Value of a Common Electrical Modulation Scheme for CDAUI-8 from a System Perspective

Gary Nicholl, Cisco; Vasu Parthasarathy, Broadcom; John D'Ambrosia, Dell, Adam Healey, Avago; David Ofelt, Juniper;
Joel Goergen, Cisco; Kapil Shrikhande, Dell;

IEEE 802.3bs TF, Atlanta, Jan 14-16, 2015

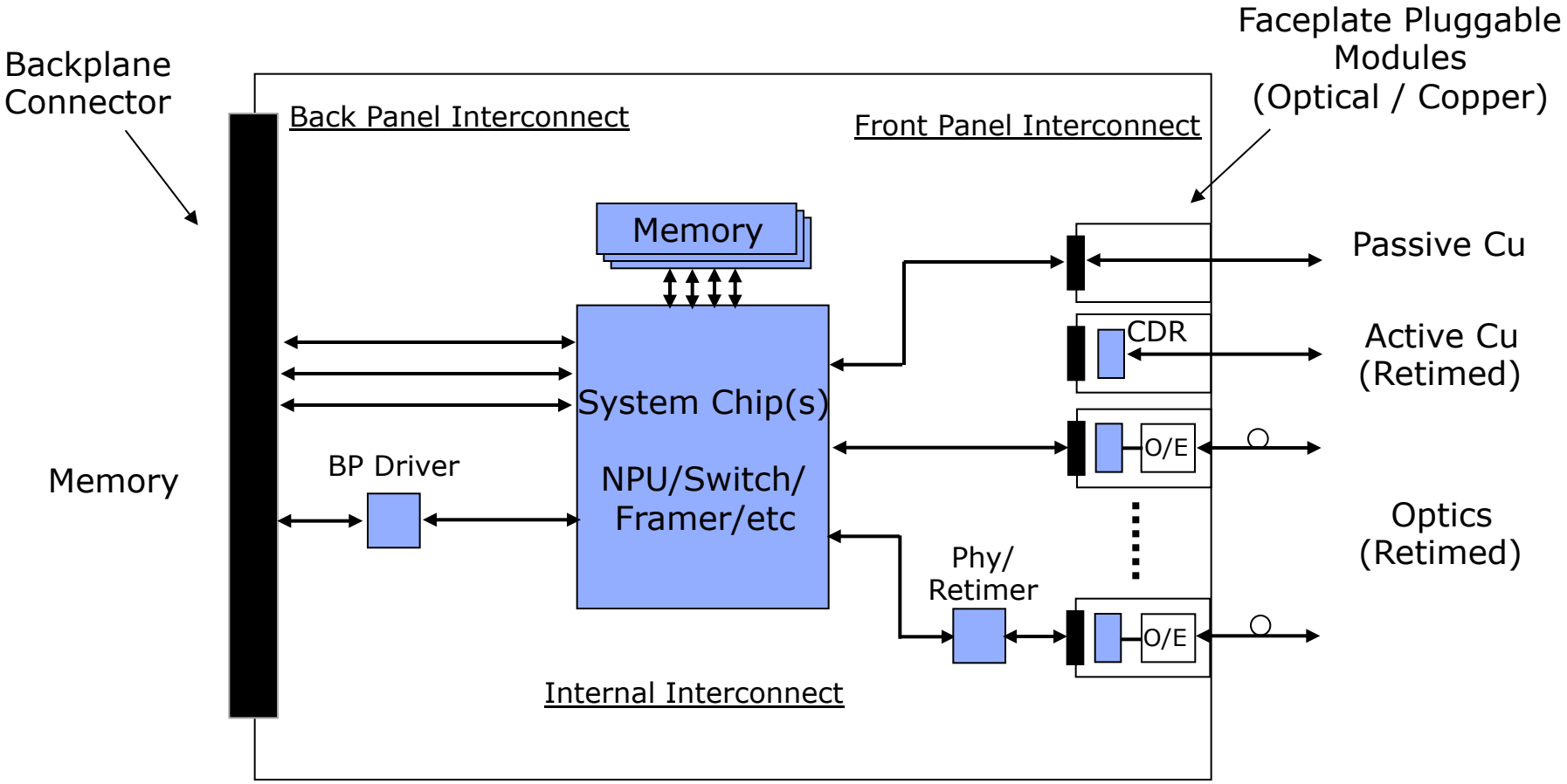
Supporters

- Mark Nowell, Cisco
- Vivek Telang, Broadcom
- Jeff Maki, Juniper
- Eric Baden, Broadcom
- Richard Mellitz, Intel
- Mike Li, Altera
- David Brown, Semtech
- Ram Rao, Oclaro
- Vipul Bhatt, Inphi
- Sudeep Bhoja, Inphi
- Brad Booth, Microsoft
- Mark Gustlin, Xilinx
- Vineet Salunke, Cisco
- Sam Sambasivan, AT&T
- Keith Conroy, Multi-Phy
- Neal Neslusan, Multi-Phy
- Paul Brooks, JDSU
- Winston Way, Neophotonics
- Ian Dedic, Fujitsu
- Scott Kipp, Brocade
- Thananya Baldwin, Ixia
- Ran, Adee, Intel
- John F Ewen, IBM
- Kent Lusted, Intel
- Rob Stone, Broadcom
- Helen Xuyu, Huawei
- Xinyuan Wang, Huawei

Introduction

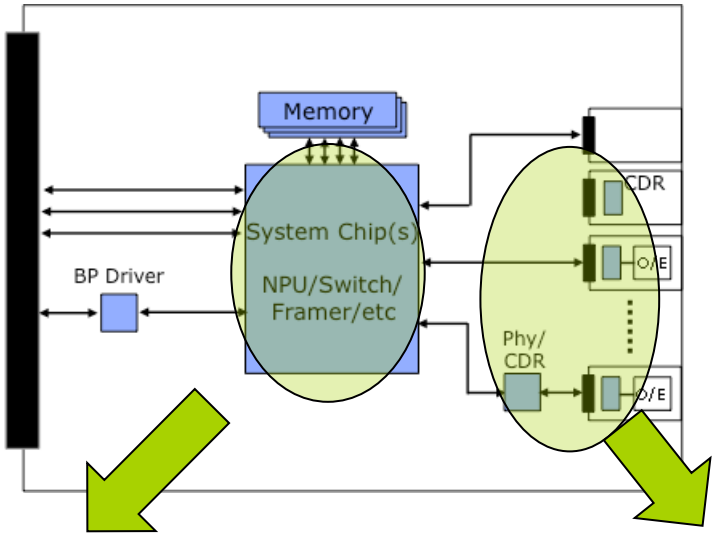
- Generic Serdes Eco-system
  - Line card architecture, chips and interconnect
- One Serdes for all applications
  - Power/area implications
- Summary and Final thoughts

“Generic” Line Card Architecture



- Multiple chips and multiple electrical interfaces
- Typically one (or more) big system chip(s), and multiple smaller ancillary chips (memory, phys, buffers, CDRs, etc)
- But not all chips are created equal …..

A Tale of Two Chips ….



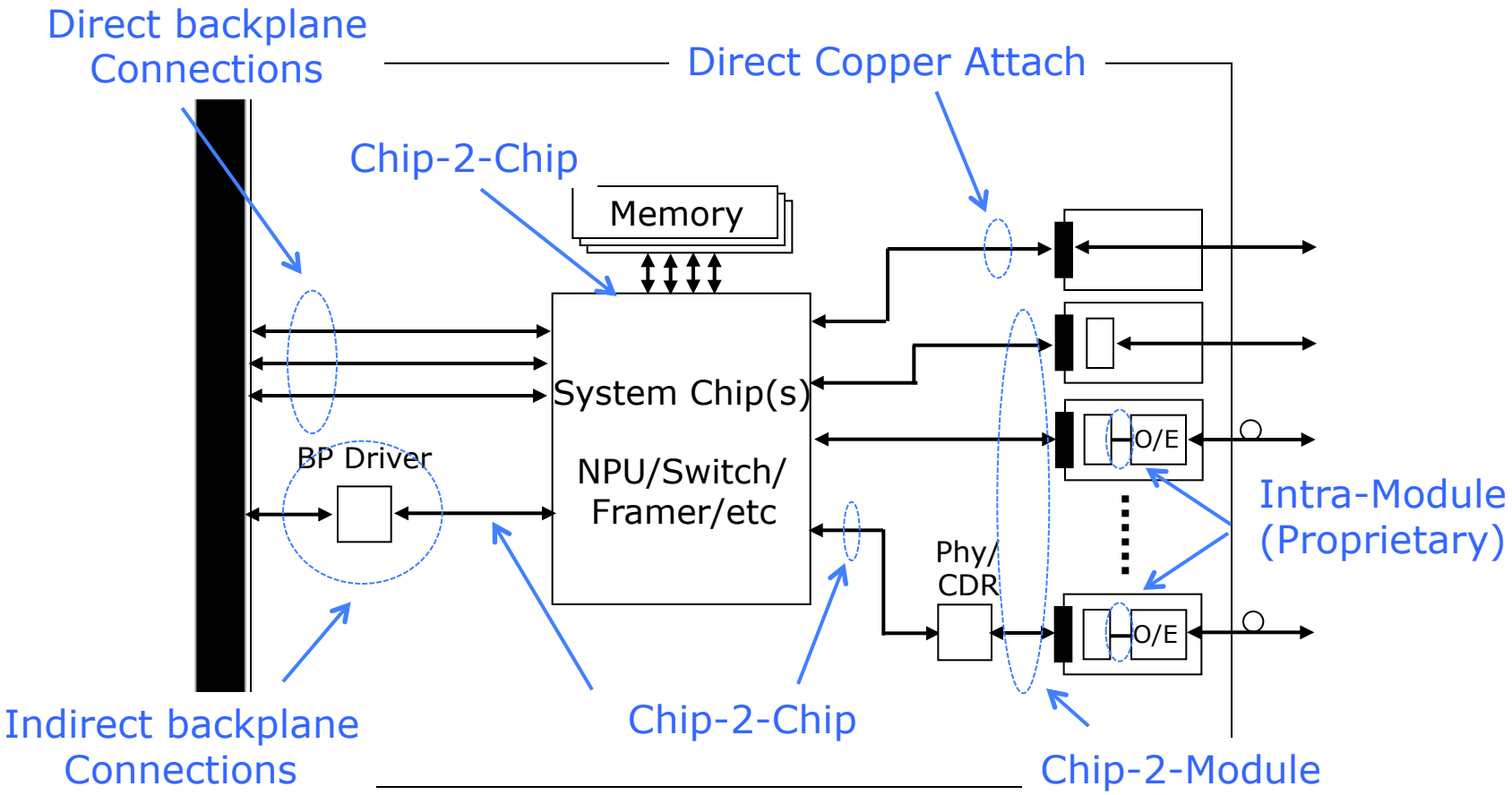
Big System Chip(s)

- Big !
- Large port counts (32+)
- High serdes counts (128+)
- Large die size
- Very expensive to develop
- Increasingly expensive mask costs
- Single design must address multiple applications
- Universal nature dictates a super-set serdes design, with multiple personalities

Ancillary Chips

- Typically much smaller
- Low port counts (1-4)
- More cost sensitive (more of them)
- Narrow application focus (specialized)
- Specialized nature allows for optimized (single personality) serdes design

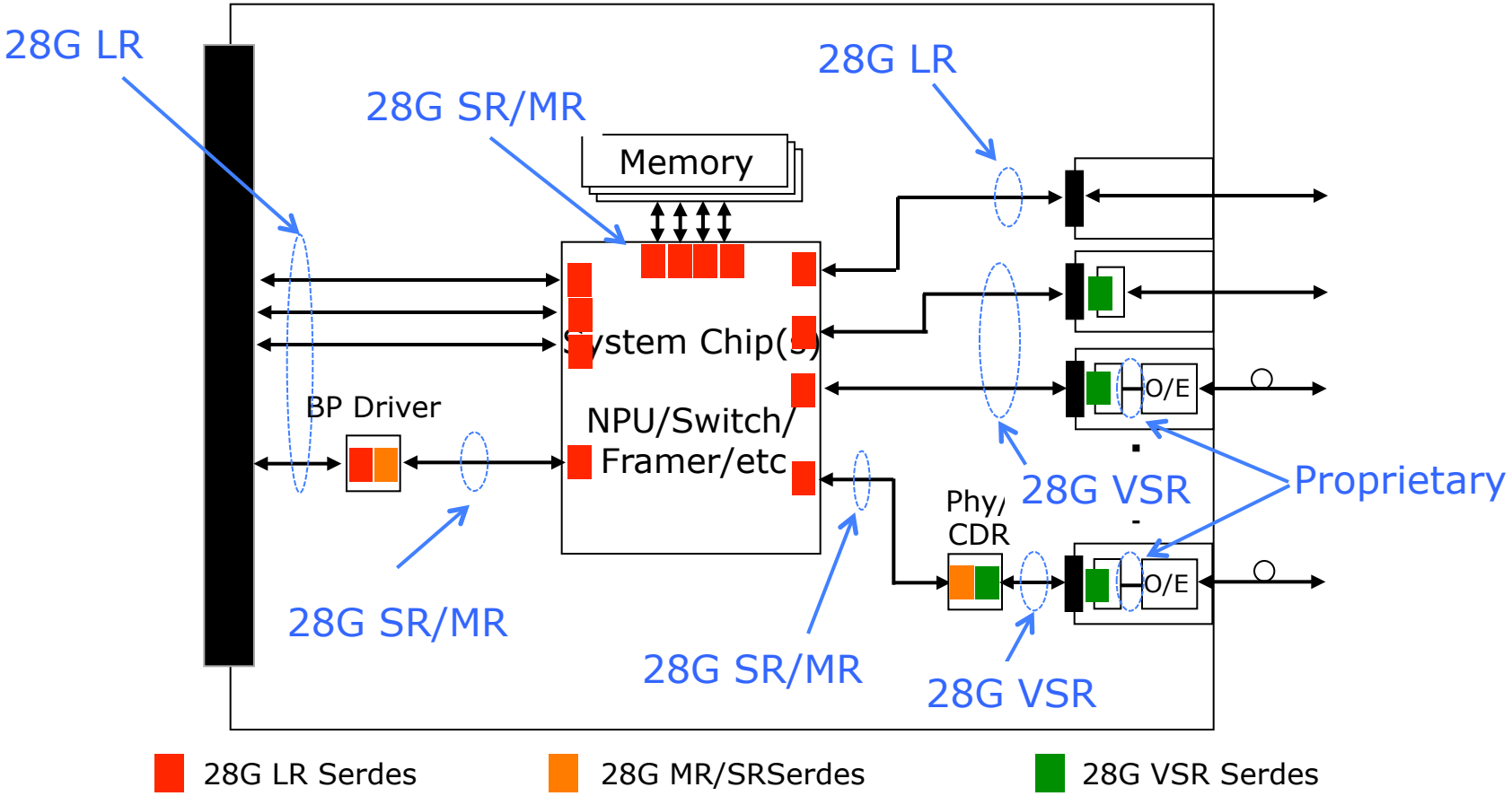
Typical Electrical Interface Types*



- Multiple electrical interfaces, each with different requirements
- A universal switch chip has to interface to all of them
- Different applications require switch ports to be wired differently
- Surgical use of Retimers (add cost/port and decrease system density)

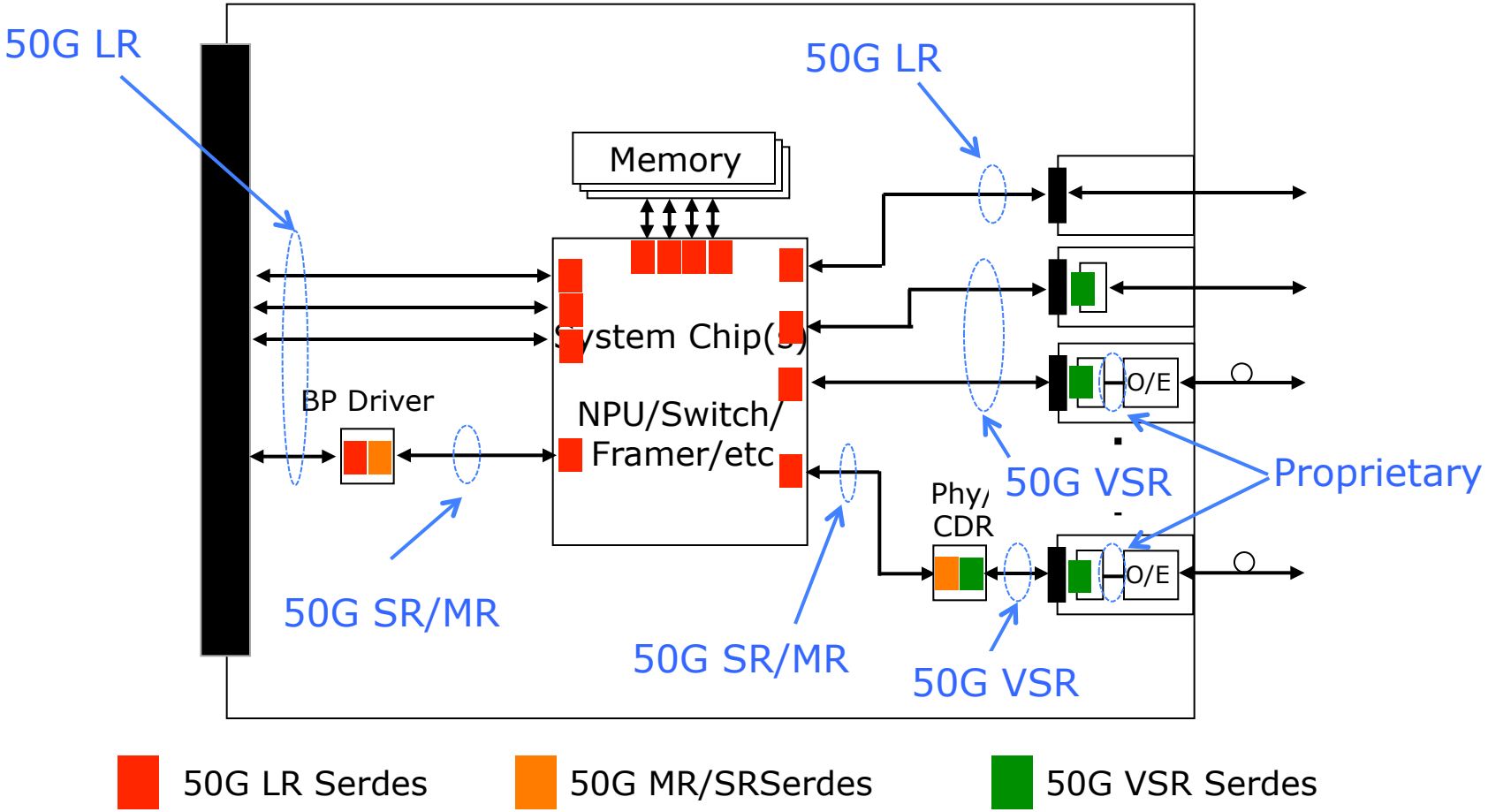
*Ignoring emerging XSR applications

28G Eco-system Example



- Switch chip typically implements 'super-set' serdes
- Switch chip serdes configurable for multiple personalities, depending on what is on the other end of the wire

How does this map to 50G Electrical ?



- Switch chip again wants to connect to multiple other serdes type
- More challenging if different modulation used for different interfaces ?

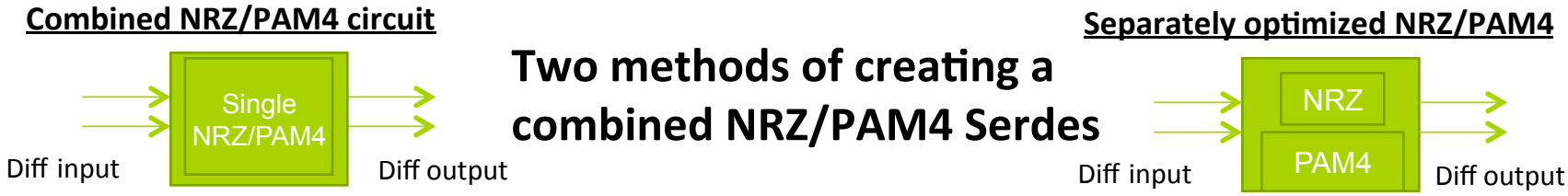
Universal Serdes: Implications

Universal Serdes: PLL implications

- What does it cost to have a Serdes that does both NRZ and PAM4?
  - PAM4 operates at ~25 GBaud while NRZ at ~50Gbaud for 802.3bs C2M/C2C
- PLL implications
  - **NRZ**: PLL and clock distribution requires full rate clock at 50GHz or multiple phase clocks at lower rate (requires well matched differential clock running at 25GHz or quadrature clock running at 12.5GHz)
  - **PAM4**: PLL and clock distribution requires only single phase clock running at full rate 25GHz, or differential clock running at 12.5GHz or quadrature clock running at 6.25GHz.
  - **PAM4***: Due to relaxed mismatch requirement, circuits using smaller transistor size can be applied; coupled with reduced frequency of operation, power saving can be 50% to 75% for PLL and clock distribution over NRZ
  - **SUMMARY**: Supporting both with a single core creates unnecessary power/area overheads due to different baud rates

*No matter what the implementation is, PAM4 would have more relaxed matching requirement than NRZ. For instance, let's assume NRZ is using half rate structure and PAM4 is using full rate structure. Then both have PLL and clock tree running at 25GHz. NRZ would require the clock to have duty cycle very close to 50%, due to its half rate architecture while PAM4 doesn't have any such requirement. Similar arguments can be made when we consider the case of both NRZ and PAM4 use half-rate implementation due to lower PAM4 clock frequency.

Universal Serdes: Transmitter implications



- **NRZ Transmitter**: Bandwidth requirement is much higher than PAM4 while PAM4 requires better linearity
  - If both NRZ/PAM4 modes are supported, the much higher bandwidth requirement (close to 100% higher) for NRZ mode would be the deciding factor for the input data path power resulting in more than 150% power overhead for PAM4
- **NRZ De-emphasis** : Very different from PAM4
  - Using a single combined data path to support both means one or both of the NRZ/PAM4 modes will not have an optimum design (difficult to optimize bandwidth/linearity simultaneously)
  - If we use separate data paths to optimize the operation of NRZ/PAM4 and create a single Serdes core/chip, then ….
    - Output pads will see almost double the load capacitance, which will degrade the S11 performance significantly (by 3-5dB). TX area will increase by 40%-80%

Universal Serdes: Receiver implications

- **NRZ Receiver**: Bandwidth requirement for the input data path is 100% higher compared to an input data buffer that supports PAM4
  - If a single combined data path is employed to support both NRZ/PAM4, the much higher bandwidth requirement for NRZ mode would be the deciding factor for input data path power resulting in more than 100% power overhead for the PAM4 mode of operation
  - If using separate data paths to optimize the operation of each mode, the input pads will see almost double the load capacitance, which will degrade the S11 performance significantly (by 3-5dB)
  - If using separate data paths to optimize each mode operation, the RX area will increase by 40%-80%.

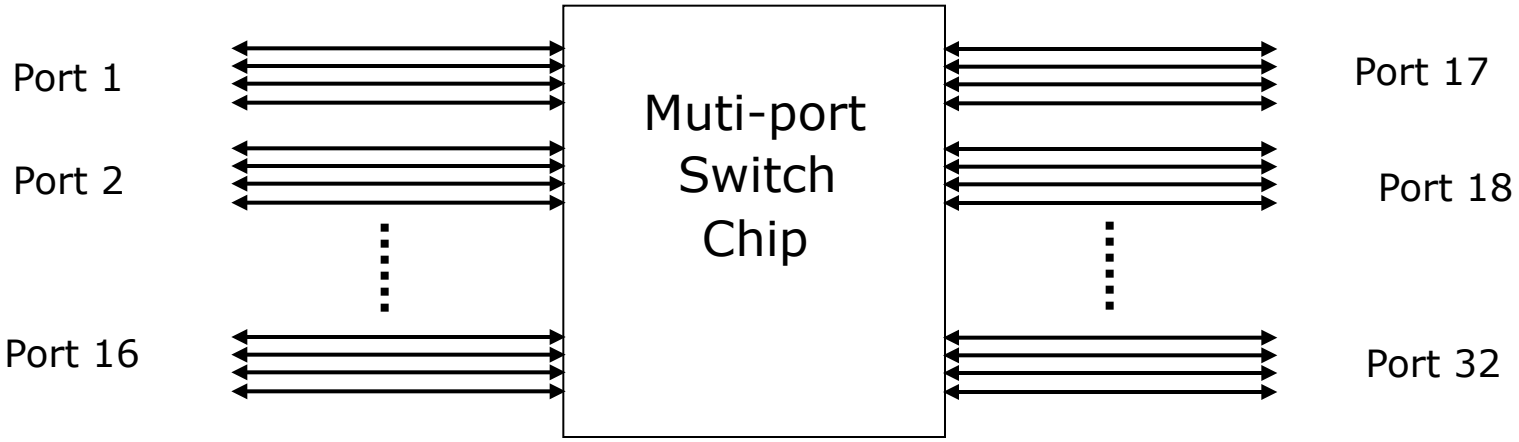
- **NRZ CTLE**: Peaking amplitude and peaking frequency requirement of the NRZ mode and PAM4 mode is drastically different (due to different baud rate)
  - Using a single data path to support both means one or both of the modes will not be able to have an optimum CTLE.

Summary

- A common signaling rate and modulation format is an enabler to building a ‘super-set’ serdes for a switch chip
- The consequences of not enabling such a super-set serdes are:
  - Need for many more ‘format translation’ chips on a given line card (drivers up power and cost)
  - Development of more specialized switch chips (fragments application space, drives up cost)
- The choice of solution for a given electrical interface is as much a system level optimization exercise, as it is an individual link/serdes optimization exercise

Backup

“Generic” Switch Chip Architecture



- Industry direction: Universal, high port count Ethernet Switch chips
  - e.g. 32 x 100G ports (192 x 25G serdes)
- Switch chips are used in multiple different applications
- Switch chip connectivity is different in different applications
  - TOR (all front panel connectivity)
  - Spine Switch (mix of front and back panel connectivity)
- Switch Serdes solution needs to be flexible
  - **Cannot be customized** for a given application