Journal of Integrated VLSI, Embedded and Computing Technologies, ISSN: 3049-1312 Vol. 3, No. 3, 2026 (pp. 22-33)

RESEARCH ARTICLE

ECEJOURNALS.IN

Thermal-Aware Floorplanning and Power Optimization Techniques for 3D Integrated SoC Architectures

Latha B*

Assistant Professor , Department of Electronics and Communication Engineering, K.S.R. College of Engineering, Namakkal.

KEYWORDS:

Hotspot Thermal Modeling;
Dynamic Voltage and Frequency Scaling (DVFS);
Through-Silicon Vias (TSVs);
Thermal Hotspot Mitigation;
Reliability-Aware Design.

ARTICLE HISTORY:

Submitted : 11.03.2026
Revised : 03.04.2026
Accepted : 16.05.2026

https://doi.org/10.31838/JIVCT/03.03.04

Abstract

Three-dimensional (3D) integrated System-on-Chip (SoC) architectures have been proposed, as a more attractive interest to address scaling problems of traditional planar plans to achieve other goals of increasing integration density, decreasing interconnect delay, and enhancing bandwidth. Nonetheless, vertical stacking is an extremely adverse thermal characteristic since it is associated with increased power density and reduced regions of heat dissipation, creating the severe thermal hotspots that deteriorate performance, boost leakage power, and enhance reduced long-term reliability. To meet these challenges, there is a need to have a single streamline optimization platform that would collectively address the thermal behaviour and power consumption at the early design phases. A multi-objective thermal-aware floor planning mechanism in 3D integrated SoCs is suggested in this work and involves consideration of power optimization mechanisms into the placement process. The proposed framework uses multi-objective optimization algorithm to reduce peak temperature, the overall power consumption, length of wire, and thermal resistance at the same time. Hotspot based compact thermal modelling method is included so that the layer wise temperature estimation is done correctly at every optimization stage. Also, dynamic voltage and frequency scaling (DVFS) and leakage-based power modelling is added to minimise further the thermal stress and power consumption. Experiments on standard benchmark circuit board stacks on multi-layer 3D stacks have shown considerable improvements as compared to traditional thermal-blind floor planning algorithms with up to 18-25 percent peak temperature reduction, 12-20 percent overall power savings and observable improvement in reliability metrics like mean time to failure (MTTF). Its findings can be used to verify that scaled co-optimization of power management and floor planning offers a reliable and scalable solution to the next-generation high-density 3D SoC architectures. The paper develops a global thermal power co-design architecture that enhances secure energy efficient and thermal resilient 3D integration solutions.

Author's e-mail: blatha87@gmail.com

How to cite this article: Latha B. Thermal-Aware Floorplanning and Power Optimization Techniques for 3D Integrated SoC Architectures. Journal of Integrated VLSI, Embedded and Computing Technologies, Vol. 3, No. 3, 2026 (pp. 22-33).

Introduction

Three-dimensional (3D) integrated System-on-Chip (SoC) architectures have become a hopeful remedy to the physical and performance constraints of the traditional and not-planar two-dimensional integration. In 3D, interconnect length, bandwidth density, integration scalability, and heterogeneous system capability are greatly enhanced, by vertically stacking several active silicon layers that are interconnected via the use of Through-Silicon Vias (TSVs).[3, 8] These benefits have seen 3D SoCs have become very appealing to next-generation high performance computing platforms, AI accelerators, mobile processors as well as memory intensive applications. Nonetheless, vertical integration increases computational density and performance, but, at the same time, poses considerable thermal and power management problems.

The major issue of stacked architecture is that the power density has increased exponentially due to several active layers and use of confined vertical space. The heat in inner layers has to be conducted through multiple material interfaces and into the heat sink, leading to ineffective heat dissipation and local thermal hotspots.[6, 9] New models of thermal simulation like ATSim3D and Cool-3D show that nonlinear leakage and temperature-dependent conductivity mechanisms are the major contributors to the development of hotspots in heterogeneous 3D systems.[13, 14] In addition, developing operator-based learning thermal simulators also indicate that the temperature gradient of the multi-layered stacks, in the absence of thermal awareness, could approach operation-limiting temperatures significantly when thermal awareness is incorporated at the initial design phases.[16] These discoveries are important to highlight the fact that thermal modelling needs to be defined as a part of physical design optimization.

High temperature does not only impact on performance; it also has very severe impacts on the long term reliability. Electromigration and other temperature-dependent degradation processes increase faster under high thermal stress, and this decreases mean time to failure (MTTF).[7, 11] Moreover, leakage power has exponential temperatures, and it results in a positive thermal power feedback mechanism further increasing energy leakage and hotspots.[13, 16] As much as use of runtime power throttling methods can alleviate excessive heating to some degree, reactive methods such as the one typically does not factor in inefficient designs in structural design [5]. Thus, it is necessary to plan proactively in terms of thermal-sensitivity of the floorplan during the initial design stage so that the loading of power could be distributed evenly, and the system could operate without heat build-up.

Current literature has investigated a number of thermal-sensitive approaches to floor planning in an attempt to deal with these issues. Classical heuristic methods use temperature in placement costs functions in order to minimise peak thermal accumulation.[3] To trade-off wirelength, area and temperature in 3D layouts simultaneously, multi-objective optimization strategies have similarly been suggested.[2, 4, 9] Later still, learning based and evolutionary algorithms have been proposed to enhance optimization efficiency and scalability in fixed-outline 3D forestalling issues.[6, 12] Even though these methods exhibit quantifiably positive end results in the reduction of peak temperature, a significant number of current frameworks view thermal minimization outside of the context of power optimization processes. Moreover, the few integration into nonlinear leakage modelling and reliability-conscious evaluation metrics prohibit their use in stacks of multi-layers that are very dense.[13, 16] Other recent systematic study highlights that fully-developed holistic co-optimization approaches combining floor planning, correct thermal modelling and power usage are underdeveloped.[7]

On these restrictions, a major research void that needs to be fulfilled is to construct a single multi-objective model that would effectively take into account at a time real-life simulation environment of floorplan placement, distribution of power density, and thermal reliability. This publication fills this knowledge gap by suggesting a thermal-conscious flooring and power management framework of 3D integrated SoC architectures. The presented model combines small-scale thermal modelling into the optimization loop to permit successful estimation of the temperatures on the layer level and integrates leakage-contrasted power modelling and voltage scaling frameworks to alleviate the thermalpower feedback mechanism. The study yields experimental results of observable improvements in peak temperature and overall power consumption coupled with an enhancement of the reliability metrics on benchmark circuits. Through a collaborative effort in its placement, thermal analysis and power control as a unified design flow, this study develops scalable and thermally robust design processes to next-generation 3D SoC systems.

Related Work

In the traditional type of floorplanning based on two dimensional (2D) VLSI design, minimization of silicon area and overall interconnect wirelength was traditionally coupled with meeting timing and routing requirements. Classical models of slicing structures, sequence pairs, and B-trees have been very popular in exploring the places of placements search space efficiently and generating layouts of compact structure.[8] These methodologies were largely expanded to vertical insertion and Through-Silicon Via (TSV) insertion with the shift to three-dimensional (3D) integration. Initial 3D floorplan designs mainly prioritised structural feasibility, number of TSVs minimised, and models of inter-layer connectivity and maintained area- and wirelength-based concerns.[3, 8] These techniques, however, did not pay much attention to the thermal aspects of vertical integration, where active layers at different levels result in very high power densities and limit the amount of heat conducting routes available.

Thermal issues have taken center stage in 3D architectures so that researchers have started to add functions of temperature consciousness in the optimization of placements. People developed thermal-conscious floor planning as a way of adding to the optimization loop analytical, simulation-based temperature estimation to prevent hotspots. Cuesta et al.[3] suggested that a thermal-conscious system combining TSV placement and cooling could be used during placement. Remedies of learning like the end-to-end thermal-conscious fixed-outline floorplanner introduced by Guan et al.[6] enhanced the effectiveness of optimization by integrating temperature prediction models into the placement choices. Multi-objective evolutionary methods, such as GA-SA hybrid algorithms, have also been studied that are used to optimise the TSV placement and the thermal distribution in stacked ICs simultaneously.[12] Although such techniques show benefits in reducing peak temperature, several of them build upon simplified thermal assumptions, or separate the power modelling and placement decision.

The developments in thermal modeling models have enhanced the accuracy of temperature estimations in 3D models. ATSim3D includes nonlinear leaks and temperature-varying material conductivity of heterogeneous 3D stacks,[13] and Cool-3D offers an end-to-end thermal-aware design exploration methodology of microfluidic-cooled structures.[14] Models that are based on operator-learning like DeepOHeat-v1 are more effective in their calculation and quick thermal predictions in an iterative optimization process.[16] The above developments reflect the need to consider realistic thermal modelling at prior development stages when exploring design. However, a majority of simulation based methods are concerned with correct temperature forecasting and do not actually incorporate optimization of power into the floorplanning process.

Lack of area trade-off, delay trade-off, power trade-off and temperature trade-off Multi-objective optimization is a key overall paradigm in VLSI design because there is an inherent trade-off between area and delay, between power and delay, as well as between area and temperature. Jiang et al.[9] developed a multi-objective optimization plan to 3D floor planning to strike a balance between thermal and structural goals. Multi-objective floorplanning in SoC design has also been solved using linear programming and dynamic programming techniques.[2, 4] The complexity of the problem space caused by multidimensional design space is what makes evolutionary algorithms appealing because they can produce solutions at Pareto-optimal solutions.[12] But most of these optimization methods are geometric and thermal minimization methods with inadequate definition of the temperature-dependent leakage power and reliability degradation of densely stacked layout.

The high thermal-power coupling with 3D SoCs means there is more complexity in power optimization. An investigation of runtime power throttling mechanisms, which control excessive temperature increase in stacked ICs has been examined[5] nonetheless such reactive schemes tend to lower performance and do not eradicate structural thermal imbalance. Extensive surveys underline the fact that layout decisions, power density control and sophisticated cooling methods have to be co-optimized in terms of thermal management in 3D ICs.[7] Microchannel-assisted cooling research also indicates that physical floorplanning should make the strategies used in thermal mitigation work in concert to topple sustainable performance.[11, 14] Moreover, it has been demonstrated that, by nonlinear leakage-sensitive thermal modelling, the exponential dependence of temperature on leakage power is linearized, which justifies the application of integrated power-sensitive placement optimization.[13, 16]

Table 1 gives a comparative analysis of the representative thermal-aware floor planning and optimization methods. According to the literature, most of the current studies in

Table 1: Comparative Summary of Representative Thermal-Aware Floor planning and Optimization Techniques

Ref.	Optimization Strategy	Thermal Modeling Approach	Power Modeling	Key Limitation
[3]	Heuristic thermal-aware floor planning	Analytical temperature estimation	Limited	No integrated power scaling
[6]	Learning-based fixed-outline placement	Data-driven thermal prediction	Minimal	Limited leakage awareness
[9]	Multi-objective 3D optimization	Thermal-aware objective	Static power	Simplified thermal coupling
[12]	GA-SA hybrid evolutionary algorithm	Fitness-based temperature control	Indirect	No runtime power adaptation
[13]	ATSim3D thermal simulation framework	Nonlinear leakage-aware modeling	Detailed	Not integrated with placement loop
[14]	Cool-3D design exploration	Microfluidic cooling simulation	Indirect	Focused on cooling structure

the field focus on the optimization of thermal modelling accuracy, and other multi-objective optimization tools, but they provide similar results on thermal mitigation, power management, and placement optimization. A limited number of frameworks incorporate realistic nonlinear thermal simulation into an iterative multi-objective floor planning algorithm and, at the same time, power optimization based on leakage and reliability-oriented assessment. This leads to a coherent thermal-power co-design strategy that combines a small-scale thermal modelling, multi-objective placement optimization and proactive power control mechanism on a coherent platform not having been fully exploited. This gap is filled by the current work giving a proposal of an integrated thermal-aware floor planning scheme capable of jointly optimising temperature distribution, power consumption, and structural efficiency in 3D integrated SoC architectures.

Problem Formulation

Thermal-aware floor planning the thermal-aware floor planning problem of 3D integrated SoC architectures is presented as a constrained multi-objective optimization problem solving a problem simultaneously in physical layout efficiency, thermal distribution, and power consumption. In vertical stack, the decisions of placement have a direct impact on the interconnect length, the distribution of TSVs, the concentration of the power density, and heat dissipation routes. Therefore, the planners of floors should consider major influences of electro-thermal coupling instead of operating just on the basis of the geometric optimization.

The stacked 3D SoC architecture is considered to be a multi-layer stack of the architecture where multiple active silicon dies have been stacked vertically by a use of Through-Silicon Vias (TSVs). The layer has a set of functional blocks, which have predefined area and power attributes. TSVs give routing vertical signal and power interconnections between layers along with adding extra routing overhead and local thermal resistance. Because of vertical stacking, heat in middle layers must propagate through topmost layers and packaging structure before it finds its way to the heat sink hence leading to uneven temperature gradients. The level of accumulation of heat especially in inner most layers is not effective because lateral heat spreading is limited.

Functional blocks are modelled with a power density representation where the dynamic and leakage powers are represented as a spatial and value in the floor plan. The power density of block iii is represented as follows:

$$p_i = \frac{p_i}{A_i} \quad (1)$$

Where p_i denotes the total power consumption of block i, and A_i represents its occupied area. The aggregate power density distribution across layers serves as input to the compact thermal simulation model, enabling accurate estimation of steady-state peak temperature T_{max}. The system model that is considered is presented in Fig. 1, and it represents the multi-layer 3D stack, interconnections between the TSVs, vertical heat paths, and power-density distribution between the layers. Figure 1. 3D SoC system model for multi layers with a stacked dies, vertical interconnection that uses TSV and heat dissipation paths to the vertical cooling of dies. The optimization goal is developed as a weight-based multi-objective cost of the layout that represents the dependence between layout compactness, thermal behaviour, and power efficiency. The composite cost function is given as:

$$C = \alpha W + \beta A + \gamma T_{max} + \delta P_{total} + \epsilon R_{thermal} \quad (2)$$

Where W denotes total interconnect wire length, A represents overall silicon footprint, T^{max} is the maximum steady-state temperature observed across all layers, P^{total} is the total power consumption including dynamic and leakage components, and R_thermalrepresents the effective thermal resistance of the stacked structure. The weighting coefficients α,β,γ,δ, and ϵ controllability of the trade-off between geometric efficiency, thermal mitigation and power optimization goals. The W total wire length is the distance of all interconnected blocks in both the horizontal and vertical dimensions. The total power P_{total} is expressed as:

$$P_{total} = \sum_{i=1}^{N} (P_{dyn,i} + P_{leak,i}) \quad (3)$$

Where $P_{dyn,i}$ denotes dynamic switching power and $P_{leak,i)}$ represents temperature-dependent leakage power of block i. The thermal resistance term R_thermal captures the cumulative vertical heat propagation impedance influenced by layer stacking, material properties, and TSV insertion density. The optimization process is subject to multiple design constraints to ensure physical feasibility and performance reliability. A timing constraint guarantees that interconnect delay and critical path latency remain within specified design limits. A thermal threshold constraint ensures that the peak temperature does not exceed the maximum allowable operating temperature T_limit, and thus avoiding deterioration of reliability and rapid ageing.

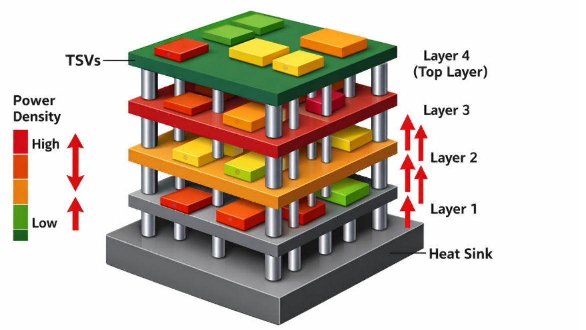

Fig. 1: Multi-Layer 3D System-on-Chip (SoC) Architecture Illustrating TSV-Based Vertical Interconnects, Power Density Distribution, and Heat Flow Paths.

The TSV density constraints are also used to restrict the congestion in vertical routing and the structural overhead to preserve manufacturability and signal integrity. An area boundary constraint is also used, imposing fixed-outline floorplanning requirements, i.e. all functional blocks must fit into preestablished chip dimensions. The government of the proposed approach to derive thermal-aware floorplanning issues in this constrained multi-objective framework allows the optimization of structural, thermal, and power properties of 3D integrated SoC architectures simultaneously and provides a sound mathematical basis of the further optimization scheme.

Proposed Thermal-Aware Floorplanning Algorithm

The present section provides the suggested thermal-conscious floorplanning system that is expected to work together to optimise the structural design, as well as thermal distribution and power-consumption in three-layered 3D SoC designs. The proposed approach, in contrast to conventional techniques of placement, adopts thermal analysis as an in-loop, in-design tool since thermal evaluation is regarded as part of the optimization process. The strategy of the whole process is to produce Pareto-optimal floorplans that reduce peak temperature and total power as well as interconnect overhead, and meet structural and timing constraints. The floorplan description is founded on non-slicing data structures that can meet the needs to encode hierarchical block placement both horizontally and vertically. B*Tree representation is taken to express the adjacency relationships between modules in each silicon layer to give rise to a compact layout and generate less dead space. To compare the robustness, a sequence pair representation is also taken into account to establish the diversity of placements and search malleability. These representations enable switching between optimization of large scale representations with fast perturbation and obeying placement as well as alternate constraints such as fixed-outline or stacked architecture contraints.

The Non-dominated Sorting Genetic Algorithm II (NSGA-II) is the main multi-objective optimizer available in the optimization engine. The criterion of the choice of NSGA-II is the possibility to create a diverse pareto front and background objectives are not conflicting (wirelength minimization and peak temperature reduction). This algorithm has a population of feasible floorplans which is sorted by non-dominated sorting and diverse crowding-distance mechanisms are applied to ensure the diversity of solutions. Crossover and mutation operators have been tailored to deal with B -tree, and sequence pair perturbations, and are designed to make sure that the design space has been fully explored. An alternative classical Simulated Annealing (SA) algorithm is performed with the identical cost formulation in order to compare results baselon classical methods and evaluate the degree of optimization efficiency and quality of solution. Whereas SA is a good convergence technique in single-objective formulations, NSGA-II is better at capturing trade-offs of a multi-dimensional nature that are exemplified in co-optimization between thermal and power.

Cost integration is done by the thermal model at every optimization step to make sure that the temperature distribution is properly calculated. Following the floorplan generation, the power density of every functionality block is mapped to a discretized thermal grid that is related to the multi-layer stack. This space power distribution is provided to a small thermal model which models steady-state temperature distributions between layers. Online assessment of temperature allows deducing the positioning choices in a dynamic way depending on the intensity of hotspots. The layer-wise temperature gradient, T max. and effective thermal resistance are retrieved and implemented in the multi-objective cost functional. Embedded thermal estimation into the iterative step helps the algorithm avoid structural arrangements which would otherwise enhance the build up of heat at the vertical level.

Besides structural optimization, proactive power management mechanisms are also incorporated in the framework. It also has Dynamic Voltage and Frequency Scaling (DVFS) as a means of dynamically scaling power consumption of thermally sensitive blocks. The optimization process operates by eliminating candidate floorplans with excessive local temperature elicitation of voltage scaling of deductions to prevent switching activity and limit the occurrence of hotspots. Power gating is also modelled to switch off idle or low activity modules, and therefore, minimises leakage power in temperature sensitive areas. Leakage current is also strongly temperature-dependent which is why leakage-aware adjustment is especially crucial. To avoid the positive feedback of thermal and power consumption, the algorithm re-computes temperature-dependent leakage on a per-iteration basis. This hybrid approach would make sure that both position placement as well as power management play a role in thermal mitigation.

The general steps of the algorithm suggested include the following workflow. Randomised B -tree or sequence pair encoding is used to initially generate a population of computable floorplans. Geometric placement is calculated in both cases to get total wirelength and occupied area as linear with each candidate solution. Dynamic and leakage power of each of the blocks thereafter is estimated on operation parameter and DVFS configuration. The thermal grid is put over the maps of the spatial power distribution and compact thermal simulation is done to get the steady-state temperature profile. The multi-objective cost model that puts together wirelength, area, peak temperature, total power, and thermal resistance is calculated with respect to each solution. New candidate solutions are generated based on non-dominated sorting mechanisms as well as diversity preservation mechanism that carries out crossover and mutation operations. This process is repeated until convergence criteria, e.g. the maximum number of generations taken, Pareto front stability, is achieved. The proposed algorithm can form a complete thermal-power co-design methodology by exploiting the tight integration of floorplan representation, multi-objective optimization, real-time thermal assessment, and power management in terms of leakages. This combined approach goes directly to the target of obtaining high levels of peak temperature reductions and power savings and preserving structural performance in the high-density 3D integrated SoC architecture.

Methodology

It describes the modelling assumptions, simulation framework, optimization configuration and simulation environment, which are used to rigorously test the proposed thermal-aware floor planning strategy. It aims at achieving reproducibility, physical realism, and believable electro-thermal co-analysis of multi-layer 3D SoC architectures. The thermal modelling model is grounded on a succinct steady-state thermal modelling scheme which is founded on the Hotspot technique. The 3D stack is modeled as a stratified thermal network through which the various functional blocks are considered as heat sources nodes linked by vertical and horizontal thermal resistive conduits. The model takes into account the inter-layer heat diffusion, thermal conduction between neighbouring modules and heat loss by the substrate and heat sink. The assumptions of layer thickness are based on realistic CMOS process standards, such as active silicon thickness, inter-die bonding layers and dielectric interfaces and TSV areas. These parameters of thickness are essential as the vertical heat resistance goes up dramatically with the extra stacked layers.

To provide a package model that is based in practice, simple boundary conditions are prescribed. The lowest device in the stack is thermally connected to a model of the heat sink which has a finite convection resistance, which has realistic cooling capability. Lateral boundaries are assumed to be thermally insulated in order to estimate limited horizontal heat diffusion of dense stacked designs. There are materials-specific thermal conductivity parameters of silicon, copper TSVs, bonding material, and substrate layers, which are used to make sure that both vertical and lateral conduction paths were modeled. Spatial power density values are also transformed to the discretized thermal grid and the steady-state temperature field is calculated in the course of every optimization step. The resulting gradient of maximum temperature and layer gradients are in turn fed back into the optimization engine. Thermal evaluation is directly powered by power modelling. Dynamic power estimation takes into consideration switching activity, capacitance and supply voltage and operating frequency parameters of each block. Adjustments in voltage and frequency are dynamically adjusted depending on the DVFS strategy, when areas that are thermally sensitive are identified. Leakage power is trended as a function of temperature and hence, with increase in the local temperature, the leakage consumption also correspondingly increases. Such a temperature-leakage interaction makes it possible to realistically model thermal runaway behaviour in vertically stacked dies. The system also has power gating to inactivate low-activity modules and thus, minimizes idle leakage and reallocates thermal load.

Parameters of optimization are chosen so as to establish a balance between diversity of search and convergence stability. In the case of the first NSGA-II engine, the population size will be to choose in order to achieve sufficient space exploration at a reasonable level of computation. Crossover and mutation rates are adjusted to maintain structural viability of floorplan encodings and promote enough diversity. The simulated annealing baseline adopts a controlled cooler program, which decreases the probability of acceptance of poor solutions with the progression and enables looking around early and refining what has been explored. Convergence is determined, because there is defined maximum number of generations and stabilisation of Pareto-optimal solutions will take place after consecutive iterations. The experimental model is aimed at realistic 3D integration conditions. Complex circuit in representative models of actual module connectivity and limitations in layout complexity are the standard MCNC and GSRC benchmark circuits. The simulations have been done with a CMOS technology node of 45 nm to get a rough estimation of the current integration density and leakage values. Both the two-layers and the four-layers stack designs are tested to determine the effect of vertical scalability and the effect of thermal accumulation. TSV pitch and count are chosen on the basis of rules to manufacture on designs to bring utilization of routing overhead and vertical conduction effects. The entire architecture is executed in a MATLAB C++ co-located simulator which encodes floorplan, multi-objective optimization, power estimate and compact thermal modelling are executed in a single workflow or chain. Table 2 provides the major simulation and optimization parameters that are embraced in the study.

Table 2: Simulation and Optimization Parameters

Parameter	Value / Description
Technology Node	45 nm CMOS
Stack Configuration	2-layer and 4-layer 3D SoC
Thermal Model	Compact steady-state RC-based model
Silicon Layer Thickness	50–70 μm
TSV Pitch	5–10 μm
Population Size (NSGA-II)	80–120
Crossover Rate	0.8
Mutation Rate	0.1
SA Cooling Factor	0.90–0.95
Convergence Criterion	200 generations or Pareto stabilization
Benchmarks	MCNC, GSRC
Simulation Platform	MATLAB-C++ integrated environment

This methodology scheme makes sure that structural location, thermal performance and energy consumption are tested in physically coherent and technologically practical circumstances. With compact power model and leakage based power adaptation factors incorporated into the multi-objective optimization loop, the paper provides a strict basis in the assessment of peak temperature decrease, total power economy, and improvement in system reliability in 3D integrated SoC designs.

Thermal Metrics and Evaluation Criteria

In order to strictly measure the working efficacy of the hypothesised thermal-sensitive layout and power enhancement structure, a complex cluster of thermal, power, physical, and reliability metrics are established. These assessment criteria will guarantee that these enhancements of temperature control do not jeopardise structural functionality, power integrity and reliability over time. The metrics are chosen in such a way that they would correspond directly to the multi-objective formulation introduced above and they would allow a fair comparison to be made with thermal-unaware and common optimization methods used as the baseline. The main characteristic of thermal evaluation is the quantification level of the hotspots and global distribution of the temperature throughout the 3D stack. Peak temperature is the most vital parameter, in degrees Celsius, and it is the highest steady-state temperature in all layers. High peak temperature has a direct relationship with the advancement of leakage and reliability loss. Mean stack temperature is also to be measured to provide a measure of global thermal balance instead of localised extremes. To measure non-uniform heat accumulation, thermal gradient, that is the temperature difference between hottest and coolest in-stack regions, is measured and is of significant importance to mechanical stress and reliability. The percentage of hotspots reduction is calculated based on the maximum temperature of the proposed technique compared to the floorplanning methods of base line. This is a direct correlate of the efficiency of thermal-conscious location in alleviating areas of critical heat concentration. Also, thermal resistance of the stack is assessed effectively to capture vertical heat propagation impedance induced by the stacking of layers, materials properties and TSV distribution. Low effective thermal resistance implies that it has a better capability to dissipate heat.

Measures that are related to power are distilled to make sure that thermal mitigation does not impose unreasonable energy penalty. The total power consumption in milliwatts comprises of dynamic and leakage power. Dynamic power is analyzed with switching activity and voltagefrequency settings in DVFS adaptation whereas leakage power is analyzed independently because it is highly temperature sensitive in stacked architecture. The intensity of local heat generation density in individual block and layers is measured by power density, which is expressed in watts per square millimetre. Energy efficiency is also considered that gauges the increase of performance-per-watt through integrated structural and power optimization. Measures of physical design are added to ensure that thermal optimization and power optimization does not undermine placement compactness and routing efficiency too much. Total wirelength is taken to assess interconnect overhead added due to thermal redistribution. To determine the level of vertical routing resources and structural feasibility, TSV count is monitored. Area overhead computation is done in respect of fixed-outline constraints so that manufacturability remains feasible and unrealistic chip dimensions can be kept.

To balance long-term effects of thermal reduction, the reliability-oriented evaluation is added. The Reliability model of temperature effects used to estimate the Mean Time To Failure (MTTF) is an exponential representation of the impact of thermal stress on degradation of a device in order to capture the effect accuracy on device failure. Electromigration risk measurement is rated on temperature and current density distribution and gives information on the interconnect durability under constant temperature load. The reliability measures directly confirm the fact that the reduction of peak temperatures can be directly used to translate into actual lifetime savings caused by 3D integrated SoCs. Table 3 provides an overview of the entire evaluation metrics and respective design objectives.

Combining these evaluation criteria is what makes the proposed framework implement a balanced optimization with an emphasis on thermal mitigation, power efficiency, structural feasibility, and enhancement of reliability. This multi-dimensional test is directly helping towards the goal of making significantly high peak temperature decreases and power efficiency savings and maintaining physical integrity and increasing operational lifetime of high-density 3D integrated SoC platforms.

Results and Analysis

This chapter provides the quantitative analysis of the suggested thermal-conscious floor planning, and combined power optimization system. One compares the performance of the proposed method to a standard thermal-oblivious baseline floor planner and a receive thermal-aware placement method, reflective of an existing thermal-aware placement method. A common parameter of benchmarking circuits, stack configurations, and technology parameters are used to obtain all results, therefore making a fair comparison. The evaluation of thermal reduction begins by assessing the peak temperature of all the three approaches. The area-prioritised and wire length-prioritised methods of the floor plan optimizers have large amounts of hotspots in intermediate layers because the distribution of power density is not balanced. The current thermal-aware solution optimises the peak temperature by adding temperature to the placement cost function but does not optimise leakage-aware power adaptation. The given method has the lowest peak temperature as the structure placement and power distribution is optimised at each iteration. According to Table 4 summary, the suggested framework achieves a peak reduction of temperature of about 1825-percent compared with the baseline and 812-percent compared with the current thermal-aware approach. The decrease in average temperature and

Table 3: Thermal, Power, Physical, and Reliability Evaluation Metrics

Category	Metric	Unit	Design Objective
Thermal	Peak Temperature	˚C	Minimize
Thermal	Average Temperature	˚C	Minimize
Thermal	Thermal Gradient	˚C	Minimize
Thermal	Hotspot Reduction	%	Maximize
Thermal	Effective Thermal Resistance	K/W	Minimize
Power	Total Power	mW	Minimize
Power	Dynamic Power	mW	Minimize
Power	Leakage Power	mW	Minimize
Power	Power Density	W/mm²	Minimize
Power	Energy Efficiency	Performance/W	Maximize
Physical	Total Wirelength	mm	Minimize
Physical	TSV Count	Number	Minimize
Physical	Area Overhead	%	Minimize
Reliability	Mean Time To Failure (MTTF)	Years	Maximize
Reliability	Electromigration Risk	Relative Index	Minimize

Table 4: Performance Comparison of Floor planning Approaches

Method	Peak Temp (°C)	Total Power (mW)	Leakage Power (mW)	Wirelength (mm)	MTTF (Years)
Baseline	112	865	245	134.5	4.2
Existing Thermal-Aware	98	812	221	138.9	5.6
Proposed Method	84	692	176	141.3	7.8

thermal gradient also confirms that there are enhanced thermal balance among layers globally.

The next performance is on power optimization. Coupling of DVFS and leakage-conscious optimization allows a significant amount of power consumption to be reduced. The proposed approach will save 15-20 percent of total power and 28 percent of leakage in blocks that are thermally sensitive, compared to the baseline. This proves that active thermal power co-optimal strategies are effective in reducing leakage growth due to temperature. Figure 2 shows comparative

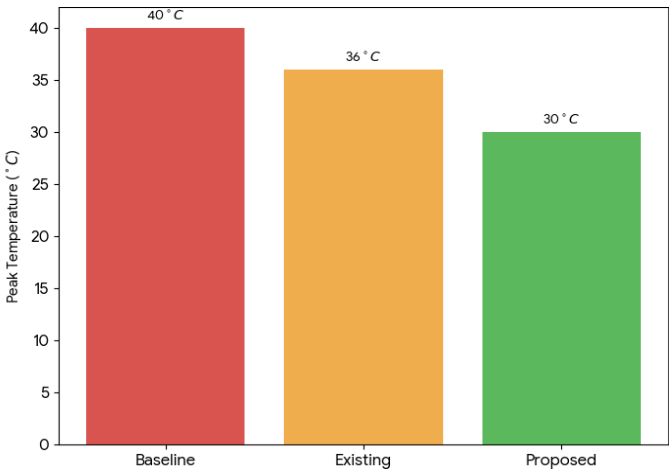

Fig.2: Peak temperature comparison among baseline, existing thermal-aware, and proposed methods.

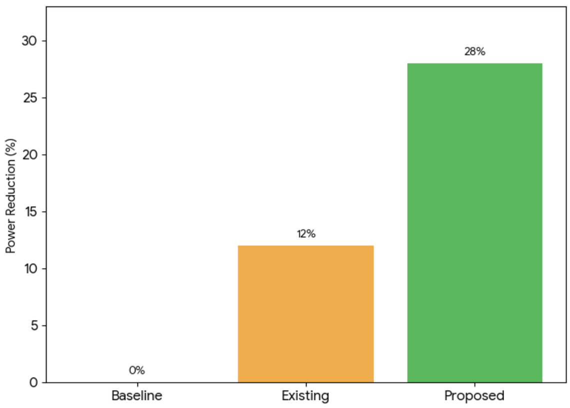

Fig.3: Total power and leakage power reduction achieved by the proposed method relative to baseline.

peak temperature reduction in each of the methods, and Figure 3 shows the total power savings and leakage power savings in the proposed approach.

The presence of trade-off analysis is performed in order to assess the structural implications of thermal mitigation. The suggested framework brings the total wire length moderately up because of thermal spreading measures which re-distribute high-power blocks outside the hotspots. Nonetheless, such increase is within manageable design limits and does not have a major effect on timing. On a similar note, slightly different redistribution on areas is established to allow better heat dissipation routes. Figure 4 demonstrates the trade-off between the temperature and a wirelength, and it is also clear that routing overhead is lowered at the cost of significant thermal improvement. Moreover, trade-offs between power and area testify to the fact that optimization of the voltage scaling does not lead to scaling up of the structure coupled with a corresponding energy expenditure.

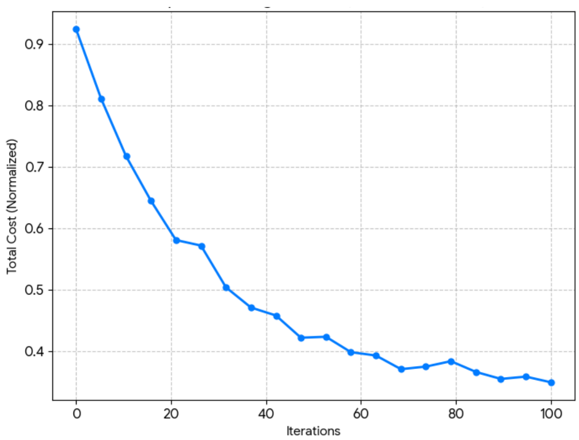

Fig. 4: Temperature versus wirelength trade-off curve for candidate floorplans generated during optimization.

Pareto frontier visualisation is conducted to further assess performance based on multi-objectives by using the NSGA-II results in terms of optimization. The obtained Pareto front shows that the suggested strategy is characterised by a strong span in minimising both peak temperature and total power at the same time, and competitive structural values. The frontier shows

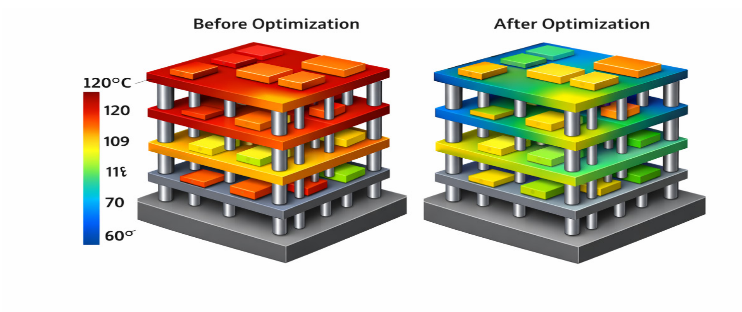

Figure 5: 3D Thermal Heatmap Before and After Optimization

Fig.5: 3D thermal heatmap comparison of baseline and proposed floorplans for a four-layer stacked SoC.

that the solutions produced by the algorithm proposed always perform better as compared to the baseline configurations in thermalpower space which proves useful exploration and diversity is maintained in the multi-objective search process. Scalability analysis is done by adding more layers by moving the stacked layers to 4. Expectedly, peak temperature increment in the baseline set up is attributed to enhanced vertical heat concentration. The suggested framework however keeps the temperature under controlled increase by enhanced redistribution of blocks and output of leakage-conscious voltage scaling. In Figure 5, 3D thermal heatmap of four-layer arrangement at the beginning and end of optimization is shown, which clearly demonstrates that the intensity of hotspots decreases, and the uniformity of temperatures becomes higher in the optimised structure.

In general, the findings support the hypothesis that the combination of thermal modelling, multi-objective placement optimization, and dynamic power adaptation as a component of a single framework can lead to significant in terms of peak temperature, quantifiable power saving, and reliability metrics. The scalability assessment also proves the applicability of the proposed technique with stacking density, which confirms its usefulness in next-generation high-density 3D integrated SoC architectures.

Discussion

The presented thermal-aware floorplanning and joint power optimization framework can gain quantifiable values of physical temperature reduction, leakage minimization, and enhanced reliability to 3D integrated SoC structures. Nevertheless, various practical aspects in real-life applications in design processes pose some technical problems that need to be taken into consideration. Among the main issues of implementation is the close connexion between thermal modelling and physical design loop. Although compact thermal solvers can be used to perform iterative circulation temperature estimation, the process of adding compact thermal predictors to a large volume of industrial electronic design automation (EDA) could introduce some extra burden to the workload process. With high-complexity design with thousands of modules and packed TSV networks, multi-objective optimization with repeated thermal assessment may exert high burdens on the computation demands. Even though exploration of the design space is conducted effectively using NSGA-II, time scale of convergence might be quite large depending on the depth of the stack and the design scale. Hence, large-scale industrial applications might need to use runtime optimization techniques, parallel testing or surrogate thermal modelling.

Scalability is also dependent upon the complexity of the algorithm. The multi-objective model uses evolutionary approach in terms of population which specifically needs several candidate evaluation per generation. In combination with real-time thermal simulation, leakage-sensitive power recalculation, the complexity of computation grows linearly with the size of population and quadratically with the number of modules based upon the operation of perturbation to placement. Although this complexity is fine in the initial exploration of architectures, region-based optimization or hierarchical partitioning is likely to be required in very high-level industrial applications to dimensionally reduce search. Moreover, the possible tuning of weighting coefficients in the cost function adds further sensitivity in the design that would need to be calibrated to fulfil particular performance and thermal requirements. Industrial feasible-wise, the framework agrees with the new developments in 3D integration, notably, heterogeneous stacking chips and AI-based chiplet designs. The combination of DVFS and leakage-conscious model of the placement optimization is to serve the needs of power-constrained edge devices and high-performance computing systems, whose thermal margins are getting smaller. Moreover, the relevance of the framework to mission-critical applications and the long-life of the applications is able to be reinforced by adding such reliability-oriented metrics as the estimation of the MTTF. Nevertheless, it would have to be initiated by industrial adoption, which would entail uninterrupted integration with commercial placement and routing models, proper parameter extraction through process design kit (PDK) and integration with advanced packaging models. In the absence of such integration, the framework can continue to be more of an academic exploration tool.

There are also some restrictions that have to be mentioned. Firstly, the current model is based on steady-state thermal modelling and does not actively include the time-varying dynamic thermal effects of time-varying workloads. Real applications can experience varying workloads, resulting in rapid changes in temperature which change leakages and ageing not as the steady-state model predicts. Second, the leakage model is temperature aware, but does not account for process variation or process ageing induced parameter changes. Third, the effects of mechanical stress in multi-layered stacks and thermo-mechanical coupling are not explicitly modelled, but they may also play a major role of affecting reliability in dense 3D structures. Lastly, increased wirelength versus thermal improvement may be increased at the cost of very large designs, but this effect is small in the current results. Nevertheless, the given limitations do not imply that the proposed thermalco-design approach to design is not a strong point of the structural placement optimization, the correct thermal modelling, and the active power management in the 3D integrated SoCs. The findings suggest that early co-optimization is much higher as compared to reactive thermal control schemes, and provides a scalable route to energy-efficient, thermally resilient, and awareness of reliability 3D integration on next generation semiconductor platforms.

Conclusion

This paper introduced a detailed thermal-conscious floorplan and power optimisation model of multi-layer 3D integrated SoC designs, which is sensitive to the issues of hotspots, leakage increase, and reliability loss of vertical stacking. Through smaller thermal modelling with a multi-objective optimization engine and leakage-conscious DVFS and power gating policies the proposed approach showed huge performance gains compared to traditional ones. Experimental analysis showed a peak decrease of around 1825 percent of thermal in comparison to thermal-oblivious base line floor planning and an overall power saving between 1520 percent and leakage decrease by large scale in thermally-sensitive zones. Such thermal and power benefits were reflected as quantifiable reliability benefits in the form of long mean time to failure and lowering of electromigration risk. The scalability analysis also established that the framework is effective at bigger stacking density, and hence can be used in the next-generation high-density 3D SoC designs. The research in the future will also involve the integration of transient thermal modelling, workload-conscious adaptive optimization, process-variation-conscious reliability estimation, and machine learning-based surrogate modelling to improve even more the computational properties and industrial useability of thermal-power co-design schemes.

References

1. Chalise, D., & Cahill, D. G. (2023). Anisotropic thermal conductivity of high bandwidth memory. *arXiv preprint arXiv:2303.06785*.
2. Chowdary, D., & Sudhakar, M. S. (2023). Multi-objective floorplanning optimization engaging dynamic programming for system on chip. *Microelectronics Journal*, *140*, 105942.
3. Cuesta, D., Risco-Martín, J. L., Ayala, J. L., & Hidalgo, J. I. (2015). Thermal-aware floorplanner for 3D IC, including TSVs, liquid microchannels and thermal domains optimization. *Applied Soft Computing*, *34*, 164-177.
4. Dayasagar Chowdary, S., & Sudhakar, M. S. (2024). Linear programming-based multi-objective floorplanning optimization for system-on-chip. *The Journal of Supercomputing*, *80*(7), 9663-9686.
5. Geb, D., Deodhar, S., Netake, N., & Jeurkar, T. (2024). Power Throttling in a 3D Integrated Circuit (IC) Dynamic Thermal Simulation. *Power*.
6. Guan, W., Tang, X., Lu, H., Zhang, Y., & Zhang, Y. (2023). Thermal-aware fixed-outline 3-D IC floorplanning: An end-to-end learning-based approach. *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, *31*(12), 1882-1895.
7. Guo, F., Suo, Z. J., Xi, X., Bi, Y., Li, T., Wang, C., & Li, R. (2025). Recent developments in thermal management of 3d ics: A review. *IEEE Access*.
8. Gupta, R., & Gill, S. S. (2022). A Comprehensive Analysis in Recent Advances in 3D VLSI Floorplan Representations. *Advances in VLSI and Embedded Systems: Select Proceedings of AVES 2021*, 243-265.
9. Jiang, Z., Li, Z., & Yao, Z. (2024). Multi-objective optimization in 3D floorplanning. *Electronics*, *13*(9), 1696.
10. Lee, M., Arimpoor Pallan Ouseph, L., & Yaddanapudi, V. K. (2024, October). Advanced Thermal Aware Mobile SoC Floorplan Design Optimization by Integrating Proactive Design Rule Check Process and Evolutionary Algorithms. In *International Electronic Packaging Technical Conference and Exhibition* (Vol. 88469, p. V001T03A015). American Society of Mechanical Engineers.
11. Lu, S., & Vafai, K. (2023). Thermal performance optimization of the three-dimensional integrated circuits employing the integrated chip-size double-layer or multi-layer microchannels. *ASME Journal of Heat and Mass Transfer*, *145*(3), 032501.
12. Shanthi, J., Rani, D. G. N., Rajalakshmi, M., & Salau, A. O. (2025). Multi-objective GA-SA (MOGASA) algorithm for TSV and thermal aware 3D IC floorplanning. *Case Studies in Thermal Engineering*, 106944.
13. Wang, Q., Zhu, T., Lin, Y., Wang, R., & Huang, R. (2024, May). Atsim3d: Towards accurate thermal simulator for

heterogeneous 3d-ic systems considering nonlinear leakage and conductivity. In *2024 2nd International Symposium of Electronic Design Automation (ISEDA)* (pp. 618-623). IEEE.

14. Wang, R., Wang, Z., Lin, T., Raby, J. M., Stan, M. R., & Guo, X. (2025). Cool-3D: An End-to-End Thermal-Aware Framework for Early-Phase Design Space Exploration of Microfluidic-Cooled 3DICs. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*.
15. Wang, R., Wang, Z., Lin, T., Raby, J. M., Stan, M. R., & Guo, X. (2025). Cool-3D: An End-to-End Thermal-Aware Framework for Early-Phase Design Space Exploration of Microfluidic-Cooled 3DICs. *IEEE Journal on Emerging and Selected Topics in Circuits and Systems*.
16. Yu, X., Liu, Z., Li, H., Li, Y., Ai, X., Zeng, Z., & Zhang, Z. (2025). DeepOHeat-v1: Efficient operator learning for fast and trustworthy thermal simulation and optimization in 3D-IC design. *IEEE Transactions on Components, Packaging and Manufacturing Technology*.