Trend and Testing of 400GE Optical and Electrical Networks 400GE高速傳輸趨勢及光電測試綜覽

Joe Lin

2020.12.16

Project Manager / Keysight Technologies





Agenda

- Datacom trends fueling innovation
- Optical Interconnects in 2020
- FAQ
- Transmitter Test Discussion



Datacom trends fueling innovation



Datacom trends fueling innovation

BANDWIDTH, POWER, PRICE

INTERNET TRAFFIC CAGR

x3 in five years (2016-2021)

25%

WORKLOAD RUN OUTSIDE THE ENTERPRISE

50%

Computing moving in the cloud or at the edge

99% TRAFFIC FROM DATA CENTER

99%

Originated or terminated in Data Center

Sources: Statista, Cisco Global Source Index



A look into Mega Data Centers

TRAFFIC DOMINATED BY MACHINE TO MACHINE COMMUNICATION

71% EAST-WEST TRAFFIC



M2M in the DC

15% NORTH-SOUTH TRAFFIC



User



Cloud

14% 0000 INTER-DC TRAFFIC



Source: Cisco Global Source Index



是德科技400GE高速傳輸暨光電收發器測試研討會

Data Center challenges

POWER, SPENDING, DELAY

POWER CONSUMPTION FOR COOLING

>1% of WW el. Power consumption

50%

DATA CENTER SPENDING ON SW & HW

\$152Md

Steadily increasing. Hyperscale Data Centers account for a third of the market

DROP IN SALES FOR 0.1S LATENCY

6%

Similar results from amazon & google

Source: Statista, Northwestern University, Synergy research group, Yole development, Akamai study



Data Center challenges

TOWARD 1PJ/BIT

- Reduced footprint
 - Optimized **cooling** techniques, Server **virtualization** (VMs), etc.
 - Deploying efficient computing & networking technologies
 - The cascade effect
 - 1W reduction at the component level results in 2W+ reduction at the Data Center level



50Tb switch @ 20pj/bit = 2000Watt for populated chassis (1000Watt for optics)

NEED TO GO BELOW 5PJ/BIT



Source: Yole dev





Data Center challenges

TOWARD 1\$/GB/S OPTICAL I/O

- Reduced spending
 - Secure multi-source supply chain by
 - Enforcing interoperability
 - Fostering white box ecosystem (OCP)
 - Just good-enough design with MSAs
 - Leverage mass volume technologies
 - CMOS for consumer electronics (Silicon Photonics)
 - VSCEL for SR interconnects

Impact of optics in networking cost is growing (10% @10G, 30% @100G, 50%@400G)

1\$/GIGABIT FOR 400G OPTICAL TRX





Source: LightCounting, Dell'oro, Rockley Photonics



Optical Interconnects in 2020



Optical Interconnects in 2020

UNLEASHING 400G

25.6Tb

25.6TB SWITCH

12.8Tb currently deployed at hyperscale

400G OPTICS IN THE STARTING BLOCKS

400G

Mass deployment delayed 50% cost reduction compared to 100G

15W/module is still an issue (100G=3W)

是德科技400GE高速傳輸暨光電收發器測試研討

100G SERIAL IN THE STARTING BLOCK

100G

IEEE 80.23ck close to release 25.6Tb 100G SerDes-based switch ASIC by Innovium



400G deployment

400G SLOWY RAMPING

- Status @ Hyperscale DC's (Source LightCounting)
 - 400GbE DR4 delayed at AWS (chip)
 - 200GbE just started at Facebook
 - 2x200GbE slow start at Google
- What is slowing down 400G?
 - 100G PSM4 (<2\$/Gb) vs 400G-DR4 (~4\$/Gb)
 - <30pj/bit for CWDM4, >35pj/bit for 400G-DR4
 - Technical issues: Interoperability, margin (BER vs. FLR)



* 400G-FR4, configured 100G CWDM4 mode * 400G-SR8, configured in 2 x 100-SR4 mode



2020: Technology crossroads for R&D

SWITCH ASIC EVOLUTION



2020: Technology crossroads for R&D

100G SERIAL ELECTRICAL



2020: Technology crossroads for R&D

100G SERIAL ELECTRICAL



→ 50Gb/s interface

→ 100Gb/s interface

Standards

- IEEE 802.3ck
- OIF-CEI 5.0

Benefits

- Reduced power consumption (no gearbox)
- 800G-ready (8 lanes)

Challenges compared to 50G

- High Signal Integrity
- Higher Channel Losses & reflections
- Complex equalization (CTLE+DFE)



The rise of on-boards and co-package optics

IN A NUTSHELL

- On-Board Optics: Optical modules mounted on PCB
- Co-packaged optics: Optical engines co-packaged with switching ASIC
- Silicon photonics: Integration of electronics and photonic circuits in the same technology platform
- Industry initiatives
 - Consortium for On-Board Optics, COBO
 - Co-Packaged Optics Collaboration, CPO
 - CWDM-MSA
- Some 2020 CPO announcements
 - 12.8 Tbps switch with 1.6Tb/s optics (Intel)
 - 25.6T OptoASIC Switch system (Rockley)
 - 3.2 Tbps TOR switch (Accton, Cisco)
 - and much more





The rise of on-boards and co-package optics

EXPECTED BENEFITS AND CHALLENGES



Benefits

- Bring optics closer to ASICs
 - Removed or simplified re-timer, 75% power reduction compared to classical PCB
 - Switch bandwidth scalability: reduced size & footprint

• Silicon scale (CPO only)

- Leverage well-known and rapidly scalable fabrication CMOS fab environment
- Reduce packaging cost (80% of the total TRX cost in 2015)

Challenges

- Lack of flexibility (e.g. replacing defect optical engines, fixed optical interfaces)
- Complex thermal and mechanical design
- Manufacturing complexity and yield (CPO only)



Source: Luxtera

FAQ

1-2.50221



是德科技400GE高速傳輸暨光電收發器測試研討會

Q:400G 的測試跟過往的 100G 有什麼主要的不同?

- 第一個主要的差異是 PAM4 的使用,這讓測試增加一些測項,而對儀器的影響就是必須加買 PAM4 相關的 Option。
- 第二個主要差異發生在 CDR 的使用。對於系統廠來說,一直以來的測試都必須要 CDR,所以變化 不大,只要確認速率夠用即可。但對於模組廠來說就影響比較大了。因為 100G 之前的模組測試架 構上,並不需要CDR。但到了 400G 之後,因為模組內部很多都採用 DSP 架構,有很大的機會讓 模組的輸入與輸出之間速率不同步,因此就必須加買 CDR。
- 至於儀器頻寬的部分,請參考下一頁。



Q:400G 測試需要多少頻寬的示波器?

 事實上,400G 只是總資料頻寬,真正影響到示波器頻寬的是 Baud per lane。以 400G-FR8 為例子, 它達成 400G 方式是 25G x 2 (PAM4) x 8 (Lanes) = 400。也就是他實際上 Baud per lane 只有 25G, 所以在談示波器頻寬的時候,是看這個 25G 來判斷,跟 400G 沒有直接關係,所以測100G 的設備 繼續沿用在 400G 測試是有可能的。

	IE	OIF-CEI		
1	Electrical	Optical	Electrical	
25~28G NRZ	33 GHz	12.6 GHz	40 GHz	
25~28G PAM4	33 GHz	13.28125 GHz	40 GHz	
53~56G NRZ			43 GHz	
53~56G PAM4		26.5625 GHz	40 GHz	



Q: 為什麼是 25~28 / 53~56 GBaud而非定值?

在網路通訊的世界,OIF協會都是走最快的。他會先訂出CEI這份標準,其中速率的部分訂一個範圍,讓後面的協會可以自由地根據編碼的形式來選擇最適當的速率。以IEEE為例,在NRZ的訊號使用 64/66b coding,所以最後的真實速率是:

25G x (66/64) = 25.78125

- 而 IEEE PAM4 coding 改為256/257b · 又引入了 RS-FEC 的 coding · 所以速率最後變成了 25G x (257/256) x (544/514) = 26.5625
- 所以網路世界的規格跟其他高速數位標準很不同的地方是,他的宣稱速率是不包含 coding overhead 的。
 7.877 G
 7.877 G



Transmitter Test Discussion



Patterns

TEST PATTERNS FOR PAM4 ENCODED SIGNALS DEFINED IN IEEE 802.3BS/CD

- JP03A The JP03A test pattern is a repeating {0,3} sequence (clock)
 - No longer used in 802.3bs/cd.
- **JP03B** The JP03B test pattern is a repeating sequence of {0,3} repeated 15 times followed by {3,0} repeated 16 times (clock with a phase shift)
 - No longer used in 802.3bs/cd

PAM4 Test Patterns		
Pattern	Pattern Description	Defined in Clause
Square Wave	Square wave (8 threes, 8 zeros)	120.5.11.2.4
3	PRBS31Q	120.5.11.2.2
4	PRBS13Q	120.5.11.2.1
5	Scrambled Idle	119.2.4.9
6	SSPRQ	120.5.11.2.3

- PRBS13Q The PRBS13Q test pattern is a repeating 8191-symbol sequence formed by Gray coding pairs of bits from two repetitions of the PRBS13 pattern into PAM4 symbols as described in 120.5.7. (Note: PRBS13Q is different from QPRBS13 defined in IEEE 802.3-2015 (bj) Clause 94). Used for victim channel in TX tests.
- **PRBS31Q** The PRBS31Q test pattern is a repeating 2^31-1 symbol sequence formed by Gray coding pairs of bits from two repetitions of the PRBS31 pattern defined in 49.2.8 into PAM4 symbols as described in 120.5.7. Used for victim channel in RX tests and aggressor lanes in TX/RX tests.
- **SSPRQ** <u>Short Stress Pattern Random Quaternary</u>.

The SSPRQ pattern is a repeating 2^16–1 PAM4 symbol sequence. Comprised of 4 sequences, each based key "stressors" from PRBS31. Stressful pattern, but short enough to use advanced analysis tools available on today's T&M tools (e.g. Equalization, Jitter/Noise analysis, etc.). Used for Optical TX test.



Optical TX Test

1.50.22



Optical Measurement Differences: NRZ vs PAM4

USE OF FORWARD ERROR CORRECTION (FEC) RESULTS IN MAJOR CHANGES TO TX TEST

Primary NRZ transmitter tests:

- OMA (optical modulation amplitude)
- Extinction Ratio (ER)
- Eye-mask
- NEW!: Transmitter Dispersion and Eye Closure (TDEC) replaces Transmitter Dispersion Penalty (TDP) for new 25G and 50G TX



- Primary PAM4 transmitter tests:
 - Outer OMA
 - Outer Extinction Ratio
 - o NEW!: <u>No</u> eye-mask
 - o NEW!: TDECQ replaces TDP
 - o NEW! Transition Time (IEEE 802.3cd)





TDECQ

TRANSMITTER DISPERSION AND EYE CLOSURE QUATERNARY

- Tells you the performance of your transmitter relative to an ideal transmitter
- For NRZ TDP, we literally used a BERT to measure the BER performance of the transmitter compared to an actual (real) golden transmitter
 - > Determine how much extra power was required at the receiver to compensate for non-ideal performance
- For TDECQ we indirectly measure SER (symbol error rate) using a scope, no BERT required.



Reference: IEEE P802.3bs™/D3.5, 10th October 2017, Figure 121-4, page 222.



TDECQ measurement process (from IEEE 802.3bs)

- SSPRQ test pattern (~2^16 length) (Short Stress Pattern Repetitive Quartenary)
- Includes test fiber dispersion
- Oscilloscope noise measured and mathematically 'backed out'
- Virtual 5 tap, T spaced FFE reference equalizer optimizes eye openings (to minimize TDECQ penalty).
- Histograms constructed to assess eye closure relative to OMA and compute an effective power penalty in dB. This is the TDECQ result







Figure 121-5—Illustration of the TDECQ measuremen

Reference: IEEE P802.3bs™/D3.5, 10th October 2017, Figure 121-5, page 222, 224.



TDECQ Measurement Setup

REQUIRES ACCURATE OPTICAL REF RX AND ROBUST CLOCK RECOVERY DESIGN



TDECQ: Measurement Tips

CHOOSE A MEASUREMENT SOLUTION THAT IS FAST, FLEXIBLE, AND ACCURATE

- Compliant TDECQ measurements require:
 - Histogram optimization (adjust left and right) (IEEE 802.3bs D3.5, IEEE802.3cd D3.2)
 - Threshold optimization (adjust up and down) (IEEE 802.3cd D3.4, likely will be added to 802.3bs in a maintenance release)
 - Recommendation: keep these enabled at all times (they have a minimal impact on throughput, but can have a large impact on result).

Recommendations for throughput optimization:

 Iterative "tap" optimization – typically results in a 0.1dB to 0.2dB difference in TDECQ (signal dependent). Consider disabling this feature to increase throughput (e.g. in production test).

o "Seed" equalizer tap values for faster tap optimization

 Consider using a more powerful CPU to perform your TDECQ measurements, especially for parallel TDECQ testing (4 channels)







TDECQ: Does it work as designed?

POWER PENALTY VALUES PROVIDED BY THE FINAL IMPLEMENTATION OF TDECQ CORRELATE DIRECTLY TO OBSERVED RECEIVER SENSITIVITY

Consider a typical PAM4 receiver and two transmitters A and B. If the TDECQ of transmitter A is 2 dB and for transmitter B is 3.2 dB, should I be able to make any predictions about the power levels required at the receiver to achieve a specific BER/SER with each transmitter?

Latest results from IEEE 802.3cd project:

http://grouper.ieee.org/groups/802/3/cd/p ublic/July18/tamura_3cd_01c_0718.pdf (used with permission from the author)





Transmitter Transition Time for Optical PAM4

NEW MEASUREMENTS ADDED TO IEEE 802.3CD

Transmitter transition time is defined as the **slower** of the time interval of the rising/falling transitions of a PAM4 signal.

Test Conditions:

- 20% to 80% of Outer OMA levels
- Measured with Square Wave or SSPRQ pattern.
 - SSPRQ: Use specific symbol sequences
 - 00000333333 (rise time)
 - o 33333000000 (fall time)
- Measured with specific Ref RX BW and 4th Order BT response



Electrical TX Test



Key PAM4 Measurements for <u>Electrical</u> Transmitters

IEEE 802.3BS & CD



What are the key PAM-4 TX parameters that get measured?

- xAUI C2C (TP0a) & CR (TP2)
 - Level Separation Mismatch Ratio
 - Steady State Voltage
 - Linear Fit Pulse Peak
 - Signal-to-noise-and-distortion ratio (SNDR)
 - J3u (802.3cd), J4u(802.3bs), Even-Odd Jitter (EOJ)

• xAUI C2M (TP1a & TP4)

- ESMW
- Vertical Eye Closure (VEC)



C2C v.s. C2M



Figure 120D-4-Typical 400GAUI-8 chip-to-chip application



Scope goes here ...

Transmitter compliance measured at TP0a



KEYSIGHT

Level separation mismatch ratio (RLM)

 PAM Linearity measures the linearity RLM (Ratio Level Mismatch) of all four amplitude levels (0, 1, 2, 3) of a PAM4 signal. Linearity is a measure of the variance in amplitude separation (distribution) between the different PAM4 levels. The PAM4 signal must be displayed as a single-valued waveform and all four logic levels should be visible on the display.





Linear fit pulse peak and Steady State Voltage



The waveform shall be captured with an effective sample rate that is M times the signaling rate of the transmitter under test. The value of M shall be an integer not less than 7. The steady-state voltage vf is defined to be the sum of the linear fit pulse p(k) divided by M.



SNDR

In an SNDR model, the degree that the actual pulse response matches a linear fit model is the main consideration. The ISI jitter/noise is 100% compensable by equalizers and of no real consequence to the system performance (assuming the equalizers can be implemented in the system). The residue from the linear fit is called distortion. The ratio between the signal from linear fit and distortion-plus-noise projects the system performance with equalizers. The standard deviation of the error in this model fit is called distortion, the s_e parameter in the SNDR expression. For each of the four symbol levels, the vertical signal noise is determined at a fixed low-slope point in a long consecutive run of identical PAM4 symbols. The noise parameter s_n is the average of the standard deviations of these four symbol measurements. P_{max} is the peak amplitude of the pulse response from the linear fit.

$$SNDR = 10\log_{10}\left(\frac{p_{\text{max}}^2}{\sigma_e^2 + \sigma_n^2}\right)$$



Output Jitter (J3u/J4u, J_{RMS}, and EOJ)

NEW MEASUREMENTS PERFORMED ON 12 SPECIFIC EDGES OF A PRBS13Q PATTERN

- J3u/J4u, J_{RMS}, and Even-Odd Jitter (EOJ)... you may recognize some of these acronyms from other (older) Standards
- So they should be pretty straight forward to measure, right?
- While the IEEE Output Jitter names may sound familiar, they are measured very differently!
 - Traditional Jn (e.g. J5, J9) and EOJ parameters were measured using <u>all edges</u> of an NRZ pattern.
 - IEEE 802.3bs/cd now measure J3u/J4u , J_{RMS}, and EOJ on <u>12 specific edges</u> of a PRBS13Q (PAM4) pattern!

Label	Description	Gray coded PAM4 symbols	Index of first symbol	Index transition begins	Index transition ends	Index of last symbol	Threshold level
REF	Reference for symbol index	3333333	1			7	-
R03	0 to 3 rise	10000 330	1830	1834	1835	1837	
F30	3 to 0 fall	23333 001	1269	1273	1274	1276	$(V_0 + V_3)/2$
R12	1 to 2 rise	0111111 2222221	3638	3644	3645	3651	
F21	2 to 1 fall	022222 113	1198	1203	1204	1206	$(V_1+V_2)/2$
R01	0 to 1 rise	100000 113	6835	6840	6841	6843	
F10	1 to 0 fall	21111 003	2992	2996	2997	2999	$(V_0+V_1)/2$
R23	2 to 3 rise	32222 330	6824	6828	6829	6831	
F32	3 to 2 fall	033333 2222223	7734	7739	7740	7746	$(V_2+V_3)/2$
R02	0 to 2 rise	10000 223	3266	3270	3271	3273	
F20	2 to 0 fall	122222 0000002	7282	7287	7288	7294	$(V_0 + V_2)/2$
R13	1 to 3 rise	011111 331	133	138	139	141	
F31	3 to 1 fall	23333 112	7905	7909	7910	7912	$(V_1+V_3)/2$

RBS13Q pattern symbols used for jitter measu

Reference: IEEE P802.3bs™/D3.5, 10th October 2017, page 357.

So don't just pull out your scope and press the Jn (J2, J5, J9) or EOJ button in Jitter Mode!



Why did the PAM4 jitter measurement methodology change?

NEW METHOD IMPLEMENTED IN IEEE 802.3BS/CD AND CEI-56G-MR/LR-PAM4

- Different TX Architectures are used to generate PAM4 signals
- Some TX designs may use different clock buffers for MSB and LSB; this can result in different uncorrelated jitter appearing on different edges.
- Measuring jitter only on JP03 (clock) patterns (original method) could miss potential issues.





Measure Output Jitter at TP0a

"12-EDGE" OUTPUT JITTER MEASURED ON KEYSIGHT DCA AND RT SCOPES

New "12 edge" jitter method in FlexDCA reduced test time from hours to < 1 minute.

• J3u/J4u and JRMs jitter

- Measure RJ/PJ on 12 specific transitions using a PRBS13Q pattern (exclude correlated jitter).
- o Data from all edges is combined and analyzed
- Even-Odd Jitter (EOJ)
 - Measured on PRBS13Q (3 repeats)
 - <u>Max</u> from measurements on all 12 edges
- Keysight DCA and RT Scopes report:
 - J3u/J4u, J_{RMS}, EOJ "ALL" measurement (per the Standard)
 - FlexDCA also reports individual results for each of the 12 edges
 - Rise: 0 to 3, 1 to 2, 0 to 1, 2 to 3, 0 to 2, 1 to 3
 - Fall: 3 to 0, 2 to 1, 1 to 0, 3 to 2, 2 to 0, 3 to 1

Measurement Setup:

- Receiver: 4th Order Bessel-Thomson low-pass filter with 33 GHz BW
- CR PLL BW 4 MHz and a slope of 20 dB/decade



Output Jitter 🜔	FlexDCA DCA Results				
Src: D5A Rate: 26.562500 GBd Pat. Length: 8191					
Measurement	To L0	To L1	To L2	To L3	
😑 J4u (All)	1.968 ps				
From L3	1.709 ps	1.754 ps	2.052 ps		
From L2	1.781 ps	2.026 ps		2.032 ps	
From L1	2.078 ps		2.056 ps	1.762 ps	
From L0		2.028 ps	1.762 ps	1.739 ps	



KEYSIGH TECHNOLOGIE 是德科技400GE高速傳輸暨光電收發器測試研討會

ESMW (Eye Symmetry Mask Width)

1. Capture 1e-5 probability CDF

2. Add CTLE

3. Use Center Eye Vmid as Center Line

4. Check each eye's eyewidth if it is symmetry.





VEC (Vertical Eye Closure)



Vertical eye closure is defined by Equation (120E-4) for a PRBS13Q differential equalized signal captured and processed according to 120E.4.2.

$$VEC = 20\log_{10}\left(\max\left(\frac{AV_{upp}}{V_{upp}}, \frac{AV_{mid}}{V_{mid}}, \frac{AV_{low}}{V_{low}}\right)\right) \text{ (dB)}$$
(120E-4)

is the vertical eye closure, in dB

where

VEC

1. Get every eye's voltage and eye height

2. Find the maximum of Voltage / Eye-height

3. Changes to dB.



New Standard in 802.3

802.3 CK AND 802.3 CU

- IEEE 802.3ck
 - 100 Gb/s, 200 Gb/s, and 400 Gb/s Electrical Interfaces
 - VEC PAM4 and NRZ test.
- IEEE 802.3cu
 - 100 Gb/s and 400 Gb/s over SMF at 100 Gb/s per Wavelength
 - Overshoot / Undershoot Test
 - Pk-Pk Test



VEC

802.3BS AND CK





是德科技400GE高速傳輸暨光電收發器測試研討會

Overshoot, Undershoot and Pk-Pk

802.3CU



Revision A.06.60, July 2020

Minor software revisions may not necessitate an update to the product's learning products. Similarly, changes to the learning products may not necessitate a change to the software revision.

New Features

- PAM4 eye mask testing with four supplied Open Eye MSA PAM4 masks.
- New Eye Mode Measurements:
- IEEE 802.3ck Vertical Eye Closure (VEC) PAM4 and NRZ measurement.
- IEEE 802.3cu PAM4 overshoot measurement.
- IEEE 802.3cu PAM4 undershoot measurements.
- IEEE 802.3cu PAM Pk-Pk amplitude measurement.
- New Jitter Mode Measurements:
- IEEE 802.3ck Vertical Eye Closure (VEC) PAM4 and NRZ measurement.



是德科技400GE高速傳輸暨光電收發器測試研討會

KEYSIGHT TECHNOLOGIES