

基于容时管控的高等级产品智能调度方法

周菁

(上海华力集成电路制造有限公司)

5 **摘要:** 随着集成电路制造行业的日益多元化和复杂化,传统的调度方式已经不能满足晶圆厂高精度、高效率的大规模生产需求。本文通过分析现有自动化系统的调度规则,结合容时管控(Q-Time Control)理论,提出了一种面向高等级产品的智能调度方法,实现对传统调度结果的再优化,在保证产品品质的同时,极大程度地减少了高等级在制品(High-WIP)的

10 等待时长,从而降低产品的生产周期,提升晶圆厂的整体效率。

关键词: 集成电路; 自动化系统; 容时管控; 智能调度; 高等级在制品;

中图分类号: 请查阅《中国图书馆分类法》

Intelligent Scheduling Method for High-Priority Products Base on Allowable Time Control

Zhou Jing

(Shanghai Huali Integrated Circuit Co., Ltd)

15 **Abstract:** With the increasing diversification and complexity of the integrated circuit manufacturing industry, traditional scheduling methods can no longer meet the high-precision and high-efficiency large-scale production needs of wafer fabs. This article proposes an intelligent scheduling method for High-Priority products by analyzing the scheduling rules of existing automation systems and combining Q-Time Control theory, implementing further optimization of traditional scheduling results. While ensuring product quality, it greatly reduces the waiting time for High Priority-WIP, thereby reducing the production cycle of products and improving the overall efficiency of wafer fabs.

20 **Key words:** Integrated circuit; Automation; Q-Time control; intelligent scheduling; High-WIP;

0 引言

30 随着信息技术的迅猛发展,半导体制造生产线自动派工系统的复杂程度也因制程革新呈现出了指数倍增长。在半导体生产流水线中存在着很多特殊工艺工段,需要在规定时间段内(Q-Time Zone)完成一系列的加工步骤,避免晶圆长时间暴露在空气中产生腐蚀、金属损伤等缺陷(Defect)。如后段工艺中铜表面氧化物、氟化物的析出等均会导致芯片良率受到严重影响[1]。

35 现阶段业界主要通过 KANBAN、CONWIP (Constant Work in Process)、DBR (Drum-Buffer-Rope) 等派工策略,在时间窗口外对在制品进行限额控制,保持合理的缓冲[2]。从而保证 Q-Zone 区间内设备有充足的 WIP,同时这些 WIP 又不会超时作业(Over Q-Time)。但是诸如 CONWIP 等策略均未给出获取 Q-Zone 内标准 WIP 的方法。实际流水线中 Q-Zone 区间内的标准 WIP 及缓冲容量是通过多次尝试得出的经验值,其结果无法根据

40 生产线实时状态进行调控。而 Q-Zone 区间内的缓冲容量又关系到产品良率与设备利用率这两个矛盾要素。适度的缓冲容量能够很好地吸收变异,降低设备故障等突发事件带来的超时

作者简介: 周菁(1978-),男,集成电路专业工程师. E-mail: zhoujing@hlmc.cn

作业风险，但也不可避免地会牺牲产能。具体关系如图 0.1 所示。

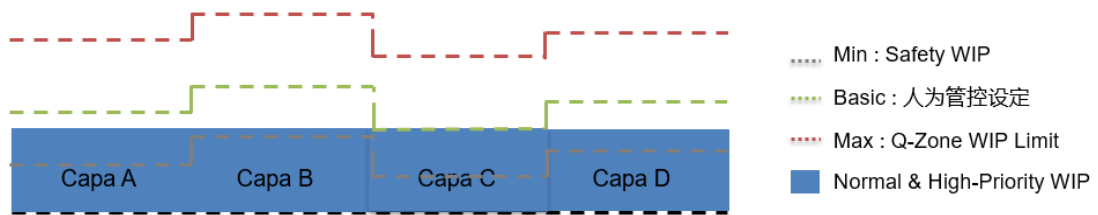


图 0.1 标准 WIP 与缓冲容量

45

现有自动化系统的调度规则是通过特定 Capability 下的高等级产品进行 Exception 管控，具有很大的局限性和风险性。本文通过提出一种面向高等级产品的智能调度方法，在传统派工策略的基础上，结合区间内机器状态、实际 WIP 值等影响因子，对区间标准 WIP 及缓冲容量进行线性拟合，得到适用于高等级产品的扩容模型。具体关系如图 0.2 所示。

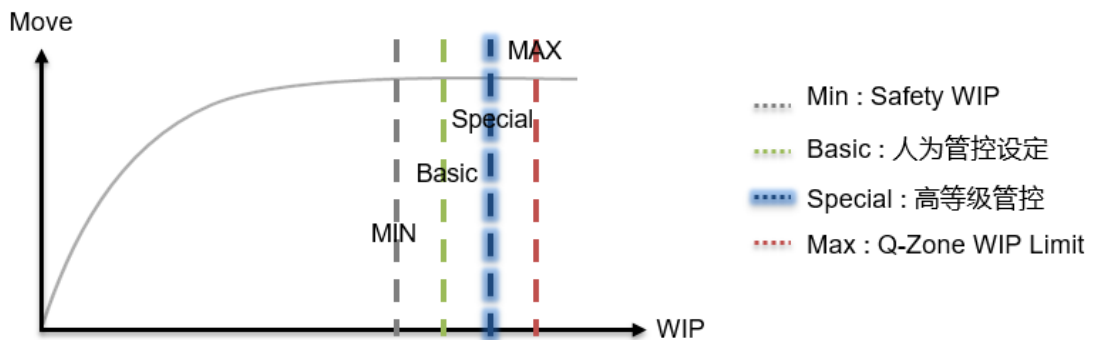


图 0.2 Move-WIP 关系图

50

该模型可以为高等级产品在复杂 Q-Zone 区间内的派工提供参考，通过合理利用缓冲容量，缩短高等级制品的等待时长，使派工策略更有针对性。

1 研究方法

55

1.1 参数选取

在半导体制造流水线中，工厂通常会根据客户需求，交期紧急度 (OTDO)，工艺限定等因素对在制品进行分级管控。对于一些复杂 Q-Zone，受到瓶颈站点的产能限制，其起始站点往往会堆积大量的 WIP；并且实际生产中大量不确定因素如产品组合变化、机器加工能力差异、机器突发故障等也会对生产线造成影响。在这一情况下，如何保证高等级制品的快速流通变得尤为关键。

60

现有派工策略中仅考虑了作业能力 (Capability) 这一变量，不足以满足实际生产需求。因此，本论文模型搭建时，综合考虑了在产品数量 (WIP)、实时生产效率 (WPH)、放货比例 (Basic Ratio)、作业能力 (Capability) 等多个类型的变量，并进行了充分的预处理，从而增强模型的拟合程度，提高模型的准确性。具体使用变量如表 1 所示。

65

表 1 扩容模型使用参数

指标类型	相关变量	单位
在制品数量 (WIP)	安全存量 (Safety WIP)	片
	最大可容许量 (WIP Limit)	片
实时生产效率 (WHP)	理论生产效率 (Peak WPH)	片/h
	派工生产效率 (DSP WPH)	片/h
放货比列 (Ratio)	基础比例 (Basic Ratio)	
	高等级特殊比例 (Special Ratio)	
作业能力 (Capability)	机台制程 (Capability)	

1.2 数据处理

为确保护容后区间内的在制品数量不会超过最大可容许量,在数据收集时需要考虑不同时间段生产线中产品组合、机器状态等差异带来的影响。由于过为久远的数据对模型搭建及验证并无太大参考价值,因此本文使用的数据均来自生产线实时数据。此外,在数据收集过程中,生产线内机器异常、人为干预等情况都会对约束函数构建以及模型计算结果产生一定影响,因此对数据使用 Z-Score^[3]统计法进行处理,并剔除|Z-Score|> 3 的异常数据。由于不同类型的变量存在单位不同、数值区间不同等问题,在进行模型拟合前对数据进行了[0, 1]归一化处理。将以上数据作为模型的输入变量。

1.3 模型搭建

1.3.1 模型搭建难点及应对策略

难点一:判定瓶颈站点是否可进行扩容。

应对策略:根据现有派工系统及生产线实时 MPS 需求,对 Basic Ratio 总和进行判断:若 Basic Ratio 总和 ≥ 1 ,认为该站点放货量已经饱和,无法进行扩容;若 Basic Ratio 总和 < 1 ,认为该站点仍存在可用缓冲容量,可以进行扩容。

难点二:判定扩容后放货量是否会超过最大放货量限制。

应对策略:使用模型计算出可进行扩容站点的 Special Ratio,通过模型优化确保所有可扩容站点的 Special Ratio 均处在缓冲区内,避免过度放货。

1.3.2 模型构建过程

本文通过求解线性约束优化 (NLP)^[4]对变量进行拟合,从而构建出适用于生产线所有高等级 Lot 的模型。以流水线中某一实际 Q-Zone 起始站点为例,利用现行 Q-Zone Result 公式计算出起始站点结果为 Not Enter 的高等级 Lot,并获取这些 Lot 在 Q-Zone 内的瓶颈站点及 Basic Ratio 等信息。通过计算 WIP Limit 与 Basic WIP 的差值即可得到当前瓶颈站点的缓冲容量。

将生产线内各瓶颈站点的缓冲容量进行统计并绘制散点图,发现其分布大致满足以下约束关系:

$$f(x) \leq \varphi_j(x) \tag{1}$$

其中 x 为 Basic Ratio, $\varphi_j(x)$ 为约束基函数, $f(x)$ 为 Special Ratio。

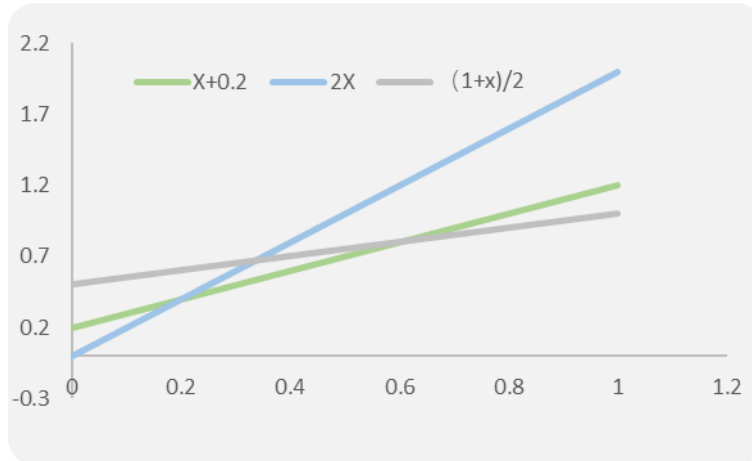


图 1.1 缓冲容量分布约束关系图

对上述线性约束关系进行最小二乘曲线拟合：

$$f^*(x) = \sum_{j=0}^n c_j \varphi_j(x) \tag{2}$$

其中 c_j 为拟合系数。利用离散函数最佳平方逼近问题求解方法，使用 MATLAB 中 Lsqcurvefit 函数计算拟合系数，得到如下最小二乘拟合曲线：

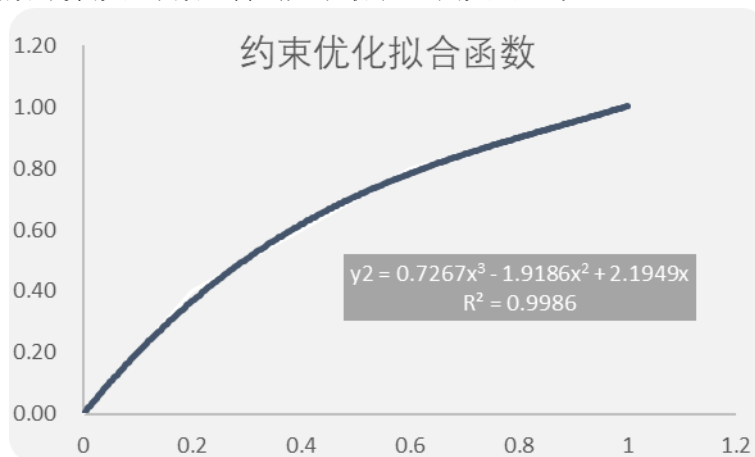


图 1.2 线性约束优化拟合函数曲线图

该拟合曲线充分考虑了大体量数据中噪声带来的影响，因此，在使用上述拟合曲线对 Basic Ratio 进行适当放大后，可得出的一套适用于高等级 Lot 的 Special Ratio，且拟合曲线适用于生产线所有瓶颈站点，能够在保证瓶颈站点放货量不超过最大限制的同时，加速高等级在制品的流通。

表 2 瓶颈站点扩容前后对比

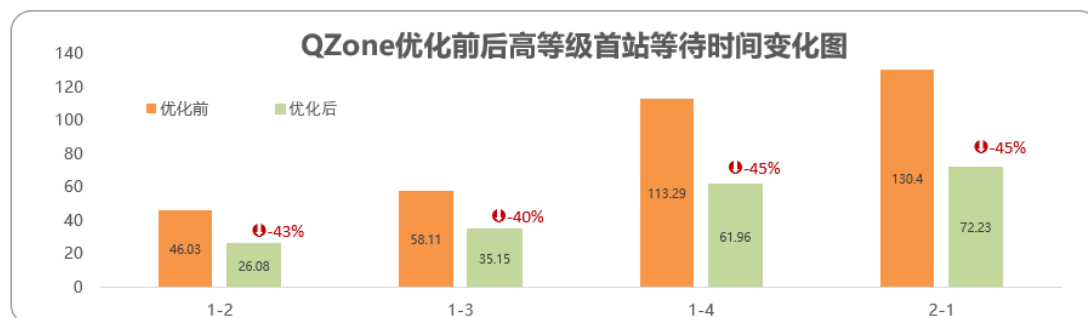
MACHINE	CAPABILITY	BASIC RATIO	$y2 = 0.7267x3 - 1.9186x2 + 2.1949x$	RATIO总和
***	PTH	0.59	0.61	0.98
	PTS	0.35	0.37	

2 研究结果

110 本文以半导体工厂实际生产流水线作为研究对象，分别统计了扩容模型搭建前后高等级在制品在 Q-Zone 首站的等待时间、各工艺站点在制品超时作业次数的详细数据，并进行了对比分析，具体数据如下：

2.1 生产效率

115 图 2.1 为模型搭建前后高等级在制品在 Q-Zone 首站等待时间对比图。从中可以看到，对于 1-2、1-3、1-4、2-1 等级的在制品，扩容模型搭建后 Q-Zone 首站等待时间有明显下降，平均下降幅度高达 44%，大幅度加快了高等级在制品的流通速度。



• 数据采用上线稳定后各等级自动派工Lot首站等待时间平均值

图 2.1 Q-Zone 优化前后高等级首站等待时间变化图

2.2 产品品质

120 此外，由于工艺段 WIP 值与实际生产效率之间的关系大致可分为线性增长区与饱和区 [5]。即在 WIP 较低时，生产效率随 WIP 值增加呈线性增长趋势；而当 WIP 增长到一定量后，生产效率逐渐达到饱和状态，不再随 WIP 增加而增长。因此，为保证产能最大化的同时，增加高等级在制品的流通速度，就不可避免的会增加一定产品品质风险。

125 本文提出的模型相较于传统 Exception 管控方式，在精确性与全局性等方面都有很大提升。因此其不仅在流通速度方面体现出极大优势，在产品品质保障方面也有一定程度的优化效果。如图 2.2 为高等级优化与 Over Q-Time Lot 数量变化趋势图。从图中能明显看到，新模型搭建后，Over Q-Time Lot 数量有明显下降趋势。

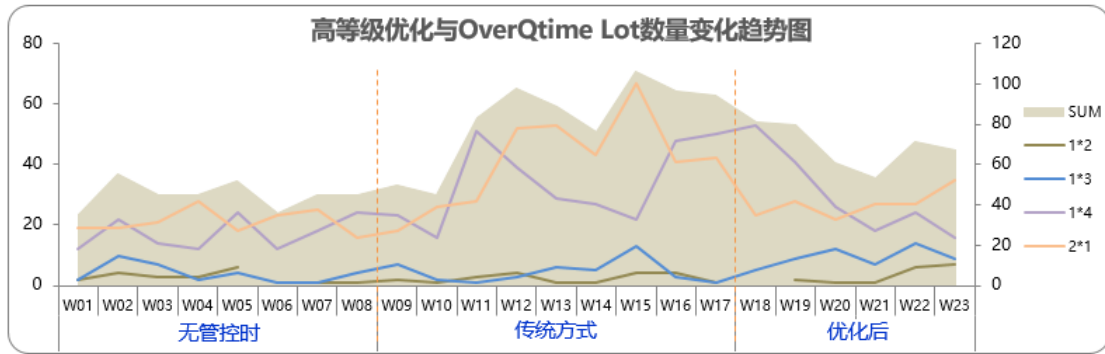


图 2.2 高等级优化与 Over Q-Time Lot 数量变化趋势图

130 **3 结语**

本文通过求解线性约束优化提出了一种面向高等级产品的智能调度方法,利用工厂生产线实际派工数搭建模型,并利用模型得到一套适用于高等级产品的 Special Ratio。

135 数据结果显示,求解线性约束优化可以对生产线复杂的派工数据进行拟合。利用该模型计算出的高等级产品 Special Ratio,有助于提高生产线高等级在制品的流通速度,且相较于传统管控模式,新模型也对降低 Over Q-Time Lot 的数量有很大帮助,为半导体制造行业自动化派工系统提供了重要参考。

[参考文献] (References)

[1] 陈建生, 杨建生. 具有有限等待时间约束的机器调度问题的模型构建[J]. 信息与优化科学杂志, 2006年, 27 (1): 225-240
 140 [2] ENG C K, SIN L K. 基于 CONWIP 的半导体终端流水线控制[J]. 工程学报, 2013, 53: 607-615
 [3] Shalabi L, Shaaban Z, Kasasbeh B. 数据挖掘: 预处理引擎[J], 计算机科学杂志, 2006, 2 (9): 735-739
 [4] 宋玉阶, 吴怀宇. 求解线性约束二次优化问题的神经计算模型. 万方论文网, 2004, 4:73-76
 [5] 范永洁, 王勇. 基于神经网络的半导体生产流水线时间窗口内派工策略优化[J]. 电子技术, 2020, 49(02): 64-66
 145