IR-Fusion: A Fusion Framework for Static IR Drop Analysis Combining Numerical Solution and Machine Learning

Feng Guo[1], Jianwang Zhai[1,†], Jingyu Jia[1], Jiawei Liu[1], Kang Zhao[1], Bei Yu[2] Chuan Shi[1,†]
[1]Beijing University of Posts and Telecommunications [2]The Chinese University of Hong Kong

***Abstract*—IR Drop analysis for on-chip power grids (PGs) is vital but computationally challenging due to the rapid growth in the integrated circuit (IC) scale. Traditional numerical methods employed by current EDA software are accurate but extremely time-consuming. To achieve rapid analysis of IR drop, various machine learning (ML) methods have been introduced to address the inefficiency of numerical methods. However, the issue of interpretability or scalability has been limiting practical applications. In this work, we propose IR-Fusion, which aims to combine numerical methods with ML to achieve the trade-off and complementarity between accuracy and efficiency in static IR drop analysis. Specifically, the numerical method is used to obtain rough solutions and ML models are utilized to improve accuracy further. In our framework, an efficient numerical solver, AMG-PCG, is applied to get rough numerical solutions. Then, based on the numerical solution, the fusion of hierarchical numerical-structural information representing the multilayer structure of the PG is employed, and an Inception Attention U-Net model is designed to capture details and interaction of features at different scales. To cope with the limitations and diversity of PG designs, an augmented curriculum learning strategy is applied to the training phase. Evaluation of IR-Fusion shows that its accuracy is significantly better than previous ML-based methods while requiring considerably less iteration on solver to achieve the same accuracy compared with numerical methods.**

I. Introduction

The on-chip power grid (PG) transfers voltage and current to each working cell, and IR drop analysis involves obtaining the voltage drop caused by parasitics between the power pads and cells [1]. Ensuring the worst-case IR drop values are within specified limits is essential, which prevents performance degradation, increased power consumption, and circuit instability [2], [3]. As chip integration density increases, traditional analysis methods face significant computational challenges due to the resolution of a high-dimensional linear equation [4]. IR drop analysis becomes very time-consuming in industrial-scale designs, often requiring hours or days. Therefore, rapid and accurate early IR drop analysis is needed to optimize the design cycle.

Many numerical methods have been proposed for this process, including direct solvers, iterative solvers, and other specialized solvers. Direct solvers such as KLU [5] and Cholmod [6] are usually employed for transient simulation with a constant time step. Iterative solvers are usually developed for static analysis utilizing the Krylov subspace method [7], [8], random walk [9], and others. Specialized methods are also proposed including spatial locality methods [10], hierarchical and macromodeling methods [11], domain decomposition [12], and some matrix algebraic methods [13]. To consider the similarity between the power grids and the discretization of Laplacian equations, while further handling irregular power grids, algebraic multigrid (AMG)-based techniques and solvers [14] are also developed. Particularly, PowerRush [15] is an efficient IR drop simulator utilizing the linear solver, named algebraic multigrid preconditioned conjugate gradient (AMG-PCG). Recently, PowerRChol [16], based on the RChol algorithm, achieves faster PCG iteration convergence in PG analysis. These numerical methods provide golden results and are applied to contemporary EDA tools. However, for early industrial-scale designs, numerical methods remain time-consuming and memory-consuming, limiting their benefits.

To address inefficiencies, machine learning (ML)-based methods have been proposed as a promising alternative for accelerating IR drop analysis. IREDGe [17] employs a U-Net architecture within its EDGe network to transform power and current images into a static IR drop image. MAVIREC [18] uses a 3D U-Net model for IR drop prediction emphasizing dynamic IR drop problems and is also applicable to static IR drop analysis. IRPnet [19] utilizes a pyramid model to capture global features and introduces a loss function with Kirchhoff's law constraints to align predictions with circuit characteristics. Recent approaches like PGAU [20] and MAUnet [21] incorporate attention mechanisms to enhance IR drop prediction by focusing on hotspot regions. Although these models improve the accuracy of hotspot identification, they still face the problem of insufficiently fine modeling granularity. Additionally, these methods struggle with issues related to model interpretability and generalizability, which can limit their adoption in practical design environments.

Numerical and ML methods have their strengths and weaknesses respectively. The question arises: can they be combined for a better trade-off in speed, accuracy, and scalability? The answer is yes. Most numerical methods solve large-scale linear systems iteratively, where more iterations yield greater accuracy but require longer runtime. By integrating ML, we can perform fewer iterations to obtain a rough solution and refine it using ML. This fusion enables a better understanding of complex physical or geometric systems, while offering more fine-grained and efficient modeling.

In this work, we propose a fusion framework, IR-Fusion, to achieve better performance in static IR drop analysis. First, the AMG-PCG solver is used to solve the system matrix rapidly, performing a small number of iterations to obtain rough numerical solutions. Furthermore, we fuse hierarchi-

†Corresponding authors.

cal numerical-structural information based on the multi-layer structure of PGs. Indeed, each layer of features needs to interact, and features at different scales within the same layer capture distinct information. Therefore, a fixed-scale convolution kernel limits scale flexibility. To address this, we incorporate the attention gate, the Inception module, and the convolutional block attention module (CBAM) within the U-Net architecture to highlight details across multiple scales and capture multi-layer connections enhancing the model. Moreover, due to the large variations in different and limited PG designs, data augmentation and curriculum learning are used to enhance model performance and scalability.

The main contributions are as follows:

- We propose IR-Fusion, an innovative fusion framework that incorporates numerical solutions to enhance ML methods for static IR drop analysis, providing an effective trade-off between accuracy and efficiency.
- We fuse hierarchical numerical-structural information to represent the influence of different metal layers of PG and propose an Inception Attention U-Net to capture hierarchical information of features from diverse scales and directions, thereby improving the performance.
- Considering the limitations and diversity of PGs, we employ augmented curriculum learning during the training phase to enhance the scalability of the ML model.
- Experimental results show that IR-Fusion outperforms SOTA ML baselines, substantially improves the F1 score, and reduces MAE and maximum IR drop error (MIRDE), critical metrics in the industry. Also, it has better accuracy and efficiency compared with the numerical method.

II. Preliminaries

A. Power Grid and IR Drop Analysis

In very large-scale integrated (VLSI) circuits, the design of PG includes analyzing and optimizing the on-chip metal network for power distribution. This process requires balances - ensuring the network delivers sufficient and reliable performance while minimizing resource consumption [2]. Typically, PG is designed from the top-level metal layer, which is connected to the power supplier, down through inter-layer vias, and finally to the active cells, as shown in Fig. 1. The current and resistance distribution in the upper layers influences the lower layers. Therefore, it is necessary to consider the characteristics of each metal layer in the PG.

IR drop analysis is critically important as it represents a fundamental step in evaluating the chip's power integrity. Due to the effects of the parasitic within the PG, particularly the partial voltage caused by the resistance of metal interconnects, the voltage drop is induced between the power pads and the cells as current flows through the PG. Consequently, it is essential to conduct thorough checks to verify the IR drop values at the working cells within specified limits.

B. Conventional Numerical Method for PG Analysis

The general circuit system can be described by Ohm's Law, Kirchoff's Voltage Law, and Kirchoff's Current Law. For the static analysis of PG, only the parasitic resistance effect of

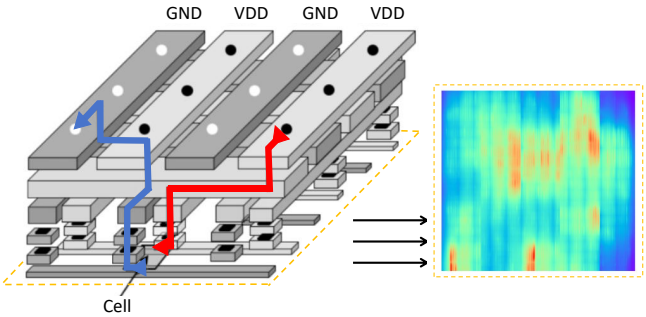


Fig. 1 The multiple-layer structure of PG.

the metal wires is considered, so the whole circuit network is a resistance network. Traditionally, using the modified nodal analysis (MNA) method [7], electrical parameters can be directly filled into the matrix by reading the elements and topological relationships in the network. The system matrix of a n-node PG network can be formulated as a linear system:

$$\boldsymbol{G}\boldsymbol{x} = \boldsymbol{I}, \tag{1}$$

where $\boldsymbol{G} \in \mathbb{R}^{n\times n}$ represents the conductance matrix, $\boldsymbol{x} \in \mathbb{R}^{n\times 1}$ is the unknown voltage vector of circuit nodes, and $\boldsymbol{I} \in \mathbb{R}^{n\times 1}$ is the vector of currents. By solving the equation, the node voltage response under static circuit conditions can be obtained. As the number of nodes in the PG grows exponentially, traditional methods struggle with longer solution times or even become infeasible due to high computational demands and memory demands. This exponential complexity not only challenges the efficiency of conventional techniques but also limits their applicability to large-scale problems. Consequently, the necessity for ML methods becomes evident.

C. U-Net Architecture and Attention Mechanism

U-Net [22], originally developed for medical image segmentation, is a widely adopted image-to-image architecture. Its encoder-decoder structure enables effective segmentation by leveraging context from more extensive overlapping regions. Skip connections between corresponding encoder and decoder layers preserve spatial information and fine-grained details during upsampling, facilitating high-resolution reconstruction.

The Attention mechanism [23] enhances ML models by selectively focusing on specific parts of the input features Global and local attention [24] target features at varying spatial scales. The CBAM [25] extends this by addressing multi-scale and directional information. In this work, we incorporate CBAM to integrate global and local attention, improving feature representation across scales and directions.

D. Problem Formulation

This work aims to fuse the numerical method with ML to achieve better performance in static IR drop analysis while focusing on the IR drop of the cell at the bottom layer. To predict IR drops, the PG is conceptualized as a 2D spatial image, with each PG design represented as a data matrix. Each feature map, denoted as $\boldsymbol{P}_{\text{map}i}$ essentially serves as a spatial representation of the inherent properties of the PG. The IR drop of working cells in the entire PG is also converted to a data matrix and represented as y. The algorithm F tries to

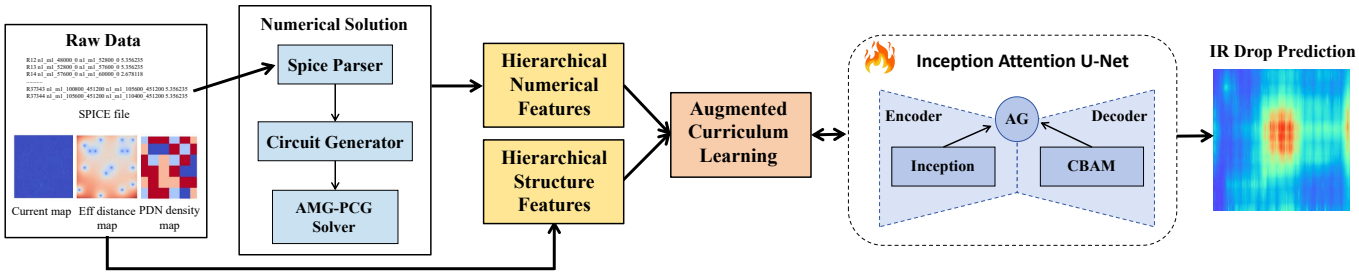


Fig. 2 The overview of IR-Fusion, a fusion framework for static IR drop analysis combining numerical solution and ML.

give the closest prediction $F*$ based on all the input features $\left(\boldsymbol{P}_{\text{map}_1}, ..., \boldsymbol{P}_{\text{map}_n}\right)$, formulated as:

$$F^* = \arg\min \text{Loss}\left(F\left(\boldsymbol{P}_{\text{map}_1}, ..., \boldsymbol{P}_{\text{map}_n}\right)\right), y\right). \quad (2)$$

III. IR-Fusion Framework

A. Overall

Our work focuses on static IR drop analysis and proposes a novel IR-Fusion framework that integrates the numerical solution with ML, as shown in Fig. 2. The numerical solution can provide a solid foundation for understanding and predicting complex PG systems. The integration with efficiency-promising ML contributes to the development of more robust and efficient analysis. In our framework, to obtain numerical solutions in a rapid and reliable process, an efficient AMG-PCG solver is applied to solve the system matrix. Then, based on the numerical solution, hierarchical numerical-structural fusion representing the multi-layer structure of the PG is employed. Subsequently, we design the Inception Attention U-Net model, which integrates the attention gate (AG), the Inception module for including both local details and surrounding context, and the CBAM that provides for both global and local attention, to achieve perceptual fields with less cost in the face of a large number of hierarchical features. Finally, data augmentation is applied to improve data diversity and robustness, while curriculum learning is used to handle the diversity of PGs, improving the scalability of IR-Fusion.

B. Numerical Solution using AMG-PCG

For the system matrix shown in Equation (1), we would like to utilize an efficient numerical method to provide as accurate an initial solution as possible in a shorter runtime, laying the foundation for understanding the PG system and subsequent ML refinement. In the numerical solution phase, the spice file of PG is regarded as an input, and a spice parser and a circuit generator for preprocessing the data are utilized. Aiming to provide reliable numerical solutions with rapid and robust processes, the algebraic multigrid preconditioned conjugate gradient (AMG-PCG) method in PowerRush [15] is applied.

The spice parser loads the spice file and creates a hash table of circuit nodes representing circuit connections. It first builds circuit elements as a sets. By traversing on these sets, the PG is stored as a nodes list and wires map, which are linked to present their topologies. Using this link table, the circuit generator constructs the circuit topology graph, enabling the extraction of the conductance matrix $\boldsymbol{G}$ for simulation.

Importantly, a linear solver based on the AMG-PCG method is used to solve the system matrix, as shown in Fig. 3. It can

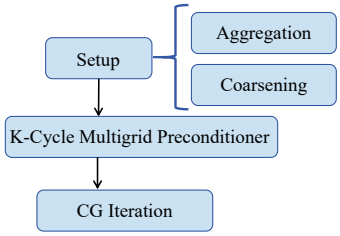


Fig. 3 The illustration of AMG-PCG solver.

be divided into three stages:

Setup Stage. The solver recursively selects coarser levels of the problem by grouping nodes and connections into progressively coarser grids. The conductance matrix $\boldsymbol{G}$ and current vector $\boldsymbol{I}$ in Equation (1) are transformed across multiple levels of coarser representations. Transfer operators connecting fine and coarse levels are represented by the preconditioning matrix $\boldsymbol{M}^{-1}$, which ensures that information, such as the residual $\boldsymbol{r}_k = \boldsymbol{I} - \boldsymbol{G}\boldsymbol{x}_k$ and search direction $\boldsymbol{p}_k$, is accurately transferred between grids.

Preconditioning Phase. The solver applies the K-cycle scheme, a multigrid cycling strategy that efficiently balances convergence speed and computational cost. The preconditioning step modifies the residual via the matrix $\boldsymbol{M}^{-1}$:

$$\boldsymbol{r}_{k+1} = \boldsymbol{r}_k - \frac{\boldsymbol{r}_k^\top \boldsymbol{M}^{-1}\boldsymbol{r}_k}{\boldsymbol{p}_k^\top \boldsymbol{G}\boldsymbol{p}_k}\boldsymbol{G}\boldsymbol{p}_k, \quad (3)$$

where $\boldsymbol{M}^{-1}\boldsymbol{r}_k$ represents the correction on multiple grid levels, ensuring fast convergence by addressing errors at various scales. This correction accelerates the reduction of the residual in each iteration.

CG Method. The solver utilizes aggregation-based AMG with the K-cycle as an implicit preconditioner for the Conjugate Gradient (CG) method. The iterative process updates the solution by combining the preconditioned residual with the search direction, using formulas:

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \frac{\boldsymbol{r}_k^\top \boldsymbol{M}^{-1}\boldsymbol{r}_k}{\boldsymbol{p}_k^\top \boldsymbol{G}\boldsymbol{p}_k}\boldsymbol{p}_k, \quad (4)$$

$$\boldsymbol{p}_{k+1} = \boldsymbol{M}^{-1}\boldsymbol{r}_{k+1} + \frac{\boldsymbol{r}_{k+1}^\top \boldsymbol{M}^{-1}\boldsymbol{r}_{k+1}}{\boldsymbol{r}_k^\top \boldsymbol{M}^{-1}\boldsymbol{r}_k}\boldsymbol{p}_k. \quad (5)$$

This ensures that the search direction is adjusted based on the multilevel corrections provided by AMG, allowing the CG method to converge more quickly to the final solution $\boldsymbol{x}$.

In general, numerical methods require more iterations to obtain an accurate solution. In this work, we use fewer

iterations to obtain fast and rough solutions and construct numerical features for ML, as detailed in Section III-C. This rough solution, although not accurate enough, still provides the IR drop values for each node and constructs detailed hierarchical numerical features, thus greatly benefiting ML in understanding and learning PG systems.

C. Hierarchical Numerical-Structural Information Fusion

Each metal layer in PG contributes to power delivery and impacts the final IR drop. Previous ML methods regard PG as a whole map missing multi-layer details. The above numerical methods can obtain rough IR drop solutions for each layer of nodes, which helps to model the whole PG more finely. Therefore, we construct hierarchical numerical features based on the numerical solution, according to the layer they belong to and their 2D spatial coordinate. Based on the row w and height l from Library Exchange Format (LEF), a design's layer of size $W_c \times L_c$ translates to an image of $W\,(=W_c//w) \times L\,(=L_c//l)$ pixels. In other words, the coordinates of each node x_n and y_n will be translated to $x = x_n//w$ and $y = y_n//l$. In this way, every node is planted into the 256×256 grid. Each metal layer corresponds to a generated feature map, allowing the PG to produce feature maps that align with the same number of grid layers in total.

Besides, given the limited representation of designs, our method extracts more hierarchical structure features using the PG spice file and cell layer features, including the current map, the effective distance map, and the PDN density map. We extract additional PG-structure-level features, such as the shortest path resistance map and resistance map to capture more details of the PDN structure. The shortest path resistance map is the average of the cumulative resistance from each node to voltage sources, while the resistance map distributes the resistance of each resistor across overlapping grids. Subsequently, we extract five kinds of hierarchical structure features based on their physical significance and the PG structure:

- **The current map** for each layer, representing the current distribution, is allocated proportionally based on the contribution from each layer, which is tied to resistance.
- **The effective distance**, calculated as the reciprocal of the sum of the reciprocals of Euclidean distances, measures proximity to voltage sources.
- **The PDN density map** is derived from the average PDN pitch within each grid as detailed in the spice file.
- **The resistance** and **shortest path resistance maps** are also computed based on their physical significance.

Hierarchical numerical and structure features together make up features for ML $\left(\boldsymbol{P}_{\text{map}_1}, ..., \boldsymbol{P}_{\text{map}_n}\right)$. This fusion of numerical-structural information enables the model to leverage theoretical insights and data-driven patterns. Numerical information offers stability and consistency, essential for interpreting complex datasets. The inclusion of structural information allows the model to adapt and learn from patterns within the data that may not be immediately apparent through numerical methods alone. This fusion-information-driven approach enables the model to uncover intricate relationships and trends, driving more accurate predictions.

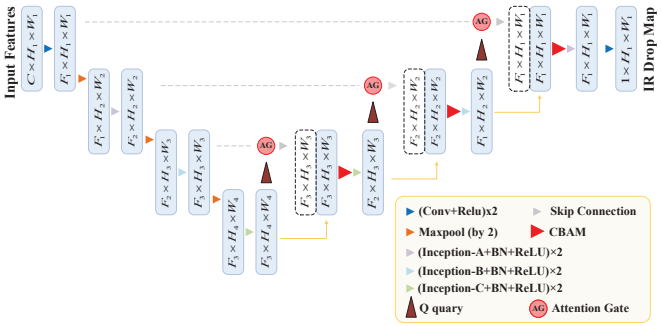


Fig. 4 The architecture of Inception Attention U-Net.

D. Inception Attention U-Net Model

To fully utilize the information from both hierarchical numerical and structure features, we design the Inception Attention U-Net. The architecture of our proposed model, illustrated in Fig. 4, is built upon the U-Net [22] framework and improves based on PGAU [20]. Input features are progressively filtered and downsampled by a factor of 2 at each scale, three times in the encoding part. Since hierarchical features provide more information, ordinary convolution methods using fixed-size kernels may lead to information loss. To enhance the network's ability to capture both local details and broader context, we incorporate Inception modules [26], [27]. Additionally, the relationship between feature layers is crucial. Moreover, the surrounding environment significantly influences special regions like hotspots. Hence, an attention mechanism with expanded scope and direction is essential for improving the model's interpretability and interaction. Thus, we incorporate the CBAM to focus on various scales and directions in subsequent decoder stages. Borrowing from the work in [18], the model leverages a regression-like layer at the end of the decoder path, for IR drop prediction at a 2D-spatial map level. In the subsequent sections, we will present a detailed explanation of the modules and blocks previously introduced.

Inception. Inception modules [26], [27] are multi-branch convolutional architectures designed to learn feature maps across different kernel sizes simultaneously. This multi-branch design helps the model to capture features at different scales, thus significantly improving its overall performance. There are three variations: Inception-A, Inception-B, and Inception-C. As detailed in [26], Inception-A is suitable for relatively early convolutional layers, while Inception-B is suitable for moderate-sized features. The Inception-C module is specifically optimized for high-dimensional feature extraction, contributing to the model's ability to process complex data representations. In our model architecture, we adopt this sequential arrangement of Inception-A, Inception-B, and Inception-C during the encoding phase. This systematic ordering not only aligns with established best practices in [27] but also minimizes information loss during downsampling. The kernel sizes of Inception modules used in our model are the same with [27], which helps capture details and broader context.

CBAM. To better focus on the information of hotspot regions and the interaction between various layers of hierarchical features, we employ CBAM [25] in our decoder. The CBAM layer

is an attention module specifically designed for convolutional neural networks. It consists of two components: the Channel Attention Module (Mc) representing global attention and the Spatial Attention Module (Ms) with the idea of local attention. While global attention attempts to capture the relevance of input features from a global perspective and direction, local attention only focuses on a local region (usually a fixed-size window) in the input feature map to calculate the attention weight. An intermediate feature map $m \in \mathbb{R}^{C\times H\times W}$ is given to the CBAM layer, it will calculate with $\mathrm{M_c} \in \mathbb{R}^{\mathrm{C}\times 1\times 1}$ and $\mathrm{M_s} \in \mathbb{R}^{1\times \mathrm{H}\times \mathrm{H}}$. The whole procedure is defined as:

$$\mathrm{m}' = \mathrm{M_c}(\mathrm{m}) \otimes \mathrm{m}, \mathrm{m}'' = \mathrm{M_s}\left(\mathrm{m}'\right) \otimes \mathrm{m}', \tag{6}$$

where m'' is the output of the CBAM layer and has the same size as m. CBAM enables the embedding of global information within the model, thereby mitigating the limitations perception filed. This module improves the network's performance without adding complexity.

E. Augmented Curriculum Learning

To enhance the model's robustness when the data is limited, we augment the training data by applying various transformations to each image-based input. Specifically, three operations are performed on each feature map: clockwise rotations of 90°, 180°, and 270°. Features originating from the same PG after these transformations are treated as new PG designs, resulting in a fourfold increase in the number of PG designs and significantly improving dataset diversity and robustness.

Since numerical methods provide a reliable prediction basis for data, the model tends to predict results related to numerical solutions. However, when it is migrated to new PGs, the model's scalability is easily lost in the prediction process, and the large number of features brought about by hierarchical operation will affect the convergence speed of the model. Curriculum learning (CL) [28] is a training strategy that trains ML models from easier samples to harder samples, which imitates the meaningful learning order in human curricula. Previous empirical studies have shown that curriculum learning strategies can improve the generalization capacity and convergence speed of various models in computer vision [29]. In our approach, we apply predefined CL for the train set. The predefined CL contains a predefined difficulty measurer and predefined continuous training scheduler [30], as shown in Fig. 5. In the predefined difficulty measurer, we set the artificially generated (fake) designs to "easier" and the real-world designs to "harder". The purpose of this setting is to strengthen the model's generalization ability.

By exposing the model to more complex "harder" data during training, the model can be forced to learn more robust and discriminating features. This helps the model maintain high performance in the face of previously unseen and potentially more complex real data. In addition, the "harder" data helps the model avoid overfitting or learning only simple features, instead capturing deeper patterns and patterns in the data. In the continuous training scheduler, the model adjusts the training data subset after each epoch.

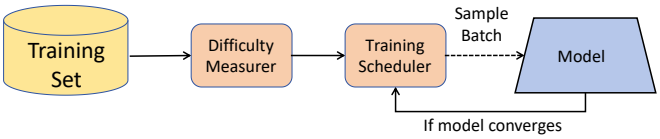


Fig. 5 The illustration of predefined curriculum learning.

IV. Evaluation

A. Experimental Settings

Baselines. IR-Fusion is compared with recent ML-based IR drop models, including IREDGe [17], MAVRIEC [18], IRPnet [19], PGAU [20], and MAUnet [21]. Besides, it is compared to the winner of the ICCAD 2023 contest [31].

Datasets. The ICCAD2023 dataset [31], specialized for the static IR drop prediction task, is used for evaluation. It contains 120 designs, 20 of which are real designs, and the rest were artificially generated based on [32], named fake designs, close to realistic PGs. The dataset provides spice and image-based data where each pixel in images represents the current, effective distance, PDN density matrix, and IR drop for a PDN node in a $1\mu m \times 1\mu m$ region. We follow the contest setup, using 10 real designs for testing and the rest for training.

Data augmentation increases the dataset size fourfold, with oversampling applied: fake designs are doubled, and real ones are quintupled. Following a curriculum learning strategy, fake designs are categorized as "easier," while real designs are classified as "harder." In particular, all baselines adopt the data after augmentation for training.

Metrics. Following the contest [31], the mean absolute error (MAE), F1 score, and runtime are selected as the evaluation metrics. The MAE is the average of the absolute difference between a prediction and the ground truth. The F1 score reflects the accuracy and comprehensiveness of the prediction for the hotspots region. In binary classification tasks, the outcomes of the predictions can be divided into four distinct categories: true positive (TP), true negative (TN), false positive (FP), and false negative (FN). TP and TN correspond to correctly identified positive and negative samples, respectively, while FN and FP indicate samples incorrectly classified as positive and negative. In this problem, IR drop values exceeding 90% of the maximum ground truth are classified as positive samples, whereas values below this threshold are considered negative. The $F1$ score is defined as:

$$P = \frac{TP}{TP+FP}, R = \frac{TP}{TP+FN}, F1 = \frac{2\times P\times R}{P+R}. \tag{7}$$

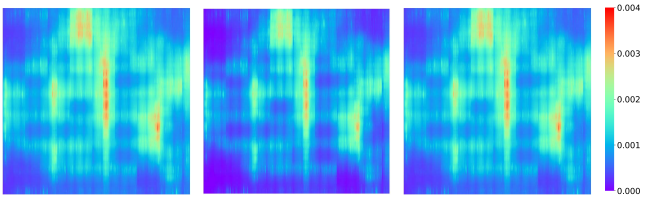
Since designers are more concerned with the worst-case of IR drop, its modeling error is extremely critical. The error in the region of maximum IR drop is evaluated, called MIRDE. Moreover, runtime is used to compare the efficiency.

B. Main Results

The IR drop analysis results are summarised in TABLE I. Our IR-Fusion surpasses all baselines and gains considerable advantages on each accuracy metric. Considering newly-proposed SOTA MAUnet [21], our approach achieves better performance with the improvement of 28.3% on MAE, 14.5% on F1, and 27.6% on MIRDE, with no significant time cost

TABLE I Main results. The unit of MAE and MIRDE is $\times 10^{-4}V$ and runtime is s.

Methods	MAE↓	F1↑	Runtime↓	MIRDE↓
IREDGe [17]	3.75	0.49	1.55	7.52
MAVIREC [18]	2.78	0.46	1.97	5.88
IRPnet [19]	1.66	0.61	2.22	5.25
PGAU [20]	1.72	0.60	2.07	5.02
MAUnet [21]	1.06	0.62	2.31	4.38
Contest Winner [31]	1.08	0.57	2.24	4.33
IR-Fusion (Ours)	**0.72**	**0.71**	6.98	**3.05**



(a) Golden (b) MAUnet (c) IR-Fusion (Ours)

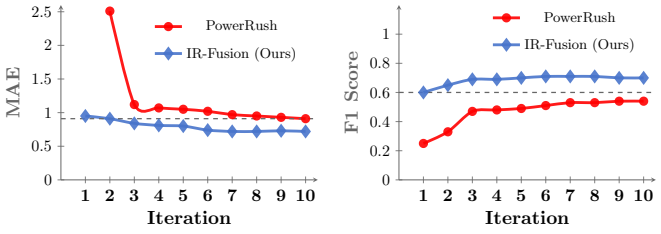
Fig. 6 Visualization of IR Drop distribution of a PG.

increase. Our prediction images have higher structural fidelity, with significant error reductions in multi-level (high and low IR drop regions, especially for significant Top 10% and 20% drop hotspot areas) and multiscale (global and local) IR drop regions. Fig. 6 visualizes the predicted IR drop map given by our model and SOTA MAUnet, comparing it with the golden label. Our model can intuitively provide more detailed predictions for the IR drop map with fewer errors and trends closer to the label. Moreover, compared with other ML baselines, our method achieves better performance in every metric. IR-Fusion still outperforms all baselines in MIRDE, representing more accuracy in the worst-case region. With the foundation of numerical solutions, the model can begin training from a point that is much closer to the target label, while CBAM increases the interpretability of the model's decisions, allowing it to perform robustly under varying conditions.

Due to the integration of numerical methods, our framework faces an inevitably higher time consumption. Nevertheless, the performance improvements achieved are substantial. In summary, our proposed fusion framework achieves more outstanding and robust performance within an acceptable runtime.

C. Trade-off Study

To evaluate the effectiveness of fusion, IR-Fusion is compared with an efficient numerical simulator PowerRush [15]. The numerical methods in both perform 1-10 iterations and the results are shown in Fig. 7. Notably, IR-Fusion surpasses



(a) MAE (b) F1 Score

Fig. 7 The analysis results of IR-Fusion and PowerRush [15].

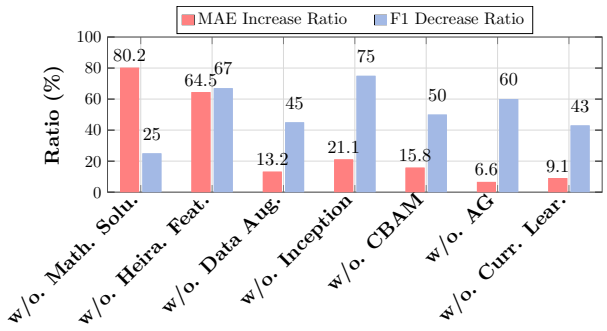


Fig. 8 Results of ablation study, showing the increase ratios in MAE (red) and decrease ratios in F1 score (blue).

PowerRush in all evaluated metrics. A key advantage of IR-Fusion is its ability to achieve the same MAE in just 2 iterations, while PowerRush requires 10 iterations to reach the same level. Furthermore, IR-Fusion consistently achieves a higher F1 score— a performance level PowerRush cannot reach at any iteration. This observation suggests that the numerical solver may partially overlook the patterns associated with hotspots, whereas ML provides an additional layer of detection, effectively capturing supplementary features. The introduction of ML drastically reduces the number of iterations required for numerical methods, which directly reduces computational costs and improves efficiency. Thanks to the fusion of numerical and ML methods, IR-Fusion achieves a better trade-off between accuracy and efficiency.

D. Ablation Study

Ablation experiments are conducted to evaluate the impact of various techniques in our IR-Fusion. Fig. 8 shows the results without a certain technique. The results demonstrate that the numerical solution (Num. Solu.) significantly reduces MAE, likely due to its precise initial point for learning. Additionally, our hierarchical features also improve performance. Replacing traditional convolution with Inception boosts F1, leveraging multi-scale information to identify hotspots better. The CBAM module, incorporating global and local attention, further enriches the model's learning. Both data augmentation (Data Aug.) and curriculum learning (Curr. Lear.) also contribute to performance gains, especially in F1 score.

V. Conclusion

In this work, we propose a novel framework IR-Fusion for static IR drop analysis, combining the numerical solution with ML. IR-Fusion exploits the advantages of both methods and can achieve a good trade-off between efficiency and accuracy. Experiments demonstrate that our framework can achieve the best performance compared to SOTA methods.

Acknowledgment

This work is supported by the National Key R&D Program of China (2022YFB2901100), the National Natural Science Foundation of China (No. U20B2045, U1936220, 62192784, 62172052, 62002029, 61772082, 62404021), the Beijing Natural Science Foundation (No. 4244107, QY24216, QY24204), and the Research Grants Council of Hong Kong SAR (No. CUHK14210723, CUHK14211824).

REFERENCES

[1] X. Zhan, P. Li, and E. Sánchez-Sinencio, "Distributed on-chip regulation: Theoretical stability foundation, over-design reduction and performance optimization," in *ACM/IEEE Design Automation Conference (DAC)*, 2016, pp. 1–6.

[2] L.-T. Wang, Y.-W. Chang, and K.-T. T. Cheng, *Electronic design automation: synthesis, verification, and test*. Morgan Kaufmann, 2009.

[3] H. Ren and J. Hu, *Machine learning applications in electronic design automation*. Springer, 2022.

[4] K. F. Yong, C. T. Lim, and W. K. Teng, "System level IR drop impact on chip power performance signoff for RISC-v system on chip," in *IEEE/ACM International Microsystems, Packaging, Assembly and Circuits Technology Conference (IMPACT)*, 2022, pp. 1–4.

[5] T. A. Davis and E. Palamadai Natarajan, "Algorithm 907: KLU, a direct sparse solver for circuit simulation problems," *ACM Transactions on Mathematical Software (TOMS)*, vol. 37, no. 3, pp. 1–17, 2010.

[6] Y. Chen, T. A. Davis, W. W. Hager, and S. Rajamanickam, "Algorithm 887: CHOLMOD, supernodal sparse Cholesky factorization and update/downdate," *ACM Transactions on Mathematical Software (TOMS)*, vol. 35, no. 3, pp. 1–14, 2008.

[7] T.-H. Chen and C. C.-P. Chen, "Efficient large-scale power grid analysis based on preconditioned Krylov-subspace iterative methods," in *ACM/IEEE Design Automation Conference (DAC)*, 2001, pp. 559–562.

[8] C.-H. Chou, N.-Y. Tsai, H. Yu, C.-R. Lee, Y. Shi, and S.-C. Chang, "On the preconditioner of conjugate gradient method—A power grid simulation perspective," in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2011, pp. 494–497.

[9] H. Qian, S. R. Nassif, and S. S. Sapatnekar, "Power grid analysis using random walks," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, vol. 24, no. 8, pp. 1204–1224, 2005.

[10] S. Köse and E. G. Friedman, "Fast algorithms for IR voltage drop analysis exploiting locality," in *ACM/IEEE Design Automation Conference (DAC)*, 2011, pp. 996–1001.

[11] M. Zhao, R. V. Panda, S. S. Sapatnekar, T. Edwards, R. Chaudhry, and D. Blaauw, "Hierarchical analysis of power distribution networks," in *ACM/IEEE Design Automation Conference (DAC)*, 2000, pp. 150–155.

[12] K. Sun, Q. Zhou, K. Mohanram, and D. C. Sorensen, "Parallel domain decomposition for simulation of large-scale power grids," in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2007, pp. 54–59.

[13] S. Cauley, V. Balakrishnan, and C.-K. Koh, "A parallel direct solver for the simulation of large-scale power/ground networks," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, vol. 29, no. 4, pp. 636–641, 2010.

[14] C. Zhuo, J. Hu, M. Zhao, and K. Chen, "Power grid analysis and optimization using algebraic multigrid," *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, vol. 27, no. 4, pp. 738–751, 2008.

[15] J. Yang, Z. Li, Y. Cai, and Q. Zhou, "PowerRush: An efficient simulator for static power grid analysis," *IEEE Transactions on Very Large Scale Integration Systems (TVLSI)*, vol. 22, no. 10, pp. 2103–2116, 2013.

[16] Z. Liu and W. Yu, "PowerRChol: Efficient Power Grid Analysis Based on Fast Randomized Cholesky Factorization," in *ACM/IEEE Design Automation Conference (DAC)*, 2024, pp. 1–6.

[17] 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 *IEEE/ACM Asia and South Pacific Design Automation Conference (ASPDAC)*, 2021, pp. 690–696.

[18] V. A. Chhabria, Y. Zhang, H. Ren, B. Keller, B. Khailany, and S. S. Sapatnekar, "MAVIREC: ML-aided vectored IR-drop estimation and classification," in *IEEE/ACM Proceedings Design, Automation and Test in Eurpoe (DATE)*, 2021, pp. 1825–1828.

[19] Y. Meng, R. Lyu, Z. Bi, C. Yan, F. Yang, W. Hu, D. Zhou, and X. Zeng, "Circuits physics constrained predictor of static IR drop with limited data," in *IEEE/ACM Proceedings Design, Automation and Test in Eurpoe (DATE)*, 2024, pp. 1–2.

[20] F. Guo, J. Liu, J. Zhai, J. Jia, K. Zhao, and C. Shi, "PGAU: Static IR Drop Analysis for Power Grid using Attention U-Net Architecture and Label Distribution Smoothing," in *ACM Great Lakes Symposium on VLSI (GLSVLSI)*, 2024, pp. 452–458.

[21] M. Wang, Y. Cheng, Y. Lin, K. Peng, S. Yang, Z. Jin, and W. Xing, "MAUnet: Multiscale attention U-Net for effective IR drop prediction," in *ACM/IEEE Design Automation Conference (DAC)*, 2024, pp. 1–6.

[22] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 2015, pp. 234–241.

[23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Annual Conference on Neural Information Processing Systems (NIPS)*, 2017, pp. 5998–6008.

[24] E. M. in Natural Language Processing (EMNLP), "Effective approaches to attention-based neural machine translation," 2015, pp. 1412–1421.

[25] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *European Conference on Computer Vision (ECCV)*, 2018, pp. 3–19.

[26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 2818–2826.

[27] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning," in *AAAI Conference on Artificial Intelligence*, 2017, pp. 4278–4284.

[28] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, "Curriculum learning," in *International Conference on Machine Learning (ICML)*, 2009, pp. 41–48.

[29] A. Pentina, V. Sharmanska, and C. H. Lampert, "Curriculum learning of multiple tasks," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 5492–5500.

[30] X. Wang, Y. Chen, and W. Zhu, "A survey on curriculum learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 44, no. 9, pp. 4555–4576, 2021.

[31] Winners of CAD Contest at ICCAD 2023. [Online]. Available: https://www.iccad-contest.org/2023/Winners.html

[32] V. A. Chhabria, K. Kunal, M. Zabihi, and S. S. Sapatnekar, "BeGAN: Power grid benchmark generation using a process-portable GAN-based methodology," in *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, 2021, pp. 1–8.