

Operations Research/Computer Science Interfaces Series

Volume 52

Series Editors:

Ramesh Sharda

Oklahoma State University, Stillwater, Oklahoma, USA

Stefan Voß

University of Hamburg, Hamburg, Germany

For further volumes:

<http://www.springer.com/series/6375>

Lars Mönch • John W. Fowler • Scott J. Mason

Production Planning and Control for Semiconductor Wafer Fabrication Facilities

Modeling, Analysis, and Systems



Springer

Lars Mönch
Department of Mathematics
and Computer Science
University of Hagen
Universitätsstraße 1
Hagen, Germany

John W. Fowler
W.P. Carey School of Business
Department of Supply Chain
Management
Arizona State University
Tempe, AZ, USA

Scott J. Mason
Department of Industrial Engineering
Clemson University
Freeman Hall 124
Clemson, SC, USA

ISSN 1387-666X

ISBN 978-1-4614-4471-8

ISBN 978-1-4614-4472-5 (ebook)

DOI 10.1007/978-1-4614-4472-5

Springer New York Heidelberg Dordrecht London

Library of Congress Control Number: 2012944171

© Springer Science+Business Media New York 2013

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

Semiconductor manufacturing is one of the most important segments of the global manufacturing sector. Today semiconductor wafer fabrication facilities, for short called wafer fabs in the rest of this monograph, can be found in the USA, Europe, and Asia. Starting in the mid-1980s, the number of people in academia that deal with modeling and analysis of wafer fab operations has constantly increased. While in the beginning the number of academics working on these problems was quite small, today there are very active research groups around the world. A growing number of academics have contributed to the literature related to modeling and analysis of wafer fabs in such different areas as simulation modeling, dispatching and shop-floor scheduling, queueing models, production planning models, supply network planning models, and design and implementation of information systems for decision support. Furthermore, a wide range of other engineering models to support yield and quality improvement have been developed (cf. Chien et al. [49]). The vast academic interest in modeling and analysis of wafer fabs is caused by the fact that wafer fabs are one of the most complex and challenging industrial environments in use today.

The number of scholarly publications in this area has also increased significantly over the years. However, there are only a few survey papers that attempt to give a complete picture of various aspects of modeling and analysis of wafer fabs. The most popular among these papers are those by Uzsoy et al. [306, 307]. Except for the monographs by Atherton and Atherton [14] and Ovacik and Uzsoy [223], there are no further books in this area. The book [14] discusses modeling and analysis issues only briefly and from a different point of view. The second related book deals mainly with certain decomposition strategies based on disjunctive graphs and the shifting bottleneck heuristic for scheduling the back-end stage of semiconductor manufacturing.

In this monograph, we are interested in covering a broader area, and we attempt to take recent research trends into account. To our best knowledge, there is no book on modeling and analysis in semiconductor manufacturing that simultaneously considers production planning, production control, and

the related information systems. In this book, after presenting basic concepts in the semiconductor manufacturing process and in basic modeling and analysis tools, we introduce production control schemes that are based on dispatching rules as they are predominately used in practice. Next, we discuss recent scheduling approaches. We continue with a description of order release strategies for wafer fabs. We then introduce different production planning approaches with a focus on capacity planning. In the second to last chapter, we present research related to the important field of automated material handling systems. Finally, we describe various aspects of decision support provided by manufacturing execution systems and advanced planning systems.

Based on our experience and research interests, we mainly suggest heuristics throughout the book. However, when practical, we also discuss methods that lead to optimal solutions. It is an important feature of this book that we consider discrete-event simulation in different situations as a modeling and analysis tool [89].

We have been helped by many people in the course of preparing this book. We would like to thank Cheryl Dwyer for carefully reading the entire manuscript and for providing many helpful suggestions for improvements. Ulrike Schmidt helped us by preparing parts of the figures and by checking the references. We would also like to thank Stefan Voß who strongly supported the inclusion of this book into the Springer Operations Research/Computer Science Interfaces Series.

Finally, we want to give special thanks to our friends and colleagues Oliver Rose, Stéphane Dauzère-Pérès, and Leon McGinnis. Many of the results in this monograph represent joint work with these scientists, and their insights, criticism, and support have been an important component of our research efforts. Furthermore, we also thank Reha Uzsoy, Robert C. Leachmann, Tae-Eog Lee, and Chen-Fu Chien among others for fruitful discussions that led to insights into semiconductor manufacturing and helped us to structure our knowledge on semiconductor manufacturing and finally to write this monograph.

Many people from the industry helped us with insights, datasets, and challenging problems. We especially thank Hans Ehm and Andreas Klemmt, Infineon Technologies AG; Volker Schmalfuß, X-Fab Semiconductor Foundries AG; Karl Kempf, Intel Corporation; You-In Choung, Samsung; Shekar Krishnaswamy, Applied Materials; and Detlef Pabst and Marcel Stehli, GLOBALFOUNDRIES.

Finally, we would like to thank Neil Levine and Matthew Amboy from Springer for their support and patience during our work on this monograph.

Hagen, Germany
Tempe, AZ, USA
Clemson, SC, USA

Lars Mönch
John W. Fowler
Scott J. Mason

Contents

1	Introduction	1
1.1	Motivation	1
1.2	Outline of the Book.....	3
2	Semiconductor Manufacturing Process Description	11
2.1	Semiconductor Manufacturing Overview	11
2.2	Front-End and Back-End Operations.....	13
2.2.1	Overall Framework for Manufacturing Systems	13
2.2.2	Description of the Base System	14
2.2.3	Description of the Base Process	20
2.3	Production Planning and Control Hierarchy	26
3	Modeling and Analysis Tools	29
3.1	Systems and Models	29
3.1.1	Representation of Systems by Models	30
3.1.2	Types of Models	31
3.2	Decision Methods and Descriptive Models	32
3.2.1	Optimal Approaches vs. Heuristics	32
3.2.2	Branch-and-Bound Algorithms	33
3.2.3	Mixed Integer Programming	34
3.2.4	Stochastic Programming	35
3.2.5	Dynamic Programming	37
3.2.6	Neighborhood Search Techniques and Genetic Algorithms	38
3.2.7	Queueing Theory	41
3.2.8	Discrete-Event Simulation Techniques.....	44
3.2.9	Response Surface Methodology.....	52
3.2.10	Learning Approaches	53
3.2.11	Summary of Decision Methods and Descriptive Models	54

3.3	Performance Assessment	54
3.3.1	Performance Assessment Methodology	55
3.3.2	Architecture for Simulation-Based Performance Assessment	62
4	Dispatching Approaches	65
4.1	Motivation and Taxonomy of Dispatching Rules	65
4.2	Simple Dispatching Rules	68
4.2.1	JS-Related Dispatching Rules	68
4.2.2	MS-Related Dispatching Rules	72
4.3	Composite Dispatching Rules	74
4.3.1	Critical Ratio Dispatching Rules	74
4.3.2	ATC-Type Dispatching Rules	75
4.3.3	Composite Dispatching Rules for the MS	77
4.4	Simulation Results for Assessing Dispatching Rules	78
4.5	Batching Rules	79
4.6	Look-Ahead Rules	81
4.6.1	Dynamic Batching Heuristic	81
4.6.2	Next Arrival Control Heuristic	84
4.6.3	Additional Look-Ahead Research	90
4.6.4	BATC-Type Rules	91
4.7	More Sophisticated Approaches	94
4.7.1	Rule-Based Systems	94
4.7.2	Determining Parameters of Dispatching Rules Based on Iterative Simulation	96
4.7.3	Construction of Blended Dispatching Rules	98
4.7.4	Automated Discovery of Dispatching Rules	101
5	Deterministic Scheduling Approaches	105
5.1	Motivation and Definitions	105
5.2	Simulation-Based Scheduling	108
5.3	Equipment Scheduling	110
5.3.1	Scheduling Jobs on a Single Batch Machine	110
5.3.2	Scheduling Jobs on a Single Cluster Tool	118
5.3.3	Scheduling Jobs on Parallel Machines with Sequence-Dependent Setup Times	123
5.3.4	Scheduling Jobs with Ready Times on Parallel Batch Machines	131
5.3.5	Scheduling Problems for Parallel Machines with Auxiliary Resources	138
5.3.6	Multiple Orders per Job Scheduling Problems	144
5.4	Full Factory Scheduling	149
5.4.1	Motivation and Problem Statement	149
5.4.2	Disjunctive Graph Representation for Job Shop Problems	150

- 5.4.3 Decomposition Approach 156
- 5.4.4 Subproblem Solution Procedures 161
- 5.4.5 Simulation-Based Performance Assessment 163
- 5.4.6 Distributed Shifting Bottleneck Heuristic 166
- 5.4.7 Multicriteria Approach to Solve Large-Scale Job Shop Scheduling Problems 169
- 6 Order Release Approaches 177**
 - 6.1 Push Versus Pull Approaches 177
 - 6.1.1 Push Approaches for Order Release 178
 - 6.1.2 Pull Approaches for Order Release 178
 - 6.1.3 Comparing Push Versus Pull Approaches 180
 - 6.2 Tailored Approaches for Wafer Fabs 180
 - 6.2.1 Starvation Avoidance 180
 - 6.2.2 Workload Regulation 181
 - 6.2.3 Subsequent Order Release Methods 183
 - 6.3 Interaction of Order Release and Scheduling 185
 - 6.3.1 Scheduling Approach and Order Release Schemes 185
 - 6.3.2 Experimental Setting and Computational Results 187
 - 6.3.3 Conclusions from the Interaction Study 190
 - 6.4 A Large-Scale Order Release Study 190
 - 6.4.1 Overall Situation 191
 - 6.4.2 Release Timing Case Study 191
 - 6.4.3 Release Quantity Case Study 195
 - 6.5 Optimization-Based Order Release 196
- 7 Production Planning Approaches 207**
 - 7.1 Short-Term Capacity Planning 207
 - 7.1.1 Motivation 208
 - 7.1.2 Spreadsheet-Based Approaches for Wafer Fabs 208
 - 7.1.3 Spreadsheet-Based Approaches for Back-End 212
 - 7.1.4 An Integrated Approach Using Simulation 214
 - 7.2 Master Planning 215
 - 7.3 Capacity Planning 219
 - 7.4 Enterprise-Wide Planning 222
 - 7.5 Modeling of Load-Dependent Cycle Times 231
 - 7.5.1 Cycle Time Throughput Curves 231
 - 7.5.2 Iterative Simulation 239
 - 7.5.3 Clearing Functions 241
- 8 State of the Practice and Future Needs for Production Planning and Control Systems 247**
 - 8.1 Motivation and State of the Art 247
 - 8.2 Requirements of Production Planning and Control Systems 249

- 8.3 Production Control Systems 251
 - 8.3.1 MES Core Functionality..... 251
 - 8.3.2 Dispatching Systems 253
 - 8.3.3 Scheduling Systems 255
 - 8.3.4 FABMAS: An Agent-Based Scheduling System 256
 - 8.3.5 Additional Application Systems as Part
of the Control System 262
- 8.4 Production Planning Systems 264
 - 8.4.1 ERP and APS Core Functionality 264
 - 8.4.2 Interaction with Other Systems 266
- References** 267
- Index** 285

Notation

The following symbols and notation will typically be used throughout the book. We have reused some of them for several purposes due to the limited supply of symbols from the alphabet and to be consistent with papers from the literature. Their usage should be clear from the context. We also introduce additional notation when needed throughout the book.

Notation	Explanation
ACT	Average cycle time
AL	Average lateness
AT	Average tardiness
AWT	Average weighted tardiness
aux	Auxiliary resource
B	Maximum batch size
BS	Base system
BP	Base process
CDT	Carrier delivery time
C_j	Completion time of job j
C_{\max}	Makespan
CP	Control process
CS	Control system
CT	Cycle time
CT_j	Cycle time of job j
d_i	Individual desirability function for objective i
d_j	Due date of job j
d_{jk}	Due date for process step k of job j
D	Combined desirability function
$DU[a,b]$	Discrete uniform distribution over the integer set $\{a, \dots, b\}$
ε	Small number or a random error $\sim N(0, \sigma^2)$
f	Number of incompatible job families
F	Number of FOUPs in a MOJ scheduling problem
$F(j)$	Family of job j
FF	Flow factor
FJm	Flexible job shop with m machine groups
h	Planning horizon
i	Machine index

Notation	Explanation
IS	Information system
j	Job index
Jm	Job shop with m machines
JS	Job processing system
k	Process step index
κ	Look-ahead parameter
L_j	Lateness of job j
L_{\max}	Maximum lateness
λ	(Arrival) rate
m	Number of machines
M_j	Set of machines that are possible for job j
MS	Material flow system
n	Number of jobs
n_j	Number of process steps of job j
NTJ	Number of tardy jobs
$N(\mu, \sigma^2)$	Normal distribution with mean μ and variance σ^2
O_j	Process flow of job j
O_{jk}	Operation k of job j
OS	Operational system
\bar{p}	Average processing time
p – batch	Parallel batching
p_j	Processing time of job j
p_{jk}	Processing time of process step k of job j
Pm	Parallel identical machines, where the number of machines is m
PP	Planning process
PS	Planning system
recrc	Recirculation, i.e., reentrant flow
r_j	Ready time of job j
Rm	Unrelated parallel machines, where the number of machines is m
\mathbb{R}_+	The set of non-negative real numbers
s – batch	Serial batching
s_{jk}	Setup time to process job k after job j
$s_{kl,ji}$	Setup time to process step l of job k after step i of job j
t	Current time
TC	Total completion time
T_j	Tardiness of job j
TP	Throughput
TT	Total tardiness
TWC	Total weighted completion time
TWT	Total weighted tardiness
τ_Δ	Planning interval
τ_{ah}	Additional planning horizon
$U(a, b)$	Continuous uniform distribution over the interval (a, b)
U_j	Indicator variable that is 1 if job j is tardy
Var(CT)	Variance of the cycle time
Var(L)	Variance of the lateness
w_j	Weight of job j
WNTJ	Weighted number of tardy jobs
x^+	$\max(x, 0)$
z_i	Weight for the individual desirability function d_i
\mathbb{Z}_+	The set of non-negative integers