Analyzing Signal Integrity at Extreme Data Rates

By Ransom W. Stephens, Ph.D.

TABLE OF CONTENTS

It's been a decade since the last major increment in networking data rate. In 2001 the ceiling was raised to 10 Gb/s and, other than a few starts and stops, it has stayed there until now. Finally, in the second decade of the high tech millennium, 100 Gb/s is coming. This is where ultra high rate serial data technology converges with mega high rate fiber optic technology. Optical engineers and electrical engineers alike need to know some of the other's trade to get ready.

In this paper, we keep a close eye on the first extreme data rate technologies that will soon come to market: 40 and 100 Gigabit Ethernet, but focus on the exceptional technologies and development challenges that will face every standard above 10 Gb/s. We'll look closely at optical phase shift key modulation, skew and crosstalk and through it all we'll keep our collective fingers on the pulse of the physical layer, the Bit Error Rate.

1.0 Quick Look at 40 GbE and 100GbE

The 100 Gb/s technology coming to market is 100 GbE (using the standard nomenclature: GbE ~ Gb/s Ethernet). In 100 GbE, 100 Gb/s is reached through a combination of parallel and serial technologies; the highest data rate of a single carrier is 25 Gb/s. Figure 1 shows the major components of 100 GbE. The components are substantially the same for 40 GbE. In Figure 1a, a single fiber carries four different wavelengths of light - a technique called Wavelength Division Multiplexing (WDM)—and in Figure 1b, each wavelength is carried on a separate fiber in parallel.

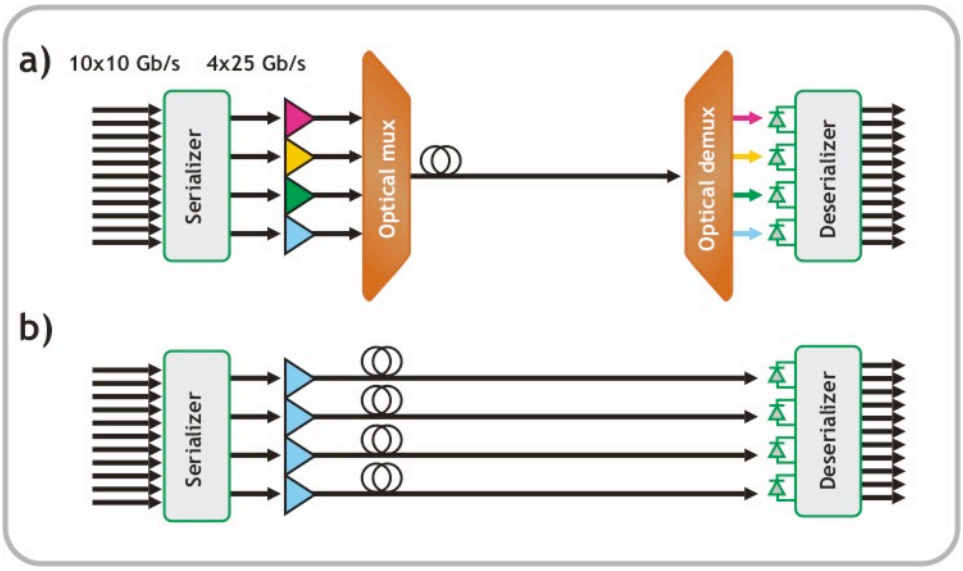

Figure 1: High level diagram of 100 Gb/s Ethernet in (a) with wavelength division multiplexing and in (b) with parallel fiber optics.

The topological options include
• 40 GbE - four parallel transmission lanes at 10 Gb/s (4×10).
• 100 GbE - ten parallel lanes at 10 Gb/s (10×10) or 4 lanes at 25 Gb/s (4×25).

On the optical side there are
• Short range multimode fiber options: up to 100 m for both 4×10=40 GbE and 10×10=100 GbE.
• Long haul single-mode fiber options: up to 10 km for 4×10=40 GbE and 4×25=100 GbE.
• Plus a very long haul single-mode fiber option: up to 40 km for 4×25=100 GbE.

The fundamental difference between “single-mode” and “multimode” fibers has to do with their character as dielectric waveguides. The optical signal in a single-mode fiber excites only the fundamental mode whereas in a multimode fiber it can excite several different modes of oscillation each of which has a different group delay resulting in “modal dispersion,” hence the diminished range capacity.

On the electrical side, the signals are either 10 or 4 lanes each carrying 10 Gb/s. In all cases they must maintain integrity over up to 10 m of copper and, for 40 GbE, 1 m of backplane.

As the technology matures it’s reasonable to expect short range 100 GbE to be deployed in data-centric applications using WDM optics. Long haul 100 Gb will take longer to develop because it requires more complex transmitters, modulators, receivers, and post-processing equalization and compensation schemes. In any case, all of the usual signal integrity problems will persist, though some are so greatly exacerbated by the data rates that old problems will require new solutions.

Moving from left to right in Figure 1, the input to the serializer consists of 4 or 10 differential lanes at 10 Gb/s. The signals are aligned, serialized and the data is encoded with the 64B/66B scheme. The transmitters consist of lasers at continuous full power with the digital signal imposed through indirect modulation. The optical signals are either multiplexed, for example, by an arrayed waveguide grating into a WDM signal on one fiber, Figure 1a, or are transmitted individually on parallel fibers, Figure 1b.

At the receiver, the optical signals are tidied up by Chromatic Dispersion (CD) and Polarization Mode Dispersion (PMD) compensators as necessary. In Figure 1a the WDM signals are demultiplexed into four individual 25 Gb/s optical signals. These sub-rate optical signals are then demodulated and converted to ten 10 Gb/s electrical signals.

The protocol dictates how signals are constructed at the transmitter and assigned to each lane, whether multiple wavelengths on one fiber or single wavelengths on multiple fibers, and then decoded at the receiver. Since there is a great deal of serializing/deserializing up and down to different rates, the Physical Medium Attachment (PMA) is called a “gearbox.”

2.0 The Overriding Requirement: BER < 10^{-12}

The Ethernet dogma is the same for 100 GbE as it has been for every other Ethernet generation: *The system will operate at a Bit Error Rate that shall not exceed 10^{-12}.*

If the system BER meets this criterion, no other signal integrity issue need be considered. It is therefore worthwhile to look closely at how BER is measured. Figure 2 shows how a Bit Error Rate Tester (BERT) works. There are three primary components: the pattern generator, the clock, and the error detector. The pattern generator produces a repeating data stream whose logic transition times are determined by the clock. In most applications, the Device Under Test (DUT) accepts the pattern, processes it in some way, and emits another pattern. In the example depicted in Figure 2, the DUT recovers the clock signal from the data and provides it to the error detector; this is just one of many possible configurations.

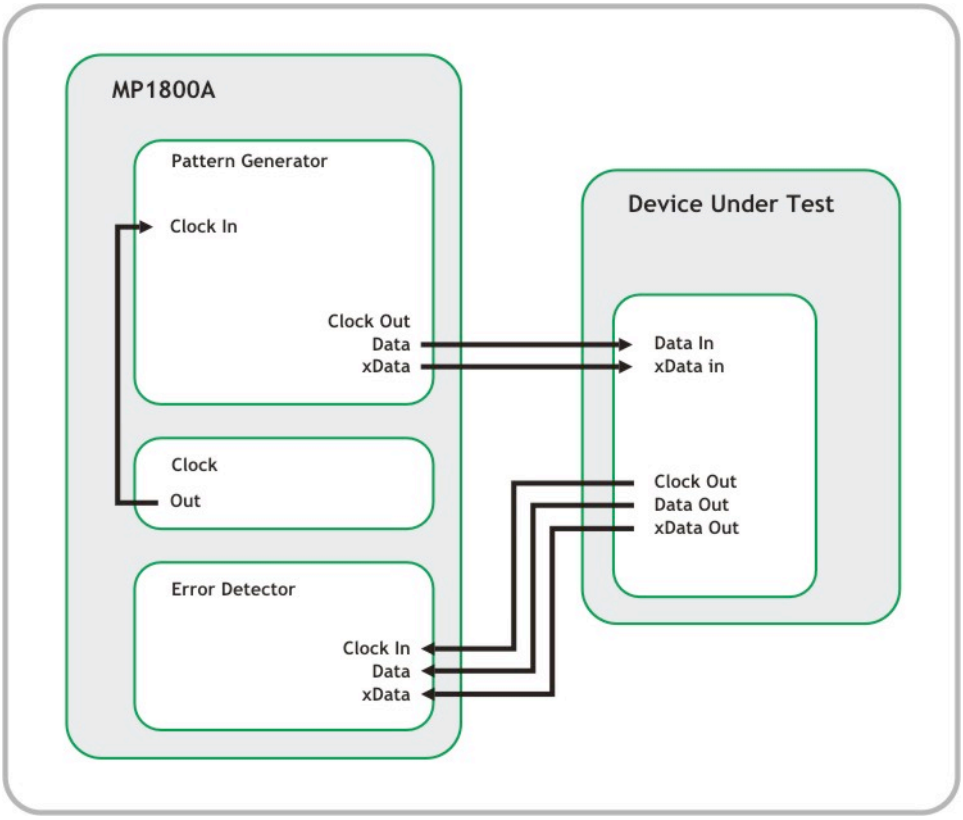

Figure 2: Bit Error Rate Testing

The error detector receives the processed pattern. It must know what pattern to expect in order to synchronize with its input. Once the error detector has synchronized, it knows what the values of incoming bits ought to be and can count errors.

The BER is defined as:

$$\mathrm{BER} \equiv \lim_{N \to \infty} \frac{N_{Errors}}{N}$$

where N_{errors} is the number of errored bits and N is the total number of bits transmitted.

Many qualities distinguish a high performance BERT. On the pattern generator side, we need nearly ideal square wave performance; that is, exceptionally fast rise/fall times that can be modified with external filters to emulate specific technologies. We also require negligible distortion, noise and jitter on the transmitted signal. A wide voltage swing is imperative for driving long haul transmitters.

Figure 3 shows the eye diagram of a 25 Gb/s signal generated by an Anritsu MP1800A. It offers 3.5 V peak-to-peak for single-ended and 7 V peak-to-peak for differential applications.

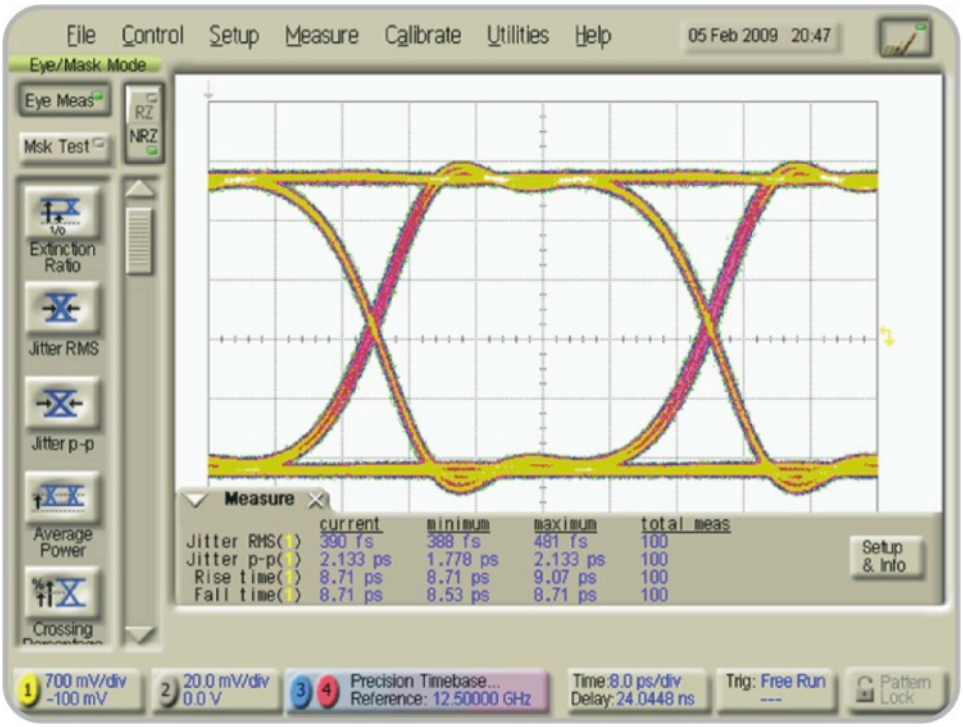

Figure 3: The eye diagram of a 25 Gb/s signal generated by an Anritsu MP1800A.

For 40 and 100 GbE testing we need multiple pattern generators capable of transmitting large user-provided data patterns whose skew is easily controlled - the MP1800A has adjustable skew delay of 128UI in 1 mUI (i.e., 1/1000th of a Unit Interval or bit period) steps up to 14 Gb/s and 2 mUI up to 28 Gb/s.

The clock source should be extraordinarily stable with plenty of divided output options. The clock must be capable of modulation so that calibrated levels of different types of jitter can be applied to generated signals; not just spread spectrum clocking, but random and periodic jitter - all necessary for stressed receiver tolerance testing.

At the other end, the Error Detectors (ED) need to have good sensitivity, phase margin, and they must rapidly synchronize with incoming data. The most important specification, as data rates grow and eye diagrams shrink, is the ED sensitivity. The ED sensitivity is the minimum voltage difference between logic levels necessary for the ED to be able to distinguish 1's from 0's. As this paper goes to press, the MP1800A's 10 mV sensitivity leads the industry.

Of course the BERT should be equipped with plenty of automation software for standard analyses of eye margin, Q-factor, bathtub plots, analysis of ISI and so forth.

2.1 Two Obvious Problems: Skew and Crosstalk

With the 40 GbE and 100 GbE parallel channels carrying data at high rates, two obvious problems jump out: skew and crosstalk. Generally, skew is the difference in propagation time between parallel lanes. It's caused by variations in path length, whether on fibers, PCB traces, or cables, and differences in propagation times on integrated circuits. Crosstalk is essentially electromagnetic interference (EMI) between neighboring electrical data lanes. Optical signals don't experience EMI, though as their power increases, they begin to affect the permittivity of the fiber which results in similar phenomena.

3.0 Physical Coding Sublayer Skew

The structure of the data itself presents two challenges to the receiver. Both occur when the transmitted data includes long consecutive repetitions of identical bits. A string of identical logic levels over a time interval longer than or comparable to the RC time constant of the receiver's AC-coupling causing baseline wander. If it is a long string of logic 1's, then it wanders high, if 0's then it wanders low. In either case, the vertical position of the receiver's sampling point, that is the slice threshold, moves away from the optimal position and the system BER increases. To prevent baseline wander, there must be an equal number of 1's and 0's over the AC coupling time scale. In other words, the fraction of logic 1's, or mark density, should be ½ over a hundred bit periods or so.

The second challenge has to do with clock recovery: the more transitions between logic levels, the easier it is for a clock signal to be recovered from and locked to the data (embedded clock technology is also called clock forwarding). If the clock drifts, then the time-delay position of the sampling point drifts and the BER increases. To assure that the frequency of logic transitions is sufficient for a receiver to reconstruct an accurate clock from the data we have to maintain a minimum transition density. The transition density is the ratio of logic transitions, 1 → 0 and 0 → 1, to the number of bits in the data.

The data is 64B/66B encoded and scrambled at the transmitter. The encoding technique maps 64 bits of data onto 66 transmitted bits in a way that assures an equal number of 1's and 0's over any two 66 bit symbols. The data is then scrambled to provide sufficient transition frequency for clock recovery. The process guarantees at least one logic transition in every 66 bits. Scrambling decreases the probability of getting long strings of identical bits but does not eliminate it.

The potential for long strings of consecutive identical bits (CID bits) is addressed by using the standard 2^{31}-1 Pseudo-Random Binary Sequence (PRBS31) test pattern. PRBS31 is over 2 billion bits and presents a test equipment challenge that requires use of a complete BERT solution. Not only is it impossible for an oscilloscope, regardless of memory depth, to analyze the full implications of such long signals, but at data rates of 25 Gb/s and higher, generating and transmitting the pattern itself is difficult. To meet compliance specifications, make sure that your equipment generates true PRBS even at extreme data rates.

Assignment of data blocks to parallel lanes and data striping, including periodic insertion of an alignment marker, is performed at the Physical Coding Sublayer (PCS). PCS encoding provides the necessary organization for data on different parallel lanes to be discriminated and reassembled at the receiver. The alignment marker enables the receiver to tolerate substantial skew between parallel lanes. The maximum permitted PCS skew is 180 ns† for all 40 and 100 GbE variations.

To test the gearbox, we need to determine the tolerance of full-rate multilane signals to PCS skew. A nice technique for this is shown in Figure 4 for the 4×25 Gb/s flavor of 100 GbE. For the transmit gearbox, parallel BERT pattern generators drive each lane. The gearbox multiplexes the lanes into the high rate signals that drive the optical transmitters. The high rate signals are each read by BERT error detectors. The ED's synchronize on the data and measure the BER. The key instrument feature for this analysis is the ability to control the timing of the parallel pattern generators. The time-delay of each lane relative to the others is increased until the BER is larger than 10^{-12}. Unlike the effect of random noise and jitter, where the BER tends to rise gradually with increasing noise, skew usually exhibits a threshold-like effect. As the skew is increased, the BER tends to remain low until the skew reaches the tolerance edge when the BER increases dramatically. The skew is then backed off a miniscule amount and the BER drops back to its stable value; this time-delay is the skew tolerance.

To test the receiver gearbox, the four 25 Gb/s signals are demultiplexed into 10 Gb/s components and each output is read by an error detector. Again, the skew is increased until the BER abruptly increases at the skew-tolerance level.

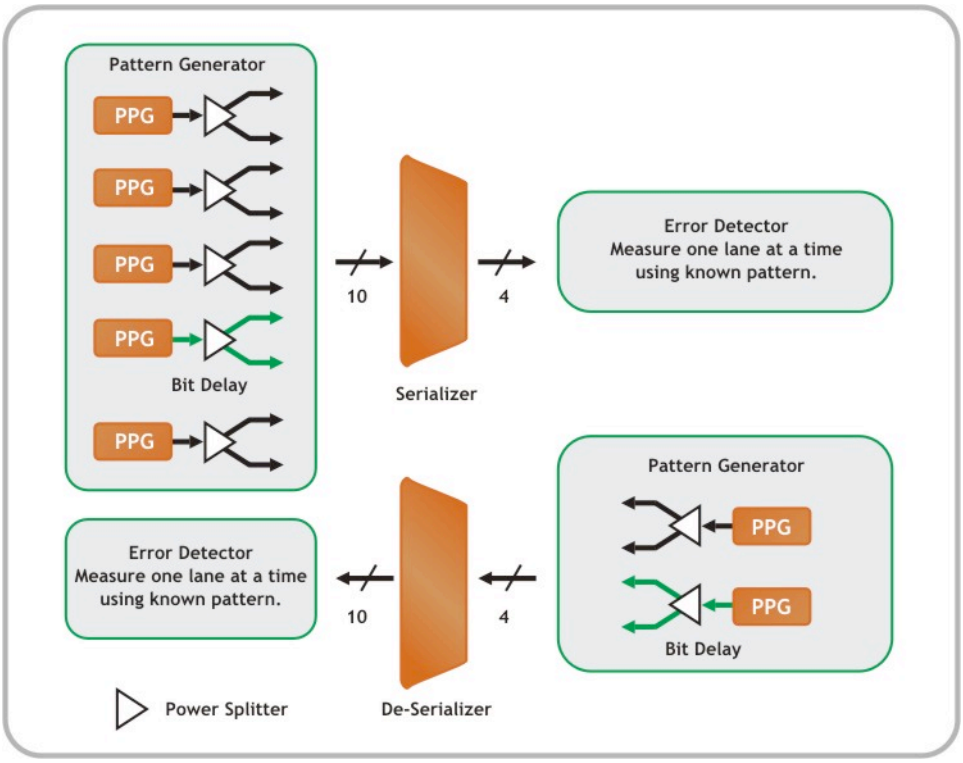

Figure 4: Configuration for measuring PCS skew tolerance of 4×25 Gb/s 100 GbE.

4.0 Optical DQPSK Modulation and I/Q Skew

Modulation of the optical signal presents several intriguing challenges. The standard on-off, Morse code-style, non-return to zero (NRZ) formatting of digital optical signals suffers as rates increase: as pulse widths get smaller, the optical bandwidth increases. As the optical bandwidth increases, so does the minimum possible wavelength spacing. Increased optical bandwidth also exacerbates both Chromatic and Polarization Mode Dispersion (CD and PMD). Plus, the increased power during high logic levels aggravates nonlinear interactions within the fibers and generates noise.

The most intriguing approach to the problem is optical Phase Shift Keying (PSK) modulation. Keep in mind that the data rate is still tiny compared to the optical frequency: 100 Gb/s, which is a 10^{11} Hz data rate compared to 10^{14} Hz optical frequencies (but only three orders of magnitude to go!). The optical phase is modulated but not wavelength by wavelength as it is for RF and microwave PSK modulation.

Optical Differential Phase Shift Key (DPSK) modulation consists of reversing the optical phase of the carrier upon certain logic transitions, Figure 5. The phase shift is achieved by use of a Mach-Zehnder interferometer. First, the carrier beam is split. The two resulting beams propagate through separate legs of the interferometer and are recombined coherently at the output. Starting with a logic 0, if there is no logic transition, that is 0 → 0, then the optical path length of both legs of the interferometer are varied a half wavelength (or, equivalently, odd multiple of half wavelengths). The recombined beam then has the opposite phase after the logic transition that it had before the logic transition. If there is a 0 → 1 logic transition, then the optical path length is not changed and the beams are recombined with no net phase change. For 1 → 1 transitions, the phase is also left the same; and for 1 → 0, the phase is reversed.

The system makes demodulation wonderfully straightforward: at the receiver, the incoming beam is split into two beams; one beam is delayed by a single bit period with respect to the other. When the beams are recombined, logic 1's experience constructive interference, logic 0's destructive interference, and the result is the convenient on-off NRZ format.

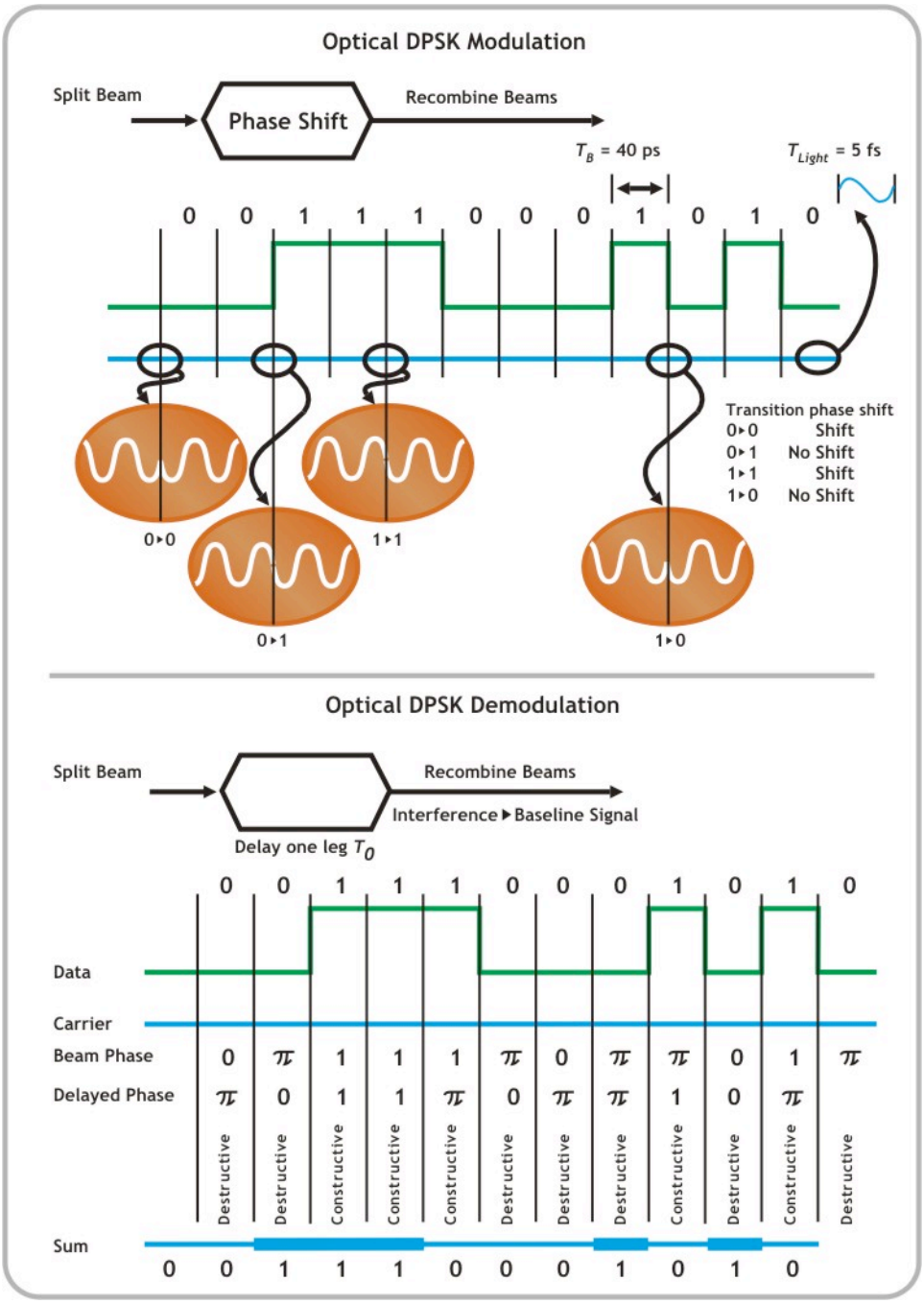

Figure 5: Optical Differential Phase Shift Key modulation and demodulation for a 25 Gb/s signal.

DPSK is extended to Differential Quadrature Phase Shift Keyed (DQPSK) modulation in the usual way so that the data rate can be maintained while halving the symbol rate. The result is a narrower optical spectrum that can tolerate more dispersion and closer channel spacing. Two bits are encoded in each symbol by adjusting the phase of the carrier sequentially. Zero shift for 00, π/2 for 10, -π/2 for 01, and π for 11 by applying I and Q to two separate modulators. The demodulator for optical DQPSK signals consists of two matched DPSK demodulators with phase offsets of ±π/4.

An even more novel approach is Dual Polarization Quadrature Phase Shift Keying (DP-QPSK) where I and Q signals are applied to each of the two orthogonal polarization states, decreasing the symbol rate by another factor of two with all attendant benefits - but at the expense of the complicated circuitry necessary to distinguish the two polarization states at the receiver and increased potential aggravation from polarization mode dispersion.

Since both I and Q signals must be synchronized in three stages—PCS coding, pattern alignment, and skew - I/Q skew tolerance is a key test of the modulation system. In Figure 6, two synchronized BERT pattern generators drive I and Q. The modulator generates the full rate DP-QPSK signal which is then demodulated back into I and Q and read by the BERT error detector. The phase between I and Q is increased until the BER starts to exceed 10^{-12}, to yield I/Q skew tolerance.

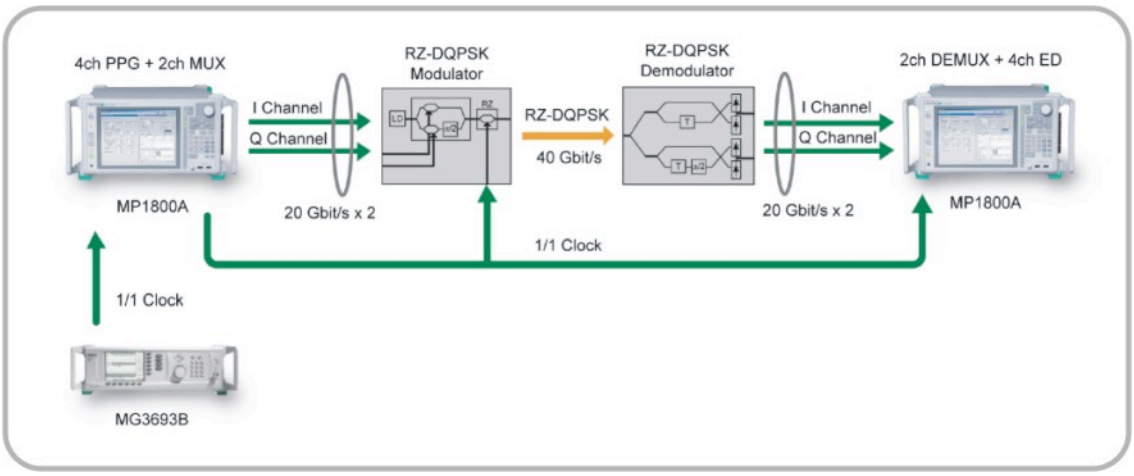

Figure 6: Setup for analyzing I/Q skew tolerance.

5.0 Crosstalk

A 100 Gb/s signal has a bit period of 10 ps and a 25 Gb/s signal has a 40 ps period - there isn't much time for the eye to open. Consequently, extreme data rates require lightening fast rise and fall times. The closer a digital signal approaches the square-wave limit, the more high frequency harmonics are introduced. The high frequencies cause ringing at the slightest impedance mismatch, but that's not the worst of it. Fast rise and fall times generate violently changing electric fields, and changing electric fields are precisely the cause of electromagnetic radiation. It is a breeding ground for crosstalk.

In crosstalk jargon, we think of victims and aggressors. The victim is the signal that is degraded by the electromagnetic field of the aggressor. If we think of separate traces on PCB, then the coupling mechanism is mutual capacitance and inductance. When the aggressor signal makes a logical transition, it fires a burst of radiation which permeates the circuit board and excites interfering currents on the signal trace.

To analyze crosstalk on 40 and 100 GbE systems it is important to keep in mind that the same clock governs data transmission on every parallel lane; this means that the victim and aggressor are frequency locked. Therefore, the noise signature on the victim signal is locked to a specific time-delay in the victim's eye diagram, as shown in Figure 7. In situations where the victim and aggressor are not frequency locked, the crosstalk signal varies randomly across the phase of the victim and is easily mistaken for random noise.

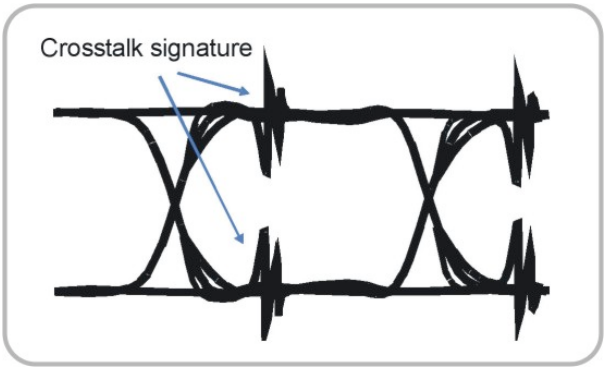

Figure 7: Eye diagram exhibiting crosstalk with victim and aggressor on a common clock.

The amplitude of crosstalk noise on the victim scales with the rise/fall time of aggressor transitions, but its width scales inversely with aggressor rise/fall time; that is, the slower the rise/fall time, the less intense the emitted radiation and so the smaller the crosstalk amplitude. However, the slower the rise/fall time, the longer the duration of emitted radiation and so the wider the degradation of the signal.

Crosstalk penalty is a measure of the extent to which crosstalk is a threat to the system BER. To measure crosstalk penalty we have to be able to transmit multiple frequency-locked signals and to vary the time-delay between those signals. With the setup depicted in Figure 8, we can measure the BER while varying the phase between the victim and aggressor. When the phase combines with the propagation delay between the two lanes so that the crosstalk distortion is in the crossing point of the victim eye diagram (since bits are sampled at the center of the eye, far from the crossing point), the effect on the BER is minimized. Conversely, when the phase plus propagation-delay aligns the crosstalk distortion with the center of the eye, the BER peaks.

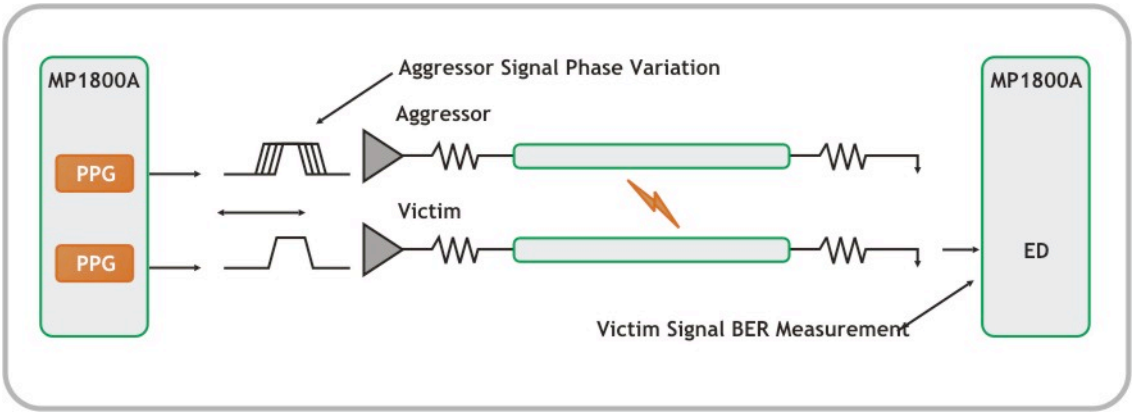

Figure 8: Crosstalk penalty analysis setup.

The question is: How does crosstalk affect the BER?

Crosstalk penalty is measured by attenuating the victim signal until it generates a steady low BER that is just high enough to measure in a few seconds, like 10^{-7} or so. Then the relative delay of the aggressor is varied. The BER is measured at each delay and we get a plot like that in Figure 9. In this example, the maximum BER of 2×10^{-6} occurs at a delay of 0.6 UI and the minimum BER, 3×10^{-8}, at a delay of 0.1 UI. The resulting crosstalk penalty is 18 dB.

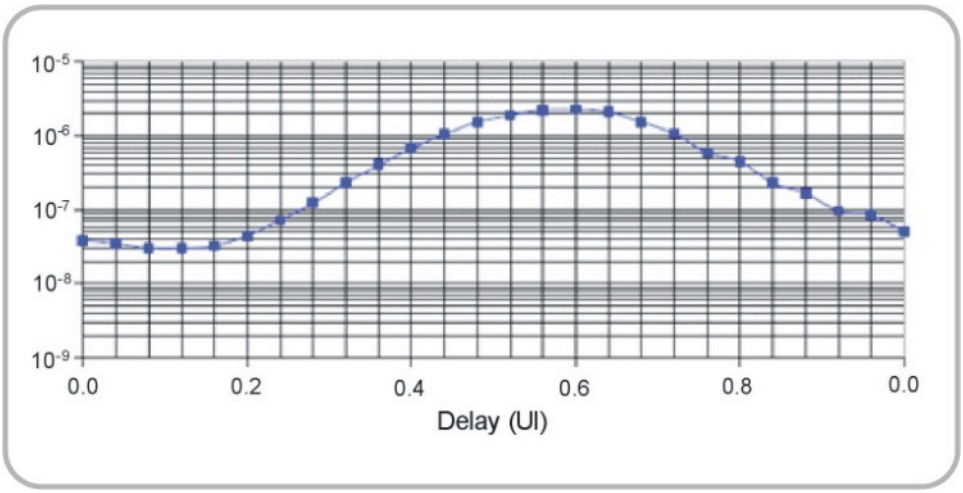

Figure 9: Crosstalk penalty for 10 Gb/s transceiver.

6.0 Conclusion

Multi-core processor technology and networked storage *as well as* IPTV and eCommerce are all driving bandwidth demand. It is inevitable that high speed serial technologies like the advanced generations of PCIe, SAS, SATA, FibreChannel, DisplayPort, RapidIO, etc. will converge with the technologies that move huge chunks of data across long distances. What was once the domain of SONET/SDH is now dominated by Ethernet. The result is 40 and 100 GbE - technology that combines the highest data-rate serial and parallel technologies while blurring the distinction between optical and electrical networks.

Two major innovations will light the way - optical DQPSK technology (with DP-QPSK waiting in the wings) to deal with optical dispersion, and PCS encoding to deal with skew. More innovations are certain to come and every one of them will bring new signal integrity analysis challenges.

In every case, the performance bottom line will be an understanding of the Bit Error Rate. A Bit Error Rate Tester will always be the key analysis tool to verify BER compliance. For 40 and 100 GbE, this means BER < 10^{-12}. Anritsu's MP1800A Signal Quality Analyzer, with its multiple channels, adjustable skew, and industry leading error detector sensitivity, is the only BERT capable of making measurements with the necessary precision for the extreme data rates we now face.

Anritsu

Specifications are subject to change without notice.

Anritsu Corporation
5-1-1 Onna, Atsugi-shi, Kanagawa, 243-8555 Japan
Phone: +81-46-223-1111
Fax: +81-46-296-1264

• **U.S.A.**
Anritsu Company
1155 East Collins Blvd., Suite 100, Richardson,
TX 75081, U.S.A.
Toll Free: 1-800-267-4878
Phone: +1-972-644-1777
Fax: +1-972-671-1877

• **Canada**
Anritsu Electronics Ltd.
700 Silver Seven Road, Suite 120, Kanata,
Ontario K2V 1C3, Canada
Phone: +1-613-591-2003
Fax: +1-613-591-1006

• **Brazil**
Anritsu Eletrônica Ltda.
Praca Amadeu Amaral, 27 - 1 Andar
01327-010-Paraiso-São Paulo-Brazil
Phone: +55-11-3283-2511
Fax: +55-11-3288-6940

• **Mexico**
Anritsu Company, S.A. de C.V.
Av. Ejército Nacional No. 579 Piso 9, Col. Granada
11520 México, D.F., México
Phone: +52-55-1101-2370
Fax: +52-55-5254-3147

• **U.K.**
Anritsu EMEA Ltd.
200 Capability Green, Luton, Bedfordshire, LU1 3LU, U.K.
Phone: +44-1582-433200
Fax: +44-1582-731303

• **France**
Anritsu S.A.
16/18 avenue du Québec-SILIC 720
91961 COURTABOEUF CEDEX, France
Phone: +33-1-60-92-15-50
Fax: +33-1-64-46-10-65

• **Germany**
Anritsu GmbH
Nemetschek Haus, Konrad-Zuse-Platz 1
81829 München, Germany
Phone: +49-89-442308-0
Fax: +49-89-442308-55

• **Italy**
Anritsu S.p.A.
Via Elio Vittorini 129, 00144 Roma, Italy
Phone: +39-6-509-9711
Fax: +39-6-502-2425

• **Sweden**
Anritsu AB
Borgafjordsgatan 13, 164 40 KISTA, Sweden
Phone: +46-8-534-707-00
Fax: +46-8-534-707-30

• **Finland**
Anritsu AB
Teknobulevardi 3-5, FI-01530 VANTAA, Finland
Phone: +358-20-741-8100
Fax: +358-20-741-8111

• **Denmark**
Anritsu A/S
Kirkebjerg Allè 90, DK-2605 Brøndby, Denmark
Phone: +45-72112200
Fax: +45-72112210

• **Spain**
Anritsu EMEA Ltd.
Oficina de Representación en España
Edificio Veganova
Avda de la Vega, n° 1 (edf 8, pl 1, of 8)
28108 ALCOBENDAS - Madrid, Spain
Phone: +34-914905761
Fax: +34-914905762

• **Russia**
Anritsu EMEA Ltd.
Representation Office in Russia
Tverskaya str. 16/2, bld. 1, 7th floor.
Russia, 125009, Moscow
Phone: +7-495-363-1694
Fax: +7-495-935-8962

• **United Arab Emirates**
Anritsu EMEA Ltd.
Dubai Liaison Office
P O Box 500413 - Dubai Internet City
Al Thuraya Building, Tower 1, Suit 701, 7th Floor
Dubai, United Arab Emirates
Phone: +971-4-3670352
Fax: +971-4-3688460

• **Singapore**
Anritsu Pte. Ltd.
60 Alexandra Terrace, #02-08, The Comtech (Lobby A)
Singapore 118502
Phone: +65-6282-2400
Fax: +65-6282-2533

• **India**
Anritsu Pte. Ltd.
India Branch Office
3rd Floor, Shri Lakshminarayan Niwas, #2726,
HAL 3rd Stage, Bangalore - 560 038, India
Phone: +91-80-4058-1300
Fax: +91-80-4058-1301

• **P.R. China (Hong Kong)**
Anritsu Company Ltd.
Units 4 & 5, 28th Floor, Greenfield Tower, Concordia Plaza,
No. 1 Science Museum Road, Tsim Sha Tsui East,
Kowloon, Hong Kong
Phone: +852-2301-4980
Fax: +852-2301-3545

• **P.R. China (Beijing)**
Anritsu Company Ltd.
Beijing Representative Office
Room 2008, Beijing Fortune Building,
No. 5, Dong-San-Huan Bei Road,
Chao-Yang District, Beijing 100004, P.R. China
Phone: +86-10-6590-9230
Fax: +86-10-6590-9235

• **Korea**
Anritsu Corporation, Ltd.
8F Hyunjuk Building, 832-41, Yeoksam Dong,
Kangnam-ku, Seoul, 135-080, Korea
Phone: +82-2-553-6603
Fax: +82-2-553-6604

• **Australia**
Anritsu Pty. Ltd.
Unit 21/270 Ferntree Gully Road, Notting Hill,
Victoria 3168, Australia
Phone: +61-3-9558-8177
Fax: +61-3-9558-8255

• **Taiwan**
Anritsu Company Inc.
7F, No. 316, Sec. 1, Neihu Rd., Taipei 114, Taiwan
Phone: +886-2-8751-1816
Fax: +886-2-8751-1817

White Paper Number EDR0210 Printed in Feb 2010 2008-09