OPEN ACCESS

ARTICLE

Journal homepage: www.the-innovation.org/energy-use

INNOVATION PRESS

Energy Use

Energy Use 3 (2026):100068

Contents lists available at INNOVATION PRESS

# Generative Residential Floor Plan Design Integrating Energy Performance Evaluation and Spatial-Functional Constraints

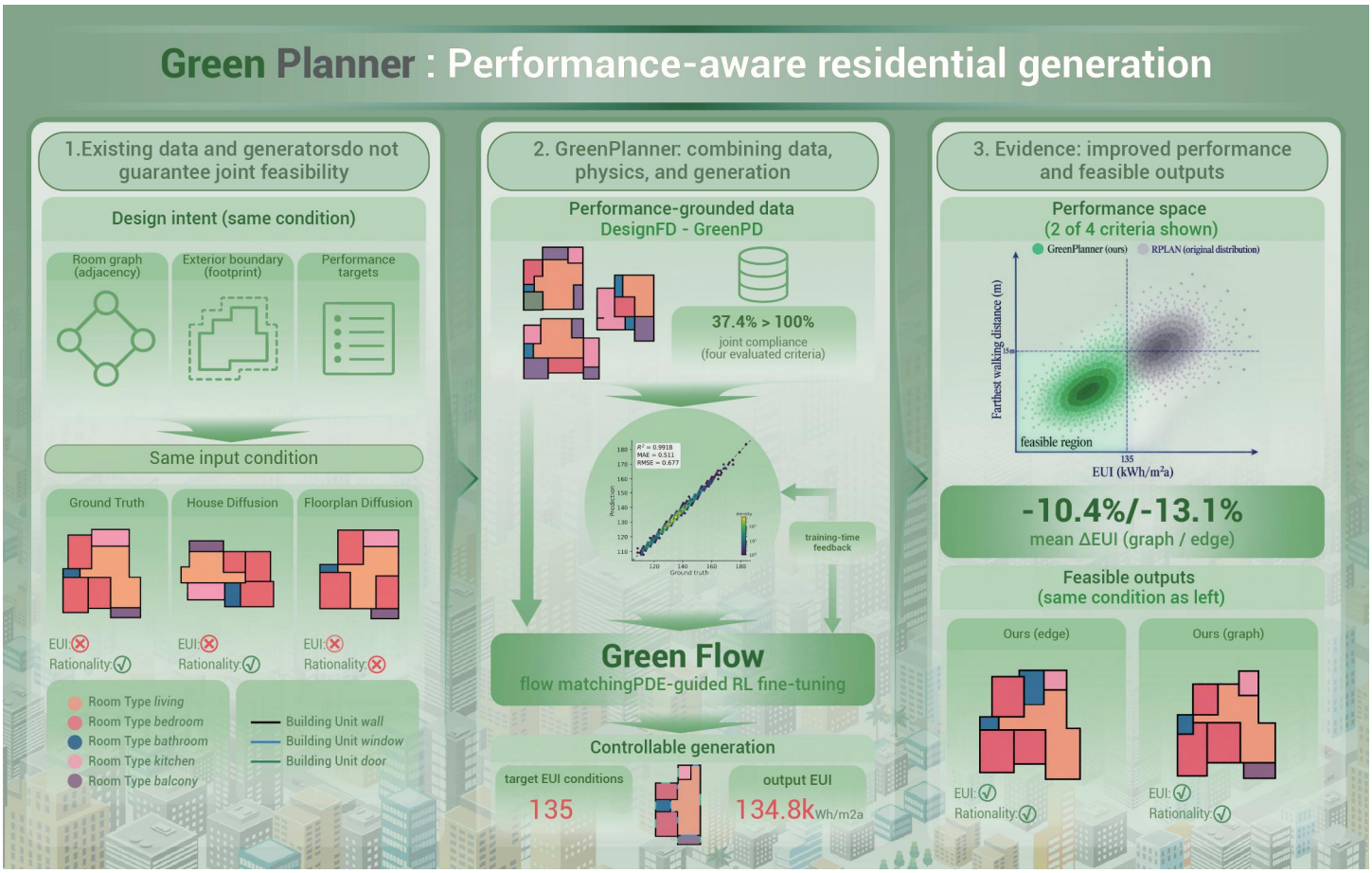
**Pengyu Zeng,[1] Tiancheng Zeng,[1] Jun Yin,[1] Jiling Chen,[1] and Shuai Lu[1,*]**

*Correspondence: shuai.lu@sz.tsinghua.edu.cn

Received: August 25, 2026; Accepted: September 20, 2026; Published Online: September 23, 2026; https://doi.org/10.59717/ipj.energy-use.2026.100068

© 2026 The Author(s). This is an open access article under the CC BY license (https://creativecommons.org/licenses/by/4.0/).

## GRAPHICAL ABSTRACT




## PUBLIC SUMMARY

- A performance-aware framework GreenPlanner is developed for early-stage residential floor-plan generation with built-in energy and spatial-function constraints.
- Surrogate evaluator PDE delivers high-accuracy multi-metric assessment, accelerating evaluation by $1.7\times10^5$.
- GreenPD filters layouts to fully satisfy energy, evacuation and functional requirements.
- GreenFlow cuts building EUI by up to 13.1% under constraints while retaining layout quality.
- This workflow reduces expert design time by 87.0% for performance-driven early design exploration.

ARTICLE

OPEN ACCESS

Journal homepage: www.the-innovation.org/energy-use

Energy Use

INNOVATION PRESS

Energy Use 3 (2026):100068

Contents lists available at INNOVATION PRESS

# Generative Residential Floor Plan Design Integrating Energy Performance Evaluation and Spatial-Functional Constraints

**Pengyu Zeng,[1] Tiancheng Zeng,[1] Jun Yin,[1] Jiling Chen,[1] and Shuai Lu[1,*]**

[1]School of Architecture, Tsinghua University, Beijing 100084, China

*Correspondence: shuai.lu@sz.tsinghua.edu.cn

Received: August 25, 2026; Accepted: September 20, 2026; Published Online: September 23, 2026; https://doi.org/10.59717/ipj.energy-use.2026.100068

© 2026 The Author(s). This is an open access article under the CC BY-NC-ND license (https://creativecommons.org/licenses/by/4.0/).

Citation: Zeng P., Zeng T., Yin J., et al. (2026). Generative Residential Floor Plan Design Integrating Energy Performance Evaluation and Spatial-Functional Constraints. *Energy Use* **2**:100068.

Early-stage residential floor-plan decisions influence both spatial functionality and operational energy use, yet existing generative approaches primarily focus on geometric plausibility and often lack efficient mechanisms for incorporating performance constraints during generation. Physics-based simulation can provide reliable energy assessment but is computationally expensive for repeated use in high-throughput generative design. To address this issue, we propose GreenPlanner, a performance-aware framework that integrates design evaluation, data refinement, and residential floor-plan generation. DesignFD augments residential layouts with automatically derived energy and spatial-functional feasibility labels. PDE learns these labels as a fast surrogate evaluator for energy performance, evacuation-related path length, floor area, and functional connectivity. GreenPD uses PDE-guided filtering and regeneration to construct demand–plan pairs satisfying the prescribed performance criteria. GreenFlow is then trained on GreenPD and further refined using PDE feedback, allowing energy and spatial-functional requirements to influence the generation process directly. Experiments show that PDE achieves an $R^2$ of 0.99 and an MSE of 0.46 while evaluating 100 layouts in 7.3 ms, providing up to a $1.7\times10^5$-fold speed-up over the reference EnergyPlus workflow. GreenPD achieves 100% compliance with the four evaluated criteria. Under performance-constrained generation, GreenFlow reduces mean EUI by 10.4% and 13.1% under graph- and edge-conditioned settings, respectively, while maintaining improved geometric and spatial-functional quality. A human expert study further shows an 87.0% reduction in design time relative to the tested manual workflow. These results demonstrate the feasibility of integrating rapid energy and functional evaluation into generative residential design for early-stage performance-aware exploration.

## INTRODUCTION

Buildings are central to both human well-being and energy-related carbon emissions. Humans spend nearly 90% of their time indoors,[1] while the built environment accounts for a substantial share of global carbon emissions.[2-4] Residential design therefore has consequences that extend beyond spatial organization. At the early design stage, decisions about plan geometry, room arrangement, exterior exposure, circulation, and openings can influence both functional usability and operational energy demand.[5] These decisions are made before many detailed envelope and system parameters are finalized, yet they establish the spatial configuration on which later design choices depend and can become costly to revise once the project progresses. Traditional residential design relies heavily on expert knowledge and repeated manual iteration,[6-8] which limits the number of alternatives that can be explored. A practical automated design system should therefore do more than generate visually plausible plans: it should support rapid exploration of layouts whose spatial organization is compatible with functional and energy-performance objectives from the outset.

Recent advances in controllable[9-11] and multimodal generation[12-14] have made automated floorplan synthesis increasingly feasible. Existing methods use conditioned generators[15-25] to translate room graphs, boundaries, sketches, or textual requirements into concrete residential layouts. This line of research has substantially improved geometric fidelity and controllability, but most methods are trained to reproduce the distribution of existing plans and satisfy only the conditions explicitly provided to the model. Practical requirements that are absent from the input—such as complete openings, circulation safety, functional adjacency, and energy performance—are therefore not guaranteed. Some representative layout representations do not explicitly include exterior windows; for example, HouseDiffusion represents floor plans as room and door polygon loops,[19] which complicates subsequent energy evaluation when exterior-opening information is required. As a result, a layout can appear geometrically plausible while remaining unsuitable for practical use. This motivates evaluating generative floorplans simultaneously in terms of geometric quality, spatial-functional rationality, and energy performance, as summarized in Figure 1A.

We identify three coupled gaps that prevent existing floorplan generators from being used directly for performance-aware residential design. First, there is a generation gap: prevailing objectives emphasize visual or geometric similarity and user-specified conditions, rather than jointly optimizing energy and spatial-functional feasibility. Second, there is an evaluation gap. Physics-based simulation and rule-based checking provide reliable assessment, but repeated full simulation inside a high-throughput generative loop is computationally expensive and requires detailed design inputs.[5,26,27] Manual design–evaluate–revise workflows are likewise too slow and inconsistent for large-scale model training and iterative generation.[28,29] A practical generative system therefore needs a learned evaluator that is sufficiently accurate to approximate key physical metrics while being fast enough to operate inside data construction and generation. Third, there is a data gap: if the training corpus itself contains infeasible examples, a more powerful generator may simply learn the same non-compliant distribution. Our assessment of RPLAN[20] illustrates this issue: under the four physical and functional criteria evaluated in this study, only 37.4% of layouts satisfy all criteria simultaneously. Thus, performance-aware generation requires not only a stronger generator, but also a scalable evaluation mechanism and a cleaner training distribution.

Accordingly, this study aims to develop a practical generative design framework that can rapidly evaluate residential layouts and explicitly incorporate energy and spatial-functional feasibility into both data construction and generation. We propose GreenPlanner, which unifies these functions through four connected components. Design Feasibility Dataset (DesignFD)

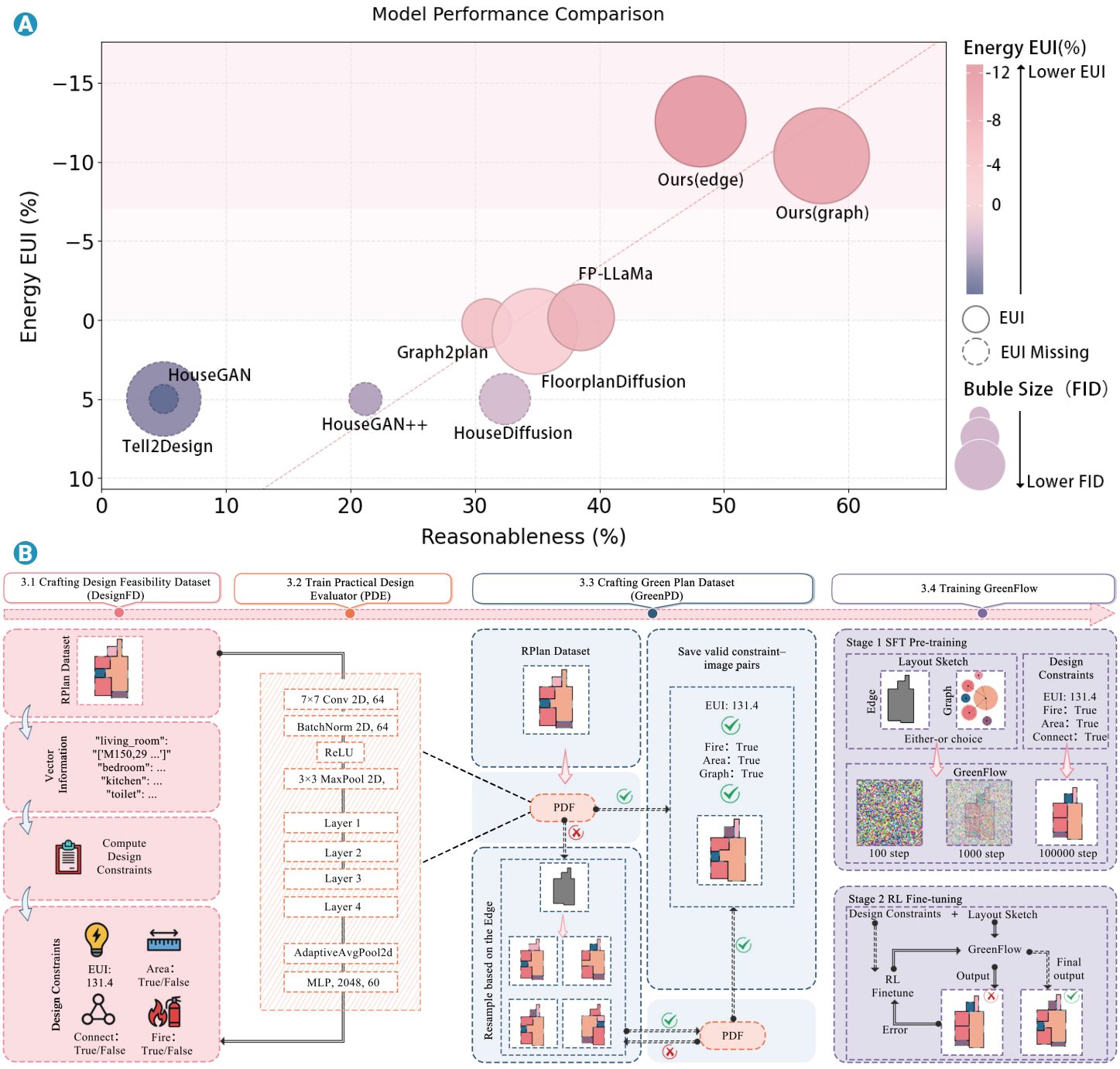



**Figure 1. (A) Comparison of model performance in terms of reasonableness (higher is better), energy EUI (lower is better), and FID (bubble size, larger is better). Our method achieves higher spatial plausibility, significantly lower energy consumption, and better generation quality, surpassing existing approaches overall. (B) Overview of our GreenPlanner, an energy- and functional-aware generative framework for residential layout design. It consists of four key components: (1) Design Feasibility Dataset (DesignFD), a dataset derived from RPLAN through quantitative energy and functional evaluations; (2) the Practical Design Evaluator (PDE), a convolutional network trained on DesignFD to predict design metrics rapidly; (3) Green Plan Dataset (GreenPD), a demand–plan dataset satisfying all prescribed criteria refined via PDE-guided resampling; and (4) the GreenFlow generator, trained and fine-tuned under PDE feedback for controllable, constraint-aware floorplan generation. Together, these components form an end-to-end pipeline from data construction to constraint-driven generative design.**

augments residential plans with automatically derived energy and functional labels; the Practical Design Evaluator (PDE) learns these labels to provide rapid surrogate assessment; Green Plan Dataset (GreenPD) uses PDE-guided filtering and regeneration to construct a compliant demand–plan distribution; and GreenFlow learns from GreenPD and is further refined with PDE feedback so that requested physical performance influences the generation distribution itself. In this way, energy performance is not treated only as a post-hoc metric used to reject finished layouts. Instead, evaluation first improves the training data and then directly supports performance-guided generation, creating an end-to-end pipeline from physical labeling to controllable layout synthesis.

The main contributions of this study are threefold: (1) We formulate GreenPlanner as a performance-aware residential generation framework that unifies evaluation, data refinement, and generation, shifting the task from visual/geometric floorplan synthesis toward practical design under energy and spatial-functional constraints; (2) We construct DesignFD with automatically derived physical and functional labels, develop the fast PDE surrogate evaluator, and use PDE-guided refinement to build GreenPD, a demand–plan dataset whose layouts satisfy the feasibility criteria considered in this work; (3) We develop GreenFlow, a flow-matching generator trained on GreenPD and refined with PDE-guided feedback, so that energy and functional requirements influence the learned generation distribution rather than being applied

only through post-hoc filtering.

Extensive experiments validate the complete pipeline. PDE reaches $R^2$ = 0.99 with an MSE of 0.46 while evaluating 100 layouts in 7.3 ms; GreenPD achieves 100% compliance under the evaluated criteria; GreenFlow improves geometric quality, spatial-functional rationality, and energy performance over representative baselines; and the expert study reports an 87.0% reduction in design time relative to the tested manual workflow.

The remainder of this paper is organized as follows. Section 2 reviews related work on floorplan generation, building-performance evaluation, and performance-guided design; Section 3 presents the GreenPlanner methodology; Section 4 reports the experimental results; Section 5 discusses practical implications and limitations; and Section 6 concludes the paper.

## MATERIALS AND METHODS

### Floorplan and architectural layout generation

Residential floorplan generation has progressed from data-driven geometric synthesis to graph-conditioned, diffusion-based, and multimodal generation. RPLAN[20] established a large-scale residential layout corpus that enabled learning-based plan synthesis. Graph-conditioned methods such as HouseGAN,[17] HouseGAN++,[18] and Graph2Plan[15] model room relationships explicitly, while HouseDiffusion[19] and FloorplanDiffusion[30] use diffusion processes to improve geometric fidelity and diversity. More recent systems broaden the conditioning interface through natural language and multimodal inputs, including Tell2Design,[16] LayoutGPT,[31] Floorplan-LLaMa,[32] CARD,[24] and unified multimodal floorplan generation.[25] These developments substantially improve controllability and reduce the amount of manual geometric drafting required at the early design stage.

Despite this progress, most floorplan generators are optimized primarily for distributional similarity, geometric agreement, or adherence to explicitly provided conditions. Practical building requirements that are not encoded in the input—such as complete openings, evacuation-related circulation, functional adjacency, and energy performance—are often outside the learning objective. Several representative layout representations focus on room geometry and door connectivity rather than complete building-physics inputs. HouseDiffusion, for example, represents a floor plan as polygonal loops for rooms and doors;[19] therefore, exterior-opening information required by an energy model may need to be supplied separately during downstream evaluation. Metrics designed for generative similarity alone may also fail to reflect engineering feasibility.[23,29] GreenPlanner is positioned differently: it treats floorplan generation as a performance-constrained design problem in which geometric realism, spatial-functional feasibility, and energy performance must be evaluated jointly.

### Building performance evaluation and learning-based surrogate models

Building performance assessment traditionally relies on physics-based simulation and rule-based checking. Automated rule-checking systems can formalize code requirements and detect design violations,[33] while EnergyPlus and related simulation workflows provide detailed estimates of operational energy use. These tools are essential for verification, but their computational and modeling costs make repeated invocation difficult inside high-throughput generative pipelines. More broadly, artificial intelligence has increasingly been used for building automation, performance prediction, and decision support,[6,8] while recent work has examined language-model-based assistance for building-energy workflows.[34] Separately, learning-based energy predictors can approximate expensive simulations and enable rapid screening of large residential design spaces.[5]

For generative design, however, a surrogate evaluator is useful only if its predictions are sufficiently accurate, fast, and aligned with the design variables that the generator can control. Prior work has combined evaluation techniques with conceptual generative design[27] and has explored energy-efficient residential layout optimization under prescribed geometric conditions.[5] These studies demonstrate the value of performance feedback, but performance evaluation is commonly treated as a downstream analysis or optimization stage. In contrast, GreenPlanner uses a learned Practical Design Evaluator (PDE) throughout the pipeline: first to support large-scale feasibility labeling and compliant data construction, and later to provide differentiable feedback for performance-aware generation.

### Constraint-aware and performance-guided generative design

Constraint-aware generative design seeks to move beyond unconstrained sample generation by incorporating engineering objectives, design rules, or evaluative feedback into the generation process. Existing strategies include optimization or evaluation in the loop,[5,6,27] revision-guided generation in general graphical-layout design,[28] explicit conditional generation,[30,35] and reinforcement-learning or reward-based optimization built on policy-gradient principles.[36,37] A related practical strategy is generate-then-filter or generate-then-select, in which many candidates are sampled and an external evaluator retains those satisfying the target constraints. These approaches can improve final solution quality, but when feasible designs occupy a small portion of the generator distribution, they may require repeated sampling and do not necessarily teach the generator to internalize the constraints.

GreenPlanner integrates these previously separated directions into a single residential design pipeline. DesignFD supplies automatically derived physical and functional labels; PDE converts expensive feasibility assessment into a fast learned surrogate; GreenPD uses this evaluator to construct a criterion-compliant demand–plan training set; and GreenFlow is subsequently optimized with PDE feedback so that energy and functional constraints influence the generation distribution itself. This combination distinguishes the proposed framework from methods that focus only on layout synthesis, only on building-performance prediction, or only on post-hoc filtering.

### Framework overview

We propose **GreenPlanner**, an energy- and functionality-aware generative framework for residential design (Figure 1B). It integrates four components: (1) **Design Feasibility Dataset (DesignFD)**, a dataset built from RPLAN[20] through automated energy and functional evaluation; (2) **Practical Design Evaluator (PDE)**, a predictor trained on DesignFD for fast simulation of energy and functional feasibility; (3) **Green Plan Dataset (GreenPD)**, a PDE-guided dataset pairing user requirements with layouts satisfying the prescribed criteria; and (4) **GreenFlow**, a generator trained on GreenPD with PDE feedback for controllable, constraint-grounded layout generation. We next describe each component of the framework in detail.

## DESIGN FEASIBILITY DATASET (DesignFD)

*Quantitative and qualitative design constraints*. We construct a Design Feasibility Dataset **(DesignFD)**, derived from RPLAN[20] with quantitative energy and spatial-functional feasibility labels derived by the automated assessment pipeline. Specifically, DesignFD defines four key constraints: **(a) Energy Use Intensity (EUI)**: calculated via annual energy simulations in EnergyPlus, where a lower EUI indicates higher energy efficiency; **(b) Fire-safety distance**: the maximum walkable distance from the entrance to the farthest accessible point, evaluated using a 15 m benchmark criterion; **(c) Floor area**: derived from pixel-to-scale conversion, expected to remain within a reasonable range following standard design practice; and **(d) Functional connectivity**: assessed through graph-based analysis of room adjacencies to ensure rational circulation and usability

The numerical thresholds adopted in this study serve as fixed benchmark criteria, informed by the cited standards and design assumptions, for examining whether heterogeneous energy and spatial-functional requirements can be propagated through evaluation, data refinement, and generative optimization. Because the framework is criterion-configurable, jurisdiction- or project-specific values can be substituted without changing the overall methodology.

*Crafting DesignFD*. DesignFD pairs each floorplan image with its corresponding energy and functional metrics, providing data support for quantitative feasibility modeling. Using the four indicators defined above, the automated assessment pipeline computes the corresponding labels and associates them with each rasterized floorplan, forming paired data for PDE training.

### Automated design feasibility assessment

The feasibility labels used in DesignFD are obtained through an automated assessment pipeline that combines deterministic spatial checks with physics-based energy simulation. This separation is important: DesignFD defines what is labeled, whereas the assessment procedures below define how those labels are computed and verified. The pipeline evaluates evacuation-related

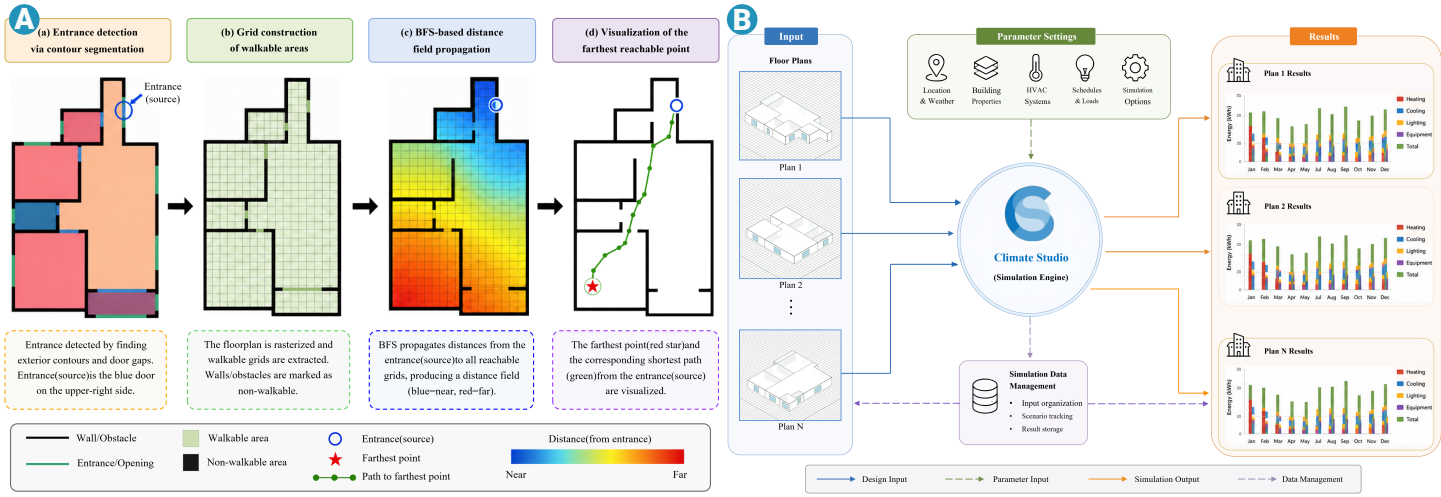



**Figure 2. (A) Pipeline of the entrance-to-interior farthest distance calculation. The process includes (a) entrance detection via contour segmentation, (b) grid construction of walkable areas, (c) BFS-based distance field propagation, and (d) visualization of the farthest reachable point. Residential energy simulation pipeline. (B) Raster floor plans are vectorized (Python: svgwrite, opencv) to generate 2D vectors for 3D modeling (Rhinoceros + Grasshopper). We assign consistent orientation (top=north) and standard geometry presets, then run monthly EUI simulations in EnergyPlus via ClimateStudio. Each room is a thermal zone.**

path length, room-topology compliance, and operational energy under a standardized simulation configuration.

*Fire-safety path calculation.* To evaluate accessibility and circulation quality within residential layouts, we compute the farthest reachable distance from the entrance door to any interior point in each plan. The algorithm automatically detects the entrance, constructs a walkable grid, and measures the maximum topological path distance.

Given a rasterized floor plan, we first run high-resolution preprocessing and scale calibration using functional cues such as wall thickness and door width. The entrance door is then detected via HSV (Hue–Saturation–Value) color segmentation, followed by contour filtering. We construct an occupancy grid for the walkable interior, treating walls as obstacles, and use breadth-first search (BFS) with 8-connectivity to compute geodesic distances from the detected entrance cell to all accessible pixels, taking the maximum. Finally, we convert this value to a physical distance in meters using the pixel-to-meter ratio derived from the plan scale.

For quality assurance, we validate the longest path with an A* search to confirm that the geodesic length is consistent with Euclidean constraints and that the farthest point lies within a closed, accessible region.

This metric evaluates spatial accessibility and is recorded in our dataset as a physical attribute for each layout (denoted as $D_{\text{path}}$). It quantifies how deep a resident must travel to reach the most distant interior space, serving as a proxy for circulation efficiency and fire-safety path analysis.

*Topological Compliance Checking.* The topological structure of a residential layout plays a crucial role in determining circulation efficiency and the coordination of functional spaces. The functional-topology criteria considered here are informed by residential design codes and occupant-oriented design practice.[38,39] Traditional evaluations mainly rely on manual inspection and expert judgment, which are often subjective, time-consuming, and inconsistent across projects.[33] To operationalize these criteria, we develop an automated topology-compliance checker grounded in computer vision and graph theory. The checker quantifies relationships among functional spaces and their door-based connections, providing an objective and reproducible metric for layout validation. Together with the entrance-to-interior farthest-distance calculation above, it forms a three-dimensional evaluation framework encompassing *function*, *circulation*, and *safety*, as illustrated in Figure 3.

The topology checker receives three structured components from the preceding vision-processing pipeline: room attributes, adjacency relations, and entry-room identifiers. Room attributes contain metadata such as room ID, type, subtype, and area (in square meters). Adjacency relations are represented as undirected room-pair edges inferred from door-segment detection, while entry rooms list those directly connected to the main entrance. Layouts are classified into three categories based on total floor area, as defined in Equation (1).

$$\text{house\_class}=\begin{cases}\text{small}, & 40\le\text{total\_area\_m2}\le 90\\ \text{big}, & 90<\text{total\_area\_m2}\le 150\\ \text{out\_of\_range}, & \text{otherwise}\end{cases} \quad (1)$$

Only layouts categorized as small or big proceed to full topology validation.

The functional-topology criteria combine code-informed relationships with residential design heuristics[39] and are grouped into required connections, forbidden connections, and entry-path constraints, as summarized in Table S1. Required connections define essential functional adjacencies such as kitchen–living room or bedroom–living room links, while forbidden connections prohibit inappropriate spatial pairings, for instance, kitchen–bathroom adjacency. Entry-path constraints ensure that circulation begins exclusively from the living room and does not directly connect the entrance to private or service spaces.

Formally, the required, forbidden, and entry-path sets are defined in Equations (2)–(4):

$$R=\{(\text{kit}_i,\text{liv}_j)\}\cup\{(\text{bed}_k,\text{liv}_j)\}\cup\{(\text{bath}_{m,\text{master}},\text{bed}_{n,\text{master}})\}\cup\{(\text{bal}_p,\text{liv}_j)\} \quad (2)$$

$$F=\{(\text{kit}_i,\text{bed}_k)\}\cup\{(\text{kit}_i,\text{bath}_q)\}\cup\{(\text{bed}_k,\text{bed}_l)\}\cup\{(\text{bal}_p,\text{bath}_{q,\text{other}})\} \quad (3)$$

$$E_{\text{in}}\subseteq\{\text{liv}_j\},|E_{\text{in}}|=1 \quad (4)$$

The checking algorithm executes in four stages: data mapping, set construction, compliance evaluation, and report generation. It identifies missing required connections, illegal adjacencies, and entry-path violations and returns a structured compliance report. Detailed pseudocode is retained in Appendix A (Algorithms A.1–A.3).

The method effectively transforms qualitative expert knowledge into a quantitative, reproducible logic framework for spatial topology assessment. When combined with the distance metric, it forms a unified "function–circulation–safety" analytical structure.

*Energy Simulation and Physical Labeling.* Full-physics simulation is computationally expensive and requires high-fidelity vector inputs, making it impractical to embed directly in a generative loop. To balance accuracy and throughput, we first vectorize RPLAN floor plans and run simulations to build the labeled dataset **DesignFD**. We then train a fast predictor **PDE** on **DesignFD** to estimate physical metrics directly from raster floor plans, enabling millisecond-level inference during generation.

**Targets (4):** (1) Energy Use Intensity (EUI), (2) Fire-safety farthest walking distance (see the fire-safety path calculation above), (3) Total floor area (12 px = 1 m), (4) Functional connectivity compliance (topology rules; see the topology-compliance procedure above).

**Automation:** Metrics (2)–(4) are computed via engineered code; EUI is obtained via simulation.

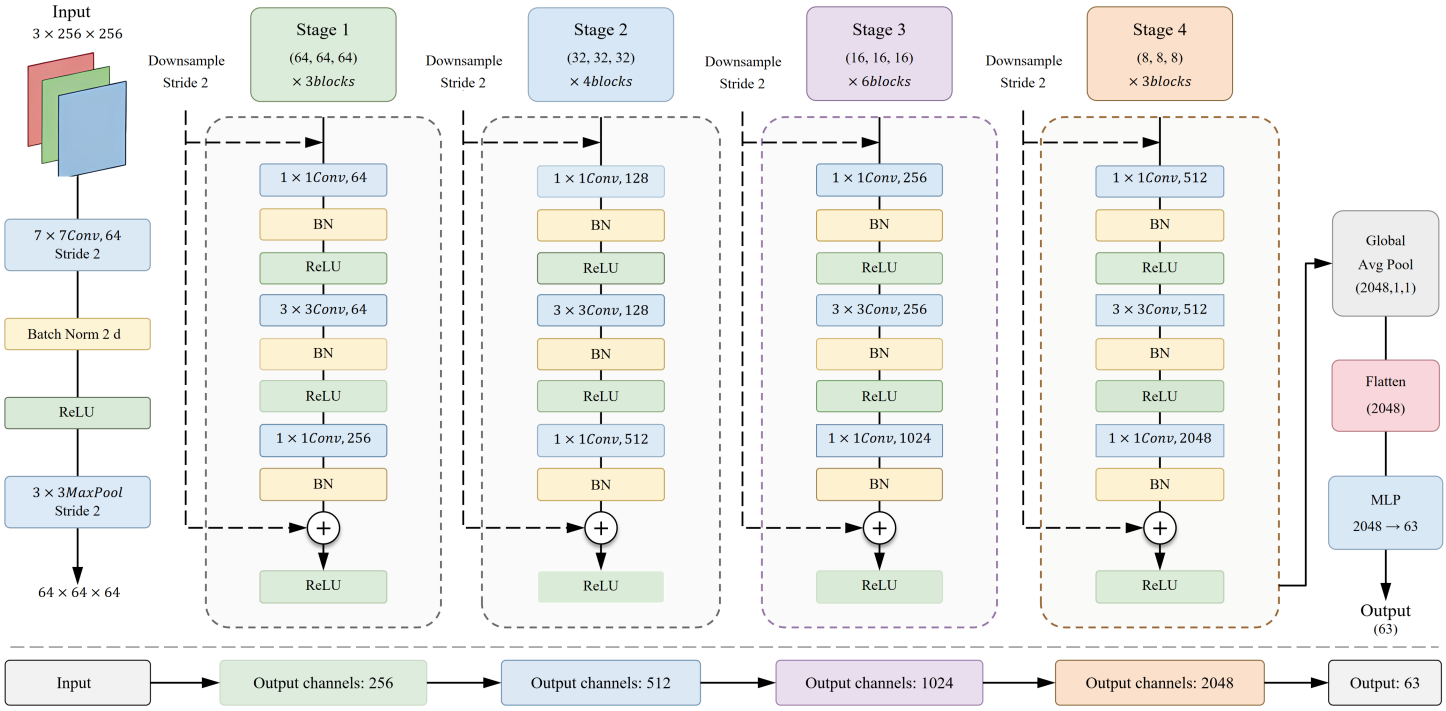



**Figure 3. PDE architecture. A ResNet backbone encodes raster plans; a lightweight head outputs physical metrics (EUI, distance, area, topology compliance). Right: example predictions versus simulation labels.**

**Usage: PDE** is trained on raster inputs with these labels for millisecond-level inference.

For energy analysis, we adopt a *monthly* simulation granularity over a full year to balance runtime with seasonal fidelity at the early design stage. Monthly EUI aligns with common regulatory benchmarks and is well-suited for concept-phase decision making.

Vector extraction converts RPLAN images into 71,125 vectorized plans. We then construct 3D models in Rhinoceros/Grasshopper with typical residential parameters: floor height 2.8 m; window width derived from the plan and a height of 1.5 m (range 0.9–2.4 m). Orientation is fixed (N up, S down, W left, E right). EnergyPlus (via ClimateStudio) computes monthly EUI for five end uses: equipment, lighting, hot water, heating, and cooling. Self-shading from balconies and overhangs is modeled; inter-building shading is excluded due to unknown surroundings, so the results reflect the plan itself.

We parameterize envelope properties, internal gains, and thermal setpoints according to national standards and prior studies, and model HVAC as ideal loads to isolate layout-driven effects. Key settings are summarized in Table S2.

The 60-dimensional energy profile used by PDE corresponds to 12 monthly values for five EnergyPlus end uses—equipment, lighting, hot water, heating, and cooling—while the remaining three outputs encode fire-safety distance, total floor area, and topology compliance.

## Practical design evaluator (PDE)

*Model Design.* The Practical Design Evaluator (PDE) predicts all DesignFD metrics directly from rasterized floorplans. Transformer architectures[40] model long-range interactions through attention, whereas convolutional networks provide locality-related inductive biases that are well suited to rasterized floor-plan geometry. Transformer-based architectures are also generally more data-demanding,[41] whereas physics-labeled residential floor-plan data remain relatively limited. We therefore adopt a convolutional regression network (illustrated in Figure 1B with architecture details provided below) for accurate and stable prediction of physical feasibility metrics.

*PDE Architecture Details.* The predictor PDE estimates the four metric groups directly from floor-plan images using a ResNet-style convolutional backbone with a lightweight fully connected head. Its 63-dimensional output comprises a 60-dimensional monthly energy-use profile and three spatial-functional outputs: farthest walking distance, total area, and topology-compliance status.

*Training Procedure.* PDE is trained using the Mean Squared Error (MSE) loss over all quantitative and qualitative targets in DesignFD, as defined in Equation (5):

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N}\sum_{i=1}^{N} ||\hat{y}_i - y_i||_2^2, \tag{5}$$

In Equation (5), $\hat{y}_i$ and $y_i$ denote the predicted and ground-truth metric vectors for the $i$-th sample, respectively. The target vector is defined in Equation (6).

$$y_i = [E_i, F_i, A_i, G_i] \in \mathbb{R}^{63}, \tag{6}$$

In Equation (6), $E_i \in \mathbb{R}^{60}$ represents the *energy-use intensity* profile comprising 60 monthly end-use values (12 months × five EnergyPlus end uses); $F_i \in \mathbb{R}$ denotes the *fire-safety distance* measuring the maximum walkable path within code limits; $A_i \in \mathbb{R}$ indicates the *floor area* reflecting geometric reasonableness; and $G_i \in \{0,1\}$ quantifies the *functional connectivity* based on room-adjacency graph consistency. Together, these 63 features comprehensively encode both energy efficiency and functional feasibility within DesignFD. As validated in previous section, PDE achieves $R^2 > 0.99$ accuracy against full EnergyPlus simulations while operating orders of magnitude faster, demonstrating strong practicality for downstream generative design tasks.

## Green plan dataset (GreenPD)

*Demand–Plan Pairing.* Green Plan Dataset (GreenPD) extends DesignFD into a paired dataset linking user requirements with residential layouts satisfying the prescribed criteria. Each sample is a pair $(d_i, x_i)$, where $d_i$ denotes the user demand and $x_i$ the corresponding floorplan. Specifically, the demand $d_i$ is composed of two parts: $d_i = (g_i, t_i)$, where (1) $g_i$ is a *layout sketch*: an image representation that encodes either the room adjacency graph (inter-room connectivity) or the exterior boundary; and (2) $t_i$ is a *design constraint*: a textual description represented as $t_i = [E_i, F_i, A_i, G_i]$, specifying the target energy-use profile, fire-safety distance, total floor area, and functional connectivity validity. This unified formulation allows GreenFlow to learn conditional generation driven jointly by spatial layout priors ($g_i$) and design constraints ($t_i$), enabling design generation that satisfies the prescribed criteria.

*Dataset Construction.* GreenPD is derived from DesignFD through PDE-

guided refinement to ensure satisfaction of all prescribed criteria. Among the 71,125 layouts in RPLAN,[20] 26,600 valid samples are directly preserved, while the remaining 44,525 undergo iterative correction. For each non-compliant footprint, FloorplanDiffusion[30] generates candidate layouts conditioned on the same boundary or room configuration. PDE then evaluates their energy and functional metrics, retaining only those that meet all feasibility criteria. This resampling process is repeated until every case satisfies the DesignFD constraints, yielding a criterion-compliant dataset for GreenFlow training. A comparative summary in Table S3 highlights the progressive evolution from RPLAN to DesignFD and GreenPD, emphasizing the dataset contributions that support feasibility-aware, criterion-compliant training through high-quality annotations.

During GreenPD construction, PDE is used as a screening evaluator rather than an optimization objective: candidate layouts are evaluated in real time and only samples satisfying all requested energy and functional criteria are retained. This filtering loop continues until a compliant replacement is obtained for each invalid case.

### GreenFlow

GreenFlow adopts a two-stage flow-matching framework for energy- and function-aware layout generation. Each demand contains a spatial condition image and a textual specification. The spatial condition is either a room-adjacency Graph or an exterior Edge/Boundary image; the Graph setting therefore does not provide an explicit Edge constraint. Text specifications encode target physical attributes, for example, "EUI: 130; Fire: True".

*Generator Architecture.* GreenFlow uses a Flow Matching generator with a DiT backbone for efficient velocity-field fitting. For image conditioning, the noisy image $x$ and conditional image $y$ are patch-embedded with positional encoding and fused according to Equation (7).

$$h_{\text{img}}=\text{MLP}\left(\left[\text{Patch}\left(x\right)+\text{PE};\text{Patch}\left(y\right)+\text{PE}\right]\right), \tag{7}$$

In Equation (7), $\text{Patch}\left(\cdot\right)$ denotes PatchEmbed and PE is positional encoding. Text conditions are encoded with OpenCLIP ViT-H/14 according to Equation (8).

$$c=\text{Embed}\left(\tau\right)+\text{MLP}\left(\text{Pool}\left(\text{CLIP}\left(\text{text}\right)\right)\right), \tag{8}$$

In Equation (8), $\tau$ denotes the diffusion/flow timestep. The DiT has hidden dimension 384, 12 transformer blocks, 6 attention heads, and patch size 2.

*Stage 1: Supervised Pre-training.*

GreenFlow is first optimized on paired GreenPD samples to learn the conditional velocity field. Let $B$ denote a mini-batch, $w_i$ the adaptive sample weight used during supervised training, and $\Delta^2_{\text{flow},i}$ the velocity-field fitting loss for sample $i$. The supervised objective is given in Equation (9).

$$\mathcal{L}_{\text{SFT}}=\mathbb{E}_{i\sim B}\left[w_i\Delta^2_{\text{flow},i}\right]. \tag{9}$$

***Algorithm 1: GreenFlow two-stage training.*** Input: paired floorplans and demands consisting of a Graph/Edge condition image and textual physical specifications; OpenCLIP text encoder; DiT flow-matching generator; PDE critic. Stage 1: encode the condition image and text specification, predict the velocity field, and optimize $\mathcal{L}_{\text{SFT}}$. Stage 2: generate layouts under the same demands, evaluate specification error with PDE, construct the reward weight, and optimize the reward-weighted flow-matching objective below.

*Stage 2: PDE-guided Reinforcement Fine-tuning.*

PDE provides a critic penalty $\varepsilon_i$, defined as the MSE between the predicted physical specifications of the generated layout and the requested specifications. GreenFlow is then optimized with reward-weighted flow matching according to Eq. (10).

$$\mathcal{L}_{\text{RL}}=\mathbb{E}_{i\sim B}\left[\text{sg}\left(\frac{1}{\left(\Delta^2_{\text{flow},i}+\varepsilon_i+c\right)^{1-\gamma}}\right)\Delta^2_{\text{flow},i}\right], \tag{10}$$

In Equation (10), $\text{sg}\left(\cdot\right)$ denotes stop-gradient, $c$ is a stabilizing constant, and $\gamma=0.5$. PDE is a differentiable neural predictor and can be evaluated inside the training graph; in this reward-weighted objective, the scalar reward weight is stop-gradient for stable optimization. During GreenPD construction, by contrast, PDE is used only as a screening evaluator to retain candidates that satisfy all feasibility criteria. This separation allows the same learned evaluator to support both large-scale dataset refinement and performance-aware generator fine-tuning.

## RESULTS

The experiments are organized around five aspects of the proposed framework: evaluator accuracy and efficiency, feasibility of the constructed dataset, comparison with existing generators, practical expert assessment, and controllability/ablation analysis. This organization separates the evidence for each component while preserving a common evaluation protocol.

### Experimental Setup

*Dataset and Baselines.* We evaluate on the RPLAN[20] dataset and our GreenPD to test model generalization and compliance. All baselines are retrained on the same split (70% train, 20% val, 10% test) for fairness. We compare GreenFlow with representative methods across major paradigms: GAN-based (HouseGAN,[17] HouseGAN++[18]), diffusion-based (HouseDiffusion,[19] FloorplanDiffusion[30]), and graph-to-image translation (Graph2Plan[15]). Language-conditioned models (Tell2Design,[16] FP-LLaMa[32]) are also included. All models are trained and tested under identical conditions.

For direct method-to-method comparison, all baselines and the GreenFlow_RPLAN variants are trained and tested on the same RPLAN split; GreenFlow_GreenPD is reported separately to isolate the effect of criterion-compliant training data. Baseline selection prioritizes representative methods with reproducible implementations that can be retrained under this common protocol. Graph denotes room-adjacency conditioning without an explicit exterior Edge constraint, whereas Edge denotes exterior boundary conditioning.

*Baseline Deficiency Handling.* Some baseline representations do not provide all exterior-opening information required for downstream energy simulation. In particular, exterior windows are not explicitly represented in the native outputs of HouseDiffusion and HouseGAN++. For the primary EUI comparison, we therefore apply minimal window completion to these outputs solely to construct the required EnergyPlus inputs. Methods (e.g., Tell2Design or HouseGAN) whose outputs lack other essential structural information needed for simulation are excluded from the corresponding energy comparison.

*Metrics.* We follow the evaluation protocol of prior layout generation works[15,17,30] and extend it with additional metrics to jointly assess functional feasibility and energy efficiency. We adopt five metrics for comprehensive evaluation: (1) **Fréchet Inception Distance (FID)**:[42] computed per layout to measure visual realism against ground truth, with scores averaged over all test samples; (2) **Intersection over Union (IoU)**:[43] quantifies geometric alignment between generated and ground-truth layouts; (3) **Functional Rationality**: fraction of layouts satisfying fire-safety, area, and connectivity constraints; (4) **Energy Use Intensity (EUI)**: estimated annual energy demand per unit area. Both Rationality and EUI are predicted by PDE. EUI predictions are validated against full EnergyPlus simulations, whereas Rationality is evaluated against the deterministic spatial and topology criteria used to construct DesignFD.

(5) Graph Edit Distance (GED) measures the edit distance between the input and generated room-adjacency graphs, following HouseDiffusion. Pass rate denotes the fraction of generated results that satisfy the specified target constraints.

*Implementation Details.* All experiments are implemented in PyTorch and trained on a single NVIDIA RTX A100 GPU. **Practical Design Evaluator (PDE)** uses a ResNet-50 backbone[44] with standard residual blocks and batch normalization. Features are processed by global adaptive average pooling and mapped to a 63-dimensional output (3 for spatial-functional feasibility and 60 for energy-related metrics). The input resolution is $64\times64$. Training uses the Adam optimizer with an initial learning rate of $1\times10^{-4}$ and exponential decay. Early stopping (patience=3) is used to avoid overfitting. **GreenFlow** uses the DiT-based Flow Matching architecture described in Sec. 3.6 and is optimized with AdamW.[45] Stage 1 (supervised pre-training) runs for 200k iterations with batch size 16 and learning rate $1\times10^{-4}$, followed by Stage 2 (reinforcement fine-tuning) for 30k iterations with batch size 16 and learning rate $1\times10^{-5}$, using PDE feedback as reward signals.

The DiT configuration is dim=384, 12 blocks, 6 heads, and patch size 2, with OpenCLIP ViT-H/14 for textual specifications. For raster wall refinement, progressive denoising proceeds from low-frequency to high-frequency color information.

### PDE accuracy and efficiency

We evaluate the Practical Design Evaluator (PDE) against physics-based *EnergyPlus* simulations. Table S4 reports three standard regression metrics: $R^2$ (coefficient of determination) measures correlation with ground-truth variance, MSE (mean squared error) quantifies overall deviation, and runtime reflects total inference time over 100 runs. As shown, PDE achieves an $R^2$ of 0.99 and an MSE of 0.46, demonstrating near-perfect consistency with EnergyPlus while being approximately $1.7\times10^5$ times faster for 100-layout evaluation (7.3 ms vs. $1.21\times10^6$ ms). These results confirm that PDE accurately predicts energy and spatial feasibility from rasterized floorplans, enabling rapid and scalable evaluation for generative design.

An additional 1,000-sample validation further tests generalization across representation types: PDE achieves $R^2$=0.96 on model-generated raster images and $R^2$=0.99 on vectorized floor-plan images, supporting its use as a fast evaluator inside the generation pipeline.

We compare PDE with representative baselines—ResNet, EfficientNet, and SwinTransformer—using $R^2$ (primary), MSE, and MAE; we also report standard deviation and 95% confidence intervals. As summarized in Table S5, PDE achieves the highest $R^2$ and the lowest errors on the test set.

Quantitative visualizations in Figure 4 highlight the alignment between predictions and simulations and help reveal any distributional bias. The scatter plot of true versus predicted EUI shows PDE closest to the ideal diagonal, while the line plot over 30 sample layouts indicates that PDE tracks simulated totals most closely.

Finally, we compare runtime with EnergyPlus to illustrate practical efficiency. As shown in Table S6, PDE delivers near-constant per-batch latency at the millisecond level and achieves orders-of-magnitude speedups over simulation, which makes it suitable for tight generative loops and large-scale screening.

### GreenPD feasibility and data quality

*Existing datasets contain extensive infeasible data.* Because the existing dataset[20] was constructed without enforcing building-code constraints and contains outdated or loosely labeled samples, our analysis reveals clear deficiencies in both energy and functional feasibility. As shown in Figure 5, about 30.9% of layouts exceed the EUI benchmark (EUI > 135), 10.0% exceed the 15 m path-length criterion, 9.7% exceed the 130 $m^2$ floor-area criterion, and 27.5% fail the functional-connectivity criterion. Consequently, only 37.4% of floor plans satisfy all four benchmark criteria simultaneously, indicating substantial infeasibility within the original training distribution.

*Proposed GreenPD effectively corrects infeasible data.* In contrast, our proposed GreenPD dataset (see Figure 7) achieves 100% compliance with the four evaluated benchmark criteria by filtering and regenerating invalid samples. This ensures a clean, physically grounded dataset for model training, substantially improving both the functional feasibility and energy efficiency of generated layouts.

### Comparison with existing floorplan generators

*Quantitative Comparison.* Several baseline representations (e.g., HouseDiffusion and HouseGAN++) do not explicitly provide exterior-window information in their native outputs, while others (e.g., Tell2Design or HouseGAN) lack other structural information required by the downstream energy model. Quantitative EUI results are therefore reported only under the standardized simulation protocol described in previous section.

Table S7 shows that GreenFlow consistently outperforms all baselines across both spatial and energy metrics. It achieves the lowest FID and the highest IoU among all methods, indicating that our generated layouts are both realistic and functional consistent. Under the graph setting, attains the best layout fidelity (FID = 13.1) and spatial accuracy (IoU = 0.84), significantly surpassing FP-LLaMa (28.2 / 0.12) and HouseDiffusion (33.2 / 0.16). It also achieves the lowest GED (2.5) and the highest Rationality (61.6%), confirming superior spatial-functional feasibility. For energy efficiency, models trained *without* explicit constraints (⋆) reduce mean EUI by 2.5% (graph) and 1.9% (edge). When green constraints are applied, the improvements grow substantially to 10.4% and 13.1%, respectively, showing that our PDE-guided flow-matching generation can effectively balance spatial organization with energy objectives. Compared with HouseDiffusion and FP-LLaMa, GreenFlow achieves consistent gains across all indicators, regardless of input modality. These results highlight that integrating criterion-compliant data (GreenPD) with performance-aware generation enables our model to jointly optimize geometric fidelity, functional feasibility, and physical performance.

*Comparison Without Post-processing.* Because exterior windows are not explicitly represented in the native outputs of several baseline methods, the primary EUI comparison applies minimal window completion only where required to construct the EnergyPlus input. To verify that the main conclusions do not depend on this completion procedure, we additionally evaluate a window-free setting that preserves the original model outputs.

GreenFlow remains clearly superior in the window-free setting, indicating that its gains are not an artifact of the opening-completion procedure used in the primary energy comparison for baselines requiring opening completion.

*Qualitative Comparison.* Several baseline representations do not provide complete opening information required by the downstream energy model. To standardize the primary energy comparison, minimal exterior-window completion is therefore applied where necessary. This preprocessing only supplies the missing simulation inputs and does not alter the generated room geometry or circulation structure. As shown in Figure 6, differences in circulation continuity, room accessibility, and façade organization remain observable after this completion. In contrast, GreenFlow jointly generates spatial organization and the opening information required by the evaluation pipeline, supporting both functional circulation and energy assessment.

### Human expert evaluation

We conducted a user study with 13 experts (4 architects and 9 graduate students) to evaluate *usability*, *soundness*, and *efficiency*. Usability and soundness were rated on a 100-point scale, and efficiency was measured by time saved compared with manual design. As shown in Table S9, GreenFlow achieved the highest ratings (89.4 in usability and 90.2 in soundness) thanks to its built-in evaluator that removes the need for external simulations. Manual design took 24.7 minutes on average, while GreenFlow completed the same task in only 3.2 minutes (1.7 generations, 61s each), resulting in an *87.0% improvement in efficiency*. These findings show that GreenFlow makes design work faster, easier, and more reliable.

### Ablation and controllability analysis

We next examine whether GreenFlow responds systematically to energy and functional targets and isolate the contributions of compliant training data, PDE-guided fine-tuning, and conditioning modality. The controllability figures establish the model response across target EUI levels, while the ablations quantify which components are responsible for the observed gains.

*Ablation Settings.* We conduct ablation studies to examine how dataset quality, RL fine-tuning, and input modality affect layout controllability:

The tested target values of 110–140 EUI were selected to cover the central portion of the approximately 100–150 range encountered in the residential simulation setting, while the 105–135 visualization in Figure 8 additionally probes the low-energy boundary where functional and energy objectives can conflict.

- **w/o RL & GreenPD (Base):** trained only on the RPLAN dataset[20] without any energy or physical feedback, serving as a weak baseline for controllability.
- **w/o RL:** trained on our criterion-compliant *GreenPD* dataset using both *Graph* and *Edge* inputs, but without PDE-based reinforcement learning. This isolates the effect of RL optimization and is evaluated under four target energy levels (110–140 EUI).
- **Ours (Full):** GreenFlow model trained on GreenPD and fine-tuned with PDE-guided reinforcement learning for joint optimization of energy and functional feasibility.

*Effect of Dataset Quality.* Replacing the noisy RPLAN dataset with our criterion-compliant GreenPD significantly improves layout stability, energy alignment, and spatial-functional feasibility. As shown in Table S7 and Figure 7, models trained on GreenPD exhibit lower variance and stronger

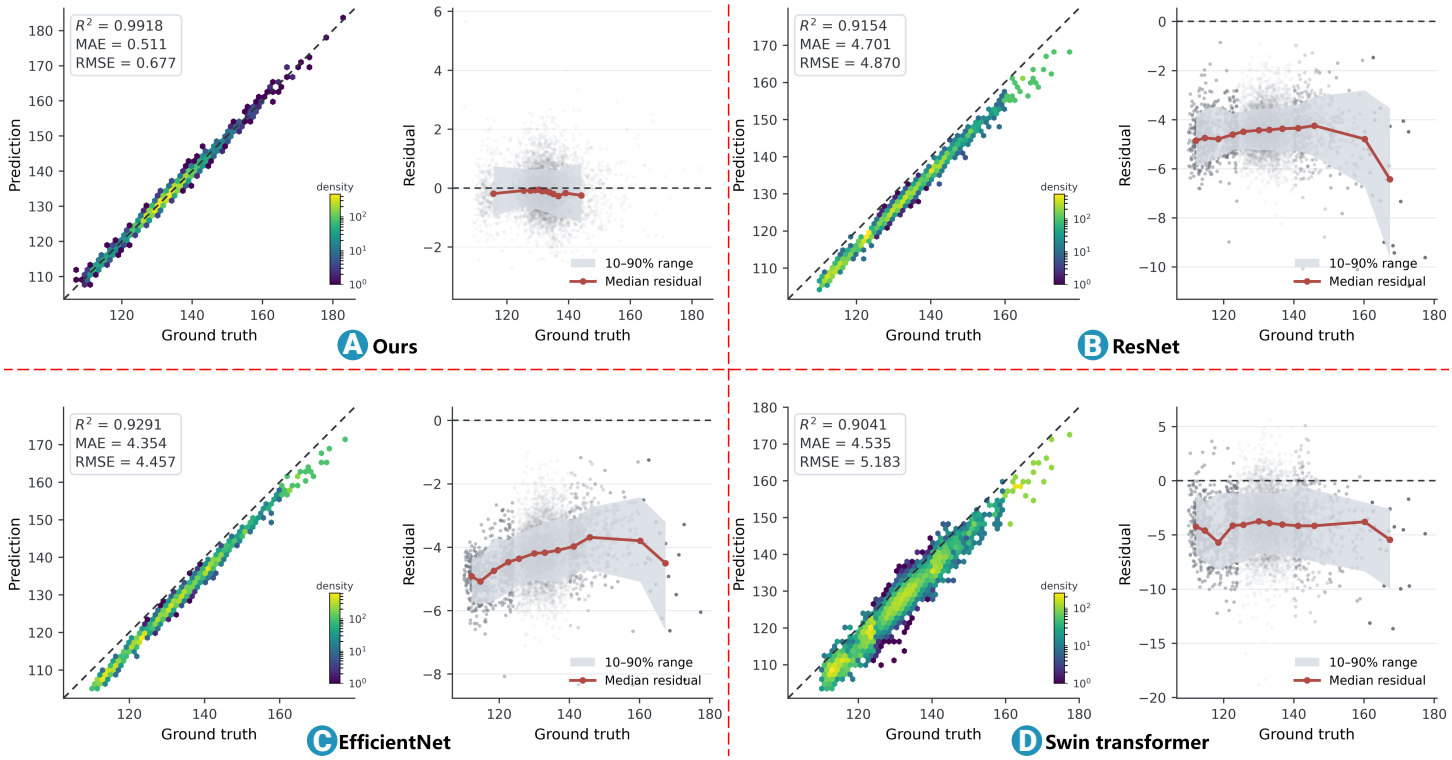



**Figure 4. Quantitative comparison of energy predictions** Left: scatter of true vs. predicted EUI (kWh/m$^2$a) with the red dotted line indicating perfect fit; PDE is closest to the ideal. Right: total EUI across 30 sample plans (index on x-axis), where PDE tracks simulation most closely.

correspondence to target EUI values compared with those trained on RPLAN. This demonstrates that consistent, criterion-compliant data are essential for reliable, controllable layout generation.

*Effect of RL Fine-tuning.* Comparing the GreenPD-only variant (w/o RL) with the RL-enhanced model (Full) shows consistent improvements in pass rate across both input modes (Table S10). Notably, for Fire, the pass rate increases from 38.75% to 79.38% under the Graph setting and from 84.38% to 96.25% under Edge, corresponding to gains of 40.6 and 11.9 percentage points, respectively. Together with the tighter EUI control shown in Figure 7, these results demonstrate that RL fine-tuning effectively enhances energy–function coupling and overall design feasibility.

*Effect of Input Modality.* To examine the influence of input modality, we analyze results under Graph and Edge settings. Both modalities achieve effective control but show distinct behaviors: Edge inputs provide more direct control over total area and better adherence to ultra-low targets (110 EUI), while Graph inputs offer richer spatial relationships and higher functional realism. As visualized in Figure 8, the model adaptively expands or suppresses high-energy functional zones (e.g., kitchens, bathrooms) as target EUI increases, maintaining a balance between physical performance and functional feasibility.

*Joint Graph-and-Edge Conditioning.* We also test Graph and Edge as joint spatial inputs under the same evaluation protocol. The joint condition combines explicit exterior geometry with room-adjacency information and therefore provides the strongest spatial prior.

The joint-input result improves the spatial metrics further, while the separate Graph and Edge settings remain useful for isolating the contribution of each conditioning modality.

## DISCUSSION

### Performance-aware generation versus post-hoc selection

A fast evaluator naturally enables a generate-then-select workflow: an existing generator can sample multiple candidates and PDE can rank or filter them. This strategy is useful as a deployment baseline and GreenPlanner itself uses high-throughput screening during GreenPD construction. However, selection acts only after samples have been proposed. When the feasible region is narrow, a generator that has not learned the physical constraints can continue assigning substantial probability mass to invalid layouts, so achieving one compliant candidate may require repeated sampling. The computational burden then shifts from evaluation to proposal generation, and the generator remains insensitive to why previous candidates were rejected.

GreenPlanner therefore uses PDE in two complementary roles. During data construction, PDE supports large-scale screening and correction so that GreenPD presents the generator with consistently feasible examples. During GreenFlow fine-tuning, PDE supplies a differentiable performance signal that changes the learned generation distribution itself. The ablation results support this distinction: compliant training data improve stability and target alignment, while PDE-guided fine-tuning further increases constraint pass rates, particularly for fire-safety compliance.

From a practical design standpoint, the difference between the two strategies has direct workflow implications. In a post-hoc selection regime, designers or automated pipelines must repeatedly generate, evaluate, and discard candidates without receiving directional feedback on how to improve the next proposal. This trial-and-error loop is especially costly during early-stage design, when the goal is rapid exploration of a broad design space under simultaneous energy and functional requirements. By contrast, a generator that has internalized performance constraints produces proposals that already concentrate within the feasible region, reducing the number of iterations required and allowing early-stage exploration to focus on comparing viable alternatives rather than searching for them. The expert study reflects this advantage: GreenFlow completed usable designs in an average of 3.2 minutes with 1.7 generation rounds, compared with 24.7 minutes for the tested manual workflow, suggesting that embedding evaluation into generation can meaningfully compress the early design cycle.

### Design trade-offs and practical implications

Energy performance and spatial-functional quality are coupled objectives rather than independent labels. Changing room proportions, exterior exposure, circulation, openings, or the allocation of high-load spaces can improve one objective while making another harder to satisfy. The controllability experiments make this interaction visible: GreenFlow responds to different EUI targets, but at very restrictive targets the model can approach a boundary where energy reduction conflicts with functional requirements. This behavior is important because a practically useful design system should

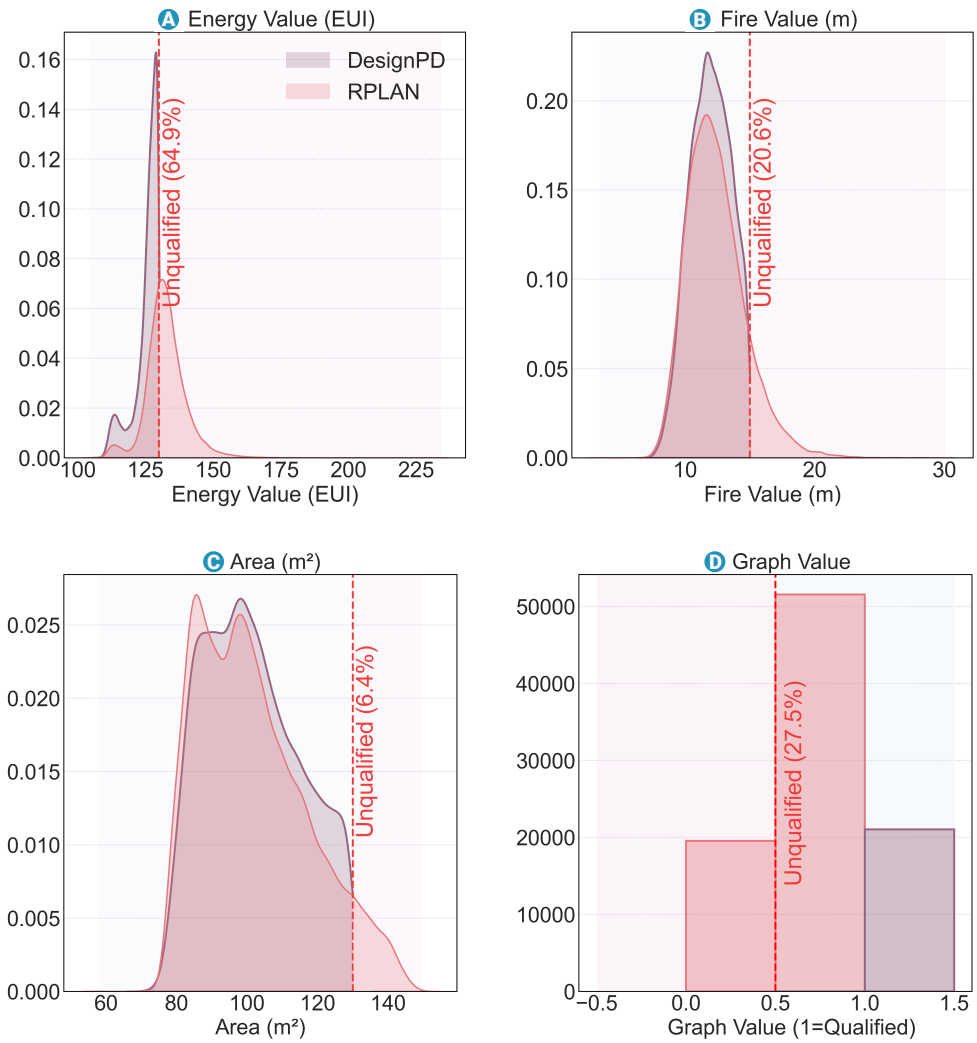



**Figure 5. Comparison between the original RPLAN dataset and our proposed GreenPD** While RPLAN contains numerous violations in energy, safety, and functional constraints, GreenPD achieves complete compliance with the four evaluated benchmark criteria.

expose such trade-offs instead of optimizing a single numerical objective at the expense of architectural usability.

Accordingly, we view GreenPlanner primarily as an early-stage design exploration and screening framework. PDE can evaluate large candidate sets in milliseconds, and the expert study indicates that GreenFlow can substantially reduce the time required to obtain usable alternatives while improving perceived usability and soundness. These properties are well suited to rapid option generation, architect-in-the-loop refinement, and preliminary performance comparison.

At the same time, the scope of PDE as a surrogate must be clearly understood. PDE approximates the physics-based simulation and rule-based checking procedures used to create its training labels; it does not replace them. Its predictions are valid within the data distribution, simulation configuration, and rule set on which it was trained. Changes in climate assumptions, envelope specifications, HVAC modeling detail, or regulatory criteria may reduce prediction reliability until the evaluator is retrained on correspondingly updated labels. In a practical deployment, PDE is therefore best positioned as a fast screening tool for iterative exploration and candidate ranking, while full EnergyPlus simulation and formal code-compliance review remain necessary for final design verification and regulatory approval.

## Scope, limitations, and future Work

Several limitations define the current scope. In each case we clarify whether the limitation reflects an inherent constraint of the proposed framework or a boundary condition of the data and simulation setup used in this study.

The most consequential limitation for energy generalizability is the use of a single climate and a fixed set of envelope and operational assumptions. All energy labels are generated with the Beijing EPW climate file, standardized U-values, ideal-load HVAC, and uniform schedules. This configuration is a deliberate experimental choice: by holding climate and envelope parameters constant, the reported EUI differences isolate the effect of layout geometry on energy performance, which is the variable that a floorplan generator can directly control. The objective of this study is to demonstrate the technical feasibility of integrating fast energy evaluation into the generative design pipeline, not to claim universal energy savings across all regions and building specifications. The framework itself is climate-agnostic: DesignFD labels can be regenerated under any EnergyPlus-compatible climate file and envelope configuration, and PDE can be retrained or extended with climate metadata as an additional input. Validating this transferability across representative climate zones and construction standards is a priority for future work.

Second, the EnergyPlus setup accounts for self-shading from elements such as balconies and overhangs but does not model inter-building shading. This is a data-availability constraint rather than a modeling limitation: RPLAN provides individual residential floor plans without surrounding urban context, so neighboring-building geometry is unavailable. In dense urban settings, mutual shading among adjacent buildings can materially affect heating and cooling loads. Incorporating contextual geometry, either from urban datasets or as parameterized surroundings, would improve the fidelity of energy estimates for real deployment scenarios and represents a natural extension once plan-level feasibility has been established.

Third, the experiments inherit the predominantly Manhattan-style geometry of RPLAN. This is likewise a characteristic of the available large-scale residential layout corpus rather than a hard architectural restriction of the framework: the generation and evaluation components operate on rasterized plans and do not assume axis-aligned walls. However, because training and evaluation are conducted entirely on Manhattan-style layouts, the present results do not establish performance on highly irregular or curved residential forms. Extending the training corpus to include non-Manhattan geometry would allow the framework to be assessed on a broader range of residential typologies.

Fourth, the current topology and functional rules encode a finite set of residential design requirements. This is a scope boundary rather than a structural limitation: the checking mechanism accepts configurable rule sets, so additional jurisdictions and project types require corresponding rule adaptation and validation rather than redesigning the evaluation pipeline.

Finally, under highly restrictive multi-objective targets, GreenFlow can occasionally trade one specification against the majority of constraints. This is a general challenge in multi-objective constrained generation rather than a deficiency specific to the proposed architecture. Future work should investigate hard-constraint projection or hierarchical objective handling to provide stronger guarantees on individual criteria satisfaction without sacrificing overall generation quality.

## CONCLUSION

This study presented GreenPlanner, a performance-aware framework for residential floorplan generation that couples design evaluation, data construction, and generative optimization. Instead of treating energy analy-

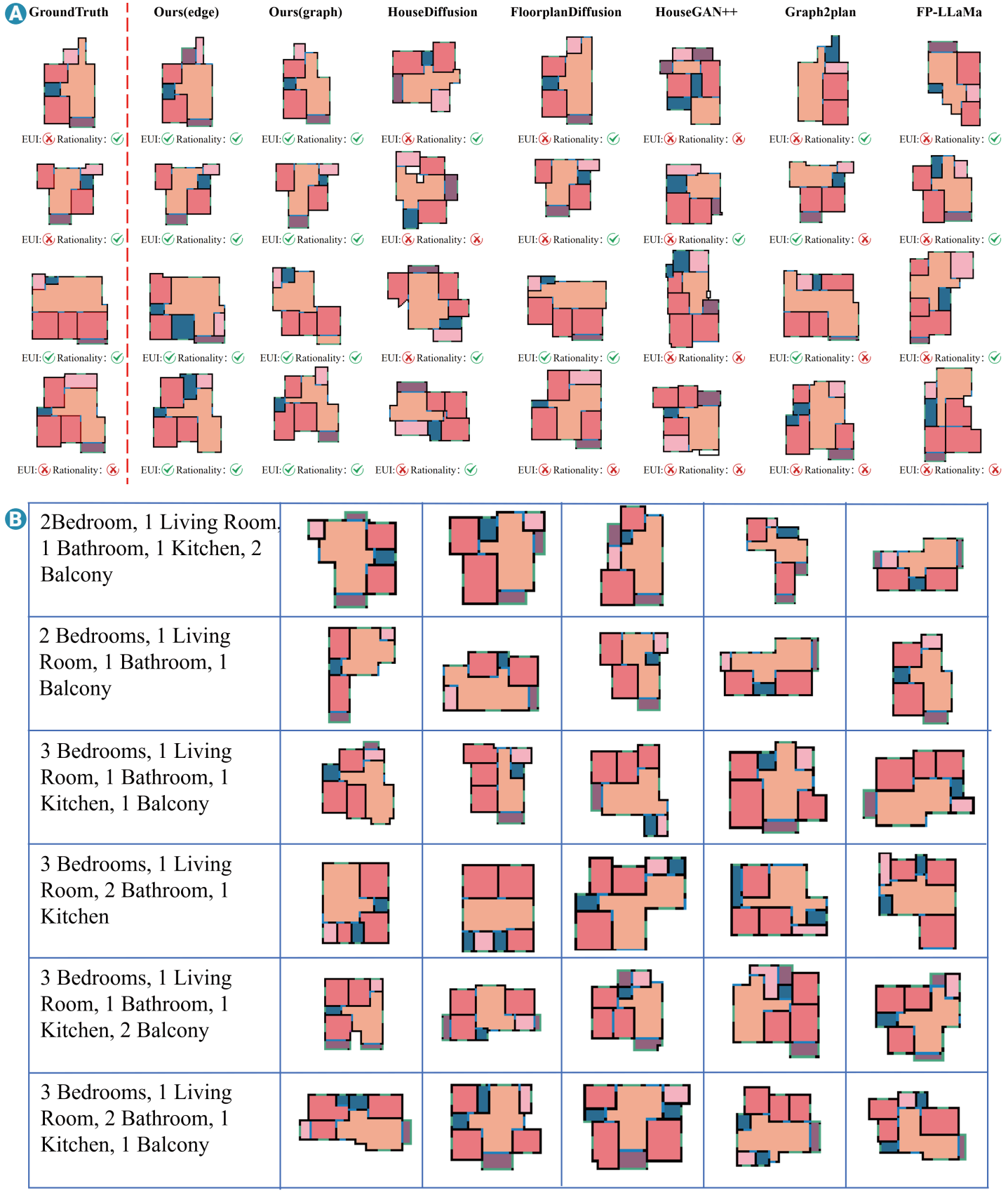



**Figure 6. (A) Qualitative comparison of layout generation results. Baseline models (e.g., HouseDiffusion, HouseGAN++) often fail to produce necessary openings (doors and windows), resulting in functional incomplete layouts and unrealistic energy performance. For fair evaluation, a small number of openings were randomly inserted into their outputs to enable energy assessment. Our GreenFlow generates layouts with coherent spatial organization and complete openings. (B) More generated results by GreenPlanner.**

sis and functional checking as downstream post-processing, GreenPlanner incorporates them throughout the learning pipeline: DesignFD provides automatically derived physical and spatial-functional labels; PDE learns a fast surrogate for feasibility assessment; GreenPD converts these assessments

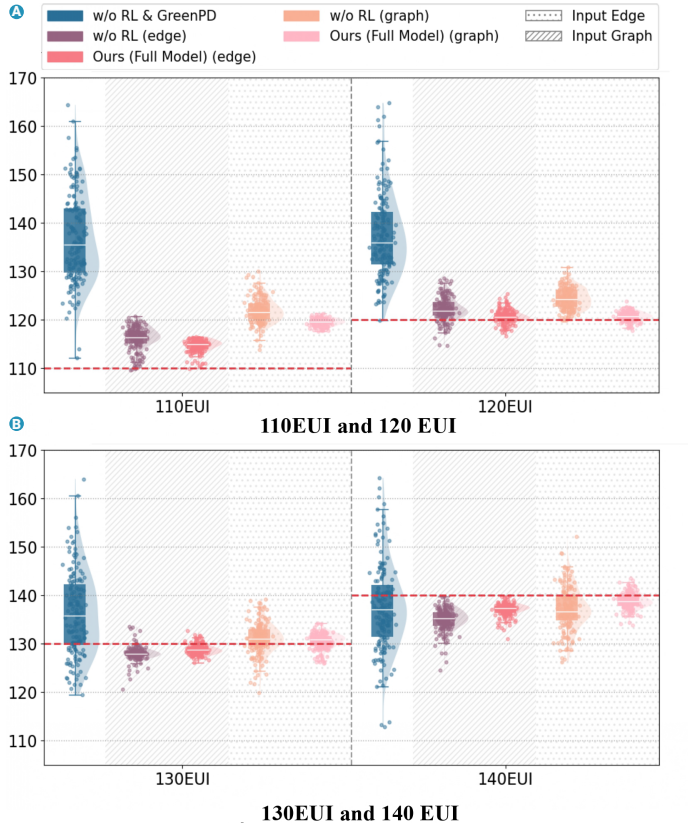



**Figure 7. Energy controllability under different energy targets. Each model is evaluated at four target EUI levels (110, 120, 130, 140)** Models trained on GreenPD (w/o RL and Full Model) are tested under both Graph and Edge inputs. Deviations at 110 EUI arise from functional constraints (e.g., kitchens).

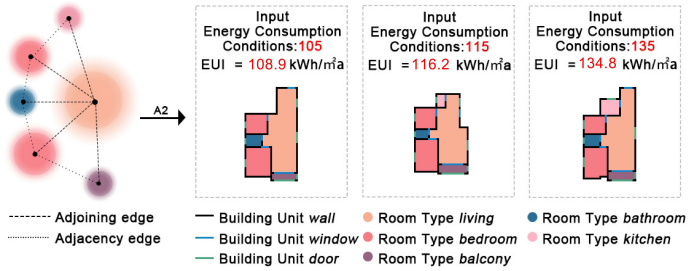



**Figure 8. Controllability of green-performance attributes** Layouts generated under increasing target energy levels (105–135 EUI) adaptively expand high-energy functional zones (e.g., kitchens, bathrooms). At ultra-low targets, conflicts between functionality and energy objectives reveal the model's adherence to both energy and functional feasibility.

into a criterion-compliant training distribution; and GreenFlow combines supervised learning with PDE-guided fine-tuning so that energy and functional constraints influence the generation process directly.

The central finding is that early-stage residential layout decisions can be systematically guided toward improved energy performance when fast, sufficiently accurate evaluation is integrated into both data preparation and generation. PDE reaches $R^2$ = 0.99 with an MSE of 0.46 while evaluating 100 layouts in 7.3 ms, making repeated feasibility assessment computationally practical at a scale that would be prohibitive with full physics-based simulation. Built on this capability, GreenPD achieves 100% compliance with the four evaluated benchmark criteria, and GreenFlow reduces mean EUI by 10.4% and 13.1% under graph- and edge-conditioned settings, respectively, while maintaining improved spatial-functional quality. The expert study further confirms practical utility, with an 87.0% reduction in design time relative to the tested manual workflow.

These results support the broader proposition that generative building design benefits from integrating rapid performance evaluation with generation rather than optimizing geometric plausibility alone. GreenPlanner is intended for early-stage exploration and rapid option screening, with full simulation and professional review retained for final verification. Extending the framework across climates, construction assumptions, urban-context effects, and stricter multi-objective handling represents the key path toward more general performance-aware architectural design support.

## Appendix A. Detailed Topology-Checking Algorithms

The main text presents the topology rules, formal constraint sets, and validation results. For reproducibility, the low-level checking procedures are retained here as pseudocode.

**Algorithm A.1:** Topology compliance checking

**Input:** room attributes, observed adjacency list, and entry-room set
**Output:** structured compliance report (missing required connections, illegal edges, and status)
**Step 1:** Map room attributes and group spaces by functional type (living room, bedroom, kitchen, etc.);
**Step 2:** Construct the required-connection set *R* and forbidden-connection set *F* using the supporting procedure;
**Step 3:** Compare the required and forbidden sets with the observed adjacency set *Areal* and evaluate the entry-path condition *Ein*;
**Step 4:** Return the structured compliance indicators.

**Algorithm A.2:** Construction of required and forbidden connection sets

**Input:** grouped room IDs (kitchen, bedroom, bathroom, master bedroom, balcony, secondary space, and living room)
**Output:** required set R and forbidden set F
$R \leftarrow \emptyset$; $F \leftarrow \emptyset$;
**foreach** *kitchen* and *living-room* pair **do**
add (*kitchen*, *living room*) to *R*;
**foreach** *bedroom* and *living-room* pair **do**
add (*bedroom*, *living room*) to *R*;
**foreach** *kitchen* and *bedroom* pair **do**
add (*kitchen*, *bedroom*) to *F*;
**return** *R*, *F*;

**Algorithm A.3:** Identification of missing required connections

**Input:** required set R, living-room IDs, and observed adjacency set A_real
**Output:** list of missing required connections
*missing* ← [ ];
**foreach** required placeholder pair (*ent*, *liv_ph*) ∈ *R* **do**
*targets* ← {(*ent*, *liv*) | *liv* belongs to living-room IDs};
**if** *targets* ∩ *A_real* = ∅ **then**
append (*ent*, "LIV") to *missing*;
**return** *missing*;

## REFERENCES


1. Dales R., Liu L., Wheeler A.J., et al. (2008). Quality of indoor residential air and health. *CMAJ* **179**:147–152. DOI:10.1503/cmaj.070359
2. Liang Y., Li C., Liu Z., et al. (2023). Decarbonization potentials of the embodied energy use and operational process in buildings: A review from the life-cycle perspective. *Heliyon* **9**:e20190. DOI:10.1016/j.heliyon.2023.e20190
3. Zhang H. (2023). Leveraging policy instruments and financial incentives to reduce embodied carbon in energy retrofits. *arXiv*:2304.03403.
4. Zhang S., Ma M., Zhou N., et al. (2024). GLOBUS: Global building renovation potential by 2070. *arXiv*:2406.04133.
5. Zeng P., Yin J., Huang Y., et al. (2025). AI-based generation and optimization of energy-efficient residential layouts controlled by contour and room number. *Build. Simul.* **18**:2777–2805. DOI:10.1007/s12273-025-1337-4
6. Baduge S.K., Thilakarathna S., Perera J.S., et al. (2022). Artificial intelligence and smart vision for building and construction 4.0: Machine and deep learning methods and applications. *Autom. Constr.* **141**:104440. DOI:10.1016/j.autcon.2022.104440.
7. Eberhardt L.C.M., Birkved M. and Birgisdottir H. (2022). Building design and construction strategies for a circular economy. *Archit. Eng. Des. Manag.* **18**:93–113. DOI:10.1080/17452007.2020.1781588
8. Himeur Y., Elnour M., Fadli F., et al. (2023). AI-big data analytics for building automation and management systems: A survey, actual challenges and future perspectives. *Artif. Intell. Rev.* **56**:4929–5021. DOI:10.1007/s10462-022-10286-2

9. Chen H., Xiang Q., Hu J., et al. (2025). Comprehensive exploration of diffusion models in image generation: A survey. *Artif. Intell. Rev.* **58**:99. DOI:10.1007/s10462-025-11110-3
10. Li J., Zhang C., Zhu W., et al. (2025). A comprehensive survey of image generation models based on deep learning. *Ann. Data Sci.* **12**:141–170. DOI:10.1007/s40745-024-00544-1
11. Wang B., Chen Q. and Wang Z. (2025). Diffusion-based visual art creation: A survey and new perspectives. *ACM Comput. Surv.* **57**:268. DOI:10.1145/3728459
12. Comanici G., Bieber E., Schaekermann M., et al. (2025). Gemini 2.5: Pushing the frontier with advanced reasoning, multimodality, long context, and next generation agentic capabilities. *arXiv*:2507.06261.
13. Hurst A., Lerer A., Goucher A.P., et al. (2024). GPT-4o system card. *arXiv*:2410.21276.
14. Xue H., Luo X., Hu Z., et al. (2025). Human motion video generation: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **47**:10709–10730. DOI:10.1109/TPAMI.2025.3594034
15. Hu R., Huang Z., Tang Y., et al. (2020). Graph2Plan: Learning floorplan generation from layout graphs. *ACM Trans. Graph.* **39**(4):118. DOI:10.1145/3386569.3392391
16. Leng S., Zhou Y., Dupty M.H., et al. (2023). Tell2Design: A dataset for language-guided floor plan generation. *Proceed. 61st Ann. Meet. Assoc. Comput. Ling.* **1**:14680–14697. DOI:10.18653/v1/2023.acl-long.820.
17. Nauata N., Chang K.-H., Cheng C.-Y., et al. (2020). House-GAN: Relational generative adversarial networks for graph-constrained house layout generation. *Computer Vision ECCV* **2020**:162–177. DOI:10.1007/978-3-030-58452-8_10
18. Nauata N., Hosseini S., Chang K.-H., et al. (2021). House-GAN++: Generative adversarial layout refinement network towards intelligent computational agent for professional architects. *Proceed IEEE CVF CVPR*:13632–13641. DOI:10.1109/CVPR46437.2021.01342.
19. Shabani M.A., Hosseini S. and Furukawa Y. (2023). HouseDiffusion: Vector floorplan generation via a diffusion model with discrete and continuous denoising. *Proceed. IEEE CVF CVPR*:5466–5475. DOI:10.1109/CVPR52729.2023.00529.
20. Wu W., Fu X.-M., Tang R., et al. (2019). Data-driven interior plan generation for residential buildings. *ACM Trans. Graph.* **38**:234. DOI:10.1145/3355089.3356556
21. Xu M., Lou Y., Gao X., et al. (2025). Floorplan-Diffusion: Automatic floor plan generation via pre-trained large latent diffusion model. *Proceed. 2025 Int. Conf. Multi. Retr.*:1617–1625. DOI:10.1145/3731715.3733343.
22. Zeng P., Gao W., Li J., et al. (2025). Automated residential layout generation and editing using natural language and images. *Autom. Constr.* **174**:106133. DOI:10.1016/j.autcon.2025.106133
23. Zeng P., Yin J., Gao Y., et al. (2025). Comprehensive and dedicated metrics for evaluating AI-generated residential floor plans. *Buildings* **15**:1674. DOI:10.3390/buildings15101674
24. Zeng P., Yin J., Zhang M., et al. (2025). CARD: Cross-modal agent framework for generative and editable residential design. *Proceed. 2025 Conf. Emp. Meth. Nat. Lang. Process.*:9304–9319. DOI:10.18653/v1/2025.emnlp-main.473.
25. Zeng P., Yin J., Zhang M., et al. (2025). Unified residential floor plan generation with multimodal inputs. *Autom. Constr.* **178**:106408. DOI:10.1016/j.autcon.2025.106408
26. Chen Y., Qin Z., Sun L., et al. (2025). GDT framework: Integrating generative design and design thinking for sustainable development in the AI era. *Sustainability* **17**:372. DOI:10.3390/su17010372
27. Lee J., Cho W., Kang D., et al. (2023). Simplified methods for generative design that combine evaluation techniques for automated conceptual building design. *Appl. Sci.* **13**:12856. DOI:10.3390/app132312856
28. Li T., Cheng C.-Y., Xie A., et al. (2024). Revision matters: Generative design guided by revision edits. *arXiv*:2406.18559.
29. Regenwetter L., Srivastava A., Gutfreund D., et al. (2023). Beyond statistical similarity: Rethinking metrics for deep generative models in engineering design. *Comput. Aided Des.* **165**:103609. DOI:10.1016/j.cad.2023.103609
30. Zeng P., Gao W., Yin J., et al. (2024). Residential floor plans: Multi-conditional automatic generation using diffusion models. *Autom. Constr.* **162**:105374. DOI:10.1016/j.autcon.2024.105374
31. Feng W., Zhu W., Fu T.-J., et al. (2023). LayoutGPT: Compositional visual planning and generation with large language models. *Adv. Neural Inf. Process. Syst.* **36**:18225–18250.
32. Yin J., Zeng P., Sun H., et al. (2025). FloorPlan-LLaMa: Aligning architects' feedback and domain knowledge in architectural floor plan generation. *Proceed. 63rd Ann. Meet. Assoc. Comput. Ling.* **1**:6640–6662. DOI:10.18653/v1/2025.acl-long.331.
33. Eastman C., Lee J.-M., Jeong Y.-S., et al. (2009). Automatic rule-based checking of building designs. *Autom. Constr.* **18**:1011–1033. DOI:10.1016/j.autcon.2009.07.002
34. Liu M., Zhang L., Chen J., et al. (2025). Large language models for building energy applications: Opportunities and challenges. *Build. Simul.* **18**:225–234. DOI:10.1007/s12273-025-1235-9
35. Zeng P., Yin J., Sun H., et al. (2025). MRED-14: A benchmark for low-energy residential floor plan generation with 14 flexible inputs. *Proceed. 33rd ACM Int. Conf. Multimedia*:11298–11307. DOI:10.1145/3746027.3754949.
36. Lehmann M. (2024). The definitive guide to policy gradients in deep reinforcement learning: Theory, algorithms and implementations. *arXiv*:2401.13662.
37. Schulman J., Wolski F., Dhariwal P., et al. (2017). Proximal policy optimization algorithms. *arXiv*:1707.06347.
38. Hanson J. (2003). Decoding Homes and Houses. Cambridge University Press.
39. Architectural Society of China. (2017). Architectural Design Data Collection, 3rd ed., Vol. 2: Residential. China Architecture & Building Press, Beijing, China. ISBN 978-7-112-20940-8.
40. Vaswani A., Shazeer N., Parmar N., et al. (2017). Attention is all you need. *Adv. Neural Inf. Process. Syst.* **30**:5998–6008. https://papers.neurips.cc/paper/7181-attention-is-all-you-need.pdf.
41. Touvron H., Cord M., Douze M., et al. (2021). Training data-efficient image transformers & distillation through attention. *Proc. Mach. Learn. Res.* **139**:10347–10357. https://proceedings.mlr.press/v139/touvron21a.html.
42. Heusel M., Ramsauer H., Unterthiner T., et al. (2017). GANs trained by a two time-scale update rule converge to a local Nash equilibrium. *Adv. Neural Inf. Process. Syst.* **30**:6626–6637. DOI:10.5555/3295222.3295408
43. Everingham M., Van Gool L., Williams C.K.I., et al. (2010). The PASCAL visual object classes (VOC) challenge. *Int. J. Comput. Vis.* **88**:303–338. DOI:10.1007/s11263-009-0275-4
44. He K., Zhang X., Ren S., et al. (2016). Deep residual learning for image recognition. *IEEE CVPR*:770–778. DOI:10.1109/CVPR.2016.90.
45. Loshchilov I. and Hutter F. (2019). Decoupled weight decay regularization. *ICLR*. DOI:10.48550/arXiv.1711.05101.
46. Radford A., Kim J.W., Hallacy C., et al. (2021). Learning transferable visual models from natural language supervision. *Proc. Mach. Learn. Res.* **139**:8748–8763. https://proceedings.mlr.press/v139/radford21a/radford21a.pdf.


## FUNDING AND ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (Grant No. 52578029).

## AUTHOR CONTRIBUTIONS

P.Z. conceived the study, developed the overall methodology and GreenPlanner framework, performed the main experiments and analyses, prepared the visualizations, and drafted the manuscript. T.Z. contributed to data processing, model implementation, experimental evaluation, and validation. J.Y. contributed to the methodology, floor-plan evaluation, data analysis, and interpretation of the results. J.C. contributed to the building-energy simulation methodology, performance evaluation, validation, and interpretation of the energy-related results. S.L. supervised the research, provided resources and project administration, acquired funding, and contributed to manuscript review and revision. All authors discussed the results, reviewed the manuscript, and approved the final version.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

## DATA AND CODE AVAILABILITY

The RPLAN dataset used in this study is publicly available as described in Ref. 20. The derived datasets and code supporting the findings of this study are available from the corresponding author upon reasonable request.

## SUPPLEMENTAL INFORMATION

It can be found online at https://doi.org/10.59717/ipj.energy-use.2026.100068