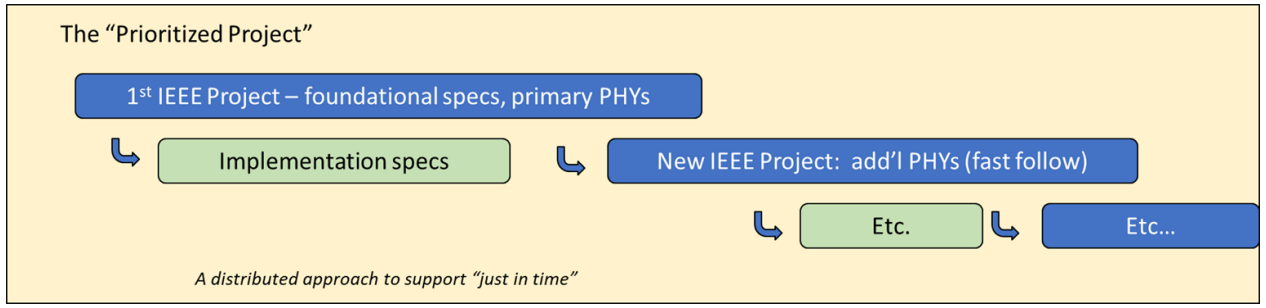
Initial Thoughts on E4AI PHYs and Interfaces, by Phase

Kent Lusted – Synopsys
Mark Nowell – Cisco

Goal Setting



Nowell_E4AI_01_250430

- Market expects a quick project
- A phased approach to provide the right PHYs and interfaces at the right time
  - Early requirements may come too late, later requirements may be too early
- Need crisp consensus on PHYs and interfaces in phase 1 (and good sense of phase 2)
- AI focused back-end networks are driving both initial requirements and some unique requirements ahead of front-end network needs

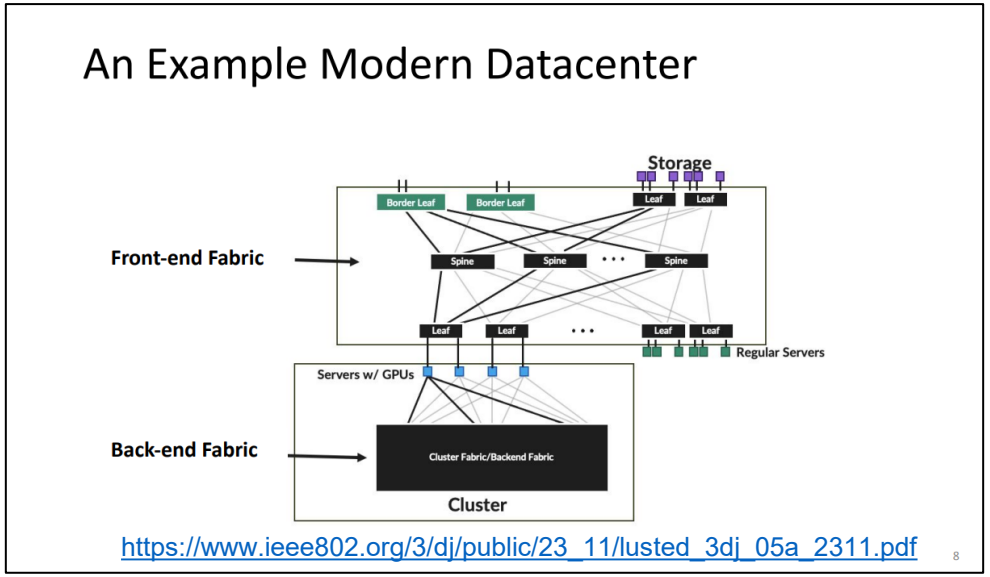
Disclaimer

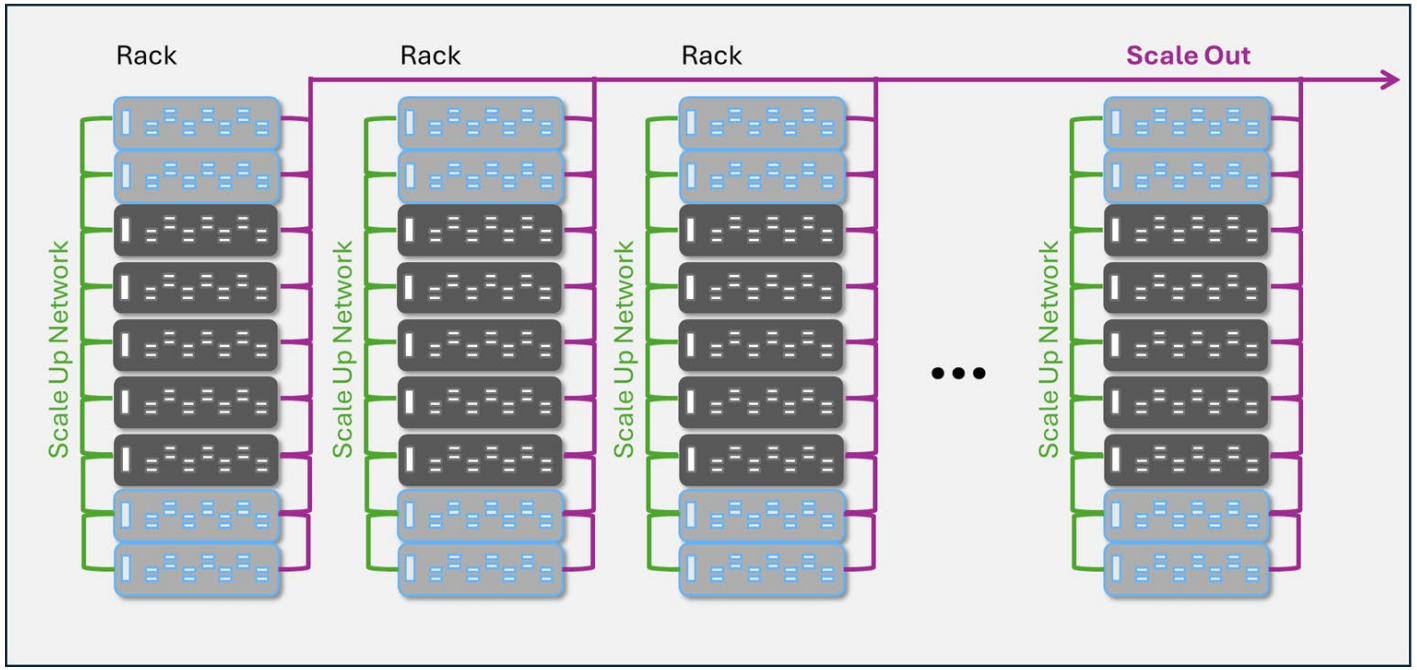
- This is an initial swag of a list by phase for the purpose of facilitating discussion
- This is **NOT** the final list. Feedback and discussion is strongly encouraged
- See first and second bullets

Nomenclature

For the purpose of this contribution, the following nomenclature is used

- **Rack** is a single server cabinet containing multiple AI servers
- **Pod** is a larger, modular unit that typically consists of one or multiple racks, sharing common resources like power and cooling
- **Scale up** = Back-end network primarily connecting devices inside of a rack unit
- **Scale out** = Back-end network primarily connecting racks within a pod





Themes that Kent and Mark hear

- Higher signaling rate is being driven first by scale-up
  - Next, higher signaling is being driven secondly by scale-out
  - Then front-end network
  - Maximize the commonality between the networks
- Stay with copper until you can’t
- CPO will have a role
- Power & “Reliability” are high on the “Care about” list

Phase 1 – “Right Now”

- Higher speed lane rate is top priority (~400 Gbps/lane)
- Radix is important for deployments: getting the highest data rate through the fewest number of lanes possible
  - x1, x2, x4 (x8?) lane widths
  - 400 GbE, 800 GbE, 1.6 TbE, (3.2 TbE?)
- AUI C2M and AUI C2C to pluggable front panel
  - Both optics and copper interconnect leads to support for pluggable modules
  - CPO and NPO implementation could yield support with “short-reach” AUI
  - CPC implementations will be common for electrical channels
- CR/KR interconnect
  - 1m DAC desired but <1m (0.75m?) may be acceptable
  - Alternatives to passive DAC?
- DR optics (maximize radix): 1 lambda per fiber and parallel fiber
  - What are the reach break points?

Phase 1 – things to think about

- FLR target when Link Layer Retry (LLR) is assumed
  - Mean time between Phy errors (MTBPE)
- Latency: find the best trade-off between error rate performance and latency
  - extended reaches can support greater FEC latency
  - “Type 1” End-end FEC
- Optics
  - Types?: Retimed, Linear (LPO) and/or half-linear (LRO)
  - Implementations to consider for optics (both pluggable and CPO).
- Should IEEE 802.3 do something differently to address active media types?
  - Active cables: AEC, ACC, AOC

Phase 2 – “Needed Soon”

- x8-lanes / 3.2 TbE (if not in Phase 1)
- Optics:
  - Longer reaches: Intra-building, inter-building
  - Segmented/Concatenated FECs for longer reaches
  - New SMF channels (Multi-core Fiber (MCF), Optical Circuit Switch (OCS) considerations)
  - Non-SMF?
- Active cables not in Phase 1

Phase 3 – “Need later”

- For further discussion
- MMF ?
- Coherent?

Wrap up

- A starting point for what could be included in a Phase 1 is proposed
  - Feedback welcome
- Key points
  - New lane rate needed, widths of 1,2,4 (and possibly 8)
  - C2M/C2C electrical, Short reach optics, cables
- Key incubation topics to dig into
  - Modulation, coding, FEC
  - Latency / FLR / LLR tradeoffs
  - Active cables