基于局部与非局部线性判别分析和高斯混合模型动态集成的晶圆表面缺陷探测与识别

余建波 [1] 卢笑蕾 [1] 宗卫周 [1]

摘 要 在复杂的半导体制造过程中, 晶圆生产经过薄膜沉积、蚀刻、抛光等多项复杂的工序, 制造过程中的异常波动都可能导致晶圆缺陷产生. 晶圆表面的缺陷模式通常反映了半导体制造过程的各种异常问题, 生产线上通过探测和识别晶圆表面缺陷, 可及时判断制造过程故障源并进行在线调整, 降低晶圆成品率损失. 本文提出了基于一种流形学习算法与高斯混合模型动态集成的晶圆表面缺陷在线探测与识别模型. 首先该模型开发了一种新型流形学习算法 — 局部与非局部线性判别分析法 (Local and nonlocal linear discriminant analysis, LNLDA), 通过融合数据局部/非局部信息以及局部/非局部惩罚信息, 有效地提取高维晶圆特征数据的内在流形结构信息, 以最大化数据不同簇样本的低维映射距离, 保持特征数据中相同簇的低维几何结构. 针对线上晶圆缺陷产生的随机性和复杂性, 该模型对每种晶圆缺陷模式构建相应的高斯混合模型 (Gaussian mixture model, GMM), 提出了基于高斯混合模型动态集成的晶圆缺陷在线探测与识别方法. 本文提出的模型成功地应用到实际半导体制造过程的晶圆表面缺陷在线探测与识别, 在 WM-811K 晶圆数据库的实验结果验证了该模型的有效性与实用性.

关键词 半导体制造, 晶圆缺陷, 模式识别, 流形学习, 高斯混合模型

引用格式 余建波, 卢笑蕾, 宗卫周. 基于局部与非局部线性判别分析和高斯混合模型动态集成的晶圆表面缺陷探测与识别. 自动化学报, 2016, **42**(1): 47−59

DOI 10.16383/j.aas.2016.c150311

Wafer Defect Detection and Recognition Based on Local and Nonlocal Linear Discriminant Analysis and Dynamic Ensemble of Gaussian Mixture Models

YU Jian-Bo[1] LU Xiao-Lei[1] ZONG Wei-Zhou[1]

Abstract The manufacturing process of wafers is very complicated, which involves thin film deposition, chemical mechanical polishing, etching, etc. Abnormal variation in the manufacturing processes could result in the generation of defect patterns on wafers. On the other hand, the defect patterns on the wafer usually indicate the abnormal sources existing in the manufacturing process. Therefore, wafer defect recognition plays a key role in finding the root causes of the out-of-control process and improving yield. This paper develops a model for wafer defect detection and recognition based on manifold learning and dynamic ensemble of Gaussian mixture models (GMMs). A new manifold learning algorithm called local and nonlocal linear discriminant analysis (LNLDA) is developed, which aims to preserve the geometric structure of samples that belong to the identical cluster in the low dimensional space while maximizing the distance of the projected samples from different clusters. Due to the randomness and complexity of wafer defect generation, a dynamic ensemble scheme for GMMs is proposed to detect and recognize various defect patterns occurring in wafers. The proposed model is applied to a real-world semiconductor manufacturing process. Experimental results on WM-811K wafer bin map database approve the effectiveness of the proposed model.

Key words Semiconductor manufacturing, wafer defect, pattern recognition, manifold learning, Gaussian mixture model (GMM)

Citation Yu Jian-Bo, Lu Xiao-Lei, Zong Wei-Zhou. Wafer defect detection and recognition based on local and nonlocal linear discriminant analysis and dynamic ensemble of Gaussian mixture models. *Acta Automatica Sinica*, 2016, **42**(1): 47−59

半导体制造是一个非常复杂的动态制造过程, 晶圆生产需经过薄膜沉积、蚀刻、抛光等多道复杂的工序. 工序异常可能导致晶圆缺陷产生, 比如薄膜沉积过程中的不均匀或者退火处理中温度的不均匀

收稿日期 2015-05-18 录用日期 2015-09-17
Manuscript received May 18, 2015; accepted September 17, 2015
国家自然科学基金项目 (51375290, 71001060), 上海市教育委员会科研创新项目 (13YZ002), 中央高校基本科研业务费专项资金资助
Supported by National Natural Science Foundation of China (51375290, 71001060), Innovation Program of Shanghai Municipal Education Commission (13YZ002), the Fundamental Research Funds for the Central Universities
本文责任编委 胡昌华
Recommended by Associate Editor HU Chang-Hua

1. 同济大学机械与能源工程学院 上海 201800
1. School of Mechanical Engineering, Tongji University, Shanghai 201800

分布都容易造成晶圆表面 Center 和 Edge-ring 缺陷[1]. 因此, 准确识别晶圆在线制造过程中的各种缺陷模式, 不仅可以避免因大批量晶圆缺陷而造成的成本损失, 而且可辅助制造过程异常源的识别, 加速半导体制造过程的异常调整, 减少晶圆生产中的返工率和废品率, 提高晶圆质量和生产效率.

早期晶圆表面缺陷探测大多采用基于统计分析的方法, 比如缺陷密度分布的估计[2]、缺陷聚类分析[3]. 这些方法往往只集中于缺陷模式的统计分析, 无法进行具体缺陷模式的分类. 近些年来, 机器学习方法成功地应用于半导体制造过程中晶圆缺陷检测与识别, 如自适应共振理论[4] (Adaptive resonance theory, ART)、支持向量机[5−6] (Support vector machine, SVM). 但是这些分类器在离线建模阶段识别模型已固定下来, 无法解决线上产生的新晶圆缺陷模式的探测与识别问题.

晶圆图像数据具有高维性、高噪音等特点, 严重影响了缺陷识别器的性能. 数据投影方法 (Data projection) 能有效地消除数据冗余信息, 可有效提取表达数据模式的低维特征. Chen 等[7] 采用主成分分析 (Principal component analysis, PCA) 检测图像表面纹理缺陷, 将空域的表面纹理图像投影到主元子空间, 通过截取主成分 (Truncated component solution, TCS) 来近似表示表面纹理, 将原始图像与 TCS 相减获得纹理表面的缺陷. Tsai 等[8] 采用独立主成分分析 (Independent component analysis, ICA) 方法识别液晶显示屏表面的周期性缺陷, ICA 可有效地分离高度相关的信号, 能够提取出相同周期的缺陷的特征. 虽然这些特征提取方法在缺陷检测中取得较好的效果, 但它们通常只能提取数据的全局结构信息 (比如 PCA), 对于非线性数据集, 全局映射很难提取出高维空间中内在流形结构的信息[9].

研究表明高维空间中数据点位于或近似位于外部空间的子流形上, 流形学习[10−12] 通过降维处理, 可揭示数据流形分布, 找出重要的低维内嵌数据结构. 因此相比于 PCA 等传统数据降维方法, 流形学习可挖掘数据更多有效的内部信息[13]. 流形学习已成功地应用于语音信号处理、图像处理[12] 等领域, 但在半导体制造领域却鲜有应用, 因此本文将流形学习方法应用到晶圆表面缺陷的探测与识别.

本文提出一个晶圆表面缺陷在线探测与识别模型, 该模型首先开发了一种新的流形学习算法 — 局部与非局部线性判别分析法 (Local and nonlocal linear discriminant analysis, LNLDA), 提取晶圆表面缺陷特征. LNLDA 通过融合数据局部/非局部信息以及局部/非局部惩罚信息, 最大化数据不同簇样本的低维映射距离, 保持特征数据中相同簇的低维几何结构, 提高其数据模式的判别性能. 相比较现有常用的数据投影算法, LNLDA 有效地提取高维晶圆图像特征数据的低维流形特征, 提高了晶圆缺陷识别系统的识别性能, 同时这也是流形学习方法在晶圆缺陷识别应用的一种探索与尝试. 由于晶圆表面缺陷产生的随机性和复杂性, 为了能够稳定地识别线上的晶圆缺陷模式, 该模型对每种晶圆模式构建特定的高斯混合模型 (Gaussian mixture model, GMM), 并提出基于高斯混合模型的动态集成方法, 在线探测与识别晶圆表面缺陷. 因此本文提出晶圆表面缺陷在线探测与识别模型的最主要贡献在于: 1) 一个新流形学习算法 (LNLDA) 解决维度缩减与特征提取问题; 2) 基于高斯混合模型的动态集成方法解决缺陷模式的探测与识别问题. 实验结果表明, 本文提出的基于 LNLDA 和 GMM 动态集成的方法能很好地应用于晶圆制造过程监控与诊断.

1 晶圆表面缺陷探测与识别模型

本文提出的晶圆表面缺陷在线探测与识别模型方案如图 1 所示, 包括离线建模和在线检测两大模块. 在离线建模阶段, 首先对收集的晶圆图进行图像预处理, 采用图像滤噪技术突出晶圆模式特征. 在原始特征产生模块提取晶圆图像的几何、灰度、纹理与投影特征. 接着采用 LNLDA 对原始特征数据进行特征提取. 针对每种晶圆模式的特征数据, 分别构建对应 GMM 识别模型. 在线检测阶段, 线上采集的晶圆图通过图像预处理、原始特征产生、特征提取模块进行处理, 被模型判定为缺陷模式的晶圆则进入新缺陷模式检测模块, 如果该缺陷晶圆属于已有缺陷模式, 则输入到 GMM 识别库进行缺陷模式识别. 若判定为新缺陷模式, 则收集新缺陷模式晶圆纳入晶圆数据库. 对每种新缺陷模式构建相应的 GMM 识别模型, 更新 GMM 识别库, 用于识别生产线上产生的新缺陷模式.

1.1 图像滤噪

由于半导体制造过程中多源随机因素的影响, 晶圆图通常混杂各种噪声. 本文首先采用滤噪技术对晶圆图进行预处理, 有效保留晶圆图中缺陷模式, 提高模型探测与识别的性能. 对晶圆图进行滤噪的方法通常有空域滤噪[14] 和最近邻滤噪[15] 法. 区别于传统的频域滤噪处理, 空域滤噪法直接对图像像素进行处理. 最近邻滤噪法则是通过选择合适的近邻值确定聚类缺陷中的缺陷晶粒数, 消除晶圆表面单个随机晶粒缺陷. 本文采用非线性空域滤噪法 (即中值滤噪), 利用像素邻域内灰度的中值代替该像素的值. 图 2 呈现了滤噪前后两种晶圆缺陷模式的表面, 可以看出滤噪处理有效地突显了晶圆表面缺陷模式.

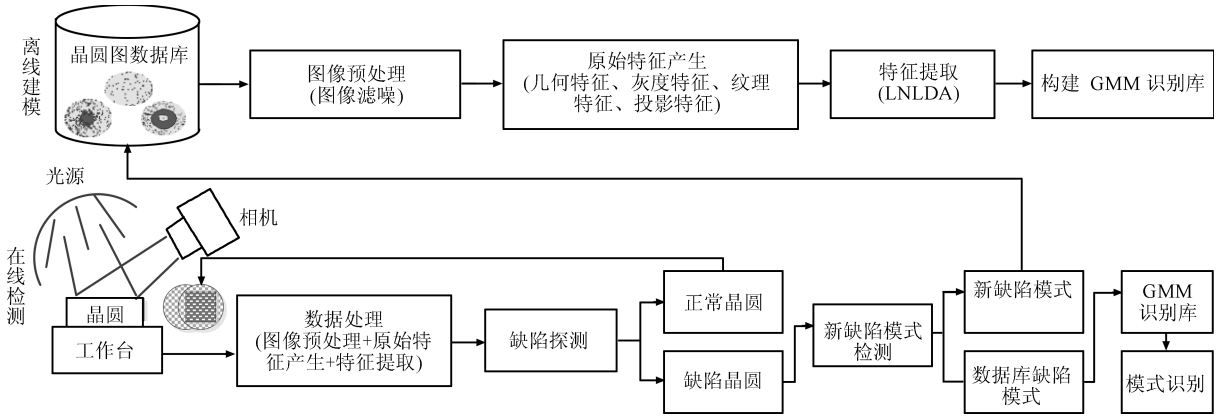


图 1　晶圆表面缺陷探测与识别方法

Fig. 1　Framework of the wafer defect detection and recognition method

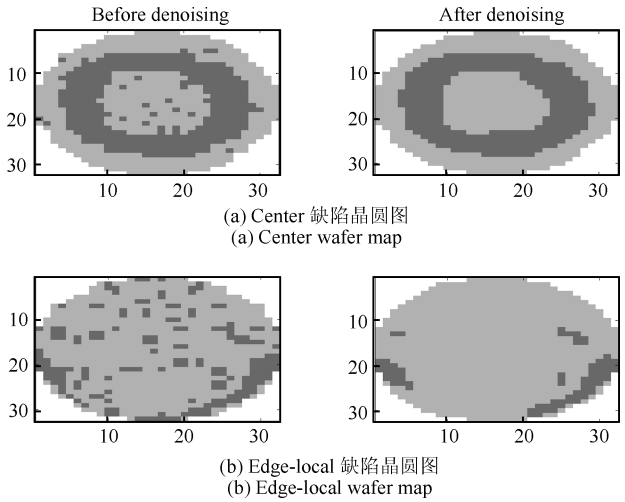


图 2　晶圆图中值滤噪结果

Fig. 2　Median filtering on wafer map

1.2　原始特征产生

从图像中提取特征能够有效地体现模式的特性, 有利于晶圆缺陷模式的识别. 因此, 本文采用四种特征产生方法来表征晶圆缺陷模式的特征: 描述形状、大小等性质的几何特征, 描述灰度分布状态的灰度特征、纹理特征和投影特征. 表 1 列出了本文选择的特征集, 总特征维度为 53, 其中几何特征数目为 18, 投影特征数目为 24. 区域特征包括缺陷边界周长、面积、重心坐标、缺陷紧凑性、矩形度、区域占空比以及与区域具有相同标准二阶中心距的椭圆的长短轴长度和离心率. 线性特征即缺陷区域中的直线数目, 通常采用霍夫变换 (Hough transform)[16] 提取缺陷目标中的直线数目作为线性特征. 霍夫变换的实质就是对图像进行坐标变换, 将图像空间的点映射到参数空间, 通过计算累计结果的局部最大值对应的参数集合作为霍夫变换结果. 图 3 展示了利用霍夫变换提取晶圆缺陷中的直线. Radon 变换[6] 可对数字图像矩阵在某一指定角度射线方向上做投影变换, 图 4 展示了六种晶圆缺陷模式图与对应的各个角度的 Radon 变换结果, 其中横坐标为从 0° 到 180° 的角度值. 基于 Radon 变换本文分别提取晶圆图像中目标区域在 0°、90°、45° 和 135° 四个方向上的投影特征.

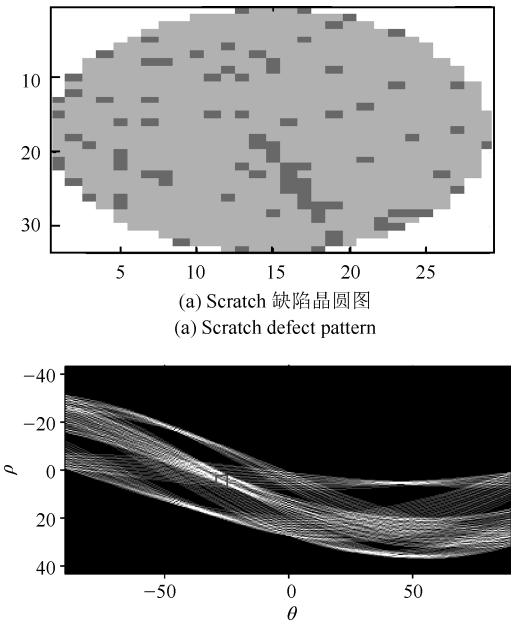
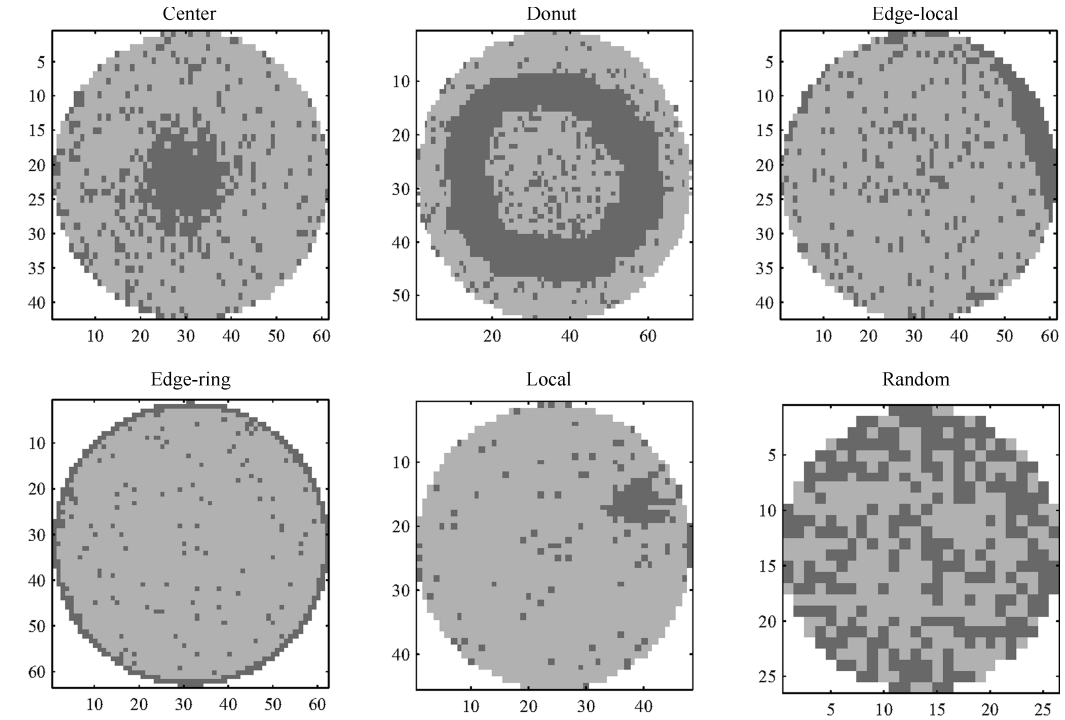


图 3　霍夫变换检测直线

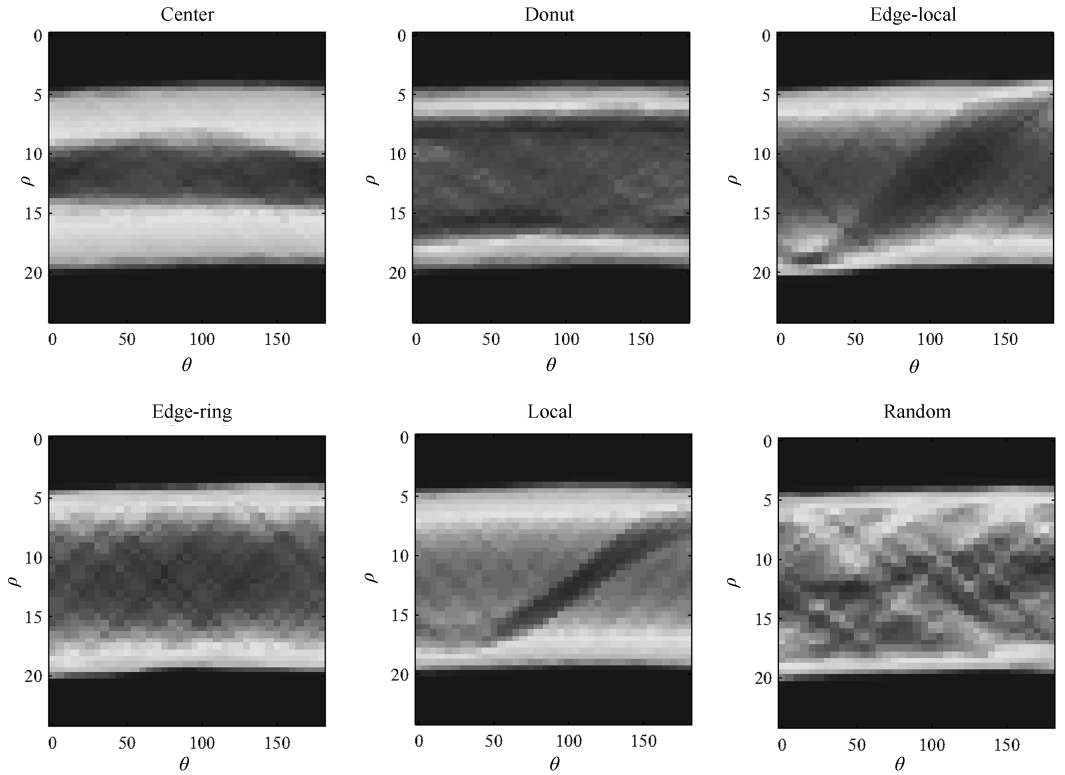
Fig. 3　Line detection by using Hough transform

1.3　局部与非局部线性判别分析算法

流形学习通过有效地提取高维数据空间中的流形结构, 对高维数据集进行降维, 获取数据集中有意



(a) 六种晶圆缺陷模式
(a) Examples of six wafer defect patterns



(b) Radon 变换
(b) Radon transform

图 4 晶圆缺陷模式图与对应的 Randon 变换

Fig. 4 Wafer defect patterns and their corresponding Radon transform

义的低维结构信息. 流形学习的方法通常可以分为两类: 一类是全局方法 (比如等距映射[10]), 数据投影时将流形上邻近点映射到低维空间中邻近点, 同时流形上距离远的点映射到低维空间中距离远的点; 另一类是局部方法 (如局部线性嵌入法[11]), 只保证将流形上邻近点映射到低维空间中的邻近点. 目前, 流形学习已经成为机器学习界的一个热门研究方向, 并已成功地应用于语音信号处理、图像处理[12] 等领域.

表 1　晶圆缺陷特征集

Table 1　Features of defect wafer pattern

特征类别	特征集
几何特征	区域特征、线性特征、Hu 不变矩
灰度特征	平均值、方差、歪斜度、峰值、能量、熵
纹理特征	能量、对比度、相关性、均匀度、熵
投影特征	峰值、平均幅值、均方根幅值、投影波形特征 投影峰值、投影脉冲

针对晶圆图像数据高冗余特征, 以及维度缩减与特征提取需求, 本文提出一种新的流形学习算法—LNLDA. LNLDA 结合局部流形学习算法的优点, 通过融合数据局部/非局部信息以及局部/非局部惩罚信息, 保证流形上同簇样本映射到低维空间中的邻近点, 同时最大化高维数据不同簇样本的低维映射距离. 相比于常规的集中于局部信息保持的流形学习算法 (如 LPP), LNLDA 算法能有效地提取数据中隐藏的低维流形结构信息, 显著提高算法的数据模式判别性能.

LNLDA 通过融合数据局部/非局部信息以及局部/非局部惩罚信息, 提取数据中隐藏的低维流形结构信息. 图 5 呈现了 LNLDA 算法考虑的数据内嵌结构信息, 即基于数据局部和非局部结构信息, 构建对应的局部/非局部图和局部/非局部惩罚图. 图 5 (a) 描述了数据局部与非局部信息, 对于类别 A, 其中相连接的点表示该类的局部信息, 周围的散点与类别 B 中的点则属于非局部信息. 图 5 (b) 描述了数据局部与非局部惩罚信息, 局部惩罚用来描述属于同类非近邻的数据, 如类别 A 中相连接的点. 非局部惩罚则描述近邻非同类的数据, 如类别 A 和类别 B 虚线相连接的点. 因此, LNLDA 通过构建局部/非局部图和局部/非局部惩罚图, 以有效地提取数据中的内嵌流形信息.

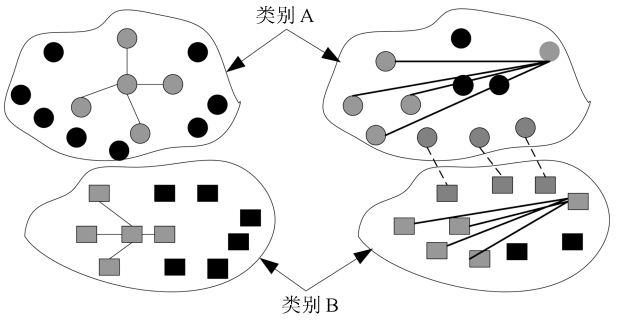
1.3.1　局部 / 非局部结构描述

假设一组 M 维样本数据 $X = \{x_1, x_2, \cdots, x_n\}$, 类别数为 K. 首先, 定义局部相似矩阵 S_l 和非局部相似矩阵 S_n 如下:

$$S_l = \begin{cases} s(x_i, x_j), & \text{若 } x_i \in N_k(x_j) \text{ 或 } x_j \in N_k(x_i) \\ & \text{且 } \pi_i = \pi_j \\ 0, & \text{其他} \end{cases} \tag{1}$$

$$S_n = \begin{cases} s(x_i, x_j), & \text{若 } x_i \notin N_k(x_j) \text{ 或 } x_j \notin N_k(x_i) \\ & \text{且 } \pi_i \neq \pi_j \\ 0 & \text{其他} \end{cases} \tag{2}$$

其中 $s(.,.)$ 表示每两个样本 (x_i, x_j) 之间的相似度 (即欧式距离), $N_k(x_i)$ 是 x_i 的 k 最近邻, p_i 和 p_j 分别是样本 (x_i, x_j) 的类标签.



(a) Local and nonlocal graph　(b) Local and nonlocal penalization graph

图 5　内嵌数据结构图

Fig. 5　Intrinsic data structure graph

通过最小化目标函数 (3) 以保持数据 X 局部信息的低维近邻关系. 从 S_l 的构造中可以看出, 两近邻同类样本的距离越大则对应权值越大, 为使目标函数达到最小, 则约束投影后的两样本距离变小.

$$\begin{aligned} J_l(w) = & \frac{1}{2}\sum_{ij}(y_i - y_j)^2 s_{ij} = \\ & \frac{1}{2}\sum_{ij}(w^{\mathrm{T}}x_i - w^{\mathrm{T}}x_j)^2 s_{ij} = \\ & \sum_i w^{\mathrm{T}}x_i D_{ii} x_i^{\mathrm{T}} w - \sum_{ij} w^{\mathrm{T}}x_i S_{ij} x_j^{\mathrm{T}} w = \\ & w^{\mathrm{T}} X D_l X^{\mathrm{T}} w - w^{\mathrm{T}} X S_l X^{\mathrm{T}} w = \\ & w^{\mathrm{T}} X (D_l - S_l) X^{\mathrm{T}} w = \\ & w^{\mathrm{T}} X L_l X^{\mathrm{T}} w \end{aligned} \tag{3}$$

其中 $y_i = w^{\mathrm{T}}x_i$, y_i 为 x_i 的低维投影向量, L_l 称为局部散度矩阵, $L_l = D_l - S_l$, D_l 为对角矩阵, 其对角线上的元素为 S_l 列向量之和, D_{ij} 为对角矩阵, 其对角线上的元素为 S_l 列向量之和, $D_{ij} = \sum_j S_{ij}$, 可以看出 L_l 为实对称矩阵.

此外通过最大化目标函数 (4), 提取样本的非局部信息.

$$\begin{aligned} J_n(w) = &\frac{1}{2}\sum_{ij}(y_i - y_j)^2 s_{ij} = \\ &\frac{1}{2}\sum_{ij}(w^{\mathrm{T}}x_i - w^{\mathrm{T}}x_j)^2 s_{ij} = \\ &w^{\mathrm{T}}X(D_n - S_n)X^{\mathrm{T}}w = \\ &w^{\mathrm{T}}XL_nX^{\mathrm{T}}w \end{aligned} \tag{4}$$

其中 $y_i = w^{\mathrm{T}}x_i$, y_i 为 x_i 的低维投影向量, L_n 称为非局部散度矩阵, $L_n = D_n - S_n$, D_n 为对角矩阵, 其对角线上的元素为 S_n 列向量之和.

1.3.2 局部/非局部惩罚描述

构建权值矩阵 F 描述近邻非同类的数据, 通过最大化目标函数 (6), 使得这类数据的低维投射距离很远.

$$F = \begin{cases} 1, & \text{若 } x_i \in N_k(x_j) \text{ 或 } x_j \in N_k(x_i) \\ & \text{且 } \pi_i \neq \pi_j \\ 0, & \text{其他} \end{cases} \tag{5}$$

$$\begin{aligned} J_f(w) = &\frac{1}{2}\sum_{ij}(y_i - y_j)^2 f_{ij} = \\ &\frac{1}{2}\sum_{ij}(w^{\mathrm{T}}x_i - w^{\mathrm{T}}x_j)^2 f_{ij} = \\ &w^{\mathrm{T}}XD_fX^{\mathrm{T}}w - w^{\mathrm{T}}XFX^{\mathrm{T}}w = \\ &w^{\mathrm{T}}XL_fX^{\mathrm{T}}w \end{aligned} \tag{6}$$

其中 $y_i = w^{\mathrm{T}}x_i$, y_i 为 x_i 的低维投影向量, L_f 称为局部惩罚矩阵, $L_f = D_f - F$, D_f 为对角矩阵, 其对角线上的元素为 F 列向量之和. 从权值 F 的构造中可以看出, 两近邻非同类的样本对应的惩罚因子为 1. 为使得目标函数 (6) 达到最大, 则约束投影后两样本之间的距离变大.

类似地, 通过构建权值矩阵 P 描述同类非近邻数据, 并最小化目标函数 (8), 保持这类数据的低维近邻关系.

$$P = \begin{cases} 1, & \text{若 } x_i \notin N_k(x_j) \text{ 或 } x_j \notin N_k(x_i) \\ & \text{且 } \pi_i = \pi_j \\ 0, & \text{其他} \end{cases} \tag{7}$$

$$\begin{aligned} J_p(w) = &\frac{1}{2}\sum_{ij}(y_i - y_j)^2 p_{ij} = \\ &\frac{1}{2}\sum_{ij}(w^{\mathrm{T}}x_i - w^{\mathrm{T}}x_j)^2 p_{ij} = \\ &w^{\mathrm{T}}XD_pX^{\mathrm{T}}w - w^{\mathrm{T}}XPX^{\mathrm{T}}w = \\ &w^{\mathrm{T}}XL_pX^{\mathrm{T}}w \end{aligned} \tag{8}$$

其中 $y_i = w^{\mathrm{T}}x_i$, y_i 为 x_i 的低维投影向量, L_p 称为非局部散度矩阵, $L_p = D_p - P$, D_p 为对角矩阵, 其对角线上的元素为 P 列向量之和.

1.3.3 目标函数

根据上述分析过程, 整合目标函数 $J_l(w)$、$J_n(w)$、$J_f(w)$ 和 $J_p(w)$ 构建最终目标函数为

$$\begin{aligned} J(w) = &\min_w \frac{w^{\mathrm{T}}XL_lX^{\mathrm{T}}w + w^{\mathrm{T}}XL_pX^{\mathrm{T}}w}{w^{\mathrm{T}}XL_nX^{\mathrm{T}}w + w^{\mathrm{T}}XL_fX^{\mathrm{T}}w} = \\ &\min_w \frac{w^{\mathrm{T}}XL_{lp}X^{\mathrm{T}}w}{w^{\mathrm{T}}XL_{nf}X^{\mathrm{T}}w} \end{aligned} \tag{9}$$

其中 $L_{lp} = L_l + L_p$, $L_{nf} = L_n + L_f$.

目标函数可以转化为广义特征值的问题求解. 采用拉格朗日乘子法, 令分母为非零常数, 即 $w^{\mathrm{T}}S_{nf}w = C \neq 0$, 定义拉格朗日函数为

$$L(w, \lambda) = w^{\mathrm{T}}S_{lp}w - \lambda(w^{\mathrm{T}}S_{nf}w - C) \tag{10}$$

式中, λ 为拉格朗日乘子. 将式 (10) 对 w 求偏导得

$$\frac{\partial L(w, \lambda)}{\partial w} = S_{lp}w - \lambda S_{nf}w \tag{11}$$

令式 (11) 偏导数为零, 得出 $S_{lp} - \lambda S_{nf}w = 0$, 即

$$S_{lp}w = \lambda S_{nf}w \tag{12}$$

式中 $S_{lp} = XL_{lp}X^{\mathrm{T}}$, $S_{nf} = XL_{nf}X^{\mathrm{T}}$. 通过求解该广义特征方程的特征值和特征向量得到最优解. 设列向量 $W = (w_1, w_2, \cdots, w_M)$ (M 为输入数据特征维度) 为式 (12) 的解, 对应特征值 $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_M$. 保留前 D 个特征向量, 组成新投影矩阵 $w = (w_1, w_2, \cdots, w_D)$ ($D \ll M$), 通过式 (13) 实施数据投影可获得低维特征, M 维样本数据 $X = \{x_1, x_2, \cdots, x_n\}$ 通过映射转变为 D 维样本数据 $Y = \{y_1, y_2, \cdots, y_n\}$.

$$y = w^{\mathrm{T}}x \tag{13}$$

1.4 基于高斯混合模型动态集成的晶圆缺陷探测与识别

由于晶圆制造过程的高度动态性, 各种晶圆缺陷模式在线上可随机出现, 并且会产生一些新缺陷

模式, 因此本文设计了一种基于 GMM 的在线动态建模与缺陷识别模型. GMM 是描述数据混合密度分布的模型, 设 d 维随机变量 $X=[X_1,\cdots,X_d]$ 服从有限个正态分布, $x=[x_1,\cdots,x_d]$ 为 X 的一组输出, 令其中一个输出样本 x 的概率密度函数为

$$p(x|\phi)=\sum_{m=1}^{M}\pi_m p_m(x|\theta_m) \tag{14}$$

其中 π_m 为混合比, 描述了第 m 个高斯模型在混合模型中的权重, 满足 $\pi_m\geq 0$, $\sum_{m=1}^{M}\pi_m=1$, $p_m(x|\theta_m)$ 是第 m 个高斯模型的概率密度函数, $\theta_m=(\mu_m,S_m)$, μ_m 是期望, S_m 是协方差矩阵. GMM 参数集可定义为 $\phi=(\pi_1,\cdots,\pi_m;\theta_1,\cdots,\theta_m)$. 本文通过贝叶斯信息准则 (Bayesian information criterion, BIC)[17] 选出最合适的 GMM 高斯子个数 M, 采用期望最大算法 (Expectation maximization, EM)[18] 对 GMM 进行参数估计. 选取式 (14) 的负对数似然概率 (Negative log-likelihood probability, NLLP) 作为 GMM 输出, 每个 GMM 代表数据的一种类别, 通过获取各个 GMM 输出, 选取 NLLP 最小值对应的类作为判别结果.

本文提出的方法实施流程如图 6 所示, 核心思想是首先在离线阶段对各种晶圆缺陷模型采用 GMM 进行单独建模, 建立基于多 GMM 集成的识别库; 在线缺陷检测阶段, 采用 GMM_0 (对应正常晶圆模式) 的输出 (即 NLLP) 作为过程状态指标, 一旦探测到线上的缺陷晶圆, 便将其输入 GMM 识别库, 进行缺陷模式新模式检测, 若缺陷晶圆属于晶圆数据库中的缺陷模式, 则进行具体缺陷模式识别. 如有新缺陷模式产生, 收集新缺陷模式晶圆并构建 GMM 识别模型, 更新 GMM 识别库. 整个系统包括离线建模和在线处理两个阶段, 具体操作流程如下:

离线建模.

步骤 1. 建立样本库, 收集各种晶圆缺陷模式数据;

步骤 2. 对样本进行图像预处理, 通过中值滤噪法消除晶圆图表面的噪声;

步骤 3. 对于正常晶圆, 提取灰度特征、投影特征和纹理特征; 缺陷晶圆在此基础上再提取几何特征;

步骤 4. 采用 LNLDA 对每种晶圆模式产生的原始特征集进行特征提取;

步骤 5. 对每种晶圆缺陷模式构建高斯混合模型 (GMM_0 为正常晶圆模式的高斯混合模型, n 种缺陷晶圆模式的对应高斯混合模型为 $GMM_1\sim GMM_n$);

步骤 6. 将每种晶圆模式分别输入到对应的 GMM 中, 将输出的 NLLP 作为监控变量, 并设定在控率为 99.73 % (3 西格玛合格率), 确定控制图上信任限 $UCL_0\sim UCL_n$ (Upper control limit).

在线检测.

步骤 1. 对线上采集的晶圆图进行中值滤噪处理;

步骤 2. 提取晶圆灰度特征、投影特征和纹理特征作为原始特征集, 采用 LNLDA 进行特征提取;

步骤 3. 将该晶圆模式的低维空间特征输入 GMM_0 中, 若 $NLLP_0>UCL_0$, 则判断线上出现缺陷晶圆;

步骤 4. 对存在缺陷的晶圆在原始特征集的基础上提取几何特征, 通过 LNLDA 降维处理后将总特征集分别输入到 $GMM_1\sim GMM_n$, 若对于所有的识别模型都满足 $NLLP_i>UCL_i\,(i=1,2,\cdots,n)$, 则判定该晶圆属于新缺陷模式, 并收集新缺陷模式构建 GMM_{new}, 更新 GMM 识别库;

步骤 5. 判定为属于晶圆数据库缺陷模式的晶圆, 将特征分别输入到 $GMM_1\sim GMM_n$, 输出 NLLP 值最小的 GMM 对应模式的类别即为缺陷模式识别结果.

2 实验与结果分析

本文采用实际半导体制造过程中采集的晶圆图像数据 (称为 WM-811K 数据) 验证提出的晶圆缺陷探测与识别方法的有效性. 该数据集包括 811 457 张晶圆图, 所有晶圆图数据都已标记晶圆模式, 有关 WM-811K 数据更多的信息及数据获取方式可参考文献 [6]. 本文从 WM-811K 数据集中选取最为常见的六种晶圆缺陷模式 (Center、Donut、Edge-local、Edge-ring、Local 和 Random) 和正常晶圆模式. LNLDA 性能分析实验每种缺陷模式各选取 200 个样本组成数据集. 基于 GMM 动态集成的晶圆缺陷探测与识别实验中, 训练样本每种缺陷模式各取 200 个样本, Normal 选取 400 个样本, 训练集共计 1 600 个样本; 测试样本中每种缺陷模式选取 100 个, Normal 选取 200 个样本, 测试集共计 800 个样本.

2.1 LNLDA 性能分析

通过对晶圆图滤噪处理并产生原始特征后, 采用 LNLDA 对高维原始特征集进行特征提取, 以提高模型的缺陷探测与识别性能. 实验中每种缺陷模式各选取 200 个缺陷样本组成数据集, 将常用特征提取算法 PCA、LDA、LPP 和 LFDA (Local fisher discriminant analysis) 与 LNLDA 进行对比测试. PCA 和 LDA 是两种常见的线性降维算法, LPP 和

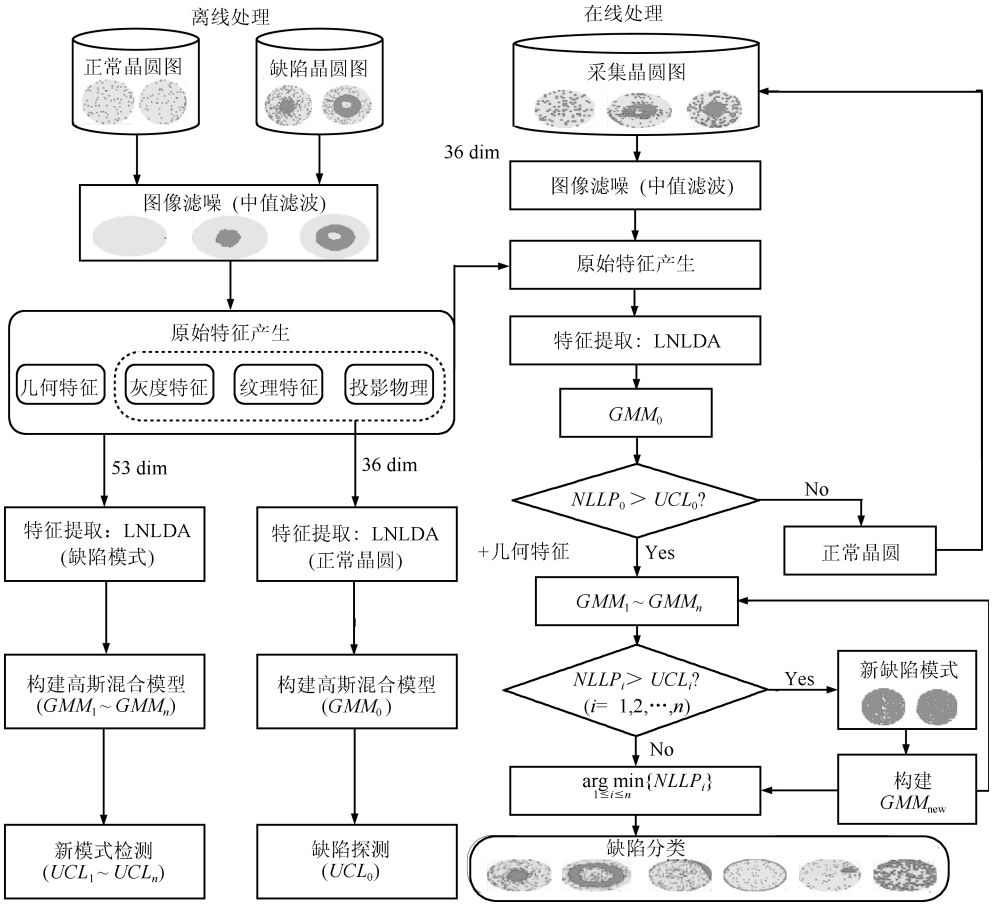


图 6 提出方案的实施流程图

Fig. 6 The flow chart of the proposed method

LFDA 属于典型流形学习算法[19]. LFDA 需要构建类内近邻图, LPP 与 LNLDA 也需要构建局部近邻图, 实验中设置近邻数为 50. LPP、LFDA 和 LNLDA 算法中 PCA 预处理维数设置为 50. 实验中每种缺陷模式样本提取特征数 53 个, 划分样本为 train:test=2:1、train:test=1:1 与 train:test=1:2 三套数据集. 采用 1NN 方法识别缺陷模式, 其中最近邻相似度测量采用欧氏距离.

为了验证基于 LNLDA 特征提取方法的有效性, 本文也给出基于所有原始特征的识别率 (命名为 Baseline). 图 7 呈现了基于各种特征提取算法的不同训练样本和测试样本数的晶圆缺陷识别效果. 表 2 给出了各种算法的最好识别率及使用维度. 从图 7 可以看出, 通过特征提取后, 晶圆缺陷的整体识别率有较大提高. LNLDA 显示了最好性能, 一直保持较高识别率, 一般在输入维度 5～10 之间获得最高识别率, 随着维度的增加, LNLDA 识别率基本处于稳定状态, 波动较小. 通过表 2 看出, 当训练和测试样本数目确定时, LNLDA 的最高识别率高于其他几种特征提取算法, 在 train:test=2:1 实验中, LNLDA 的识别率可以达到 85.44 %, 充分表明了 LNLDA 特征提取方法对于提高基于 1NN 的晶圆缺陷模式识别率的重要性.

2.2 基于 GMM 动态集成的晶圆缺陷探测与识别

2.2.1 晶圆缺陷探测

GMM 动态识别模型首先采用 NLLP 控制图进行晶圆缺陷在线探测, 实时判别线上的晶圆是否存在缺陷. LNLDA 进行特征提取时 PCA 预处理维数设置为 30, 输出维度为 7, 近邻系数 k 为 40. 表 3 列出了测试样本晶圆缺陷探测率, 图 8 呈现了 NLLP 控制图监测缺陷晶圆的情况. 对于测试样本中的正常晶圆, NLLP 控制图的误判率仅为 2.5 %; 测试样本中的缺陷晶圆探测率达到 99.33 %, Donut、Edge-ring 和 Random 这三种缺陷模式可以全部被探测出来, 充分证明本文提出的 NLLP 控制图可以应用于在线晶圆缺陷探测中.

由于正常晶圆表面存在的噪音等因素, GMM 动态识别模型会将一些正常晶圆判定为缺陷晶圆, 图 9 展示了实验中判定为缺陷晶圆的正常样本. 对

于误判为缺陷的 5 个晶圆样本, 再分别提取样本的几何特征, 输入到 GMM 识别库中进行缺陷识别, 5 个样本则被识别为 Edge-ring 缺陷模式.

2.2.2　晶圆缺陷模式识别

晶圆制造过程中, 一旦 NLLP 控制图探测到属于晶圆数据库缺陷模式的晶圆, 便将其输入到 GMM 识别库进行缺陷模式识别. 实验首先采用两套特征集进行缺陷识别: 第一套特征集数据维度为 46 (见表 1, 不包括 Hu 不变矩), 这里标识为 A; 第二套特征集数据维度为 53 (A 特征集 + Hu 不变矩), 标识为 B. 两套特征集对应的缺陷识别结果如表 4 所示, 可以看出加入 Hu 不变矩之后, 缺陷 Local 和 Edge-local 的识别率有所提升, 整体识别率提高到

表 2　基于各种投影算法的 1NN 最佳识别率 (%) 及相应输入维度

Table 2　Best recognition rate (%) of 1NN and subspace dimensionality based on the five data projection algorithms

Algorithm	train:test=2:1	train:test=1:1	train:test=1:2
Baseline	66.60 (53)	63.71 (53)	62.38 (53)
PCA	66.81 (11)	63.71 (14)	62.49 (12)
LDA	56.75 (3)	57.57 (21)	64.20 (3)
LPP	80.51 (24)	77.71 (30)	77.06 (29)
LFDA	77.94 (6)	80.29 (5)	79.42 (5)
LNLDA	85.44 (6)	84.71 (10)	84.46 (10)

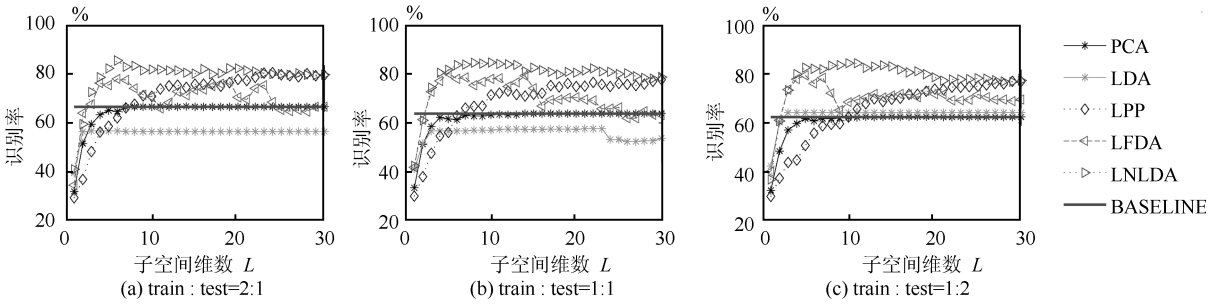


图 7　不同子空间维数下的晶圆缺陷识别率变化

Fig. 7　The recognition rate of wafer defect pattern with different reduced dimensionality

表 3　晶圆缺陷探测率 (%)

Table 3　The defect detection rate (%)

缺陷模式	Center	Donut	Edge-ring	Edge-local	Random	Local
缺陷探测率	97.0	100	100	99.0	100	100

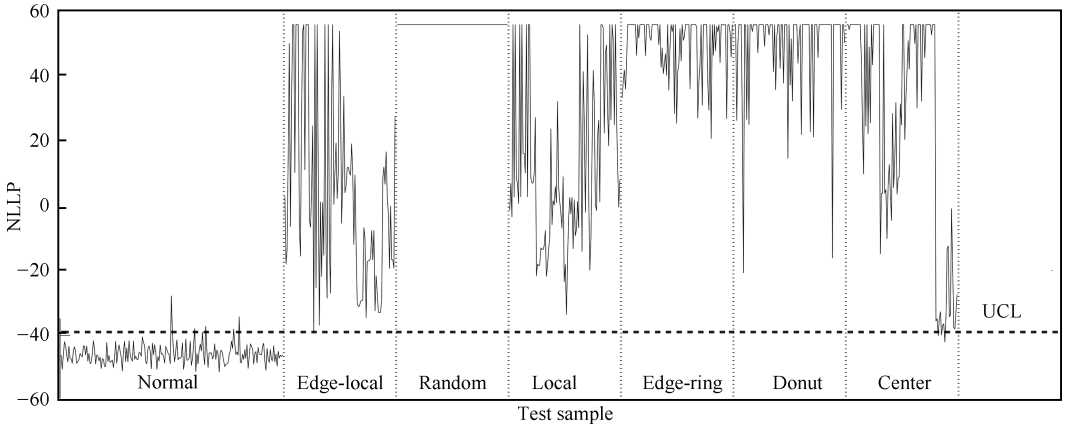


图 8　晶圆缺陷 NLLP 监控图

Fig. 8　Defect pattern detection by NLLP control chart

87.5 %, 通过混淆矩阵 (Confusion matrix) 呈现该套特征集的具体识别结果, 如表 5 所示. 从表 5 可以看出, 本文提出的识别模型能够很好地识别出 Edge-ring、Donut、Random 和 Center 缺陷, 但同时也可以发现 GMM 分类器容易将其他的缺陷模式识别为 Local 缺陷, 图 10 是两张被识别为 Local 模式的缺陷晶圆图, 实际缺陷模式分别为 Center 与 Edge-local. 可以看出, 这两个缺陷晶圆样本表面噪音严重, 而且缺陷模式不明显, 因此模型会误判它们为 Local 缺陷模式.

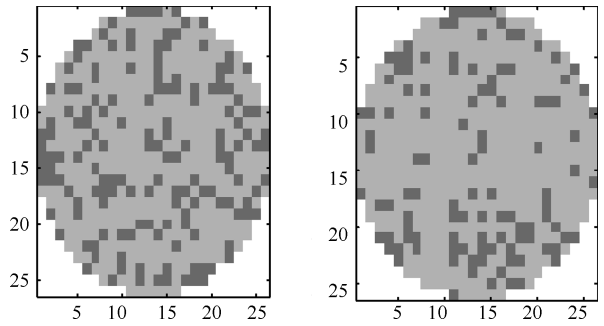


图 9 判定为缺陷模式的正常晶圆样本

Fig. 9 Examples of normal wafer that incorrectly predicted as defect pattern

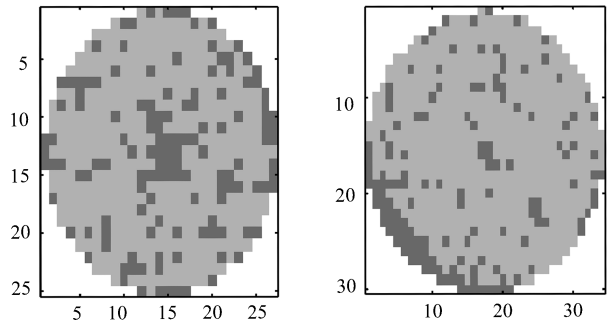


图 10 误识别为 Local 模式的缺陷晶圆样本

Fig. 10 Defect pattern that incorrectly predicted as local pattern

为了验证 GMM 动态集成系统的性能, 本实验将 GMM 与 KNN、SVM、反向传播神经网络 (Back propagation network, BPN) 和 C4.5 决策树四种分类器进行对比测试, 其中 KNN 近邻系数选取 9, SVM 选择 LIBSVM[20] 分类器, 其中核函数选择径向基函数, 其 gamma 参数设置为 2.8, 其余参数为默认值. BPN 采用 3 层网络, 中间层神经元个数 10, 最大迭代次数和学习率分别设置为 500 和 0.01. 各种分类器识别结果如表 6 所示. 可以看出, 对于一些缺陷模式, 比如 Center、Donut、Edge-ring 和 Random, 这几种分类器都能很好地识别出来, 但是 GMM 整体识别率高于其他的分类器. 晶圆缺陷探测识别中, 由于线上产生的晶圆缺陷模式是未知的, 因此事先无法确定采用哪一种分类器进行线上识别, 因此通常采用综合识别率较高的分类器. 从综合识别率上比较, 本文提供的 GMM 动态识别模型显示了最好的性能.

表 4 两套特征集对应的晶圆缺陷识别率 (%)

Table 4 The recognition rate based on two different feature sets (%)

缺陷模式	特征集 A	特征集 B
Center	86	87
Donut	86	87
Edge-ring	100	100
Local	65	76
Random	92	93
Edge-local	67	75
Average	84.0	87.5

2.2.3 扩展研究 — 新缺陷模式在线探测与识别

由于半导体制造过程中的复杂性与动态性, 线上生产的晶圆可能会出现新缺陷模式, 本文提出的 GMM 动态识别方法可检测缺陷晶圆是否出现新缺陷模式. 如图 11 所示, GMM 动态识别模型通过建立对应于每种缺陷模式的 GMM 识别库, 并建立相应的控制信任限识别新缺陷模式. 当新缺陷模式晶圆输入到识别库的各个 GMM 识别器, 如果连续超过了所有 GMM 识别器的信任限, 可认为一个新缺陷模式在线上已经产生. 在线上收集一定的该新缺陷模式样本数据, 构建对应的 GMM 识别器, 更新

表 5 晶圆缺陷模式识别率 (%)

Table 5 The recognition rate of different defect patterns (%)

缺陷模式	Center	Donut	Edge-ring	Local	Random	Edge-local
Center	87	0	0	12	0	1
Donut	0	94	0	6	0	0
Edge-ring	0	0	100	0	0	0
Local	8	3	0	76	0	13
Random	0	2	0	2	93	3
Edge-local	67	75	9	14	1	75

原有的 GMM 动态识别模型. 本文执行以下一个扩展实验以验证 GMM 动态识别模型的探测与识别新缺陷模式的性能.

该实验把 Local 模式作为新缺陷模式, 而其他五种模式 (即 Center、Donut、Edge-local、Edge-ring 和 Random) 为已有模式并采用 GMM 进行一一建模, 建立 GMM 识别库. 选取 200 个 Local 缺陷晶圆进行新缺陷模式检测, 并输入到 GMM 识别库, 图 11 展示了基于已有缺陷模式信任限 (即 UCL) 的 Local 模式检测情况. 结果显示有 179 个样本连续超过 5 个 GMM 识别器的信任限, 被判定为新缺陷模式, 检测率可达到 89.5 %, 可以看出 GMM 动态识别模型能够较好地检测出生产线上出现的新缺陷模式晶圆. 进一步采用这 200 个 Local 缺陷晶圆数据构建一个新的 GMM 识别模型 (即 GMM local), 另外选取 100 个 Local 缺陷作为测试样本, 输入到新的 GMM 识别模型中进行模式识别. 图 12 呈现了 Local 缺陷模式在六个 GMM 识别模型的识别情况. 可以看出, 绝大多数的 Local 缺陷在对应的 GMM-Local 中输出的 NLLP 最低, Local 缺陷识别率达到 76.0 %. 常规识别器 (如 KNN、SVM、BPN、C4.5 决策树) 在离线建模后其模型将固定而无法动态探测新缺陷模式并更新其模型, 本文提出的 GMM 动态识别模型的重要特点是能动态更新其识别模型. 因此, 从性能及其动态建模角度比较, 本文提出的方法具有较好的性能并更加适用于实际的工业应用.

3　结论

本文提出了一种基于局部与非局部判别分析和高斯混合模型动态集成的模型, 进行晶圆表面缺陷的探测与识别. LNLDA 通过融合数据局部/非局部信息以及局部/非局部惩罚信息, 较为显著地提高了模型的晶圆缺陷探测与识别性能. 基于 GMM 动态集成的方法可检测线上出现的新缺陷模式, 并且通过对新缺陷模式构建识别模型更新 GMM 识别库, 能够识别线上出现的新缺陷模式. 相比于离线建模阶段固定分类模型的识别器, GMM 动态集成识别方法更具适用性. 在 WM-811K 数据的实验结果中,

表 6　各种分类器的晶圆缺陷识别率比较 (%)

Table 6　Comparison of recognition rates of the five classifiers (%)

分类器	GMM	KNN	SVM	BPN	C4.5
Center	87	92	97	95	93
Donut	94	98	100	98	96
Edge-ring	100	100	100	100	100
Local	76	69	68	56	65
Random	93	93	92	94	94
Edge-local	75	66	66	78	56
Average	87.50	86.33	87.17	86.83	84.0

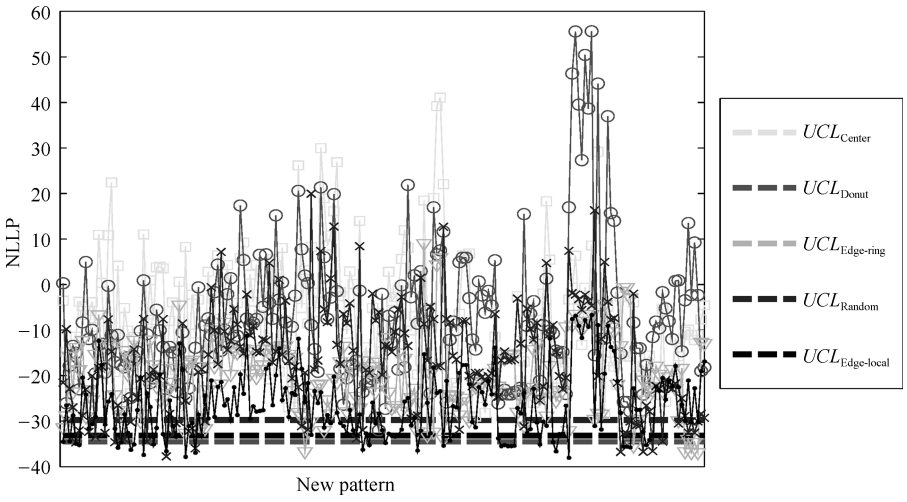


图 11　晶圆新缺陷模式 (Local) NLLP 检测图

Fig. 11　New defect pattern (local) detection by NLLP control chart

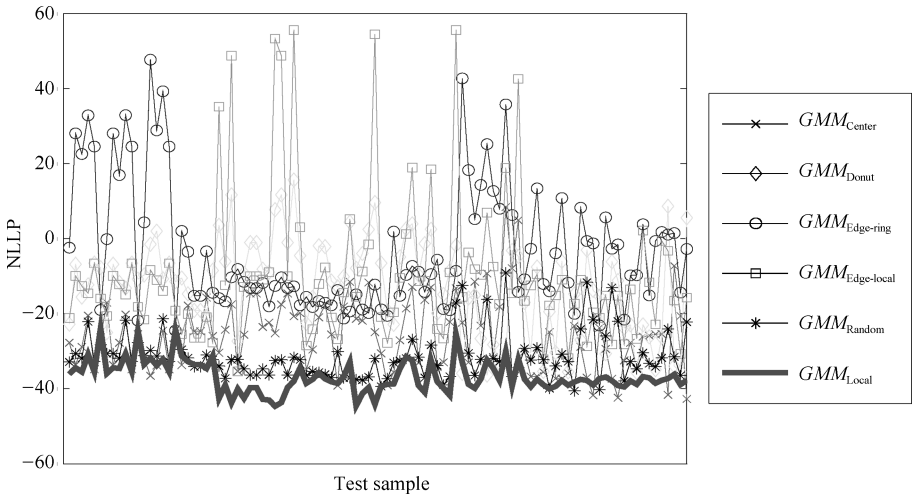


图 12 晶圆 Local 缺陷模式的 NLLP 识别图

Fig. 12 Local defect pattern recognition by NLLP control chart

本文提出的模型的整体缺陷探测率达到 99.33 %, 缺陷模式识别率可达到 87.5 %, 验证了该模型的有效性. 后续工作将集中研究基于流形学习的晶圆特征选择方法, 并应用到晶圆生产过程中缺陷模式的探测与识别.

References

1 Wang C H. Recognition of semiconductor defect patterns using spatial filtering and spectral clustering. *Expert Systems with Applications*, 2008, **34**(3): 1914−1923

2 Hess C, Weiland L H. Extraction of wafer-level defect density distributions to improve yield prediction. *IEEE Transactions on Semiconductor Manufacturing*, 1999, **12**(2): 175−183

3 Friedman D J, Hansen M H, Nair V N, James D A. Model-free estimation of defect clustering in integrated circuit fabrication. *IEEE Transactions on Semiconductor Manufacturing*, 1997, **10**(3): 344−359

4 Chen F L, Liu S F. A neural-network approach to recognize defect spatial pattern in semiconductor fabrication. *IEEE Transactions on Semiconductor Manufacturing*, 2000, **13**(3): 366−373

5 Baly R, Hajj H. Wafer classification using support vector machines. *IEEE Transactions on Semiconductor Manufacturing*, 2012, **25**(3): 373−383

6 Wu M J, Jang J S R, Chen J L. Wafer map failure pattern recognition and similarity ranking for large-scale data sets. *IEEE Transactions on Semiconductor Manufacturing*, 2015, **28**(1): 1−12

7 Chen S H, Perng D B. Directional textures auto-inspection using principal component analysis. *The International Journal of Advanced Manufacturing Technology*, 2011, **55**(9−12): 1099−1110

8 Tsai D M, Lai S C. Defect detection in periodically patterned surfaces using independent component analysis. *Pattern Recognition*, 2008, **41**(9): 2812−2832

9 Yan De-Qin, Liu Sheng-Lan, Li Yan-Yan. An embedding dimension reduction algorithm based on sparse analysis. *Acta Automatica Sinica*, 2011, **37**(11): 1306−1312
(闫德勤, 刘胜蓝, 李燕燕. 一种基于稀疏嵌入分析的降维方法. 自动化学报, 2011, **37**(11): 1306−1312)

10 Tenenbaum J B, de Silva V, Langford J C. A global geometric framework for nonlinear dimensionality reduction. *Science*, 2000, **290**(5500): 2319−2323

11 Roweis S T, Saul L K. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 2000, **290**(5500): 2323−2326

12 He X F, Niyogi X. Locality preserving projections. *Advances in Neural Information Processing Systems*, 2004, **16**: 153−160

13 Yu J B. Local and nonlocal preserving projection for bearing defect classification and performance assessment. *IEEE Transactions on Industrial Electronics*, 2012, **59**(5): 2363−2376

14 Wang C H, Kuo W, Bensmail H. Detection and classification of defect patterns on semiconductor wafers. *IIE Transactions*, 2006, **38**(12): 1059−1068

15 Byers S, Raftery A E. Nearest-neighbor clutter removal for estimating features in spatial point processes. *Journal of the American Statistical Association*, 1998, **93**(442): 577−584

16 White Jr K P, Kundu B, Mastrangelo C M. Classification of defect clusters on semiconductor wafers via the Hough transformation. *IEEE Transactions on Semiconductor Manufacturing*, 2008, **21**(2): 272−278

17 Schwarz G. Estimating the dimension of a model. *The Annals of Statistics*, 1978, **6**(2): 461−464

18 Dempster A P, Laird N M, Rubin D B. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society, Series B*, 1977, **39**(1): 1−38

19 Zhang Jian-Wei, Wang Wan-Liang, Yao Xiao-Ming, Shi Hai-Yan. Face recognition using tensor local Fisher discriminant analysis. *Acta Automatica Sinica*, 2012, **38**(9): 1485−1495 (郑建炜, 王万良, 姚晓敏, 石海燕. 张量局部 Fisher 判别分析的人脸识别. 自动化学报, 2012, **38**(9): 1485−1495)

20 Chang C C, Lin C J. LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2011, **2**(3): 27

余建波　博士, 同济大学副教授. 主要研究方向为统计分析, 智能维护和机器学习. 本文通信作者.
E-mail: jianboyu.bob@gmail.com
(**YU Jian-Bo**　Ph. D. and associate professor at the School of Mechanical Engineering, Tongji University. His research interest covers statistical analysis, intelligent condition-based maintenance and machine learning. Corresponding author of this paper.)

卢笑蕾　同济大学机械与能源工程学院硕士研究生. 主要研究方向为图像处理, 机器学习, 模式识别.
E-mail: lanmingdt@163.com
(**LU Xiao-Lei**　Master student at the School of Mechanical Engineering, Tongji University. Her research interest covers image processing, machine learning and pattern recognition.)

宗卫周　同济大学机械与能源工程学院硕士研究生. 主要研究方向为信号处理.
E-mail: zongweizhou@163.com
(**ZONG Wei-Zhou**　Master student at the School of Mechanical Engineering, Tongji University. His research interest covers signal processing.)