GridVAE: Fast Power Grid EM-Aware IR Drop Prediction and Fixing Accelerated by Variational AutoEncoder

Yibo Liu, Sheldon X.-D. Tan

***Abstract*—Electromigration (EM) remains the primary failure mechanism for copper-based interconnects in today's and future nanometer chip technologies. To ensure the longevity of on-chip power grids, effective EM-aware IR drop analysis is crucial. However, the existing sensitivity-based optimization approaches suffer from high computational costs, particularly for full-chip analysis. This paper proposes a novel and efficient framework for EM-aware IR drop prediction and fixing in full-chip power grids. Our approach leverages a variational autoencoder (VAE)-based model named *GridVAE* for accurate EM-aware IR drop predictions. Compared to state-of-the-art generative adversarial network (GAN) based methods, our VAE-based model offers improved accuracy and data efficiency. Building on the accurate predictions of GridVAE, we apply the sequence of linear programming-based optimizations to efficiently size the wires. This optimization framework takes advantage of the auto-differential capabilities of neural networks, significantly reducing computation time for sensitivity information computation. Our numerical results demonstrate the superiority of VAE over recently proposed GAN-based models, showcasing a remarkable 40% reduction in RMSE. Moreover, our proposed VAE-accelerated method achieves up to an 80X speedup (at least one order of magnitude) compared to conventional SLP-based methods for power grid EM-aware IR drop fixing. We validate our approach using synthesized power grid benchmarks from ARM Cortex-M0 processor design.**

I. Introduction

Electromigration (EM) is a physical phenomenon in which metal atoms migrate in response to various driving forces, such as the applied electrical field. In the context of modern very large-scale integration (VLSI) designs, EM remains the dominant reliability failure mechanism for copper-based interconnects, particularly in sub-nanometer technologies. Due to EM, the hydrostatic stress within the metal wire can reach critical levels, leading to resistance variations during migration. This phenomenon results in the formation of voids and hillocks at the cathode and anode, respectively, due to the accumulation and depletion of conducting electrons. The challenges posed by EM are further exacerbated as technology advances toward nanometer manufacturing processes. In these advanced nodes, the intricate interconnect layouts and the high current densities make EM a critical concern for the long-term reliability and performance of integrated circuits. Addressing EM-related issues becomes increasingly crucial in ensuring the functional integrity and operational lifespan of semiconductor devices in modern electronics.

On-chip power distribution network (PDN), as shown in Fig. 1, is a mesh-structured network that provides power to transistors from top metals, directly impacting on-chip performance and reliability. Since PDNs are usually vulnerable to EM-induced failures due to the large and unidirectional current on the PDN. To design robust PDN, designers have to properly size the PDN wires to meet the area and IR drop requirement. This task is changeling as the wires' resistance may change over time due to the EM effect, resulting in IR drops below the threshold voltage after years of aging effect.

Yibo Liu, and Sheldon X.-D. Tan are with the Department of Electrical and Computer Engineering, University of California, Riverside, Riverside, CA 92521, USA (e-mail: stan@ece.ucr.edu).

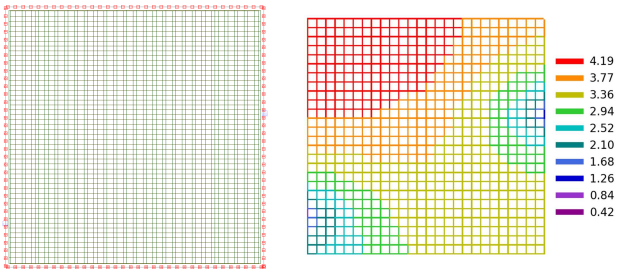


Fig. 1: (a) Power and ground networks of Cortex-M0 DesignStart; (b) Voltage drop map of the power network of (a).

Numerous past research works have investigated power grid network sizing, utilizing nonlinear or sequence of linear programming (SLP) methods [1], [2], [3], [4], [5], [6], [7], [8]. Zhou *et al.* [5], [9] proposed a power grid sizing approach based on multi-segment EM immortality check criteria. However, this EM immortality-constrained optimization proves too conservative, necessitating all interconnect trees to be immortal. In an effort to address this issue, Moudallal *et al.* [6] proposed a method that directly considers EM-induced IR drops in time-varying power grid networks. This method accounts for post-voiding resistance changes of wires through finite difference analysis of EM-induced stress in multi-segment wires, leading to a nonlinear problem solved via successive linear programming. Nonetheless, this method still incurs high computational costs, as the sensitivities of violating nodes must be computed by solving circuit matrices. More recently, Zhou *et al.* [8], [10] presented a conjugate gradient-based localized EM-aware IR drop fix for power grid networks, using the generative adversarial networks (GAN)-based deep neural network (DNN) modeling method called *GridNet* to compute gradient sensitivities.

Recently, a variational autoencoder (VAE) was proposed for generative applications [11]. In contrast to standard autoencoders, the VAE's encoder generates parameters of a predefined distribution in the latent space for each input. Moreover, it introduces additional regularization in the cost function, encouraging the resulting latent distribution to closely approximate the predefined normal distribution, thereby enhancing the reliability of generating new data. VAE has found applications

in various fields, including quantum circuit design for drug discovery [12], generative guided analog routing [13], and analog circuit sizing [14], underscoring its versatility and potential in different research areas.

Inspired by the generative capabilities of VAE, this paper aims to explore the potential of VAE for efficient prediction of EM-aware IR drop in on-chip power grid networks, to reduce computational costs in power grid design and optimization while ensuring compliance with IR drop and EM lifetime targets. The key contributions of this study are summarized as follows:

- First, we propose a deep neural network model based on VAE structure, called *GridVAE*, designed to model full-chip EM-aware IR drop data obtained from numerical EM-aware IR drop analysis tools. By comparing GridVAE with the latest GAN-based model [8], we observe a significant 40% reduction in Root Mean Square Error (RMSE) based on synthesized power grid benchmark circuits.
- Second, building upon the new VAE-based EM-aware IR drop models, we leverage its differential nature to fast compute the sensitivity of cost functions to the wire width. This led to an accelerated power grid wire sizing method for ensuring the chip's EM lifetime based on the sequence of linear program optimization framework. Numerical results on a number of synthesized power grid benchmarks from ARM Cortex-M0 processor designs show that the *GridVAE* enabled optimization can provide up to over 80X speedup over the existing analytical matrix solving-based SLP method [6].

The rest of the paper is organized as follows: Section II reviews the related preliminary works on the EM-induced IR drop analysis and current EM-aware power grid optimization strategy. Section III introduces the VAE-based EM-aware IR drop prediction method and the fast full-chip IR drop fixing strategy accelerated by our model. Experiment setup, numerical results, as long as analysis and discussions are summarized in Section IV. Section V concludes the paper.

II. Preliminaries

In this section, we first summarize and review some related preliminary works of on-chip power grid EM-induced IR drop analysis and mitigation approaches.

A. Full-chip EM-induced IR drop analysis

As mentioned in the previous section, EM is a physical phenomenon that can lead to resistance increase or even open-wire segments. The IR drop of the power grid wires may change due to the EM-induced aging effect. This means we have to consider the power girds IR drop as time-varying characters [15], [16], [17], [18]. On the other hand, the failed wire segments change the current distributions of all the interconnect wires, which may further accelerate the failure process. Hence, to emulate the on-chip power grid IR-drop after the aging effect, one has to consider the interplay between the two physics: electrical characteristics and hydrostatic stress in the interconnect wires.

EMspice [15], [19] is an open source tool that conducts the full-chip power grid network coupled EM-IR drop simulation with the dynamic interplay between the hydrostatic stress and electronic current/voltage. It solves the coupled time-varying partial differential equations in the time domain to obtain the stress evolution, and finally reports resulted IR drop and EM failure hotspots at the target aging time, such as 10 years. The tool consists of a finite difference time domain (FDTD) solver for EM stress and a linear network DC solver for IR drop. The linear network IR drop solver passes time-dependent current densities and P/G layout information to the finite difference time domain FDTD EM solver, and the FDTD EM solver provides the IR drop solver with new resistance information. These two simulations are coupled together and must be solved together, which can be described as

$$\begin{aligned} &\mathbf{C}\dot{\sigma}(t) = \mathbf{A}\sigma(t) + \mathbf{P}I(t), \\ &\mathcal{V}_v(t) = \int_{\Omega_L} \frac{\sigma(t)}{B} d\mathcal{V}, \\ &\mathbf{M}(t) \times u(t) = \mathbf{P}I(t), \\ &\sigma(0) = [\sigma_1(0), \sigma_2(0), ..., \sigma_n(0)] \ , at \ t = 0 \end{aligned} \tag{1}$$

In the above equations, $\mathbf{M}(t)$ is the time-varying power grid conductance matrix, as the resistance changes due to the EM failure process. $\mathbf{P}$ is the input matrix, and I(t) represents the current sources from the chip. $\mathbf{C}$ is the identity matrix, and $\mathbf{A}$ is the coefficient matrix. $\sigma_n(0)$ denotes the initial stress at time step $t = 0$, for node n. Thee stress from the previous simulation step is used as the initial condition for each new time step. Such iterative coupled analysis on a long target lifetime can be extremely time-consuming for large power grid networks.

There are some research efforts to facilitate the IR drop/EM-aware IR drop analysis by leveraging machine learning-based approaches to build the surrogate models, including [20], [21], [22], [23], [24] to reduce the evaluation time during the physical design process. Lin *et al.* [20] tried to extract power and physical features from cells and layouts to conduct the full-chip dynamic IR drop analysis. Fang *et al.* [21] proposed to train the models for the localized layout region to improve the scalability. A convolutional neural network (CNN) model which incorporates design-dependent features during pre-processing was proposed by Xie *et al.* [23]. Ho *et al.* [22] presented incremental IR drop prediction and mitigation by applying more electrical and physical features to train the gradient-boosting framework. Chhabria *et al.* [24] proposed *IREDGe*, a CNN-based generative network method to predict on-chip IR drop contours with image-to-image and sequence-to-sequence translation tasks. Recently Zhou *et al.* [8] considered more accurate physics-based EM effects into IR drop to build such surrogate models. It adopted the GAN-based structure to model the resulting EM-aware IR drops. This method regards the time, 2D power grid structure with input current and voltages as input images (series of images) and outputs the voltage map images. Such a surrogate model can help speed up wire sizing by fast EM-aware IR drop estimation and the sensitivity of objective functions concerning the wire geometries. It was initially applied to localized PDN optimization and then has been extended to full-chip optimization [10]. However, the GAN-based models suffer from the difficulties of training the generator and discriminator together. Also, the GAN model's deterministic latent variable encoding and decoding approach make the latent space lack of continuity and interpretability and less generative capability

(due to non-regulated latent space), eventually affecting the prediction accuracy.

B. Existing EM-aware PDN optimization

Besides the full-chip aging-aware IR-drop estimation, one also needs to fix or alleviate the excessive IR-drop to ensure a robust PDN design. Lots of past research applies nonlinear methods [1], [2], [3], [4], [5], [6] to size the PDN wires properly. These optimization strategies aim to meet the IR drop requirement at the target lifetime or extend the main time to failure (MTTF) with minimized metal routing area.

The SLP-based method was proposed first in [3] based on Black's equation. Then this method was extended to consider multi-segment wires [25]. But this method can be too conservative as it requires all the wires to be immortal after optimization. Although, this method has been extended to consider a targeted lifetime by allowing some wires to fail and optimizing the rest of the wires [5]. Recently [6] proposed to directly optimize EM-induced IR drops on the time-varying power grid networks EM caused by EM-induced aging using the SLP method. However, the EM-induced IR drop is still computed by solving Korhonen equations, and the sensitivities of the IR drop with respect to the wires are calculated through the matrix-solving method. The solving process is severely time-consuming, especially when the power grid is enormous.

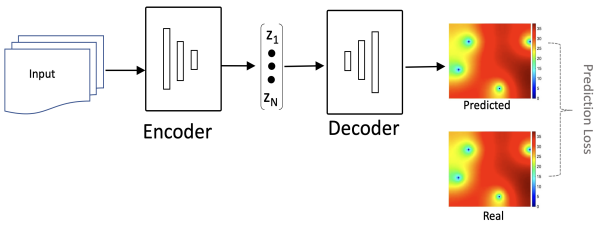
C. Variational Autoencoder

this paper adopts VAE [11] as the machine learning model for the EM-aware IR drop estimations. VAE follows a similar structure to autoencoder(AE).

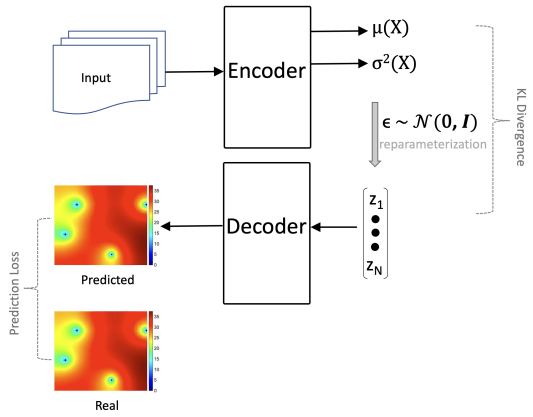
As shown in Fig.2(a), the traditional AE structure includes encoder and decoder structures. Input will be encoded to low-dimensional latent variables by the encoder and then decoded to the output by the decoder. The GAN model adopts the standard AE as a *generator* to generate the result and uses an additional binary output CNN as the discriminative network only in the training stage to better train the autoencoder.

Variational autoencoder, as shown in Fig.2(b), is a unique generative model derived from the standard autoencoder, which is used to generate new data similar to the input data it's trained on. Compared to the traditional autoencoder structure, VAE does not directly encode the input to a determined latent variable. Instead, it is encoded to a distribution over the latent space $Q(Z|X)$, and then sampled from this distribution to generate a latent variable for the decoder. This step is crucial to the VAE's characteristics over AE and AE-based GAN models, such as latent space interpretability, and training stability. Interpretability of latent space means that the latent space is smooth and interpretable, enabling friendly properties like the ability to interpolate between different points in the latent space and generate new samples that exhibit a smooth transition between the characteristics of the original points. On the other hand, AE does not explicitly enforce such a structure, making the latent space less interpretable and interpolations might not always produce meaningful outputs. Usually, the VAE latent variable z is generated using a reparameterization approach to ensure the model backpropagation with gradient descent. We denote the mean of $Q(Z|X)$ as μ and the standard deviation as σ, and sample $\epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$, then compute z as equation.(2):

$$z = \mu(X) + \sigma * \epsilon(X) \tag{2}$$



(a)



(b)

Fig. 2: (a) Architecture of an auto encoder (b)Architecture of a variational auto encoder

III. Proposed VAE-accelerated EM-aware IR-drop fixing method

A. VAE-based EM-aware IR-drop prediction model

We use VAE as the backbone for our EM-aware IR-drop prediction model. The structure is shown in Fig.3, and we use a power grid with 64 rows and 64 columns as an example.

The model input consists of four channels. They are the power grid branch conductance split into vertical conductance $\mathbb{R}^{64\times64\times1}$ and horizontal conductance $\mathbb{R}^{64\times64\times1}$, the current source $\mathbb{R}^{64\times64\times1}$ drawn to other circuit layers, and a target lifetime $\boldsymbol{t}$expanded into $\mathbb{R}^{64\times64\times1}$ by channel-wise duplication as the conditional information. Hence the total input will be 64x64x4. The conductance and current are normalized, then the input $\boldsymbol{x}$ is sent into the encoder and encoded as μ_x and σ_x. In this example, the encoder consists of four convolutional layers followed by two fully-connected layers. The detailed dimensions of the layer have been indicated in Fig.3. Next we use the reparameterization trick described in II-C to sample and obtain the latent variable $\mathbf{z}$ and send it to the decoder. The latent variable $\mathbf{z}$ is a 1x20 vector. The decoder is designed to mirror the architecture of the encoder in a symmetric manner, except the output layer has only one channel. The output is $\mathbb{R}^{64\times64\times1}$ EM-aware IR drop at target aging year. If the power grid exceeds the size of 64 x 64 but is smaller than 128 x 128, we pad the input into the standard size of 128 x 128 and add one more convolutional layer in both the encoder and the decoder network. For the smaller-sized power grid, we follow the same routine remove the convolutional layer to fit the input size.

The Loss function for a VAE contains two parts. The first part is the *reconstruction loss*, also called *prediction loss* when the input and output are expected to differ. The

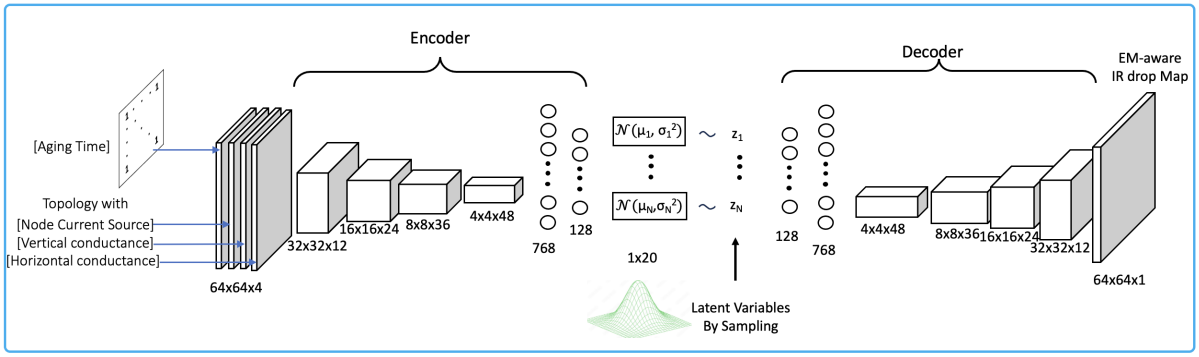


Fig. 3: VAE-based model for EM-aware IR drop prediction

prediction loss measures the difference between the decoded result $\hat{y} = P(\hat{y}|z)$ and real data y, encouraging the decoded output data to be similar to the label. As our data is continuous, we allocate the mean squared error (MSE) for the prediction loss evaluation (3), where N is the total number of output pixels.

$$MSE(y, \hat{y}) = \frac{1}{N}\sum_{1}^{N}(y - \hat{y})^2 \tag{3}$$

The second part is the *Kullback-Leibler (KL) divergence* between the encoder's latent variable distribution $Q(z|x)$ output and a chosen prior distribution, usually a standard multivariate Gaussian distribution $\mathcal{N}(\mathbf{0}, \mathbf{I})$. The *KL divergence* measures how one probability distribution differs from a second, which acts as a regularization term to force the encoded latent variable distribution from the given training dataset to be close to a standard normal distribution, and encourage the network to use the latent space efficiently. The *KL divergence* in VAE can be written as:

$$D_{KL}(\mathcal{N}(\mu_x, \sigma_x^2) \,||\, \mathcal{N}(\mathbf{0}, \mathbf{I})) = \frac{1}{2}(-\log\sigma^2 + \mu^2 + \sigma^2 - 1) \tag{4}$$

Hence the training target is to minimize the total loss:

$$\min\{MSE(y, \hat{y}) \;+\; \lambda \cdot D_{KL}(\mathcal{N}(\mu_x, \sigma_x^2) \,||\, \mathcal{N}(\mathbf{0}, \mathbf{I}))\} \tag{5}$$

where λ is the hyper parameter that adjusted similar to [26] to prevent KL vanishing.

We measure the prediction accuracy in RMSE (6), the unit is mV

$$RMSE = \sqrt{\frac{1}{N}\sum_{1}^{N}(y - \hat{y})^2} \tag{6}$$

The data preprocessing before the model training is as follows: First, the circuit layouts are automatically created by Synopsys IC compiler from a synthesized gate-level netlist and a standard cell library. Then IC Compiler output power grid information is sent to the power grid file parser, which reorganizes the information, including structure, wire layer, wire length, wire resistance values, node location, voltage, current source, etc. Next, the *EMspice* provides the EM-aware electrical information from the result above. Eventually, we parse these EM-aware electrical features and the topological information for the circuit layout and send it to our VAE-based model for training and testing.

TABLE I: Power Grid Designs Detail

circuit	# nodes	# Trees	# voltage sources	V_{DD} (V)
Design1	1024	64	2	1.05
Design2	4096	128	4	1.05
Design3	16384	256	9	1.05

B. Fast power grid EM-aware IR drop fixing framework

The proposed VAE-based fixing strategy workflow is shown in Fig. 4.

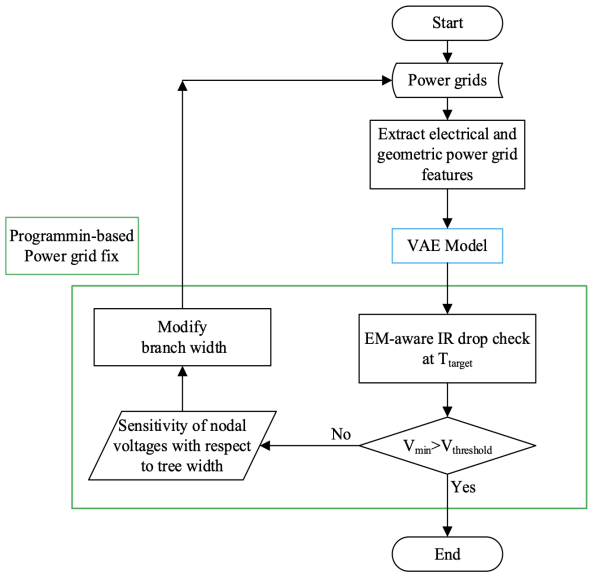


Fig. 4: Framework of power gird IR drop fixing method accelerated by the VAE model.

1) Problem formulation: The proposed method intends to solve the following problem: Given the power grid information at T=0, our VAE-based model has predicted the EM-aware IR drop at the target aging lifetime $\boldsymbol{t}$ in III-A. The power grid has EM-induced IR drop violations and demands fixing if it has node voltage drop $\boldsymbol{v}$ above the threshold V_{th}. We wish to alleviate the EM-aware IR drop failure by resizing the power grid interconnect trees' width with a minimum metal area increase.

The problem can be formulated as:

$$\begin{aligned} &\text{Minimize} && a^t s \\ &s.t. && v(t,s) \le V_{th} \\ & && s \in S \triangleq \{s \in R^{nt} : 1 \le s \le s_{up}\} \end{aligned} \tag{7}$$

In (7), $a = [a_1, a_2, \ldots, a_{nt}] = [w_1 l_1, w_2 l_2, \ldots, w_{nt} l_{nt}]$ refers to the metal areas of the power grid with n_t trees, w_i and l_i are the i_{th} tree's width and length separately. As for the constraint, $s = [s_1, s_2, \ldots, s_{nt}]$ is the resizing factor for the power grid trees. S is the feasible region, where s_{up} is the upper bound for s. Similar to [6], we assume the original power grid trees are already set to their minimum width. Hence we only increase the treewidth and $s \ge 1$. The feasible region reflects both the basic design rules and case-by-case user requirements, such as the criteria of minimum interconnect treewidth, the minimum spacing to prevent the interconnect trees overlap, and the maximum metal area usage, etc. The node IR drop $v(t,s)$ is the voltage difference between the power grid interconnect node and the power supply, it is a nonlinear function to s at aging time t. The maximum allowable voltage drop threshold V_{th} is given by the user. We assumed it to be five percent of the power supply voltage.

2) Programming-based optimization: As we mentioned in III-B1, the EM-aware PG node voltage drop $v(t,s)$ is a nonlinear function to s at aging time t. Hence (7) is a nonlinear optimization problem. We solve it by a stepping strategy, in which we linearize the voltage drop at the current latest solution point by Taylor's expansion (8) and solve it with linear programming (LP) solver. We repeat this operation until the power grid has no EM-aware IR drop violation.

$$v(t, s^{(i+1)}) \triangleq v(t, s^{(i)}) + \frac{\partial v(t, s^{(i)})}{\partial s} \cdot \delta s \tag{8}$$

where $s^{(i)}$ denotes the current power grid resizing vector and $s^{(i+1)}$ is defined as

$$s^{(i+1)} = s^{(i)} + \delta s \tag{9}$$

$\dfrac{\partial v(t,s)}{\partial s}$ is the $n \times n_t$ *Jacobian* matrix of $v(t,s)$ with respect to s, describes how the node voltage drops at target time $T = t$ respond to the corresponding treewidth rescaling.

$$\frac{\partial v(t,s)}{\partial s} = \mathbf{J}_{n \times n_t}(s,t) = \begin{bmatrix} \frac{\partial v(1,t)}{\partial s_1} & \frac{\partial v(1,t)}{\partial s_2} & \cdots & \frac{\partial v(1,t)}{\partial s_{n_t}} \\ \frac{\partial v(2,t)}{\partial s_1} & \frac{\partial v(2,t)}{\partial s_2} & \cdots & \frac{\partial v(2,t)}{\partial s_{n_t}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial v(n,t)}{\partial s_1} & \frac{\partial v(n,t)}{\partial s_2} & \cdots & \frac{\partial v(n,t)}{\partial s_{n_t}} \end{bmatrix} \tag{10}$$

3) Fast gradient computation via auto differentiation: PyTorch auto differentiation can provide the gradient of output node voltages to the input wire segment conductances, which is $\dfrac{\partial v(t,s)}{\partial g}$. Then we can quickly get (10) by chain rule, as each power grid tree g_i has its resizing factor s_i and will not be infected by other resizing factors:

$$\begin{aligned} &\frac{\partial v}{\partial s} = \frac{\partial v}{\partial g}\frac{\partial g}{\partial s}, \\ &\frac{\partial g_i}{\partial s_k} = \begin{cases} 0, & \text{if i} \neq \text{k} \\ g_i, & \text{if i = k} \end{cases} \end{aligned} \tag{11}$$

As a comparison to the sensitivity acquisition by matrix solving method [6], we briefly review its main steps. For a power grid network represented by the node conductance matrix *G(t,s)*, it can be written in the following format (12):

$$G(t,s) \cdot v(t,s) = j(t) \tag{12}$$

where $j(t)$ is $n \times 1$ vectors representing and node current vector for the power grid.

And the sensitivity is calculated as followed (13):

$$\frac{\partial v(t,s)}{\partial s_k} = -G^{-1} \cdot \frac{\partial G(t,s)}{\partial s_k} \cdot G^{-1} \cdot j(t) \tag{13}$$

To solve the above equation and obtain $\frac{\partial v(t,s)}{\partial s_k}$, one must first construct the sparse matrices G and $\frac{\partial G(t,s)}{\partial s_k}$ for all k, then perform matrix solving. Each column of the *Jacobian* matrix (10) has to be calculated through the process of (13) for the corresponding power grid tree. Hence building such a *Jacobian* matrix in the traditional method is computationally expensive.

IV. Experimental results and discussions

A. Experiment setup

The experiment is set up on a Linux server with 2 Xeon E5-2698v2 processors and Nvidia Titan X RTX GPU. The applied VAE-based model is built with the PyTorch package. The linear programming-based power grid fixing part is implemented in Python. The test cases are randomly generated following the IBM format, based on the topology we extract from the real design and the constraints that the user defines. It comprises of VDD power nets, VSS ground nets, and two external supplies. The power grid file parser output is consistent with the IBM power grid benchmarks [27]. We have three different topology designs and generate large amounts of IBM format power grid networks to ensure different workloads can be tested and verified.

Design 1 comes from Cortex-M0 DesignStart processor, which implements ARMv6-M 32-bit architecture and is routed and placed using IC Compiler with 32/28nm Generic Library from Synopsys. The Cortex power grid consists of two layers and one thousand interconnect trees, as shown in Fig. 1(a). Fig. 1(b) shows the voltage drop from the same power grid, the unit is mV. Design 2 and Design 3 follow a similar pattern, the detailed information is in Table. I. To train the model, each topology dataset contains 10000 pairs of samples (workloads and aging time, EM-aware IR drop). The maximum allowable IR drop is set to 5% V_{dd} and the target lifetime T is set to 10 years.

B. The result and performance

We implemented the excessive linear programming method in [6] and our proposed method for comparison.

TABLE II: Prediction results of VAE model and GAN model on different designs

Circuit	# nodes	RMSE (mV)	
		GAN	VAE
Design1	1024	5.697	3.144
Design2	4096	6.100	3.519
Design3	16384	3.922	3.462

TABLE III: Comparison of the proposed VAE-accelerated SLP optimization method against the existing method

Circuit	PG Size	naive SLP method [6]		*GridVAE*-accelerated SLP		Speed up
		Area Increase	Time(s)	Area Increase	Time (s)	
Design1-PG1	1024	1.61%	41.8	0.98%	2.5	16.79
Design2-PG2	4096	1.55%	165.8	1.79%	7.6	21.82
Design2-PG3	4096	0.70%	149	0.86%	7.34	20.30
Design3-PG4	16384	1.19 %	1034	1.53 %	12.3	84.06
Design3-PG5	16384	2.85 %	922	2.14%	13.7	67.2

The table II compare the prediction accuracy over then GAN-based model in RMSE, the unit is mV, and results shows we can reduce up to 40% RMSE. The table III shows the comparison of our VAE-accelerated power grid IR drop fixing strategy versus the existing SLP-based method [6] . The last column indicate the proposed method has the speedup over [6] As we can see that both SLP-based methods can lead to similar performance in terms of area, and can achieve more than 80X speedup. On the other hand, we see that our VAE-accelerated SLP method shows more speedup advantages as the sizes of the power grid increase.

V. Conclusion

In conclusion, our work introduces a novel VAE-based full-chip EM-aware IR drop estimation model, referred to as *GridVAE*, tailored specifically for on-chip power grid networks. Leveraging PyTorch automatic differentiation, *GridVAE* allows rapid sensitivity computations of the target cost function concerning wire geometries, enabling accelerated power grid fixing through a recently proposed linear programming-based optimization framework. The numerical results attest to the superiority of *GridVAE* over recently proposed GAN-based models, with a notable 40% reduction in RMSE. Additionally, our GridVAE-accelerated method demonstrates an impressive up to 80X speedup (at least one order of magnitude) compared to conventional SLP-based approaches when applied to synthesized power grid benchmarks from ARM Cortex-M0 processor design. These findings underscore the potential of *GridVAE* as a promising tool for enhancing power grid optimization in modern nanometer-scale chip technologies, providing valuable insights into efficient and accurate solutions for addressing EM-related challenges in VLSI design.

References

[1] S. Chowdhury and M. A. Breuer, "Optimum Design of IC Power/Ground Nets Subject to Reliability Constraints," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 7, no. 7, pp. 787–796, 1988.

[2] R. Dutta and M. Marek-Sadowska, "Automatic Sizing of Power/Ground (P/G) Networks in VLSI," in *Proc. Design Automation Conf. (DAC)*, Jun. 1989, pp. 783–786.

[3] X.-D. Tan, C.-J. Shi, D. Lungeanu, J.-C. Lee, and L.-P. Yuan, "Reliability-constrained area optimization of VLSI power/ground networks via sequence of linear programmings," in *Proc. Design Automation Conf. (DAC)*, June 1999, pp. 78–83.

[4] K. Wang and M. Marek-Sadowska, "On-Chip Power-Supply Network Optimization Using Multigrid-Based Technique," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 24, no. 3, pp. 407–417, 2005.

[5] H. Zhou, Z. Sun, S. Sadiqbatcha, N. Chang, and S. X.-D. Tan, "EM-Aware and Lifetime-Constrained Optimization for Multisegment Power Grid Networks," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 27, no. 4, pp. 940–953, Apr. 2019.

[6] Z. Moudallal, V. Sukharev, and F. N. Najm, "Power Grid Fixing for Electromigration-induced Voltage Failures," in *Proceedings of the 38th International Conference on Computer-Aided Design*, ser. ICCAD '19, Nov. 2019, pp. 1–8.

[7] H. Zhou, S. Yu, Z. Sun, and S. X.-D. Tan, "Reliable power grid network design framework considering EM immortalities for multi-segment wires," in *Proc. Asia South Pacific Design Automation Conf. (ASPDAC)*, Jan. 2020, pp. 1–6.

[8] H. Zhou, W. Jin, and S. X.-D. Tan, "GridNet: Fast Data-Driven EM-Induced IR Drop Prediction and Localized Fixing for On-Chip Power Grid Networks," in *Proceedings of the 39th International Conference on Computer-Aided Design*, ser. ICCAD '20, Nov. 2020, pp. 1–9.

[9] H. Zhou, L. Chen, and S. X.-D. Tan, "Robust Power Grid Network Design Considering EM Aging Effects for Multi-Segment Wires," *Integration, the VLSI Journal*, vol. 77, pp. 38–47, Mar. 2021.

[10] H. Zhou, Y. Liu, W. Jin, and S. X.-D. Tan, "Gridnetopt: Fast full-chip em-aware power grid optimization accelerated by deep neural networks," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 2021, accepted.

[11] M. W. Diederik P Kingma, "Auto-Encoding Variational Bayes," *arXiv e-prints*, p. arXiv:1312.6114v11, December 2022.

[12] J. Li and S. Ghosh, "Scalable variational quantum circuits for autoencoder-based drug discovery," in *2022 Design, Automation and Test in Europe Conference and Exhibition (DATE)*, 2022, pp. 340–345.

[13] K. Zhu, M. Liu, Y. Lin, B. Xu, S. Li, X. Tang, N. Sun, and D. Z. Pan, "Geniusroute: A new analog routing paradigm using generative neural network guidance," in *2019 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2019, pp. 1–8.

[14] K. Touloupas and P. P. Sotiriadis, "Mixed-variable bayesian optimization for analog circuit sizing using variational autoencoders," in *2022 18th International Conference on Synthesis, Modeling, Analysis and Simulation Methods and Applications to Circuit Design (SMACD)*, 2022, pp. 1–4.

[15] Z. Sun, S. Yu, H. Zhou, Y. Liu, and S. X.-D. Tan, "EMSpice: Physics-Based Electromigration Check Using Coupled Electronic and Stress Simulation," *IEEE Transactions on Device and Materials Reliability*, vol. 20, no. 2, pp. 376–389, Jun. 2020.

[16] X. Huang, A. Kteyan, S. X.-D. Tan, and V. Sukharev, "Physics-Based Electromigration Models and Full-Chip Assessment for Power Grid Networks," *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, vol. 35, no. 11, pp. 1848–1861, 2016.

[17] S. Chatterjee, V. Sukharev, and F. N. Najm, "Power Grid Electromigration Checking Using Physics-Based Models," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, vol. 37, no. 7, pp. 1317–1330, Jul. 2018.

[18] V. Sukharev and F. N. Najm, "Electromigration Check: Where the Design and Reliability Methodologies Meet," *IEEE Transactions on Device and Materials Reliability*, vol. 18, no. 4, pp. 498–507, Dec. 2018.

[19] "EMspice – Coupled EM-IR Analysis Tool for Full-Chip Power Grid EM Check and Sign-off," 2020, https://github.com/sheldonucr/EMspice.

[20] S.-Y. Lin, Y.-C. Fang, Y.-C. Li, Y.-C. Liu, T.-S. Yang, S.-C. Lin, C.-M. Li, and E. J.-W. Fang, "IR Drop Prediction of ECO-Revised Circuits Using Machine Learning," in *Proceedings of the 36th VLSI Test Symposium*, ser. VTS '18. New York, NY: IEEE Press, Apr. 2018, pp. 1–6.

[21] Y.-C. Fang, H.-Y. Lin, M.-Y. Sui, C.-M. Li, and E. J.-W. Fang, "Machine-learning-based Dynamic IR Drop Prediction for ECO," in *Proceedings of the 37th International Conference on Computer-Aided Design*, ser. ICCAD '18. New York, NY: ACM Press, Nov. 2018, pp. 1–7.

[22] C.-T. Ho and A. B. Kahng, "IncPIRD: Fast Learning-Based Prediction of Incremental IR Drop," in *Proceedings of the 38th International Conference on Computer-Aided Design*, ser. ICCAD '19. New York, NY: IEEE Press, Nov. 2019, pp. 1–8.

[23] Z. Xie, H. Ren, B. Khailany, Y. Sheng, S. Santosh, J. Hu, and Y. Chen, "PowerNet: Transferable Dynamic IR Drop Estimation via Maximum Convolutional Neural Network," in *Proceedings of the 25th Asia and South Pacific Design Automation Conference*, ser. ASP-DAC '20. New York, NY: IEEE Press, Jan. 2020, pp. 13–18.

[24] V. A. Chhabria, V. Ahuja, A. Prabhu, N. Patil, P. Jain, and S. S. Sapatnekar, "Thermal and IR drop analysis using convolutional encoder-decoder networks," in *Proceedings of the 26th Asia and South Pacific Design Automation Conference*, 2021, pp. 690–696.

[25] H. Zhou, Y. Sun, Z. Sun, H. Zhao, and S. X.-D. Tan, "Electromigration-Lifetime Constrained Power Grid Optimization Considering Multi-Segment Interconnect Wires," in *Proceedings of the 23rd Asia and South Pacific Design Automation Conference*, ser. ASP-DAC '18. New York, NY: IEEE Press, Jan. 2018, pp. 399–404.

[26] R. Fu, J. Chen, S. Zeng, Y. Zhuang, and A. Sudjianto, "Time Series Simulation by Conditional Generative Adversarial Net," *arXiv e-prints*, p. arXiv:1904.11419, Apr. 2019.

[27] S. R. Nassif, "Power grid analysis benchmarks," in *Proc. Asia South Pacific Design Automation Conf. (ASPDAC)*, 2008, pp. 376–381.