

# 网格技术在半导体生产调度中的应用研究<sup>1)</sup>

王中杰 吴启迪 李亚捷

(同济大学控制科学与工程系 上海 200092)  
(E-mail: wang\_zhongjie@mail.tongji.edu.cn)

**摘要** 网格技术是一种新型的分布计算技术, 致力于解决复杂度很高的新应用问题. 随着全球半导体生产规模的日益扩大, 半导体生产线的优化调度问题成为学术界及工程界研究的热点. 半导体生产线具有许多特殊的特点, 诸如生产规模大、工件数量多、随机性大、加工成本高、高度的可重入性等, 这些特点决定了原有的调度策略已不能满足半导体生产线的要求. 鉴于网格技术在处理设备可扩展性和资源平衡性上的优势, 主要研究将网格技术的思想用于半导体生产线的调度中. 利用网格计算中的负载向量和失衡因子的概念, 来控制半导体生产线上各加工机器处工件块的规模以及投料规模. 通过优化算法的调度, 使得半导体生产线的各加工设备负载得到平衡, 设备的生产效率提高, 缩短加工周期, 从而达到优化生产线的目的.

**关键词** 网格技术, 半导体生产线, 优化调度  
**中图分类号** TP13

## Investigation of Grid Technology on Scheduling of Semiconductor Manufacturing Line

WANG Zhong-Jie WU Qi-Di LI Ya-Jie

(Department of Control Science and Engineering, Tongji University, Shanghai 200092)  
(E-mail: wang\_zhongjie@mail.tongji.edu.cn)

**Abstract** The grid computing, a new type of distributed computing method, is applied to scheduling of semiconductor manufacturing line due to its advantages in production scale expanding and resource balance. Using the concept of load vector and load unbalance index in the grid, the product block scale and the release are adjusted for the semiconductor manufacturing line. With the optimal scheduling policy proposed in this paper, the machine balance can be kept, the production efficiency can be increased, the cycle time can be reduced and the system can be optimized to a large extent.

**Key words** Grid technology, semiconductor manufacturing line, optimal scheduling

## 1 引言

调度问题非常具有挑战性, 半导体生产线的优化调度问题更是学术界及工程界研究的热点<sup>[1,2]</sup>. 研究半导体生产线的优化调度具有重要的经济价值和学术意义. 首先, 由于优化调度的目标主要是减少制造周期及其方差, 提高设备利用率及生产率, 因此良好的调度策略能够迅速响应市场的需求, 提高产品满足客户的能力, 缓解资本积累和缓冲存储空间, 提高产品质量. 其次, 与传统的 job-shop 与 flow-shop 不同, 半导体生产线具有许多不同的

1) 国家自然科学基金项目 (70201004, 70531020), 国家 973 项目子任务 (2002CB312200), 上海市青年科技启明星计划 (04QMX1429), 霍英东第十届教育基金项目 (104030) 资助  
Supported by National Natural Science Foundation of P. R. China (70201004, 70531020), National Project "973" of P. R. China (2002CB312200), Shanghai Young Technical Rising Star Project (04QMX1429), and The 10th Fok Ying-Tung Education Foundation Project (104030)  
收稿日期 2005-6-7 收修改稿日期 2005-11-30  
Received June 7, 2005; in revised form November 30, 2005

特点, 诸如生产规模大、工件数量多、随机性大、加工成本高、高度的可重入性等, 这些特点决定了目前的调度理论和方法尚无法有效地解决该问题. 因此, 研究半导体生产线的优化调度问题将极大地促进可重入生产系统调度领域的发展. 同时, 通讯行业的许多系统以及胶片生产系统也属于可重入生产系统, 因此研究半导体生产线的优化调度方法对这些系统将提供重要的借鉴意义.

半导体生产线的调度包括两方面内容, 即投料控制及工件调度. 每一个完整的调度策略就是一种投料控制策略与工件调度策略的组合. 投料控制用于确定物料进入生产系统的速率, 工件调度是指对竞争使用某设备的工件, 当设备出现空闲时, 按何决策从中选择下一个要加工的工件.

网格技术是一种新型的分布计算技术, 致力于解决复杂度很高的新应用问题<sup>[3~6]</sup>. 网格计算将网格上的各种资源, 比如超级计算机、大规模存储设施和各种外部设备等组织在一个统一的框架下, 从而能够以方便的方法解决各种复杂问题. 鉴于网格技术的优势, 因此研究将网格技术用于半导体生产线的调度中, 根据加工系统的负载向量和失衡因子, 调整各加工机器和系统的工件规模, 使半导体生产线的设备负载得到平衡, 生产效率得到提高, 从而达到优化系统的目的.

## 2 问题定义

一个半导体生产工件批次被分成若干个步骤, 也就是若干个子作业. 每个子作业包含一批待加工的零件, 每个子作业完成之后的作业负载信息用于调整下一步加工的作业划分, 使之适应资源性能波动, 从而使整个加工过程在运行时实现机器负载的近似平衡.

每个加工批次可分为几个不同的加工子作业, 它们被指派到不同的加工机器上去. 它们各自运行的时间构成一个负载向量, 从中可以计算出加工失衡因子. 失衡因子表示作业运行时所有机器普遍的失衡程度. 负载向量用于调整下一次子作业的划分方法, 失衡因子用于调整下一批工件进入的规模.

设一个半导体生产过程中有  $K$  道工序, 每道工序前包含的工件个数为  $Q = (q_1, q_2, q_3, \dots, q_k)$ , 在第  $i$  次调度下, 各加工机器完成各自的工件组, 则整个系统完成该次作业的时间可用向量表示  $T_i = (t_1, t_2, \dots, t_j, \dots, t_k)$ , 其中  $t_j$  为完成第  $j$  道工序所需的时间.

因为各道工序在本身加工时间上会有差异, 所以时间向量应该进行加权. 这里采用的加权方法是: 取最短的加工时间作为公约数, 其它机器的加工时间根据这台机器的公约数进行计算, 从而得出加权系数.

那么根据不同工序本身属性加权后的时间向量为

$$T'_i = (a_1 t_1, a_2 t_2, \dots, a_k t_k) \quad (1)$$

加权之后的时间向量能够更好的体现各道工序加工时间上的差异.

由此可以算出平均时间为

$$\bar{t} = \frac{1}{k} \sum_{j=1}^k a_j t_j \quad (2)$$

可设

$$L_i = \left( \frac{a_1 t_1}{\bar{t}}, \frac{a_2 t_2}{\bar{t}}, \dots, \frac{a_k t_k}{\bar{t}} \right) \quad (3)$$

记  $L_i$  为负载向量, 当  $L_i = (1, 1, \dots, 1)$  时, 表示系统中各加工机器负载平衡.  $K_{ij}$  反应了第  $j$  个步骤在第  $i$  次调度中的运行负载情况, 也即  $L_{ij} > 1$ , 负载偏重;  $L_{ij} < 1$ , 负载偏轻.

### 3 调度算法

假设第  $i$  次调度下的负载向量为  $\mathbf{L}_i = (l_1, l_2, \dots, l_k)$ , 定义

$$n_j = \begin{cases} l_j, & l_j \geq 1 \\ 1/l_j, & l_j \leq 1 \end{cases} \quad (4)$$

令  $E(n)$  为  $n_j$  的均值,  $E(n)$  为失衡因子, 记为  $\delta_i$ .

可见当  $\mathbf{L}_i = (1, 1, \dots, 1)$  时,  $\delta_i = 1$ . 那么  $\mathbf{L}_i$  中各分量之间差别越大,  $\delta_i$  越大, 所以  $\delta_i$  是对系统负载平衡程度的量化.

下一次调度子作业的划分方法为

$$q_{ij} = \frac{q_{i-1j}}{l_{i-1j}} \quad (5)$$

使得  $\mathbf{L}_i$  趋于  $(1, 1, \dots, 1)$ .

若下一次调度划分某一子作业的工件数将减少, 且小于上一步的工件数, 则可将前一步完成的工件数堆积在加工机器前的虚拟工件队列中, 否则直接接受上一步的工件数, 即

$$\begin{cases} q_{ij} = \frac{q_{i-1j}}{l_{i-1j}}, & \text{若 } l_{(i-1)j} > 1 \text{ 且 } q_{ij} < q_{ii} \\ q_{ij} = q_{ii}, & \text{若 } l_{(i-1)j} > 1 \text{ 且 } q_{ij} > q_{ii} \end{cases} \quad (6)$$

若下一次调度划分某一子作业的工件数将增加, 则直接接受上一步的工件数及虚拟队列中的工件数  $M$ , 即:

$$q_{ij} = q_{ii} + M, \text{ 若 } l_{(i-1)j} < 1 \quad (7)$$

当一次调度完成后, 系统需要在第一个加工机器群中引进一批新工件, 此时  $\delta_{i-1}$  就决定这批新工件的规模是否需要扩大

$$q_{i1} = \begin{cases} q_{i-1}, & \delta_{i-1} \geq 1 \\ \frac{q_{i-1}}{\lambda \delta_{i-1}}, & \delta_{i-1} < 1, \quad 0 < \lambda < \frac{1}{\delta_{i-1}} \end{cases} \quad (8)$$

当  $\delta_{i-1} \geq 1$  时, 认为加工系统在上次作业时负载失衡较重, 因此不宜增大规模. 反之  $\delta_{i-1} < 1$  时, 负载失衡较轻, 可适当加大任务数.

### 4 仿真研究

本文以一个具有三个加工中心、六道工序的系统为研究对象, 系统示例如图 1 所示.

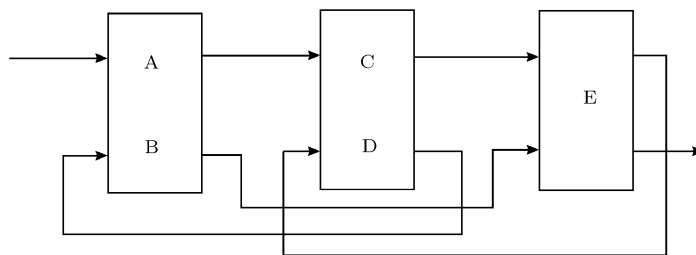


图 1 半导体生产线模型

Fig. 1 The model of a semiconductor manufacturing line

在这个模型系统中，有三个加工中心，它们一共包括五台加工机器。一号加工中心有加工机器 A 和 B，二号加工中心有加工机器 C 和 D，三号加工中心就是加工机器 E。整个生产流程分为六个步骤，也即步骤 1~6。每个加工中心可以完成其中的两步，如图 2 所示。

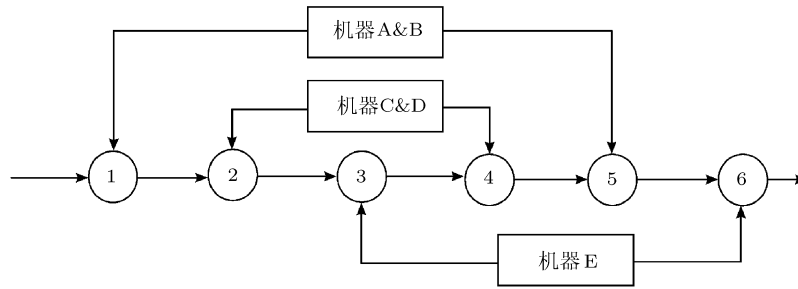


图 2 半导体加工中心的加工流程

Fig. 2 Manufacturing process of the semiconductor manufacturing line

各加工步骤的时间参数如表 1 所示。

表 1 各加工步骤的时间参数

Table 1 The time parameters for each step

	Step1	Step2	Step3	Step4	Step5	Step6
加工时间 (min)	225	30	55	255	50	10

采用 MATLAB 语言进行仿真研究。仿真研究以确定性释放策略 (DETERM)、调度规则为 LBFS(大缓冲区先输出法) 的仿真算法作为基准参考算法 (DETERM-LBFS)。

基于网格思想的改进调度算法仿真分成两步，第一步是仅调整投料控制 (GRID-RELEASE)，第二步是同时调整投料控制和工件调度 (GRID-RELEASE-DISPATCH)。

基准参考算法、投料控制改进算法以及投料控制和工件调度同时改进算法下，系统的在制品水平分别如图 3、图 4 和图 5 所示。

本文所提出的调度算法与其它几种调度算法的仿真结果见表 2 所示。

表 2 各种释放策略和调度规则下的半导体生产线仿真结果

Table 2 The Simulation result with several release policies and dispatching rules

释放策略	调度规则	生产周期 (h)	设备利用率			在制品水平	瓶颈变化
			Work1	Work2	Work3		
根据失衡因子调整 Sa(k)	LBFS	13.17	56.06	31.06	54.10	3.86	-
根据失衡因子调整 Sa(k)	改进算法	11.37	13.00	33.73	15.31	5.67	-
DETERM	LBFS	13.97	7.97	4.53	9	3.57	-
Sa(k)=0.0045	SPT	13.98	7.97	4.53	9.11	3.58	-
DETERM	LBFS	13.97	73.61	40.65	71.25	3.57	-
Sa(k)=0.45	SPT	11.97	73.61	40.65	71.05	4.12	-
CL	LBFS	43.01	99.94	45.26	92.47	12.47	Work1 Work3
	SPT	38.98	99.94	59.82	92.31	12.47	Work1 Work3
POISS	LBFS	15.98	0.6	0.46	1.41	4.03	-
Mean=0.08	SPT	15.99	0.6	0.46	1.41	4.04	-
POISS	LBFS	15.98	0.6	0.46	1.2	4.03	-
Mean=1.5	SPT	15.98	0.6	0.46	1.2	4.03	-

仿真中，Sa(k) 表示工件进入系统的速率。基准参考算法的工件释放策略是基于 DETERM，且 Sa(k)=0.0045；调度规则采用 LBFS。如表 2 所示，该算法仿真的生产周期是 13.97

小时, 一号加工中心也就是加工机器 A 和 B 的平均设备利用率为 7.97%, 二号加工中心的设备利用率是 4.53%, 三号加工中心的设备利用率为 9%.

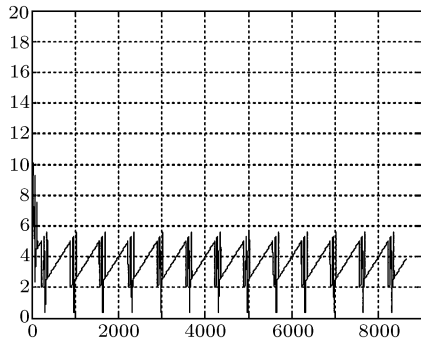


图 3 基准参考算法 (DETERM-LBFS) 下在制品水平

Fig. 3 The WIP with the benchmark algorithm (DETERM-LBFS)

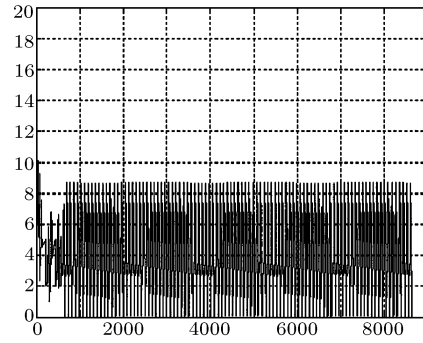


图 4 GRID-RELEASE 算法下在制品水平

Fig. 4 The WIP with GRID-RELEASE algorithm

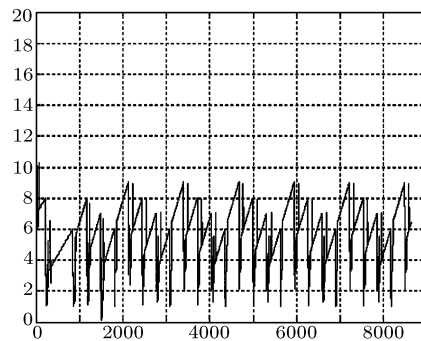


图 5 GRID-RELEASE-DISPATCH 算法下在制品水平

Fig. 5 The WIP with GRID-RELEASE-DISPATCH algorithm

根据网格负载平衡思想所得出的调度算法, 第一步是保持调度规则 LBFS 不变, 仅改变工件释放速率. 投料控制改变后仿真表明, 生产周期由原来的 13.97 小时缩短为 13.17 小时, 一号、二号、三号加工中心的设备利用率分别从 7.97%, 4.53% 和 9% 上升为 56.06%, 31.06% 和 54.10%, 设备的利用率显著上升. 在制品水平由 3.57 略微增加到 3.86, 说明生产线在制品水平有略微上升, 但变化很小.

第二步是将确定性工件释放策略变为根据失衡因子调整  $S_a(k)$ , 同时将 LBFS 的调度规则改为根据系统负载向量调整工件的划分. 从仿真结果看, 与基准算法比较, 生产周期明显缩短, 从 13.97 小时下降为 11.37 小时, 设备利用率分别由原来的 7.97%、4.53%、9% 上升为 13%、15%、33.73%、31%. 但从仿真结果也不难看出, 三个加工中心的设备利用率比第一步略微下降, 这表明生产周期的缩短和负载的均衡是以设备利用率下降为代价的. 同时在制品水平也有一定的上升, 从 3.86 上升到 5.67, 这相应表明, 生产线各性能指标之间是矛盾的.

本改进算法和基于 LBFS 或 SPT 调度规则, 释放策略为 DETERM  $S_a(k)=0.45$  的算法相比, 虽然从性能指标上看, 后者似乎更加. 但在生产过程中, 不论工况是否变化, LBFS 或 SPT 调度规则都按同一种方式进行工件调度, 不利于系统的综合优化, 而本文所提出的改进算法却能根据工况实时调整工件的释放和调度, 因此更占有优势. 和闭环调度策略 (CL) 相比, 虽然其加工中心的设备利用率相当高, 但其生产周期却长达 43.01 和 38.98 小

时. 而且由于其一号和三号加工中心的设备利用率太高, 超过了 90%, 以至于这两个加工中心成为了整个生产线的瓶颈所在. 这也是这两种算法的缺陷.

## 5 结语

本文所提出的将网格平衡思想用于半导体生产调度的改进算法还是具有一定的优势. 从仿真结果可以看出各个加工中心的设备利用率都有提高, 而且更为均衡. 虽然在制品水平略微升高, 但加工时间明显缩短, 生产的快速性得到保证. 而且跟其它调度算法, 尤其是 POISS 型输入的调度算法相比, 无论从加工时间, 设备利用率还是在制品水平, 都有较大的优势. 可见网格技术的思想应用于半导体生产调度算法是切实可行的.

## References

- 1 Wang Z J. Investigation of optimal scheduling based on hierarchical control for semiconductor manufacturing line. [Post doctoral report], Shanghai: Tongji University, 2002
- 2 Rasem Dabbas. A New Scheduling Approach Using Combined Dispatching Criteria in Semiconductor Manufacturing Systems. [Ph. D. dissertation], Arizona, USA: Arizona State University, 1999
- 3 Ian Foster, Carl Kesselman. The Grid: Blueprint for a New Computing Infrastructures, Second edition, Beijing: Mechanical Publisher, 2005
- 4 Joshy Joseph, Craig Fellenstein. Grid Computing. Beijing: Tsinghua University Press, 2005
- 5 Fujimoto N, Haqihara K. A comparison among grid scheduling algorithms for independent coarse-grained tasks. In: Proceedings of International Symposium on Applications and the Internet Workshops, Tokyo, Japan: 2004. 26~30
- 6 Wu M Y, Su W, Zhang H. Segmented min-min: A static mapping algorithm for meta-tasks on heterogeneous computing systems. In: Proceedings of 9th Heterogeneous Computing Workshop (HCW 2000), Cancun, Mexico: 2000. 375~385

**王中杰** 同济大学控制科学与工程系副教授. 研究方向为: 复杂过程的建模、优化与控制, 网络控制.

(**WANG Zhong-Jie** Associate professor in Department of Control Science and Engineering at Tongji University. His research interests include modeling, optimization and control for sophisticated process, and network control.)

**吴启迪** 教育部副部长, 同济大学控制科学与工程系教授. 研究方向为: 复杂过程的建模、优化与控制.

(**WU Qi-Di** Vice minister of Ministry of Education of P. R. China, professor in Department of Control Science and Engineering at Tongji University. Her research interests include modeling, optimization, and control for sophisticated process.)

**李亚捷** 同济大学控制科学与工程系硕士研究生. 研究方向为: 复杂过程的建模、优化与控制.

(**LI Ya-Jie** Master student in Department of Control Science and Engineering at Tongji University. His research interests include modeling, optimization and control for sophisticated process.)