

文章编号 1004-924X(2025)09-1341-12

基于自注意力神经网络的低信噪比 光谱干涉膜厚测量

王 晨, 王子政, 刘墨燃, 姚程源, 胡春光*

(天津大学 精密测试技术及仪器全国重点实验室, 天津 300072)

摘要: 为了提高薄膜厚度测量对低信噪比光谱数据的稳定性, 提出了一种基于自注意力神经网络(Self-Attention-based Neural Network, SANN)的测量方法。由于传统傅里叶变换方法在信噪比降低时噪声成分可能掩盖主要干涉频率, 难以准确提取厚度信息, 因此构建了一种以光谱数据为输入, 薄膜厚度为输出的自注意力神经网络模型, 利用自适应注意力机制对不同波长的光谱点赋予动态权重, 以增强对低信噪比光谱数据的解析能力。采用光谱干涉膜厚测量系统采集实验数据, 并通过波长偏移和光强归一化动态调整策略进行数据增强, 用以扩充训练集并提高模型的泛化能力。该系统优化了基于自注意力机制的编码器层数及隐藏节点数, 最终选定包含 8 层编码器、每层 128 个隐藏节点的模型。以晶圆为例进行验证, 测试含有异常值的光谱数据集, 结果显示该模型在低信噪比测试集上的最大相对厚度测量误差为 3.62%, 证明该方法能有效抑制噪声影响, 避免傅里叶变换方法中常见的异常值偏差, 显著提升测量稳定性。所提方法可扩展至更广泛的薄膜测量应用中。

关键词: 干涉测量; 晶圆厚度; 光谱干涉式; 自注意力神经网络; 抗噪声能力; 测量稳定性

中图分类号: TH741 **文献标识码:** A

doi: 10. 37188/OPE. 20253309. 1341

CSTR: 32169. 14. OPE. 20253309. 1341

Low signal-to-noise ratio spectral interferometry film thickness measurement based on self-attention neural network

WANG Chen, WANG Zizheng, LIU Zhaoran, YAO Chengyuan, HU Chunguang*

(State Key Laboratory of Precision Measurement Technology and Instruments,
Tianjin University, Tianjin 300072, China)

* Corresponding author, E-mail: cghu@tju.edu.cn

Abstract: To enhance the robustness of film thickness measurements from low signal-to-noise ratio (SNR) spectral data, a measurement approach based on a self-attention neural network (SANN) is developed. While the conventional Fourier transform method effectively measures thickness on high SNR data, its accuracy deteriorates as noise obscures the principal interference frequency under low SNR conditions, hindering precise thickness extraction. This study introduces a self-attention neural network model that takes spectral data as input and outputs film thickness, employing an adaptive attention mechanism to dynamically weight spectral points across different wavelengths, thereby improving analysis of low SNR

收稿日期: 2025-03-03; 修订日期: 2025-03-25.

基金项目: 国家重点研发计划资助项目(No. 2022YFF0708300); 国家自然科学基金面上项目(No. 52475566)

spectral data. Experimental data were obtained using a spectral interference film thickness measurement system and subsequently augmented through wavelength drift and adaptive intensity normalization strategies to expand the dataset and enhance the model's generalization. Model optimization identified an architecture comprising eight encoder layers and 128 hidden nodes per layer. Using wafer measurements as a case study, evaluation on spectral data containing outliers demonstrated a maximum relative thickness measurement error of 3.62% on the low SNR validation set. These results indicate that the proposed method effectively suppresses noise influence, mitigates outlier deviations common in Fourier transform approaches, and substantially improves measurement stability. the applicability of the proposed method is validated to a broader range of thin film measurement scenarios.

Key words: interferometric measurement; wafer thickness; spectral interference; self-attention neural network; anti-noise ability; measurement robustness

1 引 言

晶圆厚度测量是半导体制造、光学镀膜和微电子器件加工中的核心环节,在 3D 集成电路、极紫外光刻和高精度纳米制造中尤为关键^[1-3]。随着半导体制程进入 3 nm 时代,晶圆厚度的精确控制对于提升器件性能和良品率至关重要。然而,在实际测量中,晶圆表面划痕、污染及测量系统的固有噪声会降低光谱信号的信噪比(Signal-to-Noise Ratio, SNR),从而引入厚度测量误差。因此,开发能够适应低信噪比环境的高稳定性测量方法,是当前亟待解决的问题。

目前,已有多种厚度测量方法应用于工业生产。其中,扫描电子显微镜(Scanning Electron Microscope, SEM)通过高分辨率成像直接测量厚度,但因破坏性制样和高昂成本,不适用于在线检测^[4-7]。光谱椭圆偏法^[8]采用模型反演法,可同时表征厚度、折射率和吸收率,但对测量模型和观测数据的匹配程度的依赖性较强^[9-10],且计算复杂度高^[11-13]。傅里叶变换方法通过将光谱变换到频域进而分析提取厚度信息,计算效率高,广泛用于晶圆厚度测量^[14-15]。然而,傅里叶变换依赖高信噪比光谱数据,在噪声干扰下易导致频谱峰值偏移或失真,进而影响测量精度^[16-17]。

近年来,深度学习因其强大的非线性拟合能力,广泛应用于光谱测量领域^[18-21]。Wang 等提出 EllipsoNet 模型,利用深度学习方法从复杂多层基底的光学反射光谱中预测薄膜复折射

率^[22]。Choi 等开发了一种基于神经网络的逆向分析方法,用于从光谱数据中评估氧化物薄膜的光学常数,展示了机器学习在光学测量中的潜力^[23]。Kwak^[24]等开发了一种基于可支持向量机(Support Vector Machine, SVM)的多层半导体设备膜厚测量算法,使用测量光谱对多层结构进行了准确测量。然而,传统的卷积神经网络(Convolutional Neural Network, CNN)和循环神经网络(Recurrent Neural Network, RNN)在光谱测量任务中存在一定局限性。CNN 仅能提取局部特征,难以捕捉远距离波长点位之间的相对关系,导致测量误差;RNN 依赖于时序结构,但光谱数据并无明显时间顺序,且 RNN 存在梯度消失问题,难以处理长序列波长数据^[25-26]。此外,这些研究主要针对高信噪比数据,较少涉及低信噪比数据的评估分析。因此,探索更具测量稳定性和泛化性的神经网络结构,以提高对低信噪比光谱数据的测量精度,仍然是一个值得研究的方向。

自注意力机制(Self-Attention Mechanism, SAM)在处理长序列数据方面表现优越,已成功应用于自然语言处理(Natural Language Processing, NLP)等领域^[27-28]。Liu 等提出的 Swin-Transformer^[29]模型采用层次化自注意力机制窗口,在编码阶段提升对于输入数据的理解和全局学习能力,提高了图像识别的分类精度。相比于 CNN 和 RNN,基于 SAM 的神经网络(SANN)在光谱厚度测量中的优势包括:全局建模能力,能够同时计算所有波长点之间的权重关系,适用于光谱干涉测量;动态权重分配,能够在训练过程

中自适应调整不同波长的注意力权重,增强噪声的稳定性;计算效率高,SANN可并行计算所有光谱点的关系^[27],而RNN需要逐点计算,计算速度更快。

本文构建了SANN模型,以光谱数据为输入,晶圆厚度为输出,在噪声背景下学习测量光谱与厚度的非线性映射关系。利用光谱干涉式膜厚测量系统采集实验数据,并结合数据增强技术构建训练集和验证集。优化网络结构,通过实验比较不同编码器层数及隐藏节点数,最终选定包含8层编码器、每层128个隐藏节点的最优模型。为了评估抗噪性能,在低信噪比环境下测试该模型,并与傅里叶变换方法进行对比,验证其稳定性和测量准确性。

2 原 理

2.1 光学模型

光谱干涉膜厚测量基于多光束干涉原理,当光束在薄膜表面和底部多次反射后,叠加形成光谱振荡条纹。图1展示了入射光在膜层内的反射和透射过程。

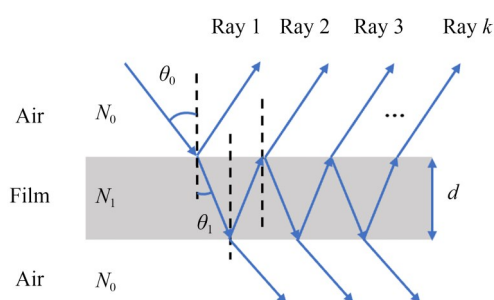


图1 由空气层、膜层和空气层组成的硅片样品示意图
Fig. 1 Schematic diagram of silicon wafer sample consisting of air layer, film layer, and air layer

相位厚度 δ_1 表示为层厚度 d 、波长 λ 和复折射指数 $N_1(\lambda)$ 的函数,即:

$$\delta_1(d, \lambda, N_1(\lambda)) = \frac{2\pi}{\lambda} d N_1(\lambda) \cos \theta_1. \quad (1)$$

多次反射后,复反射系数 r 表示为:

$$r(d, \lambda, N_1(\lambda)) = \frac{r_{01} + r_{12} e^{-j2\delta_1}}{1 + r_{01} r_{12} e^{-j2\delta_1}}. \quad (2)$$

由此得到反射率 R 的模型:

$$R(d, \lambda, N_1(\lambda)) = r \cdot r^* = \left| \frac{r_{01} + r_{12} e^{-j2\delta_1}}{1 + r_{01} r_{12} e^{-j2\delta_1}} \right|^2. \quad (3)$$

膜层厚度 d 变化时,相位厚度 δ_1 发生变化,影响反射光谱的干涉模式。采用傅里叶变换解算光谱信号时,厚度 d 依赖于光谱信号在频率域中的频率峰值 f_d 。然而,在低信噪比条件下,噪声会引入额外的频率分量并导致峰值振幅衰减,傅里叶变换对干涉信号的解析会出现混淆,降低膜厚测量的准确性和可靠性。设测量的光谱信号包含噪声项:

$$|F_{\text{measured}}(f)|^2 = |F_{\text{true}}(f)|^2 + P_{\eta}(f), \quad (4)$$

式中:噪声的功率谱密度 $P_{\eta}(f)$ 影响傅里叶变换后的信号能量分布,掩盖真实信号的频率峰值。傅里叶变换本身无法区分 $P_{\eta}(f)$ 和 $F_{\text{true}}(f)$,这是由于傅里叶变换是全局变换方法,对所有光谱点赋予相同权重,当干扰较强时,主要的频率成分会被噪声掩盖,导致膜厚测量结果失真。

相比之下,SANN通过动态调整光谱权重,能够自适应降低噪声对厚度计算的影响,增强测量稳定性。因此,本文采用SANN提高对于光谱中厚度信息的解析能力。

2.2 自注意力神经网络

为了在低信噪比光谱数据中提取有效的厚度信息,本文提出了一种基于SANN的方法。相比于CNN,RNN只能捕捉局部或相邻波长点的特征,SANN能够在全局范围内计算不同波长点之间的相关性,增强对光谱干涉模式的提取能力,特别适用于长距离依赖和噪声干扰严重的场景。

2.2.1 光谱特征嵌入

光谱数据由波长与光谱强度组成,由于光谱测量存在非线性关系,不能直接输入神经网络。因此,需要将原始光谱数据 x_i 变换到高维特征空间,以便提取更加丰富的模式信息。输入光谱数据表示为:

$$x = [x_1, x_2, \dots, x_L], \quad (5)$$

式中 x_i 表示第 i 个光谱点的光强值。为了让网络学习到光谱的潜在特征,首先对 x_i 进行线性投影,映射到 d_{model} 维的嵌入向量:

$$e_i = f_{\text{embed}}(x_i) \in \mathbf{R}^{d_{\text{model}}}. \quad (6)$$

从而形成整个光谱的输入矩阵:

$$E = [e_1, e_2, \dots, e_L] \in \mathbb{R}^{L \times d_{\text{model}}}, \quad (7)$$

其中 L 代表原始光谱数据 x 的长度。由于光谱数据是有序的, 相邻波长点之间具有一定的物理相关性, 因此在输入网络之前, 需要加入位置信息。引入可学习的位置编码:

$$P = [p_1, p_2, \dots, p_L] \in \mathbb{R}^{L \times d_{\text{model}}}, \quad (8)$$

式中: p_i 为序列中第 i 个位置的可学习位置向量。在自注意力神经网络的输入端, 将 p_i 和原始嵌入向量 e_i 相加, 得到包含位置信息的输入向量 z_i :

$$z_i = e_i + p_i, i = 1, \dots, L. \quad (9)$$

最终, 完整的输入表示为:

$$Z = E + P. \quad (10)$$

Z 作为后续多头注意力模块的输入, 用以兼顾特征信息与位置信息, 可学习的位置编码不同于固定位置编码, 能够根据数据自适应调整权重, 增强网络的稳定性和泛化能力。

2.2.2 自注意力计算

在 SAM 中, 每个输入光谱点 e_i 通过构造查询 (Q)、键 (K) 和值 (V) 向量以实现特征变换。输入序列 E 经过可学习的线性映射, 分别生成对应的查询、键和值向量集合。随后, 通过计算查询与所有键之间的相似性权重, 对值向量进行加权求和, 从而完成对输入特征的自适应建模与上下文感知的表达。

$$Q = ZW^Q, K = ZW^K, V = ZW^V, \quad (11)$$

其中: $W^Q, W^K, W^V \in \mathbb{R}^{d_{\text{model}} \times d_k}$ 为可学习参数矩阵, d_k 代表查询和键向量的维度, $Q, K, V \in \mathbb{R}^{L \times d_k}$ 分别表示查询、键和值的映射结果。

然后, 通过相似度加权求和得到最终特征表达:

$$A = \frac{QK^T}{\sqrt{d_k}}, \quad (12)$$

式中: $A \in \mathbb{R}^{L \times L}$ 是注意力分数矩阵。然后, 使用 SoftMax 归一化得到注意力权重:

$$\alpha_{ij} = \frac{\exp\left(\frac{q_i \cdot k_j}{\sqrt{d_k}}\right)}{\sum_{m=1}^L \exp\left(\frac{q_i \cdot k_m}{\sqrt{d_k}}\right)} = \frac{\exp(A_{ij})}{\sum_{m=1}^L \exp(A_{im})}, \quad (13)$$

其中 α_{ij} 反映第 i 个光谱点对第 j 个光谱点包含信息的依赖程度。由于 SoftMax 输出, 行向量和为

1, 每个位置对所有位置的注意力权重总和为 1。

最终, 计算注意力输出:

$$O = aV, \quad (14)$$

其中: $O \in \mathbb{R}^{L \times d_k}$ 是经过自注意力变换后的特征表示, 每个位置 i 的输出 o_i 是由所有波长点 j 的信息加权计算而得。

2.2.3 多头自注意力机制

为了增强模型的表达能力, 引入多头注意力:

$$\text{MultiHead} = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O. \quad (15)$$

每个注意力头 head_i 的计算方式如下:

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V), \quad (16)$$

式中: $W_i^Q, W_i^K, W_i^V \in \mathbb{R}^{d_{\text{model}} \times d_k}$ 是第 i 个头的参数矩阵。多头注意力的作用包括同时关注不同尺度的特征: 某些注意力头可关注低频信息 (如光谱大趋势), 另一些注意力头可关注高频信息 (如干涉细节); 因此, 这种聚焦模式的多样性提高了稳定性, 即使某些注意力头受噪声影响, 整体仍可稳定学习有效信息。最终, 将所有注意力头的输出拼接后, 通过线性变换恢复至 d_{model} 维度, 如式 (15) 所表示, 其中, $W^O \in \mathbb{R}^{(h \times d_k) \times d_{\text{model}}}$ 。

本文设计了 6 种不同结构的 SANN, 编码器层数取 6, 8 和 10, 每层隐藏节点数则取 128 和 256, 并通过实验选出最优结构。SANN 处理光

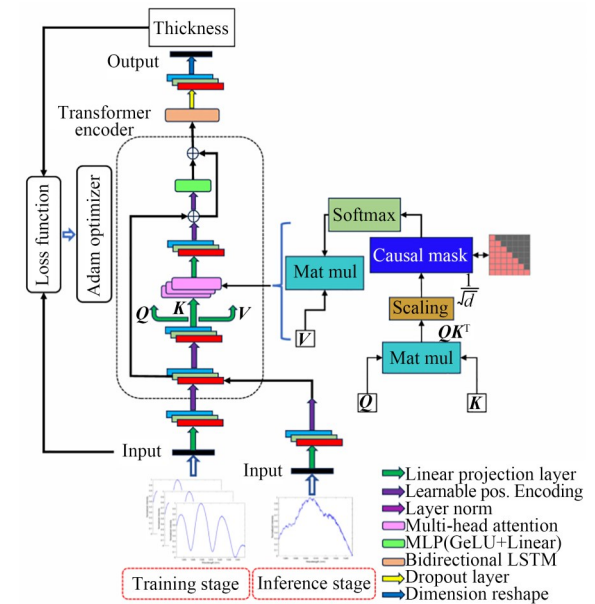


图2 自注意力神经网络处理光谱数据流程

Fig. 2 Flowchart for self-attention neural network processing spectral data

谱数据流程如图2所示。参数初始化使用Xavier初始化权重,激活函数采用高斯误差线性单元(GELU),损失函数采用均方误差(Mean Square Error, MSE),优化器采用Adam,初始学习率为 1×10^{-4} ,学习率衰减采用指数衰减,步长为100,衰减因子为0.5。

3 实验

3.1 光谱干涉式膜厚测量系统

图3展示了光谱干涉式膜厚测量系统的基本构成,该系统包括超辐射发光二极管(SLD)光源、环形器、光纤探头及光谱仪。其中,光源采用中心波长为1310 nm、光谱带宽为40 nm的SLD,提供宽带照明以增强干涉条纹的分辨能力。环形器用于减少探测光的端面反射干扰。光纤探头的作用是会聚光斑并耦合反射信号,探头的焦距和数值孔径(NA)直接影响测量区域和信号收集效率。光谱仪的工作波段为1270~1340 nm,采用512像素制冷型线阵探测器进行光谱数据采集,并利用傅里叶变换方法解析光谱信号。

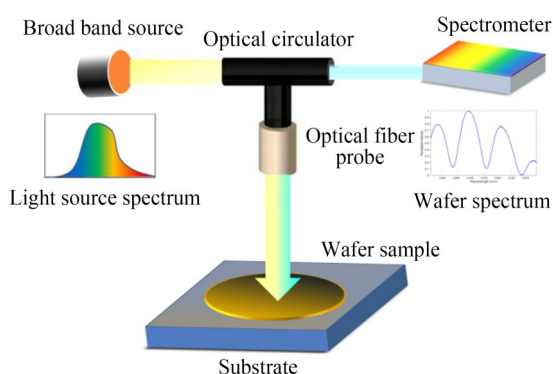


图3 光谱干涉式膜厚测量系统

Fig. 3 Configuration of spectral interferometry film thickness measurement system

晶圆在前期加工过程中产生的磨纹和表面粗糙度会导致反射光的散射增强,降低干涉信号的可见度。当散射效应增强时,光谱中的干涉条纹可能被背景噪声淹没,从而影响厚度测量的准确性。

在信号频谱分析中,功率谱反映了光谱振荡

成分的能量分布。干涉光谱信号的功率谱在特定频率处的峰值对应膜厚信息。置信度为某一频率分量的功率谱幅值占总功率的比重,定义为:

$$C_f = \frac{P(f)}{\sum P(f)}, \quad (17)$$

式中: $P(f)$ 是特定厚度的功率谱值, $\sum P(f)$ 是信号频谱的总功率。置信度 C_f 越高,代表该频率成分在光谱信号中占据主导地位,测量结果可靠性高。这种方法作为傅里叶变换方法的衍生方法,利用置信度剔除傅里叶变换解算结果的异常值,但仍然无法有效解算低信噪比的光谱数据。

为了避免上述方法在低信噪比条件下的局限性,需要利用本文提出的基于SANN的低信噪比光谱干涉膜厚测量方法。其中,SAM能够动态估计权重,针对不同波长处光谱点自适应调整注意力,增强对有效厚度信息的关注,同时抑制噪声的干扰。

3.2 自注意力神经网络训练

本文采用光谱干涉式膜厚测量系统(安装在光学平台上)进行光谱数据采集,采用SLD-1310宽光谱光源,并使用光谱仪采集。实验在 0° 入射角下,对7种标称厚度(5,6,8,10,15,20,50 μm)的晶圆进行测量,每个晶圆上采集30~50处的光谱数据。

由于实验条件限制,难以制备足够数量的不同厚度样本,采用数据增强方法扩充训练集。首先分析光谱功率谱,筛选出280个高置信度样本,并随机划分238个训练样本和42个验证样本。利用波长偏移(± 10 nm)和光强归一化动态调整($\pm 5\% \sim 20\%$)对训练集进行数据增强,每个训练样本生成20个增强数据点,最终构建4760个训练数据点。所有增强样本共享相同的厚度轮廓,以确保网络能够学习到稳定的厚度特征。其次,采集到的高置信度样本的反射光谱及其离散傅里叶变换(Discrete Fourier Transform, DFT)后的幅值如图4所示。

在NVIDIA RTX TITAN上训练所提出的SANN,总时长大约3 h。训练过程中,采用MSE作为损失函数,最终验证集误差收敛至 2.6×10^{-3} ,未观察到过拟合现象。

为了评估神经网络的预测性能,使用商用膜

厚测量仪测量 3 种晶圆, 以此作为参考值。表 1 给出实验样品的厚度信息, 表 2 为 6 种不同 SANN 的测试结果。

在 6 种 SANN 结构中, 仅 1 种网络能够较准确地预测厚度, 其结构包括 8 层自注意力编码器,

每层 128 个隐藏节点。实验发现, 编码器层数少于 8 层时, 模型对高噪声光谱数据的学习能力较弱; 而当层数超过 10 层时, 模型复杂度增加但数据集不足以支撑高复杂度模型进行充分特征学习, 导致精度未进一步提升。

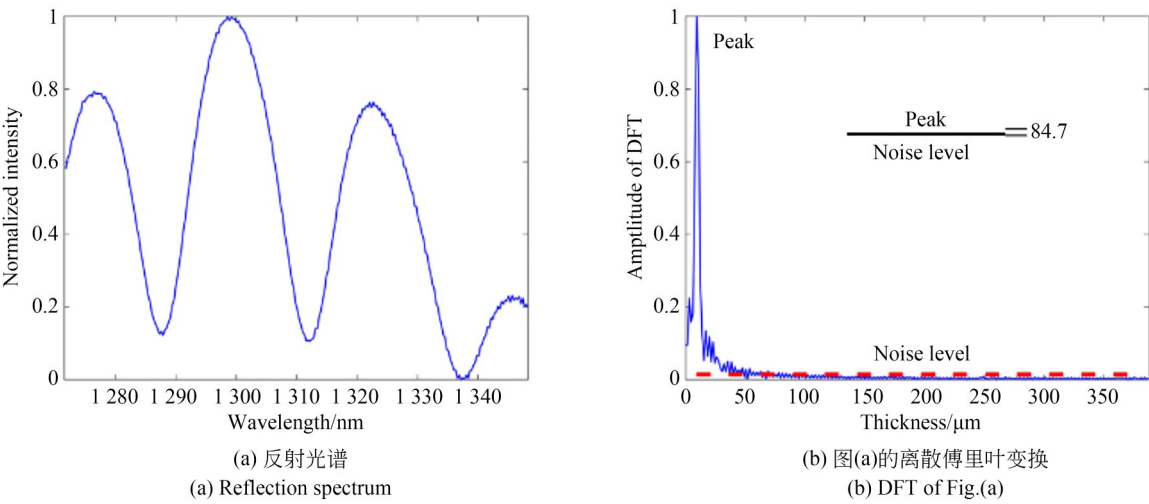


图 4 晶圆的反射光谱及其离散傅里叶变换
Fig. 4 Reflection spectrum of wafer and its Discrete Fourier Transform (DFT)

表 1 晶圆样品的厚度信息

Tab. 1 Thickness information of wafer samples (μm)				
Sample number	Nominal thickness	DFT measured thickness	Reference thickness	DFT measured deviation
No. 1	5	3.10	2.106	0.994
No. 2	8	7.19	7.216	0.026
No. 3	15	13.89	13.690	0.200

表 2 自注意力神经网络的测试结果

Tab. 2 Test results of self-attention neural network (μm)						
Node layer	$\delta_a(128)$			$\delta_a(256)$		
	No. 1	No. 2	No. 3	No. 1	No. 2	No. 3
6	0.806	0.082	0.233	0.808	0.124	0.178
8	0.579	0.014	0.102	0.857	0.048	0.084
10	0.844	0.100	0.086	0.862	0.060	0.108

3.3 自注意力神经网络评价

图 5(a)展示了傅里叶变换解算的 7 个晶圆样品的厚度分布。在测量的厚度偏差中, 离散傅里叶变换计算的最大相对测量误差是 2.1%, SANN 方法的最大相对测量误差是 1.5%。

图 5(b)展示了 DFT 的计算厚度和 SANN 预测厚度之间的均方根误差(Root Mean Square Error, RMSE)。采用 8 层编码器、每层 128 个隐藏节点的 SANN 进行预测, 结果表明, 基于 SAM 的神经网络可以准确地预测晶圆厚度。在相同的评价

标准下,SANN和DFT在整个厚度范围内表现出较为一致的测量精度。

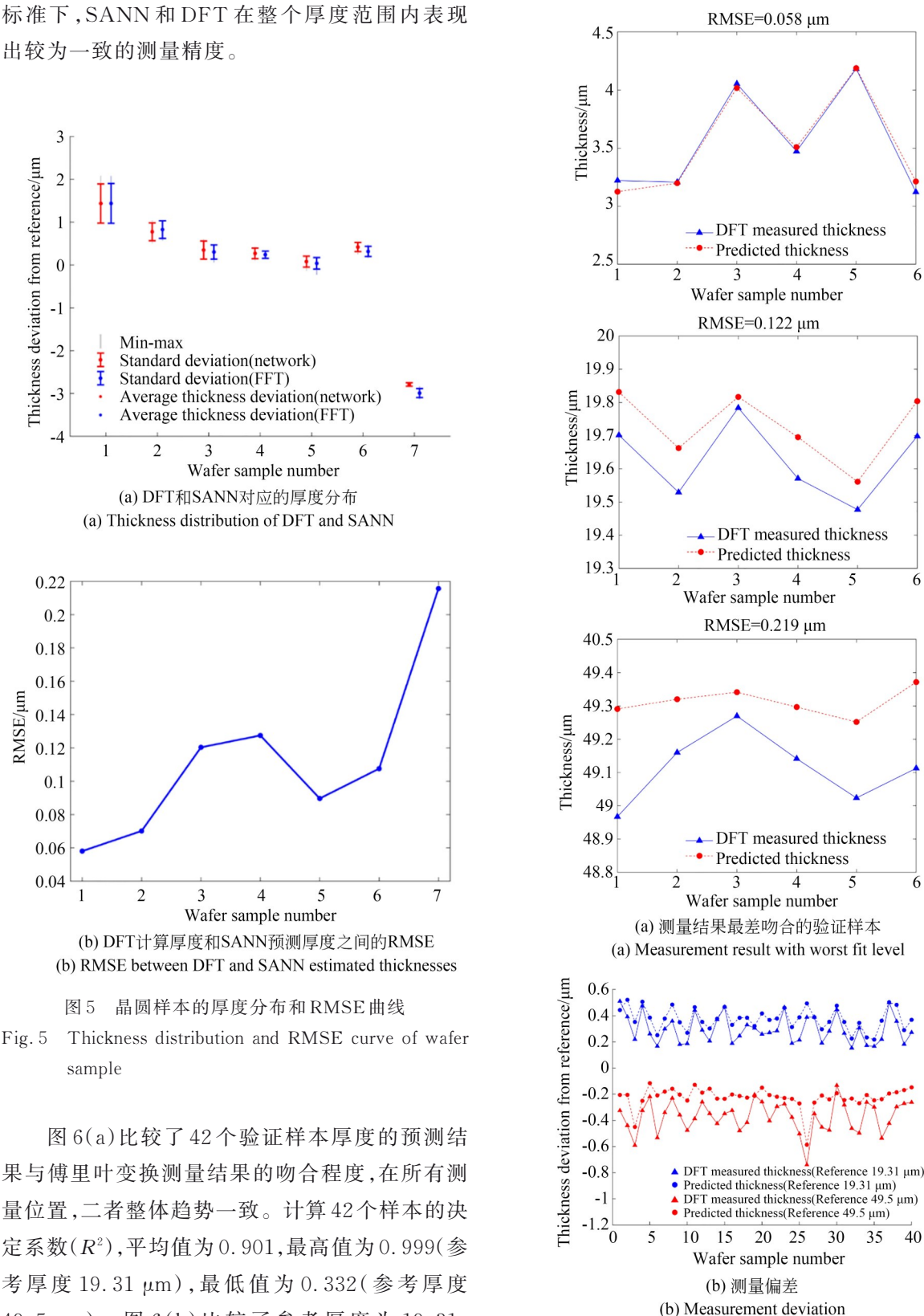


图5 晶圆样本的厚度分布和RMSE曲线
Fig. 5 Thickness distribution and RMSE curve of wafer sample

图6(a)比较了42个验证样本厚度的预测结果与傅里叶变换测量结果的吻合程度,在所有测量位置,二者整体趋势一致。计算42个样本的决定系数(R^2),平均值为0.901,最高值为0.999(参考厚度 $19.31\text{ }\mu\text{m}$),最低值为0.332(参考厚度 $49.5\text{ }\mu\text{m}$)。图6(b)比较了参考厚度为 19.31 , $49.5\text{ }\mu\text{m}$ 晶圆的测量结果。总之,在高信噪比光谱数据上,晶圆厚度的预测结果和傅里叶变换结果具有高度相关性。

图6 SANN和DFT方法对样本的测量吻合结果和测量偏差
Fig. 6 Measurement fitting results and deviations of samples using SANN and DFT methods

4 实验和结果讨论

4.1 低信噪比环境下的测量结果

为了验证 SANN 在低信噪比条件下的测量精度,对标称厚度为 5, 8 和 15 μm 的样本表面的划痕处进行测量,在每片晶圆上 20~30 个不同位置采集光谱数据,这些数据由于表面缺陷导致光强衰减,噪声水平较高。图 7 展示了这 3 片存在划痕的晶圆的测量结果。传统傅里叶变换方法的测量结果异常值多,测量数据呈现较大的波动性,说明噪声对傅里叶变换的影响大, SANN 的预测结果基本平稳且没有异常值,展现出更强的抗噪能力。在所有存在划痕的测试样本中,基于传统傅里叶变换方法的最大相对测量误差为 14.76%;基于 SANN 方法的最大相对测量误差

为 3.62%,表明 SANN 能够在低信噪比条件下进行可靠的厚度测量。

为了量化信噪比对测量精度的影响,定义峰值与噪声比(Peak-to-Noise Ratio, PNR),用于衡量光谱信号的信噪比:

$$PNR = \frac{\text{Interference peak intensity}}{\text{Noise level}}. \quad (18)$$

图 8 显示了傅里叶变换和 SANN 对于图 7(c) 正常样本 D 的分析,反射光谱的周期性振荡明显,噪声干扰较少,离散傅里叶变换的测量结果为 13.79 μm ,表明在信噪比较高的情况下,傅里叶变换提取的厚度信息较为准确,如图 8(a) 和 8(b) 所示。SANN 通过自适应权重分配,对光谱数据中的不同像素点进行选择性关注,如图 8(c) 所示,输出的预测厚度为 13.94 μm 。

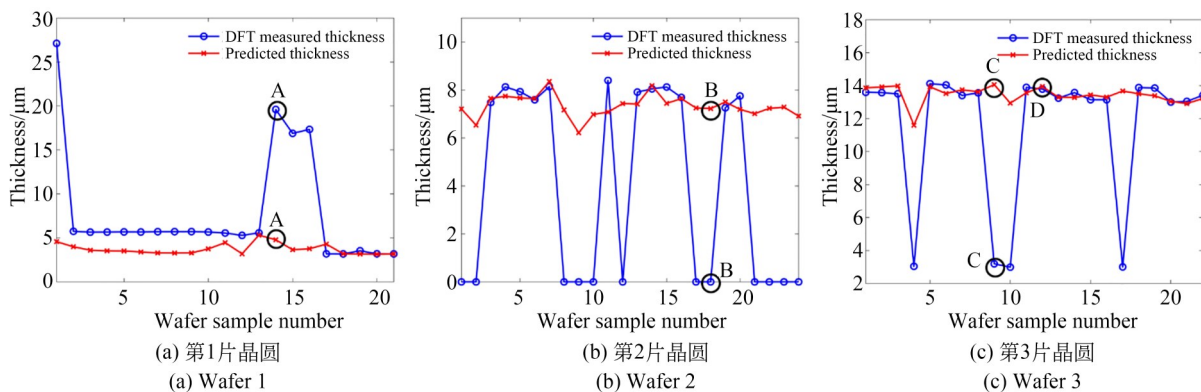


图 7 表面缺陷晶圆的厚度测量结果

Fig. 7 Thickness measurement results of wafer with surface defects

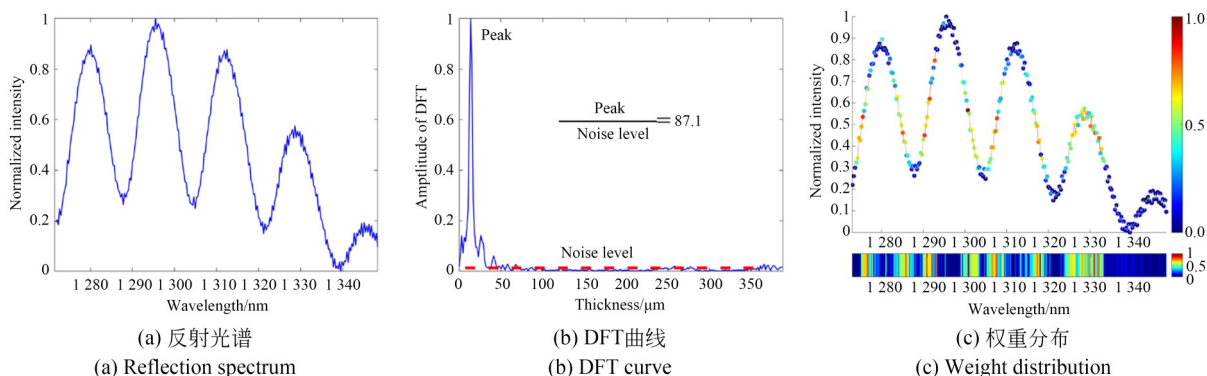


图 8 傅里叶变换和自注意力神经网络对正常样本的分析

Fig. 8 Analysis of normal sample by Fourier transform and self-attention neural network

傅里叶变换的测量精度受限于厚度分辨率和最小可测厚度,厚度分辨率指傅里叶变换解析

的最小厚度变化量。当最小波长和最大波长固定时,分辨率固定为:

$$\Delta d = \frac{1}{2 \times N \times \Delta \lambda} = \frac{1}{2 \times (\frac{1}{\lambda_{\min}} - \frac{1}{\lambda_{\max}})}. \quad (19)$$

本文中的光谱信号采样范围有限,傅里叶变换频谱可能出现旁瓣或峰值展宽现象, Δd 指傅里叶变换后频谱图中产生最小特征峰的厚度,幅值最大处的一侧应至少由2个点位构成,因此最小解析厚度要大于两倍厚度分辨率: $d > 2 \times \Delta d = 6.64 \mu\text{m}$ 。

图7(a)异常样本A中,待测厚度十分接近最小可测厚度,受到频谱泄露影响,厚度对应的特征峰和背景噪声叠加,傅里叶变换不能有效提取特征峰,如图9(a)和10(a)所示。图7(b)和7(c)异常样本B,C中,反射光谱周期性振荡显著减弱,噪声成分显著增强,厚度信息被背景噪声淹没,如图9(b),9(c)和图10(b),10(c)所示。这种情况下,傅里叶变换对所有光谱点赋予相同权重,无法区分噪声和有效信号,测量误差大。

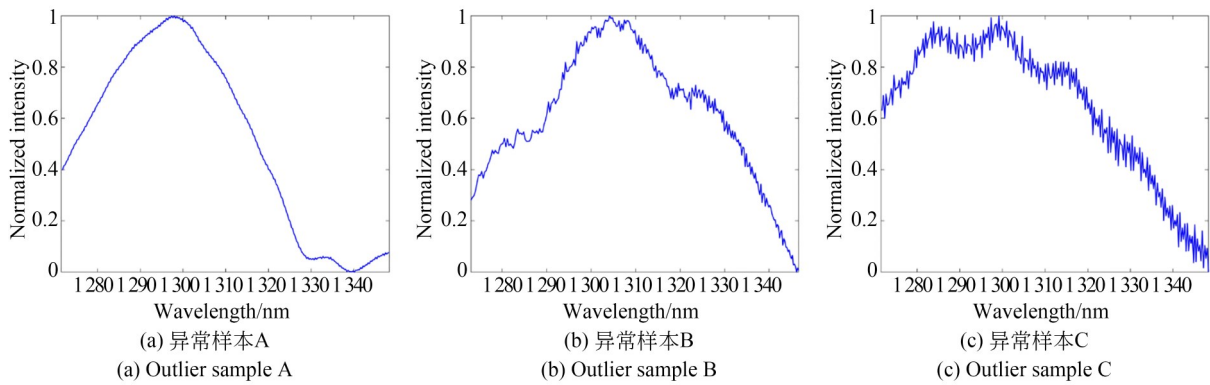


图9 异常样本的反射光谱

Fig. 9 Reflectance spectra of outlier samples

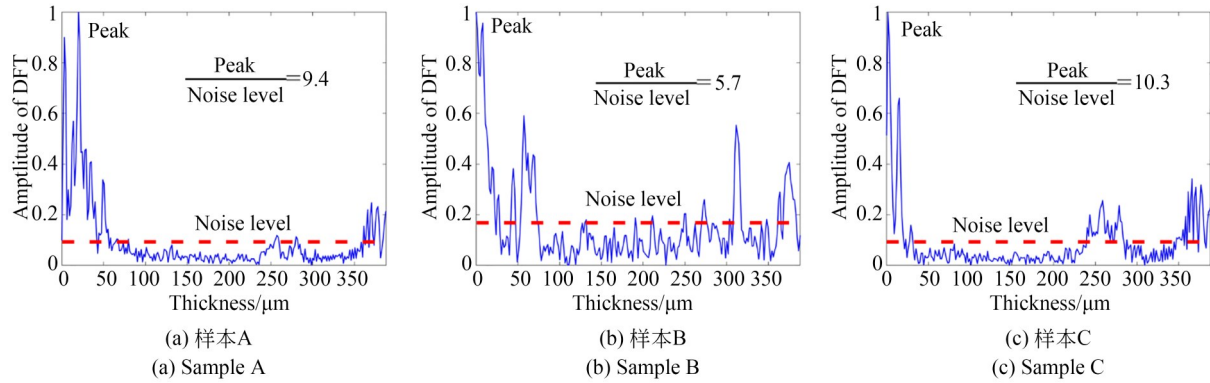


图10 异常样本的DFT曲线

Fig. 10 DFT curves of outlier samples

4.2 自注意力机制的权重分布

图11对SANN网络中的第 $x_i (i=1, 2, \dots, 8)$ 层、第 $y_j (j=1, 2, 3, 4)$ 注意力头的自注意力权重矩阵进行了可视化,能够更好地观察到不同光谱点位之间的相对关系。

图12展示了SANN计算的异常样本光谱数据注意力权重分布,色条由蓝色到红色表示注意力权重在增加。权重越大,说明该波长区域的光

谱点对厚度的预测贡献越大。SANN能够针对高置信度的光谱区域分配更高权重,增强有效厚度信息的提取。

若每个样本的光谱数据包含 L 个波长点,对于某个厚度 h 下的所有样本,定义平均注意力权重为:

$$A_{\text{mean}}^{(h)}(\lambda_j) = \frac{1}{N(h)} \sum_{i=1}^{N(h)} A_i(\lambda_j), \quad (20)$$

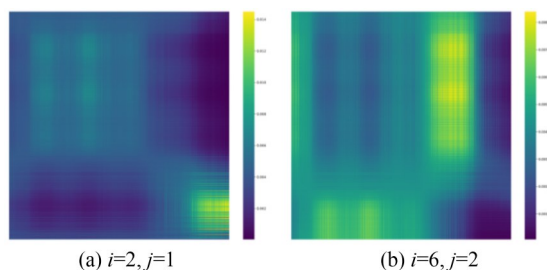


图 11 自注意力神经网络中多头自注意力机制对输入光谱计算的权重矩阵可视化

Fig. 11 Visualization of weight matrix calculated by multi-head self-attention mechanism on input spectrum in self-attention neural network

式中: λ_j 是第 j 个波长点, $A_i(\lambda_j)$ 是第 i 个样本在 λ_j 处的注意力值, 对所有的 $N(h)$ 个样本取均值, 得

到该厚度下的平均注意力分布。

图 13(a)显示了 3 片存在划痕晶圆的样本的平均注意力权重变化趋势, 代表 SANN 模型对不同类型的频谱估计权重的分布。图 13(b)显示了 SANN 模型估计的注意力权重在不同波长点处的分布。可以看到, 即使在低信噪比离群光谱集上进行测试, SANN 模型对于不同点位估计出的权重分布趋势仍然和在高信噪比正常光谱集时得到的权重分布趋势高度一致, 这表明 SANN 模型能够在训练过程中学习到光谱信号内部的通用特征并有着极强的泛化能力, 即使在从未见过的高噪声光谱上进行测试, 此方法依然能够估计出可靠权重值, 并准确估计对应膜厚。

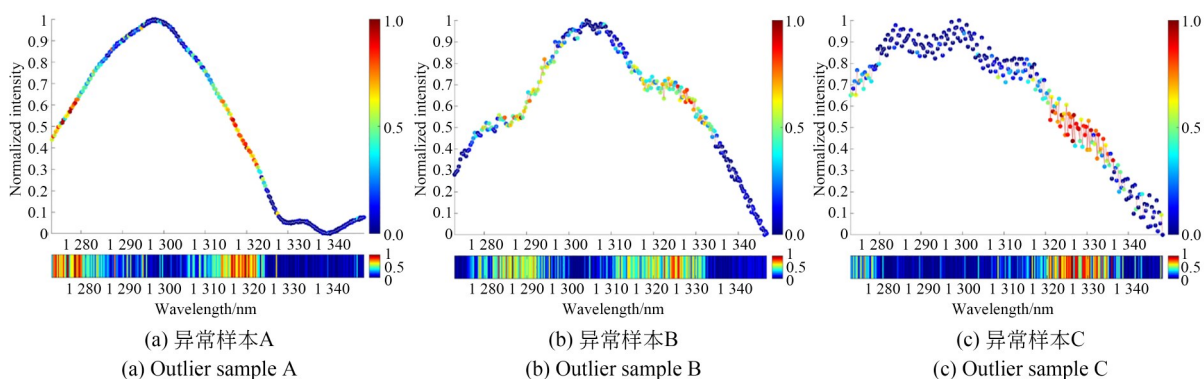


图 12 自注意力神经网络对异常样本的权重分析

Fig. 12 Weight analysis of outlier samples by self-attention neural network

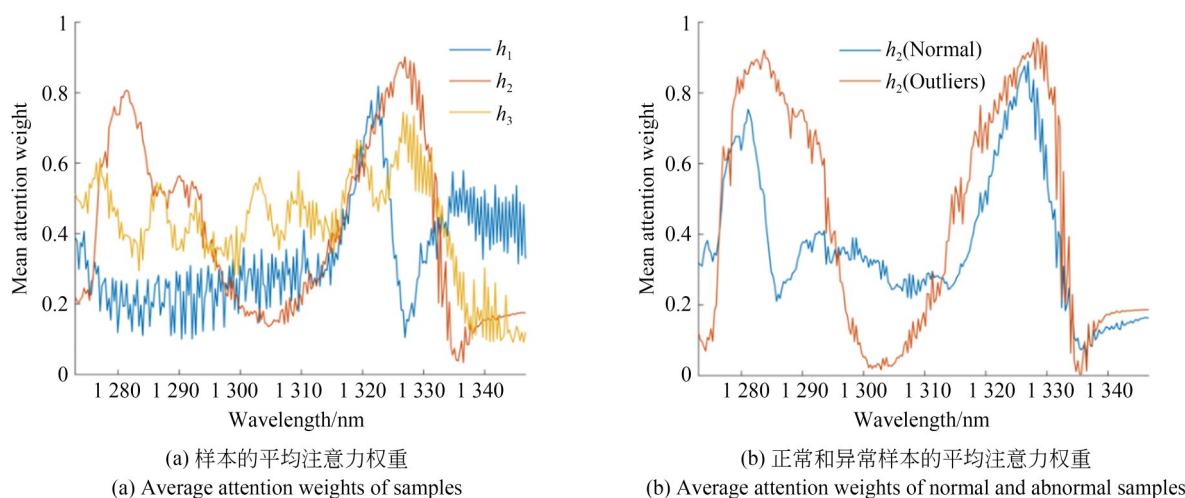


图 13 自注意力神经网络对样本的平均注意力权重曲线

Fig. 13 Average attention weight curves of self-attention neural network for samples

5 结 论

本文提出了一种基于低信噪比光谱干涉膜厚测量的SANN模型,用于提高低信噪比环境下的膜厚测量稳定性。该模型通过SAM学习光谱数据的全局相关性,能够动态调整不同光谱点的权重,从而增强对有效厚度信息的关注,减少噪声干扰。实验基于7种晶圆样本的反射光谱数据进行训练,对3种表面存有缺陷的晶圆进行测试,验证了该方法在晶圆厚度测量中的有效性和稳定性。实验结果表明,该模型在低信噪比测试集上的最大相对厚度测量误差为3.62%。模型在低信噪比光谱数据的测试中,可视化了多头自注意力机制对输入光谱计算的权重矩阵,有效抑制了厚度异常值,比傅里叶变换方法展现出更强的抗噪能力。所有光谱数据均由光谱干涉式膜厚测量系统采集,训练集中每个晶圆样品的测量点

数在30~50之间,再通过数据增强方法将模型训练数据扩充至4760组。SANN能够在较少的训练数据情况下提取出稳定的厚度信息,表现出良好的泛化能力。研究结果表明,该方法适用于噪声环境下的晶圆厚度检测,并可进一步推广至半导体制造、光学镀膜等其他薄膜测量领域。在未来工作中,可以重点优化神经网络结构,以适应更复杂、更极端的测试环境。

作者贡献说明:

王晨:测量方法提出,实验设计和数据处理,论文构思和撰写;

王子政:实验设计,数据处理;

刘翌燃:方法调查研究,论文修改;

姚程源:数据分析,资源支持;

胡春光:测量方法提出,论文修改,资源支持。

参考文献:

- [1] 李兆宇,刘子豪,王瑶莹,等. 高精度激光共焦半导体晶圆厚度测量[J]. 光学精密工程, 2024, 32(7): 956-965.
LI ZH Y, LIU Z H, WANG Y Y, *et al.* High-precision laser confocal measurement of semiconductor wafer thickness[J]. *Optics and Precision Engineering*, 2024, 32(7): 956-965. (in Chinese)
- [2] ORJI N, BADAROGLU M, BARNES B, *et al.* Metrology for the next generation of semiconductor devices [J]. *Nature Electronics*, 2018, 1(10): 532-547.
- [3] JIN Y, YU K. A review of optics-based methods for thickness and surface characterization of two-dimensional materials [J]. *Journal of Physics D Applied Physics*, 2021, 54(39): 393001.
- [4] YOON J, JO S, CHUN I S, *et al.* GaAs photovoltaics and optoelectronics using releasable multilayer epitaxial assemblies[J]. *Nature*, 2010, 465(7296): 329-333.
- [5] LU W, ROH I P, GEUM D M, *et al.* 10-nm Fin-width InGaSb p-channel self-aligned FinFETs using antimonide-compatible digital etch[C]. 2017 IEEE International Electron Devices Meeting (IEDM). December 2-6, 2017. FranciscoSan, CA, USA. IEEE, 2017: 17.7.1-17.7.4.
- [6] OHASHI T, YAMAGUCHI A, HASUMI K, *et al.* Precise measurement of thin-film thickness in 3D-NAND device with CD-SEM [J]. *Nanolithography, MEMS, and MOEMS*, 2018, 17(2): 1.
- [7] YU W J, LI Z, ZHOU H L, *et al.* Vertically stacked multi-heterostructures of layered materials for logic transistors and complementary inverters [J]. *Nature Materials*, 2013, 12(3): 246-252.
- [8] TSURU T, YAMAMOTO M. Precise determination of layer structure with EUV ellipsometry data obtained by multilayer polarizing elements [J]. *Physica Status Solidi C*, 2008, 5(5): 1129-1132.
- [9] PRICE J, HUNG P Y, RHOAD T, *et al.* Spectroscopic ellipsometry characterization of $\text{Hf}_x\text{Si}_y\text{O}_z$ films using the Cody-Lorentz parameterized model [J]. *Applied Physics Letters*, 2004, 85(10): 1701-1703.
- [10] TIKHONRAVOV A V, TRUBETSKOV M K. Online characterization and reoptimization of optical coatings[C]. *Advances in Optical Thin Films. St. Etienne, France*. SPIE, 2004, 5250: 406-413.
- [11] AKBALIK A, SOULAN S, TORTAI J H, *et al.* An inverse ellipsometric problem for thin film characterization: comparison of different optimization methods [C]. *Metrology, Inspection, and Process Control for Microlithography XXIII. San Jose, California, USA*. SPIE, 2009: 72723S.
- [12] GAO L H, LEMARCHAND F, LEQUIME M. Comparison of different dispersion models for single layer optical thin film index determination [J]. *Thin Solid Films*, 2011, 520(1): 501-509.

- [13] AMOTCHKINA T V, TRUBETSKOV M K, PERVAK V, *et al.* On the reliability of reverse engineering results [J]. *Applied Optics*, 2012, 51 (22): 5543.
- [14] BAE J, PARK J, AHN H, *et al.* Optical method for simultaneous thickness measurements of two layers with a significant thickness difference [J]. *Optics Express*, 2021, 29(20): 31615-31631.
- [15] KIM K, KWON S, PAHK H J. Fast analysis of film thickness in spectroscopic reflectometry using direct phase extraction [J]. *Current Optics and Photonics*, 2017, 1(1): 29-33.
- [16] PARK J, BAE J, JANG Y S, *et al.* Thickness measurements and uncertainty evaluation of a multilayer thin-film sample using auxiliary single-layer samples[J]. *Metrologia*, 2023, 60(2): 025007.
- [17] NA J, CHOI H Y, CHOI E S, *et al.* Self-referenced spectral interferometry for simultaneous measurements of thickness and refractive index[J]. *Applied Optics*, 2009, 48(13): 2461-2467.
- [18] FAN L J, CHEN A, LI T Y, *et al.* Thin-film neural networks for optical inverse problem [J]. *Light: Advanced Manufacturing*, 2021, 2(4): 1.
- [19] RÉDEI L, FRIED M, BÁRSONY I, *et al.* A modified learning strategy for neural networks to support spectroscopic ellipsometric data evaluation [J]. *Thin Solid Films*, 1998, 313: 149-155.
- [20] FRIED M, MASA P. Backpropagation (neural) networks for fast pre-evaluation of spectroscopic ellipsometric measurements [J]. *Journal of Applied Physics*, 1994, 75(4): 2194-2201.
- [21] 林贤锦, 吴祖伟, 叶瑞乾, 等. 基于白光三角法的芯片封装凸点高度测量[J]. *光学精密工程*, 2023, 31(13): 1890-1899.
- LIN X J, WU Z W, YE R Q, *et al.* Bump height measurement of chip packaging based on white light triangulation [J]. *Optics and Precision Engineering*, 2023, 31(13): 1890-1899. (in Chinese)
- [22] WANG Z Y, LIN Y C, ZHANG K Y, *et al.* EllipsoNet: Deep-learning-enabled optical ellipsometry for complex thin films [EB/OL]. 2022: 2210.05630. <https://arxiv.org/abs/2210.05630v1>.
- [23] CHOI J E, SONG J, LEE Y H, *et al.* Deep neural network modeling of multiple oxide/nitride deposited dielectric films for 3D-NAND flash [J]. *Applied Science and Convergence Technology*, 2020, 29(6): 190-194.
- [24] KWAK H, RYU S, CHO S, *et al.* Non-destructive thickness characterisation of 3D multilayer semiconductor devices using optical spectral measurements and machine learning [J]. *Light: Advanced Manufacturing*, 2021, 2(1): 9.
- [25] KRIZHEVSKY A, SUTSKEVER I, HINTON G E. ImageNet classification with deep convolutional neural networks [J]. *Communications of the ACM*, 2017, 60(6): 84-90.
- [26] SHERSTINSKY A. Fundamentals of recurrent neural network (RNN) and long short-term memory (LSTM) network [J]. *Physica D: Nonlinear Phenomena*, 2020, 404: 132306.
- [27] Vaswani A, Shazeer N, Parmar N, *et al.* Attention is all you need [C]. *Advances in Neural Information Processing Systems*, 2017, 30.
- [28] MA X, ZHANG P, ZHANG S, *et al.* A tensorized transformer for language modeling [C]. *Advances in Neural Information Processing Systems*, 2019, 32.
- [29] LIU Z, LIN Y T, CAO Y, *et al.* Swin transformer: hierarchical vision transformer using shifted windows [C]. 2021 *IEEE/CVF International Conference on Computer Vision (ICCV)*. October 10-17, 2021. *Montreal, QC, Canada*. IEEE, 2021: 9992-10002.

作者简介:



王 晨(2000—),女,江苏苏州人,硕士研究生,2022年于天津大学获得学士学位,主要从事光谱精密测量技术的研究。E-mail: chenwang@tju.edu.cn

通讯作者:



胡春光(1981—),男,河南光山人,博士,教授,博士生导师,2002年、2007年于天津大学分别获得学士和博士学位,主要从事微纳测试技术的研究。E-mail: cghu@tju.edu.cn