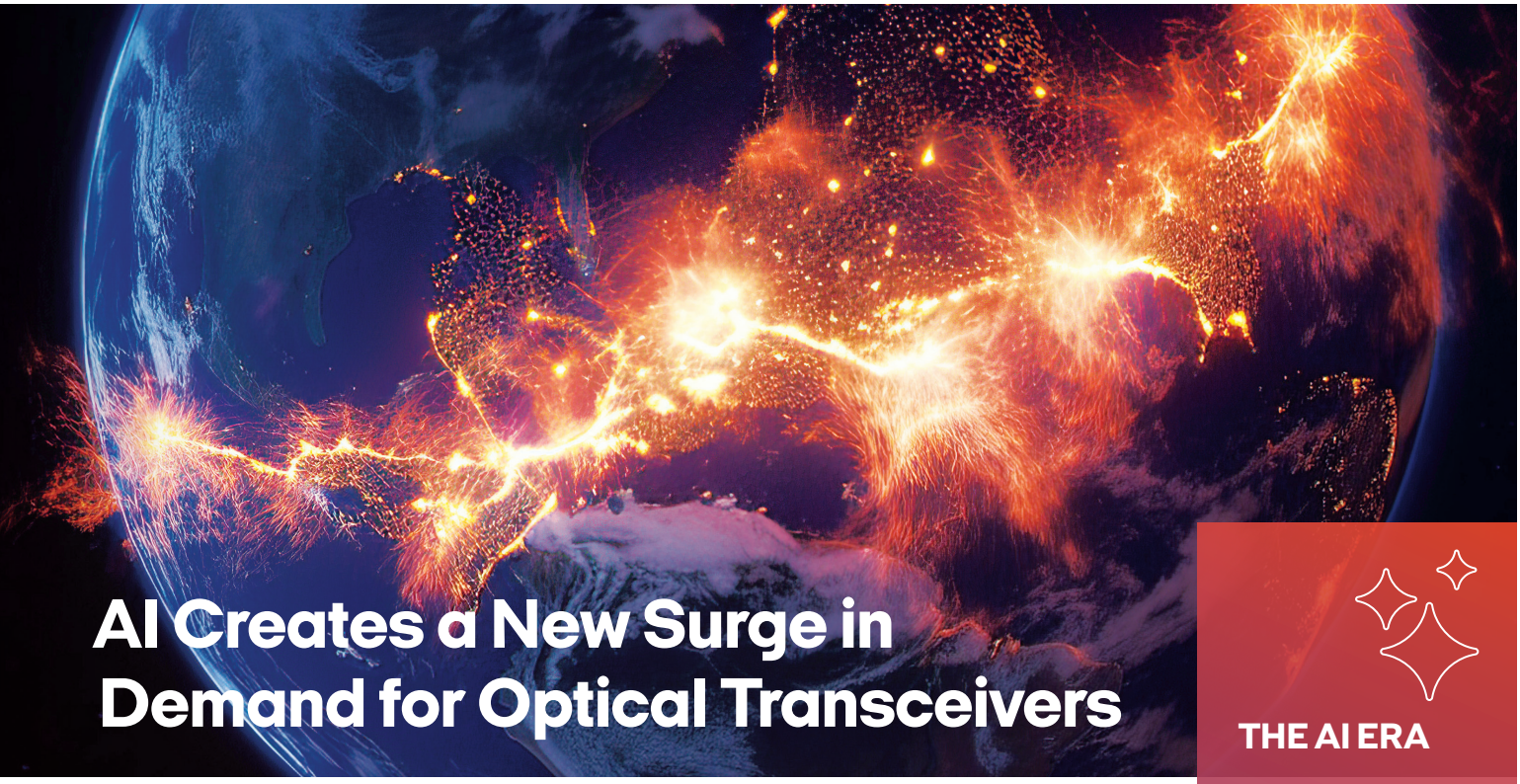
PHONONIC

AI & the Rise in Data Demands:

The Need for Cooling in Next-Gen Transceivers

TABLE OF CONTENTS



AI Creates a New Surge in Demand for Optical Transceivers

Data rate demands have skyrocketed with the growth of cloud computing for streaming movies, television and video games. Add to that the surge in remote work's use of meetings taking place over video chat platforms, and the toll on network bandwidth providers was massive. Smart planning and technology investments brought us through, but a recent technological innovation is threatening to push data center networks to a breaking point. We are entering the era of Artificial Intelligence (AI), and unless network providers start proactively transitioning to cooled optical communications, they will not be able to keep up with the high data rate demands — and that means the loss of customers — and revenue.

AI's overarching goal is to assist and improve the way humans interact with the world. While it has the potential to revolutionize the way we live our lives, it also comes with an unprecedented demand for bandwidth when compared to general computing. And as AI develops further, it will drive a faster upgrade cycle in optical transceivers. Vast volumes of data are required to train AI models, which in turn will drive a meteoric rise in data rate needs. To keep pace with AI development, data center operators are going to have to invest in powerful, high-bandwidth optical transceivers to replace electrical interconnects. In addition, AI use cases will necessitate bandwidth upgrades beyond data centers.

THE AI ERA

To keep pace with AI development, data center operators are going to have to invest in powerful, high-bandwidth optical transceivers. These next-generation components require the active cooling of TECs to optimize performance and minimize cost.

The current generation of AI chips/ GPUs use 2x transceivers per GPU, while the next generation will use 10x transceivers per GPU.

The total addressable market is expected to grow five times over the next five years. And with AI chip/GPU upgrades typically taking two years (based on past GPU launches), faster upgrades in transceiver bandwidth are imperative.

The transceiver market has already seen demand starting for 800G modules for AI applications. There's a strong ramp up for 800G in 2024, and the adoption of 1.6T transceivers is expected to arrive in 2025 alongside the new generation AI chips/GPUs. With next-gen transceiver configuration of 16, 100G lanes or 8, 200G lanes, data centers can buy fewer transceivers that have greater bandwidth over long distances that pull less power.

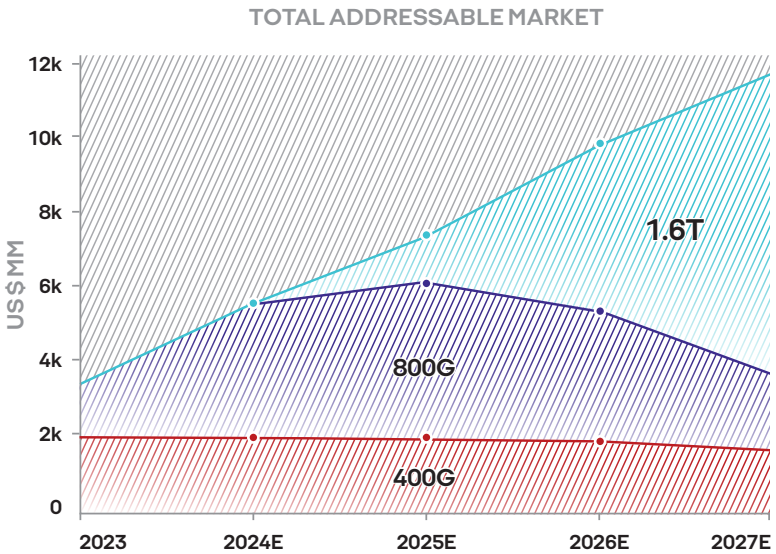
What's more, the move to upgrading optical transceivers comes with a financial incentive. Cost per lane and power consumption (on a unit-capacity basis) will drop when migrating to a larger bandwidth. Data center operators will need fewer 1.6T transceivers than 800G transceivers to deliver the same results over the same distances. This will enable networks to handle intense computing applications like AI training.

When planning infrastructure upgrades, the return on investment for high-performance transceivers will be substantial.

AI Demands High Speed Transceivers

1.6T

and Beyond Demands Cooling



Source: Goldman Sachs Global Investment Research

PHONONIC

Expected global spend to prepare for the data demands of AI:

$2,000,000,000,000

Adoption and infrastructure upgrades are needed — and inevitable. The top hyperscale data center operators are all at ~3kk pluggable transceivers as of 2024. Even Microsoft, who historically uses fewer pluggable optics, has recently purchased a significant number of Nvidia systems; in fact, Microsoft is planning an annual spend of more than $50 billion for 2024 and beyond, the purpose of which is aimed squarely at accelerating the path to AGI and bringing the intelligence of generative AI to every facet of life (per SemiAnalysis). And recently, Jensen Huang, CEO of Nvidia, has predicted that the global spend on data centers will double over the next four to five years to two trillion dollars in preparation for the data demands of AI.

Phononic has been at the forefront of this transformational technology for years, and we have proactively been preparing for an across-the-board shift to high-bandwidth optical transceivers. We've been steadily making significant investments in manufacturing, distribution and technology to prepare for this data revolution that will require TECs to advance. We're currently designing and mass producing TECs for 400G transceivers, and Phononic TECs for 800G and 1.6T transceivers are already in development to ensure our clients are ready for the upcoming demand. Our custom designed TECs will ensure that teams can achieve the high performance and efficiency needed for their applications.

Source: www.datacenterdynamics.com - February 14, 2024

To stay ahead of AI's data demands, you have to meet current demands while keeping an eye on where the market is going. By partnering with Phononic, you're investing in a solution that meets today's market needs while setting yourself up for future success.

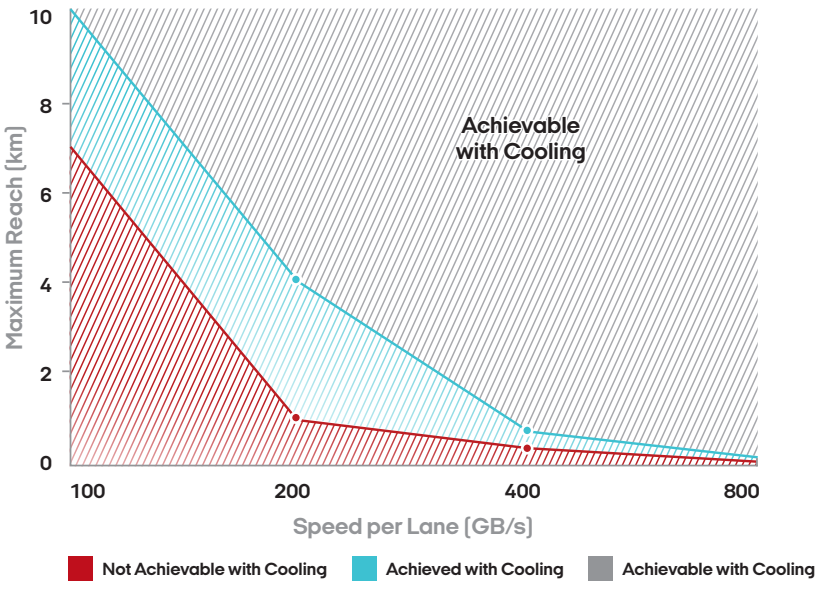
Cooling is the Future of Fiber Optical Communications

As the data rate demand continues to increase, so too does the need for extremely high performance. And while there was a time when lasers in transceivers would work within acceptable standards for short distances, that time is quickly passing. There is a clear need for an extended range, which requires higher power and thereby generates heat and more data per clock event (in which compressed data requires very accurate wavelength control in both DWDM and coherent modulation), which is highly affected by heat. This is where inexpensive, efficient thermoelectric coolers, or TECs, come into play.

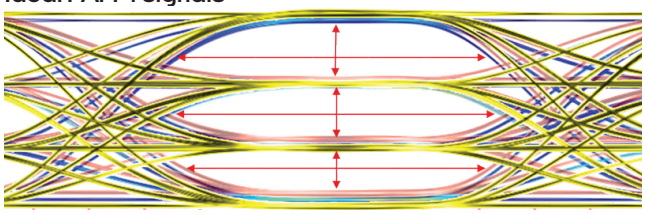
It has been well established that current, cutting-edge and next-generation optical components require the active cooling capability of TECs — this cooling is what ensures wavelength control and maintains modulation frequency specifications. What's more, cooling is necessary in order to preserve signal integrity to achieve 200G+ per lane. As the modulation rate moves beyond 100G per lane, higher-order aptitude modulation schemes must be taken into account, such as pulse-amplitude modulation 4 (PAM-4) or even coherent transmission.

The fact is, the entire fiber optic cable industry is quickly approaching a tipping point. The need to go faster and farther inside the data center is growing, but the maximum distance you can go without cooling gets shorter every time you turn up the speed per lane. Many intra-data center applications already require cooled optics – even for short distances. It is crucial that any new communication designs account for the ability to scale on demand without needing any redesigns or new products.

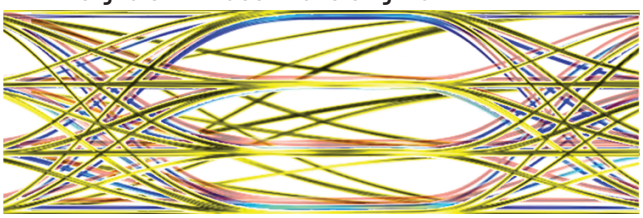
Recent presentations given by data center optics experts demonstrate that uncooled electro-absorption modulated lasers (EMLs) will not be able to transmit 200G/lane beyond one kilometer due to chromatic dispersion — unless they are cooled. To even send high-bandwidth lasers for the shortest distances, the optics will need to be cooled. Equipment manufacturers that account for scalability in real time and adaptive bandwidth expansion through active cooling control will enjoy a competitive advantage in the marketplace.



Ideal PAM 4 signals



PAM 4 signals with laser wavelength drift



Phononic TECs are the Ideal Solution for High Performance & Efficiency

The inevitable has occurred, laser packages need a TEC to control — and lock — wavelength temperature and modulation frequency. Ideal PAM-4 signals feature a modulated set of waveforms. The more exact the waveforms, the easier it is for the demodulator to clearly identify the data it needs to decode. Without TECs, wavelength drift and light interference can distort the waveforms, clouding up the data to a point where the demodulator can no longer decode the data properly.

Even more critical is coherent quadrature amplitude modulation (QAM). Some current systems employ 256 states, where 16 amplitude and 16 phase states are being transmitted at once. When the wavelengths are about 1nm apart, a drift of just 0.1nm/deg C can't be supported, and complete channel overlap will occur in just 10˚ C.

Phononic's application-specific approach to TEC design, along with the hundreds of reference designs we've developed over the past decade, presents the ideal solution for cooling lasers used in high-speed optical components. Phononic offers cost-effective solutions for multichannel laser packages through powerful TECs that deliver best-in-class power consumption to achieve high yield and low cost. Our client partners have already seen this benefit in action — Phononic TECs are already found in tens of millions of devices across the globe. Plus, our powerful manufacturing partnership with Fabrinet, combined with our strong portfolio of global distribution partners, ensures you get the TECs you require on a timetable that matches your own manufacturing schedule. With application-specific TECs by Phononic, you'll see a lower cost per lane and a significant drop in power consumption, boosting your ROI while consistently meeting your customers' needs.

When designing a laser package, you need to maximize performance while minimizing component cost to keep your solution competitive. A major tenet of good TEC design is maximizing the coefficient of performance (COP), which is the measure of a TEC's heat pumping efficiency. It calculates the amount of heat removed by a TEC as compared to the amount of work required to remove it. The higher your COP, the better in-package performance you'll achieve. Cutting-edge, high-bandwidth and uncooled EMLs are a sizable cost premium over their cooled counterparts, and they can increase the overall cost of the package if performance expectations are not met across the entire operating temperature range.

By using a cooled architecture custom designed to meet your specific needs, you don't have to sacrifice cost efficiency for performance — or vice versa. The ideal TEC design is optimized to deliver on both fronts, even at the package level.

So, whether you're already preparing for the massive upcoming data demands, or if you're just starting to realize the data surge that is coming your way, Phononic can help you stay ahead of the AI revolution. With application-specific, highly innovative TECs designed to maximize transceiver performance, Phononic makes the ideal partner to help guide and prepare you for the data pressure AI is going to put on the global data infrastructure.

PHONONIC: INDUSTRY-LEADING PERFORMANCE & RELIABILITY

Best-in-class power consumption

Rigorous standards provide top-of-the-line quality and consistency

Customized, application-specific design process ready for high-volume manufacturing

PHONONIC

Learn more:

About Phononic: As the global leader in solid state cooling technology, Phononic is driving the world to a more sustainable way to cool. Its transformational technology reduces greenhouse gas (GhG) emissions and supports climate goals, while meeting the demanding performance needs of the market. The company's thermoelectric devices and integrated products are mission critical to how people work and communicate; automobiles 'see'; the protection and effective delivery of life-saving vaccines and drugs; last mile solutions supporting e-commerce; and innovative methods to cool living and workspaces.

Learn more at: **www.phononic.com**

Address: 800 Capitola Dr #7, Durham, NC 27713 **Website:** phononic.com **Twitter:** @phononic_inc