Server I/O Considerations
By Seamus Crehan
Crehan Research, Inc.
January 2012

White Paper – sponsored by Broadcom Corporation

Market Overview:
A number of factors are falling into place to usher in some of the most significant changes and opportunities ever seen in the server I/O market. An unprecedented level of I/O demand and associated challenges are coming from a confluence of factors that includes:

- virtualization
- a step function up in the ratio of network connected devices (laptops, desktops, netbooks, smartphones and tablets) to server units
- richer traffic content (most notably video)
- increase in server to server communications
- the use and incorporation of Cloud business models
- the capture and extraction of value from massive data sets i.e., “Big Data”
- convergence of multiple traffic types that once resided in separate datacenter silos
- a major server upgrade cycle with the introduction and ramp of Intel-based Romley/Sandy Bridge servers

We are already starting to see the impact of the amount of network traffic generated and the explosion in the number of networked devices on both server unit shipments and server networking shipments. Despite the fact that virtualization enables multiple virtual servers on one physical server, server units have continued to grow. Furthermore, server-class networking ports have grown at a much higher rate and we are on track for total shipments to grow in excess of 20% for 2011.
In response to these server and storage I/O challenges, a number of new technologies and solutions have been introduced, and many of them have reached a level of maturity, performance and cost effectiveness to likely see broad adoption. These include: 10 Gigabit Ethernet (10 GbE) server networking solutions; NIC Partitioning, Data Center Bridging/Lossless Ethernet, Fibre Channel-over Ethernet (FCoE) and iSCSI (over Lossless Ethernet), and Modular/Flexible LAN-on-Motherboard (LOM)/Daughter Cards.

**10 GbE**: Although 10 GbE server-class adapters and LOM are seeing growth rates that are very strong and recently hit a port shipment run rate above one million per quarter, it still comprises a very small part of the overall server-class I/O networking market (see figure 1). However, with the confluence of steep average selling price declines, broader vendor product offerings, and higher network traffic demands, this technology should move toward mainstream adoption in conjunction with the ramp of the Romley/Sandy Bridge servers.

Furthermore, 10GBASE-T will emerge as one of the key interconnects in the enablement of broader 10 GbE adoption. The factors that facilitated the last major BASE-T upgrade – 1 GbE – are now falling into place for 10GBASE-T. These include, getting below an acceptable power threshold; an installed base of cabling that could handle that speed; LOMs on rack servers; a broad offering of server access switches with good port densities; and a bandwidth discount. As can be seen on figure 2, I anticipate that, as we enter 2015, 10GBASE-T will comprise the majority of 10 GbE connections.

**NIC Partitioning**: The advent of virtualization has resulted in many users deploying multiple network connections per server – most often numerous 1 GbE connections. The arrival of 10 GbE offered a potential solution to consolidate these connections and save on cabling but initial pricing was too high for most users.

Given the diversity of traffic that a converged 10 GbE can now handle, and taking into consideration that users may not yet require close to that full amount of bandwidth, some 10 GbE adapters (such as Broadcom’s) offer NIC partitioning or separate Virtual Fabrics enabling a single 10 GbE port to be divided into a number of separate ports. This further allows the cost of the 10 GbE connection to be amortized over numerous different applications.

**Data Center Bridging (DCB)/Lossless Ethernet**: One of the reasons that Ethernet has been so successful and has attained such widespread deployments in many different network types is due to its incorporation of features to address the needs of various markets. Past examples include Carrier Ethernet and Voice-over-Internet-Protocol (VoIP). In both of these cases users were able to take advantage of the
ubiquity, familiarity and cost effectiveness of Ethernet in addition to the specific management tools and qualities of service required for these specific networks.

The latest example of this is DCB/Lossless Ethernet which makes Ethernet a much more compelling data center fabric especially for enterprise-class Storage Area Networking (SAN). DCB/Lossless Ethernet incorporates four technology solutions: Enhanced Transmission Selection (ETS – IEEE P802.1Qaz); Priority-based Flow Control (PFC – IEEE P802.1Qbb); Quantized Congestion Notification (QCN – IEEE P802.1Qau); and DCB eXchange (DCBX, IEEE P802.1Qaz). These four features allow Ethernet – traditionally a best effort fabric subject to some packet-loss – to operate as a lossless fabric. As a result Ethernet can now deliver high quality SAN traffic.

Most new adapters and switches targeted for datacenter deployments incorporate DCB/Lossless Ethernet features (or will). Hence, anyone deploying recent server adapters and switch product offerings has the opportunity to build an end-to-end lossless and deterministic Ethernet Datacenter Fabric.

**FCoE and iSCSI over Lossless Ethernet.** The advent of DCB/Lossless Ethernet discussed above has helped to enable two high quality, enterprise-class, Ethernet SAN options – namely FCoE and iSCSI. FCoE, which saw first product introductions in 2008, encapsulates Fibre Channel frames into Ethernet frames, thus preserving the upper layers of Fibre Channel but transporting it over Ethernet. FCoE interoperates with existing Fibre Channel allowing for investment protection. FCoE-capable adapters and LOMs have seen very strong growth (and average price declines) since introduction. Furthermore, with the introduction of Romley/Sandy Bridge servers, a much higher portion of servers will ship with native FCoE capabilities.

iSCSI, the other block-based storage protocol, has been in existence much longer than FCoE. In fact, it is often seen in 1 GbE deployments whereas the FCoE starting speed is 10 GbE. iSCSI runs over TCP/IP and has, to date, gained most traction in the small-and-medium business or branch office space as a low-cost, best-effort block I/O storage protocol running over existing Ethernet LAN equipment. However, with the advent of DCB/lossless Ethernet iSCSI now offers an alternative high performance block I/O storage solution over 10 GbE.

**Modular LOMs/Daughter Cards.** With the advent of Romley/Sandy Bridge we will likely see a broad array of rack servers offer Modular LOMs/Daughter Cards. Given that there are numerous choices around server I/O technologies and various tradeoffs between these, server OEMs will offer customers choices on what kind of LOM they want in their server. Choices will vary between 1 GbE, 10 GbE, BASE-T, SFP+, FCoE, iSCSI, etc. Crehan Research expects that over the next few years this form-factor will become very popular on rack servers. Although we will see a very strong ramp in 10 GbE, the introduction of a Modular LOM/Daughter Card will have the effect of prolonging the lifecycle of 1 GbE. This is because 10 GbE won’t be the embedded default option on rack servers as it was when we saw the transition to 1 GbE in the 2001 to 2003 timeframe. Therefore, both 1 GbE and 10 GbE will be important technologies in the Romley/Sandy Bridge server upgrade.
Considerations and Applications:

Given the above backdrop, there is a lot to consider in datacenter networking choices moving forward. One of the biggest choices will be whether to deploy 10 GbE server networking. Users will also need to evaluate whether to start moving down the path toward a single converged network, and what might be the best choices here. For SAN, the potential benefits of FCoE and iSCSI over DCB/Lossless Ethernet are other important considerations.

While initial 10 GbE pricing was too high for most users, prices have fallen significantly as a result of component cost reductions, multi-port solutions and an increase in the mix of copper-based products (KR which is the internal blade connection, SFP+ Direct-Attached and more recently 10GBASE-T). Furthermore, an increase in the mix of less expensive form factors such as mezzanine cards for blade servers and LOM, as opposed to stand-up cards which dominated the early market product mix, have further reduced average selling prices. As a result, the average price of a 10 GbE server network connection is less than one quarter of what is was at the beginning of 2008.

One effective way to compare prices across speeds is to look at them on a dollar-per-gigabit-of-bandwidth basis. In looking at figure 3 we can see that 10 GbE now offers a significant bandwidth price discount over 1 GbE – across all network I/O form factors; whether it is stand-up cards, mezzanine cards or LOM. (Figure 3 doesn’t take into account additional savings from things such as fewer network cables).

It is not only the server adapters and LOMs where pricing is becoming much more compelling. A similar trend is also happening on the server access switches that connect to these adapter and LOMs, resulting in a significant reduction in the price of the total 10 GbE connection. In looking at the average selling price of a combined switch and server network connection on figure 4 we can see that what was once a very large price per port gap between 10 GbE and 1 GbE has narrowed considerably – currently offering about a 60% discount over a 1 GbE connection on a dollar-per-gigabit of bandwidth basis – and will continue to narrow going forward.

In conjunction with the ramp of Romley/Sandy Bridge Servers in 2012, we will not only see many more offerings of 10GBASE-T stand-up cards, but we will also see the first volume offerings of 10GBASE-T LOM hit the market. Using standard RJ45 Ethernet jacks and running over the now-widely deployed Cat 6-type cabling, this technology is backward compatible, auto-
negotiating with the current installed base of Gigabit Ethernet BASE-T networking infrastructure. This allows for a phased network upgrade if needed, where 10GBASE-T switches don’t have to be bought in unison with 10GBASE-T servers and vice versa. 10GBASE-T also offers greater flexibility over current Direct Attached Copper (DAC) technologies in that it can run longer distances, for example 100 meters over Cat-6a cabling. For past Ethernet technologies, such as 1 GbE, BASE-T was a key enabler of the hockey stick, in terms of broad-based deployments.

In addition to significant bandwidth discount and network flexibility, 10 GbE allows users to move toward a converged network – where traditional LAN and SAN traffic or traditional LAN and High Performance Computing (HPC) traffic can be carried over a single connection and common networking infrastructure. 10 GbE has also incorporated additional features to handle these traffic classes, for example, DCBX (Data Center Bridging eXchange)/Lossless Ethernet mentioned above for SAN and Remote Direct Memory Access (RDMA) over Converged Ethernet (RoCE) or iWARP for HPC.

Although there will be hurdles in moving toward a converged network – such as separate organizational silo structures, and technology familiarity – all signs point to this eventually happening. In the face of greater network demands from virtualization and traffic growth, the potential capital and operational cost savings are compelling. Furthermore, history tells us that, despite early obstacles users will choose to migrate toward a converged (Ethernet) Network. Enterprise networks all consolidated on to Ethernet from ATM, Token Ring and FDDI. Carrier Networks are well on their way to consolidating from Frame and ATM onto Ethernet. Enterprise voice networks also followed this path, converging onto the Ethernet network with the advent of VoIP.

Regardless of the timing of the transition, the move to a single converged infrastructure will, most likely, be a phased approach. 10 GbE and its enhanced data center feature set allows users to take that phased approach. For example, to converge separate Fibre Channel SAN and Ethernet LAN infrastructure onto a 10 GbE/FCoE infrastructure users can start at the server access layer only, in conjunction with the deployments of new Romley/Sandy Bridge servers, realizing savings from reduced adapter, switch and cabling infrastructure. They then could connect this into their existing FC SAN infrastructure. For users who want to deploy block storage but don’t want to deploy FCoE, iSCSI is a potential alternative. Unlike FCoE which first appeared on 10 GbE networks, iSCSI has been popular on 1 GbE Ethernet networks for many years, and the advent of DCB/Lossless Ethernet gives this technology more enterprise-class performance and resiliency.

It is a fair assumption that we are going to see: converged networks (with Ethernet as the fabric); we are going to see more and more virtualization; and we are going to see traffic bandwidth requirements continue to grow. 10 GbE now offers a versatile, flexible, and cost effective way to prepare our networks for this.
**Broadcom Spotlight:**

We talked earlier about the amount of connected devices (PCs, laptops, tablets, smartphones etc.) and the increase in the amount of network traffic per device. Broadcom’s semiconductors reside inside these devices, enabling them to connect to various networks, whether it is enterprise, carrier or home – wired and wireless. Furthermore, these networks, and the back-end IT infrastructure running behind them, are often powered by Broadcom Ethernet connectivity solutions.

Broadcom, which celebrated its twentieth anniversary in 2011, is a Fortune 500 company, with over 8,100 U.S and Foreign patents to its credit, and another 7,600 pending applications. Broadcom’s communication chips are behind much of the digital communication revolution that gained traction in the 90’s and continues to expand rapidly. It has shipped almost two billion chips since its inception, addressing every major segment of the broadband communications market.

Broadcom has, and continues to leverage its leading intellectual property portfolio across richly diversified product offerings. Demonstrating an excellent record of execution Broadcom has integrated numerous features and technologies in delivering “combination” or highly-integrated products. Examples include the BCM4329 for tablets, smartphones, and connected devices, and the BCM56840 for datacenter Ethernet switch connectivity. The BCM4329 integrates a complete IEEE 802.11 a/b/g/n system (MAC/baseband/radio) with Bluetooth® 2.1 + EDR (Enhanced Data Rate), and FM radio receiver and transmitter – combining several proven wireless technologies onto a single silicon die. The BCM56840 enables up to 64-ports of 10 Gigabit Ethernet (10 GbE) with the versatility to also support 40 Gigabit Ethernet (40 GbE). Furthermore, this product has integrated SerDes, and scalable services such as DCB/Lossless Ethernet and FCoE, allowing for unified storage and networking in the one switch chip. This is one of the key products enabling the lower 10 GbE datacenter switch pricing discussed earlier.

Broadcom groups its diverse portfolio of products into three broad categories, aligned with broad target market segments:

1. **Broadband Communications (Solutions for the Home)** — Highly integrated solutions for the connected home, including set-top-boxes and media servers, residential gateways, home networking, femtocells, xDSL and xPON integrated access devices, and OTT media appliances.

2. **Mobile & Wireless (Solutions for the Hand)** — Low-power, high-performance and highly integrated solutions powering the mobile ecosystem, including Wi-Fi and Bluetooth, cellular modems, personal navigation and global positioning, near field communications, multimedia and application processing, and mobile power management solutions.

3. **Infrastructure & Networking (Solutions for Infrastructure)** — Highly integrated solutions to carriers, service providers, enterprises, small-to-medium businesses and data centers for network infrastructure needs, including switches and physical layer (PHY) devices for local, metropolitan, wide area and storage networking; switch fabric solutions; and high-speed controllers.
Within Broadcom’s high-speed controller business are a broad set of 1 GbE and 10 GbE controllers for servers, storage target arrays, embedded devices, workstations, desktop and laptop/notebook computers. Broadcom took the lead in shipping the first high volume 10 GbE LOM controllers for blade servers in late 2008. As we approach a broader upgrade and transition to 10 GbE server-class networking that encompasses rack form-factor in addition to blade form-factor, it is worth acknowledging the role that Broadcom controllers have played in bringing the industry to this point.

Historically, when looking at the 100 Mbps to 1 GbE transition, it was early 2001 when Dell shipped the industry’s first servers with Broadcom’s 1 GbE LOM – NetXtreme 1 GbE embedded on the PowerEdge™ 2550 motherboard. This was the start of what would become a very quick transition to 1 GbE server networking. This was evidenced by Broadcom having shipped its 10 millionth 1 GbE-over-copper port in early 2003, fueled by 1 GbE server networking adoption and the switches connecting those servers to the network. Despite all its success, 2003 was also the year that Broadcom launched its first ever advertising campaign (A full 12 years after the company was founded, and an indication of its strong focus on “behind-the-scenes” OEM enablement and value creation).

Less than two years after its first 1 GbE server LOM, Broadcom would start to drive the 1 GbE transition on desktops and laptops, delivering products such as the BCM5705M which was offered standard on every Compaq Evo™ 620c Notebook PC from HP. Of course, since then Broadcom has significantly enhanced its 1 GbE offerings which continued to be deployed in high volumes. Enhancements include Energy Efficient Ethernet (EEE), and power reduction. EEE allows for a lower energy consumption state that can be adopted during network idle time or low utilization. Broadcom has further bolstered its EEE offerings with advanced policies to provide deeper power savings in its controllers. Broadcom’s incorporation of EEE in conjunction with controller and PHY power reduction are enabling more efficient and greener datacenters.

Recognizing early, a trend toward converged data center networks, and leveraging its integration ability, Broadcom introduced its Converged Network Adapter (CNA) controller family in 2004. Supporting the, then-new, PCI Express® bus architecture, this product family of 1 GbE controllers supported a TCP/IP offload engine (TOE), iSCSI, and Remote Direct Memory Access (RDMA) on a single-chip platform. This product gave OEM’s the ability to offer their customers what would become an ever-more compelling value proposition – the ability to simultaneously run network, storage and cluster/HPC traffic over a single Ethernet fabric.

This trend of early market OEM enablement continued as we moved toward 10 GbE. In Sept 2008, HP announced one of the first server blades specifically designed to host virtual machines, the ProLaint BL 495c. A key component of this of this platform’s virtualization optimization was enabled by Broadcom’s 10 GbE controller. (This new server line from HP also included Broadcom’s 10 GbE switching technology for its embedded blade switches).

In 2010, Broadcom sampled the industry’s first quad-port 10 GbE Controller (BCM57840) and in 2011, Broadcom shipped its fifth-generation CNA. This product (BCM57712) offers FCoE, PCI SIG Single Root IO Virtualization (SR-IOV), DCBX, along with iSCSI, on-chip TOE, and RDMA. This EMC E-Lab and NetApp Ethernet Advantage validated platform, which delivered an industry-leading 1.7M IOPs, was immediately offered by Dell and Cisco. In addition to various server deployments, this class of product enables high performance storage in 10 GbE environments. As the insatiable demand for
storage continues, and 10 GbE becomes an increasing popular choice in these implementations, this ability becomes ever more critical.

Earlier in the paper we discussed the advent of Modular LOMs/Daughter Cards and one of the earliest introductions of this form factor was Dell’s Modular Network Daughter Card for its PowerEdge M710HD Server – a Broadcom-based solution which also offers NIC partitioning.

While the feature sets and speeds of these products catch most of the attention, Broadcom’s leveraging of its broad product portfolio, diversified intellectual property, volumes, and integration ability deliver significant value in other very important areas such as cost structure, and power and space reduction. This has helped Broadcom maintain the top port shipment market share position for year-to-date 2011, in the 10 GbE + FCoE Server-class LAN-on-Motherboard & Controllers segment, according to Crehan Research Inc. market share reports.

Given Broadcom’s deep and long standing collaboration with OEMs, in addition to the rich and diversified depth of its product portfolio and technology expertise, it will be exciting to see its enablement of server and storage networking value with the ramp of Romley/Sandy Bridge platforms in 2012.