

EM-GAN: Data-Driven Fast Stress Analysis for Multi-Segment Interconnects

Wentian Jin, Sheriff Sadiqbatcha, Zeyu Sun, Han Zhou, and Sheldon X.-D. Tan
 Department of Electrical and Computer Engineering, University of California, Riverside, CA 92521
 {wjn018, ssadi003, zsun007, hzhou012, stan}@ece.ucr.edu

Abstract—Electromigration (EM) analysis for complicated interconnects requires the solving of partial differential equations, which is expensive. In this paper, we propose a fast transient hydrostatic stress analysis for EM failure assessment for multi-segment interconnects using generative adversarial networks (GANs). Our work is inspired by the image synthesis and feature of generative deep neural networks. The stress evaluation of multi-segment interconnects, modeled by partial differential equations, can be viewed as time-varying 2D-images-to-image problem where the input is the multi-segment interconnects topology with current densities and the output is the EM stress distribution in those wire segments at the given aging time. We show that the conditional GAN can be exploited to attend the temporal dynamics for modeling the time-varying dynamic systems like stress evolution over time. The resulting algorithm, called *EM-GAN*, can quickly give accurate stress distribution of a general multi-segment wire tree for a given aging time, which is important for full-chip fast EM failure assessment. Our experimental results show that the *EM-GAN* shows 6.6% averaged error compared to COMSOL simulation results with orders of magnitude speedup. It also delivers $8.3\times$ speedup over state-of-the-art analytic based EM analysis solver.

I. INTRODUCTION

Electromigration (EM) is a primary long-term reliability concern for copper-based back-end-of-the-line interconnects used in modern semiconductor chips. As predicted by International Technology Roadmap for Semiconductors (ITRS), EM is projected to only get worse in future technology nodes [1]. This, as with many other reliability effects, is due to the continued trend of feature-size reduction and rapid integration which ultimately affects the critical sizes for the EM failure process. EM-related aging and reliability will become worse for current 7nm and below technologies. As a result, it is crucial to ensure the reliability of the very large scale integration (VLSI) chips during their projected lifetimes.

Due to its growing importance, considerable recent research has focused on fast EM analysis techniques. It is well accepted that existing Black and Blech-based EM models [2], [3] are overly conservative and can only work for single wire segment [4], [5]. Recently, a number of physics-based EM model and analysis techniques have been proposed [6]–[18]. At the center of those methods is to solve partial differential equation (called Korhonen's equation) of stress in the confined metal wire segments in a general interconnect tree [19]. Although many numerical approaches such as finite method [12], [13], finite element methods [6], [15] and analytic or semi-analytic solutions [9], [11], [14], [16], [17] were proposed, these methods still suffer the high computing costs or can only apply

to some special cases, which hinder this applications for full-chip EM validation and signoff analysis.

On the other hand, deep neural networks (DNN) have propelled an evolution in machine learning fields and redefined many existing applications with new human-level AI capabilities. DNNs such as convolution neural networks (CNN) have recently been applied to many cognitive applications such as visual object recognition, object detection, speech recognition, natural language understanding, and etc. due to dramatic accuracy improvements in those tasks [20]. Recently, generative adversarial networks (GAN) [21] gained much traction as it can learn features (latent representation) without extensively annotated training data. The representations learned by GANs may be used in a variety of applications, including image synthesis, semantic image editing, style transfer, image super-resolution, and classification etc.

Recently GAN-based methods have been applied for VLSI physical designs such as generation of the various noise maps to facility the IR-drop noise sensor placement [22], for layout lithography analysis [23] and sub-resolution assist feature generation [24], for analog layout well generation [25]. However, the proposed GAN-based design and analysis techniques are mainly targeted for the statistical and static image generations (analysis). Less works have been explored to learn the time-series data. GAN models actually are not only limited to static image generation. Recently GAN models have been modified to preserve the temporal dynamics as demonstrated in recent work for time-series GAN, called TimeGAN [26], in which additional auxiliary networks called embedding and recovery are added to learn the temporal information of data. Also recently study shows one can use recurrent neural networks (RNN), which is more suitable for time series modeling, for both generator and discriminator for missing value imputation for multivariate time series data [27] and time series data augmentation [28].

In this work, we propose data-driven fast transient hydrostatic stress analysis technique for EM failure assessment of multi-segment interconnects by exploiting the conditional GAN models. We show that we time variable can be treated as continuous conditions for GAN to learn the temporal dynamics in the EM stress evolution analysis. The new contributions are as follows:

1. We first show that EM analysis, modeled by partial differential equations, can be viewed as 2D image to image transforming process. Then we propose to explore the conditional GAN structure in which the input images, which are the multi-segment interconnects topology with current densities, are treated as conditions.
2. To learn the temporal dynamics in the transient EM

This work is supported in part by NSF grants under No. CCF-1816361, in part by NSF grant under No. CCF-2007135 and No. OISE-1854276.

analysis, we further explore the conditional GAN structures to use *time* variable as the continuous condition for generator and discriminator. We show such time-conditional GAN works well for time-varying stress modeling.

3. Different hyper parameters of GAN were studied and compared. We use current densities of wire segment and aging time as the conditions for the conditional GAN. The resulting *EM-GAN* can quickly give accurate stress distribution of any multi-segment wires for a given aging time.
4. Our experimental results show that the EM-GAN has 6.6% averaged error compared to COMSOL [29] simulation results with orders of magnitude speedup. It also delivers $8.3\times$ speedup over recently proposed state-of-the-art analytic based EM analysis solver [17].

This paper is organized as follows: Section II gives the basic background of EM problem and existing solutions. Section III presents the proposed GAN solver for the EM-induced stress distribution. Section IV presents the numerical results and some related discussions. Finally section V concludes this paper.

II. PRELIMINARIES FOR TRANSIENT EM ANALYSIS

EM is the process of metal atoms migrating along the direction of the applied electric field in confined metal interconnect wires due to the momentum transfer between the conducting electrons and lattice atoms. Under EM, the aforementioned momentum transfer leads to the buildup of hydrostatic stress in the confined metal wires. When this stress reaches a critical level, the aforementioned migration of atoms is initiated. Over time, this migration leaves behind a depletion of atoms (or void) at the cathode terminal of the wire and an accumulation of atoms (or hillock) at the anode terminal. This eventually leads to failure due to an open or short circuit respectively.

Traditionally, the industry standard model to predicting the time-to-failure (TTF) under EM are based on empirical or statistical models, the most well known of which are Black's equation [2] and Blech's limit [3]. However those models have been shown to be overly conservative, applicable only to single wire segment, and therefore lead to unnecessary over-design with large overheads [5]. To mitigate this problem, EM modeling starts with the first principles of stress physics in the confined metal wires start to gain many tractions [18]. Such physics-based EM modeling analysis is centering around solving the partial differential equation with blocked terminal boundary conditions for general multi-segment interconnects as shown in Fig. 1.

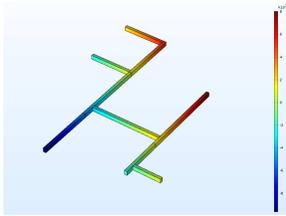


Fig. 1: Multi-segment wire with EM stress distribution. Specifically, we assume that a general interconnect wire has n nodes, including p interior junction nodes $x_r \in \{x_{r1}, x_{r2}, \dots, x_{rp}\}$ and q block terminals $x_b \in$

$\{x_{b1}, x_{b2}, \dots, x_{bq}\}$. Then the Korhonen's PDE [19] for the nucleation phase can be written in following multi-segment format:

$$\begin{aligned} \frac{\partial \sigma_{ij}(x, t)}{\partial t} &= \frac{\partial}{\partial x} \left[\kappa_{ij} \left(\frac{\partial \sigma_{ij}(x, t)}{\partial x} + G_{ij} \right) \right], t > 0; \\ BC : \sigma_{ij_1}(x_i, t) &= \sigma_{ij_2}(x_i, t), t > 0; \\ BC : \sum_{ij} w_{ij} \kappa_{ij} \left(\frac{\partial \sigma_{ij}(x, t)}{\partial x} \Big|_{x=x_r} + G_{ij} \right) \cdot n_r &= 0, t > 0 \quad (1) \\ BC : \kappa_{ij} \left(\frac{\partial \sigma_{ij}(x, t)}{\partial x} \Big|_{x=x_b} + G_{ij} \right) &= 0, t > 0; \\ IC : \sigma_{ij}(x, 0) &= \sigma_{ij,T} \end{aligned}$$

where $\sigma(x, t)$ is the hydrostatic stress for branch ij from nodes i and j , n_r represents the unit inward normal direction of the interior junction node r on branch ij , the value of which is $+1$ for right direction and -1 for left direction of branch with assumption of $x_i < x_j$, $G = \frac{Eq*}{\Omega}$ is the EM driving force, w is the width of the branch, and $\kappa = D_a B \Omega / k_B T$ is the diffusivity of stress. E is the electric field, $q*$ is the effective charge. $D_a = D_0 \exp(\frac{-E_a}{k_B T})$, which is the effective atomic diffusion coefficient. D_0 is the pre-exponential factor, B is the effective bulk elasticity modulus, Ω is the atomic lattice volume, k_B is the Boltzmann's constant, T is the absolute temperature, E_a is the EM activation energy. σ_T is the initial thermal-induced residual stress in each wire segment.

In general, numerical approaches such as finite difference, finite element based approaches are required to solve the PDE in (1), which are expensive and time consuming. The recently proposed semi-analytic solutions can still be expensive as the eigenvalues have to be computed by numerical approaches [14], [17].

III. THE PROPOSED DATA-DRIVEN FAST EM SOLVER: EM-GAN

A. Data preparation

For machine learning based approaches, one crucial aspect is sufficient training data. For our GAN-based EM stress estimation, the input data are interconnect topologies with various current densities in different wire segments while the output is the evolution of the EM-induced stress distribution across all wire segments. The proposed *EM-GAN* is trained to model the transformation scheme between these two datasets. In what follows, we present the details of the training data and how we preprocess and map them into the domain that can be leveraged by GAN-based model.

To achieve the abundance of training data, we randomly generated 2500 different topologies of multi-segment interconnects. Each of them has different number of wire segments with random widths, lengths and current densities. These generated topologies with current densities are then fed into COMSOL (an off-the-shelf finite element method (FEM) solver). For each input, COMSOL produces a series of stress distributions which reflects the stress evolution along the aging time. The data acquisition process is illustrated in Fig. 2a. The time-step between two adjacent results can be adjusted to get the best trade-off between accuracy and performance.

These data are all saved in numerical format, and as stated in Section I, to leverage the GAN model, we have to transform

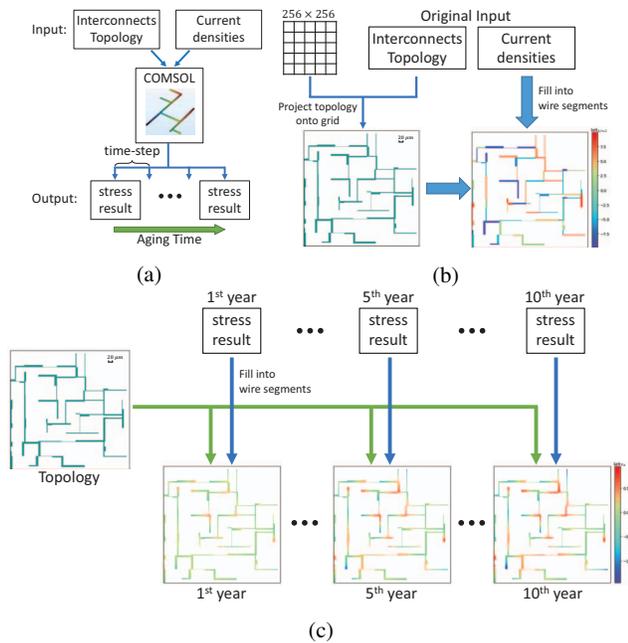


Fig. 2: Illustration of training data: (a) Raw data acquisition for training dataset (b) Input: A *design* with wire segments filled with current densities (c) Output: Evolution of EM-induced stress distribution along 10 aging years.

them into image domain so that the problem is simplified as an image-to-image task.

- **Input:** Every interconnects topology is composed of rectangular wire segments with random sizes. When generating the topologies, we set the bound for both x- and y-dimension to $256 \mu\text{m}$, and the resolution of the wire segments is set to $1 \mu\text{m}$. With such configuration, we can easily project the topology onto a 256×256 grid, which can also be seen as a single-channel image as is shown in Fig. 2b. We note that such configuration does not restrict our work to only small-size interconnects, as in real cases, bulk interconnect system may be divided into small pieces with partitioning algorithms for parallel calculation. The proposed *EM-GAN* is used as the solver for small partitions and the results can be synthesized back to form the final results for the original bulk interconnect system. Another input is the current density which is generated by applying random current sources to the interconnects. In each topology, current density varies drastically among different wire segments but is equally and uniformly distributed within the same segment. To combine these two inputs into a single image, we fill every wire segment with its current density and the resulting single-channel image is shown in Fig. 2b. In this work, we refer to every combined input of topology and current density as a *design*.
- **Output:** The results we get from COMSOL for each *design* are a time-series of gradually changing stress distributions. In this work, the maximum aging time is set to 10th year and we reserve 10 results from 1st to 10th year for training purpose. Similar with the raw data

of current densities, the stress distributions are also saved in numerical format, such that they can also be combined with topologies. The combined result is referred to as a *stress map* in this work. The combination process together with the resulting *stress maps* at 1st, 5th and 10th aging years are illustrated in Fig. 2c. Each *stress map* can also be seen as a single-channel image with the same 256×256 size as the input *design*. The difference is that each pixel in *stress maps* represents stress value in the corresponding $1 \mu\text{m}^2$ area and for each input *design*, there are 10 resulting *stress maps*.

Feeding all 2500 randomly generated *designs* into COMSOL results in 25000 stress distributions, which are then organized into a training dataset with 2500 pairs of (**Input** : 1 *design*, **Output** : 10 *stress maps*) samples. The 1-to-10 relationship within each data pair implies that a single input multiple output (SIMO) model is required while traditional GANs are only capable of single input single output (SISO) modeling. The technique we use to overcome this barrier will be detailed in Section III-B3.

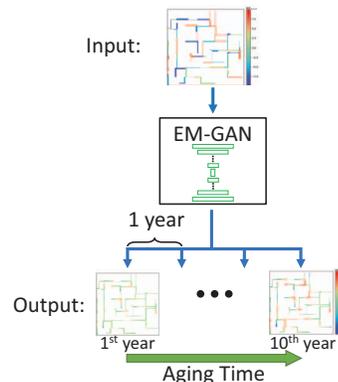


Fig. 3: EM-GAN models the stress estimation as an image-to-image task

Now that both input and output are transformed into image domain, a GAN-based model can be leveraged to solve the proposed problem as an image-to-image task as illustrated in Fig. 3. However, there are still some preprocessing needed before the data can really be fed into the model. Since there is only one channel in the image, the figures shown in Fig. 2 are depicted as heat-maps in which the colors are only for visualization purpose. Pixels in a typical color image usually have red-green-blue channels and the values are limited to the range of 0 to 255, which is not the case in our dataset. Pixels in *design* and *stress map* are filled with real values of current density and stress separately. In this work, both current density and stress can range drastically from magnitude of -10^9 to 10^9 . The positive sign here denotes the direction toward right and up, and vice versa. It is commonly accepted that values around zero are more numerically stable for neural networks, and thus, we have to scale our dataset down to such range. In this work, we rescale all samples to zero mean and unit standard deviation using data standardization method. All values are squeezed into the range of -7 to 7 with the majority of which around zero.

B. CGAN-based current density to EM stress transformation

1) *The current density image to EM stress image transformation:* We first show that we can view the PDE solving

process for a multi-segment wire shown in Fig. 1 as image synthesis process, in which the DNN can automatically extract features reflecting the physics-law of stress evolution in the confined metal wire. Then we can use the DNN network to map the input images of interconnect wires with stressing current to the stress distributions of wire segment for any given aging time.

2) *Review of GANs*: Generative Adversarial Networks (GANs), are widely used generative models which consist of two neural networks: (1) A Generator G which is trained to produce real-like data which mimics the samples in the training set and, (2) a discriminator D that takes either real or fake data as input and aims to discriminate between them. The input of G is usually random noise z which follows a certain distribution and thus, the generated output is also a random sample extracted from the distribution of fake data. The training of GAN requires both G and D to be trained simultaneously in an alternative fashion and the final goal is to let the distribution of fake data overlaps with that of the training set. The output of D measures the similarity between these two distributions and usually the Jensen-Shannon Divergence is employed as the measurement. To reduce the randomness in the generated data, Conditional GAN (CGAN) was created to provide a certain extent of control on the output of G . CGAN is a variant of GAN which introduces additional condition input so that the fake data distribution is conditioned on it. CGANs have been widely used as conditional generation method and are at the forefront of a wide-range of applications.

3) *Time Dependent Architecture*: GANs are designed for static applications where single input always lead to a single output. However, as is shown in Section II, our dataset consists of 2500 pairs of (1 *design* \rightarrow 10 *stress maps*) samples, which requires the model to be able to generate a sequence of stress distributions across all the aging years using only one *design* as input. To overcome the barrier between traditional GAN and the time-dependent data, we propose the *EM-GAN*, which is a CGAN-based model with the capability of time-variant output synthesis.

There are some recent studies trying to preserve the temporal dynamics through modifications of GAN architecture. In TimeGAN [26], additional auxiliary networks called embedding and recovery are added to learn the temporal information of data. Other researches employ recurrent neural network (RNN), which is a natural architecture for time series modeling, in both generator and discriminator for time series data augmentation [28] and missing value imputation for multivariate time series data [27]. These existing works mostly deal with simulated or size-limited synthetic data, in which employing RNNs will not cause too much overhead. However, in this work, we are dealing with a time-dependent image synthesis problem where both input and output are of quite large sizes (256×256 pixels). Such large data throughput results in a heavy model and integrating it recursively in a RNN-like architecture will lead to a bulky network that can be expressed as

$$\begin{aligned} p(z, 0) &\xrightarrow{G} p(\hat{y}_1 | z) \\ p(z, \hat{y}_{1:t-1}) &\xrightarrow{G} p(\hat{y}_t | z, \hat{y}_{1:t-1}) \end{aligned} \quad (2)$$

where z is a random *design*, G the generator model, and \hat{y}_t the estimated stress distribution produced at the t^{th} time-step which is conditioned on both *design* and history results.

This is not a practical architecture due to the significant computational overhead it would introduce in both training and inference. Additionally, considering the fact that EM-induced stress continuously evolves over 10 years, such a large time range further impedes the employment of RNN which otherwise would produce numerous intermediate results at each time-step before the final aging year is reached. In real cases, designers only care about whether the interconnects is able to last before the chip lifetime is reached, which implies that getting only the stress result at the specified aging year is enough. The intermediate results are only needed when a wire failure is spotted and further investigation into the stress evolution is required.

Basing on these observations, we propose the *EM-GAN* illustrated in Fig. 4 which employs a CGAN as the backbone. The *design* z is taken as one input and another input is the explicitly specified aging year t which serves as the time condition. Compared with the sequential network in (2), *EM-GAN* is simplified to directly map the *design* to the *stress map* at the conditioned aging year with no intermediate result generated. If Additionally stress-induced failure is found in the *stress map*, a backward investigation can be conducted by changing the input aging year to previous time-steps, such that the detailed evolution of the *stress map* can be gathered and analyzed.

With such a time-conditioned architecture, a single *design* can be projected onto multiple *stress maps* by varying the time condition input. The proposed *EM-GAN* model can be expressed as

$$p(z, t) \xrightarrow{G} p(\hat{y} | z, t) \quad (3)$$

where z is a random *design*, t the specified aging year, G the generator network, and \hat{y} the *stress map* estimated by the generator conditioned on time t .

4) *EM-GAN Architecture*: As is shown in in Fig. 4, the generator G of *EM-GAN* takes the *design* image $img_{des} \in \mathbb{R}^{256 \times 256 \times 1}$ and the aging year $t \in \mathbb{R}$ as input. The scalar value t is expanded into $\mathbb{R}^{256 \times 256 \times 1}$ through channel-wise duplication, such that img_{des} and t can be concatenated element-wise into a two-channel image x with the size of $256 \times 256 \times 2$. It is then taken as the real input of the generator G . The architecture we employed for G is an encoder-decoder network which is widely used in image-to-image applications. In such a network, the input x is first downsampled through a series of convolutional layers until a bottleneck layer in which the extracted latent features are saved. These features may contain various abstract information such as physics-law and temporal dependency. The rest part of G leverages the extracted features and generates the *stress map* by upsampling them through transposed convolutional layers. A drawback of this encoder-decoder network is that all information passes through the narrow bottleneck layer in the middle which is not necessary. In this work, both input *design* and output *stress map* share the same topology of interconnects, and the extraction and reconstruction of such geometric information leads to excessive overhead in both computation and bandwidth. To make the model focus solely on the processing of temporal and physical features, we add skip connections between the encoder and the decoder as is shown in Fig. 4. With such configuration, The bottleneck layer is bypassed and the geometric information is passed

through the shortcuts directly from encoder to decoder. Skip connections can greatly improve the output accuracy which will be discussed in detail in Section IV-C.

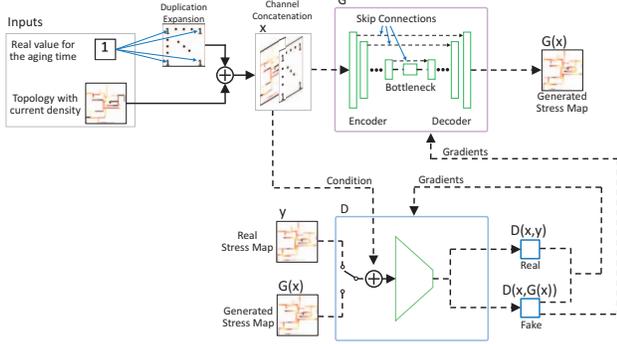


Fig. 4: EM-GAN framework for stress estimation

The output *stress map* of the generator is denoted as $G(\mathbf{x})$ and is referred to as fake *stress map* in this work. In the training process, either a fake $G(\mathbf{x})$ or a real *stress map* y from the training set is fed into the discriminator D together with its corresponding *design* and aging time x . The discriminator will then judge whether the *stress map* is real according to the given x . The output of the discriminator is a scalar score which is denoted as $D(G(\mathbf{x}), \mathbf{x})$ or $D(y, \mathbf{x})$ depending on which *stress map*, fake or real, was inputted. It reflects how confident the discriminator is that the *stress map* its being fed is a real one.

The key idea of the *EM-GAN* model is to let the generator learn the mapping method from the distribution of *designs* with aging year to that of the real *stress maps*. Such transformation is achieved by progressively training the generator according to the gradients back propagated from the loss which is based on the output of the discriminator. The generator and the discriminator are trained simultaneously but with opposite training objectives. The training goal of the discriminator is to minimize $D(G(\mathbf{x}), \mathbf{x})$ and maximize $D(y, \mathbf{x})$, which can be expressed as

$$\max_D \{ \mathbb{E}_{\mathbf{x}, y} [D(y, \mathbf{x})] - \mathbb{E}_{\mathbf{x}} [D(G(\mathbf{x}), \mathbf{x})] - \lambda_{gp} \mathbb{E}_{\hat{\mathbf{x}}} [(\|\nabla_{\hat{\mathbf{x}}} D(\hat{\mathbf{x}}, \mathbf{x})\|_2 - 1)^2] \} \quad (4)$$

where $\mathbb{E}_{\mathbf{x}, y} [D(y, \mathbf{x})]$ is the average score given by the discriminator to real *stress maps*, while $\mathbb{E}_{\mathbf{x}} [D(G(\mathbf{x}), \mathbf{x})]$ is the average score given to the fake ones. These two terms together confine the discriminator to be more confident in telling apart the real input from the fake ones. The last term in (4) is the gradient penalty adopted from WGAN-GP [30], which maintains the 1-Lipschitz continuity of the discriminator. $\hat{\mathbf{x}}$ is interpolation between the generated EM stress image and its ground truth, and λ_{gp} is the hyperparameter which controls the weight of gradient penalty.

On the contrary, the training objective of the generator is to produce real-like *stress maps* so that the discriminator is deceived to give high scores to the fake inputs. Since the generator has no influence on the scores given to the real samples, term $D(y, \mathbf{x})$ is discarded in its objective function which can be shown as

$$\min_G \{ \mathbb{E}_{\mathbf{x}} [-D(G(\mathbf{x}), \mathbf{x})] + \lambda_{L2} \cdot \mathbb{E}_{\mathbf{x}, y} [\|y - G(\mathbf{x})\|_2] \} \quad (5)$$

where only term $\mathbb{E}_{\mathbf{x}} [D(G(\mathbf{x}), \mathbf{x})]$ is reserved. We also add the average L2-norm $\mathbb{E}_{\mathbf{x}, y} [\|y - G(\mathbf{x})\|_2]$ here to further improve the objective function according to [31] in which λ_{L2} controls its weight. Introducing L2-norm into the loss function Skip connections improves the quality of generated *stress maps* which will also be discussed in detail in Section IV-C.

In both (4) and (5), we adopted the Wasserstein distance as the measure of difference between distributions of real and fake *stress maps*. Compared to the conventional Jensen–Shannon Divergence, Wasserstein distance provides higher convergence possibility and stability in the training process. The detailed architectures of both generator and discriminator in the proposed *EM-GAN* are illustrated in Fig. 5.

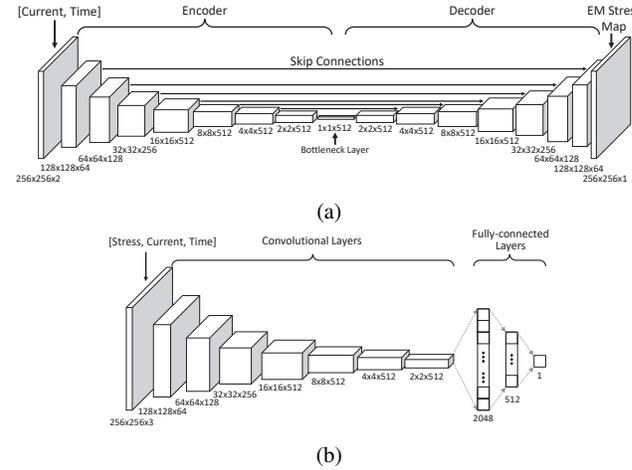


Fig. 5: The architecture of the neural networks in the proposed *EM-GAN*: (a) generator (b) discriminator.

IV. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, we present the experimental results showing both the accuracy and speed of our proposed *EM-GAN* model for time dependent EM stress estimation.

Our model is implemented in Python basing on TensorFlow(1.14.0) library [32] which is an open-source machine learning platform. As is detailed in Section IV-C, we have a dataset containing 2500 pairs of (1 *design* \rightarrow 10 *stress maps*), and to train *EM-GAN*, a random selection of 15% is set aside for testing purpose and the remaining 85% forms the training set. To accommodate the dataset to the input layer of *EM-GAN*, each 1-to-10 data pair is reorganized into 10 samples of (*design* with aging year \rightarrow *stress map*). During the training process, all samples are randomly permuted at the beginning of every epoch.

We run the training for 15 epochs on a Linux server with 2 Xeon E5-2698v2 2.3GHz processors and Nvidia Titan X GPU. The cudnn library is used to accelerate the training process on GPU. To employ mini-batch stochastic gradient descent(SGD), we set the batch size to 8 and solve it with the RMSProp optimizer. The learning rate of the optimizer is 0.0001, where the decay, momentum and epsilon parameters are set to 0.9, 0 and 10^{-10} respectively. The weight of the L2-norm distance λ_{L2} is set to 100.

A. Accuracy of EM Stress Map Estimation

Once the *EM-GAN* model is trained, only the generator part is preserved which serves as the generative model. It can

take any multi-segment interconnects topology with current densities as input and produce an estimated *stress map* at the specified aging year. To evaluate the estimation error of the trained model, we compare the estimated *stress maps* against the real ones which serve as ground truth here. We employ the root-mean-square error (RMSE) and the normalized RMSE (NRMSE) given in (6) and (7) as the metrics of error.

$$RMSE = \sqrt{\frac{\sum_{(x,y) \in S} [\sigma(x,y) - \sigma'(x,y)]^2}{|S|}} \quad (6)$$

$$NRMSE = \frac{RMSE}{\sigma_{max} - \sigma_{min}} \quad (7)$$

where σ and σ' are the real and generated *stress map* respectively. S is the set containing all pixels that are on the interconnects and $|S|$ denotes the number of elements in S . σ_{max} and σ_{min} are the maximum and minimum stress values in the ground truth *stress map* respectively.

The accuracy evaluation is conducted on the test set with 375 *designs* that were set aside during the training process. The random generation of *designs* in both training and test set guarantees that there is no overlap of either topology or current densities between these two datasets. It means that all samples used for evaluation are unseen and completely new to the trained *EM-GAN*, which makes the testing more close to real applications. When *EM-GAN* is employed to estimate stress distribution for a real *design*, it is merely possible that the given topology or current density is identical to any *design* that is from the training set. The model has to extrapolate what it learned on the training set to unseen cases, which is exactly what we are testing in this evaluation experiment.

For each of the 375 *designs* used for testing, it is fed into the generator of *EM-GAN* together with 10 scalars representing 1st to 10th aging years, and the results of which are 10 *stress maps* showing the evolution of EM-induced stress distribution. Comparing all 3750 generated *stress maps* against their corresponding ground truth (real *stress maps* derived from COMSOL), *EM-GAN* achieves an average RMSE of 0.13 GPa and NRMSE of 6.6%. Considering the large numerical range (usually several GPa and 4 GPa in this work) that typical EM stresses vary in, such accuracy is beyond enough for EM failure assessment such as critical wire identification. Some testing results are visualized in Fig. 8. Two *designs* are randomly picked from the test set and the estimated *stress maps* at 1st, 4th, 7th and 10th aging years are shown together with the ground truth as comparison.

B. Speed of Inference

In what follows, we provide a comparison of speed between our *EM-GAN* and the state-of-the-art work [17] on EM stress analysis. We set up the problem as a large multi-segment interconnects design that can be divided into 528 smaller *designs* with dimensions of $256 \times 256 \mu\text{m}^2$. We randomly generated them using same method as we did when generating the training dataset. The number of interconnect branches in each design ranges from 5 to 79. Both *EM-GAN* and the [17] method were run to estimate the EM stress distribution for all 528 *designs* at the 10th aging year. The tests were run on the same server and the accumulated time cost on all 528 *designs* are plotted in Fig. 6.

Although [17] yields more accurate results which agree well with the solution of COMSOL, *EM-GAN* demonstrates

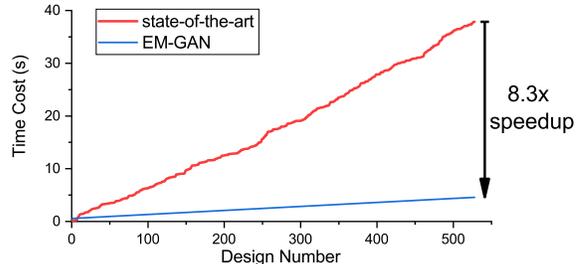


Fig. 6: Comparison of EM stress estimation speed between state-of-the-art and *EM-GAN*.

8.3 \times speedup over [17]. The total time cost of *EM-GAN* and [17] are 37.86s and 4.58s respectively. For [17], the time cost of each *stress map* prediction varies from 0.49s to 0.003s depending on the number of branches involved in the input *design*. However, for our *EM-GAN*, any given *design* is taken as a whole image with same dimensions. The inference is essentially an image transformation process dealing with fixed number of pixels regardless of how many wire segments are actually involved in the input *design*. That is to say, the inference time of *EM-GAN* is invariant to the varying number of interconnect branches, which makes it much more competent in doing estimation for large scale designs and has a better scalability.

C. Analysis of Loss and Skip Connections

As described in Section III-B4, *EM-GAN* employs skip connections in the generator to bypass the bottleneck layer and conveys the geometric information directly from the encoder to the decoder. Another technique we used to improve the estimation accuracy is adding L2-norm error in the loss function of generator. To analyze whether and how these modifications helps to improve the results, we trained two modified *EM-GAN* models. We controlled the most part of the architecture in both modified models remain the same as *EM-GAN*. The only exception is that one model removed the L2-norm from the objective function and the other one discarded all skip connections.

Both modified models are trained for 15 epochs on the same server and are tested using the same test set as above. The result turns out that both modified models suffered degradation in the accuracy of output. Specifically, model without L2-norm loss reached an average NRMSE of 8.4% and the error is even worse at 15.2% for the other model with no skip connection. Also, compared to modified models, *EM-GAN* performs better in terms of standard deviation, maximum and minimum errors as is shown in Table I. In Fig. 7, we randomly pick one *design* and show the inference results generated by all three models along with their corresponding ground truth for comparison.

We first analyze the influence of skip connections. As shown in the results above, models with skip connections outperform the one without it by a significant margin. Employing a conventional encoder-decoder architecture means that the network has to process both geometric and physics information from the input. This is completely unnecessary in this work since the input *design* and the output *stress map* share exactly the same geometric information, i.e. the interconnects topology. The extra work added to the network occupies both computational and spacial resources that could have been used for

the extraction of more meaningful latent features. The skip connections mitigate this problem by introducing shortcuts for the topology information to be directly passed from the input side to the output side. It alleviates the workload of the main network and spares more bandwidth for latent information flow, which then helps increase the output accuracy.

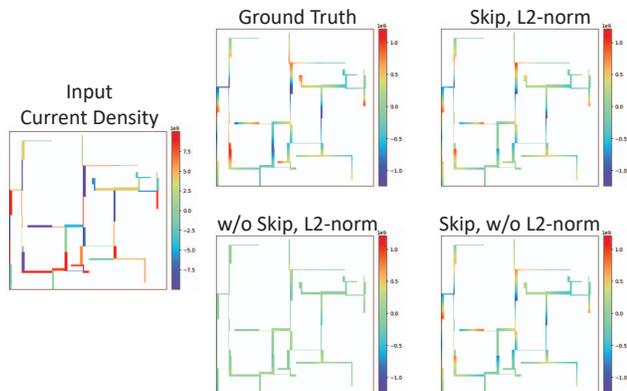


Fig. 7: Comparison of inference results between different models and the ground truth.

TABLE I: Statistics of NRMSE for EM-GAN and modified models on testing set.

Metrics	EM-GAN (Skip, L2-norm)	w/o Skip, L2-norm	Skip, w/o L2-norm
Mean	6.6%	15.2%	8.4%
Standard Deviation	1.2%	2.1%	2.1%
Max	12.9%	24.6%	18.4%
Min	3.1%	9.8%	3.8%

The influence of L2-norm loss on the result accuracy is not as large as the skip connection, but still, the removal of it degenerates the NRMSE from 6.6% to 8.4%. As is shown in Fig. 7, the *stress map* generated by *EM-GAN* is slightly closer to the ground truth than the model without L2-norm. Aside from the improvement in result accuracy, a more significant impact the L2-norm brings to *EM-GAN* is actually the speedup of training process. The modified model without L2-norm converges much slower than *EM-GAN*. This is a reasonable phenomenon since L2-norm is manually added to the objective function as a prior knowledge from human. It helps to guide the training process towards a partially defined target especially at the early stages of the training process.

The loss function is two folds, one is dynamically determined by the other part of the model itself, i.e. the discriminator, and the second one is a predefined goal, i.e. the L2-norm distance. At the very early stages of the training process, when both discriminator and generator are not well trained yet, using the loss defined by the discriminator to guide the training is more like a random walk.

the model with L2-norm has a much faster converging speed in the training process and is always closer to the ground truth than the one without L2-norm. It is a reasonable result that the

L2-norm helps the model as a prior knowledge. At the very beginning of training process, both discriminator and generator are not well trained and the discriminator is not able to provide useful guidance to the generator. This is where L2-norm can complement the discriminator and provide the generator with a meaningful learning direction. In our experiment, adding the L2-norm accelerates the convergence speed by $2\times$ and also leads to a better inference accuracy.

V. CONCLUSION

In this paper, we have proposed a GAN-based fast transient hydrostatic stress analysis for EM failure assessment for multi-segment interconnects. In our approach, we treat this traditional numerical PDE solving problem as time-varying 2D-image-to-image problem where the input is the multi-segment interconnects topology with current densities and the output is the EM stress distribution in those wire segments at the given aging time. We randomly generated the training set and trained the model with the COMSOL simulation results. Different hyperparameters of GAN were studied and compared. After the training process, the proposed EM-GAN model is tested against 375 unseen multi-segment interconnects designs and achieved high accuracy with an average error of 6.6%. It also showed $8.3\times$ speedup over recently proposed state of the art analytic based EM analysis solver.

REFERENCES

- [1] International technology roadmap for semiconductors (ITRS), 2015. <http://www.itrs2.net/itrs-reports.html>.
- [2] J. R. Black. Electromigration-A Brief Survey and Some Recent Results. *IEEE Trans. on Electron Devices*, 16(4):338–347, 1969.
- [3] I. A. Blech. Electromigration in thin aluminum films on titanium nitride. *Journal of Applied Physics*, 47(4):1203–1208, 1976.
- [4] M Hauschildt, C Hennesthal, G Talut, O Aubel, M Gall, K B Yeap, and E Zschech. Electromigration Early Failure Void Nucleation and Growth Phenomena in Cu And Cu(Mn) Interconnects. In *IEEE Int. Reliability Physics Symposium (IRPS)*, pages 2C.1.1–2C.1.6, 2013.
- [5] V. Sukharev. Beyond Black’s Equation: Full-Chip EM/SM Assessment in 3D IC Stack. *Microelectronic Engineering*, 120:99–105, 2014.
- [6] RL De Orio, Hajdin Ceric, and Siegfried Selberherr. Physically based models of electromigration: From black’s equation to modern tcad models. *Microelectronics Reliability*, 50(6):775–789, 2010.
- [7] 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*, 35(11):1848–1861, Nov. 2016.
- [8] Valeriy Sukharev, Armen Kteyan, and Xin Huang. Postvoiding stress evolution in confined metal lines. *IEEE Transactions on Device and Materials Reliability*, 16(1):50–60, 2016.
- [9] H. Chen, S. X.-D. Tan, X. Huang, T. Kim, and V. Sukharev. Analytical modeling and characterization of electromigration effects for multibranch interconnect trees. *IEEE Trans. on Computer-Aided Design of Integrated Circuits and Systems*, 35(11):1811–1824, 2016.
- [10] V. Mishra and S. S. Sapatnekar. Predicting Electromigration Mortality Under Temperature and Product Lifetime Specifications. In *Proc. Design Automation Conf. (DAC)*, pages 1–6, Jun. 2016.
- [11] H.-B. Chen, S. X.-D. Tan, J. Peng, T. Kim, and J. Chen. Analytical modeling of electromigration failure for vlsi interconnect tree considering temperature and segment length effects. *IEEE Transaction on Device and Materials Reliability (T-DMR)*, 17(4):653–666, 2017.
- [12] 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*, 37(7):1317–1330, July 2018.
- [13] C. Cook, Z. Sun, E. Demircan, M. D. Shroff, and S. X.-D. Tan. Fast electromigration stress evolution analysis for interconnect trees using krylov subspace method. *IEEE Trans. on Very Large Scale Integration (VLSI) Systems*, 26(5):969–980, May 2018.
- [14] S. Wang, Z. Sun, Y. Cheng, S. X.-D. Tan, and M. Tahoori. Leveraging recovery effect to reduce electromigration degradation in power/ground TSV. In *Proc. Int. Conf. on Computer Aided Design (ICCAD)*, pages 811–818. IEEE, Nov. 2017.

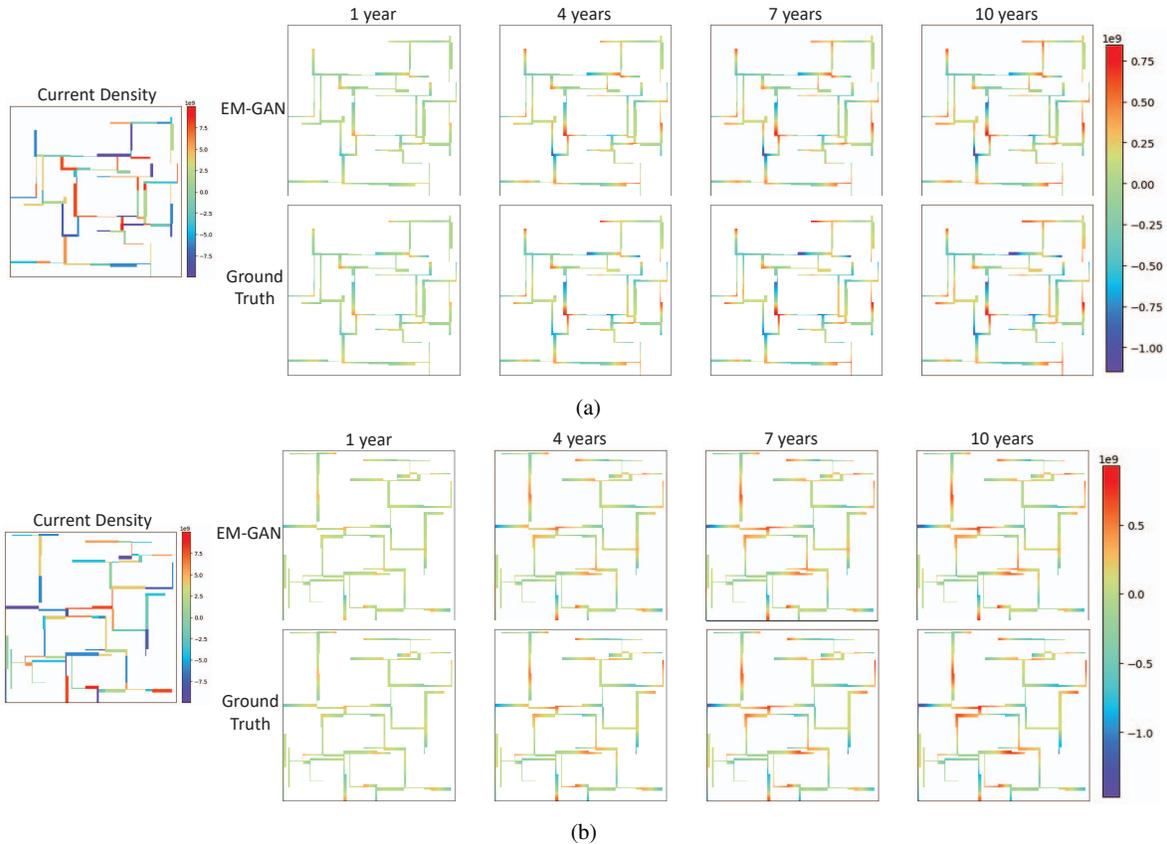


Fig. 8: Comparing the ground truth EM stress distribution with EM-GAN generated ones using two different designs.

- [15] H. Zhao and S. X.-D. Tan. Postvoiding fem analysis for electromigration failure characterization. *IEEE Trans. on Very Large Scale Integration (VLSI) Systems*, 26(11):2483–2493, Nov. 2018.
- [16] Ali Abbasinasab and Malgorzata Marek-Sadowska. RAIN: A tool for reliability assessment of interconnect networks—physics to software. In *Proc. Design Automation Conf. (DAC)*, pages 133:1–133:6, New York, NY, USA, 2018. ACM.
- [17] L. Chen, S. X.-D. Tan, Z. Sun, S. Peng, M. Tang, and J. Mao. Fast analytic electromigration analysis for general multisegment interconnect wires. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, pages 1–12, 2019.
- [18] Sheldon X.-D. Tan, Mehdi Tahoori, Taeyoung Kim, Shengcheng Wang, Zeyu Sun, and Saman Kiamehr. *VLSI Systems Long-Term Reliability – Modeling, Simulation and Optimization*. Springer Publishing, 2019.
- [19] M. A. Korhonen, P. Bojrgesen, K. N. Tu, and C.-Y. Li. Stress evolution due to electromigration in confined metal lines. *Journal of Applied Physics*, 73(8):3790–3799, 1993.
- [20] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *Nature*, 521(7553):436–444, May 2015.
- [21] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.
- [22] Jinglan Liu, Yukun Ding, Jianlei Yang, Ulf Schlichtmann, and Yiyu Shi. Generative adversarial network based scalable on-chip noise sensor placement. In *2017 30th IEEE International System-on-Chip Conference (SOCC)*, pages 239–242. IEEE, 2017.
- [23] Wei Ye, Mohamed Baker Alawieh, Yibo Lin, and David Z. Pan. Lithogan: End-to-end lithography modeling with generative adversarial networks. In *Proceedings of the 56th Annual Design Automation Conference 2019, DAC '19*, pages 107:1–107:6, New York, NY, USA, 2019. ACM.
- [24] Mohamed Baker Alawieh, Yibo Lin, Zaiwei Zhang, Meng Li, Qixing Huang, and David Z. Pan. Gan-sraf: Sub-resolution assist feature generation using conditional generative adversarial networks. In *Proceedings of the 56th Annual Design Automation Conference 2019, DAC '19*, pages 149:1–149:6, New York, NY, USA, 2019. ACM.
- [25] Biying Xu, Yibo Lin, Xiyuan Tang, Shaolan Li, Linxiao Shen, Nan Sun, and David Z. Pan. Wellgan: Generative-adversarial-network-guided well generation for analog/mixed-signal circuit layout. In *Proceedings of the 56th Annual Design Automation Conference 2019, DAC '19*, pages 66:1–66:6, New York, NY, USA, 2019. ACM.
- [26] Jinsung Yoon, Daniel Jarrett, and Mihaela van der Schaar. Time-series generative adversarial networks. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems 32*, pages 5508–5518. Curran Associates, Inc., 2019.
- [27] Yonghong Luo, Xiangrui Cai, Ying ZHANG, Jun Xu, and Yuan xiaojie. Multivariate time series imputation with generative adversarial networks. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 1596–1607. Curran Associates, Inc., 2018.
- [28] Giorgia Ramponi, Pavlos Protopoulos, Marco Brambilla, and Ryan Janssen. T-CGAN: Conditional Generative Adversarial Network for Data Augmentation in Noisy Time Series with Irregular Sampling. *arXiv e-prints*, page arXiv:1811.08295, November 2018.
- [29] Comsol multiphysics. <https://www.comsol.com/> [Oct. 16, 2013].
- [30] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein GAN. *arXiv e-prints*, page arXiv:1701.07875, Jan 2017.
- [31] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.
- [32] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. Tensorflow: A system for large-scale machine learning. In *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, pages 265–283, Savannah, GA, November 2016. USENIX Association.