white paper

Building a Better Fibre Channel Switch

by
Thomas Hammond-Doel
Director, Technical Marketing

Vixel Corporation
October 2001

VIXEL

real world solutions for storage networking
www.vixel.com | 1·888·AT·VIXEL

Introduction

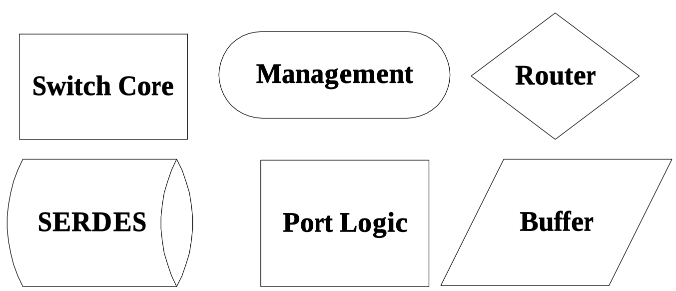
In today's technological world, there are many cases that reveal one product as empirically better than another product. One example is CDMA over TDMA technologies in cellular phones. When better performance is demonstrated through fundamental differences, and the performance can be technically explained, managers can make better-informed decisions in solving whatever problems they face in selecting a particular solution. Such end-user demands are essential to ensuring that "so-so" doesn't win over "best."

This paper will explain the general architectures used by Fibre Channel switches and the advantages and disadvantages of each. Subsequently, real world data will be presented showing how specific architectural approaches result in different performances. In particular, a comparison of the Vixel 7x00 series of Fibre Channel switches and the Brocade 2x00 series of Fibre Channel switches will be outlined. All comparisons will be made at 1 Gb/s due to no market availability of a Brocade 2 Gb/s switch.

The Fundamentals of a Switch

Each Fibre Channel switch has a basic architecture, regardless of who designs or manufactures the switch. As shown in Figure 1, each switch consists of a set of components. This includes at least 3 ports, a serializer-deserializer (SERDES) on each port, buffers to hold incoming and/or outgoing frames, a means of routing frames from one port to another, a switching element that connects the ports, and a management entity.

Figure 1 - Switch Components



So simple, yet the implementations by various vendors can radically affect switch performance. These components can be woven together to create a switch adequate for entry level applications, or can be woven together to take advantage of well known technological capabilities which increase speed, in some cases, dramatically, even in the most demanding applications.

We'll now take a look at some of the ways in which these components can be woven together and explain the advantages and disadvantages of each.

Management Entity

The management entity consists of a microprocessor, RAM, operating system, and vendor specific switch handling firmware. This firmware implements functions as diverse as FC-SW-2 protocol operations, to SNMP communications with host or server-based management applications.

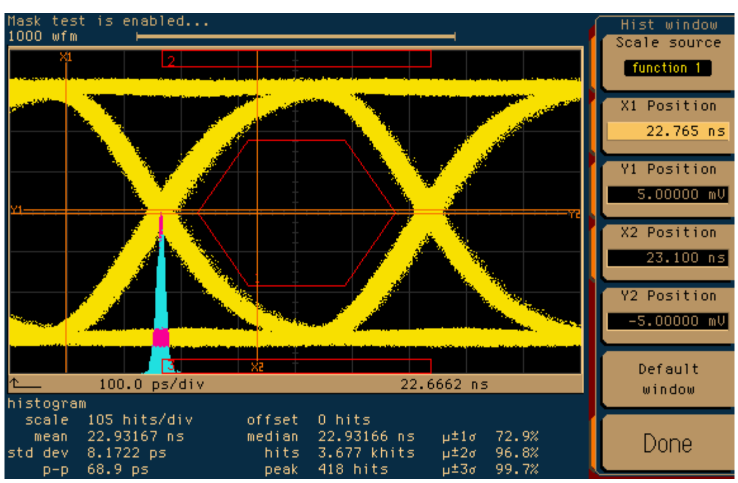
The microprocessor used generally will not affect the operation of a switch to any significant degree, unless it becomes overloaded. The efficiency of code will, however, gain some margin of performance, which speaks to the quality of the firmware available from each vendor. Vendors who use rapid prototyping processes or variations of the Software Engineering Institutes Capability Maturity Model (SEI-CMM), produce code which is more efficient, easier to troubleshoot, more extensible and generally better thought out. The results of optimized firmware will be visible in areas such as route convergence, as in the case of failed links, and also with link up/down handling.

SERDES

The purpose of a SERDES is to take the input stream of data, synchronize on that data stream, and present the port logic with a parallel version of the serial data. On the other end, a SERDES takes specifically clocked parallel data and serializes the data for transmission on the output stream.

The ability of a switch manufacturer to truly evaluate SERDES for a variety of factors, such as clock skew, random jitter, and deterministic jitter, helps increase the quality of signals that are handled by a switch. How a SERDES is designed-in with other elements of a switch - connectors, trace geometry, board capacitance, etc. - can greatly affect real world performance. Vixel is an expert in the physical layers of the Fibre Channel specifications, both as a key contributor to the ANSI standards and also as co-inventor of the GBIC. Vixel's physical layer experience is unlike that of any other switch vendor. Vixel's expertise has been leveraged to provide end-users the best possible jitter experience and explains why Vixel has the industry's "prettiest eyes" (a measure of signal integrity and quality). These "eyes" provide clean, error-free transmission of data.

Figure 2 - Typical Eye Pattern



Port Logic

Once data has been de-serialized, it is presented to the port logic circuit for processing. A port logic circuit handles a significant amount of processing in silicon. Once presented with data, the port logic circuit handles the various Fibre Channel link protocols, such as fabric port (F_Port), fabric Loop port (FL_Port), and interswitch ports (E_Ports). Vixel was not only the first vendor to support fabric loop (FL_Port) in the port logic circuit, but also the first vendor to support all three Fibre Channel link topologies in one ASIC.

Buffers

Buffers play a large role in how efficiently a switch operates, and in the case of Vixel and Brocade, two completely different tactics have been taken.

Looking at the total number of buffers on a switch, both Vixel and Brocade are similar. The key difference between the Vixel and Brocade approaches is how the buffers are allocated and tuned to common SAN deployments. Common deployments are typically a many-to-few topology: many servers to few storage subsystems. Vixel guarantees the number of buffers allocated to ports. Brocade buffers are allocated dynamically to the ports, depending on the level of traffic. In times of heavy traffic, the Brocade design will create throughput congestion.

Servers to storage pathways have evolved over time to exhibit very high performance characteristics in terms of bandwidth and I/O's per second. When a switch is placed in the middle of this pathway, it must not affect the high server to storage performance. As you will see by the data below, Vixel outperforms Brocade in these real world deployment scenarios helped by the choice made on buffer management.

Switch Core

The actual switching of frames is another key factor in the way a Fibre Channel fabric switch can be woven together. There are two fundamental methods used for switching and, again, Vixel and Brocade have chosen different paths.

Brocade uses what is called the "shared-memory" method (Figure 3). An older, effectively non-scaleable architecture, taking multi-ported RAM and sharing access to the RAM among all ports. This is a design that has been used extensively in low-end LAN switches, but abandoned in higher performance switches in favor of crossbar architectures. There are inherent latencies introduced by this technique, both in the storage and the retrieval of frames. Even more troubling with this architecture, as the port count and the traffic on those ports increase, the performance of the switch is compromised. The artifact of this can be seen in the performance of the switch in multi-device environments. Brocade's switch will experience low performance in environments where there are multiple servers attached to a few RAID subsystems. Figure 4 and 5 illustrate the performance of both switches in such an environment. This data was collected with 6 hosts and multiple targets using 128k blocks, which is a size often used in streaming and editing applications.

Figure 3 - Older Shared-Memory Architecture (Brocade choice)

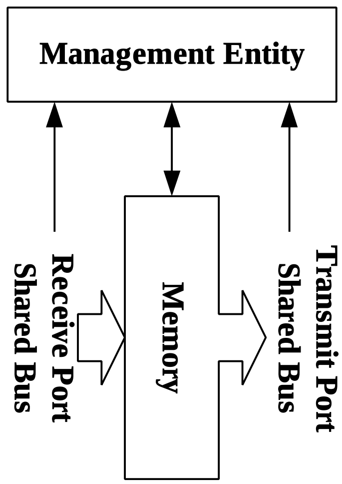


Figure 4 - IOPs for a multiple host environment, 16-port switch

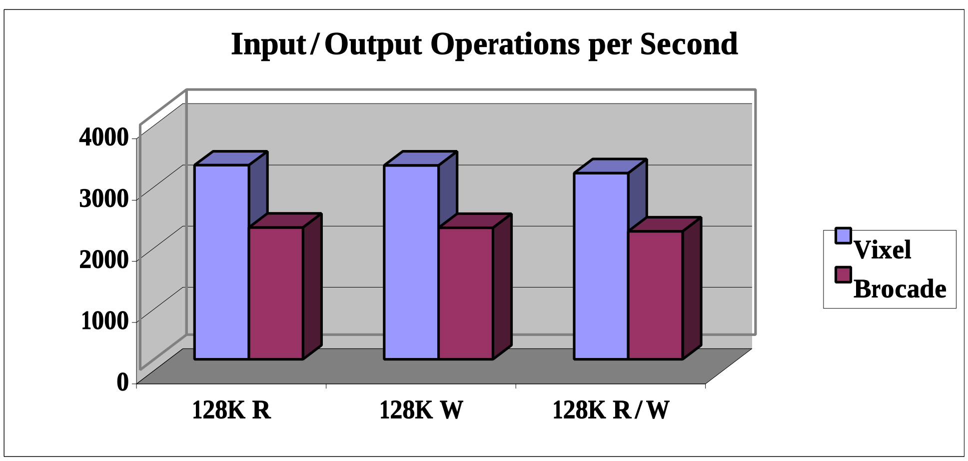
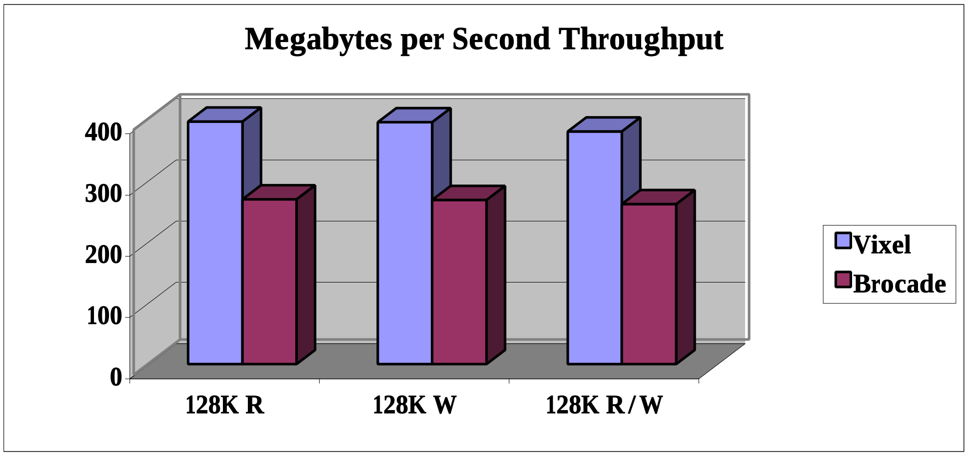
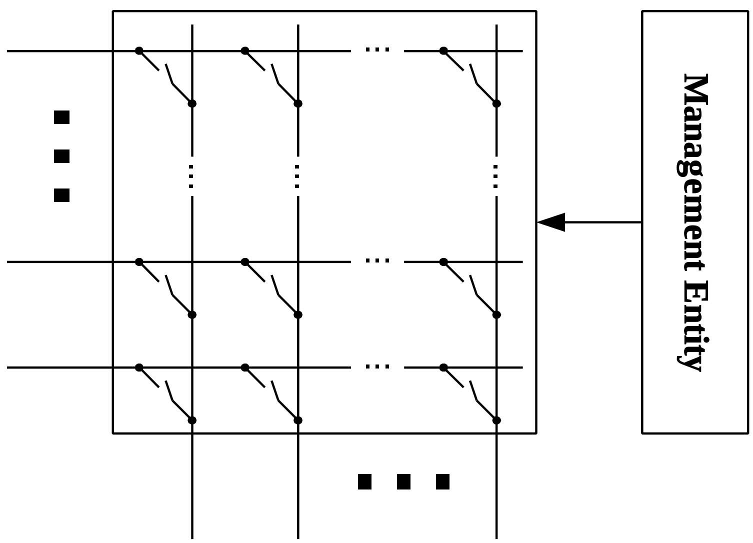


Figure 5- Total throughput for a multiple host environment (all targets added up), 16-port switch



Vixel, on the other hand, uses a crossbar switching core architecture (Figure 6) that is extremely efficient and extends performance linearly with port count. The crossbar switch core sends frames at wire speed between ports, and as additional ports are added, there is effectively no degradation of performance. Performance numbers will be shown later to illustrate this behavior. Vixel has been granted patents for its unique crossbar design.

Figure 6 - Crossbar Switch Core Architecture (Vixel choice)



Router
The router architectures are also markedly different between Vixel and Brocade. Brocade uses a distributed routing architecture that is part of the port ASIC. This architecture is rigid and not open to modification.

Vixel's routing function is located in programmable hardware. This hardware can be upgraded to introduce new features into legacy environments. This makes the Vixel switch more flexible, allowing users to customize routing protocols to specific applications, or introduce new features - such as private loop switching, zoning, and mixed speed routing - faster and more effectively than any other switch vendor. In fact, several aspects of Vixel's router have also been patented.

Real-World Environments

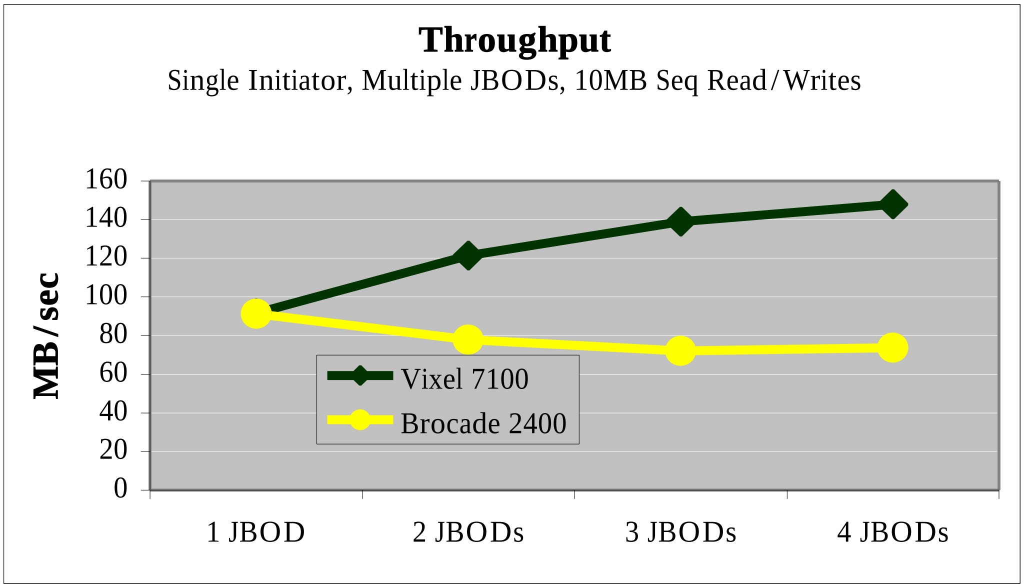
When analyzing the performance of products from different vendors, there are different types of metrics that are used in an effort to show one product as better than a competitor's. Results can be found showing Brocade switches performing better than Vixel switches, but these cases are typically those that are not representative of real world SAN topologies. In fact, it could be argued that the topologies where Brocade performs best are nearly point-to-point where a switch is not really needed.

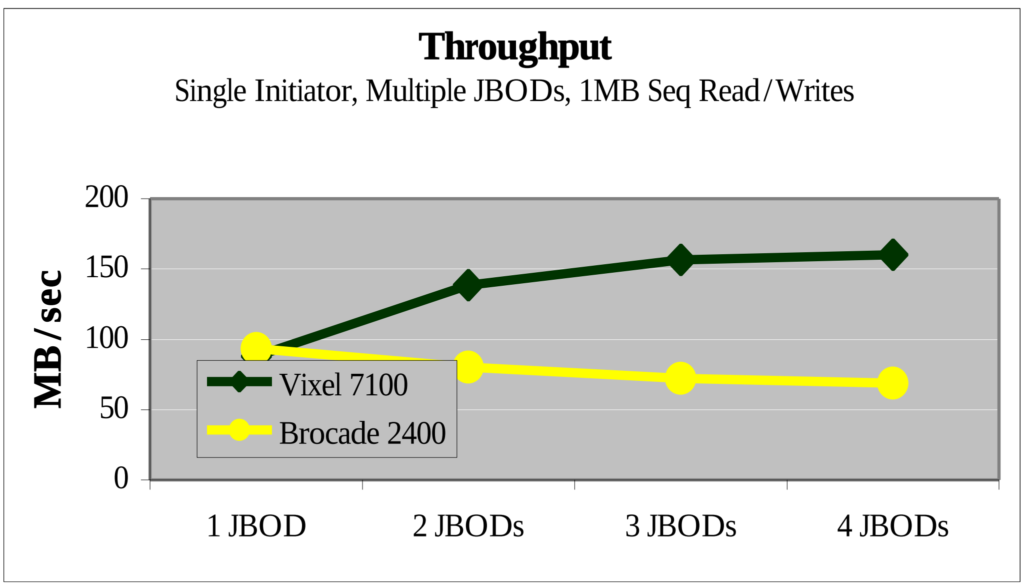
Once simple and point-to-point cases are thrown out, then real world fibre channel switching situations come into play. The situations analyzed below consist of only 50% read and 50% write operations. The 50%/50% mix of read and write operations is reasonably real world, although empirical data under a variety of solutions tends to show that 65% read and 35% write may be a bit more accurate. Vixel's switch architecture will out-perform Brocade's in both environments, but in a 65%/35% mix of read and write operations Vixel will demonstrate even further advantage over Brocade.

Under the 50%/50% scenario, different sizes of transfers are examined under the condition of a single initiator communicating with multiple JBODs/RAIDs. The single JBOD example is included only to show the sharp contrast between point-to-point and the multiple JBOD examples where real switching is needed.

Several aspects of comparative switch performance are examined under these real conditions, and each one of these demonstrates the superior performance of the Vixel switch in relation to the Brocade switch. In real world circumstances, even at 1 Gb/s, Vixel provides significantly better performance – and this performance is critical to many applications.

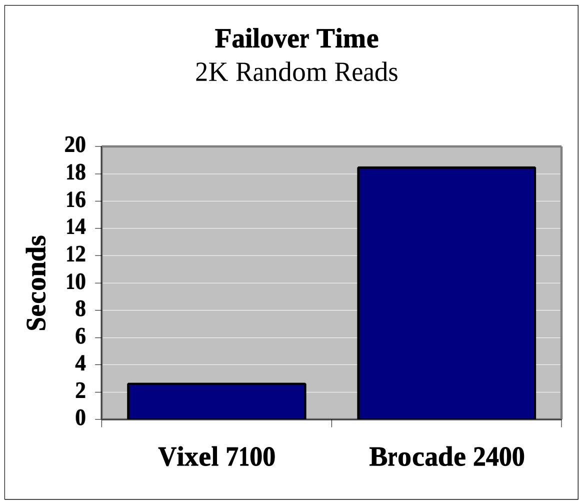
Throughput versus number of JBODs
The following graphs show the performance of Vixel versus Brocade under different transfer sizes and number of JBODs. JBODs are used here to simulate a very high performance RAID subsystem. Included is the trivial case of 1 JBOD, where the performances are essentially the same.


Throughput
Single Initiator, Multiple JBODs, 10MB Seq Read/Writes
MB/sec
160
140
120
100
80
60
40
20
0
Vixel 7100
Brocade 2400
1 JBOD
2 JBODs
3 JBODs
4 JBODs


Throughput
Single Initiator, Multiple JBODs, 1MB Seq Read/Writes
MB/sec
200
150
100
50
0
Vixel 7100
Brocade 2400
1 JBOD
2 JBODs
3 JBODs
4 JBODs

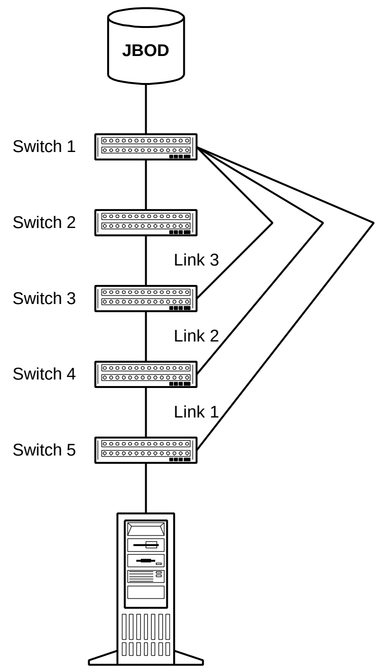
ISL Failover

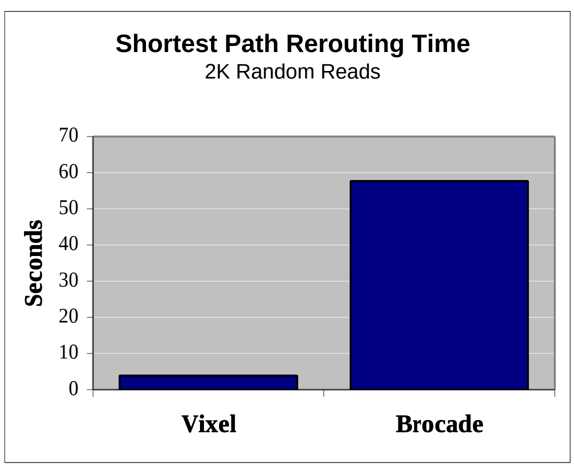
ISL Failover occurs when two switches are used with more than one Inter Switch Link and one of the links is removed. Under default conditions, the Vixel switch resolves the appropriate path to failover much more rapidly than the Brocade switch – over six times faster! The faster a switch recovers from ISL failover the faster a system can continue with the critical business of moving data. This recovery time also affects system uptime.



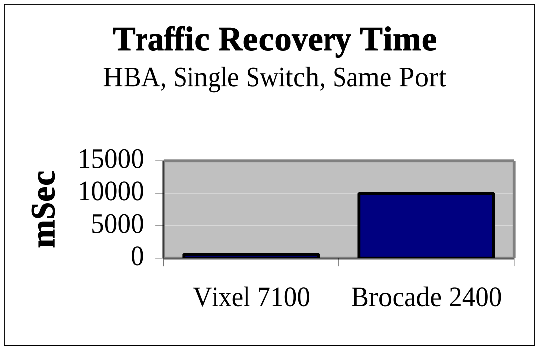
Route Convergence

When topologies change, the Vixel switches converge to the shortest path far faster than Brocade switches. The sooner a path converges to the true shortest path, the sooner a systems full capability can be utilized. This is a direct reflection of the software performance in the switch and highlights the fact that Brocade's Fabric OS contains a high level of complexity that slows down its operational performance.





Traffic Recovery
Traffic recovery time on the Vixel switch is much faster than on the Brocade switch. Traffic recovery is defined as the time from initialization after insertion (timed from a LIP on a Fibre Channel analyzer) to the first SCSI command. When performing a Traffic Recovery Test of both switches, Vixel's superior software and architecture performance will also be reflected.



Building a Better Switch

The above test results were completed using 1 Gb/s Fibre Channel switches. Brocade has no generally available 2 Gb/s product at this time. The general availability of the Vixel 9x00 series of 2 Gb/s Fibre Channel switches has provided many OEM's and end-users the opportunity to continue their development, and the level playing field has given the opportunity to test with superior – rather than incumbent – products. Many of the performance variations noted above are based upon fundamental differences in architecture. Brocade is claiming seamless backward compatibility with their unavailable 2 Gb/s switch. This would indicate that the fundamental design between the 1 Gb/s and 2 Gb/s Brocade switches are similar, and one would expect many of the differences described in this document to still be valid at 2 Gb/s. This means that new offerings from Brocade will again use older, less efficient technology. At 2 Gb/s, it is safe to assume the comparisons described in this paper will maintain validity.

Facts are facts, and the underlying technologies that make the Vixel switches better by design make the Vixel switches simply that – better by design. In the move to 2 Gb/s or with current 1 Gb/s devices, IT managers now do have a choice – and this time it does not have to be "so-so", it will be the "best."