

基于分类数据的质量控制算法研究

韩 凯, 王凯波

(清华大学 工业工程系, 北京 100084)

摘 要: 分类数据在企业当中常见、易获取。如何利用分类数据进行过程控制、提升质量水平是当前值得关注的课题。本文基于半导体制造背景, 建立了硅片研磨过程模型, 在前馈控制系统中引入分类数据, 提出了一种基于分类数据的前馈控制算法, 并以均方误作为指标, 通过理论推导及仿真来研究其性能。结果表明, 无论在理想状况下还是分类错误存在的情形下, 这种控制算法都能够显著地降低参数波动, 大幅提升硅片质量。

关键词: 统计过程控制; 前馈控制算法; 最小均方误差; 分类数据; 半导体制造

中图分类号: F 224 文章标识码: A 文章编号: 1007-3221(2010)06-0171-05

Process Quality Control Using Categorical Observations

HAN Kai, WANG Kaibo

(Department of Industrial Engineering, Tsinghua University, Beijing 100084, China)

Abstract Categorical observations are common and easily captured. Much of current work focuses on process control and quality improvement with categorical observations. This article sets up a mathematical model of wafer lapping and presents a novel algorithm by introducing categorical observations to feed-forward control system. In addition, we verify the capacity of this algorithm by simulation study, using MSE as the criterion. Results show an outstanding performance under the condition without or with misclassification of wafer thickness.

Key words statistical process control; feed-forward control; minimum mean sum of square; categorical data; semiconductor manufacturing

0 引言

分类数据在实际的生产生活当中十分常见。在企业中, 在质量控制过程中, 一般采用连续的、精确的数据, 事实上相对于这种定量数据, 分类数据有其优势所在: 分类数据的测量和记录都较为简单; 分类数据也很容易转化成货币单位, 因此有利于管理决策的制定, 便于统计人员之间的交流^[1]; 相对于连续变量的检测过程, 分类数据的测量占用的时间会少很多, 它对于人员的技术要求、对工具的要求也会降低很多。这些都能在很大程度上降低企业的成本。

在半导体制造行业中分类数据同样有很大的应用价值。硅片制造过程十分复杂, 包括拉制、滚磨、切片、脱胶、倒角、研磨、腐蚀、喷砂、抛光、封装等工序。其中, 研磨工序是切片后对硅片表面的第一次机械加工, 具有多重作用, 如使硅片达到一定厚度规格、缩小硅片之间的厚度差异、降低硅片本身的弯曲与翘曲、改善硅片表面平整度、去除刀痕等等^[2]。生产经验和实验都表明, 研磨工序基本上奠定了硅片的质量水平, 在整个生产环节属于关键工序。在企业现有生产水平和条件下, 常用抽样的方式确定一批硅片的厚度

收稿日期: 2010-01-09

基金项目: 国家自然科学基金资助项目 (70802034)

作者简介: 韩凯, 男, 硕士研究生; 王凯波, 男, 副教授, 主要研究方向: 统计质量控制、数据驱动的复杂制造系统建模/监视/控制与诊断。

© 1994-2011 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

情况, 决定研磨的去除量。如此操作对批内方差较大的一批硅片并不合适, 研磨后的厚度偏差也会较大。这样的策略造成的后果是研磨后硅片实际厚度与目标厚度差异较大, 整体波动也很大。而在企业里获取分类数据较为容易, 不需要动用昂贵的设备和消耗大量的能源, 花费的时间也较短。如果以低成本获取分类数据, 并借以调节研磨过程, 达到减小厚度波动、提升硅片整体质量水平的目的, 企业也可以获取相当大的经济效益。本文所关注的正是将基于分类数据的统计分析手段应用于研磨过程中的硅片质量控制。

如果能够在研磨过程开始之前获取分类数据, 那么可以建立一种前馈控制算法, 通过对这些分类数据的分析决定研磨过程的一些参数。一般来说, 前馈控制系统具有以下几个特点^[3]: (1)控制策略确定于控制过程开始前, 与输出无关, (2)控制及时、有效, (3)稳定性好、容易建立, (4)适用于输入已知的情况, (5)过程中有外部扰动的情况下效果不佳。前馈控制很早就应用于半导体制造行业当中, 如提高光刻过程的精度^[4], 减小 $R\mathcal{R}$ 过程的变动^[5]。

在本文案例中, 硅片研磨的目标厚度是确定的, 并且研磨技术水平基本稳定, 外部干扰(比如室温、人员)的影响可以忽略, 适合应用前馈控制。

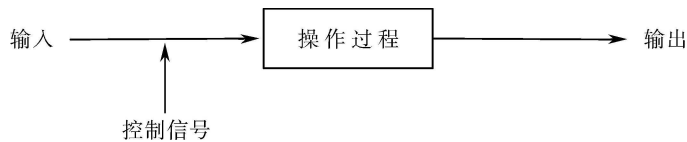


图 1 前馈控制示意图

对分类数据的统计分析、前馈控制两方面的研究工作有很多, 但还没有工作讨论将两者结合起来的情况。本文正是在硅片制造背景下, 建立了硅片研磨过程模型, 提出了一种新的基于分类数据的前馈控制算法, 并通过仿真展示这种算法在理想状态以及分类错误存在情况下的性能。

1 系统建模与控制算法开发

1.1 系统建模

如图 2 所示, 在实际生产中, 研磨过程一般可以分为以下几个基本步骤。

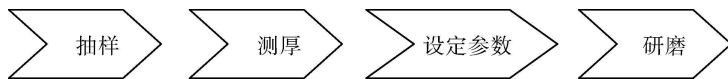


图 2 研磨工序基本步骤

- (1) 研磨前操作员首先从一盘(共数十片)硅片当中随机抽取一片;
- (2) 用硅片测厚仪测量当前硅片的厚度, 减去硅片的目标厚度, 即为本次研磨的去除量;
- (3) 根据设备的去除速度及算出研磨时长, 设置研磨过程各项参数;
- (4) 进行研磨。

研磨生产中每盘可以认为是一个批次, 每一盘加工参数的设置或者说加工作用对盘内所有硅片的作用是相同的。再者, 研磨过程属于大批量生产。一台普通研磨机床一天可以加工硅片数千片。因此, 本模型考虑的是大量生产情况下针对批次实施的控制策略。本节将基于现在的研磨过程建立统计模型, 并从理论上分析其控制效果。硅片厚度的分类值属于有序分类数据, 因此本文所研究的内容适用于有序分类数据。

若一盘硅片目标厚度为 T , 过程控制因子设为 u , 假设硅片研磨前厚度为 y , 研磨后的厚度为 y' , 在正交试验设计的基础上, 经过回归分析, 我们获得了研磨过程的模型如式(1)

$$y' = a + by + u + \varepsilon \quad (1)$$

其中 a, b 为常数, ε 为研磨过程中的噪音因子, 这里设 $\varepsilon \sim (0, \sigma_\varepsilon^2)$ 。

研磨过程的目的之一即将所在硅片研磨至目标厚度, 对此有

$$E(y') = E(a + by + u + \varepsilon) = a + bE(y) + u = T \quad (2)$$

由此可以得到控制因子的表达式为

$$u = T - a - bE(y) \quad (3)$$

控制策略的效果是期望研磨后的厚度与目标值之间的方差即均方误差越小越好, 即控制算法的效果可

以通过磨后厚度与目标厚度的均方误 (MSE, mean square error) 体现出来, 均方误的定义为

$$MSE = E((y' - T)^2) \tag{4}$$

展开, 即

$$\begin{aligned} MSE &= E((a + by + u - T)^2 + 2(a + by + u - T) + \varepsilon^2) \\ &= E((a + by + u - T)^2 + 0 + \sigma_z^2) \\ &= u^2 + 2uE(a + by - T) + E(a + by - T)^2 + \sigma_z^2 \end{aligned}$$

为求最大值, (5)式对 u 求导, 则控制因子的最优值为

$$\begin{aligned} 0 &= 2u + 2E(a + by - T) + E(a + by - T)^2 \\ u &= T - a - bE(y) \end{aligned} \tag{5}$$

(6)式中 u 的表达与 (3)一致, 即均方误最小、达到目标厚度这两种要求可以同时达到。因此可以将代入目标函数即式 (5), 则得到目标函数的最优值

$$MSE = E((by - bE(y))^2 + \sigma_z^2) = b^2\sigma_y^2 + \sigma_z^2 \tag{6}$$

1.2 基于分类数据的控制算法

由于需要获取分类数据, 研磨过程需要作一些变化, 如图 3 所示。



图 3 研磨工序基本步骤

基于分类数据的前馈控制算法示意图如下所示。

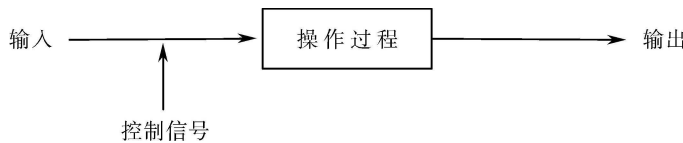


图 4 基于分类数据的前馈控制示意图

类似地, 目标函数为

$$MSE = E((y' - T)^2 | Y) \tag{7}$$

展开可得
$$MSE = u^2 + 2uE(a + by - T | Y) + E(a + by - T | Y)^2 + \sigma_z^2$$

不难得到控制因子最优值

$$u = T - a - bE(b | Y) \tag{8}$$

代入目标函数, 可得

$$MSE = E((by - bE(b | Y))^2 + \sigma_z^2) = b^2 Var(y | Y) + \sigma_z^2 \tag{9}$$

下面给出基于分类变量 Y 的控制因子的表述。厚度值被划分到第 $j(j=1, 2, \dots, m)$ 类, 可以表述为

$$Y = j \Leftrightarrow s_{j-1} < y \leq s_j \tag{10}$$

其中, s_0, s_1, \dots, s_{m-1} 为分类点, 将值域 (s_0, s_m) 分成 m 个小的区域。如果 $Y = j$ 不妨认为此时 y 满足以 (s_{j-1}, s_j) 的中心为中心的正态分布^[6], 则有

$$E(y | Y = j) = \frac{s_{j-1} + s_j}{2} \tag{11}$$

因此, 控制因子 u 可以表述为

$$u_j = T - a - b \frac{s_{j-1} + s_j}{2} (j=1, 2, \dots, m) \tag{12}$$

2 性能研究

2.1 性能研究

本文基于以上模型建立了仿真程序, 验证基于分类数据的前馈控制算法的性能。由于企业信息保密要求, 这里去除现实生产中的真实厚度范围及量纲。假定研磨目标厚度为 0 设置类别数目为 6 输入厚度

满足正态分布,其总体均值为 10Q,方差为 1Q,仿真过程共 200步,即硅片数量为 20Q,此外,需要补充说明的是:

- (1)按照传统控制策略进行研磨时,每盘的去量根据样本均值的厚度决定;
- (2)按照新型控制策略进行研磨时,可以认为大批量生产条件下,有充足的硅片使得每一盘中所有硅片都属于同一个厚度档位。

为更为明确地观察新型算法的表现,本文同时列出了未应用新型前馈控制算法时的相应参数。通过仿真得到的磨后厚度参见图 5和图 6 其中横轴为硅片序列,纵轴为硅片磨后厚度。

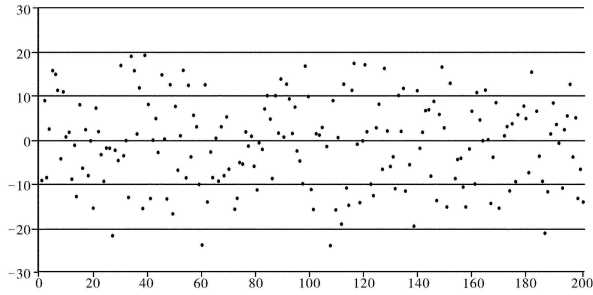


图 5 未应用新型算法时的磨后厚度

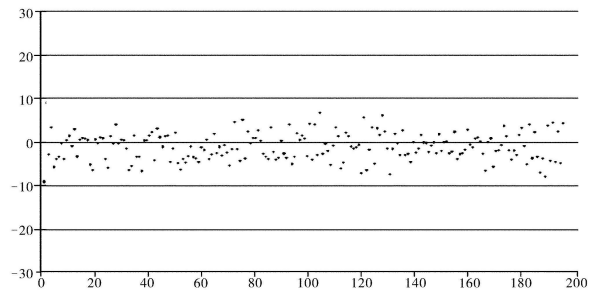


图 6 应用新型算法时的磨后厚度

可以明显的看到,未应用前馈算法时,磨后厚度值基本分布在 (-3Q, 3Q),而应用新型算法时的磨后厚度分散在 (-1Q, 1Q)范围内,过程更加稳定,波动减小,产品质量有所提高。

下面重点关注本文提出的新型算法对均方误的改善。表 1 给出了前 100片和前 200片硅片对应的方差值。

可以明显地看到,基于分类数据的控制算法能够明显地降低均方误,并且使得均方误能够较快地进入稳定状态。因此,分类控制策略可以获取显著优越的控制效果。

表 1 新型控制算法对均方误的改善

硅片数目	未应用新型控制算法	应用新型控制算法
100	99.7	11.9
200	109.1	13.4

2.2 有分类错误情况下的性能研究

本文前面讨论的情况基于一个前提:分类数据是正确无误的。在生产中,分类操作人员可以根据已知的厚度变量的分布,计算得到各个分类点的值,并以此作为分类的依据。但是若考虑到分类操作人员个人分辨能力问题,既定的分类点可能不再准确,而是在准确的分类点附近发生一些浮动,因而会出分类错误的情况。先前的一些工作对分错概率本节仅对分类错误存在情况下的算法性能进行探讨,对分错概率不作讨论,对此可参考其它一些研究^[7,8]。

由于分类点存在误差,分类过程发生了波动,实际过程可以表述为

$$Y^* = j \Leftrightarrow s_{j-1} + d_{j-1} < y \leq s_j + d_j \quad (j = 1, 2, \dots, m) \tag{13}$$

其中 $d_j (j = 1, 2, \dots, m)$ 为分类过程中由于分类人员的技能水平、辅助仪器性能等造成的噪音因子。假设

$$d_1 = d_2 = \dots = d_{m-1} = d \sim N(0, \sigma_d^2)$$

则分类点的真实值为 $\{s_1^*, s_2^*, \dots, s_{m-1}^*\}$, 其中

$$s_j^* = s_j + d (j = 1, 2, \dots, m - 1)$$

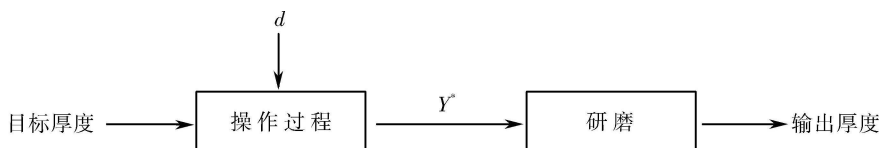


图 7 分类错误存在时的控制模型

式(12)相应地为

$$E(y|Y=j) = \frac{s_{j-1}^* + s_j^*}{2} \quad (14)$$

控制因子的表达式为

$$u_j = T - a - bE(y|Y^* = j) = T - a - b \frac{s_{j-1}^* + s_j^*}{2} \quad (j=1, 2, \dots, m) \quad (15)$$

当前情况下, 目标函数均方误差表达式为

$$\begin{aligned} MSE^* &= E(by - bE(y|Y^*))^2 + \sigma_z^2 = b^2 E \left[y - \frac{s_{j-1}^* + s_j^*}{2} \right]^2 + \sigma_z^2 \quad (j=1, 2, \dots, m) \\ &= b^2 E \left[y - \frac{(s_{j-1} + d) + (s_j + d)}{2} \right]^2 + \sigma_z^2 = b^2 E \left[\left[y - \frac{s_{j-1} + s_j}{2} \right] - d \right]^2 + \sigma_z^2 \\ &= b^2 (Var(y|Y) + Var(d)) + \sigma_z^2 = b^2 Var(y|Y) + b^2 \sigma_d^2 + \sigma_z^2 \end{aligned}$$

基于现有实际数据, 依照各个参数的相对大小, 在仿真程序中取 $d \sim N(0, 2^2)$ 。前 100 片与前 200 片的均方误差, 如表 2 新型控制算法对均方误差的改善所示。不难看出, 应用新型控制算法时均方误差远远小于较未应用新型算法的情况, 且更为稳定, 因此对于实际生产可能出现的分类错误, 此控制算法并不敏感性, 依然能够较大程度提高研磨过程质量。

表 2 新型控制算法对均方误差的改善

硅片数目	未应用新型控制算法	应用新型控制算法	应用新型控制算法 (分类出错)
100	99.7	11.9	14.1
200	109.1	13.4	14.9

3 结论

分类数据分析方面的研究已经有很长时间, 如何将分类数据应用于过程控制也是很多工作关注的焦点, 而本文尝试在前馈控制系统中引入分类数据, 建立了新的基于分类数据的质量控制算法。

在半导体制造行业, 质量的细小提升可以给企业带来巨大的收益, 尤其当前, 中国在硅片制造领域的竞争力仍然有待加强。本文基于硅片制造中的研磨过程, 提出了一种基于分类数据的前馈控制算法, 并推导出衡量质量的指标均方误差的表达式。接着本文通过仿真研究的新算法的性能, 发现新算法可以大幅降低均方误差, 显示了优越的控制效果。本文提出的基于分类数据的前馈控制算法是有效利用分类数据提升质量水平的良好范例, 对类似的生产系统有较高的参照意义。

参考文献:

- [1] Joseph V, Wu C. Failure amplification method[J]. Technometrics, 2004, 46(1): 1-12
- [2] 刘玉岭. 超大规模集成电路衬底材料性能及加工测试技术工程[M]. 北京: 冶金工业出版社, 2002
- [3] Ogata K. Modern control engineering[M]. 第四版. 北京: 清华大学出版社, 2006
- [4] Pai P F, Lin K P. Application of hybrid learning neural fuzzy systems in reliability prediction[J]. Quality and Reliability Engineering International, 2006, 22(2): 199-211.
- [5] Rueggsegger S, Wagner A, Freudenberg J S, et al. Feedforward control for reduced run-to-run variation in microelectronics manufacturing[J]. IEEE Transactions on Semiconductor Manufacturing, 1999, 12(4): 493-502
- [6] Wang K, Tsung F. Run-to-run process adjustment using categorical observations[J]. Journal of Quality Technology, 2007, 39(4): 312-325.
- [7] Shang Y, Wang K, Tsung F. An improved run-to-run process control scheme for categorical observations with misclassification errors[J]. Quality and Reliability Engineering International, 2009, 25(4): 397-407.
- [8] Tzavidis N, Lin Y. Estimating from cross-sectional categorical data subject to misclassification and double sampling: Moment-based, maximum likelihood and quasi-likelihood approaches[J]. Journal of Applied Mathematics and Decision Sciences, 2006, 1: 1-13.