_1+p_2)$.

Folgerung 18:

Gilt $p_1p_3 < p_2^2$, existiert unter Stückverfügbarkeit eine optimale leer- und wartezeitfreie Lösung mit variablen Losen, die stufenweise eine geometrische Folge bilden: $y_{im} = Q \cdot q_m^{\,i-1} \Big/ \sum_{i=1}^{I} q_m^{\,i-1}$ mit $q_m = p_{m+1}/p_m$ für $m = 1,..., M\text{-}1$.

Folgerung 19:

Gilt $p_1 = p_2 = p_3$, existiert eine optimale leer- und wartezeitfreie Lösung mit konstanten Losen x = Q / J.

Die Folgerung 19 korrespondiert dabei mit Folgerung 10, die den analogen Sachverhalt für den Zweistufenfall behandelt.

5.2.2.2 Konsistente Lose im Dreistufenfall

In der bisherigen Analyse wurde angenommen, dass Stückverfügbarkeit vorliegt. Allerdings berührt diese Einschränkung die Folgerung 16, 17 und 19 nicht, da in den optimalen Lösungen eine Aufteilung in konsistente oder konstante Lose vorgenommen wird, die auch unter Losverfügbarkeit umsetzbar ist. Lediglich im Fall *iii)* – dort dominiert die Stufe 2 und es gilt ($p_1p_3 < p_2^2$) – werden Lose ermittelt, die stufenweise eine geometrische Folge bilden. Um einen derartigen Plan umsetzen zu können, muss Stückverfügbarkeit gegeben sein. Liegt im Fall *iii)* nur Losverfügbarkeit vor, stellt sich die Frage, wie dann die optimalen Losgrößen ermittelt werden sollen. Angesprochen ist somit die

Problemstellung P(0), und folglich kann auf Folgerung 2 zurückgegriffen werden. Nach dieser existiert eine optimale durchlaufzeitminimale Lösung im Dreistufenfall für P(0), die mit konsistenten Losen erreichbar ist. Allerdings ist damit nicht gesagt, wie diese Losgrößen zu ermitteln sind.

Zur Beantwortung dieser Frage können die Ergebnisse einer Untersuchung der strukturellen Eigenschaften von Lot Streaming Problemen mit konsistenten Losen im dreistufigen Flow Shop (Glass et al., 1994) genutzt werden. Dort zeigen die Autoren, dass es einen einfachen Suchalgorithmus gibt, dessen Rechenaufwand polynomial durch O(log *J*) mit *J* als Zahl der Lose beschränkt ist. Da die Alternative zur Ermittlung der optimalen konsistenten Lose bei Losverfügbarkeit in der Lösung eines linearen Programms liegt, soll dieser für den praktischen Einsatz sinnvolle Such-Algorithmus im Folgenden vorgestellt werden.

Die Idee dieses Verfahrens wird nachvollziehbar, wenn man sich zunächst den Verlauf der kritischen Pfade in den drei Fällen: $p_1p_3 = p_2^2$; $p_1p_3 > p_2^2$ und $p_1p_3 < p_2^2$ vor Augen führt. Wie Glass et al. (1994) beweisen, sind im Fall $p_1p_3 = p_2^2$ alle Pfade durch das Netzwerk kritisch. Auf diese Eigenschaft wurde bei der Herleitung von Folgerung 17 und in Abbildung 21 eingegangen, so dass hier die beiden übrigen Fälle in den Vordergrund treten. Wenn Stufe 2 dominiert wird ($p_1p_3 > p_2^2$), nehmen im Netzwerk die kritischen Pfade den in der folgenden Abbildung fett hervorgehobenen Verlauf an:

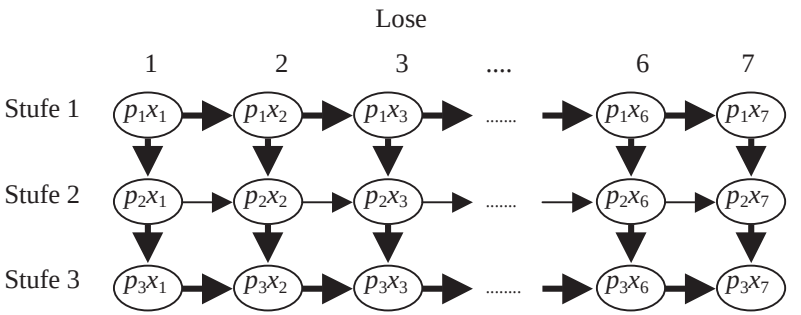


Abbildung 22: Kritische Pfade im Netzwerk, wenn Stufe 2 dominiert wird

Es fällt auf, dass auf der dominierten Stufe die horizontal verlaufenden Kanten alle nicht kritisch sind. Wenn hingegen die Stufe 2 dominiert ($p_1p_3 < p_2^2$), erhält man ein Netz, bei dem sich ein Los als sogenanntes „Übergangslos“ auszeichnet. In der folgenden Abbildung ist das Los 4 das Übergangslos. Bis zum Los 4 sind alle Kanten auf den Stufen 1 und 2 kritisch, ab Los 4 sind sie es auf den Stufen 2 und 3. Weiterhin sind alle horizontalen Kanten auf der Stufe 2 kritisch.

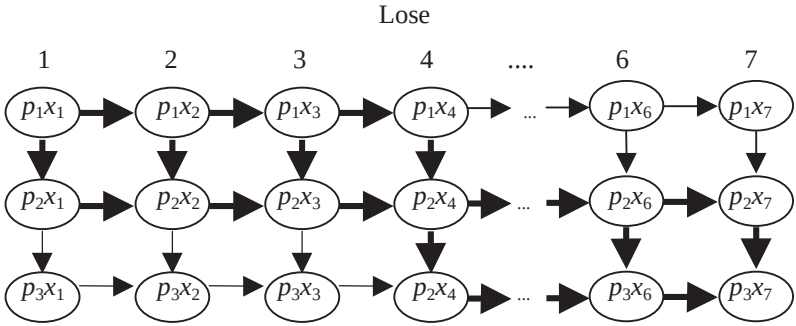


Abbildung 23: Kritische Pfade im Netzwerk, wenn Stufe 2 dominiert

Es sind damit bis zu dem Übergangslos *k* die Stufen 1 und 2, die die Größe der ersten *k* Lose determinieren; ab dem Los *k* bis zum Los *J* sind es hingegen die Stufen 2 und 3. In dem Los *k* liegt die Besonderheit, dass es sowohl im Bezug auf die Stufen 1 und 2 als auch im Bezug auf die Stufen 2 und 3 zum kritischen Los wird.

Die Idee des Suchalgorithmus besteht darin, versuchsweise ein Los *k* als Übergangslos festzulegen und zu prüfen, ob eine Verlagerung des Übergangsloses zu einer Reduktion der Durchlaufzeit führen kann. Wie Glass et al. (1994) zeigen, nimmt die minimale Durchlaufzeit in Abhängigkeit der Lage des Übergangsloses *k* einen unimodalen Verlauf an. Ergibt sich beim Wechsel des Übergangsloses von der Position *k* zu der Position *k* + 1 eine Erhöhung der Durchlaufzeit, müssen Strukturen, die sich für ein Übergangslos auf den Positionen *k* +2, *k* + 3,..., *J* ergeben, nicht mehr untersucht werden. Analog ist zu verfahren, wenn der Wechsel vom Übergangslos *k* zu *k* - 1 eine Verschlechterung der Durchlaufzeit bedeutet; dann muss das Übergangslos nicht in Richtung des ersten Loses verschoben werden. Folglich genügt eine – in Hinblick auf die Verlagerung des Übergangsloses – lokale Untersuchung, um die optimale Lage des Übergangsloses zu bestimmen.

Die folgenden Ausführungen sind so aufgebaut, dass zunächst die optimale Lösung für ein größeres Beispiel vorgegeben wird. Daran können einige Zusammenhänge, wie z.B. die Bedeutung des Übergangsloses, weiter verdeutlicht werden. Im Anschluss wird der Ablauf des Algorithmus zur Lösung dieses Beispiels vorgeführt und an Hand der vorgenommenen Berechungen die Logik der Vorgehensweise nachvollzogen.

Betrachtet wird ein Problem mit den folgenden Daten: $Q = 60$; $J = 7$ und $\underline{p} = (4; 8; 7)$. Die durchlaufzeitminimale Lösung des Problems erhält man, wenn man die folgenden Losgrößen setzt:

Tabelle 4: **Optimale konsistente Lose bei dominanter Stufe 2**

Los	Losgröße	x_{j+1}/x_j
1	3,228	
		2
2	6,456	
		2
3	12,912	
		0,875
4	11,298	
		0,875
5	9,885	
		0,875
6	8,650	
		0,875
7	7,568	

Es fällt auf, dass die konsistenten Lose 1, 2 und 3 eine geometrische Reihe bilden, bei der sich wieder der Zusammenhang $x_j = q\, x_{j-1}$ zeigt: Die Verhältnisse der Losgrößen untereinander: x_{j+1}/x_j sind so gewählt, dass sie dem Verhältnis der Ausführungszeiten der ersten beiden Stufen: q_1= 2 entsprechen, da p_1 = 8 und p_2 = 4 gilt. Ebenso bilden die Lose 3, 4, 5, 6 und 7 eine geometrische Reihe, bei der dieser Zusammenhang für das Verhältnis der Ausführungszeiten der Stufen 2 und 3 mit q_2 = 0,875 gilt. Da das Los 3 mit x_3 = 12,912 in beiden geometrischen Reihen auftritt, ist es in diesem Beispiel das Übergangslos *k*. Formal gilt dabei (vgl. Glass et al. 1994):

$$p_2\, x_j = p_1\, x_{j+1} \text{ für } 1 \le j \le k\text{-}1 \quad \text{und} \quad p_2\, x_j = p_3\, x_{j-1} \text{ für } k\text{+}1 \le j \le J \tag{31}$$

Folgerung 20:

Bei dominanter Stufe 2 existiert im Dreistufenfall unter Losverfügbarkeit ein Übergangslos *k* für das gilt: $p_2\, x_j = p_1\, x_{j+1}$ für $1 \le j \le k$-1 und $p_2\, x_j = p_3\, x_{j-1}$ für k+1 $\le j \le J$.

Stellt man diesen Zusammenhang um, indem man jeweils nach x_j auflöst, gilt zusammengefasst Formel (32), in der die übrigen Losgrößen auf die Größe des Loses *k* zurückgeführt werden:

$$x_j = \begin{cases} \left(\frac{p_1}{p_2}\right)^{k-j} x_k & \text{für } 1 \le j \le k \\ \left(\frac{p_3}{p_2}\right)^{j-k} x_k & \text{für } k \le j \le J \end{cases} \tag{32}$$

Es fällt auf, dass in Formel (32) im oberen Term der Quotient aus $p_1/p_2 = q_1$ und im unteren Term mit p_3/p_2 der Kehrwert von $q_2 = p_2/p_3$ auftritt. Werden die mit Formel (32) ermittelten Losgrößen addiert, erfasst man die Losgröße x_k doppelt, da sie im oberen und im unteren Term erfasst wird. Aus Formel (16) ist bekannt, dass sich die Lose zu der vorgegebenen Losgröße Q addieren müssen:

$$Q = \sum_{j=1}^{J} x_j$$

Daher wird in Formel (33) einmal das Los k subtrahiert. Man erhält:

$$\sum_{j=1}^{k} \left(\frac{p_1}{p_2}\right)^{k-j} x_k + \sum_{j=k}^{J} \left(\frac{p_3}{p_2}\right)^{j-k} x_k - x_k = Q \tag{33}$$

In einer durchlaufzeitminimalen Lösung mit konsistenten Losen kann kein Los die Größe Null aufweisen. Daher ist es zulässig, den Ausdruck weiter zu vereinfachen:

$$\sum_{j=1}^{k} \left(\frac{p_1}{p_2}\right)^{k-j} + \sum_{j=k}^{J} \left(\frac{p_3}{p_2}\right)^{j-k} - 1 = \frac{Q}{x_k}$$

Nach Vereinheitlichen der Summationsgrenzen ergibt sich:

$$\sum_{j=0}^{k-1} \left(\frac{p_1}{p_2}\right)^{j} + \sum_{j=0}^{J-k} \left(\frac{p_3}{p_2}\right)^{j} - 1 = \frac{Q}{x_k}$$

Daraus folgt:

$$x_k = Q \Big/ \left(\sum_{j=0}^{k-1} \left(\frac{p_1}{p_2}\right)^{j} + \sum_{j=0}^{J-k} \left(\frac{p_3}{p_2}\right)^{j} - 1 \right) \tag{34}$$

Bei gegebener Position des Übergangsloses k kann seine Größe einfach bestimmt werden. Im oben gezeigten Beispiel ergibt sich, dass das Übergangslos das Los 3 ist und es eine Größe von 12,912 hat, wenn $Q = 60$; $J = 7$ und $\underline{p} = (4; 8; 7)$ gegeben sind. Das Einsetzen dieser Werte in Formel (34) liefert für den Nenner:

1 + 0,5 + 0,25 + 1 + 0,875 + 0,7656 + 0,6699 + 0,5861 – 1 = 4,6466.
Insgesamt ergibt sich die gesuchte Losgröße von 60 / 4,6466 = 12,912.

Betrachtet man die Berechnung, fällt auf, dass sich die Betrachtung auf die Länge der beiden geometrischen Folgen reduzieren lässt. Die Anzahl der Summanden in der ersten Summe in Formel (34) entspricht der Länge der geometrischen Folge, in der die Losgrößen gemäß des Verhältnisses der Ausführungszeiten der ersten zur zweiten Stufe festgelegt werden. Es sind k-1 Übergänge zu berücksichtigen. Die übrigen Losgrößen bilden ebenfalls eine geometrische Folge, entsprechend der Relation von p_3/p_2. Für die zweite Folge sind J-k Übergänge zu berücksichtigen. Inhaltlich relevant ist die Zahl dieser Übergänge, da sie das Ausmaß der Überlappung und damit die Durchlaufzeitverkürzung bestimmen. Prinzipiell ist mit Formel (34) die Bestimmung der optimalen Losgrößen für den Fall der dominanten Stufe 2 abgeschlossen, sofern nicht $p_2 = p_1$ oder $p_3 = p_2$ gilt. Der Produktionsplaner kann versuchsweise ein Los k als Übergangslos vorgeben, mit Formel (34) seine Größe bestimmen und die übrigen Lose mit Formel (32) so festlegen, dass sich die gewünschten geometrischen Folgen einstellen. Danach wird ein anderes Los k zum Übergangslos bestimmt und der Vergleich der Durchlaufzeiten zeigt, ob sich die Lösung verbessert hat. Allerdings kann diese rechenaufwändige Bestimmung des optimalen Übergangsloses noch erheblich vereinfacht werden.

Wie Glass et al. (1994) zeigen, genügt es, das Vorzeichen der Veränderung zu bestimmen, die die Durchlaufzeit erfährt, wenn anstatt des Loses k das Los k+1 zum Übergangslos wird. Diese Veränderung wird mit $\Delta(k)$ bezeichnet und kann für ein beliebiges k mit $1 \le k \le J$-1 an Hand der folgenden Formel bestimmt werden:

$$\Delta(k) = \begin{cases} \left(\left(\frac{p_2}{p_1}\right)^{k+1} - 1\right) \Big/ \left(\frac{p_2}{p_1} - 1\right) - \left(\left(\frac{p_2}{p_3}\right)^{J-k+1} - 1\right) \Big/ \left(\frac{p_2}{p_3} - 1\right) & \text{für } p_1 \neq p_2 \neq p_3 \\ k + 1 - \left(\left(\frac{p_2}{p_3}\right)^{J-k+1} - 1\right) \Big/ \left(\frac{p_2}{p_3} - 1\right) & \text{für } p_1 = p_2 \neq p_3 \\ \left(\left(\frac{p_2}{p_1}\right)^{k+1} - 1\right) \Big/ \left(\frac{p_2}{p_1} - 1\right) - (J - k + 1) & \text{für } p_1 \neq p_2 = p_3 \end{cases} \tag{35}$$

Findet man für die Position k heraus, dass das Vorzeichen von $\Delta(k)$ positiv ist, dann führt die Verlagerung des Übergangsloses von Position k auf die Position k+1 zu einer Verlängerung der Durchlaufzeit. Ist $\Delta(k)$ hingegen negativ, dann verringert sich die Durchlaufzeit, wenn das Übergangslos verlagert wird. Gesucht ist somit das Los k, für das $\Delta(k) \le 0$ und $\Delta(k+1) \ge 0$ ist. Sollte $\Delta(k) = 0$ sein, verändert sich die Durchlaufzeit nicht, wenn das Übergangslos verlagert wird.

Inhaltlich wird in Formel (35) die Konsequenz bestimmt, die sich aus der Verlagerung des Übergangsloses für die Durchlaufzeit ergibt. Wenn das Übergangslos nicht mehr an der Position k, sondern an der Position k+1 auftritt, dann

verlängert sich die geometrische Reihe zwischen den Stufen 1 und 2, dafür verkürzt sich die geometrische Reihe zwischen den Stufen 2 und 3. Um alle Fälle abzudecken, enthält Formel (35) eine Fallunterscheidung, die sich auf die Größenverhältnisse der Ausführungszeiten bezieht. Sind alle Ausführungszeiten unterschiedlich groß, muss die Position des Übergangsloses unter Berücksichtigung der beiden Relationen p_2/p_1 und p_2/p_3 ermittelt werden. Stimmt die Ausführungszeit auf der ersten Stufe hingegen mit der auf der zweiten überein, ergeben sich in den optimalen Lösungen für die ersten k Lose konstante Losgrößen. Dementsprechend vereinfacht sich die Bestimmung des Effekts, den eine Verschiebung des Übergangsloses vom Los k auf die Position k+1 hat. Analog treten bei $p_2 = p_3$ ab dem Los k konstante Lose auf.

Wendet man Formel (35) für das Beispiel mit $Q = 60$; $J = 7$ und $\underline{p} = (4; 8; 7)$ an und legt als Übergangslos versuchsweise das Los $k = 4$ fest, erhält man:

$$\Delta(4) = ((2^{5} - 1) / (2 - 1)) - ((1{,}1428^{4} - 1) / (1{,}1428 - 1)) = 26{,}058$$

Die Durchlaufzeit wird größer, wenn man statt dem vierten das fünfte Los zum Übergangslos macht. Daher wird $k = 3$ betrachtet. Die Berechnung ergibt:

$$\Delta(3) = ((2^{4} - 1) / (2 - 1)) - ((1{,}1428^{5} - 1) / (1{,}1428 - 1)) = 8{,}353$$

Die Durchlaufzeit verlängert sich ebenfalls, wenn statt dem dritten das vierte Los zum Übergangslos wird. Daher wird $k = 2$ gesetzt:

$$\Delta(2) = ((2^{3} - 1) / (2 - 1)) - ((1{,}1428^{6} - 1) / (1{,}1428 - 1)) = -1{,}596$$

Da $\Delta(2)$ negativ ist, reduziert die Verlagerung des Übergangsloses von der Position 2 auf die Position 3 die Durchlaufzeit. Das optimale Übergangslos ist somit das dritte Los. Wie oben gezeigt, liefert Formel (34) die gesuchte Größe von $x_3 = 12{,}912$. Mit Hilfe von Formel (32)

$$x_j = \begin{cases} \left(\frac{p_1}{p_2}\right)^{k-j} x_k & \text{für } 1 \le j \le k \\ \left(\frac{p_3}{p_2}\right)^{j-k} x_k & \text{für } k \le j \le J \end{cases} \tag{32}$$

kann die Größe der übrigen Lose für das Beispiel mit $\underline{p} = (4; 8; 7)$ direkt aus der Größe des Übergangsloses x_3 bestimmt werden:

$x_1 = (4/8)^2 \cdot 12{,}912 = 3{,}228$; $\quad x_2 = (4/8)^1 \cdot 12{,}912 = 6{,}456$; etc.

$x_4 = (7/8)^1 \cdot 12{,}912 = 11{,}298$; $\quad x_5 = (7/8)^2 \cdot 12{,}912 = 9{,}885$; etc.

Glass et al. (1994) haben mit diesem Vorgehen ein Verfahren vorgestellt, das mit einfachen Mitteln erlaubt, die optimale Größe der konsistenten Lose festzulegen, wenn die zweite Stufe dominiert und Losverfügbarkeit gilt. Es kann eine beliebige Zahl an Losen berücksichtigt werden, ohne dass der Rechenauf-

wand überproportional anwächst. Die Bestimmung der optimalen Losgrößen gestaltet sich so einfach, dass sie manuell durchführbar ist. Fallen diese Berechnungen häufiger an, können sie problemlos mit einer Tabellenkalkulation unterstützt vorgenommen werden.

Folgerung 21:

Bei dominanter Stufe 2 und unterschiedlichen Bearbeitungszeiten im Dreistufenfall bilden unter Losverfügbarkeit die optimalen konsistenten Losgrößen zwei geometrische Reihen, die im Los k ineinander übergehen.

Folgerung 22:

Sind unter Losverfügbarkeit bei dominanter Stufe 2 im Dreistufenfall die Bearbeitungszeiten der ersten und der zweiten Stufe gleich, bilden die optimalen konsistenten Losgrößen k bis J eine geometrische Reihe. Die ersten k Lose sind konstant.

Folgerung 23:

Sind unter Losverfügbarkeit bei dominanter Stufe 2 im Dreistufenfall die Bearbeitungszeiten der zweiten und der dritten Stufe gleich, bilden bis zum Los k die optimalen konsistenten Losgrößen eine geometrische Reihe, die restlichen Lose sind konstant.

Ist das Übergangslos k bestimmt, kann die Durchlaufzeit $C_{\min}$ wie folgt berechnet werden. Es wird ausgenutzt, dass die dominante Stufe 2 durchgängig genutzt wird und sich deshalb die minimale Durchlaufzeit als Summe der Bearbeitung des ersten Loses auf der ersten Stufe, der leerzeitfreien Bearbeitung der Q Stück auf der zweiten Stufe und der Bearbeitung des letzten Loses auf Stufe 3 ergibt:

$$C_{\min} = p_1 x_1 + Q p_2 + p_3 x_J$$

Aus diesem Zusammenhang können mit Hilfe von Formel (32) die Losgrößen x_1 und x_J auf die Losgröße x_k zurückgeführt werden. Es gilt:

$$C_{\min} = Q p_2 + p_2 x_k \left(\left(\frac{p_1}{p_2} \right)^k + \left(\frac{p_3}{p_2} \right)^{J-k+1} \right) \tag{36}$$

Im Beispiel erhält man: $60 \cdot 8 + 8 \cdot 12{,}912 \cdot ((4/8)^3 + (7/8)^5) = 545{,}893$ ZE. Der Produktionsplaner kann somit – ohne die einzelnen Losgrößen explizit bestimmen zu müssen – direkt ermitteln, wie stark sich die Durchlaufzeit beim Einsatz von J Losen verkürzen lässt.

5.2.2.3 Diskrete Lose im Dreistufenfall

Sollen im Dreistufenfall diskrete statt kontinuierliche Losgrößen festgelegt werden, kann auf bereits vorgestellte Verfahren und Formeln zurückgegriffen werden (vgl. Baker, 1998). Wie im kontinuierlichen Fall hängt die Vorgehensweise wesentlich davon ab, ob die zweite Stufe dominiert wird oder nicht.

i) Stufe 2 wird dominiert ($p_1 p_3 \geq p_2^2$)

In diesem Fall wird das für den Zweistufenfall vorstellte Suchverfahren (Abschnitt 5.2.1.2) auf den Dreistufenfall erweitert. Den Ausgangpunkt bildet wieder die Überlegung, dass sich durch die zusätzliche Berücksichtigung der Ganzzahligkeit $C_{\min}$ nicht reduzieren wird. Insofern kann die für kontinuierliche Losgrößen ermittelte Durchlaufzeit $C^{K}_{\min}$ als Schranke für $C^{D}_{\min}$ genutzt werden. Hinsichtlich des Startzeitpunkts des j-ten Loses auf der dritten Stufe (er ist in Formel (37) durch $C_{j3} - p_3 x_j$ gegeben), muss die folgende Beziehung gelten:

$$p_1 X_j + p_2 x_j \leq C_{j3} - p_3 x_j = C^{D}_{\min} - p_3 (Q - X_{j-1}) \tag{37}$$

Sie gibt im Wesentlichen den für den Zweistufenfall bereits vorgestellten Zusammenhang wieder, allerdings mit der Ergänzung, dass die über die ersten j Lose kumulierte Menge X_j auf der Stufe 1 und der Stufe 2 bearbeitet werden muss. Da $x_j = X_j - X_{j-1}$ folgt aus (37):

$$p_1 X_j + p_2 (X_j - X_{j-1}) \leq C^{D}_{\min} - p_3 (Q - X_{j-1})$$

woraus sich die folgende Ungleichung ergibt:

$$X_j \leq [C^{D}_{\min} + p_2 X_{j-1} - p_3 (Q - X_{j-1})] / (p_1 + p_2) \tag{38}$$

Aus (38) erhält man analog zur Formel (21) einen rekursiven Zusammenhang zwischen X_j und X_{j-1}. Wieder ist X_j so groß zu wählen, wie die größte ganze Zahl, die (38) erfüllt. Als Startwert für $C^{D}_{\min}$ wird – wie im Zweistufenfall – $C^{K}_{\min}$ genutzt. Sind alle Ausführungszeiten ganzzahlig, kann $C^{D}_{\min}$ auf den nächsten ganzzahligen Wert aufgerundet werden. Die sukzessive Berechnung der kumulierten Losgröße ist dann abzubrechen, wenn $X_J = Q$ oder $X_J < Q$ wird.

Im ersten Fall ($X_J = Q$) ist die optimale Lösung gefunden. Die aus den X_j ermittelbaren Losgrößen x_j stellen eine optimale diskrete Lösung des Problems dar, die Durchlaufzeit entspricht dem Wert $C^{D}_{\min}$.

Im zweiten Fall ($X_J < Q$) wurde $C^{D}_{\min}$ zu klein gewählt. Bei rein ganzzahligen Ausführungszeiten wird $C^{D}_{\min} = C^{D}_{\min} + 1$ gesetzt und das Verfahren erneut durchgeführt. Sollten die Ausführungszeiten nicht ganzzahlig sein, ist (wie im Zweistufenfall) der kleinste Bruchteil zu dem bisherigen Wert $C^{D}_{\min}$ zu addieren, der zum Aufrunden der mittels Formel (38) festgelegten X_j auf die nächste ganze Zahl erforderlich wäre (vgl. Baker, 1998).

Die Vorgehensweise wird nun an Hand des Beispiels demonstriert, das für den kontinuierlichen Fall im Abschnitt 5.2.2.1 (vgl. Abbildung 20) bereits eingeführt wurde. Für dieses sind die folgenden Daten gegeben: $p_1 = 8$; $p_2 = 4$; $p_3 = 3$; $Q = 40$. Für $J = 3$ wurde $C^K_{min} = 369{,}53$ ermittelt. Wegen der ganzzahligen Ausführungszeiten $\underline{p} = (8; 4; 3)$ folgt $C^D_{min} = 370$. Es wird nun sukzessiv Formel (38) verwendet.

$X_1 \leq [370 + 4 \cdot 0 - 3\ (40\text{-}0)] / 12 = 20{,}8$ => Abrunden liefert: $X_1 = 20$

$X_2 \leq [370 + 4 \cdot 20 - 3\ (40\text{-}20)] / 12 = 32{,}5$ => Abrunden liefert: $X_2 = 32$

$X_3 \leq [370 + 4 \cdot 32 - 3\ (40\text{-}32)] / 12 = 39{,}5$ => Abrunden liefert: $X_3 = 39$

Da $X_3 < Q$ gilt ist der Startwert $C^D_{min} = 370$ zu klein gewählt. Als nächster Wert wird $C^D_{min} = 371$ geprüft und schließlich $C^D_{min} = 372$ vorgegeben. Für diesen Wert erhält man die folgenden Berechungen:

$X_1 \leq [372 + 4 \cdot 0 - 3\ (40\text{-}0)] / 12 = 21$ => $X_1 = 21$

$X_2 \leq [370 + 4 \cdot 21 - 3\ (40\text{-}21)] / 12 = 33{,}08$ => Abrunden liefert: $X_2 = 33$

$X_3 \leq [370 + 4 \cdot 33 - 3\ (40\text{-}33)] / 12 = 40{,}08$ => Abrunden liefert: $X_3 = 40$

Aus den gegebenen kumulierten Losgrößen lassen sich die optimalen ganzzahligen Losgrößen bestimmen: $\underline{x} = (21; 12; 7)$. Das folgende Gantt-Diagramm stellt diese Lösung dar.

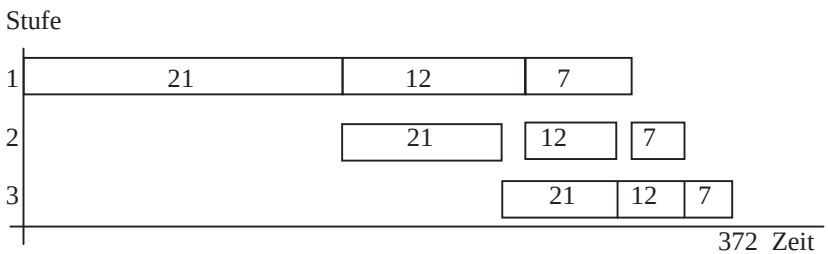


Abbildung 24: Optimale diskrete Lösung mit Leerzeit auf Stufe 2

ii) Stufe 2 dominiert ($p_1 p_3 < p_2^2$)

In diesem Fall kann das Problem – sofern Stückverfügbarkeit gegeben ist – in zwei aufeinanderfolgende und separat zu lösende Probleme dekomponiert und mit Hilfe des Verfahrens für den Zweistufenfall gelöst werden (vgl. Trietsch/Baker 1993).

Zur Verdeutlichung der Vorgehensweise wird das Beispiel aus Abschnitt 5.2.2.1. betrachtet. $Q = 60$; $M = 3$; $p_1 = 3$; $p_2 = 5$; $p_3 = 7$ und $J = 4$ Teillosen. Als kontinuierliche Lösung wurde $x_{11} = 5{,}959$; $x_{21} = 9{,}926$; $x_{31} = 16{,}544$; $x_{41} = 27{,}573$ ermittelt. Die Transferlose von Stufe 2 zur Stufe 3 sowie die Frei-

gabelose auf Stufe 3 haben die Größen: x_{13} = 8,445; x_{23} = 11,824; x_{33} = 16,544; x_{43} = 23,175. Die Durchlaufzeit im kontinuierlichen Fall beträgt 480,1 ZE.

Die optimale diskrete Lösung für den Zweistufenfall mit p = (3; 5) ergibt sich für $\underline{x}$ = (6; 10; 17; 27) und für die Stufen 3 und 4 mit $\underline{p}$ = (5; 7) erhält man $\underline{x}$ = (9; 12; 16; 20). Zur Berechnung der Durchlaufzeit im diskreten Fall kann ausgenutzt werden, dass die Lösung auf allen drei Stufen leerzeitfrei sein muss, da Stufe 2 dominiert. Die Durchlaufzeit ergibt sich aus der Zeitspanne zur Fertigstellung des ersten Loses auf der ersten Stufe, der Bearbeitungszeit für das erste Transferlos der zweiten Stufe und der gesamten Losgröße auf der dritten Stufe. Für das Beispiel: $3 \cdot 6 + 9 \cdot 5 + 60 \cdot 7 = 483$ ZE. Dieser Zusammenhang ist in der folgenden Abbildung durch Schraffieren hervorgehoben.

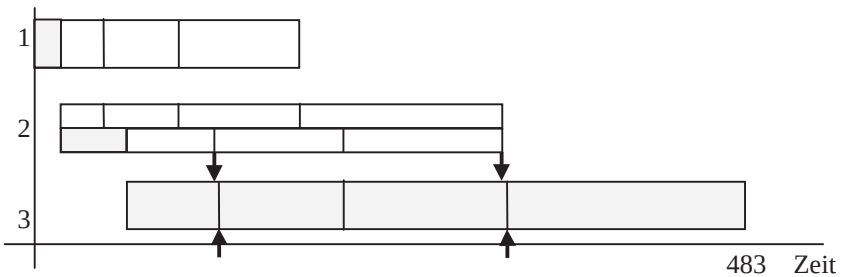


Abbildung 25: Leerzeitfreie, optimale Lösung mit diskreten Losen, Stückverfügbarkeit und Wartezeiten

Die in der Abbildung 25 zusätzlich eingetragenen vertikal verlaufenden Pfeile zeigen, dass durch die Forderung nach diskreten Losgrößen die Wartezeitfreiheit des Plans verloren gehen kann, während sie im kontinuierlichen Fall immer gegeben ist.

Fasst man die Ergebnisse für den Zwei- und den Dreistufenfall zusammen, stehen für die Losüberlappungsprobleme bei kontinuierlichen und diskreten Losen sowohl unter Stück- als auch unter Losverfügbarkeit einfach zu handhabende Formeln bzw. ein Suchverfahren mit polyonomial begrenztem Aufwand zur Verfügung. Die Ansätze eignen sich sehr gut, den Produktionsplanern die nachträgliche Anpassung von Losgrößen mit Hilfe der Losüberlappung zugänglich(er) zu machen.

5.3 Losüberlappung mit konsistenten Losen im Mehrstufenfall

In diesem zweiten Schwerpunkt des fünften Kapitels wendet sich die Untersuchung den konsistenten Losen zu. Die besondere Betonung, die dieser Losart damit zukommt, ist unter anderem aus den folgenden Gründen geboten:

Zur Umsetzung eines Plans mit konsistenten Losen muss lediglich die Losverfügbarkeit der Teile verlangt werden. Die Verfahren zur Bestimmung konsistenter Lose können somit unabhängig von den Restriktionen hinsichtlich der Form der Verfügbarkeit verwendet werden, d.h. konsistente Lose sind auch unter Stückverfügbarkeit zulässig. Ihr Anwendungsbereich ist entsprechend groß.

Die bisher aufgearbeiteten Ergebnisse haben für den Zwei- und den Dreistufenfall gezeigt, dass eine Einschränkung auf konsistente Lose die optimale Lösung nicht ausschließt. Für den Mehrstufenfall gilt die abgeschwächte Eigenschaft, dass die minimale Durchlaufzeit erreichbar ist, wenn die Losgrößen auf den ersten beiden und auf den letzten beiden Stufen konsistent sind. Konsistente Lose und optimale Lösungen schließen sich daher in vielen Fällen nicht aus; eine nähere Betrachtung für den Mehrstufenfall ist schon allein deshalb geboten.

Auch wenn mit variablen Losen (als nicht restringierter Losart) die größte mögliche Reduktion der Durchlaufzeit bei Losüberlappung möglich wird, stellen die konsistenten Lose einen sinnvollen Kompromiss dar. Hinsichtlich der Kriterien Rechenbarkeit und Erweiterungsfähigkeit der Modelle, Umsetzbarkeit der Pläne und der erreichbaren Lösungsgüte sind konsistente Lose für den praktischen Einsatz sehr gut geeignet. So ist beispielsweise die Umsetzung eines Plans mit konsistenten Losen (wegen der Losintegrität) organisatorisch einfacher, als wenn variable Lose verwendet werden. Der Vergleich der Lösungsgüte bei konsistenten gegenüber der bei variablen Losen im Mehrstufenfall kann allerdings erst im dritten Schwerpunkt dieser Arbeit erfolgen. Dort wird ein Verfahren vorgestellt, das mit variablen Losen unter Losverfügbarkeit arbeitet und somit das restliche Potenzial der Losüberlappung zugänglich macht, welches mit konsistenten Losen nicht nutzbar ist. Zuvor wird aber zu zeigen sein, wie vergleichsweise leicht ein lineares Modell zur Ermittlung konsistenter Lose formuliert, interpretiert und erweitert werden kann.

5.3.1 Zwei konsistente Lose im Mehrstufenfall

In diesem Abschnitt wird gezeigt, wie für eine beliebige Stufenzahl die optimale Lösung gefunden werden kann, wenn nur zwei konsistente Lose verwendet werden. Behandelt wird dabei ein rechentechnisch einfaches Vorgehen mit dem z.B. die in Abbildung 26 gezeigte Lösung ermittelbar ist. Das hier verwendete Beispiel geht auf Potts/Baker (1989) zurück: Im 4 Stufenfall wird für $Q = 6$; $\underline{p} = (1; 1; 15; 3)$ eine Lösung mit zwei Losen gesucht. Die optimale Lösung mit konsistenten Losen ($x_1 = 5$; $x_2 = 1$) führt zu einer Durchlaufzeit von 103 ZE:

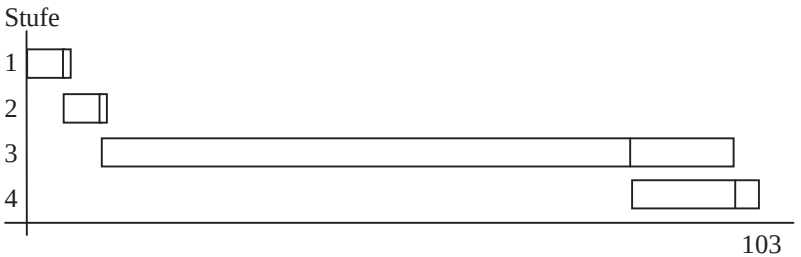


Abbildung 26: Zwei konsistente Lose im Mehrstufenfall

Mit der Problematik, für beliebige Stufenzahl optimale konsistente Losgrößen für den Zweilosfall festzulegen und dabei nicht auf die lineare Programmierung zurückzugreifen, haben sich zunächst Baker/Pyke (1990) befasst. Die zweite Arbeit, in der das Problem der Losüberlappung mit zwei Losen behandelt wird, stammt von Topaloğlu et al. (1994). Das dort vorgeschlagene Modell wurde allerdings primär entwickelt, um den Einfluss von alternativen Zielfunktionen zu untersuchen; eine Variante der Problemstellung, auf die an dieser Stelle nicht näher eingegangen wird. Der von Williams et al. (1997) als Alternative zu Baker/Pyke (1990) präsentierte Ansatz erlaubt zudem eine Erweiterung für den Fall mit drei konsistenten Losen und beliebiger Stufenzahl. Ergänzend stellen Williams et al. eine Heuristik vor, mit der sie Lösungen für beliebige Stufenzahl und beliebige Loszahl generieren. Williams et al. (1997) sehen den wesentlichen Beitrag ihres Ansatzes darin begründet, dass der Rechenaufwand für ihr Verfahren mit $O(m^2)$ polynomial begrenzt ist. Eine ähnliche Motivation findet sich bei Baker/Pyke (1990, S. 476) „we present an algorithm“ .... „that is much more efficient than linear programming“. Diese Aussagen bewertend ist festzuhalten, dass aus heutiger Sicht für die Problemstellungen mit $M > 3$ und $J \geq 3$, der Einsatz der linearen Programmierung die weitaus sinnvollere Alternative darstellt. Es ist zwar zutreffend, dass das Simplex-Verfahren zur Lösung eines linearer Programms im worst case zu exponentiellem Rechenaufwand führt, dies sollte aber kein Motiv sein, Verfahren vorzuschlagen, die zwar mit polynomial begrenztem Aufwand arbeiten, dafür aber nur heuristischen Charakter haben. Sollte man tatsächlich auf Instanzen treffen, bei denen das Simplex-Verfahren ein nicht akzeptables Laufzeitverhalten zeigt, kann man mit den Inneren-Punkt-Verfahren von Khachian oder Karmarkar leicht auf polynomiale Algorithmen zur Lösung linearer Programme zurückreifen (vgl. dazu Kistner, 2003; Dantzig/Thapa, 2003). Außerdem konnte die von Williams et al. (1997, S. 76) angeführte Behauptung, der Algorithmus von Baker/Pyke (1990) liefere bei einigen Instanzen fehlerhafte Ergebnisse, weder nachvollzogen werden, noch haben die Autoren bislang auf entsprechende Anfragen reagiert. Der von Williams et al. (1997) erarbeitete Ansatz wird daher nicht berücksichtigt.

Es ist aus mehreren Gründen ökonomisch sinnvoll, die Losüberlappung mit wenigen Losen zu untersuchen. Einerseits nimmt mit steigender Loszahl J der Koordinationsaufwand und damit die Wahrscheinlichkeit von Fehlern erheblich zu. Andererseits ergibt sich für die Reduktion der Durchlaufzeit bei Variation von J ein ertragsgesetzlicher Verlauf. Das Ausmaß, in dem sich die Durchlaufzeit verkürzt, nimmt mit steigender Zahl an Losen deutlich ab. Die größte Verkürzung ist immer die Veränderung, die mit dem Wechsel von $J = 1$ zu $J = 2$ erreichbar ist (Potts/Baker, 1989). In der betrieblichen Praxis kann es daher ausreichen, ein gegebenes Los in zwei Teile zu zerlegen. Dieser Zusammenhang wird in einer experimentellen Studie von Baker/Jia (1993) bestätigt. Auf einer Datenbasis von 6000 Instanzen wurden für den Dreistufenfall die Durchlaufzeiten bei unterschiedlichen Losarten und variierender Zahl an Losen ($J \leq 10$) verglichen. Ein wichtiges und für alle Losarten durchgängig auftretendes Ergebnis ist, dass sich im Mittel die Hälfte der mit 10 Losen erreichbaren Durchlaufzeitverkürzung bereits beim Wechsel von einem zu zwei Losen realisiert. Insofern ist es gerechtfertigt, die Losüberlappung mit zwei Losen genauer zu untersuchen.

Werden nur zwei konsistente Lose betrachtet und zur Vereinfachung die Losgröße $Q = 1$ gesetzt, kann der Stufenindex und der Losindex entfallen. Anstatt in x_1 und x_2 zu unterscheiden, genügt es, mit $x = x_1$ die Größe des ersten Loses zu modellieren, da sich x_2 aus $x_2 = 1 - x$ ergibt. An Hand der Netzwerkdarstellung des Problems wird die Vereinfachung deutlich. Die Bearbeitung des Loses beginnt im Knoten (1,1) und ist im Knoten (2,M) hier (2,5) abgeschlossen. Als Gewichtung der Knoten dienen die mit x beziehungsweise mit $(1–x)$ multiplizierten Bearbeitungszeiten $p_1,...,p_5$.

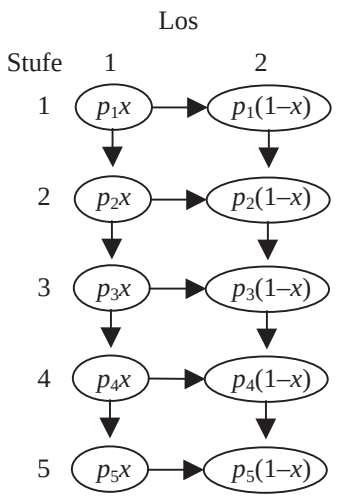


Abbildung 27: Zwei konsistente Lose in der Netzwerkdarstellung

Um den von Baker/Pyke (1990) vorgestellten Algorithmus nachvollziehen zu können, müssen zunächst einige strukturelle Eigenschaften optimaler Lösungen hergeleitet werden. Zur Vereinfachung der Darstellung wird mit $P(e,g)$ die Summe der Ausführungszeiten von der Stufe e bis zur Stufe g bezeichnet.

$$P(e,g) = p_e + p_{e+1} + ... + p_g \quad \text{mit} \quad 1 \le e \le g \le M \tag{39}$$

Die maximale Durchlaufzeit entspricht dem kürzesten der längsten Pfade durch das Netzwerk bei Variation von x mit $0 < x < 1$. Da jeder Pfad bis zu einer bestimmten Stufe über das erste und danach über das zweite Los führen muss, lässt sich die maximale Durchlaufzeit in Abhängigkeit von x und der Summe der Bearbeitungszeiten auf den Stufen ausdrücken. Die beiden Summanden in der geschweiften Klammer der Formel (40) entsprechen der Länge des Pfads über das erste Los bis zur Stufe g und über das zweite Los ab der Stufe g bis zur Stufe M:

$$C_{max} = \max_g \left\{ xP(1,g) + (1-x)P(g,M) \right\} \tag{40}$$

$$C_{max} = \max_g \left\{ x\left[P(1,g) - P(g,M) \right] + P(g,M) \right\} \tag{41}$$

dabei ergibt sich Formel (41) sich durch Ausmultiplizieren. Werden für $P(1,g)$ und $P(g,M)$ die Summen aus Formel (39) eingesetzt, hebt sich in der eckigen Klammer $p_g - p_g$ auf:

$$C_{max} = \max_g \left\{ x\left[P(1,g-1) - P(g+1,M) \right] + P(g,M) \right\} \tag{42}$$

Mit Gleichung (42) kann die kritische Stufe g eingeführt werden, mit deren Hilfe strukturelle Besonderheiten der optimalen Lösungen aufzudecken sind.

Definition: *Kritische Stufe (critical machine)*

> Eine Stufe g heißt kritisch, wenn die Durchlaufzeit nicht verkürzt werden kann, indem die Bearbeitungszeit p_g um Δ ZE vergrößert, während die Bearbeitungszeit auf den übrigen Stufen um insgesamt Δ ZE verringert wird.

Triff auf einer kritischen Stufe g eine Verlängerung der Bearbeitungszeit auf, kann diese durch eine gleichgroße Verkürzung auf anderen Stufen höchstens ausgeglichen, nicht aber überkompensiert werden. Diese Eigenschaft unterscheidet kritische Stufen von den übrigen Stufen und ist wie folgt zu begründen: In einem optimalen Plan π sei g eine kritische Stufe. Da diese Stufe sowohl auf dem Teil des kritischen Pfads liegt, der über das Los x_1 führt als auch auf dem Teil des Pfads der über das Los x_2 führt, wird eine geringfügige Veränderung der Bearbeitungszeit von p_g zu $p_g + \Delta$ ZE zu der Durchlaufzeit $C_{max} + \Delta$ ZE führen, da annahmegemäß $Q = 1$ gesetzt ist. Auf der Stufe e wird die Bearbeitungszeit $p_e - \Delta$ ZE aber lediglich mit x oder mit $(1–x)$ gewichtet erfasst, so dass sich

die auf der Stufe g gesetzte Verlängerung der Durchlaufzeit nur zum Teil ausgleichen kann. Diesen Zusammenhang verdeutlicht Abbildung 28:

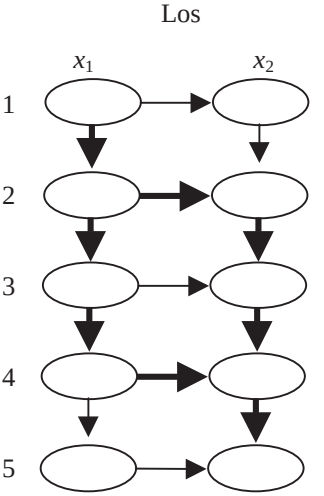


Abbildung 28: Zwei kritische Pfade

In Abbildung 28 sind zwei kritische Pfade zu erkennen: Der erste führt über die Knoten (1,1) – (1,2) – (1,3) – (1,4) – (2,4) – (2,5) und der zweite über (1,1) – (1,2) – (2,2) – (2,3) – (2,4) – (2,5). Im vorliegenden Beispiel sind damit die Stufen 2 und 4 kritisch, da ein kritischer Pfad auf ihnen horizontal verläuft. Um zu zeigen, dass die angegebene Definition tatsächlich nur kritische Stufen erfasst, werden die folgende Fälle an Hand der Abbildung 28 betrachtet:

Untersucht man die Stufe 1, kann eine Verlängerung von p_1 auf $p_1 + \Delta$ ZE z.B. durch eine Verkürzung von p_2 auf $p_2 - ½\Delta$ ZE und von p_4 auf $p_4 - ½\Delta$ ZE ausgeglichen werden. Sieht man sich aber die beiden kritischen Pfade an, hat sich ihre Länge jeweils um ½Δ ZE verkürzt. Die Durchlaufzeit sinkt, womit gezeigt ist, dass Stufe 1 in diesem Beispiel nicht kritisch ist und somit keinen Engpass darstellt.

Betrachtet man hingegen die Stufe 2 und verlängert p_2 auf $p_2 + \Delta$ ZE, kann diese Verzögerung zwar durch eine Verkürzung von p_4 auf $p_4 - \Delta$ ZE ausgeglichen werden, die Durchlaufzeit bleibt aber gleich. Wird eine der anderen Stufe zum Ausgleich der Verlängerung von p_2 auf $p_2 + \Delta$ ZE herangezogen und z.B. die Ausführungszeit auf einer nicht kritischen Stufe verkürzt, steigt die Durchlaufzeit an. Zusammengefasst kann keine Verringerung der Ausführungszeiten auf den übrigen Stufen zu einer Verkürzung der Durchlaufzeit führen, d.h. Stufe 2 ist kritisch.

Während die Argumentation bislang auf den Fall mit zwei kritischen Stufen ausgerichtet war, kann man den gezeigten Zusammenhang leicht verallgemei-

nern: Sind bei insgesamt *M* Stufen *h* Stufen kritisch und wird eine der nicht kritischen Stufen *e* betrachtet, dann kann die Veränderung von +Δ ZE auf der Stufe *e* durch *h* gleiche (negative) Anteile im Umfang von $-{}^{1}/_{h}\,\Delta$ ZE auf den h kritischen Stufen ausgeglichen werden. Da auf jedem der kritischen Pfade *h* + 1 dieser Anteile liegen müssen, verkürzt sich die Durchlaufzeit insgesamt um ${}^{1}/_{h}\Delta$ ZE, womit Stufe *e* als nicht kritisch identifiziert ist. Die restliche Argumentation erfolgt analog zum Fall mit zwei kritischen Stufen.

Es ist unmittelbar einsichtig, dass es im Zweilosfall pro kritischen Pfad immer genau eine kritische Stufe geben muss. Da in jedem Netzwerk mindestens ein kritischer Pfad auftritt, ist mindestens eine Stufe kritisch. Im Extremfall sind allerdings alle Stufen kritisch. Dieser tritt z.B. dann ein, wenn die Ausführungszeit auf allen Stufen identisch ist und damit *M* kritische Pfade existieren. Aus Sicht eines Produktionsplaners ist jede kritische Stufe als Engpass zu interpretieren, da Verzögerungen auf den kritischen Stufe unmittelbar zu einer Erhöhung der Durchlaufzeit führen.

Formal können aus der Eigenschaft, dass immer mindestens eine Stufe *g* kritisch ist, die folgenden Zusammenhänge hergeleitet werden. Für eine beliebige Stufe *e* gilt die folgende Ungleichung, die die Relation der Pfadlängen im Netzwerk angibt:

$$\begin{aligned} &x\left[P(1,g-1)-P(g+1,M)\right]+P(g,M)\geq \\ &x\left[P(1,e-1)-P(e+1,M)\right]+P(e,M) \end{aligned} \tag{43}$$

Umstellen der Formel liefert:

$$\begin{aligned} &x\left[P(1,g-1)-P(1,e-1)+P(e+1,M)-P(g+1,M)\right]\geq \\ &P(e,M)-P(g,M) \end{aligned} \tag{44}$$

Es wird nun eine Fallunterscheidung vorgenommen, da sich – je nachdem, ob die beliebige Stufe *e* vor der kritischen Stufe *g* oder hinter der kritischen Stufe *g* liegt – einige der Summanden in (44) aufheben. Es sind somit die zwei Fälle $e > g$ oder $e < g$ zu unterscheiden.

Für eine beliebige Stufe $e > g$ lässt sich Formel (44) zusammenfassen:

$x\left[-P(g,e-1)-P(g+1,e)\right]\geq -P(g,e-1)$ und damit gilt:

$$x\leq \frac{P(g,e-1)}{P(g,e-1)+P(g+1,e)} \tag{45}$$

Analog ist für die Betrachtung vor der kritischen Stufe $e < g$:

$x\left[P(e,g-1)+P(e+1,g)\right]\geq P(e,g-1)$ und damit:

$$x \geq \frac{P(e, g-1)}{P(e, g-1) + P(e+1, g)} \tag{46}$$

Die in Formel (45) und (46) angegebenen Beziehungen stellen obere und untere Schranken für x dar. Sie dienen im Weiteren zur Vereinfachung der Vorgehensweise.

Die folgende Argumentation baut darauf auf, dass der kritische Pfad über die kritische Stufe g führt und somit $xp_g + (1–x)\ p_g = p_g$ ZE zur minimalen Durchlaufzeit beiträgt. Dieser Beitrag ist aber offensichtlich unabhängig von x. Wie die Losgröße Q in die beiden Teillose aufzuteilen ist, kann nur von den Ausführungszeiten auf den übrigen Stufen abhängen; die kritische Stufe g spielt dabei keine Rolle. Betrachtet man nun den bereits in Formel (42) gezeigten Ausdruck für die maximale Durchlaufzeit näher, ist festzustellen, dass $P(g,M)$ eine Konstante ist. Nimmt man für ein gegebenes g an, dass in

$$C_{\max} = \max_g \left\{ x\left[P(1, g-1) - P(g+1, M) \right] + P(g, M) \right\} \tag{42}$$

der mit x multiplizierte Ausdruck negativ ist, dann besteht eine Tendenz, x so groß wie möglich zu wählen, da die Zielsetzung darin besteht, die maximale Durchlaufzeit $C_{\max}$ zu minimieren.

Der Fall, dass x (die Losgröße des ersten Loses) so groß wie möglich zu wählen ist, tritt auf, wenn die Summe der Ausführungszeiten bis zum Engpass kleiner ist als die Summe der Ausführungszeiten nach dem Engpass. Aus der Überlegung, dass für eine gegebene kritische Stufe g die Argumentation über beide Ungleichungen (45) und (46) zu x führen muss, kann je nach Stufe e eine der beiden Ungleichungen zur Festlegung von x gewählt werden. Die Beschränkung für x nach oben liefert dabei die Formel (45). Die obere Schranke ergibt sich als das Minimum über alle e mit $e > g$. Dieser Zusammenhang ist in Formel (47) angegeben, wobei die obere Schranke mit $\tilde{x}_g$ bezeichnet wird:

$$\tilde{x}_g = \min_{e>g} \left\{ \frac{P(g, e-1)}{P(g, e-1) + P(g+1, e)} \right\} \tag{47}$$

Analog wird vorgegangen, wenn der Koeffizient von x in Formel (38) nicht negativ ist. In diesem Fall besteht eine Tendenz, das erste Los so klein wie möglich zu wählen.

Daher wird die Losgröße x auf die untere Schranke $\underset{\sim}{x}_g$ gesetzt. Man erhält diesen Wert, indem man das Maximum über alle e mit $e < g$ aus Formel (46) ermittelt:

$$\underset{\sim}{x}_g = \max_{e<g} \left\{ \frac{P(e, g-1)}{P(e, g-1) + P(e+1, g)} \right\} \tag{48}$$

Die zugehörige maximale Durchlaufzeit ergibt sich aus Formel (42). Zusammengefasst gilt:

Falls $P(1,g-1)-P(g+1,M)<0 \Rightarrow x=\tilde{x}_g$

Falls $P(1,g-1)-P(g+1,M)\geq 0 \Rightarrow x=\underset{\sim}{x}_g$ (49)

Um die optimale Losgröße *x* zu bestimmen, ist ausgehend von der kritischen Stufe *g* zu prüfen, für welche andere Stufe *e* die Durchlaufzeit in Formel (42) minimiert wird, wenn die Losgröße *x* gemäß der Fallunterscheidung in Formel (49) festgelegt ist.

Da aber bislang nicht gesagt wurde, welche Stufe kritisch ist, müssen im Prinzip alle Stufen versuchsweise als kritische Stufe angenommen und dann mit den übrigen verglichen werden. Wie Baker/Pyke (1990, Seite 490-491) allerdings zeigen, gibt es immer mindestens zwei kritische Stufen.

Theorem 2: *Multiple Engpässe*

> In einer optimalen Lösung mit zwei konsistenten Losen sind mindestens zwei Stufen kritisch.

Zur Begründung dieser Eigenschaft genügt es, die oben angeführte Herleitung heranzuziehen und zu berücksichtigen, dass die Stufe *g* nur kritisch sein kann, wenn *x* sowohl die Formel (45) als auch die Formel (46) für jeweils mindestens eine Stufe *e* erfüllt. Die Ermittlung der Schranken hat gezeigt, dass entweder das Maximum oder das Minimum einer der beiden Ausdrücke verwendet wird. Dieses stellt sich aber für mindestens eine konkrete Stufe *e* ein und dann gilt in Formel (42) ein „=“ statt des „≤“. Im Umkehrschluss folgt aus der Gleichheit in Formel (42), dass immer mindestens eine andere Stufe *e* existiert, über die horizontal ein kritischer Pfad mit der gleichen Länge führt, die sich für die Stufe *g* als Minimum der maximalen Durchlaufzeiten ergeben hat. Es sind daher immer mindestens zwei Stufen kritisch.

Wie Baker/Pyke (1990) weiter zeigen, zählt die Stufe mit der längsten Bearbeitungszeit zu den kritischen Stufen, so dass sich der Rechenaufwand der enumerativen Bewertung aller möglichen Stufenpaare erheblich reduzieren lässt.

Außerdem vereinfacht sich die Rechnung, wenn man statt der Größe des Loses *x* das Verhältnis der beiden Lose *x* und (1 – *x*) betrachtet. Im Folgenden werden die nötigen Umformungen für die Formel (45) mit $g < e$ gezeigt und dabei angenommen, dass für *e* das strikte „=“ gilt. Wenn

$$x=\frac{P(g,e-1)}{P(g,e-1)+P(g+1,e)}, \text{ dann gilt für}$$

$$1-x=\frac{P(g+1,e)}{P(g,e-1)+P(g+1,e)}$$

Setzt man x und $(1–x)$ ins Verhältnis und bezeichnet dieses mit ρ, ergibt sich:

$$\rho = \frac{x}{1-x} = \frac{P(g,e-1)}{P(g+1,e)} \tag{50}$$

Mit diesen Umformungen gestaltet sich der Algorithmus so einfach, dass er notfalls manuell durchzuführen ist. Zunächst wird die erste kritische Stufe bestimmt, indem diejenige mit maximaler Ausführungszeit gewählt wird. Sollten mehrere Stufen in Frage kommen, kann eine dieser Stufen beliebig gewählt werden. Danach wird mit c_g geprüft, ob die Ausführungszeitsumme vor oder nach dem Engpass größer ist, was die in Formel (49) eingeführte Fallunterscheidung ermöglicht. Da nicht die Losgröße x direkt sondern das Verhältnis ρ ermittelt wird, vereinfacht sich die Betrachtung auf die Relation in Formel (50).

Schritt 1: Bestimme die erste kritische Stufe g, indem $g = \arg\max\{p_m\}$ gesetzt wird.

Schritt 2: Ermittle $c_g = \left[P(1,g-1) - P(g+1,M)\right]$

Schritt 3: Falls $c_g < 0$, setze $\rho = \min\limits_{e>g}\left\{\dfrac{P(g,e-1)}{P(g+1,e)}\right\}$

Falls $c_g \geq 0$, setze $\rho = \max\limits_{e<g}\left\{\dfrac{P(e,g-1)}{P(e+1,g)}\right\}$

Schritt 4: $x = \dfrac{\rho}{1+\rho}$

Schritt 5: $C_{\min} = Q\,[c_g x + P(g,M)]$

Dieser Algorithmus wird nun für das anfangs gegebene Beispiel mit $Q = 6$; $p_1 = 1$; $p_2 = 1$; $p_3 = 15$ und $p_4 = 3$ angewendet. Stufe 3 weist die größte Bearbeitungszeit auf und ist daher kritisch, d.h. g = 3.

Im Schritt 2 wird $c_g = (1+1) – 3 = –1$ ermittelt. Da c_g negativ ist, trifft im Schritt 3 der obere Fall zu und man erhält:

$$\rho = \min_{e>3}\left\{\frac{P(3,e-1)}{P(3+1,e)}\right\} = \frac{15}{3} = 5$$

Im Schritt 4 gilt damit: $x = \dfrac{5}{1+5} = \dfrac{5}{6}$

Berücksichtigt man, dass $Q = 6$ gegeben ist, ergibt sich $x_1 = 5$ und $x_2 = 1$.

Im Schritt 5 wird die Durchlaufzeit ermittelt:

$C_{\min} = Q\,[c_g x + P(1,M)] = 6\,[(–1)\cdot(5/6) + 18] = 103$

5.3.2 Mehrere konsistente Lose im Mehrstufenfall

Blickt man auf die bisher behandelten Fragestellungen zurück, wird deutlich, dass unter den verschiedenen Losarten den konsistenten Losen für die Praxis eine besonders hohe Bedeutung zukommt. Allerdings scheitern die bislang präsentierten Verfahren, wenn optimale konsistente Lose im Mehrstufenfall ($M > 3$) und für mehr als 2 Lose festzulegen sind. An dieser Stelle setzen die folgenden Ausführungen an. Analog zur bisherigen Vorgehensweise werden zunächst die strukturellen Eigenschaften optimaler Lösungen behandelt. Danach wird die von Trietsch/Baker (1993) vorgeschlagene Formulierung als lineares Modell formal und in der LINGO-Kodierung dargestellt. Die sich anschließende postoptimale Analyse erlaubt eine umfassende ökonomische Interpretation, mit der bislang nicht aufgezeigte Zusammenhänge verdeutlicht werden. Mit einfachen Modifikationen können andere Zielfunktionen verfolgt und warte- beziehungsweise leerzeitfreie Pläne erzeugt, sowie Transportzeiten oder diskrete Lose berücksichtigt werden. Mit den Erweiterungen der Problemstellung auf mehrere Lose und mehrere Stufen ist zunächst hinsichtlich der Struktur der optimalen Lösungen festzustellen, dass einige Eigenschaften nicht mehr gelten:

Die optimalen Losgrößen müssen keine durchgängige geometrische Folge bilden, d.h. Folgerung 9 ist nicht übertragbar.

In Analogie zu den Folgerung 21, 22 und 23 treten zwar Übergangslose auf, allerdings lässt sich der Faktor, mit dem die Reihen anwachsen oder abklingen nicht allgemeingültig aus dem Verhältnis der Ausführungszeiten ableiten. Zur Verdeutlichung der Einschränkungen dienen die folgenden Beispiele:

Beispiel A: Gegeben sind $M = 5$ Stufen mit $p_1 = 1$; $p_2 = 3$; $p_3 = 4$; $p_4 = 4$ und $p_5 = 2$. Die Losgröße ist $Q = 40$ und es wird die durchlaufzeitminimale Lösung mit $J = 12$ konsistenten Losen gesucht. Die optimale Lösung ergibt sich für die folgenden (auf zwei Nachkommastellen gerundeten) Losgrößen mit $C_{min} = 186{,}01$ ZE:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
1,78	2,37	3,16	4,22	4,22	4,22	4,22	4,22	4,22	4,22	2,11	1,05

Die Losgrößen x_1 bis x_4 bilden eine geometrische Reihe, die mit dem Faktor 1,33 anwächst. Diese Reihe endet mit dem vierten Los. Die folgenden 7 Lose sind konstant, d.h. sie bilden eine degenerierte geometrische Reihe mit dem Faktor 1. Ab x_{10} bis x_{12} tritt in diesem Beispiel die dritte geometrische Reihe auf, der der Faktor 0,5 zugrunde liegt. Ein Übergangslos, das für den Fall der dominanten zweiten Stufe im Dreistufenfall den End- beziehungsweise den Anfangspunkt von zwei geometrischen Reihen bildet, tritt hier mit x_4 und x_9 zweimal auf. Betrachtet man den Faktor, mit dem die erste geometrische Reihe an-

wächst, so ergibt sich 1,333 als Quotient aus $p_3/p_2 = 4/3 = 1{,}333$. Ebenso findet sich der Faktor $1 = 4/4 = p_4/p_3$ als Relation der Ausführungszeiten der dritten und der vierten Stufe wieder. Für die dritte geometrische Reihe gilt der Faktor 0,5, der offensichtlich der Relation der Ausführungszeiten der Stufen 4 und 5 entspricht: $p_5/p_4 = 2/4$. Basierend auf dieser Beobachtung drängt sich die Vermutung auf, dass sich das Problem wie im Zweilosfall auf eine Identifikation von Übergangslosen reduzieren lässt. Das folgende Beispiel zeigt allerdings, dass in einer optimalen Lösung auch zwei gleichgerichtete geometrische Reihen direkt aufeinander folgen können und die Quotienten der Ausführungszeiten nicht den Faktoren der Reihen entsprechen müssen.

Beispiel B: Gegeben sind $M = 8$ Stufen mit $p_1 = 10$; $p_2 = 10$; $p_3 = 2$; $p_4 = 8$; $p_5 = 4$; $p_6 = 1$; $p_7 = 4$; $p_8 = 2$. Die Losgröße ist auf $Q = 40$ festgelegt und es wird wieder die durchlaufzeitminimale Lösung für $J = 12$ Lose gesucht. Die optimale Durchlaufzeit beträgt 470,95 ZE und ergibt sich für die folgenden (auf zwei Nachkommastellen gerundeten) Losgrößen:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
4,37	4,37	4,37	4,37	4,37	4,37	3,64	3,03	2,53	2,11	1,46	1,01

Während die Losgrößen x_1 bis x_6 konstant sind, bilden die Lose x_6 bis x_{10} eine geometrische Reihe, die mit dem Faktor 0,833 abnimmt. Daran schließt sich eine weitere geometrische Reihe (x_{10} bis x_{12}) an. Sie nimmt ebenfalls ab, allerdings mit dem Faktor 0,692. Waren diese Faktoren im Beispiel A den Relationen bestimmter Ausführungszeiten gleichzusetzen, widerlegt Beispiel B, dass dieser Zusammenhang generell gilt. Weder 0,833 noch 0,692 ergeben sich als Quotient von zwei Ausführungszeiten. Setzt man allerdings $p_1 = 10$ in Relation zur Summe der Ausführungszeiten $p_4 + p_5 = 12$, ergibt sich der gesuchte Faktor von 0,833. Es ist daher zu prüfen, ob sich im Beispiel B der Faktor 0,692 ebenfalls als ein Quotient von summierten Ausführungszeiten darstellen lässt. Die Enumeration aller möglichen Quotienten zeigt, dass sich auch dieser Zusammenhang nicht bestätigt. Die Faktoren der geometrischen Reihen unterliegen offensichtlich komplizierteren Gesetzmäßigkeiten. Weiterhin belegt Beispiel B, dass in einer optimalen Lösung auch zwei gleichgerichtete Reihen unmittelbar aufeinanderfolgen können. A priori ist dabei weder erkennbar, wie viele verschiedene geometrische Reihen auftreten, noch wie lang jede der Reihen sein wird. Es ist nicht zu erwarten, dass für eine beliebige Stufen- und Loszahl ein exaktes Verfahren über die Bestimmung der Quotienten oder der Anzahl der geometrischen Reihen abzuleiten ist. Der bisher einzige Versuch, die Struktur der optimalen Lösungen für den Mehrstufenfall zu untersuchen und darauf aufbauend ein Lösungsverfahren zu entwickeln, stammt von Glass/Potts (1998). Sie analysieren den Vierstufenfall und schlagen vor, das Problem zunächst zu relaxieren. Hierfür zeigen die Autoren, dass es möglich ist, dominierte Stufen

aus der Betrachtung zu eliminieren, ohne die Struktur der optimalen Lösung zu verändern. Die Lösungsidee besteht darin, für die dominanten Stufen eine optimale Lösung mit einem Suchalgorithmus zu generieren und diese Lösung auf das ursprüngliche, nicht relaxierte Problem zu übertragen. Zwar können Glass/Potts (1998) zeigen, dass der Rechenaufwand ihres Ansatzes durch $O(M+J \log(J))$ begrenzt ist, doch hat ihr Verfahren den gravierenden Nachteil, nicht für beliebige Problemgrößen anwendbar zu sein. Da der Suchalgorithmus für maximal 3 dominante Stufen ausgelegt ist, scheitert die Vorgehensweise, wenn nach der Relaxation die Zahl der dominanten Stufen größer als 3 ist. Probleme mit mehr als 4 Stufen sind daher regelmäßig mit diesem Algorithmus nicht lösbar. Waren Glass und Potts 1998 noch zuversichtlich, ihren Ansatz auf $M > 4$ erweitern zu können, ist mittlerweile festzustellen, dass diese Forschungsrichtung nicht zum Erfolg geführt hat. Das Relaxations- und Suchverfahren kann insbesondere nicht mit dem linearen Programmierungsansatz von Trietsch/Baker (1993) konkurrieren, der für eine beliebige Los- und Stufenzahl anwendbar ist und der im Folgenden präsentiert wird:

P(1): Das Mehrstufen-Mehrlosüberlappungsproblem bei konsistenten Losen

$$\text{Min } C_{JM} \tag{51}$$

$$C_{11} \geq p_1 x_1 \tag{52}$$

$$C_{jm} \geq C_{j-1,m} + p_m x_j \qquad (j = 2,...,J; m = 1,...,M) \tag{53}$$

$$C_{jm} \geq C_{j,m-1} + p_m x_j \qquad (j = 1,...,J; m = 2,...,M) \tag{54}$$

$$\sum_{j=1}^{J} x_j = Q \tag{55}$$

$$x_j \geq 0 \qquad (j = 1,...,J) \tag{56}$$

$$C_{jm} \geq 0 \qquad (j = 1,...,J; m = 1,...,M) \tag{57}$$

Dieses lineare Programm umfasst $(2MJ–M–J+2)$ Restriktionen und $J(M+1)$ Entscheidungsvariable. Restriktion (52) ist erforderlich, da für das erste Los auf der ersten Stufe kein Vorgänger-Los existiert und Stufe 1 diejenige Stufe ist, die als erstes in der Stufenfolge durchlaufen wird. Die Einschränkungen für die übrigen Lose hinsichtlich ihrer Fertigstellungstermine ergeben sich aus den Restriktionen (53) und (54). Restriktion (53) stellt die Verfügbarkeit der Stufe sicher, d.h. auf jeder Stufe ist die Bearbeitung des vorangehenden Loses abgeschlossen, bevor das nächste Los eingelastet wird. Restriktion (54) verhindert, dass mit der Weiterbearbeitung eines Loses begonnen wird, bevor die Bearbeitung auf der Vorgängerstufe abgeschlossen ist. Restriktion (55) verlangt, dass die gesamte Losgröße Q auf die J Teillose aufgeteilt wird. Für die Entscheidungsvariablen x_{jm} und C_{jm} gelten die Nicht-Negativitätsbedingungen.

Setzt man x und $(1–x)$ ins Verhältnis und bezeichnet dieses mit ρ, ergibt sich:

$$\rho = \frac{x}{1-x} = \frac{P(g,e-1)}{P(g+1,e)} \tag{50}$$

5.3.2.1 LINGO-Kode und Beispiel

Im Folgenden wird der LINGO-Kode für den Mehrlos- und Mehrstufenfall mit konsistenten Losen angegeben. Er ist für das Beispiel C formuliert: Für ein 5 stufiges Problem mit p = (2,4,6,2,3) und Q = 30 wird die durchlaufzeitminimale Lösung in 4 Losen gesucht. Da LINGO 7.0 Groß- und Kleinbuchstaben nicht unterscheidet, wird die Zahl der Lose *J* durch *S* und die Zahl der Stufen *M* durch O ersetzt, um *j* sowie *m* als Laufindex beizubehalten. Die Losgrößen x_1 bis x_4 sind der LINGO-Syntax entsprechend mit X(1) bis X(4) bezeichnet, die Fertigstellungstermine C_{jm} mit C(j,m) etc.

```
! Eingabe der Loszahl S, Stufenzahl O und Produktionsmenge Q;
DATA: S = 4; O = 5; Q = 30; ENDDATA

! Definieren der Variablen, Indices und Konstanten;
SETS:
Stufe /1..O/: m;
Los /1..S/: j;
Ausfuehrungszeit(Stufe): p;
Fertigstellung(Los,Stufe): C;
Losgroesse(Los): X;
ENDSETS

! Zuweisen der p-Werte aller Stufen als Vektor;
DATA: p = 2 4 6 2 3; ENDDATA

! Zielfunktion;
min = C(S,O);

! Restriktion (52);
C(1,1) >= p(1)*X(1);

! Restriktion (53);
@FOR(Stufe(m): @FOR(Los(j)|j #GE# 2: C(j,m) >= C(j-1,m) + p(m)*X(j)));

! Restriktion (54);
@FOR(Stufe(m)|m #GE# 2: @FOR(Los(j): C(j,m) >= C(j,m-1) + p(m)*X(j)));

! Restriktion (55);
@SUM(Los(j):X(j))=Q;

END
```

Kommentarzeilen werden mit ! eingeleitet. Ein ; schließt eine Anweisung ab. Das Symbol @ wird als Steuerzeichen zum Aufruf vordefinierter Funktionen, wie z.B. SUM oder FOR verwendet. In der ersten DATA-Anweisung werden die Parameter gesetzt, die für die Initialisierung der Variablen erforderlich sind.

Dies geschieht in dem mit SETS überschriebenen Abschnitt. Erst nachdem der Ausführungszeitenvektor *p* hinsichtlich seiner Dimension festgelegt ist, können ihm in der zweiten DATA-Anweisung die Werte zugewiesen werden. Da die Variablen als SETS definiert sind, können die Restriktionen (53) bis (54) sehr kompakt eingegeben werden. Soll ein Index nicht den gesamten Wertebereich durchlaufen, ist die zugehörige Bedingung in die FOR Schleife aufzunehmen. Bei der Restriktion (53) sollen z.B. alle Lose *j* für $j \geq 2$ angesprochen werden: @FOR(Los(j)|j #GE# 2. In den kodierten Restriktionen (53) und (54) ist außerdem zu erkennen, wie zwei FOR Schleifen ineinandergeschachtelt werden. In Restriktion (55) wird mit der SUM-Anweisung über alle Lose *j* sichergestellt, dass mit den festzulegenden x_j die vorgegebene Losgröße *Q* abgearbeitet wird.

Wählt man die entsprechende Option, gibt LINGO 7.0 die Restriktionen explizit aus. Sie beginnen mit einer Nummer, die die Restriktion identifiziert und auf die eine schließende, eckige Klammer als Zeichen für die Restriktion folgt. Die Zielfunktion wird mit „1]“ gekennzeichnet. Die Restriktionen für das Beispiel C sind mit Erläuterungen in den folgenden Tabellen angegeben:

Tabelle 5: **Ziel und Restriktionen des Beispiels C**

Ziel und Restriktionen		Erläuterung	
1]	MIN C(4,5)	Zielfunktion	
2]	– 2 X(1) + C(1,1) >= 0	Start Los 1	
3]	– 2 X(2) – C(1,1) + C(2,1) >= 0	Los 1 – 2	Stufe 1
4]	– 2 X(3) – C(2,1) + C(3,1) >= 0	Los 2 – 3	
5]	– 2 X(4) – C(3,1) + C(4,1) >= 0	Los 3 – 4	
6]	– 4 X(2) – C(1,2) + C(2,2) >= 0	Los 1 – 2	Stufe 2
7]	– 4 X(3) – C(2,2) + C(3,2) >= 0	Los 2 – 3	
8]	– 4 X(4) – C(3,2) + C(4,2) >= 0	Los 3 – 4	
9]	– 6 X(2) – C(1,3) + C(2,3) >= 0	Los 1 – 2	Stufe 3
10]	– 6 X(3) – C(2,3) + C(3,3) >= 0	Los 2 – 3	
11]	– 6 X(4) – C(3,3) + C(4,3) >= 0	Los 3 – 4	
12]	– 2 X(2) – C(1,4) + C(2,4) >= 0	Los 1 – 2	Stufe 4
13]	– 2 X(3) – C(2,4) + C(3,4) >= 0	Los 2 – 3	
14]	– 2 X(4) – C(3,4) + C(4,4) >= 0	Los 3 – 4	
15]	– 3 X(2) – C(1,5) + C(2,5) >= 0	Los 1 – 2	Stufe 5
16]	– 3 X(3) – C(2,5) + C(3,5) >= 0	Los 2 – 3	
17]	– 3 X(4) – C(3,5) + C(4,5) >= 0	Los 3 – 4	

Tabelle 6: **Restriktionen des Beispiels C**

Restriktionen	Erläuterung	
18] – 4 X(1) – C(1,1) + C(1,2) >= 0	Los 1	Stufe 1 – 2
19] – 4 X(2) – C(2,1) + C(2,2) >= 0	Los 2	
20] – 4 X(3) – C(3,1) + C(3,2) >= 0	Los 3	
21] – 4 X(4) – C(4,1) + C(4,2) >= 0	Los 4	
22] – 6 X(1) – C(1,2) + C(1,3) >= 0	Los 1	Stufe 2 – 3
23] – 6 X(2) – C(2,2) + C(2,3) >= 0	Los 2	
24] – 6 X(3) – C(3,2) + C(3,3) >= 0	Los 3	
25] – 6 X(4) – C(4,2) + C(4,3) >= 0	Los 4	
26] – 2 X(1) – C(1,3) + C(1,4) >= 0	Los 1	Stufe 3 – 4
27] – 2 X(2) – C(2,3) + C(2,4) >= 0	Los 2	
28] – 2 X(3) – C(3,3) + C(3,4) >= 0	Los 3	
29] – 2 X(4) – C(4,3) + C(4,4) >= 0	Los 4	
30] – 3 X(1) – C(1,4) + C(1,5) >= 0	Los 1	Stufe 4 – 5
31] – 3 X(2) – C(2,4) + C(2,5) >= 0	Los 2	
32] – 3 X(3) – C(3,4) + C(3,5) >= 0	Los 3	
33] – 3 X(4) – C(4,4) + C(4,5) >= 0	Los 4	
34] X(1) + X(2) + X(3) + X(4) = 30		Menge *Q*

Das Potential, das diese Form der Modellierung zur Interpretation und Analyse der Struktur optimaler Lösungen bietet, sehen Glass et al. (1994, S. 379) als unzureichend an: „[...] *the linear programming approach provides little insight*." Auch wenn diese negative Einschätzung an anderer Stelle und von anderen Autoren, beispielsweise Chen/Steiner (1996, S. 592) oder Glass/Potts (1998, S. 624) aufgegriffen wird, trifft sie nicht zu. Die Autoren sehen offensichtlich die lineare Programmierung als ein reines Rechenverfahren an und erkennen nicht, dass ihr ein ebenso hoher Stellenwert bei der ökonomischen Analyse zukommt. Dies wird im Weiteren anhand des Beispiels C belegt.

5.3.2.2 Interpretation des Modells

Zunächst ist herauszuarbeiten, inwiefern das lineare Programm mit der Netzwerkdarstellung korrespondiert. Diese Gegenüberstellung hilft, die Interpretation des linearen Programms zugänglich zu machen.

Jede Restriktion bildet eine Kante und den Knoten ab, in den die Kante mündet. Bis auf die Knoten der ersten Stufe [(1,1), ..., (4,1)] und die Knoten des ersten Loses [(1,1), ..., (1,5)] wird jeder Knoten immer mit genau zwei Restriktionen erfasst. Die Abbildung 29 veranschaulicht den Zusammenhang zwischen

dem linearen Programm und dem Netzwerk, indem sie exemplarisch die inhaltlichen Entsprechungen für 10 Restriktionen des Beispiels C konkretisiert.

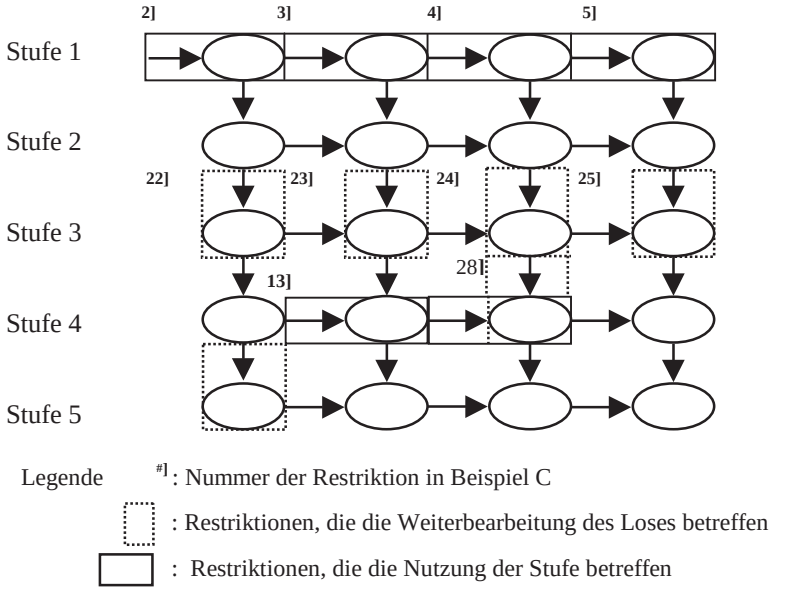


Abbildung 29: Restriktionen des Programms P(1) und ihre Zuordnung im Netzwerk

Die Restriktionen 3] bis 17] bilden die Nutzung auf den Stufen ab und stellen sicher, dass sich die Bearbeitung zweier Lose auf jeder Stufe nicht überschneidet. In Abbildung 29 wird der Zusammenhang für die Restriktionen 2] bis 5] sowie 13] gezeigt. Die Restriktionen 18] bis 33] verhindern, dass die Weiterbearbeitung eines Loses beginnt, bevor seine Bearbeitung auf der Vorstufe abgeschlossen ist. In die Abbildung 29 sind als Beispiele die Restriktionen 22] bis 25] und 28] aufgenommen worden. Während alle Restriktionen somit überlappend ineinander greifen, nutzt die Kante, die mit Restriktion 2] erfasst ist, als Ausgangspunkt den Zeitpunkt 0. Vor diesem Zeitpunkt kann nicht mit der Produktion begonnen werden.

Bestimmt man mittels LINGO 7.0 die optimale Lösung für Beispiel C, wird diese nach 41 Iterationen mit C_{min}=243,50 gefunden. Die ermittelten Werte sind im Weiteren immer auf 2 Nachkommastellen gerundet angegeben. Für die Losgrößen erhält man:

X(1): 4,87 X(2): 7,30 X(3): 10,96 X(4): 6,85

Mit diesen Losgrößen ergeben sich die folgenden Fertigstellungstermine der vier Lose auf den fünf Stufen:

C(1,1): 9,74 C(2,1): 29,23 C(3,1): 51,16 C(4,1): 102,33

C(1,2): 29,23	C(2,2): 58,47	C(3,2): 102,33	C(4,2): 129,74
C(1,3): 58,47	C(2,3): 102,33	C(3,3): 168,12	C(4,3): 209,23
C(1,4): 68,22	C(2,4): 116,95	C(3,4): 190,05	C(4,4): 222,94
C(1,5): 116,95	C(2,5): 138,88	C(3,5): 222,94	C(4,5): 243,50

Betrachtet man nur diese Werte, ist der Aussage von Glass et al. (1994) zuzustimmen. Strukturelle Besonderheiten der Lösung, wie z.B. die kritischen Pfade, lassen sich nicht unmittelbar erkennen. Zieht man aber zur weitergehenden Analyse die Dualvariablen (Schattenpreise der Restriktionen) heran, sind die Informationen leicht zugänglich. Zunächst ist aber die von LINGO 7.0 automatisch vorgenommene Transformation wichtig: Das System formt bei der Berechung die vorliegende Zielfunktion in eine zu maximierende Zielfunktion um und gibt bezüglich des dazu korrespondierenden Duals die Werte der Dualvariablen an. Da die primalen Restriktionen in „≥“ Form vorliegen, erhält man für die Dualvariablen den Wertebereich „≤ 0“. Für die weiteren Ausführungen werden die Dualvariablen auf das ursprüngliche Problem bezogen interpretiert. Für 16 Restriktionen ergeben sich von Null abweichende Schattenpreise:

2] 1,00	6] 0,64	7] 0,07	9] 0,36
10] 0,93	11] 0,64	17] 0,36	18] 1,00
22] 0,36	23] 0,57	24] 0,07	28] 0,36
29] 0,64	32] 0,36	33] 0,64	34] 8,11

Generell gibt ein Schattenpreis die relative Veränderung des Zielfunktionswerts bei einer Verschärfung der zugehörigen Restriktion um eine Einheit an, sofern dadurch kein Basiswechsel erzwungen wird (vgl. Kistner, 2003). Hier ist die Verschärfung einer der Restriktionen 2] bis 33] dahingehend zu interpretieren, dass eine Verzögerung um eine ZE auftritt, die den frühestmöglichen Start der weiteren Bearbeitung aufschiebt. Der Wert des Schattenpreises gibt an, um wie viele ZE sich die Durchlaufzeit auf Grund dieser Verzögerung verlängert. Je nachdem, ob die zugehörige Restriktion die Weiterbearbeitung eines Loses oder die Nutzung auf einer Stufe abbildet, ist eine horizontal oder eine vertikal verlaufende Kante betroffen. Beispielsweise ist für Restriktion 2] ein Schattenpreis von 1,00 ausgewiesen. Die Durchlaufzeit wird um 1,00 ZE ansteigen, wenn sich auf der Kante, die mit Restriktion 2] umfasst ist, eine Verzögerung um eine ZE ergibt. Dass in diesem Fall die Ursache „Verzögerung um eine ZE“ genau der Konsequenz „Verlängerung der Durchlaufzeit um eine ZE“ entspricht, ist unmittelbar einsichtig: Wie in Abbildung 29 gezeigt, stellt die in Restriktion 2] erfasste Kante sicher, dass die Produktion erst im Zeitpunkt Null und nicht früher beginnt. Wird mit der Produktion des ersten Loses auf Stufe 1 erst im Zeitpunkt 1 begonnen, bleibt offensichtlich die Struktur des restlichen Plans erhalten. Er verlagert sich lediglich um eine Zeiteinheit nach rechts. Folg-

lich steigt der Zielfunktionswert um eine ZE, was genau dem Schattenpreis der Restriktion 2] entspricht.

Analog kann mit Hilfe der übrigen Dualvariablen untersucht werden, welche Veränderungen des Plans eintreten, wenn sich auf einer der Kanten eine Verzögerung ergibt. Ergänzend erlauben die Dualvariablen, a priori zu ermitteln, welche Konsequenzen für die Durchlaufzeit aus der Unterbrechung der Produktion in einem der Knoten um eine ZE resultieren. Um die Bewertung der Restriktionen den jeweiligen Kanten zuzuordnen, werden sie in der folgenden Abbildung gemeinsam dargestellt.

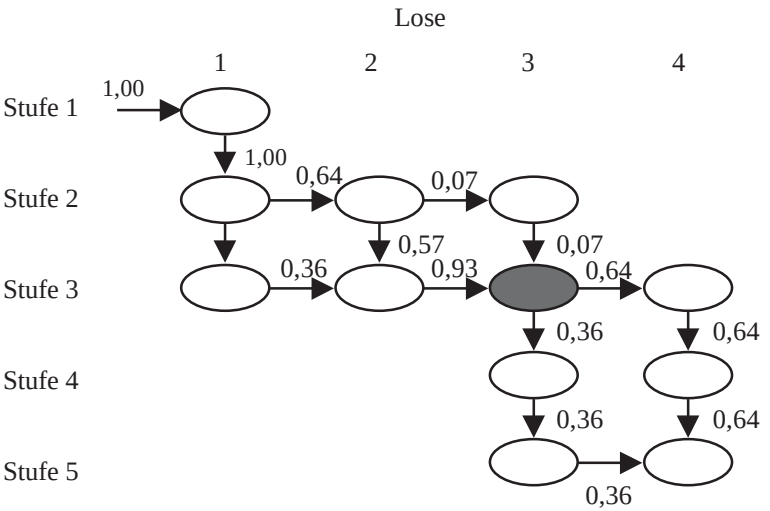


Abbildung 30: Bewertungen in der Netzwerkdarstellung

Wiedergegeben ist nur der Ausschnitt des Netzwerks, der die Kanten umfasst, für die positive Schattenpreise ermittelt wurden. Dieser Ausschnitt wird im Weiteren als das *kritische Netz* bezeichnet. Jede Kante im kritischen Netz ist kritisch: Jede Verzögerungen entlang einer dieser Kanten führt zu einer Verlängerung der Durchlaufzeit. Ebenso wirken sich die Verzögerungen aus, die in den Knoten auftreten, über die einer der kritischen Pfade führt. Betrachtet man beispielsweise die Bearbeitung des dritten Loses auf der dritten Stufe – sie ist in Abbildung 30 grau unterlegt – kann sie drei Arten von Verzögerungen erfahren:

- Die dritte Stufe kann nicht unmittelbar nach der Fertigung des zweiten Loses weitergenutzt werden, wenn zuvor beispielsweise unvorhergesehene Reinigungsarbeiten anfallen.
- Das dritte Los kann nach seiner Fertigstellung auf der zweiten Stufe nicht unmittelbar weiter bearbeitet werden, wenn z.B. das Transportmittel nicht sofort zur Verfügung steht.

- Die Bearbeitung des dritten Loses auf Stufe 3 kann sich verzögern, obwohl es rechtzeitig angeliefert und freigegeben wurde, wenn z.B. während der Produktion eine Störung auftritt.

Um die Konsequenzen der drei Verzögerungen vergleichen und sie a priori abschätzen zu können, werden die Schattenpreise genutzt. Wird die Produktion des Loses 3 auf der Stufe 3 um eine ZE verzögert freigegeben, verlängert dies die Durchlaufzeit um 0,93 ZE. Abgebildet ist dies durch die Restriktion 10] beziehungsweise durch die horizontal in den Knoten mündende Kante mit der Bewertung 0,93. Erfolgt hingegen der Transfer des dritten Loses von der zweiten zur dritten Stufe (erfasst in Restriktion 24]) nicht unmittelbar, sondern erst mit einer Verzögerung von einer ZE, verlängert sich die Durchlaufzeit um lediglich 0,07 ZE. Tritt der Fall c) ein und der Produktionsprozess des Loses 3 verzögert sich – nachdem er frühestmöglich gestartet wurde – um eine ZE, lässt sich dieser Fall als eine Kombination der beiden ersten auffassen, da es unerheblich ist, wann die Störung den Produktionsprozess unterbricht. Sie kann folglich auch schon zu Beginn der Bearbeitung des Loses 3 auf der Stufe 3 auftreten und ist daher durch ein gleichzeitiges Verschärfen beider Restriktionen 10] und 24] zu modellieren. Da sich in diesem Beispiel die Bewertung der beiden Kanten, die in den Knoten (3,3) münden, zu 1 addiert, führt eine Verzögerung des Produktionsprozesses um eine ZE zu einer Verlängerung der Durchlaufzeit um genau dieselbe Zeitspanne.

Die wichtige Beobachtung, dass sich eine Verzögerung um eine ZE – die unmittelbar vor oder während der Bearbeitung desselben Loses auf der gleichen Stufe einsetzt – derart unterschiedlich stark auswirken kann, überrascht zunächst, ist aber ökonomisch gut begründbar. Zu diesem Zweck werden in den folgenden Tabellen verschiedene optimale Lösungen gegenübergestellt.

Den Ausgangspunkt bildet die Situation ohne Verzögerung, für die die optimalen konsistenten Lose in der zweiten Spalte der Tabelle 7 angegeben sind.

Gegenübergestellt wird in der dritten Spalte der Tabelle 7 die Konsequenz, die sich bei einer Verzögerung entlang der horizontalen Kante für die optimalen Losgrößen ergibt. Die zweite Spalte entspricht damit dem Fall a), da die Verzögerung die Verfügbarkeit der Stufe betrifft (Restriktion 10]).

Die erste Spalte der Tabelle 8 gibt die optimale Lösung bei einer Verzögerung entlang der vertikalen Kante an und entspricht damit dem Fall b), da die Verzögerung den Transfer betrifft (Restriktion 24]).

Die letzten beiden Spalten der Tabelle 8 bilden den Fall c) ab. Die Verzögerung findet „innerhalb“ eines Knotens statt. Betrachtet werden die Verzögerung im Knoten (3,3) und die Verzögerung im Knoten (3,4). Der Knoten (3,4) ist der in Abbildung 29 gezeigte Knoten, für den die Überlappung der beiden Restriktionen 13] und 28] in dem Knoten (3,4) explizit hervorgehoben ist.

Tabelle 7: **Auswirkungen der Verzögerung auf den Kanten**

Losgröße	unverzögert	Verzögerung um 1 ZE	
		Fall a)	Fall b)
		10]	24]
x_1	4,873	4,807	4,930
x_2	7,309	7,210	7,408
x_3	10,964	11,066	10,862
x_4	6,852	6,916	6,789
C_{min}	243,502	244,424	243,581

Wie man anhand der Tabelle 7 erkennt, führt die Verzögerung um eine ZE zu unterschiedlich veränderten Losgrößen. Ist a priori, d.h. noch vor der ersten Einlastung bekannt, dass sich die Freigabe des Loses 1 auf der Stufe 3 um eine ZE verzögert (Verschärfung der Restriktion 10]), wird reagiert, indem die ersten beiden Lose verkleinert und dafür die beiden letzten Lose vergrößert werden. Die Menge der vor der verzögerten Weitergabe zu bearbeitenden Stücke nimmt somit ab, die danach zu fertigende Menge zu. Tritt hingegen die Verzögerung bei der Weitergabe des Loses 3 auf (dargestellt als Verschärfung der Restriktion 24]), wirkt sich dies genau entgegengesetzt aus. Die ersten beiden Lose werden größer, die letzten beiden kleiner.

Tabelle 8: **Auswirkungen der Verzögerung in den Knoten**

Losgröße	unverzögert	Verzögerung um 1 ZE	
		Fall c)	Fall c)
		10] und 24]	13] und 28]
x_1	4,873	4,873	4,852
x_2	7,309	7,309	7,279
x_3	10,964	10,964	10,918
x_4	6,852	6,852	6,949
C_{min}	243,502	244,502	243,862

Verlängert sich die Bearbeitung im Knoten (3,3) um eine ZE, dann behalten die Lose genau die Größe bei, die auch für die unverzögerte Ausgangslösung ermittelt wurde. Dies zeigt der Vergleich der Werte der fünften Spalte mit denen der zweiten Spalte.

Die letzte Spalte der Tabelle gibt schließlich die Konsequenzen einer Verzögerung im Knoten (3,4) an, d.h. die Produktion des dritten Loses auf Stufe 4. Es wird der Effekt untersucht, der sich durch gleichzeitige Verschärfung der Re-

striktionen 13] und 28] ergibt. Die ersten drei Lose werden etwas kleiner gewählt und dafür das vierte Los vergrößert.

Auf Basis der Veränderungen im Knoten (3,3) und dem Vergleich zu den Konsequenzen einer Verzögerung im Knoten (3,4) wird deutlich, dass sich der Knoten (3,4) dadurch auszeichnet, dass nur eine kritische Kante in ihn mündet. Im Gegensatz dazu sind beide in den Knoten (3,3) mündenden Kanten kritisch. Diese Eigenschaft führt zur Definition der kritischen Knoten. Für das Beispiel sind sie in Abbildung 31 grau unterlegt.

Definition: *Kritischer Knoten*

Ein Knoten (*j*,*m*) heißt kritisch, wenn sich die Schattenpreise auf den Kanten, die in den Knoten (*j*,*m*) münden, zu 1 addieren.

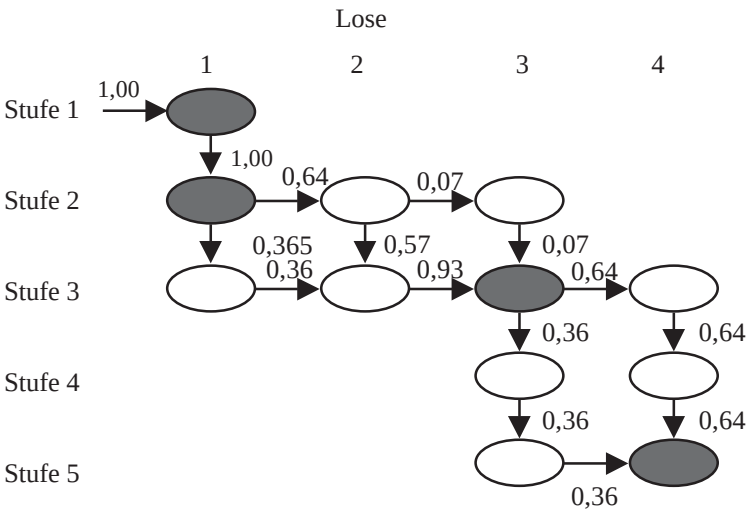


Abbildung 31: Kritische Knoten

Aus der Definition der kritischen Knoten abgeleitet und an Hand des Beispiels C motiviert, können die beiden Folgerungen aufgestellt werden:

Folgerung 24:

In jedem Netzwerk für *J* Lose auf *M* Stufen sind mindestens die Knoten (1,1) und (*J*,*M*) kritische Knoten.

Folgerung 25:

Tritt in einem kritischen Knoten eine Verzögerung auf, verlängert sich die Durchlaufzeit um diese Zeitspanne, sofern dadurch kein Basiswechsel erzwungen wird.

Löst man sich von dem Ursprung, den die Bewertungen der Kanten in diesem Netz haben, und betrachtet nur das Gefüge der Schattenpreise, hat es die Struktur eines Flussproblems. Ein Fluss der Stärke 1 tritt in der Quelle auf. Er verzweigt sich in dem Knoten (1,2) sowie (3,3) und mündet schließlich in die Senke, den Knoten (4,5).

Für jeden Knoten gilt dabei die Flussbedingung hinsichtlich der Entsprechung der Zu- und Abflüsse: In jeden Knoten fließt exakt so viel hinein wie hinaus (vgl. Neumann/Morlock, 1993, S. 257). Die Bewertung der Kante – hier interpretiert als Flussstärke – ergibt sich, indem die Konsequenz auf die Durchlaufzeit bei einer Verzögerung entlang der zugehörigen Kante um eine ZE abgetragen wird, sofern diese zu keinem Basiswechsel führt. Allerdings lassen sich die Veränderungen und ihre Interpretation ebenso in Richtung einer geringfügigen Beschleunigung interpretieren. Dieser Erweiterung, die für die praktische Anwendung wichtig ist, stand bislang eine Vereinfachung bei der Modellierung im Wege. Es wurde davon ausgegangen, dass zwischen den Stufen keine Transferzeiten anfallen und dass auf den Stufen zwischen den Losen keine Reinigungsarbeiten etc. vorzunehmen sind. Somit können sich prinzipiell die einzelnen Bearbeitungsschritte lückenlos aneinander anschließen. Lässt man diese Vereinfachung fallen, können auch die Konsequenzen von ablaufbeschleunigenden Maßnahmen betrachtet werden, die sowohl hinsichtlich des Transfers und der Bereitstellung der Stufe denkbar sind. Der Transfer eines bestimmten Loses kann schneller durchgeführt oder für Reinigungsarbeiten zusätzliches Personal eingesetzt werden. Die konkrete Modellierung dieser Erweiterungen wird im Abschnitt 5.3.4 gezeigt.

Betrachtet man die bisher präsentierten Ergebnisse aus Sicht eines Praktikers, kann er mit der Information, welche Knoten kritisch sind, die Arbeitsgänge identifizieren, die er mit besonderer Aufmerksamkeit steuern und überwachen sollte. Sowohl Verzögerungen als auch Beschleunigungen der Bearbeitung in einem kritischen Knoten wirken sich unmittelbar und in gleicher Höhe auf die Durchlaufzeit aus. Daneben zeigt die Identifikation der Schattenpreise, dass viele Möglichkeiten bestehen, wie auf Verzögerungen entlang einer Kante durch Beschleunigung entlang anderer Kanten zu reagieren ist. Wird weiter im Modell P(1) die aufzuteilende Losgröße auf $Q = 1$ normiert, ist die Aufteilung in die Losgrößen x_j als eine prozentual Aufteilung zu verstehen. Der normierte Losüberlappungsplan kann auf beliebige Losgrößen Q übertragen und allgemein die Struktur des kritischen Netzes für die Kombination aus Ausführungszeitvektor $\underline{p}$ und Loszahl J bestimmt werden.

Wie schon in den Forderungen von Little (1970) zur sinnvollen Modellbildung (vgl. den Abschnitt 2.3) dargestellt, liegen bei jedem Modell bestimmte Details außerhalb des modellierten Zusammenhangs. Es ist daher anzunehmen, dass der Produktionsplaner die verschiedenen Möglichkeiten zur Beschleuni-

gung beziehungsweise zur Reaktion auf eine Störung unterschiedlich bewertetet. Unter Berücksichtigung der Schattenpreise kann er aber den Entscheidungsprozess auf die effizienten Varianten, d.h. auf die Kanten beschränken, die bei einer Beschleunigung auch eine Reduktion des Zielfunktionswerts bewirken. Betrachtet man dazu wieder das Beispiel, zeigt sich beispielsweise:

- Eine Beschleunigung des Transfers zwischen den Stufen 2 und 3 muss nur die Lose 1, 2 und 3 betreffen, nicht aber Los 4, da dies außerhalb des kritischen Netzes liegt.
- Eine Beschleunigung um eine ZE in der Bereitstellung der Stufen muss nur für den Übergang zwischen Los 3 und 4 auf Stufe 3 und für Los 3 und 4 auf Stufe 5 nicht aber auf Stufe 4 erfolgen.

Die Schattenpreise helfen bei der Identifikation der Möglichkeiten, die als Reaktion auf eine Störung zur Verfügung stehen:

- Eine Verzögerung in Restriktion 9] um eine ZE (sie betrifft den Übergang zwischen Los 1 und Los 2 auf der Stufe 3) führt zu einer Verlängerung der Durchlaufzeit um 0,35 ZE. Darauf kann z.B. mit einer Beschleunigung um eine ZE innerhalb der Restriktionen 28] oder 32] (sie betreffen die Weitergabe des dritten Loses zwischen der dritten und vierten beziehungsweise zwischen der vierten und fünften Stufe) reagiert werden. Diese Beschleunigung führt zu einer Reduktion der Durchlaufzeit um 0,36 ZE und gleicht damit die Verlängerung um 0,35 ZE aus.
- Verzögert sich die Weitergabe aller Lose von der Stufe 2 zur Stufe 3 um eine ZE, kann dies durch eine Beschleunigung der Weitergabe der Lose 3 und 4 zwischen der vierten und der fünften Stufe kompensiert werden.

Allerdings gelten alle diese Überlegungen nur a priori, da sich z.B. bei Berücksichtigung einer verzögerten Freigabe der Stufe 3 – wie oben in der Tabelle 5 gezeigt – andere Losgrößen einstellen werden. Außerdem erfolgten die bisherigen Überlegungen immer unter der Bedingung, dass durch die Verzögerung kein Basiswechsel erzwungen wird. In dem Moment aber, in dem die Veränderung zu einem Basiswechsel führt, verändern sich auch die Dualvariablen sprunghaft. Bei starken Veränderungen kann sich somit auch die Struktur des kritischen Netzes verändern. Bisher nicht kritische Kanten werden kritisch und umgekehrt. Für die weitergehenden Analysen ist es daher wesentlich, ergänzende Aussagen über die Stabilität der Lösung und dabei insbesondere hinsichtlich des Verlaufs des kritischen Pfads abzuleiten. Genau diesen Fragen geht die postoptimale Analyse nach, die im folgenden Abschnitt dargestellt wird.

5.3.3 Postoptimale Analyse bei konsistenten Losen

Generell können zur Beurteilung der Eigenschaften einer optimalen Lösung neben den Dualvariablen die Ergebnisse der Sensitivitätsanalyse herangezogen werden (vgl. Dantzig/Thapa, 1997; Kistner, 2003). Das Konzept zur Beurteilung der Stabilität einer Lösung wird im Folgenden an Hand des Beispiels C demonstriert.

Im Zuge der Sensitivitätsanalyse sollen zunächst die nicht bindendenden Restriktionen betrachtet werden. Für folgende Restriktionen besteht in der optimalen Lösung Schlupf. Sie können um die angegebenen Werte verschärft werden, ohne dass dies zu einem Basiswechsel führt:

3] 4,87	5] 37,46	12] 34,11	13] 51,16	14] 19,19
16] 51,17	20] 7,31	25] 38,38	30] 34,11	

Tritt beispielsweise in Restriktion 3] eine Verzögerung auf, die unter dem ermittelten Wert von 4,87 bleibt, verlängert sich die Durchlaufzeit nicht. Eine Verzögerung ist mit einer Erhöhung der rechten Hand-Seite der Restriktion gleichzusetzen. Da ein Minimierungsproblem betrachtet wird und die Restriktionen als „≥“-Restriktionen vorliegen, kann eine beliebige Reduktion der rechten Hand-Seite vorgenommen werden, ohne dass sich die Struktur der Lösung verändert.

Für die weitere Interpretation nicht-bindender Restriktionen ist jedoch die Besonderheit zu berücksichtigen, dass in der vorliegenden Lösung eine mehrfache duale Degeneration auftritt. Diese Eigenschaft lässt sich aus den Daten der Problemstellung erkennen: Untersucht wird ein Minimierungsproblem mit $J(M+1) = 24$ Entscheidungsvariablen in $(2MJ–M–J+2) = 33$ Restriktionen. In der Basis der optimalen Lösung stehen die 24 Problem- und die oben genannten 9 Schlupfvariablen. Hinsichtlich der Dualvariablen wurden für 16 Restriktionen von Null abweichende Schattenpreise angegeben, d.h. in der vorliegenden Basislösung sind diese Restriktionen ausgelastet. Aus der *Complementary Slackness Bedingung* der linearen Programmierung (vgl. Kistner, 2003) ergibt sich, dass in einer optimalen Lösung für eine Restriktion entweder Schlupf ausgewiesen oder bei Knappheit ein positiver Schattenpreis anzugeben ist. Sollte sich für eine Restriktion, deren Schlupf auf Null gefallen ist, gleichzeitig ein Schattenpreis von Null einstellen, liegt duale Degeneration vor. Von den gegebenen 33 Restriktionen ist bekannt, dass 16 bindend sind und für 9 Restriktionen Schlupf ausgewiesen wird. Daher müssen die verbliebenen 8 Restriktionen dual degeneriert sein.

Für die Lösungen des linearen Programms hat das Auftreten einer dual degenerierten Basislösung die folgende Konsequenz (vgl. Kistner, 2003): Es können beliebig viele Lösungen mit gleichem Zielfunktionswert durch Konvexkombi-

nationen der dual degenerierten Basislösungen gebildet werden, die ebenfalls zulässige Lösungen des Problems darstellen. Wird in einem dual degenerierten, optimalen Tableau ein weiterer Basiswechsel durchgeführt, ist eine der bisher dual degenerierten Variablen gegen eine der bisherigen Basisvariablen auszutauschen. Die Zuordnung des ausgewiesenen Schlupfs zu den einzelnen Restriktionen ist insofern nicht eindeutig, als dass man mit dem erneuten Basiswechsel den ausgewiesenen Schlupf verlagern kann. Die Zielfunktionszeile bleibt bei diesem Wechsel der Basisvariablen unberührt. Inhaltlich ist die duale Degeneration im hier vorliegenden Kontext dadurch zu erklären, dass für die nicht-kritischen Lose Spielräume existieren. Ihre Freigabe kann zwischen dem frühest möglichen und spätest zulässigen Termin verschoben werden, ohne dass sich daraus Auswirkungen für die Durchlaufzeit ergeben. Diese zeitlichen Verlagerungen verkleinern oder vergrößern zwar den Schlupf für die nachgelagerten, nicht-kritischen Bearbeitungsschritte, die Struktur des kritischen Netzes bleibt durch diese Verlagerungen aber unberührt. Um herauszufinden, welche der Restriktionen dual degeneriert sind, muss das mit LINGO 7.0 aufgestellte Modell in den zugehörigen LINDO 6.1 Kode transferiert werden, da LINGO 7.0 keine Sensitivitätsanalyse anbietet. Inhaltlich liefert die Sensitivitätsanalyse eine Antwort auf die Frage, in welchen Grenzen die Beschränkungskonstanten der Restriktionen variiert werden können, ohne dass sich die Struktur der optimalen Lösung ändert, d.h. für den vorliegenden Fall, dass sich der Verlauf der kritischen Pfade nicht verlagert.

Die Ergebnisse der Sensitivitätsanalyse für die 16 bindenden Restriktionen: {2]; 6]; 7]; 9]; 10]; 11]; 17]; 18]; 22]; 23]; 24] 28]; 29]; 32]; 33]; 35]; 34]} und die acht degenerierten Restriktionen: {4]; 8]; 19]; 20]; 21]; 26]; 27]; 31]} sind auf zwei Nachkommastellen gerundet in der folgenden Tabelle zusammengefasst. Es fällt auf, dass sich für einige Restriktionen, trotz ausgewiesener Grenzen der Veränderung der rechten Hand-Seite der Wert 0,00 für die Dualvariable eingestellt hat. Diese Restriktionen können als dual degeneriert identifiziert werden. Bei den übrigen in der Tabelle fehlenden Restriktionen, z.B. 3] oder 5], handelt es sich um die bereits genannten 9 Restriktionen, die nicht bindend sind.

Zur Interpretation wird zunächst die Intervallgrenze für die Restriktion 34] betrachtet. Sie liefert eine Aussage über die Stabilität der vorliegenden Lösung, wenn die vorgegebene Losgröße $Q = 30$ variiert wird. Das kritische Intervall für Q beträgt $[30 - 30, 30 + \infty]$ und somit $[0, \infty]$. Inhaltlich sagt dies aus, dass bei einer beliebigen Variation der Losgröße Q ($Q \rightarrow \infty$ oder $Q \rightarrow 0$) die Lose x_1 ,..., x_4 ihre Relation beibehalten. Der Verlauf der kritischen Pfade und damit die Struktur des kritischen Netzes ändert sich nicht. Ceteris paribus bedeutet dies für den Produktionsplaner, dass er ein einmal ermitteltes optimales Größenverhältnis zwischen den Losen beibehalten kann, selbst wenn sich die Menge Q ändert. Diese Eigenschaft ist wichtig, da sie dazu beiträgt, dass die Planungs-

nervosität reduziert wird. Aus der Entscheidung, für eine bestimmte Produktionsstruktur konsistente Lose zu verwenden, resultiert eine optimale Lösung, die bei alleiniger Variation von Q in sich stabil ist. Diese Eigenschaft wurde im Verlauf der Untersuchung bereits mehrfach angesprochen. Sie gilt im Modell P(1) allgemein, da in Restriktion (55) $Q = 1$ gesetzt und entsprechend die Problemvariablen normiert werden können, wenn statt x_1 der normierte Wert x_1/Q gebildet wird. Die Sensitivitätsanalyse hinsichtlich der Restriktion (55) erlaubt aber, die Eigenschaft auch dann explizit nachzuweisen, wenn modifizierte Modelle, z.B. ohne Wartezeiten, mit stufenabhängigen Transferzeiten, mit Bereitstellungszeiten etc. betrachtet werden.

Tabelle 9: **Ergebnisse der Sensitivitätsanalyse**

Restriktion	Reduktion	Erhöhung	Dualvariable	Degeneriert
2]	9,75	∞	1,0	
4]	51,17	7,31	0,00	X
6]	2,97	33,35	0,64	
7]	4,57	39,14	0,07	
8]	37,46	38,38	0,00	X
9]	16,86	7,62	0,35	
10]	22,18	12,20	0,92	
11]	23,12	71,05	0,63	
17]	28,64	58,15	0,36	
18]	4,87	∞	1,0	
19]	7,31	4,87	0,00	X
20]	∞	7,31	0,00	X
21]	∞	37,46	0,00	X
22]	33,35	7,62	0,35	
23]	8,49	11,43	0,56	
24]	12,20	39,14	0,07	
26]	68,22	34,11	0,00	X
27]	34,11	51,17	0,00	X
28]	65,03	58,15	0,36	
29]	58,15	71,05	0,63	
31]	34,11	51,17	0,00	X
32]	28,64	58,15	0,36	
33]	58,15	28,64	0,63	
34]	30,00	∞	8,11	

Hinsichtlich der Auswirkungen einer Variation der Losgröße Q auf den Zielfunktionswert ist hier festzuhalten, dass er sich, gemäß dem ermittelten Schattenpreis der Restriktion 34] um 8,11 erhöht, wenn die Restriktionskonstante verändert wird. Werden z.B. statt 30 Stück 31 Stück produziert, bestätigt eine Kontrollrechung, dass der Zielfunktionswert tatsächlich um 8,11 ZE ansteigt.

Um die Position der degenerierten Schlupfvariablen gegenüber den Restriktionen, für die Schlupf ausgewiesen ist, übersichtlich darzustellen, werden die Ergebnisse der Tabelle 9 in die folgende Abbildung überführt. Dort sind die kritischen Kanten fett hervorgehoben. Kanten, für die Schlupf ausgewiesen ist, sind durch einen Punkt als Pfeilende gekennzeichnet und die degenerierten Kanten grau dargestellt.

Hinsichtlich der Struktur der Lösung ist zu erkennen, dass jeder Pfad von der Quelle zur Senke, der über eine der degenerierten Kanten führt, auch mindestens über eine Kante verläuft, für die Schlupf ausgewiesen ist. Der auf einer Kante des Pfads ausgewiesene Schlupf kann auf eine der benachbarten degenerierten Kanten entlang dieses Pfads (ganz oder in Teilen) verlagert werden. Inhaltlich entspricht dies dem Links- bzw. Rechtsverschieben von nicht-kritischen Losen im Gantt-Diagramm.

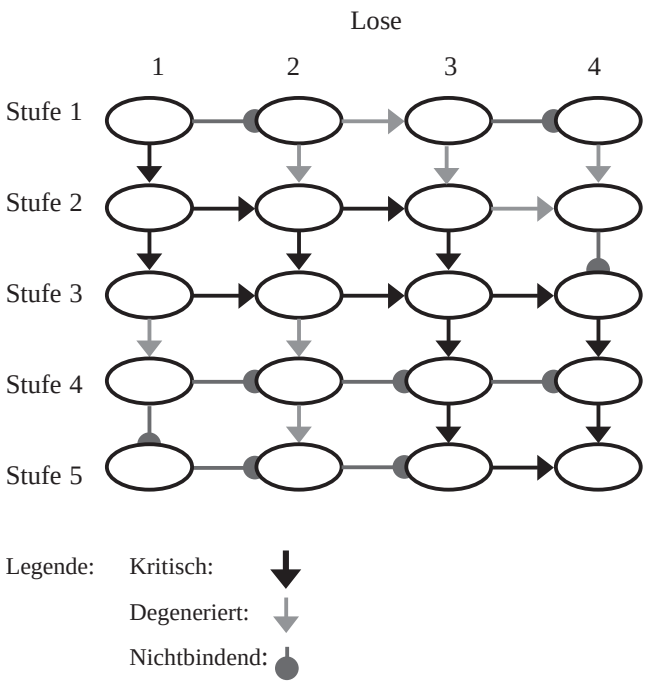


Abbildung 32: Kritische, degenerierte und nicht bindende Restriktionen

Beobachtet der Disponent, dass sich entlang nicht-kritischer Pfade Verzögerungen ergeben, sind diese noch auszugleichen, sofern auf dem Pfad bis zum Erreichen des nächsten kritischen Knotens genügend Schlupf ausgewiesen ist – oder im Falle der Degeneration – dorthin verlagert werden kann. Die mehrfache duale Degeneration macht damit die Analyse im Detail zwar unübersichtlich, ändert aber an den grundlegenden Zusammenhängen nichts.

Wendet man sich den übrigen Intervallen der Sensitivitätsanalyse zu, muss zunächst beachtet werden, dass in dem vorliegenden linearen Programm – bis auf zwei Ausnahmen – die Werte der rechten Hand-Seite der Restriktionen nicht als Datum vorgegeben sind, sondern sich aus den Entscheidungen hinsichtlich der übrigen Restriktionen ergeben. Lediglich in Restriktion 2] wird als frühester Startzeitpunkt der Bearbeitung der Zeitpunkt Null vorgegeben und in Restriktion 34] die zu fertigende Losgröße *Q* festgelegt. Alle übrigen Beschränkungen hängen voneinander ab. Betrachtet man z.B. die Restriktion 6] fällt auf, dass diese in engem Bezug zur Restriktion 18] steht:

$$6] \quad -4\,X(2) - C(1,2) + C(2,2) \geq 0 \iff -4\,X(2) + C(2,2) \geq C(1,2)$$

$$18] \quad -4\,X(1) - C(1,1) + C(1,2) \geq 0 \iff 4\,X(1) + C(1,1) \leq C(1,2)$$

$$\Rightarrow \quad -4\,X(2) + C(2,2) \geq C(1,2) \geq 4\,X(1) + C(1,1)$$

Um die in Tabelle 9 ausgewiesene Spanne zu konkretisieren, in der die rechte Hand-Seite einer Restriktion erhöht oder verringert werden kann, müssen daher zunächst die entsprechenden Werte der Problemvariablen eingesetzt werden.

Für Restriktion 6] ergibt sich: $-4 \cdot (7{,}30) + 58{,}47 \geq 29{,}23$. Gemäß der angegebenen Grenzen der Veränderung erhält man das Intervall [29,23 – 2,97; 29,23 + 33,35] = [26,26; 62,58]. Sofern die Bearbeitung des zweiten Loses auf Stufe 2 innerhalb des ausgewiesenen Intervalls abgeschlossen wird, ändert sich die Struktur der optimalen Lösung nicht. Kontrollrechnungen mit LINDO 6.1 bei denen durch zusätzliche Restriktionen der Abschluss der Bearbeitung des zweiten Loses auf Stufe 2 entsprechend eingegrenzt wird, bestätigen dieses Ergebnis. Analog wurden die übrigen Intervallgrenzen ermittelt und überprüft. Dabei fällt hinsichtlich der Restriktionen 20] und 21] auf, dass hier als Grenze der Reduktion der Wert ∞ auftritt. Diese beiden Einträge sind so zu interpretieren, dass auch beliebig große Beschleunigungen hinsichtlich der Restriktion 20] und 21] zu keinerlei Veränderungen in der Struktur der optimalen Lösung führen. Betrachtet man dazu in Abbildung 33 den Ausschnitt des Netzwerks, ist dieser Zusammenhang sofort einsichtig. Die Restriktionen 20] und 21] setzen mit dem Transfer von Stufe 1 zu Stufe 2 für die Lose 3 und 4 ein und bilden deren Bearbeitung auf der zweiten Stufe ab. Selbst wenn z.B. die Restriktion 20] beschleunigt wird, bestimmt der kritische Pfad über die Stufe 2 den frühesten Start der Bearbeitung des Loses 3 auf Stufe 2. Hinsichtlich der Restriktion 21]

ist der Zusammenhang noch klarer. Selbst eine erhebliche Beschleunigung der Fertigung des vierten Loses auf Stufe 2 hat keine Konsequenz. Die Stufe 3 als Engpass blockt den Effekt ab.

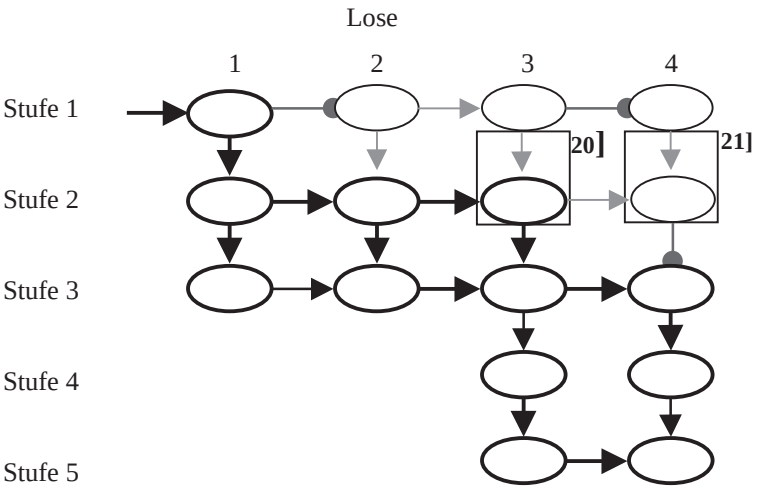


Abbildung 33: Lage der Restriktionen 20] und 21] im Netzwerk

Für den Produktionsplaner hat die Ermittlung der Intervalle den Vorteil, dass ihn kleinere Intervallgrenzen auf die Pläne bzw. Abschnitte in den Plänen hinweisen, für die Störungen schneller zu umfangreichen Änderungen führen werden. Entsprechend zeigen größere Intervalle auf, dass die zugehörigen Pläne (oder Abschnitte in den Plänen) zu Strukturen führen, die in sich stabiler sind. Es ist allerdings erneut hervorzuheben, dass dieses Modell „nur“ dafür gedacht sein kann, eine Entscheidungsunterstützung zu liefern. Der Produktionsplaner wird z.B. den Effekt (oder auch die Kosten) eines zusätzlichen Loses in Relation zur Reduktion der Durchlaufzeit setzen und damit außerhalb des Modells den zusätzlich anfallenden Konsequenzen der Losüberlappung Rechung tragen. Im Allgemeinen liegen ihm auch ergänzend Erfahrungswerte z.B. hinsichtlich der Störanfälligkeit der Maschinen oder der Leistungsfähigkeit der eingesetzten Mitarbeiter vor, die nur schwer formal zu erfassen sind. Mit Hilfe der Sensitivitätsanalyse kann er die Pläne identifizieren, die trotz Verzögerungen die Struktur des kritischen Netzes unverändert lassen.

Nimmt man weiterhin an, dass zur Bearbeitung der Lose auf den einzelnen Stufen Rohstoffe oder fremdbeschaffte Zwischenprodukte benötigt werden, kann auch die Zuverlässigkeit seines Lieferanten (im Sinne der pünktlichen Lieferung dieser Materialien) bei der Auswahl eines Plans berücksichtigt werden. Die Pläne, die genau auf den Kanten kritisch sind, für deren Einhaltung ein eher

unzuverlässiger Lieferant pünktlich liefern müsste, können a priori ausgesondert werden – oder es kann diesem Lieferanten bewusst ein früherer Liefertermin genannt und somit der Plan an kritischen Stellen „geschützt“ werden.

Da die Berechnungen der bisher dargestellten Varianten – auch für größere Probleme auf einem Standard PC – in wenigen Sekunden möglich sind, können alternative Pläne z.B. bei Variation der Anzahl der Lose ermittelt und verglichen werden. Dazu ist allerdings eine benutzungsfreundliche Aufbereitung der Zahlen, eventuell in Form der hier verwendeten Netzwerkstruktur oder in Form eines Gantt-Diagramms wesentlich.

5.3.4 Modifikationen des Modells bei konsistenten Losen

Bislang wurde davon ausgegangen, dass Eigenschaften der Problemstellung, wie

- Transferzeiten der Lose oder
- Bereitstellungszeiten der Stufen vernachlässigbar klein sind,
- Leerzeitfreiheit oder Wartezeitfreiheit nicht gefordert wird, die
- Zielfunktion zur Minimierung der Durchlaufzeit formuliert ist und der
- Wertebereich der Lose beliebig gewählt werden darf.

In diesem Abschnitt wird gezeigt, wie das Entscheidungsmodell zu modifizieren ist, wenn eine oder kombiniert mehrere der Varianten erfasst werden sollen.

Um das Modell P(1) um Wartezeit- oder Leerzeitfreiheit zu ergänzen, sind die unten genannten Modifikationen der Restriktionen (53) und (54) vorzunehmen (vgl. Baker, 1998). Restriktion (52), die die Fertigstellung des ersten Loses auf der ersten Stufe abbildet, kann bei allen Varianten unverändert bleiben, da wegen der zu minimierenden Zielfunktion ein Plan nicht optimal sein kann, wenn das erste Los nicht im Zeitpunkt Null gestartet wird. Die Restriktion (52) kann daher generell in der ursprünglichen Formulierung als „$\geq$“-Restriktionen beibehalten werden.

Für einen leerzeitfreien Plan ist in Restriktion (53) ein „=“ und analog für einen wartezeitfreien Plan in Restriktion (54) ein „=“ zu setzen. Dies führt zu den Restriktionen (53a) und (54a). Mit der Berücksichtigung von „=“ Restriktionen sind allerdings die Dualvariablen nicht mehr im Vorzeichen beschränkt (vgl. Kistner, 2003, S. 50).

Leerzeitfrei: $$C_{jm} = C_{j-1,m} + p_m x_j \;\; (j = 2,...,J;\; m = 1,...,M) \tag{53a}$$

Wartezeitfrei: $$C_{jm} = C_{j,m-1} + p_m x_j \;\; (j = 1,...,J;\; m = 2,...,M) \tag{54a}$$

Zudem besteht bei linearen Programmen mit mindestens einer „=“ Restriktion die Gefahr, dass das von Luptacik (2001) beschriebene „Get-more-for-less“ Paradoxon auftritt. Luptacik zeigt, dass trotz der Erhöhung der rechten Hand Seite einer „=“ Restriktion die zu minimierende Zielfunktion zu kleineren Werten führen kann. Für die Restriktion (53a) ergibt die rechte Hand Seite bisher den Wert 0. Wird sie z.B. auf den Wert 1 erhöht, kann sich dennoch eine Reduktion der Durchlaufzeit ergeben:

$$C_{jm} - C_{j-1,m} - p_m x_j = 0 \qquad (j = 2,\ldots,J;\ m = 1,\ldots,M) \tag{53a}$$

Insofern kann die Vorgabe einer z.B. aus technischen Gründen erzwungenen Leerzeit der Stufe dazu führen, das sich die Durchlaufzeit verringert. Analog kann dieser Fall für die Restriktion (54a) auftreten. Es kann z.B. verlangt sein, dass zwischen dem Ende der Bearbeitung auf einer Stufe und dem Beginn der Bearbeitung auf der Folgestufe eine genau einzuhaltende Zeitspanne liegen muss. Eine Verlängerung dieser Zeitspanne kann wiederum zu einer Reduktion der Durchlaufzeit führen.

Sollen Transferzeiten t_{m} oder Bereitstellungszeiten der Stufen b_m abgebildet werden, sind diese zu den rechten Hand-Seiten der Restriktionen zu addieren. Es ergeben sich die Restriktionen (53b) und (54b). Für beide Zeitspannen wird hier zur Vereinfachung angenommen, dass sie fix sind und unabhängig von der Losgröße auftreten, über die Stufen hinweg aber variieren können. Dabei bezeichnet die Transferzeit t_m die Zeitspanne für die Lieferung, die von der Stufe m zur Stufe $m + 1$ vorgenommen wird.

Bereitstellungszeit: $$C_{jm} \geq C_{j-1,m} + p_m x_j + b_m \qquad (j = 2,\ldots,J;\ m = 1,\ldots,M) \tag{53b}$$

Transferzeit: $$C_{jm} \geq C_{j,m-1} + p_m x_j + t_{m-1} \qquad (j = 1,\ldots,J;\ m = 2,\ldots,M) \tag{54b}$$

Auch bei den Ergänzungen um t_m und b_m kann Restriktion (52) unverändert übernommen werden, sofern sichergestellt wird, dass die Stufe 1 im Zeitpunkt Null betriebsbereit ist und dass das erste Los – ohne eine Lieferzeit abwarten zu müssen – im Zeitpunkt Null aufgelegt werden kann. Die praxisnahe Erweiterung um t_m und b_m hat den Vorteil, dass viele Maßnahmen, die zur Beschleunigung des Ablaufs dienen können, im Modell erfassbar werden. Für den Produktionsplaner bieten diese Erweiterungen eine genauere Modellierung der betrieblichen Abläufe, womit sich die zu erwartende Akzeptanz weiter erhöht.

Während das bislang betrachtete Modell P(1) die wichtige Eigenschaft hat, dass trotz beliebiger Setzung der Parameter immer mindestens eine zulässige Lösung existiert, gilt diese Eigenschaft für die erweiterte Problemstellung nicht mehr. Bei einer Aufteilung in mehrere Lose existiert für die meisten mehrstufigen Instanzen kein Plan, der zugleich leerzeit- und wartezeitfrei ist. Weiterhin kann a priori nicht festgestellt werden, ob eine Einschränkung der Problemstellung auf leerzeit-, wartezeitfreie oder ganzzahlige Pläne zu keiner, einer kleinen

oder einer deutlichen Veränderung des ursprünglichen Plans führen wird. Die folgenden Rechnungen zeigen die Auswirkungen für Beispiel B. Während die wartezeitfreie Bearbeitung zu keiner Änderung der oben gezeigten optimalen Lösung mit einer Durchlaufzeit von 470,95 ZE führt, steigt der Zielfunktionswert um über 70% auf 812 ZE an, wenn eine leerzeitfreie Bearbeitung verlangt ist. In diesem Fall sind die optimalen Losgrößen sogar ganzzahlig. Diese Besonderheit gilt aber nicht allgemein. Sie ergibt sich für diese Instanz und verdeutlicht den in Abbildung 10 gezeigten Zusammenhang, dass ein für die allgemeine Problemstellung entwickeltes Vorgehen eine Lösung finden kann, die zusätzlichen Restriktionen genügt.

Tabelle 10: **Wartezeitfreie Lösung für Beispiel B mit C_{min}= 470,95 ZE**

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
4,37	4,37	4,37	4,37	4,37	4,37	3,64	3,03	2,53	2,11	1,46	1,01

Tabelle 11: **Leerzeitfreie Lösung für Beispiel B mit C_{min}= 812 ZE**

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
1	4	4	4	4	4	4	4	4	4	2	1

Wie ein Vergleich der beiden Lösungen und insbesondere der (hier nicht angegebenen) Dualvariablen zeigt, ändert sich neben der Relation der Losgrößen auch der Verlauf des kritischen Pfads. Beide Modifikationen führen allerdings zu keiner nennenswerten Veränderung der Rechenzeit, da das Problem weiterhin linear ist und sich die Zahl der Variablen und Restriktionen nicht verändert. Wird für das Beispiel B nach einer Lösung gesucht, die zugleich wartezeit- und leerzeitfrei ist, tritt der Fall ein, dass keine zulässige Lösung existiert.

Kommen als Wertebereich für die Lose nur ganzzahlige Größen in Frage, wird die folgende Restriktion in den Lingo-Kode eingefügt:

```
! Ganzzahlige Losgrößen;
@FOR(Los(j): @GIN(x(j)));
```

Zur Lösung ist beispielsweise ein Branch & Bound Verfahren einzusetzen. Ermittelt man mittels LINGO 7.0 den wartezeitfreien, ganzzahligen Plan für das Beispiel B, beträgt die optimale Durchlaufzeit 485 ZE und stellt sich für die folgenden Losgrößen ein, die in Tabelle 9 den Losgrößen bei kontinuierlichem Wertebereich gegenübergestellt werden:

Tabelle 12: **Wartezeitfreie, ganzzahlige Lösung im Vergleich zu der reellwertigen**

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
4	4	4	4	4	4	4	4	3	2	2	1
4,37	4,37	4,37	4,37	4,37	4,37	3,64	3,03	2,53	2,11	1,46	1,01

Es fällt auf, dass sich die optimalen Losgrößen der ganzzahligen, wartezeitfreien Lösung nicht aus den entsprechenden Größen der reellwertigen Lösung durch Auf- und Abrunden ergeben (vgl. die Lose x_8 und x_{11}). In diesem Kontext kann auf die von Chen/Steiner (1997b) entwickelte *Balanced Redistribution* Heuristik zur Ermittlung ganzzahliger Pläne mit konsistenten Losen zurückgegriffen werden. Die Idee des Verfahrens besteht darin, ausgehend von der optimalen nicht-ganzzahligen Lösung eine möglichst gleichmäßige Umverteilung der nicht-ganzzahligen Reste über alle Lose vorzunehmen. Zur Abschätzung der Güte ihres Vorgehens vergleichen die Autoren die minimale Durchlaufzeit der optimalen ganzzahligen Lösung mit der Lösung, die durch Runden und heuristisches Umverteilen für das Problem P(1) entsteht. Für den maximalen Fehler können Chen/Steiner (1997b, S. 144) nachweisen, dass er – unabhängig von Q – unter $\sum_{m=1}^{M-1} p_m$ liegt.

Allerdings gilt dieses Ergebnis nur für die ursprüngliche Problemstellung P(1). Sollen (wie in dem oben gegebene Beispiel) neben der Ganzzahligkeit auch Eigenschaften wie Wartezeit- oder Leerzeitfreiheit erreicht werden, führt das Auf- bzw. Abrunden der entsprechenden nicht-ganzzahligen Lösungen regelmäßig zu Plänen, die die gewünschten Eigenschaften nicht mehr haben. Wie die im Zuge der vorliegenden Untersuchung durchgeführten Versuche mit LINDO 6.1 und LINGO 7.0 zeigen, sind optimale ganzzahlige Lösungen für größere Problemstellungen (z.B. mit $J = 20$, $M = 20$) auf einem Standard PC (1,3 GHZ; 512 MB Speicher, Windows 2000) nur in Ausnahmefällen mit akzeptabler Rechenzeit (innerhalb einiger Stunden) ermittelbar.

Schließlich ist es möglich, Modifikationen der Zielfunktion vorzunehmen. Bisher wurde in dieser Arbeit die Reduktion der Durchlaufzeit als Zielsetzung vorgegeben. Da sich das Losüberlappungsproblem formal als eine Variante des Flow Shop Scheduling Problems auffassen lässt (vgl. z.B. Topaloglu et al., 1994; Vickson, 1995), sind die dort gültigen Aquivalenz-Beziehungen für die regulären Zielkriterien übertragbar (Vickson/Alfredsson, 1992), sofern das Zielkriterium auf die Durchlaufzeit der vorgegebenen Losgröße Q bezogen wird. Daneben können Abweichungen von vorgegebenen Terminen berücksichtigt werden oder auch einzelne Teillose betrachtet werden. Dann allerdings ist die Regularität der Zielfunktion nicht mehr gewährleistet, da in der Losüberlappung die Teillosgröße zu einer Entscheidungsvariablen wird (vgl. Sen et al., 1998). Dennoch können in dem linearen Programm z.B. leicht die Fertigstellungszeiten je Teillos aufgegriffen werden, indem die Differenz des Fertigstellungszeitpunkts und des Beginns der Bearbeitung je Teillos in die Zielfunktion aufgenommen wird. Sollen hingegen z.B. die mit $\underline{x}$ gewichteten Durchlaufzeiten der Teillose minimiert werden, kann die Berechung mit Hilfe des linearen Programms nicht mehr vorgenommen werden, da sich die Zielfunktion dann zu einer quadratischen Funktion entwickelt. Insgesamt ist aber festzuhalten,

dass die hier vorgeschlagenen Modelle auch hinsichtlich der Zielfunktion leicht an die praktischen Anforderungen angepasst werden können.

Zusammenfassend kann festgehalten werden:

- Erweiterungen können in das Problem P(1) problemlos einbezogen werden.
- Durch Runden der nicht-ganzzahligen Lösung gehen Eigenschaften wie Leerzeit- oder Wartezeitfreiheit regelmäßig verloren.
- Eine allgemeingültige Abschätzung für den Fehler, der sich durch das Runden der nicht-ganzzahligen Lösungen ergibt, ist nicht zu erwarten.
- Der Rechenaufwand steigt bei ganzzahligen Losen sehr stark an. Große Instanzen sind daher mit prognostizierbarem Aufwand nicht optimal lösbar.

Ein Ausweg kann in der Entwicklung geeigneter Heuristiken und Meta-Heuristiken liegen.

5.3.5 Rescheduling bei konsistenten Losen

Wurde in den bisherigen Betrachtungen durchgängig davon ausgegangen, dass die Analyse a priori stattfindet, wird nun eine Erweiterung vorgenommen, die als besonders praxisrelevant einzuschätzen ist: Der Eintritt einer Verzögerung im Produktionsprozess soll nicht schon im Vorhinein bekannt und damit antizipierbar sein, sondern erst im Verlauf der Bearbeitung, also nach der Freigabe der ersten Lose stattfinden. Zu betrachten ist, wie mit einem modifizierten linearen Programm ein sogenanntes Rescheduling (vgl. Graves, 1981; Bierwirth/Mattfeld, 1999; Vieira et al., 2003), also die Anpassung des restlichen Plans an die Störung oder Verzögerung bei bereits erfolgter Freigabe der ersten Lose möglich wird. Diese Maßnahme zur dynamischen Anpassung des Plans ist insbesondere unter dem Aspekt der sich dabei gegebenenfalls ändernden kritischen Pfade von Interesse. Gesucht ist somit die optimale, durchlaufzeitminimale Lösung mit konsistenten Losen in Reaktion auf eine zwischenzeitliche Störung. Dabei ist der Zeitpunkt wesentlich, in dem die Störung eintritt. Die bereits eingelasteten Lose können nicht mehr in der Größe verändert werden, da sonst die Eigenschaft des Plans, nur konsistente Lose zu verwenden, verloren geht. Nimmt man zur Verdeutlichung an, dass im Beispiel B ($\underline{p}$ = 10; 10; 2; 8; 4; 1; 4; 2 und Q = 40 mit der optimalen Durchlaufzeit von 470,95 ZE) im Zeitpunkt 100 eine Störung auftritt, dann sind alle bis zu diesem Zeitpunkt bereits eingelasteten Losgrößen und die schon eingetretenen Fertigstellungstermine nicht mehr veränderbar. Im vorliegenden Fall sind dies die Lose x_1, x_2 und x_3. Diese sind in der Tabelle 13 kursiv hervorgehoben. Weiter wird angenommen,

dass bei der Fertigung des zweiten Loses auf der dritten Stufe eine Störung auftritt, die den Fertigstellungstermin vom Zeitpunkt 139,83 auf den Zeitpunkt 250 verzögert. Das zweite Los wird somit um 110,17 ZE später an die nächste Stufe weitergegeben als ursprünglich vorgesehen. Wie die zugehörigen Dualvariablen aufzeigen, liegt diese Weitergabe nicht auf dem kritischen Pfad. Die Verzögerung wirkt sich somit bis auf die letzte Stufe nur abgeschwächt aus. Werden versuchsweise die Lose (trotz der Störung) in der ursprünglich ermittelten Größe eingelastet, steigt die Durchlaufzeit des Plans von 470,95 ZE auf 546,04 ZE an. Die Erhöhung der Durchlaufzeit um 75,09 ZE war zu erwarten, da die betroffene Restriktion in der unverzögerten Lösung einen Schlupf von 34,95 ZE aufweist. Die Differenz 110,17 – 34,95 = 75,22 ≠ 75,09 ist darauf zurückzuführen, dass die Losgrößen $\underline{x}$ auf zwei Nachkommastellen gerundet für die Berechungen verwendet wurden.

In dieser Situation setzt nun das Rescheduling ein. Die bereits realisierten Termine werden über zusätzliche Gleichungen in das lineare Modell aufgenommen und über die noch nicht eingelasteten Lose kann die Durchlaufzeit minimiert werden.

Um die zum Rescheduling nötigen Berechnungen durchführen zu können, wird in dem Moment, in dem die Störung auftritt, ermittelt, welche Lose bereits freigegeben sind. Außerdem sind die Fertigstellungstermine für die bereits abgearbeiteten oder im Störungszeitpunkt auf den einzelnen Stufen aktuell bearbeiteten Lose vorzugeben. Die Losgrößen und die zugehörigen Fertigstellungstermine müssen im Planungsmodell mit Hilfe von „=“ bzw. „≥“ Restriktionen festgesetzt werden. Die Verwendung von „≥“ Restriktionen bietet sich für die Festsetzung der bereits realisierten Termine an, da bei „=“ Restriktionen durch Rundungsfehler leicht Unzulässigkeiten auftreten. Schließlich ist der – oft nur per Schätzung ermittelbare – Fertigstellungstermin für das Los festzulegen, das sich durch die Störung verzögert. Es sind somit zusätzliche Restriktionen in das Modell aufzunehmen. Für das vorliegende Beispiel soll die Störung im Zeitpunkt 100 aufgetreten sein, so dass folgende Restriktionen eingeführt werden:

$$x_1 = 4{,}37 \qquad x_2 = 4{,}37 \qquad x_3 = 4{,}37$$

$$C(1,1) \geq 43{,}69 \quad C(1,2) \geq 87{,}39 \quad C(1,3) \geq 96{,}13$$

$$C(2,1) \geq 87{,}39 \quad C(2,2) \geq 131{,}09 \quad C(2,3) \geq 250 \quad C(3,1) \geq 131{,}09$$

Es genügt an dieser Stelle nicht, nur den Fertigstellungstermin für das verzögerte Los einzugeben, da sich sonst auf Grund der dualen degenerierten Lösung Verschiebungen im Plan einstellen können, die für bereits abgearbeitete oder gegenwärtig bearbeitete Lose Fertigstellungstermine ausweisen, die nicht mehr realisierbar sind. Dies gilt insbesondere, wenn z.B. leerzeitfreie Pläne gesucht werden.

Es zeigt sich, dass dem Ausfall der Stufe 3 durch eine Änderung der Losgrößen im restlichen Plan begegnet werden kann. Tabelle 13 gibt eine Gegenüberstellung der ursprünglichen optimalen Losgrößen mit den nachträglich angepassten Losgrößen x_4 bis x_{12} an.

Tabelle 13: **Losgrößen vor und nach dem Rescheduling**

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}
4,37	*4,37*	*4,37*	4,37	4,37	4,37	3,64	3,03	2,53	2,11	1,46	1,01
4,37	*4,37*	*4,37*	8,58	5,94	4,11	2,85	1,98	1,37	0,95	0,65	0,45

Der Zielfunktionswert fällt in Folge des Rescheduling von 546,03 ZE auf 540,05 ZE. Durch die Anpassung der Lose x_4 bis x_{12} ist es damit gelungen, die ohne eine Anpassung auftretende Verzögerung um etwa 6 ZE zu reduzieren. Die Auswirkungen der hier gesetzten Verzögerung und damit der Erfolg des Rescheduling hängen allerdings stark vom Zeitpunkt des Auftretens der Störung und vom Verlauf der kritischen Pfade ab. Betrifft die Störung (wie im obigen Beispiel) eine Kante, für die erheblicher Schlupf ausgewiesen ist, wirkt sich die Verzögerung nur in Teilen auf die Durchlaufzeit des Plans aus. Betrifft sie eine kritische Kante oder einen kritischen Knoten, kann die resultierende Verzögerung deutlich größer ausfallen.

Während die bisher betrachtete Anpassung wesentliche Parameter des Problems unverändert ließ, wird nun zugelassen, dass im Zuge des Rescheduling auch Änderungen der Parameter zulässig sind. Insbesondere ist zu betrachten, wie über zusätzliche Lose eine aufgetretene Störung auszugleichen ist. Kommt diese Variante in Frage, muss lediglich im LINGO Kode die Zahl der Lose angepasst werden. Im ersten Data-Block wird die Loszahl *S* von bisher 12 auf z.B. 15 erhöht.

```
! Eingabe der Loszahl S, Stufenzahl O und Produktionsmenge Q;
DATA:
S = 15; O = 8; Q = 40;
ENDDATA
```

Analog zum bisherigen Vorgehen sind die bereits zur Produktion freigegebenen Losgrößen und die bereits feststehenden Fertigstellungstermine der Lose über zusätzliche Restriktionen einzugeben.

Betrachten wir zur Verdeutlichung das folgende Szenario: Gegeben ist eine Produktionsstruktur mit $M = 5$ Stufen und $\underline{p} = (2; 4; 4; 3; 1)$ auf der die vorgegebene Losgröße $Q = 300$ in konsistenten Losen hergestellt werden soll. Als Liefertermin ist ein Zeitpunkt vereinbart worden, der 2000 ZE in der Zukunft liegt. Da mit vier (zudem ganzzahligen) Losen der Größe ($x_1 = 80$, $x_2 = 80$, $x_3 = 80$; $x_4 = 60$) eine Durchlaufzeit von 1920 ZE realisierbar ist, wurde entschieden, diese Lösung zu verwenden. Nach Einlastung des ersten – aber noch vor der Einlastung des zweiten – Loses fragt der Kunde an, ob eine frühere Lie-

ferung möglich ist. Die folgende Aufstellung zeigt, wie bei einem bereits freigegebenen ersten Los der Größe 80 durch zusätzliche Aufteilung der restlichen Menge frühere Liefertermine erreichbar werden:

Tabelle 14: **Konsequenzen aus der Einführung zusätzlicher Lose**

Lose	C_{Min}	$\underline{x}$									
4	1920	80	80	80	60						
5	1816	80	80	60,54	45,40	34,05					
6	1771	80	72,11	54,08	40,56	30,42	22,81				
7	1743	80	66,90	50,18	37,63	28,22	21,17	15,87			
8	1725	80	63,47	47,60	35,70	26,77	20,08	15,06	11,29		
9	1712	80	61,11	45,83	34,37	25,78	19,33	14,50	10,87	8,16	
10	1704	80	60	45	33,75	25,31	18,98	14,23	10,67	8,00	4,02

Werden 12 Lose eingesetzt, ist bereits die Durchlaufzeit von 1700,3 ZE erreichbar. Dieses Ergebnis ist bemerkenswert, da die mit konsistenten Losen und $x_1 = 80$ minimal erreichbare Durchlaufzeit bei näherungsweise 1700 ZE liegt. Diese untere Grenze ergibt sich aus der Tatsache, dass das erste Los frühestens im Zeitpunkt 800 auf der vierten Stufe eingelastet werden kann ($80 \cdot 2 + 80 \cdot 4 + 80 \cdot 4) = 800$. Wird ab diesem Zeitpunkt die vierte Stufe ohne Unterbrechung genutzt, erhält man $300 \cdot 3 + 800 = 1700$ ZE. Wird das letzte Lose entsprechend klein gewählt, ergibt sich ein gerundeter Wert von 1700,00 ZE. Bei 16 Losen wäre z.B. $x_{16} = 0{,}0036$ Einheiten groß. Insofern existiert hier für die maximale Verkürzung durch Einfügen zusätzlicher Lose ein Grenzwert.

Wie weiter aus Tabelle 14 ersichtlich ist, kann im oben geschilderten Szenario dem Kunden ein Liefertermin im Intervall [1704; 1920] angeboten werden, indem die Zahl der Lose zwischen 4 und 10 variiert wird. Wieder wird der Disponent die Konsequenzen der zusätzlichen Lose (also z.B. zusätzliche Reinigungs- oder Rüstarbeiten) mit dem Effekt vergleichen, den er durch die Reduktion der Durchlaufzeit erreichen kann.

Eine weitere für die Praxis relevante Variante der Anpassung kann darin bestehen, dass die benötigte Menge Q durch zusätzliche Aufträge erhöht, im Nachhinein reduziert werden muss oder dass die Transport- oder Bereitstellungszeiten eine Veränderung erfahren. Auch auf diese Fälle kann durch zusätzliche Restriktionen leicht reagiert werden.

Das hier gezeigte Konzept hat den Vorteil, dass es intuitiv zugänglich ist und daher eine hohe Akzeptanz der Ergebnisse erwartet werden darf. Da die Optimierung innerhalb von Sekunden erfolgen kann, ist eine Basis für eine What-If-Analyse gelegt. Der Produktionsplaner kann verschiedene Szenarien – z.B. hinsichtlich einer unterschiedlich langen Störung – erproben und die Pläne identifizieren, die zu Strukturen führen, welche trotz Störungen weitgehend stabil blei-

ben. Es wird bewusst darauf verzichtet, in das Modell weitere Einflussgrößen aufzunehmen, da sie nur dazu führen werden, dass ein Modell entsteht, das den Anforderungen Littles nicht mehr genügen kann.

Zusammengenommen zeigen die hier erarbeiteten Ergebnisse, dass aus Sicht eines Praktikers die Verwendung konsistenter Lose mit den Möglichkeiten der postoptimalen Analyse und des Rescheduling sehr vorteilhaft erscheinen. Inwiefern allerdings durch die Verwendung der konsistenten Lose Potential für eine weitere Reduktion der Durchlaufzeit ausgeschlossen wird – welches man bei Verwendung variabler Lose noch nutzen könnte – wird im dritten Schwerpunkt des fünften Kapitels behandelt.

5.4 Losüberlappung mit variablen Losen im Mehrstufenfall

Im Zwei- und im Dreistufenfall findet die Unterscheidung in Stück- oder Losverfügbarkeit kein besonderes Gewicht, da erstens mit Folgerung 1 gezeigt worden ist, dass für $M \leq 3$ die optimale Lösung mit konsistenten Losen erreichbar ist und zweitens alle Lösungen mit konsistenten Losen unter Losverfügbarkeit auch umsetzbar sind. Dehnt man die Betrachtung allerdings auf den Mehrstufenfall ($M > 3$) aus, gilt Folgerung 1 nicht mehr, weil die Argumentation über die Invers-Äquivalenz in Verbindung mit der Randkonsistenz keine gemeinsame Stufe mehr umfasst.

Gilt Stückverfügbarkeit und wird Leerzeitfreiheit gefordert, kann im Mehrstufenfall mit Hilfe der Folgerungen 14 und 15 die optimale Lösung ermittelt werden.

Folgerung 14:

Bei Stückverfügbarkeit und Leerzeitfreiheit existiert ein durchlaufzeitminimaler Plan, wenn für die Transferlosgrößen y_{im} (i = 1,..., I) gilt: $y_{im} = Q \cdot q_m^{\,i-1} \Big/ \sum_{i=1}^{I} q_m^{\,i-1}$ mit $q_m = p_{m+1}/p_m$ für m = 1,..., M–1.

Folgerung 15:

Bei Stückverfügbarkeit und variablen Losen existiert ein optimaler Plan, der eine wartezeitfreie Bearbeitung für alle Lose ermöglicht.

Da die Anwendung dieses Zusammenhangs bisher nur für den Dreistufenfall vorgestellt wurde, wird zunächst gezeigt, wie im Mehrstufenfall vorzugehen ist. Dabei wird auf ein von Potts/Baker (1989) vorgestelltes Beispiel mit vier Stufen zurückgegriffen. Gesucht wird für Q = 6 und $\underline{p}$ = (1; 1; 15; 3) eine Lösung mit 2 variablen Losen unter Stückverfügbarkeit. Mit Hilfe der Berechnungen gemäß Folgerung 14 wird die Größe der Transferlose zwischen den Stufen bestimmt und von dieser auf die Größe der jeweiligen Freigabelose geschlossen. Auf den

beiden Randstufen müssen jeweils die Größe der Transferlose und die der Freigabelose übereinstimmen. Auf den übrigen Stufen können aus laufenden Freigabelosen kleinere Transferlose entnommen werden. Um dies darzustellen, sind in dem Gantt-Diagramm der Abbildung 34 die Balken horizontal geteilt.

Für $\underline{p}$ = (1; 1; 15; 3) ergibt sich $q_1 = 1$; $q_2 = 15$; $q_3 = 0{,}2$. Die optimalen, zwischen den Stufen zu transferierenden Losgrößen sind damit:

$y_{11} = 6 \cdot 1^0 / 2 \quad = 3$ $\qquad$ $y_{21} = 6 \cdot 1/2 = 3$

$y_{12} = 6 \cdot 15^0/16 \quad = 0{,}375$ $\qquad$ $y_{22} = 6 \cdot 15/16 = 5{,}625$

$y_{13} = 6 \cdot 0{,}2^0/1{,}2 \quad = 5$ $\qquad$ $y_{23} = 6 \cdot 0{,}2/1{,}2 = 1$

Den zugehörigen Plan als Gantt-Diagramm zeigt Abbildung 34:

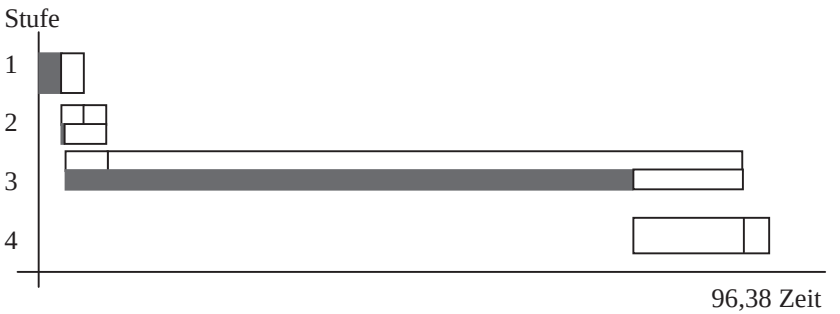


Abbildung 34: Variable Lose im Vierstufenfall bei Stückverfügbarkeit

Die minimale Durchlaufzeit beträgt 96,375 ZE. Sie ermittelt sich gemäß des folgenden Zusammenhangs: Die Leerzeit vor der Bearbeitung des gesamten Loses auf der letzten Stufe entspricht der kumulierten Bearbeitungsdauer des jeweils ersten Transferloses über die Stufen1,2,..., M–1. Diese Zeitspanne ist in der Abbildung 34 grau hervorgehoben. Formal gilt für die Durchlaufzeit bei variablen Losen unter Stückverfügbarkeit somit:

$$C_{\min} = \sum_{m=1}^{M-1} p_m y_{1m} + Q p_M \tag{59}$$

Für das vorliegende Beispiel erhält man: $3 \cdot 1 + 0{,}375 \cdot 1 + 5 \cdot 15 + 6 \cdot 3 = 96{,}375$. Diese Zeitspanne kann durch keine andere Lösung mit zwei Losen unterboten werden, solange Leerzeiten auf den Stufen unzulässig sind. Die Anwendung dieser Lösung bietet sich in der Praxis z.B. dann an, wenn der Rüstzustand auf einer Stufe durch Leerstand verloren geht und anderseits aus einer gemeinsam freigegebenen Losgröße beliebige Anteile weitergegeben werden können, d.h. die Voraussetzungen für die Stückverfügbarkeit vorliegen.

Für die weitere Diskussion der Losüberlappung im Mehrstufenfall wird nun die Annahme der Stückverfügbarkeit fallengelassen und untersucht, wie sich va-

riable Lose unter Losverfügbarkeit bestimmen lassen. Im Kern wird damit die Situation aufgegriffen, die schon im zweiten Schwerpunkt dieser Arbeit unterstellt werden konnte, allerdings mit der gelockerten Anforderung, dass statt der konsistenten Lose nun variable Lose zulässig sind. Ein Argument für den Einsatz variabler Lose ist, dass bei mehr als drei Stufen die Möglichkeit besteht, durch die Einschränkung auf konsistente Lose die optimale Lösung auszuschließen. Um diesen Zusammenhang aufzuzeigen, haben Potts/Baker (1989) das oben eingeführte Beispiel genutzt. Während unter Losverfügbarkeit die beste Lösung mit konsistenten Losen ($x_1 = 5$, $x_2 = 1$) zu einer Durchlaufzeit von 103 ZE führt, kann mit variablen Losen der Größe $x_{11} = 3$; $x_{21} = 3$; $x_{12} = 3$; $x_{22} = 3$; $x_{13} = 5$; $x_{23} = 1$; $x_{14} = 5$; $x_{24} = 1$ eine Durchlaufzeit von 102 ZE erreicht werden. Die folgende Abbildung stellt beide Lösungen gegenüber:

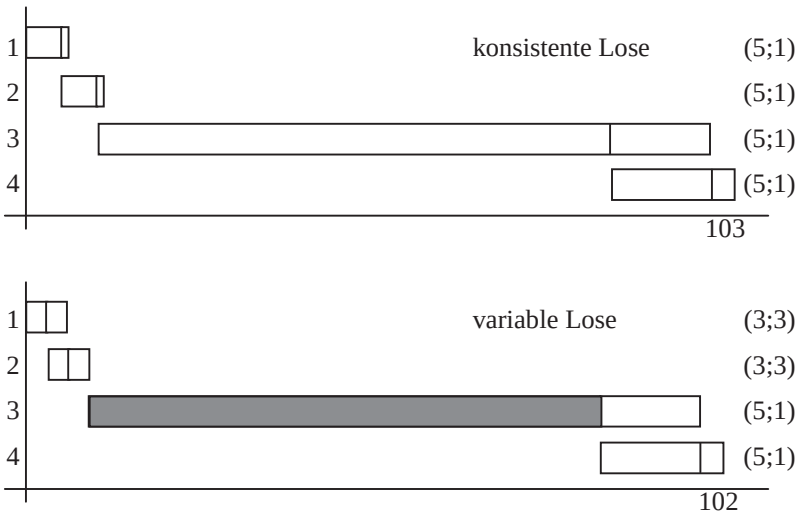


Abbildung 35: Variable versus konsistente Losgrößen bei Losverfügbarkeit

Die Lösung mit variablen Losen in Abbildung 35 zeigt die folgenden Besonderheiten auf:

- Auf Grund der Losverfügbarkeit kann das erste Freigabelos auf Stufe 3 (in Abbildung 35 grau unterlegt) erst starten, wenn die Bearbeitung des zweiten Freigabeloses auf Stufe 2 abgeschlossen ist, da dieses Los noch Teile enthält, die in das erste Freigabelos der Stufe 3 eingehen: $x_{12} < x_{13}$.
- Zwischen den Stufen 2 und 3 findet keine überlappende Fertigung statt. Entsprechend wird zwischen diesen Stufen auch nur ein Transferlos benötigt, d.h. $y_{22} = 0$.
- Aus dem Transferlos $y_{12} = 6$ werden zwei Freigabelose auf der Stufe 3 gebildet: $y_{12} = x_{13} + x_{23}$.

Diese Lösung wirkt auf den ersten Blick überraschend. Obwohl zwischen den Stufen 2 und 3 keine Überlappung vorgesehen ist und damit ein Transport ganz entfallen kann, wird die Durchlaufzeit gegenüber der Lösung mit konsistenten Losen geringfügig reduziert.

Potts/Baker (1989) konstruierten dieses Beispiel um zu zeigen, dass die optimale Lösung durch die Forderung nach konsistenten Losen ausgeschlossen werden kann. Ein Verfahren zur Lösung des mehrstufigen Losüberlappungsproblems mit variablen Losen unter Losverfügbarkeit wurde von den Autoren allerdings nicht angegeben. Diese Lücke konnte erst von Biskup/Feldmann (2005) geschlossen werden. Ihr Ansatz wird im Folgenden vorgestellt, an Hand des LINGO Kode und mehrerer Beispiele motiviert, sowie hinsichtlich der Einbeziehung von Öfen und Transporten unter Zykluszeiten erweitert. Dabei besteht eine wesentliche Zielsetzung darin, zu überprüfen, wie groß das entgangene Potential der Durchlaufzeitverkürzung ist, wenn konsistente Lose statt variabler Lose verwendet werden. Sollte sich für die hier betrachteten Instanzen zeigen, dass die Verbesserung – wie im Beispiel von Potts/Baker (1989) – mit unter einem Prozentpunkt sehr klein ausfällt, dann wäre dies ein weiteres Argument für die Verwendung konsistenter Lose: Anstatt sich auf die schwierige Abstimmung bei variablen Losen einzulassen, sollte lieber ein weiteres konsistentes Los verwendet werden. Zeigt sich aber, dass es Instanzen gibt, für die sich der Wechsel von konsistenten zu variablen Losen in einer signifikant größeren Reduktion der Durchlaufzeit bei einer unter Umständen geringeren Anzahl an Transporten niederschlägt, dann können variable Lose unter Losverfügbarkeit als sinnvolle Alternative angesehen werden. Im Folgenden wird die Modellformulierung und anschließend Ihre Umsetzung im LINGO-Kode gezeigt.

5.4.1 Variable Lose bei Losverfügbarkeit im Mehrstufenfall

Für dieses Modell wird die Notation wie folgt erweitert. Dabei ist insbesondere auf die Modellierung der Verfügbarkeit der Teile mit Hilfe der Binärvariablen δ_{jmk} hinzuweisen:

J - Zahl der Lose

j, k, u - Indices der Lose $j, k, u = 1, \ldots, J$

M - Zahl der Stufen

m - Stufenindex, $m = 1, \ldots, M$

p_m - Ausführungzeit je Stück auf Stufe m

Q - Vorgegebene Losgröße

R - Sehr große Zahl (100.000)

C_{jm} - Zeitpunkt der Fertigstellung des j-ten Freigabeloses auf Stufe m

x_{jm} - Größe des j-ten Freigabeloses auf Stufe m

δ_{jmk} - Binärvariable zur Abbildung der Verfügbarkeit der Teile:

δ_{jmk} = 1, wenn das j-te Freigabelos auf der Stufe m nicht gestartet wird, bevor das k-te Freigabelos auf der Stufe m–1 beendet ist.

δ_{jmk} = 0, sonst.

P(2): Das Losüberlappungsproblem mit variablen Losen bei Losverfügbarkeit

Min C_{JM}!

$$C_{1m} \geq p_m\, x_{1m} \qquad m = 1, ..., M \qquad (60)$$

$$C_{jm} - p_m\, x_{jm} \geq C_{j\text{-}1,m} \qquad j = 2, ..., J,\ m = 1, ..., M \qquad (61)$$

$$C_{jm} - p_m\, x_{jm} \geq C_{k,m\text{-}1} - (1 - \delta_{jmk})\, R \qquad j = 1, ..., J,\ k = 1, ..., J,\ m = 2, \ldots, M \qquad (62)$$

$$\delta_{jm1} = 1 \qquad j = 1, ..., J,\ m = 2, ..., M \qquad (63)$$

$$\delta_{jmk} \geq \delta_{jm,k+1} \qquad j = 1, ..., J,\ k = 1, ..., J-1,\ m = 2, \ldots, M \qquad (64)$$

$$x_{jm} \leq \sum_{u=1}^{k} x_{u,m-1} - \sum_{u=1}^{j-1} x_{um} + \left(1 - \delta_{jmk} + \delta_{jm,k+1}\right) R$$

$$j = 2, ..., J,\ k = 1, ..., J-1,\ m = 2, \ldots, M \qquad (65)$$

$$\sum_{j=1}^{J} x_{jm} = Q \qquad m = 1, ..., M \qquad (66)$$

$$\delta_{jmk} \in \{0,1\} \qquad j = 1, ..., J,\ k = 2, ..., J,\ m = 2, \ldots, M \qquad (67)$$

$$x_{jm},\ C_{jm} \geq 0 \qquad j = 1, ..., J,\ m = 1, ..., M \qquad (68)$$

Die Restriktion (60) legt für jede der Stufen fest, dass frühestmöglich im Zeitpunkt Null mit der Fertigung des jeweils ersten Loses auf jeder der M Stufen begonnen werden darf. Damit wird der frühestmögliche Fertigstellungszeitpunkt des ersten Freigabeloses auf allen Stufen angegeben. Restriktion (61) verhindert, dass auf einer Stufe zwei Lose gleichzeitig produziert werden: Das Freigabelos j darf erst dann starten, wenn das Freigabelose $j - 1$ auf dieser Stufe fertiggestellt wurde. Die Binärvariable δ_{jmk} nimmt genau dann den Wert 1 an, wenn das Freigabelos j auf Stufe m nicht gestartet werden darf, bevor für das Freigabelos k auf der Vorstufe ($m - 1$) die Bearbeitung abgeschlossen wurde. Die Binärvariable legt somit fest, ob Teile aus einem Freigabelos k der Vorstufe ($m - 1$) in das Freigabelos j der Folgestufe (m) eingegliedert werden dürfen. Die Variable hat den Wert 1, wenn die Teile verfügbar sind, sie nimmt den Wert 0 an, wenn das Los nicht berücksichtigt werden darf, da seine Bearbeitung auf der Stufe $m - 1$ noch andauert. Falls die Binärvariable den Wert 0 annimmt, ist die

Restriktion (62) nicht bindend, da R auf der rechten Seite der Ungleichung subtrahiert wird. Die Restriktionen (63) und (64) sind notwendig, um die Binärvariablen zu kontrollieren: In Restriktion (63) wird sichergestellt, dass für jedes Freigabelos j auf den Stufen $m = 2,..., M$ die Teile aus dem jeweils ersten Freigabelos der Vorstufe verfügbar sind, bevor über die Größe des Loses j entschieden wird. Dieser Zusammenhang lässt sich leichter nachvollziehen, wenn man sich verdeutlicht, dass sich für jedes Los j eine Sequenz von Binärvariablen ergibt. Angenommen, wir betrachten das vierte von maximal acht Freigabelosen auf der dritten Stufe: $j = 4$; $J = 8$, $m = 3$, dann könnte eine Sequenz für die $k = 8$ Lose der Vorstufe (also der Stufe 2) lauten: $\underline{\delta}_{43}$ = 1,1,1,1,0,0,0,0. Dabei ist z.B. $\delta_{432} = 1$ so zu interpretieren, dass das zweite Freigabelos der Stufe 2 für die Bildung des vierten Loses auf der Stufe 3 berücksichtigt wird. Dieses Freigabelos kann also nicht starten, bevor die Fertigung des letzten Freigabeloses der Vorstufe abgeschlossen ist, das noch Teile umfasst, die in das vierte Los der dritten Stufe eingehen sollen. Weiter ist ersichtlich, dass für den angegebenen Vektor der Binärvariablen genau ein Wechsel von 1 auf 0 zulässig ist. Im oben genannten Beispiel findet dieser Wechsel vom vierten zum fünften Freigabelos der Vorstufe statt ($\delta_{434} = 1$ und $\delta_{435} = 0$). Er ist so zu interpretieren, dass das fünfte Freigabelos der Stufe 2 nicht – wie die ersten vier Freigabelose – für die Bildung des vierten Freigabeloses der Stufe 3 verfügbar ist ($\delta_{431} = 1$; $\delta_{432} = 1$; $\delta_{433} = 1$; $\delta_{434} = 1$ und $\delta_{435} = 0$). Sollte mehr als ein Wechsel in der Sequenz der Binärvariablen eines Loses auftreten, dann hat dies eine in sich widersprüchliche Bedeutung. Der Vektor 1,1,0,0,0,1,0,0 führt beispielsweise dazu, dass das dritte bis fünfte Los nicht verfügbar ist, dafür aber das sechste. Da die Lose drei bis fünf vor dem sechsten Los produziert sein müssen, liegt hier ein Widerspruch vor, den die Restriktion (64) ausschließt.

Generell kann die Größe eines Freigabeloses nicht die Summe der Teile in den Freigabelosen auf der Vorstufe überschreiten, die zu seiner Bildung verfügbar sind. Außerdem sind von dieser Summe an verfügbaren Teile diejenigen zu subtrahieren, die auf der Stufe bereits freigegeben wurden. Dies geschieht in der Restriktion (65). Diese Restriktion ist lediglich dann bindend, wenn $\delta_{jmk} = 1$ und $\delta_{jm,k+1} = 0$ gilt. Sie bildet somit den Moment des Wechsels in der Verfügbarkeit ab. Auf jeder Stufe sind Q Teile zu fertigen (Restriktion 66). Die Wertebereiche für die Variablen werden in den Restriktionen (67 und 68) festgelegt.

Um die Zahl der möglichen Anordnungen innerhalb der Binärvariablen δ_{jmk} zu reduzieren und den Suchraum zu verkleinern kann der folgende Zusammenhang ausgenutzt werden. Wird z.B. das dritte Freigabelos auf der zweiten Stufe gestartet, nachdem das zweite Freigabelos auf der ersten Stufe fertiggestellt wurde, nimmt die Variable δ_{322} den Wert 1 an. Dann stehen die Teile des ersten Freigabeloses der ersten Stufe aber auch für die nachfolgenden Freigabelose der zweiten Stufe zur Verfügung. Es muss somit aus $\delta_{322} = 1$ folgen, dass die Vari-

ablen δ_{422}, δ_{522}, …, δ_{J22} ebenfalls den Wert 1 annehmen müssen. Dieser Zusammenhang wird mit der folgenden Restriktion erfasst:

$$\delta_{jmk} \leq \delta_{j+1,mk} \qquad j = 1, ..., J-1,\ k = 2, ..., J,\ m = 2, \ldots, M \tag{69}$$

Außerdem muss für das letzte Los J auf jeder Stufe gelten, dass alle k Lose der Vorstufe verfügbar sein müssen, bevor das Los J aufgelegt wird. Würden Lose der Vorstufe erst später verfügbar, könnten sie nicht in das Los J eingehen und folglich wären auf der Stufe m nicht alle Q Stück bearbeitet:

$$\delta_{jmk} = 1 \qquad k = 1, ..., J,\ m = 2, \ldots, M \tag{70}$$

Weiterhin kann hinsichtlich der Größe der Freigabelose auf den ersten beiden und auf den letzten beiden Stufen die für einen optimalen Plan unter Losverfügbarkeit gültige Eigenschaft der Randkonsistenz ausgenutzt werden, die als Theorem 1 im Abschnitt 5.2 formuliert und beweisen wurde.

$$x_{j1} = x_{j2} \qquad j = 1, ..., J$$

$$x_{j,M-1} = x_{jM} \qquad j = 1, ..., J \tag{71}$$

In Konsequenz der Restriktion (71) können aber die Binärvariablen auf der zweiten und der letzten Stufe entfallen, d.h. δ_{j2k} und δ_{jMk} werden nicht benötigt. Die Restriktionen (62) bis (64) und (67) bis (70) müssen somit nur noch für $m = 3,..., M-1$ formuliert werden. Um die Verrechnung der Bearbeitung der Lose auf der zweiten und der letzten Stufe richtig abzubilden, sind für die entfallenen Restriktionen die schon aus dem Modell mit konsistenten Losen bekannten Restriktionen (72) und (73) aufzunehmen:

$$C_{j2} - p_2 x_{j2} \geq C_{j,1} \qquad j = 1, ..., J \tag{72}$$

$$C_{jm} - p_m x_{jm} \geq C_{j,m-1} \qquad j = 1, ..., J;\ M-1 \leq m \leq M \tag{73}$$

Durch diese Vereinfachungen wird zwar der zu bewältigende Verzweigungsbaum erheblich reduziert, dennoch ist bei der Anwendung auf größere Probleme von einem erheblichen und mit der Problemgröße stark anwachsenden Rechenaufwand auszugehen. Im Folgenden wird das Modell in seiner Formulierung als LINGO-Kode präsentiert.

5.4.2 LINGO-Kode und Beispiel

Analog zu der Vorgehensweise im zweiten Schwerpunkt ist der LINGO-Kode für das Beispiel C formuliert. In einem fünfstufigen Problem mit $\underline{p}$ = (2, 4, 6, 2, 3) und Q = 30 wird die durchlaufzeitminimale Lösung bei variablen Losen gesucht. Da die Losgrößen über die Stufen variieren, werden sie zweifach indiziert: Losgroesse(Los,Stufe). Die Binärvariablen δ_{jmk} sind mit d(j,m,k) bezeichnet:

```
! Eingabe der Loszahl S, Stufenzahl O und Produktionsmenge Q;
DATA: S = 3; O = 5; Q = 30; ENDDATA

! Definieren der Variablen, Indices und Konstanten;
SETS: Stufe /1..O/: m; Los /1..S/: j,k,u; Ausfuehrungszeit(Stufe): p;
Fertigstellung(Los,Stufe): C; Losgroesse(Los,Stufe): X;
Binaervariable(Los, Stufe, Los): d; ENDSETS

! Zuweisen der p-Werte aller Stufen als Vektor;
DATA: p = 2 4 6 2 3; R = 100000; ENDDATA

! Zielfunktion;
min = C(S, O);

! Restriktion (60);
@FOR(Stufe(m):C(1,m) >= p(m)*X(1,m));

! Restriktion (61);
@FOR(Los(j)|j #GE# 2: @FOR(Stufe(m): C(j,m) >= C(j-1,m) + p(m)*X(j,m)));

! Restriktion (62);
@FOR(Los(j):@FOR(Los(k): @FOR(Stufe(m)|(m#GE#3)#AND#(m#LE#m_max-1):
C(j,m) >= C(k,m-1) + p(m)*X(j,m) - (1-d(j,m,k))*R ));

! Restriktion (63);
@FOR(Los(j): @FOR(Stufe(m)|(m#GE#3)#AND#(m#LE#m_max-1):
d(j,m,1) = 1));

! Restriktion (64);
@FOR(Los(j): @FOR(Los(k)|k #GE# 2:
@FOR(Stufe(m)|(m#GE#3)#AND#(m#LE#m_max-1): d(j,m,k-1) >=d(j,m,k))));

! Restriktion (65);
@FOR(Los(j): @FOR(Los(k)|k #LT# S:
@FOR(Stufe(m)|(m#GE#3)#AND#(m#LE#m_max-1):
X(j,m) <= @SUM(Los(u)|u #LE# k: X(u,m-1)) - @SUM(Los(u)|u #LT# j:
X(u,m)) + (1 - d(j,m,k) + d(j,m,k+1))*R)));

! Restriktion (66);
@FOR(Stufe(m): @SUM(Los(j): X(j,m)) = Q);

! Restriktion(67);
@FOR(Los(j): @FOR(Los(k)|k #GE# 2:
@FOR(Stufe(m)|(m#GE#3)#AND#(m#LE#m_max-1): @BIN(d(j,m,k)))));

! Restriktion(69);
@FOR(Los(j)|j #LE# S-1: @FOR(Los(k)|k #GE# 2:
@FOR(Stufe(m)|(m#GE#3)#AND#(m#LE#m_max-1): d(j,m,k) <=d(j+1,m,k))));

! Restriktion(70);
@FOR(Stufe(m)|(m#GE#3)#AND#(m#LE#m_max-1): @FOR(Los(k): d(S,m,k) = 1));

! Restriktion (71);
@FOR(Los(j): X(j,1) = X(j,2));
@FOR(Los(j): X(j,O-1) = X(j,O));

! Restriktion (72);
@FOR(Los(j): @FOR(Maschine(m)|m#LT#2:
C(j,m+1) >= C(j,m) + p(m+1)*L(j,m+1)));

! Restriktion (73);
@FOR(Los(j): @FOR(Maschine(m)|m#GE#m_max-1:
C(j,m) >= C(j,m-1) + p(m)*L(j,m)));
END
```

DOI https://doi.org/10.3790/978-3-428-51857-9

Bei der Formulierung der Summationsgrenzen werden #GE#, #GT#, #LE# und #LT# für ≥, >, ≤, < verwendet. Um in einer Summe den Index in beiden Richtungen begrenzen zu können, wird der Operator #AND# eingesetzt. Die übrige Syntax folgt dem LINGO-Kode für konsistente Lose, wobei die Binärvariablen mit @Bin festgelegt werden. Um das Modell im LINGO-Kode überschaubar zu halten, wird die Zahl der Lose auf $J = 3$ beziehungsweise $S = 3$ reduziert.

Tabelle 15: **Restriktionen 60 und 61 im ausformulierten Modell**

Restriktionen		#
– 2 X(1,1) + C(1,1) >= 0 – 4 X(1,2) + C(1,2) >= 0 – 6 X(1,3) + C(1,3) >= 0 – 2 X(1,4) + C(1,4) >= 0 – 3 X(1,5) + C(1,5) >= 0		(60)
– 2 X(2,1) – C(1,1) + C(2,1) >= 0 – 4 X(2,2) – C(1,2) + C(2,2) >= 0 – 6 X(2,3) – C(1,3) + C(2,3) >= 0 – 2 X(2,4) – C(1,4) + C(2,4) >= 0 – 3 X(2,5) – C(1,5) + C(2,5) >= 0 – 2 X(3,1) – C(2,1) + C(3,1) >= 0 – 4 X(3,2) – C(2,2) + C(3,2) >= 0 – 6 X(3,3) – C(2,3) + C(3,3) >= 0	– 6 X(2,3) – C(1,2) + C(2,3) >= 0 – 2 X(2,4) – C(2,3) + C(2,4) >= 0 – 2 X(2,4) – C(1,3) + C(2,4) >= 0 – 3 X(2,5) – C(2,4) + C(2,5) >= 0 – 6 X(3,3) – C(3,2) + C(3,3) >= 0 – 6 X(3,3) – C(1,2) + C(3,3) >= 0 – 2 X(3,4) – C(1,3) + C(3,4) >= 0 – 2 X(3,4) – C(3,3) + C(3,4) >= 0	(61)

Bereits beim Aufstellen der Restriktionen kann das System einige Vereinfachungen vornehmen. Beispielsweise werden die Binärvariablen als Integer Variablen mit dem Wertebereich ≤ 1 vereinbart und ihre Festlegung nur für die Hälfte der Binärvariablen vorgenommen, da sich für die übrigen Variablen implizit die Einschränkung ergibt.

Tabelle 16: **Restriktionen 64 bis 70 im ausformulierten Modell**

Restriktionen	#	Restriktionen	#
d(1,3,2) – d(2,3,2) <= 0 d(1,4,2) – d(2,4,2) <= 0 d(1,3,3) – d(2,3,3) <= 0 d(1,4,3) – d(2,4,3) <= 0	(69)	d(2,3,2) <= 1 d(2,3,3) <= 1 d(2,4,2) <= 1 d(2,4,3) <= 1	(67)
– d(1,3,2) >= – 1 – d(1,4,2) >= – 1 – d(2,3,2) >= – 1 – d(2,4,2) >= – 1	(64) für d(j,m,1) = 1	X(1,1) – X(1,2) = 0 X(2,1) – X(2,2) = 0 X(3,1) – X(3,2) = 0 X(1,4) – X(1,5) = 0 X(2,4) – X(2,5) = 0 X(3,4) – X(3,5) = 0	(70)
d(1,3,2) – d(1,3,3) >= 0 d(1,4,2) – d(1,4,3) >= 0 d(2,3,2) – d(2,3,3) >= 0 d(2,4,2) – d(2,4,3) >= 0	(64)		

Tabelle 17: **Restriktionen 62, 65 und 67 im ausformulierten Modell**

Restriktionen	#
X(1,1) + X(2,1) + X(3,1) = 30 X(1,2) + X(2,2) + X(3,2) = 30 X(1,3) + X(2,3) + X(3,3) = 30 X(1,4) + X(2,4) + X(3,4) = 30 X(1,5) + X(2,5) + X(3,5) = 30	(62)
– 100000 d(1,3,2) – 6 X(1,3) + C(1,3) – C(2,2) >= – 100000 – 100000 d(1,4,2) – 2 X(1,4) + C(1,4) – C(2,3) >= – 100000 – 100000 d(1,3,3) – 6 X(1,3) + C(1,3) – C(3,2) >= – 100000 – 100000 d(1,4,3) – 2 X(1,4) + C(1,4) – C(3,3) >= – 100000 – 100000 d(2,3,2) – 6 X(2,3) – C(2,2) + C(2,3) >= – 100000 – 100000 d(2,4,2) – 2 X(2,4) – C(2,3) + C(2,4) >= – 100000 – 100000 d(2,3,3) – 6 X(2,3) + C(2,3) – C(3,2) >= – 100000 – 100000 d(2,4,3) – 2 X(2,4) + C(2,4) – C(3,3) >= – 100000	(65) für d(j,m,1) = 1
– 100000 d(1,3,2) – X(1,2) + X(1,3) <= 0 – 100000 d(1,4,2) – X(1,3) + X(1,4) <= 0 – 100000 d(2,3,2) – X(1,2) + X(1,3) + X(2,3) <= 0 – 100000 d(2,4,2) – X(1,3) + X(1,4) + X(2,4) <= 0	(65) für d(J,m,1) = 1
100000 d(1,3,2) – 100000 d(1,3,3) – X(1,2) + X(1,3) – X(2,2) <= 100000 100000 d(1,4,2) – 100000 d(1,4,3) – X(1,3) + X(1,4) – X(2,3) <= 100000	
– X(1,2) + X(1,3) + X(2,3) + X(3,3) <= 100000 – X(1,3) + X(1,4) + X(2,4) + X(3,4) <= 100000 – X(1,2) + X(1,3) – X(2,2) + X(2,3) + X(3,3) <= 100000 – X(1,3) + X(1,4) – X(2,3) + X(2,4) + X(3,4) <= 100000	(67)
INTE d(1,3,2), d(1,3,3), d(1,4,2), d(1,4,3), d(2,3,2), d(2,3,3), d(2,4,2), d(2,4,3)	

Auf Grund der Vorgabe in der SETS-Anweisung legt LINGO zunächst 45 Variable d(j,m,k) an. Durch Restriktion (67) werden dann $(M - 3)(J^2 - J) = 12$ Binärvariable vereinbart, wobei durch Restriktion (63) und (70) ein Drittel dieser Binärvariablen insgesamt auf den Wert 1 festgelegt wird. Damit vereinfachen sich die Restriktionen (64) und (65). Restriktion (63) kann entfallen, da sie implizit durch diese vereinfachten Restriktionen erzwungen wird. Es verbleiben somit 8 Binärvariable, über die zu optimieren ist. Hinzu kommen $J\,(M - 2) = 9$ festzulegende Losgrößen. Da in Restriktion (66) gefordert ist, dass sich die Lose zu Q addieren, bestimmt sich zwar jeweils eine der Losgrößen über die Summe der übrigen, doch treten die Terme (z.B. x_{11}, x_{21}, x_{31}) in keiner der anderen Restriktionen so miteinander verbunden auf, dass hier eine weitere Vereinfachung möglich wäre. Löst man das gegebene Problem für drei Lose ($J = 3$), erhält man den in Abbildung 36 dargestellten Plan.

Im oberen Teil der Abbildung 36 ist die Lösung mit drei konsistenten Losen (x_1= 8,7; x_2 = 13,1; x_3 = 8,2) angegeben. Ihr wird im unteren Teil der Abbildung die Lösung mit drei variablen Losen gegenübergestellt, dabei sind die – auf

zwei Nachkommastellen gerundeten – Losgrößen im Gantt-Diagramm angegeben.

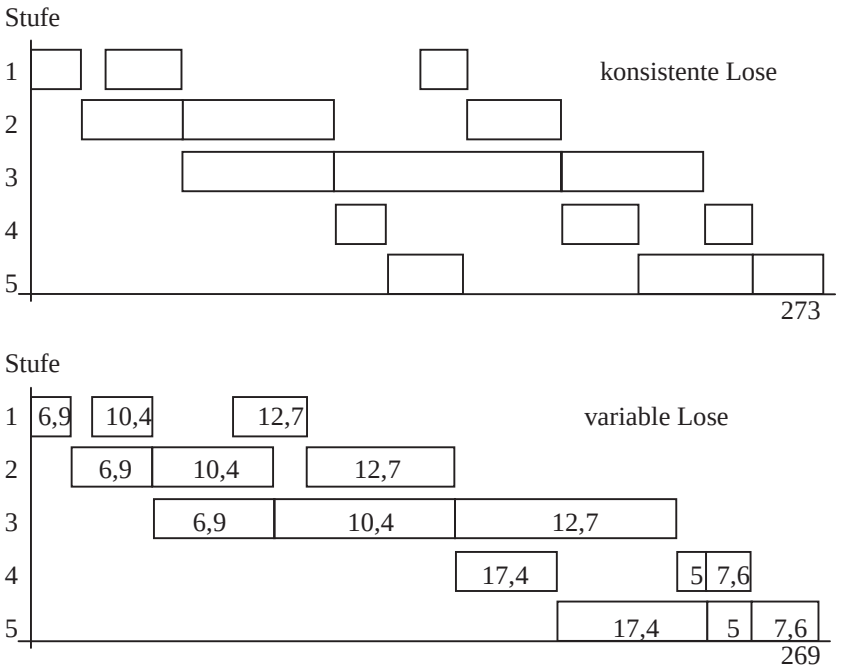


Abbildung 36: Variable Lose im Vergleich zu konsistenten Losen im Beispiel C

Die Zusammenfassung und Aufteilung im Übergang von der dritten zur vierten Stufe kann unmittelbar nachvollzogen werden. Das erste Freigabelos der vierten Stufe setzt sich aus den ersten beiden Losen der Stufe 3 zusammen, wozu die Binärvariablen die folgenden Werte annehmen: $\delta_{141} = \delta_{142} = 1$ und $\delta_{143} = 0$. Da Teile des dritten Freigabeloses der dritten Stufe in das zweite und in das dritte Freigabelos der vierten Stufe eingehen, lauten die zugehörigen Binärvariablen: $\delta_{242} = \delta_{243} = 1$ und $\delta_{342} = \delta_{343} = 1$.

Wie im Beispiel von Potts/Baker (1989) führt die Verwendung von variablen Losen im Vergleich zur Lösung mit konsistenten Losen nur zu einer geringen Reduktion der Durchlaufzeit – allerdings kann wiederum auf einen Transport (zwischen der dritten und der vierten Stufe) verzichtet werden. Im folgenden Abschnitt werden einige Erweiterungen des Modells vorgestellt und anschließend anhand einer Variation der Zahl der Lose geprüft, ob sich bei steigender Anzahl an Freigabelosen eine Vergrößerung/Verkleinerung der Verbesserung gegenüber den konsistenten Losen ergibt.

5.4.3 Modellerweiterungen

Zur Berücksichtigung von Leerzeitfreiheit wird analog zu den Modifikationen des linearen Modells in Restriktion (61) Gleichheit gefordert. Eine Berücksichtigung der Wartezeitfreiheit hingegen ist in diesem Modell weder sinnvoll noch möglich. Sie steht zum einen im Widerspruch zu der Anforderung, bestehende Lose zusammenfassen zu können, zum anderen ist Restriktion (62) – fordert man dort Gleichheit – nicht mehr lösbar, sobald eine der Binärvariablen den Wert 1 annimmt. Da Restriktion (63) die Binärvariable hinsichtlich der Verfügbarkeit des ersten Loses auf den Wert 1 festlegt, führt diese Modifikation zu einem nicht lösbaren Modell.

Die Erweiterung des Modells um antizipative Rüstvorgänge kann wie folgt vorgenommen werden:

$$C_{1m} \geq p_m x_{1m} + \omega_{jm} a_m \qquad m = 1, ..., M \qquad (74)$$

$$C_{jm} - p_m x_{jm} \geq C_{j-1,m} + \omega_{jm} a_m \qquad j = 2, ..., J,\ m = 1, ..., M \qquad (75)$$

$$\omega_{Jm} = 1 \qquad m = 1, ..., M \qquad (76)$$

$$x_{jm} - (\omega_{jm} R) \leq 0 \qquad j = 1, ..., J,\ m = 1, ..., M \qquad (77)$$

$$x_{Jm} - (\omega_{Jm} / R) \geq 0 \qquad m = 1, ..., M \qquad (78)$$

$$\omega_{jm} \in \{0,1\} \qquad j = 1, ..., J,\ m = 1, ..., M \qquad (79)$$

Dabei bezeichnet a_m die für den antizipativen Rüstvorgang benötigte Zeitspanne, die auch als Bereitstellungszeit der Stufe m für Reinigungs- oder Wartungsarbeiten interpretiert werden kann. Die Binärvariable ω_{jm} nimmt den Wert 1 an, wenn das zugehörige Los $x_{jm} > 0$ ist (Restriktion 77). Diese Binärvariable wird benötigt, damit genau dann ein Rüstprozess in den Restriktionen (74) und (75) berücksichtigt wird, wenn $x_{jm} > 0$ gilt. Weichen nämlich die Rüstzeiten über die Stufen stark voneinander ab, ist es regelmäßig vorteilhaft, einige der Lose vollständig entfallen zu lassen, d.h. ihre Größe auf Null zu setzen und damit Rüstzeiten einzusparen. In Restriktion (76) wird festgelegt, dass zumindest für das letzte der Freigabelose auf jeder Stufe zu rüsten ist. Da auf jeder Stufe mindestens ein Los aufgelegt werden muss – ansonsten kann die Menge Q nicht produziert werden – bedeutet es keine Einschränkung, wenn verlangt wird, dass das Los J auf jeder Stufe existiert ($x_{Jm} > 0$). Der in beiden Richtungen benötigte Zusammenhang: $\omega_{Jm} = 1 <=> x_{Jm} > 0$ wird in den Restriktionen (77) und (78) erzwungen. Dazu wird mit R eine sehr große Zahl (100000) verwendet.

Wird in dem oben gezeigten Beispiel mit $J = 4$ Losen $a_m = (0; 4; 5; 7; 0)$ festgelegt, d.h. nur auf den Stufen 2, 3 und 4 ist ein Rüstvorgang erforderlich, ergibt sich als Durchlaufzeit 251,53 ZE, wenn die folgenden (auf zwei Nachkommastellen gerundeten) variablen Lose verwendet werden:

$x_{11} = x_{12} = 1{,}05$ $x_{21} = x_{22} = 5{,}05$ $x_{31} = x_{32} = 9{,}41$ $x_{41} = x_{42} = 14{,}49$

$x_{13} = 6{,}10$ $x_{23} = 9{,}41$ $x_{33} = 9{,}01$ $x_{43} = 5{,}48$

$x_{14} = x_{15} = 15{,}51$ $x_{24} = x_{25} = 9{,}01$ $x_{34} = x_{35} = 3{,}59$ $x_{44} = x_{45} = 1{,}88$

Im folgenden Gantt-Diagramm ist die zum Rüsten benötigte Zeitspanne grau unterlegt.

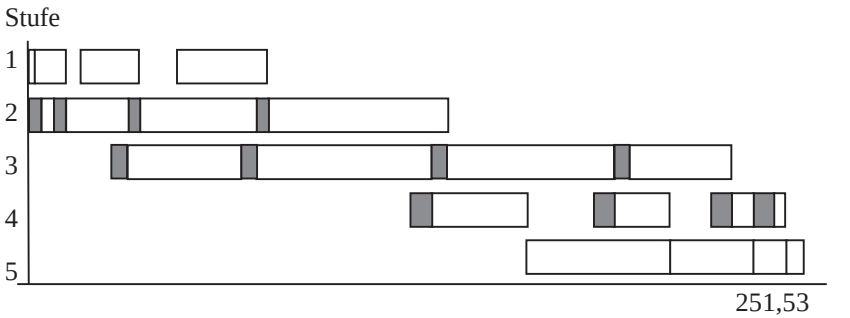


Abbildung 37: Antizipatives Rüsten und variable Lose

Stellt man dieser Lösung die beste Lösung gegenüber, die mit konsistenten Losen bei $a_m = (0; 4; 5; 7; 0)$ erreichbar ist ($x_1 = 4{,}76$; $x_2 = 7{,}4$; $x_3 = 11{,}35$ und $x_4 = 6{,}47$), ergibt sich eine Durchlaufzeit von 255,97 ZE, so dass die Verbesserung lediglich bei 1,7% liegt. Wie diverse Berechungen mit anderen Parametern zeigen, wiederholt sich dieses Muster, auch wenn die Rüstzeiten deutlich aber gleichmäßig erhöht werden. Weicht allerdings eine der Rüstzeiten gravierend von den übrigen ab, kann die Verwendung variabler Lose zu einer erheblichen Reduktion der Durchlaufzeit führen.

Wird z.B. $a_m = (0; 0; 150; 0; 0)$ gesetzt, liegt die beste mit konsistenten Losen erreichbare Durchlaufzeit bei 481,9 ZE für $x_1 = x_2 = 0{,}00$ und $x_3 = 1{,}88$, sowie $x_4 = 28{,}12$. Werden variable Lose verwendet, ergibt sich eine Durchlaufzeit von 427,39 ZE, womit die Verbesserung bei 11,3% liegt. Die dafür ermittelten Losgrößen sind:

$x_{11} = x_{12} = 0{,}00$ $x_{21} = x_{22} = 0{,}00$ $x_{31} = x_{32} = 7{,}50$ $x_{41} = x_{42} = 22{,}50$

$x_{13} = 0{,}00$ $x_{23} = 0{,}00$ $x_{33} = 0{,}00$ $x_{43} = 30{,}00$

$x_{14} = x_{15} = 3{,}69$ $x_{24} = x_{25} = 5{,}54$ $x_{34} = x_{35} = 8{,}3$ $x_{44} = x_{45} = 12{,}46$

Darüber hinaus zeigt sich, dass bei der Berücksichtigung von Rüstzeiten die Randkonsistenz nicht mehr gelten muss. Wird $a_m = (100; 0; 0; 0; 0)$ im Beispiel C gesetzt und die beste Lösung mit variablen Losen der besten Lösung mit konsistenten Losen gegenübergestellt, erhält man die in Abbildung 38 gezeigte Situation. Die optimalen variablen Losgrößen erfüllen nicht die Randkonsistenz:

$x_{11} = 0{,}00$ $x_{21} = 0{,}00$ $x_{31} = 19{,}23$ $x_{41} = 10{,}76$

$x_{12} = x_{13} = 7{,}69$ $x_{22} = x_{23} = 11{,}54$ $x_{32} = x_{33} = 6{,}16$ $x_{42} = x_{43} = 4{,}60$

$x_{14} = x_{15} = 7{,}69$ $x_{24} = x_{25} = 10{,}34$ $x_{34} = x_{35} = 7{,}36$ $x_{44} = x_{45} = 4{,}60$

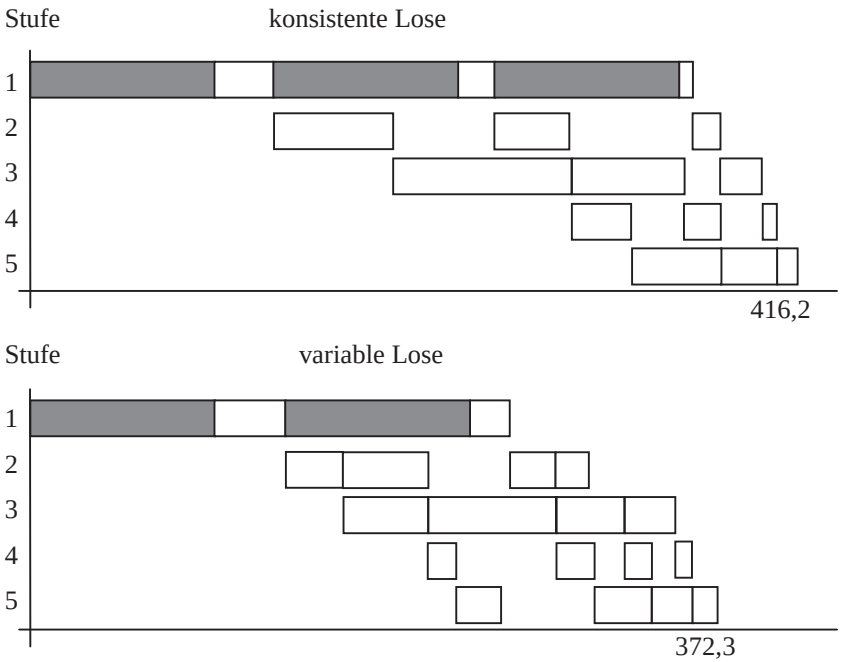


Abbildung 38: Optimale variable Lose ohne Randkonsistenz

Im Fall der konsistenten Lose (oben in Abbildung 38) ergeben sich als optimale Losgrößen $x_1 = 0{,}00$, $x_2 = 16{,}1$, $x_3 = 10{,}16$, $x_4 = 3{,}74$, d.h. es erweist sich als optimal, auf eines der Lose ganz zu verzichten. Dazu muss aber das Modell P(1) um die Restriktionen (74) bis (79) analog erweitert werden. Im Fall der variablen Lose werden sogar nur zwei der möglichen vier Lose auf der ersten Stufe eingesetzt. Auf der zweiten Stufe teilen sich diese zwei Lose in vier Lose auf, so dass die Lösung nicht randkonsistent ist. Der Wechsel von variablen zu konsistenten Losen verlängert hier die Durchlaufzeit um 43,9 ZE (11,8 %).

Folgerung 26:

> Sind Rüstzeiten zu berücksichtigen, kann die Forderung nach Randkonsistenz ($x_{j1} = x_{j2}$ und $x_{j,M-1} = x_{jM}$) die optimale Lösung ausschließen.

In ähnlicher Form wie das antizipative Rüsten ist das verbundene Rüsten in das Modell aufzunehmen. Es lässt sich dadurch charakterisieren, dass der Rüstprozess pro Los erst dann einsetzen kann, wenn das Los zur Freigabe bereit-

steht. Zu denken wäre z.B. an das Bestücken eines entsprechenden Werkstückträgers oder an Einstellungen der Maschinen, die nur in Abhängigkeit der zu fertigenden Teile möglich sind. Dieser Vorgang soll eine fixe Zeitspanne benötigen und unabhängig davon erfolgen, ob die Stufe zuvor leer stand oder benutzt wurde. Wie in der Erweiterung um antizipatives Rüsten muss mit Hilfe der Binärvariablen ω_{jm} erzwungen werden, dass der Rüstvorgang nur für die Lose eingeplant wird, für die $x_{jm} > 0$ gilt. Die Zeitspanne v_m, die für das verbundene Rüsten benötigt wird, ist wie folgt in das Modell aufzunehmen:

$$C_{1m} \geq p_m x_{1m} + \omega_{1m} v_m \qquad m = 1, ..., M \tag{80}$$

$$C_{jm} - p_m x_{jm} - \omega_{jm} v_m \geq C_{j-1,m} \qquad j = 2, ..., J,\ m = 1, ..., M \tag{81}$$

$$C_{jm} - p_m x_{jm} - \omega_{jm} v_m \geq C_{k,m-1} - (1 - \delta_{jmk})\, R \qquad j = 1, ..., J,\ k = 1, ..., J,\ m = 2, \ldots, M \tag{82}$$

Die Restriktionen (77) bis (79) gelten weiterhin, während Restriktion (71) analog zum antizipativen Rüsten nicht verwendet werden darf. Wird in dem obigen Beispiel das Rüsten nicht als antizipatives Rüsten sondern als verbundenes Rüsten mit (v_m = 0; 4; 5; 7; 0) aufgefasst, ergeben sich für J = 4 die folgenden Losgrößen, die zu einer Durchlaufzeit von 266,27 ZE führen:

$x_{11} = x_{12} = 1{,}37$ $x_{21} = x_{22} = 4{,}73$ $x_{31} = x_{32} = 9{,}41$ $x_{41} = x_{42} = 14{,}49$

$x_{13} = 6{,}10$ $x_{23} = 9{,}41$ $x_{33} = 9{,}01$ $x_{43} = 5{,}48$

$x_{14} = x_{15} = 15{,}51$ $x_{24} = x_{25} = 9{,}01$ $x_{34} = x_{35} = 3{,}59$ $x_{44} = x_{45} = 1{,}88$

Die beste Lösung mit konsistenten Losen benötigt 271,97 ZE. Die Reduktion der Durchlaufzeit bei Verwendung variabler Lose fällt mit 5,7 ZE erneut niedrig aus. Sind allerdings einzelne Rüstzeiten im Vergleich zu den übrigen sehr lang, können sich wieder größere Einsparungen ergeben. Zum Beispiel führt v_m = (0; 10; 240; 0; 15) und J = 3 zu den variablen Losen mit den folgenden Losgrößen:

$x_{11} = x_{12} = 1{,}42$ $x_{21} = x_{22} = 7{,}86$ $x_{31} = x_{32} = 20{,}71$

$x_{13} = x_{23} = 0{,}00$ $x_{33} = 30{,}00$

$x_{14} = x_{15} = 0{,}79$ $x_{24} = x_{25} = 8{,}68$ $x_{34} = x_{35} = 20{,}53$

Diese Losgrößen führen zu einer Durchlaufzeit von 709,4 ZE, wohingegen die beste konsistente Lösung (x_1 = 0,00; x_2 = 0,00; x_3 = 30,00) eine Durchlaufzeit von 775 ZE ergibt. Die Durchlaufzeit könnte durch überlappende konsistente Lose – bezogen auf die ersten zwei und die letzten zwei Stufen – reduziert werden. Das mehrfache Rüsten auf der dritten Stufe hebt diese Reduktion aber auf. Der Wechsel von variablen Losen zu konsistenten Losen führt hier zu einer Erhöhung der Durchlaufzeit um 9,2%. Es zeigt sich erneut, dass es in den Situationen mit hohen beziehungsweise sehr stark zwischen den Stufen variierenden Rüstzeiten sinnvoll sein kann, variable Lose zu verwenden. Variable Lo-

se bieten in diesem Extremfall die nötige Flexibilität, die mit konsistenten Losen nicht vereinbar ist. In Abbildung 39 werden die beiden Situationen gegenübergestellt:

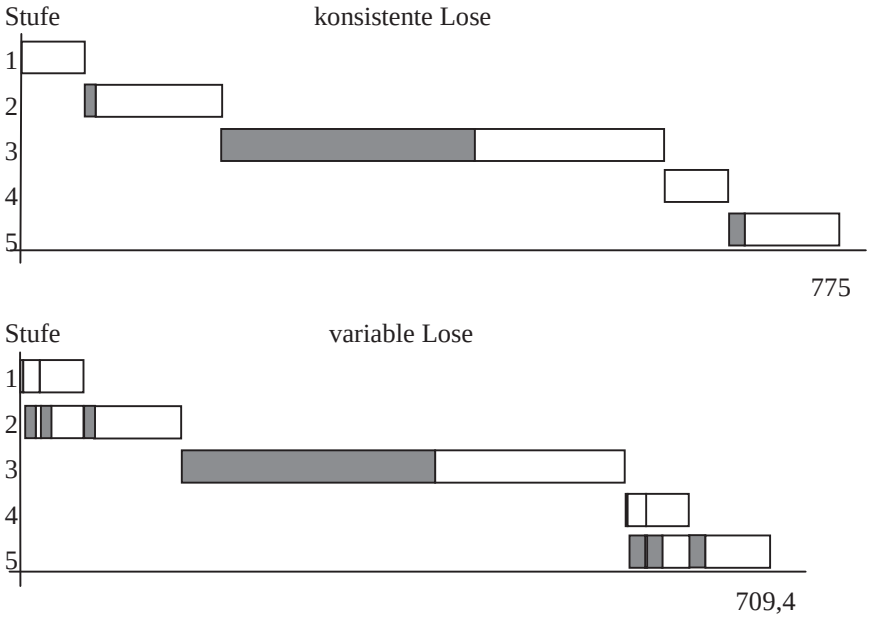


Abbildung 39: Variable versus konsistente Lose bei langen, verbundenen Rüstzeiten

Diese Lösung verdeutlicht eine für die praktische Anwendung sinnvolle heuristische Alternative. Anstatt die gesamte Problemstellung mit variablen Losen zu berechnen, kann das Problem separiert werden. Auf der Stufe m^* auf der die größte Rüstzeit auftritt, wird nur ein Los aufgelegt. Für die Stufen vor und nach m^* werden die optimalen konsistenten Lose bestimmt.

Während dieses Vorgehen für die oben gezeigte Instanz die optimale Lösung liefert, führt der Ansatz des Separierens dann zu u.U. deutlich verlängerten Durchlaufzeiten, wenn das Maximum der Rüstzeiten nicht den gesamten Ablaufplan dominiert. Wird z.B. v_m = (100; 20; 120; 0; 15) gesetzt, ist folgende Lösung mit C_{min} = 700 ZE optimal:

$x_{11} = x_{12} = x_{13} = 0 \quad x_{21} = x_{22} = x_{23} = 10 \quad x_{31} = x_{32} = x_{33} = 20$

$x_{14} = 10 \quad x_{24} = 5 \quad x_{34} = 15$

$x_{15} = 10 \quad x_{25} = 5 \quad x_{35} = 15$

Es fällt auf, dass auf der Stufe mit der längsten Rüstzeit zwei Lose aufzulegen sind. Die oben skizzierte Heuristik führt zu einer Lösung mit 728 ZE für (x_1 = 30,00; x_2 = 30,00; x_3 = 30,00; $x_{14} = x_{15}$ = 7,89; $x_{24} = x_{25}$ = 10,26;

$x_{34} = x_{35} = 11{,}85$) und bedeutet somit eine Verschlechterung von 4% gegenüber der besten Lösung mit drei variablen Losen.

Die Modellierung des verbundenen Rüstens kann in leicht abgewandelter Form auch zur Abbildung der Ofen-Typ Fertigung genutzt werden. Bezeichnet o die Stufe, auf der ein Ofen mit der Kapazität K_o (gemessen in Stück) und der Bearbeitungsdauer p_o auftritt, sind die folgenden Restriktionen zu verwenden:

$$C_{1o} \geq p_o \omega_{1o} \tag{81}$$

$$C_{jo} - p_o\, \omega_{jo} \geq C_{j\text{-}1,o} \qquad j = 2, ..., J \tag{82}$$

$$C_{jo} - p_o\, \omega_{jo} \geq C_{k,o\,\text{-}1} - (1 - \delta_{jok})\, R \qquad j = 1, ..., J,\ k = 1, ..., J, \tag{83}$$

$$x_{jo} \leq K_o \qquad j = 1, ..., J \tag{84}$$

Sofern der Ofen mit einem Los bestückt wird, nimmt die Binärvariable ω_{jo} auf Grund der Restriktion (77) den Wert 1 an. Unabhängig von der Größe des Loses wird in den Restriktionen (81) – (83) die Bearbeitungsdauer p_o berücksichtigt. Bei der Formulierung in (83) wird implizit vorausgesetzt, dass der Ofen nicht auf der ersten Stufe auftritt. Tritt er auf der ersten Stufe auf, entfällt Restriktion (83), da dann die Modellierung der Verfügbarkeit von Losen der Vorstufe nicht erforderlich ist. Die Kapazität des Ofens beschränkt die Freigabelosgröße auf der Ofenstufe auf maximal K_o Einheiten (Restriktion 84). Die übrigen Restriktionen (60) – (66) werden für die Stufen $m \neq o$ unverändert übernommen. Tritt im Beispiel C ein Ofen auf der dritten Stufe mit $K_o = 10$ und $p_o = 100$ ZE auf, ergeben sich für $J = 4$ die folgenden Losgrößen (vgl. Abbildung 40):

$x_{11} = x_{12} = 1{,}43$ $x_{21} = x_{22} = 2{,}86$ $x_{31} = x_{32} = 5{,}71$ $x_{41} = x_{42} = 20{,}0$

$x_{13} = 10$ $x_{23} = 10$ $x_{33} = 10$ $x_{43} = 0{,}0$

$x_{14} = x_{15} = 20{,}0$ $x_{24} = x_{25} = 2{,}1$ $x_{34} = x_{35} = 3{,}15$ $x_{44} = x_{45} = 4{,}73$

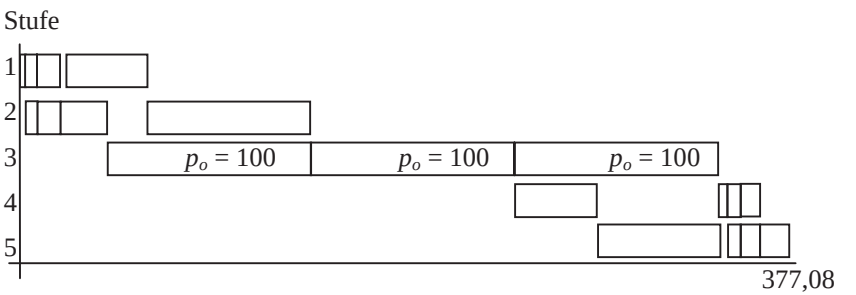


Abbildung 40: Variable Lose und ein Ofen auf der dritten Stufe

Die beste Lösung mit konsistenten Losen führt zu $x_1 = x_2 = x_3 = 10$ und verursacht eine Durchlaufzeit von 410 ZE, so dass der Wechsel von den konsis-

tenten zu den variablen Losen (mit einer Durchlaufzeit von 377,08 ZE) eine Verkürzung um 8 % ermöglicht.

Sollen Transportzeiten berücksichtigt werden, ist zu beachten, dass diese nur die Weitergabe zwischen den Stufen betreffen und daher die Kapazität auf den Stufen unberührt lassen. Es wird zunächst davon ausgegangen, dass sich die Transporte zwischen zwei Stufen überlappen dürfen. D.h. es soll zulässig sein, dass ein weiterer Transport die Stufe m verlässt, während der zuvor gestartete Transport noch nicht die Stufe m+1 erreicht hat. Das Modell ist wie folgt zu modifizieren:

$$C_{jm} - p_m\, x_{jm} \geq C_{k,m-1} + t_{m-1} - (1 - \delta_{jmk}) \qquad j, k = 1, ..., J,\ m = 2, \ldots, M \tag{85}$$

wobei t_{m-1} die Zeitspanne angibt, die zwischen der Stufe m–1 und m für m = 2,.., M als Transportzeit anfällt. Die Restriktion (85) ersetzt die Restriktion (62). Eine Modellierung mit den Binärvariablen ω_{jm} ist hier nicht erforderlich, da eine Berücksichtigung der Transportzeiten – selbst wenn diese extrem hoch ausfallen – keinen Einfluss auf die optimale Größe der Lose nehmen kann. Im Beispiel C wird für t_m = (25; 23; 15; 15) die optimale Lösung mit J = 4 variablen Losen gesucht. Sie lautet:

$x_{11} = x_{12} = 2{,}56$ $x_{21} = x_{22} = 5{,}11$ $x_{31} = x_{32} = 11{,}08$ $x_{41} = x_{42} = 11{,}25$

$x_{13} = 7{,}67$ $x_{23} = 11{,}08$ $x_{33} = 6{,}92$ $x_{43} = 4{,}33$

$x_{14} = x_{15} = 7{,}67$ $x_{24} = x_{25} = 11{,}08$ $x_{34} = x_{35} = 6{,}92$ $x_{44} = x_{45} = 4{,}33$

Mit diesen Losgrößen ist eine Durchlaufzeit von 315,43 ZE zu erreichen. Die beste Lösung mit konsistenten Losen führt zu einer Durchlaufzeit von 321,51 ZE. Um die Transportzeiten in Abbildung 40 darzustellen, sind dunkelgrau unterlegte Flächen verwendet worden. Damit die zwischen den Stufen entstehenden Rechtecke von den übrigen unterscheidbar sind, wurden die Bearbeitungszeiten hellgrau eingefärbt.

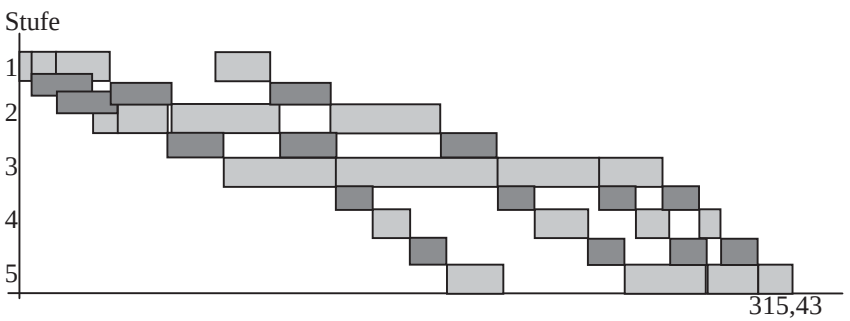


Abbildung 41: Transportzeiten und variable Lose

Bei dieser Lösung fällt die Überlappung der ersten drei Transporte zwischen der ersten und der zweiten Stufe auf. Weiterhin verdeutlicht die Abbildung,

dass sämtliche Bearbeitungen lediglich um die zwischen zwei Stufen als konstant angenommene Zeitspanne t_m verschoben weitergegeben werden. Ein Plan unter Berücksichtigung konstanter Transportzeiten kann daher unmittelbar mit dem ursprünglichen Modell P(2) ermittelt werden.

Wird allerdings t_m nicht als die Transportzeit zwischen den Stufen sondern als eine Zykluszeit des Transporters aufgefasst und angenommen, dass nur jeweils ein Transporter zwischen zwei Stufen verfügbar ist, führen die in der oben gezeigten Lösung auftretenden Überlappungen der Transporte zu einem unzulässigen Plan. Die Zykluszeit des Transporters bezeichnet dabei die Zeitspanne, die der Transporter für die Beladung, die Hinfahrt, die Entladung und die Rückfahrt benötigt. Sie gibt die Zeitspanne an, nach der frühestens der nächste Transport mit der Beladung des Transportmittels beginnen kann. Beseitigt man in der oben gezeigten Lösung die Unzulässigkeit durch eine spätere Weitergabe, steigt die Durchlaufzeit um 17,55 ZE auf 332,98 ZE an. Wesentlich vorteilhafter ist es, das Modell mit der zusätzlichen Restriktion (86)

$$C_{jm} \geq C_{j\text{-}1,m} + t_{jm} \qquad j = 2, ..., J,\ m = 1, ..., M \tag{86}$$

zu lösen. Diese Restriktion stellt sicher, dass auf keiner Stufe ein Los früher als fertiggestellt betrachtet wird, als dass das Transportmittel wieder bereitsteht. Für das Modell für $J = 4$ und $t_m = (25; 23; 15; 15; 0)$, ergibt sich die folgende optimale Lösung (vgl. Abbildung 42) mit einer Durchlaufzeit von 318,77 ZE.

$x_{11} = x_{12} = 5{,}53$ $x_{21} = x_{22} = 7{,}58$ $x_{31} = x_{32} = 11{,}37$ $x_{41} = x_{42} = 5{,}51$

$x_{13} = 5{,}53$ $x_{23} = 7{,}58$ $x_{33} = 11{,}37$ $x_{43} = 5{,}51$

$x_{14} = x_{15} = 13{,}11$ $x_{24} = x_{25} = 5{,}00$ $x_{34} = x_{35} = 6{,}37$ $x_{44} = x_{45} = 5{,}51$

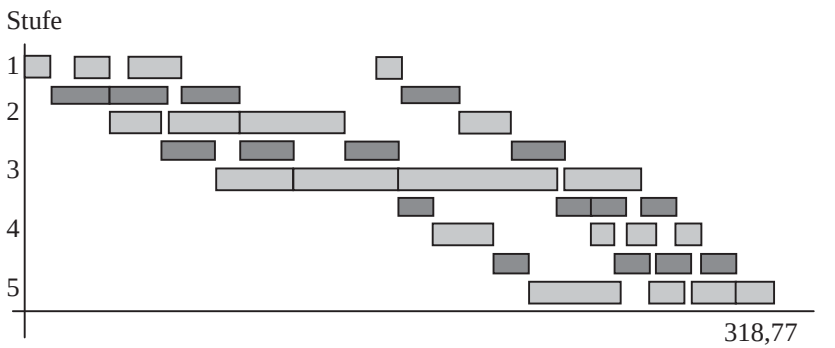


Abbildung 42: Transportzeiten als Zykluszeiten und variable Lose

Auf dem kritischen Pfad in Abbildung 42 treten zweimal zwei unmittelbar aufeinanderfolgende Transporte auf. Diese sind der erste und der zweite Transport zwischen den Stufen 1 und 2 sowie der zweite und der dritte Transport zwischen den Stufen 3 und 4. Es ist damit deutlich zu erkennen, dass die Dauer der

Zykluszeit des Transporters Einfluss auf die optimalen Losgrößen nimmt. Allerdings fällt die beste konsistente Lösung mit der Durchlaufzeit von 323,14 ZE wieder nur marginal schlechter aus.

Insofern wiederholt sich hier das bereits beobachtete Muster: Bei nahezu gleichmäßig großen Rüstzeiten der Stufen beziehungsweise Zykluszeiten der Transporter bietet ein Wechsel von konsistenten auf variable Lose nur eine relativ geringe Verbesserung, dafür werden aber regelmäßig weniger Transportvorgänge verursacht. Die variablen Lose können dann erhebliche Vorteile bieten, wenn die Höhe der Rüstzeiten (beziehungsweise die Länge der Zykluszeiten der Transporter) stark voneinander abweichen und so die Flexibilität der variablen Lose gefordert ist.

Um ganzzahlige Losgrößen ermitteln zu können, ist folgende Restriktion in den LINGO-Kode aufzunehmen:

```
@For(Los(j): @For(Stufe(m): @Gin (X(j,m))));
```

Dabei stellt sich die Frage, ob Runden der kontinuierlichen Werte eine sinnvolle heuristische Vorgehensweise sein kann. In der folgenden Tabelle sind für das Beispiel C bei $J = 5$ Losen die kontinuierlichen Losgrößen angegeben:

Tabelle 18: **Optimale kontinuierliche Losgrößen im Beispiel C (C_{JM} = 216,42)**

Stufe	x_1	x_2	x_3	x_4	x_5
1	1,43	2,87	6,22	9,33	10,12
2	1,43	2,87	6,22	9,33	10,12
3	4,3	6,22	9,33	6,86	3,26
4	10,53	9,33	2,74	4,11	3,26
5	10,53	9,33	2,74	4,11	3,26

In Tabelle 19 werden die optimalen diskreten Lose diesen gegenübergestellt. Es fällt auf, dass diese Lösung durch Runden nicht gefunden werden kann, da z.B. das dritte Los auf der vierten Stufe im kontinuierlichen Fall 2,74 und im diskreten Fall 6 Einheiten groß ist.

Tabelle 19: **Optimale diskrete Losgrößen im Beispiel C (C_{JM} = 219)**

Stufe	x_1	x_2	x_3	x_4	x_5
1	1	3	6	9	11
2	1	3	6	9	11
3	4	6	9	6	5
4	10	9	6	2	3
5	10	9	6	2	3

Werden die Werte aus Tabelle 18 dennoch gerundet, entsteht zunächst eine unzulässige Lösung, da auf mehreren Stufen $Q \neq 30$ Einheiten gefertigt werden. Eine Schwierigkeit der heuristischen Vorgehensweise liegt nun darin, dass nicht erkennbar ist, wie die abgerundeten Mengen umzuverteilen sind, beziehungsweise in welchem Los die über Q hinausgehenden Mengeneinheiten abzubauen sind. Generell ist auch nicht zu erwarten, dass hier eine allgemeingültige Regel existiert, die für beliebige Instanzen eine vorteilhafte Umverteilung sicherstellt. Ein sinnvoller Ausweg besteht darin, verschiedene Pläne mit den naheliegenden Varianten der umverteilten Mengen zu generieren und die beste auszuwählen. Diese sind in Tabelle 20 angegeben. Durch die gerundeten Werte steigt die Durchlaufzeit geringfügig von 219 ZE auf 222 ZE an. Sie liegt aber noch deutlich unter der optimalen ganzzahligen Lösung mit konsistenten Losen, die 227 ZE beträgt. Dem Produktionsplaner stehen zur Beurteilung der durch Runden gefundenen Lösungen einerseits die für variable, kontinuierliche Lose und andererseits die für ganzzahlige konsistente Lose ermittelten Pläne zur Verfügung. Beide sind mit vergleichsweise geringem Aufwand generierbar und zeigen ihm die Spanne auf, in der die optimale ganzzahlige Lösung mit variablen Losen liegen muss. Die durch Umverteilen entstehenden Pläne können somit beurteilt werden.

Tabelle 20: **Gerundete Losgrößen**

Stufe	x_1	x_2	x_3	x_4	x_5
1	2	3	6	9	10
2	2	3	6	9	10
3	4	7	9	7	3
4	11	9	3	4	3
5	11	9	3	4	3

Zieht man für diese Beurteilung ergänzend die in der Praxis häufig verwendeten konstanten Lose (Kalir, 1999) hinzu, zeigt sich, dass ein Plan mit fünf Losen der Größe 6 zu einer Durchlaufzeit von 246 ZE führt. Insofern ist die heuristisch ermittelte Lösung mit 222 ZE die deutlich bessere Wahl. Für große Instanzen ist zudem die optimale Lösung mit variablen diskreten Losen regelmäßig nicht mehr in akzeptabler Zeit rechenbar, so dass auf Heuristiken zurückzugreifen ist. Im Gegensatz zu den kontinuierlichen Losen kann sich bei diskreten Losen ein Plan hinsichtlich seiner Struktur verändern, wenn Q variiert. Um dies zu zeigen, werden im Beispiel C die optimalen Losgrößen für $Q = 30$ und $Q = 50$ ermittelt und gegenübergestellt:

Tabelle 21: **Optimale ganzzahlige Losgrößen für Q = {30, 50}**

Stufe	x_1	x_2	x_3	x_4
1	24	57	10 16	13 23
2	24	57	10 16	13 23
3	711	10 16	814	59
4	67	11 20	814	59
5	67	11 20	814	59

In jeder Zelle der Tabelle 21 sind die optimalen Losgrößen für Q = 30 und Q = 50 abgetragen. Nimmt man die jeweilige Differenz der Losgrößen, variieren die Lose um 2, 4, 6, 9 oder 10 Einheiten. Die Aufteilung in diskrete Lose ist somit nicht direkt aus der Lösung für Q = 30 abzulesen, sondern bei jeder Änderung von Q neu zu bestimmen.

Schließlich besteht noch die Möglichkeit, andere Zielfunktionen in das Modell aufzunehmen. Wenn man beispielsweise annimmt, dass für bestimmte Zeitpunkte Bestellungen vorliegen, dann sind bezogen auf die letzte Stufe die dort verwendeten Freigabelose fixiert. Für jedes Freigabelos der letzten Stufe kann ein Liefertermin vorgegeben sein. Wird die bestellte Menge z.B. später als vereinbart fertiggestellt, fallen mit der Terminabweichung gewichtet Strafkosten an. Eine Zielfunktion, die die Summe der gewichteten Terminabweichung minimiert lautet:

$$Min \sum_{j=1}^{J} \left| C_{jM} - d_j \right| \alpha_j \tag{87}$$

Die Kosten α_j können z.B. nach der Wichtigkeit des Kunden individuell gesetzt werden. Weiterhin kann nach Kundenklassen differenziert vorgegangen werden oder mit $\alpha_j = 1$ für $j = 1, ..., J$ die Gewichtung ganz entfallen und die Summe der Terminabweichung minimiert werden. Zur Umsetzung in das lineare Programm muss die Summe der Verfrühungen und die Summe der Verspätungen separat erfasst werden, damit sich nicht die negativen und positiven Abweichungen gegenseitig aufheben. Es bezeichnet E_j die Verfrühung des j-ten Loses und T_j die Verspätung. Die Verfrühungsstrafe je ZE wird mit α_j und die Verspätungsstrafe je ZE mit β_j vorgegeben. Die folgende Zielfunktion ist mit den zwei zusätzlichen Restriktionen in das Modell aufzunehmen:

$$Min \sum_{j=1}^{J} E_j \, \alpha_j + \sum_{j=1}^{J} T_j \, \beta_j \tag{88}$$

$$T_j \geq C_{jM} - d_j \qquad j = 1,..., J \tag{89}$$

$$E_j \geq d_j - C_{jM} \; j = 1,..., J \tag{90}$$

$$T_j , E_j \geq 0 \qquad j = 1,..., J \tag{91}$$

Sind für $\alpha_j = \beta_j = 1$ im Beispiel C die folgenden zwei Liefertermine gesetzt: $C_{15} = 10$ mit $d_1 = 230$ und $C_{25} = 20$ mit $d_2 = 330$, dann tritt mit zwei konsistenten Losen eine Terminabweichung von 30 ZE auf. Werden statt dessen zwei variable Lose zugelassen, sinkt die Terminabweichung um 40% auf 18 ZE. Für $d_1 = 200$ und $d_2 = 350$ können mit zwei variablen Losen sogar beide Liefertermine genau eingehalten werden, wohingegen zwei konsistente Lose eine Abweichung von 10 ZE verursachen.

Die Flexibilität variabler Lose kann damit sinnvoll eingesetzt werden, wenn Terminabweichungen – oder Verspätungen – minimiert werden sollen. Diese Varianten sind leicht in das Modell aufzunehmen.

5.4.4 Zusätzliche Lose und ihr Effekt

In diesem Abschnitt wird die für den praktischen Einsatz der Modelle wichtige Frage untersucht, wie sich das Ausmaß der Verbesserung durch variable Lose entwickelt, wenn die Zahl der Lose variiert. Die folgende Tabelle verdeutlicht den Anstieg der Rechenzeiten, wenn die Zahl der Lose zwischen 3 und 9 variiert und dabei der für konsistente Lose gefundene Zielfunktionswert als Schranke vorgegeben wird. Ergebnisse für den Fall mit $J = 2$ (vgl. z.B. die Instanz von Potts/Baker, 1989) sind nicht tabelliert, da die Rechenzeit durchgängig so klein ist, dass sie als 0,00 Sek. ausgegeben wird. Zusätzlich werden exemplarisch Laufzeiten für die Stufenzahl 6 und 8 betrachtet, die verdeutlichen, wie der Rechenaufwand mit der Anzahl der Lose J überproportional und mit der Zahl der Stufen M stark ansteigt. Im Modell P(3) beträgt die Zahl der Binärvariablen $(M–3)(J^2–J) – (M–3)(J–1) = (M – 3)(J – 1)^2$, so dass in der Instanz von Potts/Baker (1989) die Zahl der Binärvariablen 1 beträgt. Für ein Problem mit $M = 5$ Stufen und $J = 5$ Losen werden 32 Binärvariable benötigt. Diese Problemstellung kann mittels LINGO 7.0 in 15 Sekunden auf einem Standard-PC (Pentium 4, 1,8 GHz, 512 MB RAM, Windows 2000) gelöst werden. Für kleinere Probleme wird die optimale Lösung innerhalb einer Sekunde gefunden.

Tabelle 22: **Probleminstanzen und Laufzeiten**

J / M	Laufzeit	J / M	Laufzeit	J / M	Laufzeit
3 / 5	0 Sek.	3 / 6	0 Sek.	3 / 8	10 Sek.
4 / 5	1 Sek.	4 / 6	7 Sek.	4 / 8	6 Minuten
5 / 5	15 Sek.	5 / 6	30 Sek.	5 / 8	> 60 Std.
6 / 5	4 Min.	6 / 6	8 Minuten		
7 / 5	26 Min.	7 / 6	24 Stunden		
8 / 5	200 Min.	8 / 6	> 60 Std.		
9 / 5	58 Std.				

Für die Instanzen mit $M = 6$ Stufen wurden die zufällig gesetzten Werte $\underline{p}$ = (7; 9; 3; 1; 2; 5) und für die Instanzen mit $M = 8$ Stufen $\underline{p}$ = (28; 42; 22; 32; 11; 41; 14; 38) bei jeweils $Q = 30$ verwendet. Die Berechungen für den Fall mit fünf Losen auf acht Stufen wurden nach über 60 Stunden abgebrochen.

Auch wenn den Angaben in der Tabelle keine repräsentative Stichprobe zugrunde liegt, verdeutlichen sie, dass schon bei einer moderaten Erhöhung der Stufen- oder der Loszahl eine optimale Lösung nicht mehr in akzeptabler Zeit gefunden werden kann. Diese Beobachtung wird durch die komplexitätstheoretische Einordnung der Problemstellung gestützt, die Trietsch/Baker (1993, S. 1076) vornehmen. Sie äußern die Vermutung, dass das Losüberlappungsproblem mit variablen Losen und zulässiger Leerzeit mit sehr großer Wahrscheinlichkeit schwer im Sinne der Komplexitätstheorie ist. Ein Beweis dieser Aussage konnte allerdings bislang nicht erbracht werden. Die Einschätzung, dass sich bei variablen Losen eine Fragestellung ergibt, die besonders anspruchsvoll zu lösen ist, wird auch von Williams et al. (1997) herausgestellt:

> „Variable sublots“ ... „is another area of future research. However, resplitting and merging of sublots from machine to machine makes this problem a complex one.“

Als einzige bislang publizierte Studie zum Vergleich der Durchlaufzeiten bei verschiedenen Losarten ist auf die Untersuchung von Baker/Jia (1993) zu verweisen. Dort werden für den Dreistufenfall folgende Verfahren eingesetzt:

- konstante Lose;
- konstante Lose ohne Leerzeit;
- konsistente Lose;
- variable Lose unter Stückverfügbarkeit und ohne Leerzeit.

Baker/Jia (1993) betrachten $J \in \{3, 5, 8\}$ Lose, $M = 3$ und erzeugen insgesamt 6.000 Testprobleme, wobei sie die Ausführungszeiten für die Hälfte der Probleme gleichverteilt aus dem Intervall [1, 100] ziehen und für die andere Hälfte eine Exponentialverteilung ($\lambda = 50$) verwenden. Zudem untersuchen sie für einen Teil der Instanzen, wie sich die Durchlaufzeit reduziert, wenn die Zahl der Lose zwischen 1 und 10 variiert. Zusammengefasst kommen die Autoren zu den folgenden Ergebnissen:

- Im Mittel liegen die Lösungen mit konstanten Losen zwischen 10% und 20% über der jeweils besten Lösung.
- Die größte beobachtete Differenz, die bei der Verwendung konsistenter Lose unter Losverfügbarkeit gegenüber dem Fall der variablen Lose unter Stückverfügbarkeit auftritt, beträgt 9,7%.
- Im Mittel über die 6.000 Instanzen und alle Loszahlen verfehlen die konsistenten Lose die jeweils beste erreichbare Durchlaufzeit um ledig-

lich 1,4%. Dieser Wert wurde aus der Tabelle 2, Seite 564, Baker/Jia (1993) errechnet.

- Für jedes Lösungsverfahren kann die Hälfte der mit zehn Losen erzielbaren Durchlaufzeitreduktion bereits beim Wechsel von einem zu zwei Losen realisiert werden.
- 80% der mit zehn Losen erzielbaren Durchlaufzeitreduktion ist bereits mit drei Losen realisierbar.

Berücksichtigt man die dritte Beobachtung, kann sie als eine Richtgröße verwendet werden, wenn die Entscheidung zu treffen ist, ob sich ein Wechsel von den konsistenten Losen zu den variablen Losen (unter Stückverfügbarkeit) lohnt. Da sich im Mittel für den von Baker/Jia betrachteten Dreistufenfall eine Reduktion der Durchlaufzeit um 1,4% einstellt, kann man die monetären Konsequenzen abschätzen, die entstehen wenn man diesen Wechsel vornimmt. Interpretiert man ergänzend die letzte Beobachtung aus einem andern Blickwinkel, erbringen – bezogen auf die Lösung mit drei Losen – weitere sieben Lose höchstens noch ein Viertel der bereits erreichten Verbesserung. Diese Beobachtung spricht somit für die Verwendung weniger Lose.

Baker/Jia (1993) haben zwar mit 6.000 Instanzen eine mehr als ausreichend große Grundgesamtheit untersucht, dabei aber nur den Fall unter Leerzeitfreiheit auf drei Stufen behandelt. Für den Dreistufenfall gibt es, Losverfügbarkeit vorausgesetzt, mindestens eine optimale Lösung mit konsistenten Losen, so dass Baker/Jia (1993) auf die Variante der variablen Lose unter Losverfügbarkeit nicht eingehen mussten. Die einzige Studie, die sich vorrangig mit einer Abschätzung der Durchlaufzeitreduktion mittels Losüberlappung beschäftigt, stammt von Kalir/Sarin (2000). Sie untersuchen im Mehrproduktfall den Effekt der Überlappung mit gleichgroßen Losen hinsichtlich diverser Zielkriterien und können daher zu der hier angesprochenen Beurteilung der variablen Lose unter Losverfügbarkeit im Mehrstufenfall keinen Hinweis liefern. Daher werden im Folgenden Problemstellungen mit mehr als drei Stufen betrachtet und es wird zwei Fragen nachgegangen (vgl. dazu auch Biskup/Feldmann, 2005):

- *Wie stark reduziert sich die Durchlaufzeit, wenn ein weiteres Los berücksichtigt wird?*
- *Wie entwickelt sich die Differenz zwischen der Lösung mit konsistenten und der Lösung mit variablen Losen bei steigender Loszahl?*

Die Untersuchung beschränkt sich wegen der hohen Laufzeiten darauf, für ausgewählte Instanzen den Verlauf exemplarisch zu verdeutlichen. Betrachtet wird in Tabelle 17 zunächst das Beispiel C, für das die Lösungen mit konstanten, mit konsistenten und mit variablen Losen unter Losverfügbarkeit verglichen werden. Für $J = 1, 2, ..., 9$ Lose konnten optimale Pläne ermittelt werden. Für $J = 10$ Lose wurde die Berechnung nach über 200 Stunden abgebrochen und die

bis dahin ermittelte Lösung verwendet. Mit C_{JM} wird die Durchlaufzeit bezeichnet, mit Δ die marginale Verbesserung der Lösung in ZE, die durch die Einführung des letzten Loses erreicht wird. Δ% gibt diese Verbesserung in Relation zu $C_{J\text{-}1,M}$ an. Im Spaltenkopf wird jeweils die Losart angegeben und über die Zeilen die Zahl der Lose erhöht. *KoKi* bezeichnet die prozentuale Verschlechterung der Lösung mit konstanten Losen im Vergleich zu der Lösung mit konsistenten Losen und wird wie folgt bestimmt: $KoKi = (C_{Ko} - C_{KI}) / C_{KO}$. *KiV* gibt diese Differenz im Vergleich zwischen den Lösungen mit konsistenten und variablen Losen an, *KoV* dient dem Vergleich zwischen den Lösungen mit konstanten und variablen Losen. Als Ausgangpunkt dient die Situation ohne Losüberlappung, für die sich eine Durchlaufzeit von 510 ZE ergibt.

Tabelle 23: **Vergleich der Ergebnisse für drei Losarten**

J	konstante Lose			konsistente Lose			variable Lose		
	C_{JM}	Δ	Δ%	C_{JM}	Δ	Δ%	C_{JM}	Δ	Δ%
2	345	165	31,8	342	168	32,9	342	168	32,9
3	290	55	15,9	273,27	68,72	20,1	269,78	72,22	21,1
4	262,50	27,50	9,5	243,50	29,77	10,9	237,43	32,35	12,0
5	246	16,50	6,3	224,33	19,17	7,9	216,42	21,01	8,8
6	235	11	4,5	212,70	11,63	5,2	205,22	11,2	5,2
7	227,14	7,86	3,3	204,53	8,17	3,8	198,72	6,50	3,2
8	221,25	5,89	2,6	198,68	5,85	2,9	193,48	5,24	2,6
9	216,67	4,58	2,1	194,55	4,13	2,1	190,21	3,27	1,7
10	213,00	3,67	1,7	191,23	3,32	1,7	187,89	2,32	1,2

Zur Interpretation der Ergebnisse in Tabelle 24 werden zunächst die Spalten der konstanten Lose mit denen der konsistenten Lose verglichen. Es wird deutlich, dass der Grenzeffekt des Loses bei konsistenten Losen (gemessen in Δ%) immer deutlich über dem Wert bei konstanten Losen liegt. Dieser „Vorsprung“ nimmt aber mit zunehmender Loszahl ab und erreicht bei 8 Losen mit 5,89 ZE zu 5,85 ZE annähernd Gleichstand. Mit der Einführung des neunten konstanten Los verbessert sich die Lösung um 4,58 ZE, während das neunte konsistente Los eine Verbesserung um 4,13 ZE bewirkt. Da die zuvor eingeführten konsistenten Lose aber erheblich effektiver waren als die konstanten Lose, beträgt die ausgewiesene prozentuale Differenz *KoKi* bei acht und bei neun Losen 10,2%. Auf die Entwicklung der *KoKi* Werte bei steigender Loszahl wird im Anschluss an die Diskussion der Tabelle 24 näher eingegangen. Vergleicht man die ausgewiesenen prozentualen Differenzen *KoKi* und *KoV* in Tabelle 24 untereinander, fällt auf, dass z.B. für eine Verbesserung der Lösung mit konstanten Losen

um ca. 9,5 %, ein Plan mit vier variablen Losen statt einem Plan mit sechs konsistenten Losen eingesetzt werden kann.

Nimmt man nun die Spalten für die variablen Lose hinzu, zeigt sich, dass zunächst der Grenzeffekt des zusätzlichen Loses (Δ) bei variablen Losen geringfügig größer ist als bei konsistenten Losen. In dem vorliegenden Beispiel wird bei sechs Losen ein Gleichstand erreicht, der sich beim Wechsel auf sieben Lose umzukehren beginnt. Wenn acht Lose eingesetzt werden, erbringt das achte konsistente Los eine Reduktion um 5,85 ZE und das achte variable Los eine Verkürzung um 5,24 ZE. Beim neunten Los ist der Grenzeffekt bei konsistenten Losen mit 4,13 ZE schon größer als der des neunten variablen Loses, das eine Verkürzung der Durchlaufzeit um 3,27 ZE bewirkt.

Tabelle 24: **Relativer Vorteil der drei Losarten**

J	*KoKi in %*	*KoV in %*	*KiV in %*
2	0,8	0,8	0,0
3	5,8	7,0	1,3
4	7,2	9,6	2,49
5	8,8	12,0	3,53
6	9,5	12,7	3,52
7	9,9	12,5	2,84
8	10,2	12,5	2,62
9	10,2	12,2	2,23
10	10,2	11,8	1,75

Diese Entwicklung wird auch deutlich, wenn man sich den Verlauf der *KiV* Werte bei wachsendem *J* ansieht. Hier ist ein ökonomisch nachvollziehbarer Zusammenhang erkennbar: Das Potential, das die variablen Lose gegenüber den konsistenten Losen nutzbar machen können, steigt bei wachsender Loszahl zunächst an, um dann wieder zu fallen. Dieser Umkehrpunkt liegt in dem hier betrachteten Beispiel bei $J = 5$ Losen. Der Verlauf kann wie folgt begründet werden: Je mehr Lose verfügbar sind, desto mehr Möglichkeiten zur Zusammenfassung und zur Aufteilung bieten sich an. Mit steigender Loszahl stellen sich aber auch die konsistenten Lose immer besser auf die Struktur des Problems ein. Wenn man es so sehen möchte, verlieren die Unterschiede in den Ausführungszeiten der Stufen bei zunehmender Zahl an Losen ihre Bedeutung. Die Lose können zudem kleiner gewählt werden und dadurch verringert sich das Potential, das bei geringer Zahl an Losen den Erfolg der variablen Lose ausmacht. Die höhere Flexibilität der variablen Lose erlaubt die Effekte vorab auszunutzen,

die sich mit konsistenten Losen erst bei einer größeren Zahl an Losen einstellen. Wird die Zahl der Lose stark erhöht, ist zu erwarten, dass die Durchlaufzeit bei konsistenten Losen gegen den Wert konvergiert, der sich für variable Lose einstellt. Wird *KiV* gegen die Loszahl abgetragen deutet sich ein konvexer Verlauf an, der zunächst ansteigt und dann wieder fällt. Dieser Verlauf konnte für die Instanz mit sechs Stufen bestätigt werden. Die dort erreichten Reduktionen sind in der folgenden Tabelle angegeben:

Tabelle 25: **Gegenüberstellung der Werte für das Beispiel mit sechs Stufen**

J / *M*	konsistente Lose	variable Lose	*KiV* in %
2 / 6	530,77	530,77	0,00
3 / 6	431,42	431,42	0,00
4 / 6	381,78	373,33	2,21
5 / 6	353,34	345,00	2,4
6 / 6	333,53	327,85	1,7
7 / 6	320,67	318,15	0,7

Für die Instanz mit acht Stufen ergaben sich allerdings keine Differenzen. Die optimale Lösung mit variablen Losen stimmen für die betrachteten Loszahlen {2, 3, 4} durchweg mit der optimalen Lösung bei konsistenten Losen überein.

Abschließend ist noch auf den oben angesprochenen Vergleich zwischen den konstanten und den konsistenten Losen einzugehen, da sich dort die Frage ergibt, was bei einer weiteren und starken Erhöhung der Loszahl geschieht. Dazu wurde für das Beispiel C die Loszahl auf die Werte *M* = {20, 30, 40, ..., 100, 200, 300} gesetzt. Während sich für zwanzig konsistente Lose eine Durchlaufzeit von 181,15 ZE ergibt, wird dieser Wert mit konstanten Losen erst jenseits des zweihundersten Loses erreicht. Die folgende Tabelle gibt die Ergebnisse bis *M* = 100 wieder:

Tabelle 26: **Effekt sehr vieler Lose**

	20	30	40	50	60	70	80	90	100
konstant	196,5	191	188,3	186,6	185,5	184,7	184,1	183,7	183,3
konsistent	181,2	180,1	180,02	180,00	180,00	.....			

Die konsistenten Lose sind ab *M* = 60 nur noch in der vierten Nachkommastelle von 180 ZE entfernt. Für *M* = 200 Lose beträgt die Durchlaufzeit bei konstanten Losen 181,7 ZE und fällt bei *M* = 300 auf 181,1 ZE. Es zeigt sich, dass die mit konstanten Losen erreichbare Lösung zwar gegen die Lösung strebt, die mit konsistenten Losen erreichbar ist. Diese Annährung geschieht aber nur sehr langsam. Bei *M* = 20 Losen liegt die Durchlaufzeit mit konstanten Losen noch 15,35 ZE über dem mit konsistenten Losen erreichbaren Wert. Kehrt man zu

dem Beispiel C in Tabelle 17 zurück, kann die aus neun konsistenten Losen resultierende Durchlaufzeit erst mit 25 konstanten Losen unterboten werden; um die Durchlaufzeit der neun variablen Lose zu erreichen, sind bereits 33 konstante Lose notwendig etc. Es kann somit eine sehr hohe Zahl an konstanten Losen erfordern, wenn man den höheren organisatorischen Aufwand für konsistente oder variable Lose durch eine entsprechend größere Zahl an konstanten Losen vermeiden will. Die Zahl der konstanten Lose kann sogar prohibitiv groß werden, wenn man versucht, annähernd die Durchlaufzeit zu erreichen, die mit einer moderaten Zahl an konsistenten Losen realisierbar ist.

Auf Basis der Berechnungen, die in diesen Abschnitt durchgeführt wurden und denen, die zur Konstruktion der Beispiele in dem vorangehenden Kapitel dienten, ist festzuhalten, dass unter Losverfügbarkeit variable Lose nur für geringe Loszahlen zu rechenbaren Modellen führen. Für das Modell P(2) konnte exemplarisch gezeigt und ökonomisch begründet werden, dass sich für eine geringe Zahl an Losen der vergleichsweise größte Vorteil gegenüber den anderen Losarten ergibt. Allerdings bleibt dabei der maximal beobachtete Vorteil der variablen Lose mit 3,53% in einem Bereich, der den Einsatz konsistenter Lose rechtfertigt. Dies gilt insbesondere, wenn man die bessere Rechenbarkeit, die leichtere organisatorische Umsetzung und die Möglichkeiten der postoptimalen Analyse berücksichtigt, die mit konsistenten Losen realisierbar sind. Für die praktische Umsetzung ist dem Produktionsplaner somit zu empfehlen, zunächst auf Pläne mit konsistenten Losen zurückzugreifen.

Muss allerdings das Modell P(2) modifiziert werden, um z.B. hohe und ungleichmäßig verteilte Rüstzeiten, zyklende Transporter oder andere Zielfunktionen zu berücksichtigen, dann kann die höhere Rechenzeit und der zusätzliche organisatorische Aufwand für die Umsetzung eines Planes mit variablen Losen sinnvoll investiert sein. Wie im Abschnitt 5.4.2 gezeigt, sind auf Grund der höheren Flexibilität variabler Lose erhebliche Reduktionen der Durchlaufzeit möglich. Zudem können variable Lose entscheidend dazu beitragen, dass Liefertermine eingehalten werden. Für viele praktisch relevante Fällen sind variable Lose unter Losverfügbarkeit somit eine sinnvolle Alternative.

5.5 Weitere Ansätze zur Losüberlappung

Die vorangegangenen Kapitel beschäftigen sich überwiegend mit den Verfahren, die unmittelbar für eine praktische Umsetzung relevant sind. Darüber hinaus existieren einige Untersuchungen, die sich mit speziellen Fragen der Losüberlappung befassen. Auf diese wurde bisher nicht eingegangen, wenn sie hinsichtlich der Kriterien: Relevanz der Fragestellung im Kontext der MRPII-

Systeme, Entwicklungsstand und verursachter Rechenaufwand zu weit von einer praktischen Umsetzung entfernt sind. Um aber einen Überblick zu bieten und die hier erarbeiteten Ergebnisse besser einordnen zu können, wird die übrige Diskussion zur Losüberlappung im Folgenden nachgezeichnet.

Das erste Mal wird der Begriff Lot Streaming von Reiter (1966) verwendet. Er skizziert einen Vorläufer eines PPS-Systems und schlägt in dem von ihm vorgestellten Prototyp vor, Lose zwischen den Maschinen überlappend weiterzugeben. Dazu empfiehlt er, jeder Maschine eine Belastungsschranke, gemessen in ZE vorzugeben. Wird diese durch die kumulierte Bearbeitungszeit des nächsten einzulastenden Loses überschritten, ist dieses Los in so vielen gleichgroßen Teillosen weiterzugeben, dass für jedes einzelne die Belastungsschranke gerade nicht mehr überschritten wird. Reiter setzt in seiner Arbeit einen Job Shop voraus und verfolgt mit der Beachtung von Durchlaufzeiten und Terminabweichungen zeitliche Kriterien. Umsetzbare Algorithmen fehlen aber ebenso wie eine nachvollziehbare Parametrisierung. Das von Reiter (1966) präsentierte Konzept zur Entwicklung eines PPS-Systems wird allerdings noch heute regelmäßig zitiert, da er den Begriff Lot Streaming einführte und das Potential dieser Technik erkannte (vgl. Reiter, 1966, S. 376):

> „I believe that the systematic use of line-mode scheduling and lot streaming results in a greater contribution to payoff than could be expected from improved approximation to an ‚optimal' schedule that is restricted to batch processing."

Bemerkenswert ist weiterhin, dass Reiter bereits 1966 einen Vorteil des Lot Streaming in der besseren Einhaltung von Lieferterminen benennt (Reiter, 1966, S. 376), der erst Jahre später von Wagner/Ragatz (1994) oder Yoon/Ventura (2002) erneut aufgegriffen wird.

Die Diskussion der Losüberlappung ist somit eindeutig in der Ablaufplanung begonnnen worden. Wenige Jahre später wird die Idee der überlappenden Weitergabe aber auch im Rahmen der Losgrößenplanung diskutiert, so dass aus heutiger Sicht zwei Entwicklungsstränge existieren:

Einige Arbeiten befassen sich mit der überlappenden Weitergabe als einer Variante der Losgrößenplanung. Das übergeordnete Ziel dieser Ansätze besteht darin, die Losgrößen unter Beachtung der Lager- und Rüstkosten bei meist unbegrenztem Zeithorizont so festzulegen, dass eine vorgegebene Nachfrage kostenminimal befriedigt werden kann.

Der weit größere Anteil an Arbeiten wurde im Kontext des Scheduling verfasst. Dort wird angenommen, dass die aufzuteilende Losgröße von einer vorgelagerten Entscheidungsebene bestimmt wird. Optimiert wird hinsichtlich der Losart und der Größe der Freigabe- und Transferlose, wobei Zeitkriterien betrachtet werden.

5.5.1 Losüberlappung aus Sicht der Losgrößenplanung

In diesem Abschnitt wird auf die Planung der Losgrößen bei überlappender, gemeinsamer Weitergabe von Teilen in Transferlosen eingegangen. Diese Modelle behandeln die Zwischenstufe zwischen der offenen und der geschlossenen Weitergabe von Teilen. Ihre Darstellung ist bewusst knapp gehalten, da das Ziel der Arbeit darin besteht, für die bestehenden PPS-Systeme aufzuzeigen, wie die Ablaufplanung durch Losüberlappung sinnvoll unterstützt werden kann. Die Modelle der Losgrößenplanung erfüllen dieses Kriterium nicht, da sie die Entscheidung über das Ausmaß oder die Art der Überlappung vorgeben (in der Regel werden nur konstante Transferlose zugelassen) und fast durchgängig muss für jede Stufe angenommen werden, dass die Produktion ohne Unterbrechung zu erfolgen hat (no-idling).

Die erste Behandlung dieser Fragestellung wird in der Literatur Szendrovits (1975) zugeschrieben, der den Mehrstufenfall betrachtet und dabei unterstellt, dass auf allen Stufen die gleiche Produktionslosgröße festzulegen ist, für die genau ein Rüstvorgang je Stufe anfällt. Für die Weitergabe aus dem laufenden Los können gleichgroße Transferlose zwischen den Stufen verwendet werden. Er bezeichnet diese gemeinsame Weitergabe als „combined movement“. Transportkosten bleiben in seinem Modell unberücksichtigt, da er annimmt, dass ihre Höhe (z.B. der Lohn des Personals) bereits determiniert ist. Für einen unbegrenzten Zeithorizont ergibt sich ein formaler Zusammenhang zwischen der Produktionslosgröße, der Durchlaufzeit und dem Lagerbestand. Goyal (1976) greift diesen Zusammenhang auf und zeigt, dass sowohl die Losgrößen als auch die Zahl der Transporte simultan mit einem Suchverfahren optimiert werden können, wenn man fixe Kosten pro Transport ansetzt. Sein Modell wurde durch Szendrovits (1976) weiter vereinfacht, der eine Formel entwickelt, die das Suchverfahren überflüssig macht. In der Arbeit von Goyal (1977) wird für den Zweistufenfall vorgeschlagen, dass das Produktionslos nicht in gleichgroße Transportlose zu zerteilen ist, sondern deren Größe gemäß einer geometrischen Reihe festgelegt wird. Szendrovitz/Drezner (1980) erweitern die Problemstellung dahingehend, dass zwar stufenweise gleiche Transferlosgrößen, zwischen den Stufen aber unterschiedlich viele Transporte zulässig sind. In einer Simulationsstudie belegen Szendrovits/Golden (1984) die Vorteile der überlappenden Fertigung bei gleichgroßen und hinsichtlich der Größe optimierten Transferlosen gegenüber der geschlossen Weitergabe bei optimalen variablen Produktionslosen, die sie mit Hilfe des Modells von Schwarz/Schrage (1975) ermitteln. Szendrovits/Golden kommen zu dem Ergebnis, dass überlappende Lose deutliche Kostenvorteile bieten. Dieses Ergebnis wird durch die Arbeit von Drezner et al. (1984) bestätigt, die variable Produktionslose und überlappende Weitergabe kombiniert einsetzen.

Hinsichtlich der Transportkosten hebt Szendrovits (1978) hervor, dass sie in ihrer Konsequenz ähnlich zu interpretieren sind wie Kapazitätsrestriktionen, die bezogen auf das Transportmittel bestehen können. Folglich entwickeln Goyal/Szendrovits (1986) ein Modell, in das sie für den Mehrstufenfall diese Form der Kapazitierung aufnehmen. Sie lassen sowohl gleiche als auch untereinander abweichende Transferlosgrößen zu und schlagen zur Lösung des Modells eine Heuristik vor.

Graves/Kostreva (1986) untersuchen die Übertragung des von Szendrovits (1976) vorgestellten Ansatzes auf ein MRP System. Zur Vereinfachung geben sie vor, dass Produktionslose nur als ganzzahlige Vielfache der Transportlosgrößen verwendet werden dürfen und dass auf allen Stufen mit der gleichen Produktionsgeschwindigkeit gearbeitet wird. Außerdem schlagen sie vor, die Überlappung nur für jeweils zwei Stufen anzuwenden. Unter diesen sehr starken Einschränkungen ist es möglich, die optimale Produktionslosgröße und die optimale Transportlosgröße unter Kostenkriterien zu bestimmen. An Hand eines einfachen Beispiels können die Autoren durch die Überlappung eine Kostensenkung von 22% erreichen.

Ramasesh et al. (2000) setzen auf den Ergebnissen von Graves/Kostreva (1986) auf, erweitern aber die Betrachtung um die Berücksichtigung zeitlicher Aspekte, wie z.B. Transportzeiten, Wartezeiten und Rüstzeiten. Diese wurden in den bisher angesprochenen Modellen der Losgrößenplanung mit überlappenden Losen nur hinsichtlich ihrer Kostenwirkung, nicht aber mit ihrer verzögernden Wirkung auf die Produktion erfasst. Um das Modell analysierbar zu halten, müssen Ramasesh et al. (2000) allerdings von gleichgroßen Teillosen und unterbrechungsfreier Produktion ausgehen. Bogaschewsky et al. (2001) greifen schließlich das Modell von Goyal (1977b) auf und erweitern es für den Mehrstufenfall.

Während alle bisher angesprochenen Modelle davon ausgehen, dass die Produktion auf den Stufen ohne Unterbrechung ablaufen muss, untersucht Szendrovits (1987) eine mehrstufige Fertigung, bei der er Unterbrechungen mit Strafkosten gewichtet zulässt und annimmt, dass die leerstehende Maschine den Rüstzustand nicht verliert. Zur Weitergabe der Teile sieht er wieder gleichgroße Transferlose vor.

Bemerkenswert ist, dass – bis auf Ramasesh et al., 2000 – in den auf die Losgrößenplanung fokussierten Arbeiten der Begriff des Lot Streaming nicht verwendet wird. Im Kontext der Diskussion mehrstufiger Lagerhaltungsmodelle setzt Moily (1986) den Begriff zwar ein, verwendet ihn aber in einer aus heutiger Sicht missverständlichen Weise. Moily differenziert zwischen den beiden Politiken ISLM (integer split lot requirement) und IMLR (integer multiple lot requirement) und schreibt dazu (vgl. Moily, 1986, S. 117): „... This strategy, often referred to as ‚lot streaming' , is the common boundary between the ISLR

and IMLR lot-sizing policies.“ Gemeint ist in diesem Fall, dass sich die Größe eines Loses für die Komponenten eines Produktes über die Stufen betrachtet weder durch Zusammenfassen mit anderen Losen (IMLR) noch durch Zerteilen in kleinere Einheiten (ISLM) verändert. Eine überlappende Fertigung eines Loses auf mehreren Stufen ist dabei allerdings nicht vorgesehen, so dass aus heutiger Sicht der Begriff nicht zutreffend verwendet wird.

Wird die überlappende Fertigung bei der Festlegung der Losgrößen aus Kostengesichtspunkten betrachtet, können die bislang verfügbaren Modelle keine Entscheidung dahingehend liefern, welche Art von Transferlosen einzusetzen sind und wie stark die Überlappung auszufallen hat. Eine zusätzliche Berücksichtigung dieses Aspekts erscheint in den bisher vorgestellten Modellen nicht sinnvoll, da sie dann nicht mehr handhabbar werden.

5.5.2 Losüberlappung aus Sicht der Ablaufplanung

In diesem Abschnitt werden die Arbeiten kurz vorgestellt, die dem Scheduling zuzuordnen sind und auf die im Zuge der bisherigen Untersuchung – mangels praktischer Einsatzmöglichkeiten oder geringer Relevanz – nicht näher eingegangen wurde. Dabei handelt es sich um die Bereiche:

- Vorläufer der Losüberlappung,
- Rüstprozesse und Losüberlappung,
- Losüberlappung im Mehrproduktfall,
- Effekt der Losüberlappung unter verschiedenen Zielkriterien,
- Losüberlappung bei diskreten Losen.

Da sich einige der Arbeiten mehreren Bereichen zuordnen lassen, werden diese auch mehrfach angesprochen.

5.5.2.1 Vorläufer der Losüberlappung

Als die wichtigsten Vorläufer der Losüberlappung im Scheduling sind die Arbeiten von Kulonda (1984) sowie Jacobs (1983), Jacobs (1984) und Jacobs/Bragg (1988) zu nennen. Kulonda (1984) führt an Hand eines Beispiels exemplarisch den Effekt der Überlappung auf die Durchlaufzeit in einer Montagestruktur vor. Es werden für zwei Komponenten jeweils eine Überlappung mit zwei konstanten Losen und mit zwei variablen Losen gegenübergestellt und beide mit der Situation verglichen, die sich bei vier konstanten Losen ergibt. Ein Verfahren oder zumindest eine Idee, nach der die Größe der Lose festzulegen sind, fehlt allerdings. In Jacobs (1983) und Jacobs (1984) wird der Einsatz

des OPT-Konzepts (vgl. z.B. Goldratt, 1986) diskutiert. Dabei stellt Jacobs insbesondere auf eine der neun OPT-Regeln ab, nach der sich die Produktionslosgrößen und die Transferlosgrößen nicht entsprechen müssen. Jacobs kommt zu dem Ergebnis, dass weite Teile des Erfolges, der dem OPT-Konzept beziehungsweise dem Einsatz der nicht frei zugänglichen OPT-Software zugeschrieben wird, auf einfache organisatorische Veränderungen und insbesondere auf die Weitergabe von Teilen aus laufenden Losen zurückzuführen ist. Eine formale Analyse der mit Losüberlappung erzielbaren Effekte nimmt Jacobs aber nicht vor. Dafür stellt er in einer Simulationsstudie (vgl. Jacobs/Bragg, 1988) das Konzept der „repetitive lots“ vor. Dieser erste Beleg für den Vorteil der überlappenden Fertigung ist als der eigentliche Ausgangspunkt der Diskussion der Losüberlappung im Kontext des Scheduling zu beurteilen. In einem zehnstufigen Job Shop erzeugen Jacobs/Bragg (1988) zehn Produktarten. Für jede Produktart gilt eine feste aber zufällig erzeugte Maschinenfolge. Die Produkte werden in variierenden Losgrößen {100, 150, 200, 250, 300} in das System eingelastet. Die Bearbeitung auf den Stufen erfolgt mittels Prioritätsregeln (z.B. FIFO oder SPT) und setzt voraus, dass die Maschine zur Bearbeitung des Produkttyps gerüstet ist. Finden sich in der Warteschlange vor einer Maschine keine Produkte mehr, die zum Rüstzustand der Maschine passen, wird diese auf den Produkttyp umgerüstet, auf den ein Produkt in der Warteschlange am längsten wartet. Zur Weitergabe zwischen den Stufen setzten Jacobs/Bragg (1988) feste Transferlose – die sie repetitive lots nennen – mit vorgegebenen Größen {10, 50} ein. Es finden damit Weitergaben von einer Stufe statt, während auf dieser Stufe noch Teile des gleichen Typs gefertigt werden. Als Vergleich dient die geschlossene Weitergabe der Lose, wobei über die Reihenfolge zwischen den wartenden Losen mit den zwei oben genanten Prioritätsregeln entschieden wird. Als Zielkriterium wird der Effekt auf die Durchlaufzeiten betrachtet und untersucht, wie sich die für den Rüstprozess benötigte Zeitspanne durch den Einsatz der vergleichsweise kleinen Transferlose verändert. Hinsichtlich der Resultate, die mit den verschiedenen Kombinationen der Losgrößen, der Transferlose und der Prioritätsregeln erzielbar sind (vgl. Jacobs/Bragg, 1988, Table 2, S. 291), können die Durchlaufzeiten um bis zu 35% verringert werden. Die Autoren räumen allerdings ein, dass es mit Transferlosen, die nicht auf diese zwei willkürlichen Werte festgelegt werden, sondern z.B. in Abhängigkeit der Stufen variieren dürfen, zu weit höheren Einsparungseffekte kommen kann (vgl. Jacobs/Bragg, 1988, S. 293). Wie diese Losgrößen festgelegt werden können, bleibt allerdings offen. An dieser Stelle setzen später die Arbeiten von Potts/Baker (1989), Baker/Pyke (1990), Baker/Jia (1993) und Trietsch/ Baker (1993) ein, die das Phänomen der Losüberlappung zunächst im Zwei- und im Dreistufenfall analytisch untersuchen, Algorithmen vorschlagen und wichtige Eigenschaften optimaler Lösungen herleiten. Auf die wesentlichen Ergebnisse dieser Arbeiten wurde in Kapitel 5.2 bereits im Detail eingegangen.

5.5.2.2 Rüstprozesse und Losüberlappung

In Verbindung mit der Losüberlappung hat die Berücksichtigung von Rüstprozessen eine Reihe von Autoren beschäftigt. Dabei sind zwei Forschungsrichtungen entstanden, die sich dahingehend unterschieden,

- ob zwischen den Stufen eine nicht beschränkte Zahl an Transfers zulässig ist,
- oder nach einer Aufteilung in eine vorgegebene Zahl an Teillosen gesucht wird.

Die erste Forschungsrichtung findet in der Arbeit von Vickson/Alfredson (1992) ihren Ursprung. Die Autoren untersuchen den Zwei- und Dreistufenfall mit mehreren Produkten unter regulären Zielkriterien, wobei sie beliebig viele Transporte zulassen und zudem unterstellen, dass die Rüstdauer Null ZE beträgt. In der Konsequenz wird jedes einzelne Stück sofort nach seiner Fertigstellung auf der Stufe 1 zu der Stufe 2 verbracht. Gesucht ist die optimale Sequenz, in der die Produkte aufgelegt und weitergegeben werden sollen. Dabei zeigen Vickson/Alfredson (1992), dass mit einer geringen Modifikation des Algorithmus von Johnson (1954) im Zwei- und im Dreistufenfall die optimale Lösung ermittelt werden kann. Allerdings kommen die Autoren auch zu dem Ergebnis, dass eine Berücksichtigung von Rüstzeiten die Problematik erheblich kompliziert. Çetınkaya/Kayaligıl (1992) greifen diesen Punkt auf und unterstellen antizipatives Rüsten. Für den Zweistufenfall können sie zeigen, dass weiterhin eine Modifikation des Algorithmus von Johnson (1954) zur optimalen Lösung führt. Çetınkaya (1994) erweitert die Betrachtung um die Berücksichtigung der sogenannten „removal times“. Diese Zeitspanne wird nach Beendigung der Bearbeitung benötigt, um die Maschine z.B. zu reinigen, die benutzten Werkzeuge zu demontieren etc. Sie kann in dem Moment einsetzen, in dem die Bearbeitung abgeschlossen ist, verzögert aber nicht die Weitergabe sondern schiebt nur den Moment hinaus, in dem für das nächste Stück eingerüstet werden kann. Sowohl die Rüstdauer als auch die Zeitspanne zum Abrüsten wird als auftragsindividuell aber reihenfolge-unabhängig modelliert. Mit den Verfahren von Yoshida/Hitomi (1979) und Sule/Huang (1983) greift Çetınkaya (1994) auf zwei Weiterentwicklungen des Verfahrens von Johnson (1954) zurück, die antizipatives Rüsten und Abrüsten im klassischen Flow Shop mit zwei Stufen bewältigen. Er zeigt, dass diese Verfahren mit leichten Modifikationen auch bei der Losüberlappung mit einer Weitergabe Stück für Stück zum optimalen Ergebnis im Zweistufenfall führen.

Baker (1995) greift diese Ergebnisse auf, vereinheitlicht sie und zeigt, dass auch verbundenes Rüsten im Zweistufenfall mit einer leichten Modifikation des Verfahrens von Johnson (1954) optimal lösbar ist. Analog zu den Modellen der Losüberlappung ohne Rüsten können die erste oder die zweite Stufe als domi-

nierte Stufe identifiziert und zu einer Fallunterscheidung herangezogen werden. Baker (1995) führt zudem eine Netzwerk-Darstellung des Problems mit Rüsten ein. Vickson (1995) betrachtet im Zweistufenfall beide Rüsttypen und lässt ergänzend Transportzeiten anfallen, die sowohl einen fixen Bestandteil als auch eine mit der Losgröße variierende Komponente beinhalten. Sowohl für den kontinuierlichen Fall als auch für den diskreten Fall ist es möglich, die Sequenzentscheidung von der Entscheidung hinsichtlich der Größe der Teillose zu separieren, wobei allerdings vorausgesetzt werden muss, dass die Zwischenlagerkapazität nicht zum Engpass wird. Erneut ist das Ergebnis nicht auf den Mehrstufenfall übertragbar und wieder kann für die Festlegung der Sequenz auf Johnson (1954) zurückgegriffen werden.

Die zweite Forschungsrichtung ist eng mit der klassischen Fragestellung der Losüberlappung verbunden und betrachtet die Aufteilung einer gegebenen Quantität im Einproduktfall auf eine vorgegebene Zahl an Teillosen. Den Ausgangpunkt der formalen Analyse setzen hier Chen/Steiner die sowohl für den Fall des antizipativen Rüstens (vgl. Chen/Steiner, 1996) als auch im Fall des verbundenen Rüstens (vgl. Chen/Steiner, 1998) für den Dreistufenfall nach strukturellen Eigenschaften optimaler Lösungen suchen. Stark angelehnt an die Arbeit von Glass et al. (1994) gehen sie mit Hilfe der Netzwerkdarstellung des Problems vor, wobei es ihnen gelingt, für weite Teile der Ergebnisse von Glass et al. (1994) eine analoge Variante unter Berücksichtigung von Rüstvorgängen zu finden. Erneut zeigt sich, dass eine Übertragung der Ergebnisse auf den Mehrstufenfall nicht vorgenommen werden kann.

Sowohl für den Einprodukt- als auch für den Mehrproduktfall betrachten Sriskandarajah/Wagneur (1999) die Losüberlappung auf zwei Stufen mit antizipativem Rüsten. Allerdings setzen sie wartezeitfreie Pläne voraus, womit sich im Mehrproduktfall die Sequenzfestlegung vereinfacht, da die Sequenz zwischen den Maschinen nicht variieren kann, ohne dass gegen die wartezeitfreie Weiterbearbeitung verstoßen würde. Das Problem ist im Zweistufenfall sowohl für kontinuierliche Losgrößen als auch für diskrete Losgrößen mit polynomial begrenztem Aufwand optimal lösbar.

Den Einproduktfall mit verbundenem Rüsten auf zwei Stufen betrachten Bukchin et al. (2002). Sie lassen konsistente Losgrößen zu und betrachten mit der durchschnittlichen Durchlaufzeit ein Zielkriterium, das von den übrigen Arbeiten abweicht. Die Autoren geben auf Basis einer Fallunterscheidung – je nach der Maschine, die den Engpass darstellt – ein einfaches Lösungsverfahren an. Indem sie 40 Testinstanzen sowohl unter der Zielsetzung der Durchlaufzeitminimierung als auch hinsichtlich der mittleren Durchlaufzeit vergleichen, zeigt sich, dass die Struktur der Lösungen sehr stark in Abhängigkeit des Zielkriteriums variiert.

Während in den Arbeiten von Baker (1995), Vickson (1995), Chen/Steiner (1996), Chen/Steiner (1998) der Rüstprozess mit dem Los, nicht aber mit dem Teillos verbunden modelliert wird, bilden Kalir/Sarin (2003) das mit dem einzelnen Teillos verbundene Rüsten ab. Sie behandeln den Mehrproduktfall auf beliebig vielen Stufen, lassen aber nur konstante Freigabelose und Transferlose zu. Die vorgeschlagene Heuristik findet für den Zweistufenfall die optimale Lösung.

5.5.2.3 Losüberlappung im Mehrproduktfall

Die meisten Arbeiten zur Losüberlappung beschäftigen sich mit dem Einproduktfall, da die Erweiterung auf den Mehrproduktfall eine wesentlich schwieriger zu handhabende Fragestellung aufwirft. Neben der Entscheidung über die Aufteilung der Losgröße in Teillose muss über die Sequenz entschieden werden, in denen diese eingelastet werden sollen. Prinzipiell bieten sich hier vier Varianten der Vorgehensweise an.

(1) Es wird simultan über die Aufteilung und die Sequenz entschieden.

(2) Es wird zunächst die Aufteilung in die Teillose vorgenommen und im Anschluss die Sequenz bestimmt. Die Aufteilung in Teillose bleibt dabei fixiert und es wird nicht zugelassen, dass Teillose eines Produkts zwischen den Teillosen eines anderen Produkts gefertigt werden.

(3) Es wird wie unter Punkt (2) vorgegangen, nur dürfen Teillose eines Produkts zwischen den Teillosen eines anderen Produkts gefertigt werden. Diese Vermischung wird auch als „intermingling“ (vgl. z.B. Kalir, 1999) bezeichnet.

(4) Es wird erst die Sequenz bestimmt und im Anschluss die Aufteilung vorgenommen. Die Sequenz bleibt dabei fixiert, so dass keine Teillose zwischen den Teillosen eines anderen Produkts gefertigt werden können.

Wie bereits Potts/Baker (1989) im Abschluss ihres Aufsatzes für den Zweistufenfall zeigen, existieren für das Mehrprodukt-Losüberlappungsproblem Instanzen, für die die optimale Lösung nicht erreicht werden kann, wenn sequentiell über die beiden Aspekte entschieden wird. Da aber die simultane Festlegung der Losgrößen und ihrer Sequenz nicht handhabbar ist, wird in den Arbeiten, die diese Fragestellung im Flow Shop bisher aufgreifen, eine wartezeitfreie Bearbeitung der Lose als zusätzliche Restriktion gefordert. Mit dieser Einschränkung vereinfacht sich das Problem stark. Der Flow Shop wird zu einem Permutations Flow Shop (vgl. z.B. Monma/Rinnooy Kan, 1983; Bierwirth, 1993), da nur eine Sequenz zu bestimmen ist, in der die Teillose dann über alle Stufen aufgelegt werden müssen, um ihre wartezeitfreie Bearbeitung sicherzustellen. Bei Losverfügbarkeit und Wartezeitfreiheit sind zudem nur konsistente

oder konstante Lose zulässig, da mit variablen Losen die Forderung nach der wartezeitfreien Bearbeitung in sich widersprüchlich wird.

Sriskandarajah/Wagneur (1999) untersuchen die Losüberlappung im zweistufigen Flow Shop unter Wartezeitfreiheit. Für jedes Produkt wird eine vorgegebene Zahl an Losen vorgesehen und es wird davon ausgegangen, dass die Fertigung eines Produkts nicht durch die Fertigung anderer unterbrochen wird. Das Verfahren ist damit der Variante (2) zuzuordnen. Sie schlagen für den Mehrproduktfall eine Heuristik vor, wobei zur Festlegung der Sequenz die inhaltliche Nähe zum Traveling Salesman Problem ausgenutzt wird.

Kumar et al. (2000) erweitern die Fragestellung auf den Mehrstufenfall. Neben einer Heuristik, die zunächst die Größe der überlappenden Lose festlegt und anschließend die Sequenz bestimmt, stellen sie einen genetischen Algorithmus vor, der die beiden Aspekte simultan behandelt. Dieser Ansatz ist somit der Variante (1) zuzurechnen.

Hall et al. (2003) greifen die Heuristik von Kumar et al. (2000) auf und stellen ihr einen ebenfalls heuristischen Ansatz gegenüber. Sie gehen wieder davon aus, dass die Fertigung eines Produkts nicht durch die Fertigung anderer Produkte unterbrochen wird. Da sie diskrete Lose unterstellen, ergibt sich für jedes Produkt (bei wiederum vorgegebener Loszahl) eine vergleichsweise geringe Zahl an möglichen Aufteilungen in die Teillose, die zudem für den Einsatz einer Metaheuristik leichter zu kodieren sind, als die kontinuierlichen Lose. Jede dieser Aufteilungen kann einem Ort in einem generalisierten Travelling Salesman Problem (GTSP) gleichgesetzt werden. In einem GTSP sind Orte in Länder zusammengefasst und die Aufgabe besteht darin, in jedem Land genau einen Ort aufzusuchen. Übertragen auf das Losüberlappungsproblem bedeutet dies, dass Hall et al. (2003) je einem Ort eine mögliche Form der Aufteilung in die Teillose für ein Produkt gleichsetzen. Ein Land besteht dann aus den verschiedenen Möglichkeiten, ein gegebenes Produkt in die Teillose aufzuteilen und umfasst somit alle möglichen Überlappungen für ein Produkt. Mittels Tabu Search wird aus den möglichen Losaufteilungen insofern eine Auswahl vorgenommen als dass erste Rundreisen im GTSP entstehen. Die besten, mit Tabu Search gefundenen Lösungen bilden die Ausgangspopulation für einen genetischen Algorithmus. Dieses sehr aufwendige Vorgehen ist der Variante (2) zuzuordnen. Es liefert zwar besser Ergebnisse als die Heuristik von Kumar et al. (2000), zugleich steigt der Rechenaufwand aber stark an.

Während die bisher genannten Ansätze zur Losüberlappung im Mehrproduktfall alle die Vereinfachung der wartezeitfreien Bearbeitung unterstellen, lassen Dauzère-Pérès/Lasserre (1997) für das Losüberlappungsproblem bei mehreren Produkten im Job Shop Wartezeiten zu. Der dort vorgeschlagene Ansatz ist ebenfalls heuristisch und besteht letztlich aus zwei Modulen, die iterativ zur Anwendung kommen. Im ersten Modul wird zunächst die Sequenz festge-

legt, in der die Lose auf den Maschinen zu bearbeiten sind. Diese Sequenz wird festgehalten und eine Losüberlappung mit konsistenten Losen vorgenommen, womit diese Vorgehensweise der Variante (4) zuzuordnen ist. Im zweiten Modul wird zunächst eine Aufteilung in eine vorgegebene Zahl an konsistenten Losen pro Job vorgenommen. Für diese wird die Sequenz festgelegt, wobei jedes einzelne Teillos wie ein von den übrigen unabhängiger Job aufgefasst wird. Dieses zweite Modul lässt somit „intermingling“ zu, womit das Verfahren von Dauzère-Pérès/Lasserre (1997) zwischen der Anwendung der Varianten (3) und (4) alterniert. Im Ergebnis zeigt sich eine erhebliche Reduktion der Durchlaufzeit, allerdings kann die Güte der Heuristik mangels exakter Verfahren nicht bestimmt werden. Als interessante Beobachtung ist allerdings festzuhalten, dass Dauzère-Pérès/Lasserre (1997) auf Basis von 120 Instanzen erneut beobachten, dass eine geringe Zahl an Losen die vergleichsweise höchsten Effekte zeigt.

Eine andere Möglichkeit, die Schwierigkeiten der Losüberlappung im Mehrstufenfall beherrschbar zu gestalten, besteht darin, die Losart auf konstante Lose zu beschränken. Eine umfassende Analyse dieser Problemstellung findet sich bei Kalir (1999). Sie wurde in Ausschnitten in Kalir/Sarin (2001) für den Einproduktfall mit Rüstvorgängen und in Kalir/Sarin (2001b) für den Mehrproduktfall bei fixierter Sequenz der Produkte veröffentlicht, womit die Fragestellung der Variante (2) folgt. Eine Erweiterung auf den Mehrproduktfall, wenn mit dem Teillos verbundene Rüstvorgängen zu beachten sind, bieten Kalir/Sarin (2003). Diesen drei Arbeiten liegt im Wesentlichen die Idee zu Grunde, dass ein gegenläufiger Effekt auftritt, wenn die Zahl der Lose erhöht wird. Werden zusätzliche konstante Lose eingesetzt, reduziert dies zwar die Durchlaufzeit und senkt die ablaufbedingten Leerzeiten. Gleichzeitig werden für die zusätzlichen Lose aber Rüstvorgänge benötigt. Jedes zusätzliche Los senkt somit einerseits die Durchlaufzeit und reduziert die Zeitspanne, in der Stufen auf Lose warten. Andererseits erhöhen ab einem bestimmten Punkt die zusätzlichen Rüstzeiten die Durchlaufzeit mehr als dass eine weitere Aufteilung sie senken kann. In den von Kalir/Sarin veröffentlichten Arbeiten wird deshalb ein Suchverfahren vorgeschlagen, mit dem die Zahl der Lose solange variiert wird, bis sich der anfänglich fallende Verlauf der Durchlaufzeit umzukehren beginnt.

Mit der Losüberlappung im zweistufigen Open-Shop befasst sich die Arbeit von Şen/Benli (1999). Sie unterstellen Teillose, die eine geometrischen Folge bilden und untersuchen, welche Auswirkungen die Aufteilung hat, wenn für alle Teillose eines Produkts die gleiche Maschinenfolge gewählt wird (Variante (2)) oder wenn die Maschinenfolge für jedes Teillos unabhängig von den übrigen festgelegt werden darf (Variante (3)). Da nur zwei Stufen betrachtet werden, können für beide Fälle Formeln zur Bestimmung der optimalen Durchlaufzeit angegeben werden. Eine Übertragung auf den Mehrstufenfall ist allerdings wieder nicht zulässig.

Insgesamt zeigt der Entwicklungsstand der Modelle für den Mehrstufenfall, dass nur unter bestimmten, die Praxisrelevanz stark einschränkenden Bedingungen eine Optimierung erfolgen kann. Es ist offensichtlich, dass man dann die Durchlaufzeit am stärksten reduziert kann, wenn „intermingling“ zugelassen wird. Dann ist aber auch ein um Dimensionen größeres Scheduling-Problem zu lösen, so dass es wesentlich darauf ankommt, ob leistungsfähige Heuristiken für das zu betrachtende Ablaufplanungsproblem existieren oder nicht.

5.5.2.4 Losüberlappung unter verschiedenen Zielkriterien

Während die bisher behandelten Untersuchungen durchgängig reguläre Zielkriterien unterstellen, greifen Wagner/Ragatz (1994) den Ansatz der repetitive lots nach Jacobs/Bragg (1988) erneut auf und untersuchen ihn unter Zeitkriterien. In einer Simulationsstudie untersuchen sie die Auswirkungen der Weitergabe in festen Transferlosen auf die mittlere Terminabweichung und die Zahl verspäteter Aufträge. Dazu unterstellen sie einen Job Shop mit 5 Maschinen, der im Mittel 300 Lose mit je 50 bis 700 Stück bearbeitet. Sowohl die Maschinenfolge als auch der Liefertermin des Loses wird zufällig gesetzt, wobei jedes Teil im Mittel 4,16 Bearbeitungen erfährt. Zur Festlegung der Bearbeitungsreihenfolge auf den Maschinen setzen Wagner/Ragatz (1994) verschiedene Prioritätsregeln in Kombination mit unterschiedlich großen Transferlosen ein. Sie zeigen, dass durch den Einsatz der überlappenden Lose eine Reduktion der mittleren Verspätung um bis zu 39% einsetzen kann. Interessant ist dabei, dass dieser Effekt unabhängig von der Höhe der Rüstzeiten auftritt.

Wie sich nicht reguläre Zielkriterien auf die optimale Größe der Lose auswirken, untersuchen Şen et al. (1998) für den Einproduktfall im Flow Shop. Dabei berücksichtigen die Autoren die mittlere, mit der Losgröße gewichtete Durchlaufzeit der Teillose sowie auf die Fertigstellung einzelner Stücke bezogene Kriterien und geben Verfahren beziehungsweise Formeln an, mit denen die optimalen konsistenten Losgrößen zu bestimmen sind. Da sich die Herleitung als sehr anspruchsvoll erweist (vgl. Şen et al. (1998) S. 50-62) können nur Aussagen für den Zweistufenfall angegeben werden.

Beschränkt man sich hingegen auf konstante Lose, kann der Effekt der Losüberlappung auf die Durchlaufzeit, die mittlere Fertigstellungszeit und den mittleren Bestand gemäß der von Kalir/Sarin (2000) ermittelten oberen Schranken abgeschätzt werden. Diese gelten im Mehrstufenfall für den Einprodukt und für den Mehrproduktfall. Dabei zeigt sich unter anderem, dass der mittlere Bestand bei Losüberlappung im günstigsten Fall um bis zu 50% gesenkt werden kann. Die ermittelten Formeln grenzen aber nur die maximal mögliche Verbesserung nach oben ab. Eine Aussage über die wahrscheinlich zu erwartende Verbesserung des Zielkriteriums ist nicht verfügbar.

Für den Mehrproduktfall betrachten Yoon/Ventura (2002) die mittlere, mit Verfrühungs- und Verspätungsstrafen gewichtete Terminabweichung in einem Flow Shop und untersuchen den Effekt der Losüberlappung mit konsistenten Losen auf dieses nicht reguläre Zeitkriterium. Dabei modellieren sie – in Erweiterung zu den übrigen Studien im Mehrproduktfall – begrenzte Zwischenlager, die zu einer Blockierung der eigentlich freien Vorstufe führen können. Sie betrachten 2 bis 5 Stufen auf denen bis zu 20 Produkte gefertigt werden. Zur Lösung schlagen die Autoren folgendes Verfahren vor: Zunächst werden vier Sequenzen mit Hilfe einfacher Prioritätsregeln generiert. Für jede gegebene Sequenz der Produkte wird die optimale Überlappung mit konsistenten Losen erzeugt und die Lösung hinsichtlich der mittleren gewichteten Terminabweichung bewertet. Mit Hilfe des paarweisen Austauschens von Produkten wird jede der vier Sequenz solange verändert, bis erstmalig keine Verbesserung mehr auftritt; dann bricht das Verfahren ab. Der Algorithmus ist damit als eine Heuristik zu klassifizieren, die der vierten Variante nach der Unterscheidung im Abschnitt 5.5.2.3 zuzuordnen ist. Da es bisher keinen exakten Ansatz und auch keine andere Heuristik zur Lösung dieses Problems gibt, vergleichen die Autoren die Zielfunktionswerte mit denen, die man ohne Losüberlappung erreichen kann. Dabei tritt eine mittlere Verbesserung von über 14% auf.

5.5.2.5 Losüberlappung bei diskreten Losen

Mit der Fragestellung, wie diskrete Lose festzulegen sind, haben sich zunächst Chen/Steiner (1997) befasst. Sie untersuchen den Effekt, den das Runden von konsistenten Losen im Mehrstufen-Einproduktfall haben kann. Allerdings fällt die dort ermittelte Abschätzung sehr grob aus und kann nur eingeschränkt weiterhelfen. Die Autoren ermitteln, dass die Durchlaufzeit für die gerundete Lösung C^G immer in einem Intervall zwischen der optimalen diskreten Lösung C^D und der kontinuierlichen Lösung C^K liegen muss. Für dieses Intervall gilt:

$$C^D \le C^G \le C^K + \sum_{m=2}^{M} p_m + (p_{\max} - p_{\min}) J / 2 \tag{92}$$

Wird diese Abgrenzung aber auf das Beispiel C mit $\underline{p}$ = (2; 4; 6; 2; 3) und J = 5 angewendet, ergibt sich, dass die gerundete Lösung maximal 20 ZE schlechter sein kann, als die optimale diskrete Lösung: 15 + (6 – 4)5/2 = 20. Wird das Runden hingegen gemäß der von Chen/Steiner (1997) vorgeschlagenen Balanced Redistribution Heuristik vorgenommen, wird die dann auftretende Durchlaufzeit C^B in dem folgenden Intervall liegen:

$$C^D \le C^B < C^K + \sum_{m=1}^{M-1} p_m \le C^D + \sum_{m=1}^{M-1} p_m \tag{93}$$

Für das Beispiel C mit $J = 5$ Losen weiß man somit nur, dass die heuristische Lösung maximal 14 ZE von der besten Lösung entfernt liegen kann. Bezogen auf die Durchlaufzeit von 224 ZE erscheinen diese 20 beziehungsweise 14 ZE doch relativ hoch. Daher sollte – sofern die Problemgröße dies zulässt – das lineare Programm für ganzzahlige Losgrößen gerechnet werden. Spezielle, auf die Bestimmung von diskreten Losen zugeschnittene exakte Verfahren existieren bislang lediglich für den Zweistufenfall bei konsistenten Losen (vgl. Chen/Steiner, 1999). Eine Heuristik für den Mehrstufenfall mit variablen Losen unter Stückverfügbarkeit wurde zwar von Liu (2003) präsentiert, eine obere Schranke für den maximalen Fehler oder eine Evaluierung mit Hilfe eines exakten Ansatzes fehlen dort allerdings, womit eine objektive Beurteilung des Verfahrens nicht möglich ist. Auf das Verfahren von Hall et al (2003) wurde oben bereits eingegangen. Dort aber ist die optimale Festlegung von diskreten Losen nicht der primäre Untersuchungsgegenstand; vielmehr dient der Rückgriff auf die diskreten Lose der Problemvereinfachung, da der Lösungsraum „diskretisiert“ wird.

Die zusammenfassende Darstellung der übrigen Verfahren hat gezeigt, dass es bisher keine Testinstanzen gibt, mit denen studienübergreifend die Effekte der Losüberlappung validiert werden können. Dieses Fehlen einer gemeinsamen Datengrundlage macht es schwer, den Beitrag der einzelnen Arbeiten einzuschätzen. Zudem werden oft Spezialfälle behandelt, die sich wegen ihrer Spezialisierung einem Vergleich entziehen. Weiter fehlen bislang leistungsfähige – exakte oder heuristische – Ansätze zur Losüberlappung bei diskreten Losen im Mehrstufenfall. Beispielsweise wurde das Problem der exakten Festlegung von variablen, diskreten Losen unter Stückverfügbarkeit noch nie untersucht; außerdem bietet der Mehrproduktfall eine Fülle von bisher nicht betrachteten und schweren Fragestellungen. Mit Rückgriff auf den decision calculus (vgl. Little, 1970), ist es allerdings fraglich, ob Modelle zur Losüberlappung im Mehrproduktfall jemals praktisch umgesetzt werden können. Momentan erscheint daher die Berücksichtigung diskreter Lose geboten, um die dort bestehenden großen Forschungslücken zu schließen.

6. Zusammenfassung

Den Anstoß zu dieser Arbeit lieferte die Beobachtung, dass eine in der Praxis der Produktionsplanung regelmäßig verwendete und kurzfristig einsetzbare Anpassungsmaßnahme vom Mainstream kaum aufgegriffen wird. Im ersten Teil der Arbeit (Kapitel 2 und 3) wird daher versucht, den Ursachen dieser geringen Resonanz nachzugehen. Dazu werden die Produkt-Prozess-Matrix nach Hayes/Wheelwright (1979) und Spencer/Cox (1995) sowie das Konzept von Anthony (1965) als Ausgangspunkte gewählt.

Durch ihre breite Akzeptanz haben diese Konzepte sowohl die Forschung als auch die Ausgestaltung der PPS-Systeme stark mitbestimmt. Die Diskussion ihrer Konsequenzen zeigt, dass die Hierarchisierung und zeitliche Dekomposition den Einsatz von Modellen nahe legt, deren Standardannahmen eine Berücksichtigung der überlappenden Weitergabe ausschließen. Soll über das Ausmaß der Überlappung der Lose oder die Art der Lose entschieden werden, kann dieser Aspekt nicht sinnvoll in die Losgrößenmodelle der taktischen Ebene integriert werden. Diese Modelle sind notwendigerweise mit einer diskreten Zeitführung als Produktionsmengenmodelle formuliert und setzen ebenso typisch die geschlossene Weitergabe über die Lagerbilanzgleichungen voraus. Ebenso stellt sich für die operative Ebene (für die Ablaufplanung) die Berücksichtigung der überlappenden Weitergabe als eine Herausforderung dar. Dort wird mit Produktionszeitenmodellen gearbeitet, die auf eine kontinuierliche Zeitführung zurückgreifen, zugleich aber annehmen, dass die Größe der Aufträge determiniert ist. Es zeigt sich, dass die Losüberlappung eine Diskussion auslöst, die in keine der beiden Modellwelten hineinpasst.

Auf der taktischen Ebene hat die überlappende Weitergabe zudem den Charakter einer Reparaturmaßnahme. Sie kann eigentlich nur deshalb notwendig sein, weil die Vorgaben der Modelle nicht richtig umgesetzt werden, beziehungsweise es *noch nicht* gelungen ist, alle wesentlichen Aspekte angemessen zu modellieren. Damit wird aber zugleich ein Anspruch an die Modelle gestellt, dem sie aus Sicht des Autors gar nicht genügen sollten. Modelle schaffen Grundlagen für die betrieblichen Entscheidungen. Sie sind dabei zwangsläufig unvollständig, da nicht alle Zusammenhänge im Modell erfasst werden können oder aus Gründen der Rechenbarkeit sollen.

Die Frage, wie die Losüberlappung in den planerischen Ablauf aufzunehmen ist, berührt somit die Diskussion hinsichtlich der Anforderungen an Entscheidungsmodelle. Hier bietet Little (1970) mit seinem „decision calculus“ einen

geeigneten Ansatzpunkt: Einfachheit, Kommunikationsfähigkeit, Robustheit und Anpassungsfähigkeit sind auch heute noch wesentliche Anforderungen, die Modelle zu erfüllen haben. Hinzu kommt die zweifache Validierung der Modelle nach Schneeweiß (1999). Jedes Modell ist demnach auch im Rückgriff auf die Realität zu beurteilen. Wie im Abschnitt 3.2.1 gezeigt, werden im Zuge der strategischen Problemvereinfachung sehr häufig serielle Strukturen geschaffen. Folgt man den Ergebnissen von Smunt et al. (1996) und Kher et al. (2000) sind gerade diese Strukturen für einen erfolgreichen Einsatz der Losüberlappung vorteilhaft, so dass sich für diese Technik viele sinnvolle Einsatzmöglichkeiten ergeben. Wie im Abschnitt 3.3 dargestellt, ist die betriebliche Realität weiter durch den Einsatz der PPS-Systeme nach MRP-II Standard geprägt, für die ein Ersatz auf längere Zeit nicht in Sicht ist. Hinsichtlich der Idee, Lose überlappend weiterzugeben und diesen Vorgang planerisch zu unterstützen, haben viele der bisherigen PPS-Systeme noch deutliche Mängel (vgl. Gronau/Ibelings, 2002; François/Wolf, 2001). Zu untersuchen sind daher die Maßnahmen, die innerhalb des gegebenen Rahmens möglich sind. Dabei ist es dringend geboten, dass die nachgelagerte Ebene von ihren dezentralen Gestaltungsmöglichkeiten auch Gebrauch macht. Das schnell als Manko interpretierte Abweichen von Vorgaben ist nicht zu verhindern, sondern als eine wünschenswerte, dezentrale und kurzfristig einsetzbare Reaktion anzusehen, die es durch geeignete Verfahren zu unterstützen gilt.

Im Hauptteil der Arbeit (Kapitel 5) werden zunächst die Eigenschaften und Einsatzmöglichkeiten der Losüberlappung aufgearbeitet. Im Anschluss werden die Verfahren vorgestellt, die den Kriterien Praxisrelevanz, Rechenbarkeit und Erkenntnisbeitrag genügen. Dazu gliedert sich das fünfte Kapitel in mehrere Schwerpunkte, die sowohl der Darstellung als auch der Erweiterung einzelner Verfahren dienen. Zunächst werden die Ansätze vorgestellt, die für den Zwei- und den Dreistufenfall entwickelt wurden und die auch manuell umsetzbar sind. Danach wird die Aufteilung der Losgröße in konsistente Teillosen behandelt. Die zugehörigen Modelle arbeiten auf Basis der linearen Programmierung und können für die meisten praktischen Anwendungsgebiete als sehr gut geeignet empfohlen werden. Die Modelle sind – auch für größere Instanzen – in akzeptabler Zeit rechenbar. Zudem erlauben die Interpretation der Dualvariablen und der Einsatz der Sensitivitätsanalyse eine praxisrelevante, ökonomische Analyse, die über die bislang vorgeschlagene Vorgehensweise hinausgeht. Neben der Beurteilung von Plänen a priori kann auch eine angemessene Form der Reaktion auf Störungen a posteriori ermöglicht werden. Die bei der Sensitivitätsanalyse gefundenen Intervalle erlauben zudem, diverse Anpassungsmaßnahmen hinsichtlich ihrer Konsequenzen zu beurteilen, so dass dem Disponenten eine What-If-Analyse möglich ist.

In einem weiteren Schwerpunkt wird eine seit Jahren bestehende Forschungslücke geschlossen, indem ein Modell zur Festlegung variabler Lose un-

ter Losverfügbarkeit bei kontinuierlicher Zeitführung vorgestellt wird. Das Potential dieser Losart wird untersucht und gezeigt, dass sich variable Lose sehr gut zur Reduktion der Durchlaufzeit eignen, wenn z.B. hohe Rüstzeiten vorliegen oder Terminabweichungen zu berücksichtigen sind. Das fünfte Kapitel endet mit einer Charakterisierung derjenigen Verfahren, die den oben genannten drei Aspekten – zumindest bislang – nicht genügen können und bietet einen Ausblick auf die weitere Entwicklung dieses Gebiets an.

Insgesamt liegt mit dieser Arbeit die erste deutschsprachige Monographie vor, die sich der noch relativ jungen Forschungsrichtung der Losüberlappung annimmt. Sie eröffnet dem Leser die Möglichkeit, sich diesem Gebiet in einer geschlossenen Darstellung zu nähern, ohne auf das Studium zahlreicher Artikel in internationalen Fachzeitschriften zurückgreifen zu müssen. Dem Praktiker stellt sie die Verfahren im Detail vor, die unmittelbar oder mit geringem Aufwand einsetzbar sind. Den Entwicklern von PPS-Systemen zeigt diese Zusammenstellung der Verfahren auf, dass eine praxisorientierte Verbesserung der gegenwärtigen Systeme mit einem relativ geringen Aufwand umsetzbar ist.

Literaturverzeichnis

Aarts, E.H.L. / *Lenstra*, J.K. (Hrsg.): Local search in combinatorial optimization. Wiley & Sons, Chichester, 1997

Adam, D.: Produktionsplanung bei Sortenfertigung: ein Beitrag zur Theorie der Mehrproduktunternehmung. Gabler, Wiesbaden, 1969

– (Hrsg.): Fertigungssteuerung II: Systeme zur Fertigungssteuerung. Gabler, Wiesbaden, 1988

Allahverdi, A. / *Gupta* J.N.D. / *Aldowaisan*, T.: A review of scheduling research involving setup considerations. Omega, Vol. 27 (1999), 219-239

Anthony, R.N.: Planning and control systems: a framework for analysis. Harvard University Press, Cambridge, 1965

Ashby, W.R.: An Introduction to Cybernetics, Chapman & Hall, London, 1956

Bachem, A.: Komplexitätstheorie im Operations Research. Zeitschrift für Betriebswirtschaft, Vol. 50 (1980), 812-844

Baetge, J. / *von Lilienstern*, H.R. / *Schäfer*, H. (Hrsg.): Logistik – eine Aufgabe der Unternehmenspolitik. Duncker & Humbold, Berlin, 1987

Bahl, H.C. / *Ritzman*, L.P. / *Gupta*, J.N.D.: Determining lot sizes and resource requirements: A review. Operations Research, Vol. 35 (1987), 329-345

Baker, K.R.: Lot streaming in the two-machine flow shop with setup times. Annals of Operations Research, Vol. 57 (1995), 1-11

– Elements of sequencing and scheduling. Amos Tuck School of Business Administration, Dartmouth College, Hanover, NH 03755, 1998

Baker, K.R. / *Jia*, D.: A comparative study of lot streaming procedures. Omega, Vol. 21 (1993), 561-566

Baker, K.R. / *Peterson*, D.W.: An analytical framework for evaluating rolling schedules. Management Science, Vol. 38 (1979), 341-351

Baker, K.R. / *Pyke*, D.F.: Solution procedures for the lot-streaming problem. Decision Sciences, Vol. 21 (1990), 475-491

Bamberg, G. / *Coenenberg*, A.G.: Betriebswirtschaftliche Entscheidungslehre. 10. Aufl., Vahlen, München, 2000

Benli, Ö.S.: Lot streaming in production scheduling. Department of Engineering Management, University of Missouri-Rolla, Rolla, MO 65409-0370 (February 1997) presented at INFORMS San Diego Spring 1997 Meeting

Bierwirth, C.: Flowshop Scheduling mit parallelen Genetischen Algorithmen. DUV Gabler, Wiesbaden, 1993

Bierwirth, C. / *Mattfeld*, D.C.: Production scheduling and rescheduling with genetic algorithms. Evolutionary Computation, Vol. 7 (1999), 1-17

Billington, P.J. / *McClain*, J.O. / *Thomas*, L.J.: Mathematical programming approaches to capacity-constraint MRP systems: review, formulation and problem reduction. Management Science, Vol. 29 (1983), 1126-1141

Biskup, D. / *Feldmann*, M.: Lot streaming with variable sublots: an integer programming formulation. Erscheint in: Journal of the Operational Research Society, 2005

Bogaschewsky, R.W. / *Buscher*, U.D. / *Lindner*, G.: Simultanplanung von Fertigungslosgrößen und Transportlosgrößen in einstufigen Fertigungssystemen – Zwei statisch deterministische Ansätze bei unrestringierten Kapazitäten. Dresdner Beiträge zur Betriebswirtschaftslehre, Nr. 24, 1999

– Optimizing multi-stage production with constant lot size and varying number of unequal sized batches. Omega, Vol. 29 (2001), 183-191

Bretzke, W.-R.: Der Problembezug von Entscheidungsmodellen. Mohr, Tübingen, 1980

Brucker, P.: Scheduling Algorithms. 3. Aufl., Springer, Berlin, Heidelberg, New York, 2001

Brucker, P. / *Gladky*, A. / *Hoogeveen*, H. / *Kovalyov*, M. / *Potts*, C. / *Tautenhahn*, T. / *Velde*, S.: Scheduling a batching machine. Journal of Scheduling, Vol. 1 (1998), 31-54

Brucker, P. / *Heitmann*, S. / *Hurink*, J.: How useful are preemptive schedules? Operations Research Letters, Vol. 31 (2003), 129-136

Brüggemann, W.: Ausgewählte Probleme der Produktionsplanung. Physica, Heidelberg, 1995

Bukchin, J. / *Tzur*, M. / *Jaffe*, M.: Lot-splitting to minimize average flow time in a two-machine flowshop. IIE Transactions, Vol. 34 (2002), 953-970

Buscher, U.: Lossplitting in der Lossequenzplanung. Das Wirtschaftsstudium, Vol. 29 (2000), 197-204

Çetınkaya, F.C.: Lot streaming in a two-stage flow shop with set-up, processing and removal times separated. Journal of the Operational Research Society, Vol. 45 (1994), 1445-1455

Çetınkaya, F.C. / *Kayaligıl*, M.S.: Unit sized transfer batch scheduling with setup times. Computers and Industrial Engineering, Vol. 22 (1992), 177-183

Chandler, A. D.: Strategy and Structure. MIT Press, Chambridge, 1962

Chen, J. / *Steiner*, G.: Lot streaming with detached setups in three-machine flow shops. European Journal of Operational Research, Vol. 96 (1996), 591-611

– Approximation methods for discrete lot streaming in flow shops. Operations Research Letters, Vol. 21 (1997), 139-145

– Lot streaming with attached setups in three-machine flow shops. IIE Transactions, Vol. 30 (1998), 1075-1084

– Discrete lot streaming in two-machine flow shops. Information Systems and Operational Research, Vol. 37 (1999), 160-173

Chen, W.-H. / *Thizy*, J.-M.: Analysis of relaxations for the multi-item capacitated lot-sizing problem. Annals of Operations Research, Vol. 26 (1990), 29-72

Cheng, T.C.E. / *Sin*, C.C.S.: A state-of-the-art review of parallel-machine scheduling research. European Journal of Operational Research, Vol. 47 (1990), 271-292

Dantzig, G.B. / *Thapa*, M.N.: Linear programming, 1: Introduction. Springer, New York, 1997

– Linear programming, 2: Theory and extensions. Springer, New York, 2003

Dauzère-Pérès, S. / *Lasserre*, J.B.: Lot streaming in job-shop scheduling. Operations Research, Vol. 45 (1997), 584-595

Derstroff, M.C.: Mehrstufige Losgrößenplanung mit Kapazitätsbeschränkungen. Physica, Heidelberg, 1995

Dinkelbach, W.: Zum Problem der Produktionsplanung in Ein- und Mehrproduktunternehmen. Physica, Würzburg, 1964

– Modell – ein isomorphes Abbild der Wirklichkeit? In: Grochla, E., Szyperski, N.: Modell- und computer-gestützte Unternehmensplanung. Gabler, Wiesbaden, 1973, 151-162

– Entscheidungsmodelle. Walter de Gruyter, Berlin, New York, 1982

Dobson, G. / *Karmarkar*, U.S. / *Rummel*, J.L.: Batching to minimize flow times on one machine. Management Science, Vol. 33 (1987), 784-799

Domschke, W. / *Scholl*, A. / *Voß*, S.: Produktionsplanung. 2. Aufl., Springer, Berlin, Heidelberg. New York, 1997

Drexl, A. / *Fleischmann*, B. / *Günther*, H.-O. / *Stadtler*, H. / *Tempelmeier*, H.: Konzeptionelle Grundlagen kapazitätsorientierter PPS-Systeme. Zeitschrift für betriebswirtschaftliche Forschung, Vol. 46 (1994), 1022-1045

Drexl, A. / *Kimms*, A.: Lot sizing and scheduling - Survey and extensions. European Journal of Operational Research, Vol. 99 (1997), 221-235

– (Hrsg.): Beyond manufacturing resource planning (MPR II). Springer, Berlin, Heidelberg, New York, 1998

Drezner, Z. / *Szendrovits*, A.Z. / *Wesolowsky*, G.O.: Multi-stage production with variable lot sizes and transportation of partial lots. European Journal of Operational Research, Vol. 17 (1984), 227-237

Eppen, G.D. / *Martin*, R.K.: Solving multi-item capacitated lot-sizing problems using variable redefinition. Operations Research, Vol. 35 (1987), 832-848

Fandel, G. / *François*, P.: Betriebswirtschaftliche Ansprüche an die Planungsfunktionalität von PPS-Systemen und ihre Realisation. In: Werners, B., Gabriel, R.: Fortgeschrittene Planungsfunktionalität aktueller PPS-Systeme. Arbeitsbericht 83, IUU, Ruhr-Universität Bochum, 2000, 1-21

Fandel, G. / *François*, P. / *Gubitz*, K.-M.: PPS- und integrierte betriebliche Softwaresysteme. 2. Aufl., Springer, Berlin, Heidelberg, New York, 1997

Finch, B.J. / *Cox*, J.F.: Process-oriented production planning and control: factors that influence system design. Academy of Management Journal, Vol. 3 (1988), 123-153

Fisher, M.L.: The Lagrangian relaxation method for solving integer problems. Management Science, Vol. 27 (1981), 1-18

Fleischmann, B.: Operations-Research-Modelle und -Verfahren in der Produktionsplanung. Zeitschrift für Betriebswirtschaft, Vol. 58 (1988), 347-372

Fogarty, D.W. / *Blackstone*, J.H. / *Hoffmann*, T.R.: Production & inventory management. 2. Aufl., South-Western Publishing, Chincinnati, Ohio, 1991

François, P. / *Wolf*, J.: Auswahl nach Drehbuch. Industrielle Informationstechnik, Heft 1-2, (2001), 55-58

French, S.: Sequencing and Scheduling: An introduction to the mathematics of the job-shop. Wiley & Sons, New York, 1982

Fry, T.D. / *Cox*, J.F. / *Blackstone*, J.H.: An analysis and discussion of the optimized production technology software and its use. Production and Operations Management, Vol. 1 (1992), 229-242

Garey, M.R. / *Johnson*, D.S.: Computers and intractibility. A guide to the theory of NP-completeness. Freeman and Co., San Francisco, 1979

Glaser, H. / *Geiger*, W. / *Rohde*, V.: PPS-Produktionsplanung und -steuerung: Grundlagen – Konzepte – Anwendungen. 2. Aufl., Gabler, Wiesbaden, 1992

Glass, C.A. / *Gupta*, J.N.D. / *Potts*, C.N.: Lot streaming in three-stage production processes. European Journal of Operational Research, Vol. 75 (1994), 378-394

Glass, C.A. / *Potts*, C.N.: Structural properties of lot streaming in a flow shop. Mathematics of Operations Research, Vol. 23 (1998), 624-639

Goldratt, E.M. / *Fox*, R.: The race. North River Press, New York, 1986

Goyal, S.K.: Note on Manufacturing cycle time determination for a multi-stage economic production quantity model. Management Science, Vol. 23 (1976), 332-333

– Economic batch quantity in a multi-stage production system. International Journal of Production Research, Vol. 16 (1977), 267-273

– Determination of optimum production quantity for a two-stage production system. Operational Research Quarterly, Vol. 28 (1977b), 865-870

Goyal, S.K. / *Szendrovits*, A.Z.: A constant lot size model with equal and unequal sized batch shipments between production stages. Engineering Costs and Production Economics, Vol. 10 (1986), 203-210

Graves, S.C.: A review of production scheduling. Operations Research, Vol. 29 (1981), 646-675

Graves, S.C. / *Kostreva*, M.: Overlapping operations in material requirements planning. Journal of Operational Management, Vol. 6 (1987), 283-294

Gronau, N. / *Ibelings*, I.: Marktrecherche: Steuerungsstrategien in PPS-Systemen. PPS Management, 18 (2002), 46-58

Grünert, T.: Multi-level sequence-dependent dynamic lotsizing and scheduling. Shaker, Aachen, 1998

Günther, H.-O. / *Tempelmeier*, H.: Produktion und Logistik. 4. Aufl., Springer, Berlin, Heidelberg, New York, 2000

– Produktion und Logistik. 5. Aufl., Springer, Berlin, Heidelberg, New York, 2003

Haase, K.: Lotsizing and scheduling for production planning. Springer, Berlin, Heidelberg, New York, 1994

Hackman, S.T. / *Leachman*, R.C.: A general framework for modelling production. Management Science, Vol. 35 (1989), 478-495

Haehling von Lanzenauer, C.: A production scheduling model by bivalent linear programming. Management Science, Vol. 17 (1970), 105-111

Hall, N.G. / *Laporte*, G. / *Selvarajah*, E. / *Sriskandarajah*, C.: Scheduling and lot streaming in flow shops with no-wait in process. Journal of Scheduling, Vol. 6 (2003), 339-354

Hall, N.G. / *Sriskandarajah*, C.: A survey of machine scheduling problems with blocking and no-wait in process. Operations Research, Vol. 44 (1996), 510-525

Hancock, T.: Effects of lot-splitting under various routing strategies. International Journal of Operations & Production Management, Vol. 11 (1991), 68-75

Hax, A.C. / *Meal*, H.C.: Hierarchical integration of production planning and scheduling. Studies in Management Sciences, Vol. 1: Logistics, 1975, 53-69

Hayes, R.H. / *Wheelwright*, S.G.: Link manufacturing process and product life cycles. Harvard Business Review, Vol. 57 (1979), 133-140

Helber, K.: Kapazitätsorientierte Losgrößenplanung in PPS-Systemen. Wissenschaft und Forschung, Stuttgart, 1994

Inderfurth, K. / *Jensen*, T.: Planungsnervosität im Rahmen der Produktionsplanung. Zeitschrift für Betriebswirtschaft, Vol. 67 (1997), 817-843

Jacobs, F.R.: The OPT scheduling system: A review of a new production scheduling system. Production and Inventory Management, Vol. 24 (1983), 47-51

– OPT uncovered: many production planning and scheduling concepts can be applied with or without the software. Industrial Engineering, Vol. 16 (1984), 32-41

Jacobs, F.R. / *Bragg*, D.J.: Repetitive lots: flow-time reductions through sequencing and dynamic batch sizing. Decision Sciences, Vol. 19 (1988), 281-294

Jahnke, H.: Produktion bei Unsicherheit. Physica, Heidelberg, 1995

– Losgrößentheorie und betriebliche Produktionsplanung. In: *Priddat*, B.P., *Vilks*, A. (Hrsg.): Wirtschaftswissenschaft und Wirtschaftlichkeit, Metropolis, Marburg, 1998, 95-117

Jahnke, H. / *Biskup*, D.: Planung und Steuerung der Produktion. moderne industrie, Landsberg, Lech, 1999

Johnson, S.M.: Optimal two- and three-stage production schedules with set-up times included. Naval Research Logistics Quarterly, Vol. 1 (1954), 61-68

Jordan, C.: Batching and scheduling – models and methods for several problem classes. Lecture Notes in Economics and Mathematical Systems. Springer, Berlin, Heidelberg, New York, 1996

Kalir, A.A.: Optimal and heuristic solutions for the single and multiple batch flow shop lot streaming problems with equal sublots. Dissertation, State University, Virginia, 1999

Kalir, A.A. / *Sarin*, S.C.: Evaluation of the potential benefits of lot streaming in flow-shop systems. International Journal of Production Economics, Vol. 66 (2000), 131-142

– A near-optimal heuristic for the sequencing problem in multiple-batch flow shops with small equal sublots. Omega, Vol. 29 (2001), 577-584

– Optimal solutions for the single batch, flow shop, lot streaming problem with equal sublots. Decision Sciences, Vol. 32 (2001b), 387-397

– Constructing near optimal schedules for the flow-shop lot streaming problem with sublot-attached setups. Journal of Combinatorial Optimization, Vol. 7 (2003), 23-44

Kher, H.V. / *Malhotra*, M.K. / *Steele*, D.C.: The effect of push and pull lot splitting approaches on lot traceability and material handling costs in stochastic flow shop environments. International Journal of Production Research, Vol. 38 (2000), 141-160

Kilger, W.: Optimale Produktions- und Absatzplanung. Westdeutscher Verlag, Opladen, 1973

Kimms, A.: Multi-level lot sizing and scheduling: methods for capacitated, dynamic, and deterministic models. Physica, Heidelberg, 1997

Kistner, K.-P.: Koordinationsmechanismen in der hierarchischen Planung. Zeitschrift für Betriebswirtschaft, Vol. 62 (1992), 1125-1146

– Die Substitution von Umlaufvermögen durch Anlagevermögen im Rahmen der Produktion auf Abruf. OR Spektrum, Vol. 16 (1994), 125-134

– Optimierungsmethoden. 3. Aufl., Physica, Heidelberg, 2003

Kistner, K.-P. / *Steven*, M.: Neuere Entwicklungen in Produktionsplanung und Fertigungstechnik. Wirtschaftswissenschaftliches Studium, Vol. 20 (1991), 11-17

– Produktionsplanung. 3. Aufl., Physica, Heidelberg, 2001

– Betriebswirtschaftslehre im Grundstudium 1. 4. Aufl., Physica, Heidelberg, 2002

Kistner, K.-P. / *Switalski*, M.: Hierarchische Produktionsplanung. Zeitschrift für Betriebswirtschaft, Vol. 59 (1989), 477-503

Knolmayer, G.: Materialflussorientierung statt Materialbestandsoptimierung: Ein Paradigmawechsel in der Theorie des Produktions-Managements? In: *Baetge*, J., *von Lilienstern*, H.R., *Schäfer*, H. (Hrsg.): Logistik – eine Aufgabe der Unternehmenspolitik. Duncker & Humbold, Berlin, 1987, 53-77

– Advanced Planning and Scheduling Systems: Optimierungsmethoden als Entscheidungskriterium für die Beschaffung von Software-Paketen? In: *Wagner*, U., (Hrsg.): Zum Erkenntnisstand der Betriebswirtschaftlehre am Beginn des 21. Jahrhunderts. Duncker & Humblot, Berlin, 2001, 135-155

Kropp, D.H. / *Smunt*, T.L.: Optimal and heuristic models for lot splitting in a flow shop. Decisions Science, Vol. 21 (1990), 691-709

Kuhn, T.S.: The Structure of Scientific Revolutions. 2. Aufl., The University Press, Chicago-London, 1970

Kuik, R. / *Salomon*, M. / *Van Wassenhove*, L.N.: Batching decisions: structure and models. European Journal of Operational Research, Vol. 75 (1994), 243-263

Kulonda, D.J.: Overlapping operations – a step toward just-in-time production. Readings in Zero-Inventory, APICS 27th Annual International Conference (1984), 81-84

Kumar, S. / *Bagchi*, T.P. / *Sriskandarajah*, C.: Lot streaming and scheduling heuristics for m-machine no-wait flow shops. Computers and Industrial Engineering, Vol. 38 (2000), 149-172

Kurbel, K.: Produktionsplanung und -steuerung. 5. Aufl., Oldenburg, München, Wien, 2003

Lee, C-Y. / *Lei*, L. / *Pinedo*, M.: Current trends in deterministic scheduling. Annals of Operations Research, Vol. 70 (1997), 1-41

Lee, C.-Y. / *Uszoy*, R. / *Martin-Vega*, L.A.: Efficient algorithms for scheduling semiconductor burn. Operations Research, Vol. 40 (1992), 764-774

Leff, G.: William of Ockham: the metamorphosis of scholastic discourse. Manchester Univ. Press, Manchester, 1975

Leisten, R.: Iterative Aggregation und mehrstufige Entscheidungsmodelle. Physica, Heidelberg, 1996

Little, J.D.C.: Models and managers: the concept of a decision calculus. Management Science, Vol. 17 (1970), B-466-B-485

Liu, S.C.: A heuristic method for discrete lot streaming with variable sublots in a flow shop. International Journal of Advanced Manufacturing Technology , Vol. 22 (2003), 662-668

Lockwood, W.T. / *Mahmoodi*, F. / *Ruben*, R.A. / *Mosier*, C.T.: Scheduling unbalanced cellular manufacturing systems with lot splitting. International Journal of Production Research, Vol. 38 (2000), 951-965

Luptacik, M.: A note on the "more-for-less" paradoxon. In: *Kischka*, P. / *Leopold-Wildburger*, U. / *Möhring*, R.H. / *Radermacher*, F.-J. (Hrsg.): Models, methods and decision support for management, Physica, Heidelberg, 2001, 103-109

Maes, J. / *McClain*, J.-O. / *Van Wassenhove*, L.N.: Multilevel capacitated lotsizing complexity and LP-based heuristics. European Journal of Operational Research, Vol. 53 (1991), 131-148

Malinvaud, E.: Statistical methods of econometrics. 3. Aufl., North-Holland Publ., Amsterdam, 1980

Meyr, H.: Simultaneous lotsizing and scheduling by combining local search with dual reoptimization. European Journal of Operational Research, Vol. 120 (2000), 311-326

Missbauer, H.: Bestandsregelung als Basis für eine Neugestaltung von PPS-Systemen. Physica, Heidelberg, 1998

Moily, J.P.: Optimal and heuristic procedures for component lot-splitting in multi-stage manufacturing systems. Management Science, Vol. 32 (1986), 113-125

Monma, C.L. / *Rinnooy Kan*, A.H.G.: A concise survey of efficiently solvable special cases of the permutation flow-shop problem. RAIRO Recherche Opérationelle, Vol. 17 (1983), 105-119

Morey, R.: Estimating service level impacts from changes in cycle count, buffer stock, or corrective action. Journal of Operations Management, Vol. 4 (1985), 411-418

Nahmias, S.: Production and Operations Analysis. 3th Edition, Irwin, Homewood Illinois, 1997

Neumann, K. / *Morlock*, M.: Operations Research. Hanser, Wien, 1993

Ovacik, I.M. / *Uszoy*, R.: Decomposition methods for complex factory scheduling problems. Kluwer Academic Publishers, Boston, 1997

Petersen, L.: Kapazitätsorientierte Auftragsplanung bei Serienfertigung. Gabler, Wiesbaden, 1998

Petroff, J.N. / *Hill*, A.V.: A framework for the design of lot-tracing systems for the 1990‘s. Production and Inventory Management Journal, Vol. 32 (1991), 55-61

Popper, K.R.: The logic of scientific discovery. University of Toronto, 1959

Potts, C.N. / *Baker*, K.R.: Flow shop scheduling with lot streaming. Operations Research Letters, Vol. 8 (1989), 297-303

Potts, C.N. / *Kovalyov*, M.Y.: Scheduling with batching: a review. European Journal of Operational Research, Vol. 120 (2000), 228-249

Potts, C.N. / *van Wassenhove*, L.N.: Integrating scheduling with batching and lot-sizing: a review of algorithms and complexity. Journal of Operational Research Society, Vol. 43 (1992), 395-406

Proud, J.F.: Master scheduling: a practical guide to competitive manufacturing. 2. Aufl., Wiley & Sons, New York, 1999

Ramasesh, R.V. / *Fu*, H. / *Fong*, D.K.H. / *Hayya*, J.C.: Lot streaming in multistage production systems. International Journal of Production Economics, Vol. 66 (2000), 199-211

Reiter, S.: A system for managing job-shop production. The Journal of Business, Vol. 39 (1966), 371-393

Salomon, M.: Deterministic lot sizing models for production planning. Springer, Berlin, Heidelberg, New York, 1991

Schneeweiß, C.: Planung 1 – Systemanalytische und entscheidungstheoretische Grundlagen. Springer, Berlin, Heidelberg, 1991

– Planung 2 – Konzepte der Prozeß- und Modellgestaltung. Springer, Berlin, Heidelberg, 1992

– Einführung in die Produktionswirtschaft. 7. Aufl., Springer, Berlin, Heidelberg, 1999

– Hierarchies in distributed decision making. Springer, Berlin, Heidelberg, 1999b

Schwarz, L.B. / *Schrage*, L.: Optimal and system myopic policies for multi-echelon production/ inventory assembly systems. Management Science, Vol. 21 (1975), 1285-1294

Şen, A. / *Benli*, Ö.S.: Lot streaming in open shops. Operations Research Letters, Vol. 23 (1999), 135-142

Şen, A. / *Topaloğlu*, E., *Benli*, Ö.S.: Optimal streaming of a single job in a two-stage flow shop. European Journal of Operational Research, Vol. 110 (1998), 42-62

Silver, E.A.: Changing the givens in modelling inventory problems: the example of just-in-time systems. International Journal of Production Economics, Vol. 26 (1992), 347-351

Silver, E.A. / *Pyke*, D.F. / *Peterson*, R.: Inventory management and production planning and scheduling. 3. Aufl., Wiley & Sons, New York, 1998

Smunt, T.L. / *Buss*, A.H. / *Kropp*, D.H.: Lot splitting in stochastic flow shop and job shop environments. Decision Science, Vol. 27 (1996), 215-237

Spencer, M.S. / *Cox*, J.F.: An analysis of the product-process matrix and repetitive manufacturing. International Journal of Production Research, Vol. 33 (1995), 1275-1294

Sriskandarajah, C. / *Wagneur*, E.: Lot streaming and scheduling multiple products in two-machine no-wait flowshops. IIE Transactions, Vol. 31 (1999), 695-707

Stadtler, H.: Hierarchische Produktionsplanung bei losweiser Fertigung. Physica, Heidelberg, 1988

– Reformulation of the shortest route model for dynamic multi-item multi-level capacitated lotsizing. OR Spektrum, Vol. 19 (1997), 87-96

Steele, D.C.: A structure for lot-tracing design. Production and Inventory Management Journal, Vol. 36 (1995), 53-59

Steven, M.: Hierarchische Produktionsplanung. 2. Aufl., Physica, Heidelberg, 1994

Sule, D.R. / *Huang*, K.Y.: Sequencing on two and three machines with setup, processing and removal times separated. International Journal of Production Research, Vol. 21 (1983), 723-732

Switalski, M.: Hierarchische Produktionsplanung und Aggregation. Zeitschrift für Betriebswirtschaft, Vol. 58 (1988), 381-196

Szendrovits, A.Z.: Manufacturing cycle time determination for a multi-stage economic production quantity model. Management Science, Vol. 22 (1975), 298-308

– On the optimality of sub-batch sizes for a multi-stage EPQ model – a rejoinder. Management Science, Vol. 23 (1976), 334-338

– A comment on determination of optimum production quantity for a two-stage production system. Journal of the Operational Research Society, Vol. 29 (1978), 1017-1020

– An inventory model for interrupted multi-stage production. International Journal of Production Research, Vol. 25 (1987), 129-143

Szendrovits, A.Z / *Drezner*, Z.: Optimizing multi-stage production with constant lot size and varying numbers of batches. Omega, Vol. 8 (1980), 623-629

Szendrovits, A.Z. / *Golden*, M.: The effect of two classes of process organisations on multi-stage production/inventory systems. INFOR, Vol 22 (1984), 40-55

Tempelmeier, H.: Material-Logistik. 3. Aufl., Springer, Berlin, Heidelberg, New York, 1995

– Material-Logistik. 4. Aufl., Springer, Berlin, Heidelberg, New York, 1999

– Material-Logistik, 5. Aufl., Springer, Berlin, Heidelberg, New York, 2003

Tersine, R.J.: Principles of inventory and materials management. 4. Aufl., Prentice Hall, Englewood Cliffs, 1994

Topaloğlu, E. / *Şen*, A. / *Benli*, Ö.S.: Optimal streaming of a single job in an *m* stage flow shop with two sublots. Department of Industrial Engineering, Bilkent University, TR-06533 Ankara, (Revised November 1994)

Trietsch, D.: Polynomial transfer lot sizing techniques for batch processing on consecutive machines. Technical Report NPS-54-89-011. Naval Postgraduate School, Monterey, Califormia, 1989

Trietsch, D. / *Baker*, K.R.: Basic techniques for lot streaming. Operations Research, Vol. 41 (1993), 1065-1076

Truscott, W.G.: Scheduling production activities in multi-stage batch manufacturing systems. International Journal of Production Research, Vol. 23 (1985), 315-328

– Production scheduling with capacity-constrained transportation activities. Journal of Operations Management, Vol. 6 (1986), 333-348

Umble, M.M. / *Srikanth*, M.L.: Synchronous management: Profit-based manufacturing for the 21st Century - Volume II - Implementation issues and case studies. The Spectrum Publishing Company, Connecticut, Guilford, 1997

– Synchronous manufacturing: principles for world class excellence. APICS: South Western Series in Production and Operations Management, South Western Publishing Co., Cincinnati, Ohio, 1990

Uzsoy, R.: Scheduling a single batch processing machine with non-identical job sizes. International Journal of Production Research, Vol. 32 (1994), 1615-1635

Vickson, R.G.: Optimal lot streaming for multiple products in a two-machine flow shop. European Journal of Operational Research, Vol. 85 (1995), 556-575

Vickson, R.G. / *Alfredsson*, B.E.: Two- and three-machine flow shop scheduling problems with equal sized transfer batches. International Journal of Production Research, Vol. 30 (1992), 1551-1574

Vieira, G.E. / *Herrmann*, J.W. / *Lin*, E.: Rescheduling manufacturing systems: a framework of strategies, policies, and methods. Journal of Scheduling, Vol. 6 (2003), 35-58.

Vollmann, T.E. / *Berry*, W.L. / *Whybark*, D.C.: Manufacturing planning and control systems. 3. Aufl., McGraw-Hill, Homewood, 1992

Wagner, B.J. / *Ragatz*, G.L.: The impact of lot splitting on due date performance. Journal of Operations Management, Vol. 12 (1994), 13-25

Wagner, H.M. / *Whitin*, T.M.: Dynamic version of the economic lot size model. Management Science, Vol. 5 (1958), 89-96

Waters, C.D.J.: Inventory control and management. Wiley & Sons, Chichester, 1992

Webster, S. / *Baker*, K.R.: Scheduling groups of jobs on a single machine. Operations Research, Vol. 43 (1995), 692-704

Wiendahl, H.-P.: Belastungsorientierte Fertigungssteuerung. Hanser, München, Wien, 1987

– Load-oriented manufacturing control. Springer, Berlin, Heidelberg, 1995

Wildemann, H.: Produktionssteuerung nach Kanban-Prinzipien. In: Adam, D. (Hrsg.): Fertigungssteuerung II: Systeme zur Fertigungssteuerung. Gabler, Wiesbaden, 1988, 33-50

Williams, E.F. / *Tüfekçi*, S. / *Akansel*, M.: O(m^2) algorithms for the two and three sublot lot streaming Problem. Production and Operations Management, Vol. 6 (1997), 74-96

Woodruff, D.L. / *Spearman*, M.L.: Sequencing and batching for two classes of jobs with deadlines and setup times. Production and Operations Management, Vol. 1 (1992), 87-102

Yoon, S.-H. / *Ventura*, J.A.: Minimizing the mean weighted absolute deviation from due dates in lot-streaming flow shop scheduling. Computers & Operations Research, Vol. 29 (2002), 1301-1315

Yoshida, T.K. / *Hitomi*, K.: Optimal two-stage production scheduling with setup times separated. AIIE Transactions, Vol. 11 (1979), 261-263

Zäpfel, G.: Produktionswirtschaft – Operatives Produktions-Management. Walter de Gruyter, Berlin, Heidelberg, 1982

– Strategisches Produktions-Management. Walter de Gruyter, Berlin, Heidelberg, 1989

– Taktisches Produktions-Management. Walter de Gruyter, Berlin, New York, 1989b

– Grundzüge des Produktions- und Logistikmanagement. Walter de Gruyter, Berlin, New York, 1996

– Auftragsgetriebene Produktion zur Bewältigung der Nachfrageungewißheit. Zeitschrift für Betriebswirtschaft, Vol. 66 (1996b), 861-877

– Taktisches Produktions-Management. 2. Aufl., Oldenburg, München, Wien, 2000

– Strategisches Produktions-Management. 2. Aufl., Oldenburg, München, Wien, 2000b

– Grundzüge des Produktions- und Logistikmanagement. 2. Aufl., Oldenburg, München, Wien, 2001

Zhang, W. / *Liu*, J. / *Linn*, R.J.: Model and heuristics for lot streaming of one job in M-1 hybrid Flowshops. International Journal of Operations and Quantitative Management, Vol. 9 (2003), 1-16

Zschocke, D.: Modellbildung in der Ökonomie – Modell – Information – Sprache. Vahlen, München, 1995

Stichwortverzeichnis