Common Switch Interface for Fabric Independence and Scalable Switching

IEEE TUTORIAL
NOVEMBER 9, 1998

Colin Mick, The Mick Group
Russell Johnson, Power X
Praveen Kumar, XaQti
Presentation Overview

• Introduction
• The Market Issues
• The Technology Issues
• Q & A
Introduction

• What is CSIX™?
• Project status
• Product status
What is CSIX?

- CSIX: The Common Switch Interface, is a detailed interface specification between port/packet processor logic and interconnect fabric logic.
CSIX is the Common Switch Interface

- CSIX is a scalable parallel interface with separate data and control paths
- It is a generic multi-vendor specification to promote the deployment and development of highly scalable network switches
- Permits hardware and software interoperability
- Permits mix and match of interchangeable silicon components, hardware and software
- Concept can be used to expand existing switch architectures: Pt-to-Pt; Shared Memory; Shared Bus
CSIX Interoperability Goals

- **Logical or Message Level**
  - ensure that data or control message protocol exchanged over the interface are properly understood by each end and properly processed by the appropriate function.

- **Interconnection Level**
  - define all the signals with specific functions, meanings, and bit widths, input or output, signal handshake protocols, etc.

- **Physical Level**
  - specify the electrical characteristics such as voltage levels, capacitance, drive strengths, timings etc.
Benefits

• Enables the design of products that can scale from <5 Gbps to 10s of Tbts and from Layer 2 through Layer 3/4+.

• Much lower cost of development for high-speed, high aggregation systems

• Much lower cost of maintaining and operating these new networking systems

• Creates an open and dynamic marketplace for vendors/OEMs with interoperable switch logic and/or fabric expertise

• Dramatic increase in price performance of new network equipment
Current Status

• Project status:
  – Initial draft of CSIX specification will be released 4Q98
  – Target date for completion of the specification is 3Q99.

• Product status:
  – First products, based on the first draft of the CSIX interface, will ship in 2Q99.
Market Issues

- The changing switch market
- Managing growth
- Evolving switch technology
- Fabric: the point of concentration, where the interconnection happens
- Evolving switch fabric technology
- What does it mean?
The changing switch market

- Switch throughput needs to increase due to:
  - Increase in port bandwidth (10-> 100->1000->...)
  - Increase in number of ports
- Demands for QoS and better management capabilities are increasing.
- The squeeze
  - Increased functionality
  - Constrained resources
    - Stability, availability, capability
- The need for individuality
  - Responsiveness, Differentiation, Value add
Managing growth

Common Switch Interface

System management

Interconnect fabric

Media Processor

Media Processor

Media Processor

Phy

Phy

Phy

Add growth

Add features

Add users
Evolving Switch Technology

- Multiservice and policy based networks
- Soft control of extensive and complex packet processing in hardware before forwarding
- Higher port and bandwidth aggregation increase the density of interconnect fabric hardware
- Efficient scheduling and arbitration across a system of many interconnected ports with guaranteed fairness and QOS
- Increasing Management and Reliability eliminates single point of failure through redundancy.
Fabric: the point of Concentration, where the Interconnection Happens
Evolving Switch Fabric Technology

- Ability to handle increasing port speeds:
  - 10, 100, 1000, 2500...

- High data density
  - Utilization of high speed serial links

- Simplicity
  - Self Routing Switches
  - Buffer management support

- Efficiency
  - Intelligent Flow control and Congestion handling
  - Handling of priorities and QoS issues
  - Efficient handling of Multicast

- Scalability
What does it mean?

- Features and functions are added at the media processing layer
- Scalability is added at the fabric layer
- Port and Fabric processing must exchange messages on routing instructions, status, priorities, policies, and service requirements
- We NEED a common electrical and messaging interface to plug the two together
Technical Issues

• CSIX Interfaces
• CSIX Parameters
• Typical Implementation
• CSIX based Systems
Two Classes of CSIX Interfaces

- CSIX CLASS A interface supports intelligent switching fabrics with integrated routing.
  - requires forwarding and flow control messages to be delivered in line with the data
- CSIX CLASS B interface supports non-intelligent switching fabrics
  - requires connection scheduling as well as other control messages to be routed to a central or distributed set of intelligent controllers independent of, and in parallel with, the data.
- This yields Interoperability within each class
Class A Interface
Class B Interface

```
Traffic Manager
CtrlDin [n..0]
CtrlDinCtrl [2..0]
CtrlDout [n..0]
CtrlDoutCtrl [0]
CtrlDoutCtrl [2..1]
RxCtrl [1..0]
TxCtrl [1..0]
TxCtrl [3..2]

Switch Fabric Controller

Switch Fabric

TxData[n..0]
TxClk
TxPrty

RxData[n..0]
RxClk
RxPrty
```
## Parameters (1)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Specification</th>
</tr>
</thead>
<tbody>
<tr>
<td>Clock rate</td>
<td>Up to 100MHz, synchronous or asynchronous to system clock</td>
</tr>
<tr>
<td>Data path width</td>
<td>1, 2 or 4 bytes</td>
</tr>
<tr>
<td>Packet types</td>
<td>Unicast, multicast-with-mask, multicast-with-ID, broadcast, configuration, user-to-user</td>
</tr>
<tr>
<td>Destination addresses</td>
<td>Up to 4096 traffic managers. Sub addressing per traffic manager is supported through user-to-user packets</td>
</tr>
<tr>
<td>Priorities</td>
<td>Up to 16 user-definable and configurable priority levels</td>
</tr>
<tr>
<td>Urgency</td>
<td>16 levels of urgency within each priority</td>
</tr>
<tr>
<td>Header</td>
<td>4 bytes for Class A&lt;br&gt;None on data for Class B</td>
</tr>
</tbody>
</table>
## Parameters (2)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Payload</td>
<td>1-256 bytes</td>
</tr>
<tr>
<td>Flow Control</td>
<td>In-band for Class A, or through control bus on Class B, per queue, proportional or &quot;pause/resume&quot;</td>
</tr>
<tr>
<td>Error management</td>
<td>Parity</td>
</tr>
<tr>
<td>Signals</td>
<td>Data(8/16/32), Control (3), Parity (1), Clock (1) in each direction for both Class A and Class B; plus for Class B Control Data(4/8)</td>
</tr>
<tr>
<td>Electrical signaling</td>
<td>LVTTL</td>
</tr>
</tbody>
</table>
## Typical Class A implementation

<table>
<thead>
<tr>
<th>Common Switch Interface</th>
<th>Data path (bits)</th>
<th>Packet size (bytes)</th>
<th>Avail. Payload BW</th>
<th>Util. of interface BW</th>
<th>Total pin count</th>
</tr>
</thead>
<tbody>
<tr>
<td>OC-12</td>
<td>8</td>
<td>57</td>
<td>0.74</td>
<td>84%</td>
<td>26</td>
</tr>
<tr>
<td>OC-48</td>
<td>32</td>
<td>60</td>
<td>2.83</td>
<td>88%</td>
<td>74</td>
</tr>
<tr>
<td>Gigabit Ethernet</td>
<td>16</td>
<td>5-80</td>
<td>&gt;1.3</td>
<td>&lt;75%</td>
<td>42</td>
</tr>
</tbody>
</table>
A Scalable Fabric based Switch

Common Switch Interface

UTOPIA
SONET
OC-48 Uplink

GMII
PHY
1Gbps Uplink

PHY

CSIX
Port Processor
MEMORY

Port Processor
MEMORY

Port Processor
MEMORY

Port Processor
MEMORY

Controller

Port Processor
MEMORY

Port Processor
MEMORY

Port Processor
MEMORY

Port Processor
MEMORY

8-100Mbps & 1Gbps Switch

1Gbps Ethernet

2 port 1Gbps Switch

8-100Mbps FDDI Switch

8 100Mbps MAC ports

1Gbps Uplink

8 100Mbps FDDI ports

8-100Mbps FDDI Switch

SONET

PHY

GMII

PHY

Serial port

CPU

Controller

MII

FC-0

PHY

PHY

PHY

CSIX

CSIX

CSIX

CSIX

CSIX

CSIX

CSIX

CSIX

CSIX

CSIX

CSIX

CSIX

CSIX

CSIX

CSIX

CSIX

CSIX

CSIX

CSIX

CSIX
A Switch Line Card

Common Switch Interface

Management subsystem

TeraPOWER™

Congestion buffer
Parameter SRAM

MAC

StarRouter™

Serial Data

Control

VOQ SRAM

CSI

Txcvrs

Txcvrs

Txcvrs

Txcvrs

Txcvrs

Txcvrs

Txcvrs

Txcvrs
Common Switch Interface Fabric Card

Fabric card

- Management subsystem
- Data
- Configuration

StarCeptor™

StarMaster™
Open Issues

- OC192--pin speed
- Electrical signaling
- Refinement
For more information:

www.csix.org