

# 100G Ethernet Technology and Applications

**Zhang Yuanwang**

(ZTE Corporation, Shenzhen 518057, P. R. China)



## Abstract:

The increasing requirements for bandwidth drive the 100G Ethernet into use as quickly as possible. The major technologies supporting 100G Ethernet interface include the physical layer channel convergence technology, multi-fiber channel and Wavelength Division Multiplexing (WDM) technology. The key technology for high-speed optical devices at the interface requires a breakthrough. The demands driven by the increase of interface rate require better packet processing and storage, system switching, and backplane technology. In addition, in the network, the transport problem of the new interface should be solved, including defining transport standards and pushing the development of key transport technologies. As far as the cost and requirements are concerned, the commercial application of 100G Ethernet is viable in Metropolitan Area Networks (MANs).

**T**he basic demand for upgrading the Ethernet interface rate to 100 Gbit/s is the increase of bandwidth.

The precondition for promoting the 100G Ethernet application is the establishment of relevant standards. The standard corresponding to 100 Gbit/s Ethernet interface is IEEE802.3ba<sup>[1]</sup>, whose draft 2.3 is formed<sup>[2]</sup>. In the standard, the medium interface, rate and Physical Code Sublayer (PCS), and Media Access Control (MAC) architecture are defined. In addition, the standard organizations relevant with 100GE include ITU Telecommunication Standardization Sector (ITU-T) and Optical Interconnection Forum (OIF).

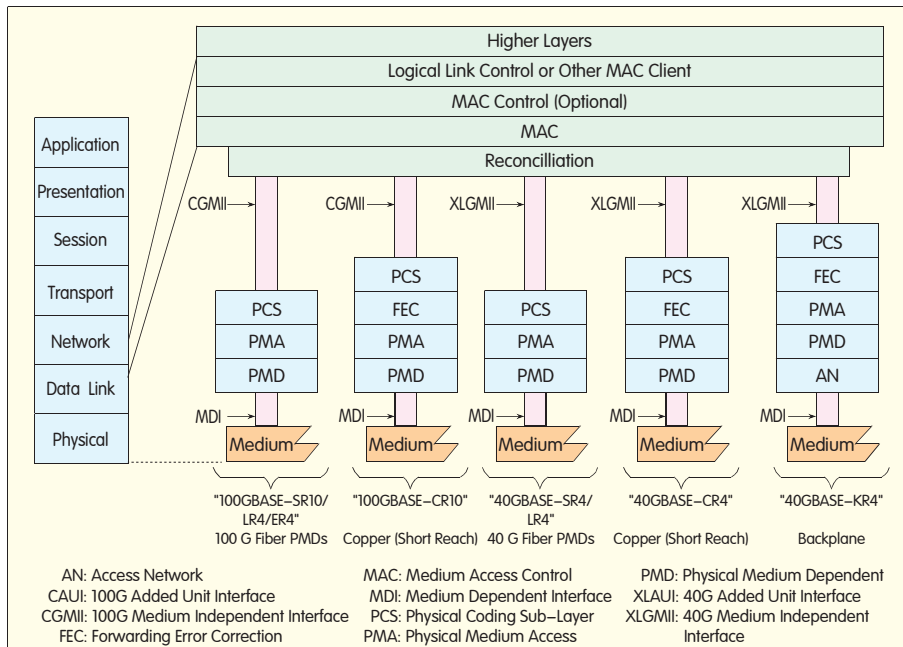
Upgrading Ethernet to 100 Gbit/s interface requires the support of key technology. The mature and commercial

application of key technology takes certain time. Technical breakthrough and time are required for chip, system, network and standard.

## 1 100G Ethernet Technology and Standards

The key technologies supporting 100G Ethernet interface include physical channel convergence technology, multiple-fiber channel and WDM technology. The Physical Media Dependent (PMD) sublayer meets the 100 Gbit/s rate. New chip technology support 40 nm technique. These provide the possibility of developing next generation high-speed interface. For the interface, the parallel multiple-mode interface of fiber interface PMD has the problem of intensive encapsulation and great power consumption. The Wavelength Division Multiplexing (WDM) interface of single-mode 4X25 Gbit/s should break through the technology of serial-parallel conversion circuit (SERDES) technology and the non-cooled optical components. For the system, the increase of interface rate brings new threshold for packet processing, storage, system switching, and backplane technology. For the network, the transport problem of the new interface should be solved. New Optical Transmission Unit (OTU) frame structure should be defined. In addition, ultra-high-speed transport, the problems including signal processing, optical signal modulation, physical coding, dispersion compensation, non-linear processing, compatibility and consistency with Fast Ethernet (FE)/Gigabit Ethernet (GE)/10GE frame structure and physical sub-layers should be solved. The 100G transport features should meet the relevant features of 10G transport network. Otherwise, the reconstruction will impact the progress of new technology.

The next-generation Ethernet



▲ Figure 1. Various medium architectures defined in IEEE802.3ba.

technology standard includes 40 Gbit/s and 100 Gbit/s, which are designed for different requirements of server and network. The 40 Gbit/s rate is mainly for the calculation application but the 100 Gbit/s rate is mainly for the core and convergence application. Two rates are provided because IEEE wants to ensure that the Ethernet can meet requirements of different applications effectively and economically to promote the network convergence based on Ethernet technology. The standard defines the modules and connection interface bus of Physical Coding Sublayer (PCS), Physical Medium Access (PMA) sublayer, PMD sublayer, and Forwarding Error Correction (FEC). The inter-chip bus between MAC and PHY use XLAUI (40 Gbit/s) and CAUI (100 Gbit/s). The intra-chip bus uses XLGMII (40 Gbit/s) and CGMII (100 Gbit/s). The architecture of each medium is shown in Figure 1<sup>[3]</sup>.

The standard supports only full-duplex operation. The 802.3MAC Ethernet frame format is reserved. Multiple physical medium interface regulations are defined, including 1 m backplane connection (definition of 100GE interface no-backplane connection), 7 m copper cable, 100 m parallel multiplex-mode fiber and 10 km single-mode fiber (based on WDM technology). 100 Gbit/s interface defines

40 km transport distance at most. The standard defines the architecture of Multi-Lane Distribution (MLD) in PCS. It also defines the regulation of electrical interface for inter-chip connection. 40 Gbit/s and 100 Gbit/s respectively use four and ten 10.3125 Gbit/s channels. It uses polling mechanism to allocate data to obtain the 40G and 100G rates. In addition, the definition of virtual channel solves the problem of adapting to different physical channel of long optical wavelength. It specifies that the physical layer coding uses 64B/66B.

The standard provides the architecture and interface definition of 100 Gbit/s Ethernet, but many problems remain unsolved. First, PMD is one key part of 802.3a. 40G/100G optical module contains the parallel interface of short wavelength corresponding to 40GBASE-4SR and 100GBASE-10SR. The technical difficulty lies in the intensity of encapsulation. For the wavelength division interface of long wavelength, the difficulty lies in the SERDES and encapsulation technology of 25 Gbit/s corresponding to PMA. For 100G WDM optical module, non-cooled laser technology is a key technology related with the standard. The encapsulation mode is specified to CFP by the CFP Multiple Source Agreement (MSA)<sup>[4]</sup>. The definition of copper Medium Dependent

Interface (MDI) standard uses SFF-8436 and SFF-8642. The detailed architecture size and pin assignment are provided. As we know, the major suppliers will not provide the 100G WDM optical module until 2010.

For the relevant chips of 100G interface, the MAC layer is ready. The electrical interface regulation of PMA service interface requires that each channel should work at the rate of 10.3125 Gbit/s. Use the Application-Specific Integrated Circuit (ASIC) after the standard is mature. The MAC implemented based on the Field Programmable Gate Array (FPGA) should support the rate of 10.3125 Gbit/s. Only a few FPGA companies support this rate<sup>[5]</sup>. In the previous evaluation system, the SERDES Mux devices<sup>[6]</sup> are added. It is transitive measure converting 8/20 5.156 25 Gbit/s channels to 4/10 10.3125 Gbit/s standard interfaces<sup>[7]</sup>.

For the 100G Ethernet device system, the preceding technical problems of 100G Ethernet interface are hard to remove. The relevant packet processor is required. For the distributed large-capacity switching system, the problem of chipsets of large-capacity packet switching system should be solved.

For the 100G packet processing capability, no universal solution is available in the industry. The solution in progress is to be evaluated. For the Content Address Memory (CAM) of the network processor, the bandwidth of search interface should be doubled. There is a bottleneck for the data bus width and rate. It encourages the application of serial high-speed bus interface such as Interlaken LA. Owing to the restriction of the single chip processing capability and the pileup of multiple chips caused by bus interface conversion, the board area and power consumption are unacceptable. The solution based on FPGA-customization requires comprehensive technology. The provided service processing capability is restricted.

The packet switching system chipset, including switching network and switching network interface chips probably contain the Traffic Management (TM) chip. Previously, most systems do

not support valid data bandwidth greater than 100 Gbit/s for each line. In the recent new solution, the backplane bandwidth of each line card is about 100–200 Gbit/s. The backplane SERDES bus rate supports up to 6.5 Gbit/s. Supporting 100 Gbit/s interface, the bandwidth of each line card should be upgraded to 200–500 Gbit/s bandwidth. The backplane SERDES rate should reach 10.3125 Gbit/s or higher. The requirements for backplane design, technic, material, and meeting the bus length are more critical than before. For the system meeting the carrier-class requirements, the Virtual Output Queue (VOQ) and Hierarchy Quality of Service (HQoS) should be met. It requires greater processing bandwidth, queue processing capability, and buffer, which greatly increases the difficulty of system design.

With the increase of system requirements, the power of the system is also increasing. 100 Gbit/s long wavelength PMD requires four 25 Gbit/s channels. The increase of SERDES rate and channels require greater power supply. The 100 Gbit/s processor requires larger number of memories and greater power. The micro-processing technology also requires greater power. Consequently, we need to find a solution.

If the high-speed Ethernet wants to bring real benefits to users, the transport network must carry the transport network service. It should not only apply the technology to the large data center or small LAN. Except the modulation technology, the transport over optical transport network of high-speed Ethernet and the Operation and Maintenance (OAM) also determine the success of the technology. In the Jeju Island conference of ITU-T SG15 Q11 in September 2008, the definition of OTU mapping of 40G/100G Ethernet interface is formed<sup>[9]</sup>. G.709 defines that 40GE is mapped to OPU3. It uses 1,024B/1,027B transport codes. 100GE is mapped to ODU4/OTU4, and the bit rate is 111.809973 Gbit/s ( $=255/227 \times 2.488320 \text{ Gbit/s} \times 40$ ). The standard is expected to get mature in 2011/2012. For the high-speed services including 100 Gbit/s Ethernet, the virtual concatenation can implement the adaptation. But to improve the utilization

of fiber, virtual concatenation is not an efficient technology. It only improves the bit rate of each wavelength.

The serial 100G Dense Wavelength Division Multiplex (DWDM) technology is used to adapt the  $10 \times 10\text{GE}/4 \times 25\text{GE}$  100GE service to the OTU4 of 111.809973 Gbit/s through ODU4. The rate single-wave 100G is quite high. The requirements for physical-damage tolerance, such as Optical Signal Noise Ratio (OSNR) and PMD, are also increased. Special technology is required to decrease the baud rate of optical signals transmitted in the fiber to increase the damage tolerance. For example, advanced code modulation technology such as Quadrature Phase Shift Keying (QPSK), 8PSK, QAM, Orthogonal Frequency Division Multiplexing (OFDM), and polarization multiplexing/demultiplexing are used. The single-wave 100 GE has more restrict requirements for PMD and Color Dispersion (CD). Therefore, the receiving end may use the receiving/electrical processing mode to increase the tolerance of physical damage, including non-linear effect suppression, PMD and CD compensation. As a result, the single-wave 100GE can be transmitted in a mixed mode and upgraded smoothly in the 10G/40G network.

From a long run, the 100GE DWDM transport uses an integrated solution including polarization multiplexing, high-level coding modulation, coherent reception/electrical processing, ultra FEC. Then, the 40G optical network can be upgraded to the 100G system smoothly. The 100GE transport requires the support of high-speed optical devices. It is estimated that the high-speed optical devices will become mature in 2012.

For the key technologies mentioned above, 100 Gbit/s Ethernet is not only the technology for solving interface construction, the processing capability of the system should also be improved for the large-capacity switching system. High bandwidth traffic management and packet processing capability can provide line speed processing and forwarding. As a result, the carrier-class features are provided. ZXCME 9500 Metropolitan Area Network (MAN) Ethernet hardware platform developed

by ZTE Corporation currently supports 100 Gbit/s interface bandwidth. Sufficient redundancy and acceleration ratio are reserved. Line speed forwarding of a single 100 Gbit/s is supported without changing the switching network and system heat dissipation. Only a new line card is required. After the system upgrading, the single line card bidirectional 400 Gbit/s or above bandwidth is supported. At the same time, though the problem of a single system is solved, the application of 100G Ethernet interface is also restricted by the improvement of transport network technology. We still have a long way to go before the commercial application of 100GE.

## 2 Applications of 100G Ethernet

According to IEEE802.3ba Task Force (TF), the standard will be completed in the middle of 2010. The commercial application is determined by many factors.

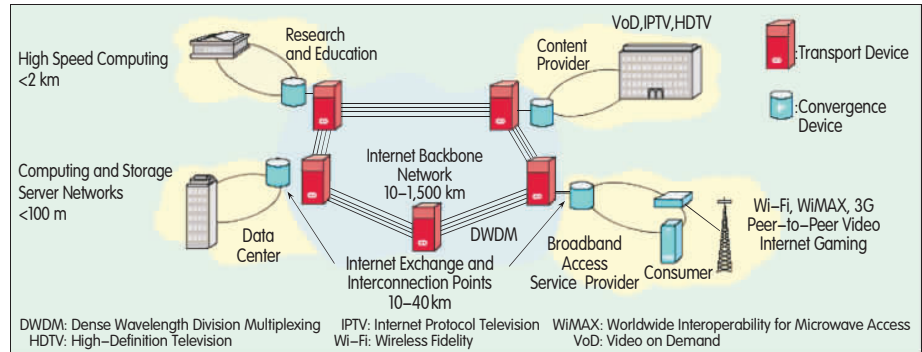
First, if the standard is mature, it should be driven by actual network demands. In addition, it should comply with the benefits of carriers. The major factors of bandwidth demands include: the increasing services are based on IP, like described of ALL IP. All IP packets are encapsulated in Ethernet frames from the source to the destination. The TDM over Ethernet technology is mature. It is compatible with traditional voice service. Ethernet encapsulation is simpler than Synchronous Optical Network (SONET)/ Synchronous Digital Hierarchy (SDH), and the cost is lower. These determine that the demands of upgrading Ethernet interface rate to 100 Gbit/s is urgent. In 100 Gbit/s Ethernet, the objective of "network communication is accelerated; the application efficiency is improved" can be realized. Fast access to applications stored in the data center, frequency bandwidth management, fast retrieval, compression, path optimization, and protocol acceleration can be implemented. Figure 2 illustrates the detailed scenarios<sup>[9]</sup>. For the application of convergence layer, the downlink port is switched to 10 Gbit/s, and the uplink can only use the link convergence of

10 Gbit/s port. If the 100 Gbit/s Ethernet interface is used, the management, assignment, and efficiency of data flow can be improved. For the data center, with the increase of 10 Gbit/s interfaces, the demands for uplink and internal interconnection high-speed interfaces also exist. The efficient transport of backbone network also expects the mature of 100G high-speed interface and transport.

When the P802.3ba standard is made, the relevant standards of electrical interfaces and the mature condition of the technology are taken into consideration. The inter-chip transport channel of 10.312 5 Gbit/s is used. Multiple-mode parallel optical interface supports the distance of 100m or more in OM3 fiber. Single-mode 40GB ASE-LR4 uses Coarse Wavelength Division Multiplexing (CWDM), which is cost-effective. 100GB ASE-LR4 uses DWDM, each wavelength transmits 25.781 25 Gbit/s, using 1,295–1,310 nm wavelength. Therefore, the original fiber can be used. Take the technology and cost into consideration, the technologies selected by the standard is practicable, which helps the commercial application of 100G interface partially or in the MAN.

For the application in overall network, before the 100GE transport standard and technology are mature, you can use the inverse multiplexing technology. Adapt the 100GE service of 10x10GE or 4x25GE interface to OTU2/OTU3 through ODU2/ODU3. Transmit the service in the 10G/40G optical network through multiple wavelengths. The current 10G/40G DWDM optical network need not be re-designed or changed. The transport code pattern is Optical Double Binary (ODB)/Differential Return-Zero (DRZ)/Electrical Return-Zero Differential Quadrature Phase Shift Keying (eRZ-DQPSK). The mode can use the current mature optical devices of 10G/40G. The performance index of the entire system is consistent with that of 10G/40G system. Through the solution, the smooth upgrade of the network can be implemented. It meets the expectation of the carrier for cost. In addition, the devices are mature<sup>[10]</sup>.

According to the current cost and demands, first applying the 100 Gbit/s Ethernet in the MAN is a practicable



▲ Figure 2. 100 Gbit/s Ethernet applications.

solution. In the MAN, a great deal of data should be uploaded and downloaded. A transport system requiring no compensation devices will simplify the network design. 100 Gbit/s Ethernet can meet the demand. At the same time, the high bandwidth meets the requirements of 40% increase in traffic of the MAN. In a word, the demands for developing 100 Gbit/s Ethernet is obvious. The advantage of cost will be enhanced. But the technology of modulation mode and OAM etc. in the Ethernet transport should be improved continually. We have a long way to go before the large-scale commercial application.

In addition, in the technical upgrade, besides 100 Gbit/s Ethernet, fiber channel, InfiniBand, and SONET are also involved in the 40/100 Gbit/s network. In the late 1990s, the decreasing rate of price of Ethernet port devices is double of that of the ATM and FDDI devices. Meanwhile, 40 Gbit/s and faster network share similar FPGA, SERDES, and coding technology. As a result, the devices corresponding to any protocol are hard to obtain the cost advantage through volume production. 100 Gbit/s Ethernet will not dominate the industry like before.

### 3 Conclusions

Generally speaking, the 100 Gbit/s Ethernet technology is vital and widely concerned. Except the challenge of technology and commercial application, the opportunity for involved parties is great. For the study organizations, it is an opportunity of discovery and innovation; for the device and module suppliers, it is a new and high-repayment; for the system provider, it is an opportunity of

dumping the tray and leading the market.

### References

- [1] IEEE 802.3ba 40Gb/s and 100Gb/s Ethernet Task Force, IEEE P802.3ba/D1.1 [S/OL]. 2008. (2009-07-20) [2008-12-12]. <http://www.ieee802.org/3/ba/>.
- [2] IEEE 802.3ba 40Gb/s and 100Gb/s Ethernet Task Force, IEEE P802.3ba/D2.1 [S/OL]. 2009. (2009-05-29) [2008-12-12]. <http://www.ieee802.org/3/ba/index.html>.
- [3] IEEE 802.3ba 40Gb/s and 100Gb/s Ethernet Task Force, BaselineSummary\_0908 [S/OL]. 2008. [2008-12-12]. [http://grouper.ieee.org/groups/802/3/ba/BaselineSummary\\_0908.pdf](http://grouper.ieee.org/groups/802/3/ba/BaselineSummary_0908.pdf).
- [4] CFP MSA, CFP-MSA-DRAFT, rev-1-0 [S/OL]. (2009-03-23) [2008-12-12]. <http://www.cfp-msa.org/Documents/CFP-MSA-DRAFT-rev-1-0.pdf>.
- [5] Altera 40-100Gbps Ethernet Solutions, Version 9.0 [EB/OL]. 2009. [2008-12-12]. [http://www.altera.com/literature/wp/wp-01080-stratix-iv-gt-40g-100g.pdf?WT.mc\\_id=ut\\_pr\\_al\\_ne\\_tx\\_213](http://www.altera.com/literature/wp/wp-01080-stratix-iv-gt-40g-100g.pdf?WT.mc_id=ut_pr_al_ne_tx_213).
- [6] Netlogic Calico NLP10142 100Gbps MLD Quad Channel Transceiver Preliminary Data Sheet [EB/OL]. (2008-11-03) [2008-12-12]. <http://www.netlogicmicro.com/4-news/pr/2009/09-03-19.htm>.
- [7] TORZA Anthony. Using FPGA Technology to Solve the Challenges of Implementing High-End Networking Equipment: Adding A 100 GbE MAC to Existing Telecom Equipment [S/OL]. 2008 (2008-09-23) [2008-12-12]. [http://www.xilinx.com/support/documentation/white\\_papers/wp280.pdf](http://www.xilinx.com/support/documentation/white_papers/wp280.pdf).
- [8] JONES M L. Report of Joint Q9/15 and Q11/15 Meeting [R/OL]. 2008. (2008-11-26) [2008-12-12]. <http://www.itu.int/md/T09-SG15-081201-TD-GEN-0037/en>.
- [9] ZHAO Yongpeng. 100G: Opportunities and challenges, and enabling technologies [EB/OL]. (2008-10) [2008-12-12]. [http://www.c-fol.net/focus/aoe/file/100GOpportunities%20and%20challenges,%20and%20enabling%20technologies\\_AOE2008.pdf](http://www.c-fol.net/focus/aoe/file/100GOpportunities%20and%20challenges,%20and%20enabling%20technologies_AOE2008.pdf).
- [10] 100G传输时代来临IT前沿 [EB/OL]. (2008-10-27) [2008-12-12]. <http://www.cnii.com.cn/20080623/ca510849.htm>.

### Biography

#### /// Zhang Yuanwang



Zhang Yuanwang graduated from Southeast University. He is engaged in Carrier Ethernet (CE) switch system design at ZTE Corporation. His research interest is 100G Ethernet. He has participated in the draft of IEEE802.3ba as a representative of ZTE. He has published five papers.