A scheme of optical interconnection for super high speed parallel computer

Youju Mao (毛幼菊), Yi Lü (吕 翊), Jiang Liu (刘 江), and Mingrui Dang (党明瑞)

Optical Communication Institute, Chongqing University of Posts & Telecommunications, Chongqing 400065

Received June 10, 2004

An optical cross connection network which adopts coarse wavelength division multiplexing (CWDM) and data packet is introduced. It can be used to realize communication between multi-CPU and multi-MEM in parallel computing system. It provides an effective way to upgrade the capability of parallel computer by combining optical wavelength division multiplexing (WDM) and data packet switching technology. CWDM used in network construction, optical cross connection (OXC) based on optical switch arrays, and data packet format used in network construction were analyzed. We have also done the optimizing analysis of the number of optical switches needed in different scales of network in this paper. The architecture of the optical interconnection for 8 wavelength channels and 128 bits parallel transmission has been researched. Finally, a parallel transmission system with 4 nodes, 8 channels per node, has been designed.

OCIS codes: 200.4650, 060.4230, 060.1810, 060.2360.

It is well known that for a parallel computer system, the higher the parallelism is, the more complicated the cross interconnection link will be. If the parallelism (i.e. parallel interface) is n and byte length of data is m, the fiber number of system is $n^2 \times m$. And the number for a super system is as high as tens of millions, even hundreds of millions, so that the switching control is very complicated and clock synchronization is also difficult[1]. In this paper, we aim to design an optical interconnection network based on coarse wavelength division multiplexing (CWDM) and data packet technology, which can be adopted to parallel computers with super high speed and large capacity. Compared with dense wavelength division multiplexing (DWDM) and bit transmission, the important difference lies in that the requirement of wavelength division multiplexer (WDM) performance for multiple wavelength optical system is much lower yet the transmission efficiency is guaranteed.

Applying CWDM technology, cross connection network[2] can be simplified greatly. What we apply is, firstly, to make bit streams from packaged CPU, with packet length of 16 bits. And then, 128 bits can be divided into 8 data packages. Then making use of CWDM with 8 channels, these parallel data packets can be multiplexed to a single optical fiber. Owing to speed of each CWDM channel for direct modulation can reach 2.5 Gb/s[3], the switch capacity of parallel computer systems can reach 20 Gb/s. Figure 1 is the transmission block diagram of our system.

CWDM is a multiple wavelength transmission technology with a larger wavelength interval than that of DWDM and is suitable for network with short transmission distance. So it is an optimal multi-wavelength multiplexing device for optical interconnection of parallel computer. The channel interval of CWDM is usually 20 nm.

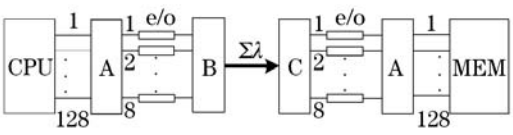

Fig. 1. CWDM & data packet uni-directional transmission. A: data packet processing; B: CWDM (MUX); C: CWDM (DEMUX).

Changing parallel bits into parallel data packets (IP packet) is propitious to apply IP transmission technology. Owing to applying WDM optical interconnection, digital wrapper technology aiming at IP over WDM should be adopted in our system. The structure of data packet is shown as Fig. 2.

Owing to added overhead bits appearing when applying data packet transmission, it is important to reduce added overhead of packet and simplify wrapper technology.

Since IP packet transmitted directly in optical fiber can simplify network configuration, if adopting optical cross interconnection (OXC) system, the transmission efficiency will be raised. In fact, OXC system is a kind of switching network[4,5], which can be realized by optical switch matrix. OXC can serve as an optical router and is capable of wavelength routing selection. Moreover, it can achieve non-blocking communication and let all wavelengths route intelligently between multi-CPU and multi-MEM.

To design an optical switch matrix, we should consider topology of network first of all. The topology will decide switching performance of the whole system, data throughput, and the design of protocols for system transmission, synchronism, and routing. A superior topology can provide flexible routing tactics, shorten delay of routing, and avoid transmission congestion to the utmost. Furthermore, the topology of network can decide the number of optical switches needed. In other words,

sign	A	B	C	D	E	sign
0+7e	8 bits	8 bits	data	16 bits	8 bits	0+7e

Fig. 2. Data packet format. A: destination address; B: control byte; C: information section; D: check sequence; E: reserved section.

it decides the cost of the whole system.

The virtue of 2×2 optical switch[6] lies on low insertion loss, low level of crosstalk, and high reliability. High order OXC network designed based on 2×2 optical switches features wide bandwidth, low time delay, and less congestion. It is an optimizing network using less optical switches than other networks.

A $n \times n$ high order optical switch adopting 2×2 switches to realize ordering can be considered as a binary-tree used to compare n numbers. And the number of exchange sorting is $n!$. Then, the minimum number of the needed 2×2 switches is $S(n) = [\log_2(n!)]$.

Binary-tree is an expandable connecting network topology, if there are K layers, then the number of nodes is $2^k - 1$, as shown in Fig. 3. An $n \times n$ optical switch has n inputs and n light paths that can be taken as n numbers which need to be sorted. The process of route switching is a process of sorting according to destination address. Each 2×2 optical switch can be used as a comparator for sorting, and n inputs behave as sorting for n numbers, if there is no repeated comparison, in accordance with graph theory, the number of sorting is $n!$ in all. According to graph theory, the least number of 2×2 optical switches is $S(n) = [\log_2(n!)]$. As a result, the number of routing for $n \times n$ optical switch and the number of 2×2 optical switches needed can be calculated, as listed in Table 1.

In order to verify aforesaid network structure, simulation for sorting is performed. Supposing there is no outputs conflicting (i.e., it will not happen that two signals want to reach the same output at the same time), as far as 8×8 sorting network is concerned, there are 8 input signals, and the number of permutations for input will be 40320. According to constructing principle of minimum permutation of network node, simulative network is realized by VC programs. In term of "Class" of C++ program, Switch Class is constructed and taken as a 2×2 optical switch in logic. It is the basic construction unit, which is in accordance with each input permutation,

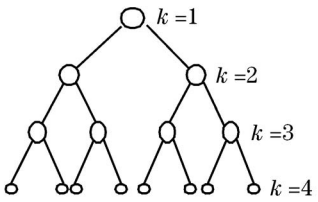

Fig. 3. Binary-tree topology.

Table 1. $n \times n$ OXC Route and Number of Switches

Input Nodes	Number of Input Rank $n!$	$\log_2(n!)$	Number of Switches
2	2	1	1
3	6	2.585	3
4	24	4.585	5
5	120	6.907	7
6	720	9.492	10
7	5040	12.299	13
8	40320	15.299	16

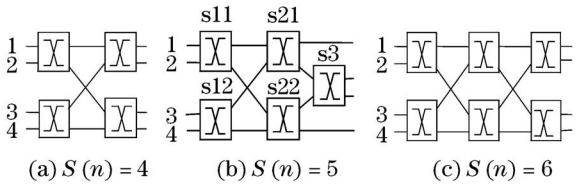

Fig. 4. 4×4 optical switch array.

makes input signals imported into simulative network and checks the output of the network. Each construction mode and the number of needed optical switches are separately discussed. When n is 4, in conformity to the above theory, it can be calculated that $S(n) = [\log_2(n!)]$ is 4.585 approximately. Different values can be assigned for $S(n)$ by increasing or decreasing the integer. Three cases of $S(n) = 4$, 5, and 6 are compared, as shown in Fig. 4.

During the process of simulation for network, optical switches serve as comparators. Supposing there is no outputs conflicting, four inputs are numbered according to their destination address (inputs marked 1, 2, 3, and 4), then the optical switch compares the number value of two input ports, the bigger one is sent to the lower output port and the other to the upper output port.

In Fig. 4(a), $S(n) = 4$, $2^4 = 16$ is the routing number of optical switch, which is less than the number of non-blocking routing for 4×4 network. Consequently, if adopting 4 optical switches, communication blocking will happen possibly, so that not all input permutations can pass through the network once. Optical switch array shown in Fig. 4(b) and (c) can provide 2^5 and 2^6 kinds of routing, respectively. On the basis of network structures shown as Figs. 4(b) and (c), if adopting certain strategy for routing selection, all input permutations can pass through the network once synchronistically. The topology based on binary-tree is applied, which can offer $n!$ kinds of non-blocking routing, as far as 4×4 network is concerned, 4! is 24. Thus it can be seen that 5 and 6 switches can also satisfy switching strategy of 4!, but 5 switches provide least invalid routing and simplify the switching network.

Five switches can offer 32 kinds of routing, but in 4×4 network, routing is decided by the number of input and output ports. As a result, only 24 different routings are provided for 4×4 optical switching network. That is to say, even though the number of optical switches changes, the routing based on topology of binary-tree will not alter. Accordingly, although five switches can offer 32 kinds of routing, only 24 different routings can be accordant with switching strategy for 4×4 network. The rest 8 kinds of routing will not appear to cause repetition.

According to the structure of optical interconnection system we researched, a parallel computer system adopting optical interconnection with 4 nodes, 8 channels (8-wavelength CWDM) and 8 data packets is designed for uni-direction transmission, as shown in Fig. 5. In Fig. 5, the system is made up of 3 parts: A) optical switch routing subsystem; B) optical IP packet WDM transmission subsystem; and C) electrical arbitral logical control subsystem. The 8 channel wavelengths applied are 1470, 1490, 1510, 1530, 1550, 1570, 1590, and 1610 nm.

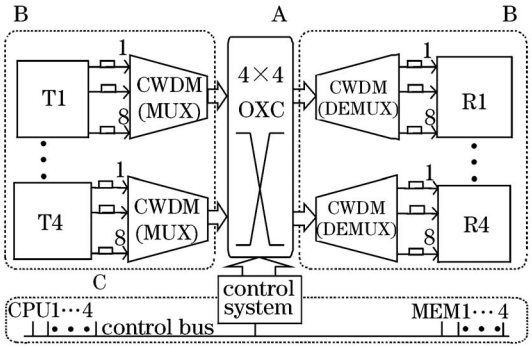

Fig. 5. Parallel transmission optical interconnection system.

The switching time of OXC is less than 20 ms. Optical switches are controlled concentratedly and unitedly, so the time delay of swap operations will not be accumulated, which can be applied in construction of super large scale optical switch array[7,8].

Test result indicates that all input permutations can reach the expected output without conflict, which is the same as our simulation. The control module can assign routing properly in terms of signaling from receiver, and the receiver module can accept data from transmitter module correctly.

In this paper, we proposed the scheme combining CWDM and packets parallel transmission. Being an application of IP over WDM in parallel computer, it not only decreases the technology difficulty and the cost for the system, but also improves the transmission performance. And it is propitious to realize transmission and switching for synchronous multiprocessing (SMP) shared memory parallel computer. Furthermore, it is meaningful for high efficiency data storage and super transmission for optical data in grid node of the next generation of Internet (NGI).

This work was supported by the Key Plan Item of Science & Technology of the Ministry of Information Industry, P. R. China under Grant No. 01XK511003. Y. Mao's e-mail address is maoyj@cqupt.edu.cn.

References

1. M. R. Dang, M. T. Zhou, and Y. J. Mao, Chin. J. Lasers (in Chinese) **28,** 932 (2001).
2. Y. J. Mao and M. R. Dang, *Technology of Wave Division Multiplexing* (People's P&T Press, Beijing, 1996) pp. 112−129.
3. W. J. Dai, H. Y. Zhang, and Y. Q. He, Chin. J. Lasers (in Chinese) **30,** 1095 (2003).
4. X. Y. Yang, M. R. Dang, Y. J. Mao, and L. M. Li, Chin. Opt. Lett. **1,** 266 (2003).
5. J. Wang, Q. Zeng, Z. Zhang, and H. Chi, Chin. Opt. Lett. **1,** 392 (2003).
6. L. W. Dong, X. N. Yan, K. Y. Shi, and J. K. Yan, Acta. Opt. Sin. (in Chinese) **7,** 787 (2003).
7. J. Y. Yang, X. Q. Jiang, F. H. Yang, M. H. Wang, and R. A. Chen, Chin. J. Lasers (in Chinese) **30,** 137 (2003).
8. M. P. Earnshaw, J. B. D. Soole, M. Cappuzzo, L. Gomez, E. Laskowski, and A. Paunescu, IEEE Photon. Technol. Lett. **15,** 810 (2003).