基于概率模型的高维集成电路良率估计

戴国浩（2019******）

摘 要：随着集成电路技术的发展，微电子元件已经缩小到纳米级的尺寸，准确估计电路的生产良率也愈发成为一个挑战．然而，对于高维度的电路来说，传统的良率估计方法的时间成本过大．为了加速良率估计，两种基于概率模型的方法被提出．第一种方法基于代理模型，叫做绝对值缩减深度核学习法．该法使用深度核高斯过程来代替电路仿真器，利用了基于希尔伯特-施密特独立准则的方法来裁剪电路参数，并通过最大化信息熵寻找电路故障区域的边界．然而，该法可能会导致估计结果的不稳定．因此，第二种基于重要性采样的新方法被提出，也就是最优流形重要性采样法．该法使用洋葱采样法进行预采样，通过归一化流模型来拟合电路故障区域，并采用了动态的更新框架．为了验证提出方法的效果，两个方法在 18 维、108 维、569 维和 1093 维的存储器电路上进行了实验．结果显示，所提出的第一和第二个方法，比传统的良率估计法最多快上 232.1 和 273.8 倍，比其他前沿方法最多快上 11.1 和 12.7 倍，并具有更高的精度．在电路设计阶段准确估计良率，能够减少电路故障率、降低成本和缩短微电子产品的上市时间．

关键词：计算机应用技术；良率估计；重要性采样；代理模型；归一化流；高维电路

Probabilistic Model-Based Yield Estimation for High-Dimensional Intergrated Circuits

Guohao Dai（2019******）

Abstract： With the development of integrated circuit technology, microelectronic devices have reached nano-scale dimensions, posing challenges for accurate yield estimation (YE). However, traditional YE methods are computationally expensive for high-dimensional circuits. To address this, two statistical YE methods were proposed. The first method, Absolute Shrinkage Deep Kernel Learning (ASDK), used a Hilbert-Schmidt independence criterion for feature selection, a deep kernel learning Gaussian process surrogate, and an optimization strategy based on information entropy. To further enhance stability, the second method proposed, Optimal Manifold Importance Sampling (OPTIMIS), was based on importance sampling. It used onion sampling for parameter space pre-sampling, fitted the circuit's failure region using a normalizing spline flow model and continuously updated it with a dynamic framework. The proposed methods were evaluated on 18, 108, 569, and 1093-dimensional circuits. ASDK and OPTIMIS achieved up to 231.1x and 273.8x speedup over the traditional method, respectively, and up to 11.1x and 12.7x speedup over other state-of-the-art YE models. Accurate YE during circuit design can reduce failure rates, lower costs, and expedite time-to-market for microelectronic products.

Key words: Computer Applied Technology; Yield Estimation; Importance Sampling; Surrogate; Normalizing Flow; High-dimensional Circuits

随着集成电路技术的发展，微电子器件已经缩小到纳米的级别，这使得随机工艺变化（如晶体管的长宽度波动）成为电路设计中必须考虑的关键因素．在现代电路设计中，某些电路单元在电路中会被复制上百万次，例如静态随机存取存储器（Static Random-Access Memory，SRAM）电路，在这种情况下随机工艺变化对电路性能的影响变得更加严重．为了减小这种工艺变化带来的电路不稳定性，需要准确地估计电路良率．

蒙特卡洛法（Monte Carlo，MC）是最传统的良率估计方法，MC 的估计结果被认为是电路良率的真实值．在 MC 的良率估计过程中，MC 从电路工艺变化空间中随机采样参数，然后将参数送入电路仿真器（Simulation Program with Integrated Circuit Emphasis，SPICE）中进行仿真，最后，这些参数的对应电路性能达标的比例就是电路的良率．然而，对于高维度的电路来说，MC 时间成本很高，在高良率问题中不可行．例如，65 纳米 SRAM 单元阵列的电路失败率处于 1×10^{-6}的量级[1]，这意味着 MC 要至少进行一百万次采样才能捕获到一个故障的电路样本，而一百万次的高维电路仿真是无比耗时的．

为了提高良率估计的速度，基于重要性采样（Importance Sampling，IS）的良率估计方法被提出了．IS 方法首先找到电路上的失效区域，然后建立一个用于采样的概率分布，并将该概率分布的采样中心移动到电路的失效区域，从而在失效区域上进行采样．由于 IS 方法主要在电路的失效区域上进行采样，而不是像 MC 方法那样随机采样，因此采到电路失效点的效率大大提高，加快了良率估计的收敛速度．比如，最小正态 IS 方法[2]将采样分布的中心移动到离原点最近的失效区域上．而超球体聚类采样法[3]使用高斯混合分布对多个失效区域的中心进行采样．这两个方法的采样概率分布都是静态的，一旦这些概率分布被定义好后就不能再改变．然而，随着采样不断的进行，越来越多的失效点被发现．如果不对这些新的失效数据加以利用，就会造成信息的浪费．所以，研究者提出了动态的采样方法．比如，自适应重要性采样[4]能够动态地将最新的失效点运到混合高斯采样分布中．另一个 IS 方法[5]则是选用了动态的混合冯米塞斯分布．尽管 IS 方法能够提升良率收敛速度，但也存在着一定局限性．IS 方法高度依赖于概率分布形式的选取．并且，当电路的维度提升时，IS 方法的复杂性会指数地上升．

另一种实现良率估计的主要方法是构建数据驱动的代理模型来代替未知的 SPICE，并使用主动学习方法来逐步减少模型误差．这种良率估计方法叫做代理模型法．比如，由 Jian 等人提出的方法[6]使用了高斯过程模型作为代理模型来评估电路．Xiao 等人提出的代理模型方法[7]则使用混乱多项式展开的低秩张量估计来代替 SPICE；为了提高代理模型的训练精度，Xiao 还提出了基于错误率权重的训练数据采样法[7]．而 Jiannan 等人则提出使用矩阵值传递函数作为代理模型[8]．Shuo 等人则提出一种了基于深度过程的良率估计方法[9]，并且该方法能够只选取最重要的区域作为训练集，节省 SPICE 调用成本．然而，尽管代理模型的方法能够节约 SPICE 仿真次数，但也存在一定的局限性．代理模型的良率估计结果高度依赖于代理模型的训练精度，对代理模型的训练精度非常敏感．对于高维度的电路来说，如果代理模型的训练策略不够精巧，代理模型往往会面临“维度灾难”的问题．

为了克服上述方法的缺点，本文提出了两个全新的基于概率模型的良率估计方法．本文提出的第一套方法是一个基于代理模型的方法，叫做绝对值缩减深度核学习法，该方法的贡献和创新之处如下：

1）该方法使用了深度核学习高斯过程模型作为 SPICE 的代理模型，能够有效提取电路的关键特征；

2）该方法使用基于希尔伯特-施密特独立准则的最小绝对值收敛和选择算子的特征选择法，裁剪掉和电路性能最不相关的特征；

3）该方法提出了一个全新的采集函数，创新性地使用最大化信息熵来寻找电路失效区域边界点作为训练数据；

4）该方法在 SRAM 电路上进行良率估计的结果比其他前沿方法最多快上快 11.1 倍，并且更精确．

本文提出的第二套方法基于 IS，叫做最优流形重要性采样法，该方法的贡献和创新之处如下：

1）该方法提出了一个新的失效点预采样方法，称为洋葱采样．洋葱采样可以在对电路毫无先验知识的时候，快速地定位到电路的失效区域；

2）该方法使用最前沿的概率模型归一化流来进行重要性采样．归一化流中的神经网络结构能够非常有效地拟合电路的失效区域；

3）该方法使用了一套自适应的动态训练框架；

4）该方法在高至 1093 维的 SRAM 电路上的良率估计结果比其他的前沿方法快，并且更加精确．

1 背景

1.1 问题定义

电路的良率由电路的工艺变化参数决定．电路中的工艺变化参数是指在电路制造或生产过程中会发生变化的参数，通常包括工艺步骤的特定参数，比如晶圆制备过程中的沉积速率和蚀刻深度等；又包含电路元器件的制造参数，如一个晶体管的长度和宽度等．工艺变化参数$\boldsymbol{x}$的定义如式（1）所示：

$$\boldsymbol{x}=\left[x^{(1)},x^{(2)},\cdots,x^{(d)}\right]^{T}\in\mathcal{X} \tag{1}$$

式中：

$x^{(j)}$—— 向量$\boldsymbol{x}$中的第j个工艺变化参数变量；

d—— 工艺变化参数的维度；

$\mathcal{X}$ —— 高维的工艺变化参数的参数空间．

（1）中，$\boldsymbol{x}$里的变化参数x大小标示着参数的波动幅度大小，x的波动通常遵从一个正态分布，并且x之间是相互独立．将x标准化后， $\boldsymbol{x}$的概率密度函数$p(\boldsymbol{x})$如式（2）所示：

$$p(\boldsymbol{x})=\prod_{i}^{D}\exp\left(-x^{(i)}/2\right)/\sqrt{2\pi} \tag{2}$$

注意，电路制造过程中$\boldsymbol{x}$的具体大小是不能够被人为控制，因为$\boldsymbol{x}$是一个随机变量，是由电路生产过程中的误差决定的．但是，对于任意的$\boldsymbol{x}$，可以对其进行 SPICE 仿真，从而得到电路的性能指标y，比如 SRAM 的读取时间和模拟电路的电压增益大小等指标，来评判电路优劣，如式（3）所示：

$$y=z(\boldsymbol{x}) \tag{3}$$

式中：

y—— 电路性能指标，是一个标量；

$z(\cdot)$—— 表示 SPICE 仿真器，会输出电路的性能指标，调用一次$z(\cdot)$较为昂贵．

为了划分电路的性能是否合格，通常会人为地给电路的性能指标设定一个阈值t．划分电路是否合格的指示函数$I(\cdot)$如式（4）所示：

$$I(\boldsymbol{x})=\begin{cases}1, & z(\boldsymbol{x})>t\\ 0, & z(\boldsymbol{x})<t\end{cases} \tag{4}$$

式（4）的含义为，当电路的指标$f(\boldsymbol{x})$大于阈值t时，则认为该电路的性能不符合生产要求，$\boldsymbol{x}$被认为是失败样本，此时$I(\cdot)$的取值为 1；反之，电路的性能指标$f(\boldsymbol{x})$小于t时，则认为该电路符合生产要求，$\boldsymbol{x}$被认为是正常样本，此时$I(\cdot)$的取值为 0．所以，$I(\cdot)$就是一个用来指示电路是否合格的函数．

为了方便读者理解，后文一概从电路失败率P_f的角度去描述电路良率估计问题．失败率P_f和良率P_y的关系为$P_y=1-P_f$．

有了以上的工具后，便可以通过概率统计的方法计算电路失败率P_f，如式（5）所示：

$$P_f=E_{p(\boldsymbol{x})}[I(\boldsymbol{x})]=\int_{\mathcal{X}}I(\boldsymbol{x})p(\boldsymbol{x})d\boldsymbol{x} \tag{5}$$

式中：

$E_{p(\boldsymbol{x})}[\cdot]$—— 对于具有$p(\boldsymbol{x})$分布的随机变量$\boldsymbol{x}$，以括号内的函数$[\cdot]$为权重取期望值．

式（5）中的求期望值本质上就是一个积分计算．在整个工艺变化参数空间$\mathcal{X}$下对$\boldsymbol{x}$进行积分计算，便可以得到电路失败率P_f．

1.2 传统蒙特卡洛法

在式（3）中， $z(\cdot)$表示 SPICE 仿真器，该仿真器相当于一个端到端的黑盒函数，没有精确的数学表达式，这也就导致了计算P_f的积分（5）无法被计算出公式解．但它的数值解可以被计算出来．传统的 MC 方法可以被用来计算该积分．MC 方法采用了大数定理的思想[10]：当事件的发生的次数趋于无穷时，该事件发生的频率等于该事件发生的概率．同理，对于某一定积分，只需要对随机变量采样无穷次，然后对该随机变量对应的函数值求平均，便可以近似得到该定积分的值．使用 MC 方法估计电路失败率$\widehat{P}_f$的数学表达如式（6）所示：

$$\widehat{P}_f=\frac{1}{N}\sum_{i}^{N}I(\boldsymbol{x}_i)\xrightarrow{N=\infty}P_f \tag{6}$$

式中：

$\widehat{P}_f$——MC 方法得到的电路失败率的估计值；

$\boldsymbol{x}_i$—— 通过$p(\boldsymbol{x})$采样得到的第 i 个样本$\boldsymbol{x}_i$；

P_f——电路失败率的真实值．

式（6）的含义是，通过概率密度函数$p(\boldsymbol{x})$，采样得到N个样本$\boldsymbol{x}$，得到样本集$\{\boldsymbol{x_i}\}_i^N$．然后，计算这些样本的$I(\cdot)$并取平均，便可以得到估计的电路失败率$\widehat{P}_f$．并且，当$N\to\infty$时，$\widehat{P}_f\to P_f$．也就是说，当 MC 方法的采样数N越多时，通过 MC 方法得到的失败率估计值$\widehat{P}_f$越接近真实值P_f．

可以看到，传统 MC 方法需要调用N次 SPICE 仿真器$z(\cdot)$才可以计算出良率，但是N往往是个非常大的数字．并且，对于高维度电路来说，调用一次 SPICE 的成本较大．因此，传统 MC 方法的估计成本非常高昂．

1.3 基于代理模型的良率估计方法

由于传统 MC 方法过于昂贵，为了减少调用 SPICE 的次数，研究者们提出了基于代理模型的良率估计方法．代理模型方法的核心思想就是，从 SPICE 中获取少量的数据，然后用这些数据去训练一个廉价的代理模型$f(\cdot)$，然后使用代理模型$f(\cdot)$去代替昂贵的 SPICE 仿真器$z(\cdot)$，基于代理模型的指示函数$I_F(\cdot)$如式（7）所示：

$$I_F(\boldsymbol{x}) = \begin{cases} 1, & f(\boldsymbol{x}) > t \\ 0, & f(\boldsymbol{x}) < t \end{cases} \qquad (7)$$

式中：

$I_F(\cdot)$—— 使用代理模型后的指示函数；

$f(\cdot)$—— 代理模型，代替 SPICE 返回电路指标．

此时，使用传统的 MC 方法便可以计算出电路的失效率，如式（8）所示：

$$\widehat{P}_f = \frac{1}{N}\sum_i^N I_F(\boldsymbol{x_i}) \qquad (8)$$

如式（8）所示，和传统 MC 方法类似，通过概率密度函数$p(\boldsymbol{x})$，采样得到样本集$\{\boldsymbol{x_i}\}_i^N$后，送入到廉价的代理模型$f(\cdot)$中，然后，计算这些样本的$I_F(\cdot)$的平均值，便可以得到估计的电路失败率$\widehat{P}_f$．

代理模型方法的优点是大大减少了调用 SPICE 仿真器$z(\cdot)$的次数，极大的节约了成本．并且，可以控制估计的精度．当调用$z(\cdot)$的次数越多时，训练数据越多，训练出来的$f(\cdot)$也就越精准．然而缺点是$\widehat{P}_f$的结果精度将高度依赖于$f(\cdot)$的精度．并且，在高维电路下，可能会发生模型欠拟合的问题．

1.4 基于重要性采样的良率估计方法

为了避免巨量的 SPICE 仿真次数，研究者们提出了重要性采样方法．重要性采样法不再用$p(\boldsymbol{x})$进行采样，转而使用一个向电路的失效区域倾斜的概率分布 $\mathrm{g}(\boldsymbol{x})$进行采样，以此来提高采到失效点的效率．此时，使用重要性采样方法来计算电路失败率的公式如式（9）所示：

$$P_f = \int_{\mathcal{X}} I(\boldsymbol{x}) \frac{p(\boldsymbol{x})}{g(\boldsymbol{x})} g(\boldsymbol{x}) d\boldsymbol{x} \qquad (9)$$

式中：

$g(\boldsymbol{x})$—— 被偏移到电路失效区域的概率分布的概率密度函数；

式（9）在整个工艺变化参数空间$\mathcal{X}$下对$\boldsymbol{x}$进行积分计算，便可以得到电路失败率P_f．但由于$I(\boldsymbol{x})$无具体表达式，所以依旧不能通过（2-10）求得P_f的公式解．但利用类似 MC 的数值方法可以得到失败率估计值，如式（10）所示：

$$\widehat{P}_f = \frac{1}{M}\sum_i^M I(\boldsymbol{x}_i) w(\boldsymbol{x}_i) \xrightarrow{N=\infty} P_f \qquad (10)$$

式中：

$\widehat{P}_f$—— 电路失败率的估计值；

$w(\boldsymbol{x}_i)$—— $p(\boldsymbol{x}_i)/g(\boldsymbol{x}_i)$；

$\boldsymbol{x}_i$—— 通过$g(\boldsymbol{x})$采样得到的第 i 个样本$\boldsymbol{x}_i$．

式（10）中，通过$g(\boldsymbol{x})$采样得到M个样本$\boldsymbol{x}$，得到样本集$\{\boldsymbol{x}_i\}_i^M$．然后，计算这些样本的$I(\boldsymbol{x}_i)w(\boldsymbol{x}_i)$的平均值，便可以得到估计的电路失败率$\widehat{P}_f$．并且，当$M \to \infty$时，$\widehat{P}_f \to P_f$．值得注意的是，$\{\boldsymbol{x}_i\}_i^M$的样本数要比 MC 采样的样本数$\{\boldsymbol{x}_i\}_i^N$小得多，因为电路失效在$g(\boldsymbol{x})$中并不是个极小概率事件．因此，重要性采样方法估计良率要比 MC 方法高效得多．

虽然，比起 MC 方法，IS 方法能够节省大量的 SPICE 仿真次数，加快良率估计值的收敛．但是，IS 方法也存在一定缺点．IS 方法对$g(\boldsymbol{x})$的选取极为敏感；并且，目前大多数 IS 方法参数的选择依赖于电路；同时，IS 方法的复杂度会随着电路维度的增加而指数型地增加．

2 绝对值缩减深度核学习法

本章主要介绍所提出的第一个基于代理模型的良率估计方法，叫做绝对值缩减深度核学习法（Absolute Shrinkage Deep Kernel Learning，ASDK）．ASDK 基于深度核高斯过程的代理模型，希-施特征选择法以及模型的最大信息熵动态更新法．本章节将详细介绍 ASDK 的每一个模块．

2.1 深度核学习高斯过程

高斯过程[11]（Gaussian Process，GP）是代理模型的常见选择，由于其模型精度和对不确定性的量化，GP 常被用于空间探索任务．

首先引入一个数据集$\boldsymbol{D} = (\boldsymbol{X}, \boldsymbol{Y})$．其中$\boldsymbol{X} = (\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n)$，$X$包含$n$个维度为$d$的向量$\mathbf{x}$；$\boldsymbol{Y} = (z(\boldsymbol{x}_1), z(\boldsymbol{x}_2), \cdots, z(\boldsymbol{x}_n))^T$，是一个$1 \times n$的向量，噪声为$\sigma$．拥有数据集$\boldsymbol{D}$后，便可使用负对数似然值$L$作为损失函数对 GP 进行训练，如式（11）所示：

$$L = -\boldsymbol{Y}^T\left(\boldsymbol{K}_{X,X} + \sigma^2\boldsymbol{I}\right)^{-1}\boldsymbol{Y} - \log\left|\boldsymbol{K}_{X,X} + \sigma^2\boldsymbol{I}\right| \qquad (11)$$

式中：

$\boldsymbol{K}_{X,X}$—— 协方差矩阵，大小为$n \times n$；

$\boldsymbol{I}$—— 单位矩阵，形状为$n \times n$．

当GP模型$f(\cdot)$训练完毕后，使用$f(\cdot)$对测试集集$\boldsymbol{X}_*$进行预测．$\boldsymbol{X}_*$包含了n_*个测试样本$\boldsymbol{x}_*$．GP在$\boldsymbol{X}_*$上的结果如式（12）所示：

$$f(X_*)\sim\mathcal{N}(\mathrm{E}[f(\boldsymbol{X}_*)],\mathrm{cov}[f(\boldsymbol{X}_*)]) \quad （12）$$

式中：

$\mathrm{E}[f(\boldsymbol{X}_*)]$——平均值向量，形状为$n_*\times 1$，每行分别每个为$\boldsymbol{x}_*$的$f(\boldsymbol{x}_*)$的平均值；

$\mathrm{cov}[f(\boldsymbol{X}_*)]$—— 协方差矩阵，形状为$n_*\times n_*$，由每个$\boldsymbol{x}_*$之间的协方差组成．

式（12）的含义是，GP的每个预测点都可以看为一个高斯分布，每个点有各自的均值和方差．

回顾了传统GP后，接下来介绍深度核学习高斯过程模型[12]（Deep Kernel Learning Gaussian Process，DKLGP），本论文里提出的第一个良率估计方法ASDK使用了DKLGP模型作为代理模型，代替SPICE仿真器．DKLGP主要对核函数做了修改，DKLGP的核函数如式（13）所示：

$$k_{\gamma,w}(\boldsymbol{x}_i,\boldsymbol{x}_j)=k_\gamma\left(n_w(\boldsymbol{x}_i),n_w(\boldsymbol{x}_j)\right) \quad （13）$$

式中：

$n_w(\cdot)$—— 表示一个深度神经网络，比如多层感知机，对d维的$\boldsymbol{x}$进行特征提取；

w—— 表示神经网络里的权重参数；

γ—— 表示核函数本身的参数．

在式（13）中，可以看到，DKLGP在核函数中多接了一个神经网络结构$n_w(\cdot)$，变量$\boldsymbol{x}$会先经过神经网络$n_w(\cdot)$，被提取出特征，然后这些特征再经过普通的GP．神经网络的加入使得DKLGP具有提取特征的强大能力．

2.2 希-施独立准则最小绝对值收敛特征选择

电路的工艺变化参数是决定电路性能的重要因素．然而，实际电路往往拥有大量的工艺变化参数．如果直接使用所有工艺变化参数进行代理模型建模可能会导致“维度灾难”的问题．幸运的是，良率研究者们在先前的实验[13]里揭示了：不是所有的参数都会对电路性能产生影响．因此，良率估计可以进行特征选择，只用核心参数进行建模．

基于希尔伯特-施密特独立准则（Hilbert-Schmidt Independence Criterion，HSIC）的最小绝对值收敛和选择算子回归（Least Absolute Shrinkage and Selection Operator，LASSO）[14]结合了核矩阵的HSIC准则和LASSO回归，能够通过建立一个代理模型，有效地选择出非线性的特征．ASDK使用HSIC-LASSO进行特征选择．

首先，LASSO通过优化一个带L1正则化项的损失函数来选择特征，函数如式（14）所示：

$$\boldsymbol{\alpha}=argmin_{\alpha}\frac{1}{2}\left|\left|Y-X\boldsymbol{\alpha}\right|\right|_2^2+\lambda\left|\left|\boldsymbol{\alpha}\right|\right|_1 \quad （14）$$

式中：

$\boldsymbol{\alpha}$——d维向量，表示$\mathbf{x}$里每个特征的重要性；

λ—— 惩罚项的惩罚因子；

$\boldsymbol{X}$，$\boldsymbol{Y}$—— 数据集，包含n个$\boldsymbol{x}$和y．

式（14）中，得到了特征权重$\boldsymbol{\alpha}$的各个维度的权重值后，便可获知$\mathbf{x}$哪个维度是最重要的．值较大的维度被留下，其他的维度则直接丢弃．然而，LASSO的缺点也很明显．对于具有复杂结构的数据集，LASSO可能无法选择出最优的特征子集，因为LASSO在选择特征时只考虑了线性相关性，而对于非线性相关性较强的数据集，LASSO的效果可能不佳．而工艺变化参数和电路的性能指标之间的关系又是呈强非线性的．

而HSIC-LASSO能克服LASSO只能应对线性数据的缺点．HSIC-LASSO借助了核函数$k(\cdot)$，能够在再生核希尔伯特空间建立输入和输出间的联系，具有很强的非线性建模能力．HSIC-LASSO特征选择的目标函数如式（15）所示：

$$\boldsymbol{\alpha}=argmin_{\alpha}\frac{1}{2}\left|\left|\boldsymbol{L}-\sum_l^d\boldsymbol{K}^{(l)}\boldsymbol{\alpha}^{(l)}\right|\right|_2^2+\lambda\left|\left|\boldsymbol{\alpha}\right|\right|_1 \quad （15）$$

式中：

$\boldsymbol{L}$—— $\boldsymbol{L}=\boldsymbol{H}\boldsymbol{K}_y\boldsymbol{H}^T$，且$[\boldsymbol{K}_y]_{ij}=k(y_i,y_j)$；

$\boldsymbol{K}^{(l)}$—— 核矩阵，$[\boldsymbol{K}^{(l)}]_{ij}=k\left(x_i^{(l)},x_j^{(l)}\right)$；

$\boldsymbol{H}$—— 中心化矩阵，$\boldsymbol{H}=\boldsymbol{I}-(1/\mathrm{n})\mathbf{1}\mathbf{1}^{\mathbf{T}}$．

式（15）中，该算法的关键结果是权重向量$\boldsymbol{\alpha}$．$\boldsymbol{\alpha}$指示了在再生核希尔伯特空间中，每个输入特征维度的重要性．通过HSIC-LASSO，ASDK得以选择出对电路性能影响最大的工艺变化参数．

2.3 最大化信息熵

对于基于代理模型的良率估计方法来说，其实代理模型$f(\boldsymbol{x})$不需要完全精准地描绘出整个工艺参数空间$\boldsymbol{\mathcal{X}}$下的$z(\boldsymbol{x})$准确形状．代理模型$f(\boldsymbol{x})$需要重点训练的区域是失效区域的边界．对失效区域的边界拟合足够精确后，便能够精确分类出电路的失效样本和非失效样本．

$f(\boldsymbol{x})$是个GP，GP将每个$\boldsymbol{x}$的输出看作一个高斯分布．先定义$f(\boldsymbol{x})\geq t$的概率$l(\boldsymbol{x})$如式（16）．

$$l(\boldsymbol{x}) = P[f(\boldsymbol{x}) \geq t] = \Phi\left(\frac{\mu(\boldsymbol{x}) - t}{\sigma(\boldsymbol{x})}\right) \quad (16)$$

式中：

$\Phi(\cdot)$ —— 标准高斯分布的累计分布函数；

$\mu(\boldsymbol{x})$ —— $f(\boldsymbol{x})$的均值函数；

$\sigma(\boldsymbol{x})$ —— $f(\boldsymbol{x})$的标准差函数.

式（16）中，若$l(\boldsymbol{x}) = 0.5$，则说明点$\mathbf{x}$处于代理模型$f(\boldsymbol{x})$的失效区域边界. $l(\boldsymbol{x})$越接近 0.5，说明$\boldsymbol{x}$离代理模型$f(\boldsymbol{x})$的失效边界越近.

ASDK 借用二项分布的信息熵的定义，定义 ASDK 的采集函数如（17）所示：

$$\begin{aligned}\mathrm{acq}_*(\boldsymbol{x}) =& -l(\boldsymbol{x})\log\big(l(\boldsymbol{x})\big) \\ & -\big(1-l(\boldsymbol{x})\big)\log\big(1-l(\boldsymbol{x})\big)\end{aligned} \quad (17)$$

式（17）中，当采集函数$\mathrm{acq}_*(\boldsymbol{x})$有最大值时，$l(\boldsymbol{x}) = 0.5$，$\boldsymbol{x}$处于代理模型$f(\boldsymbol{x})$的失效区域边界. ASDK 寻找新训练点$\boldsymbol{x}_\#$的方法如式（18）所示：

$$\boldsymbol{x}_\# = \mathrm{argmax}_{\mathbf{x}}(\mathrm{acq}_*(\boldsymbol{x})) \quad (18)$$

ASDK 要进一步训练代理模型$f(\boldsymbol{x})$时，通过式（18），寻找到位于电路失效区域的边界点$\boldsymbol{x}_\#$，然后将$\boldsymbol{x}_\#$送入 SPICE 评估器得到$z(\boldsymbol{x}_\#)$. 然后更新原有训练集$\boldsymbol{D} \leftarrow \boldsymbol{D} \cup \big(\boldsymbol{x}_\#, z(\boldsymbol{x}_\#)\big)$，使用$\boldsymbol{D}$再次训练代理模型$f(\boldsymbol{x})$. 重复此过程，便可以让$f(\boldsymbol{x})$在失效区域的边界的拟合愈来愈精准. 本文称之为最大化信息熵法.

2.4 算法总流程

前文介绍了 ASDK 的基本组成部分：DKLGP 代理模型，HSIC-LASSO 特征裁剪，最大化信息熵选择边界点方法. 本小节将前文介绍的组件集装在一起，总体介绍 ASDK 的方法流程.

ASDK 流程的具体介绍如下所示：

1）首先，在整个工艺变化参数空间$\mathcal{X}$下进行拉丁超立体采样，采样M个$\boldsymbol{x}$，并对$\mathbf{x}$进行评估得到$z(\boldsymbol{x})$，从而得到初始训练集$\boldsymbol{D} = (\boldsymbol{x}_i, z(\boldsymbol{x}_i))_{i=1}^{M}$；

2）根据$\boldsymbol{D}$使用 HSIC-LASSO 对$\boldsymbol{x}$进行特征裁剪；

3）使用$\boldsymbol{D}$建立一个 DKLGP 代理模型$f_{\mathrm{DKLGP}}(\boldsymbol{x})$；

4）进行 MC 采样. 使用工艺变化参数的概率分布$p(\boldsymbol{x})$进行采样，得到N个点（n很大），采样所得的数据集为$\boldsymbol{x}_{i}{}_{i=1}^{N}$. 然后，使用代理模型$f_{\mathrm{DKLGP}}(\boldsymbol{x})$计算电路的失效率$\widehat{P}_f$如(2-9)所示，$\widehat{P}_f = 1/N\sum_i^N I_F(\boldsymbol{x}_i)$；

5）利用优化方法寻找失效区域的边界点$\boldsymbol{x}_\#$，如式（18）所示，$\boldsymbol{x}_\# = \mathrm{argmax}_{\mathbf{x}}(\mathrm{acq}_*(\boldsymbol{x}))$；

6）将$\boldsymbol{x}_\#$送入 SPICE 进行评估，得到$z(\boldsymbol{x}_\#)$，对训练数据集$\boldsymbol{D}$进行更新，有$\boldsymbol{D} \leftarrow \boldsymbol{D} \cup \big(\boldsymbol{x}_\#, z(\boldsymbol{x}_\#)\big)$；

7）如果电路失败率的估计值$\widehat{P}_f$未达到要求精度，则继续重复第 3~6 步，直到$\widehat{P}_f$达到精度要求为止.

3 最优流形重要性采样法

ASDK 有时在特定的电路上效果并不稳定，这是所有代理模型方法的通病. 所以，本章节将完全抛弃代理模型，转而把目光转向 IS. 本章节将介绍本文提出的第二个良率估计方法. 该方法是一个基于 IS 的方法，该方法叫做最优流形重要性采样法（Optimal Manifold Important Sampling，OPTIMIS）.

3.1 洋葱预采样方法

在良率估计阶段的最开始，IS 方法对一个电路的 SPICE 没有任何的先验知识. 但是，IS 可以通过预采样的方法对电路的失效区域进行探寻，从而为 IS 模型的构建提供训练集. 本文提出了洋葱采样法来进行电路的预采样.

为了阐述洋葱采样法的原理，对于一个半径为r的超球面，首先定义这个超球面的累积分布函数（Cumulative Distribution Function，CDF）的值$\mathcal{F}(r)$如式（19）所示：

$$\mathcal{F}(r) = \int p(|\boldsymbol{x}| < r)\, d\boldsymbol{x} \quad (19)$$

类似于拉丁超立体采样法把参数空间分割成若干个超立体的做法，洋葱采样法把参数空间分割成若K个空心超球面，第k个超球面的 CDF 满足（20）关系式：

$$\mathcal{F}(r_k) = k/K \quad (20)$$

式中：

r_k —— 第k个超球面的半径；

k —— 表示第几个超球面；

K —— 空心超球面的总数.

在洋葱采样中，K的具体值需要提前被人为地设定好，然后洋葱采样便可根据K求r_k值. 由于$\mathcal{F}(\cdot)$是一个显式函数，$\mathcal{F}(\cdot)$ 的逆函数具有明显的表达式，所以给出K后可以很轻易地求出每个超球面所对应的r_k的具体值.

求出每个超球面的半径r_k后，洋葱采样按从外到内的顺序在每个超球面上进行均匀采点. 在每个超球面进行均匀采集$\mathcal{J}$个样本，采到的点为

$(\boldsymbol{x}_i, z(\boldsymbol{x}_i))_{i=1}^{\mathcal{J}}$. 然后把其中所有的失效点$\boldsymbol{x}_f$存放在失效数据集$\boldsymbol{D}_f$里，如式（21）所示：

$$\boldsymbol{D}_f \leftarrow \boldsymbol{D}_f \cup \left(\boldsymbol{x}_f, z\left(\boldsymbol{x}_f\right)\right) \tag{21}$$

在该超球面采到失效样本的概率U_k如式（22）：

$$U_k = 1/\mathcal{J}\sum_{i=1}^{\mathcal{J}} I(\boldsymbol{x}_i) \tag{22}$$

洋葱采样从外往内在各个超球面上进行采样，当遇到失效区域的边界时，U_k在数值上会突然骤降. 为了判断什么时候中止洋葱采样，OPTIMIS 设定了一个制停阈值τ，当$U_k < \tau$时，洋葱采样便停止.

洋葱采样的过程就像洋葱被一层一层地剥开，所以叫做洋葱采样法. 洋葱采样法能够在对电路无先验知识的情况下快速找到电路失效样本点.

3.2 归一化流模型

IS 方法通过建立一个$g(\boldsymbol{x})$概率分布，使用$g(\boldsymbol{x})$在电路失效区域内进行采样，从而提高采集失效点的效率. 为了能够良好地拟合出失效点的分布趋势，$g(\boldsymbol{x})$的选择至关重要. 目前大多数 IS 方法的$g(\boldsymbol{x})$选用高斯混合分布. 然而，高斯分布波形过于简单，即使多个高斯分布混合也不能完美地表示复杂的波形. 所以，OPTIMIS 没有使用普通的分布，而是选择了归一化流模型作为$g(\boldsymbol{x})$.

归一化流模型是一种可以使用神经网络构建复杂分布的模型[15]，神经网络使归一化流模型能够超越简单概率分布形式的限制. 它使用一系列的可逆变换将一个简单的分布变换为一个复杂的分布，从而实现对目标分布的建模，如式（23）所示：

$$\boldsymbol{x} = q(\boldsymbol{z}) \tag{23}$$

式中：

$\boldsymbol{x}$—— d 维随机向量，其概率密度为$g(\boldsymbol{x})$；

$q(\cdot)$—— 可逆函数，映射为$R^d \to R^d$；

$\boldsymbol{z}$—— 服从多元高斯分布$p(\boldsymbol{z})$的 d 维随机向量.

式（23）中，符合高斯分布$p(\boldsymbol{z})$的$\boldsymbol{z}$经过可逆函数$q(\cdot)$的映射后，得到了$\boldsymbol{x}$. 此时$\boldsymbol{x}$的已经不服从普通高斯分布$p(\boldsymbol{x})$，$\boldsymbol{x}$现在服从复杂的概率分布$g(\boldsymbol{x})$，如式（24）所示：

$$g(\boldsymbol{x}) = p(\boldsymbol{z})\left|\det \mathcal{D}q\left(\hbar(\boldsymbol{x})\right)\right|^{-1} \tag{24}$$

式中：

$\hbar(\cdot)$—— $q(\cdot)$的逆函数；

$\mathcal{D}\hbar(\boldsymbol{x})$—— $\partial\hbar(\boldsymbol{x})/\partial\boldsymbol{x}$，$\hbar(\boldsymbol{x})$的雅可比矩阵；

$\mathcal{D}q(\boldsymbol{x})$—— $\partial q(\boldsymbol{x})/\partial\boldsymbol{x}$，$q(\boldsymbol{x})$的雅可比矩阵.

使用$p(\boldsymbol{z})$进行采样后，对$\boldsymbol{z}$进行映射，得到$\boldsymbol{x} = q(\boldsymbol{z})$，这相当于直接使用$g(\boldsymbol{x})$进行采样. OPTIMIS 的目标是构建一个复杂的$g(\boldsymbol{x})$来拟合电路失效点的概率分布，然后使用$p(\boldsymbol{z})$进行采样，得到采样点$\boldsymbol{z}$，再对$\boldsymbol{z}$所做映射得到$\boldsymbol{x} = q(\boldsymbol{z})$，便可以得到符合$g(\boldsymbol{x})$的采样点$\boldsymbol{x}$.

经过不同模型的测试实验，最终发现归一化样条流模型[15]（Normalizing Spline Flow，NSF）最适合良率估计问题. NSF 具有强大灵活的映射能力，能拟合复杂的多元概率分布. 并且，NSF 特别擅长处理高维度分布的问题. 所以，在高维电路良率估计问题中，NSF 特别适合充当重要性采样方法里的采样分布$g(\boldsymbol{x})$.

3.3 算法总流程

前文介绍了 OPTIMIS 的基本组成部分：洋葱预采样法，NSF 概率模型. 本小节将前文介绍的 OPTIMIS 组件集装在一起，总体介绍 OPTIMIS 的方法流程. OPTIMIS 具体流程如下所示：

1）令 $j = 1$，j用于指示目前的循环次数；

2）进行洋葱预采样，得到初始的失效样本集$\boldsymbol{D}_f = (\boldsymbol{x}_{fi}, z(\boldsymbol{x}_{fi}))_{i=1}^{n_f}$；

3）使用$\boldsymbol{D}_f$训练 NSF 模型，作为 IS 的$g(\boldsymbol{x})$；

4）使用$g(\boldsymbol{x})$采样n_{IS}个点，得到$(\boldsymbol{x}_{ISi})_{i=1}^{n_{IS}}$；

5）将$(\boldsymbol{x}_{ISi})_{i=1}^{n_{IS}}$送入 SPICE，得到$(z(\boldsymbol{x}_{ISi}))_{i=1}^{n_{IS}}$；

6）计算临时失效率. 临时失效率的定义为$(U_f)_j = (1/n_{IS})\sum_i^{n_{IS}} I(\boldsymbol{x}_{ISi})p(\boldsymbol{x}_{ISi})/g(\boldsymbol{x}_{ISi})$；

7）从 IS 采样得到的数据集挑出所有失效点$(\boldsymbol{x}_{fi}, z(\boldsymbol{x}_{fi}))_{i=1}^{n_f}$，然后，令$\boldsymbol{D}_f \leftarrow \boldsymbol{D}_f \cup (\boldsymbol{x}_{fi}, z(\boldsymbol{x}_{fi}))_{i=1}^{n_f}$；

8）计算电路失效率$\widehat{P}_f = (1/j)\sum_{i=1}^{j}(U_f)_i$ ；

9）若$\widehat{P}_f$未达到精度要求，令索引值$j \leftarrow j+1$，并跳转到第 3 步.

4 实验验证

4.1 实验对象

本文的实验对象是四个不同规模的 SRAM. SRAM 是一种基于静态电路的半导体存储器件，常用于计算机内存中. SRAM 常由一组存储单元（也称为存储单元、位单元或单元）组成，每个存储单元可以存储一个比特（0 或 1）.

而由 6 个晶体管组成的 SRAM（6-Transistor SRAM，6-T SRAM）是一种常见的 SRAM 设计，如图 1 的右半部分所示. 其中，M1-M6 是 6 个晶体管，M1 和 M2 组成一个反相器，M3 和 M4 组成另一个

反相器，M5 和 M6 是访问电路的传输门．WL 是字线，是 word line 的缩写；BL 和$\overline{\text{BL}}$是输入、输出数据线，是 bit line 的缩写．在读取数据时，WL 使能，传输门打开，将反相器中的数据放到数据输出线上．在写入数据时，将数据和其互补位输入到 6-T SRAM 中，控制字线和传输门以写入数据．这就是最基本的一个 SRAM 电路．

而 6T-SRAM 是组成 SRAM 阵列的基本单元．SRAM 阵列是由一组 6-T SRAM 单元按列排列而成的 SRAM 存储结构，其电路结构图如图 1 左半部分所示，CELL 表示一个 6T-SRAM 存储单元，每个 SRAM 阵列包含多个 CELL，每个存储单元 CELL 位共享同两条位线．

本文将 SRAM 电路的每个晶体管的阈值电压、迁移率和栅极氧化物宽度被作为工艺变化参数$\boldsymbol{x}$，将 SRAM 的读取和写入时间作为评估性能指标y．所以y越大，表示性能越差．本文的实验对象分别为 18 维、108 维、569 维和 1093 维的 SRAM 电路．

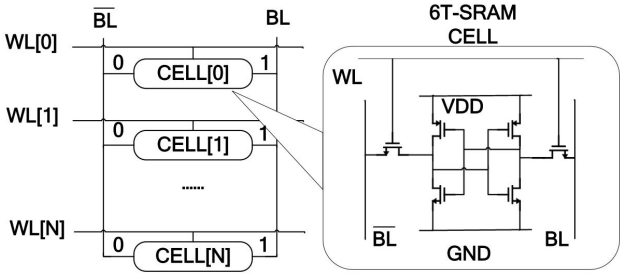

图 1　SRAM 阵列（左）和 6T-SRAM（右）

Fig 1 SRAM array column (left) and 6T-SRAM cell (right)

4.2 评估准则

本小节将介绍用于评估良率估计方法好坏的指标，介绍如何从精度和速度上评估一个方法的好坏．

MC 方法的估计结果被认为是电路良率的真实值．所以，良率估计的结果与 MC 方法的结果之间偏差的大小可以作为评判一个良率估计方法的准确度的标准．当良率估计的结果越靠近 MC 的结果时，说明这个良率估计方法越准确．本文使用相对误差$e_{\%}$来评估精准度，如式（25）所示：

$$e_{\%}=\frac{\left|P_{f\text{MC}}-\widehat{P}_f\right|}{P_{f\text{MC}}}\times 100\% \qquad (25)$$

式中：

$P_{f\text{MC}}$ —— 使用 MC 方法得到的失效率；

$\widehat{P}_f$ —— 使用其他良率估计方法得到的失效率．

在式（25）中， 当$e_{\%}$越小时，说明这个良率估计方法的效果越精准．

速度方面，从调用 SPICE 仿真器的次数$\mathcal{N}$来评估．由于进行一次 SPICE 仿真的时间很长，长到可以直接忽略掉良率估计模型训练的时间．所以本文直接使用调用 SPICE 的总次数作为评估一个良率估计方法速度的指标．当一个良率估计方法的$\mathcal{N}$越小时，说明这个方法的速度越快．

为了决定良率估计何时停止，本文引入质量因数（Figure of Merit，FOM）的概念．FOM 表示了良率估计值的收敛情况．当 FOM 到达一定值时，便停止良率估计．FOM 的定义如（26）所示：

$$\rho\left(\widehat{P}_f\right)=\frac{\text{std}\left(\widehat{P}_f\right)}{\widehat{P}_f} \qquad (26)$$

式中：

ρ —— 质量因数 FOM；

$\text{std}(\cdot)$ —— $\text{std}(\cdot)$表示取标准差；

$\widehat{P}_f$ —— 某个良率估计方法当前的良率估计值．

当$\rho\left(\widehat{P}_f\right)\le\varepsilon\sqrt{log(1/\delta)}$时，可以认为估计值$\widehat{P}_f$拥有 $(1-\varepsilon)\times 100\%$的估计精度，并且这个估计精度的置信度至少为$(1-\delta)\times 100\%$．

比如，当$\rho\left(\widehat{P}_f\right)\le 0.1$ 时，可以认为良率估计值$\widehat{P}_f$的精度和置信度都为 90%．所以，在良率估计过程中，每次得到估计值$\widehat{P}_f$后都进行质量因数$\rho\left(\widehat{P}_f\right)$的计算．当$\rho\left(\widehat{P}_f\right)<0.1$ 时，认为估计值已经达到了精度要求，良率估计方法便马上停止．这便是决定良率估计方法什么时候停止运行的准则．

4.3 比较对象

为了评估本文提出的良率估计方法性能究竟如何，本文要将提出方法和其他学界里最前沿的针对高维度电路问题的良率估计模型进行对比．下面将逐个介绍比较对象．

1）蒙特卡洛法．如前文所述，MC 是最传统的良率估计方法，MC 通过在原点按照高斯分布进行随机采样，来评估电路的良率．MC 非常稳定，只要采样量足够大就不会存在数值上的错误．在良率估计问题中， MC 方法的结果被视为电路的真实良率；

2）高维贝叶斯法[16]（High-Dimensional Bayesian Optimization，HDBO）．HDBO 是一个基于代理模型的良率估计方法．HDBO 同时也能处理高维度电路问题．HDBO 使用随机嵌入裁剪法来进行特征选择，并且使用贝叶斯优化来寻找失效点；

3）最小化正态重要性采样[2]（Minimized Norm Importance Sampling，MNIS）．MNIS 是一个基于重

要性采样的良率估计方法．MNIS 通过将采样中心转移到离原点最近的失效点上，进行基于单个高斯分布的重要性采样；

4）超球体聚类和采样法[3]（Hyperspherical Clustering and Sampling，HSCS）．HSCS 是 IS 方法，其使用一个聚类方法来识别多个电路失效区域，并且使用最小规范点来采样失效点；

5）自适应重要性采样法[4]（Adaptive Importance Sampling，AIS）．AIS 是一个 IS 方法，其动态地选择离原点最近的失效点作为训练集，然后高斯混合分布进行重要性采样；

6）自适应性聚类采样法[17]（Adaptive Clustering and Sampling，ACS）．ACS 是一个 IS 方法．ACS 使用聚类方法来辨识不同的失效区域，然后再使用一个高斯混合分布来拟合这些失效区域；

7）低秩张量近似法[7]（Low-Rank Tensor Approximation，LRTA）．LRTA 是一个代理模型方法，其使用多项式混乱展开式模型来代替了 SPICE；

8）在 OPTIMIS 的实验中，还使用了本文提出的第一个方法 ASDK 来作为比较对象．

4.4 实验结果

本文在 18 维和 569 维的 SRAM 电路上对 ASDK 进行了测试实验，其比较对象为 HSCS，HDBO 和 LRTA．电路的真实良率被设置在10^{-4}的量级．实验结果如表 1 所示，ASDK 比其他比较对象更精准，速度更快，最多比 HSCS 要快上 11.1 倍，比传统 MC 方法最多要快上 232.1 倍．实验结果证明了 ASDK 的精准性和快速性．

然后，本文在 108 维、569 维和 1093 维的 SRAM 电路上对 OPTIMIS 进行了测试实验，其比较对象为 MNIS，HSCS，AIS，ACS，LRTA 和 ASDK．电路的真实良率设置在10^{-5}的量级．实验结果如表 2 所示，OPTIMIS 比其他比较对象更精准，且速度更快，比 HSCS 最多要快上 12.7 倍，比传统 MC 方法最多要快上 273.8 倍．实验结果证明了 OPTIMIS 的准确性和快速性，其能够适用于高维的电路．

表 1 良率估计方法 ASDK 和比较对象在 18 维和 569 维 SRAM 电路上的实验结果

Table 2 Numerical results of ASDK and its comparison methods on the 18 and 569-dimensional SRAM circuits

	18 维 SRAM			569 维 SRAM		
方法	$\widehat{P}_f/10^{-4}$	$e_{\%}/\%$	$\mathcal{N}$/次	$\widehat{P}_f/10^{-4}$	$e_{\%}/\%$	$\mathcal{N}$/次
MC	4.83	0	265000	4.70	0	928500
HSCS	5.14	6.62	8100	5.82	23.83	44400
HDBO	6.25	29.40	3500	3.87	17.66	6100
LRTA	6.40	19.46	2200	5.60	19.14	5400
ASDK	**4.60**	**4.14**	**1350**	**4.39**	**6.60**	**4000**

表 2 良率估计方法 OPTIMIS 和比较对象在 108 维、569 维和 1093 维 SRAM 电路上的实验结果

Table 2 Numerical results of OPTIMIS and its comparison methods on the 108, 569 and 1093-dimensional SRAM circuits

	108 维 SRAM			569 维 SRAM			1093 维 SRAM		
方法	$\widehat{P}_f/10^{-5}$	$e_{\%}/\%$	$\mathcal{N}$/次	$\widehat{P}_f/10^{-5}$	$e_{\%}/\%$	$\mathcal{N}$/次	$\widehat{P}_f/10^{-5}$	$e_{\%}/\%$	$\mathcal{N}$/次
MC	5.01	0	699000	2.50	0	931000	4.80	0	1189000
MNIS	4.15	17.07	47500	2.07	17.33	59000	4.21	12.32	81000
HSCS	4.84	3.36	26500	2.86	14.27	46500	4.30	10.47	66000
AIS	4.75	5.21	12300	2.38	4.99	25700	4.43	7.75	38000
ACS	5.68	13.40	10400	2.73	9.19	22500	4.42	7.83	30400
LRTA	4.50	10.18	13000	2.26	9.60	18500	5.52	9.38	24000
ASDK	4.50	10.18	9200	2.30	8.00	11800	6.10	27.08	14550
OPTIMIS	**5.02**	**0.21**	**5300**	**2.49**	**0.25**	**3400**	**4.67**	**2.71**	**6400**

结 语

本文提出了两个在高维电路场合下的良率估计方法．第一个方法是基于代理模型的 ASDK，其由 DKLGP 代理模型，HSIC-LASSO 特征选择和信息熵最大化信息熵优化方法组成，速度最多比 MC 方法快上了 232.1 倍；第二个方法是基于 IS 的 OPTIMIS，由 NSF 分布，洋葱采样法和动态更新框架组成，在速度最多比 MC 方法快上了 273.8 倍．然而，这些方法也具有一定局限性，他们训练时需要依赖于高性能的显卡．

致谢:衷心感谢王兴政老师和邢炜老师的悉心指导！感谢概伦电子公司对实验的鼎力支持！

参考文献/References:

[1] Ciampolini L, Lafont J-C, Drissi F T and et al．Efficient yield estimation through generalized importance sampling with application to NBL-assisted SRAM bitcells [C]．*International Conference on Computer-Aided Design (ICCAD)*．2016：1-8.

[2] Dolecek L, Qazi M, Shah D and et al．Breaking the simulation barrier: SRAM evaluation through norm minimization [C]．*International Conference on Computer-Aided Design (ICCAD)*．2008：322-329.

[3] Wu W, Bodapati S and He L．Hyperspherical clustering and sampling for rare event analysis with multiple failure region coverage [C]．*International Symposium on Physical Design (ISPD)*．2016：153-160.

[4] Shi X, Liu F, Yang J and et al．A fast and robust failure analysis of memory circuits using adaptive importance sampling method [C]．*Design Automation Conference (DAC)*．2018：1-6.

[5] Shi X, Yan H, Li C and et al．A non-gaussian adaptive importance sampling method for high-dimensional and multi-failure-region yield analysis [C]．*International Conference on Computer-Aided Design (ICCAD)*．2020：1-8.

[6] Yao J, Ye Z and Wang Y．An efficient SRAM yield analysis and optimization method with adaptive online surrogate modeling [J]．*IEEE Transactions on Very Large Scale Integration Systems*，2014，23（7）：1245-1253.

[7] Shi X, Yan H, Huang Q and et al．Meta-model based high-dimensional yield analysis using low-rank tensor approximation [C]．*Design Automation Conference (DAC)*．2019：1-6.

[8] Zhang J, Yan S, Feng F and et al．A novel surrogate-based approach to yield estimation and optimization of microwave structures using combined quadratic mappings and matrix transfer functions [J]．*IEEE Transactions on Microwave Theory and Techniques*，2022，70（8）：3802-3816.

[9] Yin S, Jin X, Shi L and et al．Efficient bayesian yield analysis and optimization with active learning [C]．*Design Automation Conference (DAC)*．2022：1195-1200.

[10] Graham C, Talay D, Graham C and et al．Strong law of large numbers and Monte Carlo methods [J]．*Stochastic Simulation and Monte Carlo Methods: Mathematical Foundations of Stochastic Simulation*，2013（01）：13-35.

[11] Seeger M．Gaussian processes for machine learning [J]．*International journal of neural systems*，2004，14（02）：69-106.

[12] Wilson A G, Hu Z, Salakhutdinov R and et al．Deep kernel learning [C]．*Artificial intelligence and statistics (AISTATS)*．2016：370-378.

[13] Zhai J, Yan C, Wang S-G and et al．An efficient Bayesian yield estimation method for high dimensional and high sigma SRAM circuits [C]．*Design Automation Conference (DAC)*．2018：1-6.

[14] Yamada M, Jitkrittum W, Sigal L and et al．High-dimensional feature selection by feature-wise kernelized lasso [J]．*Neural computation*，2014，26（1）：185-207.

[15] Durkan C, Bekasov A, Murray I and et al．Neural spline flows [J]．*Advances in neural information processing systems*，2019，32.

[16] Hu H, Li P and Huang J Z．Enabling high-dimensional bayesian optimization for efficient failure detection of analog and mixed-signal circuits [C]．*Design Automation Conference (DAC)*．2019：1-6.

[17] Shi X, Yan H, Wang J and et al．Adaptive clustering and sampling for high-dimensional and multi-failure-region SRAM yield analysis [C]．*International Symposium on Physical Design (ISPD)*．2019：139-146.