Key Engineering Materials
ISSN: 1662-9795, Vol. 1054, pp 115-120
doi:10.4028/p-av2Xg7

Submitted: 2025-09-07
Revised: 2025-12-27
Accepted: 2026-02-28
Online: 2026-05-18

Design Space Exploration of SiC Power Module Package via Surrogate Model

Lu Wan[1,a*], Nicolo Ripamonti[1,b], Yanfei Zhao[1,c], Reza Soleimanzadeh[1,d], Arne Schröder[1,e]

[1]Hitachi Energy Research, Segelhofstrasse 1A, Baden-Dättwil, Switzerland

[a]lu.wan@hitachienergy.com, [b]nicolo.ripamonti@hitachienergy.com, [c]yanfei.zhao@hitachienergy.com, [d]reza.soleimanzadeh@hitachienergy.com, [e]arne.schroeder@hitachienergy.com

Keywords: power module, electromagnetic, packaging, surrogate model, optimization.

Abstract. This study introduces a surrogate model-based optimization methodology to explore a wide design space of power module packages for achieving user-defined electromagnetic design objectives, such as minimizing commutation loop stray inductance, gate loop stray inductance and balancing mutual inductance. A half-bridge module with four parallel SiC devices per switch position is analyzed, incorporating 17 design variables across terminals and substrate dimensions. Using Sobol sampling, 4096 design variations were simulated in Ansys Q3D to train the surrogate model, enabling efficient gradient-based single- and multi-objective optimization. Results show that the proposed methodology significantly accelerates exploration in a wide design space and outperforms traditional expert-driven methods by identifying superior electromagnetic performance.

Introduction

The advancement in transportation, such as electric vehicles, has increased the demand for silicon carbide (SiC) power modules due to their superior switching loss, speed, and current density over silicon modules. However, managing parasitic elements in SiC power module design is crucial due to the high di/dt and dv/dt feature and often challenging due to compact packages [1]. Expert-based iterative design can only cover a limited design space and may result in a local optimum as it is a time-consuming manual process. Recent research has focused on automatic layout design methods which are, however, limited in the number of design variables to achieve desired performance [2-5].

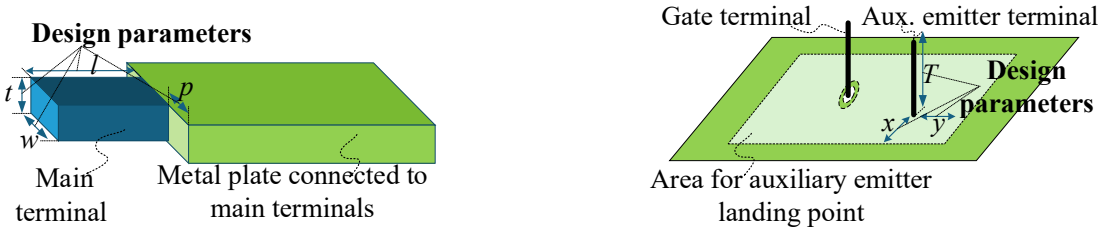
Methodology

Study Case of SiC Module and Design Variables. A half-bridge power module with four parallel-connected SiC devices at both high-side (HS) and low-side (LS) switch positions is investigated with a total number of 17 design variables. Without losing generality of the proposed methodology, the selected design variables determine three common structures of any power modules, that is, main terminals, auxiliary emitter terminals and thickness of substrates. Specifically, as shown in Fig.1a, for three main terminals (AC, DC+ and DC-), their width (w), length (l), thickness (t) and distance w.r.t. the connected metal plate (p) are selected as design variables. For the auxiliary emitter terminals, their positions (x, y) and height (h) are selected (see Fig. 1b). Besides, the thickness of substrates (d) and distance between source of chips and the metal plate (s) are selected.

The workflow of the proposed surrogate model-based optimization includes three stages.

Stage 1: Preparation of Training Database. The first stage prepares a training dataset, which should cover a variety of electromagnetic behaviors. The dataset includes the design geometry of power modules and its electromagnetic features. For the case study, 4096 design variations of this power module are evaluated in Ansys Q3D [6], where the input is the 17 design variables, and the output is the evaluated parasitic inductances: stray inductance of the commutation loop (L_s), stray inductance of the gate signal loop (L_g) and the mutual coupling between commutation loop and signal loop (M). The design variables are sampled by the Sobol sequence algorithm to achieve randomness in a large design space. The parametric simulations in Ansys Q3D were carried out at 1 MHz and the resulting

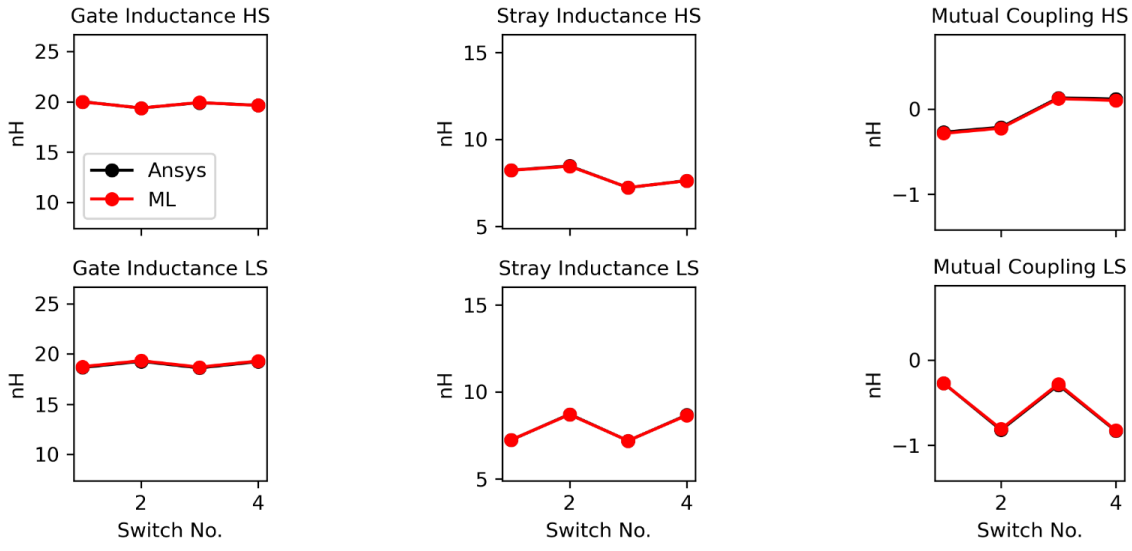
inductance matrices were exported by a Python client library that interacts with Ansys (PyAEDT). The simulation took around 125 hours for 4096 design variations. All evaluated parasitic inductances are used to create the training database for the surrogate model. The dataset was randomly split into 90% training and 10% testing. Model performance was evaluated using mean squared error on the test set at each epoch.



(a) (b)

Fig. 1. Main design parameters to determine the optimal dimensions and locations of the main terminals (a), and the location and length of the auxiliary emitter terminal of a power module (b).

Stage 2: Build-up Surrogate Model. In the second stage, a surrogate model is trained using either an Attention Network or Gaussian Process regressor, both taking 17 design variables as inputs and predicting parasitic inductances (L_s, L_g, and M). The trained surrogate model is capable of evaluating a single power module design within milliseconds (typical 3 ms), whereas Ansys Q3D typically requires approximately 30 mins for the same assessment. Thus, the computational cost reduction achieved by the surrogate model compared to full electromagnetic simulations is 99.99983%. Additionally, the surrogate model delivers rapid evaluations with predictions of parasitic inductance that closely align with those produced by Ansys Q3D. The prediction errors between the surrogate model and Ansys Q3D results are below 0.01%. One example of the comparison of the surrogate model with Ansys is shown in Fig. 2.



(a) (b) (c)

Fig. 2. Comparison of EM features per chip: (a) gate inductance, (b) stray inductance, (c) mutual coupling between commutation loop and gate signal loop.

Stage 3: Optimization of Power Module. In the last stage, the trained surrogate model is used to optimize the power module by adopting a flexible gradient-based optimization. In general, the target is to minimize stray inductance and mutual coupling and to reduce the variation of these EM features among chips on the same side. To this end, an objective function is formulated as a weighted combination of EM features (L_s, L_g and M) for both HS and LS, and deviations between chips on the same side. Specifically, the objective function can be described as:

$$F = \sum w_i \bar{f}_i + \sum w_j \Delta f_i. \quad (1)$$

where, $\bar{f}_i \in \{\overline{L_g^{HS}}, \overline{L_g^{LS}}, \overline{L_S^{HS}}, \overline{L_S^{HS}}, \overline{M^{HS}}, \overline{M^{LS}}\}$, and Δf_i being its variance over different chips, and w_i, w_j are the corresponding weights specified by the user.

Since the surrogate model is a differentiable map between design variables and EM features, the gradient-based optimization is suitable for this optimization. The optimization process starts with a randomly selected initial feasible configuration, then the algorithm optimizes the 17 design variables to minimize the objective function F following its gradient. The optimization follows the geometry constraints which are formulated as dependence among design variables.

Results and Discussion

Here we show three test cases. The first two test cases target at optimizing a single EM feature and the last case shows the capability of the surrogate model to explore a large design space to optimize multiple objects simultaneously.

Single-Objective Optimization for Gate Inductance. The goal of this test case is to minimize the gate inductance, which is reflected in the objective function $\bar{f} \in \{\overline{L_g^{HS}}, \overline{L_g^{LS}}\}$. The optimization process firstly creates a population of random initial configurations as starting points. During the optimization iterations, the algorithm leads towards optimal design to minimize the gate inductance. The illustration of several optimization iterations in one configuration depicted in Fig. 3, in terms of the position of the auxiliary emitter terminal (blue dot) with respect to the gate terminal (circle). The leftmost figures show the initial position of the auxiliary emitter terminal for a feasible configuration of the power module. During each iteration, the auxiliary emitter terminals are moved into the direction that decreases the objective function, therefore reducing the gate inductance. The rightmost figures display the optimal positions of auxiliary emitter terminals determined by the optimization process to minimize the gate inductance.

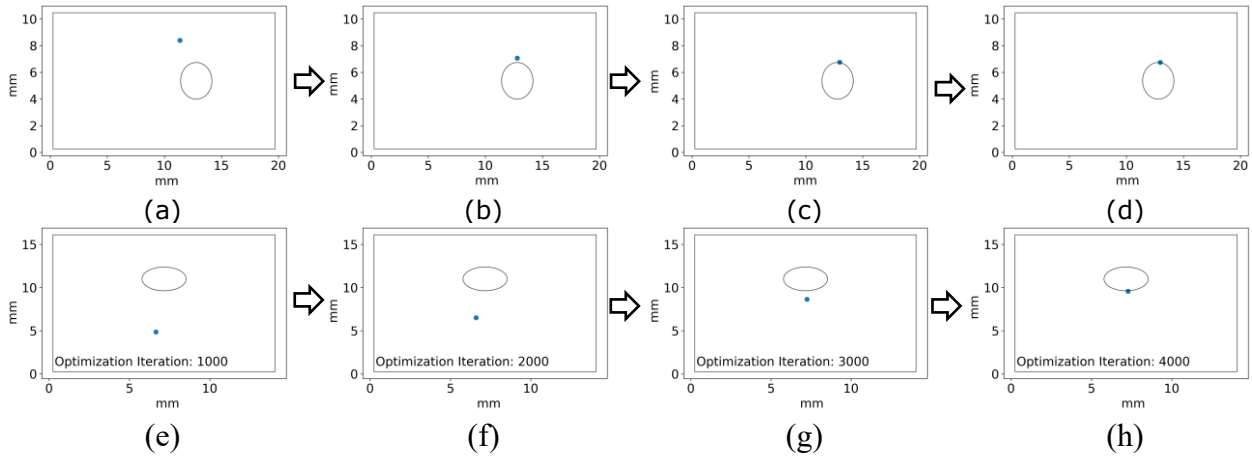


Fig. 3. This figure depicts the optimization process undertaken to reduce gate inductance, illustrating the topside interconnects of HS (a-d) and LS (e-h). The process is repeated at intervals of 1000 cycles from (a) to (d), and from (e) to (h). Blue dots indicate the locations of auxiliary emitter terminals, while circles denote the positions of gate terminals. The optimization procedure achieves minimized gate inductance by relocating the auxiliary emitter terminals in proximity to the gate terminal.

A population of random initial configurations is generated by the Sobol sequence algorithm as starting points (see blue dots in Fig. 4a-b). The optimal locations (red cross marks in Fig. 4a-b) for minimized gate inductance for these configurations are determined using the same gradient-based optimization as in Fig. 3. The red traces show the trajectories of auxiliary emitter terminals throughout the optimization process. The results indicate that the optimal designs are located near the gate terminal positions, which demonstrates that shorter gate loops yield smaller gate inductance. To further examine this, a parametric Ansys Q3D simulation is conducted, where only the position of auxiliary

emitter terminal is varied uniformly on the topside interconnects (see Fig. 4c-d). The simulation results validate the conclusion that the minimum gate inductance can be achieved when the auxiliary emitter terminal is positioned close to the gate terminal.

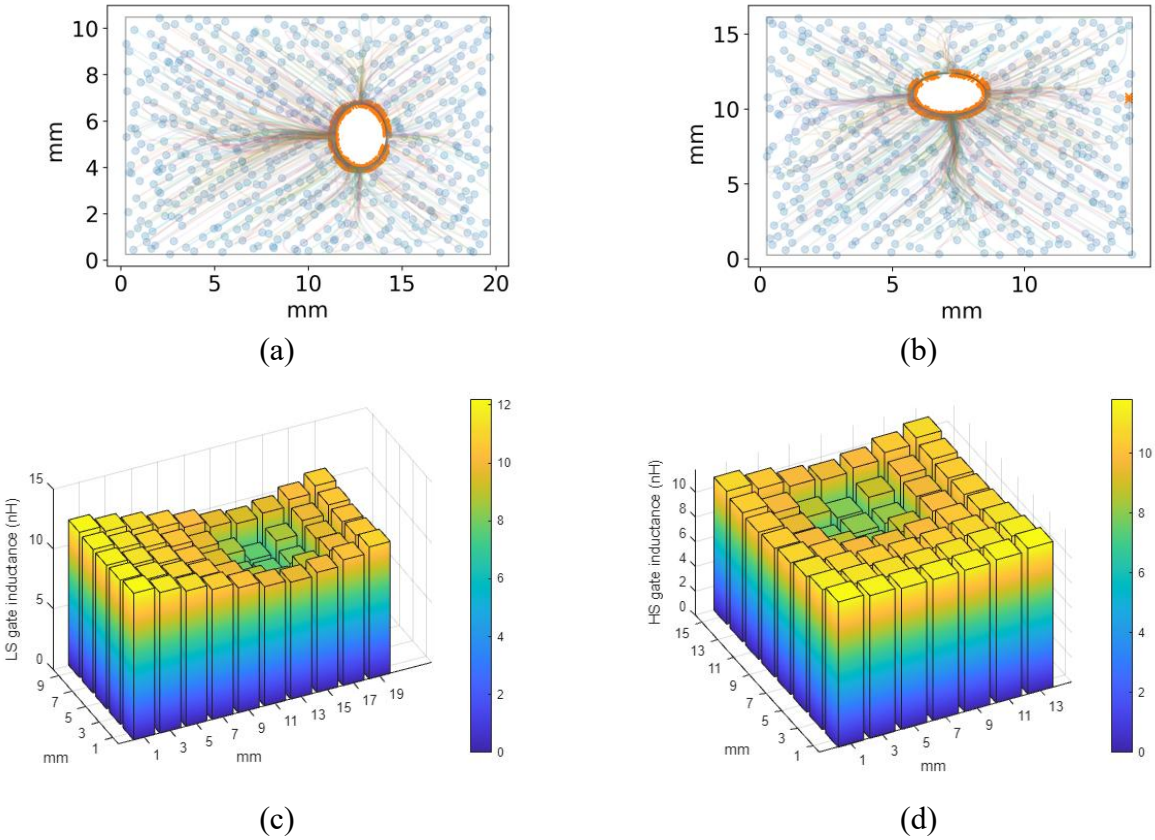


Fig. 4. Optimization of gate inductance based on our surrogate model for (a) LS and (b) HS, and the validation in Ansys Q3D parametric simulation for (c) LS and (d) HS.

Single-Objective Optimization for Mutual Coupling. The goal of this test case is to minimize the maximum mutual coupling among chips on each side, which is reflected in the objective function as $\bar{f} \in \{\overline{M^{HS}}, \overline{M^{LS}}\}$. Similar to the first test case, the optimization process creates a population of random initial configurations as starting point (see blue circles in Fig. 5a-b), and the optimal locations (see red cross marks in Fig. 5a-b) are determined by the gradient-based optimization. The red traces indicate the trajectories of auxiliary emitter terminal during the optimization process. The determination of the optimal position of auxiliary emitter terminals depends on the analysis of both the trajectories and final positions in red cross marks. For the LS, the trajectories for most samples end up in the bottom left corner. Some trajectories end up around the gate terminal because the penalty of violation of geometry constraints is set so high that the optimization process is terminated. Thus, the optimal position to obtain negative mutual coupling is in the bottom left corner on the topside interconnects for the LS auxiliary emitter terminal. Similarly, for the HS auxiliary emitter terminal, the optimal designs are clustered in the bottom right corner. Some trajectories terminated around the gate terminal due to the same reason as LS. To validate the surrogate model results, the parametric simulations in Ansys are performed as shown in Fig. 5 c-f, and the distribution of the mutual inductance validates the conclusion of the surrogate model.

Multi-Objective Optimization. The proposed optimization framework allows multiple objectives to be included in the objective function. Optimization of various EM features can occur concurrently by assigning specific weights to each feature within the objective function. By increasing the number of combinations of weights, a large design space can be efficiently treated by the surrogate model.

For instance, Fig. 6 shows the Pareto plot of the optimization for two EM features: gate inductance and mutual coupling (i.e., in the objective function, $\bar{f}_i \in \{\overline{L_g^{HS}}, \overline{L_g^{LS}}, \overline{M^{HS}}, \overline{M^{LS}}\}$), where distinct weight combinations are represented by differently colored circles. To demonstrate the breadth of the design

space, previous designs from Ansys parametric simulations, which consider only two parameters (location of auxiliary emitter terminal (x, y)), are shown together with the results of the surrogate model. The designs in Ansys are depicted as black cross, and the black curve indicates the Pareto front derived from these designs. By exploring a larger design space, the surrogate model-based optimization achieves an expanded Pareto front (blue cure in Fig. 6), indicating a better electromagnetic performance can be achieved compared to the expert-based iterative design.

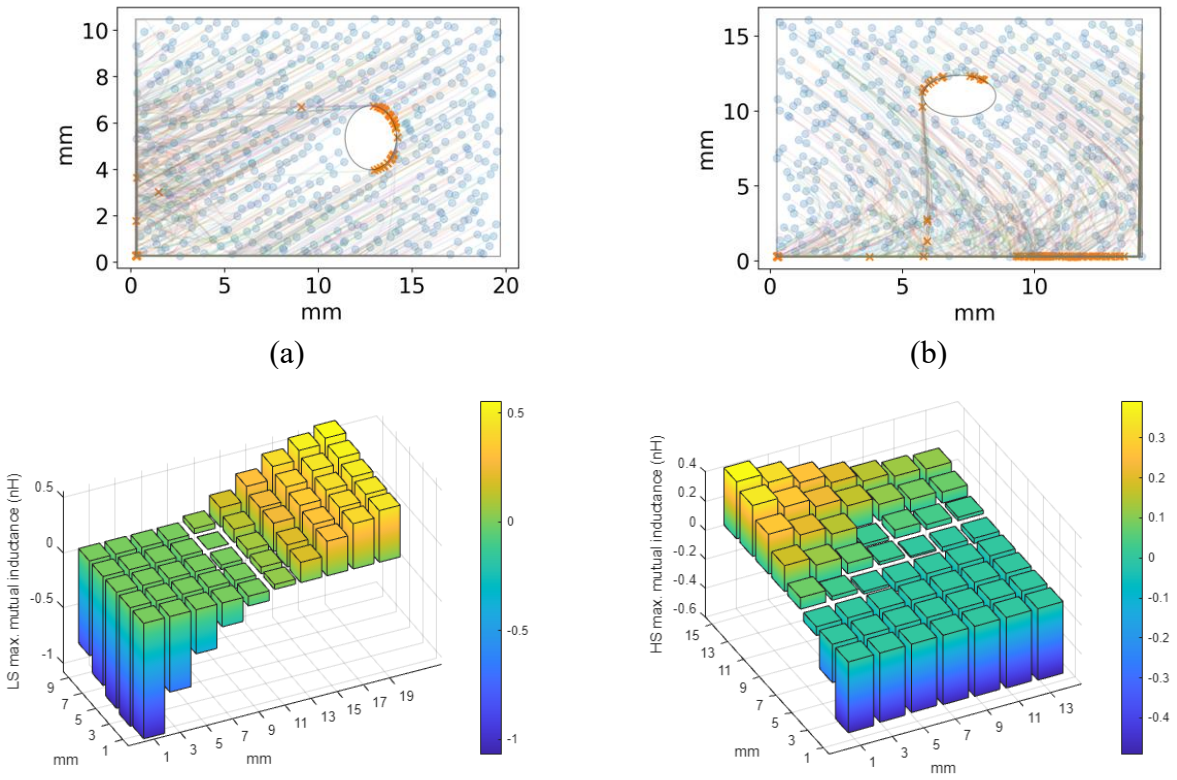


Fig. 5. Optimization of mutual coupling based on our surrogate model for (a) LS and (b) HS, and the validation in Ansys Q3D parametric simulation for (c) LS and (d) HS.

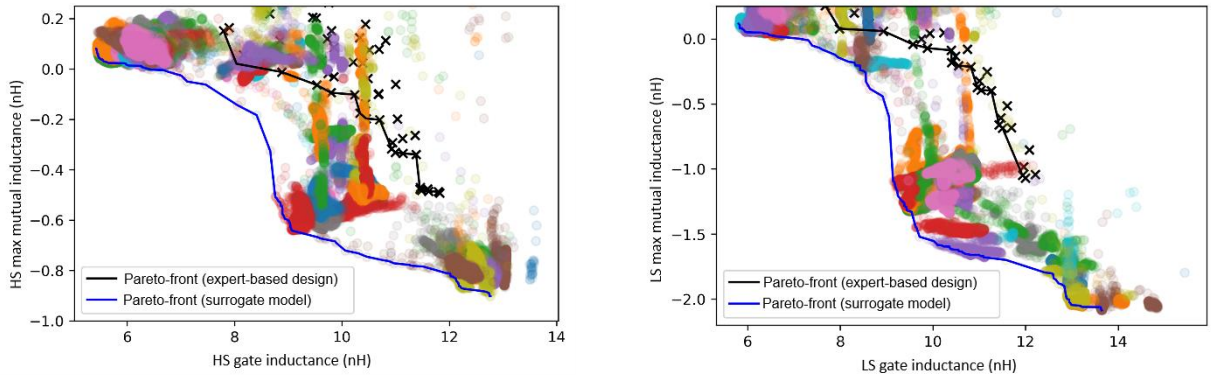


Fig. 6. Pareto-front of two objectives (gate inductance L_g, maximum mutual coupling inductance M) for (a) HS chips and (b) LS chips, obtained by the surrogate-model based optimization and expert-based parametric simulation in Ansys.

Summary

A robust surrogate model-based optimization methodology is introduced for exploring power module design spaces, with the capability to optimize multiple objectives involving several design variables simultaneously. The surrogate model reduces computational cost by 99.99983% compared to full

electromagnetic simulations, cutting computation time from 30 minutes to just 3 milliseconds. The prediction errors between the surrogate model and reference simulation results are below 0.01%. This approach provides a fast optimization of key electromagnetic parameters and enables efficient evaluation of large design spaces, which is challenging and time-consuming to discover through knowledge-based approaches alone. Future work will focus on increasing design variables, extending the surrogate model to capture frequency-dependent effects and dynamic switching transient. Additionally, active learning strategies could be studied to reduce the number of high-fidelity simulations and transfer learning could be applied to adapt the surrogate across different module topologies.

References

[1] C. Chen, F. Luo and Y. Kang, A review of SiC power module packaging: Layout, material system and integration, CPSS T. Power Elec. Appl. 02, 03 (2017).

[2] V. Parque, A. Nakamura, T. Miyashita, A study on the inductance and thermal regression and optimization for automatic layout design of power modules, IEEE CPMT Sympo. Jap. 117, 120, (2023).

[3] Y. Abe, A. Hirao, R. Kato, Y. Ikeda, V. Parque, M.K. Faiz, M. Yoshida, T. Miyashita, Design optimization of copper patterns and location of power semiconductors and terminals, Int. Conf. Electron. Pkg (ICEP). 157, 158, (2021).

[4] Y. Zhou, Y. Jin, Y. Chen, H. Luo, W. Li, X. He, Graph-model-based generative layout optimization for heterogeneous SiC multichip power modules with reduced and balanced parasitic inductance, IEEE Trans. Power Electron. 37, 8, (2022).

[5] T.M. Evans, Q. Le, S. Mukherjee, I.A. Razi, T. Vrotsos, Y. Peng, H.A. Mantooth, PowerSynth: A power module layout generation tool, IEEE Trans. Power Electron. 34, 6, (2019).

[6] ANSYS, ANSYS Electronics Desktop 2024 R2, (2024).