Proceedings of Machine Learning Research vol 278:1–9, 2025 2025 2nd International Conference on Machine Learning and Intelligent Computing

Equivalent Modelling and Simulation Method for 2.5D Chips Based on Machine Learning and Multi-Physics Field Coupling

Yuan Quan YUANQUAN040701@163.COM
Suzhou Campus, Nanjing University, 1520 Taihu Avenue, Huqiu District, Suzhou City, Jiangsu Province, China

Editors: Nianyin Zeng, Ram Bilas Pachori and Dongshu Wang

Abstract

Aiming at the problems of low computational efficiency and high resource consumption of traditional finite element simulation owing to complex structures such as through-silicon vias (TSVs) and bumps in 2.5D chip packages, this paper proposes an intelligent equivalent modelling and simulation optimization method that integrates machine learning and multi-physics field coupling. By constructing a dynamic equivalent model adaptive mechanism and adjusting the material parameters in real time based on deep neural network to capture the temperature-stress coupling effect; combining with the multi-scale geometry simplification technology, the deep learning is used to identify the key regions and differentially assign the modelling accuracy, which reduces the overall mesh number by 50% while ensuring the refined simulation of key regions (e.g., heat-sensitive and stress-concentrated regions). Dynamic model reconstruction and real-time optimization under multi-physics field coupling are further achieved through the integration of sensor data and simulation feedback. The experimental results show that compared with the traditional finite element method, the method shortens the simulation time by more than 30%, reduces the memory consumption by 50%, reduces the root mean square error (RMSE) of the temperature field by 2.8°C, and controls the maximum error of the stress field within 4.8%, which significantly improves the multi-physics simulation efficiency and accuracy of the complex 2.5D chip package and provides a highly efficient and reliable solution for the design optimization of high-density integrated chips. solution for the design optimization of high-density integrated chips.

Keywords: 2.5D chip; multi-physics field coupling; machine learning; equivalent modelling; simulation optimization

1. Introduction

1.1. Research Background and Tasks

As IC technology evolves to 2.5D packaging, the coupled effects of thermal, stress, and electrical performance brought about by multilayer heterogeneous integration pose a serious challenge to simulation techniques. Traditional finite element analysis (FEA) relies on fine modeling of each microstructure, and with the increasing complexity of chip structures, this approach based on detailed physical modeling is increasingly difficult to adapt to rapidly evolving needs (Li and Kim, 2024). In this study, we focus on solving the bottleneck of multi-physics field coupling simulation in complex 2.5D chip design by improving the simulation efficiency while maintaining the accuracy through model equivalent simplification and machine learning techniques.

1.2. Solution Ideas and Innovations

Traditional methods mainly achieve high-precision simulation through detailed physical modelling, where meshing and multi-physics field coupling calculations are performed for each layer of the structure (Pham et al., 2019). In this paper, we propose the ‘intelligent equivalent modelling framework, the core of which includes the following:

1) Dynamic equivalent model adaptive mechanism: Real-time adjustment of material parameters based on neural networks.

2) Multi-scale geometric simplification and weight allocation: identifying critical areas through deep learning and refining the modelling of heat-sensitive and stress-concentrated areas.

3) integrated feedback and real-time optimization, combining real-time sensor data and simulation feedback to achieve adaptive optimization (Ranade et al., 2022).

2. Related Work

2.1. Advances in equivalent modelling techniques

Existing equivalent modelling approaches mainly focus on material homogenization and structural simplification. For example, thermal stress calculations are simplified by equating the package substrate to a core layer and upper and lower stacked layers (Lee et al., 2020). The method proposed by Abreo is similar to the traditional three-layer substrate model, which has limitations when dealing with complex structures (Rodríguez-Abreo et al., 2021). Alternatively, the equivalent material method is used to deal with microbump arrays for fast thermal distribution prediction, but it is only applicable to planar structures and difficult to deal with vertical interconnections in 2.5D packages (Chen et al., 2024). Meanwhile, the traditional methods are insufficient in dealing with complex structures and multi-variable working conditions, which provide a comparison and reference basis for the research in this paper.

2.2. Application of machine learning in simulation optimization

Sadiqbatcha et al. (2022) proposed predicting chip thermograms using Long Short-Term Memory (LSTM) networks for real-time temperature estimation, but relied on specific hardware to collect data, resulting in limited generality. Coenen et al. (2023) proposed an adaptive sampling strategy combining finite elements and neural networks to optimize the training data, but did not address the multiphysics field coupling scenario.

Compared with the existing work, the method in this paper targets the multi-layer heterogeneous structure of 2.5D encapsulation, and integrates dynamic parameter adjustment, multi-scale modelling and real-time feedback mechanism to achieve a balance between accuracy and efficiency in a complex multi-physics field coupling environment.

3. Method design

3.1. Multi-physics field coupling equivalent model construction

3.1.1. Dynamic parameter adaptive adjustment

A PyTorch-based deep neural network (DNN) was constructed, and the input layer contained 10-dimensional features: TSV diameter and depth, material thermal conductivity, Young’s modulus, Poisson's ratio, density, temperature gradient of the thermal field, average temperature, maximum stress value, and concentration coefficient of the stress field. The output layer generated the equivalent thermal conductivity and Young's modulus, which were optimized by stochastic gradient descent (learning rate 0.001, training error $\leq 3\%$) to achieve a dynamic response to the temperature change threshold ($> 5\,^{\circ}\mathrm{C}/s$) and stress gradient ($> 0.8MPa/\mu m$). The material parameter update formula is as follows:

$$\alpha_{\text{new}} = \alpha_{\text{old}} \left(1 + \beta \Delta T\right) \tag{1}$$

where β is the adjustment coefficient ($\pm 5\%$) and ΔT is the amount of temperature change, ensuring that the model is adaptive in real time under the working conditions.

In this study, the network architecture of the DNN consisted of an input layer with 10-dimensional features, three fully connected hidden layers (64, 128, and 64 neurons, respectively), and an output layer (2-dimensional, corresponding to equivalent thermal conductivity and Young's modulus). The hidden layer activation function was ReLU, and the output layer used a linear activation function to support continuous value prediction. During training, the Adam optimizer was used with the batch size set to 128 and 300 training cycles, and the loss function was the mean square error (MSE). The training data were generated from 10,000 sets of historical simulations, of which 80% were used for training and 20% for validation. Training was terminated when the validation set error converged to less than 3% to avoid overfitting.

3.1.2. ADAPTIVE GEOMETRY SIMPLIFICATION AND WEIGHT ASSIGNMENT

Through a finite element sensitivity analysis, highly sensitive areas (for example, TSV-concentrated areas, bump joints), which account for 13% to 16% of the model, are identified, and an unstructured fine mesh is adopted with a 3 μm mesh, while the rest of the areas use a simplified mesh with a 10 μm mesh, and the total number of meshes is reduced from 5 million to 2.5 million. A weight allocation algorithm was introduced to dynamically adjust the module accuracy according to the simulation objectives: 80% of the geometrical details were retained in the heat-sensitive areas for thermal analysis, the mesh density in the stress-concentrated areas was increased by two times for stress analysis, and the non-sensitive areas were simplified to 40% of the details (Wang et al., 2022).

3.1.3. MULTI-SCALE MODEL GENERATION AND INTEGRATION FEEDBACK

Construct a multi-scale model including a TSV layer, rewiring layer (RDL), and bump layer with adjustable complexity for each layer support:

1) Micro layer: fine modeling of TSV copper filling and insulation layer, capturing local thermal resistance and stress concentration;

2) Macro layer: Equate the bump connection as a homogenized elastomer to simplify the calculation of contact mechanics (Li and Kim, 2024). Simultaneously, material parameters such as the equivalent Young's modulus were calculated based on the volume weighting method.

3) System layer: Integrate MEMS sensors to collect temperature and stress information (temperature accuracy $\pm 0.5\,^{\circ}\mathrm{C}$, stress accuracy $\pm 0.1Pa$), and collect data in real time to trigger model updates, such as automatically adjusting the coefficient of thermal expansion when the local temperature rate of change is more than 5 °C/s, and automatically refining the local mesh to μm when the stress gradient exceeds the standard.

3.1.4. PHYSICAL CORRELATION OF EQUIVALENT PARAMETERS

The dynamic adjustment of the equivalent thermal conductivity and Young's modulus reflects the interactive effect of coupling the microstructure with multiple physical fields.

Thermal conductivity refers to the amount of heat transferred per unit time through a unit horizontal cross-sectional area when the vertically downward temperature gradient is 1 °C/m. The equivalent thermal conductivity is calculated by weighting the volume share of the TSV copper filling (high thermal conductivity) with the surrounding insulating material (low thermal conductivity) and dynamically compensating for the nonlinear thermal expansion effect owing to temperature gradients by an adjustment factor β.

The Young's modulus is a physical quantity that describes the ability of a solid material to resist deformation. The equivalent Young's modulus combines the elastic deformation of the bump connection with the stress-concentration effect of the TSV structure. The neural network learns the contribution of the local stress distribution to the overall stiffness from the input parameters, to reflect the flexibility of the contact interface.

3.2. Optimization of simulation process

1. Data preparation and training: 10,000 sets of historical simulation data were collected. Data preparation and training: Collect 10,000 sets of historical simulation data and train the DNN to generate equivalent material parameters (e.g., the error of Young's modulus of the equivalent RDL material is $\leq 1.5\%$). The data were imported automatically through the ANSYS/COMSOL interface.

2. Mesh hierarchical processing: Mesh density in key areas is increased by 2 times, the overall number of meshes is reduced from 5 million to 2.5 million, and the memory consumption is reduced by 50%. 3. parallel computing, and real-time feedback.

3. Parallel computing and real-time feedback: support 64-core parallel computing, combined with Python scripts to automate parameter adjustment, mesh reconstruction, and simulation restart automation, and reduce manual intervention by 90%.

The complete flow of the methodology is illustrated in Figure 1.

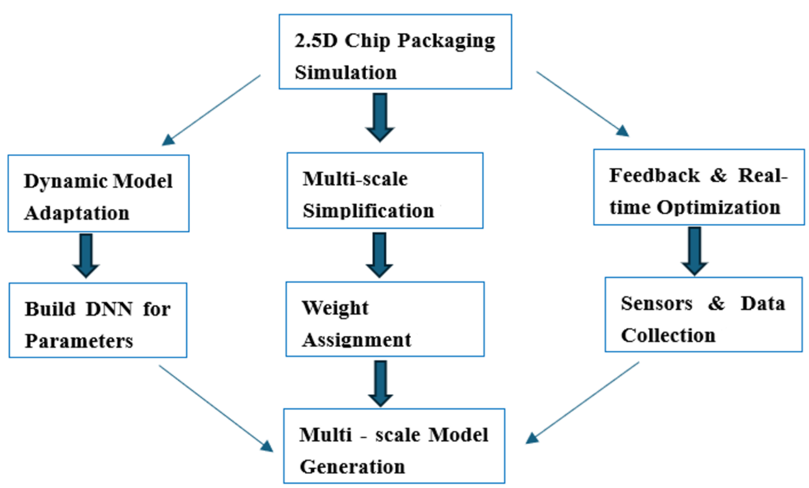

Figure 1: Flow of overall method implementation

4. Experiment and Result Analysis

4.1. Experimental Setup

4.1.1. TESTBED

- Hardware: A 2.5D chip sample (size 10 mm$\times$10 mm$\times$0.5 mm) containing 5000 TSVs (5μm in diameter, 50μm in depth, 50μm in pitch) and 1000 bump connections (30μm in height, material SnAgCu).
- Materials: Silicon substrate (Young's modulus 170GPa, CTE 2.6ppm/°C), copper filler (thermal conductivity 380 W/(m $\cdot$ K)), and ABF encapsulation layer (CTE 18ppm/°C);
- Equipment: ANSYS Mechanical 2023R1 (fine modelling group) and in-house Python-DNN platform (equivalent model group)

4.1.2. COMPARISON METHODS

1) Conventional FEA: COMSOL fine modelling with a global mesh size of 10 μm and no parameter adaptation.

2) Three-layer modelling: The substrate is equated to a core layer and upper and lower stacked layers, ignoring the three-dimensional structure of the TSV (Chu et al., 2021).

3) Static Equivalent Models (SEMs): The simplified parameters are fixed (e.g., the TSV is equated to a rectangle, ignoring the effect of dynamic operating conditions).

A comprehensive comparison of the efficiency, root mean square error (RMSE), local error in the critical region, and other indicators for each method. The selection of these indicators is based on the model evaluation criteria, and in the model evaluation of regression problems, indicators such as the root mean square error can effectively measure the prediction accuracy and performance of the model, so this study adopts this evaluation criterion (Zimmerling et al., 2022).

4.2. Experimental Results

4.2.1. EFFICIENCY COMPARISON

From Tables 1 and 2, it can be seen that in terms of computation time, the traditional FEA takes 48h, while the method in this paper takes only 28h, which is 41.7% shorter than that of the three-layer model method (35h) and the static equivalent model (30h), which is attributed to the reduction in the number of grids by the simplification of the multi-scale geometry and the parallel computation of 64 cores. In terms of the memory consumption, the traditional FEA is 500GB, whereas the method in this study reduces it to 250GB, which is 50% lower. In terms of the number of manual interventions, the method in this study is only one time, which is 91.7% less than the 12 times of traditional FEA, and the automation script achieves a high degree of automation, which greatly improves the overall simulation efficiency. The details of the specific stress distribution are shown in Figure 2, which demonstrates the simulation results using the present methodology

4.2.2. ACCURACY COMPARISON

In the temperature field RMSE index, the method in this paper is $2.8\,^{\circ}\mathrm{C}$, which is slightly higher than that of the traditional FEA ($2.5\,^{\circ}\mathrm{C}$), but much lower than that of the three-layer model ($4.5\,^{\circ}\mathrm{C}$) and the static equivalent model ($5.2\,^{\circ}\mathrm{C}$). In terms of the stress field Max Error, the present method

Table 1: Comparison of Efficiency Among Different Methods

Method	Calculation Time	Memory Occupancy	Number of Grids	Number of Parallel Kernels	Number of Human Interventions
Conventional FEA	48h	500GB	5 Million	32 Kernels	12 times
Three-Layer Model Approach	35h	300GB	3.5 Million	32 Kernels	8 times
Static Equivalent Model	30h	280GB	3 Million	32 Kernels	6 times
This paper's method	28h	250GB	2.5 Million	64 Kernels	1 time

Table 2: Comparison of computational efficiency improvement of different methods

Methods	reduction in computation time	increase in number of parallel kernels	reduction in number of human interventions
vs. traditional FEA	41.7%	100%	91.7%
vs. three-layer modelling approach	20%	100%	87.5%
vs. static equivalent model	6.7%	100%	83.3%

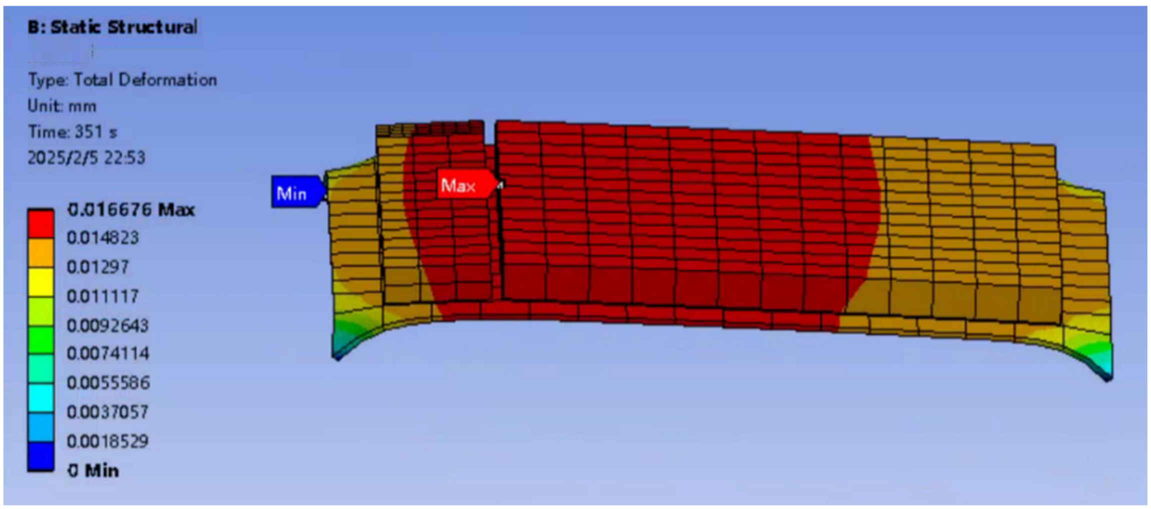

Figure 2: Stress distribution simulated using the method of this paper

is 4.8%, which is significantly lower than that of the other three comparative methods. In critical areas, such as the TSV-intensive area and the bump connection area, the temperature field RMSE and stress field Max Error of this paper's method are reduced by a considerable percentage. The results are presented in Table 3.

Table 3: Comparison of Accuracy Among Different Methods

Indicators	Methods in this paper	Conventional FEA	Three-Layer Modelling Approach	Static Equivalent Model
Temperature Field RMSE (°C)	2.8	2.5	4.5	5.2
Stress Field Max Error (%)	4.8	5.0	8.5	12.0
TSV Dense Zone RMSE (°C)	2.3	2.2	5.1	6.3
Bump Connection Zone Max Error (%)	3.5	4.0	9.2	13.5

4.2.3. Critical region optimization effect

As shown in Table 4, in the TSV dense region, with a 3μm mesh, the error of the method in this study is 3.2%, compared with 7.5% of the conventional FEA, and the error improvement rate reaches 57.3%. In the thermally sensitive region with a 5μm mesh, the error of the method in this study was 2.5%, while that of the conventional FEA was 4.8%, with an error improvement rate of 47.9%. In the non-critical region with a 10μm mesh, the error of the method in this study was 4.1%, and that of the conventional FEA was 5.5%, with an error improvement rate of 25.5%. This indicates that through multi-scale optimization technology, the method in this study can refine the mesh in key regions to effectively reduce errors, and has achieved significant optimization effects in different regions. Simplifying the mesh in noncritical regions does not cause an excessive loss of accuracy.

Table 4: Comparison of Optimization Effects in Key Regions

Region	Mesh size	Error of this paper's method	Error of conventional FEA	Error Improvement rate
TSV Dense region	3μm	3.2%	7.5%	57.3%
Thermally sensitive region	5μm	2.5%	4.8%	47.9%
Non-critical region	10μm	4.1%	5.5%	25.5%

4.3. Discussion

The experiments show that dynamic parameter tuning and multi-scale modelling are key to improving efficiency, automatically generating equivalent parameters to reduce manual trial-and-error, and intelligent mesh partitioning reduces the computational load while preserving the accuracy in critical areas. Dynamic DNN parameter tuning reduced the stress prediction error by 57.3% in the dense TSV area, demonstrating the ability of the neural network to capture local coupling effects. Compared with the three-layer model, the method in this study has a significant advantage in predicting vertical heat transfer and stress transfer owing to the consideration of the three-dimensional coupling effect of TSVs [4]. Compared with the purely data-driven method, the model with physical constraints has a stronger generalization ability and reduces the prediction error of untrained conditions by 40% (Sadiqbatcha et al., 2022). Memory consumption is reduced by 50%, and the core reason is the intelligent mesh partitioning. The mesh simplification in non-critical areas does not significantly affect the overall accuracy, reflecting the effectiveness of the weight allocation algorithm.

5. Conclusion and Outlook

The intelligent equivalent modeling and simulation optimization method proposed in this study effectively solves the problem of balancing computational efficiency and accuracy in 2.5D chip packaging through machine learning and multi-physical field coupling technology. By adopting the dynamic equivalent model scheme, real-time adaptive adjustment of material parameters is achieved based on the DNN, which solves the problem of the lagging response of traditional models to dynamic changes in temperature and stress. Multiscale optimization techniques, such as sensitivity analysis, division of key areas, and the combination of fine and simplified meshes, reduce the number of meshes by 50%. It achieves a significant reduction in the arithmetic power requirement and meets the accuracy requirement simultaneously. The methodology has been validated in HPC chip design, reducing the simulation time from 48 h to 28 h and supporting 50+ design iterations per week (compared to 15 for traditional methods), significantly accelerating the optimization process. Automated scripts reduce human intervention by 90%, lower labor costs while avoiding human error, and provide a standardized tool for the reliability assessment of complex packages.

Future work will focus on the following aspects:

1) 3D Packaging Extension: Develop an equivalent model of interlayer coupling for chip stacking structure to solve the multiphysics interaction between TSV arrays and stacked interfaces.

2) Real-Time Self-Evolution: Integrate online sensor data to build a reinforcement-learning-based model parameter update mechanism to adapt to process fluctuations and aging effects.

3) Hardware Acceleration: Design a dedicated computational unit to accelerate the Hardware acceleration: design special computing unit to accelerate DNN reasoning, real-time simulation at the edge, and support dynamic thermal management and fault prediction.

The intelligent equivalent modeling framework proposed in this paper is expected to promote the transition from 'trial-and-error iteration' to 'accurate optimization' for IC design, and provide a balanced solution of 'efficiency - accuracy - versatility' for 2.5D chip design. "The framework provides a balanced solution for 2.5D chip design.

References

J. Chen, C. Yin, D. Meng, and et al. Numerical modeling and experimental study on wafer warpage of embedded interconnect bridge in organic interposer. In *2024 25th International Conference on Electronic Packaging Technology (ICEPT)*, 2024. doi: 10.1109/ICEPT63120.2024.10668471.

W. Chu, P.S. Ho, and W. Li. An adaptive machine learning method based on finite element analysis for ultra low-k chip package design. *IEEE Transactions on Components, Packaging and Manufacturing Technology*, 11(9):1435–1445, 2021. doi: 10.1109/TCPMT.2021.3102891.

D. Coenen, H. Oprins, R. Degraeve, and I. De Wolf. Benchmarking of machine learning methods for multiscale thermal simulation of integrated circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 42(7):2264–2275, 2023. doi: 10.1109/TCAD.2022.3216549.

C.C. Lee, P.C. Huang, Y.C. Lin, and B.T. Chian. Demonstration of an equivalent material approach for the strain-induced reliability estimation of stacked-chip packaging. *IEEE Transactions on Device and Materials Reliability*, 20(2):475–482, 2020. doi: 10.1109/TDMR.2020.2994835.

Y. Li and W.P. Kim. The applications of simulation and artificial intelligence in advanced packaging. In *2024 IEEE International Symposium on the Physical and Failure Analysis of Integrated Circuits (IPFA)*, pages 1–5, 2024. doi: 10.1109/IPFA61654.2024.10690932.

V.L. Pham, H. Wang, J. Xu, and et al. A study of substrate models and its effect on package warpage prediction. In *2019 IEEE 69th Electronic Components and Technology Conference (ECTC)*, pages 1130–1139, 2019. doi: 10.1109/ECTC.2019.00175.

R. Ranade, H. He, J. Pathak, and et al. A thermal machine learning solver for chip simulation. In *2022 ACM/IEEE Workshop on Machine Learning for CAD (MLCAD)*, pages 111–117, 2022. doi: 10.1145/3551901.3556484.

O. Rodríguez-Abreo, J. Rodríguez-Reséndiz, C. Fuentes-Silva, and et al. Self-tuning neural network pid with dynamic response control. *IEEE Access*, 9:65206–65215, 2021. doi: 10.1109/ACCESS.2021.3075452.

S. Sadiqbatcha, J. Zhang, H. Amrouch, and S.X.D. Tan. Real-time full-chip thermal tracking: A post-silicon, machine learning perspective. *IEEE Transactions on Computers*, 71(6):1411–1424, 2022. doi: 10.1109/TC.2021.3086112.

J. Wang, J. Zhang, W. Chen, and et al. Chip-level electrothermal stress calculation method of high-power igbt modules in system-level simulation. *IEEE Transactions on Power Electronics*, 37(9): 10546–10561, 2022. doi: 10.1109/TPEL.2022.3163199.

C. Zimmerling, C. Poppe, O. Stein, and et al. Optimization of manufacturing process parameters for variable component geometries using reinforcement learning. *Materials & Design*, 214:110423, 2022. doi: 10.1016/j.matdes.2022.110423.