VIAVI Solutions

白皮书

QSFP-DD 模块测试

400G 和 QSFP-DD

QSFP-DD 光模块是 400G 客户侧接口的主流封装规格。本白皮书为模块开发人员、网络元件制造商和最终用户分享了 QSFP-DD 模块成功测试、故障排查和验证的关键因素。

客户侧接口速度稳步增长，典型的速率每十年至少增加十倍。100GE 已经通过 QSFP28 接口广泛部署，我们正处于 400G 部署的早期阶段。作为 2017 年 12 月正式标准化的 802.3.bs 的一部分，IEEE[1] 开发了 400G 以太网客户侧接口标准。早期采用者使用的是 CFP8[2] 封装规格，但更广泛的市场关注的是 QSFP-DD[3]，它允许与广泛采用的 QSFP28 实现一定程度的向后兼容。

由于以太网具有广泛的应用范围，并可提供一系列 PMD（物理介质相关）选项，因此允许一个“QSFP-DD”插槽支持大量应用，覆盖范围从几米长的无源铜缆 DAC 电缆到 80 千米相干 ZR。还有少数公司专注于 OSFP[(4)] 封装规格。虽然不是那么广泛和向后兼容，但它确实在电信号完整性和热管理方面提供了一些优势。以下有关 QSFP-DD 的大部分内容适用于 OSFP 和 VIAVI ONT 系列，该系列支持许多基于 OSFP 的应用。[4]

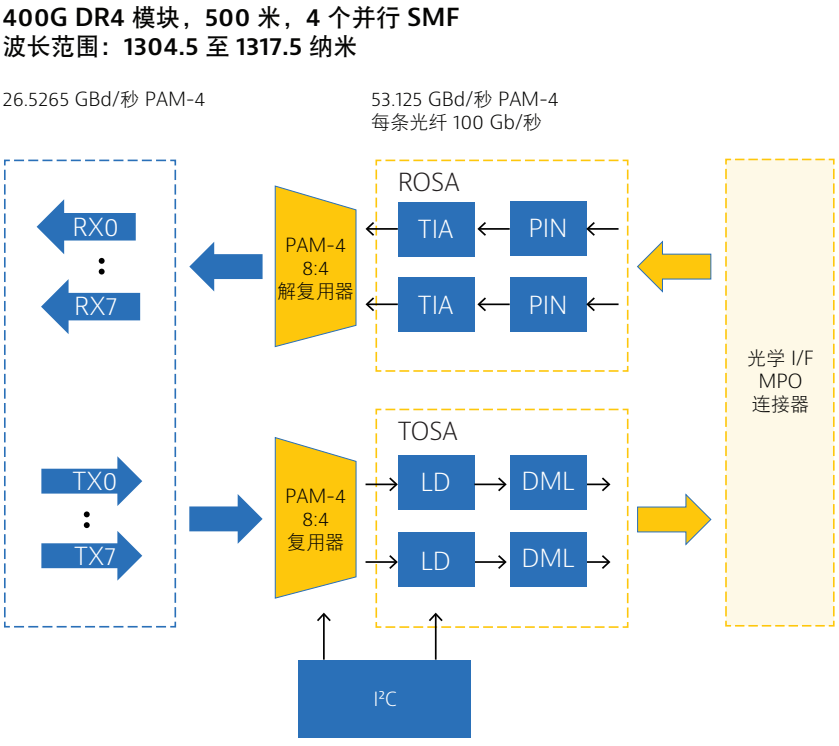
400G 对电模块到主机接口以及电口或光层 PMD 都依赖于高阶 (PAM-4) 调制。采用 PAM-4 调制是为了在给定带宽下最大限度地提高数据容量，但它在复杂性和性能方面带来了巨大挑战，这也意味着链路需要前向纠错 (FEC) 编码才能实现可靠的数据传输。

1. http://www.ieee802.org/3/bs/
2. http://www.cfp-msa.org/documents.html
3. http://www.qsfp-dd.com/
4. https://osfpmsa.org/

为什么选择 QSFP-DD?

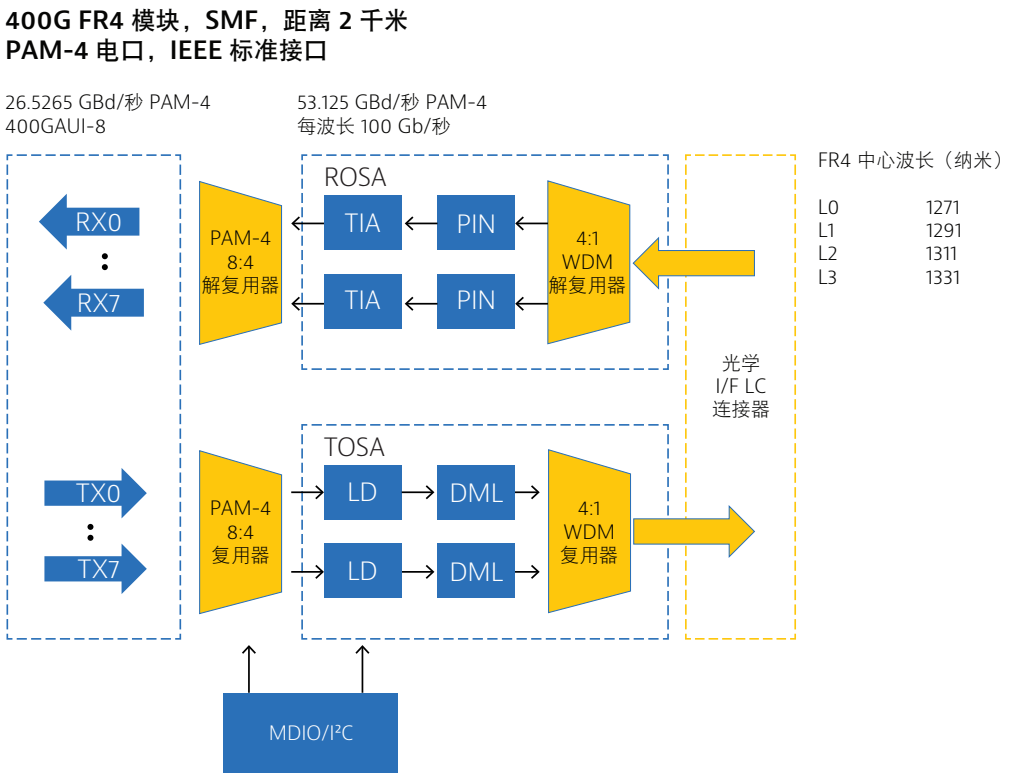
100G 以太网于 2008 年开始部署，早期设计基于 CFP 可插拔模块。第二代系统迁移到 CFP2（或一家大型设备制造商的 CPAK），然后选择了 QSFP28，这推动了广泛且经济高效的批量采用。CFP4 对 QSFP28 来说是稍早的挑战，但由于多种因素，QSFP 28 推动了 100G 的大幅增长。业界注意到了“封装规格”之战，并希望将 400G 的多步封装规格演变带来的额外复杂性和成本挑战降至最低。CFP8 允许非常早期的采用者开发和验证 400G。然而，它没有满足密度、功率、成本和“兼容性”的需求，因此业界很快将 QSFP-DD 作为目标封装规格。有人提出了替代方案 – OSFP。它提供了卓越的技术解决方案，但无法满足对传统遗留模块接口支持的迫切需求。原则上，QSFP-DD 插座可以支持传统的 QSFP-28 光模块 – 这将允许供应商发货“400G 就绪”网元，这些网元今天可以与 100G 模块一起发货，现场升级将是简单的模块更换。

为了满足向 400G 迈进所带来的更高带宽、功率和冷却需求，对现有的 QSFP28 概念进行了几项增强。这些增强包括将高速电口通道增加一倍（从 4 通道 25 Gbps 的 NRZ增强到 8 通道 56 Gbps 的 PAM-4），以及延长模块的“前端”，以提供更大的内部体积和增强的热性能。 此外，还进行了进一步的工作，以增强模块控制接口，从而产生 CMIS 4.0[5] 标准。

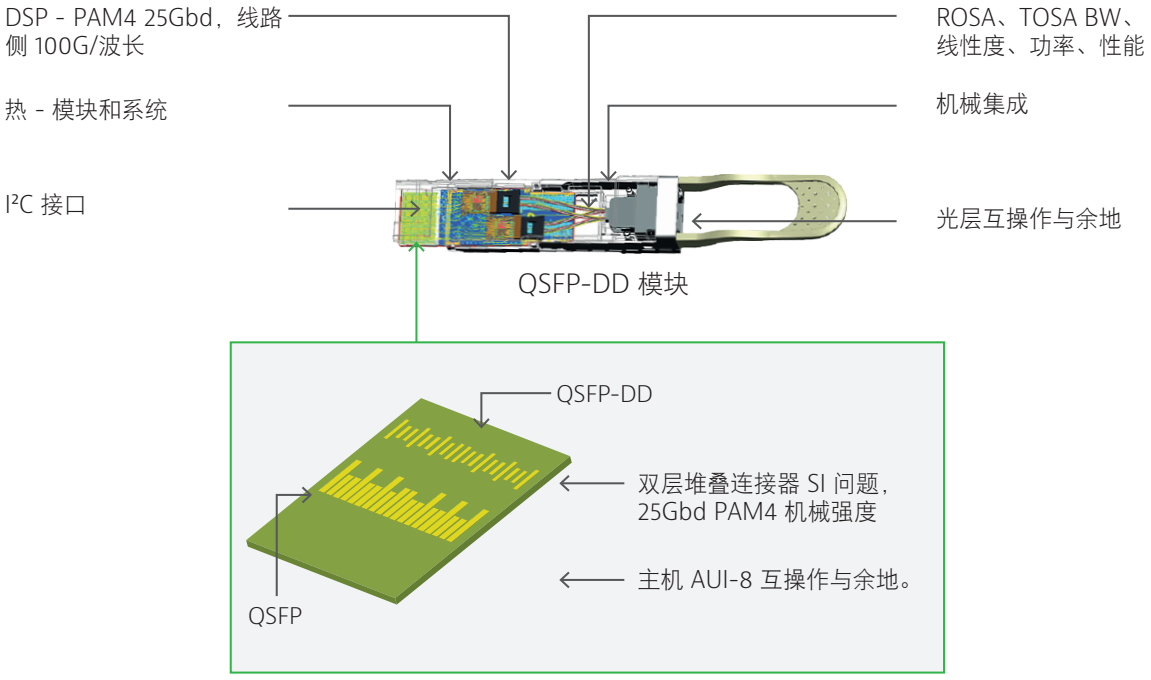


5. http://www.qsfp-dd.com/qsfp-dd-msa-group-announces-updated-specificatio/

DR4 将是 2020 年部署的最常见的 400G 客户侧光接口之一。它在单独的单模光纤上以四个 100G 信号的形式传输 400G。它在企业中有着广泛的应用。它支持 500 米的覆盖范围，并且能够连接到单独的 100G 以太网链路，作为高密度的 100G 解决方案很有吸引力，这可以使端口计数密度提高四倍。



FR4 接口也将有广泛的应用，包括电信领域。它通过一根单模光纤提供 2 千米的更长链路预算。400G 由四个 100G 信号承载，每个信号的波长略有不同。



400G PMD 模块（物理介质相关）

PMD	范围	应用	技术
DAC	2 到 3 米	机架内和服务器	无源铜缆、50G PAM-4 电口
SR8	100 米	企业	并行多模。50G/λ – PAM-4
DR4	500 米	数据中心和企业	并行单模，100G/λ – PAM-4
FR4	2 千米	大规模数据中心	单模，100G/λ，PAM-4
LR8	10 千米	电信距离	单模，100G/λ，PAM-4
ZR	80 千米	城域网和 DCI	单模/相干，PAM-4

QSFP-DD 模块 – 标准和主题

正如我们从上面的参考文献中看到的，许多标准和 MSA 都是适用的。了解开发周期每个阶段（从基本 IC 评估到模块硬件集成、软件和固件，再到供应商选择和验收）的关键测试也很重要。生产也有自己的一套关键测试要求。

要成功设计、测试、验证、制造和部署可插拔光学模块和器件，需要对 IEEE、CMIS、QSFP-DD、MSA 和 OIF 等关键文档有扎实的理解。QSFP-DD 是电子、光学、机械、热管理和固件的完美集成组合。在成功部署模块之前，所有组件必须协同工作。

互操作性

以太网客户侧接口生态系统的最大优点在于，我们拥有一套由 IEEE 和其他标准驱动的强大而清晰的标准，这些标准允许多供应商生态系统在不求助于“工程”链路的情况下实现互操作。

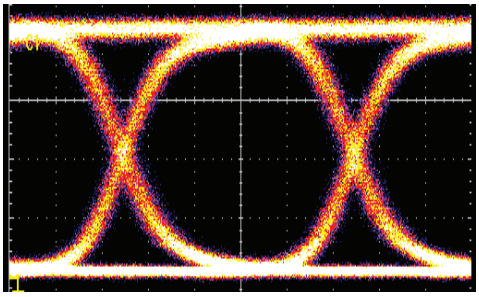
模块到主机接口和光纤接口都是这种互操作的关键。在主机到模块接口上，我们主要关注三个方面：

- 从芯片到模块 (C2M) 构建的高速数据路径 (AUI) 面临多重挑战，包括信号完整性和信号均衡。虽然 FEC 预算的一部分被分配给这部分链路，但此接口的任何问题都可能导致链路出现重大问题。“调谐”不好的链路（就均衡器和信道而言）可能会导致难以解决的问题，例如随机突发或最糟糕的偶然比特滑动情况。
- 模块管理– 这种基于 I²C 的接口已从 SFF-8636 的基本内存映射管理发展到 100G 的 QSFP28，再到复杂的状态完整的 CMIS 4.0。这一演变对生态系统来说极具挑战性，扎实的 CMIS 4.0 文档工作知识是健壮稳定的模块管理的关键。
- 模块功率 -- 对于 DCI 应用的可插拔相干 (QSFP-DD ZR) 模块，模块的功率需求已从 100G 时的几瓦攀升至可能接近 20W。这就对电源的健壮性和稳定性提出了很高的要求。此外，它还必须能够在模块唤醒时提供电力需求的动态和瞬变特性。

这些领域都是紧密交织在一起的，需要作为一个整体来对待（特别是在 CMIS 4.0 模块管理方面），以确保模块运行无故障。

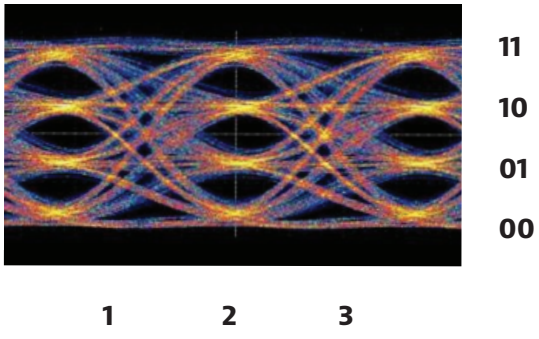
PAM-4

电（模块到主机接口）和光（电）链路均采用 PAM-4 调制。这种更高阶的调制方案允许在单位时间内发送的比特数翻倍。虽然 NRZ 技术在高速方面已经广泛应用和成熟，但 SERDES PAM-4 是一项相对较新的技术，更复杂和更具挑战性。我们在 NRZ 链路的误码分析方面有丰富的经验。但是我们仍然看到在 100GE 使用的从 10G 到 25G NRZ 通道的问题。因此，转向 PAM-4 预计将是整个行业的重大挑战。使用基于 FEC 的链路（总是有后台误码率）和复杂得多的信道均衡使这一点变得更加复杂。公平地说，PAM-4 比广泛使用的 25G NRZ 复杂一个数量级。



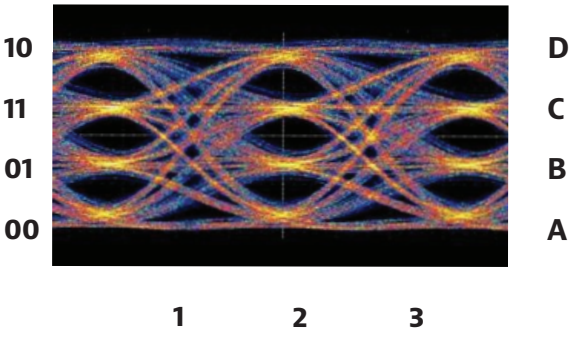
NRZ 调制：

- 每个时钟周期 1 比特：
  - 电压 A= „0"
  - 电压 B = „1"



PAM-4 调制，线性（non-Gray）编码：

- 每个时钟周期 2 比特
  - B 和 C 之间的错误判决
    -> 2 个错误比特！



PAM-4 调制，Gray 编码：

- 每个时钟周期 2 比特
  - B 和 C 之间的错误判决
    -> 只有 1 个错误比特

图像显示了 NRZ 和 PAM-4 信号

FEC

由于开发能够提供无差错 PAM-4 传输的组件极具挑战性，开发人员使用了一种 FEC，它可以同时保护电气模块接口以及光学模块到模块接口。我们花费了大量精力，仔细理解了传输信道和组件中的误码机制，以及如何以 FEC 逻辑（编码和接收）端的“成本”来进行均衡设定。FEC 的“成本”包括额外的电路，这些电路会消耗电力，并会增加任何链路的延迟。

DSP 和均衡器

在采用 400G 时，决定使用“强大的”电接收均衡器的概念，来面对“最坏情况”发射机和“最坏情况”信道的性能。这可能导致 PAM-4 接收器输入端的 PAM-4 眼图闭合，因此 PAM-4 接收器需要一个强大且可能复杂的接收器来平衡发射和信道影响，以便恢复清晰的眼图来实现对给定码元的正确解码。均衡器的复杂性意味着，在大多数情况下，必须实施基于 DSP 的解决方案，这可能会对功率、延迟、复杂性、误码性能和管理/控制产生影响。虽然 DSP 均衡器功能强大，但功能的复杂性可能会导致在寻找抽头的最佳设置等方面遇到挑战。此外，均衡器通常隐藏在 DSP 固件和控制 API 之后，对用户而言高度抽象。TDECQ[6] 的测量面临更多挑战 – 此测量复杂且可能不一致，这进一步增加了自由互操作、多供应商生态系统的挑战。

要点

总会有误码 – 链路现在总是有后台误码率。误码统计数据“指纹”至关重要。真正的随机误码流通常与用于保护链路的 FEC 兼容。但是突发、滑动和其他确定性问题可能会严重降低 FEC 纠错能力。在真实链路中，误码可能是电和光信道噪声、串扰、信号完整性问题、突发、比特滑动甚至错误设置均衡器导致的误码增殖的复杂混合。

最终重要的是，当给定一个特定的误码指纹时，FEC如何执行。什么是余地？我们还要多久才能收到丢弃的数据包？我们能否预测长期性能以了解链路退化情况？误码的根本原因是什么？

可以使用几种工具来调查误码指纹，从单个 PAM-4 码元中的错误偏差到比特滑动性质突发分析。通过时钟变化和 skew 等工具，可以进一步增强对误码偏差的理解。

PAM-4 码元分析可以用来确保在误码分布中没有“电平”偏差。关键光子元件（例如接收器光子 AGC）的稳定性可以通过观察 PAM-4 误码分布时光功率的变化（通过衰减器）来进一步验证。

重要的是要全面调查误码突发，并确认它们是突发而不是比特（或码元）滑动。滑动通常与 DSP（和相关固件）有关，无法通过 FEC 进行纠正。一般的测试无法区分由经典信号完整性或噪声问题引起的突发问题，以及与时钟和相位灵敏度相关的突发问题。因此，在调查 QSFP-DD 误码的性质和根本原因时，必须部署大量的新工具和技术。

6. https://ieeexplore.ieee.org/document/7937468

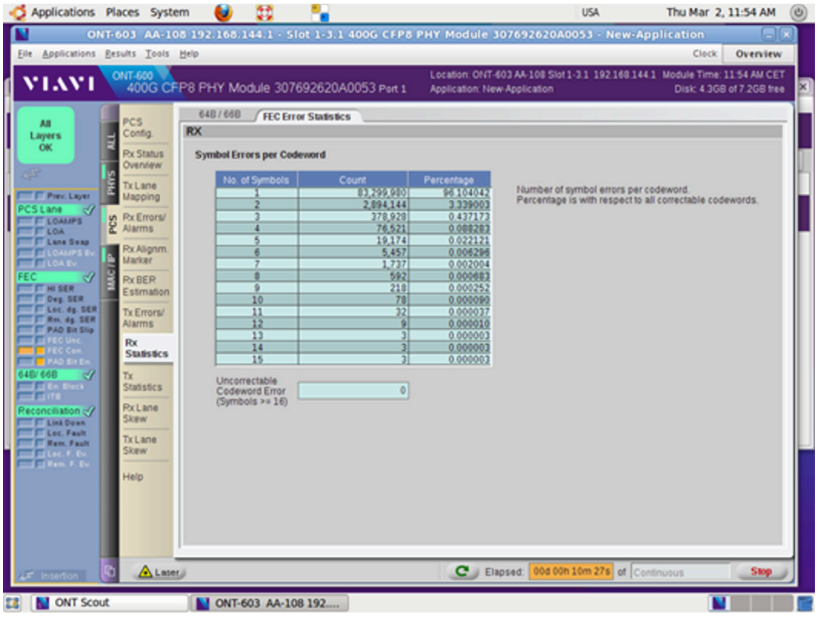
通过查看每 5440 比特 FEC 码字 (KP4 FEC) 的 10 比特码元错误的数量，可以获得最简单的顶层视图。我们通常期望单调分布的每码元计数下降大约 10。也就是说，每增加一个错误的码元/码字，我们期望误码数会减少 10 个。任何长尾或孤立的峰值都表明了某种非随机的（系统）原因。我们还期望错误码元的数量将在测量时间内增加 10 倍。因此，如果我们在 10 秒后观察到每个码字有 10 个错误码元的计数，我们期望在大约 100 秒后会看到 11 个错误码元计数。

这样的经验法则可用于估计出现不可纠正误码的时间（每个码字出现 16 个或更多错误码元）。例如，在 100 小时的测试时间之后，如果我们观察到最多 12 个错误码元/码字，则可以预期以下近似值：

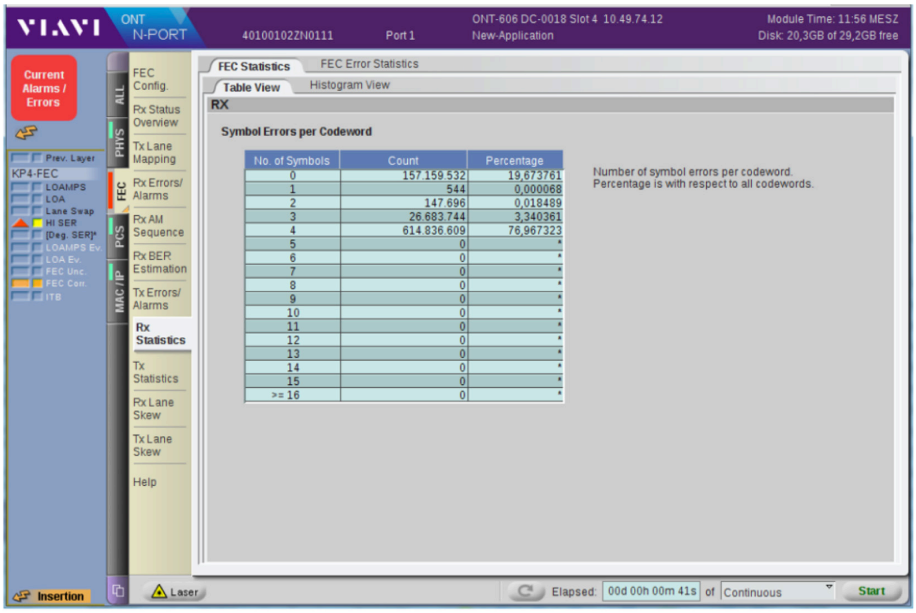
错误码元	时间	注释
12	100 小时	测量
13	1000 小时	估计
14	约 420 天	
15	约 11 ½ 年	
16（不可纠正的码元）	约 114 年	第一个丢弃的数据包 > 一个世纪

FEC – 错误码元/码字

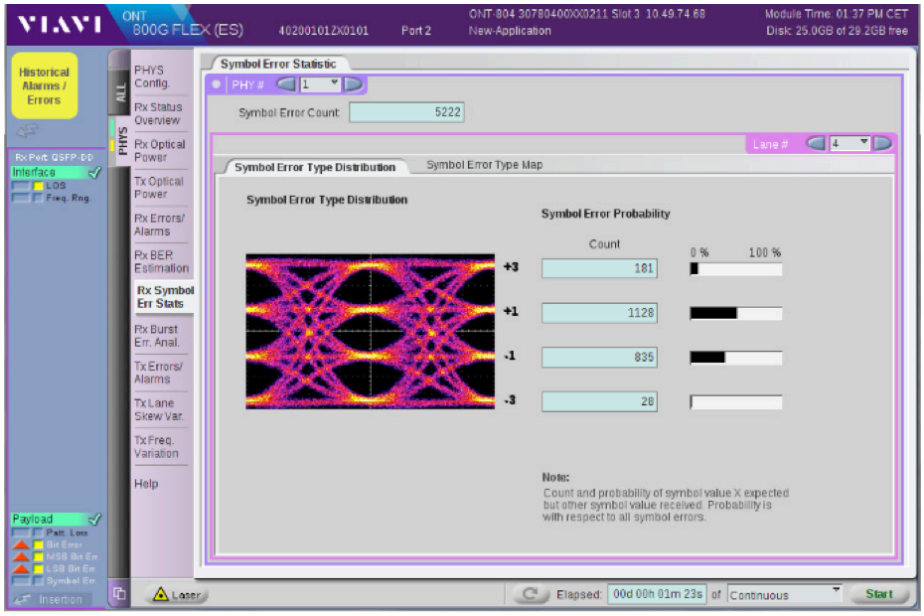
在下面的情况下，ONT 使用严重衰减的 400G 光纤链路运行，以至于在 10 分钟间隔内发生重大误码。这是合规链路可能发生的预期情况。如您所见，分布一般是单调的。每个错误码元的计数下降，但它确实显示了比 12 个错误码元/码字稍长的尾部。在这种情况下，链路极有可能由于有未校正的码字而丢弃数据包。



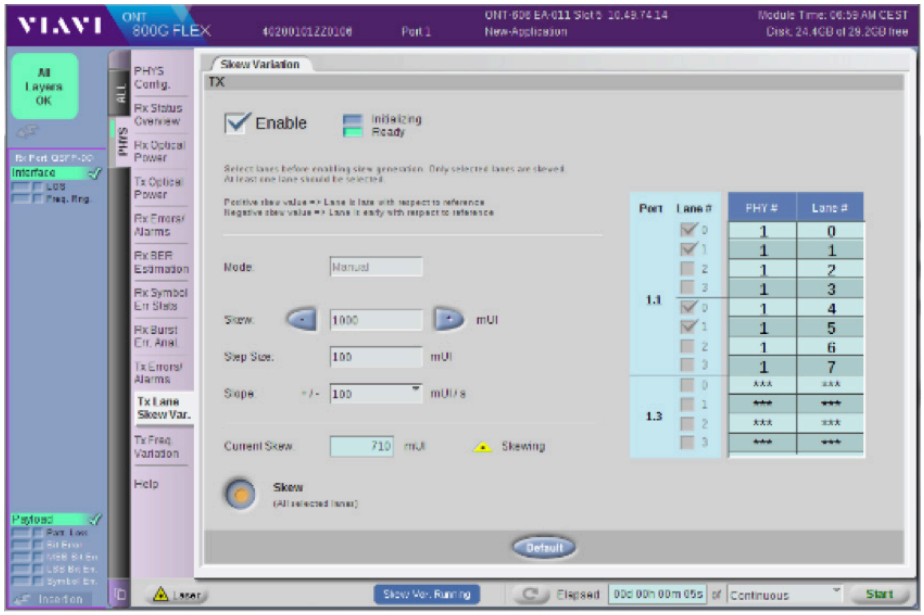
下面的屏幕截图显示了发生严重问题的情况。虽然 FEC 有很大的余量（我们在一个码字中最多可以看到 4 个错误码元），但分布不是单调的，这表明该系统存在潜在的误码源。请注意，这个 100G 链路的示例是由一个特殊的 VIAVI ONT 应用程序生成的，该应用程序可以创建广泛的 FEC 误码分布，以对 FEC 逻辑和电源完整性进行压力测试和验证。



ONT 既能够分析整个序列上的误码分布和码型，又能够在每个 PAM-4 码元的基础上跟踪误码特性。



动态 skew 变化是对 QSFP-DD 模块进行压力测试和验证的强大工具。它可用于验证是否符合 IEEE 802.3 标准，以及 DSP 和相关固件的总体稳定性。这在 DR4 模块中尤其重要，因为在 DR4 模块中， 一对单个电气和光学通道可能位于完全不同的时钟域！



上面的屏幕截图显示了 PAM-4 的动态 skew 应用。 它能够精确控制传输通道相对于 UI 的相对时序，同时仍然保持“无中断”相移，这是解决串扰和基于 DSP 的固件时序问题等挑战性问题的关键。

动态 skew（或 skew 变化）是对任何并行通道通信系统的关键测试。它可用于信号完整性测试和验证（串扰），也可用于对 PAM-4 SERDES 中 FIFO 和 CDR 的性能进行压力测试和验证。

不同的 skew 程度也可用于调查信号完整性和串扰问题，这在硬件和 SI 团队中有广泛的应用。通道定时可以调整，以确保干扰源通道转换发生在受干扰对象通道的 PAM-4 眼图的中间。

PAM-4 信号（因为信号余量较低）比经典 NRZ 更容易受到串扰的影响。在 QSFP-DD 的密集范围内（尤其是在主机连接器周围），高速 PAM-4 通道的布线非常接近，必须小心地避免出现信号串扰问题。正常情况下，BER 测试仪以固定的相位运行平行通道，因此 SI 压力测试下可能不会出现“最坏对齐情况”。利用动态 skew，干扰源通道可以在相对相位中被扫描，以充分验证问题不会发生，即使在最坏的相移情况下也是如此。最终用户只需观察特定相位偏移处是否出现误码（通常是当干扰源通道在受干扰对象“眼图”中间有电平转换时）。

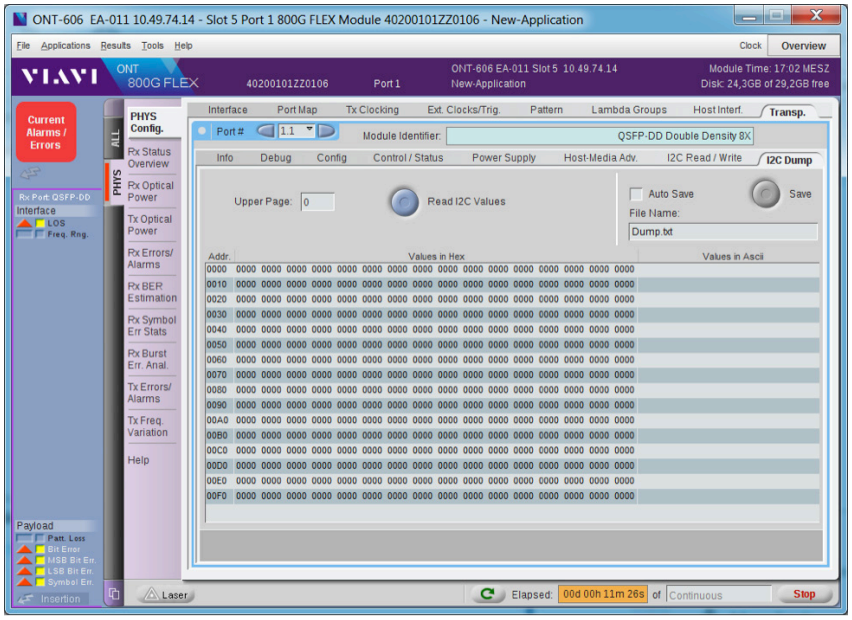
现代 SERDES 使用一系列 FIFO 缓冲器对信号重新计时和重新对齐，然后再在 IC 结构内进一步处理。重新对齐使用一系列 FIFO 缓冲器，这些缓冲器从主时钟源（通常是通过 CDR 的主通道）恢复时钟。

如果系统设计或实现方法不正确，则可能是主通道（CDR 参考通道）和其他通道之间的相位变化和变化导致 FIFO 未对齐甚至滑动。这将表现为一个比特滑动，ONT 高级错误分析可以将其作为一个比特滑动来跟踪，而不是像传统测试设备所看到的那样作为一个突发误码来跟踪。利用动态 skew 应用程序，ONT 可以有意对 SERDES 中 CDR/FIFO 的性能进行压力测试，并试图通过 skew（范围和速率）强制产生失效模式。这与 ONT 先进的误码分析相结合，为 SERDES 测试提供了一个非常强大和完整的测试系统，并可用于快速解决 400GE 链路中偶尔导致比特滑动的极具挑战性的问题。ONT PAM-4 动态 skew 可以强制产生这些误码来帮助诊断和解决根本原因。

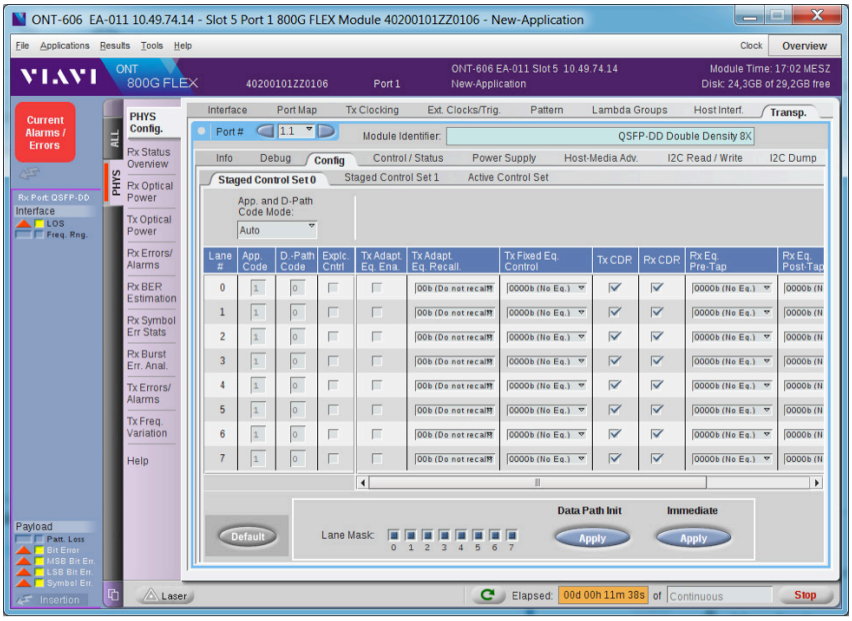
通用 QSFP-DD 控制画面

随着时间的推移，模块管理已经从非常基本的基于寄存器的系统 SFF 8636 发展到 CMIS 4.0，这是一个全面的有完整的模块状态的管理系统，旨在满足 400GE 及以上更复杂模块的需求。

模块通过 I²C 控制接口、电源和控制引脚以及数据路径之间的紧密交互对模块的稳健、稳定运行至关重要。模块的复杂性更高，尤其是在模块 DSP 中的数据路径均衡方面，需要对主机和模块之间的控制设置和执行有更全面的了解。在 CMIS 4.0 下，命令、操作和时序行为必须按照正确顺序紧密编排。如果不小心，一个模块可能看起来在一个主机插槽中运行没有问题，但另一个（在命令、电源和数据路径周围的时序上有细微差别）上面可能运行不稳定。或者更糟糕的是，误码率增加，出现罕见且难以解决的问题，很可能是比特滑动。ONT 等工具集成了 I²C 上的 CMIS 命令，以及模块电源控制和数据路径状态，不仅有助于调试和解决问题，还有助于对模块在不同主机中的稳健性进行压力测试和验证。



上面的屏幕显示了内存第一页的内存转储。这可以快速检查 QSFP-DD EEPROM 中是否存储了正确的值。空白或随机数据可能表示设备尚未初始化。



模块管理应用中的一些更高级的应用允许以清晰和明确的方式精确地控制模块的电口参数。

总结

QSFP-DD 模块是电子、光子、机械和热工程与复杂固件结合在一起的奇迹。健康的多供应商 QSFP-DD 生态系统对于 400G 网络技术的广泛部署至关重要。它代表了传统 100G 模块技术的发展和革命，同时也带来了新的挑战，包括 PAM-4 信号（电气和光学）、FEC 用于链路误码控制以及 CMIS 4.0 新的复杂性。

这些挑战之所以更大，是因为超大规模用户的规模和部署需求推动了价格预期的变化。生产必须满足产量和吞吐量要求，以达到价格预期，同时还必须具备覆盖和分析能力，以应对 PAM-4 的新挑战。

VIAVI ONT 系列对模块验证和测试应用已经有了二十多年深入透彻的研究。它通过先进的误码分析和动态 skew 等经典 100G 应用以及 CMIS 4.0 调试和 PAM-4 码元分析等 VIAVI 最新创新成果，迅速成为 400G 光学模块和器件开发、验证和部署的标杆。

在寻求全面覆盖 400G QSFP-DD 的所有挑战时，无论是传统客户侧接口的需求，还是新兴的新型相干接口的需求，ONT 都有合适的应用。

PAM-4 对于 50Gbps 信道和单波长 100G 以太网仍然是一个新的信号。对于成熟和低成本的 10Gbps 和 25Gbps NRZ 信道的使用肯定不会在一夜之间消失。但是，PAM-4 技术（模拟和数字实例化）的成熟使这种新的信号方法走在了最新的高速以太网实施的前沿。事实上，电气和电子工程师协会 (IEEE) 已经批准 PAM-4 作为综合 802.3bs、802.3cd 和 802.3ck 标准中所有 50Gbit、100Gbit、200Gbit 和 400Gbit 以太网标准的首选信号。此外，“单波 100G”组等 MSA 正在组织起来，为整个数据中心生态系统填充 PAM-4 信号的含义。

VIAVI Solutions

北京 电话: +8610 6539 1166
上海 电话: +8621 6859 5260
上海 电话: +8621 2028 3588
（仅限 TeraVM 及 TM-500 产品查询）
深圳 电话: +86 755 8869 6800
网站: www.viavisolutions.cn

qsfp-dd-moduletesting-wp-opt-nse-zh-cn
30192848 901 0420