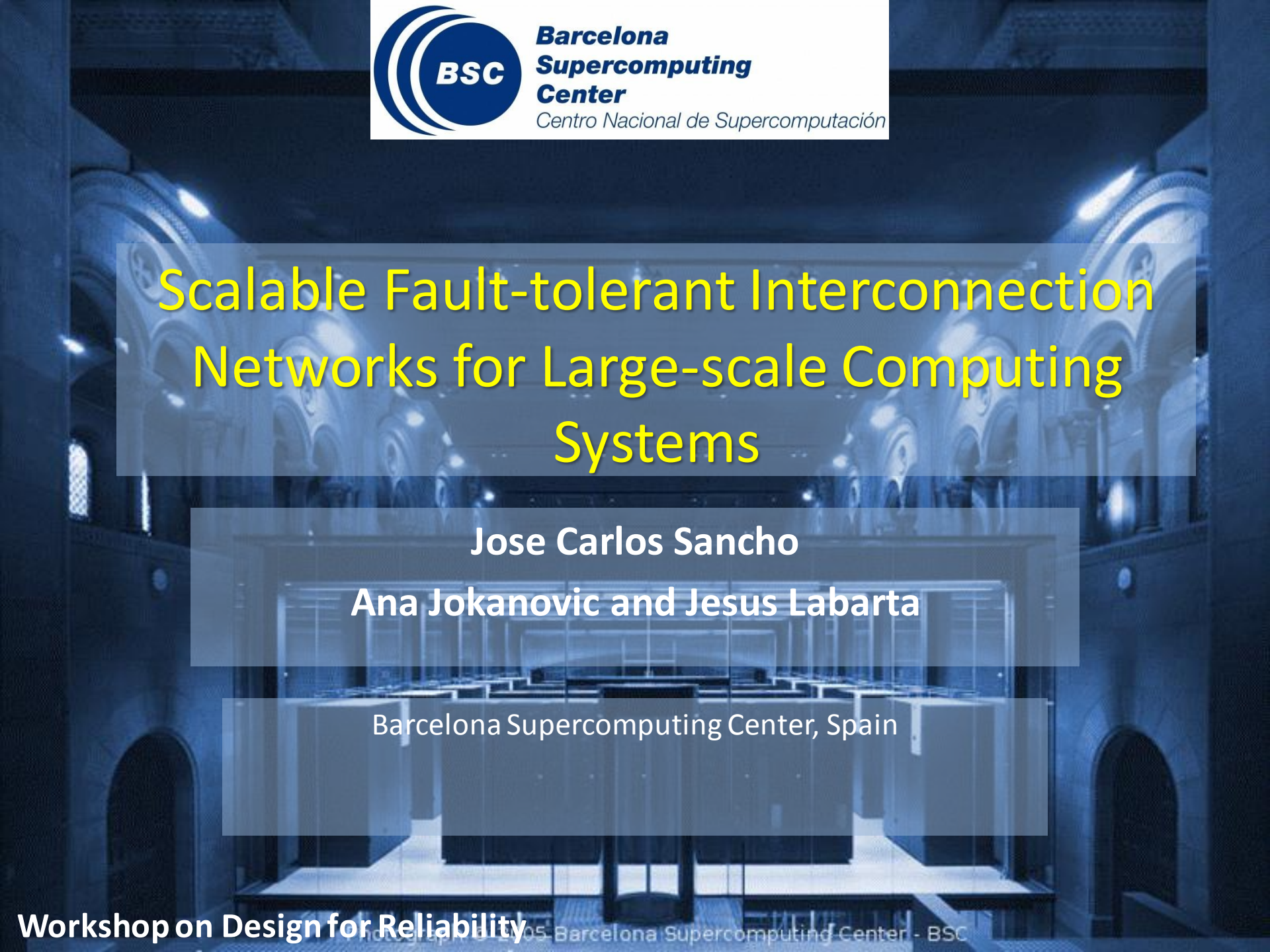


Scalable Fault-tolerant Interconnection Networks for Large-scale Computing Systems

Jose Carlos Sancho

Ana Jokanovic and Jesus Labarta

Barcelona Supercomputing Center, Spain

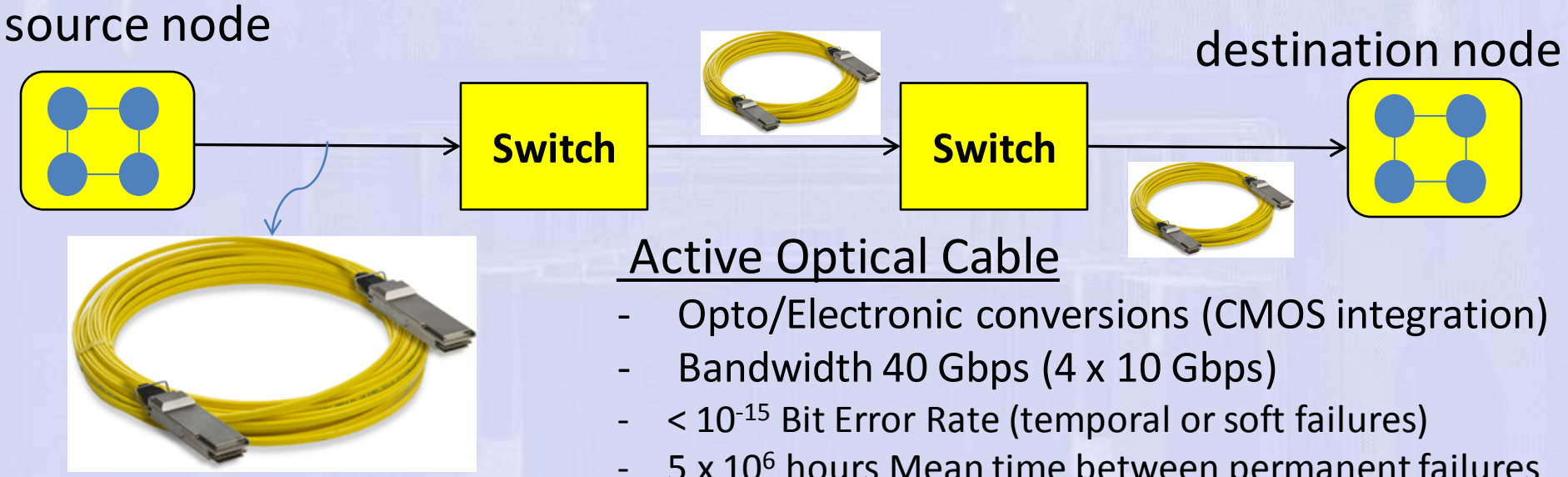
Workshop on Design for Reliability

Outline

- HPC network interconnects trends
- Future network failure expectations
- Handling errors in today's HPC networks
- Proposed scalable fault-tolerant approaches
  - Permanent or hardware failures
  - Transient failures or bit errors
- Conclusions

HPC interconnect technologies

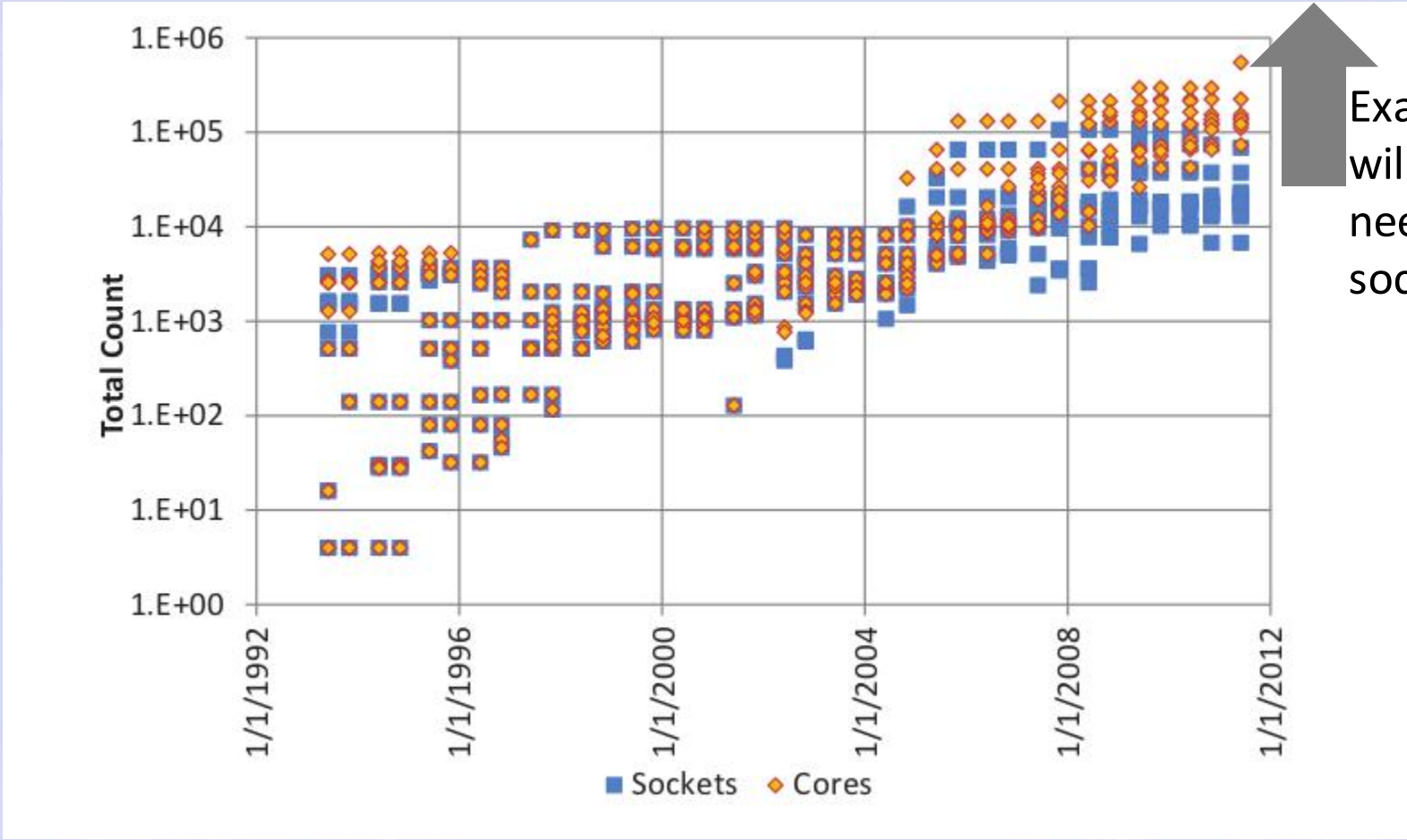
- **Optical interconnects are preferred over copper**
  - Size: Lower cable size & weight, Smaller connector size
  - Attenuation: Longer distance, more power-efficient transmission at higher BW*distance
  - Isolation: Lower EMI, less crosstalk, more reliable



Active Optical Cable

- Opto/Electronic conversions (CMOS integration)
- Bandwidth 40 Gbps (4 x 10 Gbps)
- < 10^{-15} Bit Error Rate (temporal or soft failures)
- 5 x 10^{6} hours Mean time between permanent failures

HPC networks size



Top10 historical data

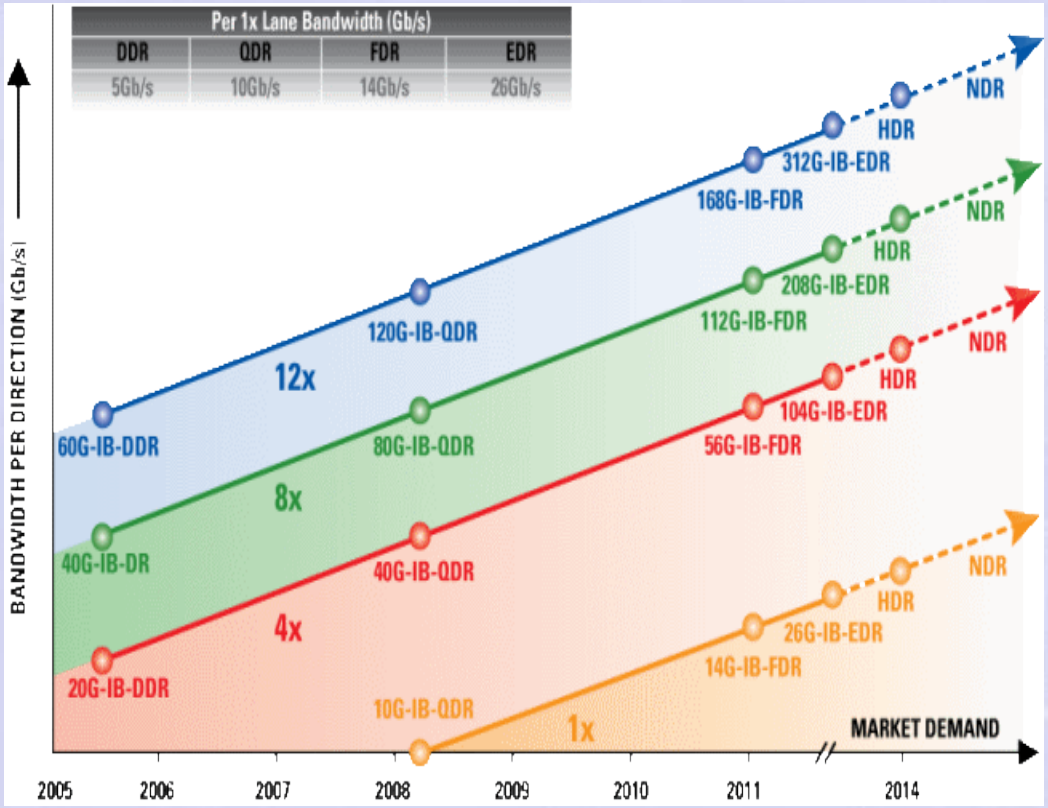
Source:

Peter M. Kogge and Timothy J. Dysart, “Using the TOP500 to Trace and Project Technology and Architecture Trends”, SC11

HPC network bandwidth

- **HPC network bandwidth is increasing**

#Cores growing faster than BW!!



Source:
InfiniBand Roadmap, available from http://www.infinibandta.org

Power wall

- **The Energy issue is well understood**
  - **Systems are constrained by power, exascale < 20MW**

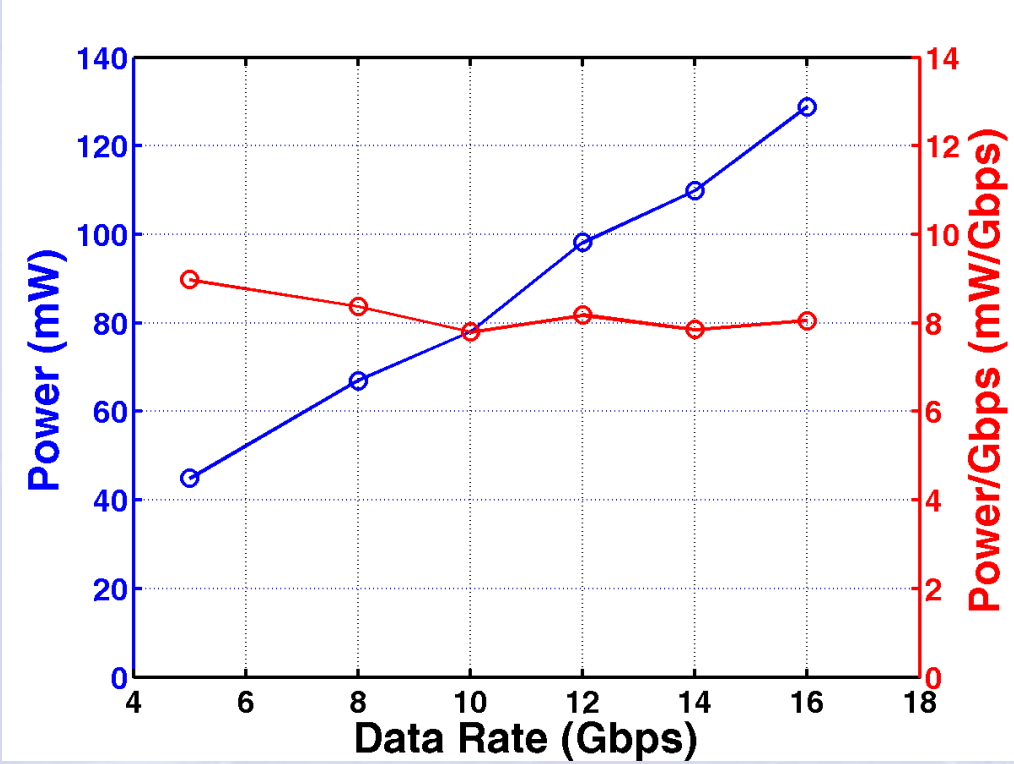
Coal energy plant

- **Data movement is the key to save energy**
  - **Up to 200 x more energy needed to transport a bit from a nearest neighbor chip than to operate on it:**
    - Energy needed for a floating point operation (0.1-0.05 pJ/bit)
    - Energy needed for (electronic) data-transport on card (2-10 pJ/bit)

Sources:
DARPA/IPTO study, by Peter Kogge, et. al. available on http://www.nd.edu/~kogge/reports.html

Optical transceiver power

- **Power scales with data rate**

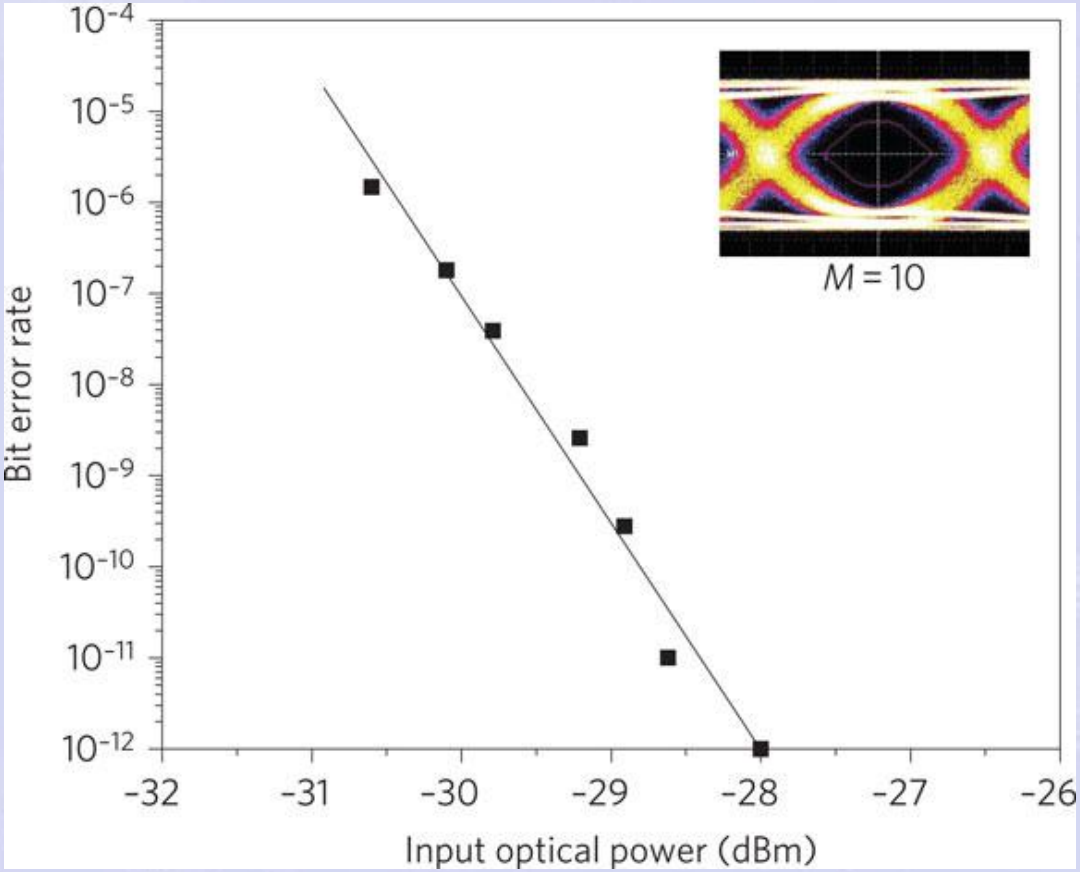


BER < 10^{-10}

Source:
Samuel Palermo, “Energy Efficient CMOS Optical I/O”, CMOS Emerging Technologies Conference

Optical BER

- **Low power increases BER**



$$\text{Power (in dBm)} = 10 \log_{10} \frac{\text{Power}}{1\text{mW}}$$

1 mW = 0 dBm
1 µW = -30 dBm

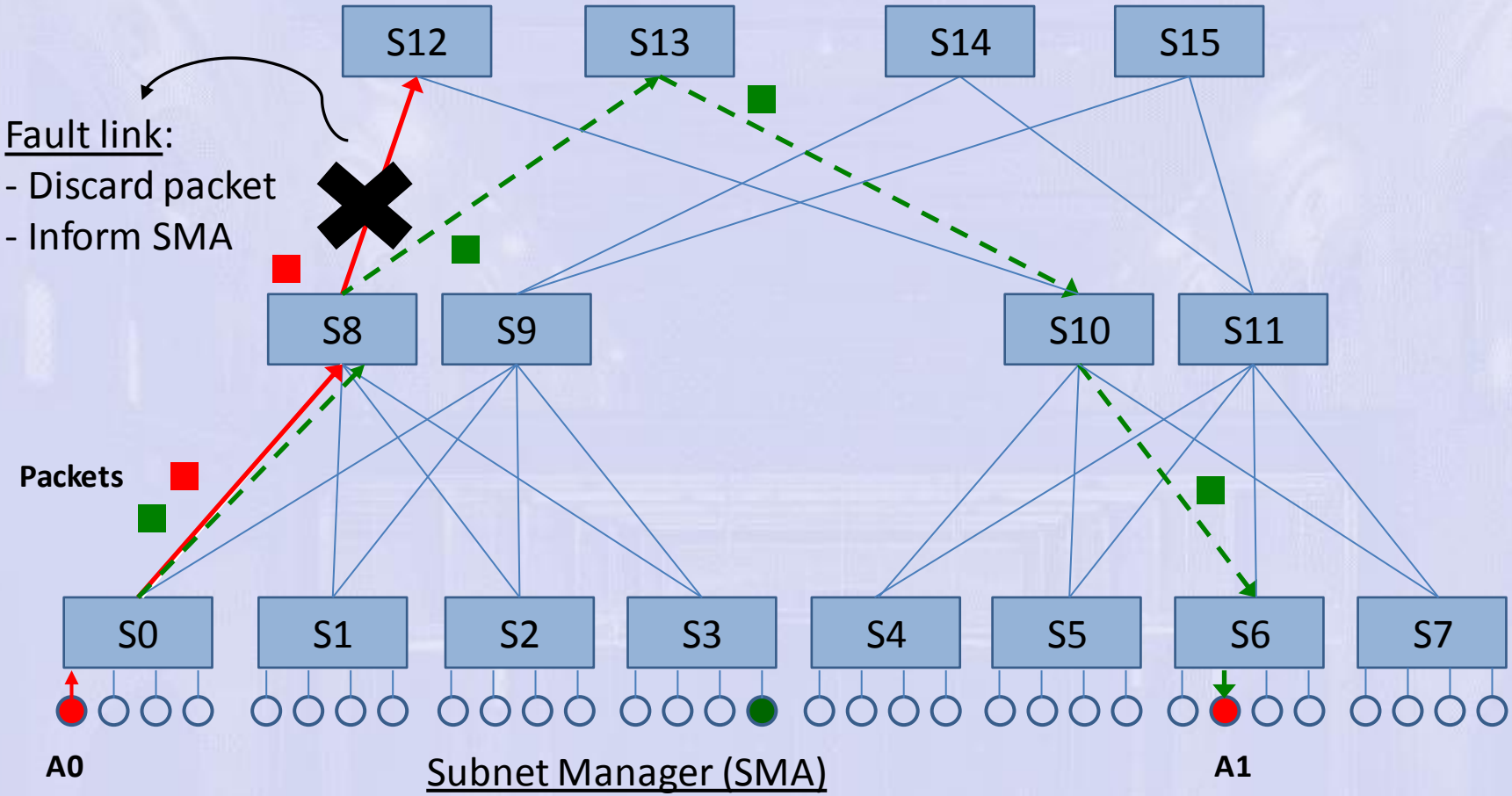
Source:
Yimin Kang, “Monolithic germanium/silicon avalanche photodiodes with 340 GHz gain–bandwidth product”, Nature Photonics 3, 59 - 63 (2009)

Power wall implications

- **Trading off power with BER @ high data rates**

Handling errors in InfiniBand (1)

Hardware failures

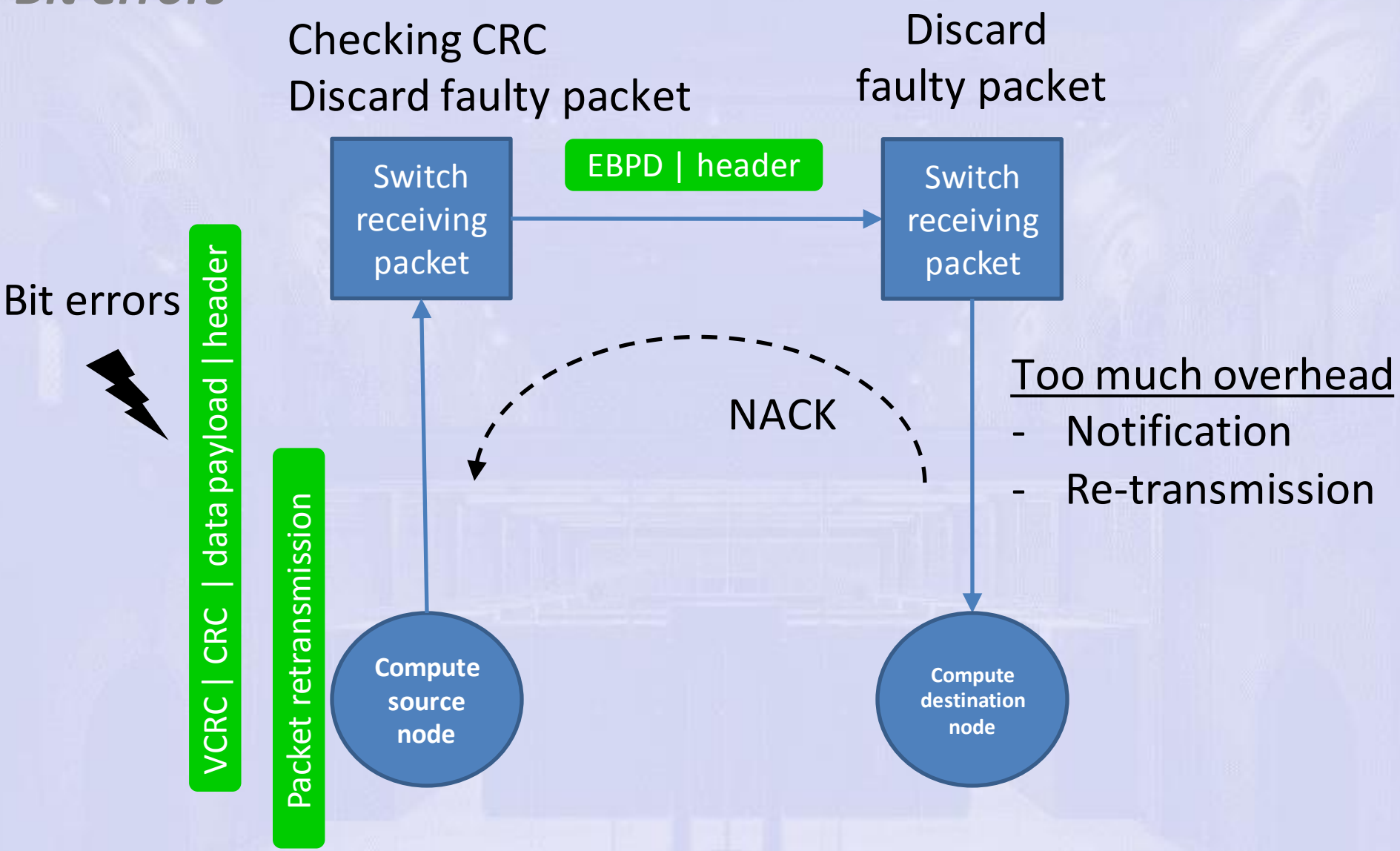


Subnet Manager (SMA)

- Re-program routing tables at S8
- Inform A0

Handling errors in InfiniBand (2)

Bit errors



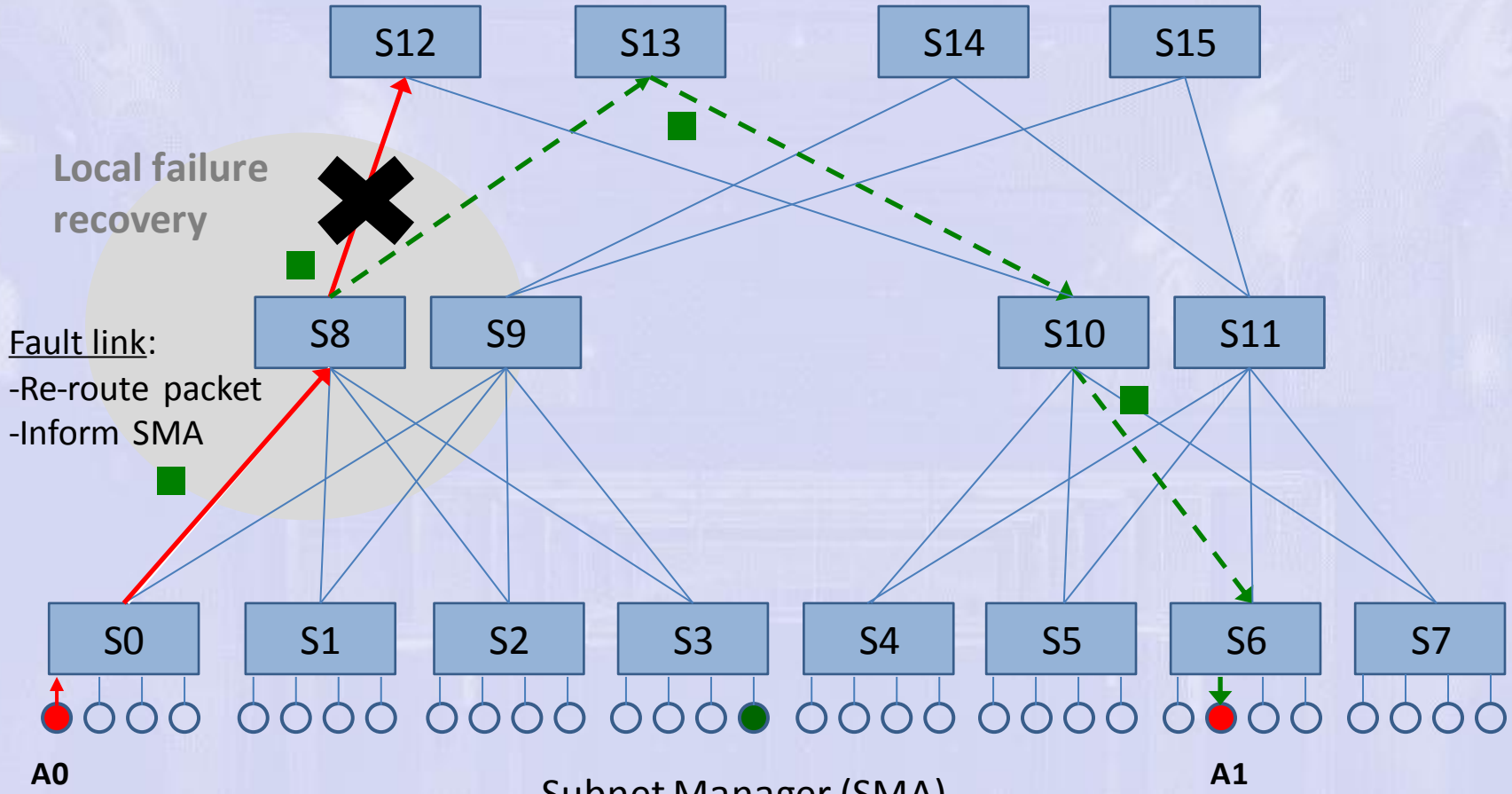
Scalable fault-tolerant approaches

- Longer distances between end nodes prevent to handle errors efficiently

Handle it locally !

Local hardware failure recovery

Open issues
- Deadlock?
- Who to re-route packet?

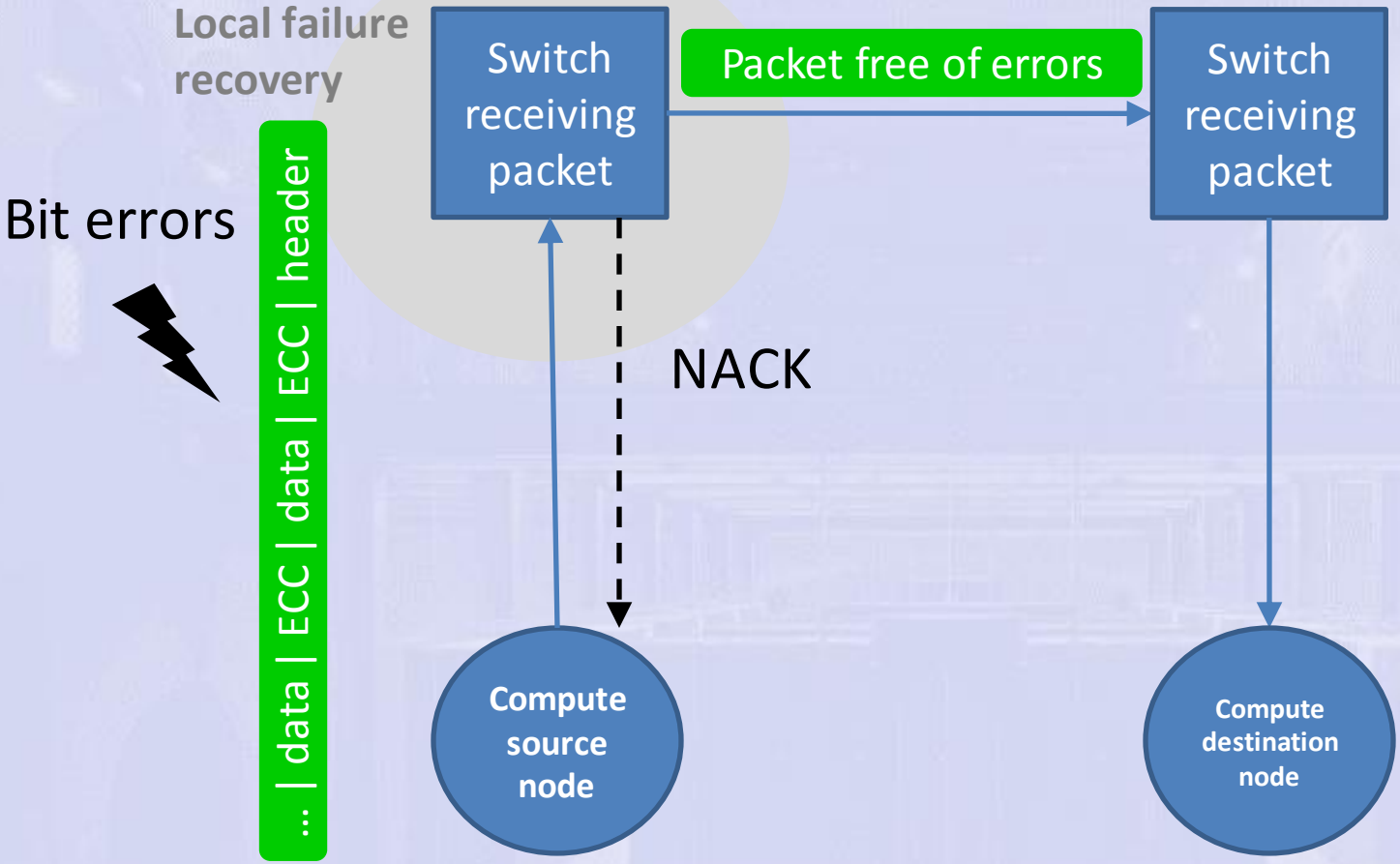


Subnet Manager (SMA)
- Re-program routing tables at S8

Local bit error recovery

- Use Error Correcting Codes (ECC)
- Populate packets with ECCs

Open issues
- Which ECC?
- Communication protocol?



Conclusions

- **Bit errors and hardware failures may happen more often in future computing systems**
  - **The number of network components is growing**
  - **Power wall increases BER**
- **Current networks are still relaying on end-node recovery which adds too much overhead**
- **Locally handling errors is promising but still requires efficient techniques to be explored which may be different depending of the type of failure**

Acknowledgements

- **Spanish Ministry of Science and Innovation**