Dealing with Aging and Yield in Scaled Technologies

Wei Ye, Mohamed Baker Alawieh, Che-Lun Hsu, Yibo Lin, and David Z. Pan

1 Introduction

The aging and yield issues arise with aggressive scaling of technologies and increasing design complexity [51, 53]. These issues impact the circuit performance and functionality throughout the product life cycles. The sources of aging and yield concerns lie in different aspects, getting more severe with technology scaling.

Modern VLSI designs have to cope with unreliable components and processes. Device aging and interconnect electromigration effects are likely to cause unexpected performance degradation and even malfunctions at the end of circuit life cycles. Meanwhile, process variations may lead to manufacturing defects and inconsistent device characterization, causing yield issues. Ignoring these effects leads short lifetime of designs and low yield, eventually increases the costs in volume production and maintenance.

Thus, for the robustness of VLSI design methodology and cycles, reliability and yield need to be accurately modeled, systematically optimized, and seamlessly integrated into the existing design flow. This chapter will survey critical aging and yield issues, and then review the state-of-the-art techniques to tackle them, including both modeling and optimization strategies which reside across the Physics and Circuit/Gate layers as part of the overall dependability scheme shown in Fig. 1. The strategies often involve synergistic cross-layer optimization due to the complicated VLSI design procedures nowadays. Novel modeling techniques leveraging machine learning are analyzed along with analytical optimization approaches.

W. Ye · M. B. Alawieh · C.-L. Hsu · D. Z. Pan (✉)

University of Texas at Austin, Austin, TX, US

e-mail: weiye@utexas.edu; mohdbaker@utexas.edu; chsul@utexas.edu; dpan@ece.utexas.edu

Y. Lin

Peking University, Beijing, China

e-mail: yibolin@pku.edu.cn

© The Author(s) 2021

J. Henkel, N. Dutt (eds.), *Dependable Embedded Systems*, Embedded Systems,

https://doi.org/10.1007/978-3-030-52017-5_17

409

Fig. 1 This chapter covers reliability and yield issues across the borders between Physics and Circuit/Gate levels

The chapter starts by investigating the device and interconnect reliability issues for modern VLSI designs in Sect. 2. It covers different device aging effects, such as bias temperature instability and hot carrier injection, as well as electromigration effects on power/ground and signal interconnections. The section introduces the modeling techniques along with optimization strategies to increase the design robustness under these effects. Section 3 dives into the state-of-the-art practices in yield issues for both analog and digital circuits. This section examines the impacts of process variations on circuit performance and manufacturing defects followed by effective modeling techniques to capture these issues early in the design flow. In the end, the chapter is concluded with Sect. 4.

2 Reliability Modeling and Optimization

With the continued feature size shrinking, reliability issue becomes increasingly severe. This section covers recent researches on aging modeling and analysis and divide the aging concerns into two sub-categories: aging at the device level and aging at the interconnect level.

2.1 Device Aging

As CMOS technologies continue to shrink, device reliability becomes a major challenge for high performance computing (HPC) and automotive applications which require robust circuit design. This section presents the device reliability modeling and optimization techniques along with mitigation strategies in advanced CMOS technologies.

Device reliability can be divided into time-independent and time-dependent categories. Time-independent reliability issues are caused by manufacturing variations

Fig. 2 Bathtub curve that illustrates the device life cycle

or noise such as random telegraph noise (RTN) or soft errors. Time-dependent reliability issues, also known as aging effects, can be illustrated using the bathtub curve in Fig. 2 which has high but decreasing failure rate in early life, low and constant failure rate in normal operation, and increasing high failure rate at the end of life wear-out period. This section focuses on modeling time-dependent reliability issues including *bias temperature instability* (BTI) and *hot carrier injection* (HCI).

BTI is an aging mechanism characterized by an increase in the device threshold voltage and a decrease in its mobility which eventually lead to an increase in the gate delay, and thus performance degradation [9, 58]. The two major factors contributing to the BTI phenomenon are the voltage bias and temperature. The term bias refers to the gate-to-source voltage bias applied to the transistor gate which is mostly a negative bias for PMOS, and a positive bias for NMOS. The theory behind BTI can be jointly explained by the reaction–diffusion (R-D) model and the charge trapping (CT) model [58]. The R-D model describes the degradation process when hole accumulation dissolves Si-H bond (reaction) and hydrogen diffuses away (diffusion), whereas the recovery stage takes place when voltage bias and duty factor stress is not present [25, 26]. The CT model explains the threshold voltage degradation by the trapped charge in the defected gate dielectrics. Early studies focused on BTI mitigation partially due to the fact that BTI dominates aging in early stages; however, HCI is more important at later stages where HCI contributes 40%–80% of device aging after 10 years of deployment [21, 47].

HCI is an aging phenomenon that degrades device drain current and is caused by the accumulation of carriers (electrons or holes) under the lateral electric fields, which can gain enough energy to damage and degrade the device mobility [16]. The traditional theory behind HCI was called lucky electron model, which is a field-based model [12, 57]. However, with the scaling of the supply voltage, the reduced electric field made HCI prediction based on field-based models a challenging task.

Fig. 3 BTI threshold voltage shift vs time for sub-45nm CMOS. BTI effect has stochastic nature in deep-sub-micro devices. Averaging each sample across large sample size of 800 (a) can achieve a well-defined voltage vs stress time curve while the voltage vs stress time trend in a much small sample size (b) contains larger variation [34]

Recent researches have proposed energy-driven theories to generalize HCI effects when devices are in low supply voltage [27, 52].

Characterizing aging degradation on circuit performance using aging model is a crucial step prior to optimization. Researchers can build deterministic models for BTI and HCI-related aging in old technologies such as 180 nm node. However, Kaczer et al. studied the threshold voltage shift vs time under the BTI effect and found its stochastic nature in deep-sub-micron nodes as shown in Fig. 3 [34]. Lorenz et al. proposed the first gate-level timing analysis considering NBTI and HCI [43]. Huard et al. [32] characterized a digital library gates under NBTI and HCI aging effects. Ren et al. discovered that BTI and HCI-related aging effects have layout dependencies [54]. In [21, 22], Fang et al. proposed frameworks to analyze BTI and HCI impacts on large digital circuits and [59] used ring oscillator-based sensors to estimate HCI/BTI induced circuit aging. Moreover, flip-flop based sensor was introduced in [2] to predict BTI aging circuit failure.

Recent researches not only model the aforementioned aging issues, but also propose design methods and optimizations for more reliable designs. Reliability optimization can be done at architecture level, logic synthesis level, and physical design level. At the architecture level, [48] demonstrated an aging analysis framework that examines NBTI and HCI to predict performance, power, and aging in the early design phase. Firouzi et al. [23] alleviated NBTI effects by using NOP (No operation) assignment and insertion in the MIPS processor. At synthesis level, Kumar et al. introduced standard cell mapping that considers signal probabilities to reduce BTI stress [35]. In [20], both HCI and BTI were considered during logic synthesis stage and put tighter timing constraint on paths with higher aging rate. Chakraborty et al. [13] optimized NBTI-induced clock skew in gated clock tree. Gate sizing [55, 64] and pin-reordering/logic restructuring [68] are also

implemented to minimize BTI effects. At the physical design level, Hsu et al. [29] proposed a layout-dependent aging mitigation framework for critical path timing during standard cell placement stage and [81] introduced aging-aware FPGA placement. Gate replacement techniques were used in [65] to co-optimize circuit aging and leakage.

2.2 Interconnect Electromigration

As IC technologies continue to scale, complex chip functionalities have been made possible by virtue of increasing transistor densities and aggressive scaling of interconnects. Besides, interconnects are getting thinner and running longer. These factors bring along higher current densities in metal wires, a phenomenon that further exacerbates *electromigration* (EM). The failure time from EM is worsened even further by the local temperature increase caused by self-heating of underlying FinFETs.

EM is the gradual displacement of atoms in metal under the influence of an applied electric field and is considered the primary failure mechanism for metal interconnects. After the migration of atoms with electrons in a metal line for a certain period, a void grows on one side, which increases the resistance of the metal line and may eventually lead to open circuits. Hillock is formed on the other side and may cause short circuits. Figure 4 shows the scanning electron microscopy (SEM) images of void and hillock.

2.2.1 Power EM Modeling

An empirical model for the mean time to failure (MTTF) of a metal line subjected to EM is given by Black's equation [11]

$$\text{MTTF} = \frac{A}{J^n} \exp\left(\frac{E_a}{kT}\right), \quad (1)$$

Fig. 4 A void and a hillock generated by electromigration [10]

where A is a constant which comprises the material properties and the geometry of the interconnect, J is the current density, E_a is the activation energy, k is the Boltzmann constant, and T is the temperature. n is the constant exponent of the current density and is usually set to 2. With Black's equation in Eq. (1), the relation between interconnect lifetime and both current and temperature can be readily estimated.

Power grid is one of the interconnect structures most vulnerable to EM due to its high unidirectional currents. Lower-level metal layers of power grids are more susceptible to EM failures due to smaller wire width. Besides, EM violations are most likely to occur around weak power grid connections, which deliver current to high power-consuming regions.

Hsu et al. [30] proposed an average power-based model to evaluate power grid static EM at placement stage. Ye et al. [78] further modified the model by considering the sum of the dynamic and leakage currents for a standard cell at this stage, which is given by:

$$I = \alpha \cdot C \cdot V_{DD} \cdot f + I_{\text{leak}},$$

where α is the cell activity factor, V_{DD} is the supply voltage, and f is the system clock frequency. C is the sum of the load capacitance and the output pin capacitance. Load capacitance further includes downstream gate capacitance and interconnect capacitance. Since nets have not been routed at this stage, half-perimeter wirelength (HPWL) [8] is widely adopted to estimate interconnect capacitance in placement. Power tile is defined as the region between two adjacent VDD (or VSS) power stripes and the adjacent power rails. Figure 5 demonstrates how to calculate the maximum current in the local power rails within a power tile. P_l and P_r are the left and right endpoints of the VDD power rail. d_i^l and d_i^r are the distances from the midpoint of the i -th cell to P_l and P_r , respectively. R_i^l and R_i^r are the wire resistances of the corresponding metal segments, which are proportional to d_i^l and d_i^r . The following equations hold

$$I_i^l = d_i^r / (d_i^l + d_i^r) I_i, \quad I_i^r = d_i^l / (d_i^l + d_i^r) I_i.$$

The currents drawn by all the cells in the power tile from P_l and P_r are computed as:

$$I^l = \sum_i I_i^l, \quad I^r = \sum_i I_i^r.$$

Therefore, there exists an EM violation in a particular power tile if $\max\{I^l, I^r\} > I_{\text{limit}}$. In this way, the EM failures in the local power rails can be estimated at the placement stage; thus, enabling an EM-aware placement that can effectively reduce the EM violations.

Fig. 5 The power grid model for current calculation in a power tile [78]

2.2.2 Signal EM Modeling

Previously, electromigration on signal interconnects does not draw great attention. Alternating current (AC) flows inside signal interconnects; when the direction of the current in an interconnect is reversed, the direction of EM diffusion is also reversed. The damage caused by EM can be partially cleared due to this compensation by material backflow. This effect is known as a self-healing, which can significantly extend the lifetime of a wire. Black's equation for AC is given by [40, 63]:

$$\text{MTTF} = \frac{A}{(J_+ - \gamma J_-)^n} \exp\left(\frac{E_a}{kT}\right), \quad (2)$$

where J_+ and J_- are the current densities during positive and negative pulses. γ is the self-healing coefficient which is determined by the duty factor of the current and other factors influencing the scale of self-healing, such as the frequency [39]. Previously, signal electromigration has attracted little attention due to the benefits of healing effect. However, EM failures in signal interconnects are no longer negligible due to higher clock frequencies, large transistor density, and the negative impact of FinFET self-heating at advanced nodes.

In [76], a set of features from the placement is extracted to train a machine learning model for EM detection before routing. Despite the fact that the current profile for the design is not available at the placement stage, multiple features that are highly correlated with the current can be crafted. These features can be divided into net-specific features and neighborhood related features. The net-specific features—including HPWL, the number of net pins, etc.—capture the net attributes. On the other hand, neighborhood related features are used to capture information about possible congestion around net pins.

The pre-routing signal EM hotspot prediction can be reduced to a classification problem [76]. A two-stage detection approach based on logistic regression shown in Fig. 6 is introduced to reduce the number of false alarms. In the first stage, a classification model M1 is trained to predict EM hotspots using all the nets in the training dataset. After the first stage, all nets with NH (Non-hotspot) prediction will be labeled as NH without further processing. For nets labeled H (Hotspot) by M1, a new model, M2, is trained to prune out false alarms. With an accurate classification model to detect signal EM hotspots based on the information available

Fig. 6 The flow of the two-stage signal EM hotspot detection approach [76]

at the placement stage, early stage EM handling is enabled, which reduces iterative EM fixing cost.

2.2.3 EM Optimization Flow

Through the preceding EM modeling in Eqs.(1) and (2), EM failures can be detected after the physical design stage, and then be fixed through layout modification. Xie et al. [69] proposed control logics to balance currents in both directions of power rails to mitigate the EM effects. Lienig [38] suggested the exploitation of several EM inhibiting measures, such as bamboo structure, short-length, and reservoir effects. Other studies [14, 33] considered global routing for EM optimization. In [49] de Paris et al. adopted a design strategy using non-default routing (NDR) rules to re-route the wire segments of EM-unsafe signal nets that present high current densities.

Conventionally, EM checking is invoked after the routing stage [36]. Current densities in metal wires are computed and compared with foundry-specified limits to detect EM failures. Next, the failures are fixed with engineering change order (ECO) efforts. EM checking leverages post-routing information to detect violations, which consequently limits the efficiency of addressing techniques. In the routing phase, the locations of standard cells and the corresponding current distribution are already fixed and the traditional fixing approaches such as wire widening and cell resizing are not effective enough to handle the ever-growing number of EM violations [1]. It is of vital importance to incorporate EM detection and fixing techniques into earlier stages of physical design (PD).

Two clear benefits are associated with such early stage EM handling. First, the number of EM violations can be decreased further by using various techniques at different design stages. Second, introducing early stage mitigation techniques can help reduce the resulting overhead when compared to post-routing fixing techniques. Thus, moving the EM detection and resolving steps to earlier stages of the physical design can help in reducing runtime or the number of iterations needed for design closure. In [78], a series of detailed placement techniques was proposed to mitigate power grid EM. Ye et al. [76] proposed a multistage EM mitigation approach at placement and routing phases to address the problematic nets detected by the classification model.

3 Yield Modeling and Optimization

3.1 Performance Modeling

With technologies descending deep into the sub-micron spectrum, process variation manifests itself among the most prominent factors limiting the product yield of analog and mixed-signal (AMS) circuits. Thus, it is indispensable to consider this variation in the design flow of modern ICs [42]. Conventionally, performance modeling has been adopted to capture this variability through analytical models that can be used in various applications such as yield estimation and design optimization [4].

Given a set of samples, the performance model coefficients are conventionally obtained through least-squares regression (LSR). However, LSR can build accurate models only when the number of samples is much greater than the number of unknown coefficients. Thus, given the high dimensionality of the performance models in complex AMS circuit designs, the simulation cost for building accurate models can be exorbitant. Hence, most recent performance modeling techniques incorporate additional information about the model to reduce the number of simulations needed [3, 5, 7].

3.1.1 Sparse Modeling

Although the number of basis functions representing the process variability is large, a few of these basis functions are required to accurately model a specific performance of interest (PoI). Hence, the vector of coefficients contains a small number of non-zero values corresponding to important basis functions [37]. This information can be incorporated in the modeling by constraining the number of non-zero coefficients in the final model.

While constraining the number of non-zero coefficients accurately reflects the sparse regression concept, the optimization problem is NP-hard. Besides heuristic approaches that select important basis functions in a greedy manner, Bayesian approaches have been widely applied to address this challenge [37]. In practice, a shrinking prior on the model coefficients is used to push their values close to zero. Examples of this include applying a Gaussian or Laplacian prior which results in Ridge and Lasso regression formulations, respectively. This allows incorporating sparse prior knowledge; however, such approaches do not perform explicit variable selection and they penalize high coefficients values by pushing all coefficients close to zero instead of selectively setting unimportant ones to zero.

On the other hand, a Bayesian spike and slab feature selection technique can be employed to efficiently build accurate performance models [7]. Spike and slab models explicitly partition variables into important and non-important, and then solve for the values of the important variables independently of the feature selection mechanism. A hierarchical Bayesian framework is utilized to determine both the

importance and value of the coefficients simultaneously. At its highest level, the hierarchy dictates that a particular coefficient is sampled from one of the two zero-mean prior Gaussian distributions: a low variance distribution centered around zero, referred to as the **spike**, and a large variance distribution, referred to as the **slab**.

This mixture of priors approach has demonstrated superior results compared to traditional sparse modeling schemes while also providing a feature selection framework that can easily select important features in the model [7].

3.1.2 Semi-Supervised Modeling

Traditionally, performance modeling has been approached from a purely supervised perspective. In other words, performance models were built by using labeled samples obtained through expensive simulations. However, as the complexity of designs increased, obtaining enough samples to build accurate models has become exorbitant. Recently, a new direction, derived from semi-supervised learning, has been explored to take advantage of unlabeled data to further improve the accuracy of performance modeling for AMS designs [3, 5].

In practice, the hierarchical structure of many AMS circuits can be leveraged to incorporate unlabeled data via Bayesian co-learning [5]. In particular, such an approach is composed of three major components. First, the entire circuit of interest is partitioned into multiple blocks based on the netlist hierarchy. Second, circuit-level performance models are built to map the block-level performance metrics to the PoI at the circuit level. Such a mapping is often low-dimensional; thus it can be accurately approximated by using a small number of simulation samples. Third, by combining the aforementioned low-dimensional models and an unlabeled data set, a complex, high-dimensional performance model for the PoI can be built based on semi-supervised learning.

To implement this modeling technique, a Bayesian inference is formulated to integrate the aforementioned three components, along with the prior knowledge on model coefficients, in a unified framework. Experimental results shown in [5] demonstrate that the proposed semi-supervised learning approach can achieve up to $3.6\times$ speedup when compared to sparse regression-based approaches.

While many AMS circuits exhibit a hierarchical structure, this feature is not always present. Hence, a more general semi-supervised framework which makes no assumption about the AMS circuit structure is desirable [3]. This can be achieved by incorporating a co-learning technique that leverages multiple views of the process variability to efficiently build a performance model. The first is the device level variations such as ΔV_{TH} or Δw_{eff} , while the second view is the underlying set of independent random variables, referred to as process variables. Traditionally, performance modeling targets expressing the PoI as an analytical function of process variables; however, capitalizing on information provided by the device level variability as an alternative view can help efficiently build the performance model for the PoI [3].

Fig. 7 An iteration of the semi-supervised co-learning modeling framework is illustrated [3]

As shown in Fig. 7, the key idea is to use a small number of labeled samples to build an initial model for each of the views of the data ($\mathbf{x}$ and $\mathbf{v}$), then attempt to iteratively bootstrap from the initial models using unlabeled data. In other words, initial models can be used to give pseudo-labels for unlabeled data, then the most confident predictions from a particular model are used as pseudo-samples for the other model. In each iteration step, *highly confident* pseudo-samples are fused with the small number of available labeled samples to build a new model. Experimental results demonstrated up to 30% speedup compared to sparse regression-based approaches [3].

3.1.3 Performance Optimization

Besides capturing the major sources of variability in AMS designs, one of the main applications of performance modeling is yield estimation and optimization. In practice, performance optimization can make use of trained models towards optimizing the performance of the design. This is established by first capturing correlations between the performance variability and the device sizes or reconfiguration knobs, then adjusting these parameters to improve the parametric yield [4, 6].

Moreover, with the increase in AMS circuits complexity, increasing nonlinearity stands out as major factor limiting the capabilities of performance modeling and optimization. Hence, performance optimization techniques relying on non-parametric surrogate models and Bayesian optimization frameworks have been recently proposed [31, 83]. These surrogate models are typically Gaussian Processes, and Bayesian optimization is used to find optimal values given a black-box function.

Bayesian Optimization is a sequential sampling based optimization technique for optimizing block-box objective functions. At each step, a set of optimal sampling locations are selected based on a chosen acquisition function. Then, queries of the

objective function to be optimized, e.g. performance of an AMS circuit, which can be costly, are only made at these optimized locations, e.g. via circuit simulations for AMS verification. The new data collected at each step augments the training dataset to retrain a probabilistic surrogate model that approximates the black-box function. Such iterative sampling scheme contributes directly to the accuracy of the surrogate model and guides the iterative global optimization process [31, 83].

3.2 Hotspot Detection

As the feature size of semiconductor transistors continues shrinking, the gap between exploding design demands and semiconductor manufacturability using current mainstream 193 nm lithography is becoming wider. Various designs for manufacturability (DFM) techniques have been proposed; however, due to the complexity of lithography systems and process variation, failures to print specific patterns still happen, which are referred to as lithography hotspots. Examples of two hotspot patterns are shown in Fig. 8.

The hotspot detection problem is to locate the lithography hotspots on a given layout in physical design and verification stages. Conventional simulation-based hotspot detection often relies on accurate yet complicated lithography models and therefore is extremely time-consuming. Efficient and accurate lithography hotspot detection is more desired for layout finishing and design closure in advanced technology nodes.

Pattern matching and machine learning based techniques have been proposed for quick and accurate detection of hotspots. Pattern matching forms a predefined library of hotspot layout patterns, and then compares any new pattern with the patterns in the library [70, 79]. There are some extensions that use fuzzy pattern matching to increase the coverage of the library [41, 66]. However, pattern matching,

Fig. 8 Example of two hotspot patterns. Core corresponds to the central location where a hotspot appears

Fig. 9 An example of a neural network for hotspot detection [74]

including fuzzy pattern matching, is insufficient to handle never-before-seen hotspot patterns. Recently, machine learning based approaches have demonstrated good generalization capability to recognize unseen hotspot patterns [17, 18, 45, 50, 80, 82].

3.2.1 Lithography Hotspot Detection with Machine Learning Models

Various machine learning models have been used as hotspot detection kernels with the goal of achieving high accuracy and low false alarms, including support vector machine (SVM) [18, 80], artificial neural network (ANN) [18], and boosting methods [45, 82]. Zhang et al. [82] have also proposed an online learning scheme to verify newly detected hotspots and incrementally update the model. Recently, deep neural networks (DNNs) have been adopted for hotspot detection [46, 60]. DNNs are able to perform automatic feature extraction on the high-dimensional layout during training, which spares the efforts spent on manual feature extraction. Promising empirical results have been observed with DNNs in several papers [46, 60, 73, 74]. Figure 9 shows a typical configuration of the structure of a DNN.

The performance of DNNs usually relies heavily on manual efforts to tune the networks, e.g., the number and types of layers. Matsunawa et al. [46] proposed a DNN structure for hotspot detection that can achieve low false alarms. Yang et al. [74] proposed Discrete Cosine Transform (DCT) based feature representation to reduce the image size for DNNs with a biased learning to improve accuracy and decrease false alarms.

3.2.2 Evaluation of Hotspot Detection Models

One special characteristic of lithography hotspot detection tasks is the imbalance in the layout datasets. Those lithography defects are critical, but their relative number is significantly small across the whole chip. Among various machine learning

Fig. 10 (a) An overlapping distribution of predicted scores for positive and negative samples and (b) the ROC curves of two example classifiers. As the threshold in (a) moves to the left, both FPR and TPR in (b) go up accordingly [77]

models at hand, the one with a highest true positive rate (TPR) and a lowest false positive rate (FPR) is preferred, but in real-world scenarios, there is always a trade-off between the two metrics. As Fig. 10a demonstrates, if the predicted score implies the belief of the classifier that a sample belongs to the positive class, decreasing the decision threshold (i.e., moving the threshold to the left) will increase both TPRs and FPRs.

The receiver operating characteristic (ROC) curve is considered a robust performance evaluation and model selection metric for imbalanced learning problems. For each setting of the decision threshold of a binary classification model (Fig. 10a), a pair of TPR and FPR values is obtained. By varying the decision threshold over the range $[0, 1]$, the ROC curve plots the relationship between TPR and the FPR (Fig. 10b).

The area under the ROC curve (AUC) is a threshold-independent metric which measures the fraction of times a positive instance is ranked higher than a negative one [62]. The closer the curve is pulled towards the upper left corner, the better is the ability of the classifier to discriminate between the two classes. For example, in Fig. 10b, classifier 2 has a better performance compared to classifier 1. Given that AUC is a robust measure of classification performances especially for imbalanced problems, it is useful to devise algorithms that directly optimize this metric during the training phase.

It has been proven that AUC is equivalent to the Wilcoxon–Mann–Whitney (WMW) statistic test of ranks [28, 44, 67]. However, AUC defined by the WMW metric is a sum of indicator functions which is not differentiable, to which gradient-based optimization methods cannot be applied. In order to make the problem tractable, it is necessary to apply convex relaxation to the AUC by replacing the indicator function with pairwise convex surrogate loss function. There are different forms of surrogate functions: pairwise squared loss [19, 24], pairwise hinge loss [61, 84], pairwise logistic loss [56], and piecewise function given in [71]. Ye et al. [77] compare these surrogate functions and show that those new surrogate loss functions are promising to outperform the cross-entropy loss when applied to the state-of-the-art neural network model for hotspot detection.

3.2.3 Data Efficient Hotspot Detection

Despite the effective machine learning models for hotspot detection, most of them rely on a large amount of data for training, resulting in huge data preparation overhead. Thus, it is necessary to improve the data efficiency during model training, i.e., to achieve high accuracy with as small amount of data as possible.

Chen et al. [15] proposed to leverage the information in unlabeled data during model training, when the amount of labeled data is small. They develop a semi-supervised learning framework, using a multi-task network with two branches to train the classification task for hotspot detection and the other unsupervised clustering task at the same time. The network will label those unlabeled data samples with pseudo-labels in each iteration. The pseudo-labeled data will be selected and added to training with different weights in the next iteration, where the weights here are determined by the clustering branch. The experimental results demonstrate over 3–4% accuracy improvement with 10%–50% amount of labeled training data.

Sometimes, there is additional flexibility to the learning problem where labels for unlabeled data be can queried. This extra capability enables the use of active learning which can actively select the data samples for training a better model. Yang et al [72] propose to iteratively query the actual labels for unlabeled data samples with low classification confidence in each training step and add these samples for training in the next step. The experiments on ICCAD 2016 contest benchmarks show similar accuracy with only 17% of training data samples.

One should note that semi-supervised learning and active learning are two orthogonal approaches to tackle the insufficient of labeled training data. Semi-supervised learning assumes the availability of unlabeled data, while active learning assumes the capability of querying the labels for unlabeled data. They can even be combined to achieve better data efficiency [85].

3.2.4 Trustworthiness of Hotspot Detection Models

Conventionally, hotspot detection approaches have been evaluated by judging upon the detection accuracy and the false alarm rate. While these metrics are indeed important, model trustworthiness is yet another metric that is critical for adopting machine learning based approaches. Addressing this concern requires machine learning models to provide confidence guarantees alongside the label predictions.

In practice, methods for obtaining confidence guarantees when using deep neural network are costly and not yet mature. However, Bayesian-based methods are the typical option when confidence estimation is needed. This can be achieved by adopting a Gaussian Process (GP) based classification that can provide a confidence metric for each predicted instance. With this approach, a label from a trained model is only valid when its confidence level matches a user-defined metric, otherwise, the prediction is marked as untrusted and lithography simulation can be used to further verify the results [75].

Fig. 11 Overall flow of Litho-GPA including data preparation with active sampling and hotspot detection with Gaussian process [75]

The flow of Litho-GPA, a framework for hotspot detection with Gaussian Process assurance, is illustrated in Fig. 11. In addition to addressing the issue of trust, Litho-GPA adopts active learning to reduce the amount of training data while favoring balance between classes in this dataset.

As a first step, an iterative weak classifier-based sampling scheme is leveraged to prepare a training set containing enough hotspots. Next, a Gaussian Process Regression (GPR) model is trained for the classification task with the selected data samples. This learned model is then used to make predictions with confidence estimation on the testing set. If GPR demonstrated high confidence in the predicted label, the result is trusted; otherwise, the unsure testing samples are verified with lithography simulations.

Experimental results shown in [75] demonstrate Litho-GPA can achieve comparable accuracy to the state-of-the-art deep learning approaches while obtaining on average 28% reduction in false alarms.

4 Conclusion

In this chapter, different important aging and yield issues in modern VLSI design and manufacturing have been discussed. These issues include device aging, interconnect electromigration, process variation, and manufacturing defects are likely to cause severe performance degradation or functionality failure, and thus need to be addressed early in the physical design flow. The chapter has surveyed recent techniques to not only build models for capturing these effects, but also to

develop strategies for optimizing them with the proposed models. These practices demonstrate that synergistic optimization and cross-layer feedback are encouraged to resolve the aforementioned aging and yield issues for robust VLSI design cycles.

References

1. Addressing signal electromigration (EM) in today's complex digital designs. <https://www.eetimes.com/document.asp?docid=1280370>
2. Agarwal, M., Paul, B.C., Zhang, M., Mitra, S.: Circuit failure prediction and its application to transistor aging. In: 25th IEEE VLSI Test Symposium (VTS'07), pp. 277–286 (2007)
3. Alawieh, M.B., Tang, X., Pan, D.: S²PM: semi-supervised learning for efficient performance modeling of analog and mixed signal circuit. In: IEEE/ACM Asia and South Pacific Design Automation Conference (ASPDAC), pp. 268–273 (2019)
4. Alawieh, M.B., Wang, F., Kang, R., Li, X., Joshi, R.: Efficient analog circuit optimization using sparse regression and error margining. In: IEEE International Symposium on Quality Electronic Design (ISQED), pp. 410–415 (2016)
5. Alawieh, M.B., Wang, F., Li, X.: Efficient hierarchical performance modeling for analog and mixed-signal circuits via bayesian co-learning. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* **37**(12), 1–13 (2018)
6. Alawieh, M.B., Wang, F., Tao, J., Yin, S., Jun, M., Li, X., Mukherjee Tamal Negi, R.N.: Efficient programming of reconfigurable radio frequency (RF) systems. In: IEEE/ACM International Conference on Computer-Aided Design (ICCAD), pp. 772–779 (2017)
7. Alawieh, M.B., Williamson, S.A., Pan, D.Z.: Rethinking sparsity in performance modeling for analog and mixed circuits using spike and slab models In: ACM/IEEE Design Automation Conference (DAC) (2019)
8. Alpert, C.J., Mehta, D.P., Sapatnekar, S.S.: Handbook of Algorithms for Physical Design Automation. CRC press, New York (2008)
9. Amrouch, H., Khaleghi, B., Gerstlauer, A., Henkel, J.: Reliability-aware design to suppress aging. In: ACM/IEEE Design Automation Conference (DAC), pp. 1–6 (2016)
10. Arnaud, L., Tartavel, G., Berger, T., Mariolle, D., Gobil, Y., Touet, I.: Microstructure and electromigration in copper damascene lines. *Microelectron. Reliab.* **40**(1), 77–86 (2000)
11. Black, J.R.: Electromigration—a brief survey and some recent results. *IEEE Trans. Electron Devices* **16**(4), 338–347 (1969)
12. C. Hu S. Tam, F.H.P.K.T.C., Terrill, K.W.: Hot-electron-induced MOSFET degradation—model, monitor, and improvement. *IEEE J. Solid-State Circuits* **20**(1), 295–305 (1985)
13. Chakraborty, A., Ganesan, G., Rajaram, A., Pan, D.Z.: Analysis and optimization of NBTI induced clock skew in gated clock trees. In: IEEE/ACM Proceedings Design, Automation and Test in Europe (DATE), pp. 296–299 (2009)
14. Chen, X., Liao, C., Wei, T., Hu, S.: An interconnect reliability-driven routing technique for electromigration failure avoidance. *IEEE Trans. Dependable Secure Comput.* **9**(5), 770–776 (2012)
15. Chen, Y., Lin, Y., Gai, T., Su, Y., Wei, Y., Pan, D.Z.: Semi-supervised hotspot detection with self-paced multi-task learning. In: IEEE/ACM Asia and South Pacific Design Automation Conference (ASPDAC), pp. 420–425 (2019)
16. Chen Ih-Chin, C.J.Y., Chenming, H.: The effect of channel hot-carrier stressing on gate-oxide integrity in MOSFETs. *IEEE Trans. Electron Devices* **35**(12), 2253–2258 (1988)
17. Ding, D., Torres, J.A., Pan, D.Z.: High performance lithography hotspot detection with successively refined pattern identifications and machine learning. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* **30**(11), 1621–1634 (2011)

18. Ding, D., Yu, B., Ghosh, J., Pan, D.Z.: EPIC: efficient prediction of IC manufacturing hotspots with a unified meta-classification formulation. In: IEEE/ACM Asia and South Pacific Design Automation Conference (ASPDAC), pp. 263–270 (2012)
19. Ding, Y., Zhao, P., Hoi, S.C.H., Ong, Y.S.: An adaptive gradient method for online AUC maximization. In: AAAI Conference on Artificial Intelligence, pp. 2568–2574 (2015)
20. Ebrahimi, M., Oboril, F., Kiamehr, S., Tahoori, M.B.: Aging-aware logic synthesis. In: IEEE/ACM International Conference on Computer-Aided Design (ICCAD), pp. 61–68 (2013)
21. Fang, J., Sapatnekar, S.S.: The impact of hot carriers on timing in large circuits. In: 17th Asia and South Pacific Design Automation Conference, pp. 591–596 (2012)
22. Fang, J., Sapatnekar, S.S.: The impact of BTI variations on timing in digital logic circuits. IEEE Trans. Device Mater. Reliab. **13**(1), 277–286 (2013)
23. Firouzi, F., Kiamehr, S., Tahoori, M.B.: NBTI mitigation by optimized NOP assignment and insertion. In: 2012 Design, Automation Test in Europe Conference Exhibition (DATE), pp. 218–223 (2012)
24. Gao, W., Jin, R., Zhu, S., Zhou, Z.H.: One-pass AUC optimization. In: International Conference on Machine Learning (ICML), pp. III906–III914 (2013)
25. Grasser, T., Kaczer, B., Goes, W., Reisinger, H., Aichinger, T., Hehenberger, P., Wagner, P., Schanovsky, F., Franco, J., Roussel, P., Nelhiebel, M.: Recent advances in understanding the bias temperature instability. In: 2010 International Electron Devices Meeting, pp. 4.4.1–4.4.4 (2010). doi:[10.1109/IEDM.2010.5703295](https://doi.org/10.1109/IEDM.2010.5703295)
26. Grasser, T., Kaczer, B., Goes, W., Reisinger, H., Aichinger, T., Hehenberger, P., Wagner, P.J., Schanovsky, F., Franco, J., Luque, M.T., et al.: The paradigm shift in understanding the bias temperature instability: from reaction–diffusion to switching oxide traps. IEEE Trans. Electron Devices **58**(11), 3652–3666 (2011)
27. Guerin, C., Huard, V., Bravaix, A.: The energy-driven hot-carrier degradation modes of nMOSFETs. IEEE Trans. Device Mater. Reliab. **7**(2), 225–235 (2007)
28. Hanley, J.A., McNeil, B.J.: The meaning and use of the area under a receiver operating characteristic (roc) curve. Radiology **143**(1), 29–36 (1982)
29. Hsu, C., Guo, S., Lin, Y., Xu, X., Li, M., Wang, R., Huang, R., Pan, D.Z.: Layout-dependent aging mitigation for critical path timing. In: IEEE/ACM Asia and South Pacific Design Automation Conference (ASPDAC), pp. 153–158 (2018)
30. Hsu, M.K., Katta, N., Lin, H.Y.H., Lin, K.T.H., Tam, K.H., Wang, K.C.H.: Design and manufacturing process co-optimization in nano-technology. In: IEEE/ACM International Conference on Computer-Aided Design (ICCAD), pp. 574–581 (2014)
31. Hu, H., Li, P., Huang, J.Z.: Enabling high-dimensional Bayesian optimization for efficient failure detection of analog and mixed-signal circuits. In: ACM/IEEE Design Automation Conference (DAC) (2019)
32. Huard, V., Parthasarathy, C., Bravaix, A., Guerin, C., Pion, E.: CMOS device design-in reliability approach in advanced nodes. In: 2009 IEEE International Reliability Physics Symposium, pp. 624–633 (2009)
33. Jiang, I.H.R., Chang, H.Y., Chang, C.L.: WiT: optimal wiring topology for electromigration avoidance. IEEE Trans. Very Large Scale Integr. Syst. **20**(4), 581–592 (2012)
34. Kaczer, B., Mahato, S., Valduga de Almeida Camargo, V., Toledano-Luque, M., Roussel, P.J., Grasser, T., Cathoor, F., Dobrovolsky, P., Zuber, P., Wirth, G., Groeseneken, G.: Atomistic approach to variability of bias-temperature instability in circuit simulations. In: 2011 International Reliability Physics Symposium, pp. XT.3.1–XT.3.5 (2011)
35. Kumar, S.V., Kim, C.H., Sapatnekar, S.S.: NBTI aware synthesis of digital circuits. In: ACM/IEEE Design Automation Conference (DAC), pp. 370–375 (2007)
36. Li, B., Muller, P., Warnock, J., Sigal, L., Badami, D.: A case study of electromigration reliability: from design point to system operations. In: IEEE International Reliability Physics Symposium (IRPS), pp. 2D.1.1–2D.1.6 (2015)
37. Li, X.: Finding deterministic solution from underdetermined equation: large-scale performance modeling by least angle regression. In: ACM/IEEE Design Automation Conference (DAC), pp. 364–369 (2009)

38. Lienig, J.: Electromigration and its impact on physical design in future technologies. In: ACM International Symposium on Physical Design (ISPD), pp. 33–40 (2013)
39. Lienig, J., Thiele, M.: Fundamentals of Electromigration-Aware Integrated Circuit Design. Springer, Berlin (2018)
40. Liew, B.K., Cheung, N.W., Hu, C.: Projecting interconnect electromigration lifetime for arbitrary current waveforms. *IEEE Trans. Electron Devices* **37**(5), 1343–1351 (1990)
41. Lin, S.Y., Chen, J.Y., Li, J.C., Wen, W.Y., Chang, S.C.: A novel fuzzy matching model for lithography hotspot detection. In: ACM/IEEE Design Automation Conference (DAC), pp. 68:1–68:6 (2013)
42. Lin, Y., Alawieh, M.B., Ye, W., Pan, D.: Machine learning for yield learning and optimization. In: IEEE International Test Conference (ITC), pp. 1–10 (2018)
43. Lorenz, D., Georgakos, G., Schlichtmann, U.: Aging analysis of circuit timing considering NBTI and HCI. In: 2009 15th IEEE International On-Line Testing Symposium, pp. 3–8 (2009)
44. Mann, H.B., Whitney, D.R.: On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.* **18**(1), 50–60 (1947)
45. Matsunawa, T., Gao, J.R., Yu, B., Pan, D.Z.: A new lithography hotspot detection framework based on AdaBoost classifier and simplified feature extraction. In: Proceedings of SPIE, vol. 9427 (2015)
46. Matsunawa, T., Nojima, S., Kotani, T.: Automatic layout feature extraction for lithography hotspot detection based on deep neural network. In: SPIE Advanced Lithography, vol. 9781 (2016)
47. Nigam, T., Parameshwaran, B., Krause, G.: Accurate product lifetime predictions based on device-level measurements. Proceedings of the 2009 IEEE International Reliability Physics Symposium, pp. 634–639 (2009)
48. Oboril, F., Tahoori, M.B.: ExtraTime: modeling and analysis of wearout due to transistor aging at microarchitecture-level. In: IEEE/IFIP International Conference on Dependable Systems and Networks (DSN), pp. 1–12 (2012)
49. de Paris, L., Posser, G., Reis, R.: Electromigration aware circuits by using special signal non-default routing rules. In: IEEE International Symposium on Circuits and Systems (ISCAS), pp. 2795–2798 (2016)
50. Park, J.W., Torres, A., Song, X.: Litho-aware machine learning for hotspot detection. *IEEE Trans. Comput. Aided Design Integr. Circuits Syst.* **37**(7), 1510–1514 (2018)
51. Pecht, M., Radojic, R., Rao, G.: Guidebook for managing silicon chip reliability. CRC Press, New York (2017)
52. Rauch, S.E., La Rosa, G.: The energy-driven paradigm of nMOSFET hot-carrier effects. *IEEE Trans. Device Mater. Reliab.* **5**(4), 701–705 (2005)
53. Reis, R., Cao, Y., Wirth, G.: Circuit design for reliability. Springer, Berlin (2015)
54. Ren, P., Xu, X., Hao, P., Wang, J., Wang, R., Li, M., Wang, J., Bu, W., Wu, J., Wong, W., Yu, S., et al.: Adding the missing time-dependent layout dependency into device-circuit-layout co-optimization: new findings on the layout dependent aging effects. In: IEEE International Electron Devices Meeting (IEDM) (2015)
55. Roy, S., Pan, D.Z.: Reliability aware gate sizing combating NBTI and oxide breakdown. In: International Conference on VLSI Design, pp. 38–43 (2014)
56. Rudin, C., Schapire, R.E.: Margin-based ranking and an equivalence between AdaBoost and RankBoost. *J. Mach. Learn. Res.* **10**(Oct), 2193–2232 (2009)
57. Tam, S., Ko, P.K., Hu, C.: Lucky-electron model of channel hot-electron injection in MOSFET's. *IEEE Trans. Electron Devices* **31**(9), 1116–1125 (1984)
58. Sapatnekar, S.S.: What happens when circuits grow old: aging issues in CMOS design. In: 2013 International Symposium on VLSI Technology, Systems and Application (VLSI-TSA), pp. 1–2 (2013)
59. Sengupta, D., Sapatnekar, S.S.: Estimating circuit aging due to BTI and HCI using ring-oscillator-based sensors. *IEEE Trans. Comput. Aided Design Integr. Circuits Syst.* **36**(10), 1688–1701 (2017)

60. Shin, M., Lee, J.H.: Accurate lithography hotspot detection using deep convolutional neural networks. *J. Micro/Nanolithogr. MEMS MOEMS* **15**(4), 043507 (2016)
61. Steck, H.: Hinge rank loss and the area under the ROC curve. In: *European Conference on Machine Learning*, pp. 347–358. Springer, Berlin (2007)
62. Swets, J.A., Pickett, R.M.: *Evaluation of diagnostic systems: methods from signal detection theory*. Academic Press, New York (1982)
63. Ting, L., May, J., Hunter, W., McPherson, J.: AC electromigration characterization and modeling of multilayered interconnects. In: *IEEE International Reliability Physics Symposium (IRPS)*, pp. 311–316 (1993)
64. Vattikonda, R., Wang, W., Cao, Y.: Modeling and minimization of PMOS NBTI effect for robust nanometer design. In: *ACM/IEEE Design Automation Conference (DAC)*, pp. 1047–1052 (2006)
65. Wang, Y., Chen, X., Wang, W., Cao, Y., Xie, Y., Yang, H.: Leakage power and circuit aging cooptimization by gate replacement techniques. *IEEE Trans. Very Large Scale Integr.(VLSI) Syst.* **19**(4), 615–628 (2011)
66. Wen, W.Y., Li, J.C., Lin, S.Y., Chen, J.Y., Chang, S.C.: A fuzzy-matching model with grid reduction for lithography hotspot detection. *IEEE Trans. Comput. Aided Design Integr. Circuits Syst.* **33**(11), 1671–1680 (2014)
67. Wilcoxon, F.: Individual comparisons by ranking methods. *Biom. Bulletin.* **1**(6), 80–83 (1945)
68. Wu, K.C., Marculescu, D.: Joint logic restructuring and pin reordering against NBTI-induced performance degradation. In: *IEEE/ACM Proceedings Design, Automation and Test in Europe (DATE)*, pp. 75–80 (2009)
69. Xie, J., Narayanan, V., Xie, Y.: Mitigating electromigration of power supply networks using bidirectional current stress. In: *ACM Great Lakes Symposium on VLSI (GLSVLSI)*, pp. 299–302 (2012)
70. Xu, J., Sinha, S., Chiang, C.C.: Accurate detection for process-hotspots with vias and incomplete specification. In: *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pp. 839–846 (2007)
71. Yan, L., Dodier, R.H., Mozer, M., Wolniewicz, R.H.: Optimizing classifier performance via an approximation to the Wilcoxon-Mann-Whitney statistic. In: *International Conference on Machine Learning (ICML)*, pp. 848–855 (2003)
72. Yang, H., Li, S., Tabery, C., Lin, B., Yu, B.: Bridging the gap between layout pattern sampling and hotspot detection via batch active sampling (2018). arXiv preprint:1807.06446
73. Yang, H., Luo, L., Su, J., Lin, C., Yu, B.: Imbalance aware lithography hotspot detection: a deep learning approach. In: *SPIE Advanced Lithography*, vol. 10148 (2017)
74. Yang, H., Su, J., Zou, Y., Yu, B., Young, E.F.Y.: Layout hotspot detection with feature tensor generation and deep biased learning. In: *ACM/IEEE Design Automation Conference (DAC)*, pp. 62:1–62:6 (2017)
75. Ye, W., Alawieh, M.B., Li, M., Lin, Y., Pan, D.Z.: Litho-GPA: Gaussian process assurance for lithography hotspot detection. In: *IEEE/ACM Proceedings Design, Automation and Test in Europe (DATE)* (2019)
76. Ye, W., Alawieh, M.B., Lin, Y., Pan, D.Z.: Tackling signal electromigration with learning-based detection and multistage mitigation. In: *IEEE/ACM Asia and South Pacific Design Automation Conference (ASPDAC)*, pp. 167–172 (2019)
77. Ye, W., Lin, Y., Li, M., Liu, Q., Pan, D.Z.: LithoROC: lithography hotspot detection with explicit ROC optimization. In: *IEEE/ACM Asia and South Pacific Design Automation Conference (ASPDAC)*, pp. 292–298 (2019)
78. Ye, W., Lin, Y., Xu, X., Li, W., Fu, Y., Sun, Y., Zhan, C., Pan, D.Z.: Placement mitigation techniques for power grid electromigration. In: *IEEE International Symposium on Low Power Electronics and Design (ISLPED)* (2017)
79. Yu, Y.T., Chan, Y.C., Sinha, S., Jiang, I.H.R., Chiang, C.: Accurate process-hotspot detection using critical design rule extraction. In: *ACM/IEEE Design Automation Conference (DAC)*, pp. 1167–1172 (2012)

80. Yu, Y.T., Lin, G.H., Jiang, I.H.R., Chiang, C.: Machine-learning-based hotspot detection using topological classification and critical feature extraction. *IEEE Trans. Comput. Aided Design Integr. Circuits Syst.* **34**(3), 460–470 (2015)
81. Zhang, H., Kochte, M.A., Schneider, E., Bauer, L., Wunderlich, H., Henkel, J.: Strap: Stress-aware placement for aging mitigation in runtime reconfigurable architectures. In: 2015 IEEE/ACM International Conference on Computer-Aided Design (ICCAD), pp. 38–45 (2015)
82. Zhang, H., Zhu, F., Li, H., Young, E.F.Y., Yu, B.: Bilinear lithography hotspot detection. In: ACM International Symposium on Physical Design (ISPD), pp. 7–14 (2017)
83. Zhang, S., Lyu, W., Wang, F., Yan, C., Hu, X., Zeng, X.: An efficient multi-fidelity Bayesian optimization approach for analog circuit synthesis. In: ACM/IEEE Design Automation Conference (DAC) (2019)
84. Zhao, P., Hoi, S.C.H., Jin, R., Yang, T.: Online AUC maximization. In: International Conference on Machine Learning (ICML), pp. 233–240 (2011)
85. Zhu, X., Lafferty, J., Ghahramani, Z.: Combining active learning and semi-supervised learning using Gaussian fields and harmonic functions. In: ICML 2003 workshop on the continuum from labeled to unlabeled data in machine learning and data mining, vol. 3 (2003)

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.