

文章编号: 1001-0920(2011)03-0339-07

晶圆生产线瓶颈区并行机半在线调度方法

吴莹, 乔非, 李莉, 古攀

(同济大学电子信息工程学院, 上海 201804)

摘要: 为提高晶圆生产线瓶颈区生产率, 提出一种瓶颈区并行机半在线调度方法. 该方法基于晶圆制造的“多重入”特点, 充分利用生产历史信息、可预知信息及实时信息, 分析了瓶颈区短期内的重入流来源, 给出了瓶颈区任务集的预测方法. 针对非零初始状态的生产线, 设计了以最小化瓶颈区平均加权拖期时间为优化目标的蚁群优化算法, 定义了并行机任务集析取图, 并给出了半在线调度方法的应用算例, 为实际生产管理提供了可行的优化方法.

关键词: 半导体制造; 瓶颈; 并行机; 半在线调度; 蚁群优化算法

中图分类号: TP278

文献标识码: A

Semi-online scheduling of parallel machines in bottleneck station of wafer production line

WU Ying, QIAO Fei, LI Li, GU Pan

(School of Electronics and Information Engineering, Tongji University, Shanghai 201804, China. Correspondent: WU Ying, E-mail: 7y@tongji.edu.cn)

Abstract: To minimize the average tardiness of the jobs processed by the bottleneck station, a semi-online scheduling method is proposed. Based on the re-entrant manufacturing essence, historical information, predictable information about future arrival jobs and real-time information are fully used. Firstly, the sources of the jobs processed is analyzed by the bottleneck station, and a method is provided to establish the task set of the bottleneck station in scheduling horizon. Then the re-entrant flows on parallel tools in bottleneck area are described with disjunctive graph, and an ant colony optimization(ACO) algorithm is developed to find a satisfactory solution in a reasonable computation time. Finally, analytical experiments show the effectiveness of the proposed method.

Key words: semiconductor manufacturing; bottleneck; parallel machines; semi-online scheduling; ant colony optimization

1 引言

在晶圆制造过程中, 生产线上任意时刻都有 35%~45% 的在制品集中在以光刻为中心的加工区^[1]. 光刻区是重入流最集中的生产区域, 往往是生产线的物理瓶颈区, 制约着整个系统的生产周期、在制品水平和准时交货率等生产绩效指标^[2].

瓶颈资源调度是 NP 难题之一, 早期通过投料策略调控瓶颈区^[3-5], 但随着产品集成度提高、重入流增加, 使得制造过程复杂度剧增, 投料策略对光刻区的调控能力大幅减弱. 很多研究者在动态在线调度瓶颈资源方面进行了大量的研究, 主要有以下 3 种方法: 1)

采用启发式派工规则对排队工件赋以优先级^[1,6-8]. 2) 实时调度时动态选择派工规则^[9-12]. 3) 结合半导体生产线固有“多重入”特点和“DBR (drum-buffer-rop)”思想产生“层”控制方法. 将相邻两次光刻之间的操作看作“层”, 在各层操作中, 工序重入现象很少, 可利用这一特点进行调度, 如文献 [13-14]. 文献 [15-16] 也提出了针对瓶颈资源的调度方法, 计算在制品位于瓶颈设备前的时间缓冲区, 以此为基础确定所有订单工件的投料时间. 这种情况比较适用于较为理想的零初始状态. 而实际生产中, 在制品分布在生产线上的不同位置, 到达瓶颈的时间缓冲差异很大, 且生产投料

收稿日期: 2009-12-14; 修回日期: 2010-07-12.

基金项目: 国家自然科学基金项目(50905129, 70531020); 教育部新世纪优秀人才支持计划项目(NCET-07-0622); 教育部博士点基金新教师课题(20070247007); 上海科技创新行动计划项目(09DZ1120602, 10DZ1120102); 同济大学优秀青年教师培养计划项目.

作者简介: 吴莹(1979-), 女, 博士生, 从事复杂制造系统调度的研究; 乔非(1967-), 女, 教授, 博士生导师, 从事复杂制造系统建模与调度、系统集成与优化等研究.

会根据实际情况发生变化。

上述动态在线调度方法对部分即将到达的工件信息以及一些全局信息的利用不够充分。最大程度地利用这些信息可以获得更大的调度优化空间,因而可以产生性能更好的调度方法^[17-18]。更重要的是,这种信息利用模式非常符合一般生产习惯和实际情况。基于此,本文提出了晶圆生产线瓶颈区半在线调度方法,以提高瓶颈资源调度的质量。

2 基本思路

生产线上交货期紧迫的加工任务通常为拖期来源,半在线调度的可能性在于生产线上存在一些交货期不紧迫的加工任务,可在满足生产约束条件下,将交货期紧迫工件的光刻工序提前进行。该短期优化方法以平均拖期为优化目标,在考虑交货期的同时,综合考量等待时长、设备负载、每层光刻任务的加工时长、工件最早到达时间和最晚完工时间等因素。

本文将调度时刻晶圆生产线可用信息划分为 3 类: 生产线历史信息(生产线过去一段时间内的一些统计分析数据); 可预知信息(全局静态信息); 实时信息(调度时生产线的实时状况数据)。这些信息在动态调度方法中很难考虑全面,但其对调度性能的提高有很大影响。根据上述信息可以对一段时间内工件的工序、生产线节奏作出合理的预测和安排。该方法的实现过程如图 1 所示。

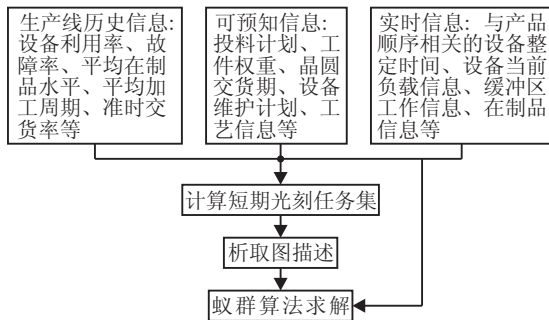


图 1 晶圆生产线瓶颈区并行机半在线调度方法

3 问题描述

3.1 问题描述及符号约定

1) 生产系统运转的某任意时刻为决策时刻 t , 即非零初始状态。

2) 预测调度期为 T_s 天。

3) 不同工件的工序之间没有顺序约束。某一光刻工序一旦开始加工便没有中断。

4) 每台设备在同一时刻只能加工一道工序。每道工序所需时间已知, 若该工序可由互替设备加工, 则取平均值。

5) 调度的目标是确定 T_s 期间, 每天每台瓶颈设备上工件的加工顺序及开工时间, 使得调度期内瓶颈

加工区工件的平均拖期值较小。

6) 研究瓶颈问题突出的生产线, 设非瓶颈设备的加工能力远大于瓶颈设备的加工能力。基于此, 松弛非瓶颈设备的能力约束, 非瓶颈设备上的工件到达后无需等待即可加工, 即忽略工序在非瓶颈设备上的等待时间。非瓶颈设备采用 FIFO (first in first out) 规则调度工件。

7) 根据投料计划, T_s 天还将投入新工件数为 $NewR$; 生产线上在制品数记为 WIP ; T_s 内经过瓶颈区的工件数记为 n : $n \leq NewR + WIP$ 。

8) 设工件 i 的工艺流程包含 N_i 步光刻工序。调度期内, 工件 i 第 j 次光刻表示为 h_{ij} , $i=1, 2, \dots, n, j=0, 1, \dots, H_i, n \leq H_i \leq N_i, H_i$ 为 T_s 天内工件 i 的光刻次数。

9) 光刻加工区有 m 台并行互替设备, 共用一个缓冲区。

10) 在调度期 T_s 天, 光刻区将要完成的任务集为新投料工件及线上所有在制品工件的 T_s 天内光刻任务, 记为 $Task_{T_s} = \{h_{11}, h_{12}, \dots, h_{1H_1}, h_{21}, h_{22}, \dots, h_{2H_2}, \dots, h_{i1}, h_{i2}, \dots, h_{iH_i}, h_{n1}, h_{n2}, \dots, h_{nH_n}\}$ 。

11) 半在线调度方法关注 $t + T_s$ 内生产线上所有工件的光刻任务在并行瓶颈设备上的安排, 其来源有如下 4 部分: 每个新投料工件的光刻步数; t 时刻在瓶颈缓冲区排队工件的光刻步数及后续光刻步数; t 时刻在瓶颈设备上加工工件的后续光刻步数; t 时刻在其他设备上加工及在其他设备缓冲区中等待加工的工件的光刻步数。

12) “光刻层”的描述性定义: 假设某一产品, 根据工艺流程完成生产需 N 次光刻, 从投料至第 1 次光刻开始前, 编为第 0 层光刻; 将每次光刻工序开始至下一次光刻工序开始之间的工序视为一光刻层, 依序编号; 最后一次光刻至最后一道工序完成, 编为第 N 层光刻。

13) 集合 $Task_{T_s}$ 中的每个任务到达光刻区的时间记为 $A_{h_{ij}}$, 权重记为 $w_{h_{ij}}$, 光刻工序交货期记为 $D_{h_{ij}}$, 光刻工序完成时间记为 $C_{h_{ij}}$ 。

14) 工件 i 在预测调度期 T_s 内首次光刻开始的时间为 $A_{h_{i1}}$; 工件 i 的权重记为 w_{h_i} : $w_{h_i} = w_{h_{ij}}$; 工件 i 在调度期内末次光刻的工序交货期为 D_{h_i} : $D_{h_i} = \max(D_{h_{ij}})$, 末次光刻完成时间记为 C_{h_i} : $C_{h_i} = \max(C_{h_{ij}})$ 。

15) 工件 i 第 j 次光刻的工序交货期计算。工件 i 完成加工共需经过 P 道工序, 总的净加工时间为 PT_i , 从第 1 步到 h_{ij} 光刻的净加工时间为 pt_{ij} , 工件 i 的交货期为 D_i , 则 h_{ij} 的工序交货期为

$$D_{h_{ij}} = D_i(pt_{ij}/PT_i).$$

16) 工件 i 在调度期 T_s 内的拖期时间用末次光刻工序交货期拖期时间表示, 记为

$$Td_{h_i} = \max\{(C_{h_i} - D_{h_i}), 0\}.$$

综上, 瓶颈区调度问题可表达为

$$m|A_{h_{ij}}| \min \left(\left(\sum_{i=1}^n w_{h_i} Td_{h_i} \right) / n \right).$$

3.2 调度期内瓶颈区工序集预测方法

根据3.1节11)所述, Task_{T_s}任务集由4部分组成, 下面给出计算方法:

1) 根据投料计划, 在调度期T_s天新投料NewR个工件, 工件*i*重入光刻设备次数为

$$H_i = \min \left\{ \left\lceil \frac{T_s}{(D_i - R_i)/N_i} \text{co}_{rh} \right\rceil, N_i \right\}. \quad (1)$$

其中: R_{*i*}为工件*i*的计划投料时间, co_{rh}为任务集调整系数. 则光刻任务集为 Task_{T_s,NewR} = {*h_{ij}* | *i* = LotID, *j* = 1, 2, ..., H_{*i*}}. 工件第1次到达光刻设备的时间为 A_{*h_{i1}*} = R_{*i*} + p_{*t_{i0}*}co_{rt}. R_{*i*}为该工件投料时刻, p_{*t_{i0}*}为第0层光刻的净加工时长, 均可从数据库中得到; co_{rt}为微调系数, 依生产线载荷而定.

2) 决策时刻在瓶颈缓冲区的工件数为 WIP_{bb}, 工件*i*在调度期T_s天重入光刻设备次数为

$$H_i = \begin{cases} \min \left\{ \left\lceil \frac{T_s}{\frac{D_i - A_{h_i}}{N_i - N_{ti}} \text{co}_{bb}} \right\rceil, N_i - N_{ti} + 1 \right\}, \\ D_i - A_{h_i} > 0; \\ N_i - N_{ti} + 1, (D_i - A_{h_i}) \leq 0. \end{cases} \quad (2)$$

其中: N_{*ti*}为工件*i*所等待加工的光刻步, 即将进入第N_{*ti*}层光刻; co_{bb}为任务集调整系数. 则在T_s天内, 光刻任务集为 Task_{T_s,WIP_{bb}} = {*h_{ij}* | *i* = LotID, *j* = N_{*ti*}, N_{*ti*} + 1, ..., N_{*ti*} + H_{*i*} - 1}. 工件*i*第1次到达光刻设备的时间为 A_{*h_{i1}*} = t_{inbf} + wait_{inbf}co_{bbt}. t_{inbf}为工件到达当前瓶颈缓冲区时刻, 可从数据库中得到; wait_{inbf}为预期等待平均时长, 可根据历史统计数据得到; co_{bbt}为微调系数, 依生产线载荷而定.

3) 决策时刻正在瓶颈设备上加工的工件数为 WIP_b, 工件*i*在调度期T_s天还将重入光刻设备次数为 (不包括当前正在进行的光刻工序)

$$H_i = \begin{cases} 0, N_i = N_{ti}; \\ \min \left\{ \left\lceil \frac{T_s}{\frac{D_i - A_{h_i}}{N_i - N_{ti}} \text{co}_b} \right\rceil, N_i - N_{ti} \right\}, \\ D_i - A_{h_i} > 0; \\ N_i - N_{ti}, (D_i - A_{h_i}) \leq 0. \end{cases} \quad (3)$$

其中: N_{*ti*}为工件*i*正在进行的光刻工序, 将其作为该工件调度期第1次光刻, 在T_s天内, 光刻任务集 Task_{T_s,WIP_b} = {*h_{ij}* | *i* = LotID, *j* = N_{*ti*}, N_{*ti*} + 1, ..., N_{*ti*} + H_{*i*}}. 工件*i*第1次到达光刻设备的时间为 A_{*h_{i1}*}, 是当前光刻工序开始时间, 可从数据库中得到; co_b为任务集调整系数.

4) 决策时刻正在非瓶颈设备或非瓶颈设备缓冲区的工件数为 WIP_{nb}, 工件*i*在调度期T_s天重入光刻设备次数为

$$H_i = \begin{cases} \min \left\{ \left\lceil \frac{T_s}{\frac{D_i - A_{h_{i(j-1)}}}{N_i - N_{ti}} \text{co}_{nb}} \right\rceil, N_i - N_{ti} + 1 \right\}, \\ D_i - A_{h_{i(j-1)}} > 0; \\ N_i - N_{ti} + 1, (D_i - A_{h_{i(j-1)}}) \leq 0. \end{cases} \quad (4)$$

其中: A_{*h_{i(j-1)}*}为上一次光刻开始时间, N_{*ti*}为将要进入的光刻层, co_{nb}为任务集调整系数. 则在T_s天内, 光刻任务集 Task_{T_s,WIP_{nb}} = {*h_{ij}* | *i* = LotID, *j* = N_{*ti*}, N_{*ti*} + 1, ..., N_{*ti*} + H_{*i*} - 1}. 工件*i*在调度期内第1次到达光刻设备的时间为 A_{*h_{i1}*} = t + p_{*t*}(now, N_{*ti*})co_t. p_{*t*}(now, N_{*ti*})表示当前工序至第1次光刻工序之间的净加工时长, co_t为微调系数.

至此, 可得到调度期内经过光刻区的加工任务集

$$\text{Task}_{T_s} = \text{Task}_{T_s, \text{NewR}} \cup \text{Task}_{T_s, \text{WIP}_{bb}} \cup \text{Task}_{T_s, \text{WIP}_b} \cup \text{Task}_{T_s, \text{WIP}_{nb}}.$$

上述过程中工件*i*在调度期内第1次到达光刻设备的时间 A_{*h_{i1}*} 作为初始化时蚂蚁概率选择的启发信息之一, 调度期内剩余到达光刻设备时间在寻优过程中随加工顺序不同实时计算, 同样在蚂蚁进行概率选择时起到启发作用.

下面利用析取图和蚁群算法, 得到调度期T_s天内每台瓶颈设备上工件的加工顺序及开工时间, 使得调度期内瓶颈加工区工件的平均拖期较小.

4 求解方法

4.1 析取图定义

析取图模型是调度问题的一类常用描述形式, Balas等人^[19]已将其应用于Job-shop调度问题. 这里将其改进, 用于多重瓶颈区并行机调度问题描述.

定义1 析取图G为三元组: G = (Q, A, E), Q为节点集, A和E为弧集. 分别定义如下.

节点集Q: 设瓶颈设备集为M, 有m台设备, 记为 M = {k₁, k₂, ..., k_m}, 每台设备虚设起始节点集O及终止节点集X, O = {O₁, O₂, ..., O_m}, X = {X₁, X₂, ..., X_m}. Q = Task_{T_s} × M ∪ O ∪ X.

有向连接弧集A: 连接同一工件的相邻光刻工序; 从虚设起始节点集O指向第1步可选工序; 从可能的最后一步工序集指向终止节点集X. 弧(h_{*ij*₁}, h_{*ij*₂}) ∈ A的长度等于工件*i*的*j*₁次光刻的加工时间与第*j*₂次光刻前的非瓶颈设备加工时间之和.

双向析取弧集E: 连接光刻中心相邻两道不同工件的加工工序. 弧(h_{*i*₁_{*j*₁}}, h_{*i*₂_{*j*₂}}) ∈ E (*i*₁ ≠ *i*₂) 的长度为 h_{*i*₁_{*j*₁}} 的完工时间 (t_{*h_i₁_{*j*₁}*}) 与紧接其后加工任务

$h_{i_2j_2}$ 的最早到达时间 ($A_{h_{i_2j_2}}$) 之差, 即

$$\text{Len}(h_{i_1j_1}, h_{i_2j_2}) = \begin{cases} 0, & t_{h_{i_1j_1}} \geq A_{h_{i_2j_2}}; \\ A_{h_{i_2j_2}} - t_{h_{i_1j_1}}, & t_{h_{i_1j_1}} < A_{h_{i_2j_2}}. \end{cases} \quad (5)$$

假设生产线上有工件 1 号和工件 2 号, 按 3.2 节计算方法得到在调度期 T_s 将产生 5 个光刻任务 $\text{Task}_{T_s} = \{h_{11}, h_{12}, h_{21}, h_{22}, h_{23}\}$; 瓶颈区有 3 台互替设备可用, $M = \{k_1, k_2, k_3\}$, 则 $\text{Task}_{T_s} \times M$ 有元素 15 个, $\text{Task}_{T_s} \times M = \{(h_{11}, k_1), (h_{11}, k_2), (h_{11}, k_3), (h_{12}, k_1), (h_{12}, k_2), (h_{12}, k_3), (h_{21}, k_1), (h_{21}, k_2), (h_{21}, k_3), (h_{22}, k_1), (h_{22}, k_2), (h_{22}, k_3), (h_{23}, k_1), (h_{23}, k_2), (h_{23}, k_3)\}$, 笛卡尔积包含了光刻任务分配至并行设备的每一种可能. 按上述定义, 可得到 2 个工件 5 步光刻工序 3 台瓶颈设备调度问题的析取图表示, 如图 2 所示.

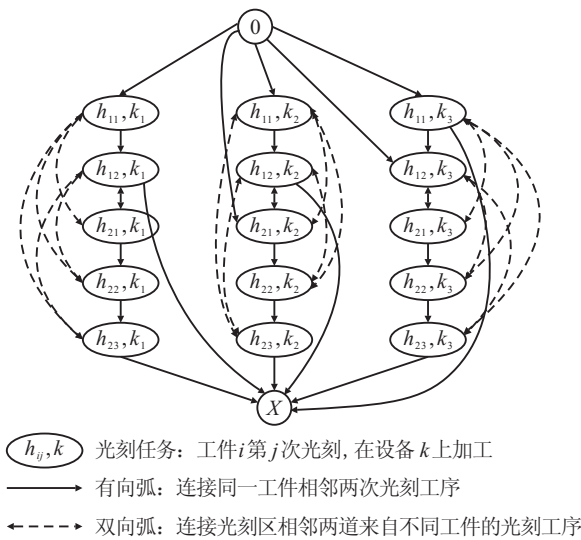


图 2 2 个工件 5 步工序 3 台设备调度问题的析取图

假设某只蚂蚁进行概率选择, 其初始结点选择 (h_{11}, k_2) , 选择完一个节点后需作如下删除和更新动作: 1) 删除 h_{11} 在图中同一行的其他节点, 表示已安排 h_{11} 在 k_2 上加工, 同时删除与之相连的双向弧; 2) 将原先指向被删除结点的有向连接弧, 重新指向被删除结点的后续结点, 与 h_{11} 结点相连的双向析取弧方向确定为从自身出, 指向其他节点; 3) 更新结点该次光刻工序的开始时间, 并据此更新该工件后续光刻步的最早到达时间.

重复以上过程, 直至所有工序均分配到设备上. 至此便建立了该调度问题的析取图模型.

4.2 蚁群优化算法求解

针对生产线上光刻区的 m 台并行设备, 蚂蚁每一步为每台设备选择 1 个节点, 设计两类禁忌表:

1) 总禁忌表 tabu . 每选择 1 个节点, 便将其插入 tabu 表, 直至无节点可选时. 该表记录了该蚂蚁所有访问过的节点, 便于计算目标函数值.

2) m 个设备表 $\text{tabu}_k, k = 1, 2, \dots, m$. 记录了每台设备所安排的工序及其开工时间.

4.2.1 信息素初始化

从虚设节点至其他节点的有向连接弧上的信息素初始值为

$$\tau'_0 = \begin{cases} 1000, & h_{ij} \text{ 正在瓶颈设备上加工;} \\ 0, & h_{ij} \text{ 不是调度区间第 1 次光刻;} \\ \frac{100(t - A_{ij})}{\max_{i,j}(t - A_{ij})} + \frac{10(D_i - A_{ij})}{N_i - N_{ti} + 1}, & \\ h_{ij} \text{ 为调度区间第 1 次光刻.} \end{cases} \quad (6)$$

其他双向析取弧上的信息素初始值为一很小值 $\tau_0 = 1/K(1000\text{TWT}_{\text{ER}})$. 这里: K 为蚂蚁数目, TWT_{ER} 为生产线总拖期的历史数据统计值. 蚂蚁为 m 台设备分别选择 l 值最大的节点作为起始节点.

$$l = \tau'_0 + \max(\text{rand}(0, 1) \times \tau'_0). \quad (7)$$

4.2.2 信息素更新

1) 局部信息素更新. 当蚂蚁每次选取一个工序节点时, 便当即更新当前完工节点至所选工序节点路径上的信息素

$$\tau_{c_0l} = (1 - \xi)\tau_{c_0l} + \xi\tau_{c_0},$$

其中 ξ 为一很小的常数, 如 0.1. 其作用为减少路径上的信息素浓度, 避免算法局部收敛.

2) 全局信息素更新. 提高算法寻找最优路径的概率, 当一次迭代结束之后, 对比当前迭代中各蚂蚁搜索到的目标值, 更新如下:

$$\begin{aligned} \tau_{xy}(t+1) &= \\ (1 - \rho)\tau_{xy}(t) + \rho\Delta\tau_{xy}^{bs}, & \forall (x, y) \in T^{bs}; \\ \Delta\tau_{xy}^{bs} &= \\ 1 / \min \left(\left(\sum_{i=1}^n w_{h_i} \times Td_{h_i} \right) / n \right), & 0 < \rho < 1. \end{aligned}$$

其中: ρ 为挥发系数, 是一较小常数, 如 0.1; T^{bs} 为当前迭代中最优方案.

4.2.3 选择节点

蚂蚁根据可选路径上的信息素浓度以及启发信息计算每一个可选路径的选择概率, 计算如下:

$$l = \begin{cases} \arg \max_{c \in L_{\text{task}}^k} \left\{ \frac{\tau_{c_0c}^\alpha \eta_{c_0c}^\beta}{\sum_c \tau_{c_0c}^\alpha \eta_{c_0c}^\beta} \right\}, & q \leq q_0; \\ \max_c \left(\text{rand}(0, 1) \times \frac{\tau_{c_0c}^\alpha \eta_{c_0c}^\beta}{\sum_c \tau_{c_0c}^\alpha \eta_{c_0c}^\beta} \right), & q > q_0. \end{cases} \quad (8)$$

其中: η_{c_0c} 为启发式信息, 与下一节点的加工时间、最早到达时间、设备负载相关; c_0 表示设备上正在加工的工序, c 表示所选择的下一道工序. η_{c_0c} 与 ΔW_c 分别计算如下:

$$\eta_{c_0c} =$$

$$\left(1 - \frac{P_c + \max\{(A_c - C_{c_0}), 0\}}{\max_c(P_c) + \max\{(\max_c(A_c) - C_{c_0}), 0\}}\right) + \Delta W_c;$$

$$\Delta W_c = \begin{cases} \frac{W_c}{\max_m(W_m)}, & W_c \leq \max_m(W_m); \\ \frac{W_c}{\max_m(W_m)} - 1, & W_c > \max_m(W_m). \end{cases}$$

其中： P_c 为后续任务的加工时长， ΔW_c 为选择加工该任务后设备增加的载荷， $\max_m(W_m)$ 为当前设备最大载荷。

5 仿真算例

以HP24模型为生产线对象，该模型曾由Wein等人^[4]所采用。数据来源于真实的HP研究型FAB：TRC (Hewlett-Packard technology research center silicon FAB)，包括3种类型生产线，其中Fab1模型为瓶颈突出生产线，本文便在Fab1模型上验证所提出的方法。HP24-Fab1的光刻区PHGCA为瓶颈区(3台并行设备)，加工1种产品，光刻区重入12次，加工流程172步，总计净加工时间549.3h，即22.8875d。

本文以HP24-Fab1模型为基础搭建3个比较典型的生产仿真环境来验证所提出的算法，并将算法结果与同样环境用启发式规则EDD(earliest due date first)和FIFO规则得到的结果进行30组抽样比较，每组用于比较的数据为多次仿真结果的平均值。之所以选择这2个启发式规则作为参照，是因为在蚁群寻优过程中，启发信息包括了交货期紧迫度以及最早到达设备的时间因素。

3个算例的生产线均采用固定间隔投料策略，通过调整投料速度调整生产线的负荷。具体流程如下：首先在仿真平台上采用启发式规则仿真较长一段时间(如半年)，以使生产线达到稳定；在某一决策时刻 t (本文取某天00:00:00时刻)产生2个动作：抽取数据到算法数据库，继续仿真 T_s 天得到仿真方案；运行半在线调度方法，依照前两节所述过程，计算瓶颈区 T_s 天的任务集(本文取 $T_s = 1$)，通过蚁群优化算法求得调度优化方案，与仿真方案进行性能比较分析。

算例1 HP24-Fab1 轻载平衡。

投料时间间隔为36h，平均生产周期为23~25d，工件交货期为23~35d均匀分布，瓶颈设备平均利用率为84%~86%。在任意时刻 t 求解第2天调度方案，对瓶颈任务集分别采用半在线调度方法、FIFO调度和EDD调度，结果如图3所示。横坐标为决策时刻 t ，

即仿真某天00:00:00时刻。

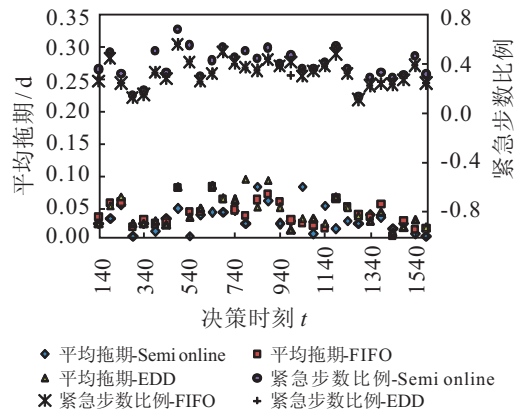


图3 轻载平衡时半在线调度方法与启发式规则对比

从图3可以看出，30次抽样有3次逊于FIFO规则，3次瓶颈区工序拖期率为0，且瓶颈区设备利用率略有提高，见表1。

从图3和表1数据可以看出，半在线调度方法能从整体考虑安排未来一段时间的瓶颈任务，在保证瓶颈利用率的情况下，能明显改进瓶颈区工序平均拖期性能指标。

算例2 HP24-Fab1 生产线平衡产能饱和。

经多次长时间仿真得到HP24Fab1产能趋于饱和且平稳时的基本数据如下：投料时间间隔32h，平均生产周期27~28d。由于产能几近饱和，导致生产周期与算例1相比较长，虽然工件交货期仍为23~35d均匀分布，但比轻载情况更为紧迫。此时HP24-Fab1瓶颈利用率高达98%，接近于产能完全利用的理想状态。同样在任意时刻 t 求第2天的瓶颈区调度方案。结果如图4所示，横坐标为决策时刻 t ，即仿真某天00:00:00时刻，平均性能比较见表2。

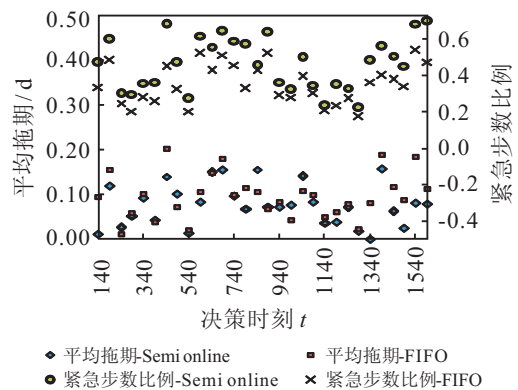


图4 饱和稳定时半在线调度方法与FIFO规则比较

表1 HP24-Fab1 轻载平衡时半在线调度方法平均性能

调度方法	平均移动/步	对比FIFO	平均拖期/d	对比FIFO	紧急步数比例平均值	对比FIFO
Semi online	57	1.79%	0.03	21.75%	0.4028	19.85%
EDD	56	1	0.04	-9.60%	0.3404	1.30%
FIFO	56	1	0.04	1	0.3361	1

表2 HP24-Fab1 饱和稳定时半在线调度方法平均性能

调度方法	平均移动/步	对比 FIFO	平均拖期/d	对比 FIFO	紧急步数比例平均值	对比 FIFO
Semi online	58	-7.9%	0.08	21.88%	0.463 5	30%
FIFO	63	1	0.1	1	0.356 5	1

表3 HP24-Fab1 不同负荷下半在线调度性能比较

HP24-Fab1	WIP	平均拖期/d			平紧急步数比例			移动步数		对比 FIFO/%
		Semi online	FIFO	对比 FIFO/%	Semi online	FIFO	对比 FIFO/%	Semi online	FIFO	
轻载	16	0.029	0.04	21.75	0.4	0.34	19.85	57	56	1.79
饱和	19	0.076	0.1	21.88	0.46	0.36	30.00	58	63	-7.90
过载	30	2.522	2.87	12.11	0.44	0.3	44.65	56	65	-13.80

图4表明,在产能趋于饱和和生产线稳定情况下,半在线调度方案瓶颈工序平均延误性能指标较之FIFO有明显提高,有5次情况略逊于FIFO.从表2可见,平均延误性能指标平均提高21.88%,其中紧急工件被调度的步数比按FIFO调度情况下平均多出30.00%.即平均拖期性能并未由于紧急工件被调度的步数比例大而下降,而是改善了很多.但瓶颈利用率为91%左右,下降了约7.9%.

算例3 HP24-Fab1 生产线过载.

投料间隔时间缩短为29h,随着仿真时间增加,生产线在制品逐渐堆积,投料速度增加,堆积速度随之增加.表3关于过载情况的数据是在WIP堆积不非常严重的情况下取抽样平均值所得,“轻载”和“饱和”情况的根据是算例1和算例2的平均值.从表3可见,“过载”情况下,虽然平均拖期性能有所提高,但瓶颈设备利用率损失较大.

分析上述算例结果,可得如下结论:

1) 在产能饱和和情况下,瓶颈设备几乎满载时(算例2瓶颈设备利用率98%~99%)的平衡态非常脆弱,任何调整性的动作都会带来扰动.因此,半在线调度方法并不适用,虽然调度性能会有所提高,但以损失一定的产能为代价,此时对生产线的控制应配合投料策略及其他策略加以改进.生产线过载时也存在同样的问题.

2) 在生产线轻载稳定时(算例1瓶颈设备利用率为84%~86%),生产线本身的平衡态具有一定鲁棒性,有一定的调度自由度,由调度调整所带来的扰动可以被吸收并保持加工的平稳流动.此时通过半在线调度方法,不仅可以较明显地提高未来一段时间瓶颈区任务的平均拖期性能,而且可以提高瓶颈设备利用率,相应提高产能.

6 结 论

实际晶圆生产线是不停机生产线,即工件加工的同时,各种工件不断进入系统,完成加工的工件也不

断离开,“非零初始状态”这一点在很多优化方法中被忽略,从而使得这些方法在实际生产应用中遇到很大困难,甚至无法使用.本文讨论非零初始状态、瓶颈突出生产线的并行瓶颈设备调度问题,提出了一种新的解决思路,并进行了初步研究和尝试.本文充分利用生产信息、晶圆制造的“多重入”特点以及实际生产中已知信息模式,提出了光刻区并行机半在线调度方法:分析短期内光刻区重入流的来源,计算可能产生的光刻任务集,定义光刻区并行机任务集析取图,通过蚁群算法求解瓶颈区平均加权拖期时间的优化调度方案.最后通过大量仿真实验,得出了半在线调度方法在半导体生产线的适用场景.

本文所提出的方法尚有很多待完善之处,比如未来一段时间的任务集确定方法有待进一步研究,仿真实验所用模型规模较小,实验场景有限等,这些工作都需进一步探索并加以完善.

参考文献(References)

- [1] Young H L, Jongkwan P, Sooyoung K. Experimental study on input and bottleneck scheduling for a semiconductor fabrication line[J]. IIE Trans, 2002, 34(2): 179-190.
- [2] 蔡志弘,曾综源,李荣贵.半导体制造业投料与瓶颈机器指派管理资讯系统之分析与设计[J].中华管理评论, 1999, 2(2): 51-66.
(Cai Z H, Zeng Z Y, Li R G. An analysis and design of the information system of releasing and bottleneck tools dispatching in semiconductor manufacturing[J]. Web J of Chinese Management Review, 1999, 2(2): 51-66.)
- [3] Oliver Rose. CONLOAD—A new lot release rule for semiconductor wafer fabs[C]. Proc of the 1999 Winter Simulation Conf. Phoenix, AZ, 1999: 850-855.
- [4] Wein L M. Scheduling semiconductor wafer fabrication[J]. IEEE Trans on Semiconductor Manufacturing, 1988, 1(3): 115-130.
- [5] Glassey C R, Resende M G C. Closed-loop job shop release control for VLSI circuit manufacturing[J]. IEEE Trans on Scheduling and Logistics, 1992, 24(4): 47-58.

- [6] 李莉, 乔非, 姜桦, 等. 半导体生产线动态调度方法研究[J]. 计算机集成制造系统, 2004, 10(8): 949-954.
(Li L, Qiao F, Jiang H, et al. Research on dynamic dispatching rule for semiconductor wafer fabrication[J]. Computer Integrated Manufacturing Systems, 2004, 10(8): 949-954.)
- [7] 王遵彤, 乔非, 吴启迪. 半导体硅片加工过程复合控制策略研究及仿真[J]. 系统仿真学报, 2005, 17(8): 1924-1927.
(Wang Z T, Qiao F, Wu Q D. Research and simulation of compound controlling strategy in semiconductor wafer fabrication[J]. J of System Simulation, 2005, 17(8): 1924-1927.)
- [8] Zuntong Wang, Qidi Wu, Fei Qiao. A lot dispatching strategy integrating WIP management and wafer start control[J]. IEEE Trans on Automation Science and Engineering, 2007, 4(4): 579-583.
- [9] 李晓红, 周炳海. 晶圆制造过程中动态瓶颈设备的实时调度[J]. 上海交通大学学报, 2008, 42(4): 599-602.
(Li X H, Zhou B H. Real-time scheduling of dynamic bottleneck facilities of wafer fabrication processes[J]. J of Shanghai Jiaotong University, 2008, 42(4): 599-602.)
- [10] 张怀, 江志斌, 郭乘涛. 面向瓶颈的半导体晶圆制造系统派工策略及参数优化[J]. 上海交通大学学报, 2007, 41(8): 1252-1257.
(Zhang H, Jiang Z B, Guo C T. The bottleneck-oriented dispatching strategy and parameter optimization in semiconductor wafer fabrication system[J]. J of Shanghai Jiaotong University, 2007, 41(8): 1252-1257.)
- [11] Zhang Huai, Jiang Zhibin, Lee Yen-Fei, et al. An approach of dynamic bottleneck machine dispatching for semiconductor wafer fab[C]. Int Symposium on Semiconductor Manufacturing. Santa Clara, 2007: 1-4.
- [12] 余红霞. 基于Petri网与免疫算法的半导体生产线建模与优化调度[D]. 上海: 同济大学电子信息与工程学院, 2009.
(Yu H X. Petri-net and IA-based modeling and scheduling of semiconductor wafer fabrication facilities[D]. Shanghai: School of Electronics and Information Engineering, Tongji University, 2009.)
- [13] 丁小进, 乔非, 李莉. 基于DBR的半导体晶圆制造生产调度研究[J]. 机电一体化, 2008, (3): 29-31.
(Ding X J, Qiao F, Li L. Research of DBR scheduling method for semiconductor manufacturing[J]. Mechatronics, 2008, (3): 29-31.)
- [14] 丁小进. 基于DBR的半导体制造生产调度研究[D]. 上海: 同济大学电子信息与工程学院, 2008.
(Ding X J. Research on scheduling of semiconductor production based on DBR Theory[D]. Shanghai: School of Electronics and Information Engineering, Tongji University, 2008.)
- [15] 郭永辉, 钱省三. 基于DBR理论的半导体晶圆厂生产作业控制[J]. 工业工程与管理, 2006, 11(5): 70-75.
(Guo Y H, Qian X S. Production control in wafer fabrication based on DBR theory[J]. Industrial Engineering and Management, 2006, 11(5): 70-75.)
- [16] Wu H-H, Yeh M-L. A DBR scheduling method for manufacturing environments with bottleneck re-entrant flows[J]. Int J of Production Research, 2006, 44(5): 883-902.
- [17] 何勇, 杨启帆, 谈之奕. 平行机半在线排序问题研究I[J]. 高等应用数学学报: A辑, 2003, 18(1): 105-114.
(He Y, Yang Q F, Tan Z Y. Semi on-line scheduling on parallel machines (I) [J]. Applied Mathematics A J of Chinese Universities, 2003, 18(1): 105-114.)
- [18] 张居阳, 孙吉贵, 杨轻云. 半在线调度中约束求解算法研究[J]. 自动化学报, 2007, 33(7): 765-767.
(Zhang J Y, Sun J G, Yang Q Y. A constrained solving algorithm for semi on-line scheduling problem[J]. Acta Automatica Sinica, 2007, 33(7): 765-767.)
- [19] Balas E. Machine scheduling via disjunctive graphs: An implicit enumeration algorithm[J]. Operations Research, 1969, 17(6): 941-957.

(上接第338页)

- [9] Willy H, Roel L. Project scheduling under uncertainty: Survey and research potentials[J]. European J of Operational Research, 2005, 165(2): 289-306.
- [10] 包振强, 李长仪, 周鑫. 基于知识的动态调度决策机制研究[J]. 中国机械工程, 2006, 17(13): 1366-1370.
(Bao Z Q, Li C Y, Zhou X. A knowledge—Based dynamic scheduling decision system[J]. China Mechanical Engineering, 2006, 17(13): 1366-1370.)