



WP280 (v1.0) 2008 年 9 月 23 日

## 高性能ネットワーク機器導入における 課題を FPGA で解決： 既存の通信機器への 100GbE MAC の追加

著者 : Anthony Torza

---

本書では、業界で急速に増加する高バンド幅のインターフェイス (特に 100GbE) への需要、システム構築時に 100GbE をプラットフォームへ追加する場合の主なリスクと問題点、およびこれらの問題に対応するための FPGA 独自のインプリメンテーション オプションについて検証します。

## 業界における 100GbE への需要の高まり

イーサネットはここ数十年間、業界で重要な位置を占めています。イーサネットは 10Mb/s から始まり、100Mb/s、1Gb/s、10Gb/s と、数年おきにそのバンド幅を順調に増加してきました。

インターネット上でのビデオトラフィックの激化によってデータセンタ内の既存の 10Gb/s ラインに過大な負荷がかかり、基幹ネットワークでのアップリンクおよびデータの伝送が停滞する結果となると同時に、マルチコアプロセッサの拡張によって各サーバが 10GbE (近い将来には 40GbE) まで拡大しています。このようなビジネスおよびテクノロジーでの傾向から、データセンタ内のトラフィックを統合およびアップリンクし、伝送ネットワークへアップリンクできるような高バンド幅 (100GbE) インターフェイスへの需要が増加しています。主要企業およびテレコムプロバイダからの高バンド幅のインターフェイスを求める声は、近年非常に高まっています。

LAG (Link Aggregation Group) に複数の 10Gb/s 配線を統合するという応急手段は、約 4 リンク (40Gb/s) まで有効なことが証明されていますが、LAG を用いるとハードワイヤおよびソフトワイヤに制限が生じるため、10GbE が 4 本以上になると効率的な拡張が妨げられます。

2006 年後半に、IEEE は次世代イーサネットを定義する HSSG (Higher Speed Study Group) を立ち上げました。この HSSG は 2007 年前半に、次世代イーサネットに対して 40GbE (エンタープライズでのラック間通信向け) および 100GbE (ネットワークアップリンクおよび伝送アプリケーション向け) の 2 つの速度を提案しました。これらの仕様は IEEE Std 802.3ba に定められています。この規格は、MLD (Multi Lane Distribution) の定義によって大部分が形成される一方で、いくつかの問題点 (バックプレーンアプリケーションに FEC を含めるかなど) が残っており、いまだ確定していません。ザイリンクスでは、40GbE および 100GbE の規格確定に向けて、イーサネットアライアンスおよび IEEE と共同で作業しています。

本書では、規格として策定前の 100GbE の例に焦点を当て、新規格に対応するためのアップグレード、および規格変更による問題に対処する FPGA 特有の機能を利用した解決手法について説明します。ここで扱うデザイン例は仮説ではなく、すでにユーザーのシステムで使用されているものです。また、本書に沿って行ったインプリメンテーションは、規格の策定後にもそれに準拠するものとなります。

## 100GbE MAC の設計

新しいテクノロジーを用いたデザインの作成を考えている場合、複数のアーキテクチャを検証、およびリスク、Time-To-Market とコスト間のトレードオフを分析することが有益です。本書では、ザイリンクス FPGA を用いた 100GbE インプリメンテーションのオプションを 2 つ示します。最初のオプションでは一般的な既存のインプリメンテーションについて説明し、次のオプションでは将来的な使用に向けて最適化したアーキテクチャについて紹介します。

### 新しいテクノロジーを用いたデザインでの主要なリスク

新しいテクノロジーを用いたデザインには、次のようなリスクが常に伴います。

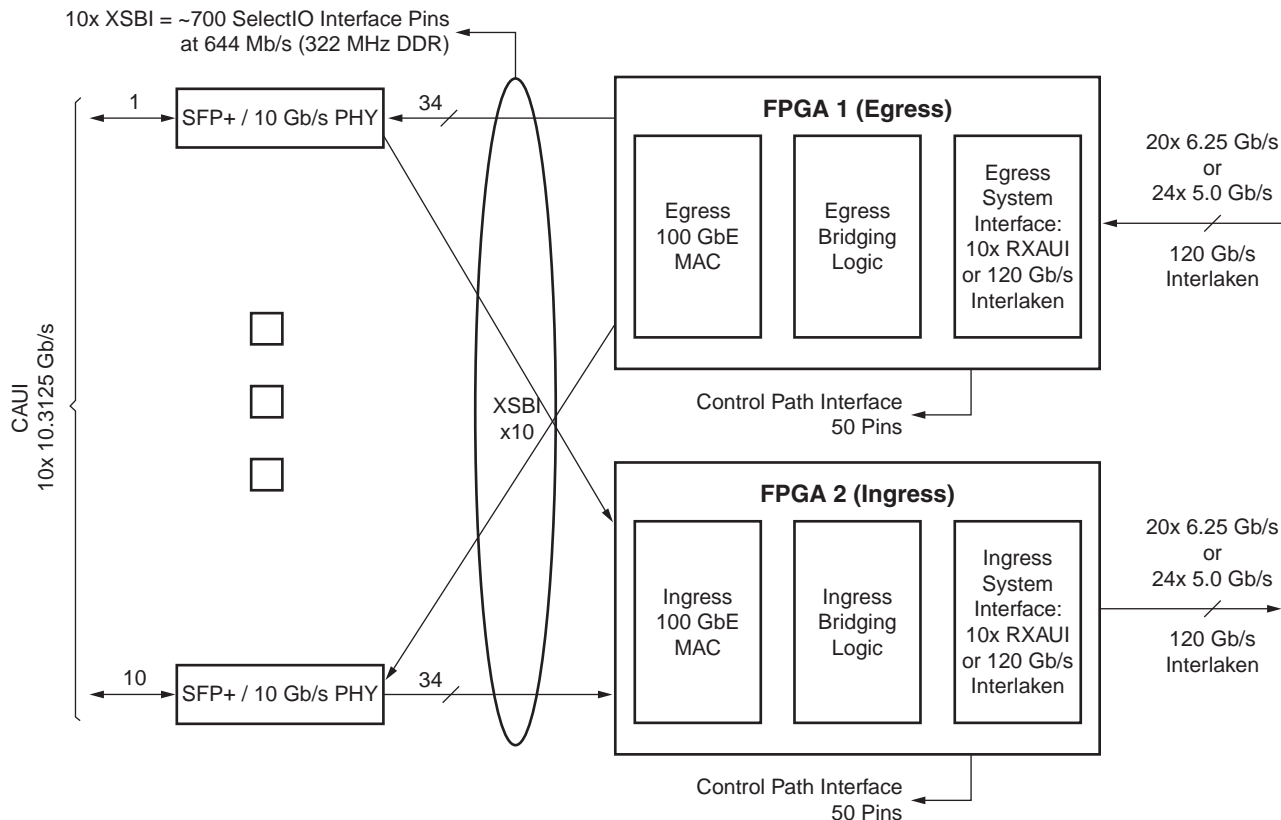
- TTM (Time-To-Market) リスク
  - ◆ 適切なコストおよびタイミングでの製品の市場投入は、市場での製品の成功において非常に重要です。
  - ◆ 多くの場合、初期導入 (最初に導入することで市場でのリーダーシップを確立) から主流テクノロジー (製品量が増加し、コストの圧力からテクノロジーの選択肢に制限が出始める) への推移を検証するには、複数世代に渡る製品が必要です。可能な限り多くの世代にわたるソリューションの提供がベンダーにとって重要となります。
- コンポーネント リスク
  - ◆ スケジュールでのリスク
    - デザインの完成に必要なコンポーネントにスケジュール上のリスク (納品 / 供給力など) がどれだけあるか。
    - 必要なコンポーネントのプロセスロードはスケジュールに遅れを生じさせる可能性のある新規で大部分のテストが未実施のものか。
    - ベンダーは I/O (特に SERDES) の特性評価データをリリース済みか。

- ◆ パフォーマンスでのリスク
  - シリコン コア ロジックのパフォーマンスおよび容量はアプリケーション要件を満たしているか。
  - シリコン I/O (SERDES ベースの場合が多い) は、パフォーマンス要件を満たしているか。
- ◆ 消費電力でのリスク
  - デザインは全体の消費電力要件を満たしているか。
- サードパーティの IP でのリスク ( サードパーティの IP を用いるデザインの場合 )
  - ◆ スケジュールでのリスク
    - IP は開発中であるか。
    - IP を既に採用しているユーザーがいるか ( いる場合、IP は 1 つ以上のプロセスで検証済みとなる ) 。
- 機能面でのリスク
  - ◆ 機能面で大きな変更がないと確信が持てるほど、規格が発展しているか。
- 環境面でのリスク
  - ◆ 外部コンポーネント
    - すべての外部コンポーネントが入手可能か。
    - 必要なコンポーネントはすべて、使用する環境でテストまたはシミュレーション済みか。
    - ベンダーはシミュレーション、テスト、ハードウェア特性評価をサポートする資料を提供可能か。
  - ◆ テスト装置
    - ハードウェア デバッグ用のテスト装置があるか。

デザインによってリスクに対する耐性が異なるため、トレードオフの分析が重要となります。次の 2 つのオプションで、これらのトレードオフの一部を説明します。

### オプション 1：従来のソリューション - XSBI を用いた外部 10Gb/s PHY への接続

複数のユーザーが XSBI をベースとした外部 10Gb/s PHY を用いるソリューションを導入しています。このオプションは、今までに利用可能なテクノロジーで最もリスクが低いものとして焦点が当てられており、2 つの Virtex-5 FXT FPGA および 10 個の外部 XSBI ベースの 10Gb/s PHY デバイスを使用しています。図 1 に示すように、MAC (Media Access Controller) は 2 つの FPGA 間で半分に分割されています。



WP280\_01\_091508

図 1：業界初の 100GbE MAC：2 つの Virtex-5 FPGA および 10 個の外部 SERDES

入り口となる FPGA には、RX 100GbE MAC (MLD を含む)、複数のブリッジ/バッファ ロジック、およびユーザーのシステム アーキテクチャ特有の TX システム インターフェイスが備わっています。ユーザーの ASIC または ASSP イーサネット スイッチが RXAUT (XAUI の 2 倍速版) を介して接続されるか、ASIC または ASSP NPU が Interlaken インターフェイスを介して接続されます。また、ユーザー特有のインターフェイスがオプションで利用可能です。出口となる FPGA には、RX システム インターフェイス、複数のブリッジ/バッファ ロジック、および TX100GbE MAC (MLD を含む) があります。

光モジュールへの CAUI インターフェイスは、10 個の外部 10Gb/s PHY によって作成されています。これらのデバイスは、変更された XSBI インターフェイス (322MHz の DDR で 16 レーン) によって FPGA に接続されます。XSBI 規格は 644MHz の SDR で 16 レーンを指定されているため、このインターフェイスは厳密には DDR ではなく、正確にはハーフレートで動作する XSBI となります。322MHz クロックを使用することで、ザイリンクス SelectIO™ テクノロジーに内蔵された DDR ロジックが利用でき、デザイン マージンの増加と消費電力の低減を実現できます。多くの主要 PHY ベンダーが、このハーフレート (322MHz) のクロックをサポートしています。

オプション 1 のリスクには次が含まれます。

- TTM リスク (最も低い)
  - ◆ このソリューションは、数多くのユーザーによって実装およびテスト済みです。
- コンポーネント リスク (非常に低い)
  - ◆ すべての部品が量産製品として入手可能です。
- スケジュールでのリスク (非常に低い)
 

Virtex-5 FXT FPGA および XSBI SERDES は広く提供されています。

  - ◆ パフォーマンスでのリスク (非常に低い)
 

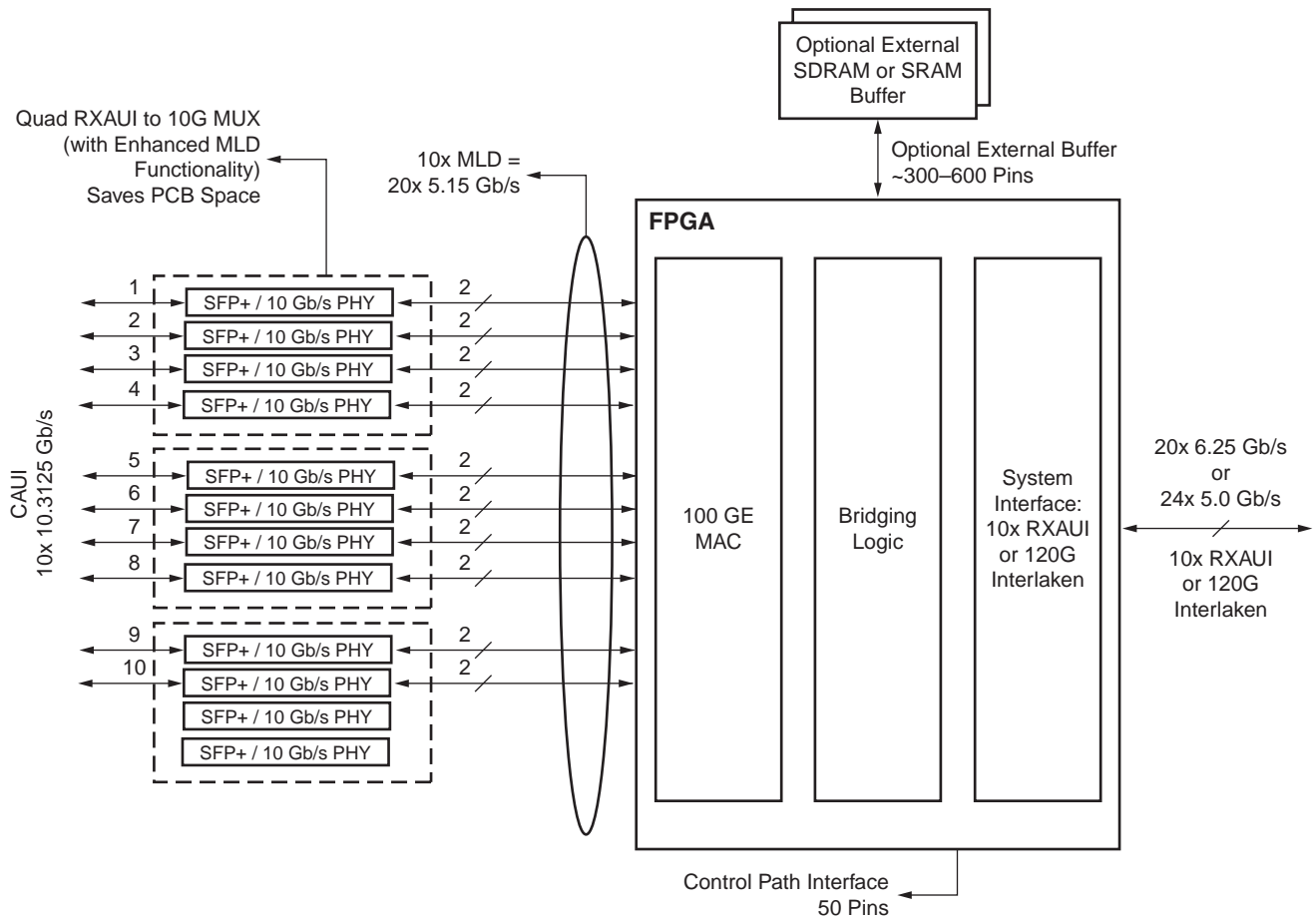
Virtex-5 FPGA のロジックのパフォーマンスは、多様なデータパス幅に十分に対応します。  
Virtex-5 FPGA GTX シリアルトランシーバは、最大 6.5Gb/s のレートをサポートします。
  - ◆ 消費電力でのリスク (低 ~ 中程度)
 

XSBI インターフェイスの消費電力は最少ではありませんが、標準的な通信機器でのワーストケース (NEBS) を考慮すると、デバイスは許容範囲内で動作可能です。
- サードパーティの IP でのリスク (非常に低い)
  - ◆ 100GbE MAC、Interlaken、および RXAUI IP コアはユーザーの環境でデバッグ済みであり、すでにプロトタイプハードウェアで動作しています。
- 環境面での リスク (最も低い)
  - ◆ 外部コンポーネントはすべて入手可能です (10Gb/s PHY は複数のベンダーから提供中)。
  - ◆ テスト装置は入手可能です (例：機能強化されたトラフィック ジェネレータが主要ベンダーから提供中)。

## オプション 2：MLD インターフェイスを用いた SERDES MUX

ザイリンクスでは現在、100GbE 装置を設計するユーザーに対してオプション 2 の使用を推奨しています。このオプションでは、近い将来利用可能な最もリスクの低いテクノロジーを用いています。

このオプションでは、Virtex-5 TXT プラットフォームと 4 個で 1 組の外部 SERDES MUX を使用します。したがって、デザインは 2 つの FPGA および 10 個の外部 PHY を含むものから、1 つの FPGA と 4 個で 1 組の外部 PHY 3 個で構築されるものへ縮小され、コストおよび消費電力を大幅に削減します。[図 2](#) に示すように、Virtex-5 TXT プラットフォームには、48 個の GTX シリアルトランシーバが備わっているため、このインプリメンテーションではロジックを 1 つの FPGA に収めることができます。



WP280\_02\_09150E

図 2：オプション 2：コスト削減向け：1 つの Virtex-5 TXT FPGA と 3 個の外部 Quad SERDES MUX

この 1 つの FPGA には、RX 100GbE MAC (MLD を含む)、複数のブリッジ/バッファ ロジック、およびユーザーのシステム アーキテクチャ特有のシステム インターフェイスが備わっています。ユーザーの ASIC または ASSP イーサネット スイッチが RXAUT (XAUI の 2 倍速版) を介して接続されるか、ASIC または ASSP NPU が Interlaken インターフェイスを介して接続されます。FPGA にはユーザー特有のインターフェイスが追加できる柔軟性が備わっています。

光モジュールへの CAUI インターフェイスは、シリアル インターフェイスを介して 3 個の外部 10Gb/s PHY (4 個で 1 組) で作成されています。3 個のデバイスは、MLD に類似したインターフェイスによって FPGA に接続されます (5.15Gb/s で 20 レーン)。VL から SERDES へのマッピングは 1:1 であり、MLD レイヤは VL を自動的にスキュー調整するため、このインプリメンテーションでは PCB のデザインが大幅に単純化され、消費電力が削減されます。

さらに、100GbE 規格では各仮想レーンに 64B/66B エンコーディングを定義しているため、ザイリンクス FPGA を使用するデザインのみがロジックを大幅に節約する (概算では全体で 10,000 個の LUT を節約) 64B/66B ギャップの利点を享受できます。この SERDES ベースのインターフェイスはロジックおよび PCB スペースを節約するだけでなく、配線の複雑性を軽減し、消費電力を削減します。ザイリンクスは主要な PHY ベンダーと、次世代マルチ コア PHY (2009 年前半に提供予定) に MLD インターフェイスを実装する契約を結んでいます。サードパーティのベンダーおよび連絡先については、e-mail にて [Virtex\\_marketing@xilinx.com](mailto:Virtex_marketing@xilinx.com) まで英語でお問い合わせください。

オプション 2 のリスクには次が含まれます。

TTM リスク (低い)

- このソリューションでは、Virtex-5 FPGA テクノロジを活用して、100GbE デザインを 1 つの FPGA および 3 個の外部 PHY を含むアーキテクチャに最適化するために十分な SERDES を備えたデバイスを使用します。
- コンポーネント リスク (低い)
  - ◆ スケジュールでのリスク (低い)  
Virtex-5 TXT デバイスは入手可能であり、サードパーティの Quad MUX 製品は、2009 年初頭に入手可能となる予定です。
  - ◆ パフォーマンスでのリスク (非常に低い)  
Virtex-5 FPGA のロジックのパフォーマンスは、多様なデータパス幅に十分に対応できることが検証済みです。
  - ◆ 消費電力でのリスク (低い)  
SERDES ベースの MDL インターフェイスに変換すると、必要なロジック数が減少します。パラレル (XSBI) からシリアル (MLD) インターフェイスへの変換に加え、消費電力が 20% 以上削減されます。
- サードパーティの IP でのリスク (非常に低い)
  - ◆ 100GbE MAC、Interlaken、および RXAUI IP コアはユーザーの環境でデバッグ済みであり、すでにプロトタイプハードウェアで動作しています。
  - ◆ MLD インターフェイスのサポートによって生じる変更は少なく、ロジック数を大幅に節約できる 64B/66B ギャップボックス (ザイリンクス製品のみで利用可能) の利点を享受できます。
- 環境でのリスク (最も低い)
  - ◆ サードパーティの Quad MUX 製品は、2009 年初頭に入手可能となる予定です。
  - ◆ テスト装置は入手可能です。

## 将来的なソリューション

ザイリンクスは高性能通信機器市場に尽力しており、新製品を継続的にリリースするにあたり、100GbE 製品のインプリメンテーションにおける推奨手法の改善を続けてまいります。最新のテクノロジーについては、e-mail にて [Virtex\\_marketing@xilinx.com](mailto:Virtex_marketing@xilinx.com) まで英語でお問い合わせください。

## まとめ

FPGA ベースのプログラマブルソリューションは、ワイヤード通信市場での多様なアプリケーション開発に対し、低コスト、低リスクの手法を提供します。FPGA ファブリックおよび I/O のパフォーマンスは、最も困難とされる 100Gb/s アプリケーションをサポートするまでに発展しています。ザイリンクスは FPGA 市場のリーダーとして、ユーザーおよび市場の要求を予見してシリコン、ソフトウェア、および FPGA IP を設計し、適切な低リスクのソリューションを適時に提供していきます。これらのことから、業界の 100GbE への移行に必要な規模の経済、柔軟性、およびサービスの質の向上を実現します。ワイヤード通信市場に対するザイリンクスのソリューションについては、<http://japan.xilinx.com/esp/wired.htm> を参照してください。

## 補足：Virtex-5 TXT FPGA

Virtex-5 TXT プラットフォームは、検証され量産準備が整ったプロセスおよびブロック機能で構築されています。このプラットフォームは、開発者からの幅広いフィードバックを得た上で、新たに非常に高バンド幅のシステム向けに最適化されています。TXT デバイスは、市場を先導する 65nm Virtex-5 ファミリの最新メンバです。ザイリンクスは、Virtex-5 ファミリのカラムベースの ASMBL™ アーキテクチャを採用し、6.5Gb/s シリアルトランシーバのカラムを 2 つ備える 2 種類のデバイスを作成しました。TXT プラットフォームのメンバには、48 個のトランシーバを備える XC5VTX240T および 40 個のトランシーバを装備した XC5VTX150T の 2 つがあります。最新の TXT プラットフォームの仕様については、[DS100](#)：『Virtex-5 ファミリー概要』を参照してください。



## 改訂履歴

次の表に、この文書の改訂履歴を示します。

日付	バージョン	内容
2008/09/23	1.0	初版リリース

## Notice of Disclaimer

The information disclosed to you hereunder (the “Information”) is provided “AS-IS” with no warranty of any kind, express or implied. Xilinx does not assume any liability arising from your use of the Information. You are responsible for obtaining any rights you may require for your use of this Information. Xilinx reserves the right to make changes, at any time, to the Information without notice and at its sole discretion. Xilinx assumes no obligation to correct any errors contained in the Information or to advise you of any corrections or updates. Xilinx expressly disclaims any liability in connection with technical support or assistance that may be provided to you in connection with the Information. XILINX MAKES NO OTHER WARRANTIES, WHETHER EXPRESS, IMPLIED, OR STATUTORY, REGARDING THE INFORMATION, INCLUDING ANY WARRANTIES OF MERCHANTABILITY, FITNESS FOR A PARTICULAR PURPOSE, OR NONINFRINGEMENT OF THIRD-PARTY RIGHTS.