DesignCon 2014

The state of IEEE 802.3bj 100 Gb/s Backplane Ethernet

Matthew Brown, APM
Mike Dudek, QLogic
Adam Healey, LSI
Elizabeth Kochuparambil, Cisco
Liav Ben Artsi, Marvell Israel Ltd.
Richard Mellitz, Intel
Charles Moore, Avago Technologies
Adee Ran, Intel
Pavel Zivny, Tektronix

Abstract

100 Gigabit Ethernet over backplane is a subject of the current IEEE P802.3bj standardization effort. The 4 x 25 Gb/s design presents significant challenges to today's technology and several new methodologies were developed for the task. We explain the philosophy driving the design of the standard, present a summary of problems solved, and provide details on the most important methodologies. In particular we will discuss: the trade-off between PAM4 and PAM2 signaling; forward error correction; transmitter to channel to receiver boundaries; COM channel description and its advantages in closing the specification; jitter and noise budget; and measurement methodology.

Author Biographies

Matthew Brown: Matt is a Principal Systems Engineer at Applied Micro, responsible for architecture and specifications definition for a range of products including mixed signal and digital transceivers, Ethernet PHYs, packet processors, and OTN framers. He has over 30 years of experience developing system and device products addressing data communications and telecom applications at companies including Bell Northern Research, Northern Telecom (Nortel), Cambrian, Ceyba, and Quake. He has been a steady participant and contributor in various physical layer standards including 802.3ap, 802.3ba, 802.3az, 802.3bm, and SFP+, and is currently the Editor-In-Chief and clause editor for the IEEE 802.3bj 100 Gb/s Backplane and Copper Cable task force. He is a holder of ten US patents.

Adam Healey: Adam is a Distinguished Engineer at LSI Corporation where he supports the development and standardization of high speed serial interface products. Adam is a regular contributor to the development of industry standards through his work in the IEEE 802.3 Ethernet working group and INCITS T11.2 Fibre Channel Physical Variants task group. Adam was chairman of the IEEE P802.3ap Task Force chartered to develop the standard for Ethernet operation over electrical backplanes at speeds of 1 and 10 Gb/s. He also served as editor-in-chief for the IEEE P802.3bj Task Force chartered to extend the Backplane Ethernet family to 100 Gb/s as well as define Energy Efficient Ethernet capabilities for 40 and 100 Gb/s Ethernet. Adam is currently chairman of the IEEE P802.3bj Task Force and vice chairman of the IEEE 802.3 Ethernet Working Group. Adam has also previously served as Secretary of the IEEE 802.3 Ethernet Working Group as well as technical committee chairman and Vice President of Technology for the Ethernet Alliance.

Liav Ben Artsi: Liav is a Senior Signal Integrity Manager at Marvell Israel Ltd. Liav has worked at Marvell for the last 13 years, most of the time as a signal integrity engineer, focusing on signal integrity methodologies, guidelines, test plan definitions and training. He holds several patents in the field as well as several others pending. For the last two years he has taken an active part in IEEE802.3bj standard committee.

Mike Dudek: Mike is Director of Signal Integrity at QLogic Corporation. He has worked in the high-speed communications field for over 30 years and holds numerous patents in the field. He has been active in Ethernet standards since 1997. He has previously worked for JDSU, Picolight, Cielo, Vixel, ADC Fibermux, PCO, and the GEC Hirst Research Centre. He has an MA in Physics from Oxford University.

Elizabeth Kochuparambil: Elizabeth is a Signal Integrity engineer at Cisco Systems with background in PCB materials and manufacturing along with high speed channel design. The breadth of her work ranges from the system level on switches and routers to PCB and substrate trace geometry to 25G copper standards (IEEE & OIF) since 2007.

Richard Mellitz: Richard is a Principal Engineer in the Data Center Group at Intel. Since joining Intel in 1998, Richard has been key member of various processor and I/O bus teams including Itanium®, Pentium® 4, , Xeon®, Microserver, and PCI Express®, SAS®, and Fabric. Additionally he has been a key contributor for the channel sections of the 10 Gbps, 40 Gbps, and 100 Gbps IEEE802.3 backplane standards. He founded and chaired an IPC committee delivering IPC's first PCB loss test method. Richard holds a number of patents in signal integrity, design, and test and has delivered many signal integrity papers at PC industry design conferences. Richard wrote the original IPC TDR standard which is used throughout the PCB industry.

Charles Moore: Charles is Chief SerDes Architect at Avago Technologies. He worked at Hewlett Packard, then Agilent Technologies from its start, then Avago since its start. He holds numerous patents circuit design and optics. He has worked on Ethernet standards since 2004.

Adee Ran: Adee is a high-speed communication physical-layer architecture and design expert at Intel's Communication and Storage Infrastructure Group. He is a participant and contributor to the IEEE 802.3 Ethernet working group. Adee holds an M.Sc. in Electrical Engineering from the Technion, Israel Institute of Technology, has co-authored papers and was granted several patents in the field.

Pavel Zivny: Pavel is a communications measurements domain expert with the high performance oscilloscopes group of Tektronix. Pavel was granted oscilloscope related patents, authored industry articles and papers, and represents Tektronix to high-speed serial data standards committees.

Content

1. 100 Gigabit Ethernet over backplane and copper cable introduction

1.1. Background – 10 Gb/s development

The standard for 10 Gb/s Ethernet, including optical Physical Layer specifications such as 10GBASE-SR, 10GBASE-LR, and 10GBASE-ER, was developed by the IEEE P802.3ae Task Force[1] and approved in 2002. Two years later, a new standard for 10 Gb/s Ethernet over four differential lanes of twin-axial copper cable (10GBASE-CX4) was approved. With the widespread deployment of 10 Gb/s Ethernet on the front-panel, the need for more backplane capacity grew.

To address the need for backplane capacity, and to establish standards-based Ethernet operation within blade servers, modular switches, and telecommunications equipment, the first IEEE standard for "Backplane Ethernet" was developed by the IEEE P802.3ap Task Force[1] and approved in 2007. This new standard included 10GBASE-KR which specified 10 Gb/s serial operation over a backplane medium. Meanwhile, the Small Form Factor (SFF) committee defined 10 Gb/s serial chip-to-module electrical specifications and direct attach copper cabling specifications.

[1] Nomenclature of IEEE 802.3 standards prefixes a 'P' when the standard is under development, e.g. P802.3ae. When the standards is ratified the 'P' is removed, e.g. 802.3ae.

1.2. 40 and 100 Gb/s development

In response to the growing need for bandwidth, the standard for 40 Gb/s and100 Gb/s Ethernet was developed by the IEEE P802.3ba Task Force[1] and approved in 2010. This standard included Physical Layer specifications for 40 Gb/s operation over backplane and copper cabling as well as 100 Gb/s operation over copper cabling based on parallel 10 Gb/s links. The requirements for each lane of the electrical interface were based on 10GBASE-KR.

The standard also included 100 Gb/s optical Physical Layer specifications based on 4 lanes at 25 Gb/s per lane (100GBASE-LR4 and 100GBASE-ER4). However, it did not include a specification for 100 Gb/s Backplane Ethernet.

With the emergence of new technologies and the ever-growing demand for bandwidth, the means and motivation for 25 Gb/s operation over copper media were available. In 2011, the IEEE P802.3bj Task Force was formed to extend the Backplane Ethernet family to 100 Gb/s (4 lanes at 25 Gb/s per lane) and also define operation over 4-lane twinaxial copper cables at 25 Gb/s per lane.

1.3. Status in January of 2014

Development of the standard for 100 Gb/s Ethernet over four differential lanes of copper is nearing completion. It brings in several new concepts in standards definition, such as a concept and accompanying algorithm for determining an overall margin estimate called channel operating margin (COM), mandatory forward error correction (FEC) encoding, and coexistence of two standards at the same aggregate bit-rate over backplane – 100GBASE-KP4 for multi-level signaling (PAM4) and 100GBASE-KR4 for the traditional PAM2 (aka NRZ). It also includes significant updates to test methods for transmitters and receivers, reflecting the new concepts.

The supported loss for the backplane channel is 35 dB at 12.9 GHz for 100GBASE-KR4 and 33 dB at 7 GHz for 100GBASE-KP4, a considerable feat in any technology and especially considering that ASIC vendors use CMOS processes.

The cabled version of the standard, 100GBASE-CR4, shares its methodology with the backplane standard (100GBASE-KR4) to a significant degree.

2. Wide Range of System Applications Challenges

The success of the 10 Gb/s Ethernet backplane standard resulted in a wide variety of Ethernet applications running on a wide range of designs, including some that are significantly different from the baseline objective, which was simply one meter of PCB made up of two line cards and a backplane. The goal for the 100 Gb/s backplane standard is to allow similar flexibility.

As with any technology stepping of this type, beyond just the bandwidth increase there also is an expectation to improve aggregate system throughput and port density. Lingering from some of

the original 10 Gb/s backplane objectives was the desire to support approximately 1 meter of backplane. There was also the need to support the market's now wider and more diverse offering with physical and design topology trade-offs. This contributed to a wide range of electrical and physical challenges. Some examples of challenges follow:

- Insertion loss at the fundamental frequency can reach 40 dB (die to die, 100GBASE-KR4).
- Response settling times range from 1 to 100 ns with an equivalent range of channel memory.
- Integrated PHYs and switches require longer package trace routing in the neighborhood of 30 mm, while other designs could present package trace routings as short as a few mm. Packages cannot be ignored at this signaling rate.
- Applications need tradeoff options for the use of low or high loss boards, conglomerate cable-PCB designs, low or high performance connectors, and mitigation choices for PCB vias as most appropriate to their design.

3. Backplane decisions – PAM2 or PAM4

Some contributions to the task force showed that backplanes can be designed to support PAM2[2] signaling at 25 GBaud, using low-loss materials and advanced manufacturing technology or shorter backplanes. PAM2 signaling at this speed had already been demonstrated in optical and high-bandwidth electrical links, and was considered a fast path to market. It was also appreciated that PAM2 signaling would facilitate commonality between electrical and optical interfaces.

However, other contributions stated that the materials and technology required for a PAM2 physical layer (PHY) aren't suitable for high volume and cost-sensitive products and expressed a desire to be able to upgrade existing systems and designs to the next speed. An alternative proposal was to use PAM4 signaling at about 13 GBaud, which requires a modest extension of the bandwidth required for existing 10 GBaud signaling. Though a PAM4 implementation did not exist at the time, the requirements were consistent with current digital signal processing technologies [1].

The conclusion of the task force was to recognize two different markets for 100 Gb/s backplane. The first market pertains to high-end systems designed appropriately for PAM2 signaling. The second market pertains to existing systems for which 100 Gb/s is a desirable upgrade and systems designed to use lower cost PCBs [2]. The corresponding PHYs are denoted

[2] Note that the PAM2 modulation method used in prior backplane Ethernet standards is often called NRZ (non return to zero), however the term NRZ also applies to the PAM4 signaling defined by the IEEE P802.3bj task force.

100GBASE-KR4, using PAM2 signaling, and 100GBASE-KP4, using PAM4 signaling, respectively.

4. Architecture

The IEEE P802.3bj draft standard specifies the following functional attributes for the 100GBASE-KR4 and 100GBASE-KP4 PHYs. First, the transmitter includes a feed-forward equalizer (FFE) modeled as a 3-tap finite impulse response (FIR) filter structure whose coefficients can be tuned during a training period using a training back-channel. Second, the data is encoded with a Reed-Solomon (RS) forward error correction code.

Two additional attributes are included for the 100GBASE-KP4 PHY to improve the efficiency and effectiveness of digital receivers. First, the PAM4 data is precoded with 1/(1+D) encoding to permit effective recovery of data using conventional digital processing techniques. Second, a deterministic block termination symbol is inserted by the transmitter every 46 PAM4 symbols. This block termination symbol permits the receiver to determine the value of the end points of a block of data and facilitates efficient design of enhanced block-based data detection algorithms such as MLSD [3].

4.1. Forward Error Correction

The IEEE P802.3bj draft standard defines a Forward Error Correction (FEC) function for the new copper cable and backplane Physical Layers to enable them to meet their application requirements. This FEC function is based on the Reed-Solomon (RS) error correcting code.

RS codes are well understood and perform well for both random and burst errors. Burst error correction capability is of particular interest for copper media since decision feedback equalization (DFE) is typically used to combat inter-symbol interference (ISI) introduced by the medium without undesirable noise enhancement. Burst errors may be observed at the output of the DFE since a decision error leads to a higher propensity to make mistakes detecting subsequent symbols.

RS codes operate over the Galois Field GF(2^m) where the m is the symbol size in bits. The encoder processes k message symbols to generate 2t parity symbols which are then appended to the message to produce a codeword of n=k+2t symbols. The decoder can then use this information to correct any combination of up to t symbol errors in a received codeword. For the purposes of this clause, a particular Reed-Solomon code is denoted RS(n, k).

The code RS(528, 514) is defined for use with 100GBASE-CR4 and 100GBASE-KR4, while RS(544, 514) is defined for use with 100GBASE-KP4. In both cases, the RS symbol size is m=10 bits. These choices considered trade-offs between coding gain, over-clocking to maintain consistent throughput with the overhead of the code, and added latency. The coding gain is the reduction in the signal-to-noise ratio (SNR) that can be accommodated while maintaining the

target probability of error. The ability of the interface to operate at reduced SNR is a key enabler for operation over high loss or high noise channels.

The overhead of the code, embodied by the parity check symbols, is mitigated by transcoding. 100 Gb/s Ethernet employs the 64B/66B block code which has an overhead of 3.125%.Groups of four 66-bit blocks are mapped into a 257-bit block to free 7 bits of overhead for use by the parity check symbols. In fact, for the RS(528, 514) code, the signaling rate of the interface is unchanged from 64B/66B encoding due to this reassignment of overhead. While the transcoding incurs some latency, it can be preferable to operating the interface at a higher signaling rate which can magnify impairments and compromise the coding gain.

Latency is a critical parameter for a number of applications. The latency added by the FEC function is mitigated in several ways. First, the FEC encoding and decoding is performed at the 100 Gb/s rate. Since the error correction latency is, at least in part, related to the time to receive and process an FEC codeword, there is a significant latency advantage to performing such operations at the higher speed. The codewords are encoded at 100 Gb/s and then striped across the four FEC lanes one 10-bit symbol at a time. The receiver aligns and serializes the four FEC lanes so that the decoder may process the codewords at the aggregate rate. This enables the operation of the FEC to incur less than 100 ns of latency per link.

If the channel performance is such that FEC is not required to achieve the target probability of error, the error correction operation may be bypassed to reduce the latency further and also to reduce the power dissipation. Even with error correction bypassed, the FEC decoding function still monitors the occurrence of errors and will inhibit the operation of the link when the error rate is high enough to compromise the integrity of the transcoded blocks.

Finally, it should be noted that performance measurements must be carefully considered for links protected by FEC. The RS codes included in the proposed standard can correct any combination of up to t symbol errors in any given codeword. If a codeword contains more than t errors, it may be difficult to discern which symbols are in error, so it is prescribed that the contents of the entire codeword be invalidated. The loss of any given codeword can affect multiple Ethernet frames (up to 9 under specific conditions) and one must consider how this relates to traditional bit error ratio (BER) requirements. An example is shown in Figure 1 for the case of the binary symmetric channel i.e. uncorrelated errors. Similar calculations can be performed for correlated errors such as burst errors at the output of a DFE.

The concept of frame loss ratio (FLR) was introduced as a better measure of the performance for these links. FLR is defined as the number of frames not received as valid divided by the total number of frames sent. A relationship between traditional BER requirements and FLR has been established which is then related to the probability of an uncorrected codeword. The relationship between the probability of error at the input to the decoder and the probability of an uncorrected

codeword can be calculated to enable the evaluation of the performance of components prior to the benefit of error correction.

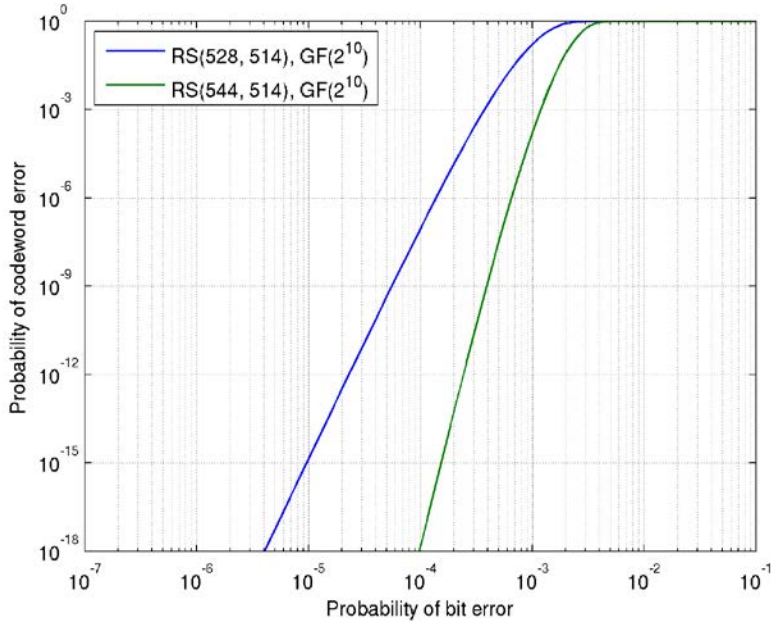


Figure 1 Reed-Solomon code performance

5. Channel Operational Margin – COM

5.1. The need for a unifying budget tool

A significant challenge to the task force's effort was the objective of supporting channels with up to 35 dB loss at 12.9 GHz, and the additional support for PAM4 signaling with its channel objective of 33 dB at 7 GHz. Broad market potential of the standard is reflected in the composition of the task force, which had representatives from a broad range of interests such as system, board, component, chip, and test equipment providers. The market scope envisioned during the development of the standard included higher volume servers, microservers, or network processors. The challenges lay in reconciling different constraints in different market segments under one or two (PAM2, PAM4) designs.

The answer to the challenge was to allow the trade-off between loss, reflections and crosstalk, as well as allowing the trade-off between channel permutations and chip capabilities.

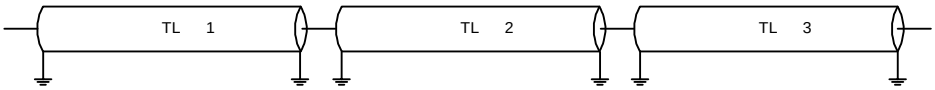
Put in another way, the effort needed a guiding "unifying budget" to make trade-off decisions.

5.2. The 10GBASE-KR example

A unifying budget in itself is not a new concept. For example the IEEE P802.3ap Backplane Ethernet Task Force attempted this with a set of frequency masks for insertion loss (IL), fitted

attenuation of the insertion loss, insertion loss deviation (ILD), return loss (RL), and insertion loss to crosstalk ratio (ICR). ICR is a useful methodology to allow the tradeoff between crosstalk and insertion loss, but it was not apparent how to budget and trade-off the rest of the constraints. Consider a channel with high ILD and low IL, low ICR but high ILD, just meeting all the specification, or the minimum of all specifications. This situation and many more combinations are common for the broad server and data center market. Transmitter and receiver specifications were also defined, but were not clearly linked with the channel specifications, and did not describe how chips are expected to cope with specified channel characteristics. Specifically, the expected receiver capabilities are not explicitly defined.

It is not always possible to easily distinguish between working and failing channels in the frequency domain, taking into account assumed receiver capabilities. As an example consider Figure 2, concatenating three interconnect sections a, b, and c, two of which are of high quality and one which has some reflective characteristics. The delay of the reflection relative to the main cursor may be highly dependent on the reflective section's location within the channel. In one case the reflection may reside within the region that the DFE can handle, while in the second case the reflection may arrive "too late" for the DFE to compensate.



	TL1	TL2	TL3
Case1	Smooth ISI section	Reflective section	Smooth ISI section
Case2	Smooth ISI section	Smooth ISI section	Reflective section

Figure 2 Example of three-section channels

Examining the insertion loss magnitudes of both interconnects in the frequency domain (see Figure 3) suggests that they are extremely similar. However, performing an end-to-end time domain simulation and examining the SNR margin at the Rx slicer (see Figure 4) running a 5-tap DFE receiver, we get a noticeable difference. The "Case 2" (blue) interconnect has a ~4 dB margin, while the Case1 (purple dashed) interconnect has only a 2.8 dB margin. The pulse responses of the two interconnects reveal that the reflections following the main pulse have different amplitudes and delays.

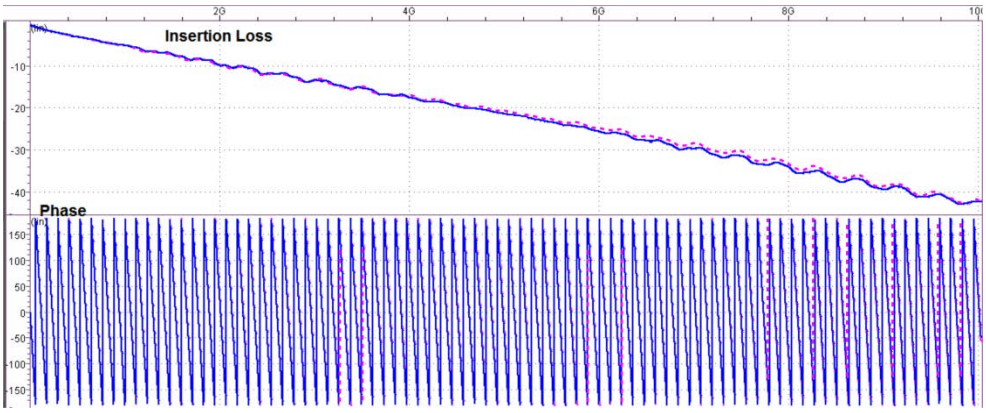


Figure 3 Comparison in frequency domain of the three-section channels

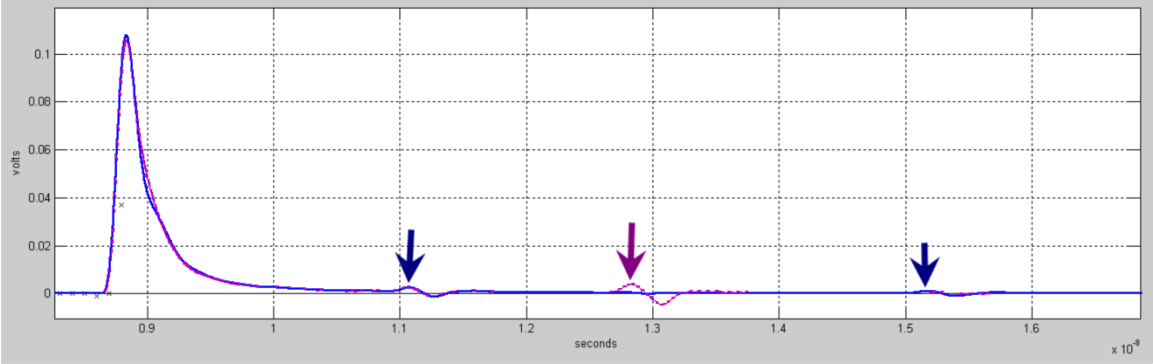


Figure 4 Time domain simulation of the two cases above. Dotted (purple) trace, Case 1; blue trace, Case 2.

Since a DFE is a non-linear device, it cannot be modeled in the frequency domain. Therefore, frequency domain analysis of an interconnect does not take into account the DFE effect in the receiver.

As demonstrated, the “through” interconnect frequency-domain magnitude analysis does not always convey the available margin. Crosstalk channels pose an additional problem to frequency domain analysis. A crosstalk channel has a well-defined time domain response which results in noise statistics that may be “spiky” or spread, and that may not be obvious in the frequency domain. Since frequency domain magnitude analysis does not yield an exact margin result, the combined effect of loss, ISI and crosstalk is not always captured accurately by the frequency-domain ratio of insertion loss to crosstalk (ICR).

To avoid over-constraining the design, at 10 Gb/s the IEEE P802.3ap Task Force did not make the channel specification normative. In spite of this, the informative specification did become a de-facto standard for channel performance. The availability of a specification – albeit an informative one – for the channel became a design imperative for a channel design.

Still the problem of budgeting the design persisted; based on the informative channel specification the imperative for minimum chip design capability was not as clear as desired by

some implementers of the standard. The “unifying budget” challenge facing the IEEE P802.3bj Task Force was the coordination between the components of the specifications; that is, between the transmitter, receiver, and channel. How does changing one part of a system affect other parts? Design imperatives that ripple through transmitter, receiver, and channel specifications need to be the foundation.

5.3. Addressing the unifying budget need

Moving into a time domain specification of the system opened the door for the IEEE P802.3bj Task Force to embrace the concept of a “channel operating margin” (COM) specification, which then provides the unifying budget of the whole system. Statistical signal analysis using channel scattering (S) parameters along with transmitter and receiver models has become an accepted method to determine link margin in the field, so it did appear to be a promising approach for making the trade-offs between all the specification parameters of the system. The challenge was to design this new simulation algorithm as one that would not have to rely on commercial simulators, since commercial simulators are proprietary; it would not be within IEEE 802.3 mandate to favor one commercial EDA simulator over another.

So the idea of a custom algorithm for the budget work was born. This became COM, a unified budgeting algorithm that is built on the concept of using the unit interval pulse response (aka single-bit response, or SBR).

The first agreement reached in the process of building the tool was the reference architecture, in which the transmitter and receiver have a certain ideal baseline behavioral capability regardless of actual design implementation. On the transmitter side a feed forward equalizer (FFE) with a pre-cursor, a post-cursor, and a main cursor tap, is specified for a compliant transmitter. On the receiver side the equivalent capability of an ideal DFE of a certain tap length is assumed. A continuous linear time equalizer (CTLE) was also added to the minimum equivalent capability of the receiver.

The COM method is based on victim and aggressor pulse responses. An established method [4] was used in which noise profiles are extracted from all the SBRs into a noise profile for the entire system of channels. All combinations of FFE and CTLE are tried to determine the best achievable performance. The COM description also includes a simple criterion, similar to the one proposed by Mueller and Müller [5], which defines the sample point in the equalized SBR. Finally, the decibel ratio of the voltage sample point over the noise at a specified BER becomes the COM figure of merit.

The fact that practical receivers are not ideal is addressed by requiring the calculated COM for a channel to be larger than a minimum threshold. This threshold represents a budget allowance for receiver implementation.

5.4. Jitter to Noise translation

It is usual to consider both noise immunity and jitter immunity when defining performance metrics. Some tools and specifications call for measuring both eye height and eye width. Equalization can be tuned to maximize one of these metrics at the expense of the other. A question that follows is how to weigh the two in order to create a single metric useful for defining the reference equalization procedure, and subsequently define the pass criteria for channel specification.

An additional complexity encountered in ISI-limited communication is that jitter effect is dependent on channel and equalization, since mathematically its effect depends on the derivative of the pulse response. Therefore, jitter assumptions can affect the optimal choice of equalization.

The approach taken in COM and receiver specifications related to it is to use the noise immunity as the main metric, and model the effect of transmitter jitter as additional noise terms which are used in the total noise calculations.

The total noise profile which is used to compute COM subsequently includes transmitter noise and jitter terms, ISI, crosstalk, and a constant-density noise floor representing unmeasured physical effects. The transmitter worst case noise and jitter are derived from specified transmitter parameters. Receiver noise and jitter are not explicitly accounted for in COM, and are considered to be part of the budget allowance for receiver implementation.

5.5. Package considerations

At the high signaling speed and tight noise budget required for 100 Gb/s backplane communication, performance becomes highly dependent on the packages that terminate the channel. Interconnect behavior would change dramatically between the “ideal” termination, the VNA measured termination, or the 3D tool extracted termination. Taking into account that an interconnect would “interact” with non-ideal loads located at its ends, the voltage transfer function of a signal passing through an interconnect was therefore defined in IEEE P802.3bj as in equation

$$H_{21} = \frac{S_{21}(1-\Gamma_1)(1+\Gamma_2)}{1-S_{11}\Gamma_1-S_{22}\Gamma_2-S_{21}S_{12}\Gamma_1\Gamma_2+S_{11}\Gamma_1S_{22}\Gamma_2} \qquad (1)$$

The above voltage transfer function takes into account reflections and re-reflections "between" the interconnect and the package return loss. Also see [6].

A plot generated by one implementation of COM (Figure 5) based on Annex 93A gives an example of the influence of the package return loss on the overall channel insertion loss. The data for the analyzed channel is shared on the IEEE P802.3bj Task Force web site[3].

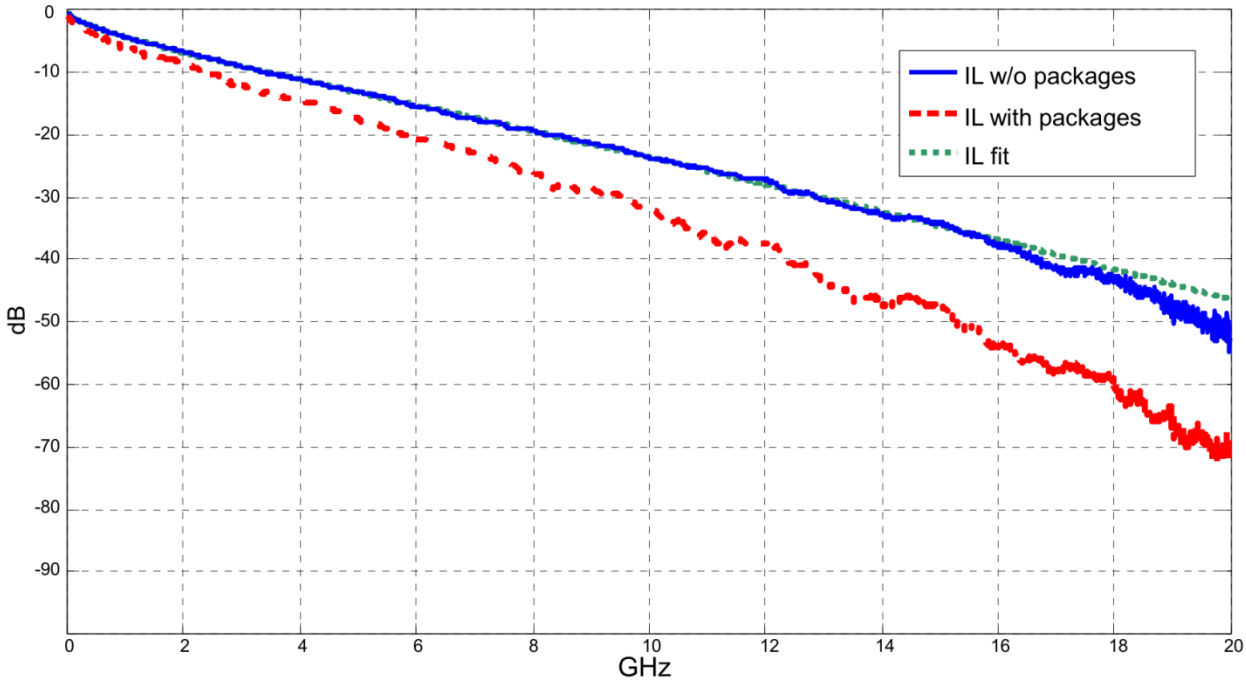


Figure 5 Influence of the packages' return loss on insertion loss. Dark blue solid line: Insertion loss w/o packages; Red line: insertion loss with packages; Green line: fitted insertion loss.

Realizing the above, it is erroneous to simulate the time domain behavior of an interconnect while neglecting the influence of non-ideal loads at its ends. Therefore, COM requires concatenation of a modeled package/die pad response to the base channel response. Hence the COM for a channel is a "die to die" figure of merit.

The support for a range of package trace lengths was an important consideration for some applications and chip vendors. Short package traces may have low insertion loss at frequencies up to the Nyquist frequency, but have poorer return loss. Long package traces on the other hand may have much higher insertion loss, but most likely have a better return loss.

In order to support a range of package sizes, two different test cases were included to account for two different SNR degradation factors:

[3] Available at http://www.ieee802.org/3/100GCU/public/ChannelData/ibm_11_0909/patel_03_0911.zip.

– a short package trace length is required to “interact” with reflective channels.
– a long package trace length is required to represent the loss one may have in a large package making it more susceptible to noise, higher loss interconnects and/or lower frequency crosstalk coupling.

A compliant channel is required to have sufficiently high COM results in both test cases.

A challenge existed regarding the representation of the package model. Given that the interconnect data may be supplied to the COM analysis at unknown frequency points, a mathematical representation of the package was provided to enable calculation of the behavior of the package at any frequency point according to the interconnect frequency resolution. Pure capacitive discontinuities were taken to represent the parasitics on the die side and board side of the package. A frequency domain representation of a lossy transmission line was used to account for the package trace connecting these two discontinuities. The resulting package model consists of a set of equations and parameters included in the standard.

5.6. Receiver test and COM

COM was tied into the receiver compliance test. In theory, any channel that passes COM can be used to test receiver compliance. The standard describes a method to determine how much noise needs to be added to the channel with a measured or derived COM and measured noise coupler characteristic, in order to create a calibrated receiver stress. In addition, the transmitter inputs for the COM calculation are determined from the transmitter compliance measurement of the actual transmitter used in the receiver test. Provisions within the standard describe how to make noise and amplifier adjustments which bring the transmitter down to minimum compliance.

The receiver compliance specification ties transmitter compliance and COM together, and does subsequently close the standard’s budget. COM, using parameters from all components of the link, presents a “unifying budget”.

5.7. Other applications of COM

Given the utility of COM for 100 Gb/s designs, one can expect COM to appear in other standards beyond 100GBASE-CR4, 100GBASE-KR4, and 100GBASE-KP4. And indeed, COM is now being incorporated into another standard, specifically the CAUI-4 chip-to-chip specification in the IEEE P802.3bm 40 Gb/s and 100 Gb/s Fiber Optic Task Force. In [7] it was demonstrated using COM that in order to support a 20 dB chip-to-chip channel at a BER of 10^{-15} for CAUI-4, a receiver capability of a 4-tap DFE or equivalent is required.

6. Fixturing for 25 Gb/s signaling

As data rates and frequencies of interest increase, it becomes more and more difficult to make accurate measurements on the transmitter and receiver. Frequency dependent loss and reflections

in test fixtures necessary to connect the components to test equipment begin to affect and potentially even dominate the result of measurements.

The first step in dealing with this problem is to standardize the test fixture used for measuring the parts. A common test fixture for receivers and transmitters in the 100GBASE-KR4 and 100GBASE-KP4 has been specified in terms of limits on return loss and insertion loss as a function of frequency. In a similar way the Host Compliance Board, the test fixture for measuring the 100GBASE-CR4, has limits on its return loss and insertion loss as a function of frequency. Return loss specification of the receivers and transmitters is specified in such a way as to take into account the frequency dependent attenuation of reflections in the test fixture (twice the insertion loss) and reflections added in the test fixture.

Note that the transmitter specification shows the effect of fixture loss especially well, see section 7 below.

7. Transmitter specification

Specification of the IEEE P802.3bj transmitters, beyond basic functionality, can be broken down into bandwidth and unequalizable perturbations.

7.1. Bandwidth

The bandwidth effect is measured as the ratio of linear fit pulse peak to steady-state voltage using a transmitter pulse response computed from the measured output of the transmitter while transmitting a PRBS pattern. The linear fit pulse method, first introduced in 2010 for the 40GBASE-CR4 specification, is beginning to replace eye masks as a way to specify transmitters in systems with heavy equalization since it clearly distinguishes between equalizable and unequalizable impairments.

7.2. Unequalizable perturbations in the transmitted signal

Unequalizable perturbations are measured several ways: most non-linearity and ISI beyond the range of equalization are measured as fitting error during the linear fit pulse extraction. Non-linearity, which results in unequal spacing of PAM4 levels, is measured with a special pattern which holds the output in one state long enough for the signal to settle and be measured.

Previously used jitter measurement methods do not accurately separate out the effect of ISI on transition timing (known as data-dependent jitter or DDJ), although DDJ can be largely mitigated by equalization. As a result, transmitter jitter specifications of previous standards were loosened to allow some DDJ. With the tight budget available for 100 Gb/s Ethernet, jitter specifications had to become tighter. Since ISI is already limited by the fitted pulse specification, it is desirable to uncouple its effect from the measurement to allow accurate measuring of other effects. The solution is an improved measurement method, in which jitter for the two PAM2 specifications is

measured on a PRBS pattern using only two well-isolated transitions out of the pattern. A histogram is generated of the number of samples as a function of jitter deviation, and a cumulative distribution function (CDF) generated from the histogram. Points near the ends of the measured CDF are fitted using an Inverse Complementary Error Function and an extrapolation of the fitted lines is used to determine the effective bounded uncorrelated jitter and effective jitter defined in terms of the fit.

Since there are 12 different possible transitions in PAM4 and they are of different amplitudes and have different thresholds, using a PRBS pattern for measuring would be complicated. Therefore we used a two-level clock like pattern switching between the highest and lowest level. The data is fitted to a dual-Dirac model and both random and deterministic parts are used for compliance. This method also prevents ISI from affecting the jitter measurement results.

Transmitter noise is measured on a repeating waveform, such as a multi-bit square wave or PRBS9, which includes an occurrence of at least 8 consecutive identical bits. Noise is measured in a region of the unchanging bits where the output change due to inter-symbol interference is minimal. The variation in level over a number of samples represents transmitter noise which could be caused by power supply noise, thermal noise, crosstalk, etc. This noise is root sum of squares added to the linear fit pulse fitting error to get the total noise RMS, which is compared to the transmitter peak pulse to get transmitter signal to noise ratio.

7.3. Conformance of Transmitter equalizer to specification

The minimum required pre- and post- cursor equalization at the Transmitter’s maximum setting and the step size for the FFE tap weights are explicitly specified. The cursor values are measured by using the same linear fit pulse method used in determining bandwidth. A base pulse is found with maximum amplitude cursor and zero pre- and post- cursors. Any other set of FFE tap cursor weights are found by fitting the measured linear fit pulse to the sum of three weighted, baud-spaced, versions of the base pulse. The weights which produce the best fit are the measured cursor values.

8. Summary and Conclusion

The IEEE P802.3bj Task Force, being challenged in many divergent ways and markets, ended up with several notable decisions. Mandatory FEC encoding was adopted in support of the long reaches. The supported insertion loss at the fundamental frequency is higher than in prior commercial backplanes. The coding gain for the FEC allowed for an extra 5 dB of channel insertion loss which effectively increased channel reach. COM was adopted for the channel specification and receiver tests, resulting in tight budgeting of noise and reflections and thus allowing trade-offs between the highest loss, quiet channel and lower loss, noisier channels. A component of COM computation was developed to precisely and efficiently account for the practical fact that most chips are mounted in packages. In addition to COM minimizing guard

bands, the adoption of PAM4 as an alternate signaling architecture widens the market potential even further. With expected ratification of the standard in 2014, the authors are wishing the standard at least as much adoption as there was in its predecessors.

Finally, we would like to thank to all of those who directly or indirectly contributed to the work of the IEEE P802.3bj Task Force.

Bibliography

[1] H. Frazier, V. Parthasarathy and J. Wang, "Feasibility of 100 Gb/s Operation on Installed Backplane Channels," *IEEE 802.3bj May 2011 (http://www.ieee802.org/3/100GCU/public/may11/parthasarathy_01a_0511.pdf).*

[2] P. Patel, J. Goergen and H. Frazier, "Two Backplane Markets, Two Backplane Channels, Two Signaling Schemes, Two PHYs," *IEEE 802.3bj March 2012 (http://www.ieee802.org/3/bj/public/mar12/patel_01_0312.pdf).*

[3] D. Dabiri, "MLSD/DFE Based Transceivers with a Partially Terminated Trellis for IEEE 802.3bj 100Gb/s Backplane and Copper Assemblies," *IEEE 802.3bj January 2012 (http://www.ieee802.org/3/bj/public/jan12/dabiri_01b_0112.pdf).*

[4] A. Sanders, M. Resso and J. D'Ambrosia, "Channel Compliance Testing Utilizing Novel Statistical Eye Methodology," *DesignCon 2004.*

[5] K. H. Mueller and M. Müller, "Timing Recovery in Digital Synchronous Data Receivers," *IEEE Transactions on Communications,* vol. COM.24, no. 5, pp. 516-531, May 1976.

[6] R. Mellitz and L. H. A. M. C. Arsi, "Adding a Simple Package Model to the Channel Response," [Online]. Available: http://www.ieee802.org/3/bj/public/jan13/mellitz_3bj_01b_0113.pdf.

[7] R. Mellitz and A. Ran, "Proposal for CAUI-4 Package, Reference Receiver, and COM Parameters In Support of Comment 105,106, 107,108,111, 112," *IEEE 802.3bm, September 2013 (http://www.ieee802.org/3/bm/public/sep13/mellitz_01_0913_optx.pdf).*