VIAVI Solutions

ホワイトペーパー

QSFP-DD トランシーバーのテスト

400G および QSFP-DD

QSFP-DD 光トランシーバーは、400G クライアントインターフェイス用のメインストリームフォームファクターです。このホワイトペーパーでは、トランシーバー開発/設計、ネットワーク要素メーカー、エンドユーザー向けに、QSFP-DD トランシーバーのテスト、トラブルシューティング、検証を成功させるための主要素について説明します。

クライアントインターフェイスの速度は、10 年に 10 倍以上の割合で一定して増加し続けています。現在(2020年Q1)、QSFP28 インターフェイスを介して 100GE が広くサービス展開されており、400G のサービス展開の初期段階にあります。IEEE[1] は、2017 年 12 月に正式に規格化された 802.3.bs の一部として 400G イーサネットクライアントインターフェイス規格を開発しました。早期導入者は CFP8[2] フォームファクターを使用しましたが、より広い市場では、広く採用されている QSFP28 とある程度の下位互換性がある QSFP-DD[3] の採用に集中しています。

イーサネットは幅広いアプリケーションに対応し、さまざまな PMD (物理メディアに依存) 選択肢に利用でき、1 つの「QSFP-DD」スロットで幅広いアプリケーションをサポートし、数メートルのパッシブメタル線 DAC ケーブルから 80km のコヒーレントベース ZR にまで到達できます。また、OSFP[4] フォームファクターに注力している企業も少数ですがあります。この製品は、広く普及しているわけでも下位互換性があるわけでもありませんが、電気信号のインテグリティと熱管理という点ではいくつかの利点があります。以下の QSFP-DD に関する記載内容の大部分は、OSFP および OSFP ベースアプリケーションの多くをサポートする VIAVI ONT ファミリーにも当てはまります。[4]

400G は、電気トランシーバーとホストインターフェイス、および電気または光 PMD の両方に対して、高次 (PAM-4) 変調を採用しています。PAM-4 変調は、特定の帯域幅のデータ容量を最大化するために採用されましたが、複雑さとパフォーマンスに大きな課題をもたらします。これは、信頼性の高いデータ伝送を可能にするために、リンクに前方誤り訂正 (FEC) 符号化が必要であることを意味します。

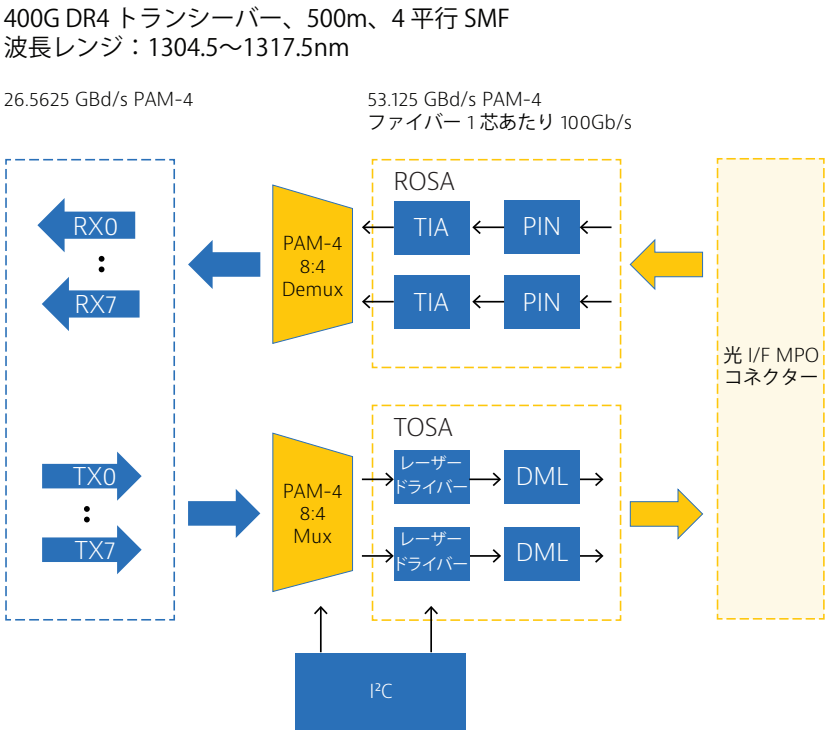
1. http://www.ieee802.org/3/bs/
2. http://www.cfp-msa.org/documents.html
3. http://www.qsfp-dd.com/
4. https://osfpmsa.org/

QSFP-DD を選ぶ理由

100G イーサネットは、CFP プラグ可能トランシーバーベースの初期設計にて 2008 年にサービス展開が開始されました。第 2 世代システムは、CFP2（またはある大手機器メーカー製の CPAK）に移行した後、広く普及しコスト効果が高い QSFP28 の大量採用に落ち着きました。CFP4 は QSFP28 の少し前の課題でしたが、数々の要因により QSFP28 は 100G の大幅な増加をもたらしました。業界は「フォームファクター」戦争を意識しており、400G のマルチステップのフォームファクターの進化に伴う複雑さとコストの問題を最小限に抑えたいと考えていました。CFP8 の早期採用者は 400G の開発と検証が可能でした。しかしながら、密度、電力、コスト、および「互換性」のニーズを満たしていなかったため、業界はすぐにターゲットフォームファクターとして QSFP-DD に注力するようになりました。それに取って代わるものとして OSFP が提案されました。それは、優れた技術ソリューションを提供しましたが、レガシートランシーバーのサポートという切迫したニーズを満たすことができませんでした。原理上、QSFP-DD ソケットは従来の QSFP-28 光トランシーバーをサポートできます。これにより、ベンダーは、現在 100G プラグで出荷可能な「400G 対応」ネットワーク要素を出荷でき、フィールドでのアップグレードは簡単なトランシーバー交換になります。

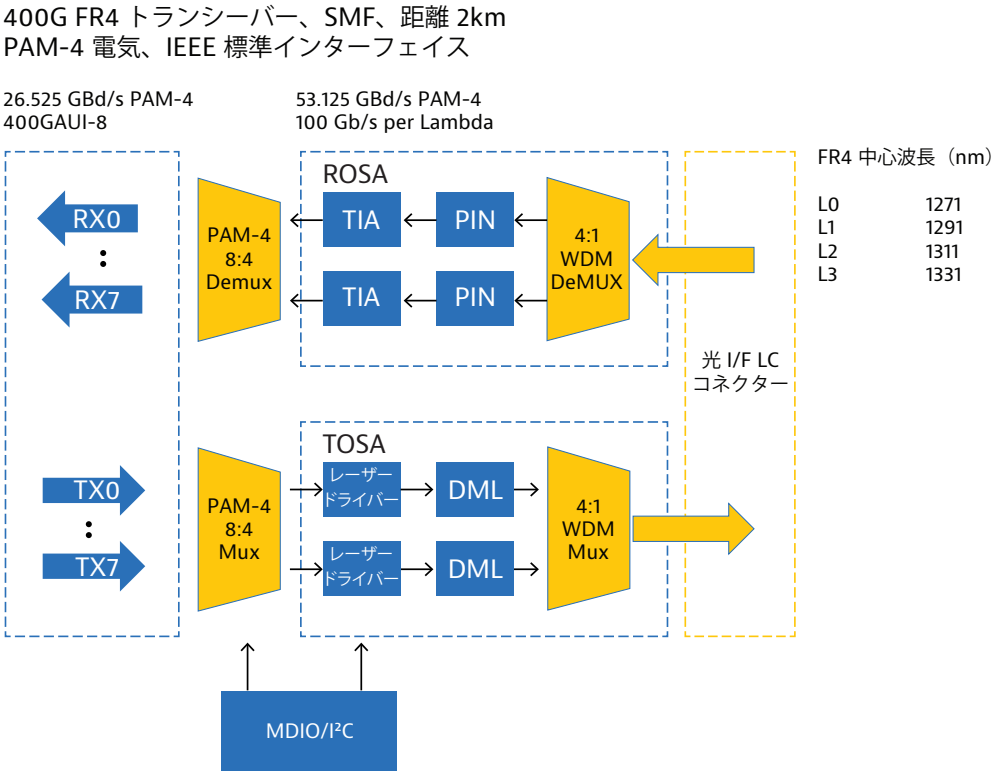
400G への移行に必要な帯域幅、電力、冷却の増加ニーズに対応するために、既存の QSFP28 コンセプトにいくつかの機能拡張が行われました。これには、高速電気回線の倍増（25Gbps NRZ の 4 レーンから 56Gbps PAM-4 の 8 レーンまで）と、内部容量の増加と熱性能の向上を実現するためのトランシーバー「ノーズ」の延長が含まれます。さらに、トランシーバーコントロールインターフェイスを強化し、CMIS 4.0 [5] 規格へアップグレードするための作業も行われています。

DR4 は、2020 年にサービス展開された最も一般的な 400G クライアント光インターフェイスの 1 つとなるでしょう。これは、400G

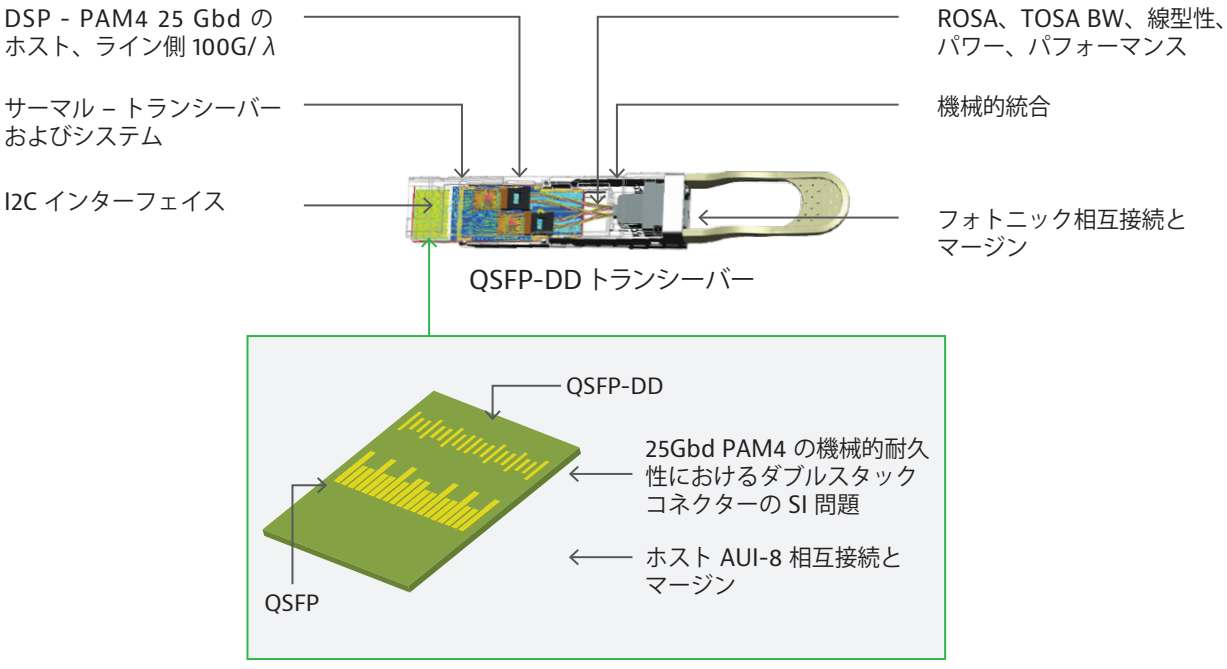


5. http://www.qsfp-dd.com/qsfp-dd-msa-group-announces-updated-specificatio/

を各シングルモードファイバー上での 4 つの 100G 信号として伝送します。これは、企業内での幅広いアプリケーションがあります。500m の到達距離をサポートし、個々の 100G イーサネットリンクに分割する機能は、高密度 100G ソリューションとして魅力的です。これにより、ポートカウント密度を 4 倍にすることができます。



FR4 インターフェイスは、通信事業者を含む幅広いアプリケーションにも対応します。1 本のシングルモードファイバーで 2km という長いリンクバジェットが可能になります。400G は、それぞれわずかに異なる波長の 4 つの 100G 信号にて伝送されます。



400G PMD トランシーバー（物理メディア依存）

PMD	到達距離	アプリケーション	テクノロジー
DAC	2〜3m	ラック内およびサーバー	パッシブメタル線ケーブル、50G PAM-4 電気
SR8	100m	エンタープライズ	パラレルマルチモード。50G/λ– PAM-4
DR4	500m	データセンターとエンタープライズ	並列シングルモード、100G/λ– PAM-4
FR4	2km	大規模データセンター	シングルモード、100G/λ PAM-4
LR8	10km	通信範囲	シングルモード、100G/λ PAM-4
ZR	80km	メトロと DCI	シングルモード/コヒーレント、PAM-4

QSFP-DD トランシーバー – 規格とテーマ

上記の参考資料のとおり、多くの規格および MSA が適用されます。また、基本的な IC 評価からトランシーバーハードウェア統合、ソフトウェア、ファームウェア、ベンダーの選択と認定まで、開発サイクルの各段階でクリティカルなテストが何かを理解することも重要です。本番環境にも、独自の主要テスト要件があります。

プラグ可能な光ファイバーの設計、テスト、検証、製造、配備を成功させるには、IEEE、CMIS、QSFP-DD、MSA、OIF などの主要資料を確実に理解する必要があります。QSFP-DD は、電子機器、光学機器、機械、熱管理、ファームウェアを統合した優れた製品です。トランシーバーの配備を成功させるには、すべてが連携して動作する必要があります。

相互運用性

イーサネットクライアントインターフェイスエコシステムの優れた利点は、IEEE などの規格団体が推進する確固として明確な規格があることです。これにより、マルチベンダエコシステムを「エンジニアリングされた」リンクに頼ることなく相互運用することができます。

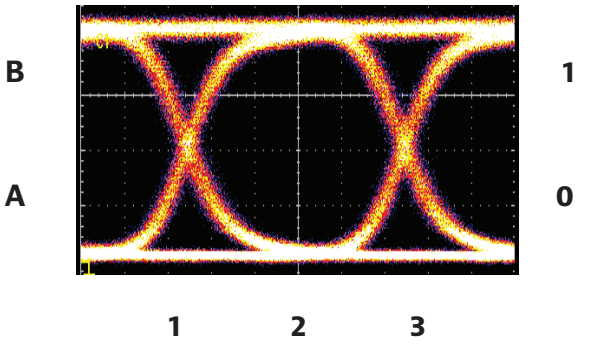
この相互運用で鍵となるのは、ホストトランシーバーと光インターフェイスの両方です。ホストからトランシーバーへのインターフェイスは、主に次の 3 つの領域に関係します。

- チップからトランシーバー（C2M）までの高速データパス（AUI）には、信号インテグリティや信号イコライゼーションなど、複数の課題があります。FEC バジェットの一部はリンクのこの部分に割り当てられますが、このインターフェイスに問題があると、リンクに重大な問題が発生する可能性があります。不適切な「調整済み」リンク（イコライザとチャネルの観点から）を使用すると、ランダムバーストや、たまに発生するビットスリップの最悪のケースなど、トラブルシューティングが困難になる問題を引き起こす可能性があります。
- トランシーバー管理 – この I^2C ベースのインターフェイスは、SFF-8636 から 100G QSFP28 までの基本的なメモリマップ管理から、高性能ステートフル CMIS 4.0 まで進化しました。この進化は、エコシステムにとって非常に困難なものであり、CMIS 4.0 資料の記載内容を十分に理解していることが堅固で安定したトランシーバー管理の鍵となります。
- トランシーバー電源 -- トランシーバーの電力需要は、100G での数ワットから、DCI アプリケーション用のプラグ可能なコヒーレント（QSFP-DD ZR）トランシーバー用の 20W に近い値まで徐々に増加してきました。このため、電源の堅牢性と安定性が求められます。さらに、トランシーバー起動時の電力需要の動的かつ一時的な性質に対応できる必要があります。

これらの領域はすべて密接に関連しており、問題のないトランシーバーの動作を保証するために全体として（特に CMIS 4.0 トランシーバー管理に関しては）扱う必要があります。

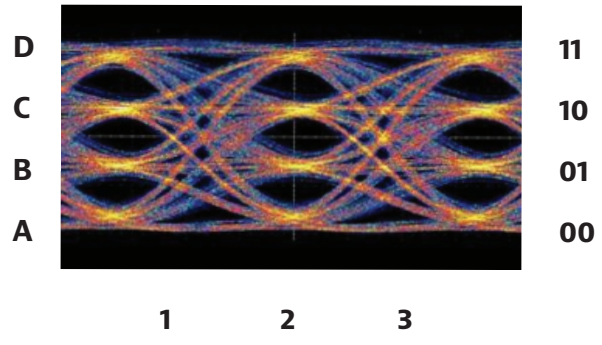
PAM-4

電気（トランシーバーからホストへのインターフェイス）リンクと光（電気）リンクの両方に、PAM-4　変調が採用されました。この高次変調方式では、単位時間に送信されるビット数を　2　倍にすることができます。NRZ　テクノロジーは広く採用され、高速化用に成熟していますが、SERDES PAM-4 は比較的新しいテクノロジーであり、複雑で困難なものです。当社は NRZ リンクの誤り解析に豊富な経験があります。しかし、それでも 100GE で使用されている 10G から 25G NRZ レーンへの移行には問題がありました。したがって、PAM-4への移行は、業界全体の大きな課題となると予想されます。これは、FECベースのリンク（常にバックグラウンドエラー率）およびはるかに複雑なチャネルイコライゼーションの使用によってさらに複雑になります。PAM-4 は、広く普及している 25G NRZ よりも桁違いに複雑であると言えます。



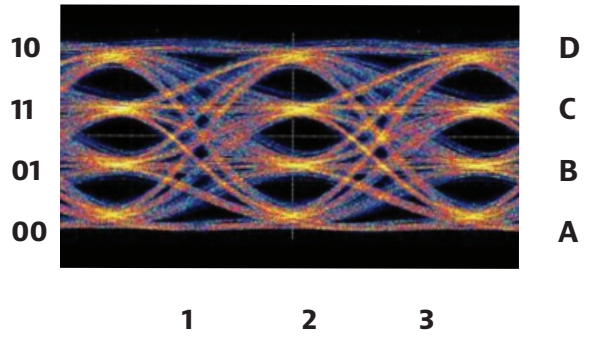
NRZ 変調：

- 1 クロックサイクルあたり 1 ビット：
  - 電圧 A =「0」
  - 電圧 B =「1」



PAM-4 変調、線形（非グレイ）コーディング：

- 1 クロックサイクルあたり 2 ビット
  - B と C の間の誤った決定 -> 2 エラービット



PAM-4 変調、グレイコーディング：

- 1 クロックサイクルあたり 2 ビット
  - B と C の間の誤った決定 -> エラービットは 1 ビットのみ

NRZ および PAM-4 シグナリングを示す画像

FEC

エラーのない PAM-4 伝送を提供できるコンポーネントを開発することは非常に困難なので、開発者は電気トランシーバーインターフェイスと光トランシーバー間インターフェイスの両方を保護するFECを使用しました。伝送チャネルとコンポーネントの両方のエラーメカニズムを理解し、FEC ロジック（エンコーディングと受信の両方）側の「コスト」のバランスを設定するように細心の注意を払っています。FEC の「コスト」には、電力を消費し、任意のリンクにレイテンシーを追加する追加回路が含まれます。

DSP およびイコライザ

400G を採用するにあたり、「強力な」電気レシーバーイコライザーの概念を使用して、「最悪ケース」のトランスミッタと「最悪ケース」のチャネルのパフォーマンスを低減することが決定されました。これにより、PAM-4 レシーバーの入力で PAM-4 アイが閉じられる可能性があります。したがって、PAM-4 レシーバーには、TX とチャネルの影響を均等化し、クリアアイを回復して特定のシンボルの正しいデコーディングを実現できるようにするために、強力かつ複雑なレシーバーが必要となります。イコライザーの複雑さは、ほとんどの場合、DSP ベースのソリューションを実装する必要があることを意味します。これは、電力、レイテンシー、複雑さ、エラー性能、および管理/制御に影響を与える可能性があります。DSP イコライザーは強力ですが、機能の複雑さがタップの最適な設定などを見つけるのに課題を引き起こす可能性があります。さらに、イコライザーは DSP ファームウェアと制御APIの背後に隠されていることが多く、ユーザーから見えない抽象的なものになっています。TDECQ[6]の測定では、さらに課題が発生します。この測定は複雑で一貫性がないため、自由に相互運用するマルチベンダーエコシステムの課題がさらに増えることになります。

重要なポイント

常にエラーが発生します。リンクには常にバックグラウンドエラー率があります。エラー統計の「フィンガープリント」は非常に重要です。真のランダムエラーフローは、リンクの保護に使用される FEC と一般的に互換性があります。しかし、バースト、スリップ、およびその他の確定的な問題により、FEC誤りエラー訂正機能が大幅に低下する可能性があります。実際のリンクでは、エラーは、電気チャネルノイズと光チャネルノイズ、クロストーク、信号インテグリティの問題、バースト、ビットスリップ、および誤って設定されたイコライザーでのエラー増殖も含む複雑な組み合わせになる可能性があります。

最終的に重要なのは、特定の誤りフィンガープリントが起こった場合の FEC の動作です。マージンはどのくらいありますか？パケットがドロップされるまでの時間はどれくらいですか。リンクの劣化を確認するために、長期的なパフォーマンスを予測できますか。エラーの根本原因は何ですか？

個々の PAM-4 シンボルの誤りバイアスからビットスリップ特性のバーストの分析まで、誤りフィンガープリントを調べるのにはいくつかのツールを使用できます。誤りバイアスの理解は、クロックバリエーションやスキューなどのツールによってさらに深めることができます。

PAM-4 シンボル解析を使用して、誤り分布に「レベル」バイアスがないことを確認できます。レシーバーフォトニック AGC などの主要なフォトニック素子の安定性は、PAM-4 の誤り分布を観察中の光パワーの変動（アッテネータ経由）によってさらに検証することができます。

誤りバーストを完全に調査し、ビット（またはシンボル）スリップではなくバーストであることを検証することが重要です。スリップはDSP（および関連ファームウェア）に関連していることが多く、FECでは修正できません。通常のテストセットでは、従来の信号インテグリティまたはノイズの問題によって発生するバーストと、クロックおよび位相の感度に関連するバーストを区別することはできません。そのため、QSFP-DD でエラーの性質と根本原因を調査する際には、多数の新しいツールと技術を導入する必要があります。

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

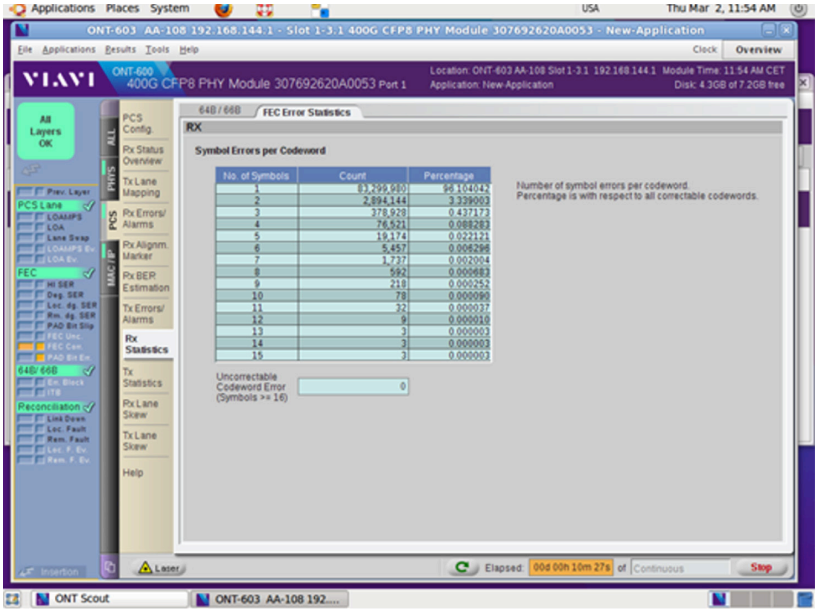
最も単純なトップレベルビューは、5440 ビット FEC コードワード（KP4 FEC）あたりのエラー 10 ビットシンボルの数を調べることによって得ることができます。通常、単調な分布はシンボル数あたり約 1 桁減少すると予想されます。つまり、エラーシンボル/コードワードが増加するたびに、エラーカウントが一桁減少することが予想されます。長いテールピークまたは分離されたピークは、一部の非ランダム（システム）な原因を示唆しています。また、エラーが発生したシンボルの数も、測定時間の x10 倍で増加することが予想されます。したがって、10 秒後にコードワードあたり 10 個のエラーシンボル数を確認した場合、100 秒以内に 11 個のエラーシンボルカラムにカウントが表示されることが予想されます。

このような経験則を使用して、修正不可能なエラーが発生するまでの時間を推定できます（コードワードあたり 16 個以上のエラーシンボルで発生）。例えば、テスト時間が 100 時間経過した後、最大 12 個のエラーシンボル/コードワードが検出された場合、次のような近似値が予想されます。

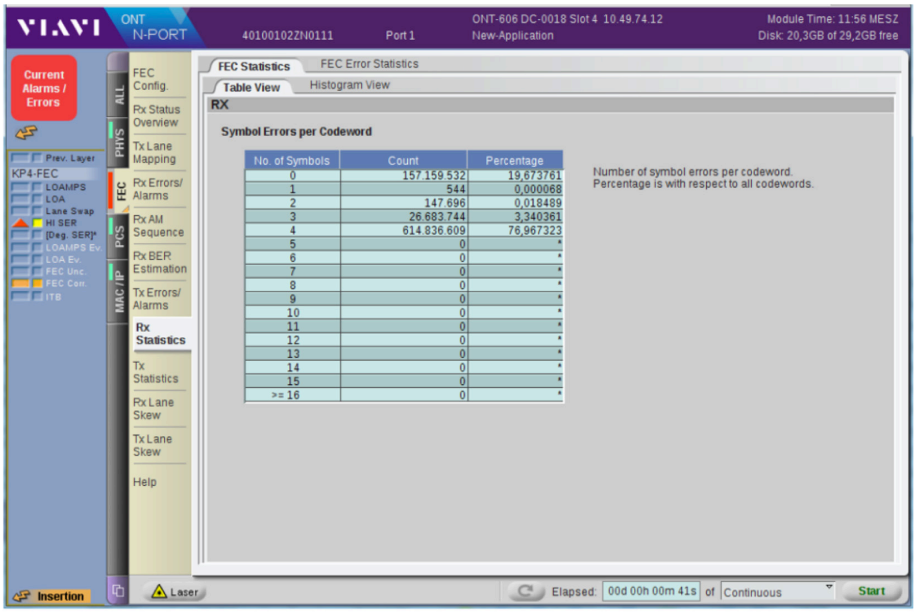
エラーシンボル	時間	注記
12	100 時間	測定
13	1000 時間	推定
14	約 420 日	
15	約 11 ½ 年	
16（修正不可能なエラー）	約 114 年	100 年を過ぎて初めてパケットをドロップ

FEC - エラーシンボル/コードワード

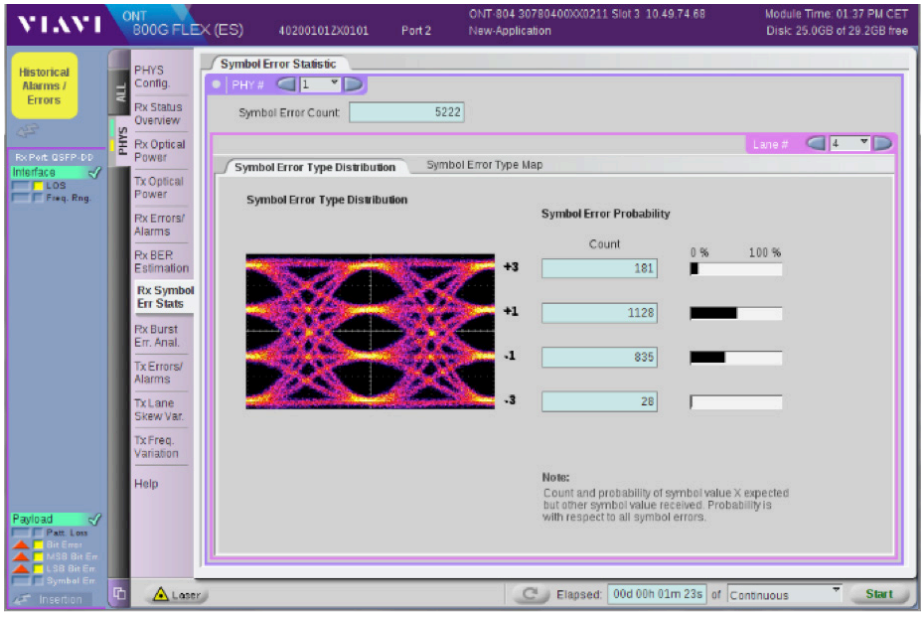
次の例では、10 分間隔で重大なエラーが発生するように、大幅に減衰された 400G の光リンクで ONT を実行したままにしています。これは、準拠リンクで期待されることです。ご覧のように、分布は一般的に単調です。エラーシンボルごとのカウントはドロップしますが、12 個のエラーシンボル/コードワードからのテールがわずかに長く表示されます。この場合、修正されていないコードワードが原因で、リンクがパケットをドロップする可能性が非常に高くなります。



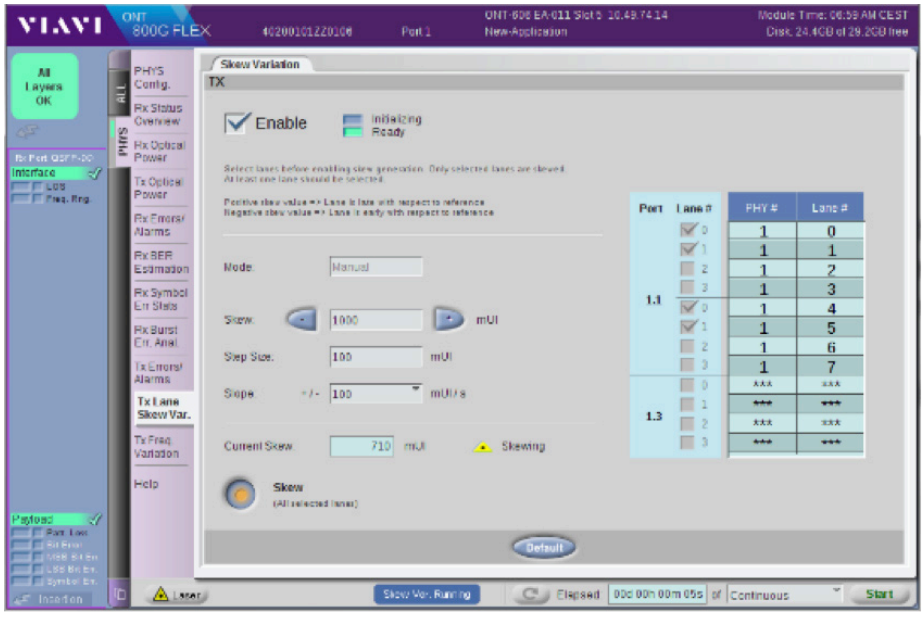
次のスクリーンショットは、重大な問題が発生しているケースを示しています。FEC には大きなマージンがありますが（コードワードには最大4個のエラーシンボルが表示されます）、このディストリビューションは単調ではなく、このシステムにはエラーの原因になるものがあることを示しています。100G リンクのこの例は、FEC ロジックと電源のインテグリティに負荷をかけて検証するための幅広い FEC エラー分布を作成できる特殊な VIAVI ONT アプリケーションによって生成されたものであることに注目してください。



ONT には、シーケンス全体のエラー分布とパターンを分析し、PAM-4 シンボルごとにエラープロファイルを追跡する機能があります。



ダイナミックスキューバリエーションは、QSFP-DDトランシーバーに負荷をかけて検証するためのパワーツールです。IEEE802.3への準拠、および DSP と関連ファームウェアの一般的な安定性を検証するために使用できます。これは、個々の電気と光のレーンペアがまったく異なるクロックドメイン上にある可能性がある DR4 トランシーバーでは特に重要です。



上のスクリーンショットは、PAM-4のダイナミックスキューアプリケーションを示しています。送信レーンの相対的なタイミングをUIの一部に正確に制御し、「ヒットレス」位相移行を維持する機能は、クロストークやDSPベースのファームウェアタイミング問題などの困難な問題を解決するための鍵となります。

ダイナミックスキュー（またはスキューバリエーション）は、あらゆるパラレルレーン通信システムにとって重要なテストです。信号の完全性テストおよび検証（クロストーク）にアプリケーションを備えており、PAM-4 SERDES 内の FIFO および CDR の性能に負荷をかけて検証するためにも使用できます。

また、さまざまなスキューレートを使用して、信号の完全性とクロストークの問題を調査することもできます。これには、H/W チームおよび SI チームと連携した幅広いアプリケーションがあります。レーンのタイミングは、攻撃側のレーン遷移が犠牲レーンの PAM-4 アイの中央で発生するように調整できます。

PAM-4 シグナリング（信号マージンが低いため）は、従来の NRZ よりもクロストークの影響をはるかに受けやすくなっています。QSFP-DD（特にホストコネクターの周囲）の密に梱包された範囲では、高速 PM-4 レーンは近接して配線されており、注意しないと信号クロストークの問題が発生する可能性があります。通常、BER テストセットは固定位相でパラレルレーンを実行するため、SI ストレスに対してで「ワーストケース」アライメントが発生しない場合があります。ダイナミックスキューでは、アグレッサーレーンを相対位相でスイープして、ワーストケースの位相移行時でも問題が発生しないことを完全に検証することができます。エンドユーザーは、特定の位相オフセットでエラーが発生したかどうかを確認するだけです（通常、アグレッサーレーンがビクティムレーンの「eye」の中央でレベル遷移している場合）。

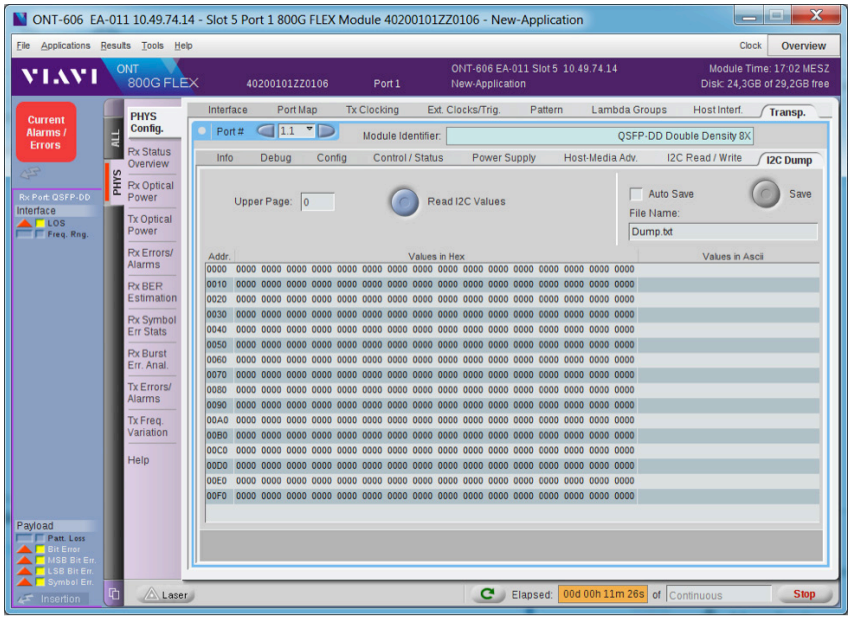
最新の SERDES は、IC ファブリック内でさらに処理を行う前に、さまざまな FIFO バッファを使用して信号をリタイミングおよびリアライメントします。リアライメントは、マスタークロックソース（通常は CDR を介したマスターレーン）からリクロックされる一連の FIFO バッファを使用します。

システムが正しく設計または実装されていない場合、マスター（CDR リファレンスレーン）と他のレーンの間で位相の変動や変化が原因で、ミスアライメントや FIFO 内のスリップが発生する可能性があります。これはビットスリップとして現れます。これは、ONT の高度なエラー分析では、従来のテスト装置で見られるエラーバーストではなくビットスリップとして追跡できます。ダイナミックスキューアプリケーションを使用すると、ONTはSERDESのCDR/FIFOの性能に意図的に負荷をかけ、スキュー（範囲とレート）によって強制的に不具合を発生させようとします。これは、ONT の高度なエラー分析と組み合わせることで、SERDES テスト用の非常に強力で完全なテストシステムを実現し、400GE リンクで時折ビットスリップが発生する非常に困難な問題を迅速に解決するために使用できます。ONT PAM-4 のダイナミックスキューは、これらのエラーを強制的に診断および根本原因の解決に役立てることができます。

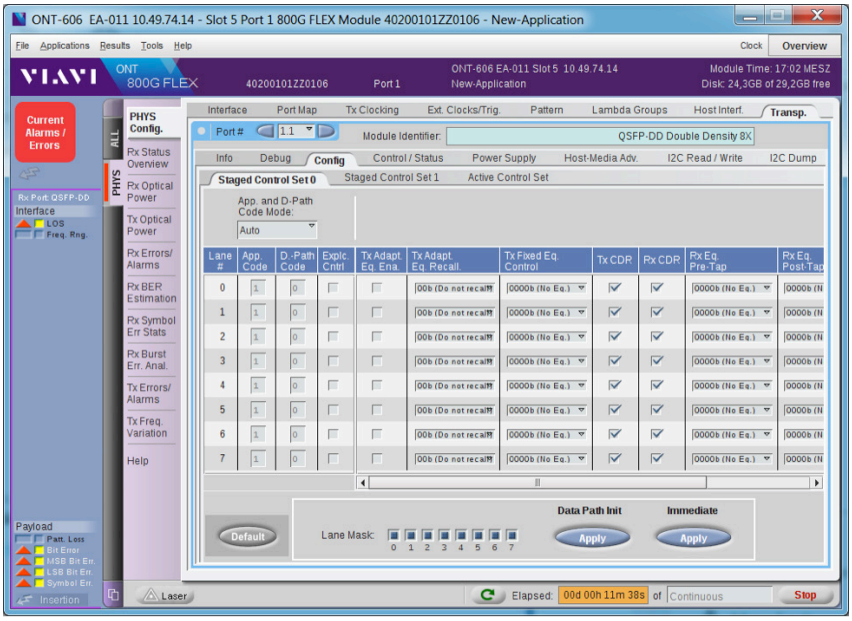
一般的な QSFP-DD 制御画面

トランシーバー管理は、非常に基本的なレジスタベースのシステムである SFF 8636 から CMIS 4.0 へと徐々に進化してきました。CMIS 4.0 は、400GE 以上の複雑なトランシーバーのニーズに対応するように設計された、包括的でステートフルなトランシーバー管理システムです。

I^2C 制御インターフェイスを介したトランシーバーと、電源ピンおよび制御ピン、およびデータパス間の緊密な相互作用は、堅牢で安定したトランシーバー動作に不可欠です。トランシーバーの複雑さ、特にトランシーバー DSP のデータパスイコライゼーション関連の複雑さが増すと、ホストとトランシーバー間の制御のセットアップと実行についてより包括的な知見が必要になります。コマンド、操作、およびスロット動作の正しい順序は、CMIS 4.0 で厳密に調整する必要があります。注意しないと、トランシーバーは問題なく 1 つのホストスロットで動作しているように見えても、別のホストスロット（コマンド、電源、およびデータパスのタイミングにわずかな違いがあります）では動作が不安定になる場合があります。また、さらに悪い状態では、エラーレートが増加し、まれに発生し解決が困難な問題や、ビットスリップが発生する可能性があります。I^2C を介して CMIS コマンドを統合する ONT やトランシーバーの電源制御およびデータパスの状態などのツールは、問題のデバッグと解決だけでなく、異なるホストでのトランシーバーの堅牢性に負荷をかけて検証する上でも、非常に役に立ちます。



上の画面は、メモリの最初のページのメモリダンプを示しています。これにより、QSFP-DD EEPROM に正しい値が保存されているかどうかをすばやく確認できます。ブランクまたはランダムなデータは、デバイスがまだ初期化されていないことを示している場合があります。



トランシーバー管理アプリケーションの高度なアプリケーションの中には、トランシーバーの電気パラメータをあいまいさなくはっきりと正確に制御できるものがあります。

まとめ

QSFP-DD トランシーバーは、複雑なファームウェアと一緒に用いられ、電子工学、光通信、機械工学、および熱工学の驚異と言えるものです。400G ネットワークテクノロジーの広範な導入には、健全なマルチベンダーの QSFP-DD エコシステムが不可欠です。これは、従来の 100G トランシーバーに比べて技術の進化と革命の両方を表すものです。また、電気的および光の PAM-4 シグナリング、リンクエラー制御に FEC を使用すること、および CMIS 4.0 の新たな複雑さにより、新たな課題が生じています。

これらの課題をさらに大きなものにしているのは、超大規模ユーザーの規模と導入のニーズに対応して期待される価格の低下です。価格の期待に応えるためには生産量と処理能力を満たす必要があります。また、PAM-4 の新たな課題に対応するためのカバレッジと分析も必要です。

VIAVI ONT ファミリーは、20 年以上にわたってトランシーバーの検証およびテストアプリケーションをその DNA に組み込んでいます。高度なエラー分析やダイナミックスキューなどの従来の 100G アプリケーション系統に、CMIS 4.0 デバッグや PAM 4 シンボル分析などの VIAVI の最新の技術革新を併せて、400G 光伝送の開発、検証、展開のためのリファレンスとして、すぐにその地位を確立しました。

400G QSFP-DD のすべての課題に完全に対応したい場合は、従来のクライアントインターフェイスのニーズにも、新しいコヒーレントインターフェイスのニーズにも、ONT には適切なアプリケーションがあります。

PAM-4 は、50Gbps チャンネルとシングルラムダ 100G イーサネットの両方において今だ新興の信号方式です。10Gbps および 25Gbps チャンネルにおいて確立し、低コストの NRZ の使用は、一夜にして消えることはありません。しかし、PAM-4 テクノロジー（アナログおよびデジタルのインスタンス化）の確立により、この新しいシグナリング方式は最新の高速イーサネット実装の最前線に進出しています。実際、米国電気電子技術者協会（IEEE）は、50Gbit、100Gbit、200Gbit、400Gbit のすべてのイーサネット規格で、802.3bs 、802.3cd、802.3ck 規格内の推奨信号方式として、PAM-4 を承認しています。さらに、100G per Lambda MSA グループでは、データセンターエコシステム全体にわたって PAM-4 シグナリングを実装するように組織化しています。

VIAVI Solutions

〒163–1107
東京都新宿区西新宿6–22–1
新宿スクエアタワー7F

電話：03–5339–6886
FAX: 03–5339–6889
Email: support.japan@viavisolutions.com

qsfp-dd-moduletesting-wp-opt-nse-ja
30191242 901 0420