Potential Improvements to Strawman for PHY Baseline Proposal

Contribution to IEEE 802.3: 40GBASE-T PHY Baseline Proposal Ad Hoc
27 Feb 2014
Teleconference

George Zimmerman, Ph.D.
CME Consulting

Overview

- Motivation

- Recap of Strawman

- Motivations – Power & Cost

- Review of Potential Improvements

- Thoughts for Discussion

Motivation

- Strawman approach reuses technical and standards development from 10GBASE-T
  - Reduces risk and improves time-to-standard
- Does not preclude 'upgrading' possible areas
  - Lessons learned for robustness, additional power savings and ease of implementation
- This contribution expands on the suggestions in zimmerman_3bqah_02_1213.pdf, slide 6

Recap of Strawman

Baseline Proposal

- Baseline PHY proposal:
  - Use PCS, Framing and Line Coding from Clause 55
  - Increase symbol rate 4X to 3200 Mbaud
  - Drop transmit power to ~ 0 dBm at MDI
- Areas for improvement/consideration:
  - Backchannel for THP dynamic update?
  - Revised FEC to cover uncoded bits?
    - Multiple ways of doing this
  - Faster startup?
  - Negotiated patch-cord operational mode?
  - Remove PBO?

Motivations – power & cost

- Cost = ease of implementation, higher yields, lower risk
  - Reuse of 10GBASE-T technology, easing the channel go a long way towards this end
  - Interoperability is a complexity and time-to-market risk
  - Relaxations on MDI specifications could save component cost
  - Robustness to PCB layout variations could save system cost
- Power = a war of Milliwatts
  - No single magic bullets, lots of 5-10% pieces
  - Designer-specific: allocation of implementation loss
  - Standards-related: tolerance to defects
    - E.g., Impulse noise, NEXT and Echo

Remove PBO?

- WHAT IT IS: Remove power back off negotiation
- WHY WAS 10GBASE-T THE WAY IT WAS:
  - Power back off based on Insertion Loss was important to managing Alien Far-end crosstalk (AFEXT)
- WHY WE MIGHT RECONSIDER:
  - Shielded channel and reduced distance mitigate AFEXT concern
  - Ease of use: Startup time and interoperability would be aided without impacting adaptation process
- COSTS/POWER examined in contribution by Peter Wu, Wu_01a_0214_802.3bq_adhoc.pdf
  - 7-15% increase in short-link transceiver power if PBO is removed
  - Potential for simplifications in PBO without removal

Faster Startup

- WHAT IT IS: Decrease time allotted for startup

- WHY WAS 10GBASE-T THE WAY IT WAS:
  - 2-3 seconds was considered a tolerable "human time scale" for initial link
  - Training activity during startup drove peak power in early implementations

- WHY WE MIGHT RECONSIDER:
  - Faster baud rate & shorter channel may allow 2-4X startup improvement
  - Ease of use: experience has shown 10GBASE-T startup times are relatively long when testing reliability with multiple startups

- COSTS:
  - Small interoperability risk as slack for vendor-differentiation diminishes
  - Potential for small extra hardware cost to improve adaptation time

- POWER SAVINGS: NONE - Small INCREASE relative to strawman
  - 40G adaptation circuitry can be run at same rate or slower than 10GBASE-T if startup time is unchanged, not scaled 4X

Backchannel for THP update

- WHAT IT IS: Adapt the equalization on the fly based on the receiver's state
- WHY WAS 10GBASE-T THE WAY IT WAS: Variability during link up of 10GBASE-T MDI-to-MDI channel did not require it
- WHY WE MIGHT RECONSIDER:
  - Enables deep notches for narrowband interference, shown effective in 10G
  - Eliminates a need for "fast retrain" interrupting data flow
  - BUT – is this a problem with a shielded channel?
  - ON THE OTHER HAND - What about noise from the host PCB?
    - e.g., use in WAPs?
- COSTS:
  - Change in PCS framing to add back channel, small increase in bit rate
  - Small increase in complexity and risk in interoperability
- POWER SAVINGS: None significant, may cost minimal power for computing updates

FEC to cover uncoded bits

- WHAT IT IS: Change or add a coding layer to cover all bits in the PCS framing
- WHY WAS 10GBASE-T THE WAY IT WAS: Performance in stationary (non-impulsive) interference was slightly better with set-partitioning
- WHY WE MIGHT RECONSIDER:
  - Performance: Experience shows bit errors are often on uncoded bits
  - 40GBASE-T noise is likely host-electronic-noise dominated, which is often impulsive
  - Cost: Potential relaxation of MDI return loss & front end requirements
- COSTS:
  - Change in PCS framing to accommodate either code layer or coding change, possibly with a small increase in line rate
  - Minimal but nonzero increase in complexity or risk in interoperability
- POWER SAVINGS:
  - Reduction in AFE clip levels by x dB could save $100*(1-2^{-x/6.02})$% of AFE receiver power, e.g., 3dB = 29% savings in AFE RX power

Negotiated Patch Cord Operational Mode: "Direct Attach Mode"

- WHAT IT IS: Include Negotiation of link-length in startup
- WHY WAS 10GBASE-T THE WAY IT WAS:
  - "10GBASE-T Short reach test mode"
  - Confusion over possible multiple PHY types ('10GBASE-TSR vs 10GBASE-T?)
  - Minimal power savings vs. Single-ended determination
  - Switch-to-server market was in early stages of segmentation
- WHY WE MIGHT RECONSIDER:
  - Switch-to-server market has segmented much more
  - Differentiated switch and server solutions for within-rack connections?
  - Power and port-counts in within-rack applications are more critical
- COSTS: Potential market confusion of 30m vs. within-rack 40GBASE-T
- POWER SAVINGS: Vary substantially by vendor architecture
  - From 10% to 50% relative to power at 30 meters
    - Much overlap with existing power savings approaches in multiple vendors
  - Engineered architectures (e.g., ToR) should be able to realize the power benefit without standards changes
  - Most savings for this are in receiver signal processing – only the PBO savings needs communication, unless an alternate line encoding is envisioned (is this even in our scope?)

Thoughts for Discussion

- Remove PBO? – PROBABLY NOT
    - Saves power, probably don't remove, consider simplification
    - CAN BE DONE LATER, BUT CONSIDER PROPOSALS
- Faster Startup – PROBABLY NOT
    - Costs small power, don't consider unless a driver emerges
- Backchannel for THP update – PROBABLY NOT
    - Unlikely to have a need, depends on how confident we are in channel shielding & host noise (e.g., what about use in WAPs?)
- FEC for Uncoded Bits - MAYBE
    - SOLICIT CONTRIBUTION & PROPOSALS, potential for power savings, MDI RL relaxation, ease of host PCB layout
- Negotiated Patch Cord Operational Mode - MAYBE
    - Need to determine what benefits can't be achieved without making this formal – otherwise realizing benefits of power savings on direct-attach links is more of a marketing problem

THANK YOU!