IEEE 802 10GBASE-T Tutorial

Albuquerque, NM

November 10, 2003

Agenda

- **Overview**
  - **Presenter: Brad Booth; Chair, 10GBASE-T Study Group**
- **Cabling**
  - **Presenter: Alan Flatman; Independent**
- **PHY**
  - **Presenters: George Zimmerman; SolarFlare
Sailesh Rao; Intel**
- **Wrap-Up and Q&A**

IEEE 802 10GBASE-T Tutorial
Overview

Contributors: Shimon Muller, Sun Microsystems
Jeff Warren, Independent Consultant
Geoff Thompson, Nortel Networks
Bruce Tolley, Cisco Systems
Brad Booth, Intel Corporation

What – Where – Who – Why - When

- **What is 10GBASE-T?**
  - **It's a New 10GE PHY using the existing MAC**
- **Where are the 10GBASE-T applications?**
  - **Initially in the Data Center, but also the Horizontal**
- **Who will implement 10GBASE-T products?**
  - **Both Server and System Vendors (for data & storage)**
- **Why is 10GE over copper important?**
  - **Cost $$$ It's cheap relative to 10GE Optical**
- **When will it be available?**
  - **Typical standards timeline: 1st half of 2006**

What: 10GBASE-T Initial Goal

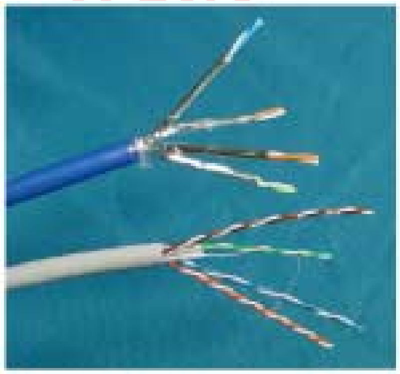
- **Initial Goal from Call-for-Interest**
  - **10 Gigabit Ethernet over horizontal structured, twisted-pair copper cabling**
  - **10 Gigabit Ethernet MAC and media independent interface as specified in IEEE 802.3ae™, 2002**
  - **Copper cabling is assumed to be ISO/IEC-11801:2002 Class D or better copper cable**

What: 10GBASE-T Objectives

- ***Keeping it Ethernet***
  - **Preserve the 802.3/Ethernet frame format at the MAC Client service interface**
  - **Preserve min. and max. frame size of current 802.3 Std.**
  - **Support star-wired local area networks using point-to-point links and structured cabling topologies**
- ***Keeping it 10 Gigabit Ethernet***
  - **Support full duplex operation only**
  - **Support a speed of 10.000 Gb/s at the MAC/PLS service interface**
- ***Compatibility with 802.3***
  - **Support Clause 28 auto-negotiation**
  - **To not support 802.3ah (EFM) OAM unidirectional operation**
  - **Support coexistence with 802.3af (DTE Power via Ethernet)**

What: 10GBASE-T Objectives (con't.)

- *Speed, Media & Reach*
  - Select copper media from ISO/IEC 11801:2002, with any appropriate augmentation to be developed through work of 802.3 in conjunction with SC25/WG3
  - Support operation over 4-connector structured 4-pair, twisted-pair copper cabling for all supported distances and Classes
  - Define a single 10 Gb/s PHY that would support links of:
    - At least 100 m on four-pair Class F (Cat 7) balanced copper cabling
    - At least 55 m to 100 m on four-pair Class E (Cat 6) balanced copper cabling



- *Environmental*
  - Meet CISPR/FCC Class A
  - Support a BER of 10^{-12} on all supported distances and Classes

Where: 10GBASE-T Applications

Application	10GBASE Fiber (802.3ae)	10GBASE-T	10GBASE-CX4 (802.3ak)
Data Center Server Clustering	**Yes**	**Yes**	**Yes** **(< 15m)**
Horizontal In Building (inc. wiring closet)	**No**	**Yes**	**No**
Vertical (Risers BB Links) Within Building	**Yes**	**No**	**No**
Campus & Metro	**Yes**	**No**	**No**

Where: 10GBASE-T Markets

- **1st – The Data Center**
  - **Density of compute devices (modular platforms)**
  - **Need more bandwidth per link than 1000BASE-T or link aggregation can provide**
  - **Less constrained by installed base or structured cabling standards**
  - **"If there is no way to verify the circuit, I will install new cabling"… M. Bennett, Lawrence Berkeley Lab**

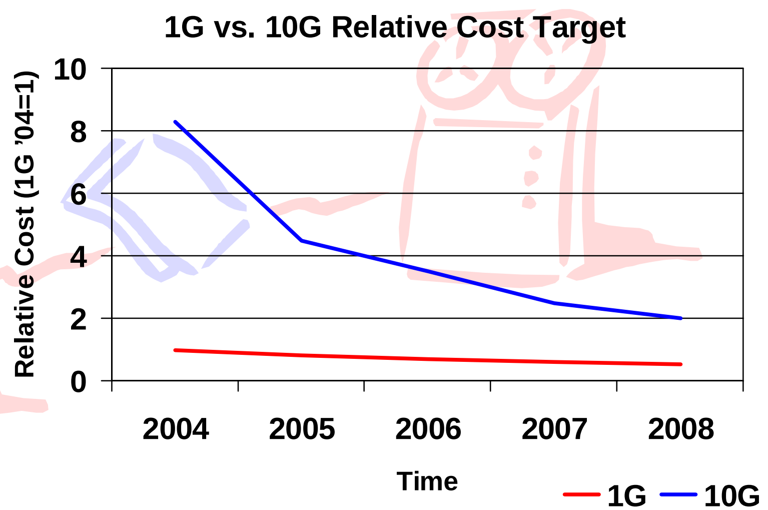
- **2nd – Horizontal Enterprise Networks**
  - **Higher speed aggregation points in the wiring closets**
  - **Needs to conform to structured cabling standards**
    - **Future build outs will utilize enhance cabling specifications**
  - **"Today's server is tomorrow's desktop"… S. Muller**
    - **Not by 2006, but eventually it will happen as costs drop and bandwidth intensive applications increase**

Who: 10GBASE-T Usage/Development

- **Manufacturers of 10GBASE-T Products**
  - **Data center & wiring closet LAN switching vendors**
  - **Data and storage server vendors**
  - **Module vendors**
- **End-User Profile: Lawrence Berkeley Lab**
  - **14,000 network attached devices 2003 year end**
  - **1,000 devices attached via 1000BASE-T by 2006**
  - **Horizontal limited to 100m – follow TIA/EIA-568-B**
  - **If necessary, willing to pull new cabling for 100m of 10GBASE-T**
- **IEEE 10GBASE-T Standards Participation**
  - **Nov 2002: CFI Attendance included 69+ vendors and users**
  - **July 2003: Technical development of Std. inc. 34+ companies**
  - **Sept. 2003; Portonovo, Italy: Attendance of 40 people**
    - **Almost as many as EFM**

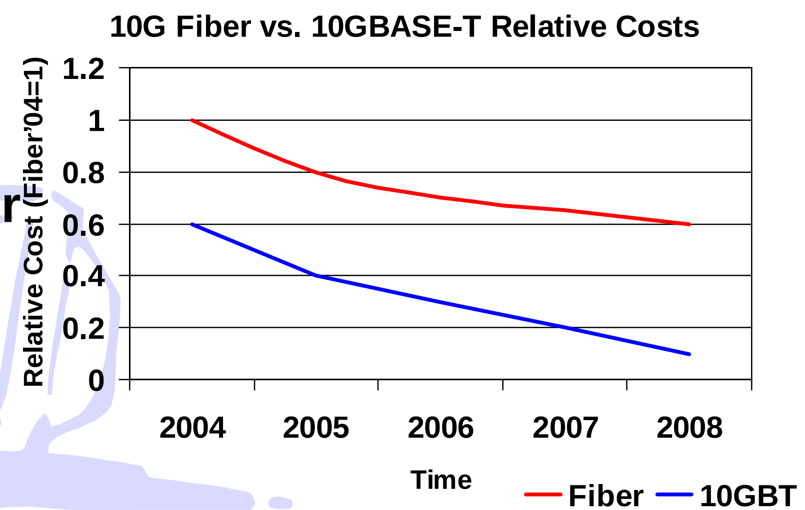
Why: 10GBASE-T Relative Costs

- **10GBASE-T vs. 1000BASE-T**
  - **Absolute cost will be 8-9x and trend toward 2-3x**
  - **Cost per gigabit will start at 0.8x**



- **10GBASE-T vs. 10GE Fiber**
  - **E-only vs. EOE**
  - **Absolute cost will target 0.6x fiber (short reach - SR) and trend to 0.15x**

Moore's Law in Action!

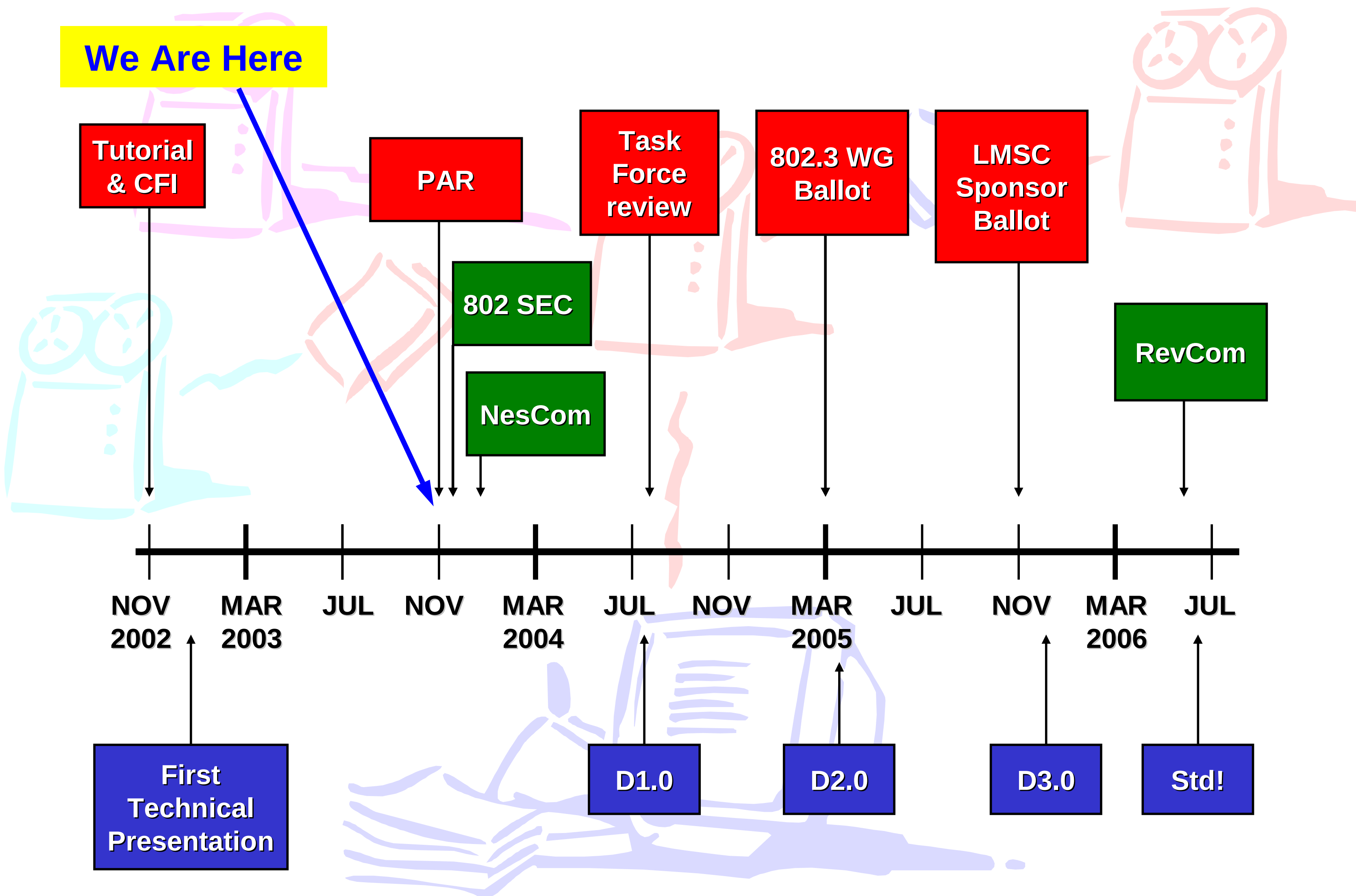


Source: Cahners In-stat, CFI Group

Why: 10GBASE-T Importance

- Faster network link speeds provide new generation of systems
  - Modular switches and servers
    - Backplanes and switch fabrics aggregate to support multiple 10GBASE-T ports
  - Servers with faster I/O subsystems (i.e. PCI Express™)
- Low cost solutions are market stimulus
  - 10GBASE-CX4 is a step in the right direction, but limited reach
  - 10GBASE-T:
    - Addresses PHY costs concerns in Enterprise market
    - Enhances reach and conforms to structured cabling environments
- Lower cabling costs
  - Installation practices are well-known
  - Ease of installation
  - Cost of termination

When: 10GBASE-T Timeline



IEEE 802 10GBASE-T Tutorial

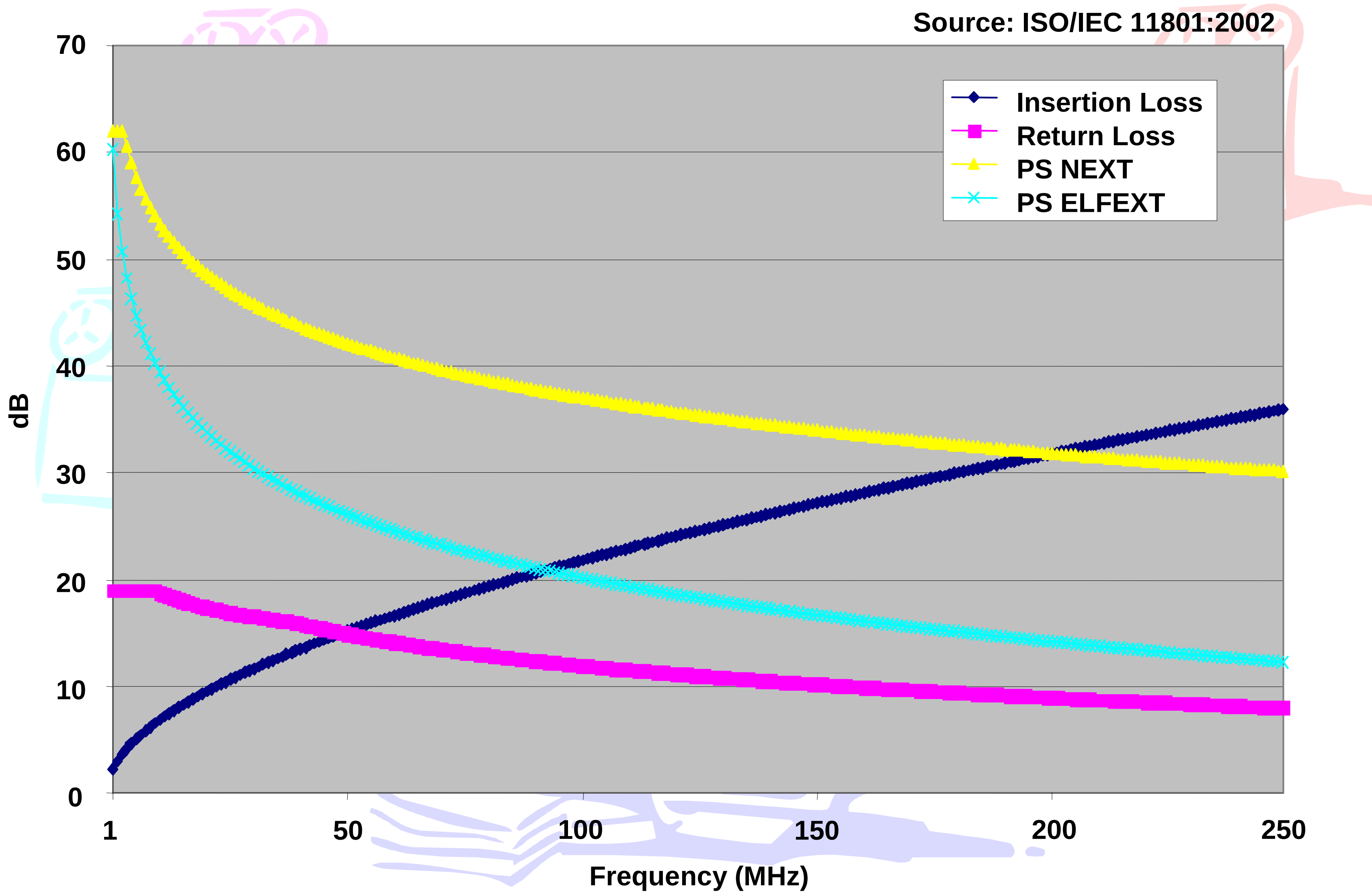
Cabling

Contributors: **Shadi AbuGhazeleh, Hubbell Premise Wiring**
Randy Below, The Siemon Company
Chris DiMinico, MC Communications
Alan Flatman, Independent Consultant
Valerie Rybinski, Hitachi Cable Manchester
Bruce Tolley, Cisco Systems
Sterling Vaden, Superior Modular Products

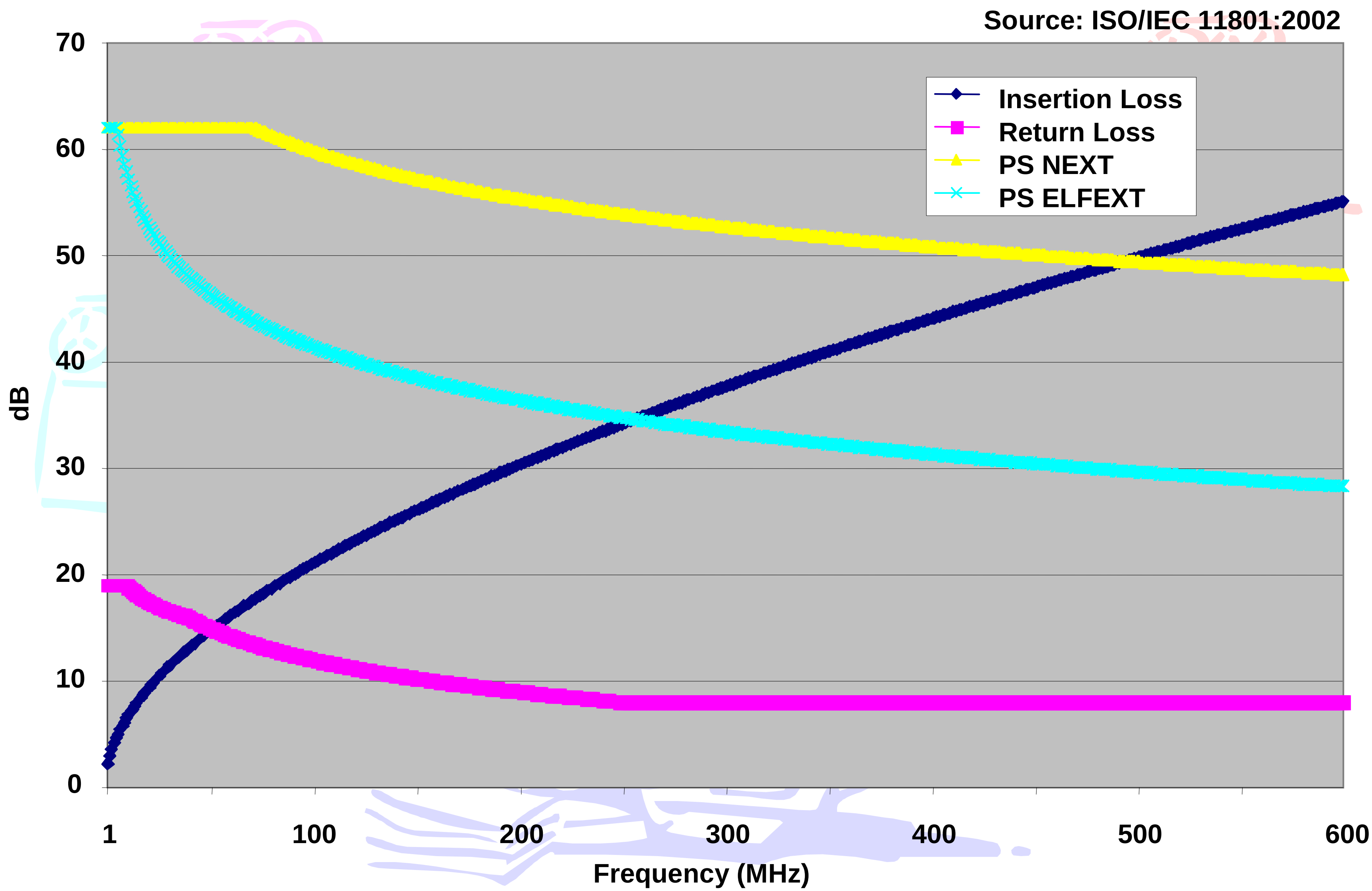
10GBASE-T Cabling Objectives

- **Support operation over 4-connector structured 4-pair, twisted-pair copper cabling for all supported distances and classes**
- **Define a single 10 Gbit/s PHY that would supports links of:**
  - **at least 100m on four-pair Class F balanced copper cabling**
  - **at least 55m to 100m on four-pair Class E balanced copper cabling**
- **Support star-wired local area networks using point-to-point links and structured cabling topologies**
- **Select copper media from ISO/IEC 11801:2002, with any appropriate augmentation to be developed through work of 802.3 in conjunction with ISO/IEC SC25 WG3**
- **Meet CISPR/FCC Class A EMC limits**
- **Support a BER of 10^{-12} on all supported distances and classes**

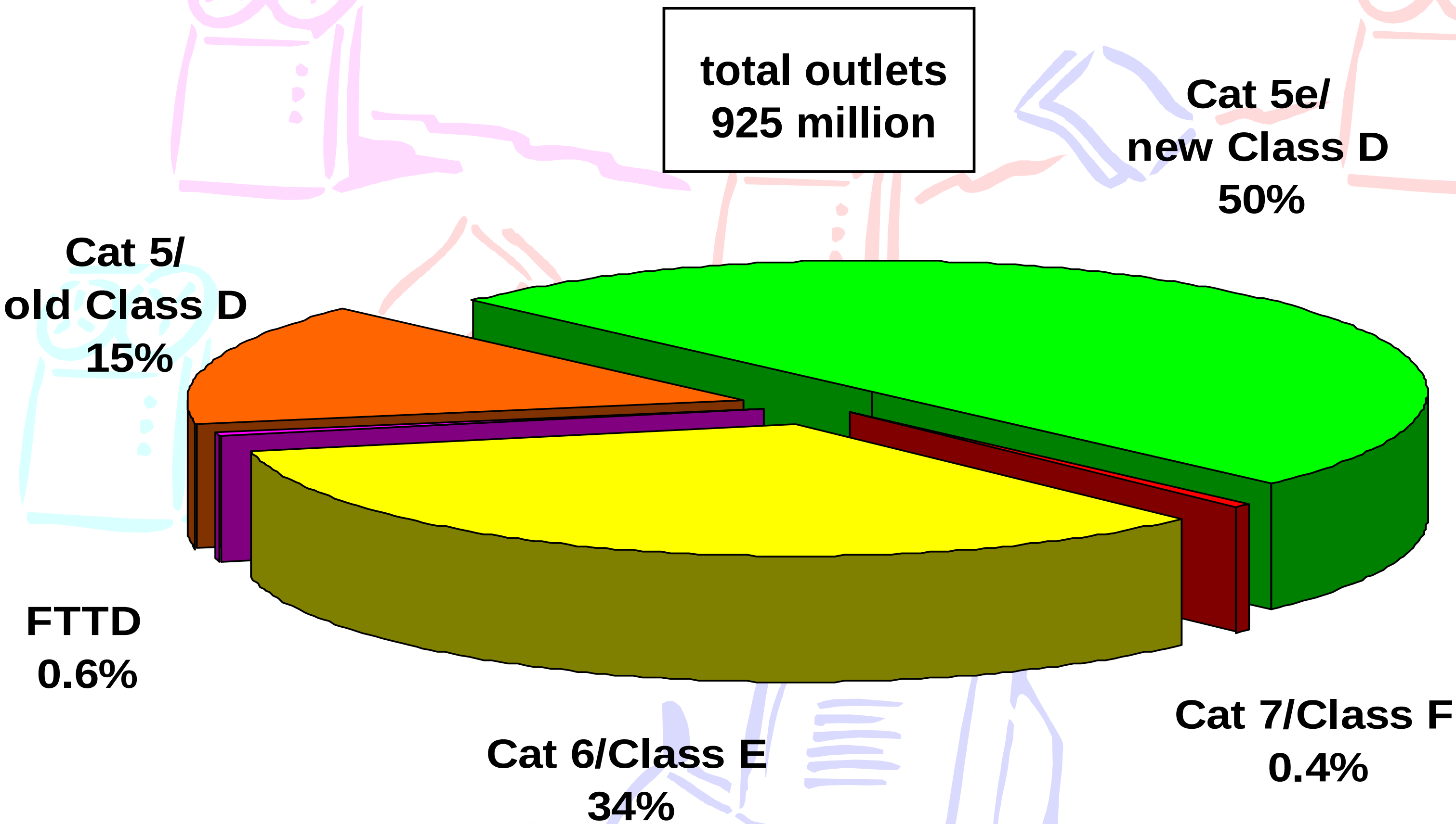
Class E Channel Performance



Class F Channel Performance

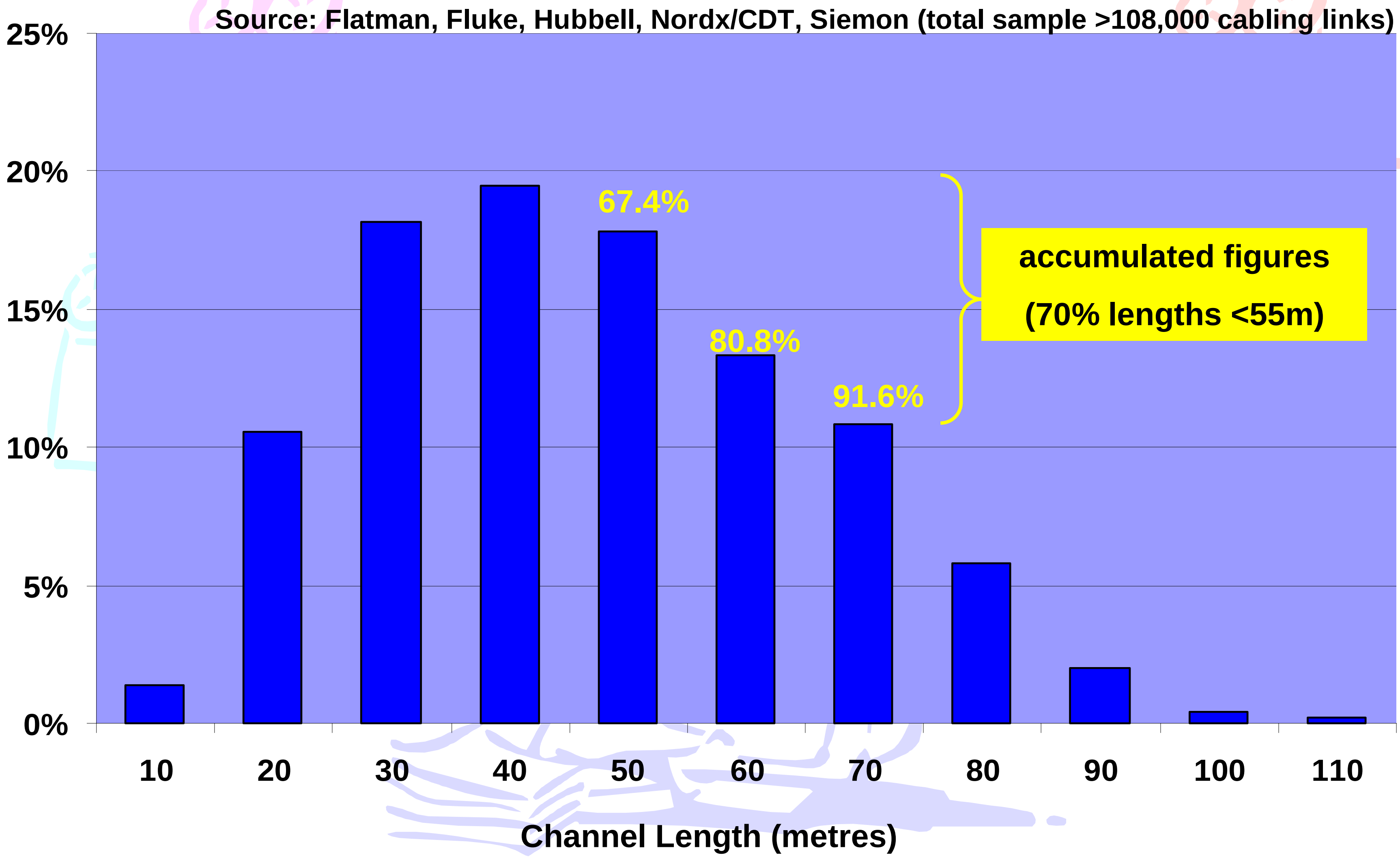


Worldwide Structured Cabling Deployment: Installed Base Forecast for Dec 2005

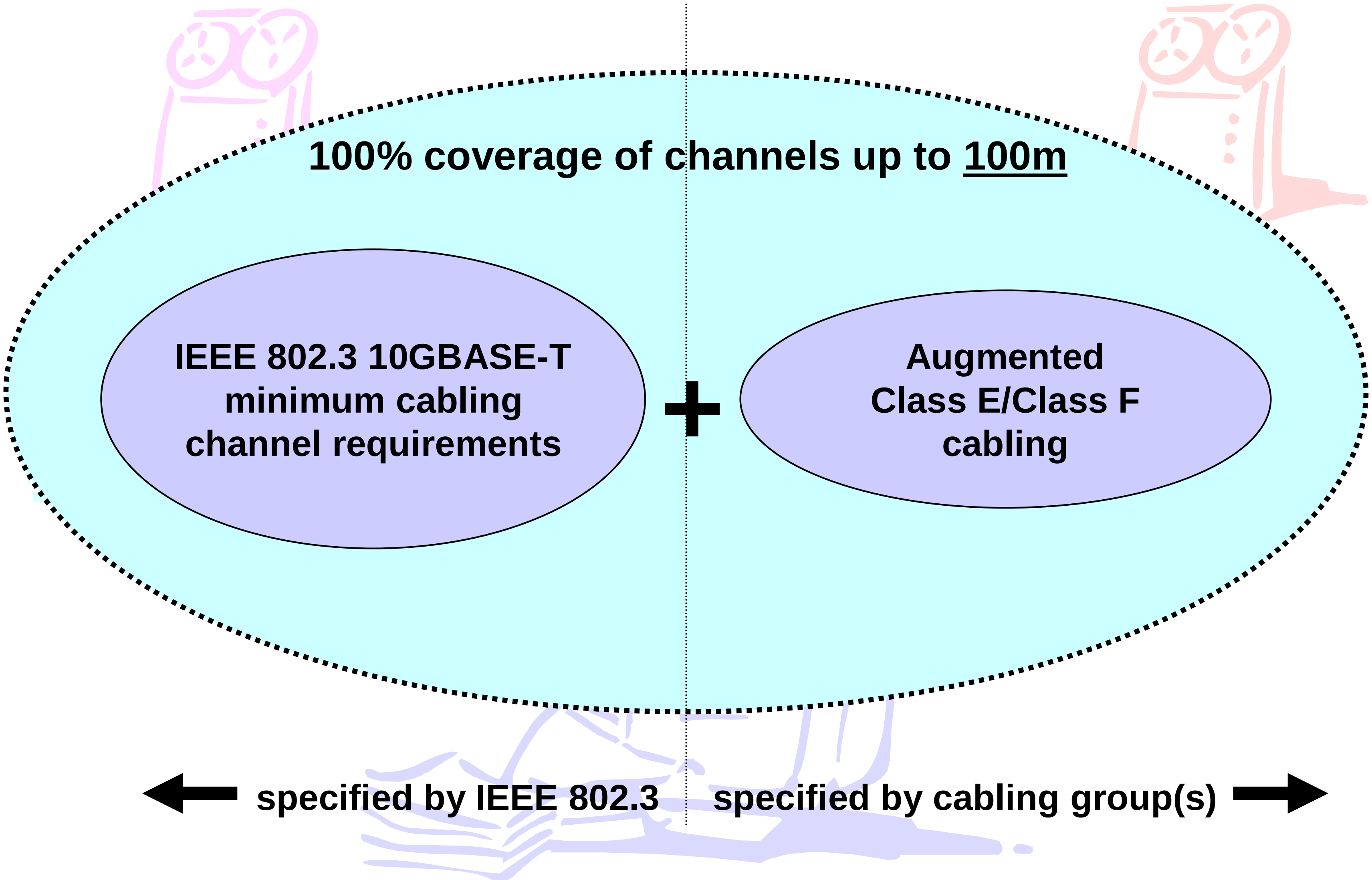


Source: LAN Technologies (Jan 2003)

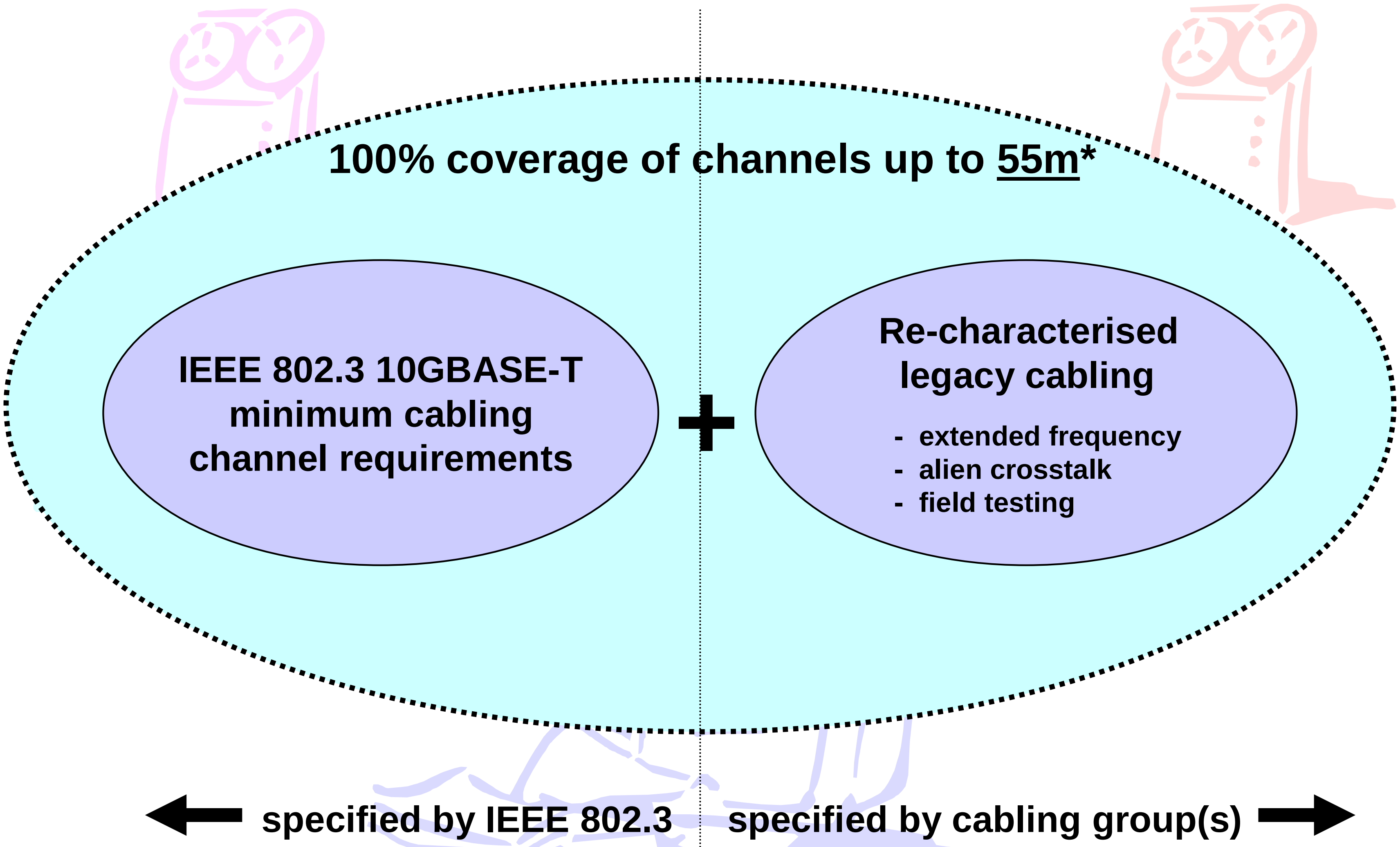
Commercial Office Structured Cabling: Channel Length vs % Distribution



Installing New Cabling for 10GBASE-T



Using Installed Class E Cabling for 10GBASE-T



*** 70% of all Class E/Cat 6 channels are less than or equal to 55m.**

10GBASE-T Cabling Characterisation

- **Performance models established by 10GBASE-T Study Group**
  - **Measurement data for Class D - F cabling to 625 MHz**
  - **Data includes screened & unscreened cabling systems**
  - **Measured data has been scaled to established limits**
  - **Data captured for Cat 5e/Class D, Cat 6/Class E, Cat 7/Class F:**
    - **Insertion Loss**
    - **Return Loss**
    - **Pair-to-Pair NEXT**
    - **Power Sum NEXT**
    - **Pair-to-Pair FEXT**
    - **Pair-to-Pair ELFEXT**
    - **Power Sum ELFEXT**
- **Alien Crosstalk also investigated by 10GBASE-T Study Group**
  - **Valuable measurement data established**

10GBASE-T Cabling - Key Tasks

- **Define worst case cabling channel requirements for 10GBASE-T**
- **Request cabling standards group(s) to verify channel requirements**
  - **This may be an iterative process**
- **Request cabling standards group(s) to develop industry specs**
  - **augmented Class E & Class F**
  - **re-characterised legacy cabling**
  - **alien crosstalk mitigation methods**
  - **alien crosstalk test method**

Worth noting:

- **ISO/IEC & TIA cabling groups proactively engaged in SG**
  - **ISO/IEC SC25 WG3 working on alien crosstalk management & offer to assist IEEE 802.3 in augmentation work necessary for Class E & Class F cabling.**
  - **TIA TR-42 has outlined projects to assist IEEE 802.3 in augmentation and extended frequency characterisation. This work will also establish the relationship of transmission parameters and alien crosstalk plus their field testing and mitigation.**

IEEE 802 10GBASE-T Tutorial

PHY

Contributors: **Joseph Babanezhad, Plato Labs**
Sanjay Kasturia, Teranetics
George Zimmerman, SolarFlare Communications
Sailesh Rao, Intel Corporation

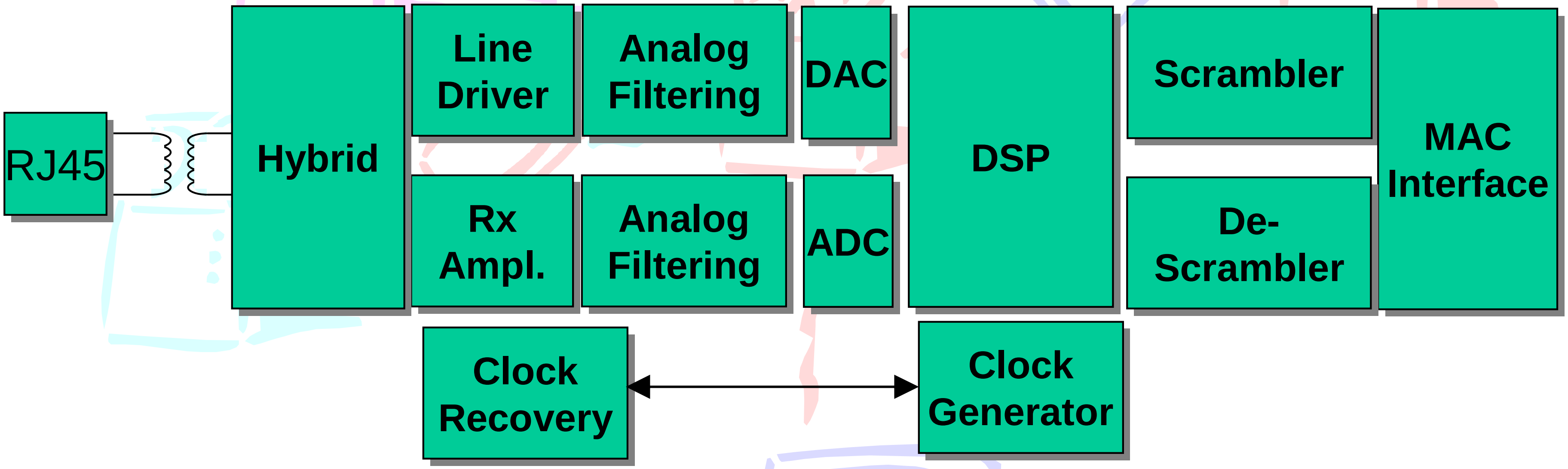
Comparison with 1000BASE-T

- **Full duplex operation only**
  - **Half duplex operation is not supported in 802.3ae MAC**
  - **1000BASE-T supported "carrier extension" for 1G repeaters**
- **The tutorial assumes signaling methodology which was the basis for most study group discussion**

1000BASE-T	10GBASE-T
5-level coded PAM signaling (2 information bits/symbol)	10-level coded PAM signaling (3 information bits/symbol)
8-state 4D Trellis code across pairs	8-state 4D Trellis code across pairs
Full duplex echo-cancelled transmission	Full duplex echo-cancelled transmission
125 Mbaud, ~80 MHz used bandwidth	833 Mbaud, ~450 MHz used bandwidth
No FEXT Cancellation	FEXT Cancellation required

- **Throughput is 4 (lanes) x 833 Mbaud x 3 bits/baud = 10Gb/s**

10GBASE-T Block Diagram



Key issues

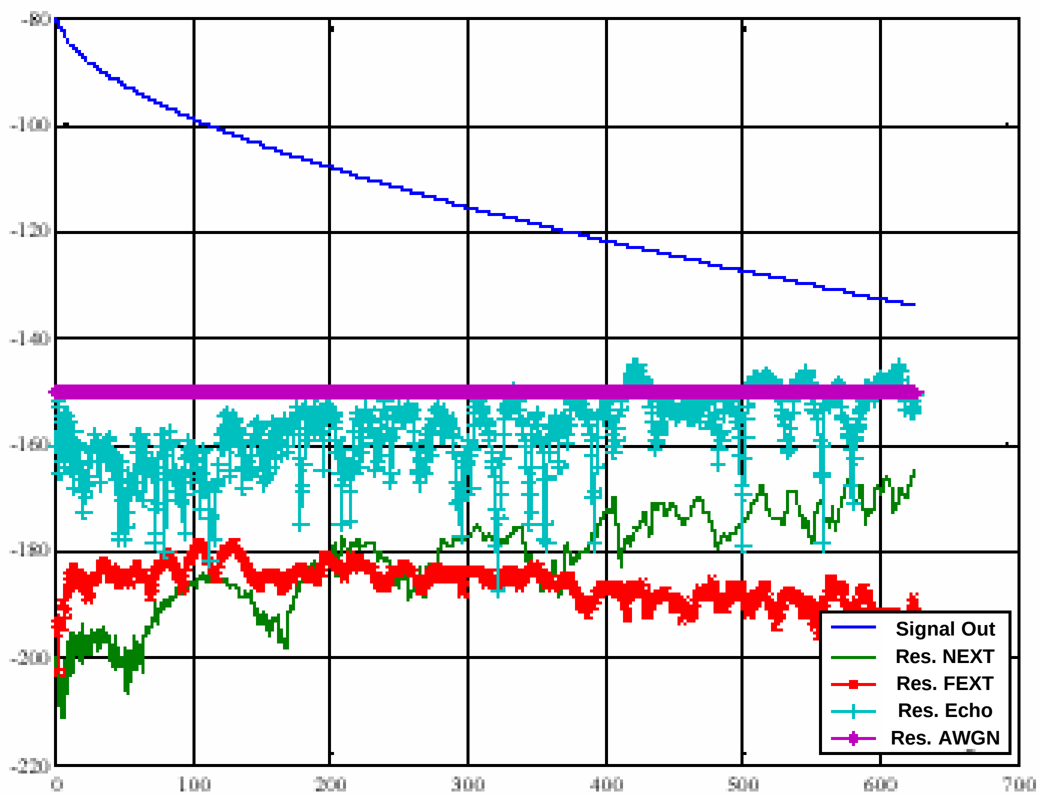
- Higher symbol rate requires higher signal bandwidth
  - Class D (Cat 5e), if used, will be utilized beyond its specified frequency range
  - Class E (Cat 6) will have to have it's performance characterized beyond 250MHz and up to 625MHz
    - TSB being prepared by TIA
  - Class F (Cat 7) is adequately specified
  - TIA and ISO are engaged in extended frequency and alien crosstalk augmentation and characterization of Class E & F
- Higher symbol rate and higher level modulation imply
  - Higher performance requirements on the Analog Front End
  - More complex signal processing
  - Cancellation of FEXT
  - Aggressive timing requirements
- Alien Crosstalk is a significant factor in capacity on UTP

Achievable Performance

- **With the 4 connector model and proposed signaling:**
  - **100m on Class F (Cat 7)**
  - **> 55m on Class E (Cat 6) operating beyond the specified frequency range**
  - **100m on the new cabling being defined by cabling standards groups (derivative of Class E/Cat 6)**
  - **20 to 60m on Class D (Cat 5e) was discussed**
    - **Requires operation beyond the specified frequency range**
    - **No consensus achieved on extending the specification**
- **Increase in system margin and/or reach are possible:**
  - **Several techniques have been presented in the SG:**
    - **Analog signal conditioning**
    - **Alien noise suppression**
    - **Improvements in the cabling specification**

Class F/Cat 7: 100m Capacity & Margin

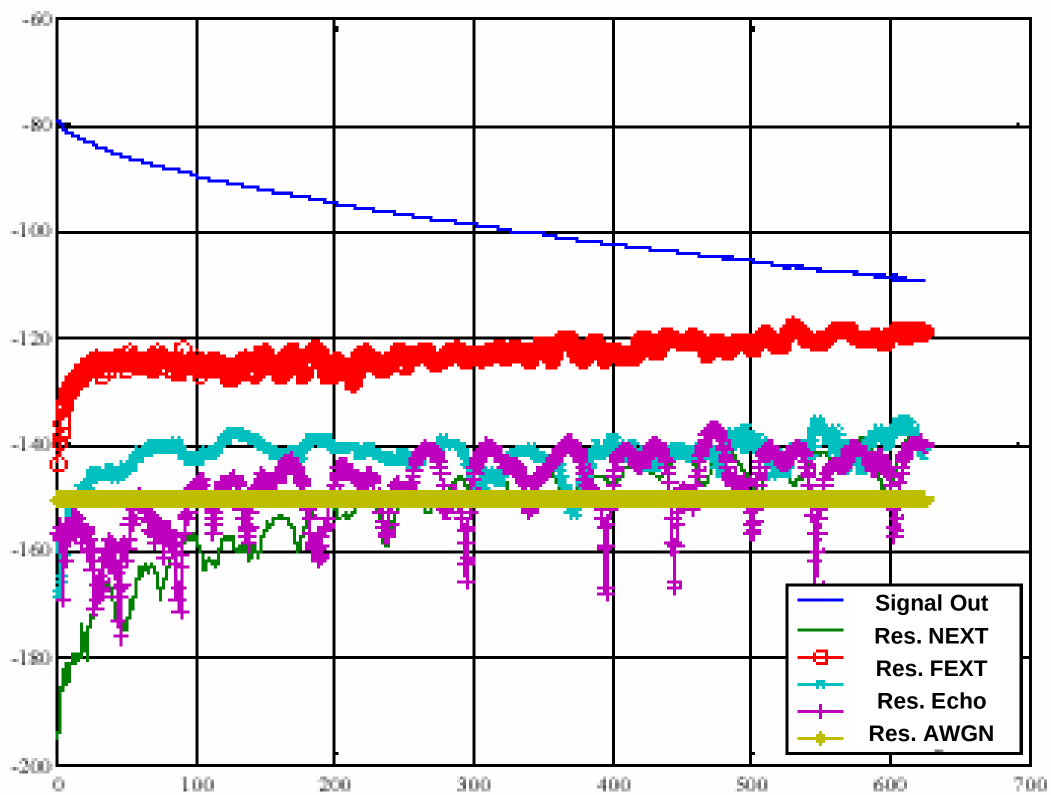
- Capacity > 28.6 Gbps
- > 16 dB PAM-10 margin at 1e-12 BER



Source: diminico_1_0903.pdf

Class E/Cat 6 UTP: 55m Capacity & Margin

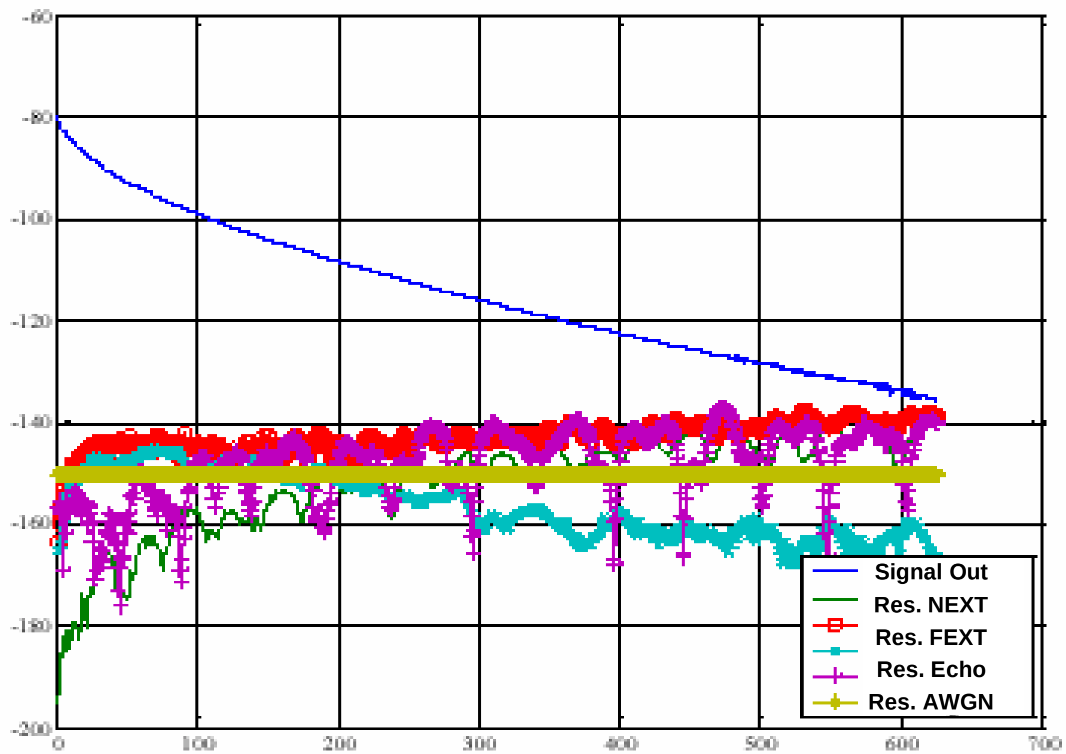
- Capacity > 19.8 Gbps
- >3.4 dB PAM-10 margin at 1e-12 BER



Source: diminico_1_0903.pdf

Improved Class E/Cat 6: 100m Capacity & Margin

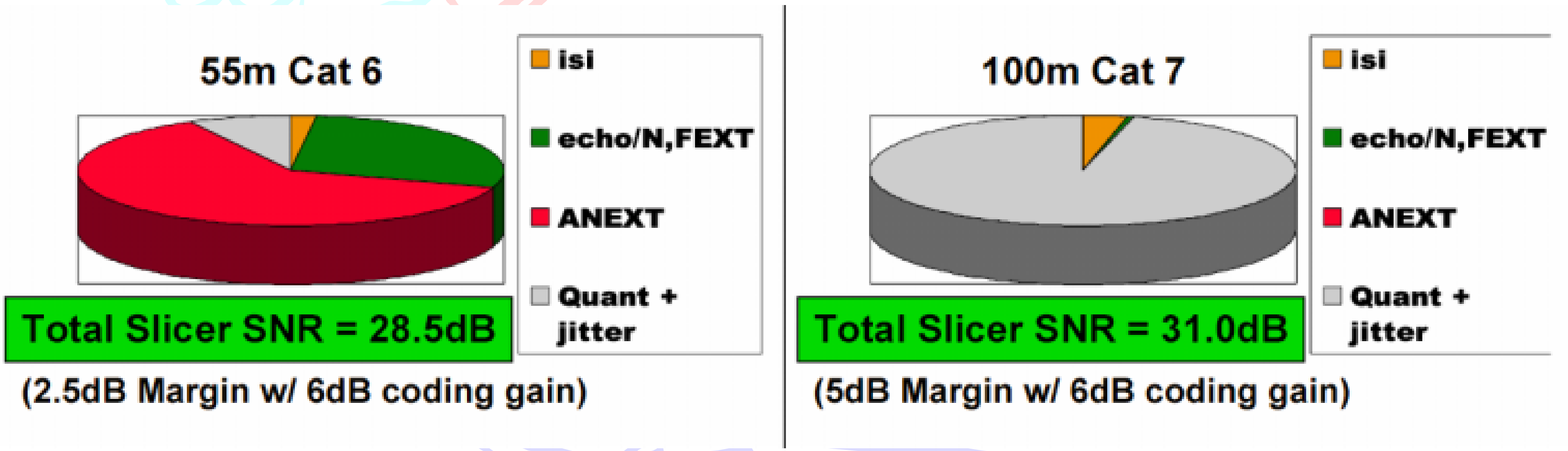
- (ref cohen_1_0703)
- Capacity > 20.6 Gbps
- >7 dB PAM-10 margin at 1e-12 BER



Source: diminico_1_0903.pdf

Example: 10Gbps over 100m Cat 7 and 55m Cat 6 UTP (mmse analysis)

- **PAM-10, Fbaud = 833 Msps**
  - **Uncoded SNR target = 32 dB, coded = 26 dB**
  - **This analysis was reproduced from powell_1_0903.pdf (Sept 03 Interim) using cabling adhoc models**



AFE Challenges

- **Demanding ADC requirements**
  - **Presentations made:**

Company	Rate	Jitter	ENOB
Broadcom, Vativ, Marvell	833 MS/s	4 ps	10.5 bits
Cicada	833 MS/s	N/a	9 bits
NEC Electronics Corporation	833 MS/s	3 ps	9.5 bits
Plato (PAM-5)	1250 MS/s	0.1 ps	10 bits
A. Vareljian, Independent	833 MS/s	N/a	9.5 bits
Solarflare	833 MS/s	6 ps	>8 bits

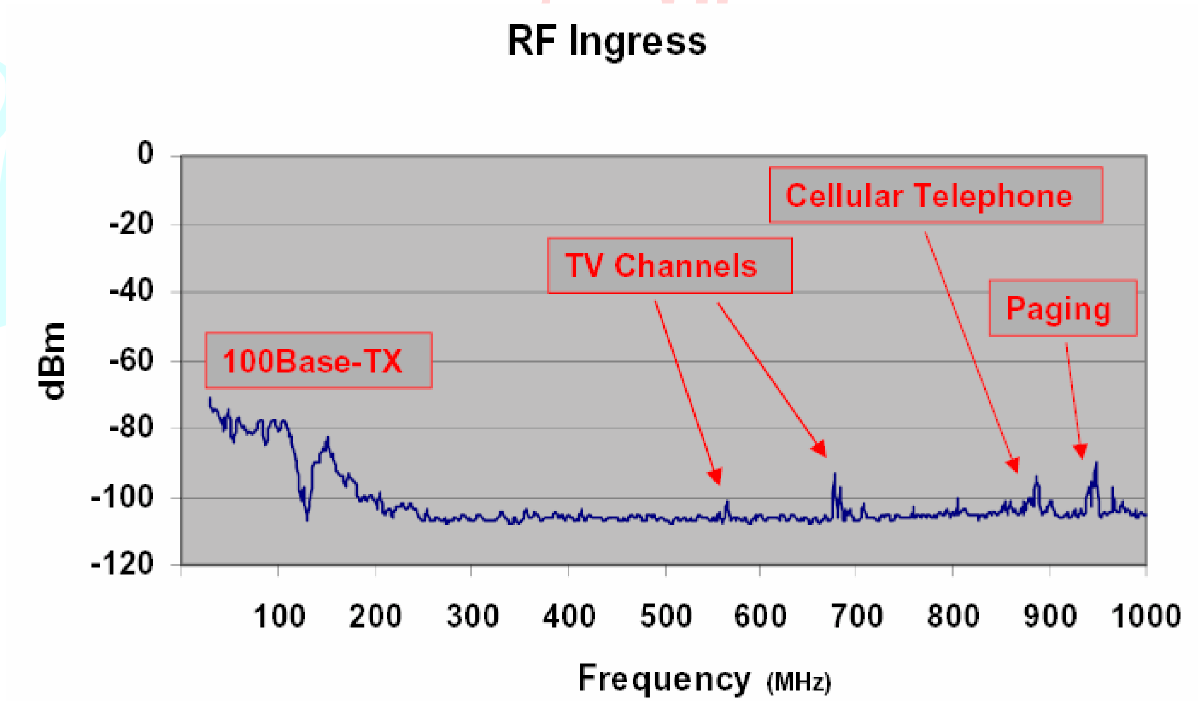
  - **Some variation due to differences in signal conditioning**
  - **Presentations have included known technology to implement these requirements**

DSP Challenges

- **Primary issue is complexity**
- **Driven by number of taps required in echo and NEXT cancellers**
- **Operation speed higher than in 1000BASE-T**
- **Presentations have been made detailing methods to reduce the complexity**
- **Complexity & power will benefit from process advances**

EMI

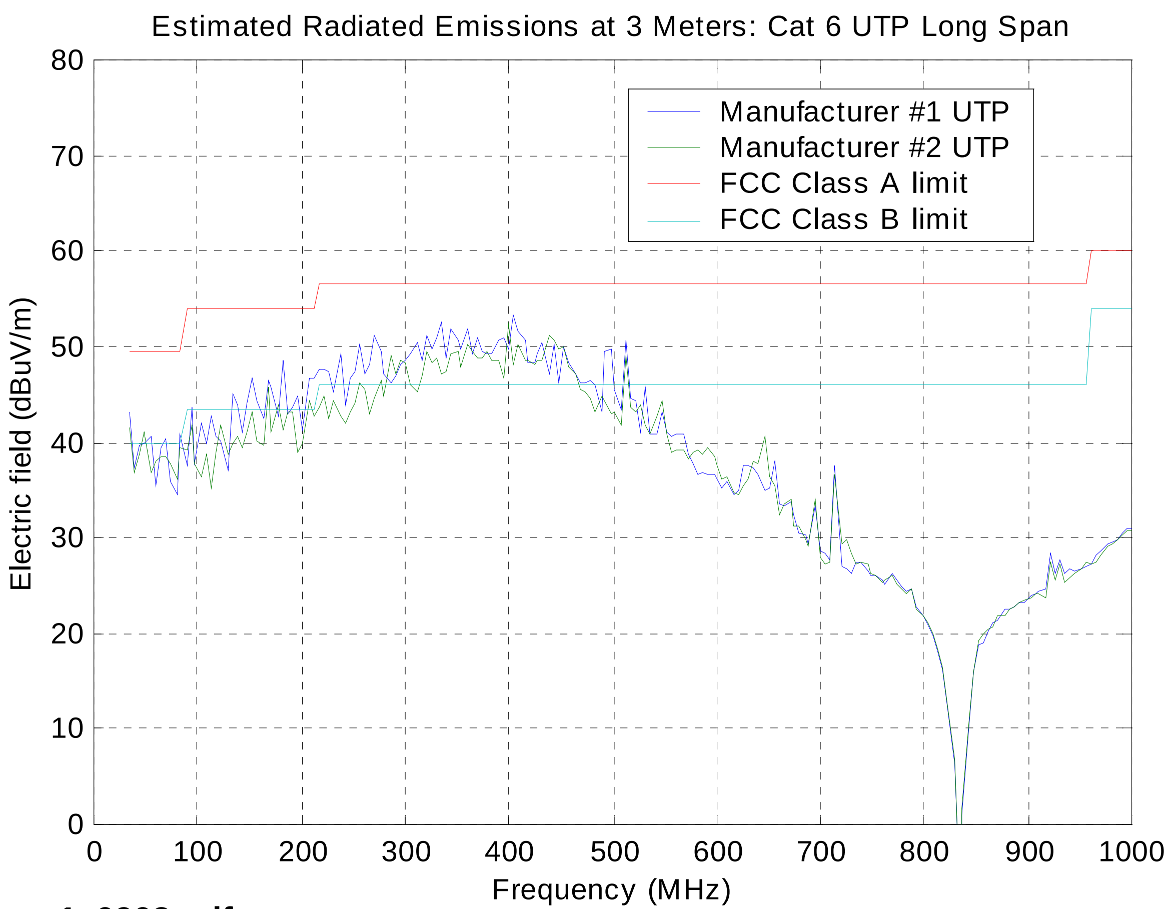
- **Presentations on EMI test data have included RF Ingress testing for interference and EMC compliance testing**
  - **Class D and Class E can be EMC compliant to FCC/CISPR Class A**
  - **RF ingress is not a limiting factor**



Source: cobb_1_0503.pdf

EMI Comparison of Cat 6 Cables

- **Unshielded Cat 6 cabling can meet FCC Class A**
- **Expect Class F (Cat 7) to meet FCC Class B**



Source: cohen_1_0903.pdf

Power

- **Several presentations have been made:**

Company	Process	AFE Power	DSP Power	Total Power
Broadcom, Marvell, Cicada, Vativ	65nm	8-16W	2.2W	10-18W
Solarflare	90nm	<5W	2W	<7W
NEC Electronics	90nm 65nm	3.7W 2.2W	8.0W 4.0W	11.7W 6.2W

- **Variations due to assumed linear extrapolation of existing technology versus implementations optimized for 10GBASE-T**

Areas for Complexity & Performance Improvement

- **Complexity**
  - **Advanced Filter Architectures**
  - **Analog Signal Processing**
- **Performance**
  - **Alien NEXT mitigation**
  - **Improved Coding gain**

Advanced Filter Architectures

	ECHO	NEXT	FEXT	FF EQ	**Total FIR**
FIR length	500	300	100	80	
BlockSize or net samples	524	724	156	176	
FFTsize	1024	1024	256	256	
log2N	10	10	8	8	
Real operations/sample for FIR	**500**	**300**	**100**	**80**	**7120**
Total operations/block for DFT FIR (4*(N/2)log_2(N)*2+4*N)/2	22528	22528	4608	4608	
Real operations/sample for FFT	**43**	**31**	**30**	**26**	**1005**
Approx Savings	**91%**	**90%**	**70%**	**67%**	**86%**
Gain	**11x**	**10x**	**3x**	**3x**	**7x**

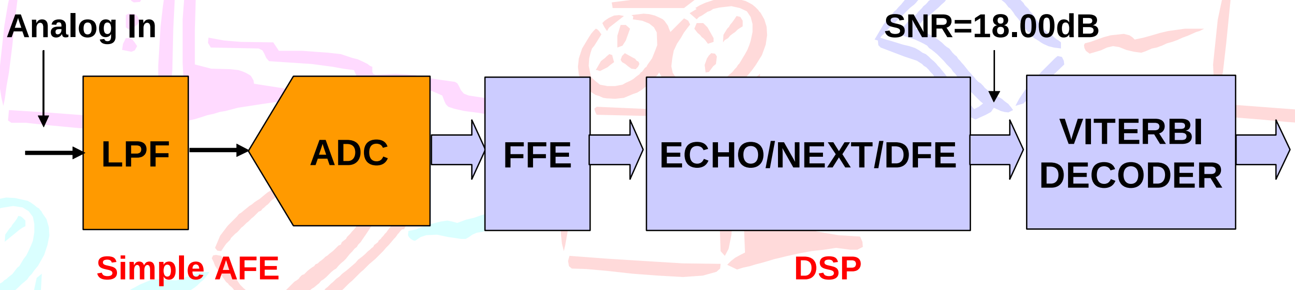
New Issues

- Block processing Latency
- Increased memory
- Increased precision

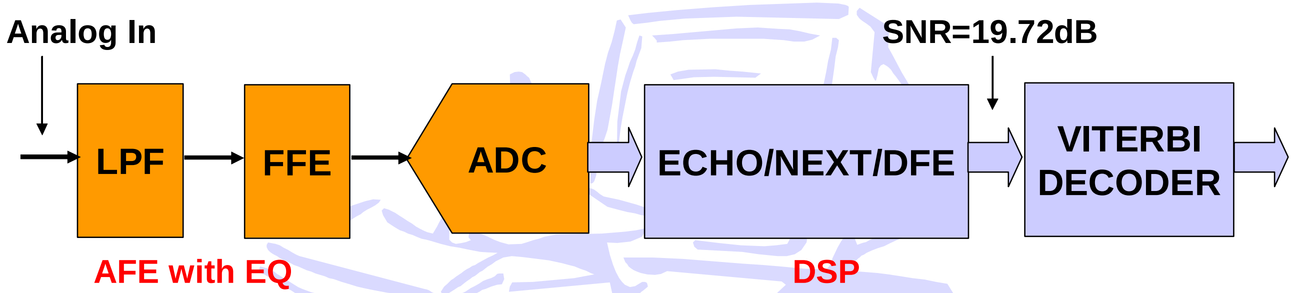
Source: kasturia_1_0903.pdf

Analog Signal Processing

- **Classic 1000BASE-T Architecture**
  - Significant quantization noise boosting at the Viterbi input



- **Alternative 1000BASE-T Architecture**
  - Reduced quantization noise boosting at the Viterbi input
  - Can use ADC with a lower ENOB

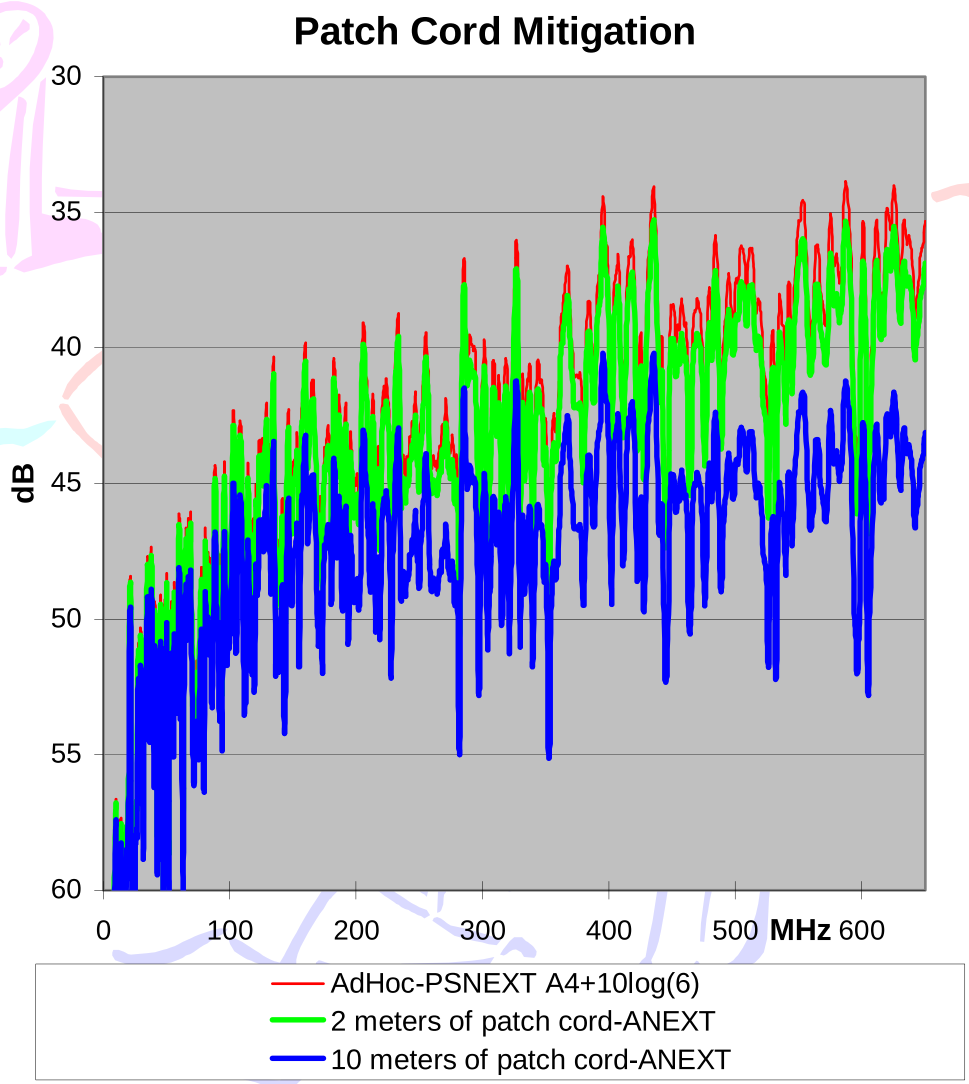


Reference: spencer_1_0703.pdf

Alien NEXT Challenges

- **Alien NEXT is agreed to be the limiting noise source for the objective reaches**
- **10GBASE-T is self-disturbing (from other links)**
  - **Worst case configuration is metal conduit filled to capacity with 10GBASE-T cabling links**
- **Various mitigation techniques proposed**
  - **Installation-practices based**
  - **Signal processing based**
  - **ISO & TIA cabling standards groups considering augmented cabling specs and testing in this area**

Example of Alien NEXT Mitigation



Background

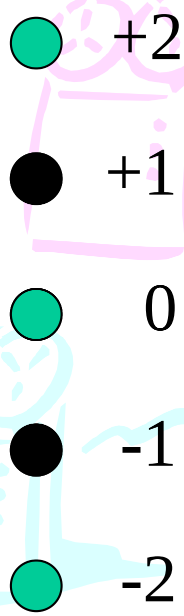
- **Full duplex operation only**
- **Much better system than 1000BASE-T can be designed with less stringent latency requirement**
  - **880ns MDI-MDI round trip latency specified in 1000BASE-T for back-to-back operation.**
  - **if MDI-MDI round trip latency budget is on the order of 1us for 10GBASE-T, we can use significantly more powerful techniques to reduce complexity of 10GBASE-T**
- **1000BASE-T line code and Forward Error Correction (FEC) code were designed for half-duplex operation**
  - **Strict latency budget requirements of CSMA/CD necessitated simple FEC codes**
    - **4D TCM used in 1000BASE-T shows weak Bit Error Rate reduction as a function of receiver Signal to Noise Ratio**

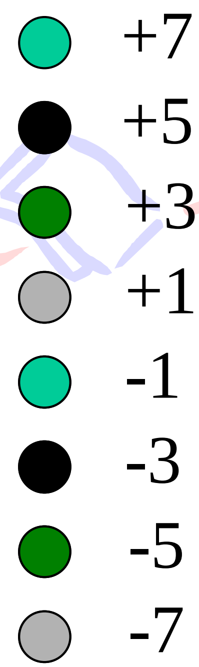
Improving the Coding

- Capacity Approaching FEC Codes
  - Gallagher's Low Density Parity Check (LDPC) Block Code
    - Achieves strong BER reduction as a function of SNR
  - Concatenated convolutional codes (Turbo Codes)
- 12dB co-set partitioning for improved noise tolerance
  - 6dB co-set partitioning used in 1000BASE-T
  - Doubles noise tolerance over 6dB partitioned codes
- Tomlinson-Harashima pre-coding to reduce receiver complexity
  - Allows spectral shaping in the transmitter to reduce alien cross-talk coupling
  - Eliminates Decision Feedback Equalizer (DFE) error propagation, even with large DFE coefficients

- Can be combined to target full 100m Cat 6 operation over extended Cat 6 specifications

12dB Co-set Partitioning



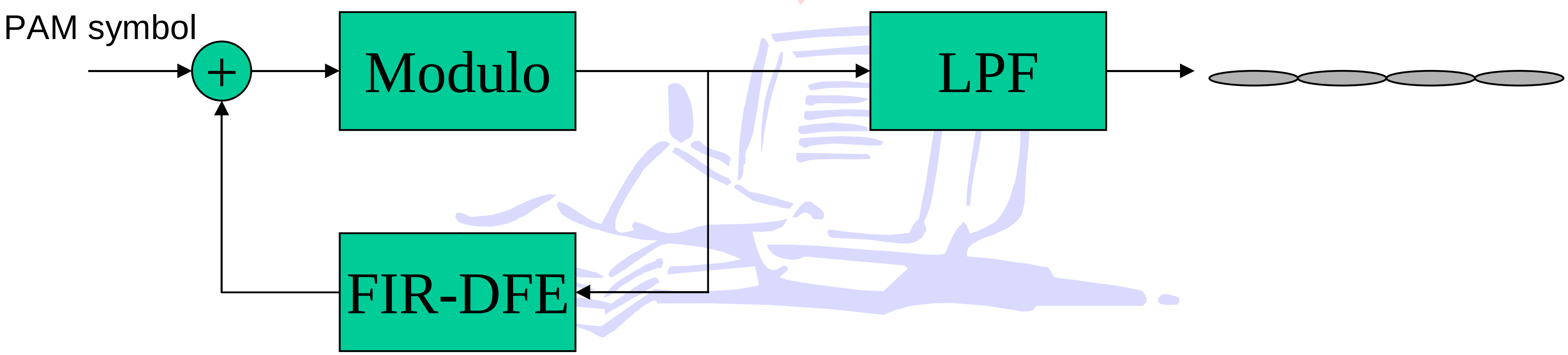


6dB co-set partitioning
in 4DPAM-5 1000BASE-T
(transmit 5 levels, but achieve noise immunity of 3 level signaling)

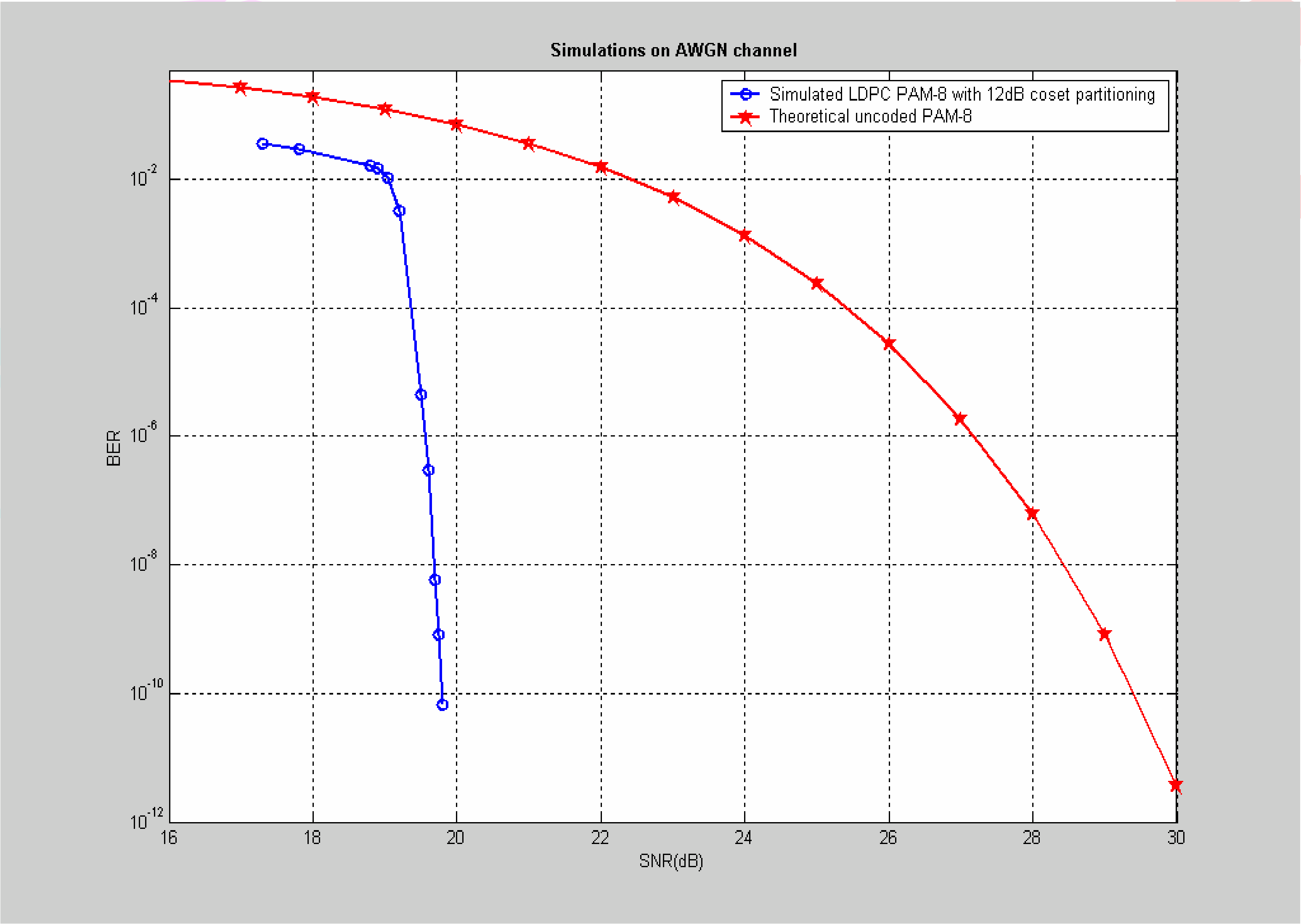
12dB co-set partitioning
in a 4DPAM-8 10GBASE-T
(transmit 8 levels, but achieve noise immunity of 2 level signaling)

Tomlinson-Harashima Pre-coding

- Independently developed by Tomlinson and Harashima in 1971.
- Uses a Decision Feedback Equalizer at the transmitter instead of the receiver
  – receiver computes DFE coefficients during startup and sends coefficients over to transmitter
  – advantage - allows for block processing and decoding at the receiver.
  – advantage - reduces complexity of receiver analog front end.
  – drawback - increases complexity of transmitter.



Example: Simulation Results



Wrap-Up

- 10GBASE-T is a new PHY for 10 Gigabit Ethernet
  - Compatible with existing 802.3ae MAC, XGMII and XAUI
- Primary Target Market is initially the Data Center
  - Greater reach than CX4
  - Lower cost than 10GbE fiber
- Provides 10GbE for the horizontal structured cabling market
  - To 100m on Class F (Cat 7)
  - To at least 55m on Class E (Cat 6)
- Early review shows EMI and power within acceptable range
- Further gains in reach, power and EMI
  - Complexity and performance improvements
  - Augmented cabling specifications

Next Step

The 10GBASE-T Study Group would like your support in moving forward on the 10GBASE-T project.

Supporters

Brad Booth; Intel
Jeff Warren; Independent
Shimon Muller; Sun
Geoff Thompson; Nortel
Bruce Tolley; Cisco
Alan Flatman; Independent
Shadi AbuGhazaleh; Hubbell Premise Wiring
Randy Below; Siemon
Chris Di Minico; MC Communications
Valerie Rybinski; Hitachi
Sterling Vaden; Superior Modular
George Zimmerman; SolarFlare
Joseph Babanezhad; Plato
Sailesh Rao; Intel
Sanjay Kasturia; Teranetics
Scott Powell; Broadcom
Clint Early, Jr.; Independent
Ron Nordin; Panduit
Terry Cobb; Avaya
Luc Adriaenssens; Avaya
Hugh Barass; Cisco
Albert Vareljian; Independent
Carrie Higbie; Siemon
Kiyoshi Fukuchi; NEC Corporation
Ben Brown, Independent
Tetsu Koyama, NEC Electronics
Petre Popescu; Quake
Dan Dove; HP ProCurve
Jose Tellado; Teranetics
Richard Mei; Avaya
Bernie Hammond; KRONE
Joe Dupuis; Ortronics
Mike Bennett; Lawrence Berkeley Labs
Barry O'Mahony; Intel
Wael William Diab; Cisco
Kevin Daines; World Wide Packets
Paul Vanderlaan; Belden Wire and Cable
Henri Koeman; Fluke
Rick Rabinovich; Spirent
David Law; 3Com
Rob Hays; Intel
Steve Carlson; High Speed Design
Ted Rado; Analogix
Mike McConnell; KeyEye
Rehan Mahmood; Hubbell Premise Wiring
Steve Haddock; Extreme Networks
Wayne Eng; Mysticom

Reflector and Web Site

- **To subscribe to any of the 10GBASE-T reflectors send an email to:**

  majordomo@ieee.org

  with the following in the <u>body of the message</u>:

  subscribe stds-802-3-10GBT <your email address>

  subscribe stds-802-3-10GBT-Modeling <your email address>

  subscribe stds-802-3-10GBT-Cabling <your email address>

- **10GBASE-T Study Group web page URL:**

  http://www.ieee802.org/3/10GBT/

Q&A

Thank you!!

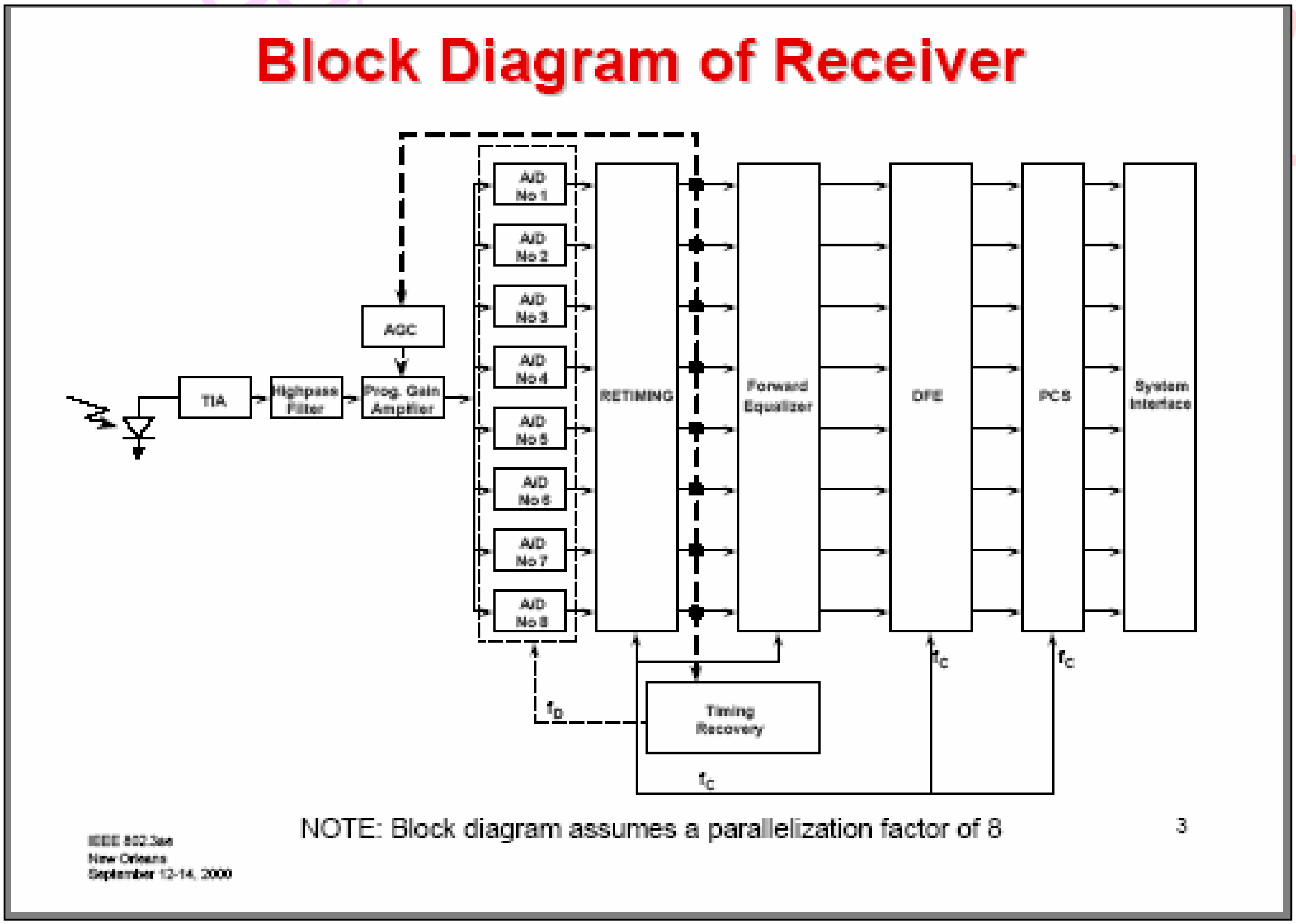
Backup

Capacity Approaching Codes

- Gallagher's Low Density Parity Check (LDPC) Code
  - First introduced in Robert Gallagher's MIT PhD thesis in 1960
  - Re-discovered by Mackay and Neal in 1995. Since then,
    - used in high performance optical networking systems
    - Proposed by JPL for use by Consultative Committee for Space Data Systems (CCSDS)
  - Receiver uses an iterative belief propagation decoder to achieve waterfall reduction of BER as a function of SNR.
  - uses sparse matrix techniques to minimize decoder complexity
    - reasonably low latency requirement in the decoder
    - allows block processing to reduce receiver complexity
  - can be based on provably good block codes to maximize Hamming distance between code words.
- Concatenated Convolutional codes (Turbo codes)
  - Uses two interleaved convolutional codes to create "turbo" effect
    - 1000BASE-T uses a single convolutional code due to latency budget constraint
    - Excessive latency requirement and decoder complexity

Example of Time Interleaved ADC

- “DSP Based Equalization for Optical Channels”, Sept. 2000



6bit
10GHz
800mvP-P
8 parallel ADCs
.18u
1.8V
450mW