S^{3}THERM: A SOURCE-SHEET NEURAL OPERATOR FOR 3D STEADY HEAT CONDUCTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Repeated thermal analysis requires predicting three-dimensional steady temperature fields for changing volumetric heat sources. A neural predictor must capture how a source's shape and position affect the full field and how the responses to multiple sources combine. For fixed geometry, material properties, and Robin boundary conditions, temperature rise is linear in the heat source. This makes a thin source layer a useful response unit: it retains an in-plane heat pattern and a representative vertical position, while responses to different layers can be combined. We introduce S^{3}Therm, which uses one shared neural operator to predict the three-dimensional response to each layer and sums these responses to predict the field of the full source. The operator is trained from the governing equation and boundary conditions without paired temperature fields. Energy-balance rescaling and symmetry averaging refine the combined prediction. On ten held-out source configurations evaluated at matched COMSOL coordinates across seven planes, S^{3}Therm achieves a three-seed mean per-layout RMSE of 0.397 K, compared with 1.529 K for supervised FactFormer.

1 INTRODUCTION

Thermal analysis informs component placement, power planning, and reliability decisions in chip design. This paper studies repeated prediction of the three-dimensional steady temperature field produced by a changing volumetric heat source. We consider a fixed geometry, material, and cooling condition, while the shape, position, and strength of the heat source vary.

Numerical solvers can compute the temperature field for each new source, but repeated high-resolution solves are costly. Neural operators learn to predict fields for new inputs and have been applied to thermal analysis Liu et al. (2023); Guan et al. (2025). Physics-informed training can also learn a thermal operator without paired temperature fields Liu et al. (2023). A common approach supplies the complete heat source to the model. This gives the model the source geometry, but leaves it to learn from training how the effects of separate heating regions combine. The technical challenge is to learn a response that retains the shape and position of a heating region and can be reused when several regions form a new three-dimensional source.

Steady heat conduction provides the principle for this reuse. Under fixed geometry, material, and linear boundary conditions, the temperature rise caused by several sources is the sum of their individual responses Stakgold & Holst (2011). We use a thin, finite-thickness part of the source as the unit whose response is learned. We call it a source sheet: it has an in-plane heat pattern, a representative height, and a strength. S^{3}Therm predicts the three-dimensional response to each sheet with one shared neural operator and adds these responses to obtain a field for the complete source. In this way, the operator learns how heat spreads from an extended source region, while the governing equation determines how responses to different regions combine.

The complete prediction also uses a second property of steady heat conduction: the heat generated inside the domain must be balanced by heat leaving its boundary. We train the shared operator from the heat equation and boundary conditions, without paired temperature fields. After combining the sheet responses, we use the global heat balance and the spatial symmetries of the problem to refine the predicted field. Figure 1 shows the full procedure.

Our contributions are as follows:

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.

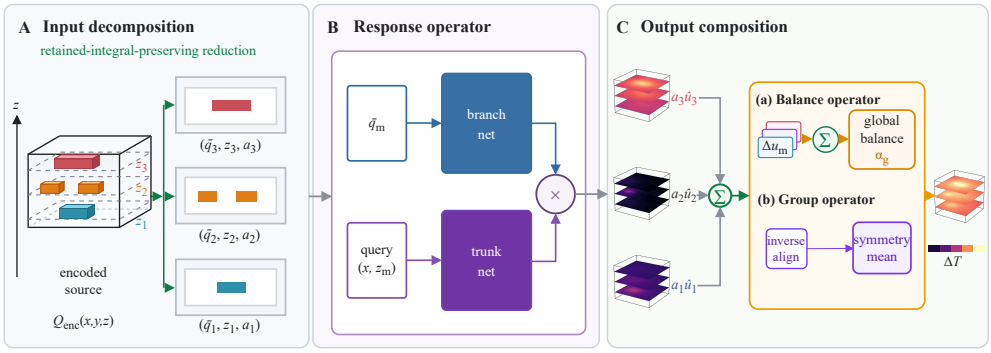


Figure 1: Overview of S^{3}Therm. (A) A three-dimensional heat source is divided into thin source sheets, each described by an in-plane heat pattern, a representative height, and a strength. (B) One shared operator predicts the three-dimensional temperature rise caused by each sheet. (C) The sheet responses are added, adjusted using the overall heat balance, and averaged under four in-plane symmetries.

- We introduce, to our knowledge, the first neural operator that uses finite-thickness source sheets as reusable response units for predicting three-dimensional steady temperature fields from volumetric heat sources. We make each sheet's in-plane heat pattern and representative height explicit while combining sheet responses according to source superposition.
- We develop a procedure that reduces a three-dimensional source to sheets and train one shared operator for all sheets using the heat equation and boundary conditions. We combine its predictions and use global heat balance and spatial symmetry to refine the full temperature field without paired temperature-field supervision.
- We evaluate S^{3}Therm on ten held-out heat-source configurations at matched COMSOL coordinates across seven planes, using three independent seeds per method. We obtain a mean per-layout RMSE of 0.397 K, compared with 1.529 K for FactFormer, the strongest evaluated baseline.

2 Related Work

Thermal prediction. Compact simulators such as HotSpot and 3D-ICE support repeated thermal analysis Skadron et al. (2003); Sridhar et al. (2010). Learned methods predict temperature fields from design inputs in integrated circuits and photonic systems Liu et al. (2023); Guan et al. (2025); Zhu et al. (2024). DeepOHeat is particularly relevant because it uses the governing equations to train a thermal operator without paired temperature fields Liu et al. (2023). Our work studies changing three-dimensional heat sources under fixed physical conditions and learns responses to parts of a source that can be combined for a new configuration.

Neural operators and training from physical equations. DeepONet and the Fourier neural operator provide general ways to learn maps between input functions and solution fields Lu et al. (2021); Li et al. (2021); Kovachki et al. (2023). FactFormer and LANO offer different architectures for predicting PDE solutions and serve as baselines in our evaluation Li et al. (2023); Hou et al. (2026). Training an operator from equation and boundary residuals is established in physics-informed DeepONets and PINO Wang et al. (2021); Li et al. (2024). We apply this training approach to the response caused by one source sheet and use the heat equation to combine responses across sheets.

Learning source responses. For a linear heat equation, a Green function expresses the temperature field as an integral of point responses over the heat source Stakgold & Holst (2011). Learned Green-function methods develop this idea further Gu et al. (2025); Melchers et al. (2026); Yoo et al. (2025); the integral accounts for the full spatial shape of the source. S^{3}Therm learns the response to

a finite-thickness source sheet whose in-plane heat pattern and height are inputs to the operator. It then combines the predicted sheet responses to obtain the field for a three-dimensional source.

3 GOVERNING PROBLEM

Let $\Omega = [0, L]^3$ and $u = T - T_{\mathrm{amb}}$ denote temperature rise. We consider steady conduction with uniform scalar conductivity k, volumetric source Q_{phys}, and a linear Robin boundary condition with effective Robin conductance h_{phys}. Setting $\xi = x/L \in \widehat{\Omega} = [0, 1]^3$, $Q_*(\xi) = L^2 Q_{\mathrm{phys}}(L\xi)$, and $h_* = h_{\mathrm{phys}} L$ gives

$$\begin{aligned} -k\Delta_\xi u &= Q_* \quad \text{in } \widehat{\Omega}, \\ -k\nabla_\xi u \cdot n &= h_* u \quad \text{on } \partial\widehat{\Omega}. \end{aligned} \tag{1}$$

Here n is the outward unit normal, and Q_* and h_* have units $\mathrm{W\,m^{-1}}$ and $\mathrm{W\,m^{-1}\,K^{-1}}$, respectively. The benchmark fixes $L = 1$ mm, $k = 1.4\,\mathrm{W\,m^{-1}\,K^{-1}}$, $h_{\mathrm{phys}} = 500\,\mathrm{W\,m^{-2}\,K^{-1}}$, and $T_{\mathrm{amb}} = 273.15$ K, giving $h_* = h_{\mathrm{phys}} L = 0.5\,\mathrm{W\,m^{-1}\,K^{-1}}$. Both the training residuals and the energy-balance correction use Eq. (1). We henceforth use (x, y, z) for the components of the normalized coordinate ξ and $z_{\mathrm{phys}} = Lz$ for reported physical depths.

The source-sheet representation below maps the dimensionless encoded input Q_{enc} to the PDE source Q_*. It permits repeated heights and overlapping footprints, while the learned parameter count remains independent of the number M of retained height intervals. Absolute temperature is recovered as $\widehat{T} = T_{\mathrm{amb}} + \widehat{u}$.

3.1 LINEARITY AND SOURCE SUPERPOSITION

Positive k and h_* make the Robin problem coercive, so it has a unique weak solution and defines a bounded linear operator $\mathcal{S}_* : H^1(\widehat{\Omega})' \to H^1(\widehat{\Omega})$ Evans (2010). Hence, for any finite sum $\widetilde{Q} = \sum_{m=1}^{M} Q_m$,

$$\mathcal{S}_*(\widetilde{Q}) = \sum_{m=1}^{M} \mathcal{S}_*(Q_m). \tag{2}$$

Equation (2) is the additivity the design exploits: the predictor sums its interval responses explicitly, and the final map then applies the layout-dependent balance and symmetry operations. COMSOL samples from layouts not used for training evaluate this complete prediction Mishra & Molinaro (2023).

4 SOURCE-SHEET RESPONSE OPERATOR

Equation (2) expresses a retained multi-sheet source as a weighted sum of reusable responses. Figure 1 shows how S^{3}Therm realizes this structure through source-sheet reduction, one shared response operator conditioned on representative source height, and pre-correction composition across height intervals, followed by balance and symmetry operations.

The decomposition assigns distinct roles: the normalized footprint is the functional input, centroid height conditions propagation, and amplitude remains an outer scalar. The retained-interval count changes response terms, not parameters. The network learns the response to a normalized footprint lifted to finite thickness, while positive homogeneity with respect to sheet amplitude and cross-interval addition are explicit; interactions within a merged footprint are learned jointly.

4.1 SOURCE-SHEET REPRESENTATION

Each component carries a source-strength label P_{label} in mW that indexes strength rather than integrated power. With footprint area A in $\mathrm{mm^2}$, normalized thickness $t = t_{\mathrm{phys}}/L$, and the encoder constant $\phi_{\mathrm{char}} = 4000\,\mathrm{mW\,mm^{-2}}$, the encoder is $\beta = (P_{\mathrm{label}}/A)/\phi_{\mathrm{char}}$ and $Q_{\mathrm{enc}} = \beta/t$ on each component block and zero elsewhere, with $Q_0 = 4 \times 10^9 \beta\,\mathrm{W\,m^{-3}}$ the corresponding volumetric source density.

The encoded heat sources are nonnegative. On an input tensor with z-cell centers z_j and $\Delta z = 1/N_z$, we retain z-bins whose in-plane sums exceed 1% of the peak and split each contiguous run into intervals B_m of width at most 0.05. Let $p_j = \sum_{i,k} Q_{\mathrm{enc}}(x_i, y_k, z_j)$ and let $\mathcal{J}_{\mathrm{ret}}$ denote the retained bins. Each interval gives

$$\begin{aligned} q_m(x,y) &= \sum_{j \in B_m} Q_{\mathrm{enc}}(x,y,z_j)\,\Delta z, \\ z_m &= \frac{\sum_{j \in B_m} z_j p_j}{\sum_{j \in B_m} p_j}, \\ a_m &= \max_{x,y} q_m(x,y), \qquad \bar{q}_m = q_m / a_m, \\ \sum_{m=1}^{M} q_m(x,y) &= \sum_{j \in \mathcal{J}_{\mathrm{ret}}} Q_{\mathrm{enc}}(x,y,z_j)\Delta z. \end{aligned} \tag{3}$$

The last equality preserves the retained Riemann sum and, under piecewise-constant interpolation, equals $\int_0^1 Q_{\mathrm{enc}}^{\mathrm{ret}}(x,y,z)\,dz$. The 1% rule excludes subthreshold tails. For one retained component, $q_m(x,y) \approx \int_{B_m} Q_{\mathrm{enc}}(x,y,z)\,dz = \beta$ on its footprint. Here, β is a normalized encoding amplitude, rather than a physical integrated power.

The factorization $q_m = a_m \bar{q}_m$ places the global amplitude of each extracted sheet outside the response operator. Every retained interval has $a_m > 0$, so the branch receives a bounded, unit-peak pattern. Relative spatial variation within a merged interval footprint remains encoded in $\bar{q}_m$ and is modeled jointly by the shared response operator.

Thresholding and interval splitting are deterministic. Footprints within one interval sum into q_m; distinct intervals remain separate calls even when their projections overlap. Centroid assignment collapses vertical variation within each interval, defining the source-representation approximation before PDE superposition.

Residual evaluation requires a differentiable volumetric source. For a nominal width W, let $W_{\mathrm{eff}}(z_m) = |[z_m - W/2, z_m + W/2] \cap [0,1]|$. We define

$$\begin{aligned} w_0(z; z_m) &= \tfrac{1}{2}\left[\tanh \frac{z - z_m + W/2}{\sigma} \right. \\ &\qquad \left. - \tanh \frac{z - z_m - W/2}{\sigma}\right], \\ w(z; z_m) &= \frac{W_{\mathrm{eff}}(z_m) w_0(z; z_m)}{\int_0^1 w_0(\zeta; z_m)\,d\zeta}, \\ Q_{*,m}(x,y,z) &= s_{\mathrm{src}} a_m \bar{q}_m(x,y) w(z; z_m). \end{aligned} \tag{4}$$

Thus $\int_0^1 w(z; z_m)\,dz = W_{\mathrm{eff}}(z_m)$ when a source intersects the domain boundary. For $L = 1$ mm, the density convention gives $s_{\mathrm{src}} = 4000\,\mathrm{W\,m^{-1}}$. The factor $a_m \bar{q}_m$ supplies the encoded amplitude, and w supplies the nominal lifted thickness. We use $W = 0.05$ (50 μm) and $\sigma = 0.1W$.

Figure 2 expands the height grouping, interval integration, and finite-thickness lifting steps.

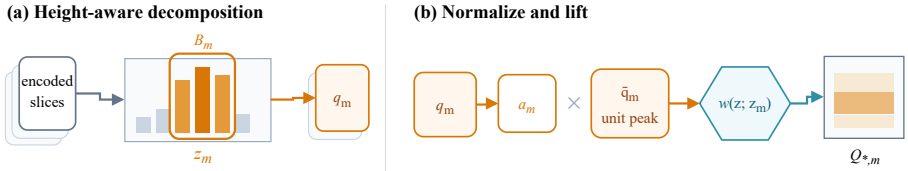


Figure 2: Source-sheet construction. (A) Plane-wise encoded source strengths p_j identify the retained height interval B_m and its centroid z_m; vertical integration across the retained slices gives the footprint $q_m(x, y)$. (B) The footprint is factorized as $q_m = a_m \bar{q}_m$, where a_m is the peak encoded amplitude and $\bar{q}_m$ is a unit-peak footprint. Applying the finite-width profile $w(z; z_m)$ produces $Q^*_{,m}$, with boundary truncation when necessary.

4.2 Shared Response Operator Conditioned on Source Height and Pre-Correction Composition

Let $Q_{*,\mathrm{lift}} = \sum_m Q_{*,m}$. The exact lifted-source response, its linear interval assembly, the learned response, and the pre-correction prediction are

$$\begin{aligned} \mathcal{H}(q, z_0) &= \mathcal{S}_*[s_{\mathrm{src}} q(x, y) w(z; z_0)], \\ \mathcal{S}_*(Q_{*,\mathrm{lift}}) &= \sum_{m=1}^{M} a_m \mathcal{H}(\bar{q}_m, z_m), \\ \mathcal{H}_\theta(q, x, y, z, z_0) &= \sum_i b_i(q) t_i(x, y, z, z_0) + c, \\ \widetilde{\mathcal{S}}_\theta(Q_{\mathrm{enc}}) &= \sum_{m=1}^{M} a_m \mathcal{H}_\theta(\bar{q}_m, \cdot, z_m). \end{aligned} \tag{5}$$

The outer assembly is a weighted sum over retained height intervals. For each interval, a_m supplies a global amplitude outside the response operator, while $\bar{q}_m$ retains relative in-plane variation, which $\mathcal{H}_\theta$ models jointly across the merged footprint. A transformer branch maps the footprint on a fixed 500×500 sensor grid to $b(q)$, while Fourier features feed a residual fully connected trunk that maps (x, y, z, z_0) to t. The resulting 256-dimensional feature vectors enter an inner product, and a scalar bias completes the temperature-rise prediction at a query point. Supplying z_0 to the trunk allows a single parameter set to serve all source heights instead of using layer-specific networks or a composite-volume branch input.

The parameter count is independent of M; adding intervals only adds terms to the pre-correction sum. For N queries and four inference-time transforms, M intervals require $4M$ shared-response evaluations and $O(4MN)$ assembly. Because the field can be accumulated one sheet at a time, increasing M increases evaluation work but not network width or parameter count.

4.3 Physics-Informed Training

For a normalized sheet q, height z_0, interior point $X = (x, y, z)$, and boundary point X_b, write $u_\theta(q, z_0, X) = \mathcal{H}_\theta(q, x, y, z, z_0)$. With the adopted sign convention, the residuals and training objective are

$$\begin{aligned} r_\theta(q, z_0, X) &= \gamma \left[k \Delta_\xi u_\theta(q, z_0, X) \right. \\ &\qquad \left. + s_{\mathrm{src}} q(x, y) w(z; z_0) \right], \\ R_{\mathrm{bc}}(q, z_0, X_b) &= \nabla_\xi u_\theta(q, z_0, X_b) \cdot n + \frac{h_*}{k} u_\theta(q, z_0, X_b), \\ \mathcal{L}(\theta) &= \mathbb{E}_{q, z_0, X} \left[r_\theta(q, z_0, X)^2 \right] \\ &\qquad + \lambda_{\mathrm{bc}} \mathbb{E}_{q, z_0, X_b} \left[R_{\mathrm{bc}}(q, z_0, X_b)^2 \right]. \end{aligned} \tag{6}$$

The residual symbols in Eq. (6) denote numerical values after the interior and boundary expressions are evaluated in $\mathrm{W\,m^{-1}}$ and K, respectively. Here, $\gamma = 0.1$ scales the interior value, and λ_{bc} is 100 for the first 30,000 steps and 200 for the remaining 70,000 steps. The loss is evaluated on 2,000 synthetic sheets, drawn equally from Gaussian random fields and nonoverlapping rectangles with continuous height sampling, and on collocation points using the PDE and Robin boundary condition. It is a fixed-unit numerical objective, not a unit-invariant physical norm; γ and λ_{bc} are therefore numerical weights rather than physical parameters.

Training uses individual sheets, including synthetic rectangles. Equation (5) then provides explicit composition across height intervals, while the spatial response within each merged interval footprint is learned jointly.

4.4 Integral-Balance Rescaling and Symmetry Averaging

For the raw prediction $\widetilde{u}_g$ associated with transformed source g, the divergence theorem applied to Eq. (1) gives global balance for $Q_{*,\mathrm{lift}}$. We approximate this balance on the unit-cube grid as follows. Let $\mathcal{F}$ denote the six boundary faces.

$$\begin{aligned} \widehat{S}_{\mathrm{src}} &= s_{\mathrm{src}} \sum_m W_{\mathrm{eff}}(z_m)\ \mathrm{mean}_{x,y}\, q_m, \\ \widehat{S}_{\mathrm{bdry},g} &= h_* \sum_{F\in\mathcal{F}} \mathrm{mean}_{p\in F}\, \widetilde{u}_g(p), \\ \alpha_g &= \frac{\widehat{S}_{\mathrm{src}}}{\widehat{S}_{\mathrm{bdry},g}}, \\ \mathcal{C}[\widetilde{u}_g] &= \alpha_g \widetilde{u}_g. \end{aligned} \tag{7}$$

The boundary sum weights the six unit-area faces equally. The correction $\mathcal{C}$ applies Eq. (7) for finite positive sums and uses $\alpha_g = 1$ otherwise. The scalar factor is an integral-balance correction: it closes the global source–boundary budget while preserving the spatial structure of the composed response.

The cube, scalar conductivity, and Robin boundary condition are also invariant under the Klein four-group $\mathcal{G} = \{e, \sigma_x, \sigma_y, \rho_\pi\}$ generated by axial reflections, with $\rho_\pi = \sigma_x\sigma_y$. Let $\rho_Q(g)$ and $\rho_T(g)$ apply the corresponding exact grid permutations to the source and temperature field while leaving z unchanged. Applying the Reynolds projection to the corrected predictions gives the final predictor

$$\widehat{\mathcal{S}}_\theta(Q_{\mathrm{enc}}) = \frac{1}{|\mathcal{G}|}\sum_{g\in\mathcal{G}} \rho_T(g)^{-1}\mathcal{C}\Big[\widetilde{\mathcal{S}}_\theta(\rho_Q(g)Q_{\mathrm{enc}})\Big]. \tag{8}$$

The Reynolds projection is idempotent and makes the final predictor equivariant to $\mathcal{G}$ Cohen & Welling (2016); Kondor & Trivedi (2018). For finite positive boundary sums, rescaling enforces discrete integral balance, while group averaging enforces in-plane equivariance.

Figure 2 details the height-associated representation. Figure 3 shows the raw–balance–symmetry sequence used at inference; Table 2 reports the correction ablation and its cost.

Exact source superposition applies only to the pre-correction interval sum. The subsequent balance rescaling and symmetry averaging make the complete predictor non-additive.

4.5 Approximation and Stability Considerations

The source-sheet construction introduces a controlled approximation before the neural operator is evaluated. Retained height bins preserve the in-plane Riemann sum, while centroiding and finite-width lifting replace variation within one interval by a representative vertical profile. Thresholding changes the retained source by removing low-amplitude tails; centroiding changes how the retained power is distributed vertically. The same representation is used for training and testing, so the reported comparison measures the complete representation–operator pipeline.

The interval width trades representation accuracy against inference cost. Narrower intervals preserve more vertical structure and add terms to Eq. (5); wider intervals do the reverse. Conditioning the

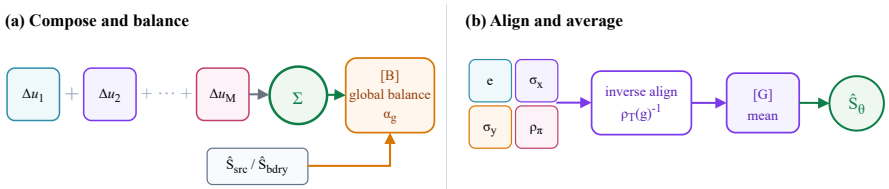


Figure 3: Response composition and inference-time corrections. (A) For a transformed source g, the interval responses are summed to form the raw field $\widetilde{u}_g$. The balance factor $\alpha_g = \widehat{S}_{\mathrm{src},g}/\widehat{S}_{\mathrm{bdry},g}$ applies the integral-balance correction $C[\widetilde{u}_g]$. (B) The four predictions generated by $G = \{e, \sigma_x, \sigma_y, \rho_\pi\}$ are inverse-aligned and averaged to produce the final predictor $\widehat{S}_\theta(Q_{\mathrm{enc}})$.

Table 1: Layout RMSE (K) ↓ of the temperature rise $u = T - T_{\mathrm{amb}}$ over all valid COMSOL samples of each held-out layout, pooled across the seven planes so that a plane carries its COMSOL sample count; the mean weights layouts equally. All entries are arithmetic means over three independent seeds. Bold and underlining mark the best and second-best mean.

Held-out layout	S^{3}Therm	LANO	FactFormer
Edge-large	0.151	0.627	1.539
Sparse hotspot	0.659	1.496	1.107
Dense hotspots	0.317	3.465	1.327
Edge high-power	0.958	2.545	3.782
Elongated	0.141	0.590	0.497
Shifted heights	0.170	0.529	0.538
Corner pair	0.744	2.441	4.127
Vertical offset	0.168	0.482	0.396
Dense small blocks	0.446	2.995	1.202
Aspect extremes	0.215	1.429	0.773
Mean (10)	**0.397**	1.660	<u>1.529</u>

shared operator on z_0 refines vertical placement without introducing height-specific networks. Thus separated source heights produce separate conditioned responses, while a broad contiguous source is represented by nearby centroids. The parameter count is fixed, although the number of response evaluations follows the retained vertical complexity.

The two inference corrections target different errors. Local residual training can produce plausible curvature while leaving a global source–boundary mismatch; Eq. (7) removes this mismatch without changing the raw field's relative spatial pattern. Symmetry averaging suppresses components outside the invariant solution space. A low layout RMSE therefore measures agreement with matched COMSOL samples after source reduction, composition, balance correction, and symmetry projection, but does not isolate the error of any one stage. Further separation would require volumetric references and controlled sweeps over the retention threshold, interval width, and lifted thickness.

5 Experiments

5.1 Experimental Setup

The benchmark contains ten held-out layouts of fixed-thickness components, up to 36 per layout, and COMSOL Multiphysics 6.4 provides stationary reference samples at seven planes from $z_{\mathrm{phys}} = 100$ to 900 μm. We score each layout by the root-mean-square error (RMSE) between the predicted and the reference temperature over the valid COMSOL samples of that layout and compare methods by the equal-weight mean over the ten layouts; all methods predict at the same coordinates. Each method is scored from its native output grid through the same trilinear sampler. S^{3}Therm is trained from PDE and Robin residuals on synthetic sources and uses no paired temperature fields, whereas

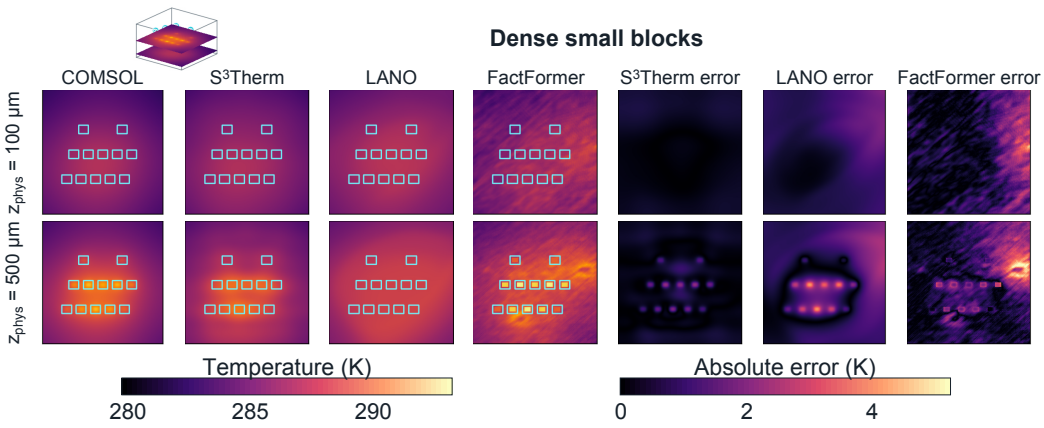


Figure 4: Held-out *Dense small blocks* layout at $z_{\mathrm{phys}} = 100$ and 500 μm: the COMSOL reference, the S^{3}Therm, LANO, and FactFormer predictions, and their absolute errors; cyan contours are projected source footprints. Temperature panels share one scale and absolute-error panels one nonnegative scale capped at 5.2 K, the joint 99th percentile of the two slices. COMSOL samples cover $52.7/96.3\%$ of the display grid at $100/500$ μm, and interpolation is used only for display.

LANO and FactFormer are supervised on twelve paired COMSOL layouts under a nominal wall-clock budget of roughly ten hours per run. This is a budgeted native-resolution comparison, not a resolution-matched convergence study. All method values in Table 1 are arithmetic means over three independent seeds; the qualitative panels show one representative run.

Ten held-out stress layouts contain 1–36 blocks at one or three source heights (309 fixed-thickness components in total). COMSOL Multiphysics 6.4 provides stationary reference samples at $z_{\mathrm{phys}} = 100$, 250, 400, 500, 600, 750, and 900 μm. Each method's native output grid is sampled at these coordinates using trilinear interpolation. For the valid samples $\mathcal{I}_\ell$ of layout ℓ, we compute

$$E_\ell = \left[\frac{1}{|\mathcal{I}_\ell|} \sum_{i \in \mathcal{I}_\ell} \left(\widehat{T}_\ell(x_i) - T_\ell^{\mathrm{COMSOL}}(x_i) \right)^2 \right]^{1/2}. \tag{9}$$

Table 1 reports equal-weight summaries of the layout-level RMSE. Every method is scored at the same frozen COMSOL coordinates. The metrics therefore measure agreement with these plane samples, not reference-solver convergence throughout the volume. Coordinate matching and sample pooling are detailed in the supplement.

5.2 Main Results

On the ten held-out layouts, S^{3}Therm attains the lowest RMSE in every layout, with an equal-weight three-seed mean of 0.397 K against 1.529 K for FactFormer, the strongest baseline (Table 1).

Spatial error. We place one representative qualitative comparison in the main text. Figure 4 shows S^{3}Therm on the dense 36-block layout at $z_{\mathrm{phys}} = 100$ and 500 μm, where the error concentrates near steep source neighborhoods. The other nine held-out layouts are shown together in the appendix gallery. The panels are a representative run; the main-comparison numbers use the three-seed means in Table 1.

5.3 Ablations

In a separate four-layout representation control, collapsing a layout's height content into a single power-weighted centroid that preserves the integrated in-plane footprint and total source power, replacing three retained sheet responses by one, raises the mean layout RMSE from 0.296 to 1.020 K on Corner pair, Vertical offset, Dense small blocks, and Aspect extremes.

Table 2: Inference-time correction ablation, toggling global balance and in-plane symmetry averaging from the raw composition. RMSE (K) ↓ is over the ten held-out layouts at the shared COMSOL coordinates; the latency and memory columns are per-layout inference on four of them, over ten repeats after three warmups on one GPU. All entries are arithmetic means over the same three independent S3Therm checkpoints.

			Latency (s)		Peak memory (MiB)	
Variant	Mean RMSE ↓	Worst RMSE ↓	median	p95	allocated	reserved
Raw composition	2.030	6.643	3.228	4.251	5101.9	5848
Balance only	0.678	2.273	3.128	3.896	5101.9	5848
Symmetry only	1.878	7.040	3.427	4.958	6186.2	7110
Full predictor	**0.397**	**0.955**	3.626	5.054	6186.2	7110

On the ten held-out layouts, balance, an integral rescaling, provides the larger part of the gain and the symmetry projection adds the rest (Table 2); the full predictor is about $1.1\times$ the raw composition latency under the same warm-started protocol. Applying both corrections to LANO raises its mean and worst layout RMSE by 0.04 and 0.13 K, so the gain is not reproduced as generic post-processing. All rows in Table 2 are inference-only evaluations of the same three independent S3Therm checkpoints used for the main comparison, with arithmetic means taken across seeds.

Limitations. All training and evaluation layouts share one geometry, conductivity, ambient temperature, and Robin boundary condition. Consequently, the experiments test layout generalization within a fixed physical problem, not transfer to new materials, boundary coefficients, domain shapes, or source strength regimes.

Evaluation uses the original COMSOL coordinates on seven horizontal planes. These samples support paired comparison between methods, but they do not bound the error throughout the three-dimensional volume or establish convergence with respect to COMSOL mesh density and solver tolerance. The current predictor should therefore not be treated as a certified estimator of the continuous-domain maximum temperature.

Finally, exact superposition belongs to the underlying linear PDE and to the explicit pre-correction interval sum. Sources assigned to the same retained interval are merged and modeled jointly, while the layout-dependent balance factor and symmetry averaging act on the assembled field. The complete predictor is therefore not strictly additive. Future work should evaluate denser volumetric references, mesh and tolerance convergence, and conditioning on material, geometry, and boundary parameters.

6 Conclusion

We introduced S^{3}Therm, a source-sheet neural operator that exposes composition across retained source intervals while learning the three-dimensional response of each finite-thickness sheet. On ten held-out fixed-physics layouts, the complete predictor achieves a three-seed mean RMSE of 0.397 K, compared with 1.529 K for the strongest evaluated supervised baseline, while using no paired temperature fields during training. The results support source-sheet decomposition as a useful inductive bias for steady thermal prediction, but they should be interpreted within the fixed physical configuration and plane-sampled evaluation considered here.

7 AI Assistance Disclosure

The authors did not use large language models (LLMs) or any other AI tools at any stage of this work, including research ideation or execution, literature retrieval or discovery, data generation, mathematical proof development, or the drafting, editing, and polishing of the manuscript. All research, analysis, and writing were conducted by the authors.

REFERENCES

Taco S. Cohen and Max Welling. Group equivariant convolutional networks. In *Proceedings of the 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pp. 2990–2999, 2016. URL https://proceedings.mlr.press/v48/cohenc16.html.

Lawrence C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, second edition, 2010. doi: 10.1090/gsm/019.

Jianghang Gu, Ling Wen, Yuntian Chen, and Shiyi Chen. An explainable operator approximation framework under the guideline of Green's function. *Journal of Computational Physics*, 539: 114244, 2025. doi: 10.1016/j.jcp.2025.114244.

Weilong Guan, Li Huang, Yuxuan Lin, Yuchao Wu, Yeyu Tong, and Yuzhe Ma. ThermoPhoton: Fast 3D thermal simulation of photonic integrated circuits via operator learning. In *Proceedings of the IEEE/ACM International Conference on Computer-Aided Design*, pp. 1–8, 2025. doi: 10.1109/ICCAD66269.2025.11240705.

Jingren Hou, Hong Wang, Pengyu Xu, Chang Gao, Huafeng Liu, and Liping Jing. Learning neural operators from partial observations via latent autoregressive modeling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 390–398, 2026. doi: 10.1609/aaai.v40i1.37001.

Risi Kondor and Shubhendu Trivedi. On the generalization of equivariance and convolution in neural networks to the action of compact groups. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pp. 2747–2755, 2018. URL https://proceedings.mlr.press/v80/kondor18a.html.

Nikola Kovachki, Zongyi Li, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Neural operator: Learning maps between function spaces. *Journal of Machine Learning Research*, 24(89):1–97, 2023. URL https://jmlr.org/papers/v24/21-1524.html.

Zijie Li, Dule Shu, and Amir Barati Farimani. Scalable transformer for PDE surrogate modeling. In *Advances in Neural Information Processing Systems*, volume 36, pp. 28010–28039, 2023. doi: 10.52202/075280-1216.

Zongyi Li, Nikola Kovachki, Kamyar Azizzadenesheli, Burigede Liu, Kaushik Bhattacharya, Andrew Stuart, and Anima Anandkumar. Fourier neural operator for parametric partial differential equations. In *International Conference on Learning Representations*, 2021. URL https://openreview.net/forum?id=c8P9NQVtmnO.

Zongyi Li, Hongkai Zheng, Nikola Kovachki, David Jin, Haoxuan Chen, Burigede Liu, Kamyar Azizzadenesheli, and Anima Anandkumar. Physics-informed neural operator for learning partial differential equations. *ACM/IMS Journal of Data Science*, 1(3):1–27, 2024. doi: 10.1145/3648506.

Ziyue Liu, Yixing Li, Jing Hu, Xinling Yu, Shinyu Shiau, Xin Ai, Zhiyu Zeng, and Zheng Zhang. DeepOHeat: Operator learning-based ultra-fast thermal simulation in 3D-IC design. In *Proceedings of the 60th ACM/IEEE Design Automation Conference*, pp. 1–6, 2023. doi: 10.1109/DAC56929.2023.10247998.

Lu Lu, Pengzhan Jin, Guofei Pang, Zhongqiang Zhang, and George Em Karniadakis. Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021. doi: 10.1038/s42256-021-00302-5.

Hugo Melchers, Joost Prins, and Michael Abdelmalik. Neural Green's operators for parametric partial differential equations. *Computer Methods in Applied Mechanics and Engineering*, 455: 118893, 2026. doi: 10.1016/j.cma.2026.118893.

Siddhartha Mishra and Roberto Molinaro. Estimates on the generalization error of physics-informed neural networks for approximating PDEs. *IMA Journal of Numerical Analysis*, 43(1):1–43, 2023. doi: 10.1093/imanum/drab093.

Kevin Skadron, Mircea R. Stan, Wei Huang, Sivakumar Velusamy, Karthik Sankaranarayanan, and David Tarjan. Temperature-aware microarchitecture. In *Proceedings of the 30th Annual International Symposium on Computer Architecture*, pp. 2–13, 2003. doi: 10.1145/859618.859620.

Arvind Sridhar, Alessandro Vincenzi, Martino Ruggiero, Thomas Brunschwiler, and David Atienza. 3D-ICE: Fast compact transient thermal modeling for 3D ICs with inter-tier liquid cooling. In *Proceedings of the 2010 IEEE/ACM International Conference on Computer-Aided Design*, pp. 463–470, 2010. doi: 10.1109/ICCAD.2010.5653749.

Ivar Stakgold and Michael J. Holst. *Green's Functions and Boundary Value Problems*. Wiley, third edition, 2011. doi: 10.1002/9780470906538.

Sifan Wang, Hanwen Wang, and Paris Perdikaris. Learning the solution operator of parametric partial differential equations with physics-informed DeepONets. *Science Advances*, 7(40):eabi8605, 2021. doi: 10.1126/sciadv.abi8605.

Seungwoo Yoo, Kyeongmin Yeo, Jisung Hwang, and Minhyuk Sung. Neural green's functions. In *Advances in Neural Information Processing Systems*, volume 38, 2025. URL https://papers.nips.cc/paper_files/paper/2025/hash/448abd486677165ceedfa790e9a61802-Abstract-Conference.html.

Tianxiang Zhu, Qipan Wang, Yibo Lin, Runsheng Wang, and Ru Huang. FaStTherm: Fast and stable full-chip transient thermal predictor considering nonlinear effects. In *Proceedings of the 61st ACM/IEEE Design Automation Conference*, pp. 1–9, 2024. doi: 10.1145/3676536.3676738.

A APPENDIX

A.1 EVALUATION PROTOCOL

The ten held-out layouts contain fixed-thickness components, up to 36 per layout, and are disjoint from the twelve layouts that supervise the baselines. Reference temperatures are stationary COMSOL Multiphysics 6.4.0.293 solutions of Eq. (1) under the material and boundary parameters of Section 3, sampled on seven planes at $z_{\mathrm{phys}} = 100$, 250, 400, 500, 600, 750, and 900 μm.

Reference samples within 0.5 μm of a plane are assigned to that plane. Each method predicts on its native output grid and is evaluated at the reference coordinates by trilinear interpolation. S^{3}Therm predicts on the $500 \times 500 \times 120$ grid.

Samples are pooled within each layout before scoring, and the ten layouts are weighted equally. The layout error is the root-mean-square error over its valid COMSOL samples.

A.2 MODEL CONFIGURATIONS

The ten held-out COMSOL layouts are not used for checkpoint selection. For the supervised baselines, the selected checkpoint is the minimum logged training loss on the twelve labeled training layouts; no held-out evaluation is used for selection. The baselines use the official LANO Hou et al. (2026) and FactFormer Li et al. (2023) implementations under MIT licences. LANO uses two blocks of width 128 with a latent width of 96, four attention heads, and one latent layer; FactFormer uses feature dimension 64, depth 2, and four heads of width 16. Their peak learning rates are 8.0×10^{-4}; LANO uses FP32 and FactFormer BF16. FactFormer normalises over the spatial domain in a way that does not preserve the tensor shape; we restore the shape of its output, leaving the normalisation itself unchanged.

The per-seed overall layout-RMSE standard deviations are 0.099 K for S^{3}Therm, 1.557 K for LANO, and 0.087 K for FactFormer (sample standard deviation across three seeds); the main table reports the corresponding arithmetic means.

S^{3}Therm is trained from PDE and Robin residuals alone, with interior residual weight $\gamma = 0.1$ and boundary weight 100 for the first $30{,}000$ steps and 200 thereafter. Its source vocabulary is $2{,}000$ synthetic sheets, and each step samples five of them together with $80{,}000$ interior and $12{,}500$ boundary points, and the points are resampled every $10{,}000$ steps. Checkpoint selection uses residual-validation loss on a fixed three-sheet source. Training has two stages. The learning rate warms up linearly to 10^{-3} over the first $1{,}000$ steps and follows a cosine schedule to step $30{,}000$, where the optimiser changes from Adam to AdamW and the rate warms up over $2{,}000$ steps to 5×10^{-5} before a final cosine decay to 5×10^{-7}.

A.3 QUALITATIVE COMPARISON ON THE HELD-OUT LAYOUTS

Figures 5–6 show the nine held-out layouts not shown in Figure 4 at $z_{\mathrm{phys}} = 100$ and 500 μm, in the order of Table 1. The audited temperature and error scales are shown for each source-layout group and are not shared across groups.

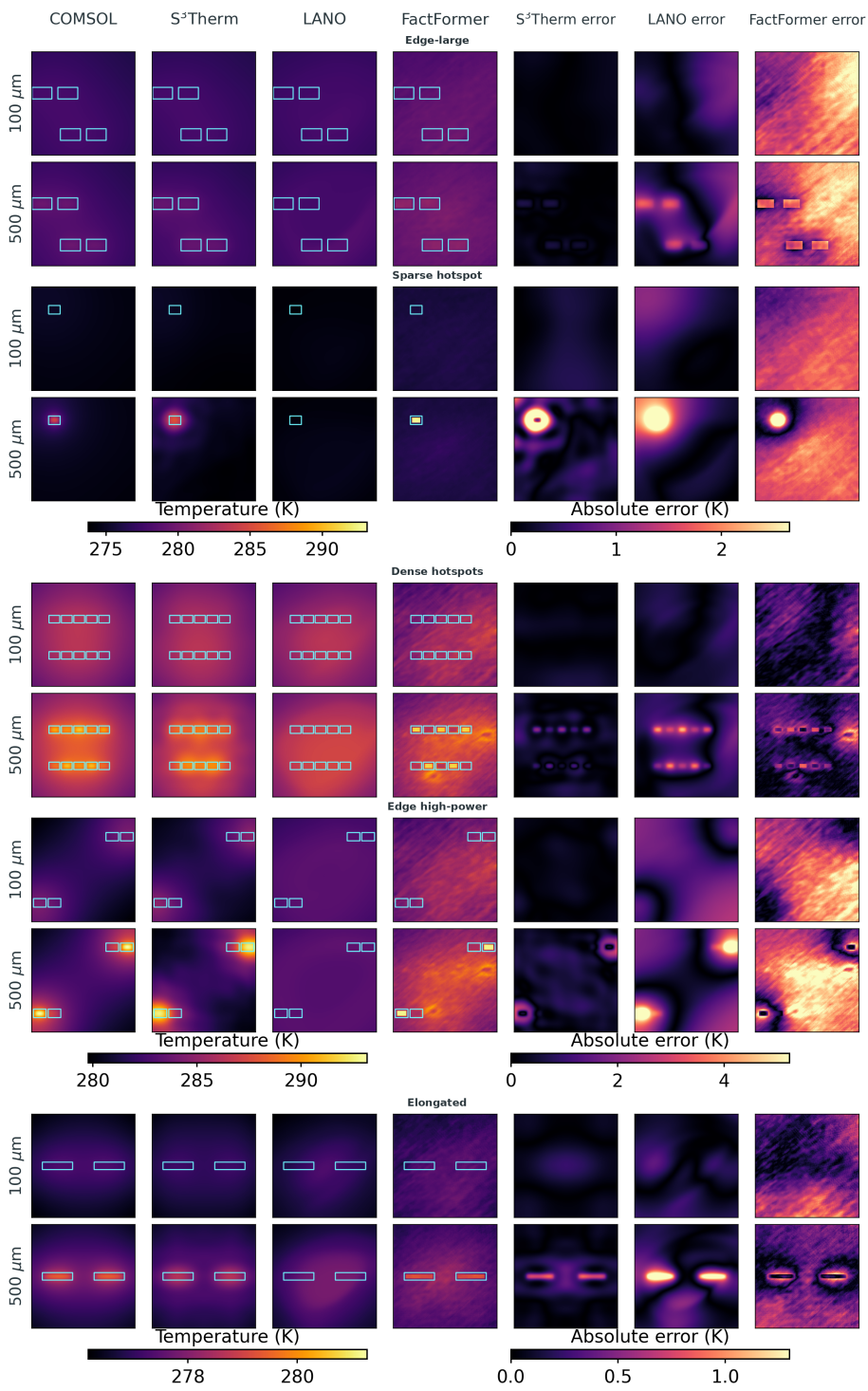


Figure 5: Qualitative overview of five held-out layouts: *Edge-large*, *Sparse hotspot*, *Dense hotspots*, *Edge high-power*, and *Elongated*. Each layout has rows at $z_{\text{phys}} = 100$ and 500 μm; columns are the COMSOL reference, the S^{3}Therm, LANO, FactFormer, and the absolute error of each predictor. Cyan contours mark projected source footprints. The audited error caps are 2.6 K, 5.2 K, and 1.3 K for the three source-layout groups, with each group's temperature and error scales shown beneath that group.

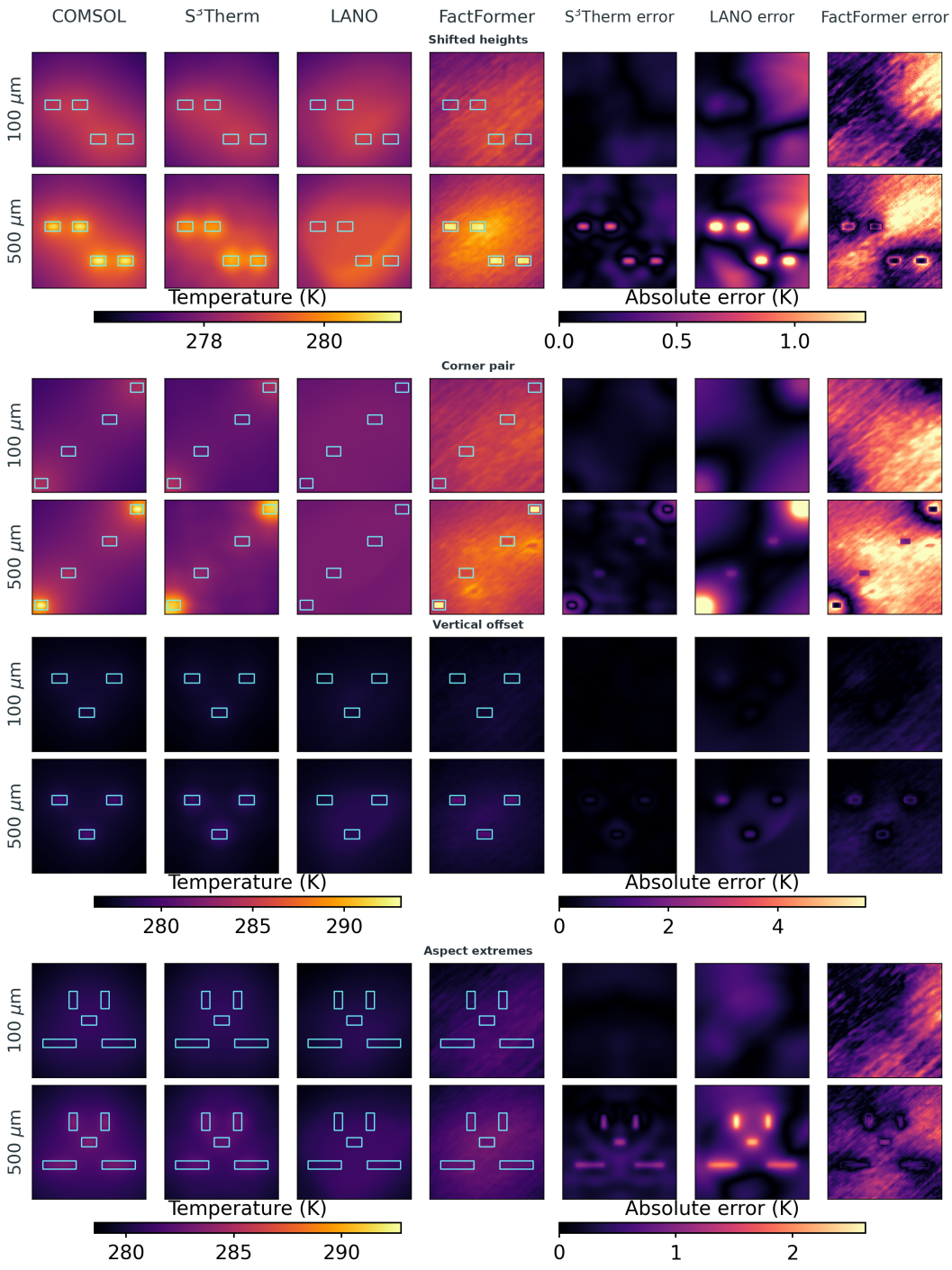


Figure 6: Qualitative overview of the remaining four held-out layouts: *Shifted heights*, *Corner pair*, *Vertical offset*, and *Aspect extremes*, drawn as in Figure 5. The audited error caps are 1.3 K, 5.6 K, and 2.6 K for the three source-layout groups, with each group's temperature and error scales shown beneath that group.

Algorithm 1 S^{3}Therm inference: sheet-response composition followed by integral-balance and symmetry correction.

Require: Encoded source Q_{enc}, query points X, trained response $\mathcal{H}_\theta$, in-plane symmetry group $\mathcal{G}$

$\widehat{u} \leftarrow 0$
for $g \in \mathcal{G}$ **do**
  $Q^g \leftarrow \rho_Q(g) Q_{\mathrm{enc}}$
  Retain the active height intervals and group them into $\mathcal{B} = \{B_m\}_{m=1}^{M}$
  $u^g \leftarrow 0$
  for $m = 1, \ldots, M$ **do**
    Compute $(\bar{q}_m, z_m, a_m)$ using Eq. (3)
    $u^g \leftarrow u^g + a_m \mathcal{H}_\theta(\bar{q}_m, X, z_m)$
  end for
  Compute $\widehat{S}_{\mathrm{src}}$, $\widehat{S}_{\mathrm{bdry}}$, and α using Eq. (7)
  $\widehat{u} \leftarrow \widehat{u} + \rho_T(g)^{-1}(\alpha u^g)$
end for
return $\widehat{u}/|\mathcal{G}|$

A.4 Inference Algorithm and Lifted Source

Algorithm 1 summarises the inference procedure: the encoded source is transformed over the in-plane symmetry group, reduced to the retained height intervals, composed from the response operator, and corrected by the integral-balance rescaling of Section 4.4.