Real-Time Precision Prediction of 3-D Package Thermal Maps via Image-to-Image Translation

Michael Joseph Smith[1*], Seunghyun Hwang[2*],Vinicius Cabral Do Nascimento[2],
Qiang Qiu[2], Cheng-Kok Koh[2], Ganesh Subbarayan[1], Dan Jiao[2]
[1]*School of Mechanical Engineering,* [2]*Elmore Family School of Electrical and Computer Engineering*
Purdue University, West Lafayette, IN 47907, USA
smit4789@purdue.edu, hwang229@purdue.edu

***Abstract*—Thermal optimization plays a crucial role in the design of advanced packaging technologies. Due to the large number of thermal simulations needed for optimal design, reductions in simulation run-time are critical. Here, we cast the temperature solution process into an image-to-image translation problem. We model the power generation map, conductivity map, and boundary conditions into separate channels of an image. We then generate temperature solutions by training a conditional image generative model, composed of a U-Net shaped generator and a discriminator, using deep neural networks (NNs). The resultant NN model exhibits superior accuracy for unseen inputs. Speed wise, it enables near real-time design, providing a 1663x and 14,885x speedup over a sparse matrix optimized finite element method (FEM) and ABAQUS® respectively.**

***Index Terms*—Image-Based Learning, Heterogeneous Integration, Thermal Integrity**

I. Introduction

With the growing thermal challenges of advanced packaging and the thousands of simulations required for effective thermal design optimization, speeding up these simulations is critical. Recently, this has been attempted by creating a neural network (NN) replacement for a conventional 3D steady-state thermal solver. This approach is ideal for design optimization tasks because, once trained, the neural network can provide high-resolution thermal solutions in milliseconds. However, due to their large training times, neural network based methods are only useful when they can accurately capture the entire design space [1], [2].

Neural network models for 3D steady-state heat conduction can be split into two categories: Physics Informed and Data Driven. In physics informed NN models, a domain's governing partial differential equations are calculated during training and used as the loss function. Liu et al. [3] recently developed a physics informed model capable of evaluating the thermal solution of homogeneous objects with arbitrary boundary conditions and a power generation map applied to the top surface. In data driven approaches, simulation results or data have been used for training to generate an NN-based surrogate model. Chen et al. [4] used a lumped element approximation of 3D steady state heat conduction combined with graph neural networks as a replacement simulation. This problem has also been solved with a convolutional neural network. Wang et al. [5] encoded material and power generation input data as 3D matrices and leveraged a U-net architecture for training. However, previous methods have yet to demonstrate full generalization of the heat conduction equation.

In this paper, we introduce the first fast image-to-image translation technique to generate an NN-model of high fidelity to replace general 3D steady state conduction simulation. Our generalized method allows for arbitrary conductivity and power generation anywhere in the domain as well as free choice of Neumann and Robin conditions. Our model resolves all material, power generation, and boundary condition data into just three channels of an image and can evaluate a thermal map in 0.0035 seconds on an RTX A5000 GPU, providing a 1663x speedup over sparse matrix optimized FEM.

II. Methodology

A. Steady State Heat Conduction

The strong form of steady state heat conduction with generalized boundary conditions is shown in (1-3),

$$\nabla \cdot \vec{q} - s = 0 \text{ in } \Omega. \tag{1}$$

$$\vec{q} = -k\nabla T \text{ in } \Omega. \tag{2}$$

$$\hat{n} \cdot (-k\nabla T - \bar{q}) = h(T - \bar{T}) \text{ on } \Gamma. \tag{3}$$

In equations (1-3), $\vec{q}$ is heat flux, s is heat generation, k is isotropic thermal conductivity, and h is the convection coefficient. Equation (3) is a generalized boundary condition that can be used to represent Dirichlet, Neumann, and Robin boundary conditions. Dirichlet boundary conditions are not represented directly but are approximated by raising h to a very high value. By the usual derivation of the weak form, and applying the finite element approximation, the discretized versions of equations (1-3) can be obtained,

$$\begin{aligned}&\left[k\int_\Omega \{\nabla N\}[\nabla N]\,d\Omega + h\int_\Gamma \{N\}[N]\,d\Omega\right]\{T\}\\ &= \int_\Omega s\{N\}\,d\Omega - \int_\Gamma \bar{q}\{N\}\,d\Gamma + \int_\Gamma h\bar{T}\{N\}\,d\Gamma\end{aligned} \tag{4}$$

By condensing (4), the standard linear equation (5) appears.

$$[K]\{T\} = \{f\} \tag{5}$$

This work was supported by Rapid-HI (Heterogeneous Integration) Design Institute (an Elmore ECE Emerging Frontiers Center) and an NSF Future of Semiconductors (FuSe) grant under award No. 10002201. (*Michael Joseph Smith and Seunghyun Hwang are co-first authors.)

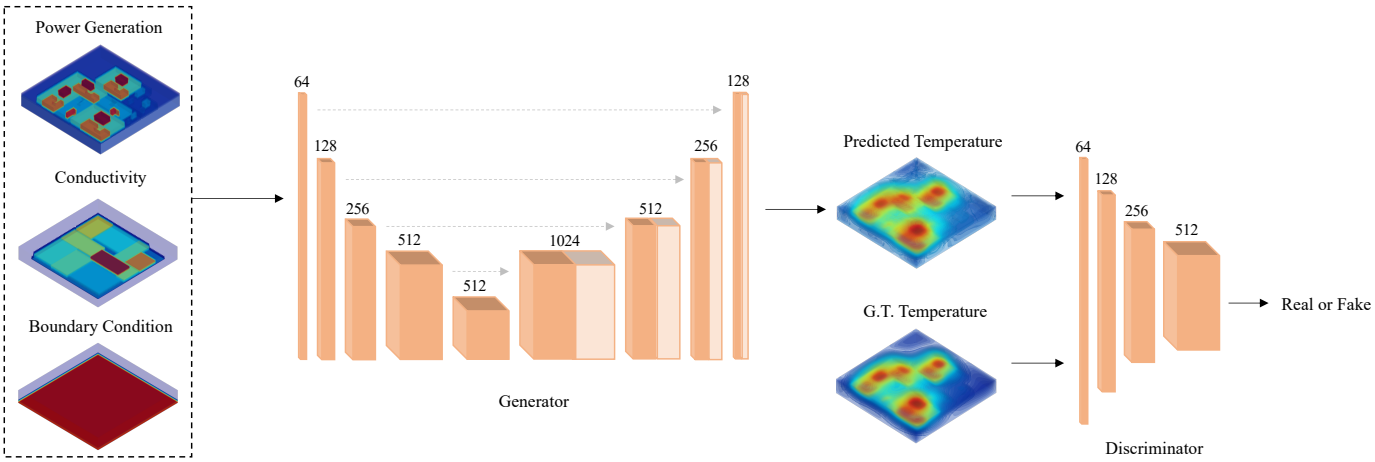

Fig. 1. Overview of the proposed image-based learning approach: input conditions (left) are stacked as respective channels of an input image to the U-Net shaped generator, which predicts a temperature solution. The discriminator distinguishes the real temperature solution from the data created by the generator. Each block in the generator and the discriminator represents different feature maps, where the number on top denotes channel dimension.

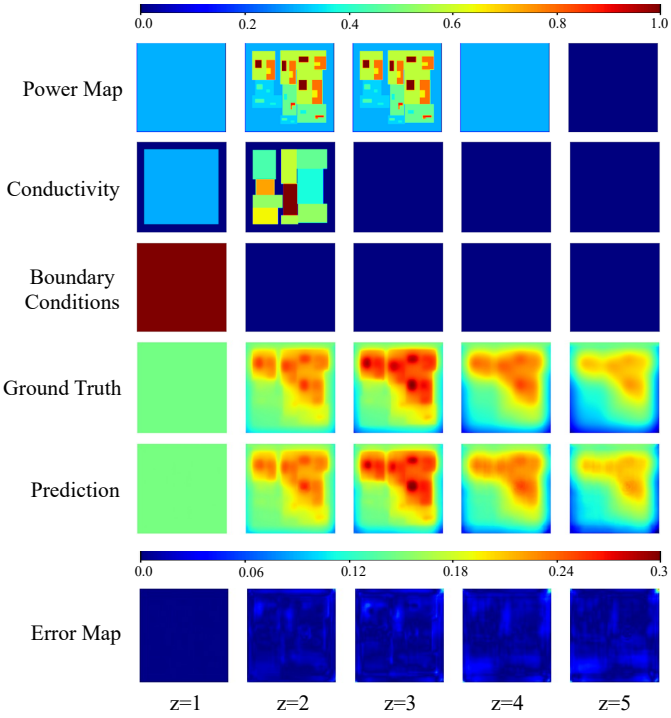

Fig. 2. 2D visualizations of 3-D results. Vertical slices of the xy plane from z=1 (left) to z=5 (right). Bottom colorbar indicates color scale of the error map, while the top one indicates color scale of all other rows.

Equation (5) provides the basis of our method. As $\{f\}$ is a node based quantity, it can be directly mapped to pixels in a 3D image. As the combined $[K]$ matrix is a non-nodal quantity, we feed k and h to the network as separate 3D image inputs. The network is then trained to calculate $\{T\}$.

B. Data Generation

Data was generated for training using a custom sparse matrix optimized finite element solver (TStack3D) described in [6]. Simulations contained 10 chiplets, each with a different conductivity, in different positions on top of a silicon interposer encased in mold compound. A floorplan algorithm was developed to randomize the placement of the 10 chiplets to generate training data. 1200 samples were generated by varying chiplet floorplan, power maps, and boundary condition placement and magnitude. The bottom plane of the substrate (z=0) and ambient temperature were held at 293K during the simulations. 1200 samples were split into a 9:1 ratio for training and testing, which is a ratio frequently used for training deep learning models in a small scale dataset.

C. Learning Algorithm

Inspired by conditional image generation using deep NNs [7], we take an image-based approach to 3D temperature solution prediction. We model the power generation map, conductivity map, and boundary conditions as separate channels of an image. Then, we use this image as the input condition to our learned NN to generate an output temperature solution image. In this way, we cast temperature solution prediction into an image-to-image translation problem. The input and output image size to the NN can be flexible according to the problem, e.g, an output image size 121x121x5 is adopted in this paper. An illustration of our overall method is presented in Fig. 1.

More specifically, given image-based input conditions, we generate temperature solutions by training a conditional image generative model, composed of a U-Net [8] shaped generator and a discriminator. To handle stacks of 3D-shaped inputs, we adopt 3D convolutional filters in our model, instead of 2D convolutional filters in typical image generative networks. All input conditions, i.e., power generation map, conductivity map, and boundary conditions, are stacked together as respective channels of an input image to the generator. Note that all the inputs are normalized to the range of [0-1]. To accommodate various sizes of chiplets, we adopt the size of 5 for convolutional kernels. Based on our empirical observations, with such larger-than-usual receptive fields, temperature distribution of large-size chiplets is better captured by our model. At every iteration of training, the generated temperature solution

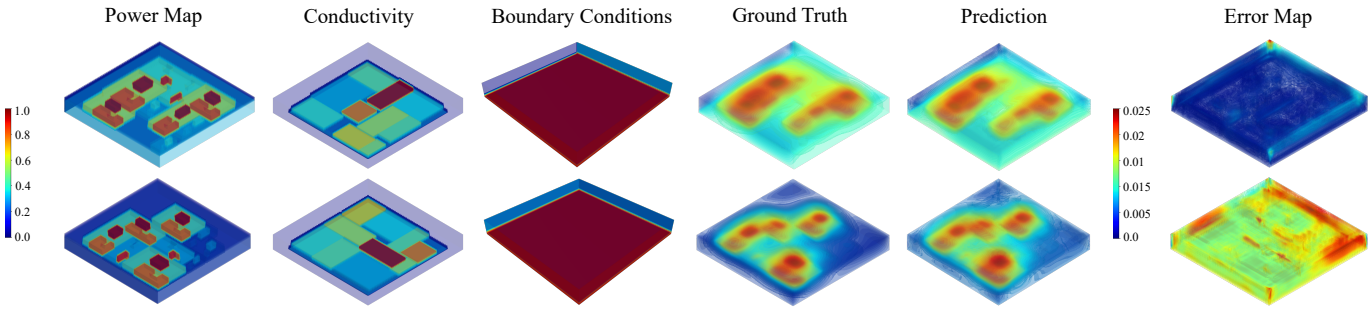

Fig. 3. Predicted thermal solutions of two 10-chiplet packages. All 3D visualizations except MSE are normalized to [0,1]. The Right colorbar indicates color scale of the error map, while the left one indicates the color scale of non-error quantities. The boundary conditions plots are in log scale for visualization.

is processed by the discriminator. The discriminator tries to distinguish the real temperature solution from the one generated by the model. Following the min-max optimization method for training of GANs [9], the training objective of the discriminator and the generator can be reached by:

$$L_{GAN} = \arg\min_{G}\max_{D}[\mathbb{E}_{x,y}[\log(D(x,y))] + \mathbb{E}_{x}[\log(1-D(G(x)))]], \tag{6}$$

where x denotes input 3D images and y denotes the real temperature solution. Following [7], an additional L_1 loss is employed in (7) to ensure the match between the predicted and the ground truth temperature map. The total training Loss is shown in (8).

$$L_1 = \mathbb{E}_{x,y}[||y - G(x)||_1] \tag{7}$$

$$L_{total} = L_{GAN} + \lambda L_1 \tag{8}$$

Here, λ is set to 100 during training. After iterations of min-max optimization of the generator and discriminator, only the generator is used at the inference stage.

For better generation, we added skip connections to the U-Net shaped generator between mirrored layers in the encoder and decoder. The model is trained for 180 epochs using the Adam optimizer. The initial learning rate is set to be 2e-4 and linearly decays for the first 100 epochs. A single RTX A5000 GPU is used for training and testing.

III. Results

After being trained on 1080 training samples, our model is validated on 120 testing samples. We use mean squared errors (MSEs) and peak errors (PEs) to measure the performance presented in Table I. The average inference time is also included. Our method provides 1663x speedup over sparse matrix optimized FEM [6].

2D and 3D visualizations of the model inputs and outputs are presented in Fig. 2, and 3, respectively. In Fig. 3, the top and bottom rows had peak temperatures of 413K and 421K respectively, showing that a better floorplan can help reduce hot-spots. It is noteworthy that the error maps produced by measuring absolute discrepancy of the true and predicted temperature solutions nearly converge to zero.

TABLE I
Inference time and average errors normalized [0-1]

Method	*MSE [K]^2*	*PE [K]*	*Inference Time [s]*
Ours	0.0002	0.0808	0.0035
TStack3D [6]	Reference	Reference	5.8235
ABAQUS [10]	Reference	Reference	52.1

IV. Conclusion

For efficient learning of 3D steady-state heat conduction, we introduced a versatile image based learning system. We predict 3D thermal solutions using image-to-image translation by stacking arbitrary power generation, conductivity, and boundary conditions into separate channels of input images. To the best of our knowledge, we are the first to propose a general image-based approach that can accommodate most chiplets' positions, conductivity, power generation, and a free choice of Neumann and Robin conditions. Evaluation of unseen chiplet layouts verified the accurate prediction ability of our method with negligible testing time.

References

[1] L. Zhang and G. Subbarayan, "An evaluation of back-propagation neural networks for the optimal design of structural systems: Part I. training procedures," Computer Methods in Applied Mechanics and Engineering, vol. 191, no. 25–26, pp. 2873–2886, 2002.
[2] L. Zhang and G. Subbarayan, "An evaluation of back-propagation neural networks for the optimal design of structural systems: Part II. numerical evaluation," Computer Methods in Applied Mechanics and Engineering, vol. 191, no. 25–26, pp. 2887–2904, 2002.
[3] Z. Liu et al., "DeepOHeat: operator learning-based ultra-fast thermal simulation in 3D-IC design," IEEE Design Automation Conference (DAC), July 2023, arXiv:2302.12949.
[4] L. Chen, W. Jin and S. X. -D. Tan, "Fast thermal analysis for chiplet design based on graph convolution networks," 2022 27th Asia and South Pacific Design Automation Conference (ASP-DAC), Taipei, Taiwan, 2022, pp. 485-492.
[5] Y. Wang, J. Zhou, Q. Ren, Y. Li and D. Su, "3-D steady heat conduction solver via deep learning," in IEEE Journal on Multiscale and Multiphysics Computational Techniques, vol. 6, pp. 100-108, 2021.
[6] C. Chen, "Enriched Isogeometric Analysis for Parametric Domain Decomposition and Fracture Analysis," Ph.D. dissertation, Dept. Mech. Eng., Purdue Univ., West Lafayette, IN, USA, 2020.
[7] P. Sola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," Proceedings of the IEEE conference on computer vision and pattern recognition, 2017.
[8] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation." Med. Image Comput. Comput.-Assist. Interv., 2015, pp. 234-241.
[9] I. J. Goodfellow et al., "Generative adversarial networks," Proceedings of Advances in Neural Information Processing Systems, 2014, pp. 2672–2680.
[10] Abaqus FEA. (2020). Dassault Systemes.