

面向寄存器传输级设计阶段的高效高精度功耗预测模型

李康* 师瑞之 陈嘉伟 史江义 潘伟涛 王杰

(西安电子科技大学 西安 710000)

摘要: 功耗已成为电路设计的关键性能目标之一, 现有商业工具PrimeTime PX(PTPX)的功耗预精度高, 但是运行时间长, 且仅面向已经生成网表的逻辑综合或者物理实现阶段。因此, 降低功耗分析时间, 且前移功耗预测在芯片设计中的环节变得尤为重要。该文提出一种面向千万门级专用集成电路(ASIC)的寄存器传输级(RTL)功耗预估方法, 可在RTL设计阶段实现快速且准确的周期级功耗预测: 根据输入信号的功耗相关性原则使用基于平滑截断绝对偏差惩罚项(SCAD)的嵌入法对输入信号自动筛选, 从而解决大信号特征输入数量对预估性能的影响; 通过时序对准方法对仿真波形数据进行校正, 解决了sign-off级功耗与RTL级仿真波形之间的时序偏差问题, 有效提升了模型预测的精度; 建立了仅拥有两个卷积层和1个全连接层的浅层卷积神经网络模型, 学习相邻位置和相邻时间上的信号活动与功耗的相关性信息, 充分降低部署开销, 使训练速度得到显著提高。该文使用开源数据集、28 nm工艺节点的 3×10^7 门级工业级芯片电路作为测试对象, 实验结果表明, 功耗预测结果和物理设计后PTPX分析结果相比, 平均绝对百分比误差(MAPE)小于1.71%, 11k时钟周期的功耗曲线预测耗时不到1.2 s。在场景交叉验证实验中, 模型的预测误差小于4.5%。

关键词: 功耗预估; 卷积神经网络; 寄存器传输级; 超大规模集成电路

中图分类号: TN406

文献标识码: A

文章编号: 1009-5896(2023)09-3166-09

DOI: [10.11999/JEIT230359](https://doi.org/10.11999/JEIT230359)

An Efficient and High-precision Power Consumption Prediction Model for the Register Transfer Level Design Phase

LI Kang SHI Ruizhi CHEN Jiawei SHI Jiangyi PAN Weitao WANG Jie

(Xidian University, Xi'an 710000, China)

Abstract: Power consumption is identified as a critical performance objective in circuit design. Existing power estimation tools, such as PrimeTime PX (PTPX), provide high accuracy but are hampered by lengthy execution times and are confined to logic synthesis or physical implementation stages with an already generated netlist. As a result, the need to reduce power analysis time and stress the importance of forward power prediction in chip design has been recognized. A power estimation model for early-stage large-scale Application Specific Integrated Circuit (ASIC) is introduced, which can achieve fast and accurate cycle-level power prediction at the Register Transfer Level (RTL) design stage. The model applies the Smoothly Clipped Absolute Deviation (SCAD) embedding method based on the power correlation principle of input signals for automatic signal selection, addressing the impact of large input feature numbers on estimation performance. A timing alignment method is employed to correct the timing deviation between sign-off level power and RTL-level simulation waveform, enhancing prediction accuracy. The strong nonlinearity of a shallow two-layer convolutional neural network is utilized by the model for power training, consisting of two convolutional layers and one fully connected layer, which reduces computational overhead. Power labels use backend Sign-off level power output data to enhance the accuracy of prediction results. This power estimation model is evaluated on a 28 nm Network Processor (NP) with more than 30 million gates. Experimental results demonstrate that the Mean Absolute Percentage Error (MAPE) of this model for predicting total circuit power consumption is less than 1.71% when compared with the PTPX analysis results following physical design back-annotation. The model takes less than 1.2 s to predict the power curve for 11k clock cycles. In cross-validation experiments with different scenarios, the prediction error of the model is found to be less than 4.5%.

Key words: Power estimation; Convolutional Neural Network(CNN); Register Transfer Level (RTL); Very Large Scale Integration circuit (VLSI)

1 引言

数字集成电路的集成密度和计算复杂性遵循摩尔定律的不断发展。随着工艺的不断改进,由于晶体管电压阈值限制电源电压无法继续降低,同时短沟道效应导致亚阈值特性退化、泄漏电流增大两个主要原因,功耗已成为电路设计的关键性能目标之一^[1,2]。对于数字电路设计的功耗分析,目前最主流解决方案是使用PTPX (synopsys PrimeTime PX)工具,在逻辑综合或者物理实现环节,其将网表和网表的仿真波形作为输入工具,基于标准单元工艺库中的参数进行计算得到高精度的功耗分析结果。PTPX工具的功耗预估有着高精度度,但是,运行时间长,且仅面向已经生成网表的逻辑综合或者物理实现阶段。当前,数字芯片对早期设计规划(Early Design Planning, EDP)有迫切需求,若在寄存器传输级(Register Transfer Level, RTL)设计阶段就可准确预测芯片的功耗,则可在芯片设计早期调整和修复设计中不合理的架构,降低芯片功耗。因此,面向RTL阶段的功耗预估方法紧迫且重要^[3]。

针对上述问题,目前,学术界尝试利用机器学习技术来提高功耗预测的准确性和降低流程的复杂性。Dhotre等人^[4]使用多层感知机(Multi-Layer Perceptron, MLP)实现了针对电路模块的功耗预测。Nasser等人^[5]使用了活动因子和静态概率作为功耗特征,实现了电路工作频率对功耗的影响的评估。文献^[4,5]虽实现了功耗的准确预估,但由于MLP仅处理向量特征,面对大规模电路的大特征量时计算成本过高,仅针对小规模电路模块进行原理性验证。Zhou等人^[6]提出将寄存器信号进行3通道翻转作为特征,并通过使用卷积神经网络(Convolutional Neural Network, CNN)构建功耗模型得精确的功耗预估结果,但由于并未对电路的信号特征进行筛选,面对大规模电路时,模型需要较长运行时间。Kim等人^[7]针对开源处理器Rocket提出了利用基于Lasso惩罚项的正则化方法实现信号特征筛选,较好地提升了针对RTL级电路功耗预测性能。Xie等人^[8]针对Arm Cortex-A77微处理器提出了基于极小化极大凸惩罚(Minimax Concave Penalty, MCP)的信号选择方法,实现了更加精确的功耗预估。但文献^[7,8]都是针对处理器的指令进行功耗预估,且并未考虑工业级芯片的布局布线对于功耗的影响,与芯片的实际功耗值存在一定误差。

为了弥补上述缺陷,本文提出一种面向千万门级ASIC电路的RTL级功耗预估方法,基于机器学习方法建立了其功耗分析模型,利用信号筛选策略解决了大信号特征输入数量对功耗预测模型的性能影响,并通过对sign-off级功耗数据的时序预处理技术得到功耗模型的标签数据,有效提升了模型预测的精度。本文的主要贡献如下:

(1) 通过时序对准方法对仿真波形数据进行校正,解决了sign-off级功耗与RTL级仿真波形之间的时序偏差问题,从而可以利用物理设计反标的功耗结果作为RTL级电路功耗模型的标签数据,得到更准确后仿功耗预测结果。

(2) 利用正则化方法进行功耗相关性评价的特征信号筛选,使用基于平滑剪切绝对偏差惩罚函数(Smoothly Clipped Absolute Deviation, SCAD)惩罚项的正则化方法对输入特征信号进行筛选处理,最终得到与功耗相关度更大的关键特征信号,显著提升了功耗预测的效率。

(3) 提出一个仅拥有两个卷积层和1个全连接层的浅层卷积神经网络模型,本模型在进行功耗训练与推理时,可以学习相邻位置和相邻时间上的信号活动与功耗的相关性信息,在实现高精度预测的同时,降低部署开销,使训练速度得到显著提高。

(4) 上述方法在开源数据集、28 nm工艺节点的 3×10^7 门级工业级网络处理器芯片电路上进行了测试,最终实现了误差小于4.5%的跨场景功耗波形预测,验证了所提出方法的有效性。

2 面对RTL级电路的功耗预测框架

2.1 面对RTL级仿真的功耗预估流程

图1给出了功耗预估方法流程,主要包括训练数据获取、功耗建模、功耗预测3个部分。

首先通过电路仿真获取训练数据,包括RTL电路仿真生成信号波形文件(.vcd/.fsdb文件),从中提取特征数据,即神经网络模型的输入数据;以及对于RTL电路逻辑综合、布局布线后生成版图级网表,仿真生成VCD文件与PTPX的功耗曲线文件(.out文件),从中获取标签数据,即用于校正模型的功耗数据。

在获得仿真的信号活动数据以及功耗数据后,需要对数据进行处理。首先对信号的仿真波形进行特征提取;获得信号特征后,进行特征与标签的匹配对齐;之后进行特征筛选,减少负样本数量,筛选出关键信号,增强特征与标签的关系;再对最终

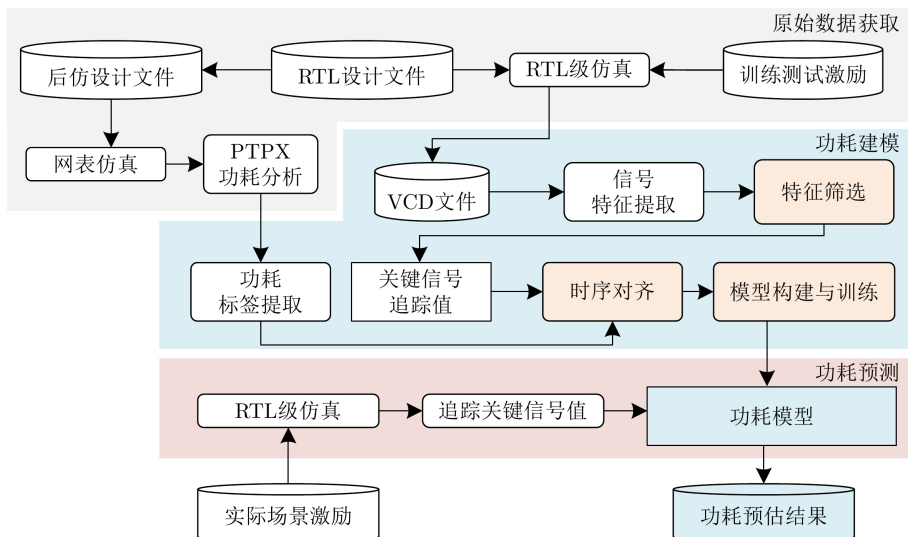


图1 功耗预估方法示意图

的数据集进行模型构建，实现多周期的输入数据合并等。

获得处理后的特征与标签数据后，建立卷积神经网络模型，包括两层卷积层以及1层全连接层，进行训练，获得功耗模型。

最后，在功耗预测阶段，对RTL电路提取关键信号的仿真波形，通过功耗模型快速地得到高精度功耗预估结果。

2.2 基于SCAD惩罚项的关键信号筛选方案

2.2.1 数据集特征构建

功耗建模的特征数据来自VCD文件^[9]。在特征采样过程中，信号的仿真结果将按照仿真时间顺序被分割成多个时间窗口。本文采取汉明距离作为信号数据的特征，其计算公式为

$$D_H(x, y) = \frac{1}{N} \sum_{i=1}^N |X_i - Y_i| \quad (1)$$

其中， X_i 与 Y_i 为信号值， N 为信号位宽。汉明距离是功耗建模中最常用的特征^[10]，计算简单，仅需异或操作，也可以较好表征信号的翻转次数。

2.2.2 基于惩罚项的特征选择方法

本文使用机器学习中的特征选择方法处理数据集，实现基于电路仿真的信号值的信号筛选。去除负样本，即部分干扰信号，如无翻转信号，或是活动较少的信号；同时筛选出对于功耗的关键信号，增强特征与标签之间的关系。

采取基于SCAD惩罚项的特征选择法，SCAD惩罚项^[11]即平滑截断绝对偏差惩罚项。在回归模型的损失函数中加入惩罚项，使得函数权重在训练过程中具有偏向0的趋势。模型训练完成后，保留权重不为0的特征作为特征选择结果。其计算为

$$\theta^{(t)} = \theta^{(t-1)} - \eta \cdot \frac{\partial J(\theta^{(t-1)})}{\partial \theta^{(t-1)}} \quad (2)$$

其中， θ 是需要迭代的参数向量， t 代表迭代次数， η 代表学习率， $J(\theta)$ 为损失误差。

在回归模型的原有优化器中加入惩罚项后的损失函数计算为

$$J'(\theta) = J(\theta) + P(\theta) \quad (3)$$

其中， $P(\theta)$ 为惩罚项， $P(\theta)$ 的计算为

$$P_{\text{SCAD}} = \begin{cases} \lambda |\theta_i|, & |\theta_i| \leq \lambda \\ \frac{2\gamma\lambda |\theta_i| - \theta_i^2 - \lambda^2}{2(\gamma-1)}, & \lambda \leq |\theta_i| \leq \gamma\lambda \quad (\gamma > 2) \\ \frac{1}{2}(\gamma+1)\lambda^2, & |\theta_i| \geq \gamma\lambda \end{cases} \quad (4)$$

其中，参数向量 θ 的大小为 $n \times 1$ ， λ 表示正则化系数。

同时，基于相同方法有其他不同的惩罚项，包括L1惩罚项^[7]、L2惩罚项^[12]、MCP惩罚项^[8]等。L1, MCP, SCAD 3个惩罚项的对比如图2所示，其中 $\gamma = 3$, $\lambda = 1$ 。L1惩罚项对任何情况下的参数都进行恒定量的惩罚，因此为了实现信号的有效筛选，需要设定非常大的 λ ，而其余的信号权重也会受到过多的惩罚^[8]；MCP与SCAD惩罚项在保持了对较小的权重信号的惩罚力度的同时，对于权重大的信号影响较小，其中，SCAD惩罚项保留了区间 $\lambda \leq x \leq \lambda r$ 内的惩罚率的1阶微分，保证了对于小权重信号足够的惩罚力度。

2.3 功耗标签数据处理

2.3.1 信号特征与后仿功耗标签的对齐方法

使用物理设计反提后的PTPX分析结果可以得到更加精确的功耗波形。该信号仿真和功耗波形的

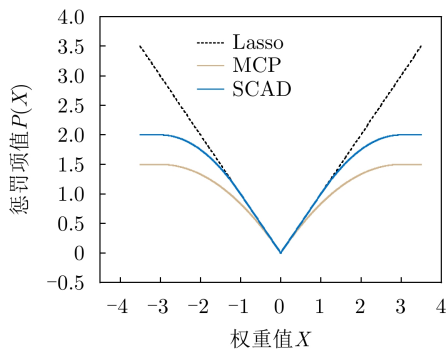


图2 Lasso, MCP和SCAD的惩罚项示意图

仿真时间是一致的。但是，由于缺失连线延时等信息，RTL级的仿真时间无法与后仿的功耗波形保持一致^[13]。如果不进行数据对齐，用于训练功耗模型的数据集将是错误的。这种错误的数据集将无法训练出正确的功耗模型。

首先对两组数据进行最大最小归一化处理，计算方法为

$$x_i^* = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}} \quad (5)$$

本文中，使用一种常见的信号处理方法——最小均方匹配方法^[14]对信号翻转特征进行处理，使信号特征和功耗在均方根误差意义下最接近。最小均

方匹配方法通常用于预处理步骤，以提高信号的质量和准确性。通过平移波形，同时计算其与目标波形的均方根误差，寻找最小误差最小值，得到最佳匹配结果。左移(或右移)的最大限度为3个样本数，在单周期采样的情况下，为3周期。均方误差(Mean Squared Error, MSE)的计算公式为

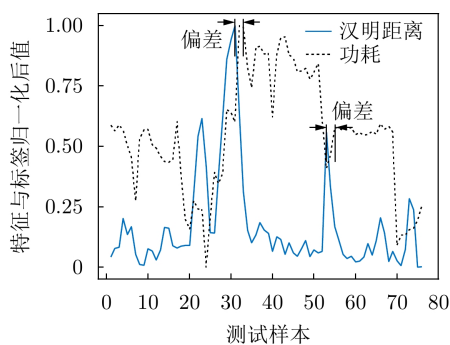
$$MSE = \frac{1}{N} \sum_{i=1}^N (X_i - Y_i)^2 \quad (6)$$

其中，MSE表示均方误差， N 表示样本数量， X_i, Y_i 分别表示原始信号(即信号特征)与目标信号(即功耗标签)的第 i 个样本值。

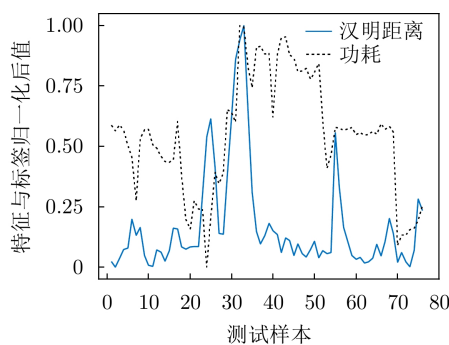
图3为网络处理器中的分组处理与调度模块的汉明距离与功耗标签在最小均方匹配前后的对比，通过对比汉明距离和功耗波形的方法进行数据对齐，保证了数据的正确性和有效性。

2.3.2 数据集基于时间相关性的特征构建

完成特征选择后，构建功耗模型所需要的数据集，图4给出了数据集构建示意。每列代表着按照仿真时间排序的时间窗口，共有 T 个时间窗口； S_1, S_2, \dots, S_n 代表着电路中的信号；表中的数值代表则对应时间窗口下各个信号的特征值，即汉明距离； P 指对应时间窗口下电路的实际功耗。



(a) 对齐前汉明距离与功耗波形对比

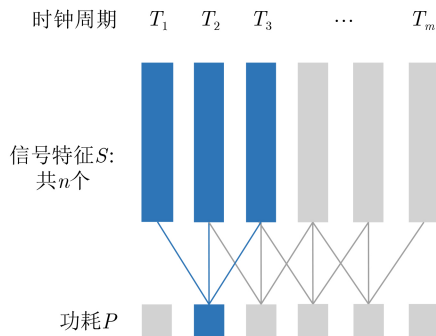


(b) 对齐后汉明距离与功耗波形对比

图3 数据对齐前后汉明距离与功耗值局部放大图

	1	2	3	...	T
信号 S_1	5	1	0	...	1
信号 S_2	0	0	1	...	2
信号 S_3	4	2	0	...	3
⋮	⋮	⋮	⋮	⋮	⋮
信号 S_n	1	0	1	...	1
功耗 P	0.04	0.01	0.02	...	0.02

(a) 数据文件示意



(b) 多输入构建示意

图4 数据集构建示意图

该数据集的每列都包含着一组信号特征和相对应的功耗标签。在正常的监督式神经网络模型中,可以对训练样本进行随机采样,使用该数据集的每行作为一个训练样本对模型进行训练。但是这种训练样本丢失了时间关系,模型的功耗预测每次只根据1个时间窗口下的信号情况。而在实际功耗分析中,电路在1个时间窗口内的功耗波动与上个时间段的功耗是存在相关性的。

为保留时间上的相关性,采用多输入单输出的神经网络形式,即使用多组特征数据预测1个功耗标签。如图4(b)展示了3输入单输出的数据集。以 T_1, T_2, T_3 3个时间窗口为例,前3个时间窗口的特征数据被当作一组输入数据,用来预测第2时间窗口的功耗值,功耗模型可以观测到第2时间窗口的前后时间窗口的特征情况,保留样本的时间相关性。

2.4 基于浅层卷积神经网络模型的功耗回归模型

多输入单输出的数据使得模型的输入从原本的1维变成了2维。便于卷积神经网络的使用。本文构建了一种基于卷积神经网络的功耗模型,图5给出了网络具体结构,模型由两层卷积层和1层全连接层构成。

模型输入层为 $n \times m$ 的2维矩阵,其中 n 是信号特征的数量, m 是时间窗口的数量。数据进入第1个卷积层C1后会被纵向卷积。再经过ReLU激活函数,进入第2个卷积层C2。会对数据进行横向的卷积。然后,数据会再次经过ReLU激活函数,并且会被展平成1维数据。最后,数据通过全连接层,得到输出 P' ,即预测的功耗值。

在该模型中,第1个卷积层C1只进行纵向卷积,这是因为特征提取过程是顺序执行的,相邻的信号特征大概率是在一个电路模块中,即特征与特征之间存在着一定相关性。因此,使用纵向的卷积核对数据进行处理,学习信号间的部分结构关系。第2个卷积层C2只进行横向卷积。该卷积层学习信号特征时间上的相关性。由于功耗预测是一种回归问题,所以全连接层不设置激活函数,避免激活函数对预测值的约束。

3 实验与分析

3.1 测试集

本实验构建了两组电路规模不同的数据集,其中之一按照目前其他功耗预估方法的实验,选用功能相同或是规模相似的电路,如表1所示,包括浮点加法器,高级加密标准(Advanced Encryption Standard, AES)算法电路,以太网媒体访问控制(Media Access Control, MAC)器,简单的第5代精简指令集(Reduced Instruction Set Computer Version 5, RISC-V)电路等构建数据集,避免电路结构化对于模型的功耗预估结果的误差影响。另外一组使用工业级的28 nm网络处理器芯片的各个模块以及整体电路系统构建,进行方法在更大规模电路的有效性验证。

网络处理器芯片采用了联电28 nm工艺,设计规模约 3×10^7 门。包括帧处理模块、调度模块、crossbar总线模块、插入捕获指令模块、配置模块等。使用电路的物理设计反提后的PTPX分析结果作为功耗标签。

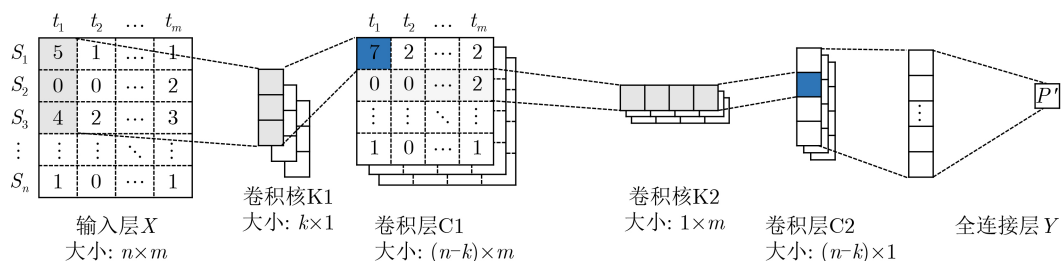


图5 卷积神经网络模型示意图

表1 神经网络模型的测试用例

电路名称	电路描述	电路规模	仿真场景(样本数)
float-adder	32位浮点加法器	约15000门	随机激励(220k)
RISC-V Core	简易的RISC-V核心	约100000门	单一场景测试(850k)
AES	AES加密算法	约300000门	随机测试(84k)
MAC	媒体访问控制层协议	约400000门	随机测试(95k)
FIFO	8位宽16深度的同步FIFO	约10000门	随机测试(80k)
light8080	简易的8080处理器	约100000门	随机测试(4 800k)

3.2 实验评价指标

对于模型对电路测试集的功耗预估结果，本文选取平均绝对百分比误差(Mean Absolute Percentage Error, MAPE)以及标准化均方根误差(Normalized Root-Mean-Squared-Error, NRMSE)作为评估指标。MAPE的计算公式为

$$\text{MAPE} = \frac{1}{n} \cdot \sum_{t=1}^n \left| \frac{A_t - F_t}{A_t} \right| \quad (7)$$

其中， n 为样本数， A_t 为实际值， F_t 为预测值。NRMSE的计算公式为

$$\text{NRMSE} = \frac{1}{\bar{A}} \sqrt{\frac{1}{n} \cdot \sum_{t=1}^n (A_t - F_t)^2} \quad (8)$$

其中， n 为样本数， A_t 为实际值， F_t 为预测值， \bar{A} 为实际值平均值。

3.3 实验流程

本实验的模型算法由tensorflow框架以及scikit-learn库搭建，其中scikit-learn库提供了模型中的特征筛选方法和结果评估方法。实验运行的硬件为Intel(R) i5-12600F @4.20 GHz, Intel UHD Graphics 770和32 GB运行内存。

针对常见的小规模电路测试集，除了本文所提出的卷积神经网络(Convolutional Neural Network, CNN)模型外，还构建了其他多个神经网络功耗模型作为对比，包括逆传播(Back Propagation, BP)神经网络模型^[15]、ResNet18模型^[16]，以及长短时记忆网络(Long Short Term Memory, LSTM)模型^[17]。

针对工业级的网络处理器芯片，除了本文所提出的基于SCAD惩罚项的特征筛选方案外，还实现了其他多个特征筛选方案，包括线性相关性F检测(Corr.)^[18]、递归特征消除法(Recursive Feature Elimination, RFE)^[19]、随机森林法(Random Forst, RF)^[20]等。对各个信号筛选方法对模型精确度的影响进行对比后，再将信号筛选后的各个机器学习模型的功耗预估结果进行对比。

实验采用5折交叉验证的方法，将数据集划分为5份后，每份轮流作为测试集，对模型进行验证。

3.4 功耗预估误差与建模耗时比较

表2给出了不同模型对于电路的预估误差结果MAPE，表3给出了不同模型对电路的训练耗时。本文提出的CNN模型的预估结果的平均误差为1.79%，仅次于LSTM模型的1.69%。但CNN模型对所有电路的训练总耗时为6.03 s，训练速度是LSTM的80倍，在面对大规模电路的功耗预估场景时，这一优势更加明显。

表2 不同模型的预测误差(MAPE)(%)

电路名称	CNN	LSTM	线性回归	BP	ResNet18	CNN_PRIMAL
float-adder	4.29	4.32	4.39	4.38	6.44	2.83
RISC-V Core	0.23	0.08	0.78	0.66	4.96	0.39
AES	0.19	0.56	0.34	0.81	1.75	1.85
MAC	1.15	0.34	3.44	1.89	4.63	-
FIFO	4.78	4.77	4.80	4.85	5.06	-
light8080	0.11	0.05	0.14	0.15	2.95	-
平均误差	1.79	1.69	2.31	2.12	4.30	-

表3 不同模型训练耗时(s)

电路名称	CNN	LSTM	线性回归	BP	ResNet18	CNN_PRIMAL
float-adder	1.47	8.25	1.64	1.45	10.34	23.41
RISC-V Core	1.03	63.55	0.80	2.52	43.39	19.52
AES	1.03	263.69	0.67	4.98	56.19	25.80
MAC	0.55	128.53	0.31	1.70	18.28	-
FIFO	0.27	5.20	0.22	0.31	1.92	-
light8080	1.69	31.64	1.30	1.70	24.67	-
总耗时	6.03	500.86	4.94	12.67	154.80	-

同类型的文献[6] PRIMAL (Power Inference using Machine Learning)所提出的CNN模型的规模更大。其实验环境与实验对象与本文不同，本文将PRIMAL的特征构建方法与CNN模型应用在了测试集上，以各自环境下的线性回归模型训练耗时为基准进行对比。预估结果的误差方面，本文的CNN模型与PRIMAL的模型的预测结果大致相当，均维持在5%以内；但本文的CNN模型的训练速度明显优于PRIMAL模型，建模速度快约10倍。

3.5 网络处理器电路功耗预估结果

3.5.1 时序对齐结果对比

对于NP各模块进行时序对齐方法的功耗模型预估结果如图6所示，对于所有模块的预估误差均有一定程度的下降，这说明功耗标签与信号特征之间的时序误差对模型的准确度的影响较大，验证了时序对齐方法的必要性。

3.5.2 信号筛选结果对比

图7给出了对于网络处理器电路各个模块，使用线性回归、随机森林、支持向量机、惩罚项添加等特征筛选方法后，进行BP神经网络模型建模的功耗预估误差结果。RFE的分类器模型选用的是回归模型；L1的正则化系数为0.05；L2的正则化系数为7；MCP与SCAD的惩罚系数分别设置为 $\gamma = 3$, $\lambda = 1$ ；将重要性评分最大值的0.1倍作为筛选目标，对重要性评分较大的特征进行筛选。

各个筛选方法对于网络处理器电路所有模块的运行时间对比如图8所示。RFE方法所筛选的特征构建的模型虽然具有最小的误差结果，然而，RFE和RF方法的耗时情况非常严重，以inst_fp_and_sch_top模块的特征选择为例，RFE和RF特征选择的耗时分别为156 min和8 min，这说明在功耗预估角度，RFE与RF方法无法处理大规模电路的大量信号特征。

SCAD惩罚项与MCP惩罚项的运行时间基本相同，与线性相关性检测、Lasso惩罚项相比，这两种方法的运行时间都较长，但是均可以在160 s内对4个模块完成特征筛选。使用SCAD惩罚项方法筛选结果构建的神经网络模型的预估误差结果与

其他方法相比，误差较小，并且保证了有效的运行时间。

3.5.3 功耗预估结果对比

图9(a)给出了对于网络处理器电路，采用SCAD惩罚项特征筛选方法，构建不同模型的功耗预估误差。图9(b)则给出了电路整体的预估误差。对于系统的功耗预测误差均保持在3%以内，这证明了线性功耗叠加方案的可行性。对比各模块的预估误差与电路整体系统误差，可以看出电路整体的电路功耗预估误差较低，这是因为不同功能模块在电路整体系统的占比不同，对大功耗占比的模块实现高精度的功耗预估，可以保证电路整体系统功耗预估的高预测精度。

3.5.4 模型迁移结果对比

使用网络处理器的5个场景进行场景交叉验证本文的功耗预估方法的场景迁移能力。表4给出了不同场景的验证结果。其中，动作测试的场景系统误差最大，为4.41%。这是由于动作测试场景下，测试了所有的微码处理器模块。但在其他场景下，部分微码处理器模块的功耗波动较小。

实验结果表明在场景的交叉验证下的系统误差仍然可用维持在4.5%以内。说明这种功耗预估方法是可行的。另一方面，建模耗时加上预测耗时的总体时间，都被控制在13 min(780 s)以内，这远远快于动辄数个小时或者数天的传统功耗分析流程。

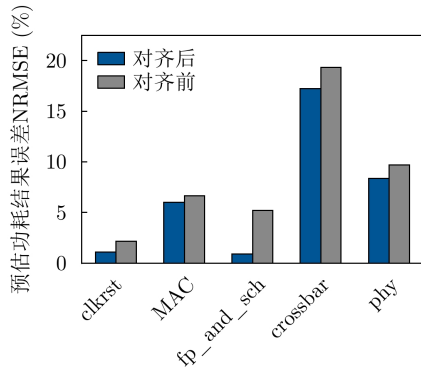


图6 数据对齐前后CNN模型预估误差对比

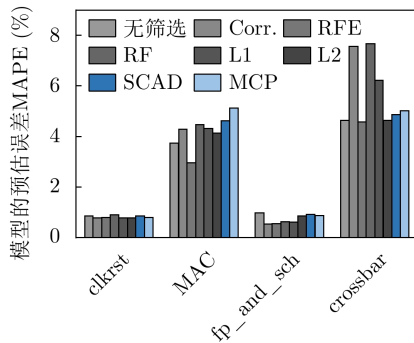


图7 不同特征筛选方法后BP模型的预估误差对比

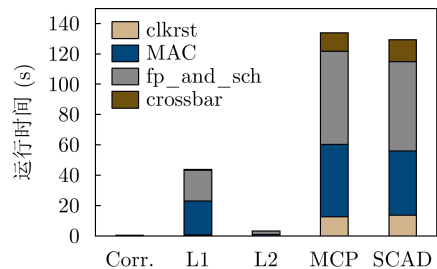
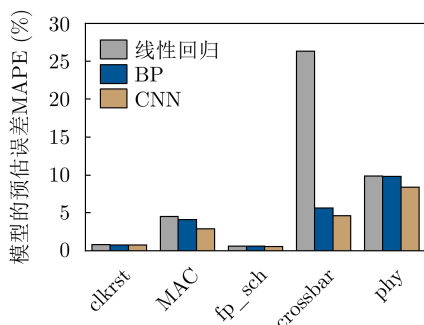
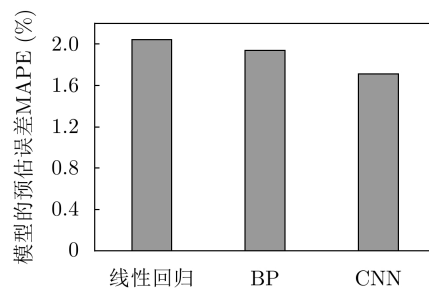


图8 特征筛选方法速度对比



(a) 网络处理器各模块不同模型预估误差对比



(b) 网络处理器整体预估误差对比

图9 网络处理器不同模型预估误差对比

表4 场景交叉验证结果

测试场景	样本数量	建模耗时(s)	预测耗时(s)	预测误差 MAPE(%)
动作测试	约11k	451.35	1.20	4.41
单播流控分类	约5.6k	661.74	0.60	1.86
单播线速	约4k	698.76	0.43	1.79
单播学习	约3.4k	713.61	0.38	1.76
组播测试	约1.5k	776.25	0.19	2.82

4 结束语

本文提出了针对RTL级电路的功耗预估方法，包括一个2层卷积神经网络功耗模型，本模型具有快速建模、快速预测和高精度的优点；以及采用SCAD惩罚项特征选择方法对RTL级电路的功耗建模进行了特征筛选；并使用该方法对ASIC芯片网络处理器的各个模块进行了功耗建模。实现了平均误差为1.71%的RTL级周期精确的功耗预测。在网络处理器的11k时钟周期中，本文提出的功耗预测方法仅需要1.2 s，而Synopsys PTPX工具需要4 h以上。今后的工作将围绕着CNN模型在电路功能仿真级模型的应用，以及CNN模型的跨电路迁移能力展开。

参考文献

- [1] RAUT K J, CHITRE A V, DESHMUKH M S, *et al.* Low power VLSI design techniques: A review[J]. *Journal of University of Shanghai for Science and Technology*, 2021, 23(11): 172–183. doi: [10.51201/JUSST/21/11881](https://doi.org/10.51201/JUSST/21/11881).
- [2] REN Haoxing and HU Jiang. Machine Learning Applications in Electronic Design Automation[M]. Cham: Springer, 2023.
- [3] SROUR M. Data-dependent cycle-accurate power modeling of RTL-level IPs using machine learning[D]. [Master dissertation], The University of Texas at Austin, 2018.
- [4] DHOTRE H, EGGERSGLÜB S, CHAKRABARTY K, *et al.* Machine learning-based prediction of test power[C]. 2019 IEEE European Test Symposium (ETS), Baden-Baden, Germany, 2019: 1–6. doi: [10.1109/ETS.2019.8791548](https://doi.org/10.1109/ETS.2019.8791548).
- [5] NASSER Y, SAU C, PRÉVOTET J C, *et al.* NeuPow: A CAD methodology for high-level power estimation based on machine learning[J]. *ACM Transactions on Design Automation of Electronic Systems*, 2020, 25(5): 41. doi: [10.1145/3388141](https://doi.org/10.1145/3388141).
- [6] ZHOU Yuan, REN Haoxing, ZHANG Yanqing, *et al.* PRIMAL: Power inference using machine learning[C]. The 56th Annual Design Automation Conference 2019, Las Vegas, USA, 2019: 39. doi: [10.1145/3316781.3317884](https://doi.org/10.1145/3316781.3317884).
- [7] KIM D, ZHAO J, BACHRACH J, *et al.* Simmani: Runtime power modeling for arbitrary RTL with automatic signal selection[C]. The 52nd Annual IEEE/ACM International Symposium on Microarchitecture, Columbus, USA, 2019: 1050–1062. doi: [10.1145/3352460.3358322](https://doi.org/10.1145/3352460.3358322).
- [8] XIE Zhiyao, XU Xiaoqing, WALKER M, *et al.* APOLLO: An automated power modeling framework for runtime power introspection in high-volume commercial microprocessors[C/OL]. MICRO-54: 54th Annual IEEE/ACM International Symposium on Microarchitecture, 2021: 1–14. doi: [10.1145/3466752.3480064](https://doi.org/10.1145/3466752.3480064).
- [9] PUNDIR N, PARK J, FARAHMANDI F, *et al.* Power side-channel leakage assessment framework at register-transfer level[J]. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 2022, 30(9): 1207–1218. doi: [10.1109/TVLSI.2022.3175067](https://doi.org/10.1109/TVLSI.2022.3175067).
- [10] HUANG Guyue, HU Jingbo, HE Yifan, *et al.* Machine learning for electronic design automation: A survey[J]. *ACM Transactions on Design Automation of Electronic Systems*, 2021, 26(5): 40. doi: [10.1145/3451179](https://doi.org/10.1145/3451179).
- [11] FAN Jianqing and LI Runze. Variable selection via nonconcave penalized likelihood and its oracle properties[J]. *Journal of the American statistical Association*, 2001, 96(456): 1348–1360. doi: [10.1198/016214501753382273](https://doi.org/10.1198/016214501753382273).
- [12] TIAN Yingjie and ZHANG Yuqi. A comprehensive survey on regularization strategies in machine learning[J]. *Information Fusion*, 2022, 80: 146–166. doi: [10.1016/j.inffus.2021.11.005](https://doi.org/10.1016/j.inffus.2021.11.005).
- [13] SCHÜRMAN S, ONNEBRINK G, LEUPERS R, *et al.* ESL power estimation using virtual platforms with black box processor models[C]. The 2015 International Conference on Embedded Computer Systems: Architectures, Modeling, and Simulation (SAMOS), Samos, Greece, 2015: 354–359. doi: [10.1109/SAMOS.2015.7363698](https://doi.org/10.1109/SAMOS.2015.7363698).
- [14] ZHANG Xianda. Modern Signal Processing[M]. Tsinghua University Press, 2022: 497–564. doi: [10.1515/9783110475562](https://doi.org/10.1515/9783110475562).
- [15] ZHOU Guochang, GUO Baolong, GAO Xiang, *et al.* A FPGA power estimation method based on an improved BP neural network[C]. 2015 International Conference on Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), Adelaide, Australia, 2015: 251–254. doi: [10.1109/IIH-MSP.2015.76](https://doi.org/10.1109/IIH-MSP.2015.76).
- [16] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, *et al.* Deep residual learning for image recognition[C]. The 2016 IEEE

- Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770–778. doi: [10.1109/CVPR.2016.90](https://doi.org/10.1109/CVPR.2016.90).
- [17] CHHABRIA V A, AHUJA V, PRABHU A, *et al.* Thermal and IR drop analysis using convolutional encoder-decoder networks[C]. The 26th Asia and South Pacific Design Automation Conference, Tokyo, Japan, 2021: 690–696. doi: [10.1145/3394885.3431583](https://doi.org/10.1145/3394885.3431583).
- [18] FARAWAY J J. Linear Models with R[M]. 2nd ed. New York: CRC Press, 2014. doi: [10.1201/b17144](https://doi.org/10.1201/b17144).
- [19] JEON H and OH S. Hybrid-recursive feature elimination for efficient feature selection[J]. *Applied Sciences*, 2020, 10(9): 3211. doi: [10.3390/app10093211](https://doi.org/10.3390/app10093211).
- [20] XUAN Yi, SI Weiguo, ZHU Zhu, *et al.* Multi-model fusion short-term load forecasting based on random forest feature selection and hybrid neural network[J]. *IEEE Access*, 2021, 9: 69002–69009. doi: [10.1109/ACCESS.2021.3051337](https://doi.org/10.1109/ACCESS.2021.3051337).
- 李 康: 男, 博士, 副教授, 研究方向为EDA、SoC设计方法学.
师瑞之: 男, 硕士生, 研究方向为EDA、SoC设计.
陈嘉伟: 男, 硕士生, 研究方向为EDA、低功耗设计.
史江义: 男, 博士, 教授, 研究方向为SOC设计与设计方法学、低功耗设计、物理实现、硬件安全.
潘伟涛: 男, 博士, 副教授, 研究方向为SoC设计与设计方法学.
王 杰: 男, 硕士生, 研究方向为EDA、SoC物理实现.

责任编辑: 余 蓉