Power Considerations For 100GBASE-KR4 PHY

Oren Sela
Jonathan Malkiman
Samuel Attaly
Nov 2011

Mellanox®
TECHNOLOGIES

Supporters/Contributers

- Pravin Patel
- Mike Dudek
- Scott Kipp
- Piers Dawe
- Mike Peng
- Stephen Bates

Overview

- The Economic value of enabling the 100GigE over KR grade backplane has been proposed by the group
  - chalupsky_02_0311
  - parthasarathy_01_0111
- To Enable 100GigE over "legacy" systems need to meet their power budget
- Most challenging component power-wise is the switch blade
- Analysis overview-
  - Show the power budget for the switch blade on the majority of the 40GBASE-KR4 compatible systems
  - Accumulate all the power component excluding the 100GBASE-KR PHY
  - Show the estimated power budget for the PHY
- This analysis covers >70% KR4 compatible blade systems

Assumptions And Metrics

- **Best case approach**
  - Lowest power estimation for each component
  - Provide a max power budget for the PHY

- **Only blade systems which support KR4 or are KR4 ready**
  - Since KR lanes are single lane many KR compliant system used in the broad market potential are excluded

- **16/18 Blades systems –**
  - Switch blade is either 32/36 ports switch –
    - 16/18 Internal KR4 ports
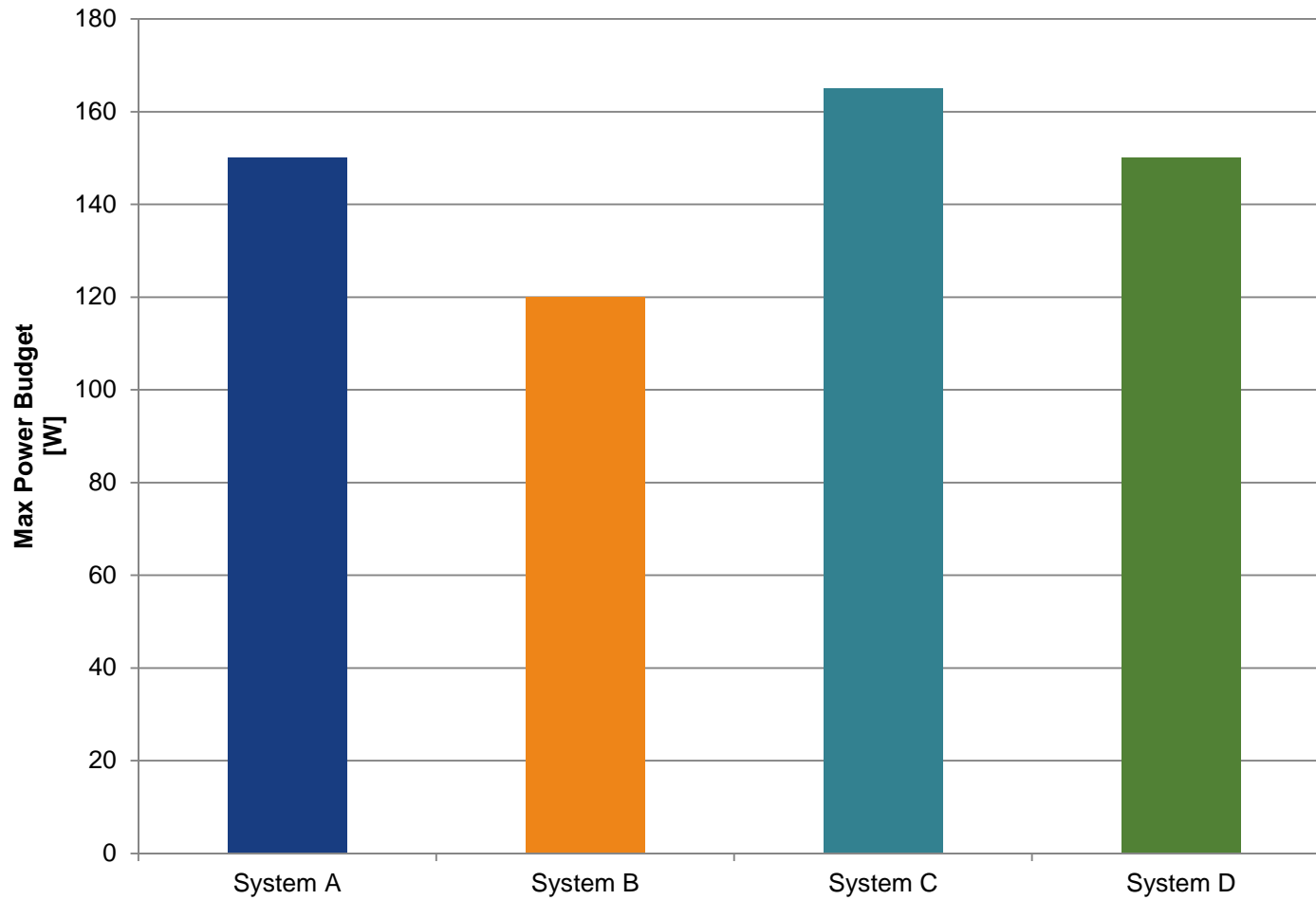   - 16/18 External CR4/SR4/LR4

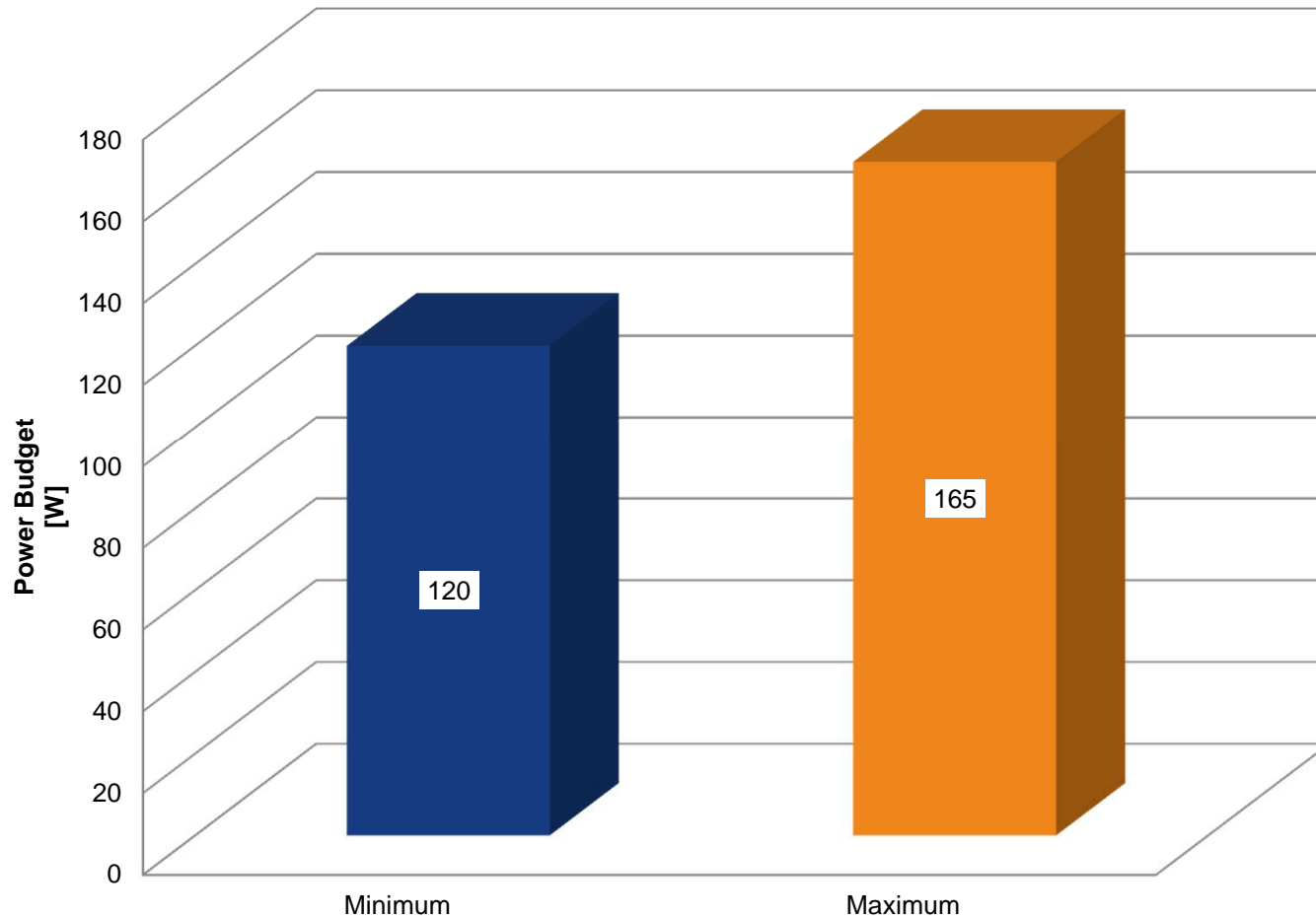# Switch Blade
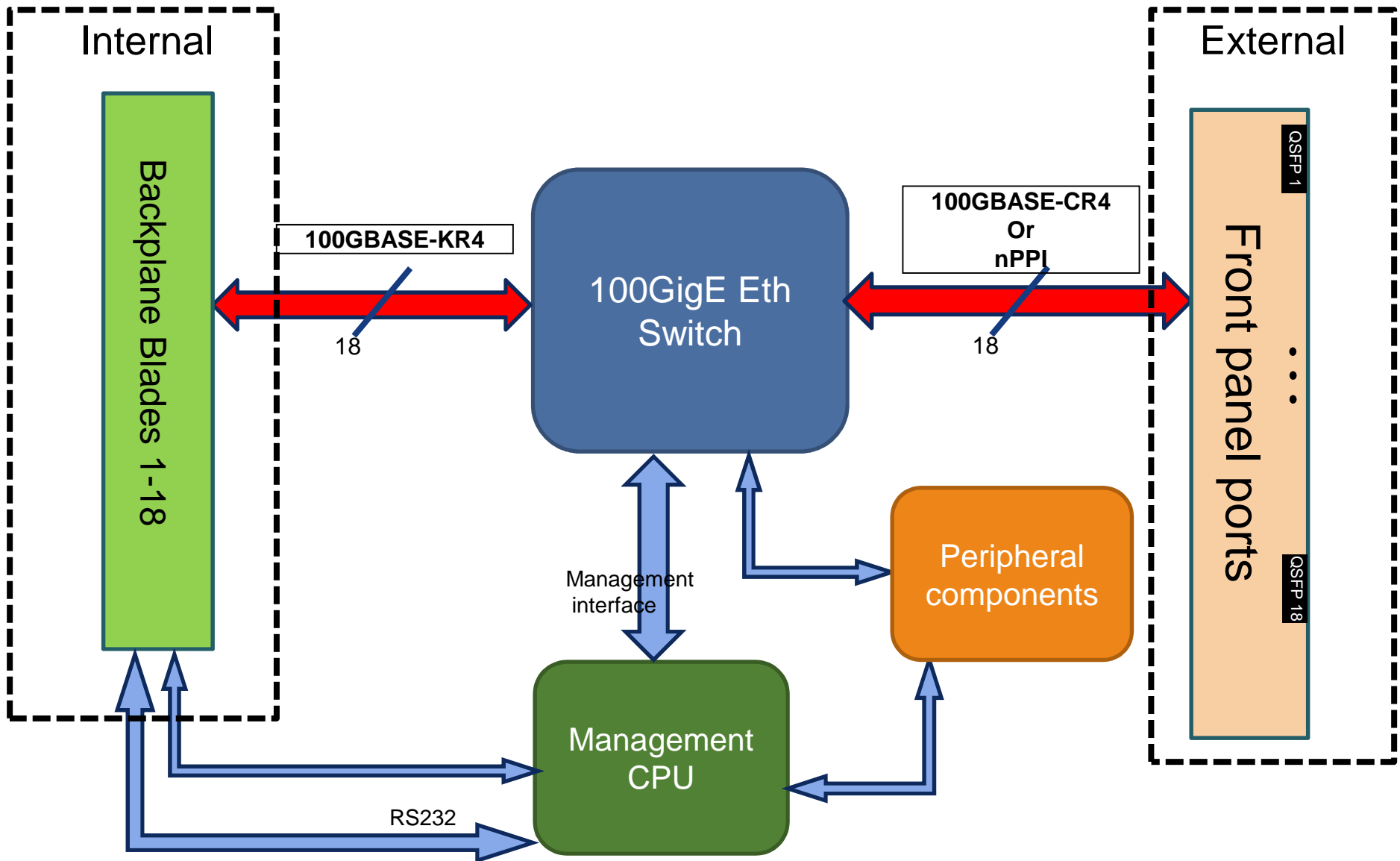
Switch Blade Power Budget

Different Power Budgets

Maximum And Minimum Power Budget

Min-Max Power Budget

NOTE – The Minimum system is a 16 blade system
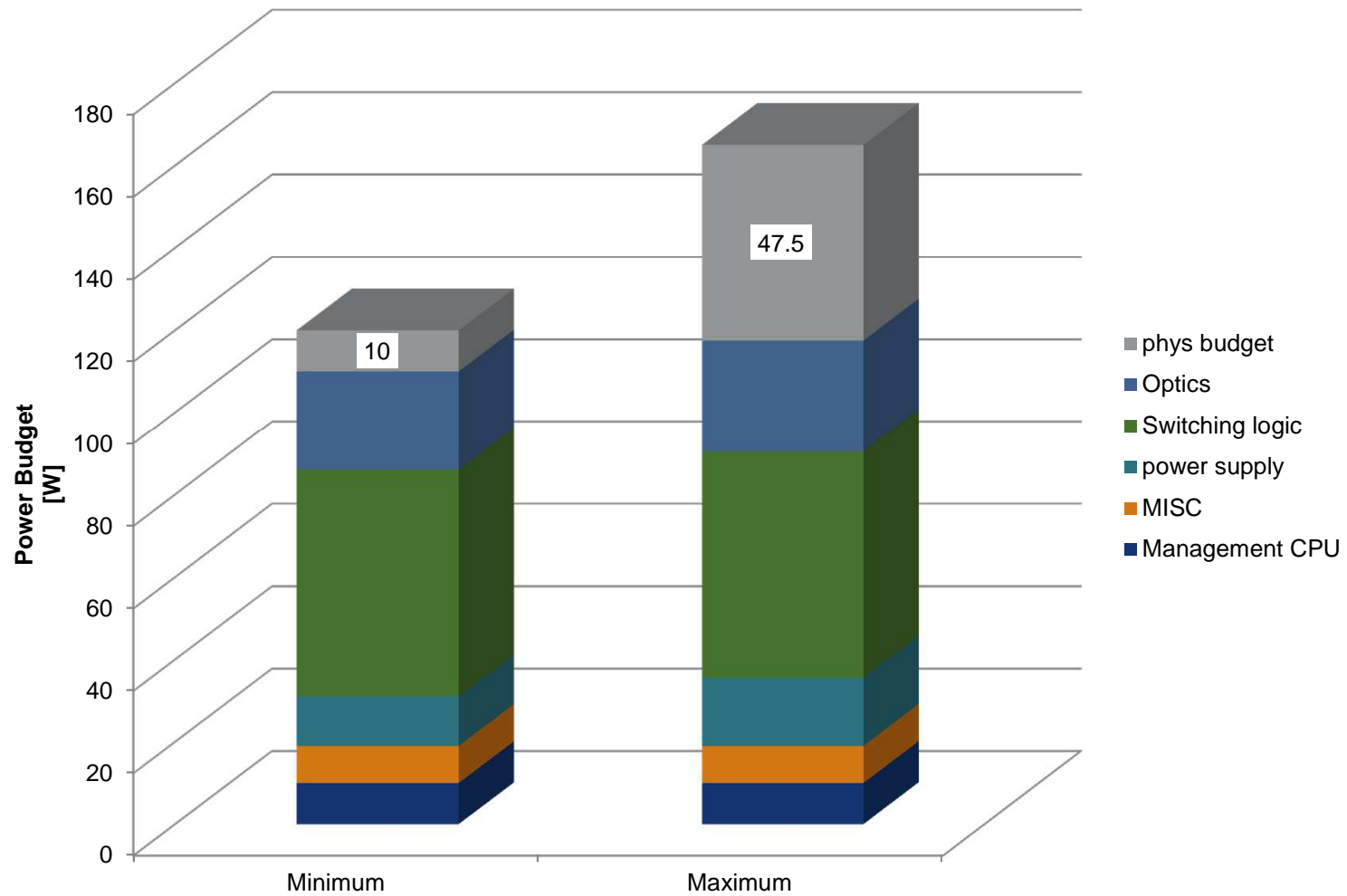
Power Components For Switch Blade

Component	40G Max [W]	100G Max [W]
Switching logic (excluding PMD)	55	55
Active Optics	1.5 (per port)	1.5 (per port)
Management CPU	10	10
Misc	9	9
Power supply efficiency	10%	10%

■ Misc- All power consuming components:
- LEDs, Oscillators, EEPROMs, low speed signals

Max Power Budget

PHY Power Requirements

- PHY is a 4 lane full duplex as defined in 802.3bj objective
- PHY power includes:
  - PLLs
  - Transmitter power
  - Receiver power – including all equalization analog and digital logic
  - CDR
  - FEC
  - PMD logic

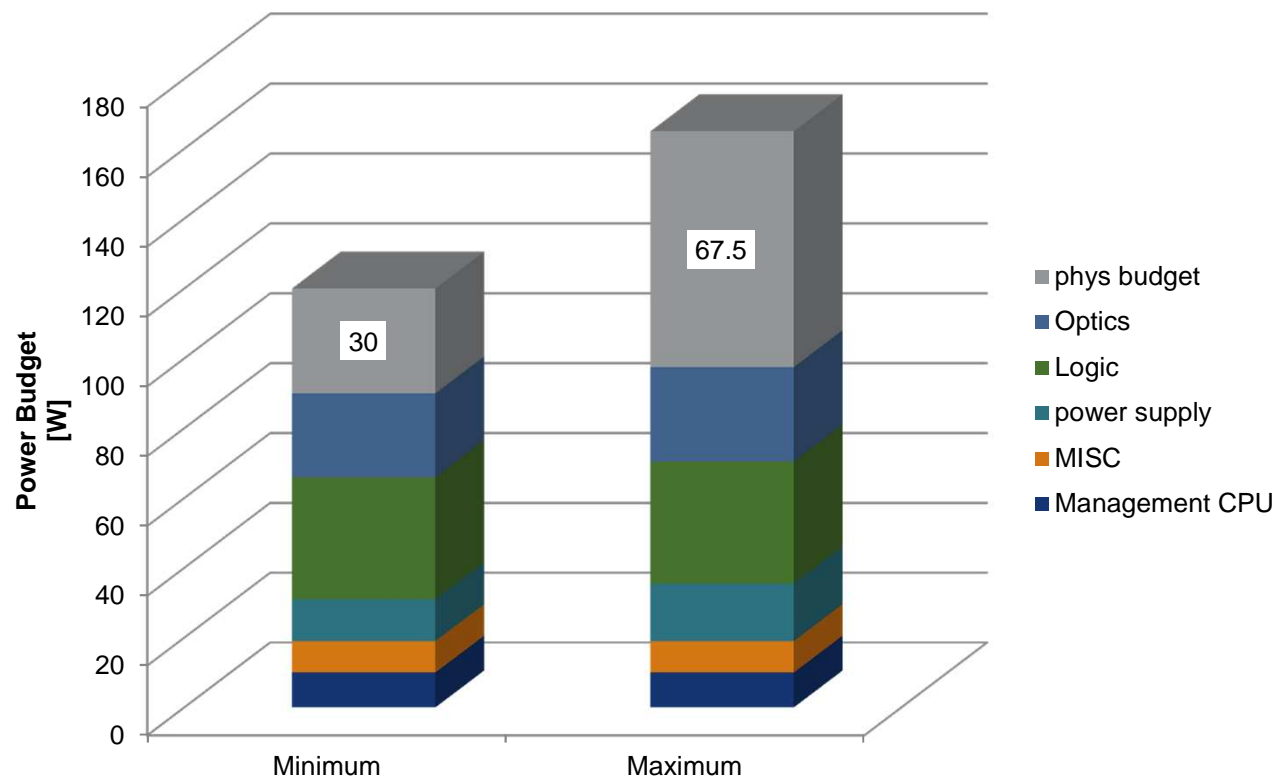| System	Total Budget [W]	Per x4 [W]	Per Lane [W]
Minimum	10	0.31	0.08
Maximum	47.5	1.32	0.33

Assume same power for KR4, CR4, nPPI PHY

- Removes the core switching power consumption, lower processor power
- Increase the total power and cost due to additional rack switching
- Assume same encoding

Pass-Through Power Budget

Pass-Through Module - PHY Power Requirements

- PHY power includes:
  - PLLs
  - Transmitter power
  - Receiver power – including all equalization analog and digital logic
  - CDR

System	Total Budget [W]	Per x4 [W]	Per Lane [W]
Minimum	30	0.94	0.23
Maximum	67.5	1.88	0.47

Assume same power for KR4 and CR4/SR4 PHY

More Power Consumers

- **Different modulation on CR4 and SR4 PMDs**
  - Requires conversation in the optical module – consumes more power

- **Retimed optical interface**

- **Failover CPU**

- **External PHYs**
  - Either on the internal or external switch side

- **External switch memory**

Summary

- **To Support KR4 compatible system; Max PHY power requirements -**
  - To support the majority of the systems - 1W
  - To support any systems – 1.3W
  - To support all systems – 0.3W
  - Including PLLs, TX and RX

- **Using a Pass-Through only option; Max PHY power-**
  - To support the majority of the systems- 1.5W
  - To support any systems – 1.8W
  - To support all systems – 0.94W
  - Total Power and cost will be higher
  - Industry trend is moving from Pass-through to switches

- **This is a best case analysis**
  - Extra Margins Require

- **Due to cooling constraints – limited options to extend power budget**

Conclusions

- To support the legacy 40GBASE-KR4 system aggressive power targets are needed

- PHY power consumption should be considered when analyzing system cost-
  - Example – more expensive material can result in lower power architecture and as a result cheaper cooling and power supply
  - Green Data Center Require Low power design

- Material and PHY defined by 802.3bj must meet power requirements
  - PHY power target of ~1W

Thank You

Mellanox®
TECHNOLOGIES