

# Evolution of Thermal Management for AI Data Centers

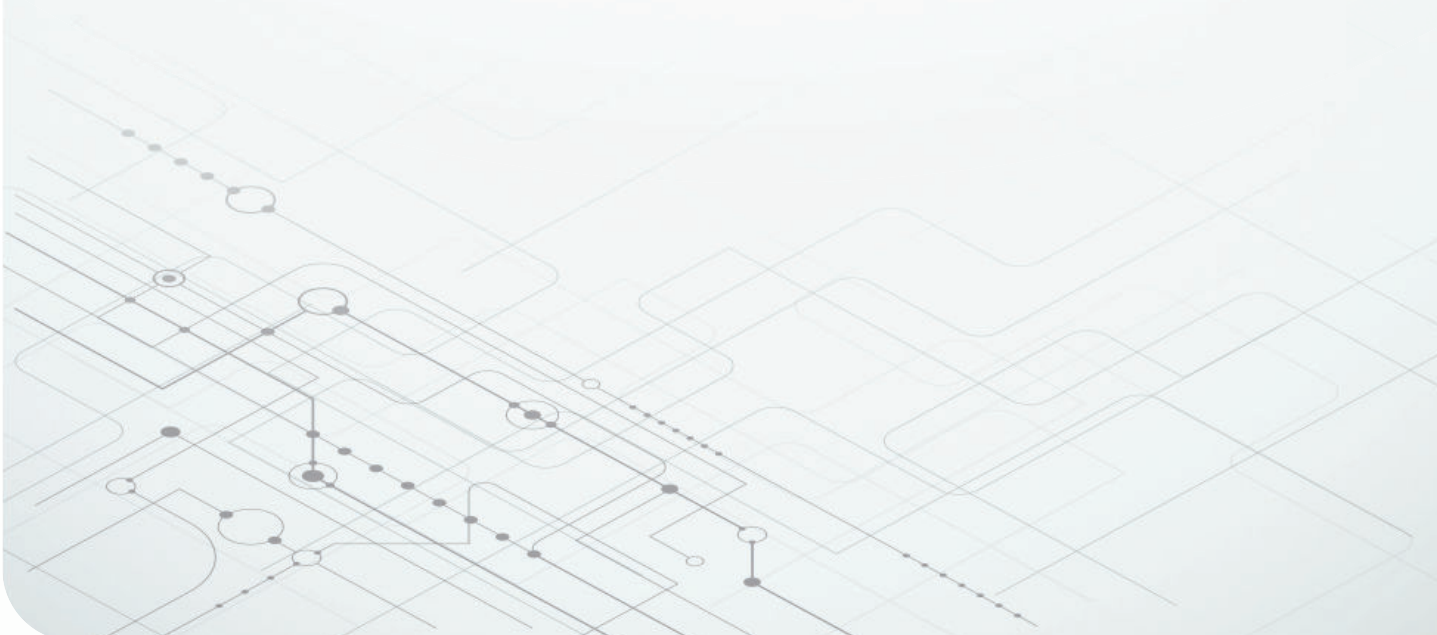
---



**David Stout**

Systems engineer

Building automation - Data center thermal management





## At a glance

- **Rising heat loads:** The rack power density in artificial intelligence (AI) data centers has climbed from about 5kW to more than 100kW, with some AI clusters expected to reach 1.5MW.
- **Cooling architecture evolution:** As racks exceed 100kW, data centers are moving from air-based cooling to liquid cooling and immersion cooling.
- **Predictive thermal management:** Server-tray-level liquid flow control and edge AI-based sensor fusion will help data centers detect leaks faster and predict failures.

## Introduction

Power levels for compute racks in data centers have increased over the last several years, driven by the more advanced central processing unit (CPU) and graphics processing unit (GPU) architectures that large language models and other AI workloads require. As a result, the importance of thermal management has grown as well.

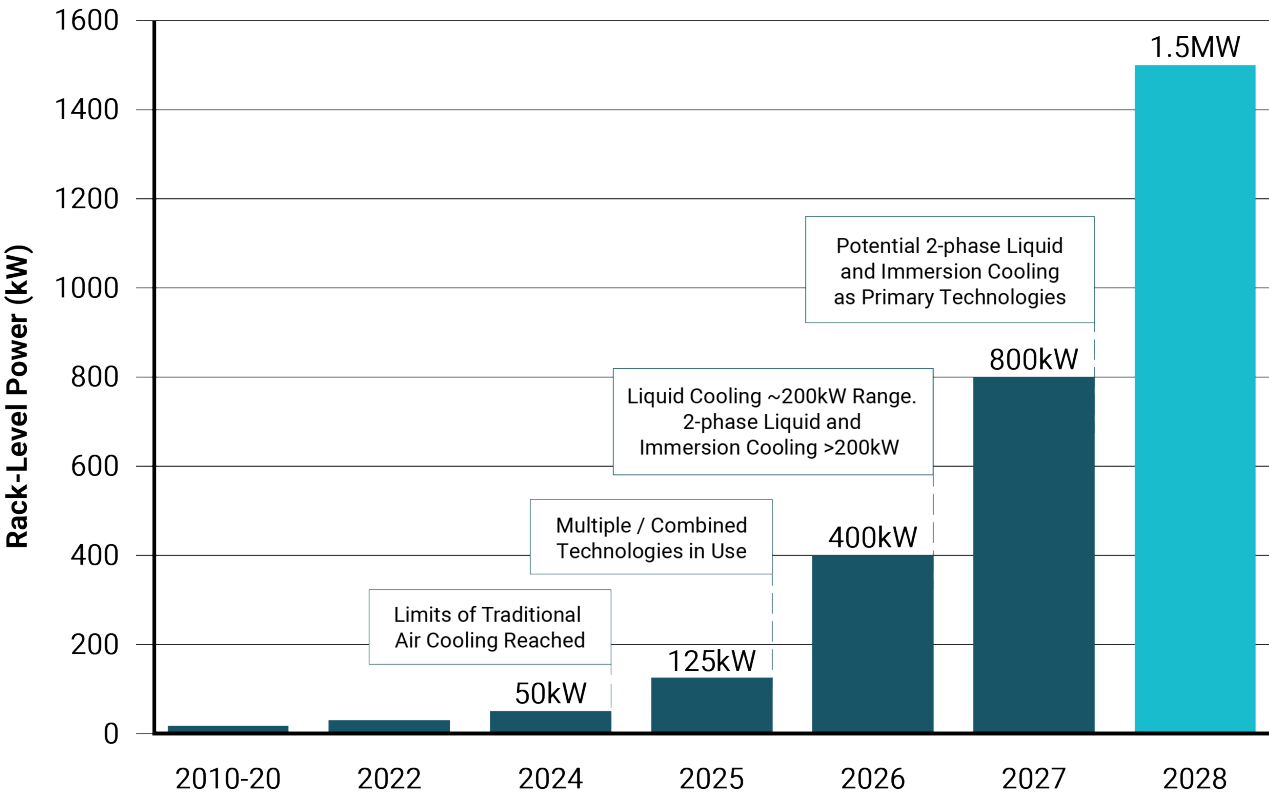
The main job of any **thermal management system** is simple in concept: absorb heat from areas where it isn't wanted and transfer it where it can be dispersed safely. As compute racks generate more heat, keeping the processors inside those racks operating within an acceptable temperature range requires different thermal management techniques to optimize performance. This white paper provides an overview of how thermal management systems have evolved as power levels have increased and looks at the thermal management systems that may be needed to handle the power levels of the not-so-distant future.

**Related reading:** "[Data centers are generating more heat than ever. Cooling has to keep up.](#)"

One way to visualize the power levels of compute racks, and the amount of heat they generate, is to compare a rack to a traditional wood- or gas-burning fireplace found in homes. These fireplaces typically put out the equivalent of about 10kW of heat in roughly the same footprint as a compute rack, a load any typical air conditioning system can handle

without issue. At a 50kW rack power level, though, that's 10 fireplaces packed into that footprint, and air cooling is no longer capable of regulating the room temperature with that amount of heat load. A 1.5MW rack, the equivalent of 150 fireplaces in that space, really puts the challenge of data center thermal management into perspective. Fortunately, liquid cooling offers a heat-absorption capability that's 3,000 times greater than air for the same volume.

**Figure 1** illustrates how quickly the challenge of cooling has escalated, plotting rack power levels of the past, present and future against the architectures capable of handling them, and where each architecture reaches its limit. Rack power density in AI data centers has climbed from roughly 5kW to more than 100kW. Some AI clusters now approach 200kW per rack, with projections in the next few years to levels as high as 1.5MW. That's why cooling architectures are shifting from air cooling to liquid cooling to keep pace



**Figure 1.** Rack-level power projections and cooling technologies

**Comparing today's thermal management architectures**

While liquid cooling represents the future of thermal management, mixed technologies will continue to serve the industry for years. Engineers typically reserve the most advanced liquid cooling systems for racks dedicated to AI computing, while other architectures handle standard workloads.

There are three main air-based cooling architectures in data centers, listed here in order of increasing rack power density capability. Each architecture describes the medium used to absorb heat generated in the compute rack, followed by the medium used to transfer heat out of the data center.

## Air-to-air

An air-to-air cooling architecture operates much like the central air system in your home, except that in a data center, cool air circulates through compute racks to absorb generated heat. Hot air then passes over coils containing cold refrigerant that absorbs the heat and cools the air for recirculation. The refrigerant transfers the absorbed heat outdoors, where it's released to the environment and re-cooled through compression, condensing and evaporation. This process can be summarized as a cycle of heat absorption, heat transfer, and heat release and cooling. To control airflow, racks are typically mounted on raised floors and arranged in back-to-back rows to form cold aisles and hot aisles.

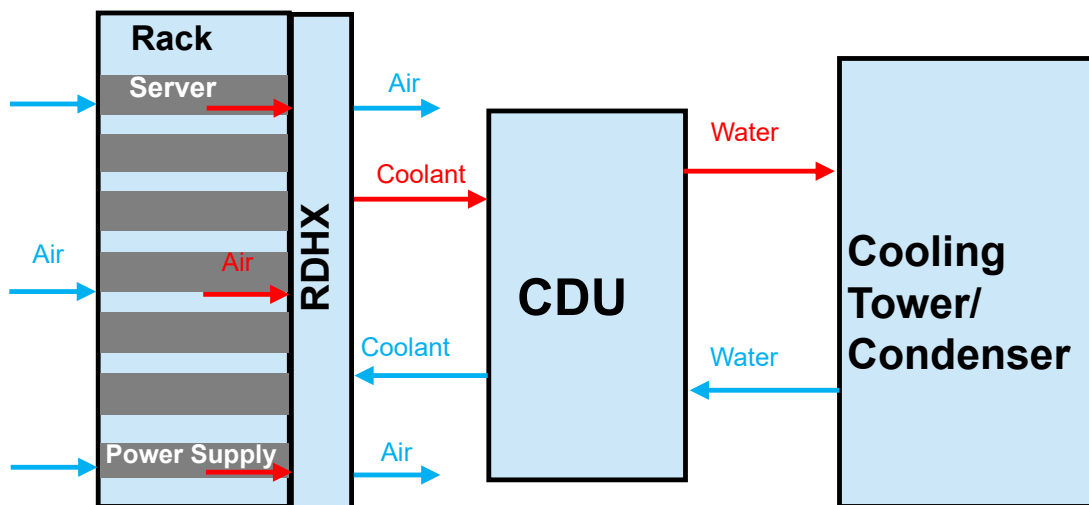
As power levels rise, hot aisle containment prevents hot air from bleeding into and raising cold-aisle temperatures controlled by computer room AC units (CRACs). Computer room air handlers (CRAHs) manage airflow and the transfer of heat to refrigerant in the evaporator coils. Increased heat load requires larger blower fans and evaporator coils, along with increased server fan airflow. In the 20kW to 50kW rack power range (assuming that all racks in the computer room are similar), efficiency and operational demands require a different heat transfer method.

## Air-to-liquid

An air-to-liquid cooling architecture uses the same method as the air-to-air cooling architecture but adds a rear door heat exchanger (RDHX), shown in **Figure 2**. The purpose of the RDHX is to transfer heat from the air that has moved through the compute rack to a cold liquid while also cooling the air as it passes through. The RDHX can be passive or active, meaning the absence or presence of additional fans to aid airflow through the exchange.

A coolant distribution unit (CDU) circulates the liquid that has absorbed heat from the RDHXs and moves it outdoors using pumps where cooling towers, dry coolers or other means are used to release heat to the environment and cool the liquid for recirculation. The CDU typically includes a heat exchanger that isolates the liquid flowing through a computer room from liquid flowing outdoors (generally, a facility's water supply). This isolation allows the use of different types of liquid to be used in the compute room versus what is used outdoors.

As power levels continue to increase, using air as the heat absorbing medium becomes a bottleneck, prompting a transition toward cold plates that enable the absorption of heat directly from the components generating it.

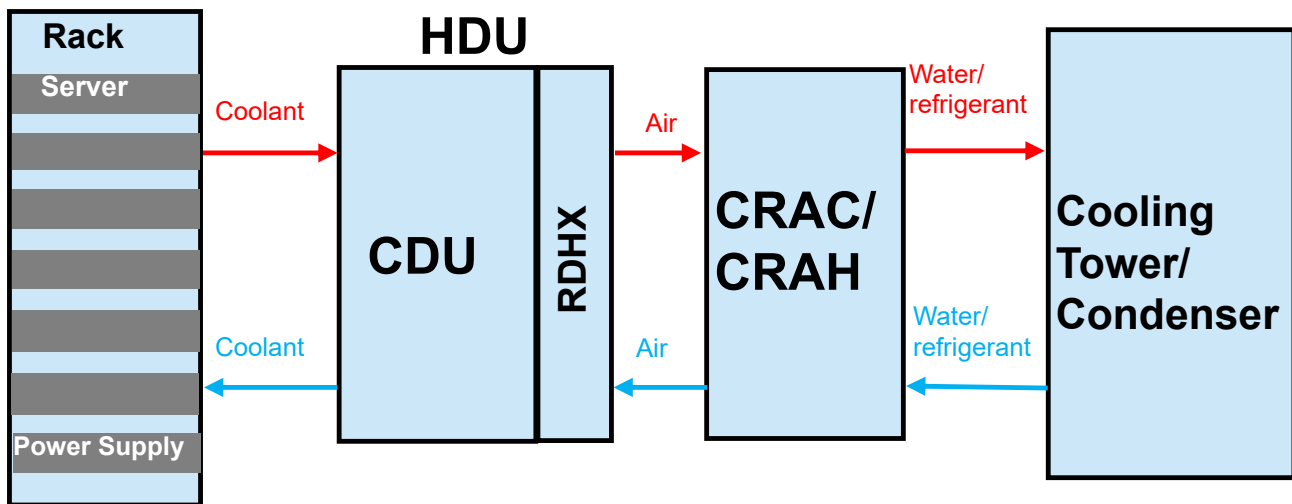


**Figure 2.** Air-to-liquid cooling architecture

## Liquid-to-air

Because cold plates greatly improve heat absorption, the airflow needed to transfer heat is reduced, enabling traditional CRAC/CRAH units to continue transferring heat through air. A liquid-to-air architecture, shown in [Figure 3](#), employs a heat dissipation unit (HDU) that is a combination of a CDU and an active RDHX. In an HDU, the CDU circulates liquid through cold plates, which are mounted against the sources of heat throughout the rack for heat absorption as close as possible to where the heat is generated. The liquid flows through the active RDHX, where fans draw cool computer room air across the coils to transfer the heat to air that flows into the hot aisle. The HDU can reside in-rack, in-row or as a sidecar, depending on power levels.

As power levels climb into the 150kW to 200kW range, using air as a heat transfer method again becomes a bottleneck, leading to the liquid-to-liquid architecture becoming prevalent today.

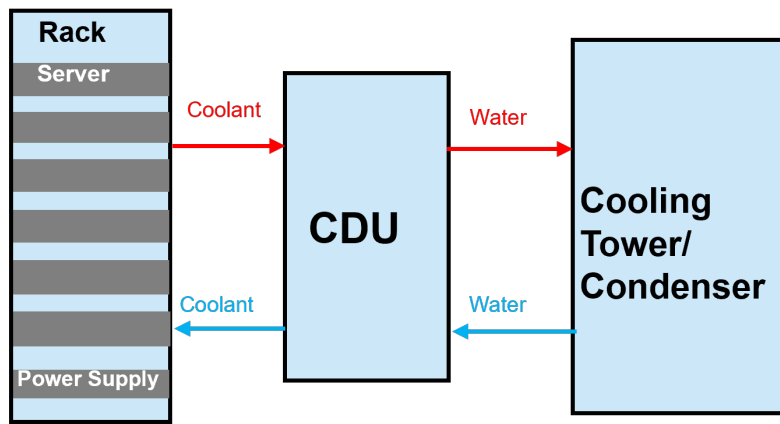


**Figure 3.** Liquid-to-air cooling architecture

Beyond thermal performance, engineers must also consider facility requirements. The air-to-liquid architecture requires a facility water infrastructure and significant plumbing throughout the white space, while the liquid-to-air architecture works in facilities without existing water systems, making it an attractive option for retrofitting older data centers. Together, the air-to-liquid, liquid-to-air and liquid-to-liquid architectures account for the majority of thermal management systems in use today.

## Evolution of liquid cooling technology

The liquid-to-liquid architecture, shown in [Figure 4](#), uses direct-to-chip cold plates to absorb heat at the source and coolant distribution units (CDUs) to circulate fluid and transfer heat from one liquid to another and from the white space to the outdoors, where it's released.



**Figure 4.** Liquid-to-liquid cooling architecture

The CDU is the most important component in the liquid-to-liquid architecture, keeping fluid flowing so heat is absorbed effectively, and components stay within their intended temperature range. The heat exchanger inside the CDU is equally vital, isolating the low-impurity, high-specific-heat liquid that absorbs heat from the cold plates from the facility water used in the outdoor cooling systems.

### The need for leak sensing

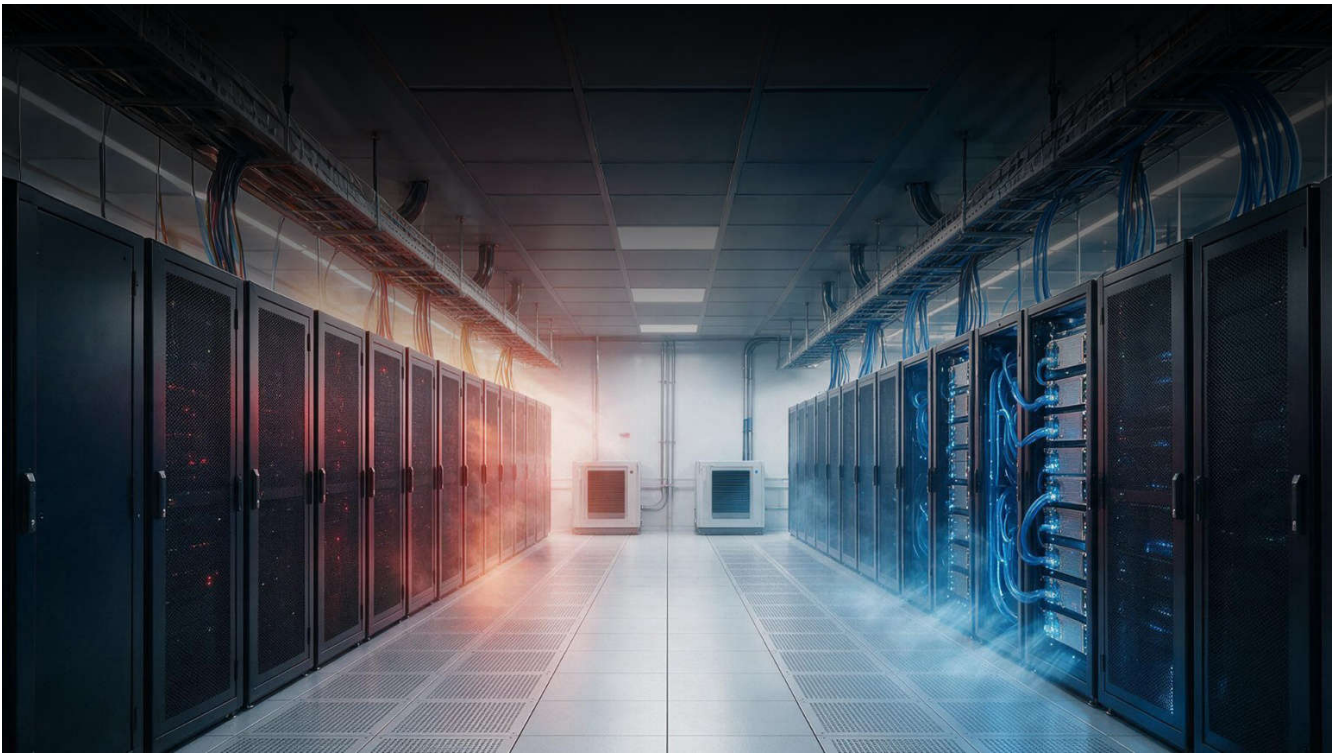
Introducing liquid so close to electronic components makes leak sensing extremely important. Detecting leaks reliably, with few false triggers, is as important as detecting them at all, since server uptime depends on it. Today's rack installations use on-and-off valves that stop liquid from flowing when a sensor detects a leak, shutting down the entire rack until technicians can find and fix the source. Moisture-sensitive resistive ropes are the most common types of leak sensors, but they're notorious for accuracy problems, and installation requires care to avoid pinches and sharp bends. Engineers are researching alternatives, including humidity sensors that detect micro-leaks and capacitive leak sensors.

Other approaches address leaks at the CDU level. One method reconfigures the CDU to pull fluid through the system under negative pressure instead of pushing it. This has the effect of drawing in air, rather than forcing fluid to escape through a puncture or loose connection. Another approach uses non-conductive liquid so that a leak just creates a mess, rather than a conductive short on expensive equipment.

### Immersion cooling

Immersion cooling further simplifies the cooling process and meets the needs of the higher rack power levels experts predict. This architecture orients the traditional rack horizontally and submerges it in a tub of non-conductive fluid (not including the metal frame), eliminating downtime from leaks altogether. Immersion tanks typically integrate the heat exchanger and smaller pumps to keep fluid circulating, making the architecture highly area-efficient, since it eliminates the manifolds and plumbing found on the sides of traditional racks.

With the pumps and heat exchanger built in, the CDU function becomes self-contained. Immersion cooling absorbs heat roughly two to three times more effectively than direct-to-chip cooling, enabling data centers in colder climates where facility water inherently runs cold, to potentially eliminate external cooling towers for moderate rack power levels.



## Future AI data center cooling trends

Cooling systems built for power levels up to 1.5MW will almost certainly rely on the liquid-to-liquid and immersion cooling architectures. Both techniques can convert to two-phase cooling, where low-temperature evaporation of a fluorocarbon dielectric, such as a refrigerant, improves heat absorption and transfer efficiency, extending the useful lifespan for these architectures well into the future.

Two other trends aim to keep pace with rising power densities and make control systems more responsive:

- Controlling liquid flow at the server tray level enables more granular control of GPU and CPU die temperatures and helps maximize server uptime elsewhere in the rack when leaks occur.
- Edge AI computing offers faster response times and lower latency by monitoring and processing all system sensors together via sensor fusion, catching subtle changes that the monitoring of individual sensors or parameters would miss. Edge AI computing opens the door to a range of possibilities, including improving the reliability and speed of leak detection and predictive maintenance for vital mechanical components.

Together, these trends give data center operators a clear view of emerging issues well before they become failures, pointing toward cooling systems that sense conditions, predict issues, and adjust automatically. Thermal management becomes an active part of how a data center runs, rather than something addressed after the fact.

## Conclusion

Cooling architectures share many of the same core components. Temperature sensors play an important role in controlling temperature. Motor control is equally important, powering the fans, pumps, compressors and electronic valves found throughout these systems. Microcontrollers and microprocessors keep everything working together. These are only a few examples of the semiconductors that cooling systems need, and TI is well positioned to provide

analog technologies and embedded processing products that address every layer of the cooling system, regardless of architecture. See the Additional Resources section for details on the specific product areas used in these cooling architectures.

## About the author

David Stout is a systems engineer focused on building automation systems, including data center thermal management at Texas Instruments. He has more than 30 years of experience designing analog, mixed-signal and radio-frequency circuits and systems for the industrial semiconductor market. David has a bachelor's degree in electrical engineering from Louisiana State University and a master's degree in electrical engineering from the University of Texas at Dallas.

## Additional resources

Explore related TI product categories relevant to data center thermal management systems:

- Check out [GaN power stages for inverters and variable frequency drives](#) and [motor drivers](#) for fans, pumps, compressors and electronic valves.
- Discover [temperature and other sensors](#) and [precision analog-to-digital converters](#) for leak detection and thermal monitoring.
- Learn more about [microcontrollers and microprocessors](#) that keep cooling subsystems coordinated and responsive.

**Important Notice:** The products and services of Texas Instruments Incorporated and its subsidiaries described herein are sold subject to TI's standard terms and conditions of sale. Customers are advised to obtain the most current and complete information about TI products and services before placing orders. TI assumes no liability for applications assistance, customer's applications or product designs, software performance, or infringement of patents. The publication of information regarding any other company's products or services does not constitute TI's approval, warranty or endorsement thereof.

All trademarks are the property of their respective owners.

## IMPORTANT NOTICE AND DISCLAIMER

TI PROVIDES TECHNICAL AND RELIABILITY DATA (INCLUDING DATASHEETS), DESIGN RESOURCES (INCLUDING REFERENCE DESIGNS), APPLICATION OR OTHER DESIGN ADVICE, WEB TOOLS, SAFETY INFORMATION, AND OTHER RESOURCES "AS IS" AND WITH ALL FAULTS, AND DISCLAIMS ALL WARRANTIES, EXPRESS AND IMPLIED, INCLUDING WITHOUT LIMITATION ANY IMPLIED WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE OR NON-INFRINGEMENT OF THIRD PARTY INTELLECTUAL PROPERTY RIGHTS.

These resources are intended for skilled developers designing with TI products. You are solely responsible for (1) selecting the appropriate TI products for your application, (2) designing, validating and testing your application, and (3) ensuring your application meets applicable standards, and any other safety, security, regulatory or other requirements.

These resources are subject to change without notice. TI grants you permission to use these resources only for development of an application that uses the TI products described in the resource. Other reproduction and display of these resources is prohibited. No license is granted to any other TI intellectual property right or to any third party intellectual property right. TI disclaims responsibility for, and you fully indemnify TI and its representatives against any claims, damages, costs, losses, and liabilities arising out of your use of these resources.

TI's products are provided subject to [TI's Terms of Sale](#), [TI's General Quality Guidelines](#), or other applicable terms available either on [ti.com](https://www.ti.com) or provided in conjunction with such TI products. TI's provision of these resources does not expand or otherwise alter TI's applicable warranties or warranty disclaimers for TI products. Unless TI explicitly designates a product as custom or customer-specified, TI products are standard, catalog, general purpose devices.

TI objects to and rejects any additional or different terms you may propose.

Copyright © 2026, Texas Instruments Incorporated

Last updated 10/2025