Carbice Reliable cooling expertise.™

The secret to increasing data center performance lies at the interface.

The future of data center performance depends on manufacturing innovation and cooling expertise at the interface.

Table of contents

PART 1: INTRODUCTION

The secret to increasing data center performance lies at the interface.

The future of data center performance depends on manufacturing innovation and cooling expertise at the interface.

The future of data center performance depends not just on compute power or cooling capacity—it hinges on what connects the two. As the demands of AI, high-performance computing (HPC), and energy-intensive applications continue to rise, data centers are confronting an overlooked bottleneck: the thermal interface.

While semiconductors become more powerful and cooling systems more sophisticated, the interface that links them remains stagnant. This small, often hidden layer—known as the thermal interface material (TIM)—plays a critical role in a cooling system's effectiveness. Yet for decades, innovation at the interface has been limited, fragmented, and undervalued across the electronics manufacturing ecosystem.

As data centers evolve to support the explosion of AI and ever-increasing power densities, thermal management is no longer just a design consideration—it's a strategic imperative. Power fluctuations, overheating, and inefficient cooling solutions aren't just technical nuisances; they drive unplanned downtime, inflate operational costs, and threaten long-term profitability. The question for data center operators isn't whether they'll face these challenges—it's how to solve them without compromising performance, sustainability, or serviceability.

This paper explores the hidden costs of legacy thermal interfaces and the systemic risks they pose to data center operators and IT suppliers.

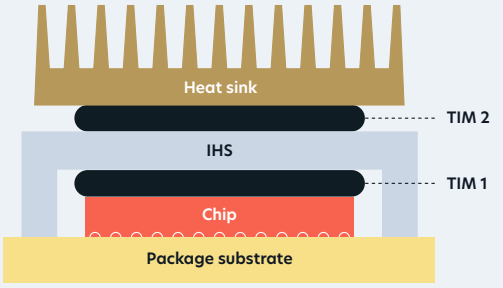
The importance of thermal interface materials in data centers

What is an interface material and why is it critical?

Thermal interface materials are critical for maintaining the thermal stability of electronic components by efficiently transferring heat from chips to heat sinks, cold plates, and other cooling devices. TIMs act as the bridge between heat-generating semiconductors and their cooling solutions, ensuring that the heat is effectively dissipated and preventing performance degradation. The device relies on the TIM to transfer heat to the cooling system at a higher temperature than the cooling medium—creating the temperature gradient that drives effective heat removal. In fact, the TIM contributes to more than 30% of the overall cooling efficacy[1], arguably making it one of the most critical components in any thermal system design. Without a reliable TIM, even the most advanced cooling technologies—whether air-cooled heat sinks or cutting-edge direct liquid cooling systems—would fail to perform as intended, leading to overheating, system throttling, and hardware degradation. In today's power-dense AI chips, poorly performing TIMs can result in up to 20°C of lost cooling potential between the chip and the cooling device—severely limiting power efficiency and compute performance in large data centers.

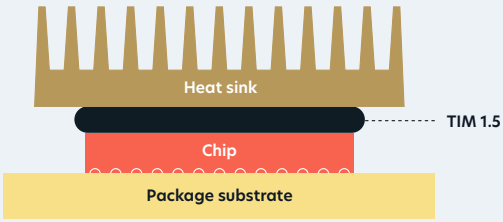
Modern data centers are pushing the boundaries of power density, with hyperscalers like Google, Amazon, and Meta deploying thousands of high-powered GPUs in massive clusters to support AI training, machine learning workloads, and cloud computing. To manage these extreme thermal loads, hyperscale data centers employ advanced cooling strategies such as chilled water cooling, direct-to-chip liquid cooling, and immersion cooling. These technologies are designed to remove vast amounts of heat efficiently, but their effectiveness is still fundamentally dependent on the performance of the TIM that sits between the processors and cooling elements. Without high-performance TIMs, even the most innovative cooling architectures cannot achieve their full potential, leading to costly inefficiencies, increased power consumption, and unexpected downtime events.

Key TIM categories in compute infrastructure:



TIM 1 – Between the chip and the lid

TIM 2 – Between the lid and the heat sink



TIM 1.5 Between the chip and the heat sink

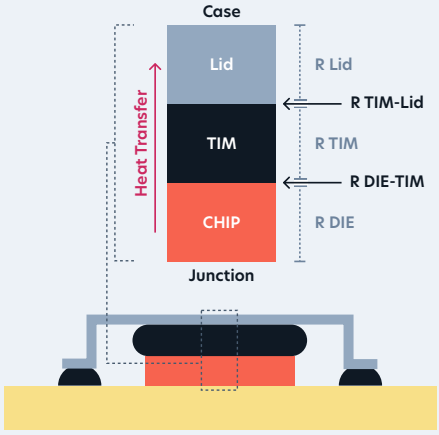


FIGURE 1: Thermal interface diagram.

Across an entire data center, there are thousands—if not millions—of TIM interfaces, all of which must function optimally to prevent thermal bottlenecks. As data centers continue to scale and deploy higher-wattage processors, ensuring the reliability, longevity, and serviceability of TIMs will be key to maintaining peak operational efficiency while minimizing costs and environmental impact.

Minimizing system risk through high-performance thermal interface design

In today's data centers and power electronics, thermal interface materials must do more than reduce temperatures at startup. As systems scale and power density increases, true high-performance is defined by how well they endure real-world operating stress—not just how they perform in early benchmarks. Yet system designers often optimize for low system steady-state temperature while giving less attention to more impactful system performance metrics driven by both dynamic thermal and mechanical stresses.

This focus on low steady-state temperature is often reinforced by the application of generalized models like the Arrhenius equation, which overemphasize steady-state temperature and overlook dynamic, real-world stressors. As Wilcoxon (2017) highlights, thermal cycling alone can reduce product lifetime by up to 8×—posing a much greater threat than modest increases in temperature[2]. Dasgupta (2024) further highlights how system degradation is driven by coupled mechanical, thermal, and material fatigue effects—none of which are captured in static models[3]. Establishing a new metric for true high-performance in electronics requires accounting for these more impactful thermal and mechanical failure mechanisms that conventional models and early benchmarks often miss.

To redefine what "true high-performance" means in thermal interface design, we must look beyond thermal conductivity alone. Many critical reliability issues—like fatigue, dry-out, pump-out, delamination, and contact stress relaxation—emerge over time due to a combination of thermal and mechanical stresses that aren't captured in conventional early-stage design and testing. These issues can show up in typical data center operations as soon as 3 months from initial startup.

Two industry recognized approaches, the Design Failure Mode and Effects Analysis (DFMEA) and Process FMEA (PFMEA) can be combined to define, analyze and address risks associated with thermal interface failure modes. While DFMEA highlights design-related vulnerabilities, PFMEA surfaces risks introduced during manufacturing and integration. On their own, each view is incomplete. Together, they offer a full-spectrum risk analysis that reveals the critical failure mechanisms in thermal interfaces that degrade performance, increase service costs, and reduce system life. True high-performance interfaces minimize these risks and allow systems to operate under their steady-state thermal limits.

A combined DFMEA/PFMEA framework identifies and ranks the risks associated with real-world failure modes in thermal interfaces. The chart below outlines common thermal interface

failure modes, the underlying mechanisms that cause them, and the corresponding design and process controls used to manage risk. Each failure mode is assessed using a combined DFMEA/PFMEA approach and scored from low to high risk—highlighting which issues are most severe, most likely to occur, and hardest to detect, and therefore require accelerated reliability stress testing to fully characterize.

Design and Process FMEA Risk Analysis

CATEGORY	FAILURE MODE	DRIVING MECHANISM	CURRENT DESIGN CONTROLS PREVENTION	RISK PRIORITY NUMBER (RPN)
Design	Contact Stress Relaxation	Mechanical creep	Thermal cycling, vibration, shock, high temperature storage, elevated humidity	High
Design	Interface Fatigue	Thermal-Mechanical gradients, cycling	Thermal cycling, vibration, shock, high temperature storage, elevated humidity	
Design	Grease Dry-out / Pump-out	Thermal cycling, viscosity changes	Thermal & mechanical cycling, vibration, shock, high temperature storage, elevated humidity	
Design	Polymers Stiffening	Elevated temps, time	Data sheet review, high temperature storage	
Process	Manufacturing Assembly Variance	Variance in dimensional tolerance	QA & visual inspection, final test	Medium
Process	Mechanical Damage During Rework	Stress concentrations from warped parts or misalignment	Visual inspection, rework procedure controls, functional testing	
Process	TIM Cleanliness	Poor quality control, environment, contamination causes scrapped parts	Visual inspection, process controls and isolation, downstream process detection	
Process	Vibration Contact Loss	Shipping/transportation, harsh environments	Shock, vibration, impact & drop testing	
Design	Metallic Connector Corrosion	Vibration, humidity, temperature cycles	High temperature, elevated humidity, salt fog	
Design	Delamination	Thermal-mechanical cycling	Thermal cycling, vibration, shock, high temperature storage, elevated humidity	Low
Process	Improper TIM application	Different application methods	Visual inspection	
Design	Electrostatic Discharge	Leakage, FOD	Data sheet review, dielectric materials, visual inspection and functional testing	
Design	Thickness out of Range	Application-defined	Engineering process development	
Design	Device Throttling	Insufficient variable or transient cooling	Initial system design, final test	
Design	Bulk Temperature Overload	Insufficient steady state cooling	Initial system design, final test	

RPN SCORE High >250 Medium 150-250 Low 1-149

FIGURE 2: FMEA Risk Priority Analysis for High-performance Thermal Interface Solutions

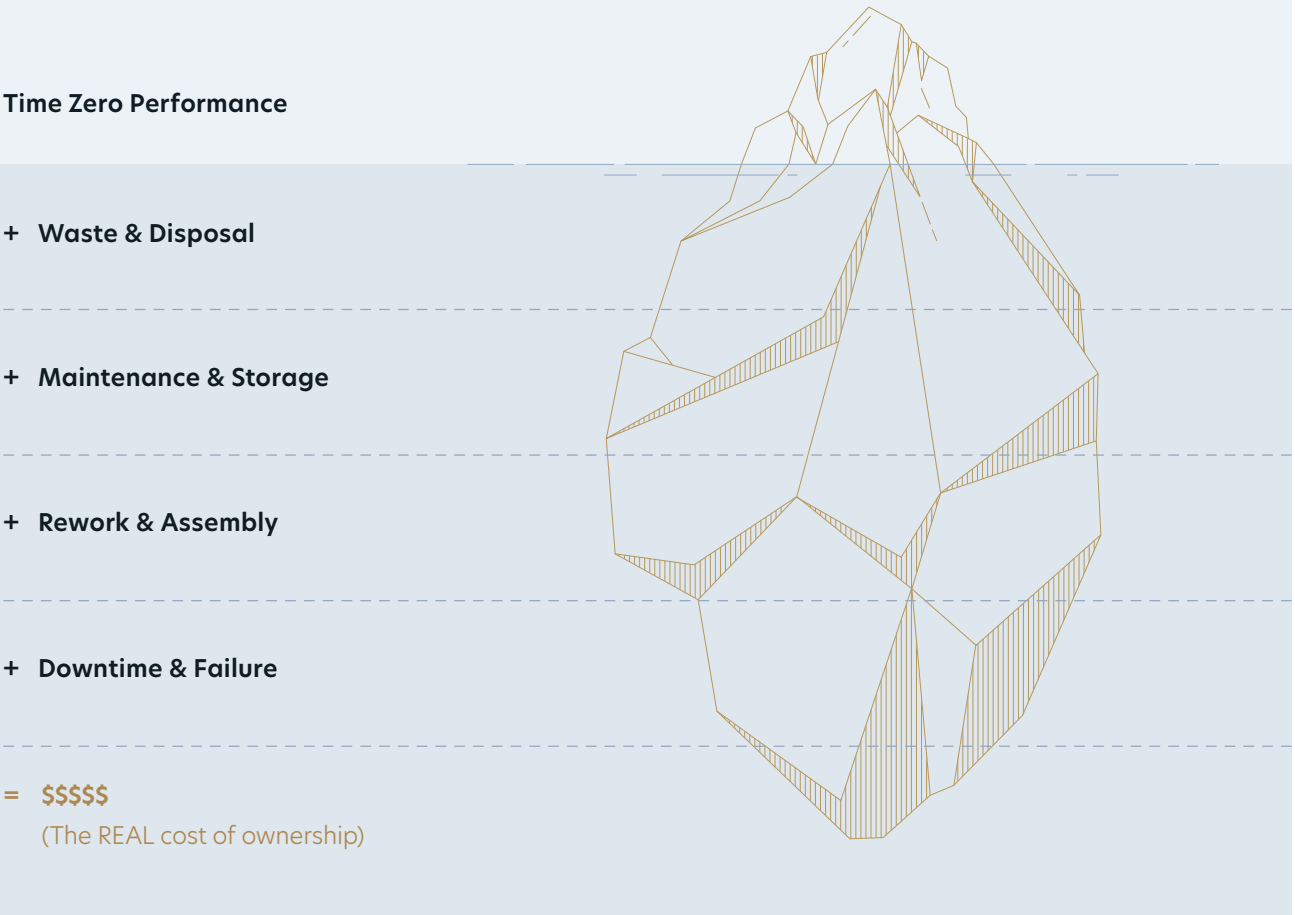
Data center operators, electronics manufacturers, and systems integrators who don't evaluate these key risk factors are likely missing the design and process controls necessary to achieve consistent, reliable, and long-term high-performance across their infrastructure.

Reducing data center costs through optimized thermal solutions

At first glance, ensuring desired temperature results at initial installation (Time Zero) might seem like the biggest challenge—solved in a laboratory evaluation with the right thermal interface material. However, as with an iceberg, the most significant costs of inadequate thermal solutions lie beneath the surface. We know this is true from just a brief search of "re-pasting" on the internet where gamers are constantly spending time, energy, and money adjusting for poor temperature stability in their systems.

FIGURE 3:

Legacy TIMs: The Iceberg of Hidden Costs



Beyond the initial performance checks, unpredictable and unreliable TIMs can trigger downtime and excessive maintenance when products leave the factory, adding layers of operational inefficiencies and financial strain. Data centers rely on uninterrupted performance, yet many are unknowingly burdened by hidden costs tied to equipment recalls during initial build outs, rework, labor-intensive maintenance, waste and premature component wear and tear—all stemming from suboptimal thermal management. By adopting a serviceable, recyclable TIM that minimizes rework, simplifies assembly, is robust in shipping and storage, and eliminates wasteful disposal practices, data centers can significantly lower their long-term total cost of ownership while also reducing environmental impact.

Thermal challenges in data centers cause downtime

At the heart of every data center is an array of high-density servers, each housing thousands of GPUs, CPUs, memory modules, and networking components working together to process vast amounts of data. As AI workloads and high-performance computing (HPC) become more prevalent, these servers are pushing the limits of thermal design. The demand for higher power densities means components must dissipate more heat than ever before, making reliable heat transfer a critical bottleneck.

Inside a single compute rack, there can be hundreds of TIM interfaces, all of which must function optimally to avoid thermal throttling, increased fan speeds, higher energy draw, and even hardware failures. When TIMs degrade over time or fail to maintain uniform contact, hotspots develop, forcing servers to operate inefficiently. This leads to throttled performance, higher cooling loads and electrical power draws, and an increased risk of downtime—all of which translates to greater operational costs and a lower return on investment.

Downtime is a major concern for data centers, not only due to maintenance and repair costs but also the risk of lost revenue. Despite millions spent on redundancy, Uptime Institute reports that 45% of operators face outages costing $100,000 to $1 million per event[4]. Therefore, identifying all potential failure points, such as TIM degradation, is crucial to protecting productivity and profitability.

In an environment where uptime, efficiency, and sustainability are paramount, TIM performance directly influences the overall profitability and longevity of a data center.

Inconsistent TIM performance creates ripple effects throughout a data center:

- **Performance loss:** Thermal throttling reduces compute efficiency, forcing workloads to be distributed inefficiently across nodes and complex computations to be interrupted.
- **Higher energy consumption:** Fans, liquid cooling systems, and overall power draw must increase to compensate for thermal inefficiencies.
- **Serviceability challenges:** Any issue with the server components, for any reason, often requires the detachment of the heat sink and removal and reapplication of the TIM, leading to downtime and increased labor costs.
- **Premature hardware degradation:** Cyclic heat stress natural to regular operation leads to accelerated wear on semiconductor packaging interfaces, reducing their lifespan and increasing replacement costs.

Durable TIMs are crucial to reliable power operations

Beyond compute nodes, the power infrastructure of a data center also faces significant thermal challenges. Power modules, such as Insulated Gate Bipolar Transistors (IGBT) power modules and other semiconductor devices within UPS (Uninterruptible Power Supply) systems and inverter-based power infrastructure, are subjected to extreme electrical and thermal loads.

These components must not only handle high-voltage switching and fast transients but also maintain consistent performance under fluctuating power demands—whether from critical backup power applications or the surging energy needs of AI-driven workloads. As data centers and industrial power systems push higher power densities and efficiency standards, these power modules must operate with exceptional reliability, enduring intense heat cycles and mechanical stresses without degradation.

Challenges of TIMs in power infrastructure:

Electrical and thermal cycling degradation:
Liquid TIMs dry out or pump out over time, leading to increasing resistance and heat buildup.

Increased power losses:
Poor heat transfer leads to inefficiencies in power conversion, causing additional stress on cooling systems.

Service disruptions:
Maintenance and replacement of power modules become time-intensive and costly, especially when entire systems need to be taken offline for rework.

Sources

1. **B. Ramakrishnan et al.,**
   "CPU Overclocking: A Performance Assessment of Air, Cold Plates, and Two-Phase Immersion Cooling," in IEEE Transactions on Components, Packaging and Manufacturing Technology, vol. 11, no. 10, pp. 1703-1715, Oct. 2021, doi: 10.1109/TCPMT.2021.3106026.

2. **Wilcoxon, R.**
   (2017, August). Does a 10°C Increase in Temperature Really Reduce the Life of Electronics by Half? Electronics Cooling. Retrieved from https://www.electronics-cooling.com/2017/08/10c-increase-temperature-really-reduce-life-electronics-half/

3. **Dasgupta, A.**
   (2024, July). How Many Ways Can Temperature Affect Performance and Reliability of Electronic Systems? Electronics Cooling. Retrieved from https://www.electronics-cooling.com/2024/07/how-many-ways-can-temperature-affect-performance-and-reliability-of-electronic-systems/

4. **Uptime Institute.**
   (2023). Annual Outage Analysis 2023. Uptime Institute. Retrieved from https://datacenter.uptimeinstitute.com/rs/711-RIA-145/images/AnnualOutageAnalysis2023.03092023.pdf

Interested in seeing how Carbice enables true high performance and reliability in data centers?

Read Part 2. Download the full white paper at Carbice.com/data-center-performance.

Carbice Reliable cooling expertise.™

Get in touch: Carbice Corporation | Carbice.com

WP-CAR2025.01