

# 次世代AIクラスター スケーリングにおける 448Gインターコネクットの パフォーマンス評価 と変調戦略

AIによって、データセンターインフラストラクチャへの投資が急速に進んでいます。特にハイパースケール事業者がAIクラスターのサイズとパフォーマンスを拡大しているため、このような動きが促進されています。トレーニングワークロードと推論アプリケーションが爆発的に増加したことで、数千もの相互接続されたコンピューティングノードをサポートする高帯域幅インターコネクストという新たな要件が生まれました。このような要件を満たすには、シリアルリンク技術をレーンあたり448Gに進化させ、将来的にポートあたり3.2Tbpsのイーサネットインターフェースを実現する必要があります。

AIクラスターをこのレベルに拡張すると、物理層に厳しい要件が発生します。この物理層では、インターコネクストファブリックによってプロセッサ、アクセラレーター、ネットワークインターフェース間で達成可能な帯域幅密度が決まります。インターコネクストファブリックとして光ソリューションと銅線ソリューションの両方が検討されていますが、パッケージング、PCBレイアウト、コネクタの移行における設計上の課題を克服できる場合は、やはり銅線インターコネクストが依然として魅力的です。候補となる448G変調フォーマットをサポートする帯域幅が必要であるため、従来のPCBインターコネクスト、フライオーバーインターコネクスト、従来のコネクタフォームファクターがこのデータレートで使用できるかどうかはいまだに未解決の問題です。

本ホワイトペーパーでは、3種類の変調スキーム（PAM-4、PAM-6、PAM-8）を用いて448Gにおける銅線インターコネクストの性能を調査した研究を紹介します。各スキームは、損失許容度、直線性、イコライゼーションの要件に対して異なる意味合いを持ち、コネクタの設計とシステムアーキテクチャーに影響を与えます。

この調査では、コンポーネントパッケージとQSFPモジュール間のインターコネクストオプションを検討し、AIデータセンターで信頼性の高い448G環境を実現するために満たす必要があるシグナルインテグリティ要件の概要を示します。



# 448Gチップ・モジュール間アーキテクチャーにおけるシグナルインテグリティの課題

AIクラスターのスケールアップによって物理層に厳しい要件が発生します。物理層では、各インターコネクタがより高い帯域幅を提供するとともに、信号劣化を最小限に抑える必要があります。PCB直接配線とフライオーバーTwinaxチャンネルを比較した、以前のMolexのシミュレーション調査では、トレースとビアを用いた従来のPCB配線の限界が示されています。PCB直接配線はシンプルですが、距離が長くなると挿入損失が大幅に増加し、その結果、PCBビアで反射損失が増加するため、80~90GHzで顕著な挿入損失ロールオフが発生します。フライオーバーTwinaxコネクタ設計でも同様の挿入損失ロールオフが見られますが、総挿入損失ははるかに低いため、フライオーバーケーブル配線のチャンネルリーチはより長くすることが可能です。

コパッケージド銅 (CPC) は、PCBビアとコネクタビアを介した信号伝送中に発生する高損失を回避するためのソリューションとして、大きな関心を集めています。これらのモジュールは、パッケージ基板内のトレース配線に接続し、帯域幅を制限する要因となるPCBビアをバイパスして、Twinaxケーブルと直接インターフェースを取ります。このアーキテクチャーは、Chip-to-Module (C2M、チップ・モジュール間) と呼ばれており、図1で概説しています。

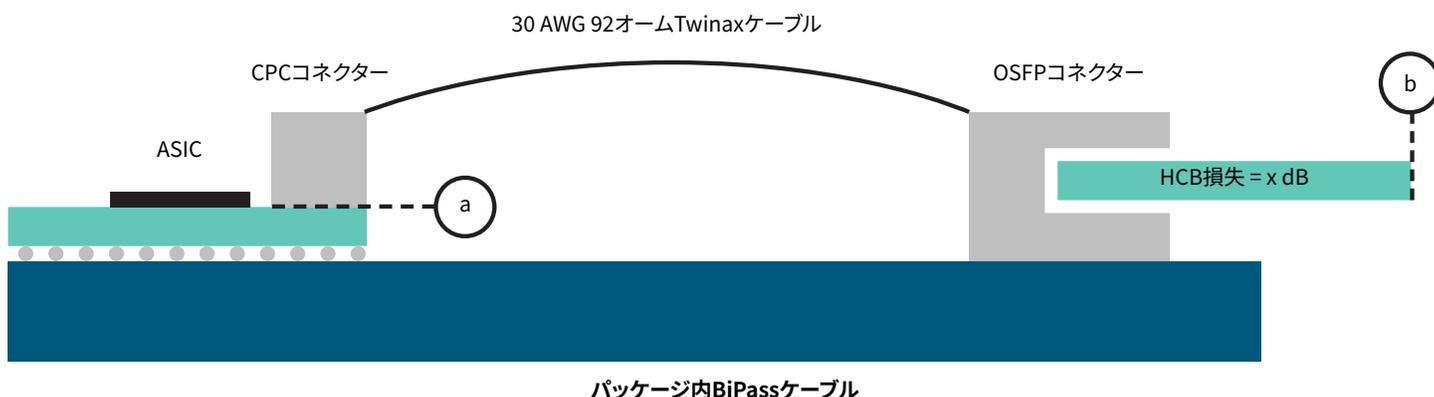


図1: コパッケージド銅コネクタと、MSA標準コネクタおよびモジュールに接続するフライオーバーケーブルを使用するC2Mアーキテクチャー

また、コパッケージド銅のコネクタでは、表面実装デバイス (SMD) のランドパターン、接点のスタブ長、パッケージからコネクタへのビアを経由した配線に関連するシグナルインテグリティの課題も生まれます。しかし、信号がコネクタを横断し、ケーブル配線に沿って通過すると、挿入損失が大幅に低減され、同軸コネクタ設計 (Twinaxなど) ではPCBインターコネクタよりもはるかに低い反射損失になる傾向があります。これは、QSFP/OSFPコネクタや銅線バックプレーンコネクタへのフライオーバーケーブルの場合に当てはまります。

シグナルインテグリティのパフォーマンス指標は変調フォーマットによって変わります。フォーマットが異なると、シリアルビットストリームの信号レベル間にそれに応じた電力マージンが必要となるためです。表1および表2は、対象とした各変調フォーマットのチャンネルリーチ、ビットエラー率 (BER)、信号対雑音比 (SNR) に対するチャンネル帯域幅の要件を比較したものです。AIデータセンターに導入されている現在の技術は、56GHzチャンネル帯域幅の224Gbps-PAM-4であるため、相対値は、112GHzのチャンネル帯域幅が必要な448Gbps-PAM-4と比較したものになります。

| 変調    | 寸法 | 寸法あたりのビット数 | シグナリングレート、GBd  | 帯域幅、GHz           | 到達距離低下量、dB |
|-------|----|------------|----------------|-------------------|------------|
| PAM-4 | 1  | 2          | 225<br>(212.5) | 112.5<br>(106.25) | —          |
| PAM-6 | 2  | 2.5        | 180<br>(170)   | 90<br>(85)        | -4.44      |
| PAM-8 | 1  | 3          | 150<br>(142.5) | 75<br>(71.25)     | -7.36      |

帯域幅要件の低減

信号検出の困難化

表1:レーンあたり448Gのインターコネクต์における変調フォーマット候補のシグナリング特性とチャンネルリーチ

| 変調    | SNR = 19dBでのBER | Δ    | BER = 2.4e-5でのSNR | Δ    |
|-------|-----------------|------|-------------------|------|
| PAM-4 | 2.4e-5          | —    | 19                | —    |
| PAM-6 | 3e-3            | 125倍 | 22.6              | +3.6 |
| PAM-8 | 1.5e-2          | 625倍 | 25.1              | +6.1 |

表2:表1に示す各変調フォーマットのBERおよびSNRの要件

表2は、PAM-4変調の目標SNR値に対して必要なBERの比較と、その逆の比較を示しています。データは、インターコネクットの受信側で観測されるSNRが次のような複数の要因の影響を受けていることを示しています。

- 挿入損失
- 反射損失 (反射)
- クロストーク
- スキュー/ジッター



高帯域幅チャンネルにおける各変調フォーマットのBERおよびSNRの要件は、多値PAMを用いた場合、低帯域幅チャンネルでも、高密度化によって特定のBER目標を達成するには反射とクロストークをさらに低減することが必要になることを示しています。検出の改善、リンクのイコライゼーション、または誤り訂正を用いることで、SNR要件を緩和できます。

これらのデータレートでは、コネクタの機械的性質によってチャンネルの全体的な性能が制約を受ける可能性があります。スタブ長、はんだ形状、パッケージとコネクタ間の伝送、コネクタとTwinax間の伝送などの要因により、低周波数での反射や、高周波数での共振が生じることがあります。それらが重なって、高周波数での挿入損失にロールオフが生じ、チャンネル帯域幅の実質的な上限が決まります。コネクタモジュールのピン密度が高いと、パッケージとコネクタ間の伝送でクロストークのリスクも生じます。挿入損失に加えて、クロストークもインターコネクタの受信側でSNRとBERの値を決定する要素となります。

---

## 448Gインターコネクタにおける コパッケージド銅の使用に関する調査

以前のシミュレーション調査で観測されたシグナルインテグリティ特性を参考に、さまざまな変調周波数で448Gインターコネクタにおけるコパッケージド銅のシグナルインテグリティ特性を調査することにしました。変調フォーマット、損失、SNRおよびBERの上限に明確な関係が見られたため、代表的な導入環境において実際のチャンネルを使用した実験的調査を行うべきであると判断しました。

本ホワイトペーパーの以降のセクションでは、標準リンクアーキテクチャーを用いた448Gチャンネルのシグナルインテグリティ指標の分析について説明します。本調査では、次の目標を達成することを目指しました。

- PAM-6およびPAM-8シグナリングのSNR、BER、挿入損失の評価
- 挿入損失のロールオーバーに基づく帯域幅制限の判断
- C2Mアーキテクチャー内のインターコネクタにおいて識別された帯域幅範囲内のクロストークの検証
- C2Mアーキテクチャーにおけるホストからモジュール方向およびモジュールからホスト方向での信号伝播の比較

C2Mアーキテクチャーでは、実際のコパッケージド銅コネクタとOSFPコネクタを使用しました。PAM-6およびPAM-8シグナリングに対する挿入損失、BER、SNRマージンを測定して比較しました。

本調査で評価したリンクアーキテクチャーと、アーキテクチャーの各セクションの推定BER上限を図2に示します。このC2Mアーキテクチャーでは、300mmと500mmの長さのVSR (Very Short Reach) チャンネルをホストからモジュール方向とモジュールからホスト方向の両方で評価し、インターコネクタに沿ったクロストークと挿入損失の結果を得ました。次に、チャンネルリーチと変調フォーマットを変化させながら、SNRとBERをクロストークの存在下で測定しました。これらの結果から、PAM-6およびPAM-8変調をサポートするチャンネルのパフォーマンスを比較するのに十分なデータが得られました。

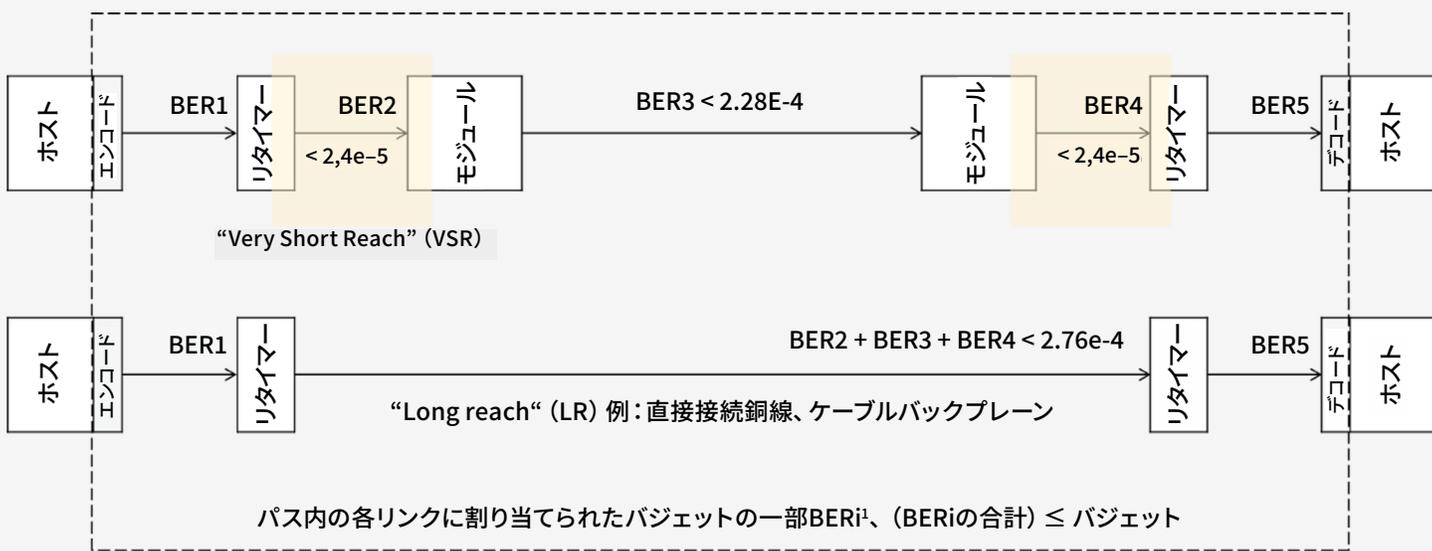
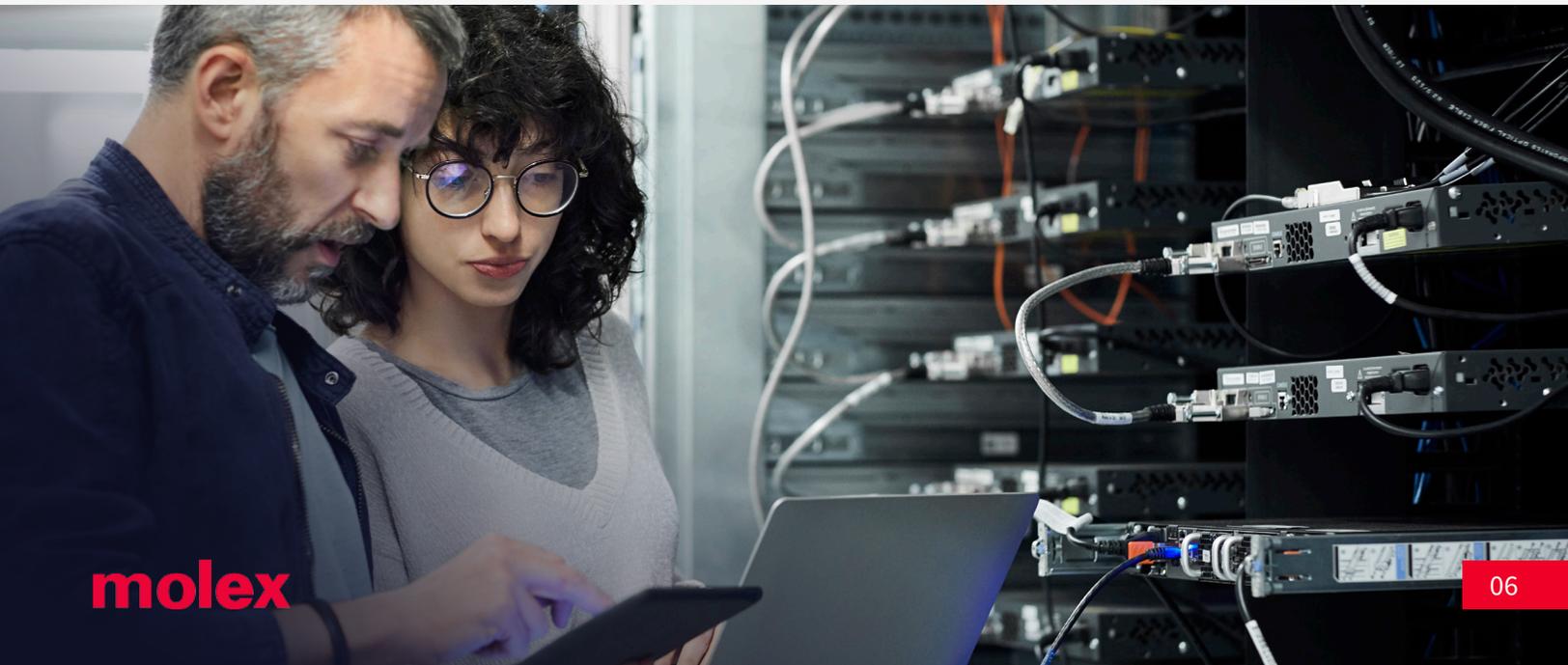


図2: 今回の調査で評価したリンクアーキテクチャー

<sup>1</sup> デコーダーの性能を損なうようなエラーが発生した場合、状況によってはリンクBERをBER<sub>i</sub>より小さくする必要があります。



# 調査結果と結論

## 挿入損失とクロストーク

図3は、ホストからモジュール方向およびモジュールからホスト方向の信号伝播の挿入損失曲線とクロストーク曲線を示しています。どちらの曲線も、90GHz付近でチャンネル帯域幅の上限が固定される強力な挿入損失ロールオフを示しています。これにより、これらのチャンネルが500mmチャンネルリーチでPAM-6およびPAM-8シグナリングをサポートできることが確認されました。

この結果は、特に最悪のケースでのクロストーク状態を示すために選択された、標準OSFPトランシーバーモジュール上のチャンネルTX6 (ホストからモジュール方向) およびRX6 (モジュールからホスト方向) に関するものです。クロストークレベルはチャンネル帯域幅全体で低く維持され、ホストからモジュール方向のPSNEXTは、80GHzに近づく周波数でも-60dBを超えた程度にしか上昇しません。ホストからモジュール方向のPSFEXTは、約75GHz以降に、時折-50dBを超える程度です。

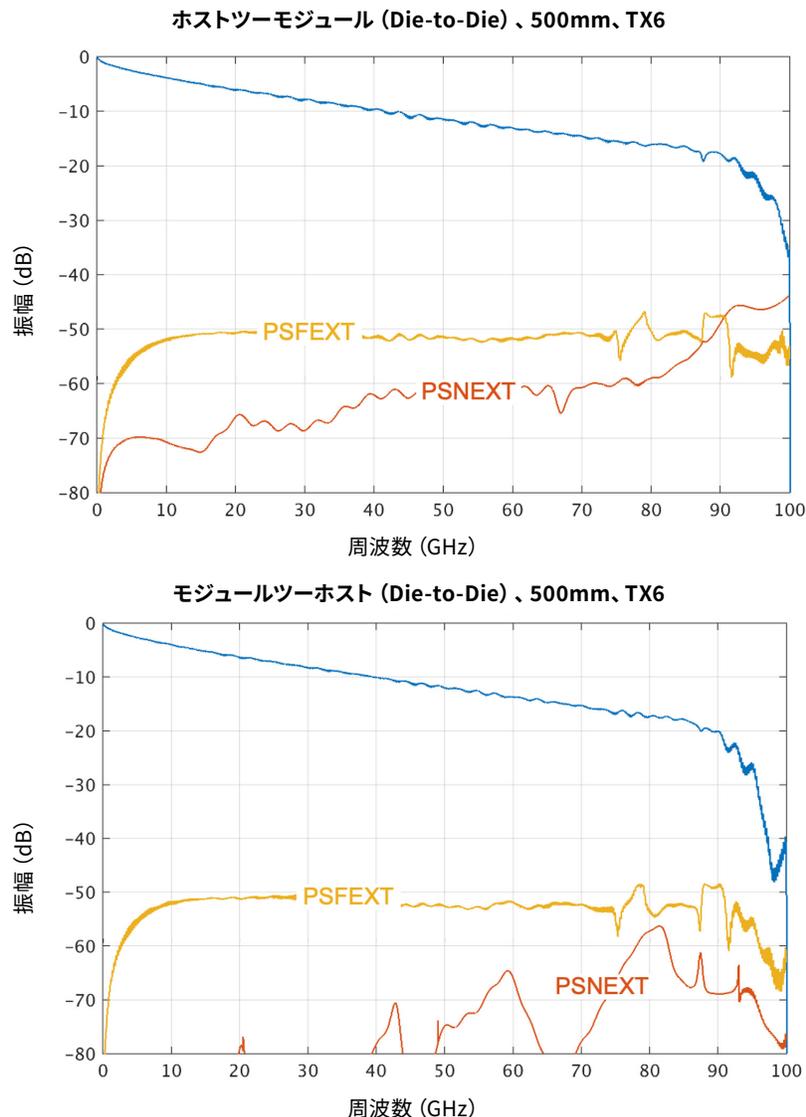


図3: ホストからモジュール方向 (上部パネル) およびモジュールからホスト方向 (下部パネル) のダイツードイ挿入損失およびPSFEXT/PSNEXTの結果。チャンネルの長さ = 500mm

## BERおよびSNRマージン

図3に示すように、これらのチャンネルで発生する可能性があるクロストークレベルを考慮すると、BERおよびSNRマージンの期待値を調査することが重要です。表3と表4は、ナイキスト周波数での挿入損失、観察されたBER、およびPAM-6およびPAM-8フォーマットで利用可能なSNRマージンを示します。この評価では、最悪のケースでのクロストークを示すチャンネルTX6/TX8およびRX6/RX7を選択しました。

| ホストツーマジュール |              | PAM-6             |             |        |                | PAM-8             |             |        |                |
|------------|--------------|-------------------|-------------|--------|----------------|-------------------|-------------|--------|----------------|
| ペア         | ケーブル長、<br>mm | シグナリング<br>レート、GBd | 挿入損失、<br>dB | BER    | SNRマージン、<br>dB | シグナリング<br>レート、GBd | 挿入損失、<br>dB | BER    | SNRマージン、<br>dB |
| TX6*       | 300          | 170               | 15.1        | 4.8e-7 | 1.7            | 145               | 13.7        | 1.8e-5 | 0.2            |
| TX8**      | 300          | 170               | 15.1        | 4.4e-7 | 1.7            | 145               | 13.7        | 1.7e-5 | 0.2            |
| TX6        | 500          | 170               | 16.5        | 5.4e-7 | 1.6            | 145               | 14.9        | 2.4e-5 | 0              |
| TX8        | 500          | 170               | 16.5        | 4.9e-7 | 1.7            | 145               | 14.9        | 2.1e-5 | 0.1            |

\*FEXTが最悪になるチャンネル

表3：ホストからモジュール方向の挿入損失、BERおよびSNRの結果

\*\*NEXTが最悪になるチャンネル

| モジュールツーマホスト |              | PAM-6             |             |        |                | PAM-8             |             |        |                |
|-------------|--------------|-------------------|-------------|--------|----------------|-------------------|-------------|--------|----------------|
| ペア          | ケーブル長、<br>mm | シグナリング<br>レート、GBd | 挿入損失、<br>dB | BER    | SNRマージン、<br>dB | シグナリング<br>レート、GBd | 挿入損失、<br>dB | BER    | SNRマージン、<br>dB |
| RX6*        | 300          | 170               | 16.5        | 3.9e-7 | 1.8            | 145               | 14.8        | 1.8e-5 | 0.1            |
| RX7**       | 300          | 170               | 16.9        | 4.2e-7 | 1.7            | 145               | 14.9        | 1.7e-5 | 0.2            |
| RX6         | 500          | 170               | 17.9        | 5.4e-7 | 1.6            | 145               | 15.9        | 2.3e-5 | 0              |
| RX7         | 500          | 170               | 18.3        | 5.3e-7 | 1.6            | 145               | 16          | 2.1e-5 | 0.1            |

\*FEXTが最悪になるチャンネル

表4：モジュールからホスト方向の挿入損失、BERおよびSNRの結果

\*\*NEXTが最悪になるチャンネル

結果は、PAM-6とPAM-8のどちらも、300mmおよび500mmの長さのOSFPコネクタを備えたこれらのチャンネルで使用することが技術的に可能であることを示しています。予想どおりPAM-8のBERは高いため、リタイマー/モジュールインターフェースでBER上限の $2.4e-5$ に対して残存するSNRマージンはほとんどありません（図2を参照）。これにより、クロストーク、外部電磁干渉 (EMI)、パワーレールの変動など、追加のノイズ源に対してリンクは脆弱になります。

従来のOSFPパドルカードインターフェースでは、必要な112GHzチャンネル帯域幅を下回る挿入損失ロールオフが発生するため、このインターフェースを使用して448Gbps-PAM-4変調を実現することはできません。このデータレートでPAM-4を使用する場合、ロールオフしきい値を増やすには、マルチソースアグリーメント (MSA) 標準の変更が必要になります。

これらの結果に基づき、PAM-6はBER、SNR、帯域幅マージンのバランスが優れているため、最も実現可能なオプションであることが明らかになりました。必要なチャンネル帯域幅は、C2Mアーキテクチャーで観察された挿入損失ロールオフと一致するため、OSFPコネクタとパドルカードモジュールのMSA標準を変更する理由になります。

### MSA標準の変更案

上記の結果に基づいて、パドルカードインターフェースに更新を加えると、レーンあたり448GのデータレートでPAM-6を実現するのにさらに役立つ可能性があります。この更新には、ファイバーモジュールPCBのコネクタとパッドのレイアウトを変更して、C2Mアーキテクチャーのチャンネル帯域幅を90GHz以上に拡張することが含まれます。変更案を表5にまとめました。

| 変更案                                   | 詳細   |
|---------------------------------------|--|
| 信号ビーム先端の短縮を容易にするため、モジュールカードのエッジを面取りする | 面取りを0.25mmから0.30mmに増やす                                   |
| モジュールカード信号パッドの長さを短くする                 | カードエッジから信号パッドエッジまでの公称距離を1.70mmから1.90mmに延ばす               |
| ワイブ長の許容差を小さくして、モジュールカード信号パッドの短縮を促進する  | 許容差を $\pm 0.395\text{mm}$ から $\pm 0.200\text{mm}$ に厳格化する |
| モジュールカード上の接地パッドの露出部の長さを増やす必要がある       | 最小接地パッド長を1.40mmから2.50mmに増やす                              |
| ホストPCBのOSFPコネクタパッドレイアウトを変更する          | 差動信号には、 $\varnothing 0.36\text{mm}$ のピアをピアインパッド構造で使用する   |
| ファイバーモジュールPCBの上面と底面のパッドレイオフセットをなくす    | モジュールカードの上面と底面のパッドレイアウトを揃える                              |

表5：448Gbps-PAM-4およびより高次の多値PAMに対応するパドルカードインターフェース向けMSA標準の変更案

# 448Gへの道を切り開く

この調査では、C2MアーキテクチャーのOSFPコネクタとコパッケージド銅コネクタを備えた従来のパドルカードインターフェースが、PAM-6またはPAM-8変調を用いた448Gシグナリングをサポートできることを実証しました。BERとSNRの結果から、パドルカードとコネクタの設計に対するMSAを更新することを条件として、PAM-6が変調フォーマットとして望ましいことが示されました。これらの変更を加えなくてもPAM-8を実現できる可能性はありますが、SNRマージンを改善するには、内部のコーディングや高度なイコライゼーションなどの機能強化が必要になります。

Molexでは、広範囲に及ぶ調査、そして112Gと224Gの速度における実績あるリーダーシップに基づく深いエンジニアリング専門知識を通じて、448Gインターコネクタテクノロジーの実現に向けた道を切り開いています。Molexは、コネクタアーキテクチャーとシグナルインテグリティを進化させることで、データセンターにおけるデータの高速伝送と高いシグナルクラリティを実現し、AIを活用する新たなデータ環境が求めるパフォーマンス要件に対応しています。

次世代データセンターの設計戦略など、この移行を可能にする基盤技術の詳細については、[224Gbps-PAM-4高速データセンター技術ページ](#)をご覧ください。



Molexは、米国におけるMolex, LLCの登録商標であり、その他の国でも登録されている場合があります。本文書に記載されている他のすべての商標は、それぞれの所有者に帰属します。