

主成分分析及聚类方法在碲镉汞晶片参数判别中的应用研究

吴佳昊^{1,2}, 乔辉¹, 李向阳¹

(1. 中国科学院上海技术物理研究所, 上海 200083;
2. 上海科技大学 信息科学与技术学院, 上海 201210)

摘要: 基于主成分分析和聚类方法提出了一种碲镉汞晶片参数筛选方法, 建立了对碲镉汞晶片参数进行筛选的数据模型, 模型中通过对初始晶片数据进行清洗和分析, 利用主成分分析(PCA)降维法和基于密度的聚类算法(DBSCAN), 确定了晶片数据中最密集的区域, 同时利用流片后得到高性能芯片的优质碲镉汞晶片参数拟合边界椭圆曲线, 并将其作为优质晶片的判断标准, 能够根据输入的晶片电学和光学参数生成晶片评级, 得到了大于 90% 的成功率。

关键词: 碲镉汞; 晶片筛选; 主成分分析; 聚类
中图分类号: O471.2; TP311 **文献标识码:** A

Application of principal component analysis and clustering methods in the discrimination of parameters in HgCdTe crystals

Wu Jia-Hao^{1,2}, QIAO Hui¹, LI Xiang-Yang¹

(1. Shanghai Institute of Technical Physics, Chinese Academy of Sciences, Shanghai 200083, China;
2. School of Information Science and Technology, ShanghaiTech University, Shanghai 201210, China)

Abstract: A method for selecting parameters in HgCdTe crystals has been proposed, utilizing Principal Component Analysis (PCA) and clustering methods, with the establishment of a data model for screening the parameters of HgCdTe crystals. Within the model, initial crystal data undergoes a cleaning and analysis process. PCA is employed for dimensionality reduction, and the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm is used to identify the densest regions within the crystal data. Furthermore, high-performance chip data, obtained post-processing, is utilized to fit boundary ellipses for high-quality HgCdTe crystal parameters. These ellipses act as criteria for identifying high-quality crystals. Capable of generating crystal ratings based on input electrical and optical parameters, the model achieves a coverage rate exceeding 90%.

Key words: HgCdTe, crystal parameter selection, principal component analysis, clustering analysis
PACS:

引言

碲镉汞材料因其具有电子有效质量小、吸收系数高以及工作温度高等优点, 被广泛认为是一种重要的窄禁带红外半导体材料。由碲镉汞制成的探测器因其低噪声、高探测率和响应波段宽等特点, 得到了广泛的应用。在碲镉汞探测器的制备工艺中, 高性能碲镉汞材料的选片是关键步骤, 相比更多依赖个人经验的定性判断, 需要建立一种更为科学、客观、高效的晶片筛选方法^{[1][2]}。

作为一种三元系合金, 碲镉汞材料由碲化镉和碲化汞两者组成, 其禁带宽度以及光学和电学特性参数与组分 x 密切相关。碲镉汞材料的电学特性参数通常包括载流子浓度、电阻率和迁移率; 光学特性参数通常包括材料的组分和透过率等。碲镉汞材料特性与器件性能参数密切相关。文献^[3]基于大量材料和芯片性能数据拟合出其生长的碲镉汞材料电学参数的经验公式, 给出了 300K 下碲镉汞材料电子浓度、电子迁移率、电阻率与组分 x 之间的关

系,并以此作为芯片制备中碲镉汞材料筛选的重要依据。国内的情况也类似,对于碲镉汞晶片参数的筛选主要基于大量流片经验的总结。这种材料筛选方式的缺点显而易见,筛选效率和准确性都受到限制。

目前随着科研活动越来越多步入数据密集型的“第四范式时代”^[4],论文提出一种基于主成分分析和聚类算法的碲镉汞晶片参数筛选方法,通过对大量晶片测试数据进行处理和分析,建立了对碲镉汞晶片参数进行筛选的数据模型,以实现碲镉汞晶片参数的有效和准确筛选。论文的第一部分介绍了碲镉汞材料的电学和光学参数以及测试方法;第二部分介绍了主成分分析和聚类算法;第三部分介绍了基于大量碲镉汞材料电学和光学特性参数利用算法得到的碲镉汞材料参数筛选模型。最后一部分是文章的总结。

1 碲镉汞材料性能数据

1.1 碲镉汞材料

碲镉汞材料的生长采用碲作熔剂的移动加热器法(THM)。生长好的链条从石英安瓿中取出后,经过切片、磨片、清洗、热处理等工艺后,得到直径为15mm、厚度为600~700微米的碲镉汞晶片。

1.2 电学参数测试

文章中碲镉汞材料的电学参数是通过范德堡法霍尔测试得到。霍尔测试中直接测试得到的参数为霍尔电压和电阻。以N型半导体为例,将正方形碲镉汞样品置于均匀磁场中,在材料中施加电流,电场方向与磁场方向相互垂直,假设电场沿 x 方向,磁场沿 z 方向,载流子在洛伦兹力的影响下将向负 y 方向偏移并聚集,从而在垂直于电场和磁场的方向上产生一个横向电场,即所谓霍尔电压。当横向电场力和洛伦兹力将达到平衡状态,可用以下方程表示:

$$qE_y - qv_x B_z = 0 \quad (1)$$

实际测试过程中,为了消除材料均匀性和电极接触等影响,需要进行电场和磁场反向取平均值。磁场沿 $+z$ 轴方向的情况下,施加从触点1到触点3的电流 I_{13} ,并测量触点2和触点4之间的电压 V_{24P} 。然后将电流反向,得到 I_{31} 并测得电压 V_{42P} 。更换电场电极,得到电流 I_{42} 下的电压 V_{13P} 和 I_{24} 下的电压 V_{31P} 。接着,将磁场反向,即为 $-z$ 磁场,重复上述测量,得到 I_{13} 、 I_{31} 、 I_{42} 、 I_{24} 相对应的 V_{24N} 、 V_{42N} 、 V_{13N} 、 V_{31N} 。之后计算出 V_C 、 V_D 、 V_E 和 V_F :

$$V_C = V_{24P} - V_{24N}, V_D = V_{42P} - V_{42N} \quad (2)$$

$$V_E = V_{13P} - V_{13N}, V_F = V_{31P} - V_{31N} \quad (3)$$

为了测量碲镉汞晶片的电阻,首先需要设定直流电流 I 。在触点1和触点2之间施加电流 I_{21} ,然后测量电压 V_{34} 。接着,将电流反向得到 I_{12} ,并测得电压 V_{43} 。通过重复这些步骤,测得 V_{41} 、 V_{14} 、 V_{12} 、 V_{21} 、 V_{23} 、 V_{32} 等电压值。在测量之前和之后,还需要分别测量温度,以确保温度的稳定性。如果两次测量的温度相差超过 1°C ,则需要等待温度稳定后,才能继续测量上述电压值。

接下来,将测得的八个电压值除以相应的电流值,得到八个正电阻值。为确保测量一致性,需要确保电流反向后得到的电阻值相等,因此可以计算出两个特征电阻值 R_A 和 R_B :

$$R_A = (R_{21,34} + R_{12,43} + R_{43,12} + R_{34,21})/4 \quad (4)$$

$$R_B = (R_{32,41} + R_{23,14} + R_{14,23} + R_{41,32})/4 \quad (5)$$

而 R_A 和 R_B 满足范德堡公式^[5]:

$$\exp(-\pi R_A/R_S) + \exp(-\pi R_B/R_S) = 1 \quad (6)$$

通过这些步骤,可以得到方块电阻 R_S ,电阻率 ρ 可以通过 $\rho = R_S d$ 计算得出。

为了测量碲镉汞晶片的载流子密度,首先施加一个正向的磁场,设为沿 z 轴方向,然后施加一个从触点1到触点3的电流 I_{13} 并测量触点2和触点4之间的电压 V_{24P} (P 指此时施加的磁场方向为正向),将电流反向即为 I_{31} 并测得电压 V_{42P} ,继续测量,得到电流 I_{42} 下的电压 V_{13P} 和 I_{24} 下的电压 V_{31P} 。将磁场反向,此时为负向磁场,重复上述测量,得到 I_{13} 、 I_{31} 、 I_{42} 、 I_{24} 和相对应的 V_{24N} 、 V_{42N} 、 V_{13N} 、 V_{31N} 。此次测量也要注意温度一致性。计算:

$$V_C = V_{24P} - V_{24N} \quad (7)$$

$$V_D = V_{42P} - V_{42N} \quad (8)$$

$$V_E = V_{13P} - V_{13N} \quad (9)$$

$$V_F = V_{31P} - V_{31N} \quad (10)$$

片状载流子密度可由 $n_s = 8 \times 10^{-8} IB/[q(V_C + V_D + V_E + V_F)]$ 计算得出,载流子密度 n 可以通过 $n = n_s/d$ 得到。

通过以上霍尔测量方法,可以获得碲镉汞晶片的电学参数,包括载流子浓度和电阻率。对18612组碲镉汞N型晶片的电学参数测量结果,如下图所示:

测量结果表明,在室温下,电子浓度和电阻率与组分 x 高度相关,因为组分 x 与碲镉汞晶片的禁带

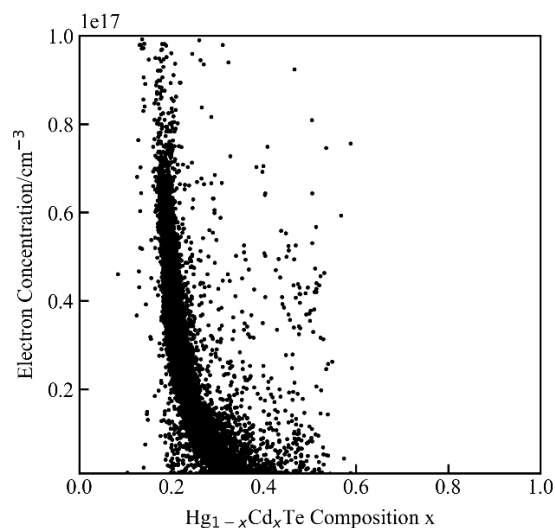


图1 室温电子浓度关于组分x的图像

Fig. 1 Electron concentration at 300K for $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ materials

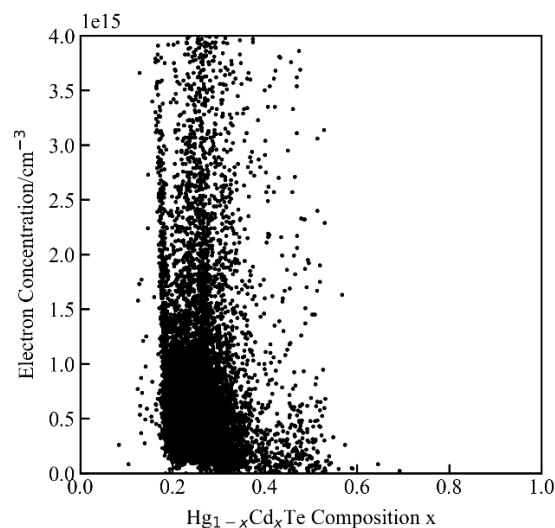


图3 低温电子浓度关于组分x的图像

Fig. 3 Electron concentration at 77K for $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ materials

宽度直接相关^[6],从而直接影响了热激发下的载流子浓度和电阻率。

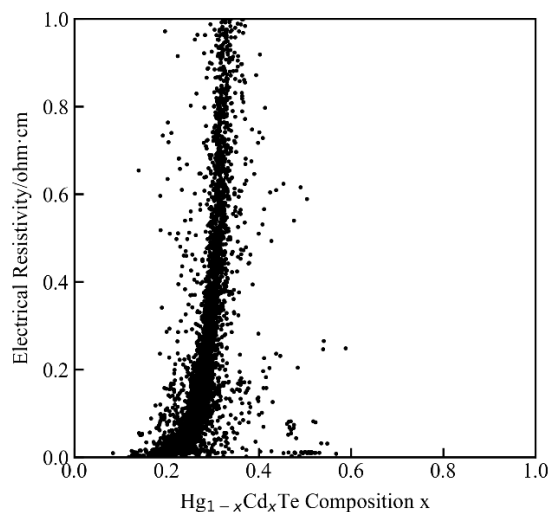


图2 室温电阻率关于组分x的图像

Fig. 2 Electrical resistivity at 300K as a function of composition x

此外,如图3和图4所示,随着组分 x 增大,低温电子浓度和低温电阻率变化基本没有规律,与组分关联较小,表明此时载流子不由热激发主导,而是主要来自杂质掺杂。

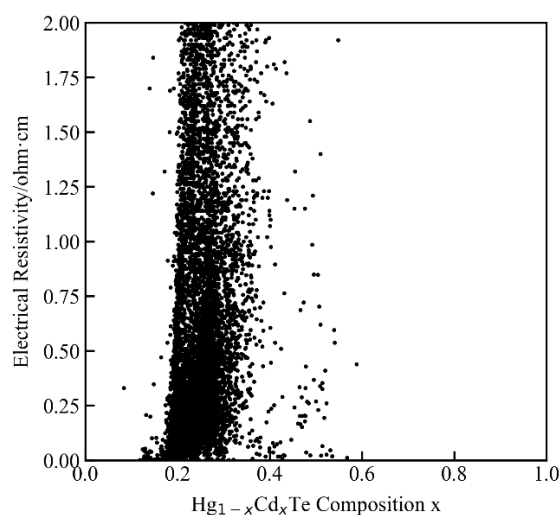


图4 低温电阻率关于组分x的图像

Fig. 4 Electrical resistivity at 77K for $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ materials

1.3 光学参数测试

傅里叶红外光谱仪(FTIR)利用傅里叶变换的原理,将入射的红外光进行频率分解,以获取每个频率成分的信息。其工作原理基于迈克尔逊干涉技术,红外光源发出的光经过待测样品后进入干涉仪,利用固定镜和移动镜的光程差形成干涉条纹。干涉后的光进入探测器,探测器将光强信号转换为电信号。最后这组电信号经过傅里叶变换,得到样品对不同波长光的吸收信息,从而反映样品的组成

信息。

由 FTIR 测量出这 18612 组碲镉汞晶片的透射率和组分 x , 得到图 5。

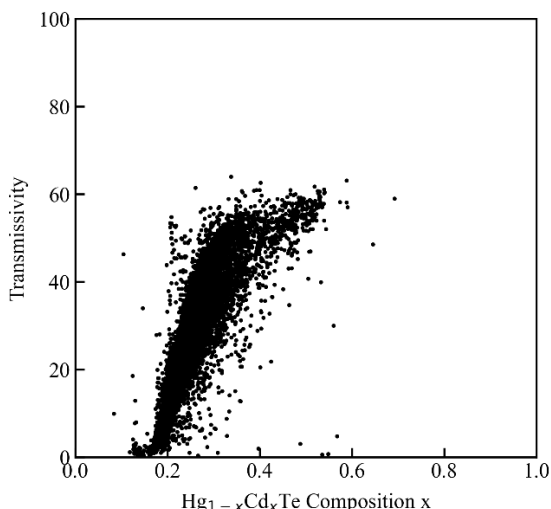


图 5 透过率关于组分 x 的图像

Fig. 5 Transmissivity for $\text{Hg}_{1-x}\text{Cd}_x\text{Te}$ materials

通过分析图像可以发现, 透过率随着组分 x 的变化呈现出一个先下降后上升最后趋于稳定的趋势。这表明透过率与组分 x 之间存在较强的相关性。

2 基于主成分分析的机器学习算法

2.1 主成分分析

主成分分析(PCA)是一种广泛使用的数据降维算法, 在多元变量分析领域, 主成分分析被用作一种统计分析技术, 以简化数据集^{[7][8]}。该方法运用正交变换对一组潜在相关的变量观测值进行线性转换, 将其投影为一系列线性无关的新变量, 这些新变量被称为主成分^{[9][10]}。

在对碲镉汞晶片样品进行电学和光学参数测量后, 对其进行进一步的数据处理, 并进行主成分分析。该分析旨在将原始 n 维特征映射到新的 k 维正交特征空间上, 这些正交特征即为主成分。主成分分析在保留信息的前提下, 将数据从原始空间转化到一组相互正交的新坐标轴上。这些新坐标轴的第一个坐标轴的方差最大, 第二个坐标轴与第一个坐标轴正交且具有最大方差, 其余坐标轴遵循相似规律。因此, 通过这种方式, 前 k 个主成分涵盖了数据绝大部分的方差, 而后续主成分的方差逐渐趋

近于零。通过保留前 k 个主成分, 保存数据的绝大部分方差信息, 实现数据降维。

对于数据集 $X = \{x_1, x_2, x_3, \dots, x_n\}$, 为实现降维到 k 维, 首先对数据进行去平均值操作, 即去中心化, 将每一维特征减去其各自的平均值。其次计算数据的协方差矩阵 $\frac{1}{n}XX^T$, 利用特征值分解方法对协方差矩阵 $\frac{1}{n}XX^T$ 进行分解, 得到特征值与相应的特征向量。对所得特征值进行从大到小的排序, 并选择其中最大的 k 个特征值。将其对应的 k 个特征向量分别作为行向量组成特征向量矩阵 P 。最终得到降维后的数据 $Y=PX$, 其中 Y 为数据在新的 k 维特征空间中的表示^{[11][12]}。

2.2 DBSCAN 聚类算法

经主成分降维后的数据需要使用 DBSCAN 聚类算法找到数据点最密集的区域, 并以此作为判别依据。

DBSCAN 是一种基于密度的空间聚类算法, 特点在于能够有效地在含有噪声的数据集中发现各种形状的聚类结构。该算法的核心思想在于根据数据点的密度情况来进行聚类, 从而适应不同密度分布的数据集。它依赖于两个参数: epsilon 和 minPts。Epsilon 确定聚类的最大半径, minPts 定义形成聚类所需的最少样本数。DBSCAN 从每个点出发, 以 epsilon 为半径绘制圆, 若圆内点数不小于 minPts, 则形成一个聚类。随后扩展该聚类, 直至新加入点数不足 minPts。通过调整 epsilon 和 minPts, DBSCAN 能有效划分样本点密集区域作为聚类, 并过滤噪声点^{[13][14][15]}。

2.3 Alpha shape 算法

通过聚类算法确定数据集中最密集区域的簇类后, 需要找到该簇类的边界, 通过判断数据点是否在边界内确定碲镉汞晶片评级。

Alpha shape 是描述点集合在空间中的形状轮廓的一种方法。其基本思想是给点集中的每个点放置一个球, 球的半径可调整。当半径很小时, 球不相交, alpha shape 就是点集本身。当半径增大, 球开始相交, 形成联通区域。Alpha shape 由这些区域边界组成, 反映了点集在此半径下的形状轮廓。具体来说算法可以从点集出发构建一种特殊的三角剖分。当两个点所在球相交时, 连接这两个点; 当三个点所在球有公共交集时, 用这三点构成一个三角形。随着 α 增大, 三角剖分中包含的三角形越来越

越多, alpha shape 就是其中的一部分。Alpha shape 反映了点集在不同尺度下的拓扑结构, 是理解和分析点集形状的有效工具^{[16][17][18]}。

3 计算结果

3.1 降维

将组分范围设置成 0.18 到 0.3, 因为光谱仪的截止波长是 25 微米, 能测到的最小组分是 0.18, 而组分位于 0.3 以上的晶片, 用 P 型片做光伏片, T 组分位于 0.3 以下的晶片, 用 N 型片做光导片, 这里只对 N 型片进行分析。接着对初始数据去除粗大误差值, 得到 12815 组碲镉汞晶片, 每组包括六个属性, 分别是室温电子浓度、室温电阻率、低温电子浓度、低温电阻率、透过率和组分。对数据进行主成分分析并计算每个主成分的方差, 得到主成分的累计方差图如图 6。

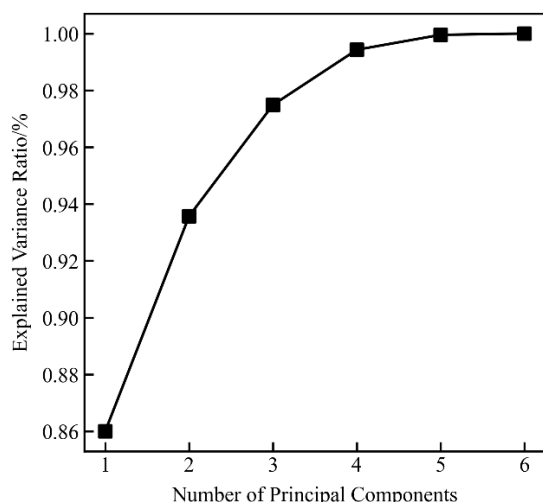


图 6 主成分的累计方差图

Fig. 6 Scree plot of principal components

降维的初始数据有 6 类, 因此横坐标的主成分最大数量为 6, 当主成分数为 6 时, 得到的方差为 100%。选取的主成分数越多, 数据组的方差越大, 数据组经过降维后保留的信息也越多, 在尽可能保留更多信息的前提下选取合适的主成分数是 PCA 降维法的关键, 一般来说选择方差在 90% 到 95% 之间的主成分数比较合理。图 6 表明, 当该数据组降为 2 维时, 方差达到了 94%, 此时既保留了足够多的信息量, 又将数据降到了合适的维度以便后续处理, 因此选择将 6 维数据组降至 2 维。降为 2 维后的

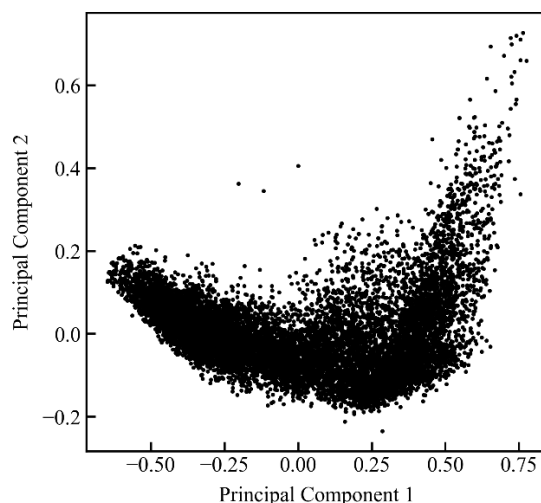


图 7 第一个主成分关于第二个主成分的形象

Fig. 7 Score plot of first principal component versus second principal component

数据如图 7。此时数据点主要呈现 V 形分布, 绝大多部分数据位于 V 形两侧, 只有一小部分数据分布于 V 形之外。得到主成分分析的降维系数, 写出经过尺度缩放后主成分计算公式:

$$PC1 = -0.692n_1 + 0.0717\rho_1 + 0.0196n_2 + 0.194\rho_2 + 0.685T + 0.0923x \quad (11)$$

$$PC2 = 0.507n_1 + 0.0202\rho_1 + 0.0718n_2 + 0.812\rho_2 + 0.279T - 0.00530x \quad (12)$$

其中 n_1 是室温电子浓度, ρ_1 室温电阻率, n_2 是低温电子浓度, ρ_2 是低温电阻率, T 是透过率, x 是组分。

3.2 聚类算法

对经过降维处理的数据运用 DBSCAN 聚类算法时, 这一过程如图 8 所示。聚类算法将这些数据点划分为多个簇, 其中蓝色的 Cluster 0 呈现出最大的聚集规模, 共计有 9687 个数据点。该聚类过程有效地将大部分分散的噪声数据成功隔离。接着将经过流片验证的高质量晶片数据按照相同的模型进行缩放和 PCA 降维处理, 将其整合到同一图像中, 得到图 9。

由图 9 可以观察到, 长波段 ($0.194 \leq x \leq 0.231$) 和中波段 ($0.259 \leq x \leq 0.283$) 晶片的聚类结果与实际情况高度吻合, 这些晶片主要分布于图像中 V 字形状的两侧。

3.3 判别边界

运用 alpha shape 算法来界定聚集规模最大的簇

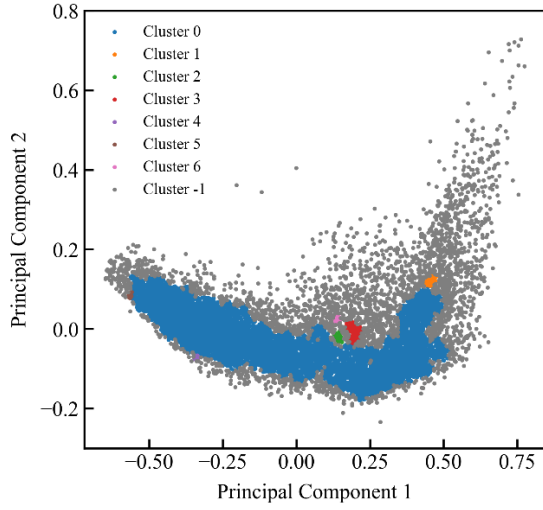


图 8 对初始数据进行 DBSCAN 聚类算法后的图像
Fig. 8 Cluster plot after DBSCAN clustering algorithm on original data

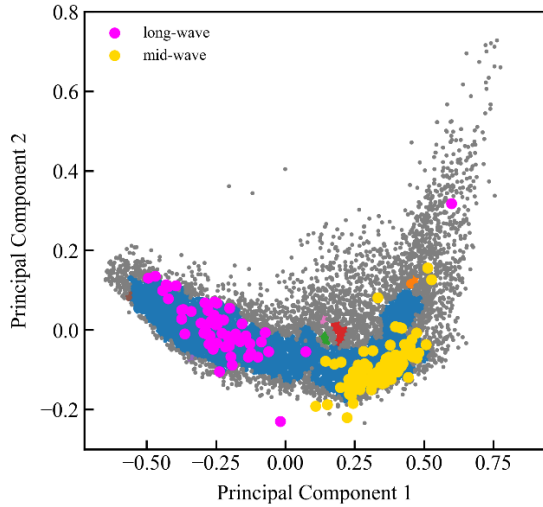


图 9 优质长波和中波材料的主成分图像
Fig. 9 Principal component plot of high-quality long-wave and mid-wave materials

的边界,并根据 x 和 y 方向的平均值和标准差画出判别椭圆,将其作为判定高质量晶片的标准,具体示例如图 10 和图 11 所示。图 10 呈现了长波段体晶材料所在的边界,而图 11 则描绘了中波段体晶材料的边界。长波材料和中波材料的椭圆方程分别为:

$$\begin{aligned} & ((x\cos(45) + y\sin(45)) \times 0.254 - 0.292)^2 / \\ & 0.505^2 + ((-x\sin(45) + y\cos(45)) \times 0.143 + \end{aligned}$$

$$0.0261)^2 / 1.321^2 = 1 \quad (13)$$

$$\begin{aligned} & ((x\cos(45) + y\sin(45)) \times 0.230 + 0.333)^2 / \\ & 1.284^2 + ((-x\sin(45) + y\cos(45)) \times 0.139 - \\ & 0.0935)^2 / 0.594^2 = 1 \quad (14) \end{aligned}$$

基于优质晶片数据点的聚集现象,把在椭圆内的数据点定义为 A 类晶片,在椭圆外的数据点定义为 B 类晶片,基于椭圆曲线和降维公式,可以把初始晶片电学参数和光学参数对应起来,得到长波材料和中波材料的判别公式分别为:

$$\begin{aligned} M = & 3.92 \times (-0.00398 T + 2.54n_1 \times 10^{-17} + \\ & 2.33n_2 \times 10^{-18} - 0.0162\rho_1 + 0.0788\rho_2 - 0.108x \\ & - 1)^2 + 6.93 \times (0.0484 T + 6.48n_1 \times 10^{-18} + \\ & 6.25n_2 \times 10^{-18} + 0.0950\rho_1 + 0.204\rho_2 + 0.174x \\ & - 1)^2 - 1 \quad (0.194 \leq x \leq 0.231) \quad (15) \end{aligned}$$

$$\begin{aligned} M = & 0.339 \times (-0.0436 T + 2.27n_1 \times 10^{-16} + \\ & 1.99n_2 \times 10^{-17} - 0.161\rho_1 + 0.674\rho_2 - x \\ & - 0.680)^2 + 5.21 \times (0.0262 T + 2.52n_1 \times 10^{-18} + \\ & 3.28n_2 \times 10^{-18} + 0.0517\rho_1 + 0.107\rho_2 + 0.0976x \\ & - 1)^2 - 1 \quad (0.258 \leq x \leq 0.284) \quad (16) \end{aligned}$$

当 $M < 0$, 判别为 A 类晶片,当 $M > 0$, 判别为 B 类晶片。经过验证,模型对优质长波晶片的覆盖率为 91.4%,对优质中波晶片的覆盖率为 90.6%。通过对整个晶片数据组进行筛选,得到长波段 A 类晶片占比为 67.8%,中波段 A 类晶片占比为 54.3%。

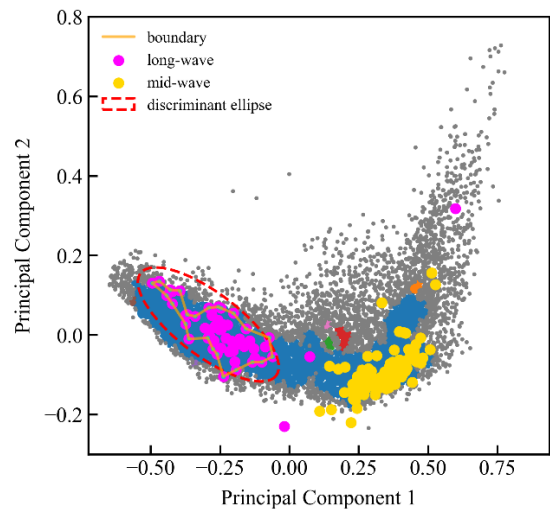


图 10 优质长波材料的判别椭圆图像
Fig. 10 Discriminant ellipse plot of high-quality long-wave materials

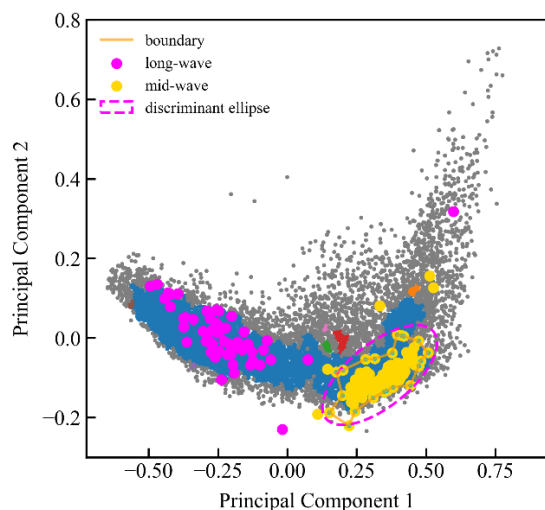


图 11 优质中波材料的判别椭圆图像

Fig. 11 Discriminant ellipse plot of high-quality mid-wave materials

4 总结

本文提出了一种基于主成分分析和聚类方法的碲镉汞晶片参数筛选方法,针对包括电学参数和光学参数的碲镉汞晶片测试数据,利用主成分分析对高维度的数据进行降维,并通过 DBSCAN 聚类算法找到数据分布最密集的区域。最后使用 alpha shape 算法确定聚类区域的边界椭圆,作为判别优质晶片质量的标准,得到了较好的准确率。该方法利用经典算法,通过数据驱动实现了晶片参数的筛选,与传统经验方法相比,提高了筛选效率,使碲镉汞晶片筛选更具科学客观性。今后可继续收集更多碲镉汞晶片数据,扩充样本规模,并从更多维度考虑影响碲镉汞晶片质量的因素,进一步提升筛选的准确性。本研究也为后续大规模快速筛选碲镉汞芯片性能数据并在碲镉汞材料参数与芯片性能之间建立映射关系提供了基础。

References

[1] Li X Y, Sang M S, Xu G Q, *et al.* Simulation on the saturation properties of room-temperature mid-wave infrared HgCdTe detectors[J]. (李向阳, 桑茂盛, 徐国庆, 乔辉, 等. 室温中波红外碲镉汞探测器激光辐照饱和特性的仿真. 红外与毫米波学报), Journal of Infrared and Millimeter Waves, 2023, **42**(02): 143-148.

[2] Sang M S, Xu G Q, Qiao H, *et al.* High speed uncooled MWIR infrared HgCdTe photodetector based on graded bandgap structure[J]. (桑茂盛, 徐国庆, 乔辉, 等. 基于梯度能带结构的高速非制冷中波红外 HgCdTe 探测器.

红外与毫米波学报), Journal of Infrared and Millimeter Waves, 2022, **41**(6): 972-979.

[3] Higgins, William M. *et al.* "Standard relationships in the properties of Hg_{1-x}Cd_xTe." *Journal of Vacuum Science and Technology* **7** (1989): 271-275.

[4] Zhonghua D, Zhifang L. The evolution of scientific research paradigm: The fourth paradigm of scientific research in the Era of big data[J]. *Information and Documentation Services*, 2013, **34**(4): 19-23.

[5] Philips' Gloeilampenfabrieken O. A method of measuring specific resistivity and Hall effect of discs of arbitrary shape [J]. *Philips Res. Rep.*, 1958, **13**(1): 1-9.

[6] Yang J R. Physics and Technology of HgCdTe Materials [M]. National Defense Industry Press, Beijing. (杨建荣碲镉汞材料物理与技术. 北京: 国防工业出版社) 2012, -808431-3.

[7] Vidal R, Ma Y, Sastry S S, *et al.* Principal component analysis [J]. *Generalized principal component analysis*, 2016: 25-62.

[8] A. Weingessel and K. Hornik, "Local PCA algorithms," in *IEEE Transactions on Neural Networks*, vol. **11**, no. 6, pp. 1242-1250, Nov. 2000, doi: 10.1109/72.883408. Vidal R, Ma Y, Sastry S S, *et al.* Principal component analysis [J]. *Generalized principal component analysis*, 2016: 25-62.

[9] Greenacre M, Groenen P J F, Hastie T, *et al.* Principal component analysis [J]. *Nature Reviews Methods Primers*, 2022, **2**(1): 100. Vidal R, Ma Y, Sastry S S, *et al.* Principal component analysis [J]. *Generalized principal component analysis*, 2016: 25-62.

[10] Sehgal S, Singh H, Agarwal M, *et al.* Data analysis using principal component analysis [C]//2014 international conference on medical imaging, m-health and emerging communication systems (MedCom). IEEE, 2014: 45-48.

[11] Groth D, Hartmann S, Klie S, *et al.* Principal components analysis [J]. *Computational Toxicology: Volume II*, 2013: 527-547.

[12] Kong X, Hu C, Duan Z, *et al.* Generalized principal component analysis [J]. *Principal Component Analysis Networks and Algorithms*, 2017: 185-233.

[13] Schubert E, Sander J, Ester M, *et al.* DBSCAN revisited, revisited: why and how you should (still) use DBSCAN [J]. *ACM Transactions on Database Systems (TODS)*, 2017, **42**(3): 1-21.

[14] Khan K, Rehman S U, Aziz K, *et al.* DBSCAN: Past, present and future [C]//The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014). IEEE, 2014: 232-238.

[15] Hou J, Gao H, Li X. DSets-DBSCAN: A parameter-free clustering algorithm [J]. *IEEE Transactions on Image Processing*, 2016, **25**(7): 3182-3193.

[16] Edelsbrunner H, Mücke E P. Three-dimensional alpha shapes [J]. *ACM Transactions On Graphics (TOG)*, 1994, **13**(1): 43-72.

[17] Edelsbrunner H. Alpha shapes—a survey [M]// *Tessellations in the Sciences: Virtues, Techniques and Applications of Geometric Tilings*. 2011.

[18] Ohbuchi R, Takei T. Shape similarity comparison of 3D models using alpha shapes [C]//11th Pacific Conference on Computer Graphics and Applications, 2003. Proceed-

ings. IEEE, 2003: 293–302.