

**DISTRIBUTION OPTICAL ELEMENTS AND
COMPOUND COLLECTING LENSES FOR
BROADCAST OPTICAL INTERCONNECT**

CROSS-REFERENCES TO RELATED
APPLICATIONS

[0001] This application claims a benefit of priority under 35 U.S.C. 119(e) from provisional patent application U.S. Ser. No. 60/813,970, filed Jun. 14, 2006 and is also a continuation-in-part of, and claims a benefit of priority under 35 U.S.C. 120 from copending utility patent application U.S. Ser. No. 11/796,133 filed Apr. 25, 2007, which in-turn is a continuation-in-part of, and claims a benefit of priority under 35 U.S.C. 120 from copending utility patent application U.S. Ser. No. 10/702,227 filed Nov. 5, 2003, which in-turn claims a benefit of priority under 35 U.S.C. 119(e) from both provisional patent application U.S. Ser. No. 60/423,939, filed Nov. 5, 2002 and provisional patent application U.S. Ser. No. 60/432,141, filed Dec. 10, 2002, the entire contents of all of which are hereby expressly incorporated herein by reference for all purposes.

BACKGROUND OF THE INVENTION

[0002] 1. Field of the Invention

[0003] The invention relates generally to the field of optical interconnects for computer systems and/or their subsystems as well as networks and/or their subsystems. More particularly, the invention relates to a free-space optical interconnect that includes a fan-out and broadcast signal link.

[0004] 2. Discussion of the Related Art

[0005] The concept of parallel-distributed processing (PDP), which is the theory and practice of massively parallel processing machines, predates the first supercomputers of the 1960s. In practice, high-performance parallel-distributed processing machines are difficult to achieve for several interrelated reasons. On the physical side of the equation, interconnections between n processors or nodes increase as the square of the number of processors (n^2); the physical bulk increases as n for the packaging and n^2 for the interconnecting wiring; latency due to capacitance increases as the average distance between nodes, which is also proportional to n ; heat-removal difficulty increases as the square root of the number of processors ($n^{1/2}$) due to the surface-to-volume ratio. On the logical side of the equation, message overhead is constant for broadcast mode and can increase as n for relay mode. The impact on software is roughly proportional to n^2 due to the increased complexity of parallel-distributed processing algorithms. The overall cost per node increases more rapidly than the number of nodes when all these factors are considered. What is needed is a method of parallel-distributed processing, design and operation that overcomes some or all of these scaling problems.

[0006] The present record holder in performance is NEC's "Earth Simulator" topping out at 35.86 teraflops (a teraflop is 1000 gigaflops and a flop is a floating-point operation while "flops" usually refers to a flop per second). While there are many interesting and novel entries in today's supercomputer marathon, the Department of Energy's Advanced Simulation and Computing Initiative (ASCI) has sponsored several of the top contenders. The latest of these

is a fifth-generation ASCI system to be built by IBM. The ASCI Purple (AP), if on time and within budget, will arrive by 2005 at a projected cost of approximately \$550 per gigaflop with an ultimate option to have a 100-teraflops performance figure in a single machine. (A gigaflop is one billion operations per second.) This is about 12 times the performance of the previous ASCI Q and ASCI White machines. By contrast, a present-day personal computer is typically priced about \$750/GF (the minimum cost is probably about \$500/GF, i.e., actually less than the ASCI Purple.) This clearly shows that economies of scale are nonexistent to marginal given the factor of nearly 13,000 increase in the number of processors required to achieve the 100 teraflop (TF) figure. (A teraflop is 1000 gigaflops.) The ASCI Purple (AP) is estimated to weight in at 197 tons and cover an area of two basketball courts (volume not specified). The AP will have 12,433 Power5 microprocessors, a total memory bandwidth of 156,000 GBs (gigabytes per seconds), and approximately 50 terabytes (million megabytes) of memory. Power dissipation will be between 4 and 8 MW (megawatts), counting memory, storage, routing hardware and processors.

[0007] IBM's Blue Gene3/L (BGL), based on that company's system-on-chip (SOC) technology, will take up four times less space and consume about 5 times less power, it is expected to perform at the 300 to 400 teraflops level. The cost per gigaflop will be about the same at about \$600/GF as above. Each of the 65,000 nodes in the BGL will contain two Power PCs, four floating-point units, 8 Mbytes of embedded DRAM, a memory controller, support for gigabit Ethernet, and three interconnect modules. The total number of transistors is expected to be around 5 million, making for a large, expensive, and relatively power-hungry node. The interconnect topology is that of a torus, where each node directly connects to six neighbors. For synchronizing all nodes in the system, hardware called a "broadcast tree" is necessary. Establishing broadcast mode to begin a computation, for example, will require several microseconds. To round out the hardware complement of a node, nine memory chips with connectors (for a total of 256 Mbytes) are foreseen. Four nodes will be placed on a 4 by 2-inch printed-circuit card.

[0008] Reliability in these existing machines is a major concern when there are from hundreds-of-thousands to millions of material interconnections (e.g., wires, connectors, solder joints, contact bonding). What is needed is an approach to super computer design that increases reliability.

[0009] Moreover, the main, unsolved problem facing today's supercomputers is how to achieve the economies of scale found elsewhere in the industrial world. Machines with tens of thousands of processors cost as much per gigaflop as commodity PCs having only a single processor. Part of the reason for this lack of progress in supercomputer scaling is that the interconnect problem has not yet found a satisfactory solution. Adopting present solutions leads to a reliance on slow and bulky, off-chip hardware to carry the message traffic between processors. A related problem is that communication delays increase as the number of nodes increases, meaning that the law of diminishing returns soon sets in. This issue drives the industry to faster and faster processing nodes to compensate for the communications bottleneck. However, using faster and more powerful nodes increases both the cost per node and the overall power consumption. Smaller, slower, and smarter processors could