GIF: A CONDITIONAL MULTIMODAL GENERATIVE FRAMEWORK FOR IR DROP IMAGING IN CHIP LAYOUTS

Anonymous authors
Paper under double-blind review

ABSTRACT

IR drop analysis is a critical step in physical chip design, ensuring the power integrity of on-chip power delivery networks (PDNs). As transistor density scales, the computational cost of IR drop analysis grows substantially, motivating scalable ML-based alternatives. Existing ML approaches formulate IR drop analysis as a spatial regression problem using power map representations, which do not jointly model the geometric and topological factors that govern voltage drop distribution across the layout. We propose GIF (Generative IR drop Framework), which reformulates IR drop analysis as a conditional image generation problem. GIF conditions a diffusion model on two complementary inputs: a 34-channel geometry-aware feature image capturing local power consumption, cell density, and PDN resistance factors; and graph tokens encoding logical connectivity from the gate-level netlist, injected via gated cross-attention. This allows GIF to jointly capture local geometric variation and long-range structural dependencies that existing IR drop methods do not model. On CircuitNet-N28 (28 nm), GIF achieves 0.79 SSIM and 0.95 Pearson correlation, and detects the top-10% IR drop hotspots with 0.80 precision. As a generative model, GIF also provides a per-pixel uncertainty map from the variance across generated samples. This uncertainty reaches 91% of the oracle in AUSE, compared with 86% for an ensemble of regressors and for MC-dropout, and is 1.6× higher on missed hotspots, so unreliable regions can be targeted for re-simulation. On the more diverse CircuitNet-N14 (14 nm), GIF achieves 0.91 Pearson and 0.83 Spearman correlation across diverse design styles and establishing baselines for future work at this technology node. Codes are available at [1]

1 INTRODUCTION

IR drop analysis is a fundamental step in physical chip design. As current flows through the on-chip power delivery network (PDN), resistive voltage drops degrade circuit timing and reliability, making accurate IR drop analysis essential to design closure Chen & Ling (1997); Nassif (2008). As technology nodes shrink and transistor density increases, the cost of full IR drop analysis grows with layout complexity, creating a bottleneck in design iteration Zhao et al. (2024); Jiang et al. (2024); Chai et al. (2023). ML-based methods have therefore been explored as a scalable alternative Chhabria et al. (2020); Xie et al. (2020b); Fang et al. (2018); Ho & Kahng (2019). Existing ML approaches treat IR drop analysis as a spatial regression problem, mapping tile-based power maps directly to IR drop values. Power maps capture current demand but omit the geometric and routing factors that determine effective PDN resistance, including routing congestion, cell placement density, and macro blockage, which vary spatially and independently of power consumption. Regions with similar power can exhibit substantially different IR drop depending on how current paths distribute through the PDN Zheng et al. (2023). Beyond feature limitations, existing approaches model either spatial layout features or circuit connectivity, but not both in a unified representation Chhabria et al. (2021); Zhao et al. (2024). A further limitation is that all existing methods produce a single deterministic IR drop map per layout, offering no indication of prediction confidence across layout regions. The core insight of this work is that IR drop analysis

[1] https://anonymous.4open.science/r/GIF-EFE2

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

is a conditional generation problem, not a regression problem. The IR drop distribution at each layout tile depends jointly on local geometric context and on global topological structure from the logical netlist, which creates correlated current demand across the layout. Learning this conditional distribution, rather than a point estimate, enables complementary geometric and topological signals to jointly shape the generation process, and provides spatial uncertainty estimates by construction. We propose GIF, a conditional diffusion framework that realizes this formulation. GIF conditions the generative process on two complementary inputs derived from the physical design: a geometry-aware feature image that captures local current demand and PDN resistance variation, and graph tokens from the gate-level netlist that encode global logical connectivity. The two inputs condition the generative process at different spatial scales, matching the physical mechanisms through which each signal influences IR drop distribution. Table 1 summarizes how GIF differs from prior methods.

Table 1: Comparison of IR drop learning approaches.

Method	Layout Power Maps	Geometry-aware Layout Features	Logical Topology	Generative Modeling
PowerNet Xie et al. (2020b)	✓	✗	✗	✗
MAVIREC Chhabria et al. (2021)	✓	✗	✗	✗
PDNNet Zhao et al. (2024)	✓	✗	✓	✗
GIF (ours)	✓	✓	✓	✓

Our contributions are:

1. **Physics-grounded geometry-aware feature construction.** We construct a 34-channel feature image grounded in the physical decomposition of IR drop into current demand and effective PDN resistance. Ten additional layout-aware channels extend conventional power maps with routing congestion and placement geometry features, each independently correlated with IR drop ground truth.
2. **Netlist graph for topology-aware conditioning.** We build a cell-level graph from the gate-level netlist where node features encode instance placement geometry and edges encode logical connectivity. A lightweight graph encoder produces a fixed set of graph tokens, enabling topology-aware conditioning that scales independently of design size.
3. **Multimodal conditioning for IR drop generation.** We fuse geometry features and netlist graph tokens into the diffusion process through complementary conditioning pathways: geometry via adaptive feature normalisation at every resolution, graph tokens via gated cross-attention at the coarsest resolution, matching the physical scales at which each signal influences IR drop distribution.
4. **Uncertainty for hotspot screening.** The variance across generated samples gives a per-pixel uncertainty map without training additional models. This uncertainty ranks errors better than an ensemble of regressors and MC-dropout, and it is 1.6× higher on missed hotspots, so engineers can target re-simulation to unreliable regions.

2 Background and Related Works

CNN-based IR-drop models. CNN-based learning IR-drop prediction is typically formulated as an image-to-image regression problem. Additional background on chip design flow and IR-drop analysis is provided in the supplementary material 6.1. PowerNet Xie et al. (2020b) employs a U-Net to map layout features to IR-drop images, MAVIREC Chhabria et al. (2021) improves hotspot localization, and PDNNet Zhao et al. (2024) integrates CNN-GNN spatial modeling. These models rely solely on image features and struggle to capture long-range dependencies in power-delivery networks.

GNN-based chip-design models. GNNs are widely applied in physical-design tasks to represent logical connectivity and structural relationships Yang et al. (2022); Thorat et al. (2025). Despite their relevance to global PDN behavior, existing IR-drop models do not incorporate topological information from the netlist. While GNNs capture netlist structure effectively, most existing approaches apply them to node-level voltage estimation or global PDN analysis rather than generating full-resolution IR-drop maps.

Transformers and generative models. Transformers capture global spatial interactions Dosovitskiy (2020); Liu et al. (2021), and layout-specific variants such as Lay-Net Zheng et al. (2023) demonstrate advantages for chip layout for congestion prediction. Generative models including VAEs Kingma

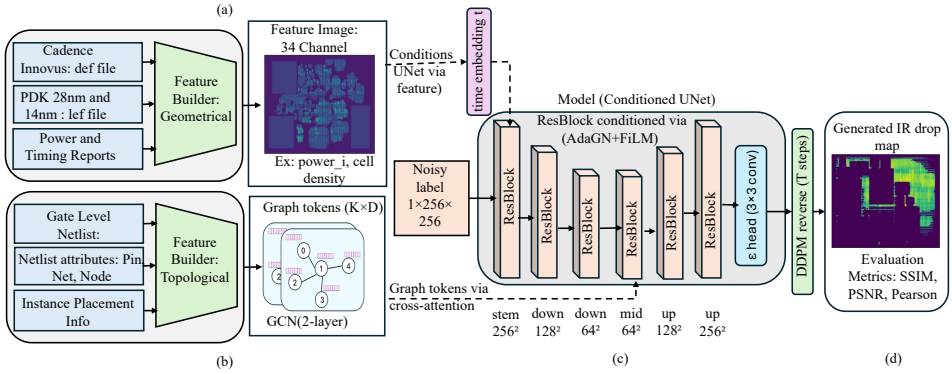


Figure 1: Overview of the proposed framework. (a) Geometric features creation from DEF/LEF files and power reports, (b) Topological features creation, and (c) A diffusion-based UNet predicts the noise ϵ, conditioned on features via AdaGN+FiLM and on graph tokens via gated cross-attention, (d) Generated IR drop map.

& Welling (2013), denoising diffusion models Ho et al. (2020), and latent diffusion Rombach et al. (2022) have achieved best performance in image synthesis. However, prior IR-drop work remains deterministic and does not exploit generative modeling or multimodal conditioning. This work introduces the first generative formulation for IR-drop, filling a gap in prior methods.

2.1 Problem Formulation

Given a placed digital design, let $\mathbf{X}_p \in \mathbb{R}^{H\times W\times C_p}$ denote the set of physical layout features, including power-map features and geometry-aware quantities such as cell density, RUDY, and overflow. Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ denote the topology-aware graph constructed from the netlist, where each node $v \in \mathcal{V}$ represents a cell instance and edges $e \in \mathcal{E}$ represent logical connectivity. We define the mapping

$$f_\theta : (\mathbf{X}_p, \mathcal{G}) \to \mathbf{Y}, \tag{1}$$

where $\mathbf{Y} \in \mathbb{R}^{H\times W}$ is the IR-drop map. In the generative formulation, we instead model the conditional distribution

$$p_\theta\left(\mathbf{Y} \mid \mathbf{X}_p, \mathcal{G}\right), \tag{2}$$

and generate samples of $\mathbf{Y}$ by drawing from this distribution. The problem is therefore to estimate the conditional distribution $p_\theta\left(\mathbf{Y} \mid \mathbf{X}_p, \mathcal{G}\right)$ and generate IR-drop maps that match the ground-truth distribution.

3 Framework

Figure 1 shows an overview of GIF. Two feature-builder modules produce the conditioning inputs: a geometry-aware feature image from DEF/LEF files and power reports (Section 3.1), and a fixed set of graph tokens from the gate-level netlist (Section 3.2). Both condition a diffusion denoising network that generates the IR drop map from Gaussian noise (Section 3.3).

3.1 Construction of Geometry-Aware Feature Maps

IR drop analysis simulates the voltage drop caused by resistive current flow through the on-chip power delivery network (PDN) Zhong & Wong (2005). The PDN is governed by:

$$G\,v = i, \tag{3}$$

where G is the conductance matrix, v is the node voltage, and i is the injected current Köse & Friedman (2011). Solving Eq. 3 at full resolution is computationally expensive, motivating tile-based ML approximations Chhabria et al. (2021).

Existing ML methods represent each layout tile using power maps derived from power and timing reports. Power reports provide instance-level internal power (p_i), switching power (p_s), leakage power (p_l), and toggle rate (r_{togg}), aggregated per tile into four static channels:

$$p_{\text{i}} \propto p_i, \qquad p_{\text{s}} \propto p_s, \tag{4}$$

$$p_{\text{sca}} \propto (p_s + p_i) \cdot r_{\text{togg}} + p_l, \tag{5}$$

$$p_{\text{all}} \propto p_s + p_i + p_l. \tag{6}$$

Timing window reports record switching intervals within each clock period from static timing analysis. The clock period is divided into 20 equal windows; each cell contributes to the time-decomposed power map p_t only during its active windows:

$$p_t[k] \propto p_{\text{sca}}, \quad k = 0, \ldots, 19. \tag{7}$$

These yield a 24-channel feature image per layout: four static components and 20 temporal components. Power maps encode current demand but not PDN resistance. From Eq. 3, voltage drop depends on both injected current i and the conductance structure G, which is shaped by routing congestion, cell placement, and macro blockage. Two tiles with identical power can exhibit substantially different IR drop if their PDN paths differ in effective resistance. Methods relying solely on power maps Chhabria et al. (2021); Zhao et al. (2024) cannot distinguish these cases.

Group	Channels	I_{eff}	R_{eff}
Cell density	C_{den}	✓	✗
Macro region	M	✓	✗
RUDY demand	RUDY, RUDY_{pin}	✓	✗
RUDY demand	$\text{RUDY}_{\text{long}}$, $\text{RUDY}_{\text{short}}$	✓	✗
Routing overflow	O_H^{eGR}, O_V^{eGR}	✗	✓
Routing overflow	O_H^{GR}, O_V^{GR}	✗	✓

Table 2: Layout-aware channel groups and their physical role (Eq. 8).

Additional layout-aware channels. We augment the 24-channel representation with ten additional channels that encode PDN resistance variation alongside current demand. The motivation follows directly from Eq. 3: voltage drop at each tile is determined jointly by local current demand and effective PDN resistance:

$$V_{\text{drop}}(x, y) \approx R_{\text{eff}}(x, y)\, I_{\text{eff}}(x, y), \tag{8}$$

where the current demand proxy I_{eff} and resistance proxy R_{eff} are:

$$I_{\text{eff}}(x, y) = \frac{P(x, y)}{V_{dd}} + R(x, y) + R_{\text{pin}}(x, y) + D(x, y) + M(x, y), \tag{9}$$

$$R_{\text{eff}}(x, y) = R_0(x, y)\big[1 + O_H^{\text{eGR}} + O_V^{\text{eGR}} + O_H^{\text{GR}} + O_V^{\text{GR}}\big]. \tag{10}$$

In Eq. 9, $P(x, y)$ is the aggregated tile power; $R(x, y)$ is RUDY (Rectangular Uniform Wire Density); $R_{\text{pin}}(x, y)$ is pin-RUDY capturing pin-driven routing density;

$D(x, y)$ is the cell-density map; and $M(x, y)$ is the macro-region indicator. In Eq. 10, $R_0(x, y)$ is the nominal PDN resistance, and O_H^{eGR}, O_V^{eGR}, O_H^{GR}, and O_V^{GR} are early global routing and global routing overflow maps in horizontal and vertical directions.

All terms are normalised to $[0, 1]$. Rather than fixing explicit weights, the diffusion model learns appropriate spatial weighting for each channel through the geometry-aware conditioning pathway described in Section 3.3. Table 2 summarises the resulting channel groups and their role in the $I_{\text{eff}}/R_{\text{eff}}$ decomposition.

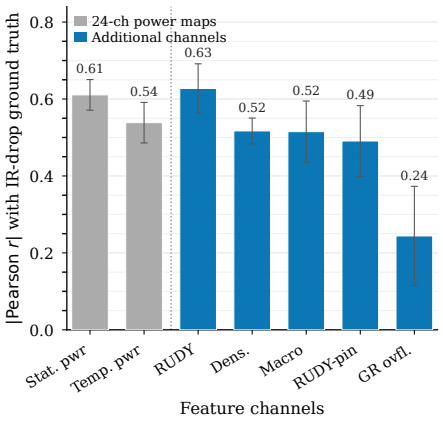


Figure 2: Pearson correlation with IR drop ground truth.

Figure 2 reports the Pearson correlation of each feature group with IR drop ground truth over 1,024 test samples on CircuitNet-N28. RUDY-based routing density achieves the strongest correlation ($|r| = 0.627$), exceeding the static power map baseline ($|r| = 0.607$), confirming that PDN resistance variation carries independent predictive signal that power maps do not capture. Cell density

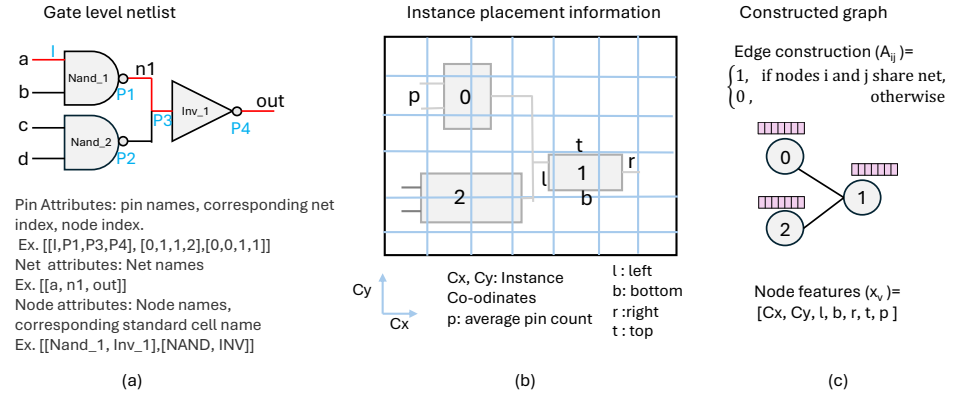


Figure 3: Graph construction: (a) Gate level netlist and graph construction attributes, (b) Instance (GCell) placement information, each instance is placed on grid (c_x, c_y), annotated with its bounding-box coordinates (l, b, r, t), and pin count p, (c) Constructed graph representation with node feature vector $\mathbf{x}_v = [c_x,\ c_y,\ l,\ b,\ r,\ t,\ p]$.

($|r| = 0.517$) and macro region ($|r| = 0.515$) also independently correlate with IR drop: dense regions can have better PDN coverage, providing a complementary signal to power consumption. Routing overflow ($|r| = 0.244$) captures directional congestion that modulates effective PDN resistance. Early global routing overflow ($|r| = 0.046$) shows weak marginal correlation independently but contributes as part of the R_{eff} representation. Together, these ten channels yield a $256\times256\times34$ feature tensor that provides both current demand and PDN resistance information to the generative conditioning process. Figure 7 in supplementary material visualises representative channels alongside the ground-truth IR drop map, showing spatial patterns that differ substantially from power maps.

3.2 Graph Creation Based on Topological Information

To incorporate logical connectivity into IR-drop generation, we construct a graph for each design using the flattened gate-level netlist and instance placement data provided in CircuitNet Chai et al. (2023); Jiang et al. (2024). Each design is represented as $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$, where $\mathcal{V}$ is the set of instances, $\mathcal{E}$ encodes connectivity through nets, and $\mathbf{X}$ contains physical design node features. Pins that share the same *net index* belong to the same net; pins that share the same *node index* belong to the same instance. For example, in Figure 3(a), pins P1 and P3 both have net index 1, so they belong to net `n1`, which connects `NAND_1` and `INV_1`. By traversing the pin attributes, we identify all nets and the set of instances connected through each net.

Two instances are connected by an edge if they appear together on at least one net:

$$\mathbf{A}_{ij} = \begin{cases} 1, & \text{if instances } i \text{ and } j \text{ share a net,} \\ 0, & \text{otherwise.} \end{cases} \tag{11}$$

The example in Figure 3(a) shows how the three attribute arrays describe the circuit. The pin-attribute array lists the pin names together with their net indices and node indices (e.g., `I`, `P1`, `P3`, `P4` belong to nets {0,1,1,2} and nodes {0,0,1,1}). The net-attribute array simply maps each net index to its net name (e.g., net 0 corresponds to `a`, net 1 to `n1`, and net 2 to `out`). The node-attribute array maps each node index to an instance and its standard-cell type (e.g., node 0 is `NAND_1` of type `NAND`, and node 1 is `INV_1` of type `INV`). Together, these three arrays fully specify which pins belong to each net and which pins belong to each instance, allowing the logical connectivity to be reconstructed directly from the attributes. Placement information is provided as bounding-box coordinates for each instance in GCell units (Figure 3(b)). For each node v, we construct the feature vector

$$\mathbf{x}_v = [c_x,\ c_y,\ l,\ b,\ r,\ t,\ p], \tag{12}$$

where (c_x, c_y) is the instance center, (l, b, r, t) are the bounding-box coordinates, and p is the pin count. The final constructed graph is shown in Figure 3(c). Each node encodes geometric layout

properties, and each edge reflects logical connectivity derived from the netlist. All graphs are stored in PyTorch Geometric format, enabling efficient integration into our multimodal generative framework.

3.3 Image–Graph Fusion for Multimodal Conditioning

GIF conditions a diffusion denoising network jointly on geometric layout features and netlist graph tokens. The conditioning design reflects the physical structure of IR drop: local current demand varies at fine spatial granularity and must be captured at every denoising resolution, while global logical connectivity creates long-range current demand correlations best captured at the coarsest spatial scale. We construct conditioning pathways that match these two regimes, rather than applying a uniform conditioning interface across the network.

Diffusion formulation. Let $Y \in \mathbb{R}^{H\times W}$ denote the IR drop map, normalised to $[-1, 1]$ via $\widetilde{Y} = 2Y - 1$. The forward process adds Gaussian noise over T steps Ho et al. (2020):

$$\widetilde{Y}_t = \sqrt{\bar{\alpha}_t}\,\widetilde{Y} + \sqrt{1-\bar{\alpha}_t}\,\varepsilon, \quad \varepsilon \sim \mathcal{N}(0, \mathbf{I}), \tag{13}$$

where $\bar{\alpha}_t$ follows a cosine schedule. The denoising network ϵ_θ learns to predict ε conditioned on both the geometry-aware feature image $X \in \mathbb{R}^{C\times H\times W}$ and the graph token set $T \in \mathbb{R}^{K\times D}$.

Geometry conditioning. IR drop varies sharply at fine spatial scales: routing congestion, cell placement, and macro blockage create tile-level variation throughout the layout. The geometry-aware feature image X conditions every resolution of the denoising network. At resolution level ℓ, X is downsampled to X_ℓ and used to produce affine scale-shift pairs that modulate intermediate activations, consistent with FiLM-style feature modulation Perez et al. (2018) and adaptive normalisation Huang & Belongie (2017):

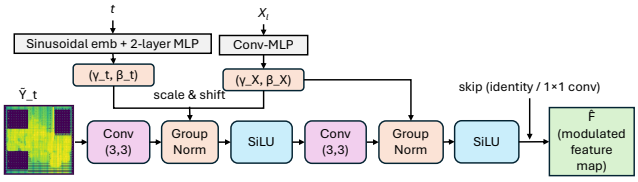


Figure 4: ResBlock with FiLM conditioning.

$$\widehat{F}^{(\ell)} = \gamma^{(\ell)}(X_\ell, t) \odot F^{(\ell)} + \beta^{(\ell)}(X_\ell, t), \tag{14}$$

where $\gamma^{(\ell)}$ and $\beta^{(\ell)}$ are produced by convolutional MLPs applied to X_ℓ and the sinusoidal timestep embedding. This is applied inside each ResBlock at every spatial scale from 256^2 down to 64^2 (Figure 4), encoding fine-grained IR drop structure tied to local geometric variation.

Topology conditioning. The gate-level netlist encodes logical connectivity that creates correlated current demand across the layout. We encode the netlist graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X})$ using a two-layer GCN Yang et al. (2022):

$$H = \phi_{\text{gcn}}(\mathcal{G}) \in \mathbb{R}^{N\times D}, \tag{15}$$

and compress node embeddings into a fixed set of K graph tokens via permutation-invariant pooling:

$$T = \rho(H) \in \mathbb{R}^{K\times D}. \tag{16}$$

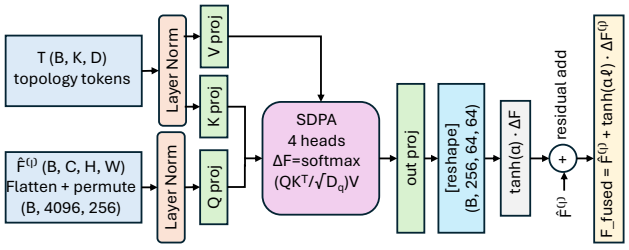


Figure 5: Cross-attention at the bottleneck. Graph tokens T serve as keys and values. Gate $\tanh(\alpha_\ell)$, initialised to zero, is learned from data.

This produces a fixed-dimensional summary of global logical structure independent of design size. Graph tokens are injected at the bottleneck resolution (64×64) via cross-attention Rombach et al. (2022) (Figure 5). The spatially compressed feature map provides queries while T provides keys and values:

$$\Delta F^{(\ell)} = \text{softmax}\left(\frac{Q^{(\ell)}K^\top}{\sqrt{D_q}}\right)V, \tag{17}$$

where $Q^{(\ell)} = W_Q \widehat{F}^{(\ell)}$, $K = W_K T$, $V = W_V T$. The bottleneck is the correct injection point: at 64×64, the network captures global IR drop structure where netlist topology provides the most complementary signal to local geometry.

Zero-initialised gate. The topology update is controlled by a learnable scalar α_ℓ initialised to zero:

$$F^{(\ell)}_{\text{fused}} = \widehat{F}^{(\ell)} + \tanh(\alpha_\ell)\ \text{reshape}\Big(\Delta F^{(\ell)}\Big). \tag{18}$$

At initialisation, $\tanh(\alpha_\ell) = 0$ and the model is a purely geometry-conditioned denoiser. The gate opens only to the extent that topology reduces the denoising objective. The learned gate values confirm this: on CircuitNet-N28, where designs share a common RISC-V architecture, $\tanh(\alpha) = 0.036$; on CircuitNet-N14, which includes diverse processor and accelerator layouts, $\tanh(\alpha) = 0.097$. Topology contributes more when design diversity is higher and global current demand is less predictable from geometry alone.

4 Evaluation

Experimental Setup. All experiments use Ubuntu 22.04 with an AMD EPYC 7763 CPU (64 cores, 128 threads) and four NVIDIA RTX A6000 GPUs (48 GB memory each, CUDA 12.6, driver 560.35.05).

Model and Training Parameters. We adopt a U-Net-based Ronneberger et al. (2015) denoising backbone with T=1000 diffusion steps. We use AdamW (β_1=0.9, β_2=0.999) with learning rate 2×10^{-4} and no weight decay. An auxiliary reconstruction loss (L1 on $\hat{x}_0$) with weight 0.1 is applied only for diffusion steps $t < 150$. Exponential moving average (EMA) weights with decay 0.999 are used for evaluation.

Datasets. We evaluate our framework on the CircuitNet-N28 and CircuitNet-N14 datasets Jiang et al. (2024); Chai et al. (2023). Additional dataset details are provided in the supplementary material 6.2.

CircuitNet-N28. CircuitNet-N28 contains six open-source RISC-V designs fabricated in 28 nm technology. We adopt a design-wise split with four designs for training, one for validation, and one for testing to prevent cross-design leakage (Table 8).

CircuitNet-N14. CircuitNet-N14 includes a broader set of designs such as RISC-V processors, NVIDIA GPUs, and ML accelerators fabricated in 14 nm technology. Compared to N28, these layouts exhibit greater variation in floorplan organization, utilization, aspect ratios, and power delivery configurations (Table 9).

4.1 Evaluation Metrics and Protocols

We compare each generated IR drop map with the ground-truth IR drop map using the metrics of prior IR drop work Xie et al. (2020b); Chhabria et al. (2021); Zhao et al. (2024): PSNR and SSIM Fardo et al. (2016); Wang et al. (2004) for pixel-level accuracy and spatial structure, MAE, NMAE, and RMSE for IR drop error, and Pearson and Spearman correlation for the IR drop distribution and the ordering of tiles. For hotspot detection, we mark the top 5% and top 10% IR drop tiles as hotspots and report precision, recall, and F1. For uncertainty, we use the per-pixel variance across N generated samples and report its correlation with the error, AUSE Ilg et al. (2018), and the miscalibration area Levi et al. (2022) (definitions in supplementary Section 6.3). On CircuitNet-N28, we use two splits. The *PDNNet protocol* Zhao et al. (2024) train and test split; prior methods report results under this split, so we use it to compare with prior methods. The *design-wise split* (Table 8) adds a validation design for model selection and uncertainty calibration.

4.2 Generation Results on CircuitNet-N28

Comparison with prior methods. Table 3 reports results under the PDNNet protocol. GIF has the lowest NMAE (0.026) and the highest PSNR (21.77 dB), SSIM (0.79), and Pearson correlation (0.95) among all methods. Results of prior methods are taken from their publications.

Geometry conditioning. All results below use the design-wise split. Table 4 reports the effect of each conditioning input. All rows use the same training configuration and

Table 4: Effect of each conditioning input on CircuitNet-N28, design-wise split (1,122 test samples). Upper rows: one generated sample per layout. †Mean of $N=4$ generated samples.

Conditioning	ControlNet	PSNR↑	SSIM↑	MAE↓	RMSE↓	Pearson↑	Spearman↑
24-ch images	✗	19.218	0.699	0.0378	0.1103	0.9069	0.4949
34-ch images	✗	19.542	0.739	0.0352	0.1063	0.9118	0.4973
24-ch images	✓	19.255	0.711	0.0374	0.1098	0.9065	0.4865
34-ch images	✓	19.583	0.766	0.0345	0.1058	0.9127	0.5033
34-ch + graph (K=32, top-k)	✗	19.519	0.611	0.0383	0.1066	0.9130	0.5107
34-ch + graph (K=32, top-k)	✓	19.307	0.762	0.0363	0.1091	0.9064	0.4976
34-ch + graph (K=64, mean)	✓	19.727	0.780	0.0335	0.1041	0.9154	0.5068
34-ch + graph (K=64, mean)†	✓	**21.296**	**0.791**	**0.0306**	**0.0870**	**0.9406**	**0.5399**

differ only in the conditioning input. Extending the input from 24 to 34 channels increases SSIM from 0.699 to 0.739 and Pearson correlation from 0.9069 to 0.9118. The ten layout-aware channels (Section 3.1) therefore add IR drop information that power maps do not contain. Adding ControlNet Zhang et al. (2023) increases SSIM to 0.766 and gives the lowest MAE (0.0345) and RMSE (0.1058) among the geometry-only configurations.

Graph conditioning. Graph tokens from the gate-level netlist improve all six metrics. Adding graph tokens with $K=64$ and mean pooling to the 34-channel configuration with ControlNet increases SSIM from 0.766 to 0.780, Pearson correlation from 0.9127 to 0.9154, and Spearman correlation from 0.5033 to 0.5068, and decreases MAE from 0.0345 to 0.0335 and RMSE from 0.1058 to 0.1041. Among the graph configurations, mean pooling with $K=64$ gives the best results, and we use it in all following experiments. The learned gate value $\tanh(\alpha) = 0.036$ confirms that geometry features are the primary conditioning signal, with graph tokens providing complementary structural regularisation at the bottleneck resolution where global IR drop coupling is strongest.

Table 3: Comparison with prior methods on CircuitNet-N28 under the PDNNet protocol.

Method	NMAE↓	PSNR↑	SSIM↑	Pearson↑
PowerNet	0.149	11.60	0.56	0.77
MAVIREC	0.039	18.27	0.68	0.91
PDNNet	0.028	19.35	0.72	0.92
GIF (ours)	**0.026**	**21.77**	**0.79**	**0.95**

Mean of generated samples. GIF generates multiple IR drop maps for each layout. The mean of $N=4$ generated samples gives the best results in Table 4 on all six metrics: PSNR 21.296 dB, SSIM 0.791, MAE 0.0306, RMSE 0.0870, Pearson correlation 0.9406, and Spearman correlation 0.5399. Compared with one generated sample, Pearson correlation increases from 0.9154 to 0.9406, and RMSE decreases from 0.1041 to 0.0870. Generating one sample takes 12.6,s with DDPM and 1.4,s with DDIM; supplementary Section 9 reports the speed and accuracy of each sampler.

Comparison with deterministic regressors. To separate the effect of the conditioning from the effect of the generative formulation, we train three deterministic regressors with the same 34-channel input and the same split: a U-Net Ronneberger et al. (2015), the same U-Net with $K=64$ graph tokens, and Restormer Zamir et al. (2022). Table 5 reports the results. GIF has higher PSNR and Pearson correlation and lower RMSE than all three regressors, and GIF also gives a per-pixel uncertainty map. The regressors have higher SSIM (about 0.84, compared with 0.79 for GIF) because the mean of generated samples smooths local detail. The strong results of all models with the same input show that the geometry and graph conditioning is the main source of accuracy, and the generative formulation adds the uncertainty map.

Hotspot detection and uncertainty. Hotspots are the tiles whose IR drop can cause timing failures. Table 6 reports hotspot detection and uncertainty. At the top-10% threshold, GIF detects hotspots with precision 0.804 and recall 0.570. The uncertainty on missed hotspot tiles is 1.6× the uncertainty on detected hotspot tiles, so the uncertainty map marks missed hotspots for re-simulation. We compare the uncertainty of GIF with MC-dropout Gal & Ghahramani (2016) and with an ensemble of the three regressors. GIF reaches 91% of the oracle in AUSE, compared with 86% for both methods, and it does not require training additional models. After we fit one scale parameter on the validation

Table 5: GIF and deterministic regressors with the same 34-channel input.

Method	PSNR↑	RMSE↓	Pearson↑	Unc.
U-Net	21.04	0.090	0.937	✗
U-Net + graph	20.92	0.091	0.936	✗
Restormer	20.95	0.091	0.936	✗
GIF	**21.30**	**0.087**	**0.941**	✓

Table 6: Hotspot detection (all test samples, N=4) and uncertainty (72 test layouts)

Hotspot	Precision	Recall	F1
Top-5%	0.715	0.288	0.368
Top-10%	0.804	0.570	0.641
Uncertainty	**Corr.↑**	**AUSE↑**	**Cost**
MC-dropout	0.50	86%	1×
Regressor ensemble	0.55	86%	3×
GIF	**0.68**	**91%**	1×

Table 7: Quantitative results on CircuitNet-N14 Jiang et al. (2024); Chai et al. (2023). †Mean of $N = 4$ generated samples on 100 test instances.

Model	PSNR↑	SSIM↑	MAE↓	RMSE↓	Pearson↑	Spearman↑
34-ch + ControlNet	14.156	0.513	0.0795	0.1996	0.8903	0.8133
34-ch + ControlNet (cfg-drop 0.1)	14.987	0.558	0.0667	0.1797	0.9106	0.8284
34-ch + Graph (K=32, top-k) + ControlNet	14.143	0.495	0.0797	0.1995	0.8892	0.8103
34-ch + Graph (K=32, top-k) + ControlNet (finetuned)	14.455	0.503	0.0758	0.1918	0.8974	0.8149
34-ch + Graph (K=32, top-k) + ControlNet (finetuned)†	16.430	0.532	0.0695	0.1523	0.9299	0.8532

design Levi et al. (2022), the miscalibration area on the test design decreases from 0.37 to 0.11 (supplementary Section 8).

4.3 Generation Results on CircuitNet-N14

Table 7 reports results on CircuitNet-N14. The image-only configuration with classifier-free guidance dropout of 0.1 achieves the strongest single-sample performance: 14.987 dB PSNR, 0.558 SSIM, and 0.9106 Pearson correlation across all 3,456 test samples. Graph-conditioned variants show lower pixel-level metrics on N14 because a large fraction of designs contain incomplete placement metadata, producing no valid graph tokens for those samples (supplementary Table 10). Where graph tokens are available, the finetuned graph model achieves competitive Pearson correlation (0.8974), consistent with N28 findings. The generative advantage transfers to N14. Averaging $N = 4$ independent samples from the finetuned graph model improves PSNR from 14.455 to 16.430 dB on a representative 100-sample subset — a gain of 1.97 dB, larger than the corresponding N28 gain (1.57 dB). The $N = 4$ result is evaluated on 100 samples due to computational cost and is not directly comparable to the full-dataset single-sample rows. Pearson correlation reaches 0.9299 at $N = 4$, with variance-error correlation of 0.517, confirming that the generative formulation produces calibrated uncertainty estimates at 14 nm as it does at 28 nm. The larger PSNR gain on N14 is consistent with its higher layout diversity producing more learnable uncertainty in the conditional distribution. GIF maintains strong correlation accuracy across both technology nodes, with Pearson above 0.89 in all single-sample configurations and above 0.92 with ensemble sampling. A representative qualitative example is provided in the supplementary material Figure 9.

5 Conclusion

GIF reformulates IR drop analysis as a conditional image generation problem, conditioning a diffusion model jointly on a geometry-aware feature image and gate-level netlist graph tokens. Geometry features condition the denoising network at every spatial resolution, capturing local current demand and PDN resistance variation. Graph tokens condition at the bottleneck via gated cross-attention, capturing global logical connectivity that geometry alone cannot represent. GIF achieves strong generation quality on CircuitNet-N28 across all evaluation metrics, outperforming prior regression-based methods. As a generative model, ensemble sampling improves generation quality beyond any single-sample configuration, with sample variance providing calibrated uncertainty estimates as a natural byproduct. GIF also establishes the first IR drop baselines on CircuitNet-N14, demonstrating that the framework generalises across technology nodes and design styles.

AI USE STATEMENT

In This work, we have not used generative AI tools to generate synthetic data sets, develop theoretical models, formulate mathematical claims or proofs, propose hypotheses, design experiments, implement methods, clean datasets, or interpret results. The research idea, method design and implementation, experiments, analysis, and figures are the authors own. We used generative AI tools to help write some supporting scripts, to draft and edit parts of the paper for readability, to identify relevant literature, and to format and verify references. We have reviewed all AI-assisted work: AI-assisted scripts were tested by the authors, all numbers in the paper were checked against our experimental logs, and every reference was checked against its original publication. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Code and trained models are available in an anonymous repository (`https://anonymous.4open.science/r/GIF-EFE2`); a public GitHub repository will be released after the review process. Both datasets, CircuitNet-N28 and CircuitNet-N14 Chai et al. (2023); Jiang et al. (2024), are publicly available.

REFERENCES

S. Borkar. Design challenges of technology scaling. *IEEE Micro*, 19(4):23–29, 1999. doi: 10.1109/40.782564.

Zhuomin Chai, Yuxiang Zhao, Wei Liu, Yibo Lin, Runsheng Wang, and Ru 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*, 42(12):5034–5047, 2023.

Howard H Chen and David D Ling. Power supply noise analysis methodology for deep-submicron vlsi chip design. In *Proceedings of the 34th annual Design Automation Conference*, pp. 638–643, 1997.

Vidya A Chhabria, Andrew B Kahng, Minsoo Kim, Uday Mallappa, Sachin S Sapatnekar, and Bangqi Xu. Template-based pdn synthesis in floorplan and placement using classifier and cnn techniques. In *2020 25th Asia and South Pacific Design Automation Conference (ASP-DAC)*, pp. 44–49. IEEE, 2020.

Vidya A Chhabria, Yanqing Zhang, Haoxing Ren, Ben Keller, Brucek Khailany, and Sachin S Sapatnekar. Mavirec: Ml-aided vectored ir-drop estimation and classification. In *2021 Design, Automation & Test in Europe Conference & Exhibition (DATE)*, pp. 1825–1828. IEEE, 2021.

Alexey Dosovitskiy. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

Yen-Chun Fang, Heng-Yi Lin, Min-Yan Sui, Chien-Mo Li, and Eric Jia-Wei Fang. Machine-learning-based dynamic ir drop prediction for eco. In *2018 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pp. 1–7. IEEE, 2018.

Fernando A Fardo, Victor H Conforto, Francisco C De Oliveira, and Paulo S Rodrigues. A formal evaluation of psnr as quality measurement parameter for image segmentation algorithms. *arXiv preprint arXiv:1605.07116*, 2016.

Bushra Fatima and Rajeevan Chandel. Analysis of ir drop for robust power grid of semiconductor chip design: a review. In *ITM web of conferences*, volume 54, pp. 04001. EDP Sciences, 2023.

Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *international conference on machine learning*, pp. 1050–1059. PMLR, 2016.

Chia-Tung Ho and Andrew B Kahng. Incpird: Fast learning-based prediction of incremental ir drop. In *2019 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pp. 1–8. IEEE, 2019.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Xun Huang and Serge Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision*, pp. 1501–1510, 2017.

Eddy Ilg, Ozgun Cicek, Silvio Galesso, Aaron Klein, Osama Makansi, Frank Hutter, and Thomas Brox. Uncertainty estimates and multi-hypotheses networks for optical flow. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 652–667, 2018.

Xun Jiang, Zhuomin Chai, Yuxiang Zhao, Yibo Lin, Runsheng Wang, Ru Huang, et al. Circuitnet 2.0: An advanced dataset for promoting machine learning innovations in realistic chip design environment. In *The Twelfth International Conference on Learning Representations*, 2024.

Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

Selçuk Köse and Eby G Friedman. Fast algorithms for ir voltage drop analysis exploiting locality. In *Proceedings of the 48th Design Automation Conference*, pp. 996–1001, 2011.

Dan Levi, Liran Gispan, Niv Giladi, and Ethan Fetaya. Evaluating and calibrating uncertainty prediction in regression tasks. *Sensors*, 22(15):5540, 2022.

Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 10012–10022, 2021.

Sani R Nassif. Modeling and analysis of manufacturing variations. In *Proceedings of the IEEE 2001 Custom Integrated Circuits Conference (Cat. No. 01CH37169)*, pp. 223–228. IEEE, 2001.

Sani R Nassif. Power grid analysis benchmarks. In *2008 Asia and South Pacific Design Automation Conference*, pp. 376–381. IEEE, 2008.

Ethan Perez, Florian Strub, Harm De Vries, Vincent Dumoulin, and Aaron Courville. Film: Visual reasoning with a general conditioning layer. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015. URL `https://arxiv.org/abs/1505.04597`.

Naveed Sherwani. *Algorithms for VLSI Physical Design Automation*. Springer, 1995.

Kiran Thorat, Hongwu Peng, Yuebo Luo, Xi Xie, Shaoyi Huang, Amit Hasan, Jiahui Zhao, Yingjie Li, Zhijie Shi, Cunxi Yu, and Caiwen Ding. Groot: Graph edge re-growth and partitioning for the verification of large designs in logic synthesis. In *2025 IEEE/ACM International Conference On Computer Aided Design (ICCAD)*, pp. 1–9, 2025. doi: 10.1109/ICCAD66269.2025.11240954.

Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. doi: 10.1109/TIP.2003.819861.

Wayne Wolf. *Modern VLSI Design: Systems on Silicon*. Pearson, 2008.

Zhiyao Xie, Hai Li, Xiaoqing Xu, Jiang Hu, and Yiran Chen. Fast ir drop estimation with machine learning. In *2020 IEEE/ACM International Conference on Computer-Aided Design (ICCAD)*, pp. 1–8. IEEE, 2020a.

Zhiyao Xie, Haoxing Ren, Brucek Khailany, Ye Sheng, Santosh Santosh, Jiang Hu, and Yiran Chen. Powernet: Transferable dynamic ir drop estimation via maximum convolutional neural network. In *2020 25th Asia and South Pacific Design Automation Conference (ASP-DAC)*, pp. 13–18, 2020b. doi: 10.1109/ASP-DAC47756.2020.9045574.

Shuwen Yang, Zhihao Yang, Dong Li, Yingxueff Zhang, Zhanguang Zhang, Guojie Song, and Jianye Hao. Versatile multi-stage graph neural network for circuit representation. *Advances in Neural Information Processing Systems*, 35:20313–20324, 2022.

Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5728–5739, 2022.

Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3836–3847, 2023.

Yuxiang Zhao, Zhuomin Chai, Xun Jiang, Yibo Lin, Runsheng Wang, and Ru Huang. Pdnnet: Pdn-aware gnn-cnn heterogeneous network for dynamic ir drop prediction, 2024. URL `https://arxiv.org/abs/2403.18569`.

Su Zheng, Lancheng Zou, Peng Xu, Siting Liu, Bei Yu, and Martin Wong. Lay-net: Grafting netlist knowledge on layout-based congestion prediction. In *2023 IEEE/ACM International Conference on Computer Aided Design (ICCAD)*, pp. 1–9. IEEE, 2023.

Yu Zhong and Martin DF Wong. Fast algorithms for ir drop analysis in large power grid. In *ICCAD-2005. IEEE/ACM International Conference on Computer-Aided Design, 2005.*, pp. 351–357. IEEE, 2005.

6 Appendix

6.1 Background: Modern Chip Design Flow and IR-Drop

Figure 6 shows modern chip design follows a standard sequence of stages including system specification, architecture, RTL, logic synthesis, physical design, and sign-off. Early stages establish the architecture and generate the logic, while physical design resolves the geometric and electrical constraints that govern on-chip behavior Wolf (2008); Sherwani (1995). This stage determines where cells are placed, how wires are routed, and how the power-delivery network (PDN) distributes supply voltage across increasingly resistive metal layers. As technology scales, wire resistance rises sharply and switching demand grows, making stable power delivery a central limitation in advanced designs Borkar (1999); Fatima & Chandel (2023). As illustrated in Figure 6, these stages collectively shape the final layout and its power-distribution characteristics.

IR-drop is the reduction in supply voltage caused by current flowing through the resistive PDN. Although conceptually simple, its consequences are severe: small voltage reductions produce measurable delay shifts, alter noise margins, and destabilize logic under high activity Nassif (2001); Chen & Ling (1997). IR-drop interacts with clock uncertainty and process variation, amplifying timing sensitivity in deeply scaled nodes. As a result, IR-drop has become a dominant factor in design closure, often determining whether a layout can meet its timing targets after routing Xie et al. (2020a). When violations are detected late typically during Static IR-Drop Analysis in the post-route stage they trigger engineering change order (ECO) loops that require re-routing, PDN reinforcement, or placement adjustments, all of which propagate through timing and congestion, significantly increasing turnaround time.

Reliable IR-drop generation is therefore essential to reduce costly iterations and guide early design choices. Fast, accurate generation of IR-drop maps enables designers to identify weak PDN regions, anticipate voltage-loss patterns, and evaluate design modifications without repeatedly invoking expensive sign-off tools. By exposing critical hotspots earlier in the flow, generative models help stabilize timing behavior and improve the likelihood that the final layout converges without disruptive late-stage modifications.

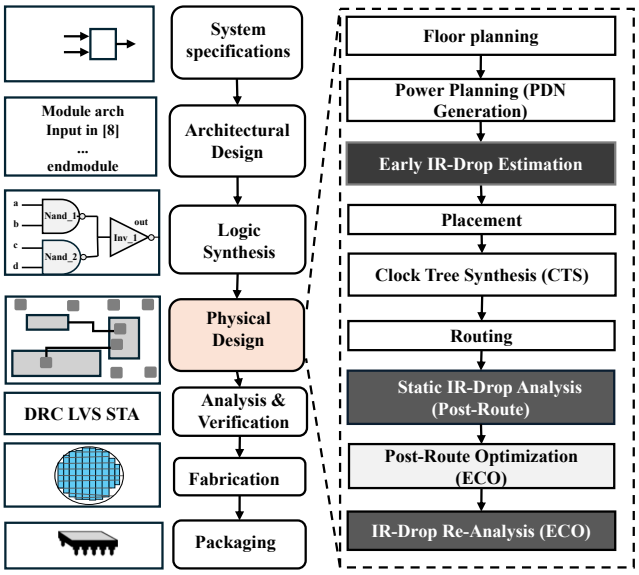


Figure 6: Modern chip design flow including physical design (orange) and IR-drop analysis (black).

Table 8: Design-wise split of the CircuitNet-N28 dataset.

Design	Split	Samples
RISCY-a	Train	2,003
RISCY-b	Train	1,858
RISCY-FPU-a	Train	1,969
zero-riscy-a	Train	2,042
RISCY-FPU-b	Val	1,248
zero-riscy-b	**Test**	**1,122**
Total		**10,242**

Table 9: Design-wise split of the CircuitNet-N14 dataset.

Design	Split	Samples
Nvidia-small	Train	85
RISCY	Train	3,162
RISCY-FPU	Train	3,456
Vortex-large	Train	61
openc910-1	Train	96
Nvidia-large	Train	32
Vortex-small	Val	96
zero-riscy	**Test**	**3,456**
Total		**10,444**

6.2 Dataset Details

CircuitNet-N28 CircuitNet-N28 provides tile-based physical design features and IR-drop maps generated from six RTL designs. Each layout is discretized into a regular grid of approximately 300×300 tiles, where each tile corresponds to a $2.25\,\mu\mathrm{m} \times 2.25\,\mu\mathrm{m}$ region of the physical layout (chip area $\sim 450\,\mu\mathrm{m} \times 450\,\mu\mathrm{m}$). This yields spatial feature maps that directly encode routing demand, placement density, switching activity, and power-related signals. We use all 10,242 samples from CircuitNet-N28 and follow the official design-wise split: four designs for training, one for validation, and one for testing. All feature and IR-drop maps are resized to 256×256 and normalized before training.

CircuitNet-N14 CircuitNet-N14 contains full-chip IR-drop maps and physical design features extracted from eight RTL designs. Compared to N28, the N14 layouts exhibit larger variation in floorplan styles, utilization, aspect ratio, and power delivery configurations. Each layout is represented by a spatial grid of comparable resolution (on the order of 300×300 tiles), covering the entire chip image at a uniform sampling density. We use all 10,444 available samples and adopt the design-wise split provided by CircuitNet: six designs for training, Vortex-small for validation, and zero-riscy for testing.

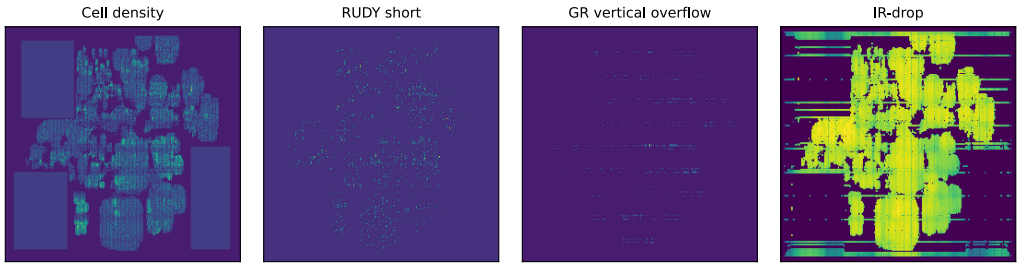


Figure 7: Additional feature channels and IR drop ground truth for a RISCY design from CircuitNet-N14 Jiang et al. (2024); Chai et al. (2023). (a) Cell density C_{den}, (b) RUDY short, (c) GR vertical overflow O_V^{GR}, (d) IR drop ground truth.

6.3 Evaluation Metrics

IR drop analysis identifies the tiles where the voltage drop can cause timing failures, so that the power delivery network can be strengthened before tapeout. We choose metrics that reflect this goal. Let $\mathbf{Y} \in \mathbb{R}^{H\times W}$ be the ground-truth IR drop map and $\widehat{\mathbf{Y}} \in \mathbb{R}^{H\times W}$ the generated IR drop map, both normalized to $[0, 1]$. We flatten both into vectors $y, \hat{y} \in \mathbb{R}^N$ with $N = HW$.

Pixel-level accuracy. MAE is the average per-tile error:

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |\hat{y}_i - y_i| \,. \tag{19}$$

Because IR drop values are normalized to $[0, 1]$, MAE is the average error as a fraction of the normalized range. NMAE follows the definition of PDNNet Zhao et al. (2024), so that GIF can be compared with published results. RMSE gives larger weight to large errors:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (\hat{y}_i - y_i)^2}. \tag{20}$$

Large errors are important in IR drop analysis, because an underestimated hotspot can hide a timing failure. PSNR is computed from the mean squared error, $\text{PSNR} = 10\log_{10}(1/\text{MSE})$, and is included for comparison with prior work Zhao et al. (2024); Chhabria et al. (2021).

Spatial structure. SSIM Wang et al. (2004) compares local mean, contrast, and structure in windows of the two maps:

$$\text{SSIM}(\mathbf{Y}, \widehat{\mathbf{Y}}) = \frac{(2\mu_y\mu_{\hat{y}} + C_1)(2\sigma_{y\hat{y}} + C_2)}{(\mu_y^2 + \mu_{\hat{y}}^2 + C_1)(\sigma_y^2 + \sigma_{\hat{y}}^2 + C_2)}, \tag{21}$$

computed with an 11×11 Gaussian window Wang et al. (2004). Current flow through the resistive power grid produces smooth IR drop gradients. SSIM measures whether the generated IR drop map reproduces these local gradients, which MAE and RMSE do not measure directly.

Correlation. Pearson correlation measures whether the generated IR drop map follows the IR drop distribution of the ground-truth map:

$$\text{Pearson} = \frac{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{N}(\hat{y}_i - \bar{\hat{y}})^2}\sqrt{\sum_{i=1}^{N}(y_i - \bar{y})^2}}. \tag{22}$$

Spearman correlation measures whether the generated IR drop map preserves the ordering of tiles by IR drop, independent of absolute values:

$$\text{Spearman} = \text{Pearson}\left(\text{rank}(\hat{y}), \text{rank}(y)\right). \tag{23}$$

Hotspot detection. A tile is a hotspot if its IR drop is above the top-$k\%$ threshold of the ground-truth map, with $k \in \{5, 10\}$. We apply the same threshold to the generated IR drop map. With TP, FP, and

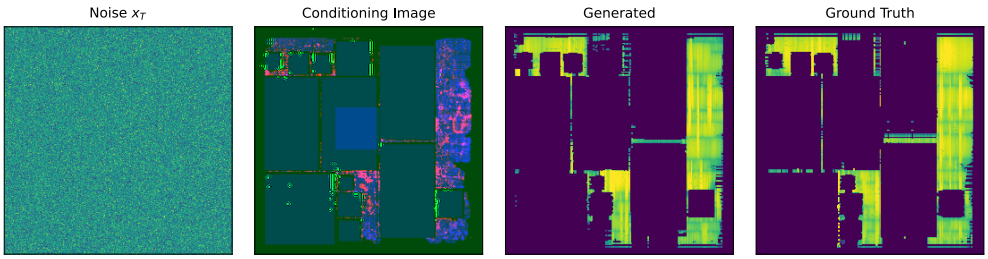


Figure 8: Qualitative IR-drop generation on CircuitNet-N28: (a) noise x_T, (b) conditioning features (3-channels shown), (c) generated IR-drop $\hat{x}_0$, (d) ground truth.

FN the numbers of true positive, false positive, and false negative hotspot tiles:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}, \quad \text{F1} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}. \tag{24}$$

Uncertainty. For N_s generated samples $\hat{y}^{(1)}, \ldots, \hat{y}^{(N_s)}$, the generated IR drop map is the mean $\hat{y}_i = \frac{1}{N_s} \sum_s \hat{y}_i^{(s)}$ and the per-tile uncertainty is the variance $u_i = \frac{1}{N_s} \sum_s \left(\hat{y}_i^{(s)} - \hat{y}_i \right)^2$. We use four measures. (1) The *variance–error correlation* is the Pearson correlation between u_i and the error $|\hat{y}_i - y_i|$. (2) The *MAE reduction* is the relative decrease in MAE when only the 50% of tiles with the lowest u_i are kept. (3) *AUSE* Ilg et al. (2018) uses the sparsification curve, which gives the MAE of the remaining tiles as tiles are removed in order of decreasing u_i. The oracle curve removes tiles in order of decreasing true error. AUSE is the area between the two curves; lower is better. (4) The *miscalibration area* is the area between the expected and the observed coverage of central intervals built from u_i. We fit one scale parameter on the standard deviation $\sqrt{u_i}$ using the validation design Levi et al. (2022) and report the miscalibration area on the test design.

Evaluation protocol. All metrics are computed on a single generated sample per test layout. In IR-drop analysis, the ground truth is produced by a single deterministic simulation from a commercial EDA tool given a fixed layout and power map. There is no distribution of ground truths. The task is to predict what that simulation would produce, and a single generated sample is the appropriate unit of comparison. This is the same protocol used by all prior discriminative methods Zhao et al. (2024); Chhabria et al. (2021), and it is the correct protocol for this task regardless of whether the model is generative or deterministic. GIF requires 1000 denoising steps per sample, so generating multiple samples per layout for the full test set is also computationally prohibitive.

7 Qualitative Visualization.

To complement the quantitative evaluation, we provide a representative CircuitNet-N28 test instance generated by our best graph-conditioned model (34-channel geometric features with K=64 mean-pooled graph tokens and ControlNet); see Figure 8. From left to right, the figure shows the initial Gaussian noise x_T, a three-channel composite of selected conditioning features, the generated IR-drop map $\hat{x}_0$, and the ground-truth IR-drop map. Because the conditioning tensor has 34 channels, we display only three representative feature maps as an RGB composite to provide a compact visualization; showing all channels is not visually interpretable. For the sample design zero-riscy (zero-riscy-b-3-c2-u0.85-m1-p6-f1), the model outputs a PSNR of 19.625, an SSIM of 0.811, an MAE of 0.0333, an RMSE of 0.1044, a Pearson correlation of 0.9320, and a Spearman correlation of 0.5039. These values align with the aggregate results in Table 4. In the visualization, the generated IR-drop map follows the overall magnitude and spatial distribution of the ground truth, including the primary high-drop regions and surrounding local variations.

7.1 Additional Analysis on CircuitNet-N14 Generation

First IR-Drop Results on CircuitNet-N14. To the best of our knowledge, GIF provides the first reported IR-drop results of any kind on CircuitNet-N14. Although the dataset includes IR-drop

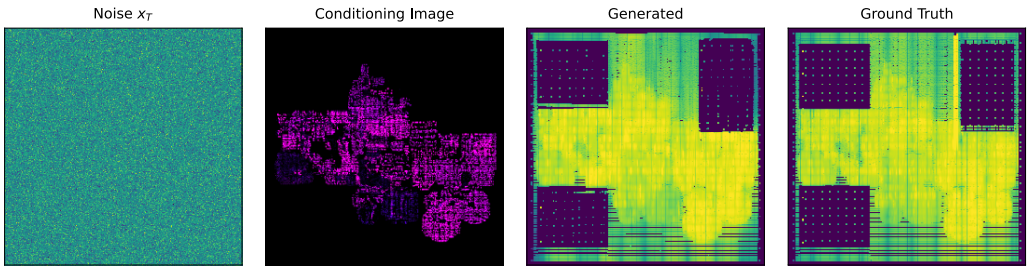


Figure 9: Qualitative IR-drop generation on CircuitNet-N14: (a) noise x_T, (b) conditioning features (3-channels shown), (c) generated IR-drop $\hat{x}_0$, (d) ground truth.

ground truth, no prior prediction (CNN, GNN, or hybrid), or generative (diffusion-based) method has published evaluation results on CircuitNet-N14 dataset. Thus, the values presented in the main paper represent the initial quantitative baselines for IR-drop map generation at 14 nm. These results demonstrate that the diffusion backbone maintains strong correlation behavior even under the much higher variability of N14 physical layouts.

Graph Availability in CircuitNet-N14. As described in the supplementary implementation details, graphs for N14 are constructed using the same procedure as for N28. However, CircuitNet-N14 contains a large number of placement snapshots whose instance bounding boxes are all zero. These samples cannot produce meaningful geometric node features and must be skipped during graph construction. This behavior arises from irregularities in the dataset metadata and is not specific to our framework.

Table 10 reports the number of valid graphs extracted for each design family. Several families (e.g., `Vortex-small`) yield no usable graphs, while others produce only a limited number. Since the full N14 feature/label set contains 10,444 IR-drop samples, many conditioning instances necessarily receive no graph tokens. This imbalance explains why graph-conditioned variants offer limited benefit or slight degradation on N14, in contrast to N28 where graph metadata are complete and consistently available.

Table 10: Number of valid graphs constructed for CircuitNet-N14 (from the official metadata). Designs with all-zero placements produce no usable graphs.

Design Family	**Valid Graphs**	**Total Placements**
RISCY	3373	3456
RISCY-FPU	3456	3456
Vortex-large	62	74
Vortex-small	0	96
nvdia-large	54	68
nvdia-small	88	93
openc910-1	96	96
zero-riscy	3456	3456
Total	10585	10795

GIF establishes the first IR-drop benchmarks on CircuitNet-N14 and maintains strong correlation consistency despite the dataset's higher layout variability. The reduced benefit of graph conditioning on N14 is driven by incomplete and inconsistent metadata, not by limitations of the fusion design. Where graph features are reliably available (e.g., N28), graph conditioning consistently improves performance.

Qualitative Visualization on CircuitNet-N14. To complement the quantitative evaluation, we provide a representative N14 test instance generated by our image-only model (34-channel conditioning with ControlNet and a classifier-free dropout rate of 0.1); see Figure 9. As in the N28 visualization, the figure shows the initial Gaussian noise x_T, a three-channel composite of selected conditioning

features, the generated IR-drop map $\hat{x}_0$, and the ground-truth IR-drop map. For the shown zero-riscy sample, the model yields a PSNR of 16.818, an SSIM of 0.518, an MAE of 0.0593, an RMSE of 0.1442, a Pearson correlation of 0.9243, and a Spearman correlation of 0.8922. These values are consistent with the overall N14 trends, where correlation metrics remain strong despite the larger design variability and the reduced spatial regularity of 14 nm layouts. Visually, the generated IR-drop map tracks the primary magnitude and spatial patterns of the ground truth, including prominent high-drop regions and surrounding gradients.

8 Uncertainty Evaluation

We evaluate the per-pixel uncertainty of GIF on CircuitNet-N28 with the design-wise split. We use three measures. The variance–error correlation is the Pearson correlation between the per-pixel variance and the error of the generated IR drop map. The MAE reduction is the decrease in MAE when we keep only the 50% of tiles with the lowest uncertainty. AUSE Ilg et al. (2018) measures how close the ordering of tiles by uncertainty is to the ordering by true error. Table 11 compares GIF with MC-dropout Gal & Ghahramani (2016) and with an ensemble of the three deterministic regressors (U-Net, U-Net with graph tokens, and Restormer) on the same 72 test layouts. GIF has the highest value on all three measures and does not require training additional models.

Table 11: Uncertainty on 72 test layouts, CircuitNet-N28, design-wise split.

Method	**Var–error corr.↑**	**MAE reduction↑**	**AUSE↓**	**Training cost**
MC-dropout (K=8)	0.50	90%	0.127	1×
Regressor ensemble	0.55	92%	0.122	3×
GIF	**0.68**	**93%**	**0.077**	1×

Calibration. We fit one scale parameter on the per-pixel standard deviation using the validation design Levi et al. (2022), and we evaluate on the test design. The miscalibration area decreases from 0.37 to 0.11.

Missed hotspots. At the top-10% threshold, the uncertainty on missed hotspot tiles is 1.6× the uncertainty on detected hotspot tiles.

9 Sampling Speed

Table 12 reports the generation time, accuracy, and uncertainty for three samplers on CircuitNet-N28. DDIM with 100 steps is 9× faster than DDPM with 1,000 steps and keeps Pearson correlation (0.930) and variance–error correlation (0.582). SSIM decreases with fewer steps, so we use DDPM with 1,000 steps for the main results. Table 13 reports the total time for N=4 and N=8 generated samples.

Table 12: Sampler comparison. Time is per sample; Pearson and SSIM are for the mean of N=4 generated samples.

Sampler	**Steps**	**Time (s)**	**Pearson↑**	**Var–err↑**	**SSIM↑**
DDIM	20	0.28	0.909	0.561	0.20
DDIM	100	1.41	0.930	0.582	0.26
DDPM	1000	12.6	0.941	0.570	0.79

Table 13: Total time (s) for N generated samples.

Sampler	N=1	N=4	N=8
DDPM, 1000	12.6	50	101
DDIM, 100	1.4	5.6	11