SYNOPSYS®

WHITE PAPER

Evolution of Data Center Networking Technology — IP and Beyond

Author

Jerry Lotto
Sr. Staff Technical Marketing Manager, Synopsys

Introduction

Ethernet is ubiquitous—it is the core technology that defines the Internet and serves to connect the world in ways that people could not imagine even one generation ago. HPC clusters are working on solving the most challenging problems facing humanity—and cloud computing is the service hosting many of the application workloads struggling with these questions. While alternative network infrastructure within datacenters often facilitates low latency, high-speed communication, they are often limited in reach to within a single facility. Additionally, there is often a prohibitive cost to building multiple fabrics and cloud providers that occupy multiple locations need to move data and even perform inter-process communication between locations. For these needs, Ethernet SoCs supporting up to 800 GbE or even beyond play a critical role and companies who can aggregate, route, and deliver this traffic with minimal latency will thrive—presenting a way to leverage massively scale-out and long-haul high-speed communication critical to multi-site high performance computing, machine learning and data analytics.

This white paper explains the Ethernet standards' evolution over the years from supporting home networking to now enabling hyperscale and cloud data center networking. The paper also highlights the need for a more comprehensive Ethernet solution beyond IP that SoC designers demand for 100G to 800G SoCs.

The Ethernet Standard and Other Critical Elements

Ethernet initially became a reality in 1972 at Xerox's Palo Alto Research Center when Bob Metcalf and Dave Boggs were challenged to share the world's first laser printer with hundreds of workstations. Inspired by a packet radio network used to communicate among the Hawaiian Islands (ALOHANet), the concept of a Carrier Sense Multiple Access protocol for Collision Detection (CSMA/CD) was developed to define the Media Access Control (MAC) layer of the Ethernet protocol. Below the MAC layer, a passive coaxial cable was used to propagate electromagnetic waves (packets) and an outstanding 3 Mbit/s Ethernet network was achieved. While this remained internal for several years until the commercialization of the Ethernet protocol in 1979. After four years of work, the IEEE approved the first standard for n CSMA/CD Ethernet communication over 10Base5 (aka "thicknet") at a rate of 10 Mbps. Challenging to use, this cabling method required a hole to be bored through the cable jacket and outer braid to the center core and installation of a "vampire tap" to connect each machine. The development of the Ethernet physical layer steadily improved both usability and speed over the next three decades until there was a literal explosion of standards to address different market segments and applications. Figure 1 illustrates the Ethernet PHY standard evolutions throughout the years.

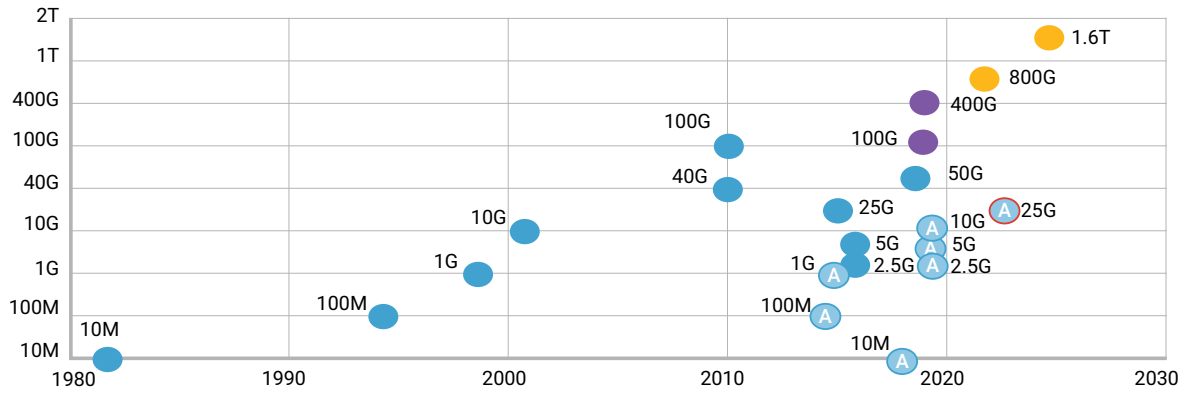

Figure 1: IEEE 802.3 Physical Interface Standards

The IEEE now plays a critical role in standardization and therefore interoperability of Ethernet devices—known in aggregate as 802.3 for wired networking and 802.11 for wireless Ethernet transport. These standards include everything from the original 10Base5 thick Ethernet cable for 10 Mbit/s Ethernet to the 800GBASE-R specification for 800G Ethernet. Additionally, the 802.1Q standards for time sensitive networking are critically important to industrial control and vehicular applications of Ethernet. In a practical sense integration of the physical layer (PHY), consisting of the physical media dependent (PMD), physical medium attachment (PMA) and physical coding (PCS) sublayers and the data link layer (MAC) of the Ethernet protocol is a challenging problem for engineers but IP designers solve these problems in clever and reliable ways, providing silicon architectures that satisfy the needs of upper protocol software developers. Power and distance remain the biggest challenges for high-speed Ethernet above 100G. While short haul connectivity solutions can often rely on zero or trivial bit error rates, as distances increase so does the ambient noise, signal degradation, and propagation delay intrinsic to any signal transmission. NRZ or simple one bit per pulse signaling no longer provides sufficient bandwidth or error free connectivity even over short distances at speeds more than 200 Gbps. Figure 2 shows the difference between the two signaling models where compared to NRZ's two voltage levels, PAM-4 has four voltage levels that result in 12 distinct signal transitions, (six rise & six fall times) creating three district eye openings.

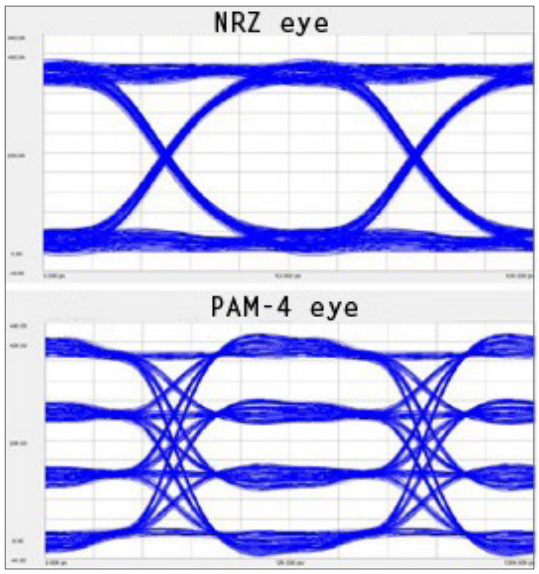

Figure 2: NRZ and PAM-4 signaling (Image curtesy of: S-parameters: Signal Integrity Analysis in the Blink of an Eye | 2017-05-14 | Signal Integrity Journal)

PAM-4 provides a way to signal two bits per symbol (00, 01, 10, and 11) and quadrature amplitude modulation (QAM) formats can extend this density from 8 to as many as 256 bits per symbol. Copper cable is limited in reach due to the power requirements of the medium, but optical fiber provides a way to combine different wavelengths to extend the width of communications. Wavelength division multiplexing (WDM) technologies include Coarse (CWDM) up to 70 km and Dense (DWDM) variants which can deliver 400

Gbps Ethernet beyond 80 km (about half the distance from Washington, D.C. to New York City) using active repeater technology and with amplification can reach distances of thousands of kilometers for 100 Gbps transmission.

Modern Ethernet networks consist of many different elements. The media over which it is carried can vary from wireless radio, cellular or Wi-Fi, to coaxial or twisted copper, to many wavelength and intensity variations over optical fiber transmission. Hubs, switches, routers, gateways, bridges, media converters, optical amplifiers, transponders and muxponders create pathways over which Ethernet can travel. The IEEE 802.3 standards over which wired Ethernet can be carried are shown in Figure 3 below.

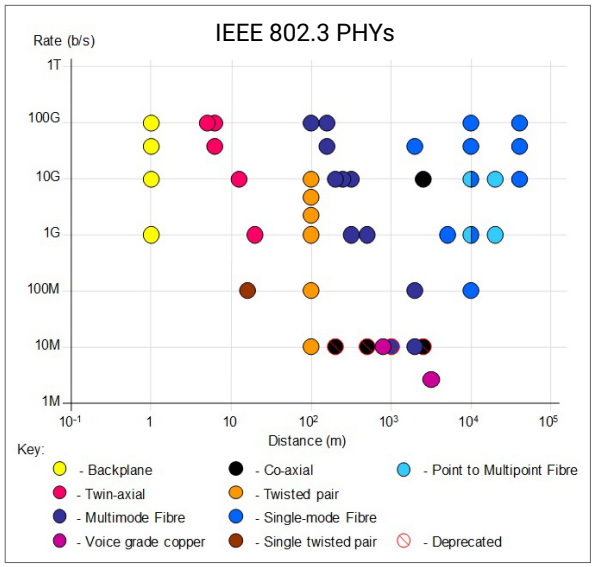

Figure 3: Ethernet PHYs by Distance (Image curesty of: Evolution of Ethernet Standards in IEEE 802.3 Working Group | IEEE Standards University)

HPC Evolution

High-performance computing (HPC) has undergone many transitions since the first monolithic mainframe behemoths were designed, giving way first to network-connected collections of personal-class computers evolving to collections of specialized computers enhanced with accelerators and specialized memory, storage, and networking components, and more recently to virtualized on-demand high-performance computing in a virtually unlimited capacity cloud, hosted by various providers. Workloads necessarily shifted along with these architectural changes, taking advantage of the unique performance characteristics of the infrastructure on which they were run. Application and communication frameworks were developed to further optimize and simplify the ever-changing world of high-performance computing applications, machine learning algorithms and big data analytics that have become so central to our lives. All of which support the advancement of science and technology, form the basis of all Internet services, continually integrate with and become central to government, industry, business and education. Within a datacenter, the scale of HPC—especially the aggregation of HPC workflows on a virtualized cloud hosting infrastructure—demand massive scale-out architectures. According to the Green 500 list for June 2021 <June 2021 | TOP500>, HPC computational efficiency is approaching 30 Gflops/watt and standard 42" rack can cool nearly 40 kW in an enclosed pod delivering up to 1.2 Pflops in that rack! A reasonable estimate of the power utilization in a data center allocates approximately 40% each to running and cooling servers, and 20% for storage and networking combined, imposing practical limits on datacenter growth and capacity. The power scenario is different for the number and capacity of Ethernet ports that need to be connected within a 19" 42U rack. Top of rack (ToR) aggregation may need to support hundreds of Ethernet connections both for primary communication and out of band management purposes.

HPC-driven Requirements

Within the datacenter, HPC depends on low latency, fast communication between computers (nodes) in a cluster. Inter-process communication for distributed applications uses communication frameworks such as MPI, SHMEM, and UPC++. These traffic patterns tend to be bursty, using relatively small packets of information. Depending on the scaling properties of an algorithm, these traffic patterns can extend to hundreds or even thousands of nodes in a cluster. Storage traffic which co-resides on the same communication fabric often is more bandwidth intensive with larger packet sizes, often leveraging the alternative jumbo frame format extending payload size from the standard 1500 bytes to 9000 bytes per packet. Machine learning frameworks can also benefit from large packet communications while training a new model as they chew through massive datasets. Big data analytics workloads are similar in nature, often leveraging trained AI framework to recognize trends and features in large datasets or streaming Internet data sources.

The Internet itself is carried over layer 3 (aka IPv4 or IPv6) of the OSI model for data communication. VLAN segregation of Ethernet traffic is a layer 2 protocol that allows management, data, storage, and inter-process communication to traverse common Ethernet infrastructure in relative safety and exclusivity even though traffic co-resides in active data flows. VXLANs move this capability to layer three and allows VLANs to coexist in multiple datacenters. Beyond the datacenter, graphics visualization/interaction, file transfer and command/control I/O extend to end user machines and terminals throughout the world. None of these technologies can work if the PHY, PCS, and MAC layers of the underlying layers 1 and 2 of Ethernet infrastructure cannot satisfy the speed and latency demands imposed at each of these upper layers. This is the focus area of system-on-chip (SoC) designers, enabling the creation of firewalls, routers, gateways, NICs, and cables that deliver high-speed Ethernet worldwide.

Summary

The Internet is growing in ways that we never anticipated and will continue to expand beyond our imagination. Communication, entertainment, education, employment, industry, and finance all depend on the Internet. The delivery of services has become integral to many people's social and interpersonal lives. This also demands high-performance computing, supporting the scale of transactional and streaming needs of the aggregate public. Estimates of bandwidth requirements for the future just going back 10 years are consistently low, often by half or even a third of today's reality! Nobody anticipated how a world-wide pandemic would impact the workforce in such fundamental ways and how working from home would drive so much higher demands on both distributed and core Internet infrastructure. Regardless, the worldwide demand for bandwidth is growing at unprecedented rates and the aggregation of that traffic is pushing metropolitan networks and commercial Internet providers to ramp up infrastructure. The anticipation of 5g cellular is another factor driving core infrastructure to new heights.

For all these reasons, Ethernet is becoming the de facto networking standard of choice for high-performance computing in the data center. The industry is working with organizations such as the IEEE to define new standards for Ethernet to address the worldwide demand for bandwidth. While having access to IP that supports various configurations for 10G to 800G Ethernet is important, SoC designers are looking for solutions that offer far more differentiations:

- Integrated MAC and PHY IP solutions that include PCS, PMD, PMA and auto negotiation functionalities for a range of data rates, reducing time-to-market
- Lowest latency Ethernet IP that can be configured to the specific needs of SoC designers
- Solutions that support several types of co-packaged optics to support the bandwidth and power increase in the data center due to high volume and complexity of data
- Successful interoperability with the ecosystem supporting data center channels, cables, and interconnects for high-speed Ethernet
- Access to package escape studies that allow tile stacking for maximum SoC density, evaluation boards to get a head start on testing and prototyping, and comprehensive services that allow for fast integration and expert support

Synopsys is leveraging its experience in developing high-quality IP solutions to help designers meet their complex and specific design challenges. Our partnerships with foundries and other ecosystem partners allow us to ensure our IP is interoperable and available in most demanding process technologies. Our team of engineering experts work closely with designers to resolve their most technical issues and minimize integration risk.

01/12/22.Evolution of Data Center Networking Technology-IP and Beyond-WP.

Pub: Dec. 2021