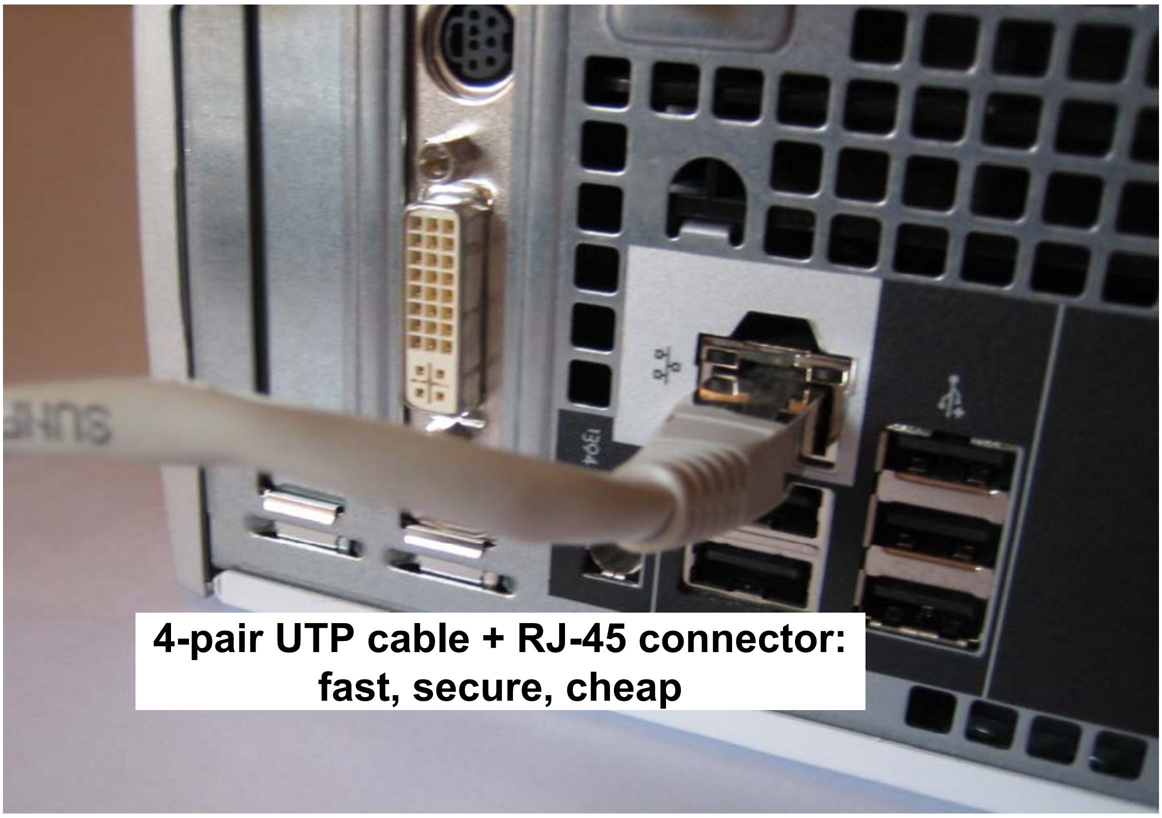
10GBASE-T:
10Gbit/s Ethernet over copper

NEWCOM-ACoRN Joint Workshop
www.newcom-acorn.org
Vienna, 20-22 September 2006

Gottfried Ungerboeck

Contents

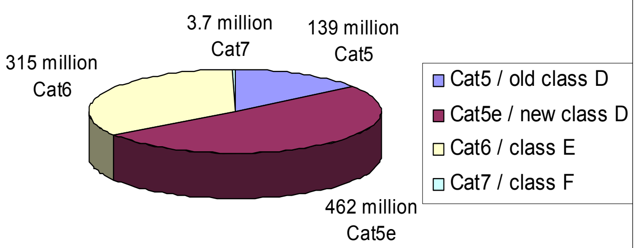
Ethernet over UTP copper is ubiquitous



Ethernet over UTP copper is ubiquitous

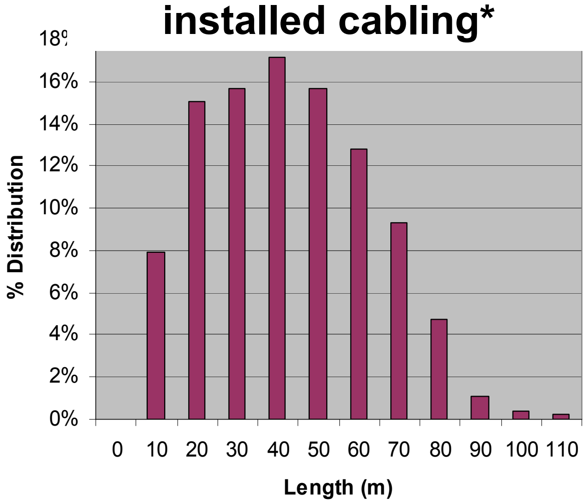
Total Ethernet ports (switch + client) shipped through 2004:
>> 2 Billion
(BRCM estimate)

Estimated WW installed base of copper links



(BSRIA)

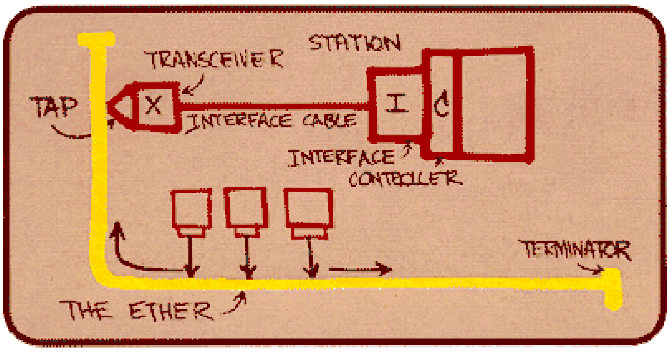
Length distribution of installed cabling*



Sources: Hubbell, Seimon Co., Nordx/CDT, Cabling Partnership, & Fluke; 120K links surveyed

* shorter distances for data centers only

Evolution of Ethernet



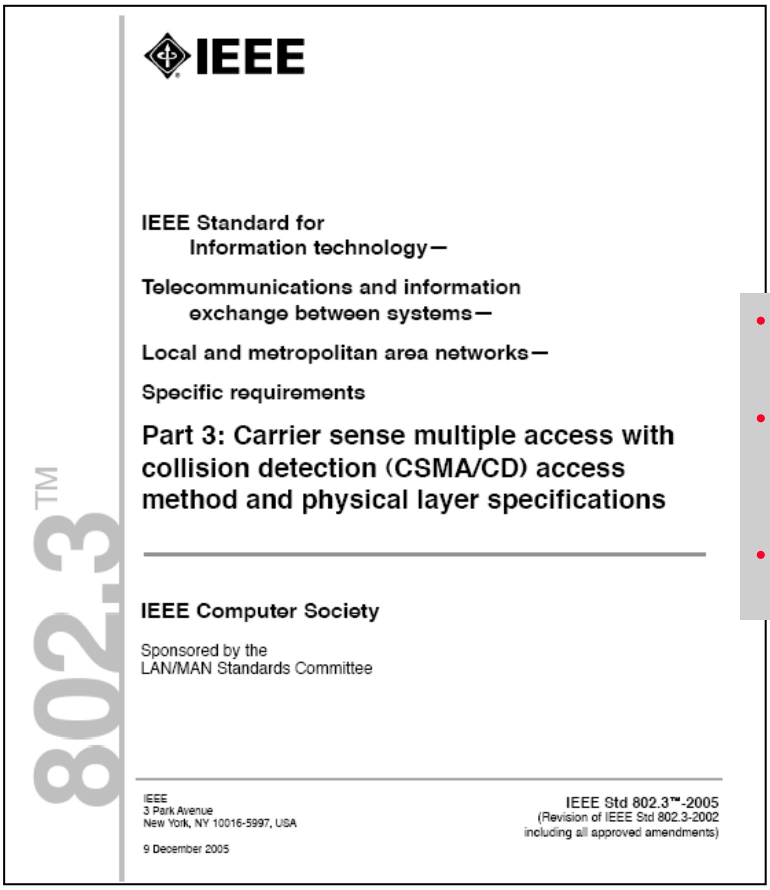
Ethernet drawing by Bob Metcalfe, around 1976

In the meanwhile, the ETHER bus has evolved into a topology of connected stars, in which stations are attached to the network nodes via point-to-point links.

Repeaters are replaced by switches. The Carrier-Sense Multiple Access with Collision Detection (CSMA/CD) protocol no longer plays a critical role.

The Ethernet Frame Format has been retained.

IEEE 802.3 (Ethernet) Standard



IEEE

IEEE Standard for
Information technology—

Telecommunications and information
exchange between systems—

Local and metropolitan area networks—

Specific requirements

Part 3: Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications

IEEE Computer Society

Sponsored by the
LAN/MAN Standards Committee

IEEE
3 Park Avenue
New York, NY 10016-5997, USA

9 December 2005

IEEE Std 802.3™-2005
(Revision of IEEE Std 802.3-2002
including all approved amendments)

802.3™

- **Latest consolidated version of 9 Dec 2005**
- **Comprises 67 clauses, 2696 pages (Clause 55 reserved for 10GBASE-T)**
- **10GBASE-T approved on 21 July 2006.**

Ethernet Physical Layers (PHYs) for Copper

1 Mbit/s: 1BASE5 (coax)

Deployed in huge quantities

10 Mbit/s: 10BASE5, 10BASE2, 10BROAD36 (coax)

10BASE-T (2-pair UTP-3, 1 x 10 Mbit/s, HDX, 1991)

100 Mbit/s: 100BASE-T4 (4 pair UTP-3, 3 x 33 Mbit/s, HDX)

100BASE-TX (2-pair UTP-5, 1 x100 Mbit/s, FDX, 1995)

100BASE-T2 (2 pair UTP-3, 2 x 50 Mbit/s, dual DX, 1997)

1Gbit/s: **1000BASE-T (4-pair UTP-5, 4 x 250 Mbits, quad DX, 1999)**

UTP-3: unshielded twisted pair – category 3 (voice grade)
UTP-5: unshielded twisted pair – category 5 (data grade)

10Gbit/s: 10GBASE-T (4-pair UTP- 6 or better, 4 x 2.5 Gbit/s, quad DX, 2006)

Ethernet Copper PHY Progression

- **As data rates increase, copper PHYs must become increasingly more sophisticated to operate over UTP cabling**

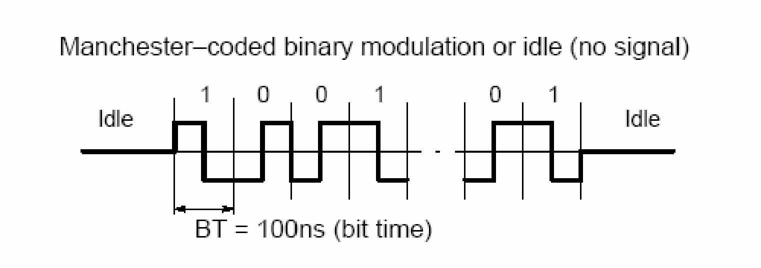
10 Mbps →	100 Mbps →	1000 Mbps →	10 Gbps
2-pair HDX	2-pair FDX	4-pair quad DX	4-pair quad DX
Manchester	Scrambling	Scrambling	Scrambling
	MLT-3	Echo & NEXT canc.	Echo & NEXT canc.
	Equalization	PAM-5	128-DSQ
		TCM (8st 4D)	LDPC + CRC
		Parallel DFE	TH precoding
			Matrix FFE
			PCS frames
			TX power control
			Complex training procedure

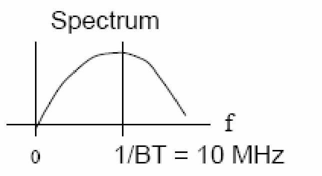
2-pair HDX/FDX ⇄

4-pair quad DX ⇄⇄⇄⇄

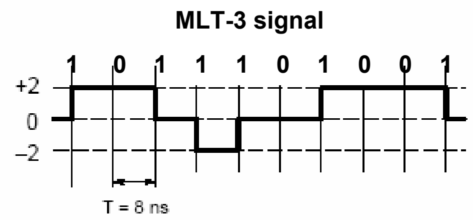
10BASE-T and 100BASE-TX modulations

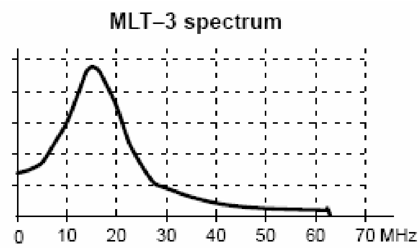
10BASE-T: 10 Mbit/s over 2-pair UTP-3 (voice grade, 1991)





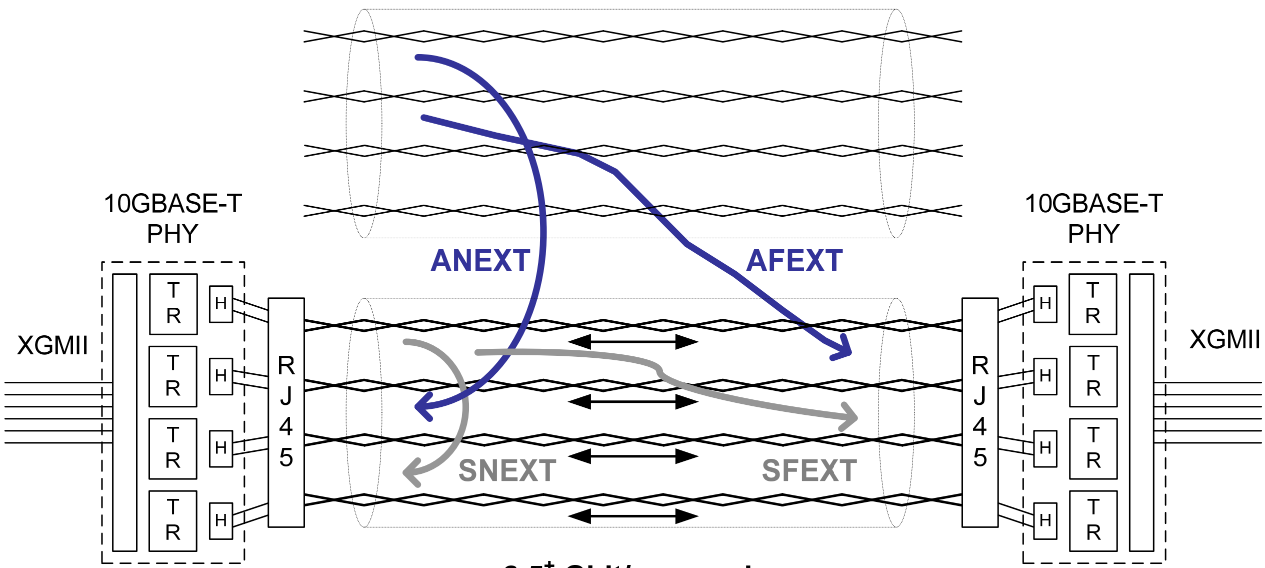
100BASE-TX: 100 Mbit/s over 2-pair UTP-5 (data grade, 1995)





10GBASE-T: 4-pair quad DX

Like 1000BASE-T, but 10 times faster and more sophisticated ...



2.5⁺ Gbit/s per pair

Echo and self crosstalk can be cancelled,
alien crosstalk acts like noise.

Link characteristics and achievable rate

Connecting everything™

BROADCOM®

Link segment characteristics specified for 10GBASE-T

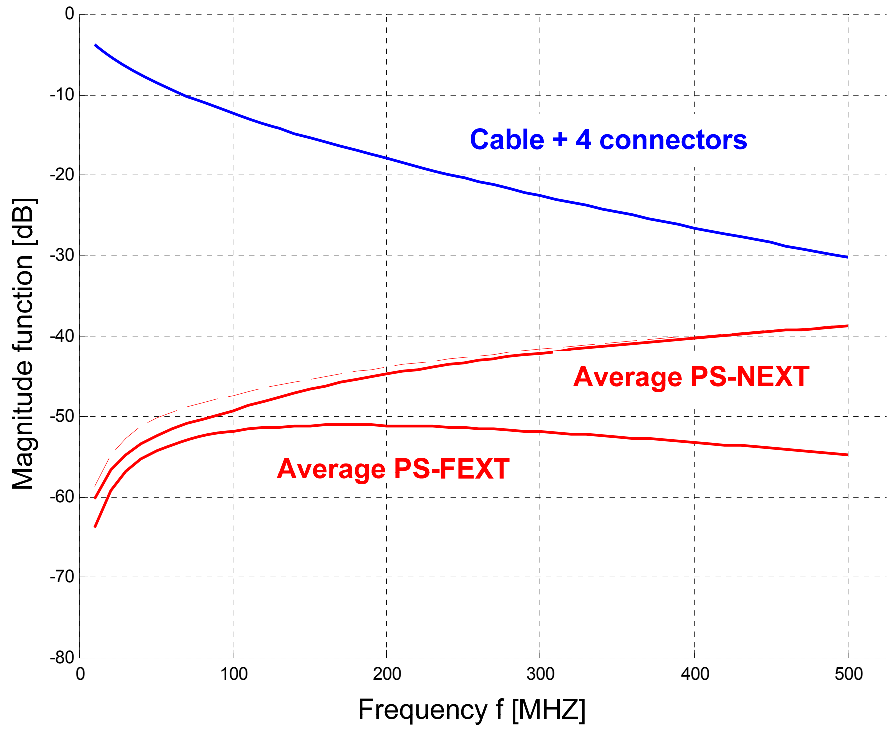
- **"The cabling system used to support 10GBASE-T requires ... ISO/IEC 11801 Class E or Class F 4-pair balanced cabling with a nominal impedance of 100 Ω"** **i.e., cabling better than Cat 5**
- **Link segment characteristics include the effects of work area & equipment cables and connectors**
- **Cabling types and distances:**
  - **Class E* / Category 6***: unscreened** **≤ 55 m**
  - **Class E* / Category 6***: screened** **≤ 100 m**
  - **Class F* and Class E_A**/ Augmented Category 6 ****** **≤ 100 m**

* ISO/IEC TR-24750, ** ISO/IEC 11801 Ed 2.1, *** TIA/EIA TSB-155, **** TIA/EIA-568-B.2-10

Connecting everything™

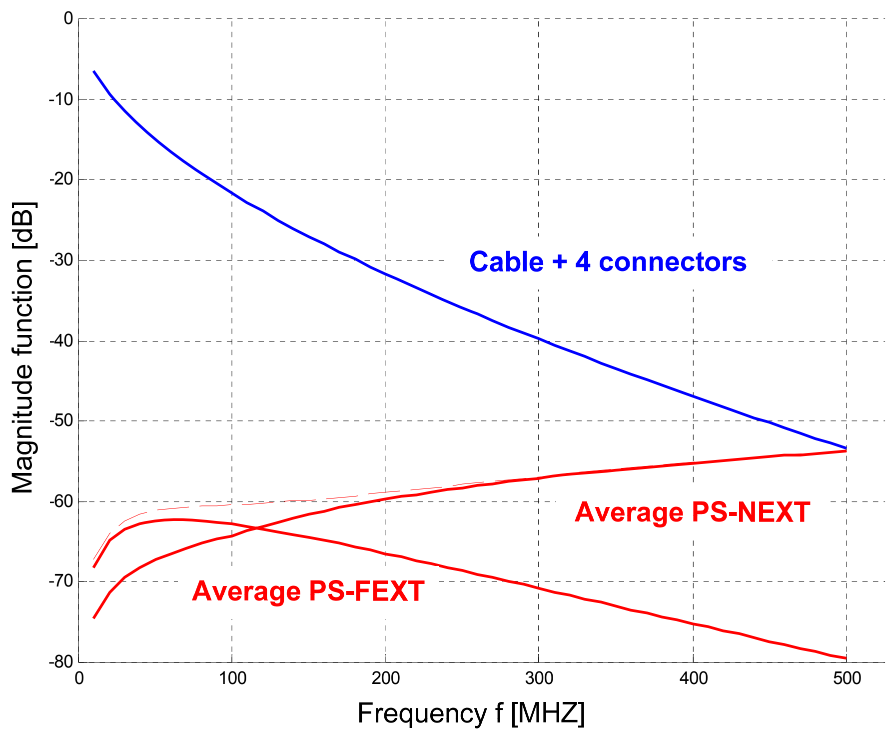
BROADCOM®

Class E / Category 6: unscreened, 55 m



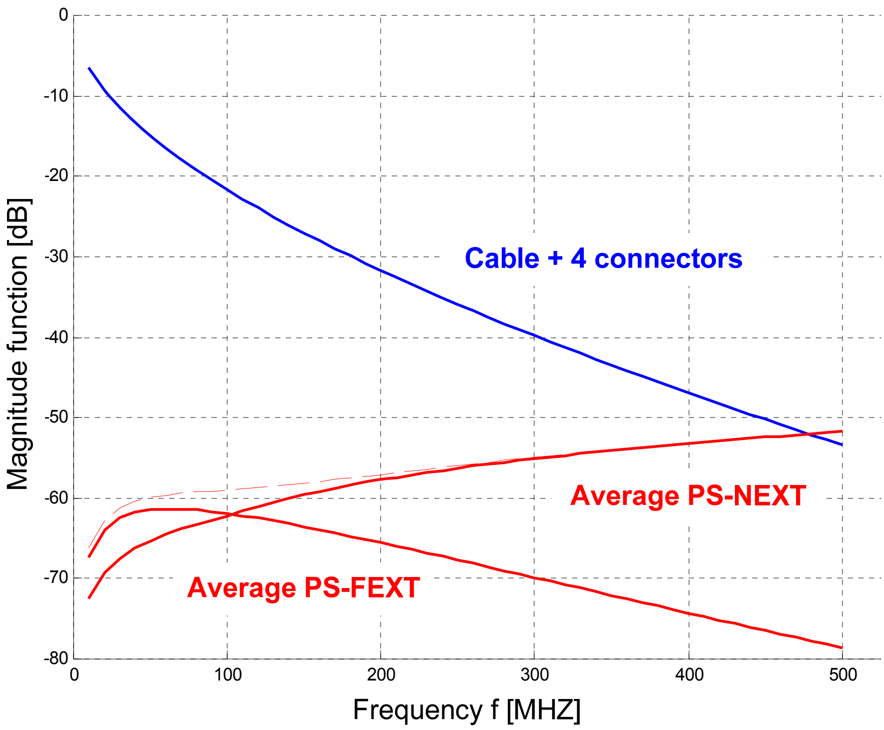
Average PS-A N/FEXT: power sum coupling from alien link pairs to one link pair on average over 4 link pairs

Class E / Category 6: screened, 100 m



Average PS-A N/FEXT: power sum coupling from alien link pairs to one link pair on average over 4 link pairs

Classes F and E_A / Augmented Category 6, 100 m

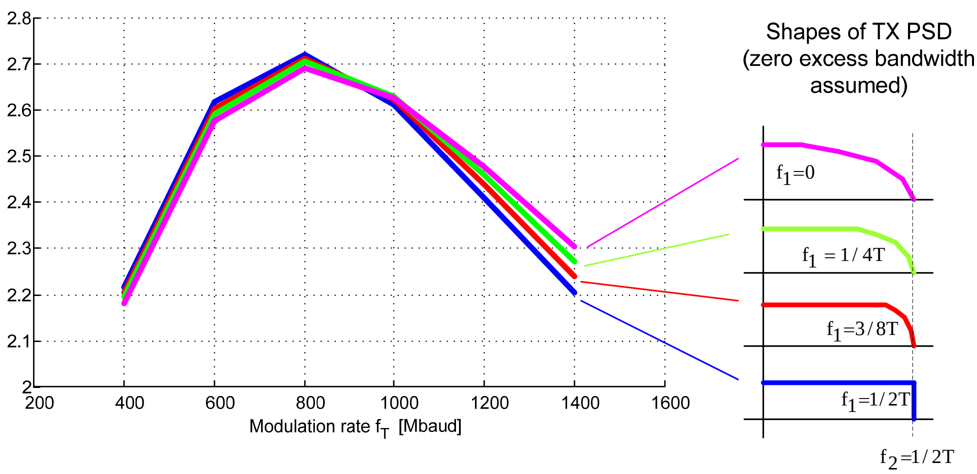


Average PS-A N/FEXT: power sum coupling from alien link pairs to one link pair on average over 4 link pairs

Achievable bit rate vs modulation rate

- **Class E / Category 6: screened, 100m**
- **Transmit power P_T = 5 dBm, background noise N_0 = -140 dBm/Hz**
- **ANEXT from same kind transmission, AFEXT ignored**

Achievable rate [Gbit/s] with 6 dB gap to capacity



This motivated the choice of 800 Mbaud.
800 Mbaud x 3.125 bit/dim x 4 pairs = 10 Gbit/s

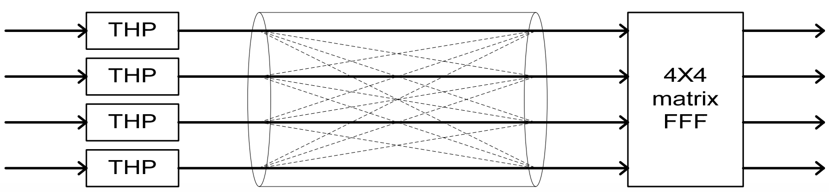
Modulation and equalization

Connecting everything™

BROADCOM

10GBASE-T modulation and equalization

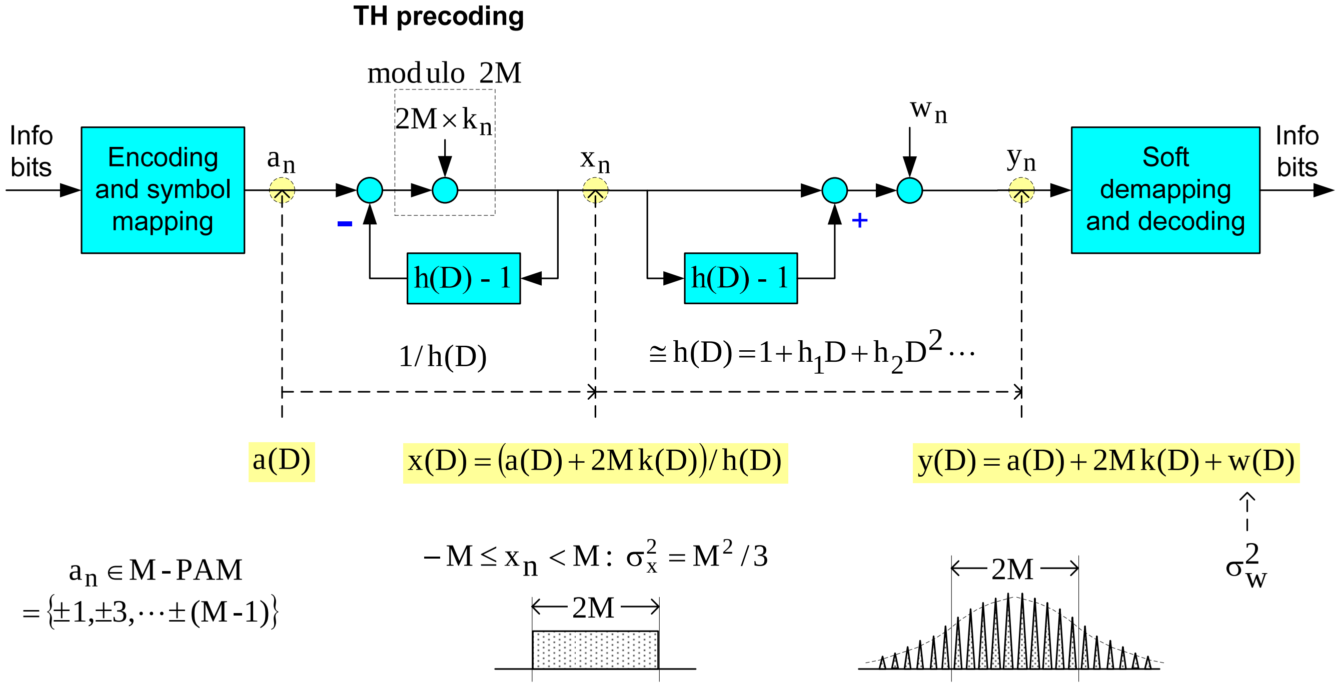
- **At 800 Mbaud, a 4-pair UTP cable acts like a MIMO–ISI channel.**
- **MIMO-OFDM cannot be used because 10GBASE-T transceiver latency is required to be $\leq$ 2.56 μsec**
- **Hence the following choice:**
  - **800 Mbaud baseband transmission using 16-PAM: 4 bit/dim, reduced to 3.125 bit/dim by 2-D alphabet partitioning and coding.**
  - **Link training: decision-feedback receiver structure with adaptive matrix feedforward filter (4x4 FFF) and scalar feedback filters (4 FBFs).** Matrix FBF not needed because cable transfer function is strongly diagonal-dominated.
  - **Data mode: Feedback filters are swapped to transmitter. Tomlinson-Harashima precoding in transmitter. Matrix FFF in receiver.**



Connecting everything™

BROADCOM

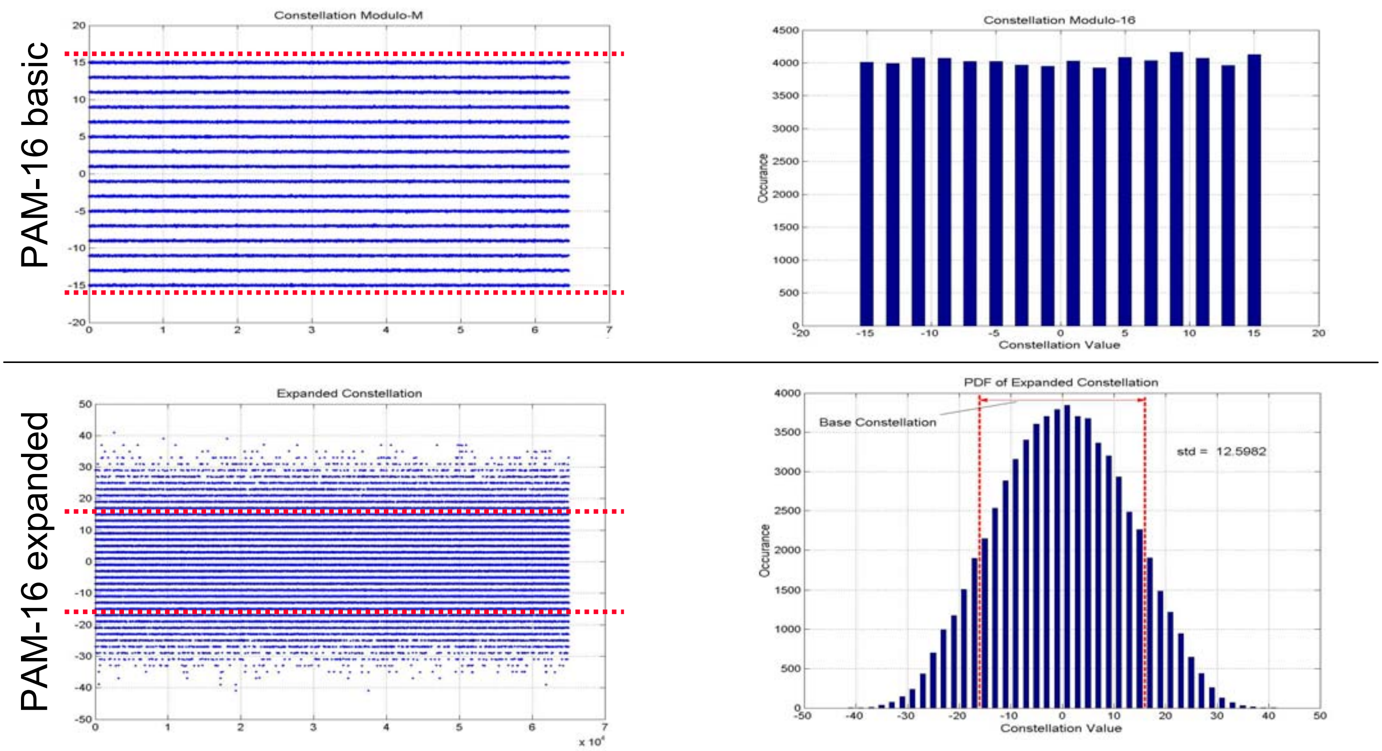
Tomlinson-Harashima (TH) precoding



$$\text{Signal-to-noise ratio } SNR = \sigma_x^2/\sigma_w^2$$

TH precoding: symbol distribution

Symbol values vs. time | Symbol distribution



Coding and framing

Towards LDPC-coded 128-DSQ modulation

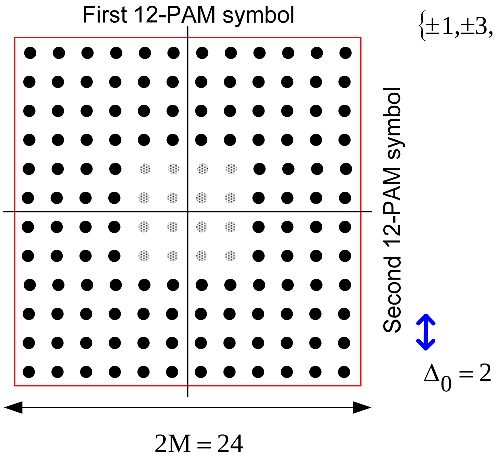
- **First proposal for 10GBASE-T: interleaved RS coding concatenated with 4-D 16-state TCM. Decoding complexity was low. Performance was OK with sufficient interleaving. Possibility of iterative decoding was considered.**
- **10GBASE-T task force considered latency caused by RS byte interleaving/deinterleaving unacceptable.**
- **The majority favored short-block LDPC coded modulation, initially with a 2-D 128-point "doughnut" constellation (3.5 bit/dim).**
- **Finally *(2048,1723) LDPC* coded 128-DSQ* was adopted.**

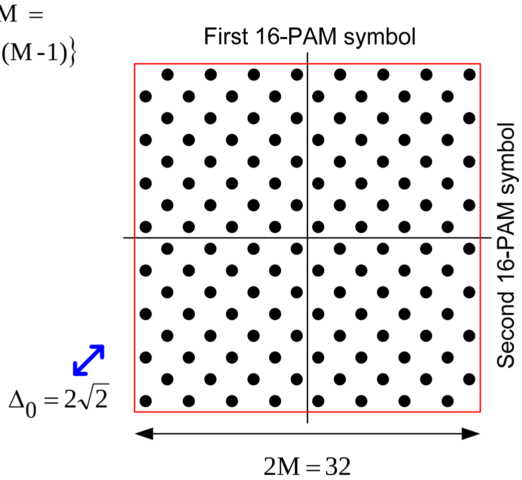
* H matrix construction is based on Generalized RS(32,2,31) code over GF(2^6)
(similar to Djurdjevic et al., "A class of low-density parity-check codes constructed based on Reed-Solomon codes with two information symbols," IEEE Commun. Letters, vol. 7, pp. 317-319, July 2003).

128-point 2-D constellations (3.5 bit/dim)

12-PAM^2 with hole (= doughnut) | **128-DSQ (Double SQuare)**

$$\text{M-PAM} = \{\pm 1, \pm 3, \cdots \pm (M-1)\}$$





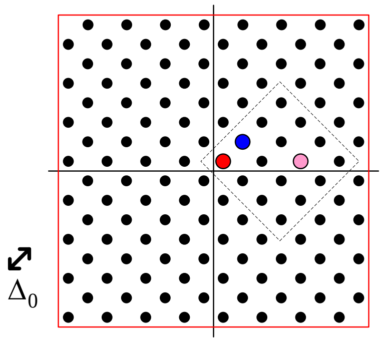
Signal energy per dimension at precoder output $E_x = (2M)^2/12$

$E_x / \Delta_0^2 = 48/4 = 12$

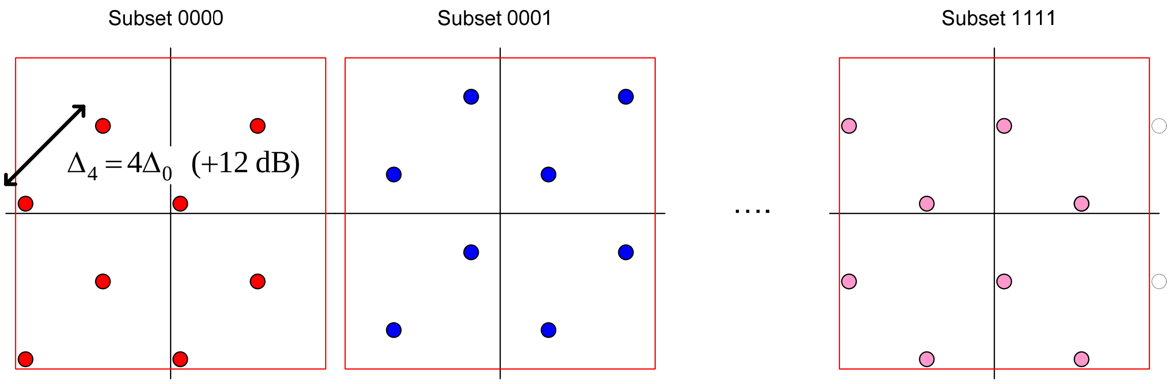
$E_x / \Delta_0^2 = (256/3)/8 = 10.666$

-0.5115 dB; ideal for THP

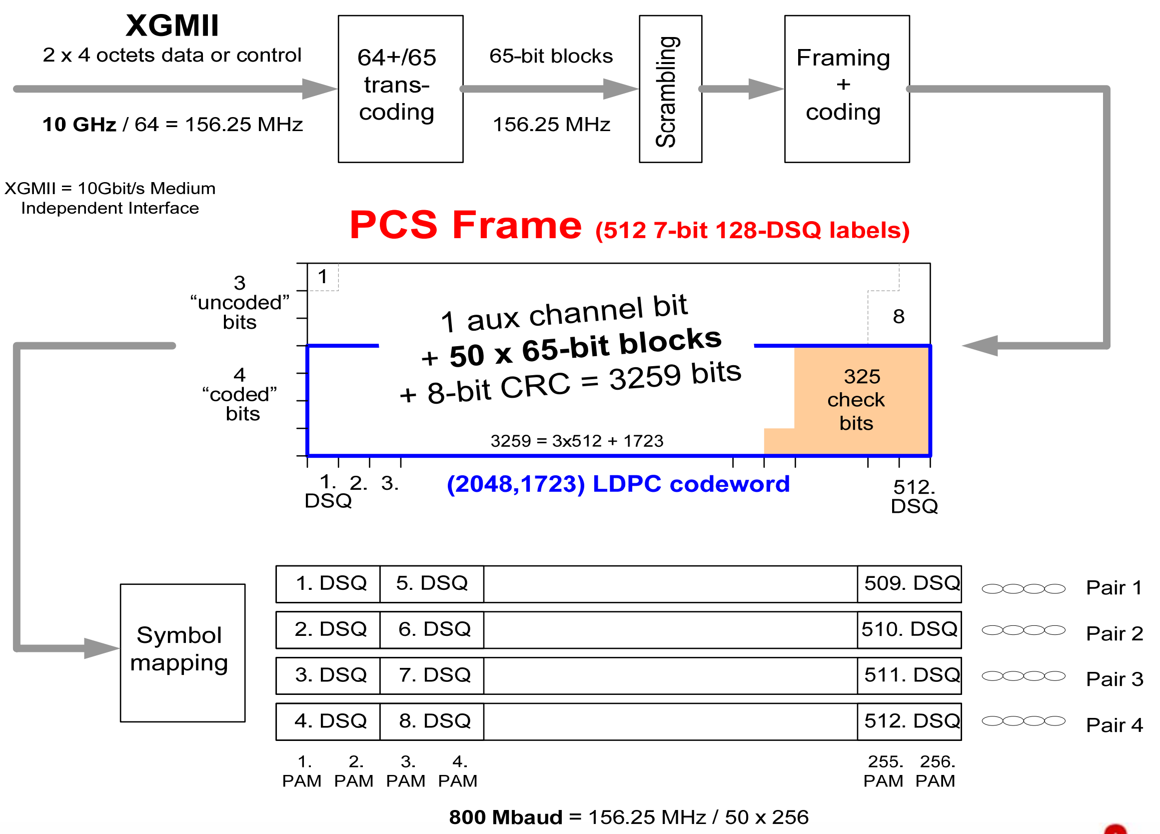
128-DSQ partitioning into 16 subsets ("12 dB" partitioning)



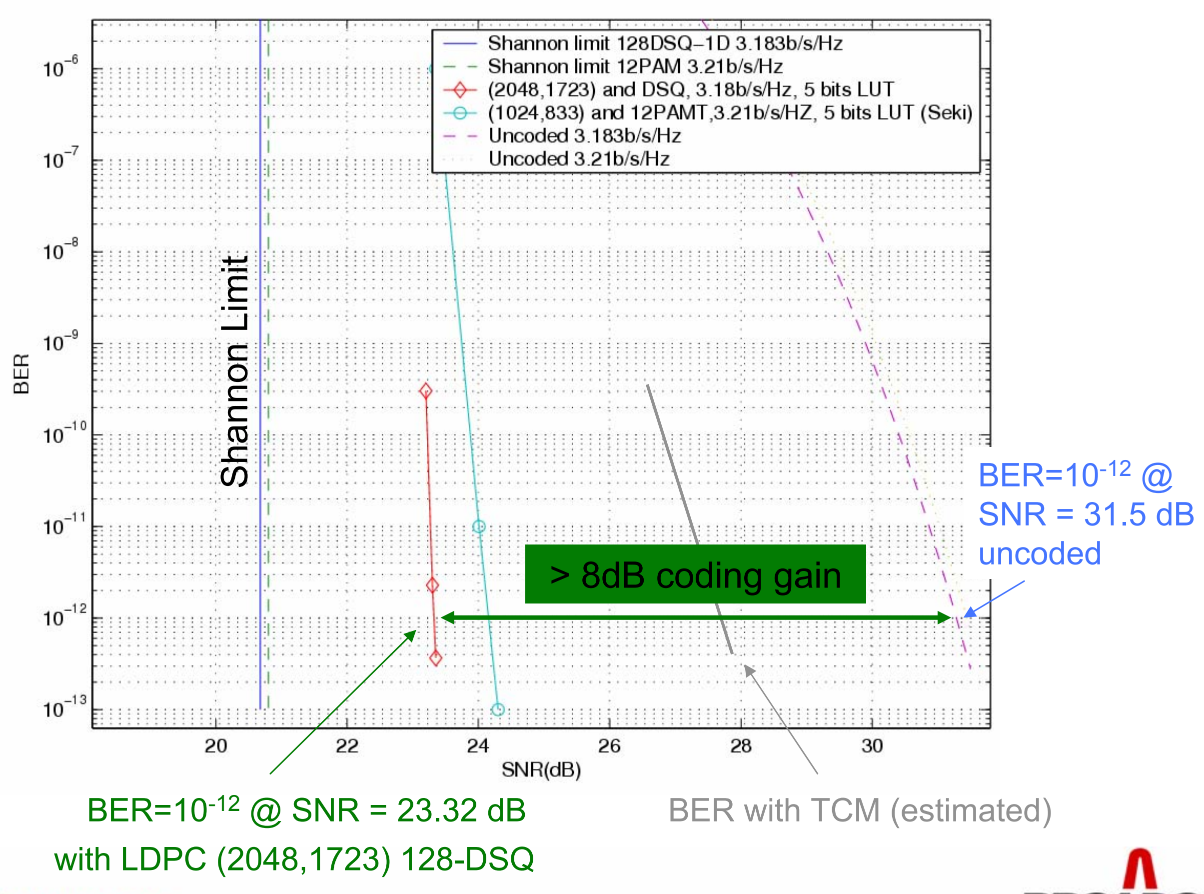
128-DSQ symbol selection
- 4 coded bits select subset
- 3 uncoded bits select symbol within subset



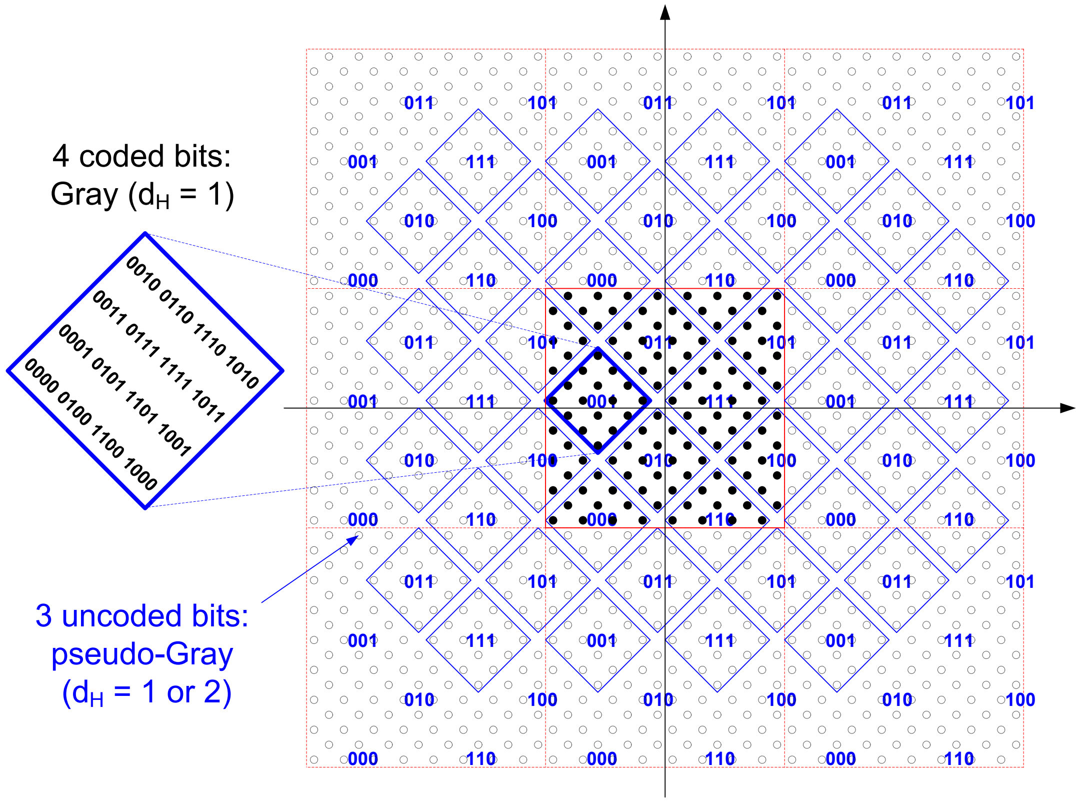
10GBASE-T coding, framing, symbol mapping



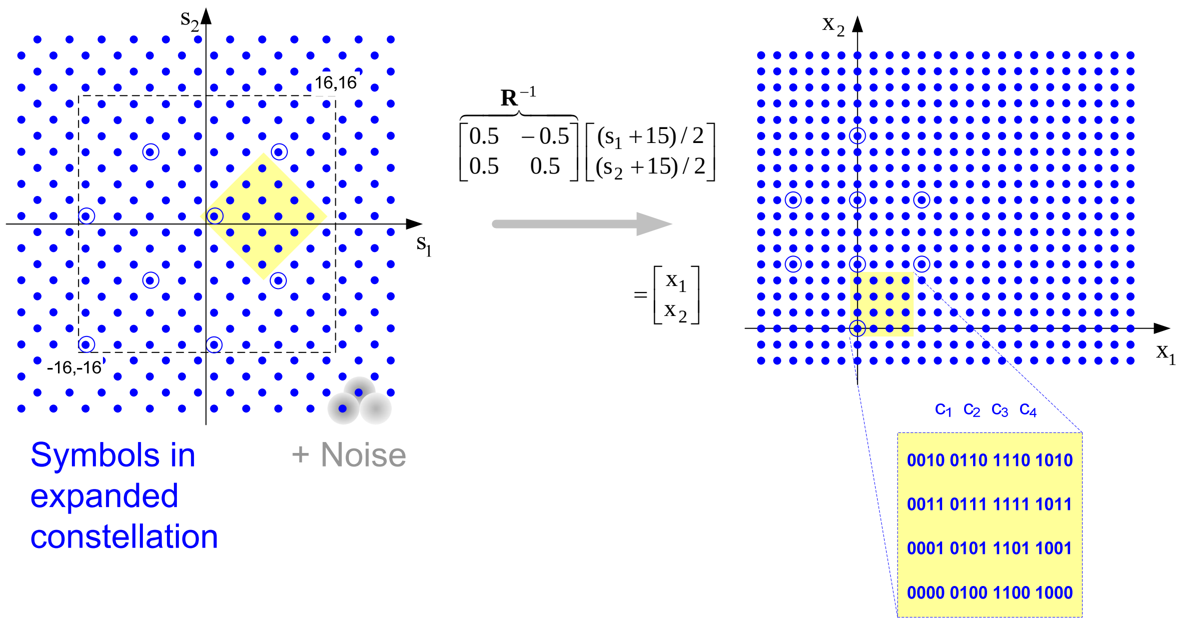
Error performance



128-DSQ constellation, modulo-32 extended



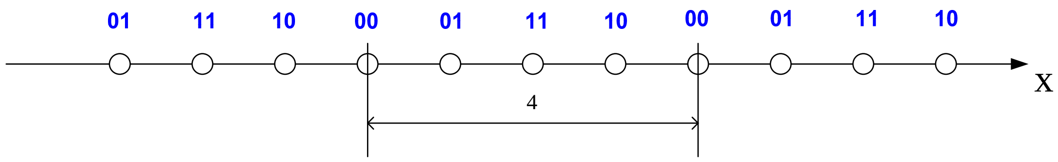
Metric calculation for 4 coded bits



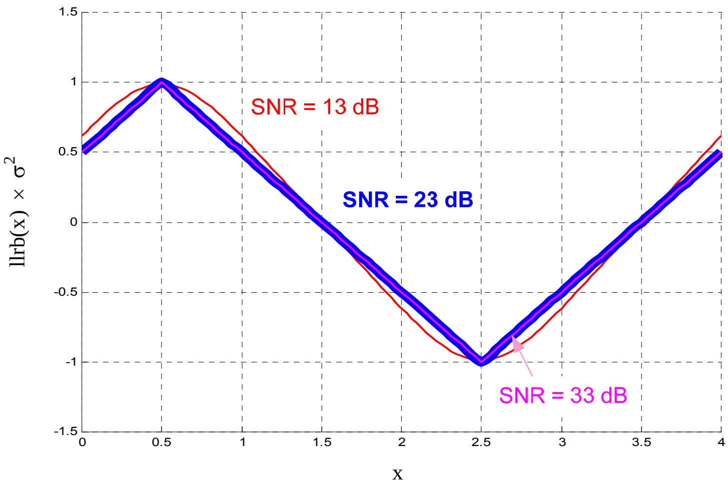
$$\log \frac{\Pr(c_1 = 0/x_1)}{\Pr(c_1 = 1/x_1)} = \text{llrb}(x_1 \mod 4) \qquad \log \frac{\Pr(c_2 = 0/x_1)}{\Pr(c_2 = 1/x_1)} = \text{llrb}(x_1 + 1 \mod 4)$$

$$\log \frac{\Pr(c_3 = 0/x_2)}{\Pr(c_3 = 1/x_2)} = \text{llrb}(x_2 \mod 4) \qquad \log \frac{\Pr(c_4 = 0/x_2)}{\Pr(c_4 = 1/x_2)} = \text{llrb}(x_2 + 1 \mod 4)$$

The function llrb(x)



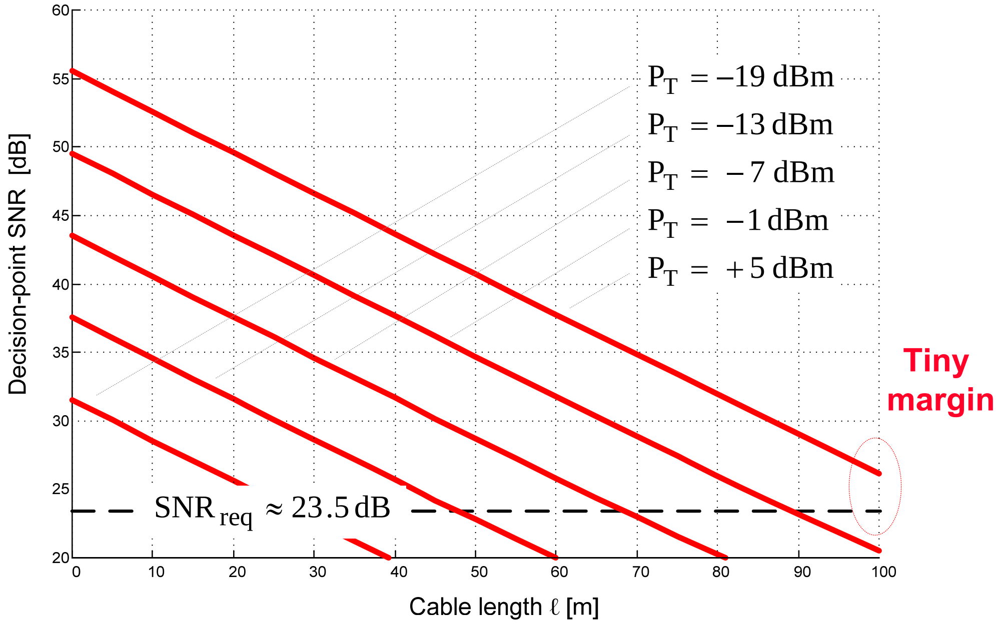
$$\mathrm{llrb}(x) = \ln \frac{\sum_{k=-\infty}^{\infty} \exp\left(-[x-(4k+0)]^2/2\sigma^2\right) + \exp\left(-[x-(4k+1)]^2/2\sigma^2\right)}{\sum_{k=-\infty}^{\infty} \exp\left(-[x-(4k+2)]^2/2\sigma^2\right) + \exp\left(-[x-(4k+3)]^2/2\sigma^2\right)} \cong \frac{1}{\sigma^2} \begin{cases} x+0.5 & : \; 0 \le x \le 0.5 \\ 1.5-x & : \; 0.5 \le x \le 2.5 \\ x-3.5 & : \; 2.5 \le x \le 4 \end{cases}$$



Decision-point SNR versus cable length

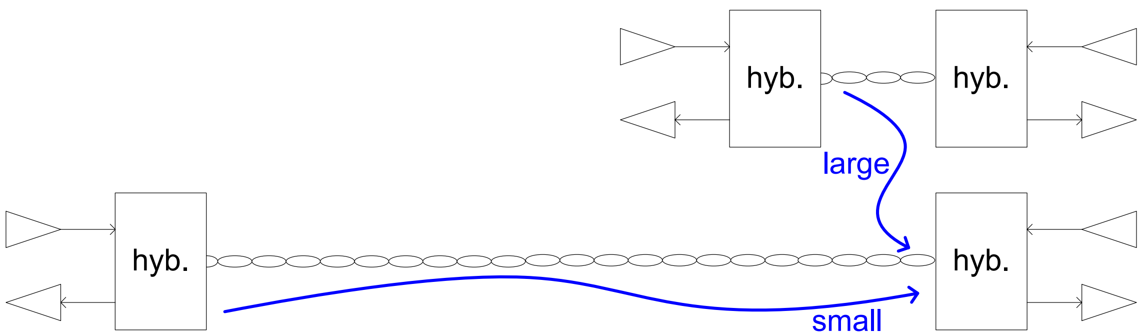
Decision-point SNR vs. cable length

Cable Class E screened; worst-case ANEXT from adjacent links transmitting at P_T = 5 dBm; AWGN = -140 dBm/Hz; ideal precoding response, ideal receiver filter and FFF equalization



Near-far problem for 10GBASE-T

- **Worst case alien crosstalk configuration: short cable bundled at the end with long cable(s)**



- **TX power of short link can be reduced without impacting short-link BER, and should be reduced to improve long-link BER.**

Decision-point SNR vs cable length: conclusions

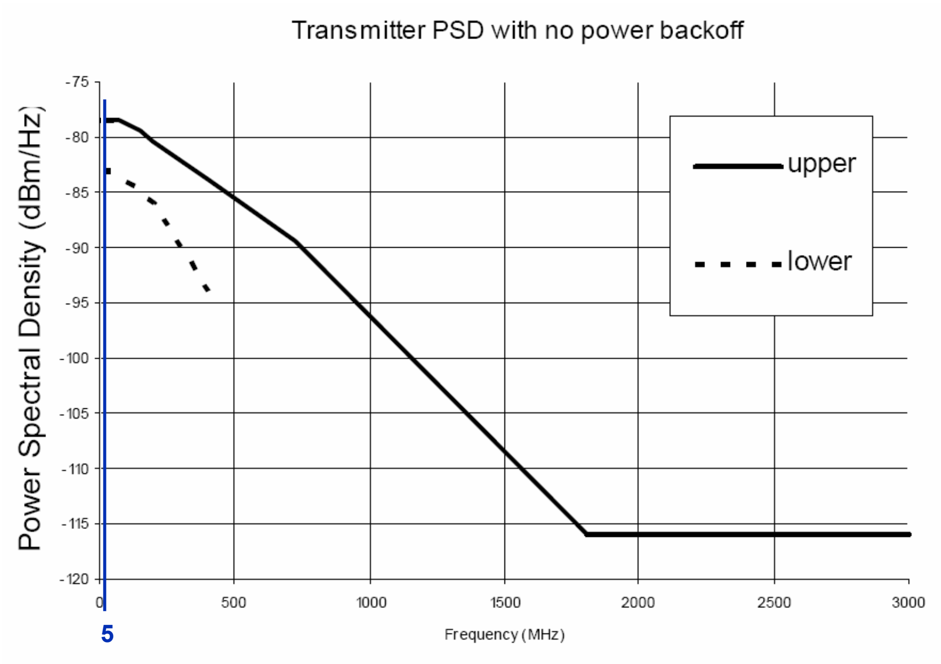
- SNR margin for 100m Class E screened cable is tiny even for ideal transceiver realization
- 10GBASE-T requires TX power control

Transmit power options adopted for 10GBASE-T

- Maximum power P_{max} = 3.2 – 5.2 dBm
- Power backoff levels: 0, -2, -4, -6, -8, -10, -12, -14 dB
- Power backoff must be used for shorter cables

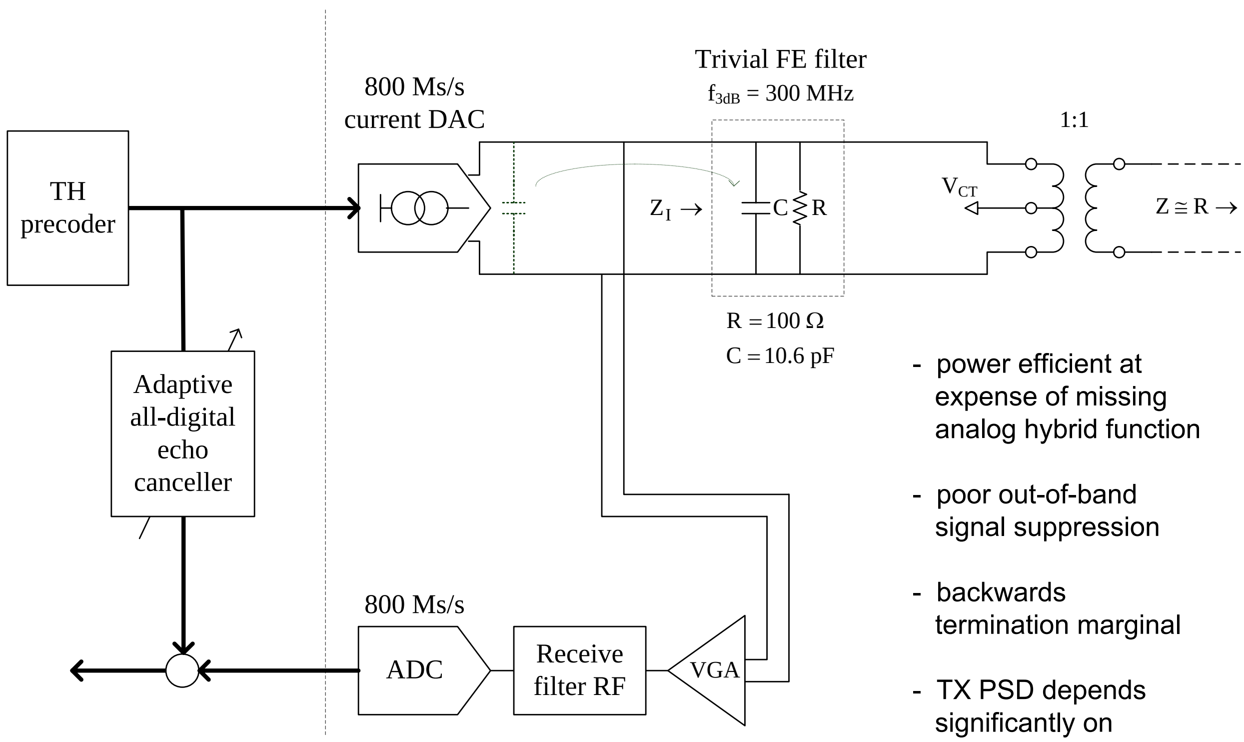
Transceiver front end and echo cancellation

10GBASE-T transmit PSD specification



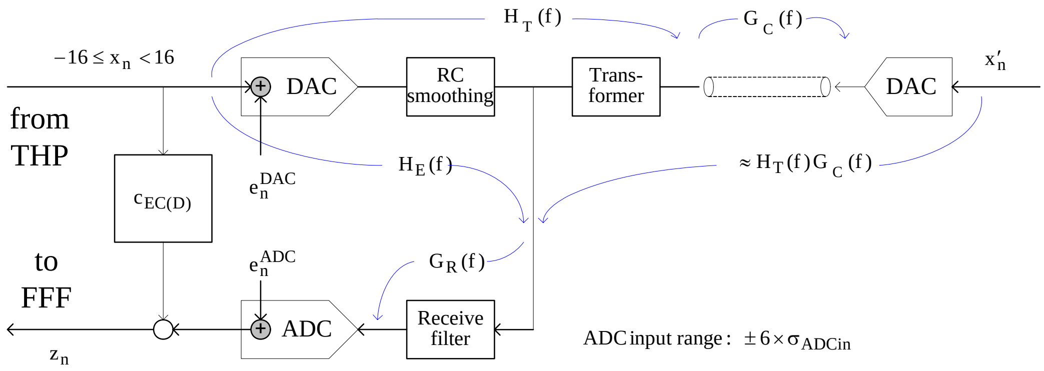
The 10GBASE-T Task Force could not agree on a tighter specification. This loose specification allows for a wide range of PSD shapes except one with a wider spectral notch around dc!

Transmitter front-end: "simple"



- power efficient at expense of missing analog hybrid function
- poor out-of-band signal suppression
- backwards termination marginal
- TX PSD depends significantly on transformer

DAC & ADC precision requirements



<u>Decision-point SNR$_{mmse}$ [dB] with DAC and ADC errors only</u> (no noise, no alien Xtalk)

50 m Class E

enob(ADC) \ enob(DAC)	8	9	10
8	<25	<25	<25
9	<25	<28	<28
10	28.17	30.25	31.05

100 m Class E

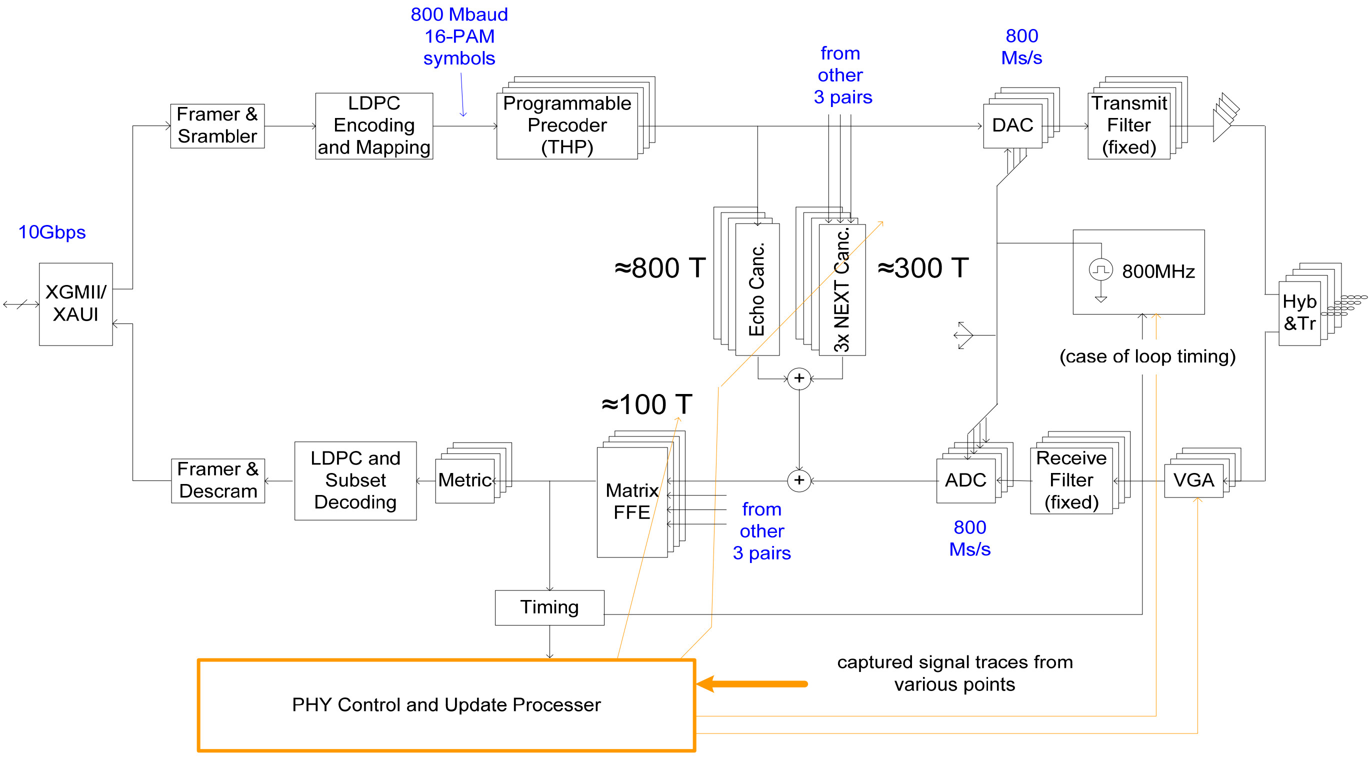
enob(ADC) \ enob(DAC)	9	10	11
11	<25	<25	<25
12	<25	<28	28.36
13	<25	28.18	31.89

100 m: "simple" front-end requires unrealistic DAC & ADC precision, <u>other approach is needed</u>.

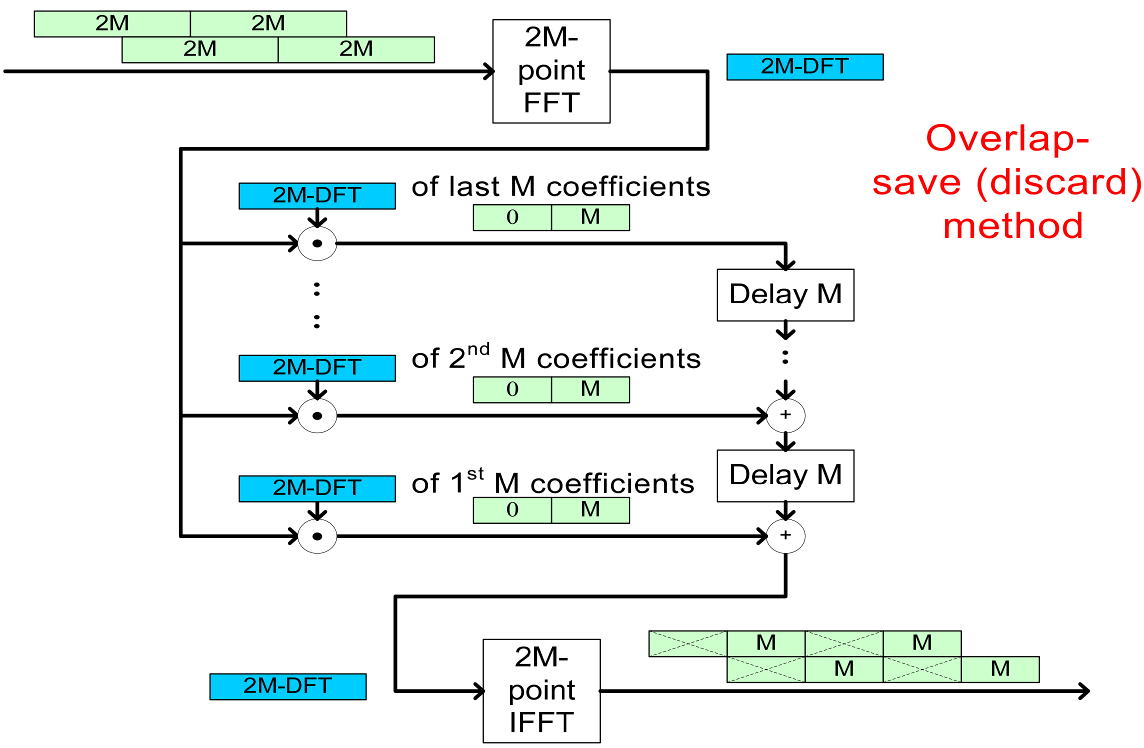
enob = equivalent number of bits

Transceiver realization

Transceiver block diagram



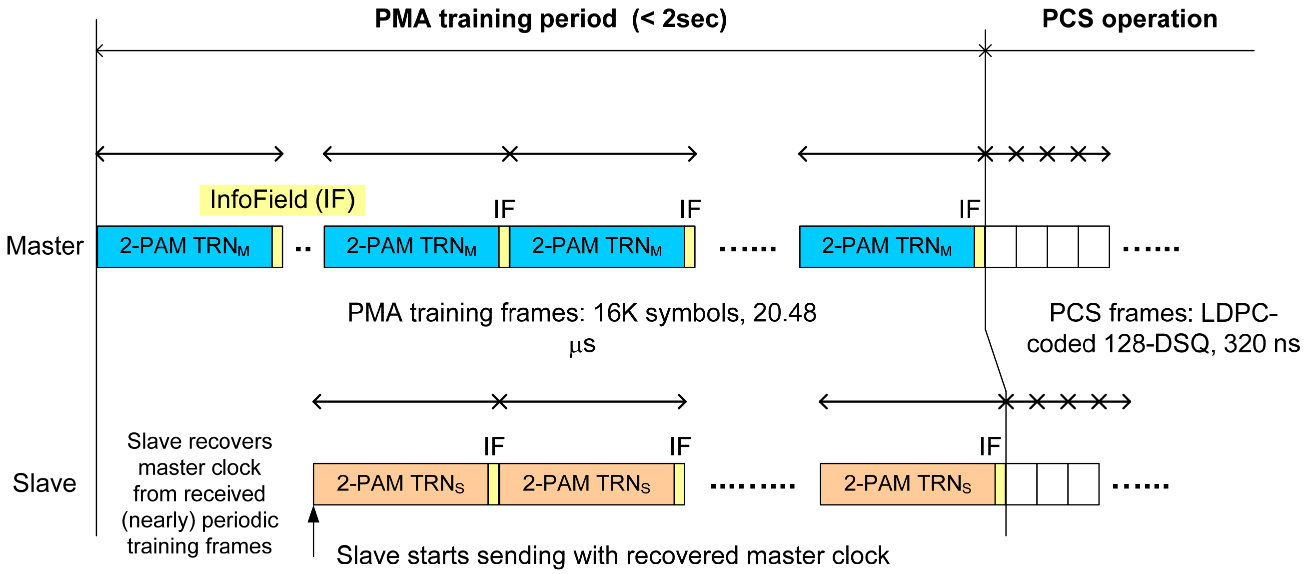
Partitioned frequency-domain filter



M. Joho and G.S. Moschytz, "Connecting partitioned frequency-domain filters in parallel or in cascade," *IEEE Trans. Circuits and Systems -II: Analog and Digital Signal Processing*, vol. 47, pp. 685-698, August 2000.

Start-up procedure

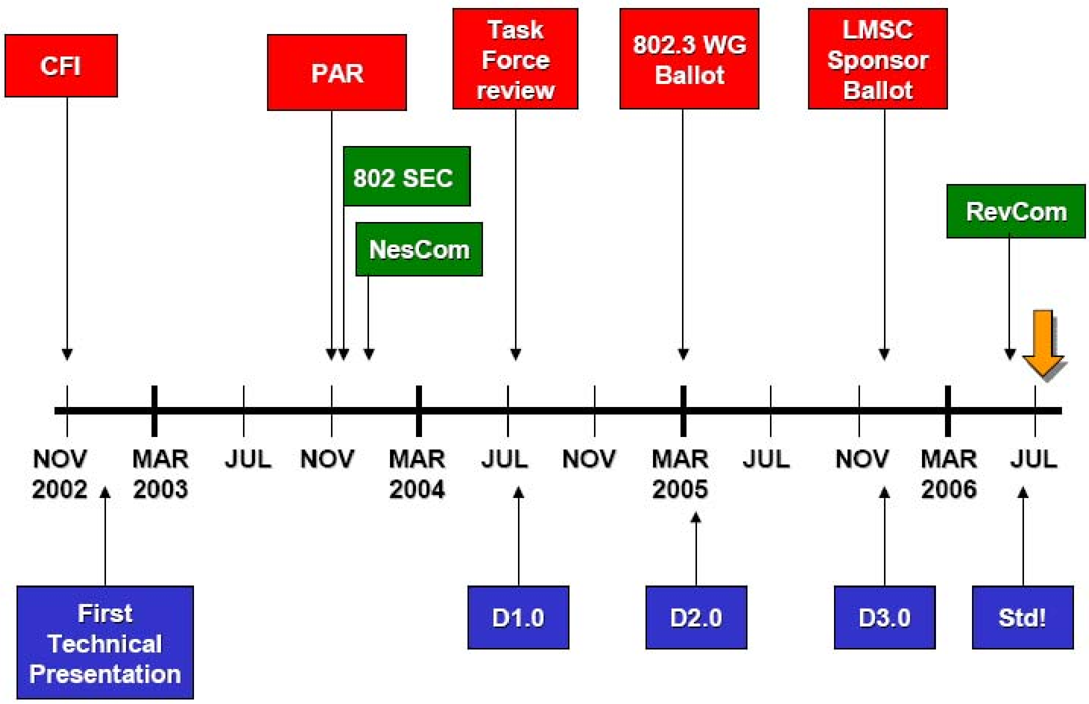
Start-up procedure



<u>PMA training frames:</u> pseudo random sequences, based on seed value determined during Auto Negotiation. Periodically transmitted during the entire PMA training period. Link partners communicate via InfoFields.

Status and outlook

The IEEE P802.3an (10GBASE-T) time line



Outlook

- **10GBASE-T: the ultimate copper PHY?** Pushing everything to the limit: data rate, cable length, transceiver front-end, signal converters, modulation and coding, length of adaptive filters, speed in every respect
- **Very large chip.** Estimated ≈ 10M gates, ≈ 10 W (65nm)
- **Big challenge for 100 m: front-end DACs, ADCs**
- **Main/initial market for 10Gbit/s over copper: short reach**
- **Intermediate solutions …**
  - Short-haul 10GBASE-T implementations: ≤ 30 m
  - “Copper FiberChannel”: 1, 2, 4 Gbit/s; 50 – 100 m Cat5/5A