



XAPP1289 (v1.0) 2016 年 6 月 20 日

PCI Express 用 Zynq UltraScale+ MPSoC コントローラーおよび DMA をルート ポートとして使用

著者 : Bharat Kumar G., Sunita Jain, Jason Lawley

概要

PCI Express® アーキテクチャの一般的なシステムでは、システム メモリとエンドポイント間でデータを転送するために、エンドポイントの多くにシステム ホストで制御される DMA エンジンが含まれます。これにより、プロセッサはこのタスクを処理する必要がなくなるため、ほかの処理に利用できるサイクルが多くなります。このようなシステムのデメリットは、システムからカード (エンドポイント) にデータを移動する際に、どこでフェッチするかをエンドポイントに通知する必要があるためにレイテンシが増加することです。

PCI Express 用 Zynq® UltraScale+™ コントローラーの内蔵 DMA エンジンは、エンドポイントだけでなく、ルートポートモードとしても使用できます。内蔵 DMA エンジンを利用するルートポートモードを使用することにより、ほかの多くの処理サブシステムでは不可能な方法で、レイテンシを削減でき、さらにシステム性能を向上できる場合があります。

このアプリケーション ノートでは、PCI Express 用コントローラーがルートポートとして構成されているときにどのようにこのコントローラーの DMA を設定して使用するのかを説明します。さらに、システムとカード間でのデータ移動のパフォーマンス モニターの結果も示します。このデザインは、ZCU102 ハードウェアプラットフォームをターゲットとしており、ルートコンプレックスとして動作する Gen1 x1 から Gen2 x4 までの PCIe システムを開発できます。

開発者やシステムアーキテクトは、このアプリケーション ノートを利用することにより、データ移動の最も効果的な方法を判別する手段をさらに得ることができます。PCI Express 用コントローラー内の DMA をそれのみで使用するか、エンドポイントとルートポートの両方で DMA エンジンを利用する複雑な方式を採用するかにかかわらず、Zynq UltraScale+ は、このようなシステムを実現するための機能と手段を提供します。

このアプリケーション ノートの [リファレンスデザインファイル](#) は、ザイリンクスのウェブサイトからダウンロードできます。デザインファイルの詳細は、[8 ページの「リファレンスデザイン」](#) を参照してください。

はじめに

この資料のデザインでは、DMA を PCI Express のルートポートとして使用し、エンドポイントに対してデータをプッシュおよびプルします。ルートポートモードでは、Zynq UltraScale+ 上の PCI Express 用コントローラーが統合された DMA ブロックと共に使用されます。

このデザインは、エンドポイントとして KCU105 ボードのベースデザインを使用しています。エンドポイントデザインには KCU105 上の DDR4 をターゲットにした Memory Interface Generator IP (MIG) が含まれ、これはザイリンクス IP の AXI Bridge for PCI Express Gen3 v2.0 を介して PCIe BAR にマップされています。デザインの概要は、[2 ページの図 1](#) を参照してください。

PCIe BAR から DDR4 および AXI Performance Monitor へのアドレス変換は、IP のカスタマイズの際に設定します。ルートポート上の DMA は、エンドポイントをコプロセッサまたはアクセラレータブロックとして使用できるカスタムアプリケーションに使用されます。

必要な環境

ハードウェア

1. ZCU102 ボード (と電源)、USB-UART ケーブル、SD カード
2. KCU105 ボード (と電源ケーブル)、USB-JTAG ケーブル

ソフトウェア

1. Vivado® Design Suite 2016.1
2. PetaLinux 2016.1

デザインの概要

図 1 にデザインの概要を示します。KCU105 エンドポイントにプログラムするビット ファイルが、リファレンス デザイン パッケージに付属しています。この資料で使用する用語について説明します。

- AXI メモリとは、Zynq UltraScale+ MPSoC 上の PS-DDR のことです。
- EP メモリまたは PCIe メモリとは、KCU105 エンドポイント上の DDR のことです。

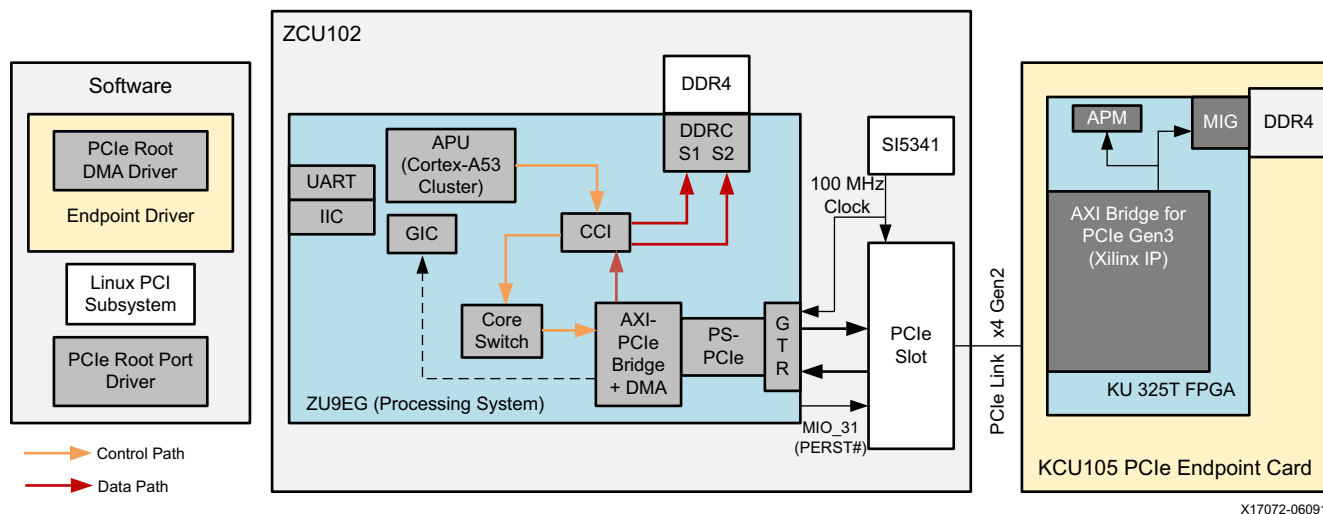


図 1: デザインの概要

注記: DDRC の S1 と S2 は、CCI からのインターフェイスが DDRC に接続されるスロットを表しています。

データフロー

ルートポートから発行された PCIe リンク上のトランザクション (メモリ読み出しおよびメモリ書き込み) は、エンドポイントで受信されるには、エンドポイントの BAR (ベースアドレスレジスタ) のいずれかにヒットする必要があります。DMA は、送信元と転送先のディスクリプター (それぞれ SRC-Q および DST-Q と呼ぶ) を提供します。DMA の詳細は、『Zynq UltraScale+ MPSoC テクニカルリファレンスマニュアル』[参照 1] を参照してください。

2 ページの図 1 に示すとおり、PCIe 用コントローラーは、CCI (キャッシュコヒーレントインターコネクト) を介して PS-DDR にアクセスします。PS-DDR に接続されている CCI から PS-DDR へのインターフェイスには、スロット 1 とスロット 2 の 2 つがあります。

注記: 以降の説明では、DMA とは、Zynq UltraScale+ MPSoC 上の PCIe 用コントローラーの一部であり、かつルートポートとして機能しているものを指します。

制御フロー

APU SMP Linux 上でルートポート DMA ドライバーが実行されると、次の動作が起こります。

1. PS-DDR メモリにディスクリプター Q (SRC と DST、およびそれぞれのステータス) がセットアップされます。
 - a. SRC エレメントと DST エレメント内のフラグによってデータ転送の方向が指定されます。
 - b. すべての DMA ディスクリプター Q が、PS-DDR (ルートポートメモリ) 内に含まれます。
2. DMA 動作に必要な、PCIe 用コントローラーの各種 DMA レジスタがプログラムされます。

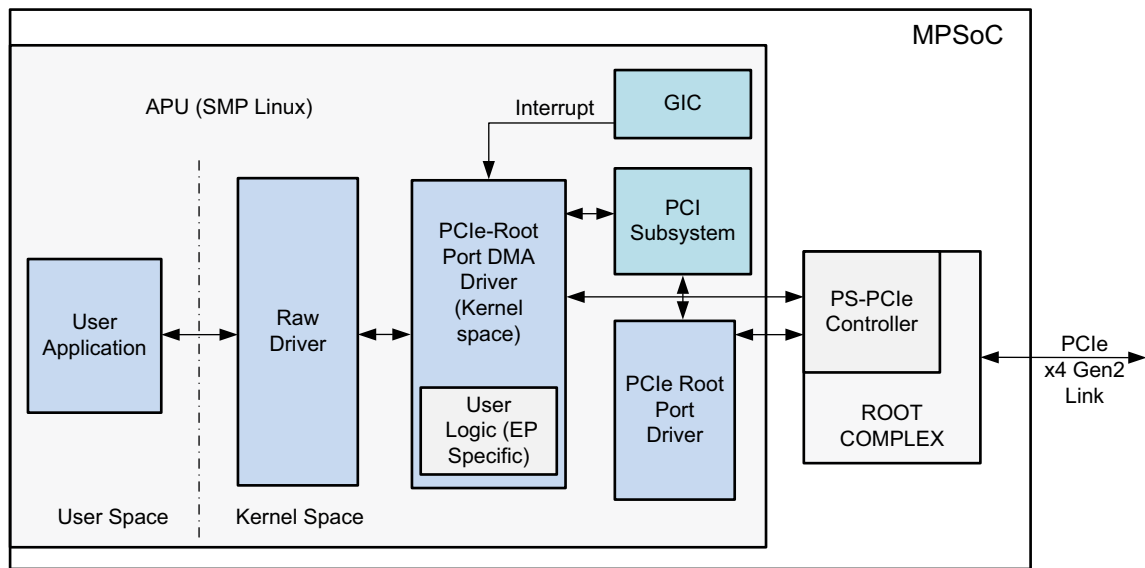
データフロー

ルートポート DMA ドライバーにより DMA がセットアップされると、次の動作が起こります。

1. S2C (ルートポートからエンドポイント) 転送の場合
 - a. DMA は、AXI ドメイン内での読み出しを (SRC-Q エレメントからのアドレスに対して) 発行し、データをフェッチします。
 - b. DMA は次に、PCIe リンク上でメモリ書き込みトランザクションを (DST-Q エレメントからのアドレスに対して) 発行し、データをエンドポイントに転送します。
 - c. この書き込みがエンドポイントで受信され、アドレスが変換されて (AXI Bridge for PCI Express Gen3 IP により PCIe BAR から EP-DDR AXI ドメインアドレスに変換)、データが EP-DDR メモリへ転送されます。
2. C2S (エンドポイントからルートポート) 転送の場合
 - a. ルート上の DMA が、PCIe リンク上で読み出しを (SRC-Q エレメントからのアドレスに対して) 発行します。このダウンストリームメモリ読み出しが、AXI Bridge for PCI Express Gen3 IP によって EP-DDR AXI ドメインアドレスに変換されます。
 - b. EP-DDR AXI ドメインアドレスから読み出されたデータが、データを含む完了として PCIe リンク経由で受信されます。
 - c. DMA が、AXI ドメインに対して書き込みを (ルートポート内で、DST-Q エレメントからのアドレスに対して) 発行し、PCIe リンク経由で受信したデータを PS-DDR メモリに転送します。

コンポーネント

図 2 にソフトウェア コンポーネントを示します。



X17071-060916

図 2: ソフトウェアの概要

このデザインの主なコンポーネントは次のとおりです。

1. **第一段階ブートローダー (FSBL) :** FSBL は、各種ペリフェラル、クロック、PS-DDR に関して ZCU102 をセットアップし、ルートモード動作用に PCI Express コントローラーをプログラムします。
2. **ルートポートドライバー :** このドライバーは、PetaLinux ビルドにより提供されるカーネルの一部です。ZCU102 を PCIe ルートポートとして使用するための詳細は、<http://www.wiki.xilinx.com/ZynqMP+Linux+PCIe+Root+Port> を参照してください。このドライバーは、MPSoC 上の PCI Express コントローラー内で AXI-PCIe ブリッジをセットアップし、エニュメレーション用に Linux PCI サブシステムに接続します。
3. **ルートポート DMA ドライバー :** このドライバーは、MPSoC の PCI Express コントローラーの DMA を管理します。
 - a. すべての DMA チャンネル Q (送信元 Q、転送先 Q、対応するステータス Q) が、このドライバーにより管理されます。
 - b. すべての Q は、AXI (PS-DDR) メモリー内に含まれます。
 - c. ダウンストリーム (ルートポートからエンドポイント) 転送の場合、送信元バッファは AXI メモリーにあり、転送先バッファはエンドポイントの PCIe メモリーにあります。
 - d. アップストリーム (エンドポイントからルートポート) 転送の場合、送信元バッファはエンドポイントの PCIe メモリーにあり、転送先バッファは AXI メモリー (PS-DDR) にあります。
 - e. MPSoC の PCI Express コントローラーからの、AXI ドメインにある DMA 割り込みが使用されます。エンドポイントからの割り込みは使用されません。
 - f. ダウンストリーム転送用のデータバッファとアップストリーム転送からデータを受信するためのフリーバッファを提供するユーザー空間アプリケーションが、付属のデザイン内にあります。

ユーザーロジック : ルート DMA ドライバーのこの部分は、エンドポイント用の基本的な初期化タスクを実行します (エンドポイントに固有)。現在のデザイン (KCU105 エンドポイント) では、このドライバーは、BAR4 にマップされる AXI Performance Monitor をセットアップします。

DMA ドライバーは PCI プロンプを 2 回呼び出します (PCIe ルートポート用に 1 回と PCIe エンドポイント用に 1 回)。DMA ドライバーの上に Raw ドライバーが配置され、ユーザー空間アプリケーションに接続されます。このドライバーは、データ転送の方向に基づいて、データ移動用の関連 DMA ドライバー API を呼び出します。ユーザー空間アプリケーションは、トラフィックジェネレーターです。

4. **デバイス ツリー**: これは PCI Express コントローラー用の関連ノードをリストしたもので、ルート ポート ドライバーとルート DMA ドライバーで使用されます。

パフォーマンス モニター

AXI Performance Monitor (APM) を使用して、読み出しと書き込みのバイト数を使用したスループット統計が収集されます。Zynq UltraScale+ MPSoC の場合、PS 内の統合 APM が使用されます。スループットは、次の場所でモニターされます。

1. Zynq UltraScale+ MPSoC の PS-DDR

PCI Express 用コントローラーからのトラフィックは、スロット 1 とスロット 2 を介して PS-DDR に送信されます。ここでの統計には、実際のデータ パケットの読み出しと書き込みのほかに、PS-DDR への APU アクセス (DMA エンジンによる Q フェッチ/更新) が含まれます。APM と PS-DDRC のスロットの場所の詳細は、『Zynq UltraScale+ MPSoC テクニカル リファレンス マニュアル』[参照 1] を参照してください。

2. KCU105 EP-DDR

これは、EP-DDR に送信される実際のデータ ペイロードのスループットを (PCIe プロトコルや変換層パケットのオーバーヘッドなしで) 測定します。

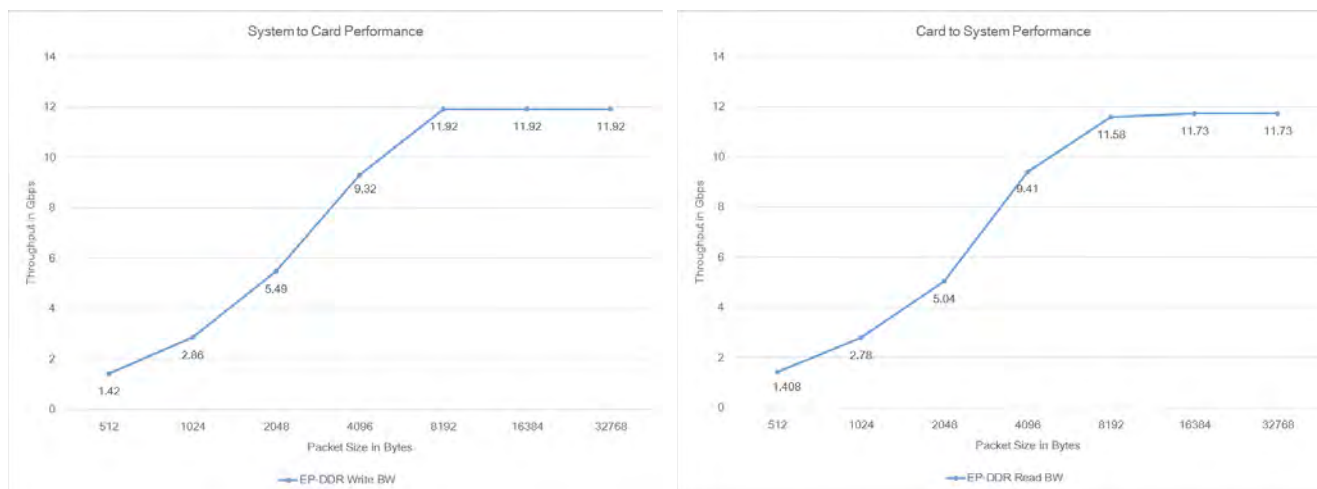
テストの説明

テストのセットアップと手順の詳細は <http://www.wiki.xilinx.com/XAPP1289+PCIe+Root+DMA> を参照してください。

結果

図 3 は、システムからカード (S2C) とカードからシステム (C2S) のパフォーマンスそれぞれを示したものです。次のことがわかります。

- パケット サイズに伴うスループットの変化は予想どおりに観測されています (パケットはユーザー空間アプリケーションで生成)。
- 特定のパケット サイズのダウンストリームトラフィック (S2C) のスループットは、同じサイズでのアップストリームトラフィック (C2S) のスループットよりも高くなっています。
 - ダウンストリームトラフィックでは DMA はメモリ書き出しトランザクションを PCIe リンク上で発行するのに対して、アップストリームトラフィックの場合は、メモリ読み出しトランザクションが PCIe リンク上に存在します。



X17138-060916

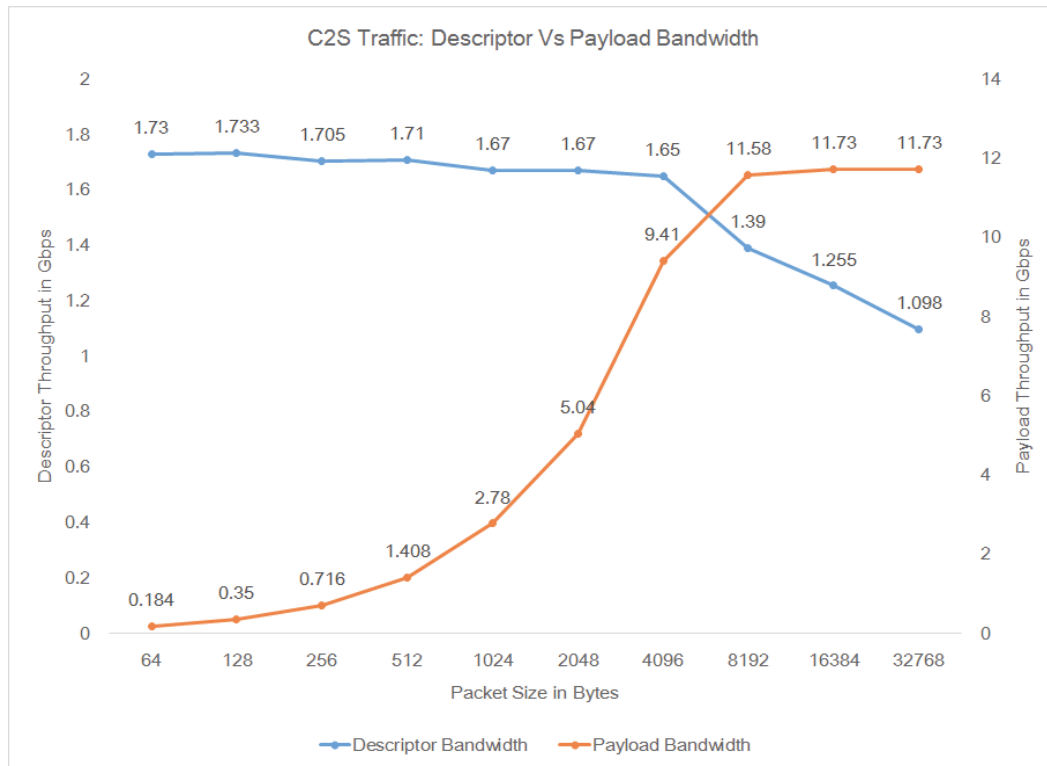
図 3: パフォーマンス サマリ

注記: エンドポイント DDR の帯域幅には、データペイロードのパフォーマンスのみが含まれます。

実際の PCIe リンクのスループットは、確認された EP-DDR 帯域幅と同様になると予想されます。

図 4 は、ルートポートでの DDR 読み出しの帯域幅を、C2S トラフィックのパケットサイズに対して示したものです。C2S の場合、PS-DDR に対する読み出しは主に送信元と転送先のディスクリプターの読み出しになります。

- 小さいパケットサイズの C2S トラフィックでは、ルートポートの PS-DDR での読み出し帯域幅が広いことがわかります。これは、C2S では PS-DDR に対する読み出しが主に送信元と転送先のディスクリプターの読み出しになるためです。パケットサイズが小さくなると、パケットサイズが大きい場合と比較して、ディスクリプターのフェッチ頻度が高くなります。

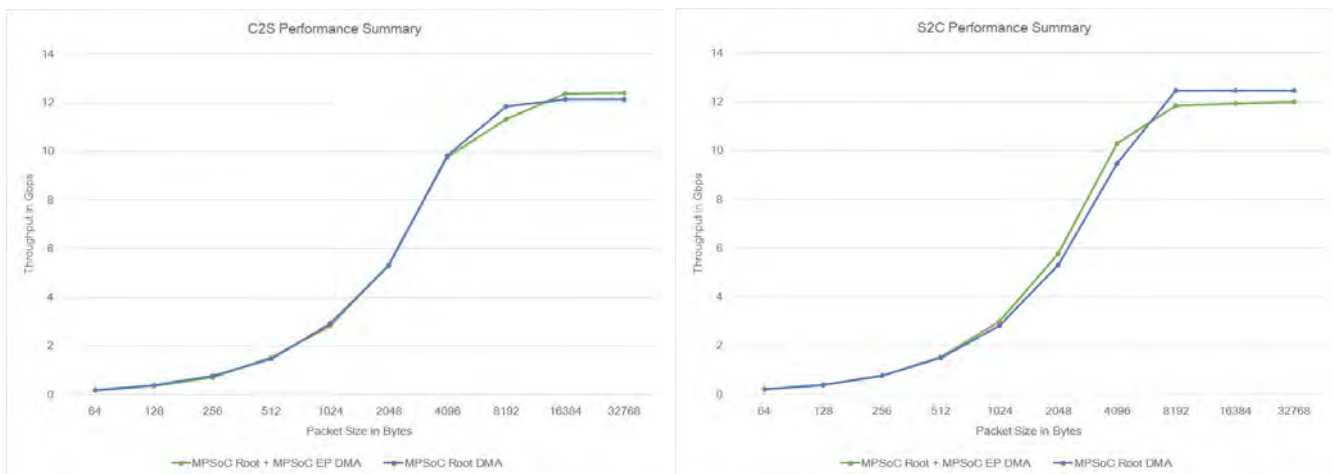


X17139-060916

図 4: RP DDR 読み出し帯域幅の変化 (C2S トラフィック)

図 5 は、ルートポートで DMA を使用した場合と EP で DMA を使用した場合の、パフォーマンスを比較したものです。次が観測されています。

- ルートポートとして MPSoC を使用する場合、EP DMA のパフォーマンス数値はルート DMA とほぼ同じです。



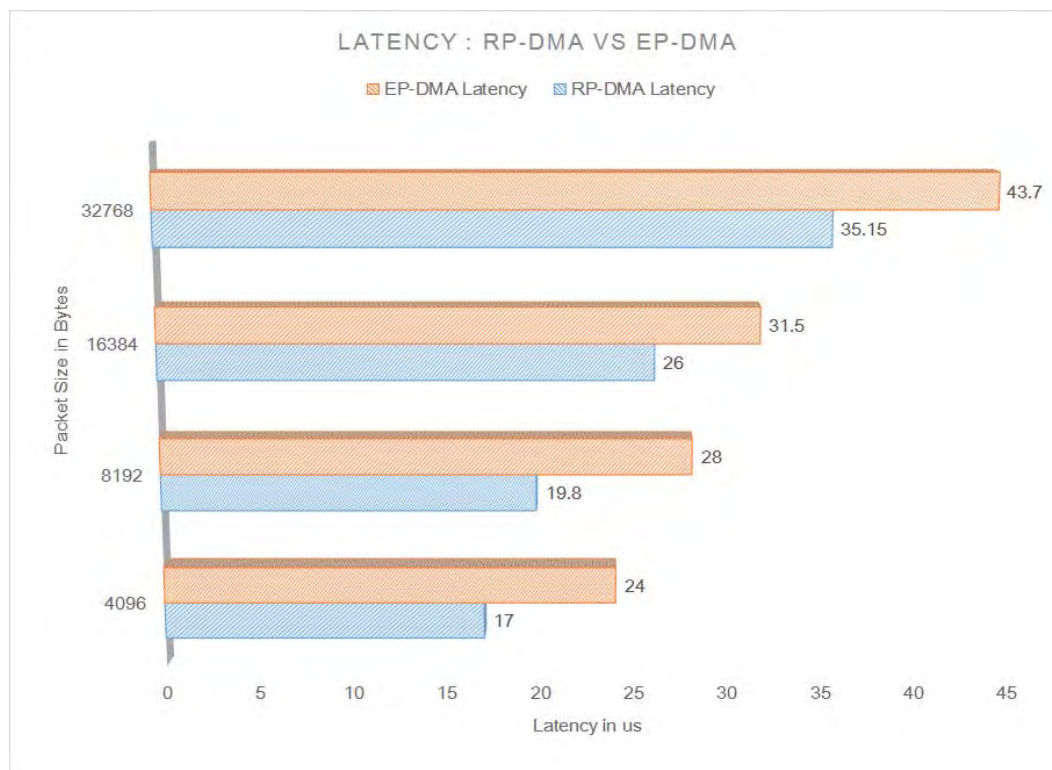
X17140-060916

図 5: ルート DMA と EP DMA のパフォーマンス比較

次に、ルート ポート DMA とエンドポイント DMA 間の単一パケット転送の完了時間の比較を示します。これらのテストは、ルート ポートとしての Zynq UltraScale+ MPSoC デバイス上で実行されています。

- 図 6 では、ルート ポート DMA での単一パケット転送のレイテンシは、エンドポイント DMA でのレイテンシよりも小さいことがわかります。これは、ルート ポート DMA に伴う DMA セットアップ オーバーヘッドが少ないためです。

注記: レイテンシの計算はソフトウェア内で実行され (カーネル空間内の転送の開始から完了割り込みの受信まで)、複数回繰り返した平均値です。



X17210-060916

図 6: 単一パケット転送のレイテンシ

リファレンス デザイン

このアプリケーション ノートの [リファレンス デザイン ファイル](#) は、ザイリンクスのウェブサイトからダウンロードできます。

表 1 に、リファレンス デザインの詳細を示します。

表 1: リファレンス デザインの詳細

| パラメーター | 説明 |
|------------|-----------------------------------------|
| 全般 | |
| 開発者 | Bharat Kumar G、Sunita Jain、Jason Lawley |
| ターゲット デバイス | XCZU9EG-FFVB1156 |
| ソース コードの提供 | あり |
| ソース コードの形式 | C |

表 1: リファレンス デザインの詳細 (続き)

| パラメーター | 説明 |
|-----------------------------------------------------------------|---------------|
| 既存のザイリンクス アプリケーション ノート/リファレンス デザイン、またはサードパーティからデザインへのコード/IP の使用 | なし |
| インプリメンテーション | |
| 使用した合成ツール/バージョン | Vivado 2016.1 |
| 使用したインプリメンテーション ツール/バージョン | Vivado 2016.1 |
| スタティック タイミング解析の実施 | N/A |
| ハードウェア検証 | |
| ハードウェア検証の実施 | あり |
| 使用したハードウェア プラットフォーム | ZCU102 |

まとめ

ルート ポート デザインの実装に PCI Express 用 Zynq UltraScale+ コントローラーと DMA を使用することで、エンドポイントに配置された DMA と比較して、システムからカード (S2C) へのデータ送信にかかる時間を大きく削減できます。このアプリケーション ノートと付随するリファレンス デザインは、DMA をルート ポート モードでどのように設定して使用するかのサンプルを提供しています。さらに、実際のハードウェア デザインのレイテンシと、ルート ポートで DMA を使用した場合の S2C 転送で可能となるレイテンシとの時間差を測定し、レイテンシが削減されることを証明しました。設計者は DMA をルート ポートで使用することを検討してみてください。エンドポイントとルート ポートの両方に DMA を利用することで、低レイテンシ デザインを実現する独自の手段を見出せる可能性があります。

参考資料

このアプリケーション ノートの参考資料は次のとおりです。

注記: 日本語版のバージョンは、英語版より古い場合があります。

- 『Zynq UltraScale+ MPSoC テクニカル リファレンス マニュアル』(UG1085: [英語版](#)、[日本語版](#))
- 『AXI Performance Monitor LogiCORE IP 製品ガイド』(PG037: [英語版](#)、[日本語版](#))
- 『AXI Bridge for PCI Express Gen3 Subsystem v2.1 製品ガイド』([PG194](#))

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 |
|------------|-------|----|
| 2016年6月20日 | 1.0 | 初版 |

法的通知

本通知に基づいて貴殿または貴社(本通知の被通知者が個人の場合には「貴殿」、法人その他の団体の場合には「貴社」。以下同じ)に開示される情報(以下「本情報」といいます)は、ザイリンクスの製品を選択および使用することのためにのみ提供されます。適用される法律が許容する最大限の範囲で、(1)本情報は「現状有姿」、およびすべて受領者の責任で(with all faults)という状態で提供され、ザイリンクスは、本通知をもって、明示、黙示、法定を問わず(商品性、非侵害、特定目的適合性の保証を含みますがこれらに限られません)、すべての保証および条件を負わない(否認する)ものとし、また、(2)ザイリンクスは、本情報(貴殿または貴社による本情報の使用を含む)に関係し、起因し、関連する、いかなる種類・性質の損失または損害についても、責任を負わない(契約上、不法行為上(過失の場合を含む)、その他のいかなる責任の法理によるかを問わない)ものとし、当該損失または損害には、直接、間接、特別、付随的、結果的な損失または損害(第三者が起こした行為の結果被った、データ、利益、業務上の信用の損失、その他あらゆる種類の損失や損害を含みます)が含まれるものとし、それは、たとえ当該損害や損失が合理的に予見可能であったり、ザイリンクスがそれらの可能性について助言を受けていた場合であったとしても同様です。ザイリンクスは、本情報に含まれるいかなる誤りも訂正する義務を負わず、本情報または製品仕様のアップデートを貴殿または貴社に知らせる義務も負いません。事前の書面による同意のない限り、貴殿または貴社は本情報を再生産、変更、頒布、または公に展示してはなりません。一定の製品は、ザイリンクスの限定的保証の諸条件に従うこととなるので、<http://japan.xilinx.com/legal.htm#tos>で見られるザイリンクスの販売条件を参照してください。IP コアは、ザイリンクスが貴殿または貴社に付与したライセンスに含まれる保証と補助的条件に従うこととなります。ザイリンクスの製品は、フェイルセーフとして、または、フェイルセーフの動作を要求するアプリケーションに使用するために、設計されたり意図されたりしていません。そのような重大なアプリケーションにザイリンクスの製品を使用する場合のリスクと責任は、貴殿または貴社が単独で負うものです。<http://japan.xilinx.com/legal.htm#tos>で見られるザイリンクスの販売条件を参照してください。

自動車用のアプリケーションの免責条項

ザイリンクスの製品は、フェイルセーフとして設計されたり意図されてはならず、また、フェイルセーフの動作を要求するアプリケーション(具体的には、(I) エアバッグの展開、(II) 車のコントロール(フェイルセーフまたは余剰性の機能(余剰性を実行するためのザイリンクスの装置にソフトウェアを使用することは含まれません)および操作者がミスをした際の警告信号がある場合を除きます)、(III) 死亡や身体傷害を導く使用、に関するアプリケーション)を使用するために設計されたり意図されたりしていません。顧客は、そのようなアプリケーションにザイリンクスの製品を使用する場合のリスクと責任を単独で負います。

© Copyright 2016 Xilinx, Inc. Xilinx, Xilinx のロゴ、Artix、ISE、Kintex、Spartan、Virtex、Vivado、Zynq、およびこの文書に含まれるその他の指定されたブランドは、米国およびその他の各国のザイリンクス社の商標です。すべてのその他の商標は、それぞれの所有者に帰属します。ARM は、ARM の欧州連合その他の国における登録商標です。