



Multi-Gigabit Transceivers

Getting Started with Xilinx's Rocket I/Os

Craig Ulmer

cdulmer@sandia.gov



July 26, 2007

Craig Ulmer

SNL/CA

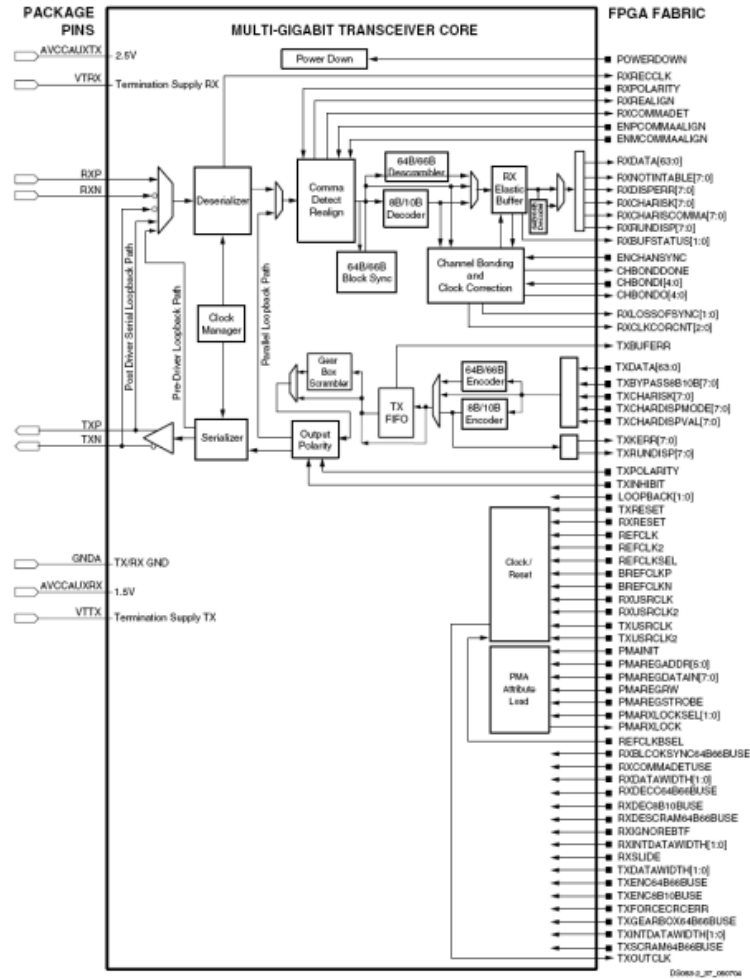


Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.





Easy, Right?





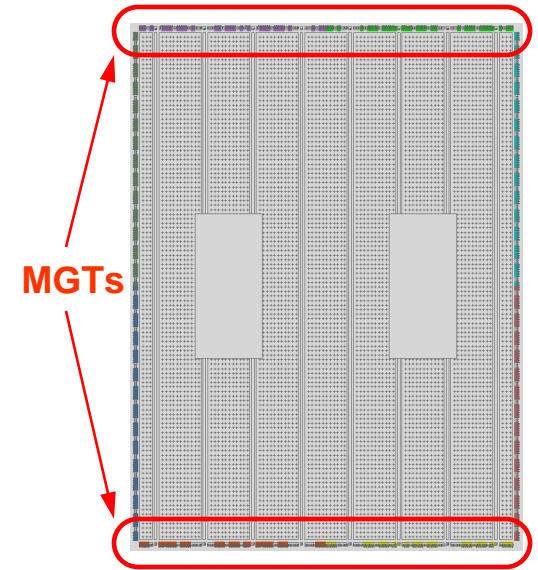
Outline

- Overview
- Clocking
- Data Interface
- Example Uses
- Summary

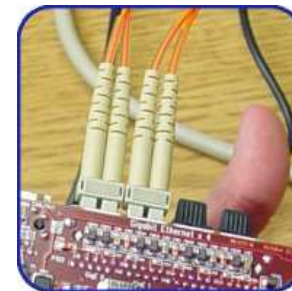


Overview

- **Xilinx Rocket I/O: Multi-Gigabit Transceivers (MGTs)**
 - Flexible units for off-chip, high-speed serial links
 - Easy communication with other hardware
- **MGT Specifics**
 - Multiple standards: GigE, IB, FC, SATA, Custom
 - Single links up to 3.125 Gb/s
 - Up to 24 MGTs per chip
 - Channel bonding
- **History**
 - Virtex II/Pro: Introduced
 - Virtex II/Pro X: Up to 6.25 Gb/s
 - Virtex 4FX: Built-in GigE MAC core
 - Virtex 5LXT: Built-in PCIe core

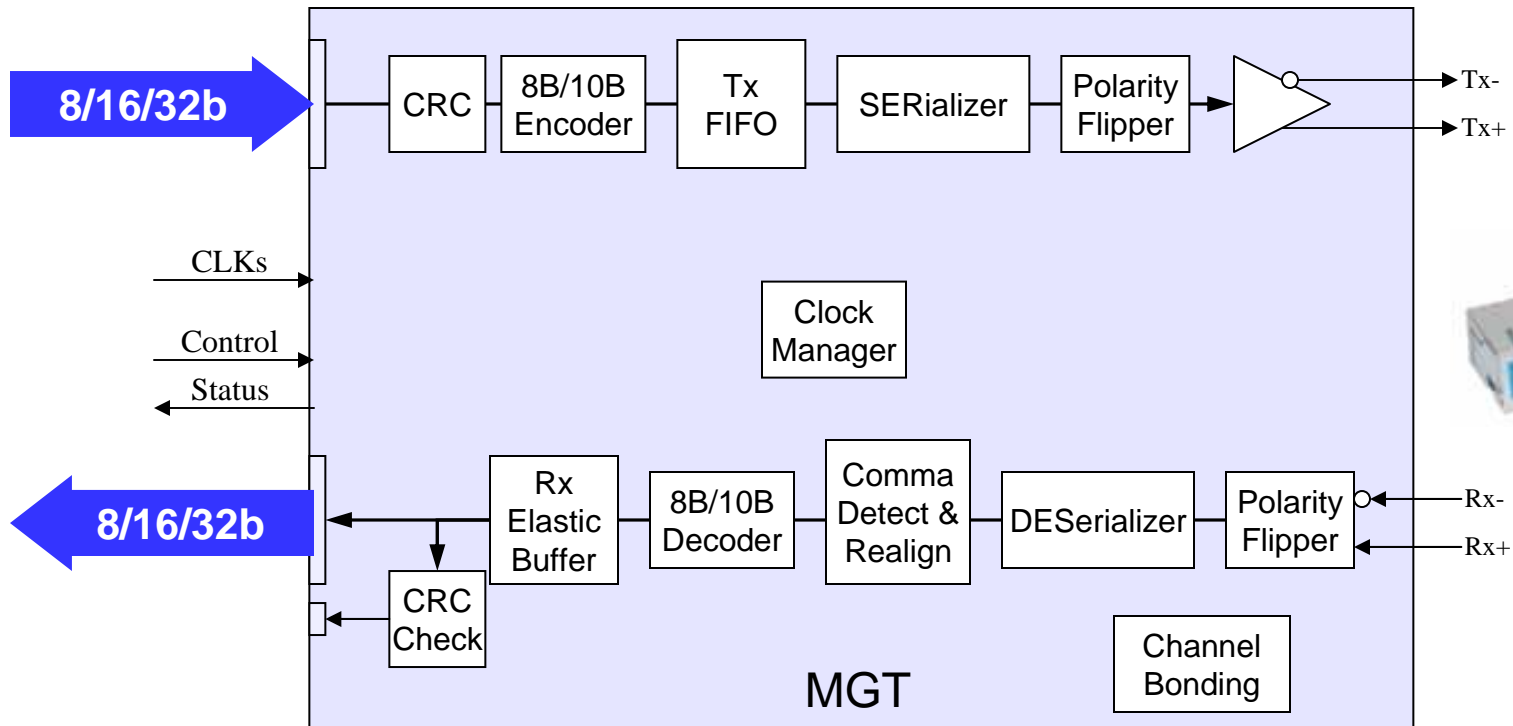


V2P20





In a Nutshell...



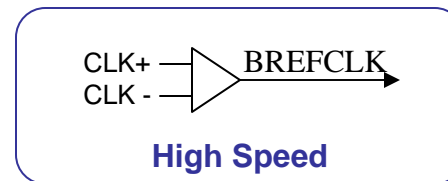
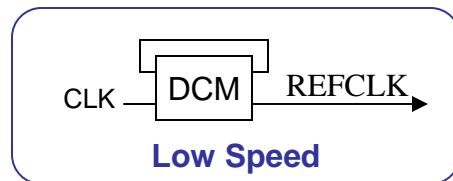
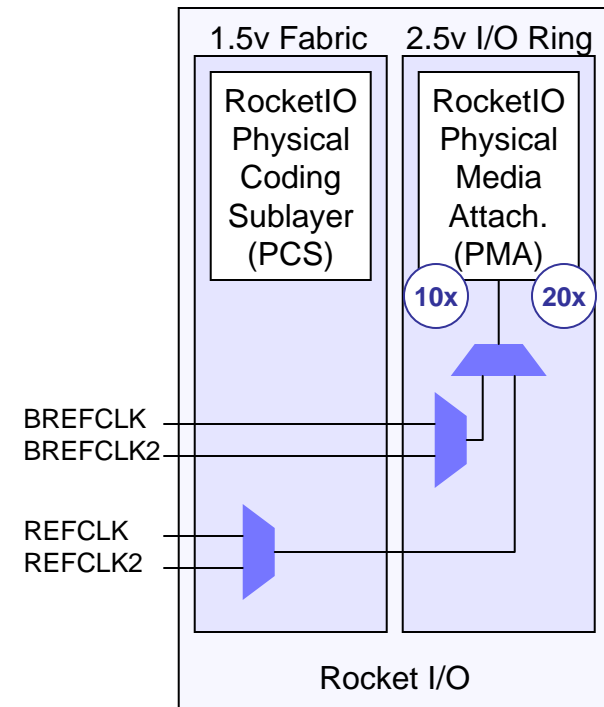


Clocking



Clock for Physical Layer

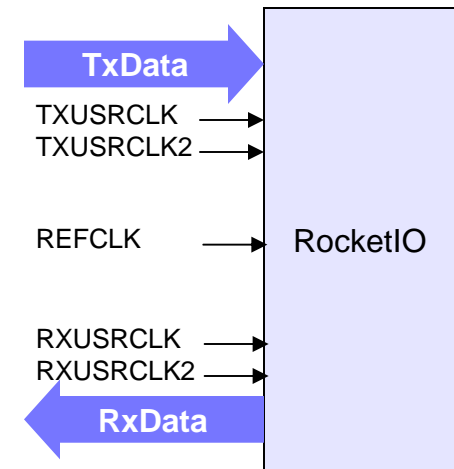
- **Base Clock drives Physical Layer**
 - 50 MHz – 156.25 MHz
 - Internal 10x or 20x multiplier (compile time)
 - Serial: 600 Mb/s – 3.125 Gb/s
 - Different clock sources
- **REFCLK: Fabric-driven clocks**
 - Up to 2.5 Gb/s rates
 - Can be driven by DCM
- **BREFCLK: Low-Jitter clocks**
 - All speeds
 - Driven by LVDS
 - Cannot route through or around chip
 - Specific pins (top and bottom)





Clocks for User Interface

- Additional clocks for Tx and Rx
 - TXUSRCLK, TXUSRCLK2
 - RXUSRCLK, RXUSRCLK2
- TX/RX Clocking depends on data width
 - Get frequency right for Phy's clock
 - Get phase right between CLK and CLK2
- Example: GigE
 - User Data Rate: 1.0 Gb/s = 16b x 62.5 MHz
 - Phy Data Rate: 1.25 Gb/s = 20 x 62.5 MHz



Clock Multiplier	Data Width	USRCLK	USRCLK2
10x	8	CLKDV (1/2)	CLK180
	16	CLKDV (1/2)	CLKDV(1/2)
	32	CLKFX180 (1/2)	CLKDV (1/4)
20x	8	CLK0	CLK2X180
	16	CLK0	CLK0
	32	CLK180	CLKDV (1/2)

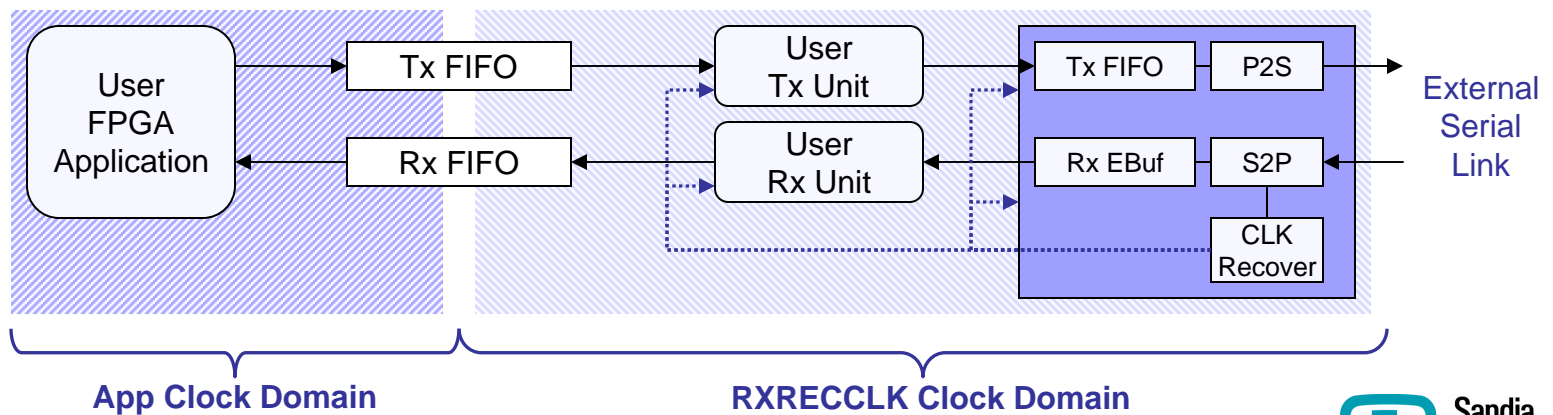
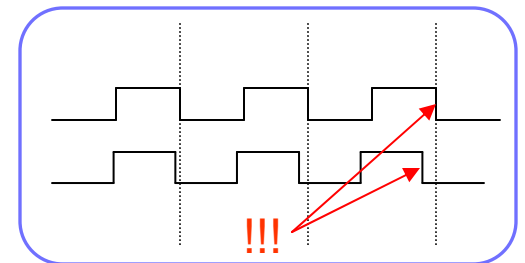
← Easy

← Easy



Handling Mismatches in Clock Frequency

- Networks: Always clock mismatch between sender/receiver
 - Slight difference in clock frequency
 - Buffer underflow or overflow
 - Under/Overflow → **Reset Unit**
- Phy does Clock Recovery on Rx stream
 - Generates *RXRECCLK*
 - Use to synch receiving hardware
 - Falls back to refclk when not locked
- Suggestion: Use recovered clock to drive user clocks





Data Interface



Data Coding for the Phy

- Coding
 - Encode/Decode an n-bit word to/from serial stream
 - Provides “transition density”, makes clock recovery easier, and allows inclusion of special symbols for controlling channel
- 8B/10B Encoding: 8-bits encoded as 10-bits
 - D-groups (regular data) and K-groups (control symbols)
 - User only gets 80% of raw physical channel
 - GigE has 1.25Gb/s SERDES speed, 1.00 Gb/s packet speed
 - 1x IB has 2.5Gb/s SERDES speed, 2.00 Gb/s packet speed
- Disparity: Want to balance number of 1s and 0s sent
 - Each input has a *positive* and *negative* 10b value
 - Phy keeps a running count and sends + or – symbol to even out

Data Example:

001 00100 → D4.1 → +1101011001, -0010101001



K-Symbols

- K-Symbols help control the serial channel
 - Some are interpreted by the physical layer
 - Make it impossible for higher levels to affect channel
- GigE Examples:
 - (K27.7) Start of Frame (SOF)
 - (K29.7, K23.7) End of Frame (EOF)
 - (K28.5, D16.2) Idle Channel (or comma)





Rocket I/O Interface

- Good News: MGT does most of the work for you
 - 8B/10B, Disparity, Word separation
 - You provide 8b data and 1b 'K' flag per byte lane
 - MGT can produce CRC for IP (leave a space)
 - Do have to worry about alignment if multiple byte lanes
- Templates for common standards
 - IB, GigE
- Suggested approach
 - Tx and Rx data every clock cycle
 - Align Rx stream based on SOF

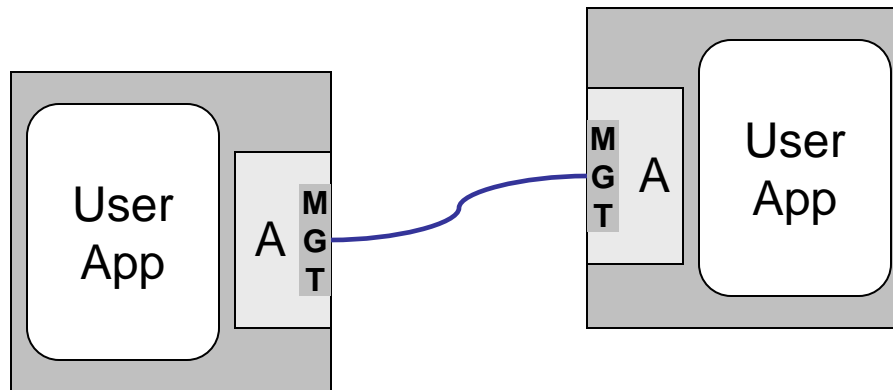


Example Uses



Raw Link

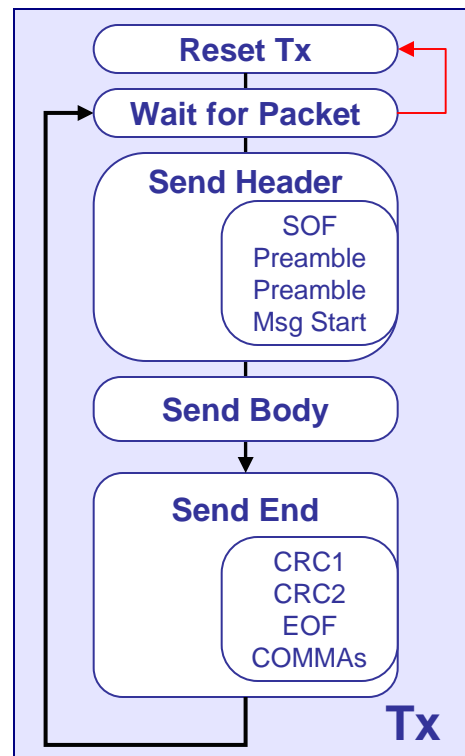
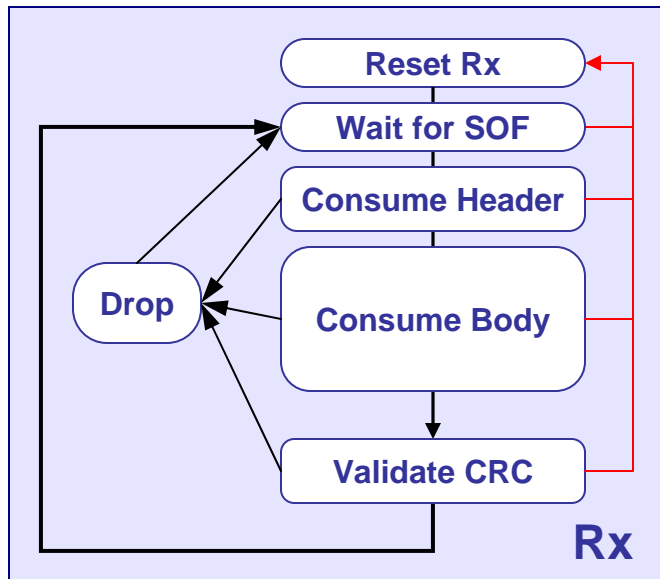
- Use the MGTs to link FPGAs together
 - Should make board design easier (fewer traces, internal buffering)
 - Doing traces similar to running network cables
- Xilinx provides Aurora core for custom link
 - Byte-stream interface (simple API)
 - Maximizes bandwidth
 - Free core





Packet Processing

- Simple GigE interface that understands IP packets
 - Rx/Tx units stream packets into/out of message queue
 - Example: Network Intrusion Detection System (NIDS)

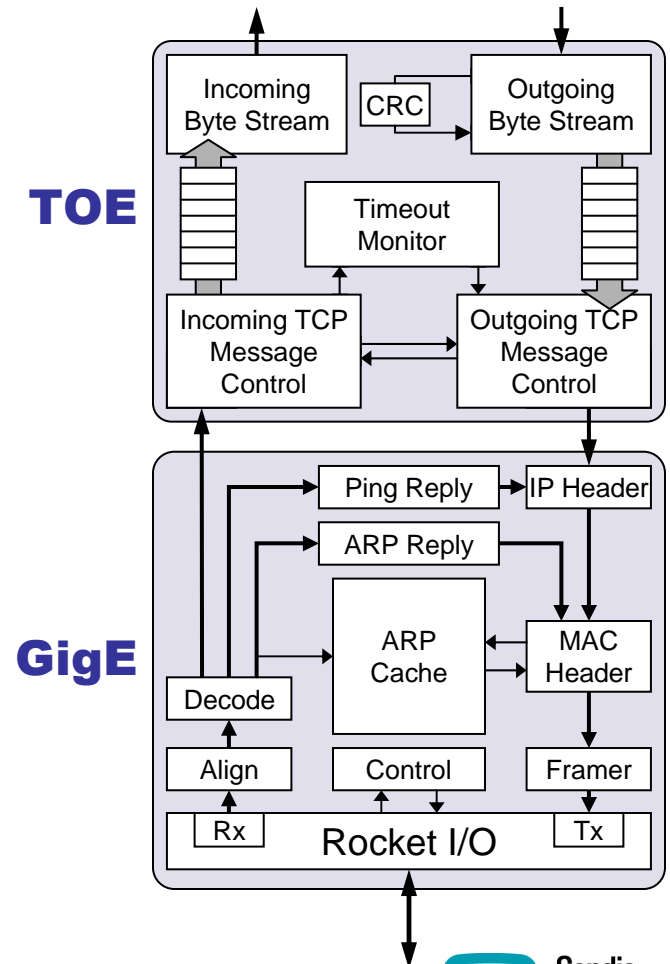




Reliable Transport

- Need for end-to-end data transfer
 - Transmission Control Protocol (TCP)
 - Ancient, complex, and nasty
- TCP Offload Engine (TOE) / GigE
 - Single stream, in-order acks
 - Slow start, nacks, retransmit, but no Nagle
- Lesson: TCP is painful
 - CRCs are at beginning of packet
 - All packets imply something
 - Multiple crossovers between Rx/Tx
 - Do it in software instead

Unit	Slices	V2P20
TOE	2,068	22%
GigE	1,102	12%





Summary

- RocketIO provides a flexible interface for off-chip communication
 - Can be tricky to setup, but easier when you understand concepts
 - Can use to make a design portable between boards
 - Can use to link multiple boards
- Advanced Topics
 - Link configuration (duplex, jumbo frames)
 - Channel Bonding: Multiple MGT links
 - Customizing the Phy