

(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7700153号
(P7700153)

(45)発行日 令和7年6月30日(2025.6.30)

(24)登録日 令和7年6月20日(2025.6.20)

(51)国際特許分類		F I		
G 0 6 F	13/36 (2006.01)	G 0 6 F	13/36	3 1 0 E
G 0 6 F	13/42 (2006.01)	G 0 6 F	13/42	3 1 0
G 0 6 F	13/14 (2006.01)	G 0 6 F	13/14	3 1 0 F
		G 0 6 F	13/14	3 1 0 H
請求項の数 20 (全23頁)				
(21)出願番号	特願2022-574299(P2022-574299)	(73)特許権者	591025439	
(86)(22)出願日	令和3年2月18日(2021.2.18)		ザイリンクス インコーポレイテッド	
(65)公表番号	特表2023-529831(P2023-529831 A)		X I L I N X I N C O R P O R A T E D	
(43)公表日	令和5年7月12日(2023.7.12)		アメリカ合衆国 カリフォルニア州 9 5	
(86)国際出願番号	PCT/US2021/018540		1 2 4 - 3 4 0 0 サン ホセ ロジック	
(87)国際公開番号	WO2021/247103	(74)代理人	ドライブ 2 1 0 0	
(87)国際公開日	令和3年12月9日(2021.12.9)		110002077	
審査請求日	令和6年2月19日(2024.2.19)	(72)発明者	園田・小林弁理士法人	
(31)優先権主張番号	16/894,446		ミッタル, ミリンド	
(32)優先日	令和2年6月5日(2020.6.5)		アメリカ合衆国 カリフォルニア 9 5 1	
(33)優先権主張国・地域又は機関	米国(US)	(72)発明者	2 4, サン ノゼ, ロジック ドライブ	
			2 1 0 0	
			ダスティダー, ジャイディーブ	
			アメリカ合衆国 カリフォルニア 9 5 1	
			2 4, サン ノゼ, ロジック ドライブ	
			最終頁に続く	

(54)【発明の名称】 直接取り付けられたディスパッチを使用した、分離されたスイッチ制御経路

(57)【特許請求の範囲】

【請求項1】

組み込まれた構成要素であって、
ホストと前記組み込まれた構成要素との間の物理接続を形成するように構成された第1のポートと、
複数のI/O機能と、
埋設されたスイッチと、
前記第1のポートを介して前記ホストからパケットを受け取り、前記パケットのタイプを識別し、かつ、前記パケットの前記タイプに基づいて、前記複数のI/O機能のうちの行先I/O機能へ直接前記パケットを経路指定するかまたは前記埋設されたスイッチを使用して前記行先I/O機能へ間接的に前記パケットを経路指定するように構成されたパススルーインタフェースと
を備える、組み込まれた構成要素。

【請求項2】

前記複数のI/O機能が互いに独立して動作し、異なるI/O機能を実施する、請求項1に記載の構成要素。

【請求項3】

前記パケットの前記タイプが、前記パケットが時間に敏感なデータパケットであるかどうかを示し、前記パケットを前記行先I/O機能へ直接経路指定することが、前記埋設されたスイッチをバイパスする、請求項1に記載の構成要素。

【請求項 4】

前記パススルーインタフェースが、前記複数の I/O 機能のうちの第 1 の I/O 機能および前記埋設されたスイッチの両方が、いつでも前記ホストに伝送することができるデータを有していることの指示を受け取り、また、データを前記ホストに伝送するために、前記第 1 の I/O 機能および前記埋設されたスイッチのうちのどちらに前記第 1 のポートの使用を許可するかを決定するために調停するように構成された調停論理を備える、請求項 1 に記載の構成要素。

【請求項 5】

前記パススルーインタフェースが、前記パケットの前記タイプを識別し、また、前記行先 I/O 機能に対応する前記パケット中の行先 ID を識別するための経路指定論理を備える、請求項 1 に記載の構成要素。

10

【請求項 6】

前記ホストが、前記複数の I/O 機能が並行して動作している間、新しい I/O 機能を前記組み込まれた構成要素の中にホット追加するように構成される、請求項 1 に記載の構成要素。

【請求項 7】

前記新しい I/O 機能をホット追加することが、

前記新しい I/O 機能を含むように前記組み込まれた構成要素のプログラマブル論理を構成することを含む、請求項 6 に記載の構成要素。

20

【請求項 8】

それまで起動が解除されていた前記新しい I/O 機能を起動することを含み、前記新しい I/O 機能が硬化論理で実現される、請求項 6 に記載の構成要素。

【請求項 9】

装置であって、

ホスト上の第 2 のポートとの物理接続を形成するように構成された第 1 のポートと、
複数の I/O 機能と、
埋設されたスイッチと、

前記第 1 のポートを介してホストからパケットを受け取り、前記パケットのタイプを識別し、かつ、前記パケットの前記タイプに基づいて、前記複数の I/O 機能のうちの行先 I/O 機能へ直接前記パケットを経路指定するかまたは前記埋設されたスイッチを使用して前記行先 I/O 機能へ間接的に前記パケットを経路指定するように構成されたパススルーインタフェースと

30

を備える装置。

【請求項 10】

前記複数の I/O 機能が独立した I/O デバイスであり、前記独立した I/O デバイスの各々が少なくとも部分的にプログラマブル論理で実現される、請求項 9 に記載の装置。

【請求項 11】

前記独立した I/O デバイスが同じ集積回路の中に形成される、請求項 10 に記載の装置。

【請求項 12】

前記複数の I/O 機能が独立した I/O デバイスであり、前記独立した I/O デバイスの各々が硬化論理を使用して実現される、請求項 9 に記載の装置。

40

【請求項 13】

前記独立した I/O デバイスが同じ集積回路の中に形成される、請求項 12 に記載の装置。

【請求項 14】

基板

をさらに含み、前記複数の I/O 機能の各々が異なる集積回路で実現され、前記異なる集積回路および前記埋設されたスイッチが前記基板の上に取り付けられる、請求項 9 に記載の装置。

50

【請求項 15】

前記装置が、前記複数の I/O 機能が並行して動作している間、新しい I/O 機能をホット追加するように構成される、請求項 9 に記載の装置。

【請求項 16】

前記新しい I/O 機能をホット追加することが、

前記新しい I/O 機能を含むように前記装置のプログラマブル論理を構成することを含む、請求項 15 に記載の装置。

【請求項 17】

前記新しい I/O 機能をホット追加することが、

それまで起動が解除されていた前記新しい I/O 機能を起動することを含み、前記新しい I/O 機能が前記装置の硬化論理で実現される、請求項 15 に記載の装置。

【請求項 18】

方法であって、

組み込まれた構成要素のパススルーインタフェースでホストから第 1 のパケットを受け取ることであって、前記組み込まれた構成要素が複数の I/O 機能と、前記パススルーインタフェースに通信結合された埋設されたスイッチとを備える、第 1 のパケットを受け取ることと、

前記第 1 のパケットがデータパケットであることを決定することであって、前記複数の I/O 機能のうちの第 1 の I/O 機能が前記データパケットの行先である、前記第 1 のパケットがデータパケットであることを決定することと、

前記埋設されたスイッチをバイパスする直接データ経路を使用して、前記データパケットを前記パススルーインタフェースから前記第 1 の I/O 機能へ直接経路指定することと、

前記パススルーインタフェースで前記ホストから第 2 のパケットを受け取ることと、

前記第 2 のパケットが構成パケットであることを決定することであって、前記第 1 の I/O 機能が前記構成パケットの行先である、前記第 2 のパケットが構成パケットであることを決定することと、

前記構成パケットを前記埋設されたスイッチを介して前記パススルーインタフェースから前記第 1 の I/O 機能へ経路指定することと

を含む方法。

【請求項 19】

前記構成パケットを前記埋設されたスイッチを介して前記パススルーインタフェースから前記第 1 の I/O 機能へ経路指定することが、

前記構成パケットを前記パススルーインタフェースから前記埋設されたスイッチへ経路指定することと、

前記埋設されたスイッチで、前記第 1 の I/O 機能が前記構成パケットの行先であることを決定することと、

前記構成パケットを前記埋設されたスイッチから前記第 1 の I/O 機能へ転送することとをさらに含み、

前記方法が、

第 3 のパケットを前記ホストから前記パススルーインタフェースで受け取ることと、

前記第 3 のパケットが異なる構成パケットであることを決定することと、

前記異なる構成パケットを前記埋設されたスイッチへ経路指定することと、

前記埋設されたスイッチで、前記埋設されたスイッチが前記異なる構成パケットの行先であることを決定することとをさらに含む、請求項 18 に記載の方法。

【請求項 20】

前記パススルーインタフェースで、前記埋設されたスイッチおよび前記複数の I/O 機能のうちの少なくとも 2 つが前記ホストにいつでも伝送することができるパケットを有していることを決定することと、

データを前記ホストに伝送するために、前記組み込まれた構成要素の共有ポートの使用

10

20

30

40

50

をどちらに許可するかを決定するべく、サービス品質（QoS）政策に基づいて、前記埋設されたスイッチと前記複数のI/O機能のうちの前記少なくとも2つとの間で調停する、請求項19に記載の方法。

【発明の詳細な説明】

【技術分野】

【0001】

本開示の例は、一般に、ホストと、組み込まれた構成要素上のI/O機能との間で伝送されるデータパケットおよび構成パケットのための個別のデータ経路の確立に関する。

【背景技術】

【0002】

Peripheral Component Interconnect Express (PCIe) によってイネーブルされるシステムなどのServer Host-Acceleratorシステム、またはCompute Express Link (CXL) およびCache Coherent Interconnect for Accelerators (CCIX) などのキャッシュコヒーレンシプロトコルは、プロトコル認識スイッチ構成要素を介して、複数のデバイスに対する増加したファンアウトを達成している。したがって単一の物理ホストポートは、ホストおよびI/Oデバイスの両方がCXL、CCIXおよびPCIeによって確立された二地点間接続を介して通信中であっても、ネットワーク機能、記憶機能、アクセラレータ機能、Direct Memory Access (DMA) 機能、等々などの異なるI/O機能を実施するデバイスを含む、書替え可能ゲートアレイ (FPGA)、グラフィック処理装置 (GPU)、ネットワークインタフェースカード (NIC) などの複数のI/Oデバイスに通信することができる。

【0003】

また、Server Host-Acceleratorシステムは、システムにおける多重デバイスカードスロットのために、同じプロトコル認識スイッチ構成要素を介してホットプラグ機構を提供している。ホット追加およびホット除去能力を含むこれらのホットプラグ機構は、特定のサーバが、これらのスロットにおける静的プラグインプロトコルカードに基づいて、一定の組合せの機能に制限されないシステムをもたらしている。その代わりに、ランタイム時に、任意の組合せのI/O機能を動的にホット追加し、ホット除去し、またはホッスワップして、所望の構成のシステムをもたらすことができる。

【0004】

しかしながらPCIeおよびCXLトポロジはツリートポロジである。ツリートポロジの欠点は、ホストからのトラフィックがスイッチの上流側のポートを介してソースルートノードからツリーの分岐まで横断しなければならないことである。逆方向のトラフィックも同じツリー横断経路の対象になる。さらに、キャッシュコヒーレンシプロトコルは、総合システム性能に対する待ち時間の均衡がとれない影響のため、待ち時間に対してより敏感な感度を有している。キャッシングエージェントの場合、先行技法では、スイッチボトルネックを介して接続された複数のキャッシングエージェントエンドポイントにコヒーレンシアクションをサービスする際の待ち時間が長くなる。コヒーレンシプロトコルに加えて、先行技法では、ホストとI/Oデバイスとの間でデータを伝送する場合、スイッチのリソースを調停し、また、スイッチを介して輸送しなければならないため、ホストと個々のデバイスとの間の待ち時間が長くなる。さらに、ホストとすべてのデバイスとの間の同時プロトコルメッセージのために、スイッチを介して帯域幅を共有するため、ホストと個々のI/Oデバイスとの間の帯域幅が狭くなる。最後に、スイッチは要求および応答を記憶し、次に、ホストへの単一の上流側接続のために、その要求および応答をすべてのI/Oデバイス間で転送しなければならないため、スイッチのリソースの効率が低下する。

【発明の概要】

【0005】

本明細書において説明される一実施形態は、第1のポートを備えるホストと、第2のポートを含む組み込まれた構成要素であって、第1のポートおよび第2のポートがホストと組み込まれた構成要素との間の物理接続を形成する、組み込まれた構成要素とを備える、コンピューティングシステムであり、組み込まれた構成要素は、複数のI/O機能と、第2のポートを介してホストからパケットを受け取り、パケットのタイプを識別し、かつ、パケットのタイプに基づいて、複数のI/O機能のうちの行先I/O機能へ直接パケットを経路指定するかまたは埋設されたスイッチを使用して行先I/O機能へ間接的にパケットを経路指定するように構成されたパススルーインタフェースとを含む。

【0006】

本明細書において説明される一実施形態は、ホスト上の第2のポートとの物理接続を形成するように構成された第1のポートと、複数のI/O機能と、埋設されたスイッチと、第1のポートを介してホストからパケットを受け取り、パケットのタイプを識別し、かつ、パケットのタイプに基づいて、複数のI/O機能のうちの行先I/O機能へ直接パケットを経路指定するかまたは埋設されたスイッチを使用して行先I/O機能へ間接的にパケットを経路指定するように構成されたパススルーインタフェースとを含む装置である。

【0007】

本明細書において説明される一実施形態は、組み込まれた構成要素のパススルーインタフェースでホストから第1のパケットを受け取ることであって、組み込まれた構成要素が複数のI/O機能と、パススルーインタフェースに通信結合された埋設されたスイッチとを備える、第1のパケットを受け取ることと、第1のパケットがデータパケットであることを決定することであって、複数のI/O機能のうちの第1のI/O機能がそのデータパケットの行先である、決定することと、埋設されたスイッチをバイパスする直接データ経路を使用して、データパケットをパススルーインタフェースから第1のI/O機能へ直接経路指定することと、パススルーインタフェースでホストから第2のパケットを受け取ることと、第2のパケットが構成パケットであることを決定することであって、第1のI/O機能がその構成パケットの行先である、決定することと、データパケットを埋設されたスイッチを介してパススルーインタフェースから第1のI/O機能へ経路指定することとを含む方法である。

【0008】

したがって上で示した特徴を詳細に理解することができる方法においては、上で簡単に要約したより特定の説明には例示的实施形態が参照されており、それらの例示的实施形態のうちのいくつかは添付の図面に示されている。しかしながら添付の図面は典型的な例示的实施形態を示したものにすぎず、したがって本開示の範囲を限定するものと見なしてはならないことに留意されたい。

【図面の簡単な説明】

【0009】

【図1】例による、I/O機能のための異なるデータ経路を有するコンピューティングシステムを示す図である。

【図2】例による、異なるデータ経路を有するパススルーインタフェースを示す図である。

【図3】例による、データパケットおよび構成パケットを異なるデータ経路を使用してホストからI/O機能に伝送するためのフローチャートである。

【図4】例による、データパケットおよび構成パケットを異なるデータ経路を使用してI/O機能からホストに伝送するためのフローチャートである。

【図5】例による、新しいI/O機能にホットスワップするためのフローチャートである。

【図6】例による、新しいI/O機能が追加されているコンピューティングシステムを示す図である。

【図7】例による、ホストがコンバインドネットワークインタフェースカードと通信しているコンピューティングシステムを示す図である。

【発明を実施するための形態】

【0010】

10

20

30

40

50

理解を容易にするために、可能である場合、図面で共通の全く同じ要素を示すために同じ参照符号が使用されている。1つの例の要素は、他の例に有利に組み込むことができることが企図されている。

【0011】

以下、様々な特徴について、図を参照して説明する。図は、スケール通りに描かれていることもあれば、スケール通りに描かれていないこともあること、また、すべての図を通して、同様の構造または機能の要素は同様の参照符号で表されていることに留意されたい。図に意図されていることは、単に特徴の説明を容易にすることにすぎないことに留意されたい。図には、特許請求の範囲の説明についての網羅的な説明であることも、あるいは特許請求の範囲に対する限定であることも意図されていない。さらに、例証されている例は、必ずしも示されている態様または利点のすべてを有している必要はない。特定の例に関連して説明されている態様または利点は必ずしもその例に限定されず、たとえそのように例証されていなくても、あるいはそのように明確に説明されていなくても、任意の他の例で実践することができる。

【0012】

本明細書における実施形態は、組み込まれた構成要素におけるI/O機能とホストとの間で伝送されるデータを個別のデータ経路に分離するための技法を記述している。一実施形態では、データパケット（例えばDMAペイロードおよび記述子、CXLスヌープまたはCCIXメッセージ）は、組み込まれた構成要素のスイッチをバイパスする直接データ経路を使用して伝送される。それとは対照的に、構成パケット（例えばホットスワップ、ホット追加、ホット除去データ、構成制御書込み、または構成状態読出し、等々）はスイッチに伝送され、次に、スイッチがそれらの構成パケットをそれらの行先に転送する。この方法によれば、スイッチ制御経路が2つの経路、すなわちデータパケットのための経路、および構成パケットのための別の経路に分離される。データパケットのための直接経路は、帯域幅または待ち時間に敏感なトラフィックをホストとI/O機能との間で輸送するのにスイッチ接続性を当てにせず、その代わりに、帯域幅、リソース、記憶／転送、およびスイッチの待ち時間特性を回避している（例えばバイパスしている）。一方、ホットプラグ属性または構成レジスタのプログラミング（これらは、待ち時間または帯域幅に敏感ではない）などのソフトウェア互換性属性は、構成データ経路を提供するために、スイッチを使用することによって継続してサポートされる。

【0013】

一実施形態では、組み込まれた構成要素は、ホストから受け取ったデータをI/O機能およびスイッチへ経路指定し、さらにはデータをホストに伝送する際に、I/O機能とスイッチの間で調停するためのパススルーインタフェースを含む。しかしながらデータを待ち行列にバッファする（したがって待ち時間が追加されて、帯域幅に影響を及ぼす）スイッチとは異なり、パススルーインタフェースの経路指定機能および調停機能はパケットを記憶せず、受け取ったパケットを直ちにそれらの行先へ転送する。上で説明したように、パススルーインタフェースは、時間に敏感なデータに対して、スイッチをバイパスする、ホストとI/O機能との間の直接経路を確立することができ、一方、構成データ（待ち時間および時間に敏感ではない）は、スイッチを使用して、ホストとI/O機能との間で経路指定される。この方法によれば、待ち時間または帯域幅に敏感ではないパケットが、待ち時間および帯域幅に敏感なデータによって使用される同じデータ経路を妨げる原因になることはない。

【0014】

図1は、例による、I/O機能140のための異なるデータ経路を有するコンピューティングシステム100を示したものである。詳細には、コンピューティングシステム100は、ホスト105とI/O機能140との間でデータパケットを伝送するための直接データ経路170、およびホスト105とI/O機能140との間で構成パケットを伝送するための間接構成データ経路180を提供している。したがって時間に敏感なデータおよび時間に敏感ではないデータが同じ物理接続を共有する従来の解決法とは異なり、コンピ

ューティングシステム 100 では、組み込まれた構成要素 160 の中で、時間に敏感なデータを時間に敏感ではないデータとは別の経路上で伝送することができる。

【0015】

示されているように、コンピューティングシステム 100 は、ホスト 105 と、I/O 機能 140 を含む組み込まれた構成要素 160 とを含む。この例ではホスト 105 は、1 つまたは複数のプロセッサ 110 およびメモリ 115 を含む。プロセッサ 110 は任意の数の処理要素を代表しており、任意の数の処理要素の各々は任意の数の処理コアを含むことができる。メモリ 115 は、揮発性メモリ要素、不揮発性メモリ要素、またはそれらの両方の組合せを含むことができる。この例ではメモリ 115 は、1 つまたは複数の仮想計算機 (VM) 120 またはテナントをホストしている。これらの VM 120 は、組み込まれた構成要素 160 にタスクの実行を依頼する機能を実施することができる。組み込まれた構成要素 160 の I/O 機能 140 は、次に、これらのタスクを実施することができる。

【0016】

ホスト 105 は、組み込まれた構成要素 160 のポート 130 に結合されているポート 125 を含む。すなわち、ホスト 105 および組み込まれた構成要素 160 の I/O 機能 140 は、同じ対のポート 125、130 を使用してデータを交換している。一実施形態では、ホスト 105 および組み込まれた構成要素 160 は、PCIe プロトコルを使用してポート 125、130 上でデータを交換する。さらに、ポート 125 と 130 の間の同じ物理接続が組み込まれた構成要素 160 の I/O 機能 140 によって共有されている。一実施形態では、ポート 125 と 130 の間の物理接続を任意の所与の時間に使用することができるのは、複数の I/O 機能 140 のうちの 1 つの I/O 機能 140 だけである。複数の I/O 機能 140 の各々が物理接続を使用してホスト 105 とデータを交換する機会を持つよう、時間多重化を使用することができる。この方法によれば、ポート 125 と 130 の間の物理接続（この接続は、典型的には最も広い帯域幅接続である）の帯域幅が I/O 機能 140 同士の間で共有される。

【0017】

組み込まれた構成要素 160 は、複数の I/O 機能 140 および埋設されたスイッチ 150 を統合することができる任意の物理デバイスであってもよい。一実施形態では、組み込まれた構成要素 160 は印刷回路基板 (PCB)（例えば基板）を含むことができ、I/O 機能 140 および埋設されたスイッチ 150 は、PCB の上に取り付けられる個別の集積回路（例えば半導体チップ）である。PCB は、これらの集積回路が PCB に差し込まれるソケットを含むことができる。この方法によれば、集積回路をホットスワップすることができる（例えば第 1 の I/O 機能を実施する 1 つの集積回路がソケットから除去され、第 2 の I/O 機能を実施する第 2 の集積回路に置き換えられる）。別の実施形態では、組み込まれた構成要素 160 は、I/O 機能 140 のための集積回路および埋設されたスイッチ 150 が 1 つまたは複数のチップキャリアパッケージ中に密閉されるシステムインパッケージ (SiP) であってもよい。SiP の場合、I/O 機能 140 をホットスワップすることができないかもしれないが、依然として I/O 機能 140 を選択的に起動し、かつ、起動を解除することができる（例えばホット追加およびホット除去することができる）。

【0018】

さらに別の実施形態では、組み込まれた構成要素 160 は、構成要素 160 のすべての構成要素が同じ集積回路またはチップの中に含まれるシステムオンチップ (SoC) である。SoC は、機能 140 を起動し、または起動を解除することができる（例えばホット追加 (hot added) またはホット除去 (hot removed) する）、I/O 機能 140 を実現するための硬化論理を含むことができる。別法としては、SoC は、I/O 機能 140 をホットスワップすることができるよう、I/O 機能 140 を実現するためのプログラマブル論理を含むことも可能であり、プログラマブル論理が第 2 の I/O 機能を実施するよう、1 つの I/O 機能のためのプログラマブル論理が再構成される。他の実施形態では、組み込まれた構成要素 160 は、組み込まれた構成要素 160 の例証され

10

20

30

40

50

ている回路機構がプログラマブル論理で実現されるFPGA、または回路機構が硬化論理を使用して実現されるASICであってもよい。

【0019】

組み込まれた構成要素160の特定の実施態様に無関係に、コンピューティングシステム100が動作している間、集積回路を物理的に除去し、硬化論理を起動解除／起動し、またはプログラマブル論理を再プログラミングすることにより、I/O機能140を起動し、または起動を解除することができる（例えばホットー追加またはホットー除去することができる）。いくつかの実施形態では、I/O機能140は、基板（例えばPCB）上で第1の集積回路を別の集積回路に置き換えることによって、またはそれまで第1のI/O機能を実施していたプログラマブル論理を第2のI/O機能を実施するように再構成することによってホットスワップすることができる。組み込まれた構成要素160の、ホットスワップ／追加／除去によって影響されない他のI/O機能140は、並行して動作を継続することができる。

10

【0020】

組み込まれた構成要素160は、ポート130、I/O機能140および埋設されたスイッチ150に結合されているパススルーインタフェース135を含む。パススルーインタフェース135は、ポート130を使用してI/O機能140、スイッチ150およびホスト105の間でパケットを送送するための経路指定機能および調停機能を実施している。例えばホスト105からパケットを受け取る場合、パススルーインタフェース135は、パケットをI/O機能140へ複数の直接データ経路170のうちの1つを横切らせるべきか、またはその代わりに埋設されたスイッチ150へ経路指定すべきかどうかを示すパケットのタイプを決定する。パケットを組み込まれた構成要素160からホスト105に伝送する場合、パススルーインタフェース135は、調停論理を使用して、パケットをホスト105に伝送するためにポート130を使用することができるソース（例えばI/O機能140または埋設されたスイッチ150のうちの1つ）を決定することができる。

20

【0021】

一実施形態では、パススルーインタフェース135は、ホスト105、I/O機能140またはスイッチ150から受け取るパケットをバッファせず、または待ち行列に加えない。その代わりにインタフェース135は、待ち時間を加えることなくパケットを「通過」させる。例えばパススルーインタフェース135がパケットを受け取ると、パススルーインタフェース135はそのパケットを直ちに先行へ転送し、したがって受け取られたパケットは、その前に受け取られたパケットがパススルーインタフェース135によって転送されるのを待機する必要はない。パススルーインタフェース135については、図2により詳細に考察される。

30

【0022】

I/O機能140は、組み込まれた構成要素160によって実施される、ホスト105によってオフロードされてもよい任意の機能であってもよい。例えばI/O機能140はアクセラレータであってもよい（例えば図形アクセラレータ、機械学習アクセラレータの人工知能、暗号方式アクセラレータ、圧縮アクセラレータ、等々）。他の例では、I/O機能140は、ネットワーク通信機能（例えばNIC機能）、DMAエンジン、ネットワーク記憶機能、等々であってもよい。

40

【0023】

I/O機能140は、互いに独立して動作することができる個別のI/Oデバイスまたは機能と見なすことができる。例えばI/O機能140Aはネットワーク記憶を実施するDMAエンジンであってもよく、一方、I/O機能140Bは人工知能アクセラレータであってもよい。I/O機能140Aおよび140Bは個別の集積回路であってもよく、または同じ集積回路中の異なる回路機構（例えば異なる硬化論理または異なるプログラマブル論理）であってもよい。いずれの場合においても、以下で考察されるように、I/O機能140は、コンピューティングシステム100が動作している間、ホットー除去する（起動を解除する）ことができ、またはホットー追加する（起動する）ことができる。例え

50

ばホスト105は、組み込まれた構成要素160が新しいI/O機能140を追加し（例えばそれまで起動が解除されていたI/O機能140Bを起動し、または第5のI/O機能（図示せず）を追加する）、またはI/O機能を除去する（例えばそれまで起動されていたI/O機能140Cの起動を解除する）のと同時に、その時点でI/O機能140Aと通信中であってもよい。

【0024】

埋設されたスイッチ150は、パケットをI/O機能140とホスト105との間で経路指定するPCIeスイッチであってもよい。また、スイッチ150は、ホスト105から、I/O機能140には転送されないパケットを受け取ることも可能である。上で言及したように、スイッチ150を使用して、I/O機能140をホットスワップし、ホット追加し、またはホット除去するために使用される構成パケットなどの、待ち時間および帯域幅に敏感ではないデータを経路指定することができる。構成パケットは、送りアクションおよび受取りアクションを開始するためにキャッシュコヒーレンシプロトコルに使用される記述子などの他の情報を同じく含むことができる。

【0025】

図1では、ホスト105によって伝送される、複数のI/O機能140のうちの1つのために意図されている構成パケットは、複数の間接構成データ経路180のうちの1つに沿ったスイッチ150を介して経路指定される。したがって構成パケットはスイッチ150中の待ち行列155に記憶される。この待ち行列155はホストスイッチバッファと呼ぶことも可能である。スイッチ150は、調停機能を実施して、待ち行列155に記憶された構成パケットが伝送され、または処理される時期を決定することができる。

【0026】

注目すべきことには、図1では、直接データ経路170は埋設されたスイッチ150、より詳細には待ち行列155をバイパスしている。したがって直接データ経路170は、待ち行列155によって導入された待ち時間を回避するバイパス経路と呼ぶことも可能である。したがってコンピューティングシステム100は、複数のキャッシュエージェントエンドポイント（すなわちI/O機能140）に対するコヒーレンシアクションをサービスする際にすべてのデータがスイッチ150を通過する従来技術と比較すると、待ち時間を短くする。さらに、本明細書における実施形態は、待ち時間および帯域幅に敏感なデータをホスト105とI/O機能140との間で伝送する際の、スイッチのリソースの調停の必要を回避する。すなわち敏感なデータは、直接データ経路170を使用して、スイッチ150によって実施される調停機能を回避することができる（調停は、以下で説明されるように依然としてパススルーインタフェース135で実施されるが、パススルーインタフェース135は待ち行列を使用しない）。

【0027】

また、ホスト105とすべてのI/O機能140との間の同時プロトコルメッセージは直接データ経路170を使用することができるため、ホスト105および個々のI/O機能140は、これらのメッセージのためにスイッチを介して帯域幅を共有する必要はない。さらに、コンピューティングシステム100は、ポート125および130によって形成されたホスト105への単一の上流側接続のために、スイッチ150を当てにして要求および応答を記憶し、次に、すべてのI/O機能140の間で転送することを回避する。したがってコンピューティングシステム100は、すべてのトラフィックがスイッチ150を介して経路指定される従来技術と比較すると、改善された性能を利用することができ、複数のエンドポイント（例えばI/O機能140）が単一の接続（例えばポート125と130の間の物理接続）を使用して、扇形に広がった形でホスト105に接続される。

【0028】

図2は、例による、異なるデータ経路を有するパススルーインタフェース135を示したものである。すなわち図2は、コンピューティングシステム200が図1に示されている直接データ経路および間接構成データ経路を有することができる、パススルーインタフェース135における回路機構の一例を示している。

【0029】

簡潔にするために、図2の組み込まれた構成要素260は2つのI/O機能140Aおよび140Bしか含んでいないが、任意の数のI/O機能またはI/Oデバイスを含むことができる。データをホスト105からI/O機能140または埋設されたスイッチ150へ経路指定するために、パススルーインタフェース135は、経路指定論理205およびデマルチプレクサ(demux)215を含む。通常、経路指定論理205は、ホスト105のポート125から受け取ったパケット(下流側トラフィックと呼ばれる)の行先を決定する。経路指定論理205は、行先に基づいて、パケットが正しい行先へ経路指定されるよう、demux215の選択線、すなわちI/O機能140または埋設されたスイッチ150のうちの1つを制御する。

10

【0030】

この例では経路指定論理205は、ホスト105から受け取ったパケットに含まれているデータを復号し、そのパケットの行先を決定する復号器210を含む。一実施形態では、復号器210はパケットのタイプならびにそのパケットの行先IDを識別する。パケットのタイプは、パケットを複数のI/O機能140のうちの1つへ直接データ経路を横切らせるべきか、あるいは埋設されたスイッチ150へ間接構成データ経路を横切らせるべきかどうかを決定する。すなわち、データパケットはI/O機能140へ直接送ることができる、一方、構成パケットは埋設されたスイッチ150へ伝送される。パケットがデータパケットであることを復号器210が決定すると、復号器210は、複数のI/O機能140のうちのそのパケットの行先であるI/O機能140を同じく決定することができる。組み込まれた構成要素260にI/O機能140を追加する場合、ホスト105は、ホスト105が復号器210に提供するIDをそのI/O機能140に割り当てることができる。ホスト105によって伝送されるデータパケットにこのIDを埋め込むことにより、復号器210はデータパケットの正しい行先を識別することができ、したがって経路指定論理205は、そのデータパケットを選択されたI/O機能140に対応する直接データ経路上に経路指定する。

20

【0031】

上流側トラフィックを組み込まれた構成要素260からホスト105へ転送するために、パススルーインタフェース135は、ポート130を使用することができる組み込まれた構成要素260の回路構成要素を決定する調停論理220を含む。示されているように、mux225は複数のI/O機能140の各々および埋設されたスイッチ150をポート130に接続している。mux225のための選択信号が調停論理220によって提供される。一実施形態では、調停論理220は、これらの回路構成要素のうちのパケットをポート130に伝送することができる回路構成要素を決定する(例えば調停論理220時間はmux225の選択線を制御する)。この例では調停論理220は、I/O機能140または埋設されたスイッチ150のうちの1つにのみ、ポート130へのアクセスを許容し、したがってデータ衝突は存在しない。調停論理220の詳細は以下でより詳細に考察される。

30

【0032】

図2が示しているように、組み込まれた構成要素260がホスト105から下流側データを受け取り、またはホスト105に上流側データを伝送するかどうかに関係なく、データは、待ち行列に加えられることなくインタフェース135を通過することができる。したがってI/O機能140とホスト105との間の直接データ経路に沿って伝送されるトラフィックは、I/O機能140がそれらとホスト105との間の仲介者として埋設されたスイッチ150を当てにしているシステムと比較すると、待ち時間が短い。

40

【0033】

パススルーインタフェース135と同様、埋設されたスイッチ150も調停論理230を同じく含む。すなわち、待ち行列155は複数のソースからの複数のパケット(例えばI/O機能140から受け取ったパケット、またはスイッチ150中の内部回路機構によって生成されたパケット)を記憶することができるため、調停論理230は、これらのパ

50

ケットのうちの、待ち行列 155 における優先権を持つべきパケットを決定することができる（単純な先入れ先出しモデルではなく）。例えば調停論理 220 および調停論理 230 は、いずれも、I/O 機能によって生成されたトラフィックにスイッチ中の内部回路機構によって生成されたトラフィックよりも上位の優先順位を付けることができ、あるいは I/O 機能 140A から受け取ったトラフィックに I/O 機能 140B から受け取ったトラフィックよりも上位の優先順位を付けることができる。これについては以下でより詳細に考察される。

【0034】

図 3 は、例による、データパケットおよび構成パケットを異なるデータ経路を使用してホストから I/O 機能に伝送するための方法 300 のフローチャートである。ブロック 305 で、組み込まれた構成要素がパススルーインタフェースでホストからパケットを受け取る。一実施形態では、組み込まれた構成要素は、組み込まれた構成要素とホストとの間の共有物理接続を当てにする複数の I/O 機能（または I/O デバイス）を備えている。

【0035】

ブロック 310 で、パススルーインタフェースの復号器が、受け取ったパケットがデータパケットであるか、または構成パケットであるかどうかを決定する。例えばパケットヘッダは、パケットのタイプを示すデータを含むことができる。この情報はホストによってパケットの中に入れることができ、あるいはこの情報は、パケットを伝送するために使用される物理輸送プロトコル（例えば P C I e）の一部であってもよい。いずれの場合においても復号器は、パケット中の情報を復号して、それがデータパケットであるかどうかを決定することができ、あるいはより一般的には、時間に敏感なデータを有するパケットであるか、または構成パケット、例えば時間に敏感ではないデータを有するパケットであるかどうかを決定することができる。

【0036】

データパケットと構成パケットの間の区別は、コンピューティングシステムの特定の実施態様に依拠して変化し得る。例えばデータパケットは、DMA ペイロード、C X L スヌープ、C C I X メッセージ、等々であってもよく、一方、構成パケットは、ホットスワップ、ホット追加またはホット除去を実施するための記述子またはコマンドを含む（例えばホストツー I/O デバイス制御メッセージ）。本明細書における実施形態は、データをパケットタイプに従って分岐することができる任意のシステムと共に使用することができる。

【0037】

パケットがデータパケットである場合、方法 300 はブロック 315 へ進行し、パススルーインタフェースがデータパケットに対応する I/O 機能へ直接経路指定する。言い換えると、パススルーインタフェースの経路指定論理は、組み込まれた構成要素の埋設されたスイッチをバイパスする直接データ経路上でデータパケットを転送する。一実施形態では、経路指定論理の復号器は、受け取ったデータパケットを復号して、そのパケットの行先を識別する（例えばデータパケット中の行先 ID を識別する）。例えばコンピューティングシステムを構成する場合（例えば I/O 機能を追加し、あるいは I/O 機能とホストとの間の通信を確立する場合）、ホストは、経路指定論理に知られている行先 ID を I/O 機能に割り当てることができる。パケットを組み込まれた構成要素に伝送する場合、ホストは行先 ID をパケットの中に埋め込むことができる。復号器は、次に、これらの ID を識別することができ、また、経路指定論理は、例えば d e - m u x を使用して、受け取ったパケットが適切な I/O 機能へ転送されることを保証することができる。

【0038】

しかしながらパケットが構成パケットである場合、方法 300 は、ブロック 315 の代わりにブロック 320 へ進行し、パススルーインタフェースがその構成パケットを埋設されたスイッチへ転送する。ブロック 325 で、埋設されたスイッチが、その構成パケットの行先がスイッチ自体であるか、あるいは複数の I/O 機能のうちの 1 つであるかどうかを決定する。すなわち方法 300 では、ホストは、スイッチに宛てられる、特定のタスク

を実施するようにスイッチを構成することができる構成パケットを伝送することができる。ホストは構成パケットをI/O機能に送ることも可能である。

【0039】

構成パケットがスイッチに宛てられている場合、方法300はブロック330へ進行し、埋設されたスイッチがスイッチの構成(c o n f i g)空間で(図2には示されていない)パケットを処理する。構成パケットは、c o n f i g空間を変えることによってスイッチの構成の操作を変更することができる。

【0040】

構成パケットが複数のI/O機能のうちの1つに宛てられている場合、方法300は、ブロック330の代わりにブロック335へ進行し、埋設されたスイッチがパケットを対応するI/O機能へ転送する。すなわちスイッチは、構成パケットが宛てられているI/O機能を識別し、間接構成データ経路を使用してパケットをそのI/O機能へ転送する。

【0041】

図4は、例による、データパケットおよび構成パケットを異なるデータ経路を使用してI/O機能からホストに伝送するための方法400のフローチャートである。すなわち方法300は、2つのデータ経路を使用してデータをホストから組み込まれた構成要素の様々な回路構成要素に伝送するための技法を説明しているが、方法400は、2つのデータ経路を使用した、組み込まれた構成要素からホストへのデータの伝送を説明している。

【0042】

ブロック405で、埋設されたスイッチが複数のI/O機能のうちの1つから第1の構成パケット(例えば構成応答メッセージ)を受け取る。例えば第1の構成パケットは、ホストによってI/O機能に既に伝送された構成パケットに対する返答であってもよい。

【0043】

並行して、または実質的に同時に、ブロック410で、埋設されたスイッチがスイッチ中のc o n f i g空間から第2の構成パケットを受け取る。あるいは別の実施形態では、埋設されたスイッチは、複数のI/O機能のうちの2つのI/O機能から実質的に同時に2つ(または3つ以上)の構成パケットを受け取ることも可能である。

【0044】

ブロック415で、埋設されたスイッチの調停論理が、埋設されたスイッチの第1の構成パケットと第2の構成パケットの間を調停する。すなわち、第1のパケットおよび第2のパケットを待ち行列に記憶し、パケットをパススルーインタフェースに伝送することができ、次にホストに伝送することができるまで、調停が完了するのを待機することができる。この調停論理は、スイッチ中のc o n f i g空間よりもI/O機能を好み得る、または他の複数のI/O機能のうちの1つまたは複数よりも、複数のI/O機能のうちの1つを好み得るサービス品質(QoS)政策に基づくことができる。

【0045】

スイッチの調停論理が第1のパケットおよび第2のパケットのうちの最初に送る方のパケットを決定しても、スイッチは、パケットをパススルーインタフェースに伝送するまで依然として待機することができる。図2に示されているように、パススルーインタフェース135は、例えばm u x 225を使用してホストにデータを伝送することが許容される回路(例えばスイッチまたは複数のI/O機能のうちの1つ)を決定する独自の調停論理220を有している。

【0046】

ブロック420で、パススルーインタフェースの調停論理(例えば調停論理220)が、スイッチがホストにいつでも伝送することができる構成パケット(例えば第1の構成パケットおよび第2の構成パケット)を有していること、および少なくとも1つのI/O機能がホストのためのデータパケットを有していることの指示を受け取る。すなわち方法400は、組み込まれた構成要素の少なくとも2つのデバイス(例えばスイッチおよび複数のI/O機能のうちの1つ、または複数のI/O機能の複数のI/O機能)がいつでもホストに送ることができるデータを有していることを仮定している。現時点でホストへのデ

10

20

30

40

50

一タの伝送を希望している構成要素が1つだけである場合、調停論理は、何ら調停を実施することなく、単純にその構成要素に物理接続（例えばホストと集積回路の間の物理接続）の使用を許可することができる。

【0047】

しかしながら複数の構成要素がホストへのデータの伝送を希望していると仮定すると、ブロック425で、パススルーインタフェースの調停論理が構成パケットとデータパケットの間で調停する。一実施形態では、調停論理は、データパケットに構成パケットよりも上位の優先順位を付けるQoS政策を使用することができる。言い換えると、QoS政策は、スイッチによって伝送されるパケットよりも、I/O機能から直接伝送されるパケットを好むことができる。別の例では、QoS政策はI/O機能を互いに対して優先順位を付けることができる。例えばホストのVM（またはテナント）は異なる優先順位を有することができる。ホストの優先順位がより高いVMによって使用される組み込まれた構成要素の1つまたは複数のI/O機能には、ホストのより優先順位が低いVMによって使用されるI/O機能よりも、パススルーインタフェースの調停論理によって使用されるQoS政策におけるより高い優先順位を与えることができる。

【0048】

ブロック430で、パススルーインタフェースの調停論理が、ホストへの選択されたパケット（調停によって決定された）の伝送を許可する。一実施形態では、調停論理は重付き調停を有しており、特定の時間の間、共有バスに調停論理がアクセスすることができること（あるいは特定の量または特定の数のデータを送ること）を、複数のI/O機能のうちの1つまたはスイッチに知らせる。この方法によれば、調停論理は、組み込まれた構成要素とホストとの間の共有物理接続を使用することができる組み込まれた構成要素の構成要素を制御することができる。

【0049】

図5は、例による、新しいI/O機能にホットスワップするための方法500のフローチャートである。説明を容易にするために、方法500は、例による、新しいI/O機能が追加されているコンピューティングシステムを示す図6と連携して考察される。

【0050】

ブロック505で、組み込まれた構成要素が新しいI/O機能を追加する要求をホストから受け取る。一実施形態では、組み込まれた構成要素のためのソフトウェアドライバ（ホストの中で実行される）が、新しいI/O機能を組み込まれた構成要素にホット追加することを決定する。例えばホスト上で実行するVMまたはテナントが新しいI/O機能のための要求を送っていてもよく、あるいはVMまたはテナントが新しいI/O機能を要求していることをハイパーバイザが決定する。

【0051】

図6では、コンピューティングシステム600は、I/O機能の追加を処理中の組み込まれた構成要素660を含む。すなわち、ダッシュ線で示されているようにAccelerator Function 0 (AF0) およびCXL Cache Xが組み込まれた構成要素660の中に追加されており、一方、AF1およびCXL Cache Yは組み込まれた構成要素660の中で既に動作中である。図6では、I/O機能すなわちAF0、AF1、CXL Cache XおよびCXL Cache Yはプログラマブル論理で実現されており、一方、AF0 Config Space、ID-X Config Space、ID-Y Config Space、AF1 Config Spaceおよび埋設されたCXL分離スイッチ615は硬化回路機構で実現されていることが仮定されている。すなわちプログラマブル論理を再構成することにより、組み込まれた構成要素660は、I/O機能すなわちAF0、AF1、CXL Cache XおよびCXL Cache Yをホットスワップする（すなわちホット追加またはホット除去する）ことができる。しかしながら別の実施形態では、I/O機能は硬化論理で実現することも可能である。その例では、I/O機能を追加または除去するのではなく、ホスト105は、I/O機能を選択的に起動し、または起動を解除することによってI/O機能をホッ

10

20

30

40

50

ト追加またはホット除去することができる。

【0052】

別の実施形態では、I/O機能すなわちAF0、AF1、CXL.Cache XおよびCXL.Cache Y、ならびにAF0 Config Space、ID-X Config Space、ID-Y Config Space、AF1 Config Spaceは、AF0およびCXL.Cache Xが組み込まれた構成要素660に追加される場合、ホット追加事象の開始に先立って、部分的に再構成されたプログラブル論理ビットストリームがAF0およびCXL.Cache Xに対して追加されるよう、プログラブル論理で実現されている。この実施形態では、AF0およびCXL.Cache Xは、いずれも、プログラブル論理ビットストリームとして先にロードされる機能性を有するホット差込みデバイスであってもよい。

10

【0053】

ブロック510で、組み込まれた構成要素がホストから新しいI/O機能のための構成情報およびバインディングを受け取る。一実施形態では、構成情報は、組み込まれた構成要素にI/O機能を追加し、または組み込まれた構成要素のI/O機能を起動するためのデータを含むことができる。さらに、ホストは、ホストによって新しいI/O機能に割り当てられた、I/O機能のためのバインディングとして使用される、パススルーインタフェース（より詳細にはパススルーインタフェースの経路指定論理）に識別データを知らせる識別データを伝送する。パススルーインタフェースの経路指定論理は、次に、受け取ったデータパケットを復号して、上記方法300で説明したようにそのパケットを新しいI/O機能に経路指定すべきかどうかを決定する際に、この識別データを使用することができる。

20

【0054】

ブロック510は、新しいI/O機能、データ経路バインディングおよび構成データバインディングのための構造を含むビットストリームを集積回路が受け取るサブブロック515を含む。一実施形態では、サブブロック515は、I/O機能がプログラブル論理で実現される場合に実施される。例えば組み込まれた構成要素660は、ビットストリームを使用して、AF0およびCXL.Cache Xを含むようにプログラブル論理を構成することができる。また、ビットストリームは、AF0およびID-X Config Spacesのレジスタのための構造を同じく含むことができる。

30

【0055】

データ経路バインディングは、パススルーインタフェースがデータパケットを新しいI/O機能へ直接経路指定するために使用する経路指定情報を提供することができる。一方、構成データバインディングは、埋設されたスイッチおよびパススルーインタフェース135が間接構成データ経路を使用して構成データパケットを新しいI/O機能へ経路指定するために使用する経路指定情報を含む。すなわちデータ経路バインディングにより、データはCXL.Cache Xにパススルーインタフェース135から直接到達することができ、一方、構成データバインディングにより、データは埋設されたCXLスイッチ615を介してAF0およびID-X Config Spacesに到達することができる。

【0056】

ブロック520で、組み込まれた構成要素が新しいI/O機能およびそのバインディングを起動する。すなわち組み込まれた構成要素が、新しいI/O機能を含むようにプログラブル論理を構成するか、またはブロック510で得た情報を使用してそれまで駆動が解除されていた硬化回路機構中のI/O機能を起動する。

40

【0057】

ブロック525で、組み込まれた構成要素が仮想Hot-Plug Eventをホストに伝送する。図6では、スイッチ615は仮想Hot-Plug Eventを生成し、その事象をホスト上で実行するHost Hot-Plugソフトウェアドライバへ転送している。新しいI/O機能は上流側ポートに直接取り付けられてはいるが、仮想Hot-Plug Eventは、AF0 Config Spaceとスイッチ615の間の仮想

50

エンドポイント接続に接続されている仮想下流側ポートに新しいI/O機能（例えば新しいI/Oデバイス）が差し込まれている（または通信結合されている）ことをホストに示す。

【0058】

ブロック530で、ホストが、構成データ経路上で送られた構成パケットを使用して新しいI/O機能を発見する。例えばHost Hot-Plug Software Driverは仮想Hot-Plug Eventに応答することができ、また、CSLルートポート（RP）605からCXL上流側ポート（USP）610へ、また、パススルーインタフェース135を介してスイッチ615へ経路指定された構成読出しメッセージを使用して、新しいエンドポイントI/O機能AF0およびCXL.Cache Xの発見へ進行することができる。スイッチ615は、次に、構成読出しメッセージをAF0およびID-X Config Spaces中の仮想エンドポイントレジスタへ転送することができる。

10

【0059】

ブロック535で、ホストが、対応するレジスタをプログラミングすることによって新しいI/O機能を列挙する。一実施形態では、ホスト105は、CXL.Cache XのデバイスIDおよびAF0のデバイスIDを使用して、AF0およびID-X Config SpacesレジスタをプログラミングすることによってAF0およびCXL.Cache Xを列挙する。ホスト105は、次に、直接データ経路および間接構成データ経路を使用して、データトラフィックを新しいI/O機能にいつでも通信することができる。

20

【0060】

一実施形態では、以上のブロックが完了すると、ブロック540で、ホストおよび組み込まれた構成要素が直接データ経路を使用してデータパケットを新しいI/O機能へ経路指定する。ブロック545で、ホストおよび組み込まれた構成要素が間接構成データ経路を使用して構成パケットを新しいI/O機能へ経路指定する。この方法によれば、ホストおよび組み込まれた構成要素は新しいI/Oデバイスをホット追加することができる。

【0061】

図7は、例による、ホストが組み込まれた構成要素760を使用して実現されたコンバージドNICと通信しているコンピューティングシステム700を示したものである。上記コンピューティングシステムと同様、コンピューティングシステム700は、PCIe RP705とPCIe USP710の間の単一の物理接続を使用して組み込まれた構成要素760に通信結合されているホスト105を含む。さらに、組み込まれた構成要素760は、ホスト105とI/O機能（すなわちDMA Engines 0～3）との間の直接データ経路、および埋設されたPCIe分離スイッチ715を含む間接構成データ経路を確立するためのパススルーインタフェース135を含む。

30

【0062】

図7は、PCIe接続Server（例えばホスト105）に接続されたPCIe構成要素（例えば組み込まれた構成要素760）を示している。組み込まれた構成要素760は、ホスト105に対する、待ち時間が短く、かつ、帯域幅が広いインタフェースを有するPCI DMA Engines 0～3、および対応するサービスを含み、対応するサービスは、それらの制御および状態構造または構成空間とは別である。この例では個々のDMA Engines 0～3は、コンバージドNIC（SmartNICとも呼ばれる）の異なるネットワーク機能に対応している。例えばDMA Engine 0はネットワークサービスに対応し、DMA Engine 1はRemote Direct Memory Access（RDMA）サービスに対応し、DMA Engine 2はNon-Volatile Memory Express Over Fiber（NVMeoF）サービスに対応し、また、DMA Engine 3は記憶サービスに対応している。Network、RDMA、NVMeoFおよびStorage ServicesのためのHigh Bandwidth Device-to-Host DMAトラフィックは直接

40

50

データ経路を追従している。さらに、これらのNetwork、RDMA、NVMe over Storage Servicesのために宛てられる待ち時間が短いHost-to-I/O Function Job Descriptorsも同じく直接データ経路を追従することができる。それとは対照的に、PCIeスイッチ715は、DMA Engines 0～3ならびに対応するNetwork、RDMA、NVMe over Storage Configuration SpacesのためのHost-to-I/O Function低性能制御経路トラフィックを間接構成データ経路に沿って経路指定することができる。

【0063】

DMA Engines 0～3は、プログラマブル論理または硬化回路機構を使用して実現することができる。さらに、上で考察したホッター追加および除去技法を使用して、DMA Engines 0～3および対応するサービスを追加し、除去することができる（例えば起動し、起動を解除することができる）。

【0064】

埋設されたスイッチを含む間接経路とは別の直接データ経路を作り出すことにより、本明細書における実施形態は、エンドポイント（例えばI/O機能）をホッター差込み、追加／除去する能力を含む、ホストに対する、待ち時間が短く、帯域幅が広いデータ経路インタフェースをもたらす。スイッチを介して調停する代わりに直接データ経路を使用することにより、CXL Cache実施形態およびCXL Cache実施形態のための待ち時間が短いスヌープ応答、CXL.mem実施形態およびCXL Home Agent and Slave Agent実施形態のための待ち時間が短く、かつ、帯域幅が広いメモリトラフィック、およびPCIe Endpoint実施形態のための待ち時間が短い記述子通信および帯域幅が広いDMA Reads/Writesなどの、多くの実施形態のための優れた性能が得られる。

【0065】

以上の説明には、本開示において提示されている実施形態が参照されている。しかしながら本開示の範囲は、説明されている特定の実施形態に限定されない。そうではなく、異なる実施形態に関連しているにせよ、あるいは関連していないにせよ、説明されている特徴および要素の任意の組合せが実現されるべく企図されており、また、企図された実施形態を實踐することが企図されている。さらに、本明細書において開示されている実施形態は、他の可能な解決法に優る、あるいは従来技術に優る利点を達成することができるが、特定の利点が所与の実施形態によって達成されるにせよ、あるいは達成されないにせよ、本開示の範囲を限定しない。したがって以上の態様、特徴、実施形態および利点は単に例証的なものにすぎず、特許請求項で明確に記載されている場合を除き、添付の特許請求の範囲の要素または限定とは見なされない。

【0066】

当業者には認識されるように、本明細書において開示されている実施形態は、システム、方法またはコンピュータプログラム製品として具体化することができる。したがって態様は、全面的にハードウェア実施形態の形態、全面的にソフトウェア実施形態の形態（ファームウェア、常駐ソフトウェア、マイクロコード、等々を含む）、またはソフトウェア態様およびハードウェア態様を組み合わせた実施形態の形態を取ることができ、これらは、すべて、本明細書においては、一括して「回路」、「モジュール」または「システム」と呼ぶことができる。さらに、態様は、コンピュータ可読プログラムコードがその上に具体化された、1つまたは複数のコンピュータ可読媒体の中に具体化されたコンピュータプログラム製品の形態を取ることにも可能である。

【0067】

任意の組合せの1つまたは複数のコンピュータ可読媒体を利用することができる。コンピュータ可読媒体は、コンピュータ可読信号媒体またはコンピュータ可読記憶媒体であってもよい。コンピュータ可読記憶媒体は、例えばそれらに限定されないが、電子、磁気、光、電磁、赤外線または半導体のシステム、装置またはデバイス、あるいはそれらの任意

10

20

30

40

50

の適切な組合せであってもよい。コンピュータ可読記憶媒体のより具体的な例としては（非排他的列挙）、1つまたは複数の配線を有する電気接続、携帯型コンピュータディスク、ハードディスク、ランダムアクセスメモリ（RAM）、リードオンリメモリ（ROM）、消去可能プログラマブルリードオンリメモリ（EPROMまたはフラッシュメモリ）、光ファイバ、携帯型コンパクトディスクリードオンリメモリ（CD-ROM）、光記憶デバイス、磁気記憶デバイスまたはそれらの任意の適切な組合せを挙げることができる。この文書の文脈においては、コンピュータ可読記憶媒体は、命令実行システム、装置またはデバイスが使用するための、あるいはこれらの命令実行システム、装置またはデバイスと関連して使用するためのプログラムを含み、または記憶することができる任意の有形媒体である。

10

【0068】

コンピュータ可読信号媒体は、中に、例えばベースバンドの中に、または搬送波の一部として、コンピュータ可読プログラムコードが具体化された伝搬データ信号を含むことができる。このような伝搬信号は、それらに限定されないが、電磁、光またはそれらの任意の適切な組合せを含む任意の様々な形態を取ることができる。コンピュータ可読信号媒体は、コンピュータ可読記憶媒体ではなく、命令実行システム、装置またはデバイスが使用するための、あるいはこれらの命令実行システム、装置またはデバイスと関連して使用するためのプログラムを通信し、伝搬し、または輸送することができる任意のコンピュータ可読媒体であってもよい。

【0069】

20

コンピュータ可読媒体の上に具体化されたプログラムコードは、それらに限定されないが、無線、配線、光ファイバケーブル、RF、等々、またはそれらの任意の適切な組合せを含む任意の適切な媒体を使用して伝送することができる。

【0070】

本開示の態様のための操作を実施するためのコンピュータプログラムコードは、Java、Smalltalk、C++、等々などのオブジェクト指向プログラミング言語、および「C」プログラミング言語などの従来の手続きプログラミング言語、または同様のプログラミング言語を含む1つまたは複数のプログラミング言語の任意の組合せで作成することができる。プログラムコードは、ユーザのコンピュータ上ですべてを実行し、ユーザのコンピュータ上で、独立ソフトウェアパッケージとして一部を実行し、ユーザのコンピュータ上で一部を実行し、かつ、遠隔コンピュータ上で一部を実行し、あるいは遠隔コンピュータまたはサーバ上ですべてを実行することができる。後者のシナリオの場合、遠隔コンピュータは、ローカルエリアネットワーク（LAN）または広域ネットワーク（WAN）を含む任意のタイプのネットワークを介してユーザのコンピュータに接続することができ、あるいはこの接続は、外部コンピュータに対して実施することも可能である（例えばインターネットサービスプロバイダを使用してインターネットを介して）。

30

【0071】

本開示の態様は、以下で、フローチャート図解、および／または本開示で提示されている実施形態による方法、装置（システム）およびコンピュータプログラム製品のブロック図を参照して説明される。フローチャート図解および／またはブロック図の個々のブロック、およびフローチャート図解および／またはブロック図におけるブロックの組合せは、コンピュータプログラム命令によって実現することができることは理解されよう。これらのコンピュータプログラム命令は、コンピュータまたは他のプログラマブルデータ処理装置のプロセッサを介して実行する命令が、フローチャートおよび／またはブロック図の1つまたは複数のブロックに明記されている機能／作用を実現するための手段を作り出すよう、汎用コンピュータ、専用コンピュータまたは他のプログラマブルデータ処理装置のプロセッサに提供して機械を製造することができる。

40

【0072】

これらのコンピュータプログラム命令は、コンピュータ可読媒体に記憶された命令が、フローチャートおよび／またはブロック図の1つまたは複数のブロックに明記されている

50

機能／作用を実現する命令を含む製造の物品をもたらすよう、特定の方法で機能するようにコンピュータ、他のプログラマブルデータ処理装置または他のデバイスを導くことができるコンピュータ可読媒体に同じく記憶することができる。

【0073】

コンピュータプログラム命令は、コンピュータまたは他のプログラマブル装置上で実行する命令が、フローチャートおよび／またはブロック図の1つまたは複数のブロックに明記されている機能／作用を実現するためのプロセスを提供するよう、コンピュータ、他のプログラマブルデータ処理装置または他のデバイスに同じくロードして、それらのコンピュータ、他のプログラマブル装置または他のデバイス上で一連の操作ステップを実施させ、それによりコンピュータ実現プロセスを生成することができる。

10

【0074】

図のフローチャートおよびブロック図は、本発明の様々な例によるシステム、方法およびコンピュータプログラム製品の可能な実施態様のアーキテクチャ、機能性および操作を示している。この点に関し、フローチャートまたはブロック図における個々のブロックは、明記されている論理機能を実現するための1つまたは複数の実行可能命令を含む、命令のモジュール、セグメントまたは部分を表すことができる。いくつかの代替実施態様では、ブロックの中で言及されている機能は、図の中で言及されている順序から外れて生じ得る。例えば連続して示されている2つのブロックは、実際、実質的に同時に実行することができ、あるいはこれらのブロックは、時によっては、含まれている機能性に依じて逆の順序で実行することも可能である。ブロック図および／またはフローチャート図解の個々のブロック、およびブロック図および／またはフローチャート図解におけるブロックの組合せは、明記された機能または作用を実施し、あるいは専用ハードウェアとコンピュータ命令の組合せを実施する専用ハードウェアベースシステムによって実現することができることに同じく留意されよう。

20

【0075】

本開示は、以下の非限定の例のうちの1つまたは複数によって説明することも可能である。

【0076】

例1．第1のポートを含むホストと、組み込まれた構成要素とを含むコンピューティングシステム。組み込まれた構成要素は、第2のポートであって、第1のポートおよび第2のポートがホストと組み込まれた構成要素との間の物理接続を形成する、第2のポートと、複数のI/O機能と、埋設されたスイッチと、第2のポートを介してホストからパケットを受け取り、パケットのタイプを識別し、かつ、パケットのタイプに基づいて、複数のI/O機能のうちの行先I/O機能へ直接パケットを経路指定するかまたは埋設されたスイッチを使用して行先I/O機能へ間接的にパケットを経路指定するように構成されたパススルーインタフェースとを含む。

30

【0077】

例2．複数のI/O機能が互いに独立して動作し、異なるI/O機能を実施する、例1のコンピューティングシステム。

【0078】

例3．パケットのタイプが、そのパケットが時間に敏感なデータパケットであるかどうかを示し、そのパケットを行先I/O機能へ直接経路指定することが、埋設されたスイッチをバイパスする、例1のコンピューティングシステム。

40

【0079】

例4．パススルーインタフェースが、複数のI/O機能のうちの第1のI/O機能および埋設されたスイッチの両方が、いつでもホストに伝送することができるデータを有していることの指示を受け取り、また、データをホストの第1のポートに伝送するために、第1のI/O機能および埋設されたスイッチのうちのどちらに第2のポートの使用を許可するかを決定するために調停するように構成された調停論理を含む、例1のコンピューティングシステム。

50

【0080】

例5．パススルーインタフェースが、パケットのタイプを識別し、また、行先I/O機能に対応するパケット中の行先IDを識別するための経路指定論理を含む、例1のコンピューティングシステム。

【0081】

例6．ホストが、複数のI/O機能が並行して動作している間、新しいI/O機能を組み込まれた構成要素の中にホット追加するように構成される、例1のコンピューティングシステム。

【0082】

例7．新しいI/O機能をホット追加することが、その新しいI/O機能を含むように組み込まれた構成要素のプログラマブル論理を構成することを含む、例6のコンピューティングシステム。

10

【0083】

例8．新しいI/O機能をホット追加することが、それまで起動が解除されていた新しいI/O機能を起動することを含み、その新しいI/O機能が硬化論理で実現される、例6のコンピューティングシステム。

【0084】

例9．ホスト上の第2のポートとの物理接続を形成するように構成された第1のポートと、複数のI/O機能と、埋設されたスイッチと、第1のポートを介してホストからパケットを受け取り、パケットのタイプを識別し、かつ、パケットのタイプに基づいて、複数のI/O機能のうちの行先I/O機能へ直接パケットを経路指定するかまたは埋設されたスイッチを使用して行先I/O機能へ間接的にパケットを経路指定するように構成されたパススルーインタフェースとを含む装置。

20

【0085】

例10．複数のI/O機能が独立したI/Oデバイスであり、独立したI/Oデバイスの各々が少なくとも部分的にプログラマブル論理で実現される、例9の装置。

【0086】

例11．独立したI/Oデバイスが同じ集積回路の中に形成される、例10の装置。

【0087】

例12．複数のI/O機能が独立したI/Oデバイスであり、独立したI/Oデバイスの各々が硬化論理を使用して実現される、例9の装置。

30

【0088】

例13．独立したI/Oデバイスが同じ集積回路の中に形成される、例12の装置。

【0089】

例14．基板をさらに含み、複数のI/O機能の各々が異なる集積回路で実現され、異なる集積回路および埋設されたスイッチが基板の上に取り付けられる、例9の装置。

【0090】

例15．装置が、複数のI/O機能が並行して動作している間、新しいI/O機能をホット追加するように構成される、例9の装置。

【0091】

例16．新しいI/O機能をホット追加することが、その新しいI/O機能を含むように装置のプログラマブル論理を構成することを含む、例15の装置。

40

【0092】

例17．新しいI/O機能をホット追加することが、それまで起動が解除されていた新しいI/O機能を起動することを含み、その新しいI/O機能が装置の硬化論理で実現される、例15の装置。

【0093】

例18．組み込まれた構成要素のパススルーインタフェースでホストから第1のパケットを受け取ることであって、組み込まれた構成要素が複数のI/O機能と、パススルーインタフェースに通信結合された埋設されたスイッチとを含む、第1のパケットを受け取る

50

ことと、第1の packets がデータ packets であることを決定することであって、複数の I/O 機能のうちの第1の I/O 機能がそのデータ packets の行先である、決定することと、埋設されたスイッチをバイパスする直接データ経路を使用して、データ packets をパススルーインタフェースから第1の I/O 機能へ直接経路指定することと、パススルーインタフェースでホストから第2の packets を受け取ることと、第2の packets が構成 packets であることを決定することであって、第1の I/O 機能がその構成 packets の行先である、決定することと、データ packets を埋設されたスイッチを介してパススルーインタフェースから第1の I/O 機能へ経路指定することを含む方法。

【0094】

例19. データ packets を埋設されたスイッチを介してパススルーインタフェースから第1の I/O 機能へ経路指定することが、構成 packets をパススルーインタフェースから埋設されたスイッチへ経路指定することと、埋設されたスイッチで、第1の I/O 機能がその構成 packets の行先であることを決定することと、構成 packets を埋設されたスイッチから第1の I/O 機能へ転送することとをさらに含み、方法が、第3の packets をホストからパススルーインタフェースで受け取ることと、第3の packets が異なる構成 packets であることを決定することと、その異なる構成 packets を埋設されたスイッチへ経路指定することと、埋設されたスイッチで、埋設されたスイッチがその異なる構成 packets の行先であることを決定することとをさらに含み、例18の方法。

【0095】

例20. パススルーインタフェースで、埋設されたスイッチおよび複数の I/O 機能のうちの少なくとも2つがホストにいつでも伝送することができる packets を有していることを決定することをさらに含み、データをホストに伝送するために、どちらに組み込まれた構成要素の共有ポートの使用を許可するかを決定するべく、サービス品質 (QoS) 政策に基づいて、埋設されたスイッチと複数の I/O 機能のうちの少なくとも2つとの間で調整する、例19の方法。

【0096】

以上では特定の例が対象にされているが、それらの基本的な範囲を逸脱することなく、他の例およびさらなる例を工夫することも可能であり、それらの範囲は、添付の特許請求の範囲によって決定される。

10

20

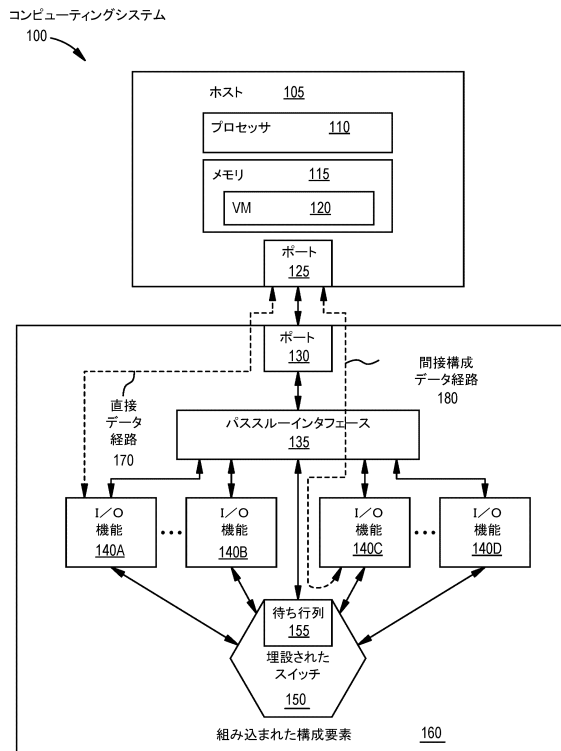
30

40

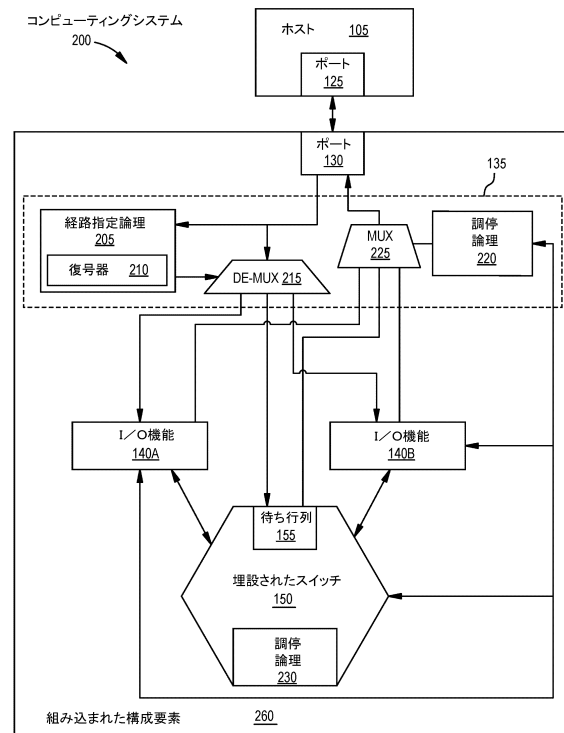
50

【図面】

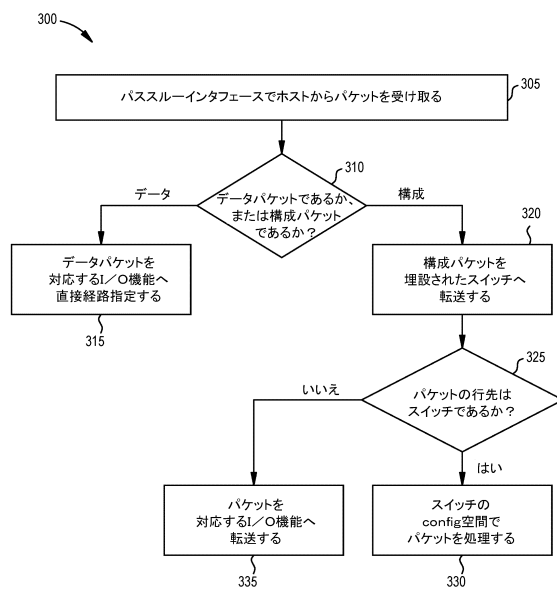
【図 1】



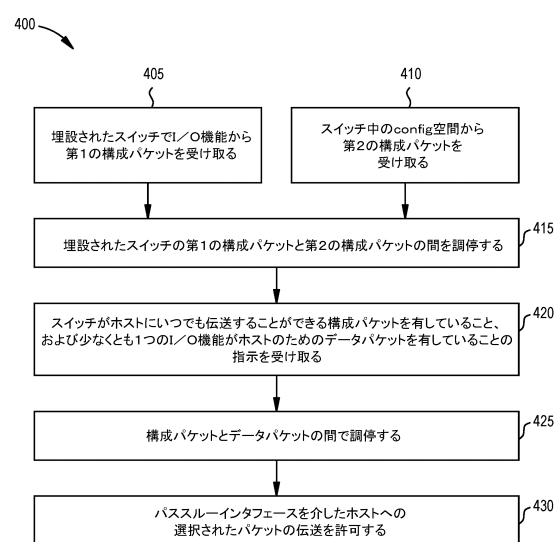
【図 2】



【図 3】



【図 4】



10

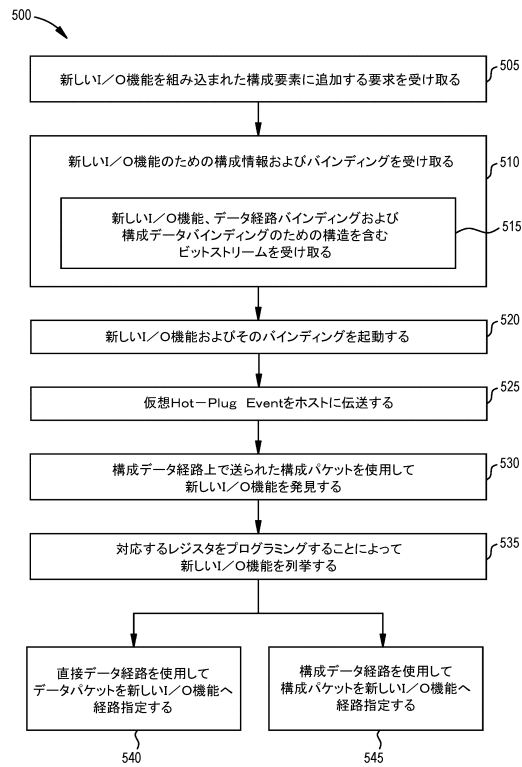
20

30

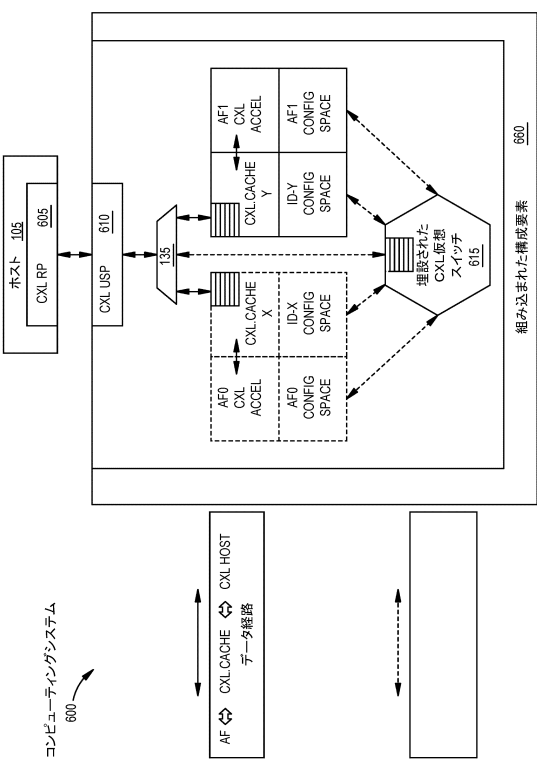
40

50

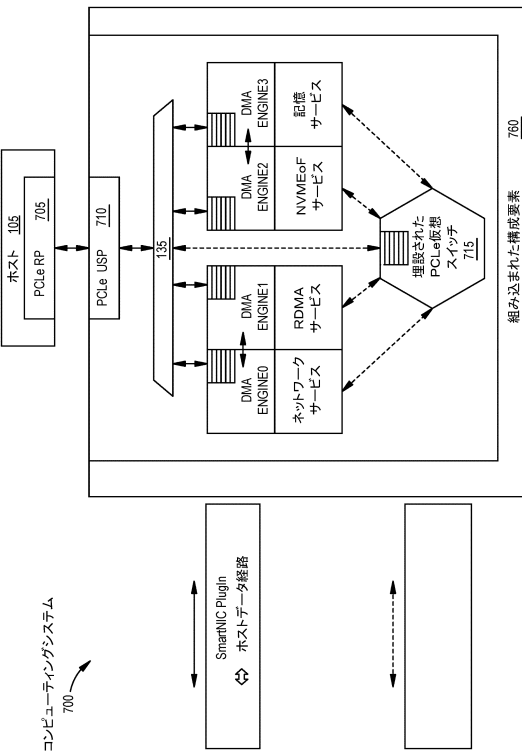
【図 5】



【図 6】



【図 7】



10

20

30

40

50

フロントページの続き

2100

審査官 豊田 真弓

- (56)参考文献 特表2019-530100 (JP, A)
国際公開第2018/003629 (WO, A1)
米国特許出願公開第2005/0005084 (US, A1)
米国特許出願公開第2009/0083471 (US, A1)
- (58)調査した分野 (Int.Cl., DB名)
G06F 13/36
G06F 13/42
G06F 13/14