

Virtex-II Pro FPGAs Enable 10 Gb iSCSI/TCP Offload

The Ipsil FlowStack enables 10 Gbps TOE solutions using a very small footprint.

by Sriram R. Chelluri
Senior Manager, Storage and Servers
Xilinx, Inc.
sriram.chelluri@xilinx.com

Enterprise data-center connectivity solutions have evolved over the years from direct attach storage (DAS) solutions (such as SCSI) to Ethernet-based file servers like network file system (NFS) and switch-based architectures (like Fibre Channel [FC]) that can encapsulate the SCSI/IP or FICON protocols.

Fibre Channel is a well-known high-performance protocol used in critical latency-sensitive applications like databases, video streaming, and tape backup. The most common data rate for Fibre Channel is 2 Gbps, with 4 Gbps slowly making headway into data centers, as the per-port cost factors for both are nearly the same. However, FC has not become a prevalent part of IT infrastructures because of its lack of interoperability between vendors and its

high connectivity costs, requiring special host bus adaptors (HBAs).

To address the interoperability issues and costs associated with FC, IETF adopted a protocol to transport SCSI data over TCP, more commonly known as iSCSI. In this article, I'll review current iSCSI implementations and Ipsil Corporation's 10 Gbps TCP Offload Engine (TOE) solution to address this rapidly growing market.

Current iSCSI Solutions

Current iSCSI solutions rely on a complete software stack or on special network interface cards (NICs) based on ASICs for handling TCP/IP processing. Like FC, ASIC-based TOEs are expensive and have many interoperability issues. Vendors were building TOE cards long before the standards were approved. This resulted in a very slow adoption rate of iSCSI solutions. Also, ASIC-based solutions are expensive, with long lead times for development, testing, and manufacturing, and with general

network performance outpacing storage device performance.

An all-software solution is acceptable for low-bandwidth applications, but high-performance applications would consume all of the CPU resources, creating a system bottleneck for critical applications. For example, every 1 Gbps in Ethernet performance requires 1 GHz of CPU resources at 100%. Also, when Ethernet was at 1 Gbps, FC had an established market place at 2 Gbps. iSCSI market penetration could not justify the price/performance.

After a few years of lackluster customer interest for iSCSI, it is slowly emerging as a viable complement to Fibre Channel. The network infrastructure is migrating to 10 Gbps, while FC is still at 4 Gbps.

To address the performance challenges of 10 Gbps iSCSI, Xilinx teamed with Ipsil Corporation to build a programmable solution for the iSCSI market based on Xilinx® Virtex™-II Pro FPGAs with RocketIO™ multi-gigabit transceivers.

FPGA-Based iSCSI/TOE Engine

With standards-compliant FC, PCIe core, and a TOE core from Ipsil, we created a technology demonstration platform for the 10 Gbps iSCSI TOE market, working with industry leaders in high-performance computing solutions. Programmable solutions enable system architects to add functionality as needed. Depending on the Xilinx product family, you can integrate multiple IP cores into a single FPGA, reducing board costs and time-to-market requirements. For example, you can:

- Combine the FC IP core with the iSCSI/TOE engine to make FC storage available to Ethernet clients
- Combine the PCIe IP core with the iSCSI/TOE engine for host-based iSCSI storage access
- Combine the TOE engine with established IPsec or the emerging 802.1AE specification for security

10 Gbps iSCSI TCP Offload Engine Core

Ipsil FlowStack is an implementation of a full-featured TCP/IP networking subsystem in the form of a “drop-in” silicon-IP core. It includes a standards-compliant TCP/IP stack implemented as a compact and high-performance synchronous state machine.

The FlowStack core also includes support for the FastPath of upper-layer application protocols such as RDDP, MPA, and RDMA. In addition, for embedded monitoring and remote management, FlowStack incorporates an embedded HTTP, FTP, and SNMP server. Ipsil FlowStack is built from a collection of synthesizable Verilog modules, which are composited to customer specifications and requirements before delivery and can therefore be customized to specific requirements.

In FPGA delivery, you can instantiate FlowStack along with customer intellectual property and configure it to operate without the need for host CPU support. The FlowStack packet processing engine operates in strictly deterministic time. FlowStack typically takes five core clock cycles to process each packet, which means that the core can be clocked at modest clock rates to achieve 1 or 10 Gbps throughput. This zero-jitter and low-overhead processing can be used to your

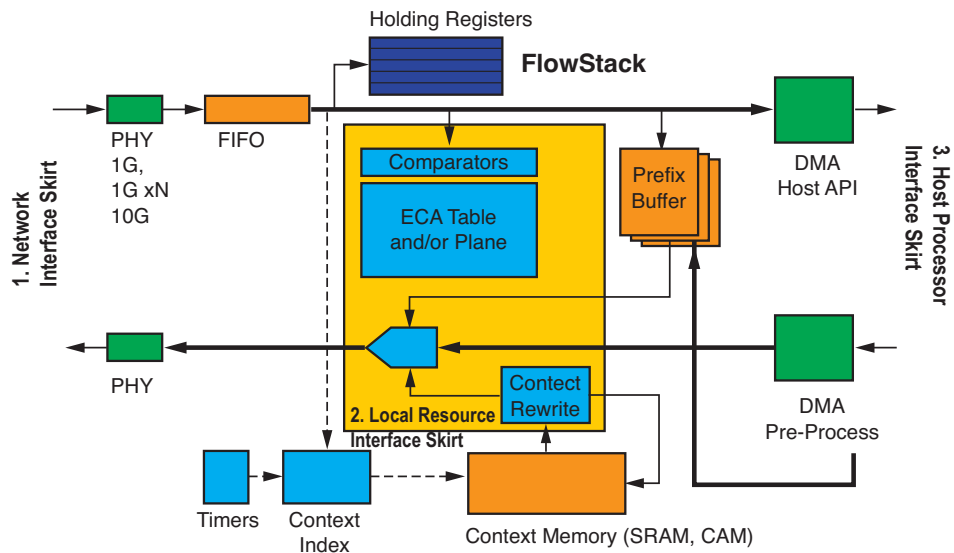


Figure 1 – Ipsil FlowStack interface

advantage in low-latency processing and on high-speed networking applications.

Interfacing to the Ipsil FlowStack Core

The customer interface to the Ipsil FlowStack core is through well-defined interconnect interfaces (Figure 1):

1. Network interface: Ethernet and other PHY interface
2. Local resource interface: off-core memory and local I/O interface
3. Host processor interface: SiliconSockets interface

The three interfaces are implemented independently in the form of a resource-frugal synchronous registered interface that you can integrate into customer FPGA designs in a straightforward manner, with the least overhead.

Network Interfaces

A variety of network interconnect options are available in a Xilinx FPGA delivery platform. These include:

- Industry-standard MII interface for Ethernet PHY interconnects, (10/100-BaseT Ethernet)
- Xilinx GMAC and 10GE MAC supporting 1 and 10 Gbps
- SPI-4.2 for OC-192 and OC-768 framers
- XAUI and 1-10Gbase CX-4 interfaces

Local Resource Interface

You can instantiate local memory resources in different ways. In an FPGA delivery platform, some memory can be realized with on-chip block RAM, while the FlowStack core also supports external SDR and DDR SDRAMs for additional buffer memory. In gigabit applications, FlowStack’s local buffer memory requirements are very low, as it is capable of performing most operations directly to and from host memory, achieving true zero-copy operation at a high speed.

TCP state memory, required for some TCP and RDMA applications, is instantiated as combinations of SRAM, DRAM, or CAM blocks. These resources are composited by Ipsil to meet requirements according to specific customer and application needs. These resources can be on-chip or off-chip, depending on the demands, complexity, and requirements.

In a Virtex II-Pro VP40 FPGA platform, as many as 2,000 connections can be supported with internal block RAM in addition to the PCIe and 10 Gb MAC core. The FlowStack’s cut-through architecture requires no data buffers and the connection count can scale, depending on the available internal and external memory.

Interface to the Host Processor: SiliconSockets

The user host processor interfaces to the Ipsil FlowStack core through the

SiliconSockets interface. The interface resources of the Ipsil FlowStack core are mapped into the processor's accessible address space, either as a set of host-accessible registers, through a coprocessor interface, or as an on-chip system peripheral (through an on-chip SoC bus) according to customer requirements. The SiliconSockets interface comprises two mechanisms: a register-based interface and a scatter/gather-based DMA.

Implementation Example – 10 Gbps TOE

The Ipsil FlowStack core was designed to be targeted to a variety of FPGA families depending on the performance and functionality requirements. It has a synchronous control path that comprises fully synthesizable Verilog modules. The fully registered data path has the flexibility of being instantiated using block RAM across the entire range of Xilinx product families.

Ipsil FlowStack SiliconSockets can be interfaced to PCI, PCI-X, or PCI-Express bus interfaces.

A sample design is illustrated with the FlowStack core targeted to a Virtex-II Pro device with a selection of Xilinx LogiCORE™ IP and AllianceCORE™ components that help to complete the application specifications.

Parameters of the Design Example

You can build a configuration with the following parameters:

- FlowStack instantiated in a Virtex-II Pro device
- 100 simultaneous TCP connections
- Connect to external gigabit PHY/GBIC using Xilinx GMAC and RocketIO transceivers

You can assume that the application will operate at the full 10 Gbps network speed and will be used in a typical Internet application. Consistent with current Internet engineering practice, a round-trip time of 200 ms is the assumed median case.

Figure 2 is a block diagram of an instance of an Ipsil IPμ FlowStack core in a Virtex-II Pro device. The FlowStack core has been composited with a selection of ancillary cores to allow it to provide an Ipsil

SiliconSockets interface at the “uncooked-BSD-socket” level to TCP consumer applications resident in custom cores, either on the same Virtex-II Pro device or off-chip.

The Ipsil FlowStack core can operate at full line speed in any process technology in which the corresponding physical input and output mechanisms are available. The core uses nominally five internal machine cycles in its packet rewrite engine to process any packet.

The Ipsil FlowStack core requires approximately 2,500 LUTs for the packet rewrite engine, including 1,600 distributed bits of flip-flops. This is a very compact core that performs all TCP/IP packet pro-

In Figure 2, the FlowStack core has been supplied with a 10GMAC interface to an external 10 Gigabit Ethernet CX4 interface. In this example, based on resource requirements for the Xilinx LogiCORE 10GMAC, this module would need approximately 4,000 slices for a 10 Gbps PHY interconnection.

Conclusion

The Ipsil FlowStack is a very compact, fully synthesizable design. FlowStack is a processor-free design (that is, not containing within itself an embedded CPU) based on a hardware state machine; thus the clock-speed requirements are very modest. These

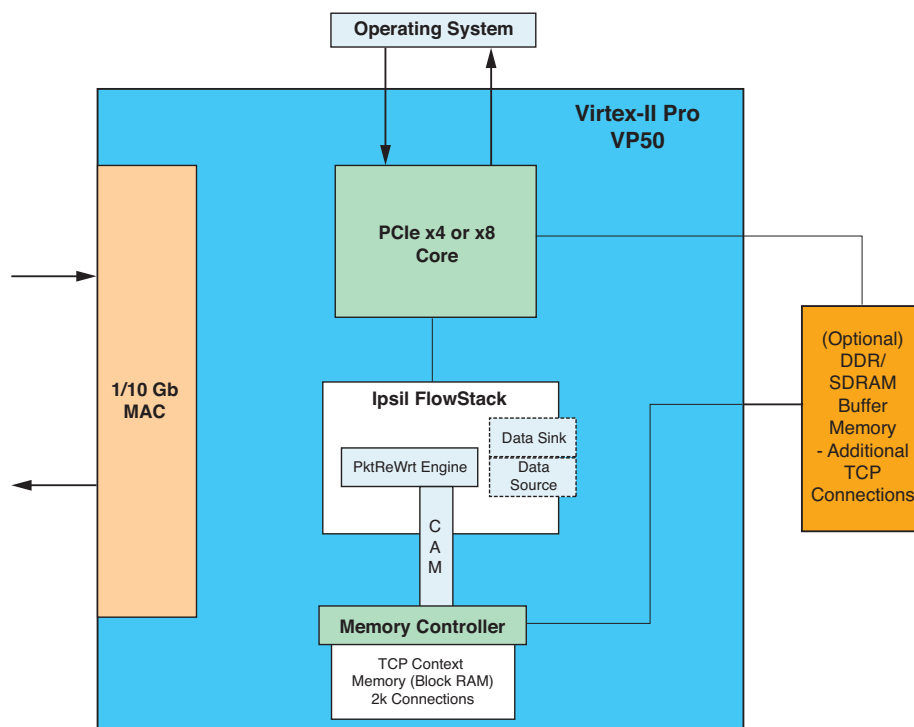


Figure 2 – FlowStack implementation on a Virtex-II Pro device

cessing procedures, but requires the support of several ancillary modules to manage raw memory resources and the overhead of the interface with custom cores.

TCP data buffer memory has been implemented as off-chip SDRAM. This has been sized assuming a 10 Gbps PHY connection and a 200 ms maximum TCP RTT (round-trip time). In this example, this has been realized in the form of a 32 MB DDR/SDRAM memory bank. This memory is interfaced using an SDRAM controller, which is budgeted at about 364 CLB slices.

characteristics lend themselves well to be realized in an FPGA delivery platform.

Higher level protocols such as iSCSI and RDMA are currently in the early stages of market acceptance. To ensure compatibility and interoperability with all versions of operating systems and hardware in a constantly evolving market, the high-density Virtex FPGA families with built-in multi-gigabit transceivers are an ideal choice for reducing board density while scaling to your performance requirements. ●●●