Disentangled Multimodal Learning for Scalable Dynamic IR-Drop Analysis

Yihan Wen[1], **Su Zheng**[1], **Tianshu Hou**[1], **Shiping Wu**[2], **Yuzhou Feng**[2], **Bei Yu**[1*]
[1]The Chinese University of Hong Kong [2]Independent Researchers

Abstract

The robust delivery of power is a non-negotiable prerequisite for the operational reliability of modern chips. Even a single unanticipated IR-drop violation can cause system-level crashes, making IR-drop a strict criterion in chip signoff. However, current methodologies ranging from physics-based simulators to recent learning-based surrogates remain bottlenecked by an inherent coupling between spatial topology and temporal evolution. This paradigm is intrinsically unscalable when billion-gate spatial complexity converges with the extensive workloads of full-testbench verification. In this work, we reformulate dynamic IR-drop assessment as a disentangled multimodal learning problem and introduce Layout Waveform Fusor (LWFusor), which decouples invariant spatially static chip layouts from transient switching waveforms and fuses them through physics-aligned end-to-end learning. By projecting activity sequences directly into the spatial potential domain, the proposed approach departs from traditional time-stepping bottlenecks and makes long-horizon power integrity analysis computationally efficient. Experiments on industrial-scale designs demonstrate orders-of-magnitude acceleration while maintaining high fidelity in hotspot capture, highlighting a new trajectory for power integrity sign-off in large-scale EDA workflows.

1 Introduction

The rapid scaling of chips has become the cornerstone for the massive scale-up of modern artificial intelligence, yet it simultaneously imposes challenges on electronic design automation (EDA) [3]. As transistors continue to shrink and integration density increases, many physical phenomena become important factors affecting the chip yield. One of the most critical issues is the dynamic voltage drop (IR-drop) in the power delivery network (PDN). Driven by surging current densities, such excessive IR-drop compromises the integrity of the supply voltage at circuit cells, causing severe timing degradations, and ultimately leading to unpredictable logic failures or system-to-silicon mismatches [15]. Consequently, the ability to effectively assess worst-case dynamic IR-drop has become a fundamental requirement for ensuring the functional correctness of high-performance computing applications.

Dynamic IR-drop analysis is increasingly constrained by the coupled growth of spatial and temporal complexity in modern physical design. On the spatial side, the complexity of chip layouts characterized by large scale renders traditional analytical methods computationally challenging. On the temporal side, ensuring functional coverage requires analyzing long switching waveforms, dramatically expanding the analysis horizon. The combination of these two factors results in a prohibitively large state space where the repetitive solution of high-dimensional governing equations becomes the primary bottleneck for conventional analysis.

Recent machine learning approaches [8, 11, 26, 6] attempt to alleviate this burden, but often inherit similar scalability limitations due to a sub-optimal problem formulation. The typical flow of these methods is shown in Figure 1(a), which tightly couples static layout geometry with dynamic switching waveforms into a single monolithic input. As a result, invariant physical structures must be repeatedly

*Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

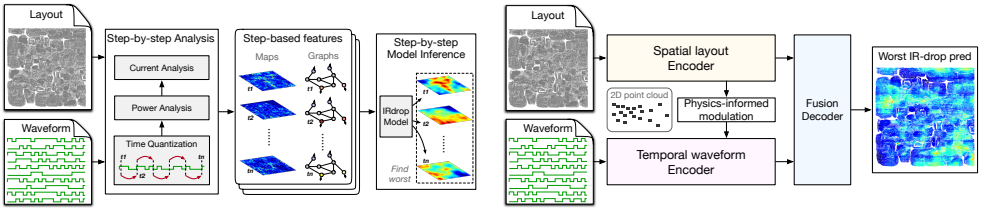


Figure 1: (a) Traditional dynamic IR-drop prediction flow; (b) Proposed LWFusor overview.

re-encoded for every temporal step, leading to substantial computational redundancy and severely limiting scalability. Moreover, prior methods rely on fine-grained step-by-step inference driven by time-varying power features, where each step incurs picosecond-resolution dynamic power analysis and expensive waveform-level feature extraction. Together, these limitations call for a more principled learning paradigm that respects the physical separation between static physical layout and dynamic activity modeling and fuses them through physics-aligned interactions without relying on costly intermediate power feature extraction.

In this paper, we propose LWFusor, a physically grounded multimodal learning framework for scalable dynamic IR-drop prediction. As demonstrated in Figure 1(b), the LWFusor reformulates dynamic IR-drop analysis as a cross-modal inference problem, where invariant layout structures and dynamic switching waveforms are encoded separately and coupled through physics-aligned interactions. This design enables efficient, end-to-end prediction of worst-case IR-drop over long temporal windows by eliminating the need for iterative power estimation. The main contributions are as follows:

- We introduce a PDN-aware point-cloud formulation for large-scale chip layouts, which enables cell-level modeling without sacrificing scalability. By embedding PDN-aware priors through resistance-weighted neighborhood construction and structure-guided hierarchical sampling, the proposed representation provides a physically consistent geometry for capturing spatial electrical correlations.
- We propose a disentangled multimodal architecture that separates invariant spatial layout encoding from dynamic waveform processing. This formulation reframes dynamic IR-drop prediction from step-wise regression to cross-modal projection, eliminating redundant computation on static structures and enabling near-linear scaling with respect to the temporal window size.
- We develop a physics-aligned fusion paradigm via feature-wise modulation, which enables direct mapping from logic activity to voltage drop. This design allows the model to condition dynamic behavior on physical context in an end-to-end manner, bypassing the traditional power-to-IR-drop analysis stage and its associated computational overhead.

The experimental results on industrial cases demonstrate that the proposed framework achieves impressive performance compared with the state-of-the-art baselines in both accuracy and scalability for dynamic IR-drop prediction. LWFusor maintains high fidelity in capturing transient hotspots while delivering significant speedups, validating its potential for full-chip sign-off acceleration.

2 Preliminaries

2.1 Problem Formulation

The objective of dynamic IR-drop analysis is to determine the transient voltage fluctuations across a chip's PDN under specific operational workloads. This phenomenon is governed by the complex interplay between the fixed physical implementation of the chip and its dynamic switching activities.

IR-drop on Power Delivery Network. The PDN is a massive, multi-layer resistive-capacitive (RC) network designed to distribute power from external supply pins (PAD) to standard cell instances on the die. The spatial characteristics of this network include wire geometries and via placements, which are defined during the power planning stage. As current (I) flows through resistance (R) on the PDN to reach a cell, a portion of the voltage is dropped along the path according to the Kirchhoff equation.

Activity Waveform. The actual workload on a chip is characterized by the switching activities of its circuit cells, typically recorded in a Value-Change Dump (VCD) format file in the VLSI industry.

Table 1: Layout and waveform features.

Modality	Shape	Features	width	Type	Description
Layout $\mathbf{L}$ (Spatial)	$\mathbb{R}^2$	(x, y)	2	float	Cell 2-D coordinate on chip layout
	$\mathbb{R}^{18}$	D_{PAD}	1		The weighted Manhattan distance to the nearest power PAD
		P_{rise}	3		Avg., max., and variation of cell total rise power distribution in the PDK
		P_{fall}	3		Avg., max., and variation of cell total fall power distribution in the PDK
		C_{load}	2		Cell and net capacitance from parasitic parameter extraction
		$T_{A,rise}$	2		Rise arrival timing window from timing analysis
		$T_{S,rise}$	2		Rise transition timing window from timing analysis
		$T_{A,fall}$	2		Fall arrival timing window from timing analysis
		$T_{S,fall}$	2		Fall transition timing window from timing analysis
		$freq$	1		The frequency of the clock domain in which the cell resides
Waveform $\mathbf{W}$ (Temporal)	$\mathbb{R}^T$	$[v_i^1, \cdots, v_i^T]$	T	int	Logical status (0 or 1) in each clock cycle

These waveforms represent the logical transitions (e.g., $0 \rightarrow 1$ or $1 \rightarrow 0$) of cells over millions of clock cycles. Each time a cell toggles its logic state, it draws a transient surge of current from the PDN. They create massive, time-varying current demands that fluctuate at a picosecond resolution.

Analysis Paradigms. Industrial workflows rely on a two-step simulation: first converting logic toggles into instantaneous current vectors, and then solving global systems of differential equations to calculate nodal voltages. Although these numerical solvers are highly accurate, their tremendous runtime becomes computationally prohibitive as workloads extend to thousands of cycles. More recently, learning-based surrogates have been introduced to bypass these solvers by directly mapping PDN and power features to IR-drop values, thus avoiding costly numerical solving.

2.2 Limitations of Previous Works

Existing approaches to dynamic IR-drop prediction span rasterization-based, graph-based, and point-based learning paradigms, as well as spatio-temporal modeling techniques for waveform-driven analysis. A comprehensive review of these methods is provided in Appendix A. Despite their progress, current solutions exhibit two fundamental limitations that hinder scalability to modern design and workload regimes.

First, existing approaches largely entangle spatially static layout structures with dynamically varying waveform excitations. Most methods jointly process spatial and temporal information through concatenated image channels or fused node features, overlooking their fundamentally different roles. Even recent dual-path transformer designs [8] still inject all temporal frames into the spatial pathway at each iteration, causing static layout information to be repeatedly re-encoded. This entanglement leads to significant computational redundancy and poor scalability to long temporal windows.

Second, prior works rely on fine-grained temporal iteration driven by frame-wise dynamic power or current maps. Such formulations require expensive dynamic power analysis at every time step, followed by repeated model inference. Even rasterization-based approaches, which achieve high inference throughput, remain constrained by per-step feature preparation and therefore can only sample a small fraction of temporal windows from million-step waveforms [6]. As a result, for testbench-level workloads, this paradigm becomes computationally prohibitive, rendering comprehensive worst-case IR-drop assessment impractical.

Motivation. To overcome these limitations, we adopt an explicitly disentangled formulation that separates static layout encoding from dynamic waveform processing, eliminating redundant spatial computation and enabling scalable worst-case IR-drop prediction. Instead of rasterized or graph-based representations, the framework leverages a hierarchical point-based layout embedding and fuses it with temporal features through a physics-aligned FiLM modulation mechanism [20]. By reformulating the task from dense, step-by-step temporal prediction to direct worst-case estimation over long waveform windows, the proposed approach bypasses fine-grained temporal iteration and achieves efficient testbench-level analysis.

3 LWFusor Overview

This paper represents the worst-case IR-drop prediction as a multimodal regression problem containing a spatially static chip layout and a temporally dynamic waveform.

Physical Layout as 2-D Point Cloud. The physical layout $\mathbf{C}$ of a chip consisting of N standard cells is defined in Equation (1). Each cell c_i is represented as a point in a 2-D space corresponding to its physical location on the chip die.

$$\mathbf{C} = \{c_1, c_2, \ldots, c_N\}. \tag{1}$$

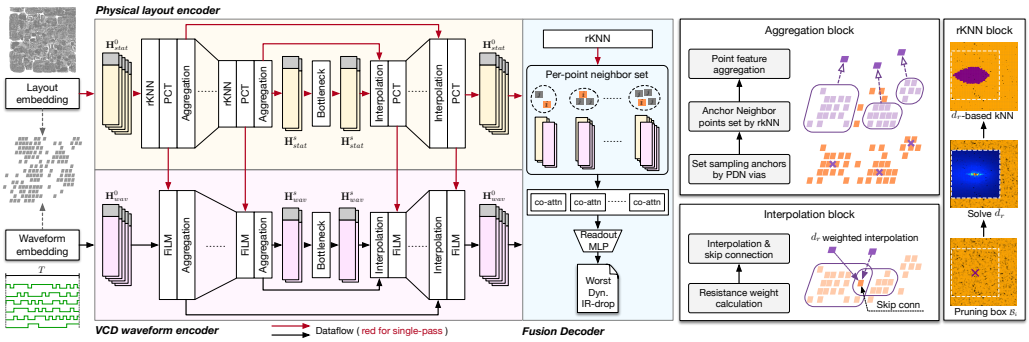


Figure 2: The LWFusor architecture for IR-drop prediction.

We define the layout input as a set of pairs $\mathbf{L}$ in Equation (2), where $\mathbf{p}_i \in \mathbb{R}^2$ represents the spatial coordinates and $\mathbf{f}_i \in \mathbb{R}^l$ denotes a vector of intrinsic properties on cells such as library cell type, parasitic capacitance, power and timing metrics. Detailed features are summarized in Table 1.

$$\mathbf{L} = \{\mathbf{p}_i, \mathbf{f}_i\}_{i=1}^{N} \in \mathbb{R}^{N\times(2+l)}. \tag{2}$$

Temporal Switching Activity as Waveform. We define the logical signal on cell c_i that triggers power consumption and IR-drop as waveform $\mathbf{w}_i$ in Equation (3). For a time slot of T clock cycles, the raw logic state of cell c_i is a binary sequence $[v_i^1, \ldots, v_i^T]$, where v_i^t denotes the logical state of cell c_i at clock cycle t. The signal tensor $\mathbf{W}$ denotes the complete waveform input for all cells in the chip.

$$\begin{aligned} \mathbf{W} &= \{\mathbf{w}_i\}_{i=1}^{N} \in \{0,1\}^{N\times T}, \\ \mathbf{w}_i &= [v_i^1, v_i^2, \ldots, v_i^T] \in \{0,1\}^{T}. \end{aligned} \tag{3}$$

Learning Objective. Given the multimodal spatial-temporal input $(\mathbf{L}, \mathbf{W})$, LWFusor learns a mapping function $\mathcal{F}$ that predicts the peak dynamic voltage drop $\hat{y}_i$ for each cell c_i across the entire window T, as shown in Equation (4):

$$\hat{y}_i = \mathcal{F}(\mathbf{L}, \mathbf{W}). \tag{4}$$

By predicting the slot-T-level maximum IR-drop directly, we avoid the high computational cost of traditional step-by-step iterations at picosecond granularity.

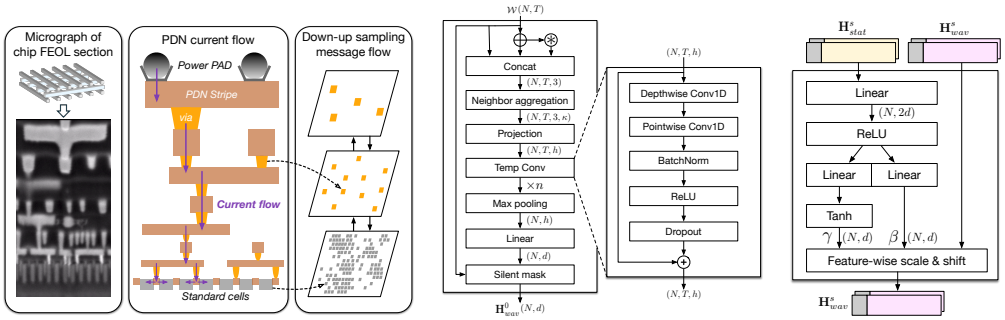
Physics-modulated Dual-stream Architecture. As illustrated in Figure 2, LWFusor adopts a dual-stream architecture designed to disentangle the invariant layout physical features from the varying inductive queries in the waveform. This design enables the amortized inference strategy: the computationally expensive processing of the static layout features is performed only once (red flow in Figure 2), while a lightweight dynamic branch processes varying waveforms. Finally, a fusion decoder integrates the disentangled representations to predict the worst-case IR-drop at the cell level.

The upper branch in Figure 2 serves as the physical layout encoder, which is designed to capture the static context of the layout's inherent electrical properties. To capture the multi-scale nature of the layout, we utilized a hierarchical Point Cloud Transformer (PCT) [29]. This encoder processes $\mathbf{L}$ through a series of aggregation and interpolation stages. The lower branch serves as the waveform encoder, which encodes and modulates the waveform information at each hierarchy. Instead of simple concatenation, we introduce a physics-inspired interaction mechanism using Feature-wise Linear Modulation (FiLM). The pre-computed static features $\mathbf{H}_{stat}^s$ from the layout branch are used to generate scale γ and shift β parameters to modulate the dynamic waveform features $\mathbf{H}_{wav}^s$. This modulation physically grounds the logical activity by leveraging features to inject the physical information that is absent in the binary waveform. The fusion decoder fuses the disentangled representations to predict the worst-case IR-drop. Given that IR-drop exhibits strong spatial locality, we employ a localized fusion decoder to optimize both accuracy and efficiency.

4 Methodology and Model Design

4.1 Physical Layout Encoder

The physical layout encoder is designed to extract multi-scale spatial features from the chip layout. To maintain physical fidelity while scaling up to large designs, we adopt a hierarchical point



(a) (b) (c)

Figure 3: (a) Physically aligned hierarchical via point sampling; (b) Waveform embedding block; (c) FiLM block.

transformer architecture. The encoder follows a U-Net-like structure, progressively aggregating the cell-level features into coarse-grained representations. The coarsest level features are processed by the bottleneck network, and then upsampled back to the original resolution through interpolation layers. At each hierarchy level, the layout state $\mathbf{H}_{stat}$ is processed through the PCT block to capture localized electrical interactions. The field of each point's PCT calculation is constrained to neighbors obtained by the resistance-weighted k-nearest neighbors (rkNN) block to ensure scalability.

Resistance-weighted kNN (rkNN). Standard geometric distance fails to capture the anisotropic conductivity of power grids. We define a resistance-weighted metric d_r as an approximation of the equivalent resistance on the PDN between cells. By substituting Euclidean distance with d_r, rkNN forces the model to aggregate features from electrically relevant aggressors with low-resistance paths, rather than mere geometric neighbors. The details of rkNN are illustrated in Section D.1.

Hierarchical Via Point Sampling (VPS). Instead of generic downsampling, hierarchical Via Point Sampling (VPS) utilizes the physical locations of PDN power vias as sampling anchors to align the model's architecture with the chip's multi-tier PDN topology. As illustrated in Figure 3(a), the PDN is a hierarchical structure where transient currents are constrained by vertical via bottlenecks and horizontal metal stripes. VPS selects anchors from different layers to capture the immediate current paths, regional spreading, and macro-scale distribution across the die. This physics-aligned hierarchy ensures that the model's message-passing flow mirrors actual electrical behavior: the aggregation stage represents the physical confluence of current into power trunks, while interpolation models the anisotropic distribution of supply voltage back to cells. By grounding the receptive field in the resistive topology rather than mere Euclidean distance, VPS effectively captures the short and long-range electrical interactions governing dynamic IR-drop.

Hierarchical Message Passing. The layout encoder performs bidirectional information flow between cells and via anchors, as shown in Figure 2. Aggregation abstracts cell features into coarser via representations, mirroring the physical confluence of currents from local logic gates to power trunks. Conversely, Interpolation restores spatial granularity by upsampling features back to cell resolution, modeling the anisotropic distribution of supply voltage from the grid. By employing rkNN as the neighborhood operator, the layout encoder encourages feature propagation to follow the resistive topology, thereby approximating the physical behavior of current flow within the PDN. The details of aggregation and interpolation are illustrated in Section D.2.

In the inference stage, the physical encoder is a one-time cost since the chip layout remains constant across different workloads. The encoded feature maps $\mathbf{H}_{stat}$ at all levels and the rkNN result $\mathbf{C}_{\mathcal{N},i}$ are cached and reused by the waveform encoder and fusion decoder.

4.2 Waveform Encoder

The Waveform encoder is a hierarchical encoder that integrates temporal feature extraction and multi-scale feature modulation using FiLM to generate point-level embeddings aligned with physical layout encoder outputs.

Multichannel Activity Tensor. While raw waveform signals provide precise logical status, the localized intensity of signal flips is often a more informative indicator of the IR-drop phenomenon. To capture this, we process the binary flip vector $\mathbf{w}_i$ into a multi-channel toggle activity representation. First, we define the binary flip indicator Δw_i^t to represent signal transitions. For a cell c_i at time t

with logic state $v_i^t \in \{0, 1\}$, the flip activity is calculated using the XOR operation Equation (5):

$$\Delta w_i^t = v_i^t \oplus v_i^{t-1}. \tag{5}$$

This operation yields 1 if a transition (either $0 \to 1$ or $1 \to 0$) occurs, and 0 otherwise, which effectively captures both rising and falling edges of the digital signal.

To further extract temporal context, we define the local switching density σ_i^t by aggregating these transitions over a sliding window of size $\mathcal{T} = 2\tau + 1$. The local switching density is computed by Equation (6):

$$\omega_i^t = \sum_{\delta=-\tau}^{\tau} \Delta w_i^{t+\delta}, \tag{6}$$

where τ is the half-window size. The final activity tensor $\mathcal{W}$ after preprocessing is constructed by concatenating the raw logic states, flip indicators, and switching densities, as shown in Equation (7):

$$\mathcal{W} = \Big[\mathbf{v}_i, \Delta\mathbf{w}_i, \boldsymbol{\omega}_i\Big] \in \mathbb{R}^{N\times T\times 3}. \tag{7}$$

This multi-channel input allows the model to simultaneously perceive instantaneous switching events and the cumulative activity levels within a local temporal window.

Waveform Temporal Embedding. To transform the binary activity tensor $\mathcal{W}$ into a dense temporal representation, we design a wave embedding module that integrates local switching dynamics with neighborhood contexts, as shown in Figure 3(b). The embedding process begins by aggregating the states of neighboring cells to capture the temporal coupling. For each cell c_i, the neighbor aggregation block concatenates $\mathcal{W}$ over its spatial neighbors $\mathbf{C}^0_{\mathcal{N},i} \in \mathbb{R}^{\kappa}$, resulting in a context-aware tensor of dimension $\mathbb{R}^{N\times T\times 3\times\kappa}$. The $\mathbf{C}^0_{\mathcal{N},i}$ is already calculated and cached by the rkNN in the physical layout encoder and can thus be reused here without extra cost. The concatenated tensor is then projected into a hidden space $\mathbb{R}^{N\times T\times h}$ to facilitate subsequent feature extraction. Then, the convolutional layers extract temporal patterns from the projected tensor. After sequence processing, the max-pooling is applied to extract the most critical patterns within the window, followed by a linear projection to align the embedding.

In a typical chip workload, a significant portion of cells remains idle. To explicitly focus the model on active switching regions and prevent the model from being distracted by non-switching cells, we implement a silent mask layer. A binary mask derived from the original input is applied for each cell, as shown in Equation (8), where $\mathbb{I}(\cdot)$ is the indicator function. This mask ensures that the final temporal embedding $\mathbf{H}^0_{wav} \in \mathbb{R}^{N\times d}$ focuses on active regions, effectively filtering out noise from silent logic gates and enhancing the fidelity of the peak IR-drop prediction.

$$\mathbf{H}^0_{wav} \leftarrow \mathbf{H}^0_{wav} \odot \mathbb{I}\left(\sum_{t=1}^{T} \Delta w_i^t > \mathbf{0}\right). \tag{8}$$

Hierarchical Modulation with FiLM. To effectively bridge the gap between abstract logical signals and their physical properties, we employ a hierarchical modulation strategy using Feature-wise Linear Modulation (FiLM). The core intuition is that while the waveform $\mathbf{W}$ dictates cells' switching activities, the physical layout $\mathbf{L}$ governs the magnitude of the resulting current consumption due to varying cell drive strengths and local placement densities. To capture this interaction explicitly, we define the modulation process at each hierarchy level k, as shown in Equation (9).

$$\text{FiLM}(\mathbf{H}^k_{wav}|\mathbf{H}^k_{stat}) = \gamma^k(\mathbf{H}^k_{stat}) \odot \mathbf{H}^k_{wav} + \beta^k(\mathbf{H}^k_{stat}), \tag{9}$$

where $\gamma^k(\cdot) \in \mathbb{R}^{N\times d}$ and $\beta^k(\cdot) \in \mathbb{R}^{N\times d}$ are learned scaling and shifting functions at level k respectively, and $\odot$ denotes the Hadamard product.

The detailed internal structure of FiLM is shown in Figure 3(c). The static feature matrix $\mathbf{H}^k_{stat}$ is first processed through a bottleneck to extract high-level physical context. This representation is then chunked into two specialized linear heads corresponding to the scaling γ and shifting β parameters. The scaling factor γ is further constrained by the $\tanh$ activation to ensure numerical stability during training. The shifting factor β represents the bias. Performing a feature-wise scaling and shifting, this modulation strictly follows the physical intuition that identical switching sequences result in heterogeneous IR-drop responses depending on their spatial and electrical context.

Table 2: Worst-case dynamic IR-drop prediction accuracy comparison between the proposed method and baselines IRXGBoost [11], DuST [8], IGHF [19], and Cloud LSTM [28].

Method	Regression				Critical cell identification			Risk prioritizing (NDCG)		
	MAE↓	R^2↑	PCC↑	Spea↑	Accuracy↑	F1↑	ROC-AUC↑	10%↑	5%↑	1%↑
Proposed	**2.4778**	**0.8167**	**0.9088**	**0.9164**	**0.9242**	**0.6693**	**0.9503**	**0.9349**	**0.9189**	**0.8956**
IRXGBoost	4.9565	0.3261	0.6488	0.6589	0.8470	0.4040	0.8263	0.7799	0.7264	0.6187
DuST (ViT)	3.2328	0.7006	0.8376	0.8399	0.9163	0.6050	0.9212	0.9080	0.8856	0.8202
IGHF (CNN+GNN)	3.0875	0.7530	0.8833	0.8886	0.9215	0.6352	0.9403	0.9255	0.9071	0.8720
Cloud LSTM	4.2929	0.5192	0.7310	0.7339	0.8960	0.4340	0.8815	0.8483	0.8229	0.7637

4.3 Fusion Decoder

The Fusion Decoder integrates disentangled static and dynamic representations to predict the peak IR-drop value. Given that peak IR-drop arises from the intricate spatial-temporal coupling between physical impedance and simultaneous switching, we propose a masked co-attention mechanism to model these cross-modal interactions. The core of the decoder is a symmetric cross-modal attention operator. The updated layout embedding $\mathbf{H}^0_{stat}$ is computed by querying the waveform context of its rkNN neighbors $\mathbf{C}^0_{\mathcal{N},i}$ shown in Equation (10a), and the waveform embedding $\mathbf{H}^0_{wav}$ is updated symmetrically in Equation (10b):

$$\mathbf{H}^0_{stat} \leftarrow \sigma\left(\frac{\mathcal{P}_q(\mathbf{H}^0_{stat}) \cdot \mathcal{P}_k(\mathbf{H}^0_{wav})^\top}{\sqrt{d}} + \mathbf{M}\right) \mathcal{P}_v(\mathbf{H}^0_{wav}) \tag{10a}$$

$$\mathbf{H}^0_{wav} \leftarrow \sigma\left(\frac{\mathcal{P}_q(\mathbf{H}^0_{wav}) \cdot \mathcal{P}_k(\mathbf{H}^0_{stat})^\top}{\sqrt{d}} + \mathbf{M}\right) \mathcal{P}_v(\mathbf{H}^0_{stat}) \tag{10b}$$

$$\mathbf{M} \in \{0, -\infty\}^{N \times N}, M_{ij} = \begin{cases} 0, & \text{if } j \in \mathbf{C}^0_{\mathcal{N},i} \\ -\infty, & \text{otherwise} \end{cases} \tag{10c}$$

Here $\mathcal{P}_q, \mathcal{P}_k, \mathcal{P}_v$ denote the query, key, and value projection operators applied row-wise to the feature matrices. The spatial coupling is enforced by the neighbor mask matrix $\mathbf{M} \in \mathbb{R}^{N \times N}$ in Equation (10c), where $\mathbf{C}^0_{\mathcal{N},i}$ denotes the rkNN-identified neighbor indices for cell c_i, which is already calculated and cached in the physical layout encoder. This formulation ensures that the attention mechanism is constrained by the rkNN-defined topology. In practice, rather than instantiating the full dense matrix, the implementation relies on a sparse attention operator. This effectively prunes the global attention space and reduces the computational complexity from $O(N^2)$ to $O(Nk)$.

The resulting multi-modal embeddings are concatenated and mapped to the target space via a Read-out MLP Φ to produce the final prediction in Equation (11):

$$\hat{y}_i = \Phi(\text{concat}(\mathbf{H}^0_{stat}[i], \mathbf{H}^0_{wav}[i])). \tag{11}$$

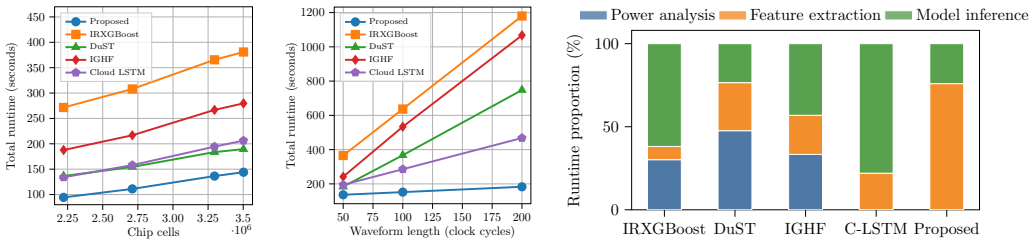
By constraining the high-dimensional interaction to localized clusters, Masked coAttn maintains the scalability necessary for billion-scale physical designs while preserving high-fidelity modeling of localized electrical failures.

5 Experiments

In this section, we evaluate the performance of the proposed method on various cases. Four baseline methods, including three state-of-the-art (SOTA) IR-drop prediction models and one point cloud sequence model, are selected for performance comparison. IRXGBoost [11] is the SOTA of cell-level IR-drop prediction using gradient boosting decision trees; DuST [8] is a recent image-based IR-drop prediction method based on a spatio-temporal vision transformer (ViT); IGHF [19] is the SOTA of the CNN-GNN fusion method for IR-drop prediction; Cloud LSTM [28] is a point cloud-based fusion method in the stream forecasting task. Since its question formulation and data modality are similar to those of IR-drop prediction, we adapt Cloud LSTM to our task for comparison. All baselines and the proposed method are evaluated on the same test set, which comprises four industrial-scale chip designs ranging from 2.2 to 3.5 million cells. Each design features a layout and three waveforms with lengths ranging from 50 to 200 clock cycles. The implementation details of all methods and dataset configurations are provided in Section B.

5.1 Accuracy of Worst-case IR-drop Prediction

We evaluate the accuracy of worst-case IR-drop prediction from three aspects: (i) cell-level regression performance, (ii) accuracy of critical cell identification, and (iii) cell risk prioritizing accuracy. In regression performance, mean absolute error (MAE), R^2, Pearson correlation coefficient (PCC), and



(a) (b) (c)

Figure 4: Runtime comparison on chip test design with (a) Different cell counts; (b) Different waveform lengths; and (c) Proportion of the runtime of each step

Spearman's rank correlation coefficient (Spea) are metrics at the cell level. For the image-based baseline DuST, we apply interpolation to obtain cell-level IR-drop values from tiles. In critical cell identification, the accuracy, F1 value, and the area under the ROC curve (ROC-AUC) are adopted as metrics to evaluate the performance of identifying critical cells with top 10% IR-drop. Industrially, designers also care about the risk prioritizing of cells to optimize the chip design. A correct ranking of cells based on their IR-drop risk is also important. We introduce the normalized discounted cumulative gain (NDCG) of top 10%, 5%, and 1% IR-drop as the metric to evaluate the risk prioritizing performance of different methods.

The average results of all test designs and waveforms are summarized in Table 2. Compared with the four baselines, the proposed method achieves the best performance in all three aspects. In regression performance, the proposed method achieves the best MAE of 2.4778, 0.8167 for R^2, 0.9088 for PCC, and 0.9164 for Spearman correlation. In critical cell identification, the proposed method achieves the best accuracy of 0.9242, F1 score of 0.6693, and AUC of 0.9503. In risk prioritizing, the proposed method achieves the best NDCG of 0.9349 for the top 10%, 0.9189 for the top 5%, and 0.8956 for the top 1% of cells. These results demonstrate the effectiveness of the proposed method in accurately predicting worst-case IR-drop at the cell level. Compared with the image-based DuST, the proposed method achieves higher accuracy since it can directly process cell-wise features without information loss during image rasterization. Compared with cell-level IRXGBoost and IGHF, the proposed method achieves a significant improvement because it more effectively extracts layout and waveform features.

5.2 Runtime and Scalability

In this section, the runtimes of the proposed method and the baselines are compared under different chip scales and waveform lengths. We measure runtime as the sum of method-specific feature extraction and model inference. Shared preprocessing for layout features, including design file parsing, parasitic resistance-capacitance extraction, and static timing analysis, is common to all methods and therefore excluded from per-method runtime. As illustrated in Fig. 4(c), the runtime is distributed among dynamic power analysis, feature extraction, and model inference. Unlike previous IR-drop prediction methods, which require costly dynamic power analysis, our approach removes this computational burden. Most of the runtime is instead attributed to a one-time, reusable feature extraction step, enabling efficient processing of long waveforms and ensuring strong scalability.

Layout scalability. The runtime comparison under different chip scales is shown in Figure 4(a), in which the scales of the four design cases are 2.2, 2.7, 3.3, and 3.5 million cells, respectively. All four cases have waveforms with 50 cycles each. The proposed method achieves the lowest runtime among all methods. For the largest design with 3.5 million cells, the proposed method takes only 144.2 seconds, which is obviously faster than baseline methods. This demonstrates the efficiency of the proposed method in handling large-scale chip designs.

The runtime comparison across waveform lengths is shown in Figure 4(b), where all methods are evaluated on the largest design (3.5M cells) with waveform lengths of 50, 100, and 200 cycles. Our method consistently achieves the lowest runtime. This runtime advantage arises from two key design choices: (i) we process waveform windows directly and extract worst-case IR-drop without costly per-cycle power analysis or iterative feature reconstruction, and (ii) we disentangle spatial and temporal encoding, so layout-dependent computations are performed once rather than repeated for every time step, reducing inference cost on long sequences.

Waveform scalability. To evaluate the scalability of dynamic IR-drop analysis at the testbench level, we conduct experiments on sustained workloads using a design case with 3.3 million cells and

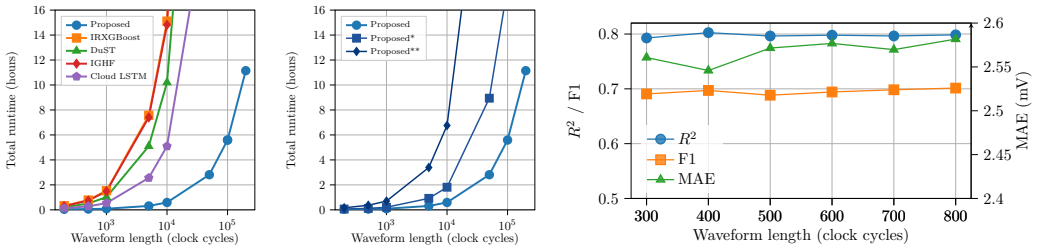


Figure 5: The scalability to testbench-level waveforms: (a) Runtime comparison of the proposed method and baselines; (b) Ablation studies of the proposed method; (c) Extrapolation consistency across different waveform lengths.

a waveform window of up to 200,000 cycles. It is important to note that generating ground-truth IR-drop results for sequences of this magnitude is practically intractable within industrial design cycles. Even with cutting-edge commercial EDA tools, dynamic IR-drop simulation on a 200k-cycle window would require several months of continuous computation due to inherent complexity, where each picosecond step demands a full matrix solve. To bridge this evaluation gap, we adopt a two-pronged validation strategy. First, we focus on runtime scalability. Then we provide an extrapolation consistency study. By verifying that LWFusor maintains stable predictive performance as the horizon extends, we provide empirical evidence that the model captures steady-state physical invariants, ensuring its reliability for the 200k-cycle regime where golden simulators fail to deliver results.

Figure 5 reports the runtime comparison under increasing waveform lengths, where the x-axis is log-scaled for clarity. For the longest 200k-cycle waveform, the proposed method completes full IR-drop assessment in 11.5 hours, enabling testbench-level analysis within half a day. In contrast, all baseline methods fail to finish within one week due to their reliance on frame-wise dynamic power analysis and per-step model inference, resulting in millions of steps of power analysis and over 50k inference passes. By contrast, the proposed LWFusor directly infers worst-case IR-drop over waveform windows of $T = 50$ cycles, requiring only 4000 model inferences and eliminating dynamic power analysis entirely. To isolate the contribution of design choices, we conduct ablation studies shown in Figure 5(b). "Proposed*" removes the disentangled encoding and re-computes the layout representation for each waveform window. This modification increases runtime from 11.5 hours to 35.7 hours on the 200k-cycle waveform, demonstrating the importance of separating static and dynamic modalities. "Proposed**" additionally introduces dynamic power analysis prior to model inference. In this setting, the runtime again exceeds one week, demonstrating that the elimination of per-step dynamic power analysis is critical for achieving testbench-level scalability.

Extrapolation Consistency Analysis. To evaluate the reliability of LWFusor in scenarios involving extremely long waveforms where commercial golden simulators become computationally prohibitive, we conduct an extrapolation consistency study on the minimal design with 2.2 million cells. By varying the input waveform length from 300 to 800 cycles, we examine whether the model maintains stable predictive fidelity as the temporal horizon extends. As summarized in Figure 5(c), LWFusor demonstrates performance stability across increasing window sizes. Specifically, the R^2 fluctuates within a negligible margin, while the MAE remains nearly horizontal, showing no signs of error accumulation or performance degradation. The consistency in F1-score further suggests that the model's ability to capture transient hotspots is independent of the sequence length. This verifies that LWFusor effectively captures invariant spatio-temporal correlations rather than overfitting to specific waveform granularities or local temporal patterns. This steady-state behavior provides strong confidence in the model's accuracy and robustness when scaling to full-testbench workloads in industrial sign-off flows.

6 Conclusion

This paper presents a scalable spatio-temporal multimodal learning framework for dynamic IR-drop prediction. By disentangling static layout representation from dynamic switching waveforms, the proposed method eliminates redundant re-encoding of invariant physical structures. The layout and waveform features are fused via physics-aligned feature modulation, enabling worst-case IR-drop inference without fine-grained temporal iteration or costly dynamic power analysis. Experiments demonstrate superior accuracy with substantial runtime reduction over existing baselines, highlighting the effectiveness of disentangled multimodal modeling for large-scale power integrity analysis.

Acknowledgments and Disclosure of Funding

The Research Grants Council of Hong Kong SAR (No. CUHK14211824, No. CUHK14210723, and T46-415/25-R). The authors declare no competing interests.

References

[1] Y. Cai, Y. Liang, Z. Luo, B. Xie, and X. Li. PCT-Cap: Point cloud transformer for accurate 3D capacitance extraction. In *International Symposium of Electronics Design Automation (ISEDA)*, pages 421–426. IEEE, 2024.

[2] Z. Chai, Y. Zhao, W. Liu, Y. Lin, R. Wang, and R. Huang. CircuitNet: An Open-Source Dataset for Machine Learning in VLSI CAD Applications With Improved Domain-Specific Evaluation Metric and Learning Strategies. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, 42(12):5034–5047, 2023.

[3] L. Chen, Y. Chen, Z. Chu, W. Fang, T.-Y. Ho, R. Huang, Y. Huang, S. Khan, M. Li, X. Li, et al. The dawn of AI-native EDA: Opportunities and challenges of large circuit models. *arXiv preprint arXiv:2403.07257*, 2024.

[4] Y.-H. Chen, Y.-C. Cheng, Y.-F. Chang, Y.-C. Lee, J.-W. Lin, H.-W. Pao, P.-W. Chen, P.-Y. Chen, H.-Y. Chen, Y.-C. Chen, C.-Y. Wang, and S.-C. Chang. Dynamic IR-Drop Prediction Through a Multi-Task U-Net with Package Effect Consideration. In *Design, Automation and Test in Europe Conference (DATE)*, pages 1–7. IEEE, 2025.

[5] 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)*, pages 690–696. ACM, 2021.

[6] 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 *Design, Automation and Test in Europe Conference (DATE)*, pages 1825–1828, 2021.

[7] H. Fan, X. Yu, Y. Ding, Y. Yang, and M. Kankanhalli. PSTNet: Point Spatio-Temporal Convolution on Point Cloud Sequences. In *International Conference on Learning Representations (ICLR)*, 2021.

[8] B. Fu, L. Liu, Q. Wang, Y. Wang, M. D. F. Wong, and E. F. Y. Young. Fast dynamic IR-Drop prediction with dual-path spatial-temporal attention. In *Design, Automation and Test in Europe Conference (DATE)*, pages 1–7. IEEE, 2025.

[9] F. Guo, Y. Xi, J. Zhai, J. Jia, J. Liu, K. Zhao, and C. Shi. IRGNN: A Graph-based Framework Integrating Numerical Solution and Point Cloud for Static IR Drop Prediction. In *ACM/IEEE Design Automation Conference (DAC)*, pages 1–7, 2025.

[10] X.-X. Huang, H.-C. Chen, S.-W. Wang, I. H.-R. Jiang, Y.-C. Chou, and C.-H. Tsai. Dynamic IR-Drop ECO Optimization by Cell Movement with Current Waveform Staggering and Machine Learning Guidance. In *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 1–9, 2020.

[11] Y.-C. Lee, Y.-H. Chen, Y.-C. Cheng, Y.-F. Chang, J.-W. Lin, H.-W. Pao, Y.-C. Chen, Y.-T. Li, W. Tang, S.-C. Chang, et al. Real-time dynamic IR-drop prediction for IR ECO. In *ACM/IEEE Design Automation Conference (DAC)*, pages 1–7. IEEE, 2025.

[12] W. Li, G. Chen, H. Yang, R. Chen, and B. Yu. Learning point clouds in EDA. In *ACM International Symposium on Physical Design (ISPD)*, pages 55–62, 2021.

[13] W. Li, Y. Qu, G. Chen, Y. Ma, and B. Yu. TreeNet: Deep point cloud embedding for routing tree construction. In *IEEE/ACM Asia and South Pacific Design Automation Conference (ASPDAC)*, pages 164–169, 2021.

[14] Y. Li, R. Yu, C. Shahabi, and Y. Liu. Diffusion convolutional recurrent neural network: Data-driven traffic forecasting. In *International Conference on Learning Representations (ICLR)*, 2018.

[15] S. Lin and N. Chang. Challenges in power-ground integrity. In *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 651–654, 2001.

[16] X. Liu, M. Yan, and J. Bohg. MeteorNet: Deep learning on dynamic 3D point cloud sequences. In *IEEE International Conference on Computer Vision (ICCV)*, pages 9246–9255, 2019.

[17] Z. Liu and W. Yu. pGRASS-Solver: A Parallel Iterative Solver for Scalable Power Grid Analysis Based on Graph Spectral Sparsification. In *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 1–9, 2021.

[18] K. Ma, Z. Wang, H. He, Q. Xu, T. Chen, and H. Geng. LMM-IR: Large-scale netlist-aware multimodal framework for static IR-Drop prediction. In *ACM/IEEE Design Automation Conference (DAC)*, pages 1–7. IEEE, 2025.

[19] D. Niu, D. Zhang, Y. Cao, Z. Jin, C. Wang, Y. Dong, and C. Sun. A novel image-graph heterogeneous fusion framework for static IR Drop prediction. In *ACM/IEEE Design Automation Conference (DAC)*, pages 1–7, 2025.

[20] E. Perez, F. Strub, H. de Vries, V. Dumoulin, and A. C. Courville. FiLM: Visual reasoning with a general conditioning layer. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2018.

[21] J. Shi, Y. Cai, W. Hou, L. Ma, S. X.-D. Tan, P.-H. Ho, and X. Wang. GPU friendly Fast Poisson Solver for structured power grid network analysis. In *ACM/IEEE Design Automation Conference (DAC)*, pages 178–183, 2009.

[22] M.-H. Tsai, W.-S. Ding, H.-Y. Hsieh, and J. C.-M. Li. Transient IR-Drop analysis for at-speed testing using representative random walk. *IEEE Transactions on Very Large Scale Integration Systems (TVLSI)*, 22(9):1980–1989, 2014.

[23] 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)*, pages 1–6. ACM, 2024.

[24] C.-J. Wei, H. Chen, and S.-J. Chen. Design and Implementation of Block-Based Partitioning for Parallel Flip-Chip Power-Grid Analysis. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, 31(3):370–379, 2012.

[25] Z. Wu, S. Pan, G. Long, J. Jiang, and C. Zhang. Graph WaveNet for deep spatial-temporal graph modeling. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pages 1907–1913, 2019.

[26] 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 *IEEE/ACM Asia and South Pacific Design Automation Conference (ASPDAC)*, pages 13–18, 2020.

[27] J. Yang, Z. Li, Y. Cai, and Q. Zhou. PowerRush : Efficient transient simulation for power grid analysis. In *IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pages 653–659, 2012.

[28] C. Zhang, M. Fiore, I. Murray, and P. Patras. CloudLSTM: A recurrent neural model for spatiotemporal point-cloud stream forecasting. In *AAAI Conference on Artificial Intelligence (AAAI)*, volume 35, pages 10851–10858, 2021.

[29] H. Zhao, L. Jiang, J. Jia, P. H. S. Torr, and V. Koltun. Point Transformer. In *IEEE International Conference on Computer Vision (ICCV)*, pages 16259–16268, 2021.

[30] Y. Zhao, Z. Chai, X. Jiang, Y. Lin, R. Wang, and R. Huang. PDNNet: PDN-Aware GNN-CNN Heterogeneous Network for Dynamic IR Drop Prediction. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems (TCAD)*, 44(6):2253–2263, 2025.

[31] Q. Zhou, K. Sun, K. Mohanram, and D. Sorensen. Large power grid analysis using domain decomposition. In *Design, Automation and Test in Europe Conference (DATE)*, volume 1, pages 1–6, 2006.

[32] Z. Zhu, B. Yao, and C.-K. Cheng. Power network analysis using an adaptive algebraic multigrid approach. In *ACM/IEEE Design Automation Conference (DAC)*, pages 105–108, 2003.

[33] J. Zou, X. Wang, J. Guo, W. Liu, Q. Zhang, and C. Huang. Circuit as set of points. *Advances in Neural Information Processing Systems*, 36:32468–32480, 2023.

A Related works

This paper covers several research areas, including IR-drop prediction, point cloud representation in EDA, and spatial-temporal fusion for point cloud data structures. We review the related works in these areas below.

A.1 IR-drop Analysis and Prediction

In order to efficiently solve the large-scale linear systems arising from IR-drop analysis, numerous works have focused on matrix decomposition-based methods and accelerating runtime through the block-based partitioning [24], domain decomposition [31], and random walk [22]. On the other hand, iterative solvers such as preconditioned conjugate gradient (PCG) [17, 21] and multigrid methods [27, 32] have been widely adopted for simulation-based IR-drop analysis.

However, simulation-based methods face challenges in handling the increasing scale of modern chips. To address these challenges, machine learning-based approaches have been proposed for dynamic IR-drop prediction. Typically, existing learning-based methods fall into two categories: the rasterization methods and cell-level methods. The rasterization methods represented by [26] discretize the chip and PDN layout into 2D images, and convert the IR-drop prediction into an image-to-image translation task. Early learning-based approaches formulate IR-drop prediction as an image regression problem, where convolutional neural networks are employed to infer full-chip IR-drop maps from rasterized layout features [26]. Subsequent works enhance spatial modeling capacity by introducing multi-scale encoder–decoder architectures to jointly capture global context and local layout variations [5]. To incorporate temporal dynamics, 3D convolutional operators have been adopted to model switching activities across time for dynamic IR-drop analysis [6]. Beyond single-objective prediction, multi-task learning frameworks have been proposed to simultaneously estimate worst and average IR-drop metrics [4]. To improve generalization to unseen designs, transfer learning techniques based on low-rank approximation have been introduced [23]. More recent methods leverage heterogeneous network structures to jointly model local and global characteristics [30], or adopt spatio-temporal transformer architectures to extract static spatial features together with dynamic, time-varying switching activities [8].

One major drawback of image-based methods is the information loss caused by the discretization of continuous PDN structures [12]. To address this issue, cell-level regression methods have been proposed for more precise and fine-grained IR-drop prediction [10, 11]. Graph-based formulations have also been explored to explicitly model the power delivery network, where graph neural networks are used to capture topological dependencies among standard cells and PDN stripes [9].

Recent research on an image-graph heterogeneous fusion framework [19] jointly captures global PDN spatial features and cell-level topological dependencies for accurate IR-drop prediction.

A.2 Point Cloud Representation in EDA

Point cloud representations are increasingly adopted in EDA as they preserve native layout geometry without discretization-induced information loss [12]. In physical design, point-based embeddings have been applied to model net pins for routing tree construction [13], and to represent circuit components for congestion and DRC violation prediction using transformer-based perception models [33]. Point clouds have also been shown effective for lithography hotspot detection [12]. Beyond routing and manufacturability, point cloud architectures have been extended to parasitic and power integrity analysis, including transformer-based capacitance extraction from interconnect structures [1] and static IR-drop prediction using point-based representations of cells and PDN elements [9, 18].

A.3 Spatial-Temporal Fusion on Point Clouds

In many EDA tasks, including dynamic IR-drop prediction, input data naturally comprises static spatial structures and dynamic temporal excitation waveforms, whose effective fusion is essential for modeling vector-driven physical behaviors. Spatio-temporal learning over point cloud sequences has been extensively studied. MeteorNet [16] constructs spatio-temporal neighborhoods to aggregate local point features across time. PSTNet [7] explicitly decouples spatial and temporal modeling via spatial and temporal convolutions, respectively. CloudLSTM [28] employs DConv for spatial aggregation and LSTM for temporal dependency modeling. When explicit connectivity is introduced, point sequences can be formulated as graph sequences, where adaptive dependency learning [25] and diffusion-

Table 3: The scales of the design cases and waveform lengths in the dataset.

Chip design	D1			D2			D3			D4			D5			VD1			TD1			TD2			TD3			TD4		
Number of cells	2549213			2553525			2863688			3069827			3159778			3015853			2221272			2710635			3295812			3502135		
VCD length (cycles)	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	50	100	200	50	100	200	50	100	200	50	100	200
Purpose	Training set															Valid set			Test set											

Table 4: Comparison between the proposed method and IR-drop prediction state-of-the-art baselines.

Method	Property				
	Model type	Cell-level accuracy	Power analysis -free	Spatial-temporal Disentangling	Iteration
IRXGBoost [11]	XGBoost	✓	×	None	Power frame
DuST [8]	ViT	×	×	Partial	Power frame context
IGHF [19]	GNN+CNN	✓	×	None	Power frame context
Proposed	Point cloud	✓	✓	Complete	Logical waveform window

based graph convolutions with RNN encoders [14] are used to capture coupled spatial–temporal dependencies.

B Experimental Details

B.1 Dataset Characterization

To evaluate the proposed method, a chip physical design dataset comprising 10 distinct CPU designs was implemented using the cutting-edge EDA tool with advanced process nodes. The detailed specifications of each design case are summarized in Table 3. While the current dataset is centered on high-performance CPU cores, it should be noted that the underlying physical mechanism of dynamic IR-drop, which is governed by the interaction between PDN resistive networks and transient switching currents, remains consistent across different functional units (e.g., GPUs or NPUs). By modeling these fundamental electrical properties, LWFusor is designed to be agnostic to the specific logic functionality of the design.

Each design case includes multiple Value Change Dump (VCD) format waveforms to capture signal switching activities. The dataset is partitioned into three subsets: chip cases D1 to D5 serve as the training set, while each case contains three VCD waveforms of 50 cycles in length. Case VD1 is designated for the validation set with three VCD waveforms of 50 cycles in length. Cases TD1 to TD4 are designated as the test set for accuracy and runtime evaluation, each containing three VCD waveforms of varying lengths (50, 100, and 200 cycles). Ground-truth IR-drop labels for these cases were obtained via full-waveform dynamic IR-drop analysis using the industry-standard tool Voltus.

Moreover, seven additional extremely long waveform lengths of 0.5k, 1k, 5k, 10k, 50k, 100k, and 200k cycles are generated for chip case TD3 to assess the scalability of the proposed method, which are excluded from Table 3. Due to the prohibitive computational resources and runtime required to analyze the IR-drop results for such massive scales, the ground truth labels are not available for these extended waveforms. These extended cases are only utilized for scalability benchmarks.

B.2 Comparison with Baseline Methods

The key differences between the proposed method and baseline IR-drop prediction methods are summarized in Table 4. All IR-drop prediction baselines rely on the dynamic power analysis to extract input features. Prior work typically uses switching, internal, and leakage power as inputs for dynamic IR-drop prediction. Leakage power is a static attribute extracted from the cell library, whereas switching and internal power are dynamic quantities computed from activity-driven power analysis. In our experiments, the dynamic power values are produced with the commercial EDA tool Voltus and evaluated at the frame level (5 frames per clock cycle). In contrast, the proposed method does not depend on precomputed dynamic power. It operates directly on raw logical waveforms and predicts the worst-case IR-drop within the input waveform window length of $T = 50$ clock cycles.

IRXGBoost [11] is a regression model designed for IR-drop prediction in the ECO stage. It operates at the frame level, taking features derived from a single dynamic power frame as input for each inference. In its original formulation, IRXGBoost includes pre-ECO IR-drop values as part of its

input features. Since such features are not available in our experimental setup, we remove this component and use the remaining power-related features for training and inference.

DuST [8] uses a 3D, spatial-temporal vision transformer (ViT) to process rasterized power maps. It consumes a context of multiple power frames, which are concatenated along the channel dimension to form the input tensor; in our experiments, each DuST step uses 20 consecutive frames (4 cycles). While DuST's temporal encoding captures short-term temporal correlations within this window, it does not fully disentangle spatial and temporal information because the spatial encoder still requires the full set of context frames. Consequently, DuST performs repeated step-level inference over long waveforms, redundantly re-encoding spatial features that remain unchanged across steps. In addition, dynamic power must be precomputed for every frame before inference, so overall runtime is dominated by feature preparation.

IGHF [19] employs a hybrid GNN+CNN architecture. Originally developed for static IR-drop prediction using total and leakage power features, we adapt IGHF for dynamic IR-drop prediction by supplying dynamic power features obtained by averaging multiple frames within each context window. IGHF processes 20 consecutive frames (4 cycles) per step, which is the same as DuST. Despite this adaptation, IGHF still requires rasterized power maps and repeated step-level inference over long waveforms, resulting in substantial computational overhead for testbench-scale workloads.

Cloud LSTM [28] is not listed in Table 4 because it was originally developed for point-cloud stream forecasting rather than IR-drop prediction. Nevertheless, given its similar problem formulation, we adapt Cloud LSTM as an additional baseline. In our adaptation, each point is fixed to the corresponding cell coordinates in the layout, and the per-cell static layout features are concatenated with the waveform features to form the input point-wise stream.

IRXGBoost, IGHF, and Cloud LSTM naturally operate at the cell level and can therefore be compared directly with our method. DuST, being an image-based method, predicts IR-drop values per pixel; we obtain cell-level predictions by sampling the predicted IR-drop map at each cell's center location.

B.3 Model Training Configurations

The training of the proposed method was performed end-to-end by the AdamW optimizer with learning rate = 1e-4 and weight decay = 1e-4. The learning rate was subsequently adjusted by an ExponentialLR scheduler with a linear warmup. Models were trained for up to 100 epochs with early stopping based on validation MAE. Both the baselines and the proposed method were implemented in PyTorch and trained on an NVIDIA RTX 5090 GPU.

DuST and IGHF rely on image-based input features that require large amounts of training data to achieve stable performance. To ensure fair and competitive baselines, these models were pretrained on the CircuitNet [2] dataset. CircuitNet provides power features generated by a toggle rate-based vectorless power analysis and does not include waveform-based information. Therefore, all baseline models were subsequently fine-tuned on our training set using dynamic power features derived from waveform-driven analysis, aligning them with the evaluation setting used for the proposed method.

B.4 Detailed Experimental Results

The detailed accuracy comparison between the baseline methods and the proposed approach across all test cases is summarized in Table 5. Overall, the proposed method achieves strong performance across most evaluation metrics and test cases, indicating its effectiveness for accurate cell-level IR-drop prediction.

On the larger chip designs TD3 and TD4, whose scales significantly exceed those of the training set, the proposed method maintains stable accuracy. This suggests that the learned representations generalize well to unseen large-scale layouts. For the image-based baseline DuST, we observe a noticeable accuracy degradation on these larger designs. This behavior is likely related to the fixed-resolution rasterization process, which may limit the ability to preserve fine-grained cell-level details as the layout scale increases.

The GNN-CNN hybrid model IGHF exhibits improved robustness compared to DuST on large designs, benefiting from its cell-level graph representation. Nevertheless, its performance remains below that of the proposed method, which more explicitly leverages both spatial layout structure and temporal waveform information through disentangled multimodal modeling.

Table 5: Detailed results comparison of the proposed method and baselines on each design case and waveform. Baselines are IRXGBoost [11], DuST [8], IGHF [19] and Cloud LSTM [28].

Chip design	Waveform	Method	Regression				Critical cell identification			Risk prioritizing (NDCG)		
			MAE↓	R^2↑	PCC↑	Spea↑	Accuracy↑	F1 ↑	ROC-AUC↑	10%↑	5%↑	1%↑
TD1	V1 (50 cycles)	Proposed	**2.4971**	**0.8208**	**0.9074**	**0.9179**	0.9186	**0.6662**	**0.9455**	**0.9291**	0.9093	0.8725
		IRXGBoost	5.3322	0.2741	0.6270	0.6206	0.8424	0.3850	0.8152	0.7796	0.7360	0.6490
		DuST	2.9913	0.7523	0.8675	0.8695	0.9182	0.6480	0.9326	0.9219	0.8991	0.8228
		IGHF	3.0670	0.7735	0.8932	0.8947	**0.9243**	0.6484	0.9440	0.9323	**0.9187**	**0.8955**
		Cloud LSTM	4.3662	0.5308	0.7301	0.7300	0.8885	0.4307	0.8709	0.8421	0.8245	0.7794
	V2 (100 cycles)	Proposed	**2.7893**	**0.7740**	**0.8898**	**0.9026**	0.8983	0.6303	**0.9290**	**0.9157**	0.8900	0.8430
		IRXGBoost	5.6348	0.2209	0.6454	0.6485	0.8439	0.3912	0.8110	0.7910	0.7462	0.6496
		DuST	3.1754	0.7296	0.8543	0.8655	0.8966	**0.6466**	0.9192	0.9018	0.8690	0.7714
		IGHF	3.1107	0.7583	0.8738	0.8834	**0.9113**	0.5875	0.9245	0.9113	**0.8910**	**0.8504**
		Cloud LSTM	4.4821	0.5075	0.7281	0.7385	0.8563	0.4147	0.8548	0.8312	0.8022	0.7484
	V3 (200 cycles)	Proposed	**2.4568**	**0.8039**	**0.8975**	**0.9015**	0.9233	0.6429	**0.9411**	**0.9304**	**0.9171**	**0.9041**
		IRXGBoost	5.2466	0.1884	0.5892	0.6033	0.8320	0.3417	0.7851	0.7554	0.6983	0.5849
		DuST	2.9570	0.7306	0.8549	0.8571	**0.9248**	**0.6488**	0.9323	0.9263	0.9085	0.8547
		IGHF	3.0266	0.7419	0.8704	0.8799	0.9152	0.6060	0.9327	0.9149	0.8939	0.8557
		Cloud LSTM	4.1336	0.5157	0.7194	0.7198	0.8980	0.4277	0.8685	0.8450	0.8260	0.7626
TD2	V1 (50 cycles)	Proposed	**2.6348**	**0.8176**	**0.9088**	**0.9180**	0.9145	**0.7159**	**0.9510**	**0.9381**	**0.9189**	**0.8810**
		IRXGBoost	5.4902	0.2857	0.6840	0.7009	0.8568	0.4449	0.8499	0.7930	0.7377	0.6277
		DuST	3.5869	0.6763	0.8244	0.8347	0.8913	0.6434	0.9114	0.8991	0.8795	0.8313
		IGHF	3.1608	0.7767	0.8862	0.8925	**0.9257**	0.6549	0.9453	0.9302	0.9128	0.8910
		Cloud LSTM	4.3564	0.5534	0.7631	0.7725	0.8608	0.4871	0.8914	0.8619	0.8360	0.7668
	V2 (100 cycles)	Proposed	**2.3576**	**0.8258**	**0.9148**	**0.9218**	**0.9290**	**0.6777**	**0.9564**	**0.9406**	**0.9256**	**0.8936**
		IRXGBoost	4.6545	0.3340	0.6453	0.6580	0.8461	0.4004	0.8197	0.7774	0.7257	0.6105
		DuST	3.1234	0.6952	0.8341	0.8360	0.9221	0.5819	0.9214	0.9070	0.8862	0.8139
		IGHF	2.9438	0.7559	0.8809	0.8809	0.9230	0.6420	0.9414	0.9271	0.9100	0.8637
		Cloud LSTM	4.0405	0.5245	0.7283	0.7277	0.9071	0.4385	0.8774	0.8452	0.8271	0.7631
	V3 (200 cycles)	Proposed	**2.2498**	**0.8340**	**0.9199**	**0.9277**	**0.9292**	**0.7027**	**0.9596**	**0.9452**	**0.9301**	**0.8950**
		IRXGBoost	4.9263	0.2755	0.6180	0.6269	0.8421	0.3836	0.8169	0.7712	0.7165	0.6026
		DuST	2.9722	0.7245	0.8513	0.8502	0.9227	0.6341	0.9329	0.9216	0.9011	0.8565
		IGHF	2.9324	0.7573	0.8869	0.8919	0.9274	0.6628	0.9478	0.9335	0.9155	0.8647
		Cloud LSTM	3.7481	0.5825	0.7662	0.7719	0.8971	0.4098	0.8884	0.8554	0.8295	0.7737
TD3	V1 (50 cycles)	Proposed	**2.4484**	**0.8211**	**0.9117**	**0.9180**	**0.9291**	**0.6635**	**0.9551**	**0.9361**	**0.9221**	**0.9175**
		IRXGBoost	4.7175	0.3779	0.6607	0.6668	0.8574	0.4473	0.8506	0.7895	0.7313	0.6130
		DuST	3.4118	0.6681	0.8180	0.8141	0.9242	0.5824	0.9187	0.9094	0.8911	0.8488
		IGHF	2.9518	0.7687	0.8882	0.8922	0.9238	0.6459	0.9436	0.9290	0.9162	0.8907
		Cloud LSTM	4.3307	0.5098	0.7162	0.7212	0.9068	0.4149	0.8858	0.8370	0.8067	0.7563
	V2 (100 cycles)	Proposed	**2.5609**	**0.8105**	**0.9012**	**0.9147**	**0.9143**	**0.6308**	**0.9362**	**0.9221**	**0.9035**	**0.8950**
		IRXGBoost	5.0622	0.3143	0.6336	0.6575	0.8368	0.3618	0.7964	0.7554	0.6965	0.5914
		DuST	3.5441	0.6605	0.8138	0.8237	0.9024	0.5491	0.8952	0.8871	0.8684	0.8045
		IGHF	3.1120	0.7519	0.8726	0.8882	0.9126	0.5938	0.9245	0.9080	0.8846	0.8473
		Cloud LSTM	4.4427	0.5087	0.7135	0.7245	0.8932	0.4247	0.8638	0.8376	0.8060	0.7208
	V3 (200 cycles)	Proposed	**2.3543**	**0.8235**	**0.9088**	**0.9166**	**0.9287**	**0.6353**	**0.9469**	0.9352	**0.9204**	**0.9018**
		IRXGBoost	4.7839	0.3433	0.6590	0.6581	0.8494	0.4141	0.8330	0.8060	0.7604	0.6705
		DuST	3.2101	0.6933	0.8340	0.8367	0.9139	0.5470	0.9130	0.9016	0.8677	0.7840
		IGHF	2.9332	0.7637	0.8857	0.8867	0.9175	0.6165	0.9354	**0.9354**	0.9149	0.8895
		Cloud LSTM	4.2353	0.5192	0.7206	0.7185	0.9061	0.4170	0.8889	0.8552	0.8309	0.7854
TD4	V1 (50 cycles)	Proposed	**2.3238**	**0.8334**	**0.9178**	**0.9205**	**0.9392**	**0.6955**	**0.9643**	**0.9461**	**0.9326**	**0.9078**
		IRXGBoost	4.4682	0.4329	0.6783	0.6889	0.8559	0.4412	0.8537	0.7893	0.7338	0.6247
		DuST	3.2562	0.6901	0.8316	0.8283	0.9268	0.5862	0.9248	0.9097	0.8903	0.8085
		IGHF	3.1262	0.7437	0.8887	0.8905	0.9272	0.6617	0.9496	0.9322	0.9126	0.8750
		Cloud LSTM	4.3637	0.5015	0.7309	0.7248	0.9148	0.4484	0.8961	0.8600	0.8359	0.7747
	V2 (100 cycles)	Proposed	**2.4516**	**0.8309**	**0.9152**	**0.9191**	**0.9354**	**0.7000**	**0.9624**	**0.9440**	**0.9331**	**0.9091**
		IRXGBoost	4.6207	0.4326	0.6760	0.6901	0.8497	0.4154	0.8390	0.7819	0.7280	0.6200
		DuST	3.2256	0.7075	0.8412	0.8389	0.9246	0.6125	0.9293	0.9110	0.8867	0.8387
		IGHF	3.2268	0.7434	0.8868	0.8893	0.9251	0.6519	0.9479	0.9276	0.9107	0.8789
		Cloud LSTM	4.4602	0.5069	0.7324	0.7323	0.9074	0.4583	0.8958	0.8604	0.8375	0.7952
	V3 (200 cycles)	Proposed	**2.6086**	**0.8048**	**0.9127**	**0.9183**	**0.9306**	**0.6706**	**0.9560**	**0.9357**	**0.9241**	**0.9265**
		IRXGBoost	4.5404	0.4341	0.6695	0.6873	0.8511	0.4214	0.8448	0.7692	0.7069	0.5799
		DuST	3.3400	0.6792	0.8259	0.8236	0.9274	0.5796	0.9237	0.8998	0.8800	0.8075
		IGHF	3.4583	0.7009	0.8859	0.8931	0.9250	0.6515	0.9474	0.9239	0.9045	0.8621
		Cloud LSTM	4.5549	0.4700	0.7227	0.7256	0.9155	0.4362	0.8964	0.8484	0.8125	0.7381

IRXGBoost, which relies on a shallow tree-based model, shows comparatively lower accuracy across most test cases. This observation suggests that capturing the complex spatial–temporal relationships underlying IR-drop behavior may require higher-capacity models with stronger representation learning capabilities.

Cloud LSTM, adapted from point-cloud time-series forecasting, serves as a relevant baseline for comparison. While it provides reasonable performance, it generally underperforms the proposed LWFusor across the evaluated benchmarks. This result indicates that directly applying generic point-cloud forecasting models to IR-drop prediction may be insufficient, and that task-specific architectural designs informed by physical characteristics are important for achieving high accuracy.

Figure 6 visualizes the worst-case dynamic IR-drop distributions over the chip layout, together with the corresponding predictions from different methods. The results are shown for the largest design TD4 under a 200-cycle waveform. The predictions produced by the proposed method exhibit a high degree of consistency with the ground-truth results. In addition to matching the overall spatial

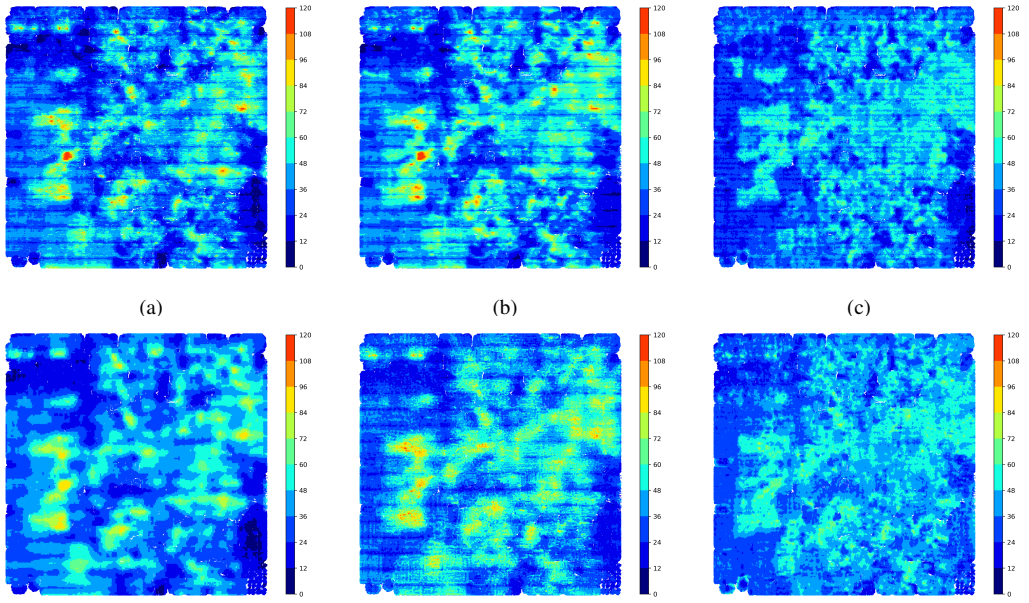


Figure 6: Worst dynamic IR-drop distribution over chip layout. (a) The ground truth result from Voltus; (b) The prediction result from the proposed LWFusor; (c) The prediction result from IRXGBoost; (d) The prediction result from DuST; (e) The prediction result from IGHF; (f) The prediction result from Cloud LSTM.

distribution, the proposed approach accurately identifies most of the critical IR-drop hotspot regions, which are particularly important for sign-off analysis.

The cell-based method IGHF also demonstrates reasonable agreement with the ground truth in hotspot localization, supporting the importance of cell-level representations for capturing localized power integrity issues. In contrast, the image-based method DuST shows noticeable discrepancies in several hotspot regions. These deviations are likely related to the inherent resolution constraints introduced by rasterization, which can limit the representation of fine-grained spatial variations in large layouts. Cloud LSTM and IRXGBoost produce comparatively conservative IR-drop patterns with less precise hotspot localization. This behavior suggests that models not explicitly designed to incorporate physical layout structure may have difficulty capturing the localized extrema that dominate worst-case IR-drop behavior.

Overall, the comparison highlights the effectiveness of the proposed LWFusor in preserving spatial detail while accurately modeling worst-case IR-drop characteristics across large-scale designs.

C Ablation Study

In this section, ablation studies are conducted to validate the effectiveness of each proposed component of the proposed LWFusor. The dataset and evaluation metrics are the same as Table 2 and the performance is averaged over all test cases.

C.1 Ablation Study on Multi-modal Fusion

We first conduct an ablation study on the multimodal fusion mechanism, as summarized in Table 6. Specifically, the layout modality and the waveform modality are removed independently to assess their individual contributions. When only the layout modality is retained, performance consistently decreases across all evaluation metrics. This result suggests that switching waveforms provide essential temporal cues that cannot be inferred from static physical information alone. Conversely, when only the waveform modality is used, static cell features are removed, and only spatial coordinates (x,y) are preserved to enable model inference. Under this setting, a noticeable performance degradation is also observed, indicating that detailed layout features play a critical role in characterizing the spatial heterogeneity of IR-drop behavior.

Table 6: Ablation study on layout and waveform modalities.

Modality		Regression				Critical cell identification			Risk prioritizing (NDCG)		
Layout	Waveform	MAE↓	R^2↑	PCC↑	Spea↑	Accuracy↑	F1↑	ROC-AUC↑	10%↑	5%↑	1%↑
✓	✓	**2.4778**	**0.8167**	**0.9088**	**0.9164**	**0.9242**	**0.6693**	**0.9503**	**0.9349**	**0.9189**	**0.8956**
✓	×	3.6377	0.5953	0.8014	0.8138	0.8984	0.4623	0.8891	0.8680	0.8345	0.7609
(x, y) only	✓	3.0440	0.7335	0.8710	0.8780	0.9144	0.5743	0.9334	0.9131	0.8933	0.8590

Table 7: Ablation study on multimodal fusion mechanisms.

Fusion model		Regression				Critical cell identification			Risk prioritizing (NDCG)		
FiLM	Masked coAttn	MAE↓	R^2↑	PCC↑	Spea↑	Accuracy↑	F1↑	ROC-AUC↑	10%↑	5%↑	1%↑
✓	✓	**2.4778**	**0.8167**	**0.9088**	**0.9164**	0.9242	**0.6693**	**0.9503**	**0.9349**	**0.9189**	**0.8956**
✓	×	2.8631	0.7538	0.8865	0.8983	0.9201	0.6469	0.9446	0.9270	0.9112	0.8865
×	✓	2.6225	0.8045	0.9018	0.9071	**0.9251**	0.6595	0.9484	0.9325	0.9166	0.8826
×	×	2.9940	0.7498	0.8893	0.8964	0.9191	0.6387	0.9431	0.9250	0.9102	0.8843

Table 8: Ablation study on physically informed aggregation and sampling.

kNN&sampling method		Regression				Critical cell identification			Risk prioritizing (NDCG)		
rkNN	VPS	MAE↓	R^2↑	PCC↑	Spea↑	Accuracy↑	F1↑	ROC-AUC↑	10%↑	5%↑	1%↑
✓	✓	**2.4778**	**0.8167**	**0.9088**	**0.9164**	**0.9242**	**0.6693**	**0.9503**	**0.9349**	**0.9189**	**0.8956**
×	✓	2.9469	0.7534	0.8733	0.8775	0.9160	0.6093	0.9393	0.9142	0.8964	0.8693
×	×	3.0474	0.7423	0.8690	0.8714	0.9145	0.6281	0.9371	0.9173	0.9007	0.8727

Table 9: Impact of aggregation and interpolation range on performance.

rkNN-k	GPU Memory	Regression				Critical cell identification			Risk prioritizing (NDCG)		
		MAE↓	R^2↑	PCC↑	Spea↑	Accuracy↑	F1↑	ROC-AUC↑	10%↑	5%↑	1%↑
150	31350 M	2.5567	0.8075	0.9057	0.9120	**0.9277**	0.5524	0.9497	0.9364	0.9211	0.8946
128	25243 M	**2.4778**	**0.8167**	0.9088	0.9164	0.9242	0.6693	0.9503	0.9349	0.9189	0.8956
64	18557 M	2.6186	0.8068	0.9030	0.9069	0.9270	0.6748	0.9510	0.9376	0.9230	0.8948
32	12599 M	2.5153	0.8135	**0.9093**	**0.9176**	0.9244	**0.6787**	**0.9513**	**0.9386**	**0.9254**	**0.9037**
16	–	2.7102	0.7838	0.9038	0.9084	0.9148	0.6341	0.9497	0.9041	0.8867	0.8629
8	–	2.7619	0.7892	0.8964	0.8989	0.9209	0.6649	0.9485	0.9322	0.9172	0.8890

Overall, these results highlight the complementary nature of spatial layout and temporal waveform information, and validate the necessity of jointly modeling both modalities for accurate dynamic IR-drop prediction.

The effectiveness of the proposed multimodal fusion mechanisms, including FiLM-based modulation and the masked co-attention fusion decoder, is further examined in Table 7. Replacing both components with a simple concatenation-based fusion results in the largest performance degradation, highlighting the limitations of naive feature fusion. Removing either FiLM or the masked co-attention decoder individually generally reduces performance, while retaining either component improves most metrics over direct concatenation. This observation suggests that each component contributes meaningfully to multimodal interaction and that their combination provides complementary benefits for accurate IR-drop prediction. Overall, the ablation results indicate that both FiLM modulation and masked co-attention fusion play important roles in effectively integrating spatial layout and temporal waveform information.

C.2 Ablation Study on Physically Informed Aggregation and Sampling

We further investigate the impact of physically informed design choices in the spatial encoder, including resistance-aware kNN (rkNN) aggregation and power-via-based point sampling (VPS), as summarized in Table 8. Replacing rkNN with a standard Euclidean kNN leads to a consistent performance degradation across all evaluation metrics. This result suggests that incorporating PDN-aware connectivity information is beneficial for identifying electrically relevant neighbors, beyond purely geometric proximity. Similarly, substituting VPS with farthest point sampling also results in a noticeable drop in accuracy. This observation indicates that aligning hierarchical sampling with dominant current flow paths helps the model learn more meaningful multi-scale representations for IR-drop prediction. Overall, these ablation results highlight the importance of embedding physical priors into both neighborhood aggregation and hierarchical sampling when modeling power integrity phenomena.

C.3 Ablation Study on rkNN Spatial Pruning Scope

We further analyze the effect of the spatial pruning scope by varying the number of neighbors k in resistance-weighted kNN (rkNN), as summarized in Table 9. Specifically, we evaluate k from 8 to 150, where $k = 150$ represents the practical upper bound constrained by current GPU memory

Table 10: Ablation study on activity tensor channels.

Waveform channel	Regression				Critical cell identification			Risk prioritizing (NDCG)		
	MAE↓	R^2↑	PCC↑	Spea↑	Accuracy↑	F1↑	ROC-AUC↑	10%↑	5%↑	1%↑
Full	**2.4778**	**0.8167**	**0.9088**	**0.9164**	**0.9242**	**0.6693**	**0.9503**	**0.9349**	**0.9189**	**0.8956**
w/o Δw	2.5659	0.8050	0.9050	0.9148	0.9237	0.6325	0.9494	0.9332	0.9170	0.8937
w/o $\Delta w, \omega$	2.7192	0.7815	0.8929	0.9000	0.9235	0.5915	0.9430	0.9303	0.9165	0.8950

(32GB). Table 9 reveals that the predictive performance follows a non-monotonic trend, reaching an optimization between $k = 32$ and $k = 128$. While increasing k initially improves accuracy by capturing broader spatial contexts, expanding the attention scope beyond $k = 128$ actually leads to performance degradation and increased memory overhead. This suggests that excessively large neighborhoods introduce long-range spatial noise that hinders model convergence. Ultimately, $k = 64$ provides the optimal balance between accuracy and memory efficiency. From a physical perspective, rkNN acts as a resistance-aware filter that focuses on electrically relevant neighbors. In contrast, very small values of k (8 or 16) result in consistent performance degradation across all metrics, confirming that overly restrictive pruning fails to capture necessary electrical coupling effects. This trade-off analysis demonstrates that LWFusor achieves industrial-scale feasibility by maintaining $O(Nk)$ complexity without compromising the integrity.

C.4 Ablation Study on Activity Tensor Channels

The design of the three-channel activity tensor $\mathbf{W}$ is a principled decision aimed at capturing the complex temporal dependencies and localized switching intensities that govern dynamic IR-drop. To evaluate the contribution of each component, we conducted an ablation study by progressively removing the density and flip channels. The experimental results, as detailed in Table 10, confirm that the "Full" three-channel configuration provides the highest predictive fidelity across all evaluated metrics. The removal of the binary flip channel ($\Delta\mathbf{w}_i$) leads to a measurable degradation in all criteria, indicating that the model loses a vital prior for recognizing temporal clusters. A more substantial performance drop occurs when the local switching density (ω_i) is also excluded. Without this channel, the model struggles to distinguish between static logic levels and active switching power, resulting in a significant performance decrease. These findings empirically validate that our multi-channel activity tensor effectively preserves the essential temporal features required for robust dynamic IR-drop assessment, ensuring the model remains sensitive to the hazardous switching bursts that drive functional failures.

D Expanded Formulation of Methodology

D.1 Resistance Weighted k-Nearest Neighbors

In typical machine learning tasks, k-nearest neighbors (kNN) are applied using the Euclidean or Manhattan distance, implicitly assuming an isotropic and homogeneous spatial manifold. However, in modern VLSI designs, on-chip power delivery networks (PDNs) are inherently anisotropic and spatially heterogeneous due to: (i) Directional Routing, where horizontal and vertical metal layers possess distinct dimensions and current-carrying capacities; and (ii) Resource Heterogeneity, where power stripes and VDD bumps are non-uniformly distributed. Consequently, two standard cells that are geometrically proximate may be electrically distant if separated by high-impedance PDN bottlenecks.

Calculating the exact transfer impedance between any two points in a billion-node PDN requires inverting a massive, ill-conditioned conductance matrix $\mathbf{G}$, scaling at worst $O(N^3)$. This numerical inversion is computationally prohibitive for dynamic neighborhood searches in large-scale ML frameworks. To balance physical fidelity with the extreme efficiency required for graph construction, we propose a Resistance-Weighted kNN (rKNN). We approximate the effective electrical proximity by deriving an analytic distance metric from a localized quasi-3D homogenization theory, serving as a tractable surrogate for PDN-constrained current propagation.

Instead of treating the PDN as discrete metal segments, we model the multi-layer PDN surrounding a target cell c_i as an anisotropic, locally stationary conductive medium. The physical topology of the PDN is instantiated from the Design Exchange Format (DEF) file, which specifies the spatial bounding boxes of metal stripes and vias, along with the square resistance $R_{sq,L}$ for each layer L and the via resistance $R_{via,l}$ between adjacent layers. To map these discrete geometric constraints into continuous physical parameters, we define a local spatial sensing window W_i^{PDN} centered at the coordinates $\mathbf{x}_i$ of the cell. To bridge the gap between discrete routing geometries and a

continuous physical manifold, we homogenize the metal stripes into a directional sheet conductance field $\sigma_{L,d}(\mathbf{x}_i)$. For each metal layer L and routing orientation $d \in \{x, y\}$, the effective conductance is governed by the local metal utilization ratio $D_{L,d}(\mathbf{x}_i)$, which represents the density of available conductive paths. Specifically, within a sensing window W_i^{PDN} centered at $\mathbf{x}_i$, we compute $D_{L,d}$ as the area-weighted sum of all stripes $\mathcal{S}_{L,d}$ intersecting the window. Following the principle of parallel resistive aggregation, the homogenized sheet conductance is formulated as Equation (12):

$$\sigma_{L,d}(\mathbf{x}_i) = \frac{D_{L,d}(\mathbf{x}_i)}{R_{squa,L}}, \quad \text{where} \quad D_{L,d}(\mathbf{x}_i) = \frac{\sum_{s \in \mathcal{S}_{L,d} \cap W_i^{PDN}} \text{Area}(s)}{|W_i^{PDN}|} \tag{12}$$

where $|W_i^{PDN}|$ denotes the total area of the sensing window, and $R_{squa,L}$ is the intrinsic square resistance of layer L. $D_{L,d}(\mathbf{x}_i) \in [0, 1]$ represents the local metal utilization ratio. This formulation represents the effective sheet conductance along direction d. Consequently, regions with fragmented or sparse PDN stripes are naturally mapped to low-conductance zones, correctly capturing the resistive bottlenecks in the underlying design.

The vertical coupling between adjacent routing layers is quantified by the local via conductance within the sensing window, as shown in Equation (13):

$$g_{v,l}(\mathbf{x}_i) = \frac{N_{v,l}}{R_{via,l}} \tag{13}$$

where $N_{v,l}$ denotes the exact count of vias connecting layer l and $l+1$ within the given window. The effective conductance tensor $\mathbf{\Sigma}_i = \text{diag}(\Sigma_{xx}^{(i)}, \Sigma_{yy}^{(i)})$ perceived by the bottom standard cell layer is an aggregation of all metal layers, heavily constrained by the vertical impedance of the via network. Treating the multi-layer PDN as a vertically distributed resistive ladder, the current penetration into upper layers suffers from a characteristic voltage decay. The contribution of layer L attenuates exponentially with an accumulated dimensionless inter-layer attenuation term. We formalize this mapping in Equation (14):

$$\Sigma_d^{(i)} = \sum_L \sigma_{L,d}(\mathbf{x}_i) \exp\left(- \sum_{l=1}^{L-1} \sqrt{\frac{\bar{\sigma}_l(\mathbf{x}_i)}{g_{v,l}(\mathbf{x}_i)}} \right), \quad d \in x, y \tag{14}$$

where $\bar{\sigma}_l = \sqrt{\sigma_{l,x}\sigma_{l,y}}$ represents the geometric mean of the horizontal conductance at layer l. This dynamically penalizes the utilization of higher-layer metals in regions with sparse via distributions or inherently high via resistances, adhering to the local physical structural limitations provided by the DEF geometry. The governing physical behavior of the voltage drop diffusion on this homogenized 2D medium is described by the anisotropic screened Poisson equation, as shown in Equation (15):

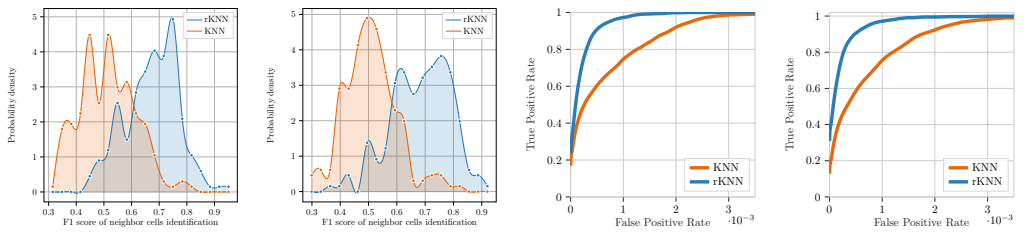
$$\nabla \cdot (\mathbf{\Sigma}_i \nabla \Phi) - G_{s,i} \quad \Phi = -I\delta(\mathbf{x} - \mathbf{x}_i) \tag{15}$$

where $G_{s,i}$ is the local equivalent conductance to the ideal voltage sources, directly extracted from the top-layer power connection density within W_i^{PDN}. By extracting the argument of this asymptotic solution, we define the physics-informed resistance-weighted distance metric d_r as the effective scaled distance in the mapped space, as shown in Equation (16):

$$d_r(c_i, c_j) = \kappa_i \sqrt{(x_j - x_i)^2 \frac{\Sigma_{yy}^{(i)}}{\bar{\Sigma}_i} + (y_j - y_i)^2 \frac{\Sigma_{xx}^{(i)}}{\bar{\Sigma}_i}} \tag{16}$$

where $\bar{\Sigma}_i = \sqrt{\Sigma_{xx}^{(i)}\Sigma_{yy}^{(i)}}$ normalizes the tensor magnitude, and $\kappa_i = \sqrt{G_{s,i}/\bar{\Sigma}_i}$ acts as the local screening constant. This parameter κ_i captures the PDN density constraints: in regions with dense PDN, κ_i increases, mapping geometrically close nodes to a larger electrical distance d_r and effectively confining the IR-drop receptive field to prevent erroneous over-aggregation.

The proposed metric d_r is not intended to replace rigorous numerical sign-off resistance extraction or full-chip SPICE simulations. Instead, it serves exclusively as an efficient topological similarity metric for feature aggregation. The underlying homogenization is applied strictly within a local spatial window W_i^{PDN}, where the PDN geometry is sufficiently dense and statistically stationary, ensuring that the continuous approximation remains physically meaningful even in the presence of global non-uniformities such as macro blockages or irregular PDN. Unlike $\mathbf{G}^{-1}$ based solvers that require global matrix inversions, d_r is analytically evaluated in nearly $O(1)$ time given the local DEF features. In practice, the metal density $D_{L,d}$ and via counts $N_{v,l}$ are efficiently obtained using spatial



(a) (b) (c) (d)

Figure 7: The accuracy comparison of neighbor identification by the Manhattan distance-based kNN and the proposed resistance-weighted rkNN (a) F1 score on uniform PDN; (b) F1 score on irregular PDN; (c) ROC curve on uniform PDN; (d) ROC curve on irregular PDN.

indexing structures, enabling constant-time queries per cell after a one-time linear preprocessing pass over the layout. This ensures overall near-linear scalability for large designs.

The neighbor indices $\mathbf{C}_{\mathcal{N},i}$ for each cell c_i are obtained by selecting the k cells with the minimum resistance-weighted distance, as shown in Equation (17):

$$\begin{aligned} \mathcal{I}_i &= \text{argsort}[d_r(c_i, c_j) \mid c_j \in \mathbf{C}], \\ \mathbf{C}_{\mathcal{N},i} &= \mathcal{I}_i[1:k]. \end{aligned} \tag{17}$$

Calculating the resistive distance d_r between each cell and all candidates yields $O(N^2)$ complexity. To ensure scalability, we implement a spatial pruning strategy. Given that d_r evaluates a scaled semi-positive definite quadratic form, it establishes a mathematically rigorous lower bound with respect to the Euclidean distance. For any target cell c_i, its k electrically-closest neighbors are strictly bounded within a local geometric bounding box $\mathcal{B}_i$ defined by a maximum resistive threshold t_r, as shown in Equation (18):

$$\mathbf{C}_{\mathcal{N},i} \subseteq \{c_j \mid d_r(c_i, c_j) \leq t_r\} \subseteq \mathcal{B}_i. \tag{18}$$

By locally determining t_r based on κ_i and the minimum eigenvalue of $\mathbf{\Sigma}_i$, we safely prune the vast majority of the chip area from the search space without excluding any true k-nearest neighbors under the conservative lower-bound constraint. Through this spatial pruning, the neighbor search complexity collapses from $O(N^2)$ to near-linear $O(N \log k)$, maintaining the extreme efficiency of our framework.

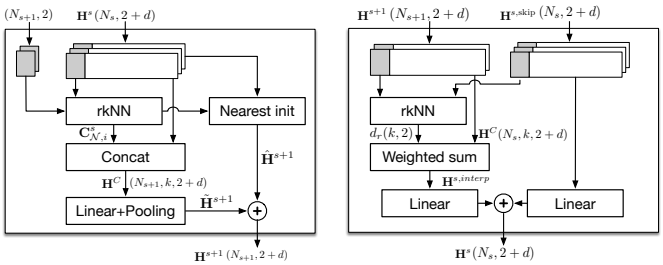
To evaluate the fidelity of the proposed d_r metric in capturing true electrical coupling, we conduct a comparative validation against the Manhattan distance-based kNN baseline. Ground truth neighborhoods are established via SPICE-level simulations. For a given source cell, we inject a unit current and measure the steady-state voltage drop at all other cell pins. A cell is classified as a ground truth neighbor if its pin voltage drop is greater than 10% of the source voltage, indicating a significant resistive coupling and susceptibility to IR-drop interference. We evaluate the neighbor identification performance across two distinct scenarios: a strictly uniform PDN and a highly irregular PDN with variant stripe densities.

Figure 7(a) and Figure 7(b) illustrate the probability density of F1 scores for neighbor identification across hundreds of sampled instances. In both scenarios, the proposed rKNN distribution exhibits a pronounced rightward shift compared to the baseline Manhattan distance-based kNN, demonstrating that the physics-informed metric d_r successfully aligns the model's receptive field with the true resistive paths of the PDN. In contrast, the baseline KNN often fails to capture electrically relevant neighbors, with its F1 distribution peaking at significantly lower values. The performance advantage of rKNN is further underscored by the ROC curves shown in Figure 7(c) and Figure 7(d). The rKNN achieves a substantially higher true positive rate at low false positive rate (FPR) levels.

Even under the resource heterogeneity of the irregular PDN (Figure 7(b) and Figure 7(d)), the d_r metric remains highly effective in identifying impactful neighbors. These results confirm that d_r provides a physically-consistent foundation for feature aggregation, ensuring that the model remains accurate regardless of the underlying PDN's uniformity.

D.2 Hierarchical Aggregation and Interpolation

Aggregation. The aggregation block in Figure 8(a) abstracts fine-grained cell-level features $\mathbf{H}^s$ into a coarser representation $\mathbf{H}^{s+1}$ anchored by power via nodes, a process that mirrors the physical



(a) (b)

Figure 8: (a) The aggregation block; (b) The interpolation block.

confluence of currents from local standard cells to global power trunks. For each via anchor i at level $s+1$, we first define its electrical attraction domain through the rkNN module, which identifies a set of neighbors $\mathbf{C}^{s}_{\mathcal{N},i}$ from the finer level s. This domain ensures that the information flow respects the anisotropic electrical reachability of the underlying PDN. The features of these neighbors are concatenated with relative positional information to form a neighborhood physical context $\mathbf{H}^{C}_{i}$. We define the confluence operator Φ_{agg} as a learned non-linear mapping followed by a permutation-invariant pooling. Let $\mathbf{e}_{ij} = \phi_{pos}(d_r(c_i, c_j))$ be the resistive distance embedding; the aggregated feature $\tilde{\mathbf{h}}^{s+1}_{i} = \Phi_{agg}(\mathbf{h}^{s}_{j}, \mathbf{e}_{ij})$ is computed in Equation (19):

$$\tilde{\mathbf{h}}^{s+1}_{i} = \text{MAX}_{j \in \mathbf{C}^{s}_{\mathcal{N},i}} \left(\sigma \left(\mathbf{W}_{agg}[\mathbf{h}^{s}_{j} \parallel \mathbf{e}_{ij}] + \mathbf{b}_{agg} \right) \right), \tag{19}$$

where $\mathbf{W}_{agg}$ and $\mathbf{b}_{agg}$ are the learnable weight matrix and bias vector of the operator, σ denotes the activation function, and MAX represents element-wise max-pooling over the k neighbors. To maintain the stability, the final state is refined via a residual update as shown in Equation (20):

$$\mathbf{H}^{s+1} = \tilde{\mathbf{H}}^{s+1} + \hat{\mathbf{H}}^{s+1}, \tag{20}$$

where $\hat{\mathbf{H}}^{s+1}$ is initialized by the nearest-neighbor feature to provide a robust physical baseline.

Interpolation. The interpolation block in Figure 8(b) restores spatial granularity by upsampling coarse via-level features $\mathbf{H}^{s+1}$ back to the fine cell-level resolution $\mathbf{H}^{s}$, simulating the distribution of supply voltage from the global grid to individual logic gates. For a target cell i at level s, its relative IR-drop is significantly influenced by its surrounding via anchors in the coarser level $s+1$. We model this influence as a conductance-weighted attribution process, where the contribution of each via anchor $j \in \mathbf{C}^{s+1}_{\mathcal{N},i}$ is inversely proportional to the resistive distance d_{ij}.

Specifically, we compute normalized interpolation weights α_{ij} based on the resistance-weighted distance, mirroring the additive nature of conductance in a resistive network, as shown in Equation (21):

$$\alpha_{ij} = \frac{1/d_{ij}}{\sum_{m \in \mathbf{C}^{s+1}_{\mathcal{N},i}} (1/d_{im})}. \tag{21}$$

The upsampled potential context $\mathbf{h}^{s,\text{interp}}_{i}$ is then derived through a weighted summation in Equation (22):

$$\mathbf{h}^{s,\text{interp}}_{i} = \mathbf{W}_{up} \left(\sum_{j \in \mathbf{C}^{s+1}_{\mathcal{N},i}} \alpha_{ij} \mathbf{h}^{s+1}_{j} \right), \tag{22}$$

where $\mathbf{W}_{up}$ is a trainable projection matrix that transforms the aggregated anchor context into the target cell's feature space. This interpolation mechanism ensures that anchors with lower resistive paths exert a dominant influence on the cell's potential state. To account for both the global IR-drop trend and local transient switching fluctuations, the framework employs a dual-branch fusion strategy. The final output $\mathbf{H}^{s}$ is obtained by integrating the upsampled context with the skipped high-resolution features:

$$\mathbf{H}^{s} = \mathbf{H}^{s,\text{interp}} + \mathbf{H}^{s,\text{skip}}. \tag{23}$$

This multi-scale integration allows the model to capture high-frequency spatial details while simultaneously benefiting from the expanded receptive field of the coarser levels.

Table 11: Sensitivity Analysis on Patch Size.

Patch size	Regression				Critical cell identification			Risk prioritizing (NDCG)		
($B_x = B_y$)	MAE↓	R^2↑	PCC↑	Spea↑	Accuracy↑	F1↑	ROC-AUC↑	10%↑	5%↑	1%↑
120	**2.4778**	**0.8167**	**0.9088**	**0.9164**	**0.9242**	**0.6693**	**0.9503**	**0.9349**	**0.9189**	**0.8956**
80	2.6066	0.7975	0.8998	0.9049	0.9220	0.6666	0.9465	0.9324	0.9168	0.8947
60	2.7733	0.7711	0.8879	0.8912	0.9180	0.6662	0.9399	0.9279	0.9131	0.8930
40	2.9899	0.7356	0.8746	0.8752	0.9131	0.6130	0.9356	0.9243	0.9087	0.8860

D.3 Layout Patching and Spatial Weighted Fusion

Since the millions of cell-scale layouts are too large to fit into GPU memory, we add a patching strategy to sample sub-regions from the full layout. We specify a patch size of (B_x, B_y) and an overlap ratio $\lambda \in [0, 1)$. The strides S_x and S_y for the sliding window are computed as Equation (24):

$$S_x = B_x(1 - \lambda), \quad S_y = B_y(1 - \lambda). \tag{24}$$

The patches are generated by sliding a window across the chip layout domain starting from $(X_{\min}, Y_{\min})$. The top-left coordinates (x^u, y^v) for the patches are given by a discrete grid in Equation (25):

$$\begin{aligned} x^u &= X_{\min} + u \cdot S_x, \quad u = 0, 1, \ldots, \left\lfloor \frac{X_{\max} - X_{\min}}{S_x} \right\rfloor; \\ y^v &= Y_{\min} + v \cdot S_y, \quad v = 0, 1, \ldots, \left\lfloor \frac{Y_{\max} - Y_{\min}}{S_y} \right\rfloor. \end{aligned} \tag{25}$$

To mitigate the edge effects introduced by partitioning, we employ a spatial weighted fusion strategy during inference. For each patch p covering the rectangular region $[x^p_{\min}, x^p_{\max}] \times [y^p_{\min}, y^p_{\max}]$, we define a confidence weight that decays linearly from the patch center to its boundaries. Given a margin ratio η, the margin widths along the x and y axes are defined in Equation (26):

$$m_x = \eta B_x, \quad m_y = \eta B_y. \tag{26}$$

For a point i with coordinates (x_i, y_i) within patch p, we calculate its minimum distance to the patch boundaries in Equation (27):

$$\begin{aligned} d^p_{x,i} &= \min(x_i - x^p_{\min}, x^p_{\max} - x_i), \\ d^p_{y,i} &= \min(y_i - y^p_{\min}, y^p_{\max} - y_i). \end{aligned} \tag{27}$$

The spatial weight sw^p_i represents the reliability of the prediction for point i within patch p. It is computed by normalizing the boundary distances by the margin widths in Equation (28):

$$sw^p_i = \text{clip}\left(\min\left(\frac{d^p_{x,i}}{m_x}, \frac{d^p_{y,i}}{m_y} \right), 0, 1 \right). \tag{28}$$

This function assigns a weight of 1.0 to points in the central core region and linearly decays to 0 at the boundaries. The final global prediction $\hat{y}_i$ for point i is obtained by fusing the local predictions $\hat{y}^p_i$ from all overlapping patches $p \in \mathbf{P}_i$ that cover point i in Equation (29):

$$\hat{y}_i = \frac{\sum_{p \in \mathbf{P}_i} sw^p_i \hat{y}^p_i}{\sum_{p \in \mathbf{P}_i} sw^p_i}. \tag{29}$$

The choice of patch size is a trade-off between the model's receptive field and hardware memory constraints. We evaluated patch sizes from 40 to 120 units, as summarized in Table 11. In the tested range, increasing patch size improves all reported metrics, as larger patches capture broader spatial correlations within the PDN. We recommend using the largest tested patch size that fits within the target GPU's memory (e.g., patch-120 for 32GB+ GPUs). Smaller patches reduce memory usage but can truncate spatial interactions and reduce predictive accuracy.

E Future Work

While LWFusor provides a scalable and high-fidelity solution for monolithic industrial designs, it currently focuses on 2D and 2.5D physical layouts. Future extensions could adapt our hierarchical via point sampling and resistance-weighted kNN topology to 3D Integrated Circuits (3DIC) to model vertical electrical dependencies such as Through-Silicon Vias (TSVs). Furthermore, incorporating package-level parasitics (for example, the package inductance) or multi-chiplet spatial-temporal modeling would allow the framework to capture macro-scale power noise and inter-die interference that currently fall outside the die-level PDN analysis. These advancements would enable a more comprehensive power integrity assessment for heterogeneous and vertically integrated systems.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and Section 1 state the paper's contributions and scope, which are evaluated in Section 5.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The paper discusses the limitations of the work in Appendix E.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [N/A]

   Justification: The paper does not include theoretical results.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The model and experimental procedures are described in Sections 3 and 4, and Appendices B and D.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: We will open-source our code and data after publication.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The data splits and training settings are described in Appendices B.1 and B.3.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: The paper does not need to provide the statistical significance of the experiments.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.

- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The compute hardware is described in Appendix B.3, and runtime measurements are reported in Section 5.2.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The research conducted in the paper conforms, in every respect, with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: There is no societal impact of the work performed.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.

- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper poses no such risks.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The original owners of assets, used in the paper, are properly credited.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The paper does not release new assets.

    Guidelines:

- The answer [N/A] means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: The paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

    Answer: [N/A]

    Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

    Guidelines:

    - The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.