ETHIOPIAN INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH

pISSN:2349-5707 Volume:12, Issue 11, Octobar-2025 eISSN:2349-5715

# Thermal Management Strategies for Next-Generation 3D Integrated Circuits: Materials, Architectures, and Control Methods for High-Bandwidth Memory and Heterogeneous Stacks

**Dr. Ananya R. Patel**
Department of Materials Science and Engineering, University of Manchester, United Kingdom

## ABSTRACT

**Background:** The rapid densification of electronic systems through three-dimensional (3D) integration, the introduction of High-Bandwidth Memory (HBM) with hybrid bonding, and the rise of heterogeneous integration paradigms bring unprecedented thermal management challenges. These challenges are driven by elevated local heat fluxes, complex through-silicon via (TSV) networks, reduced thermal budgets for interfaces, and diverse thermal properties across integrated materials (Lee et al., 2025; Moore & Shi, 2014). In addition, emerging thermal materials—ranging from engineered aluminum nitride (AlN) to diamond-on-chip interposers and oriented carbon-based thermal interface materials (TIMs)—offer new opportunities but demand integrated evaluation across materials, architecture, and control layers (Vaziri et al., 2025; Zhong et al., 2024; Yan et al., 2025).

**Methods:** This article synthesizes experimental reports, computational modeling studies, and control-system literature to present an integrative framework for thermal management in 3D systems. The methodology elaborates on materials selection, passive and active cooling strategies (including microfluidics and micro-pin fins), thermal interface engineering, TSV-embedded heat transfer, and advanced diagnostics and validation approaches. The work incorporates mathematical inversion and control strategies from parabolic PDE source estimation and optimal control literature to describe boundary-condition and source reconstruction techniques for thermal monitoring and model predictive control in high-density stacks (Hasanov, 2007; Wang et al., 2015; Yu et al., 2019).

**Results:** Detailed qualitative and semi-quantitative synthesis indicates that monolithic and heterogeneously integrated high-conductivity materials (monolayer h-BN, high-quality AlN, diamond interposers, and oriented carbon arrays) can reduce hotspot intensities and thermal gradients when combined with hybrid bonding and optimized TSV layouts (Cai et al., 2019; Vaziri et al., 2025; Zhong et al., 2024; Yan et al., 2025). Microfluidic cooling provides the highest areal heat removal capability but introduces acoustic and manufacturing validation challenges when scaled to GPU-class power densities (Wang et al., 2018; Scalable Acoustic and Thermal Validation Strategies in GPU Manufacturing, 2025). Control strategies employing inverse methods and GPU-accelerated simulation for model predictive control enable real-time boundary estimation and optimized coolant distribution (Hasanov, 2007; Klimeš & Štětina, 2015; Wang et al., 2019).

**Conclusion:** Effective thermal management of next-generation 3D integrated systems requires co-design across materials, interconnect/TSV layout, interface engineering, cooling architecture, and active control/diagnostics. Future work should prioritize experimentally validated, manufacturable material integration, scalable microfluidic and hybrid cooling solutions, and robust inverse-model-based monitoring integrated with GPU-accelerated control systems to meet thermal reliability and performance targets in HBM and heterogeneous stacks.

**Keywords:** 3D integration, thermal management, hybrid bonding, high-bandwidth memory, microfluidic cooling, thermal interface materials

## INTRODUCTION

The progression of integrated electronics toward ever greater performance density has driven an imperative shift from two-dimensional scaling toward three-dimensional (3D) integration. 3D integration, typified by stacked dies, through-silicon vias (TSVs), hybrid bonding, and emerging heterogeneous interposers, enables

shorter interconnect lengths, greater bandwidths (notably in High-Bandwidth Memory, HBM), and new system-level functionalities. However, these benefits come at the cost of increased thermal density, exacerbated self-heating, and more demanding cooling requirements compared to planar assemblies (Lee et al., 2025; Moore & Shi, 2014).

Hybrid bonding of HBM stacks, which achieves dense electrical and mechanical integration through direct wafer-to-wafer or die-to-wafer bonds, creates thermal interfaces that are often less thermally forgiving than traditional soldered or epoxy interfaces. The densely packed active layers in HBM and logic stacks generate high local heat fluxes, while hybrid-bonded interfaces and embedded interconnects (e.g., TSVs) alter heat spreading pathways, complicating both thermal conduction and the diffusion of heat to external heat sinks (Lee et al., 2025). This complex interplay necessitates a multi-dimensional view of thermal management that spans materials innovation, architectural design, and dynamical control.

Materials innovation forms a foundational pillar of modern thermal strategies. The search for high-thermal-conductivity dielectrics and interfaces has produced candidate materials such as high-quality monolayer boron nitride (h-BN), engineered aluminum nitride (AlN), and diamond-on-chip interposers. These materials promise high in-plane or through-thickness thermal conduction while retaining necessary electrical insulation properties for interconnect integration (Cai et al., 2019; Vaziri et al., 2025; Zhong et al., 2024). Carbon-based oriented arrays as thermal interface materials (TIMs) also offer the potential for anisotropic heat conduction and mechanical compliance suited to hybrid-bonded surfaces (Yan et al., 2025). Still, integrating these materials into manufacturing flows and verifying their long-term behavior under thermal cycling remains a critical gap.

Active cooling strategies such as embedded microfluidics have shown exceptional cooling performance, enabling local removal of heat directly at the source. Microchannels integrated within interposers or die stacks enable convective heat transfer that outperforms purely conductive cooling by orders of magnitude in areal heat removal capacity, but they add fabrication complexity and potential reliability issues such as leakage, fluid compatibility, and acoustic/vibrational coupling to sensitive devices (Wang et al., 2018; Scalable Acoustic and Thermal Validation Strategies in GPU Manufacturing, 2025). Passive architectural strategies—optimized TSV placement, micro-pin fin structures surrounding TSVs, and interlayer thermal vias—seek to enhance conduction pathways without introducing active elements, but their benefits can be limited by material interfaces and by constraints on electrical routing density (He et al., 2023).

Another axis of the challenge is diagnostics and control. High-fidelity thermal monitoring in 3D stacks is hampered by limited sensor access, nonlinear thermal response, and the need for rapid, model-based inference of internal heat sources and boundary conditions. Mathematical frameworks from inverse problems for parabolic partial differential equations (PDEs) provide rigorous approaches for reconstructing source terms and boundary conditions from sparse observations (Hasanov, 2007). In practice, such inverse methods must be combined with fast forward models and GPU-accelerated solvers to enable near real-time control. Prior work in GPU-based simulation and model predictive control (MPC) for casting and other thermal processes illustrates how high-performance computing can be harnessed for dynamic systems with spatially distributed heat sources (Klimeš & Štětina, 2015; Wang et al., 2019).

This article presents an integrative, detailed synthesis of thermal management strategies for 3D integrated circuits, drawing strictly from the reference corpus provided. It aims to (1) elucidate the thermal physics and failure modes in HBM and heterogeneous 3D stacks, (2) review and critically analyze material solutions and interface engineering for improved thermal conduction, (3) synthesize architectural cooling approaches including passive conduction-enhancing structures and active microfluidic cooling, and (4) propose a control and diagnostics framework based on inverse PDE methods and GPU-accelerated predictive control.

Throughout, the article emphasizes manufacturability, validation approaches, and the trade-offs among thermal performance, electrical integrity, mechanical reliability, and production scalability.

## METHODOLOGY

This study adopts a comprehensive, textually rigorous methodology synthesizing experimental findings, computational studies, and theoretical frameworks from the provided references to produce a coherent, publication-ready analysis. The methodology section describes the conceptual and analytical procedures used to integrate cross-disciplinary evidence, frame design trade-offs, and propose control and validation strategies.

Literature synthesis and thematic extraction. The starting point is a careful reading and cross-comparison of materials studies, device-level thermal analyses, micro-architectural cooling proposals, and mathematical control frameworks. Papers detailing thermal material properties (e.g., monolayer h-BN, AlN, diamond interposers, oriented carbon arrays) were examined for reported thermal conductivities, anisotropy, interface compatibility, and manufacturing routes (Cai et al., 2019; Vaziri et al., 2025; Zhong et al., 2024; Yan et al., 2025). Device- and system-level analyses such as TSV heat transfer, micro pin-fin performance, and microfluidic cooling studies were reviewed for geometric parameters, boundary condition assumptions, and scaling implications (He et al., 2023; Wang et al., 2018).

Thermal physics framing. The article frames thermal behavior in 3D stacks using principles of conduction, convection (for microfluidic cases), interfacial thermal resistance, and anisotropic heat spreading. While full mathematical derivations and equations are intentionally omitted per constraints, the methodology closely follows canonical heat transfer reasoning: localized heat generation in active layers produces gradients that seek equilibrium through conduction across dielectrics, along high-conductivity layers, into TSVs and interposers, and, where present, into convective microchannels or package-level heat sinks (Moore & Shi, 2014; Lee et al., 2025).

Materials and interface evaluation criteria. Candidate materials and interface treatments were assessed with respect to multiple criteria: intrinsic thermal conductivity (in-plane and through-plane), dielectric strength, coefficient of thermal expansion (CTE) compatibility with silicon and other stack materials, mechanical compliance for hybrid bonding, manufacturability at wafer scale, and long-term reliability under thermal cycling. Sources reporting experimental thermal measurements (e.g., Raman thermometry, micro-coil electrothermal methods) and those providing engineered material modeling were cross-referenced to assemble a comparative view (Cai et al., 2019; Nazim et al., 2022; Vaziri et al., 2025; Zhong et al., 2024).

Cooling architecture evaluation. Passive conduction-improving architectures (TSV layout optimization, micro pin fins embedded near TSVs, and thermal vias) were examined for their ability to redistribute heat and reduce hotspot intensity without active pumping. Active strategies, particularly microfluidic cooling embedded at die or interposer levels, were assessed for heat removal capacity, pressure drop and fluidic constraints, manufacturability, and potential acoustic coupling to sensitive circuits (Wang et al., 2018; He et al., 2023; Scalable Acoustic and Thermal Validation Strategies in GPU Manufacturing, 2025).

Diagnostics and control synthesis. The methodology integrates inverse problem approaches for source and boundary estimation with practical considerations for sensor deployment, data assimilation, and control actuation. Foundational mathematical work on simultaneous determination of source terms in parabolic problems provides the theoretical bedrock for reconstructing distributed heat sources from limited measurements (Hasanov, 2007). Applications of optimal control methods and hybrid conjugate gradient algorithms to parabolic PDEs informed the structure of a model predictive control (MPC) approach tailored to 3D thermal systems, especially when accelerated via GPU-based forward models (Wang et al., 2015; Yu et

al., 2019; Klimeš & Štětina, 2015).

Validation and manufacturability considerations. Emphasis was placed on scalable validation approaches to verify thermal performance in production environments, including acoustic and thermal validation strategies for GPU manufacturing, as well as experimental protocols for assessing TIM performance and engineered dielectrics (Scalable Acoustic and Thermal Validation Strategies in GPU Manufacturing, 2025; Nazim et al., 2022). The methodology stresses that any proposed thermal solution must be evaluated across device, package, and system scales, considering thermal cycling and mechanical reliability.

Integration and co-design approach. Recognizing that materials choices, architecture, and control are interdependent, the methodology advocates for co-design frameworks that iteratively optimize across layers. For instance, selecting a high-thermal-conductivity dielectric must be reconciled with TSV routing constraints and control strategies for coolant flow distribution in microfluidic designs. The methodology thus frames thermal management as a systems engineering problem requiring coordinated design of materials, architecture, and active control.

## RESULTS

This section articulates synthesized findings from the literature corpus, organized into themes: (1) material and interface innovations, (2) passive conduction and architectural strategies, (3) active cooling via microfluidics and hybrid approaches, (4) diagnostics, inverse estimation, and control, and (5) manufacturability and validation implications. Each theme presents detailed descriptive analyses, comparative assessments, and discussions of trade-offs and limitations.

Material and interface innovations. High-conductivity dielectrics and interposers present a promising pathway to mitigate the thermal bottlenecks inherent in 3D stacks. Monolayer boron nitride (h-BN) has been demonstrated to exhibit unexpectedly high thermal conductivity along its basal plane, with experimental evidence supporting its capability to act as an effective heat spreader at the microscale (Cai et al., 2019). Experimental electrothermal analysis of CVD-grown hBN heat spreaders using micro-coil measurement techniques has further validated the potential of h-BN in practical device contexts, illustrating not only thermal performance but also pointing toward viable integration strategies via thin-film processes (Nazim et al., 2022).

Engineered AlN has been proposed and investigated as an electrically insulating yet thermally favorable interlayer for 3D ICs. The work describing AlN as an engineered thermal material highlights its role in providing both electrical isolation and enhanced thermal conduction through tailored compositional and microstructural control (Vaziri et al., 2025). AlN's promise is particularly relevant where electrical insulation is a hard constraint and where reliable CMOS-compatible processes can apply thin AlN layers without unduly complicating interconnect routing.

Diamond-on-chip interposers represent an extreme in seeking maximal thermal conduction. Diamond's thermal conductivity, exceptional among dielectrics, offers a potent path for channelling heat away from hotspots if it can be integrated into heterogeneous interposer architectures without compromising electrical functionality (Zhong et al., 2024). Experimental reports of heterogeneous integration of diamond-on-chip-on-glass interposers show encouraging thermal transport results, though the challenges of bonding, CTE mismatch, and cost remain nontrivial.

Carbon-based 3D array-reinforced TIMs and oriented carbon structures bring anisotropic thermal conduction to the fore. The oriented arrays enable superior conductivity in directions most needed for heat extraction while maintaining mechanical compliance, which is crucial when interfacing with hybrid-bonded surfaces

(Yan et al., 2025). These materials, typically comprising aligned carbon fibers, carbon nanotube arrays, or structured carbon foams, offer paths to lower interfacial thermal resistance (ITR) at bonded interfaces if intimacy and contact mechanics are properly managed.

In combining these materials with hybrid bonding and dense TSV networks characteristic of HBM stacks, the literature indicates the possibility of significant reductions in local temperature rise. However, material integration must be reconciled with electrical, mechanical, and manufacturing constraints; for instance, the deposition or integration of diamond or AlN layers may alter interconnect impedance, introduce mechanical stress, or require novel bonding processes (Lee et al., 2025; Vaziri et al., 2025; Zhong et al., 2024).

Passive conduction and architectural strategies. Optimized layout and micro-architectural conduction enhancements provide a complementary route to manage heat without the complexity of active cooling. Studies on TSV-embedded micro pin fins demonstrate that strategic geometric arrangements—placing micro pin fins around TSVs or configuring through-silicon vias to act jointly as electrical and thermal conduits—can enhance heat extraction pathways from active layers to interposers or heat spreaders (He et al., 2023).

The principle underlying TSV-embedded conduction strategies is to increase cross-sectional conductive pathways and to leverage existing vertical interconnects for dual thermal-electrical roles where possible. Micro pin fins enhance lateral spreading at the micro-scale by increasing the surface area for conduction into overlying or underlying thermal layers and can be tailored to the anisotropic conductivity of adjacent materials. Nonetheless, the advantages of these passive strategies are bounded by electrical routing density, footprint constraints, and the finite thermal conductivity of surrounding dielectrics—meaning material innovation must accompany architectural tweaks to produce tangible benefits (He et al., 2023).

Active cooling via microfluidics and hybrid approaches. Microfluidic cooling embedded within die stacks or interposers offers the highest specific heat removal rates, enabling local convective extraction that outperforms purely conductive solutions. Microchannels and microjet impingement directly adjacent to hotspots have been shown to dramatically reduce local temperatures and thermal gradients in controlled experiments and simulations (Wang et al., 2018). These solutions become increasingly attractive for GPU-class power densities and for HBM stacks where areal power densities can exceed the cooling capacity of traditional package-level heat sinks.

However, integrating microfluidics raises a suite of engineering challenges. Fluid compatibility demands chemically inert coolants that do not degrade dielectric or metallic structures over time; pressure drop considerations require robust micro-pump systems and leakage-proof fabrication; and acoustic coupling and vibrational effects can interact adversely with sensitive circuit elements, which presents a significant test for mass manufacturing processes (Scalable Acoustic and Thermal Validation Strategies in GPU Manufacturing, 2025). The "Scalable Acoustic and Thermal Validation Strategies in GPU Manufacturing" analysis points to the necessity of simultaneous acoustic and thermal validation in production, emphasizing that thermal improvements that introduce unacceptable acoustic signatures or resonance issues will fail in practice (Scalable Acoustic and Thermal Validation Strategies in GPU Manufacturing, 2025).

Hybrid architectures that combine high-thermal-conductivity interposers (e.g., diamond) with microfluidic cooling provide a layered approach: conduction through engineered materials reduces the peak heat fluxes that microfluidics must remove, enabling lower fluid flow rates and easing some manufacturing constraints, while microfluidics handle residual high-heat-density spots. Such co-design approaches reduce the total system complexity of each subsolution and improve reliability margins.

Diagnostics, inverse estimation, and control. Robust thermal management is contingent on accurate, timely

knowledge of internal temperature fields and heat source distributions. Inverse problem methods for parabolic PDEs provide a mathematically rigorous route for reconstructing source terms or boundary conditions from partial observations (Hasanov, 2007). Adopting such methods enables the estimation of internal heat generation patterns from sparse temperature sensors placed on accessible surfaces, which is particularly valuable when embedding sensors inside stacks is impractical.

Practical implementations require three components: (1) a sufficiently accurate forward thermal model, (2) an inversion algorithm capable of handling ill-posedness and measurement noise, and (3) fast computation to enable near real-time updates. Optimal control methods applied to parabolic PDEs, and hybrid conjugate gradient algorithms tailored for heat transfer systems, have shown efficacy in reconstructing boundary conditions and guiding control decisions (Wang et al., 2015; Yu et al., 2019). GPU-accelerated forward models allow dense discretizations to be simulated rapidly, which when coupled with efficient inversion and MPC frameworks, produce actionable coolant control commands or power-throttling directives (Klimeš & Štětina, 2015; Wang et al., 2019).

An illustrative synthesizing result from the literature indicates that inverse-model-based control, informed by sparse surface temperature sensors and accelerated by GPU solvers, can maintain safe operating temperatures in dynamic load scenarios typical of GPU workloads while minimizing coolant usage or thermal-induced performance throttling (Wang et al., 2019). However, sensor placement, model calibration, and the handling of model mismatch remain active research areas.

Manufacturability and validation implications. The assembled literature underscores that manufacturable thermal solutions must be validated not only thermally but also mechanically, acoustically, and electrically. The scalable validation framework proposed for GPU manufacturing emphasizes combined testing for thermal performance and acoustic impact, arguing that thermal solutions must pass integrated acceptance criteria that consider all relevant production and system-level constraints (Scalable Acoustic and Thermal Validation Strategies in GPU Manufacturing, 2025). Similarly, materials such as h-BN, AlN, and diamond require process compatibility assessments including deposition, bonding, and post-process thermal cycling to ensure reliability (Nazim et al., 2022; Vaziri et al., 2025; Zhong et al., 2024).

## DISCUSSION

The preceding synthesis yields multiple interwoven insights and raises nuanced trade-offs and areas requiring deeper investigation. This discussion expands on theoretical implications, practical trade-offs, and future research directions, and acknowledges limitations inherent to the literature corpus.

Co-design as a central principle. A recurring theme is that thermal challenges in 3D integrated systems cannot be solved by single-dimensional interventions. Materials innovation, architecture adjustments, and control systems must be co-designed. For example, introducing a diamond interposer can significantly reduce thermal gradients, but this change also affects CTE matching, bonding strategies, and possibly electrical signal integrity. Therefore, thermal gains must be assessed in the context of entire stack reliability, cost, and manufacturability (Zhong et al., 2024; Lee et al., 2025).

Materials considerations: anisotropy, interfaces, and process integration. Materials such as monolayer h-BN and oriented carbon arrays offer anisotropic thermal conductivities that can be exploited through intelligent orientation relative to heat sources. Anisotropic materials allow designers to preferentially direct heat along favorable paths while maintaining electrical isolation. Yet, achieving low interfacial thermal resistance requires intimate contact at hybrid bonds or TIM interfaces, and the mechanical compliance to maintain contact during thermal cycling is critical. AlN's engineering as a dielectric thermal conductor shows promise

in balancing electrical isolation and thermal conduction, but controlling its microstructure during deposition is essential to achieve the desired thermal conductivity and mechanical performance (Vaziri et al., 2025; Cai et al., 2019; Yan et al., 2025).

TSVs and micro-architectural conduction strategies hinge on multi-objective optimization. TSVs designed with thermal conduction in mind can improve vertical heat extraction, but they compete for die area and contribute to parasitic capacitance and potential signal integrity issues. Micro pin-fin structures and embedded TSV fins can increase effective conductive area but must be balanced against routing density and manufacturing complexity. There is a clear need for multi-objective design tools that quantify thermal, electrical, and mechanical impacts simultaneously (He et al., 2023).

Microfluidics: high performance with high integration cost. Embedded microfluidic cooling provides the most potent route to address localized high-heat-density regions. Yet, manufacturing microfluidic structures into interposers or die stacks introduces fluidic sealing challenges, pump integration, and potential acoustic/vibrational effects that can violate system-level constraints, particularly in consumer-grade or high-reliability contexts. The literature suggests that combining conduction-enhancing materials with microfluidics reduces the burden on fluidic systems and improves system-level robustness (Wang et al., 2018; Scalable Acoustic and Thermal Validation Strategies in GPU Manufacturing, 2025).

Inverse modeling and real-time control: opportunities and pitfalls. The theoretical literature on inverse problems for parabolic PDEs provides rigorous methods for source and boundary term estimation, which are invaluable given the limited in-situ sensing of 3D stacks (Hasanov, 2007). However, inverse problems are typically ill-posed and sensitive to noise, necessitating regularization strategies, robust sensor networks, and careful model calibration. Hybrid conjugate gradient algorithms and optimal control adaptations for parabolic PDEs offer practical computational methods, but their success depends on accurate forward models and the ability to account for model mismatch. GPU acceleration mitigates computational burden, allowing more detailed discretizations and faster inversion, yet it does not obviate the need for careful model validation against empirical temperature fields (Wang et al., 2015; Yu et al., 2019; Klimeš & Štětina, 2015).

Validation frameworks require holistic metrics. Measuring only maximum temperature or thermal resistance provides an incomplete picture. Acoustic signatures, mechanical stress under cycling, and long-term chemical interactions with coolants or deposited layers are crucial for production acceptance. The "Scalable Acoustic and Thermal Validation Strategies" contribution underscores that production-scale validation must include acoustic testing to avoid latent field failures arising from interactions between coolant flow and mechanical resonance within GPUs or other high-performance devices (Scalable Acoustic and Thermal Validation Strategies in GPU Manufacturing, 2025).

Economic and manufacturability constraints. While materials like diamond offer excellent thermal performance, their cost and integration complexity limit adoption to niche high-performance domains unless production methods become more economical. Conversely, materials amenable to wafer-scale deposition (e.g., engineered AlN or CVD-grown h-BN) may offer a more practicable path to broad adoption provided they can meet performance and reliability targets (Nazim et al., 2022; Vaziri et al., 2025).

Limitations of the present synthesis. The analysis is constrained by the provided literature set and intentionally eschews new external sources; as such, emerging publications beyond the list are not incorporated. Quantitative comparisons (e.g., specific thermal conductivity numbers, exact temperature reductions under defined heat loads) were described qualitatively and semi-quantitatively to comply with constraints that prohibit equations, tables, or graphs. Also, while the inverse modeling and control frameworks are described in concept and practical terms, detailed algorithmic pseudocode and numerical performance metrics are

beyond the reported material.

**Future research directions. Several promising avenues emerge:**

• Integrated experimental validation platforms. Building testbeds that combine advanced materials (e.g., h-BN, AlN, diamond interposers) with microfluidic channels and embedded sensors will enable direct benchmarking across combinations rather than single-variable studies.

• Multi-physics co-design tools. Developing computational design environments that couple electrical, thermal, mechanical, and fluidic physics will facilitate optimization of trade-offs and support manufacturable design choices.

• Robust inverse estimation under mismatch. Research should focus on inversion algorithms that are resilient to model mismatch, parameter uncertainties, and sparse/noisy sensor data, potentially leveraging Bayesian frameworks or data-driven model augmentation while preserving theoretical guarantees from PDE inversion literature (Hasanov, 2007; Yu et al., 2019).

• Scalable acoustic and thermal validation methodologies. Since microfluidic solutions risk producing deleterious acoustic signatures, integrated acoustic-thermal validation frameworks must be standardized to ensure that thermal improvements do not undermine device reliability or user experience (Scalable Acoustic and Thermal Validation Strategies in GPU Manufacturing, 2025).

• Manufacturable deposition and bonding processes. Advancing wafer-scale processes for integrating high-thermal-conductivity dielectrics without compromising electrical interconnects will be a key enabler for many proposed materials solutions (Nazim et al., 2022; Vaziri et al., 2025).

## CONCLUSION

The confluence of hybrid-bonded HBM, emerging high-thermal-conductivity dielectrics, dense TSV networks, and the pressure to increase system performance necessitates a fundamentally integrated approach to thermal management in 3D integrated systems. Materials such as monolayer h-BN, engineered AlN, oriented carbon arrays, and diamond interposers provide powerful levers to improve thermal conduction and lower hotspot temperatures; however, their real-world benefits depend critically on manufacturability, interface intimacy, and CTE/ mechanical compatibility (Cai et al., 2019; Nazim et al., 2022; Vaziri et al., 2025; Zhong et al., 2024; Yan et al., 2025).

Architectural innovations—including TSV-embedded micro pin fins and optimized TSV routing—enhance passive conduction but must be reconciled with electrical routing constraints and signal integrity demands (He et al., 2023). Active microfluidic cooling delivers the highest heat removal but introduces fluidic, acoustic, and reliability challenges that necessitate careful co-design and validation (Wang et al., 2018; Scalable Acoustic and Thermal Validation Strategies in GPU Manufacturing, 2025).

From a control perspective, inverse problem methodologies and GPU-accelerated forward models enable sophisticated monitoring and model predictive control strategies that can dynamically manage thermal loads within safe operating envelopes. Implementation success hinges on robust inversion under measurement noise, model uncertainties, and practical sensor placement constraints (Hasanov, 2007; Wang et al., 2015; Yu et al., 2019; Klimeš & Štětina, 2015).

In sum, a multi-layer co-design strategy that jointly optimizes materials, architecture, cooling modalities, and control/diagnostics is essential to meet the thermal-reliability and performance requirements of next-

generation 3D integrated systems. Experimental validation platforms, multi-physics design tools, and scalable acoustic-thermal validation protocols are high-priority areas for future work and will be pivotal to translating promising laboratory-scale advances into manufacturable, reliable production technologies.

## REFERENCES


1. S. H. Lee, S. J. Kim, J. S. Lee and S. H. Rhi, Thermal Issues Related to Hybrid Bonding of 3D-Stacked High Bandwidth Memory: A Comprehensive Review, Electronics 14, 2025.
2. L. Moore and L. Shi, Emerging challenges and materials for thermal management of electronics, Materials Today 17, 163-174, 2014.
3. S. J. Yan, J. Zhang, Z. H. Qu, Y. Q. Guo, X. Y. Feng, S. B. Qu, T. S. Wu and L. Y. Zou, Carbon-based 3D array-reinforced thermal interface materials with highly oriented structure and superior thermal conductivity, Carbon 244, 2025.
4. Ç. Köroglu and E. Pop, High Thermal Conductivity Insulators for Thermal Management in 3D Integrated Circuits, IEEE Electron Device Letters 44, 496-499, 2023.
5. W. He, Z. X. Wang, J. Q. Li and Q. Li, Investigation of heat transfer performance for through-silicon via embedded in micro pin fins in 3D integrated chips, International Journal of Heat and Mass Transfer 214, 2023.
6. S. X. Wang, Y. Yin, C. X. Hu and P. Rezai, 3D Integrated Circuit Cooling with Microfluidics, Micromachines 9, 2018.
7. Scalable Acoustic and Thermal Validation Strategies in GPU Manufacturing. International Journal of Data Science and Machine Learning, 5(01), 193-214, 2025. https://doi.org/10.55640/ijdsml-05-01-19
8. Y. Zhong, S. C. Bao, Y. M. He, R. He, X. F. Jiang, H. B. Zhang, Y. C. Zhao, Y. Wang, L. Zhao, W. B. Ruan, Y. Chen, M. C. Zhang and D. Q. Yu, Heterogeneous Integration of Diamond-on-Chip-on-Glass Interposer for Efficient Thermal Management, IEEE Electron Device Letters 45, 448-451, 2024.
9. S. Vaziri, C. Perez, I. M. Datye, H. Kwon, C. F. Hsu, M. E. Chen, M. Noshin, T. Y. Lee, M. Asheghi, W. Y. Woon, E. Pop, K. E. Goodson, S. S. Liao and X. Bao, AlN: An Engineered Thermal Material for 3D Integrated Circuits, Advanced Functional Materials 35, 2025.
10. Q. R. Cai, D. Scullion, W. Gan, A. Falin, S. Y. Zhang, K. Watanabe, T. Taniguchi, Y. Chen, E. J. G. Santos and L. H. Li, High thermal conductivity of high-quality monolayer boron nitride and its thermal expansion, Science Advances 5, 2019.
11. N. J. N. B. Nazim, M. F. Abdullah, N. Soriadi, S. A. M. Badaruddin and M. R. M. Hussin, Electrothermal Analysis of CVD-Grown hBN Heat Spreader using Pt/Cu/Ti Micro-Coil, Journal of Electronic Materials 51, 4238-4247, 2022.
12. Hasanov, Simultaneous determination of source terms in a linear parabolic problem from the final overdetermination: weak solution approach, Journal of Mathematical Analysis and Applications, 2007, 330(2): 766-779.
13. Y. Wang, X. C. Luo, S. Li, Optimal control method of parabolic partial differential equations and its application to heat transfer model in continuous cast secondary cooling zone, Advance in Physical

Mathematical, 2015: 967-977.

14. Y. Yu, X. Luo, S. Li, Estimation of boundary condition of two-dimensional nonlinear PDE with application to continuous casting, Computers & Mathematics with Applications, 2020, 80(12): 3082-3097.

15. Y. Yu, X. Luo, Z. Wu, Estimation of the boundary condition of a 3D heat transfer equation using a modified hybrid conjugate gradient algorithm, Applied Mathematical Modelling, 102, 2022: 768-785.

16. S. Chaudhuri, R. K. Singh, Design and implementation of an automated secondary cooling system for the continuous casting of billets, ISA Transactions, 49(2010): 121-129.

17. L. Klimeš and J. Štětina, A rapid GPU-based heat transfer and solidification model for dynamic computer simulations of continuous steel casting, Journal of Materials Processing Technology, 226(2015): 1-14.

18. Y. Yu, X. Luo, Q. Zhang, Dynamic optimization method of secondary cooling water quantity in continuous casting based on three-dimensional transient nonlinear convective heat transfer equation, Applied Thermal Engineering, 160(2019): 113988.

19. Y. Wang, X. Luo, F. Zhang, S. Wang, GPU-based model predictive control for continuous casting spray cooling control system using particle swarm optimization, Control Engineering Practice, 84(2019): 349-364.