Machine learning for digital circuit backend design

LIN Yibo¹, GAO Xiaohan¹, CHEN Tinghuan², YU Bei²
(1. School of Integrated Circuits, Peking University, Beijing 100000, China; 2. Department of Computer Science & Engineering, The Chinese University of Hong Kong, Hong Kong SAR, 999077, China)

Abstract: Backend design automation bridges logic design and manufacturing in modern integrated circuit (IC) design flow. It is crucial to the eventual design closure. Backend design in the modern design flow needs to consider constraints from both high-level design and low-level manufacturing. With the continuous increase in design complexity and aggressive shrinking of feature sizes, various new challenges have emerged in backend design, especially in modeling and optimization tasks. To tackle these challenges, machine learning has been introduced to backend design automation for efficient modeling and effective optimization. This paper will introduce the typical IC design flow, what machine learning can do in the backend design, and what the literature has explored on machine learning assisted backend design automation.

Keywords: EDA; machine learning; physical design; lithography

0 引言

现代集成电路设计依赖于电子设计自动化软件（electronic design automation, EDA），并按照一定的设计流程进行设计。数字电路设计流程一般分为前端和后端设计。前端设计负责生成功能正确的逻辑实现，后端设计负责将电路网表映射到物理版图上，并保障制造环节的鲁棒性。随着设计复杂度增加、工艺节点向物理极限推进，数字集成电路后端设计难度日趋增加。为应对后端设计的挑战，越来越多的研究将机器学习技术引入到后端算法和流程当中，用以提高建模、搜索和优化过程的性能和效率。

*基金项目：自然科学基金重点项目（62034007）、北京市科委项目（Z201100004220007）资助

林亦波（通信作者），研究员，主要研究方向为机器学习在EDA中的应用，E-mail：yibolin@pku.edu.cn
余备（通信作者），副教授，主要研究方向为机器学习在EDA中的应用，E-mail：byu@cse.cuhk.edu.hk
1 机器学习辅助数字后端设计简介

1.1 数字设计流程及挑战

数字集成电路设计通常分为前端和后端两大设计流程。如图1所示，前端设计负责生成逻辑实现并验证，包含功能设计、系统设计、逻辑综合与验证。后端设计负责生成相应逻辑设计的物理实现，即将逻辑电路网表映射到物理版图上，并确保后续制造的鲁棒性。后端设计环节包含物理设计、物理验证、掩膜设计与验证等步骤。其中，物理设计可进一步分为布图规划、布局、时钟树综合、布线等步骤，掩膜设计与验证则包含光学邻近效应修正、亚分辨率辅助图形生成、光刻仿真等步骤。在实际设计当中，由于各个步骤之间信息相互依赖，设计指标评估复杂，流程需要反复迭代才能实现设计收敛。

1.2 机器学习辅助数字后端设计

机器学习技术近年来发展迅速，特别是在图像识别、自然语言处理、强化学习等领域展现了优异的建模和搜索性能。目前，越来越多的后端设计自动化研究引入机器学习提升传统优化算法的性能和效率，为缓解上述3大挑战提供了可能性。1) 由于其强大的建模能力，它可以为流程中的早期步骤提供精准高效的性能指标评估。比如在布局阶段考虑布线拥塞、时序等，加快设计收敛的效率。2) 由于其数据驱动的特性，它可以为特定设计流程和应用场景定制设计自动化算法。比如布局阶段的拥塞评估依赖于后续布线算法的特点，不同布线算法可能得到不同的拥塞结果。传统评估方法通常考虑独立于布线算法的通用评估，而机器学习技术则可以根据布线算法得到的拥塞数据自动适配模型，实现定制化的设计流程。3) 机器学习技术的普及也降低了开发设计自动化的门槛。提供了一种数据驱动的新方法，将传统解析式的仿真和优化问题转变为一种数据到数据的映射问题，并训练具有强大表达能力的机器学习模型去学习这种映射，弱化了算法开发人员对相关知识背景的要求。比如光刻仿真问题可以抽象为一个图像转译的问题，利用神经网络去...
学习转译映射函数，不再要求一般算法开发人员了解光刻仿真中涉及到的偏微分方程求解等细节。

本文总结了近期机器学习技术在数字后端设计各步骤的应用与问题抽象，特别是在布局布线、时钟树综合、掩模综合与验证等步骤中发挥的作用，并分析这些成功应用背后的共性问题与对应的机器学习技术，为进一步深入研究提供基础。

2 物理设计

物理设计是数字集成电路后端设计中的一个重要步骤。将设计从逻辑综合后的图表示（标准单元和互连）转换为由逻辑门物理形状组成的几何表示。在现代设计中，图神经网络非常适合解析基于图表示的数据。几何表示可以被描述为图像。而基于图像的计算机视觉是机器学习的成熟应用之一。应用于物理设计的典型机器学习模型是卷积神经网络及其变型，例如卷积条件对抗网络等。下面将详细阐述机器学习方法在物理设计各个阶段上的应用。

2.1 布局


除上述工作外，布线阶段的可布线性预测任务也得到了广泛的研究。因为可布线性的优劣会很大程度影响后续的布线性能。深度学习因其在计算机视觉和其他相关任务中展现的性能与效率而被引入到可布线性预测任务当中。Xie等[4]通过布线后特征预测实际布线后设计规则违例（design rule violation）的数量和位置。如图2所示，由于二维布局版图可以表示为图像，因此计算机视觉领域成熟的图像分类问题和可布线性问题具有高度相似性。这使得它们不仅可以采用类似的方法，如卷积神经网络，甚至可以从不同数据集的模型进行迁移学习。Xie等提出的方案是利用ImageNet数据集上预训练的卷积神经网络，在布局数据集上进行微调，得到可布线性预测模型。在布线期间使用该模型的预测来主动避免难以布线的布局结果。Tabrizi等[5]也提出通过预测给定布局设计的规则违例来引导布局。相比于文献[7]，他们从布局中提取了更多特征，而前者是直接将布局设计的原始版图像传到神经网络中。

图2 物理设计与图像分类具有相似性

Fig.2 The similarity between physical design and image classification

大多数传统布局器设计用于处理通用布局而很少关注数据路径布局。实际上自动生成数据路径驱动的布局器是一个开放问题。但是在过去几年中也取得了一系列进展。Ward等[6]提出了一种统一的布局流程，可同时处理随机逻辑和数据路径上的标准单元。其方案为从网表中提取图结构特征和物理特征，并将其输入到一些有效的分类器（例如神经网络）对数据逻辑路径的布局模式进行分类，再根据分
类进行针对数据路径的布局。


2.2 时钟树综合

时钟偏差 (clock skew) 是估计时钟性能的基本指标。现有工作已经表明，修改锁存器布图位置是一种能够有效减少局部时钟树电容的方法[17]。目前存在 3 种锁存器布图修改技术：锁存器移位、锁存器聚类和锁存器组合。为了减少由于局部时钟树导致的额外功耗，Ward 等[18]为每个工艺提供优化过的锁存器布图配置。之后在物理设计阶段采用机器学习模型快速选择最优的配置对锁存器进行布局。

Lu 等[19]提出了 GAN-CTS，它使用生成对抗网络和强化学习进行时钟树预测。如图 3 所示，将触发器分布、时钟网络分布和实验布线结果作为输入图像，利用在 ImageNet 数据集上预训练的 ResNet-50 网络串接上若干全连接层进行特征提取。该框架利用条件生成对抗网络优化时钟树，其中生成器由回归模型所组成，并采用策略梯度算法优化时钟树综合。

![图 3 时钟树综合的特征提取流程](image)

*Fig.3  The feature extraction of clock tree synthesis*

2.3 布线


布线过程是一项非常复杂且耗时的任务，很难通过纯机器学习方法来解决。因此将机器学习模型和传统算法相结合是很有前景的方向，例如采用机器学习模型识别出的软规则来指导布线[25]。通过这种方式可以在保留传统算法稳健性的同时获得更好的
性能。本文还观察到，在数字后端流程中，多个不同实现方式都可能达到相同的性能。考虑到几乎没有从输入到输出的一对一映射监督学习方法可行性较低。这也是广泛使用生成方法（例如，生成对抗网络）的原因，因为它很好地保留了上述一对多映射的自由度。


机器学习不仅在经典的设计流程中有效，它还可以在分体制造等安全制造场景下实现反向工程。Li 等[28]和 Zeng 等[29]成功利用低层的完整信息重建了较高金属层的互连关系。相关技术可以预测两种引脚在较高金属层互连的可能性，有助于重建整个芯片。

3 光刻仿真与掩膜综合

光刻是芯片制造的重要环节。由于流片成本高昂，在流片前对掩膜进行可制造性验证是必不可少的环节。传统光刻仿真与验证的计算代价高昂，成为了设计流程的瓶颈。比如 Synopsys Sentaurus Lithography[30]这类商用物理级别严格光刻仿真软件，在 2 μm×2 μm 掩模切片上单次仿真，需要耗时几分钟（考虑厚掩膜近似模型的仿真时间）。而一颗处理器芯片可以分解为百万至千万个同样大小的切片。随着近年来的掩膜学习领域的突飞猛进，应用机器学习方法替代或辅助光刻领域的传统算法成为一种趋势。同时，本文注意到光刻领域的问题与计算机视觉领域的问题有很高的相似性，将计算机视觉领域的神经网络模型应用到光刻问题，已在光刻仿真、
优化、验证等方面取得了不错的结果。目前机器学习在光刻相关的应用主要包括光刻模型、掩膜优化、版图可制作性预测和版图图形生成。

3.1 光刻模型

光刻模型一般包含两个阶段，第1阶段是光学模型（optical model），第2阶段是光刻胶模型（resist model）。如图5所示，光学模型仿真光通过掩膜照射到光刻胶形成的光强分布（空间像），光刻胶模型仿真光强分布在光刻胶上形成的图案（光刻胶图案）。光学模型一般分为薄掩膜近似和厚掩膜近似。其中薄掩膜近似即Kirchhoff近似，不考虑掩膜的3D效应。物理级别的厚掩膜近似虽然更精准，但仿真耗时长，因此Ye等\cite{38}提出将掩膜视为像素图，将光学模型类比为计算机视觉中的图像生成问题。前人构建了基于条件生成对抗网络（conditional generative adversarial network, CGAN）的光刻仿真模型，通过在放大的光刻胶图案上训练生成对抗网络，LithoGAN能够输出超分辨率的光刻胶图案。相比于传统的物理级别严格仿真，LithoGAN在7 nm的接触层（contact layer）上实现了1 800倍速度的优势。

3.2 掩膜优化

掩膜优化问题是光刻仿真的逆问题。其目标是寻找合适的掩膜图案，使得通过光刻生成的光刻胶图案最接近目标版图。掩膜优化技术包括亚分辨率辅助图形（sub-resolution assist features, SRAF）、光学邻近效应修正（optical proximity correction, OPC）以及逆向光刻图形技术（inverse lithographic techniques, ILT）。

光学邻近效应修正技术通过改变目标图案的边缘来增强光刻性能，保证正确的关键尺寸（critical dimension）。早期OPC主要研究如何设计调整目标图案边缘的方案，比如Awad等\cite{43}采用一种同时移动关键图案邻边的优化算法。GAN-OPC\cite{44}将问题建模为图像到图像的转换问题，使用生成对抗网络直接生成OPC结果，为解决这一类问题提供了新思路。

亚分辨率辅助图形技术在目标图案附近插入亚分辨率图形，以干涉来增强目标图案。早期SRAF主要基于规则或模型，如商用工具Mentor Graphics Calibre中采用的方案\cite{39}。近来出现了基于机器学习模型的SRAF插入技术，即采用类似于GAN-OPC的思路，GAN-SRAF\cite{45}将插入亚分辨率图形的问题建模为图像到图像的转化问题，使用生成对抗网络来生成亚分辨率辅助图形。

逆向光刻图形技术从光刻仿真逆问题的角度同时解决亚分辨率辅助图形插入和光学邻近效应修正的问题。如图6所示，ILT过程以目标版图为输入，优化掩膜图案使得生成的版图最接近目标版图。相比于SRAF和OPC，ILT拥有更大的搜索空间。Poonawala等\cite{46}最早将光刻优化的问题表述为逆成像问题。沿着这条思路，MOSAIC\cite{47}引入U-Net\cite{48}神经网络模型作为求解器，生成优化的掩膜结果。
3.3 版图可制造性预测

热点检测问题是版图可制造性的关键问题。如图7所示，将版图中可能无法正常制造的区域称为版图的热点。提前检测出版图中的热点区域可以帮助设计者及时改进版图。由于光刻仿真计算开销高，运行时间长，版图热点检测旨在版图通过光刻仿真之前，预测版图中可能出现的问题区域。

版图热点检测的两大主流算法为模式匹配和机器学习。模式匹配方法通过比对版图热点数据库的版图拓扑来识别可能存在问题的版图。在先进工艺节点下，模式匹配方法受限于热点数据库的通用性，更多研究选择了使用机器学习来预测版图热点。早期将传统机器学习模型应用于热点检测问题的尝试包括：使用支持向量机检测区域内是否存在热点，使用贝叶斯模型检测。Yang等首次将版图热点检测问题类比为图像分类问题并考虑热点检测数据中的严重数据不均衡，并引入计算机视觉领域的卷积神经网络训练预测模型。Chen等进一步研究如何同时检测出一个版图区域内可能存在多个热点的场景，并将其问题抽象为计算机视觉领域的目标检测问题。

3.4 版图图形生成

机器学习也被应用在版图图形数据库的构建中。有效的版图图形数据库需要满足：1）不违背设计规则；2）模式尽可能多样。Yang等提出将版图图形生成问题抽象为带约束的随机图像生成问题，使用transforming convolutional auto-encoder编码版图到高维空间，扰动高维空间的向量获得版图的变体。Zhang等进一步提出一种尽量生成不违背设计规则的合法版图的方法，即使用VCAE自编码模型生成版图经过高斯扰动得到的变体，并使用条件对抗生成网络合法化生成的版图，再经由对抗自编码器过滤掉不合法的版图图形。

4 结 论

本文回顾了数字集成电路设计自动化的主流流程，以及在先进工艺下数字后端设计中建模和优化的挑战。在此基础上，介绍了如何运用机器学习技术辅助后端设计中的建模和优化问题，并就物理设计和光刻环节的设计自动化问题抽象和相应的机器学习技术展开了讨论。可以看到，机器学习在数字后端设计自动化中主要处理图和图像等两类数据，这些数据分别来自于电路网表和版图特征。同时，机器学习的任务也较为多样，包括分类、回归、图像生成、参数搜索等。相信随着研究的不断深入，机器学习将与数字后端设计的传统算法更加有机地结合，实现更精准高效的建模与优化。

参考文献

[3] FU J, ZHENG H, MEI T. Look closer to see better: Recurrent attention convolutional neural network for fine-grained image recognition[C]. IEEE Conference on Computer Vision and Pattern Recognition, 2017: 4438-
4446.


