Optimum Connectivity in the Multi-core Environment

Gilad Shainer, M.Sc., Mellanox Technologies
John Benninghoff, Intel
Lutfor Bhuiyan, Intel

"In 1978, a commercial flight between New York and Paris cost around $900 and took seven hours. If the principles of Moore's Law had been applied to the airline industry the way they have to the semiconductor industry since 1978, that flight would now cost about a penny and take less than one second." (Source: Intel)

In 1965, Gordon Moore predicted that the number of transistors that could be integrated into a single silicon chip would approximately double about every two years. For more than forty years Intel has been transforming that law into reality (See *Figure One*). The increase in transistor density enables more transistors on a single chip and therefore increases in the CPU performance. However, it is not the only factor driving the CPU performance, as the increase of the CPU clock frequency, a bi-product of the transistor density was an important factor in the overall performance improvement.

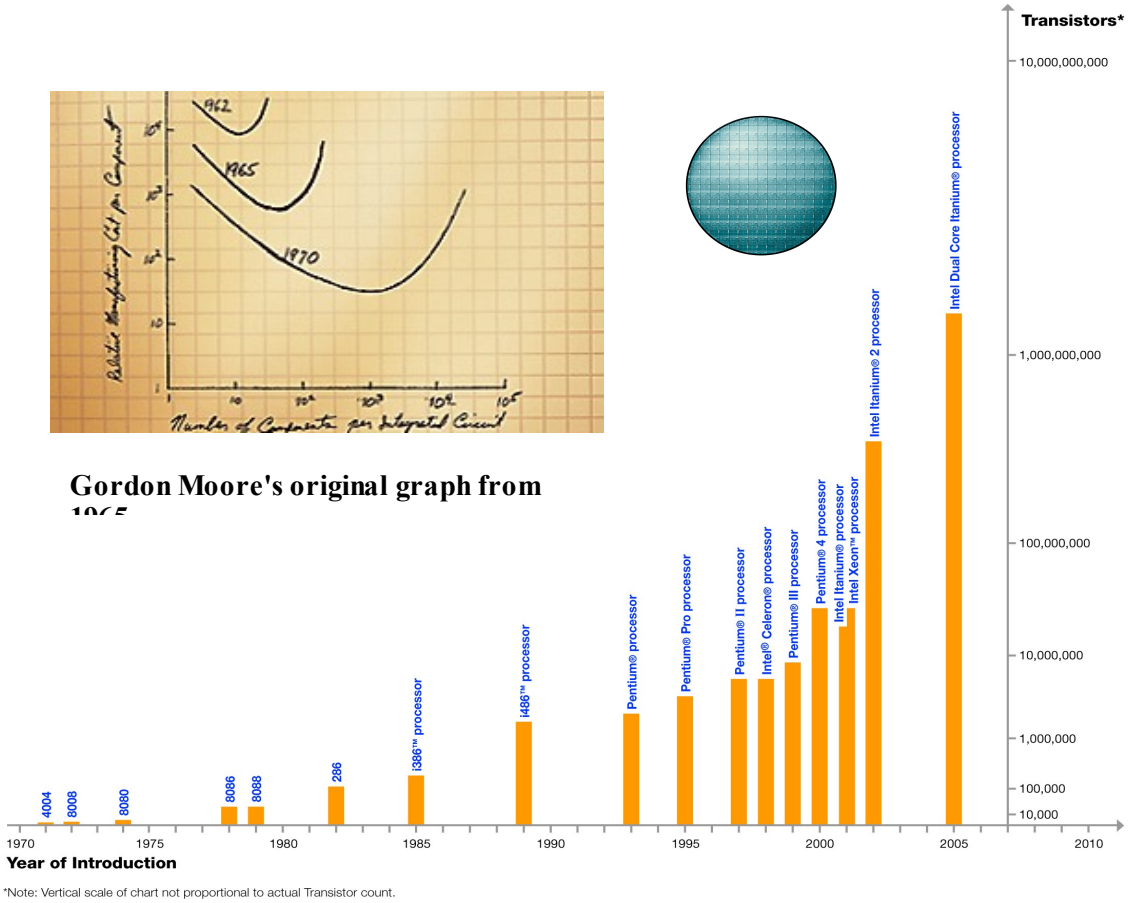

Figure One: Growth of transistors has followed Moore's Law for forty years

"Another decade is probably straightforward. There is certainly no end to creativity", Gordon Moore, 2003. Moore's Law is expected to deliver increasing transistor densities for at least the near future but power consumption and heat generation, which rise exponentially with clock frequency, will limit the increase in the CPU clock frequency.

High-performance computations are rapidly becoming a critical tool for conducting research, creative activity, and economic development. In order to provide intense computing platforms and still maintain the historic rates of performance and price/performance improvements, more execution cores are being integrated into each CPU. With multiple cores executing simultaneously, CPU clock frequency can be reduced in order to contain heat generation, while still increasing total system performance. This mega-trend, shown below in *Figure Two, is* one of three trends that are shaping the technical computing market – clusters, multi-core environments, and high-performance industry standard interconnects.

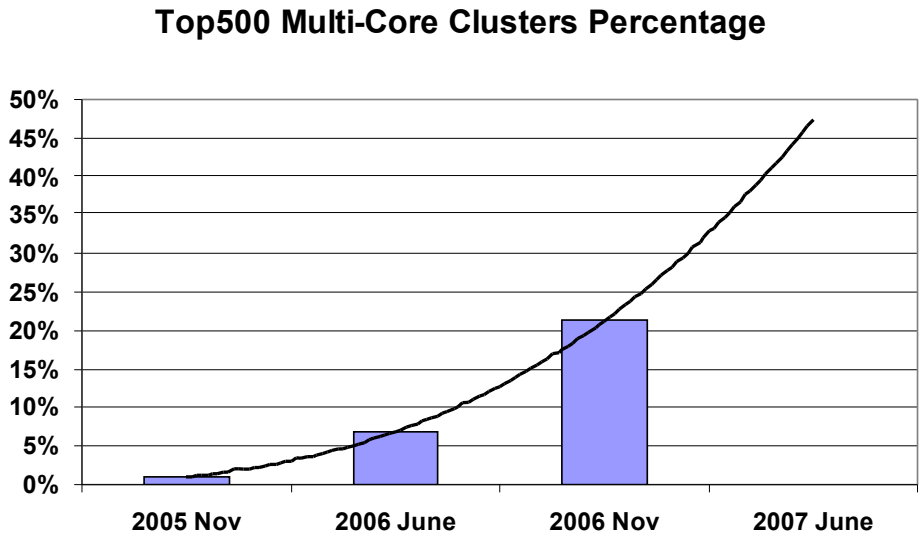

Figure Two: Growth of Multi-core Clusters

Connecting Multi-core Platforms

Efficient data transfer between clustered compute nodes is critical for balanced system performance. In a balanced system, the overall performance is equal to or greater than the sum of its components, while in a non-balanced system, the performance is less than the sum. The challenge of achieving balanced performance becomes more evident in multi-core environments. A multi-core environment introduces high demands on the cluster interconnect and the interconnect needs to be able to handle multiple I/O streams simultaneously.

By providing low-latency, high-bandwidth and extremely low CPU overhead, InfiniBand is emerging as the most deployed high-speed interconnect, replacing proprietary or low-performance solutions. In a multi-core environment, it is essential to avoid interconnect

protocol processing in the CPU cores. In order to maximize the overall compute cluster efficiency and to allow performance-hungry applications to efficiently utilize the CPU's core resources, a fully hardware transport-offload solution is needed. Furthermore, un-necessary overhead on the CPU cores reduces the ability of balanced computing between the various cores, leading to higher degradation in real application performance.

Interconnect flexibility is another requirement for multi-core systems. As various cores can perform different tasks, it is necessary to provide remote direct memory access (RDMA) along with the traditional semantics of Send/Receive. RDMA and Send/Receive in the same network provides the user with a variety of tools that are crucial for achieving the best application performance and the ability to utilize the same network for multiple tasks, such as compute, storage and management.

Mellanox InfiniBand provides both the flexibility and a full hardware transport-offload implementation. Transport-offload capabilities enable various applications and software interfaces, such as Message Passing Interface (MPI) to use overlapping of CPU computations with the interconnect communication cycles to reduce run time of MPI-based applications and to increase application performance. Mellanox's hardware implementation also provides quality of service (QoS), so different I/O streams could be served as required by the application.

Applications Demand a High-speed Interconnect

Computational Fluid Dynamics (CFD) is one of the branches of fluid mechanics that uses numerical methods and algorithms to solve and analyze problems that involve fluid flows. At the core of any CFD calculation is a computational grid, used to divide the solution domain into thousands or millions of elements where the problem variables are computed and stored.

FLUENT, a leading commercial software provider for solving fluid flow problems, implemented flexible parallel processing capabilities in order to effectively utilize the multi-core environments. Dynamic load balancing automatically detects and analyzes parallel performance and adjusts the distribution of computational cells among the processors and the server nodes. The following chart compares Mellanox InfiniBand and Gigabit Ethernet using FLUENT FL5L benchmark, on Intel dual-core Xeon 3GHz 5100 series (code name Woodcrest) server

Mellanox InfiniBand delivers superior performance than Gigabit Ethernet - up to 155% higher performance on 128 CPU cores - due to InfiniBand's proven efficiency and super-linear scaling capabilities.

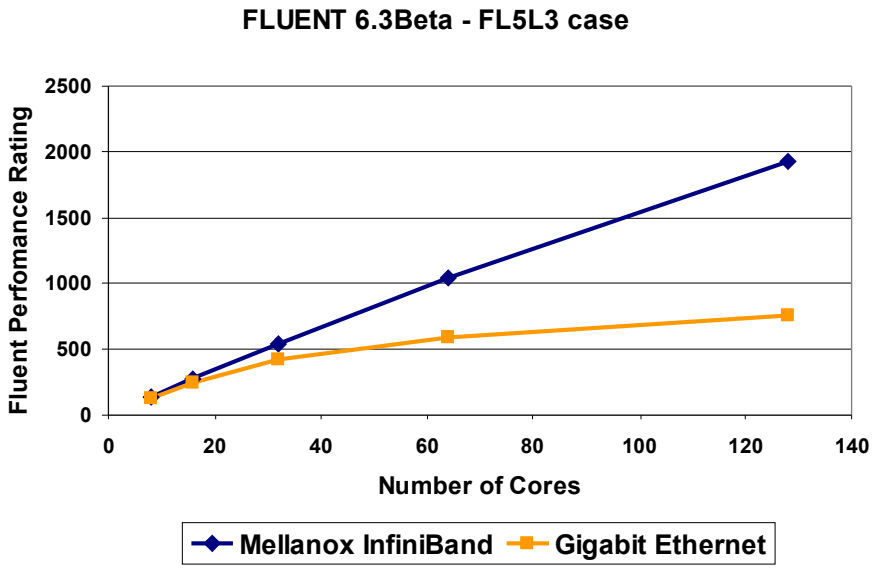

Figure Three: Effect of interconnect over increasing numbers of nodes

In order to determine the importance of the interconnect architecture for multi-core environments, the same benchmark was used to compare between single-core Xeon 3.4GHz and dual-core Xeon 5100 series 3GHz (Woodcrest). In both cases, InfiniBand shows higher performance, but the difference between Mellanox InfiniBand and Gigabit Ethernet increases on the multi-core setting (See *Figure Four* Below). In order to meet the requirements of each CPU core, multi-core servers demand higher I/O throughput from the interconnect solution. InfiniBand is proved to provide the aggregate CPU cores demands, while Gigabit Ethernet fails to do so.

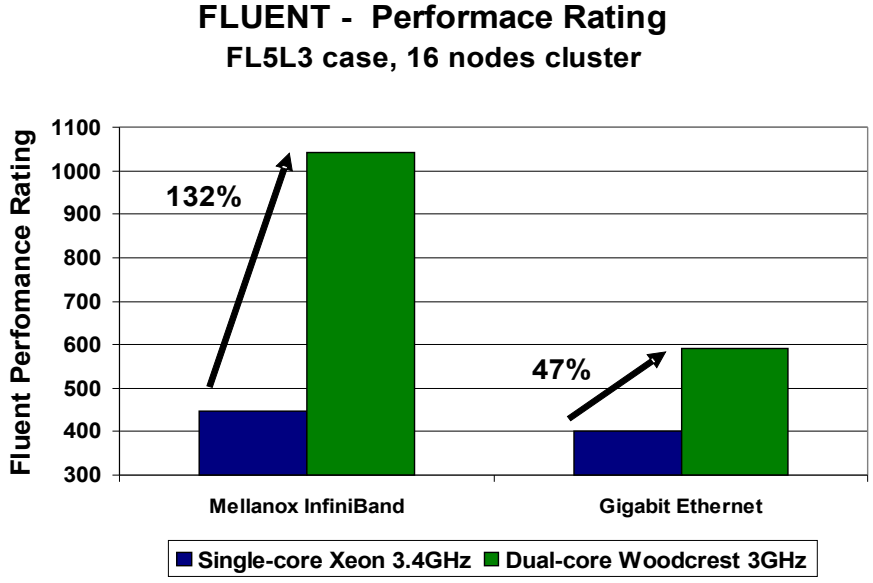

Figure Four: Effect of Interconnect on Fluent perforamnce rating

As dual-core environments introduce higher I/O requirement than single-core systems a high throughput interconnect with low CPU overhead is vital in order to maintain high CPU and application efficiency. The recent introduction of Intel quad-core environments will further increase this demand.

Multi-core environments increase the demand for I/O throughput, low-latency, low CPU overhead, flexibility and high-efficiency in order to maintain a balanced system and to achieve high application performance and scaling. Low-performance interconnect solutions, or lack of native hardware support, will result in degraded system performance. Mellanox high-speed InfiniBand meets the multi-core system requirements and provides a balanced compute solution with Intel multi-core technology.

The author would like to thank John Benninghoff (Intel Corporation) and Lutfor Bhuiyan, (Intel Corporation) for their contributions during reviews of this article.

Gilad Shainer is a senior technical marketing manager at Mellanox technologies focusing on high performance computing. He joined Mellanox Technologies in 2001 to develop Mellanox's InfiniHost PCI-X Host Channel Adapter (HCA) device and later led the development of Mellanox's InfiniHost III Ex PCI Express HCA device. Gilad Shainer holds MSc. degree (2001, Cum Laude) and a BSc. degree (1998, Cum Laude) in Electrical Engineering from the Technion Institute of Technology in Israel. He is also a member of the PCISIG PCI-X and PCI Express Working Groups and has contributed to the definition of the PCI-X 2.0 specifications.

