

第3章 晶圆制造并行机调度问题建模

3.1 引言

晶圆制造系统是典型的离散时间动态系统,存在各种各样的不确定性。设备情况变化、客户需求变化、加工瓶颈变化、新产品试制、临时工艺更改、返工等都会影响生产线正常运行,造成性能指标波动。晶圆制造过程集成混合了各种生产调度模型,大大增加了生产线复杂性和调度难度。为了设计有效的晶圆制造系统并行机调度方法,首先需要建立合理的调度模型对不同设备调度的本质特点进行分析。

具体来说,该模型应该具有以下功能:①对并行设备的内在运行变化过程和机理进行分析,确定系统的静态和动态因素对系统运行性能的影响,进而为建立有效的调度方法提供基础;②确定评价调度性能的评价指标,进而对调度方法的性能进行快速有效的评价,以验证调度方法的有效性。

目前,晶圆制造系统中加工系统并行机调度问题的建模方法主要有数学规划模型和排队论模型;物流运输系统并行机调度问题的建模方法主要有复杂网络模型、马尔可夫模型及仿真模型等。

3.2 晶圆制造加工系统建模

3.2.1 基于数学规划模型的晶圆制造加工系统建模

3.2.1.1 基本概念

数学规划是运筹学的一个重要分支,也是现代数学的一门重要学科。其基本思想出现在19世纪初,并由美国哈佛大学的Robert Dorfman于20世纪40年代末提出。数学规划的研究对象是数值最优化问题,这是一类古老的数学问题。古典的微分法已可以用来解决某些简单的非线性最优化问题。直到20世纪40年代以后,由于大量实际问题的需要和电子计算机的高速发展,数学规划才得以迅速发展起来,并成为一门十分活跃的新兴学科。今天,数学规划的应

用极为普遍,它的理论和方法已经渗透到自然科学、社会科学和工程技术中。

数学规划问题实际上都是在满足某些约束条件的情况下求函数极值的问题。一个数学规划问题通常包含以下要素:一组决策变量,变量的值就是要求解的对象;一个目标函数,就是我們希望通过改变决策变量而最大化或最小化的对象;一组约束条件,决策变量的取值必须满足这些约束条件。通常满足所有约束条件的解被称作可行解,所有可行解的集合叫作可行域。

数学规划模型一般可表示为

$$\begin{aligned} \min(\max) \quad & f(x, \alpha, \beta) \\ \text{s. t.} \quad & g(x, \alpha, \beta) \leq 0 \end{aligned} \quad (3-1)$$

其中, f 表示数学规划问题的目标函数, g 为数学规划问题的约束函数, x 为数学规划问题中的变量, α 表示模型中的已知参数, β 为模型中的随机参数。

数学规划主要包括线性规划、整数规划、混合整数规划以及动态规划等,按照最优化理论可将数学规划问题划分为无约束优化问题和约束优化问题两大类,可通过最优化理论和算法对数学规划模型进行化简和求解。

3.2.1.2 数学规划模型在晶圆制造并行机调度中的应用

晶圆制造系统中的加工设备划分为:单件加工型设备、批处理型设备等。Kumar^[43]等学者曾利用线性规划的方法解决晶圆制造并行机调度问题,给出了生产过程中保证生产稳定进行的 WIP 水平界限和开环带重入生产线最小延迟的界限,对实际生产起到一定的指导作用。但总体而言,由于调度问题的复杂度为 NP-难,所以数学规划方法的应用受到了限制。为此,提出了分支定界法、拉格朗日松弛法以及过滤束搜索等方法以降低问题的难度。Sung 和 Choung^[55]曾使用分支定界法解决半导体生产工厂中的一类批处理设备——炉管的动态调度问题,其优化目标是最大化所有工件的完成时间(Makespan)。Kaskavelies 和 Caramanis^[56,57]使用拉格朗日松弛算法对超过 10 000 个资源约束的半导体测试工厂进行调度。他们介绍了两类拉格朗日乘子更新过程的新特征,以帮助运算的进行。De 和 Lee^[58]使用过滤束搜索方法,构建一个基于知识的调度系统,以调度半导体测试生产线。其调度目标是最小化总生产周期。在调度过程的搜索树上,所需要搜索的节点的数量取决于过滤宽度和束的宽度。

批处理机一般为瓶颈机,它制约着晶圆制造系统的绩效。这里介绍数学规划模型在具有重入特性的批处理机重组批调度问题中的应用,具体描述如下。

加工工艺流程简图(图 3-1)主要包括两个设备组,设备组 1 和设备组 2,且这两个设备组均是批处理机,设备组 2 的最大批处理能力大于设备组 1;设备组的产品组批包括两种类型:产品相容和不相容:设备组 1 具有产品不相容性

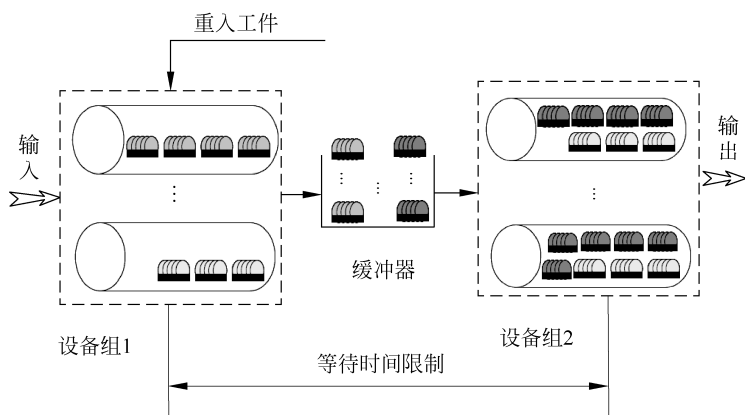


图 3-1 具有重入特性的批处理组批加工工艺流程简图

组批、重入特性；设备组 2 具有产品相容性组批，但由设备组 1 提供的批不拆分，具有最大批处理能力限制，需要对进入缓冲器中产品批进行重组批、排序^[59]。主要假设如下：

(1) 设备组 2 由于加工时间长，故此工位是生产瓶颈，假设设备组 2 不会出现饥饿。

(2) 抢占不允许，即任意设备一旦开始加工，加工过程就不允许中断。

(3) 缓冲器里产品批到达是动态的。

(4) 设备组 2 的缓冲器里产品批存在等待时间限制，即为了避免由于工件表面接触空气时间过长而对产品的质量造成影响，产品批在设备组 1 中加工完成后必须在规定的时间内进入设备组 2 中加工。

(5) 三层混合调度算法中第一层完成缓冲器里产品批实时状态信息和设备组 2 实时状态信息的记录、保存和传送；第二层完成进入缓冲器中产品重组批、排序；第三层完成重组优先批的装载分配及其分配完毕后缓冲器里产品批实时状态信息和设备组 2 的实时状态信息更新、保存。

(6) 不计工件从设备 1 组到缓冲器，从缓冲器到设备组 2 和在缓冲器重组批的过程时间。

为了使术语具有通用性，对后面需要的术语进行统一如下定义。

K ：当设备组 2 出现设备空闲可用时，缓冲器里来自设备组 1 的不同批的数量集；

K' ：当设备组 2 出现设备空闲可用时，重组批的数量集；

S ：设备组 2 的最大加工能力；

M ：一个足够大的正整数；

t ：设备组 2 出现设备空闲可用时的当前时间；

q_k : 批 k 在设备组 1 和设备组 2 之间等待限制时间;

P : 设备组 2 的加工时间;

s_k : 批 k 包含的工件数量;

r_k : 批 k 到达缓冲器的时间;

d_k : 批 k 的交货时间;

ZS_k : 批 k 总加工工序数量集;

DS_k : 批 k 在设备组 2 上所属的加工工序数;

P_{wk} : 批 k 在第 w 个加工工序中的加工时间;

PR_k : 批 k 在设备组 2 后纯剩余的加工时间;

$dr_{k'}$: 重组批 k' 的交货时间;

$Cr_{k'}$: 重组批 k' 的完工时间;

$PRr_{k'}$: 重组批 k' 在设备组 2 后纯剩余加工时间;

$\sum T$: 总拖延时间和。

注: $k=1,2,\dots,K$; $k'=1,2,\dots,K'$; $w=1,2,\dots,ZS_k$ 。

以拖延时间和最小为目标,且具有等待时间限制和工件动态到达的批处理机重组批的 NP-难调度问题,在分解规则下,将总的调度时间窗分解为许多时间窗,每个时间窗的长度等于相邻两个触发事件之间的时间间隔,其值是变化的,即变时间窗;在滚动时域策略下,每个时间窗之间不断更新优化调度结果。

每个时间窗内的子问题,由于规模被减小,精确算法可以适用。在整数规划模型求解重组批及其排序优化调度问题中,重组批状态的 0-1 变量和重组批的排序位置序号是决策变量;满足设备加工时间,同一设备不能同时加工两个重组批,同一重组批只能占有一个排序位置序号等是约束条件;总拖延时间和最小为目标函数,具体整数规划模型如下:

目标函数

$$\sum T = \sum_{k'=1}^{K'} \max[(Cr_{k'} - dr_{k'}), 0] \quad (3-2)$$

约束条件

$$Cr_{k'} = t + \text{Position}(k') \times P + PRr_{k'}, \quad \forall k' \in \{1, 2, \dots, K'\} \quad (3-3)$$

$$\text{Position}(k') \in \{1, 2, \dots, K'\}, \quad \forall k' \in \{1, 2, \dots, K'\} \quad (3-4)$$

$$\text{如果 } k' \neq l', \text{ 那么 } \text{Position}(k') \neq \text{Position}(l'), \quad (3-5)$$

$$\forall k' \text{ 和 } l' \in \{1, 2, \dots, K'\}$$

$$x_{kk'} = \begin{cases} 1, & \text{当缓冲器的批 } k \text{ 被分配到重组批 } k' \text{ 中} \\ 0, & \text{其他} \end{cases}, \quad (3-6)$$

$$\forall k' \in \{1, 2, \dots, K'\}, \forall k \in \{1, 2, \dots, K\}$$

$$\sum_{k'=1}^{K'} x_{kk'} = 1, \quad \forall k \in \{1, 2, \dots, K\} \quad (3-7)$$

$$PR_k = \mu_w \sum_{w=DS_k+1}^{ZS_k} P_{wk}, \quad \forall k \in \{1, 2, \dots, K\} \quad (3-8)$$

$$\mu_w = \begin{cases} s_k, & \text{如果第 } w \text{ 的工序数不是批处理设备} \\ \frac{s_k}{B_w}, & \text{如果第 } w \text{ 的工序数是批处理设备, 且其加工能力是 } B_w \end{cases} \quad (3-9)$$

$$K' = \left\lceil \left(\sum_{k=1}^K s_k \right) / S \right\rceil \quad (3-10)$$

$$\sum_{k=1}^K x_{kk'} \cdot s_k \leq S, \quad \forall k' \in \{1, 2, \dots, K'\} \quad (3-11)$$

$$PR_{k'} = \min_{1 \leq k \leq K} (x_{kk'} \times PR_k), \quad \forall k' \in \{1, 2, \dots, K'\} \quad (3-12)$$

$$dr_{k'} = \min_{1 \leq k \leq K} (x_{kk'} \times d_k), \quad \forall k' \in \{1, 2, \dots, K'\} \quad (3-13)$$

$$t - r_k + x_{kk'} \times \text{Position}(k') \times P \leq q_k, \quad \forall k' \in \{1, 2, \dots, K'\}, \forall k \in \{1, 2, \dots, K\} \quad (3-14)$$

式(3-2)是每个时间窗内的总拖延时间和最小的目标函数。约束条件式(3-3)用于计算重组批 k' 的完工时间。约束条件式(3-4)和式(3-5)用于确保每个重组批只占有一个排序位置号,且每个排序位置号只有一个重组批。约束条件式(3-6)引入重组批状态的 0-1 变量。约束条件式(3-7)确保缓冲器中每个批都被进行重组批,且只被进行 1 次重组批。约束条件式(3-8)用于计算批 k 在设备组 2 后剩余的加工时间,其中 μ_w 是调整系数。约束条件式(3-9)用于确定调整系数。约束条件式(3-10)用于计算当设备组 2 出现设备空闲可用时,重组批的数量集 K' 。约束条件式(3-11)用于确保任一重组批包含工件数量不大于设备组 2 的最大加工能力 S 。约束条件式(3-12)用于计算重组批 k' 在设备组 2 后剩余的加工时间,其值等于重组批中所包含产品族中在设备组 2 后剩余的加工时间的最小值。约束条件式(3-13)用于计算重组批 k' 的交货时间,其等于重组批中所包含产品族中交货期的最小值。约束条件式(3-14)用于确保被调度批 k 在缓冲器里的等待时间不超过设备组 1 和设备组 2 之间等待限制时间 q_k 。

注: 约束条件式中 $\lceil a \rceil$ 代表返回到不小于 a 的最小整数,例如, $\lceil (\sum_{k=1}^K s_k) / S \rceil$

表示返回不小于 $\sum_{k=1}^K s_k$ 除以 S 的最小整数,当 $\sum_{k=1}^K s_k = 51$ 和 $S = 12$,那么

$$\left\lceil \left(\sum_{k=1}^K s_k \right) / S \right\rceil = \lceil 51/12 \rceil = \lceil 3.5 \rceil = 4. (\text{Position}(1), \text{Position}(2), \dots, \text{Position}(K'))$$

是优化排序后位置序号的一维数组。

在 CPLEX 和开发程序联合平台中,线性的约束适合 CPLEX 求解。约束条件式(3-12)、式(3-13)和式(3-14)中表达关系式是非线性的,需要进行线性化。

Montagne 介绍批调度时考虑用平均加工时间 $\bar{p} = \sum_{i=1}^n p_i/n$ 计算比率系数 $\frac{p_i}{n \times \bar{p} - d_i}$ 处理总拖延时间之和最小问题的 Montagne 比率方法^[60]。这里基于 Montagne 比率方法,计算每个重组批内所有待调度批的平均加工时间,分别建立约束条件式(3-12)和式(3-13)的松弛线性约束条件式(3-15)和式(3-16)。

$$\text{PRr}_{k'} = \frac{\sum_{k=1}^K (x_{kk'} \times \text{PR}_k)}{\sum_{k=1}^K x_{kk'}}, \quad \forall k' \in \{1, 2, \dots, K'\} \quad (3-15)$$

$$\text{dr}_{k'} = \frac{\sum_{k=1}^K (x_{kk'} \times d_k)}{\sum_{k=1}^K x_{kk'}}, \quad \forall k' \in \{1, 2, \dots, K'\} \quad (3-16)$$

采用大 M 法将约束条件式(3-14)线性化:

$$\begin{aligned} t - r_k + \text{Position}(k') \times P - M(1 - x_{kk'}) &\leq q_k, \\ \forall k' \in \{1, 2, \dots, K'\}, \quad \forall k \in \{1, 2, \dots, K\} \end{aligned} \quad (3-17)$$

综上,等式(3-2)、约束条件式(3-3)~式(3-11)和式(3-15)~式(3-17)构建基于松弛的混合整数线性规划模型,适合 CPLEX 和开发程序联合平台优化求解。实验结果表明,该模型能在 CPLEX 环境下以较短 CPU 运行时间取得令人满意的目标函数值。

3.2.2 基于排队论模型的晶圆加工系统建模

3.2.2.1 基本概念

排队论,或称随机服务系统理论,是通过对服务对象到来及服务时间的统计研究,得出这些数量指标(等待时间、排队长度、忙期长短等)的统计规律,然后根据这些规律来改进服务系统的结构或重新组织被服务对象,使得服务系统既能满足服务对象的需要,又能使机构的费用最经济或某些指标最优。它是数学运筹学的分支学科,也是研究服务系统中排队现象随机规律的学科。所谓排队,指在服务系统中要求服务的“顾客”所形成的等待线。其主要研究方向是如何合理地设计与控制服务系统,使之以最经济的服务机构满足顾客的需要。由

于服务系统中顾客的到达及服务时间和次序都是随机的,因此,排队论又被称为随机服务系统理论。

D. G. Kendall 首先用 3 个字母组成的符号 A/B/C 表示排队系统,其中:A 表示顾客到达时间分布;B 表示服务时间的分布;C 表示服务机构中的服务台的个数。在此基础上,L. Takacs 等人又将组合方法引进排队论,使它更能适应各种类型的排队问题。

排队论在晶圆制造中的研究在 1950 年也已经开始,用于解决在制品排队问题的方法层出不穷,从大的方向上来看有两个显著的研究趋势:①基于运筹学的排队方法,强调数学规划的方法来建立排队模式,或利用启发式搜索法来求得最佳或近似最佳的排队方法,这里面包括派工法、线性规划、动态规划、分支界限法、启发式搜索等方法;②基于知识的方法,即通过模拟人类解决问题的思考方法或人类某部分器官的功能来解决复杂度高的排队问题,这里面包括智能调度专家系统、神经网络、基因演算法、限制条件满足技巧、基于约束的方法以及一些基于智能搜索的算法。

3.2.2.2 排队论模型在晶圆制造并行机调度中的应用

排队论模型能直观描述生产线的加工过程,因此,很多生产系统都用此方法建模。Kumar^[61]将晶圆制造系统抽象成可重入的排队网络,Meyn 等^[62]用马尔可夫模型对可重入排队网络进行建模,对模型研究了相应的动态规划算法。黄敏等^[63]提出了排队论模型和非线性整数规划模型相结合的方法,并通过遗传算法对模型求解。朱连章等^[64]在基本排队论模型的基础上加入 Monitor 监视器监视网络状态,在仿真过程中收集状态数据进行性能评价,该方法在分析过程中不会对系统结构进行限制,具有很大的灵活性。

现代晶圆制造业正在朝着更高密度的集约化生产和更大规模的量产方向发展。在现代的半导体晶圆制造厂中,经常能看到晶圆在进行某些工艺流程的加工制造中,通常会有多台晶圆制造设备来对应同一工艺和工序。于是,这就给晶圆制造中的排队策略带来了新的问题:如何达到晶圆制造并行设备的最优控制——是采用多制造设备、多队列的排队系统方案,还是应该采用多制造设备、单队列的排队系统方案。为了解决这个问题,通过建立符合实际情况的数学模型来求解^[65]。

首先考虑当某个批次的晶圆加工到某一工艺和工序时,有 M 台晶圆制造设备正在独立的并行制造,这里只考虑晶圆制造设备状态处于正常和待机两种情况下,即有 M 台晶圆制造设备处于正常或待机。当晶圆到达时,若有处于待机状态的晶圆制造设备便立刻开始加工制造,若当前所有的晶圆制造设备都处

于正常状态,则这个批次的晶圆进入排队等待系统,直到有处于待机状态的晶圆制造设备时再开始加工制造。

其次,为了使建立的数学模型更具有随机性和典型性,这里假设晶圆批次按照参数为 $\lambda (\lambda > 0)$ 的泊松流到达,每个批次晶圆需要待加工的制造时间是独立的,且服从参数为 $\mu (\mu > 0)$ 的负指数分布,不考虑排队等待系统的容量限制,而且到达与加工制造是彼此独立的。假设 $M(t)$ 表示某时刻 t , M 台晶圆制造设备中的晶圆批次数,令

$$p_{k_1 k_2}(\Delta t) = P\{M(t + \Delta t) = k_2 \mid M(t) = k_1\} \quad (3-18)$$

易得

$$p_{k_1 k_2}(\Delta t) = P \begin{cases} \lambda \Delta t + o(\Delta t), & k_2 = k_1 + 1, k_1 \geq 0 \\ k_1 \mu \Delta t + o(\Delta t), & k_2 = k_1 - 1, k_1 = 1, 2, \dots, M-1 \\ k_1 \mu \Delta t + o(\Delta t), & k_2 = k_1 - 1, k_1 = M, M+1, \dots \\ o(\Delta t), & |k_1 - k_2| \geq 2 \end{cases} \quad (3-19)$$

于是 $\{M(t), t \geq 0\}$ 为 $E = \{0, 1, 2, \dots\}$ 上的生灭过程,其中:

$$\lambda_{k_2} = \lambda, \quad k_2 \geq 0; \quad \mu_{k_2} = \begin{cases} k_2 \mu, & 1 \leq k_2 < M \\ M \mu, & k_2 \geq M \end{cases} \quad (3-20)$$

再令 $\rho = \frac{\lambda}{\mu}, \rho_M = \frac{\lambda}{M\mu}, p_{k_2} = \lim_{t \rightarrow \infty} P\{M(t) = k_2\}, k_2 = 0, 1, 2, \dots$, 则当 $\rho_M < 1$,

有 $\{p_{k_2}, k_2 \geq 0\}$ 存在,与初始条件无关,且

$$p_{k_2} = \begin{cases} \frac{1}{k_2!} \rho^{k_2} p_0, & 1 \leq k_2 \leq M-1 \\ \frac{1}{M^{k_2-M} \cdot M!} \rho^{k_2} p_0, & k_2 \geq M \end{cases}, \quad p_0 = \left[\sum_{k_2=0}^{M-1} \frac{\rho^{k_2}}{k_2!} + \frac{M \rho^M}{M!(M-\rho)} \right]^{-1} \quad (3-21)$$

因为有 M 台晶圆制造设备,所有晶圆批次到达时需要等待的概率为

$$p = \sum_{k_2=M}^{\infty} P_{k_2} = \frac{1}{1 - \rho_M} p_M \quad (3-22)$$

其中, $\rho_M = \frac{\lambda}{M\mu} < 1, p_M = \frac{\rho^M}{M!} p_0$ 。

为了求得平均队长与平均等待队长,在统计平衡下,等待队长 M_q 呈现分布状态,为

$$P\{M_q = 0\} = \sum_{k_2=0}^M P_{k_2}, \quad P\{M_q = \mu\} = p_{M+\mu}, \quad \mu = 1, 2, \dots$$

故当 $\rho_1 < 1$ 时,有以下关系存在:

$$\begin{aligned}\overline{M}_q &= E[M_q] = \sum_{k_2=M}^{\infty} (k_2 - M) p_{k_2} = \sum_{k_2=M}^{\infty} (k_2 - M) \frac{1}{M^{k_2-M} \cdot M!} \rho^{k_2} p_0 \\ &= \frac{\rho^M p_0}{M!} \sum_{k_2=M}^{\infty} (k_2 - M) \left(\frac{\rho}{M}\right)^{k_2-M}\end{aligned}\quad (3-23)$$

经过方程式的进一步化解,可得

$$\overline{M}_c = \rho - \frac{\rho^M}{(M-1)!} p_0 - \frac{\rho^{M+1}}{M!(1-\rho_M)} p_0 + \frac{\rho^M}{(M-1)!(1-\rho_M)} p_0 \quad (3-24)$$

不难发现,上面的表达式即为 ρ 的泰勒二次展开式,所以 $\overline{M}_c = \rho$ 。

从而得出,当排队等待系统平衡时,正在加工的晶圆批次与晶圆制造设备的个数(M)无关。

由式(3-23)和式(3-24)可得:

$$\text{在某时刻 } t, \text{ 平均队长 } \overline{M} = \overline{M}_q + \overline{M}_c = \rho + \frac{\rho_M}{(1-\rho_M)^2} \rho_M。$$

其次,假设晶圆批次是按照 FIFO(先进先出)的规则进行加工制造,令 $p_{k_2}^-$ 表示晶圆批次到达排队等待系统后发现已有 k_2 个晶圆批次的平稳概率,

则可以推出: $p_{k_2}^- = p_{k_2}, k_2 = 0, 1, 2, \dots$, 当 $\rho = \frac{\lambda}{M\mu} < 1$, 在统计平衡下,晶圆批次的

的排队等待时间可以用分布函数 $W_q(t) = 1 - \frac{p_M}{1-p_M} e^{-\mu(M-\rho)t}$ 来表示,则其平均等待时间为

$$\overline{W}_q(t) = \frac{\rho_M}{\lambda(1-\rho_M)^2} p_M \quad (3-25)$$

在到达的晶圆批次看到已有 $k_2 (k_2 > M)$ 个晶圆批次的条件下,由于 M 台晶圆制造设备均处于正常状态,所以该晶圆批次必须等待 $(k_2 - M + 1)$ 个晶圆批次加工完成后才能被加工。在 M 台晶圆制造设备均处于正常状态的条件下,由于每个晶圆批次的离去均是参数为 μ 的泊松流,这样相继离去晶圆批次的间隔时间服从参数为 $M\mu$ 的负指数分布,所以该晶圆批次的排队等待时间等于 $(k_2 - M + 1)$ 个晶圆批次相继离去的间隔时间总和,其分布为参数 $M\mu$ 的 $(k_2 - M + 1)$ 阶埃尔朗

分布,即 $P\{0 < W_q \leq t |_{M^-=k_2}\} = \int_0^t \frac{M\mu(M\mu x)^{k_2-M}}{(k_2-M)!} e^{-M\mu x} dx$, 由此可以得出

$$W_q(t) = W_0(t) + \sum_{k_2=M}^{\infty} \frac{p_0 \rho^{k_2}}{M! M^{k_2-M}} \int_0^t \frac{M\mu(M\mu x)^{k_2-M}}{(k_2-M)!} e^{-M\mu x} dx$$

$$= 1 - \frac{p_M}{1 - \rho_M} + \frac{p_M}{1 - \rho_M} [1 - e^{-(1-\rho_x)M\mu t}] \quad (3-26)$$

经过简化后可得排队等待时间为

$$W_q(t) = 1 - \frac{p_M}{1 - \rho_M} e^{-\mu(M-\rho)t} \quad (3-27)$$

从而平均排队等待时间为: $\overline{W_q(t)} = \int_0^\infty t dW_q(t) = \frac{\rho_M}{\lambda(1-\rho_M)^2} p_M$ 。

由于逗留时间 $W = W_q + \chi$, 且 W_q 与 χ 独立, 因此根据分布函数的卷积公式可以得出逗留时间的分布函数为

$$W(t) = \begin{cases} 1 - e^{-\mu t} \left[1 + \frac{p_M}{1 - \rho_M} \mu t \right], & \rho = M - 1 \\ 1 - e^{-\mu t} - \frac{p_M}{(1 - \rho_M)(M - \rho_M - 1)} [e^{-\mu t} - e^{-\mu(M-\rho)t}], & \rho \neq M - 1 \end{cases} \quad (3-28)$$

$$\Rightarrow \overline{W} = \overline{W_q} + \frac{1}{\mu} \quad (3-29)$$

根据上面公式推导的定性分析, 可以通过一个实际的举例来进行定量判断。

在某个批次的晶圆加工到某一工艺和工序时, 有两台晶圆制造设备正在独立地并行制造, 且批次的晶圆按照参数 $\lambda = 8$ 批次/min 的泊松流到达, 而每台晶圆制造设备的加工制造时间都服从 $\mu = 5$ 批次/min 的负指数分布。可以比较两种排队方案的运行指标, 并以此判断晶圆制造并行设备的最优控制。

(1) 多制造设备、多队列的排队系统方案。晶圆批次到达后, 以 50% 的概率分成两个队列进行排队等待, 如图 3-2 所示。

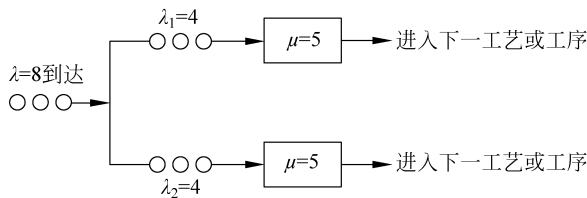


图 3-2 多制造设备、多队列的排队系统方案

在这种情况下, 可以将排队系统方案看成两个独立的 $M/M/1/\infty$ 排队系统, 每个晶圆批次的到达均为 $\lambda_1 = \lambda_2 = 4$ (批次/min) 的泊松流, 且

$$\rho = \frac{\lambda}{\mu} = \frac{4}{5} \Rightarrow p_0 = 1 - \rho = 0.2$$