WP280 (v1.0) September 23, 2008

Using FPGA Technology to Solve the Challenges of Implementing High-End Networking Equipment: Adding a 100 GbE MAC to Existing Telecom Equipment

By: Anthony Torza

This white paper examines the industry's urgent need for higher rate interfaces (particularly 100 GbE), the important risks and concerns that a system architect has when adding 100 GbE to a platform, and several implementation options that show how FPGAs are uniquely positioned to handle these challenges.

Industry Drive to 100 GbE

Ethernet has been an industry mainstay for decades. Starting with 10 Mb/s, every few years Ethernet has steadily increased in bandwidth to 100 Mb/s, 1 Gb/s, and 10 Gb/s.

The explosion of video traffic on the Internet has overwhelmed the existing 10 Gb/s lines within data centers, resulting in congestion in the uplinks and transport across the backbone network. At the same time, the expansion of multi-core processors has enabled individual servers to expand to 10 GbE (and to 40 GbE in the near future). These business and technology trends drive demand for higher rate (100 GbE) interfaces to aggregate and uplink traffic within the data center and to uplink to the transport network. Major enterprise and telecom providers have been quite vocal about their demand for a higher rate interface.

Stopgap methods for aggregating multiple 10 Gb/s lines into Link Aggregation Groups (LAG) have proven effective to about four links (40 Gb/s), but LAG implementations have hardware and software restrictions that prevent effective scaling beyond 4x10 GbE.

In late 2006, IEEE created the Higher Speed Study Group (HSSG) to define the next generation of Ethernet. In early 2007, the group emerged with a proposal for two speeds in the next generation of Ethernet: 40 GbE (focused on enterprise rack interconnect) and 100 GbE (focused on network uplink and transport applications). These requirements are captured in IEEE Std 802.3ba. While the standard is largely solidified with the definition of the Multi-lane Distribution (MLD), a few issues remain (e.g., the inclusion of FEC for backplane applications), and the standard has not yet been ratified. Xilinx is working with both the Ethernet Alliance and IEEE to finalize the 40 GbE and 100 GbE standards.

This white paper focuses on a pre-standard 100 GbE example and leverages the unique ability of FPGAs to upgrade and address changes in the emerging standard. This is not a hypothetical example—this design is already being used in customer systems. Designers can be confident that their implementations will remain compliant to the standard once it is ratified.

Designing a 100 GbE MAC

Customers considering a design with new technology can benefit from exploring multiple architectures, enabling trade-off analyses between risk, time to market, and cost. Two 100 GbE implementation options using Xilinx FPGAs are described in this white paper. Option 1 features a common legacy implementation. Option 2 features an optimized architecture for the near future.

Key Risk Areas When Designing with New Technology

Any new design contains an element of risk, which can be quantified as:

- Time To Market (TTM) Risk
  - Introducing a product with the right cost at the right time is key to product success in the market.
  - Often, multiple generations are necessary to track the market as it transitions from *early adopter* (being first establishes market leadership) to mainstream technology (volumes are higher and cost pressures start to restrict the choice of technologies). It is important for a vendor to offer solutions that span as many of these generations as possible.

- Component Risk
  - Schedule Risk
    - How much risk is in the schedule for the components required to complete the design, i.e., risk of delivery/availability?
    - Are the required components on a new, largely untested process node that might cause delays?
    - Has the vendor released characterization data for the I/O (especially SERDES)?
  - Performance Risk
    - Does the silicon core logic have the performance and capacity to support the required application?
    - Does the silicon I/O (likely SERDES based) support the required performance?
  - Power Risk
    - Will the design fit in the required total power envelope?
- Third-party IP Risk (for designs that require third-party IP)
  - Schedule Risk
    - Is the IP still under development?
    - Has the IP been deployed by other customers? (This implies that the IP has been verified by more than one process).
- Functionality Risk
  - Has the standard evolved enough that the designer is confident the functionality will not change significantly?
- Ecosystem Risk
  - External Components
    - Are all the external components available?
    - Have all required components been simulated or tested in this environment?
    - Can the vendor produce data (documentation) to support the simulation, testing, and hardware characterization?
  - Test Equipment
    - Is there test equipment that enables debugging of the hardware?

Different designs have different sensitivities to these risks, prompting trade-off analyses. The following options highlight some of these trade-offs.

Option 1: Legacy Solution - XSBI to External 10 Gb/s PHY

Several customers have implemented the *XSBI to external 10 Gb/s PHY* solution. This option is focused on the lowest risk technology available in the past year. This implementation is based on two Virtex-5 FXT FPGAs and ten external XSBI-based 10 Gb/s PHY devices. As shown in Figure 1, the Media Access Controller (MAC) is split in half across two FPGAs.

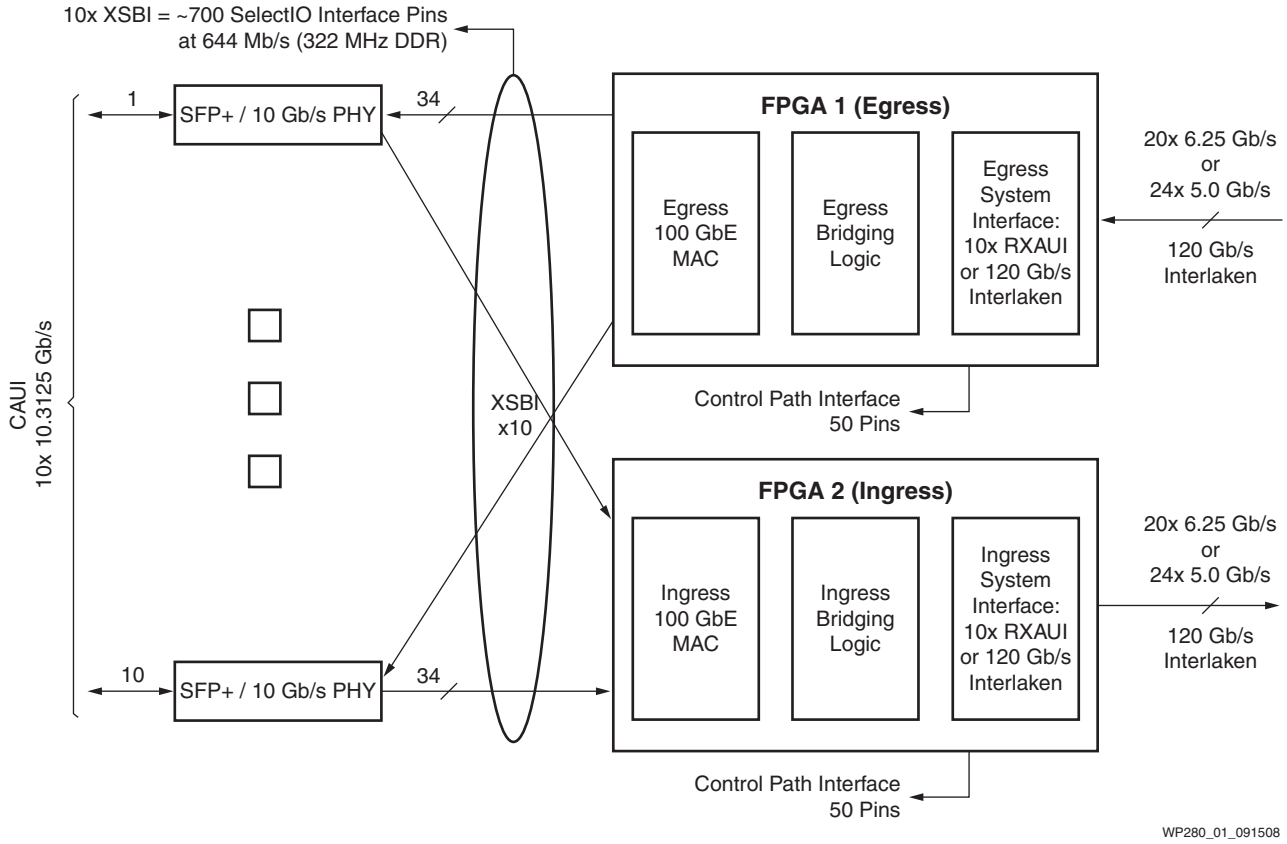

Figure 1: **Industry's First 100 GbE MAC: Two Virtex-5 FXT FPGAs with 10 External SERDES**

The ingress FPGA contains the RX 100 GbE MAC (including MLD), some bridging/buffering logic, and the TX system interface, which is specific to the particular customer's system architecture. A customer ASIC or ASSP Ethernet switch can be connected via RXAUI (a double-speed version of XAUI), or an ASIC or ASSP NPU can be connected via an Interlaken interface. Customer proprietary interfaces are also options. The egress FPGA contains the RX system interface, some bridging/buffering logic, and the TX 100 GbE MAC (including MLD).

The CAUI interface to the optics module is created via 10 external 10 Gb/s PHYs. These devices connect to the FPGAs via a modified XSBI interface (16 lanes at 322 MHz DDR). The XSBI standard specifies 16 lanes at 644 MHz SDR, so this interface is not precisely DDR—more accurately, it is XSBI with a half-rate clock. The use of a 322 MHz clock enables designers to leverage the DDR logic built into Xilinx SelectIO™ technology to give the design more margin and lower power. Many major PHY vendors support this half-rate, 322 MHz clock.

The risk profile of Option 1 is as follows:

- TTM Risk (Lowest Possible)
  - This solution has been tested in a number of customer implementations.
- Component Risk (Very Low)
  - All parts are available today—in full production.
- Schedule Risk (Very Low): Virtex-5 FXT FPGAs and XSBI SERDES are widely available
  - Performance Risk (Very Low): Virtex-5 FPGA logic has sufficient performance for a number of datapath widths. The Virtex-5 FPGA GTX serial transceivers are capable of supporting rates up to 6.5 Gb/s.
  - Power Risk (Moderate to Low): While the XSBI interface does not provide the lowest power, under standard telecom worst-case conditions (NEBS), the devices can be cooled to operate within acceptable limits.
- Third-party IP Risk (Very Low)
  - The 100 GbE MAC, Interlaken, and RXAUI IP cores have been debugged in customer environments and are passing traffic in prototype hardware.
- Ecosystem Risk (Lowest Possible)
  - External Components are fully available, e.g., 10 Gb/s PHY is available from multiple vendors.
  - Test Equipment is available, e.g., advanced traffic generators are available from a leading vendor.

Option 2: MLD Interface to SERDES MUX

Xilinx currently recommends Option 2 for customers designing 100 GbE equipment. It uses the lowest risk technology available in the near future.

This option utilizes the Virtex-5 TXT platform and external Quad SERDES MUXes to enable the design to shrink from 2 FPGAs and 10 external PHYs to 1 FPGA and 3 external Quad PHYs. This combination enables significant cost and power reductions. As shown in Figure 2, this implementation fits the logic functions in a single FPGA, because the Virtex-5 TXT platform contains up to 48 GTX serial transceivers.

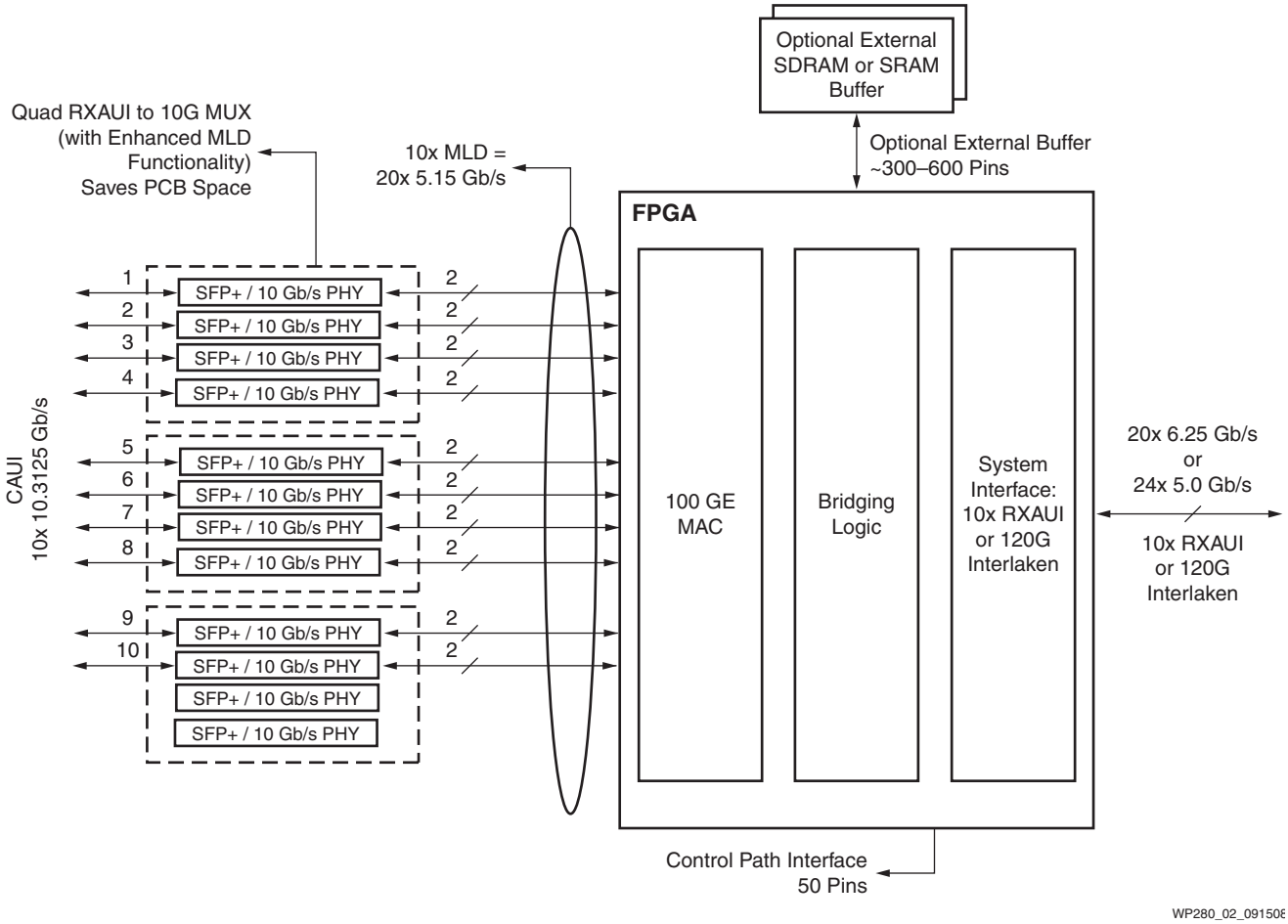

Figure 2: **Option 2: Cost-reduced Version: 1 Virtex-5 TXT FPGA with 3 External Quad SERDES MUXes**

This single FPGA contains the 100 GbE MAC (including MLD), some bridging/buffering logic, and the system interface, which is specific to a particular customer's system architecture. A customer ASIC or ASSP Ethernet switch can be connected via RXAUI (a double speed version of XAUI), or an ASIC or ASSP NPU can be connected via an Interlaken interface. The flexibility of the FPGA allows customers to add their own proprietary interfaces.

The CAUI interface to the optics module is created via a serial interface to three external Quad 10 Gb/s PHYs. These devices connect to the FPGAs via an MLD-like interface (20 lanes at 5.15 Gb/s). Because the mapping from VL to SERDES is 1:1 and the MLD layer automatically deskews the VLs, Option 2 greatly simplifies the PCB design and reduces power consumption.

Furthermore, because the 100 GbE standard defines 64B/66B encoding on a per-virtual-lane level, only designs using Xilinx FPGAs can take advantage of the 64B/66B gearbox to save significant logic (an estimated 10k LUTs total). This SERDES-based interface saves logic, PCB space, routing complexity, and power.

Xilinx has commitments from major PHY vendors to implement the MLD interface in their next generation of multi-core PHYs, which should be available in early 2009. Contact Xilinx representatives at Virtex_marketing@xilinx.com for third-party vendor and contact information.

The risk profile of Option 2 is as follows:

TTM Risk (Low)

- This solution reuses the base Virtex-5 FPGA technology to produce a device with sufficient SERDES that optimize the design into a single FPGA and three external PHYs.
- Component Risk (Low)
  - Schedule Risk (Low): Virtex-5 TXT devices are available, and the third-party Quad MUX parts will be available in very early 2009.
  - Performance Risk (Very Low): Virtex-5 FPGA logic has been proven to have sufficient performance for a number of datapath widths.
  - Power Risk (Low): The conversion of a SERDES-based MLD interface reduces the required logic. Combined with the conversion from a parallel (XSBI) to serial (MLD) interface, Option 2 reduces power consumption by over 20%.
- Third-party IP Risk (Very Low)
  - The 100 GbE MAC, Interlaken, and RXAUI IP cores have been debugged in a customer environment and are passing traffic in prototype hardware.
  - The changes to support the MLD interface are minor and take advantage of the 64B/66B gearbox (available exclusively in Xilinx products) to reduce the logic count.
- Ecosystem Risk (Lowest Possible)
  - Third-party Quad MUX parts will be available early in 2009.
  - Test Equipment is available.

Future Solutions

Xilinx is committed to the high-end telecom market. As we continuously release new products, we will continue to further refine the recommended implementations for 100 GbE products. Contact Xilinx representatives at Virtex_marketing@xilinx.com to keep current on the latest technology.

Summary

FPGA-based, programmable solutions offer a low-cost, low-risk path to developing a wide variety of applications in the wired telecom market. The performance of the FPGA fabric and I/O has evolved to support even the most challenging 100 Gb/s applications. As the FPGA market leader, Xilinx anticipates customer and market requirements, and engineers silicon, software, and FPGA IP to offer the right low-risk solution at the right time - enabling the economies of scale, flexibility, and an increase in quality of service needed for the industry's migration to 100 GbE. For more information about Xilinx Solutions for the Wired Telecom Market, visit: http://www.xilinx.com/esp/wired.htm.

Appendix: Virtex-5 TXT FPGA

The Virtex-5 TXT platform is built upon a proven production-ready process and block functions. Extensive feedback from developers was used to make the Virtex-5 TXT platform suitable to support new ultra-high bandwidth systems. The TXT devices are the latest members of the market-leading 65 nm Virtex-5 family. By leveraging the

column-based ASMBL™ architecture in the Virtex-5 family, Xilinx has created two devices with two columns of 6.5 Gb/s serial transceivers. This TXT platform has two members: the XC5VTX240T with 48 transceivers and the XC5VTX150T with 40 transceivers. For the latest TXT platform specifications, see DS100, *Virtex-5 Family Overview.*

Revision History

The following table shows the revision history for this document:

Date	Version	Description of Revisions
09/23/08	1.0	Initial Xilinx release.

Notice of Disclaimer