

Accelerating CONVERGE CFD Software with Cornelis® CN5000

White Paper

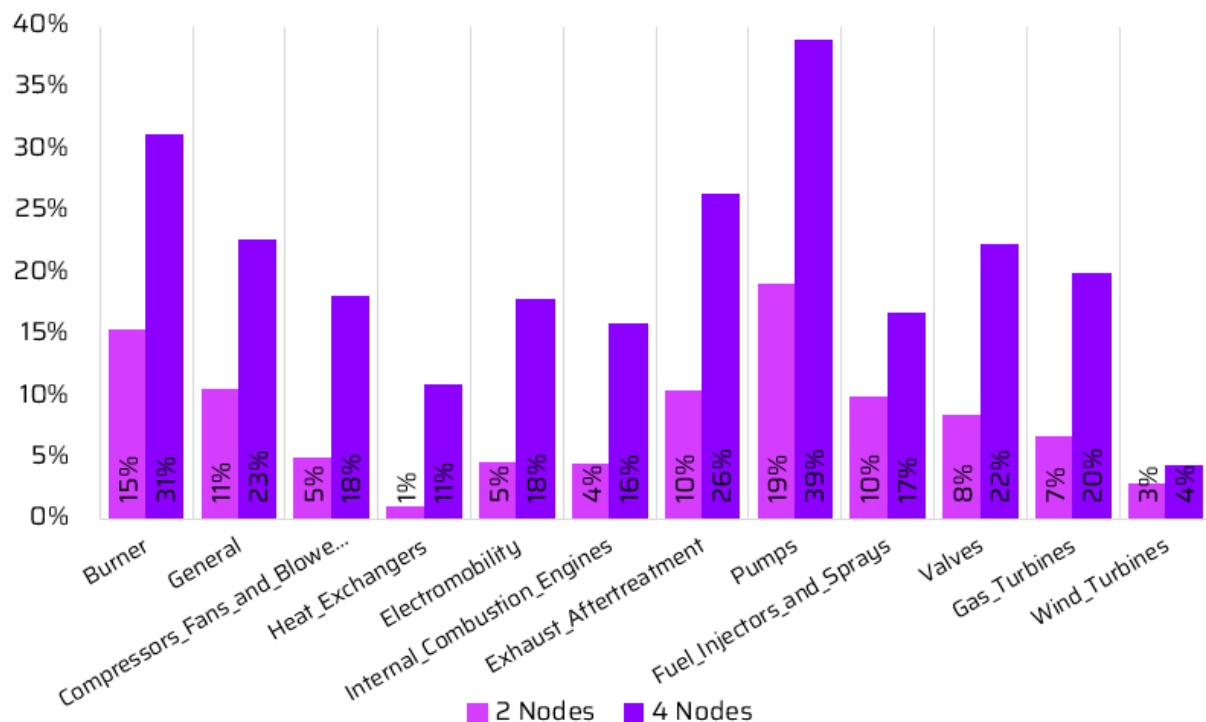
This paper evaluates the performance of CONVERGE CFD software 5.1.1 on the Cornelis® CN5000 Omni-Path® 400G fabric and compares it against NVIDIA® NDR across a broad set of CFD workloads. Benchmarking was performed on an eight-node AMD EPYC™ 48C CPU cluster using 132 representative CONVERGE example cases spanning 12 industry verticals.

Across the full workload set, CN5000 delivers consistent performance advantages over NDR. Using geometric means by industry vertical, CN5000 shows average throughput gains of 8% at two nodes and 20% at four nodes, with select verticals reaching improvements of up to 19% at two nodes and 39% at four nodes.

Scalability results show that CN5000 sustains higher efficiency as node count increases, averaging approximately 90% at two nodes and 70% at four nodes, while NDR drops below the commonly accepted 60% threshold.

MPI profiling explains these gains. Most observed CONVERGE message transfers are smaller than 16 kB, a range where CN5000 provides up to 2x higher message rate and 30% lower latency than NDR. Together, the results show that CN5000's small-message optimized data path and high message concurrency provide a strong technical fit for CONVERGE workloads.

Figure 1. CN5000 Throughput Improvement Relative to NVIDIA NDR



CONVERGE CFD Software

CONVERGE is an advanced computational fluid dynamics (CFD) software package developed by Convergent Science for simulating fluid flow, heat transfer, combustion, and chemical reactions in complex engineering systems. Unlike traditional CFD tools, CONVERGE employs autonomous meshing enhanced by Adaptive Mesh Refinement (AMR), removing the need for time-consuming manual meshing of complex moving geometries. CONVERGE is widely used in automotive, aerospace, energy, and industrial applications to optimize performance, improve efficiency, and reduce emissions. Its robust physics models and scalability on HPC platforms enable accurate prediction of real-world phenomena, supporting innovation in design, development, and engineering research.

AMD EPYC 9005 Series Processors

The AMD EPYC 9005 Series processors are designed to deliver exceptional performance for HPC and AI workloads. Built on AMD's latest Zen 5 architecture, EPYC 9005 processors combine high core counts, increased memory bandwidth, PCIe Gen5 connectivity, and strong per-core performance to address the needs of modern simulation and scientific computing environments. The AMD EPYC processor used in this study is a down-cored 48C processor based on the EPYC 9755, enabling outstanding performance for parallel engineering applications.

Cornelis CN5000 Omni-Path 400G

Cornelis CN5000 Omni-Path is an end-to-end 400 Gbps high-performance networking solution engineered to deliver low-latency, high-message-rate communication for HPC and AI workloads. Its architecture combines advanced congestion management, telemetry-driven adaptive routing, and lossless communication mechanisms to ensure predictable application performance.

The CN5000 architecture is built on Cornelis's foundational CN5000 Omni-Path products, including the CN5000 SuperNIC, 48-port Switch, and industry-leading 576-port Director Class Switch. Together, these components enable seamless scaling from a single switch to environments with over 500,000 endpoints. Through the open-source Libfabric interface, CN5000 integrates into major MPI and xCCL libraries for CPU and GPU deployments, enabling frictionless adoption without application code changes. CN5000 is a turnkey fabric solution that delivers leading performance and price/performance at any scale.

Refer to the [Cornelis CN5000 Product Family](#) for more information.

Methodology

Cluster Configuration

- **Test Platform**

Benchmarking was performed on an eight-node Dell R7725 cluster configured with dual-socket AMD EPYC CPUs and 768 GB of 6,400 MT/s RAM at 12 memory channels per socket. All compute nodes were tuned for optimal performance using guidelines published in the [CN5000 Performance Tuning Guide](#).

- **CPU Configuration**

To represent a typical CFD engineering deployment, the systems were configured with 48C per CPU (originally 128C EPYC 9755) via an appropriate choice of CCD and core pinning, i.e., six cores per CCD and two active CCDs per CPU quadrant. Processor power remained at the standard 500 W TDP found in EPYC 9755.

- **Network Configuration**

To ensure repeatable and fair comparisons, the CN5000 SuperNIC was installed in a PCIe slot attached to socket 0, while the NVIDIA ConnectX-7 NDR HCA was installed in a PCIe slot attached to socket 1. Both adapter sets were connected directly to their corresponding switches using copper cables in a single-hop configuration.

- **Software Environment**

CONVERGE version 5.1.1 was used for benchmarking and its module environment was modified to accept a system-installed IntelMPI installation.

- **OS:** Rocky Linux 9.5 (Blue Onyx). 5.14.0-503.33.1.el9_5.x86_64 kernel
- **Cornelis CN5000:** OPXS 12.1.2 software stack with in-distro OFED
- **Nvidia NDR:** UCX as packaged in HPC-X 2.25.1 with MLNX_OFED 24.10-4.1.4.0
- **MPI middleware:** Intel oneAPI MPI 2021.17, using the opx libfabric provider for CN5000 and the mlx provider for NDR

Workload Configuration

CONVERGE 5.1.1 includes 167 representative workloads in the *examples/case_setup* folder, covering 12 industry verticals. The *Internal_Combustion_Engines* and *Gas_Turbines* verticals are further organized into multiple subgroups. Of these 167 cases, 132 workloads with four-node runtimes of 12 hours or less were selected. The excluded cases are distributed mostly evenly across verticals.

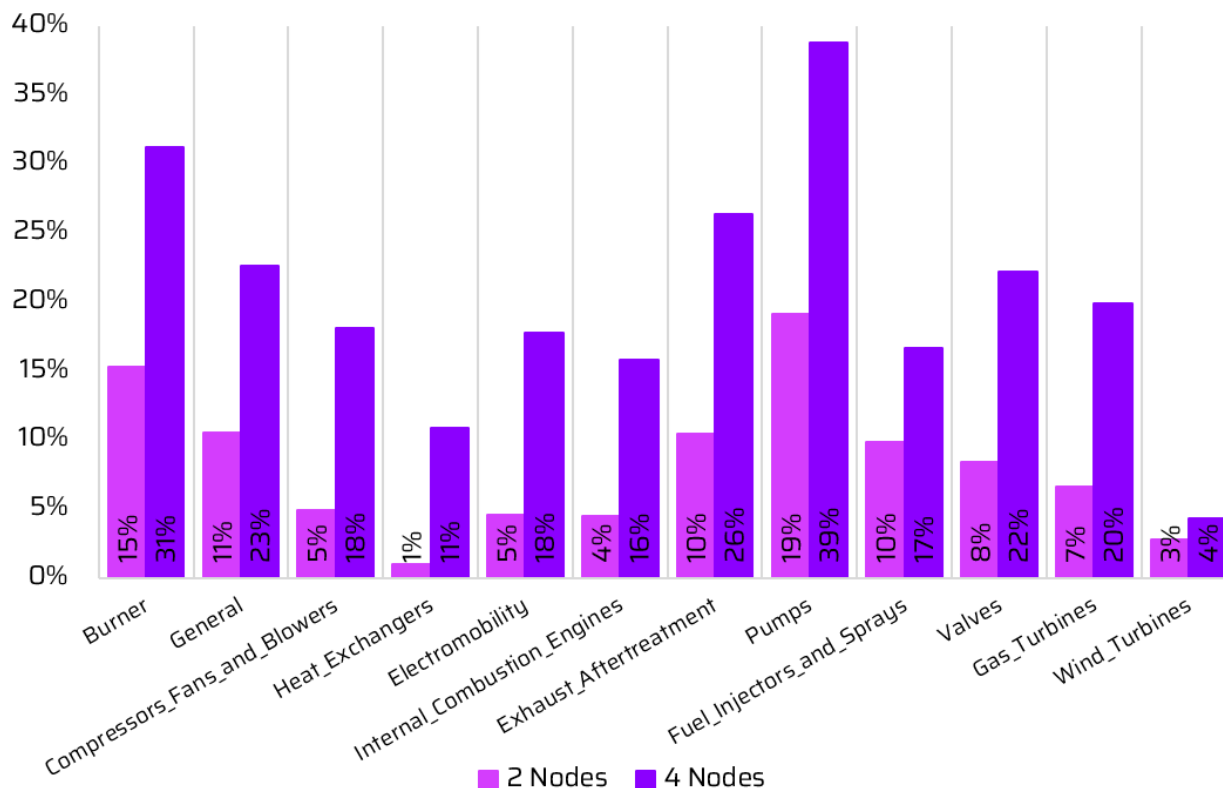
Workloads running for one to four hours were evaluated from one to four nodes, while longer four to 12-hour workloads, mostly from the *Internal_Combustion_Engines* vertical, were primarily evaluated at four nodes due to time constraints. Many workloads with runtimes under 30 minutes are functional or validation test cases that do not scale beyond a single node, mostly in the General and Pumps verticals.

Logging and file output was reduced to a minimum to avoid serialization bottlenecks.

Results

This section shows how CN5000 translates its low-latency, high-message-rate architecture into an application-level performance advantage for CONVERGE through competing NVIDIA NDR measurements taken under identical conditions.

Figure 2. Geomean of Relative CN5000 Throughput Improvement vs. NDR Per Industry Vertical

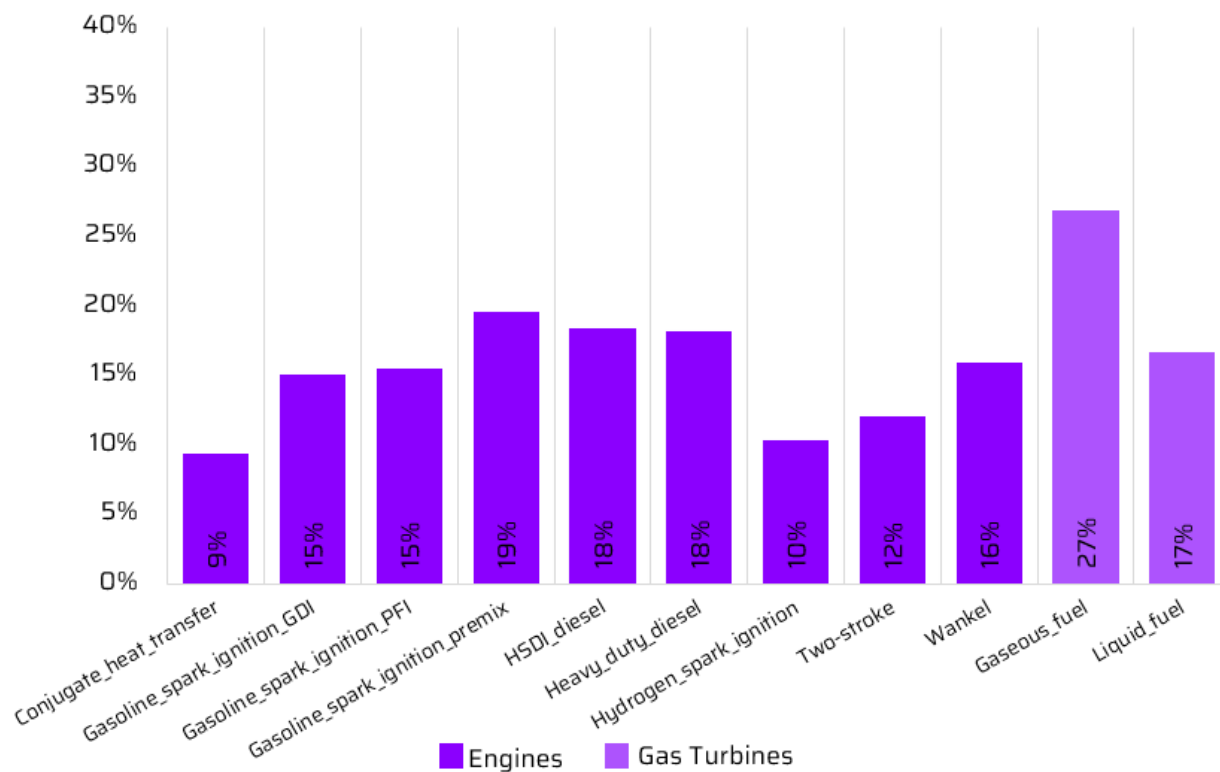


To summarize results across the 132 workloads, throughput improvement for CN5000 was aggregated using the geometric mean for each industry vertical. The two-node and four-node geomeans provide a representative estimate of expected performance gains within each vertical, as the sampled workloads cover most of the relevant physics. Throughput is defined as the number of repetitions of each workload that can be solved inside of a finite amount of time.

Customers selecting Cornelis CN5000 can expect average throughput gains of 8% at two nodes and 20% at four nodes. In select verticals, improvements can reach up to 19% at two nodes and 39% at four nodes.

A closer look at the composite *Internal_Combustion_Engines* and *Gas_Turbines* verticals further demonstrate CN5000's ability to handle fine-grained communication patterns driven by complex moving geometries, dynamic AMR, chemical reactions, and thermal conduction. Performance variation within each subvertical closely follows the overall geometric mean, highlighting CN5000's robustness across diverse workload configurations.

Figure 3. Zoom-in of Geomean Throughput Improvement for Engines and Gas Turbines at Four-Node Scale



Scalability Comparison

Many of the reference workloads shipped with CONVERGE are primarily aimed at functional and accuracy validation purposes rather than designed for scalability studies. Nonetheless, a model that is at the end of its scaling curve exhibits a stronger dependence on the performance of the fabric than a model that can hide all communication behind computation. Fabrics that are stressed beyond their limit tend to show a characteristic knee in their scaling curve and may even exhibit negative scaling, i.e., the throughput starts to decrease with increasing number of compute nodes.

The following table shows such scaling efficiency measurements for both fabrics in a filtered and unfiltered manner. Filtering here means only workloads that show a two-node scaling efficiency greater than 60%, in either fabric, are included in the geomean.

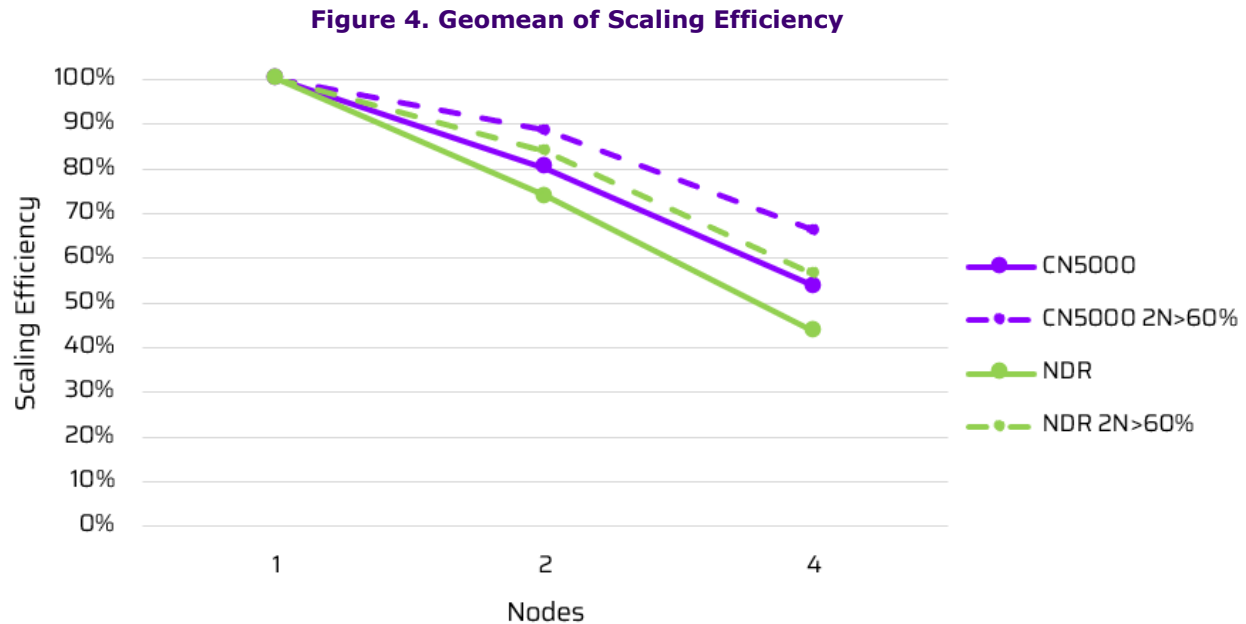
Scaling efficiency is calculated by $\frac{\text{runtime}(1N)}{\text{runtime}(nN) \cdot n}$, where n is the number of nodes. A scaling efficiency of 100% gives perfect linear scaling.

Table 1. Scaling Efficiency at Two-Node (2N) and Four-Node (4N) Scale

Fabric Nodes	CN5000		CN5000 2N>60%		NDR		NDR 2N>60%	
	2N	4N	2N	4N	2N	4N	2N	4N
Burner	77%	59%	89%	67%	69%	39%	83%	54%
General	59%	31%	77%	55%	53%	25%	74%	47%
Compressors_Fans_...	88%	67%	88%	67%	84%	57%	84%	57%
Heat_Exchangers	88%	67%	88%	67%	87%	61%	87%	61%
Electromobility	80%	56%	80%	56%	77%	48%	77%	48%
Exhaust_Aftertreatment	75%	43%	83%	57%	68%	34%	77%	45%
Pumps	65%	37%	73%	47%	55%	27%	66%	36%
Fuel_Injectors_and_Sprays	83%	43%	87%	48%	76%	37%	79%	42%
Valves	68%	44%	103%	127%	63%	36%	103%	116%
Gas_Turbines	93%	66%	93%	66%	87%	55%	87%	55%
Wind_Turbines	121%	110%	121%	110%	118%	105%	118%	105%
Geometric Mean	80%	54%	89%	66%	74%	44%	84%	57%

The results confirm the earlier statement that the workloads in the General and Pumps vertical are generally too small for 4N scaling. Nonetheless, CN5000 can extract meaningful performance gains with most workloads at scales where NDR has already topped out.

The filtered scaling efficiency of CN5000 maintains an ~90% average at two nodes and an ~70% average at four nodes, whereas NDR, even in the filtered case, has already slipped below the 60% threshold generally considered to be borderline acceptable.



The dotted lines only include cases with a two-node scalability over 60%.

Scale-up vs. Scale-out

Another way to verify the exceptional performance of Cornelis CN5000 is to compare it to the intra-node scale-up network of every dual-socket server, i.e., the inter-socket network connecting both sockets via the mainboard. This comparison can easily be performed by pitting a 2N single-socket scale-out configuration against a 1N dual-socket scale-up configuration, while maintaining identical power limits, memory bandwidth and compute. A scale-out network that can keep up with the scale-up network increases both the flexibility and composability of a CFD cluster, as well as demonstrates that the scale-out will not become a bottleneck at increasing scale.

The following workloads were compared with CN5000:

1. *Wind_Turbines/IEA_15MW_wind_turbine_transient_RANS* with super-linear scaling efficiencies of 136% at 2N and 118% at 4N
2. *Internal_Combustion_Engines/Gasoline_spark_ignition_premix/SI8_engine_premix_E85_transient_RANS* with scaling efficiencies of 84% at 2N and 66% at 4N

Workload 1 showed a case runtime of 1694 s in the 1N dual-socket scale-up case compared to 1678 s in the 2N single-socket scale-out case. Workload 2 showed a case runtime of 7284 s at 1N scale-up vs. 7132 s at 2N scale-out.

Note that using CN5000 single-socket scale-out leads to faster completion times than dual-socket scale-up, proving that CN5000 performs at the upper limit of what is possible in terms of latencies and message rates.

Scale-out can be used to improve cluster utilization in a scenario where there is a mix of scaling and non-scaling workloads on the cluster. Dense compute clusters in such a scenario require complicated multi-tenancy scheduling with often uneven utilization that consumes 100% of the available infrastructure while only delivering a fraction of its compute. Instead, the ability to handle effortless scale-out allows for composing many smaller nodes into a single “big” node at high efficiency. This opens the possibility of right-sizing your cluster for the smallest common workload and composing the compute units into bigger nodes only when needed to reach 100% cluster-wide compute delivery.

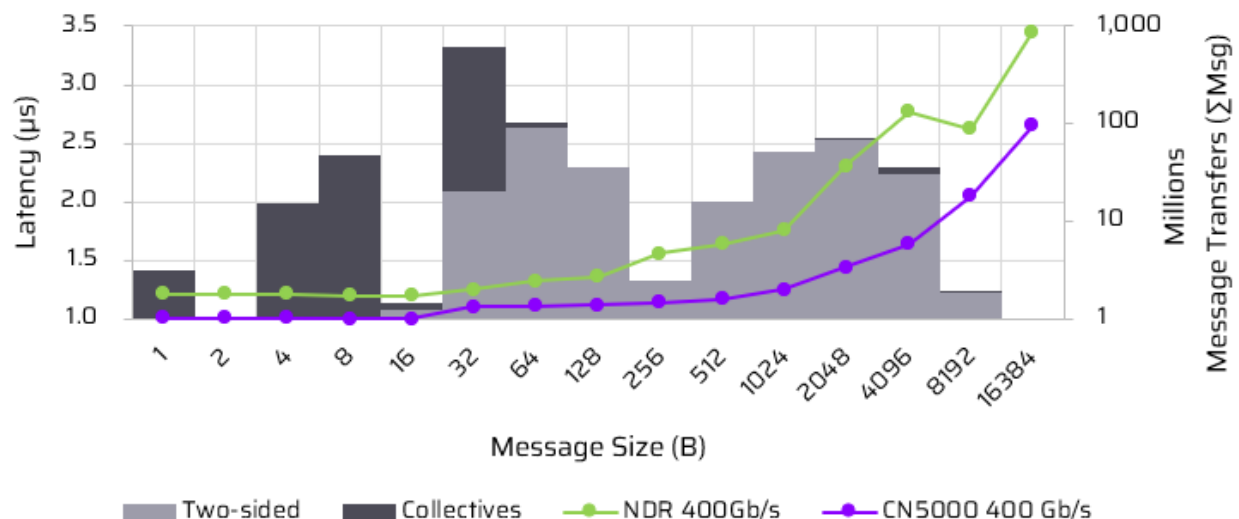
Analysis

This section explains how CN5000 achieves its excellent performance by directly mapping typical microbenchmark results such as latency, message rate, and bandwidth to representative measured communication patterns in CONVERGE.

Message Transfers vs. Microbenchmarks

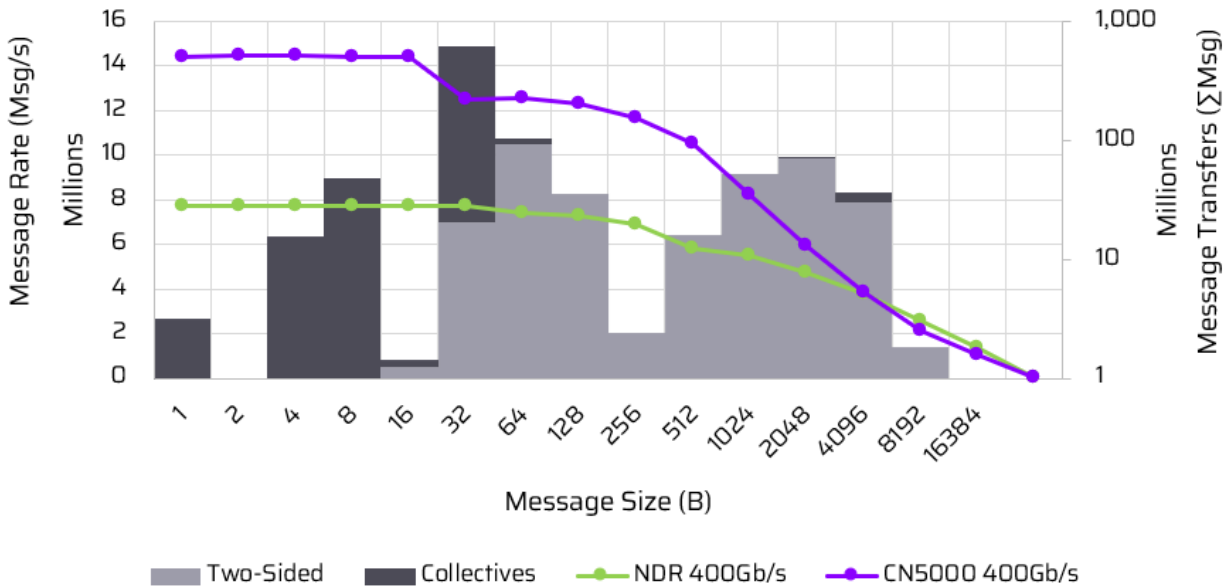
The following figures illustrate how MPI profiling shows message transfers are smaller than 16 kB, where CN5000 delivers up to 2x higher message rate per rank and 30% lower latency than NDR. This message-rate advantage accelerates fine-grained, non-blocking data exchanges in neighbor communication and global all-to-all patterns, while the lower latency speeds-up synchronization barriers, blocking two-sided communication, and successive stages of collective operations.

Figure 5. IMB-MPI1 Pingpong Results Overlaid onto Intel APS MPI Profiling Results



Background gives the message-size binned histogram of MPI transfers during a run of Internal_Combustion_Engines/Gasoline_spark_ignition_premix/SI8_engine_premix_E85_transient_RANS

Figure 6. 1 PPN IMB-MPI1 Uniband Results Overlaid onto Intel APS MPI Profiling Results

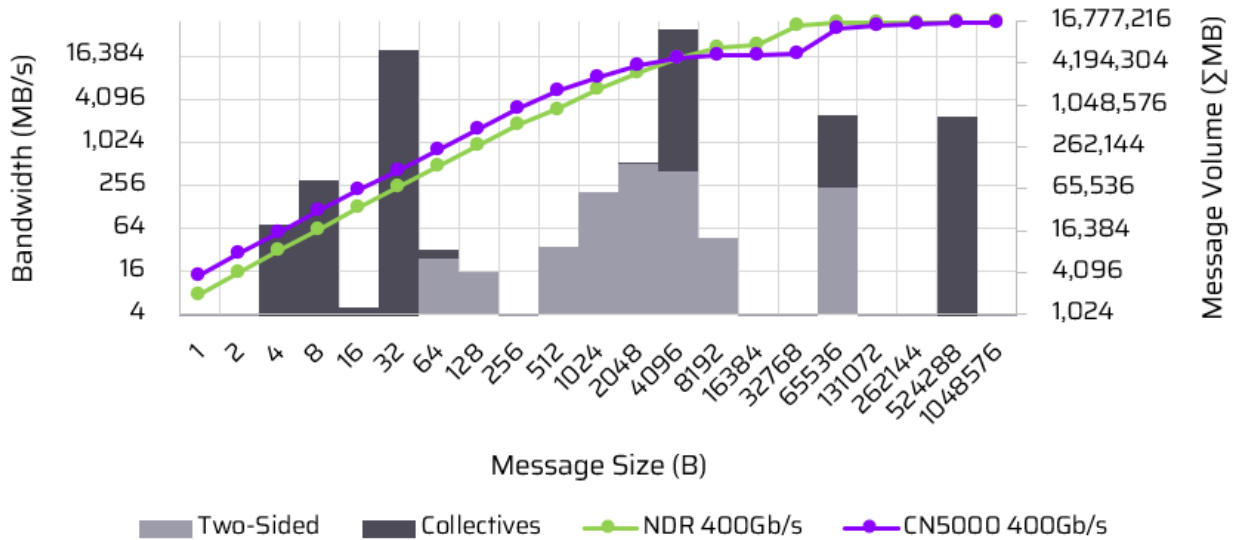


Background gives the message-size binned histogram of MPI messages during a run of Internal_Combustion_Engines/Gasoline_spark_ignition_premix/SI8_engine_premix_E85_transient_RANS

The 1.7x higher peak message rate per SuperNIC of 800 Mmsg/s on CN5000 against NDR's 460 Mmsg/s also increases MPI communication concurrency on high-core-count CPUs. This message rate advantage enables low-core count CPUs to still use the network to its fullest extent, while adding higher performance headroom than NDR for high-core count CPUs.

Message Volume vs. Microbenchmarks

Figure 7. 1 PPN IMB-MPI1 Uniband Results Overlaid onto Intel APS MPI Profiling Results



Background gives the message-size binned histogram of MPI message data volume during a run of Internal_Combustion_Engines/Gasoline_spark_ignition_premix/SI8_engine_premix_E85_transient_RANS

CN5000's higher message rate and lower latency improve the effective bandwidth of both blocking and non-blocking communications, enabling MPI applications to operate closer to line rate across a broader range of message sizes. For small messages, effective bandwidth is determined by different factors depending on the communication model:

- Non-blocking communication (Uniband): Effective bandwidth scales with message rate $\times$ message size, reflecting how quickly the software stack can enqueue messages and how efficiently the hardware can transmit them.
- Blocking communication (Pingpong): Effective bandwidth is approximated by message size $\div$ (latency + data transfer time). For small messages, latency is the dominant factor.

To optimize small-message performance, CN5000 uses a streamlined data path that minimizes latency and overhead by placing data directly on the wire, avoiding unnecessary round trips between memory and the SuperNIC.

For large messages, CN5000 leverages standard RDMA semantics and hardware-accelerated zero-copy transfers to read and write remote memory directly, eliminating CPU involvement and maximizing throughput.

Key Study Findings

- CN5000 delivers consistent CONVERGE performance gains over NDR across 132 representative workloads spanning 12 industry verticals.
- Across the full workload set, CN5000 improves average throughput by 8% at two nodes and 20% at four nodes, with select verticals reaching gains of up to 19% at two nodes and 39% at four nodes.
- CN5000 maintains stronger scaling efficiency as node count increases, sustaining approximately 90% efficiency at two nodes and 70% at four nodes.
- MPI profiling shows that most CONVERGE communication uses messages smaller than 16 kB, where CN5000's higher message rate and lower latency provide a direct application-level advantage.
- CN5000's small-message optimized data path, high message concurrency, and efficient large-message RDMA behavior make it a strong fit for communication-intensive CONVERGE simulations.

Legal Disclaimer

You may not use or facilitate the use of this document in connection with any infringement or other legal analysis concerning Cornelis Networks products described herein. You agree to grant Cornelis Networks a non-exclusive, royalty-free license to any patent claim thereafter drafted which includes subject matter disclosed herein.

No license (express or implied, by estoppel or otherwise) to any intellectual property rights is granted by this document.

All product plans and roadmaps are subject to change without notice.

The products described may contain design defects or errors known as errata which may cause the product to deviate from published specifications. Current characterized errata are available on request.

Cornelis Networks technologies may require enabled hardware, software, or service activation.

No product or component can be absolutely secure.

Your costs and results may vary.

Cornelis, Cornelis Networks, Omni-Path, Omni-Path Express, and the Cornelis Networks logo belong to Cornelis Networks, Inc. Other names and brands may be claimed as the property of others.

Copyright © 2026, Cornelis Networks, Inc. All rights reserved.