

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2013-504130

(P2013-504130A)

(43) 公表日 平成25年2月4日(2013.2.4)

(51) Int.Cl.	F I	テーマコード (参考)
G06F 15/17 (2006.01)	G06F 15/17 630A	5B005
G06F 15/80 (2006.01)	G06F 15/80	5B013
G06T 1/20 (2006.01)	G06T 1/20 B	5B045
G06T 1/60 (2006.01)	G06T 1/60 450C	5B047
G06F 9/38 (2006.01)	G06F 9/38 370C	5B057
審査請求 未請求 予備審査請求 未請求 (全 17 頁) 最終頁に続く		

(21) 出願番号 特願2012-528080 (P2012-528080)
 (86) (22) 出願日 平成22年9月3日 (2010.9.3)
 (85) 翻訳文提出日 平成24年4月27日 (2012.4.27)
 (86) 国際出願番号 PCT/US2010/047784
 (87) 国際公開番号 W02011/028984
 (87) 国際公開日 平成23年3月10日 (2011.3.10)
 (31) 優先権主張番号 12/616,636
 (32) 優先日 平成21年11月11日 (2009.11.11)
 (33) 優先権主張国 米国 (US)
 (31) 優先権主張番号 61/239,730
 (32) 優先日 平成21年9月3日 (2009.9.3)
 (33) 優先権主張国 米国 (US)

(71) 出願人 591016172
 アドバンスド・マイクロ・ディバイシズ・
 インコーポレイテッド
 ADVANCED MICRO DEVI
 CES INCORPORATED
 アメリカ合衆国、94088-3453
 カリフォルニア州、サニベイ、ピー・
 オウ・ボックス・3453、ワン・エイ・
 エム・ディ・プレイス、メイル・ストップ
 ・68 (番地なし)
 (74) 代理人 100108833
 弁理士 早川 裕司
 (74) 代理人 100111615
 弁理士 佐野 良太

最終頁に続く

(54) 【発明の名称】 汎用使用のための処理ユニット内部メモリ

(57) 【要約】

【解決手段】

汎用使用のための内部メモリを有するグラフィクス処理ユニット (GPU) 及びそのアプリケーションがここに開示される。そのようなGPUは、第1の内部メモリと、第1の内部メモリに結合される実行ユニットと、第1の内部メモリを他の処理ユニットの第2の内部メモリに結合するように構成されるインタフェースと、を含む。第1の内部メモリは積層ダイナミックランダムアクセスメモリ (DRAM) 又は埋め込みDRAMを備えてよい。インタフェースは第1の内部メモリをディスプレイデバイスに結合するように更に構成されてよい。GPUは第1の内部メモリを中央処理ユニットに結合するように構成される別のインタフェースを含んでもよい。またGPUはソフトウェアにおいて具現化され且つ/又はコンピューティングシステム内に含まれてよい。

【選択図】 図2

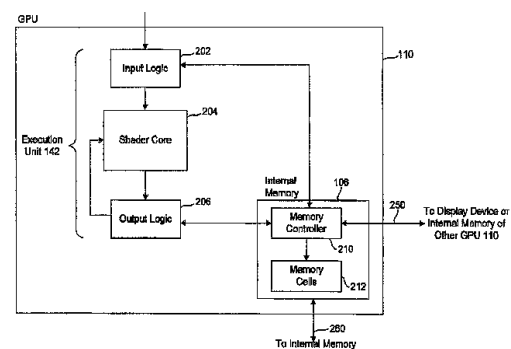


FIG. 2

【特許請求の範囲】**【請求項 1】**

第 1 の内部メモリと、
前記第 1 の内部メモリに結合される実行ユニットと、
前記第 1 の内部メモリを他の処理ユニットの第 2 の内部メモリに結合するように構成されるインタフェースと、を備えるグラフィクス処理ユニット（GPU）。

【請求項 2】

前記他の処理ユニットは GPU を備える請求項 1 の処理ユニット。

【請求項 3】

前記他の処理ユニットは中央処理ユニットを備える請求項 1 の処理ユニット。

10

【請求項 4】

前記第 1 の内部メモリは積層ダイナミックランダムアクセスメモリを備える請求項 1 の処理ユニット。

【請求項 5】

前記第 1 の内部メモリは埋め込みダイナミックランダムアクセスメモリを備える請求項 1 の処理ユニット。

【請求項 6】

前記インタフェースは前記第 1 の内部メモリをディスプレイデバイスに結合するように更に構成される請求項 1 の処理ユニット。

【請求項 7】

コンピューティングデバイス上で実行される場合にグラフィクス処理ユニット（GPU）を定義する命令が入っているコンピュータ可読記憶媒体を備えるコンピュータプログラム製品であって、前記 GPU は、

20

第 1 の内部メモリと、

前記第 1 の内部メモリに結合される実行ユニットと、

前記第 1 の内部メモリを他の処理ユニットの第 2 の内部メモリに結合するように構成されるインタフェースと、を備えるコンピュータプログラム製品。

【請求項 8】

前記他の処理ユニットは GPU を備える請求項 7 のコンピュータプログラム製品。

【請求項 9】

前記他の処理ユニットは中央処理ユニットを備える請求項 7 のコンピュータプログラム製品。

30

【請求項 10】

前記 GPU の前記第 1 の内部メモリは積層ダイナミックランダムアクセスメモリを備える請求項 7 のコンピュータプログラム製品。

【請求項 11】

前記 GPU の前記第 1 の内部メモリは埋め込みダイナミックランダムアクセスメモリを備える請求項 7 のコンピュータプログラム製品。

【請求項 12】

前記 GPU はハードウェア記述言語ソフトウェアにおいて具現化される請求項 7 のコンピュータプログラム製品。

40

【請求項 13】

前記 GPU はベリログハードウェア記述言語ソフトウェア、ベリログ A ハードウェア記述言語ソフトウェア及び VHDL ハードウェア記述言語ソフトウェアの 1 つにおいて具現化される請求項 7 のコンピュータプログラム製品。

【請求項 14】

第 1 の内部メモリと前記第 1 の内部メモリに結合される第 1 の実行ユニットと前記第 1 の内部メモリを他の GPU の内部メモリに結合するように構成される第 1 のインタフェースとを備える第 1 のグラフィクス処理ユニット（GPU）と、

第 2 の内部メモリと前記第 2 の内部メモリに結合される第 2 の実行ユニットと前記第 2

50

の内部メモリを他のGPUの内部メモリに結合するように構成される第2のインタフェースとを備える第2のGPUと、を備えるシステムであって、

前記第1の内部メモリ及び前記第2の内部メモリは互いに結合されて前記第1のGPUの前記第1の実行ユニットが前記第2のGPUの前記第2の内部メモリにアクセスすることを可能にすると共に前記第2のGPUの前記第2の実行ユニットが前記第1のGPUの前記第1の内部メモリにアクセスすることを可能にするシステム。

【請求項15】

前記第1の内部メモリは積層ダイナミックランダムアクセスメモリを備える請求項14のシステム。

【請求項16】

前記第1の内部メモリは埋め込みダイナミックランダムアクセスメモリを備える請求項14のシステム。

【請求項17】

前記第1のインタフェースは前記第1の内部メモリをディスプレイデバイスに結合するように更に構成され、

前記第2のインタフェースは前記第2の内部メモリを前記ディスプレイデバイスに結合するように更に構成される請求項16のシステム。

【請求項18】

外部メモリと、

キャッシュメモリを備える中央処理ユニット(CPU)と、

前記外部メモリ及び前記CPUの間を結合するバスと、を更に備える請求項14のシステム。

【請求項19】

前記第1のGPUは前記第1の内部メモリを前記CPUの前記キャッシュメモリに結合するように構成される他のインタフェースを更に備える請求項18のシステム。

【請求項20】

前記第2のGPUは前記第2の内部メモリを前記CPUの前記キャッシュメモリに結合するように構成される他のインタフェースを更に備える請求項18のシステム。

【発明の詳細な説明】

【技術分野】

【0001】

本発明は概してコンピューティングデバイス(例えばコンピュータ、組み込みデバイス、携帯デバイス等)に向けられている。より特定的には、本発明はそのようなコンピューティングデバイスの処理ユニットによって使用されるメモリに向けられている。

【背景技術】

【0002】

コンピューティングデバイスは、典型的には、中央処理ユニット(CPU)及びグラフィクス処理ユニット(GPU)等の1つ以上の処理ユニットを含む。CPUは、命令の正確なセットに従うことによって、コンピューティングデバイスの活動を連携させる。GPUは、エンドユーザアプリケーション(例えばビデオゲームアプリケーション)によって要求されるであろうグラフィクス処理タスク及び/又は物理的シミュレーション等のデータ並列コンピューティングタスクを行うことによってCPUを支援する。GPU及びCPUは、別個のデバイス及び/又はパッケージの一部であることがあり、あるいは同じデバイス及び/又はパッケージ内に含まれていることがある。更には、各処理ユニットは別のより大きなデバイス内に含まれていることがある。例えばGPUは、経路付けデバイス又は例えばノースブリッジ等のブリッジデバイス内にしばしば集積化される。

【0003】

エンドユーザアプリケーションとGPUの間には何層ものソフトウェアが存在する。エンドユーザアプリケーションは、アプリケーションプログラミングインタフェース(API)と通信する。APIは、GPUに依存するフォーマットでよりはむしろ標準的なフォ

10

20

30

40

50

ーマットでエンドユーザアプリケーションがグラフィクスデータ及びコマンドを出力することを可能にする。ワシントン、レッドモンドのマイクロソフト社(Microsoft Corporation of Redmond, Washington)によって開発されたダイレクトX(DirectX)(登録商標)、クロノスグループ(Khronos Group)によって維持されるオープンGL(OpenGL)(登録商標)及びオープンCL(OpenCL)を含めて様々な種類のAPIが商業的に利用可能である。APIはドライバと通信する。ドライバは、APIから受信した標準コードを、GPUによって理解されるネイティブフォーマットの命令にトランスレートする。ドライバは典型的にはGPUの製造業者によって書かれる。GPUは次いでドライバからの命令を実行する。

【発明の概要】

【発明が解決しようとする課題】

10

【0004】

従来のシステムにおいては、CPU及びGPUは典型的には各々が外部メモリに結合される。外部メモリは、CPU及び/又はGPUによって実行されるべき命令及び/又は使用されるべきデータを含むであろう。外部メモリは例えばダイナミックランダムアクセスメモリ(DRAM)であることがある。外部メモリはかなり大きく構成され得るので、それが結合される各処理ユニットに十分な記憶容量を提供することができる。残念なことに、外部メモリへのアクセスには数百クロックサイクルを必要とするであろう。従って、外部メモリは、高性能なGPUに対する十分な帯域幅又は高速メモリアクセスをメモリに提供しないかもしれない。

【0005】

20

十分なメモリ帯域幅をGPUに提供するための1つの見込みのある解決法は、GPUに内部メモリを設けることである。内部メモリは、例えば、埋め込み(embedded)DRAM又は積層(stacked)DRAMであることがある。外部メモリと比べて、内部メモリは、より大きな帯域幅、より高速なメモリアクセスを提供し、且つより少ない電力を消費する。しかし、内部メモリの容量は、高性能なGPUの記憶要求を満たすために容易には拡大され得ない。例えば、高性能なGPUは、GPUの内部メモリに含まれ得るよりも多くのメモリを必要とするであろう。

【0006】

上述に鑑み、十分なメモリ容量(外部メモリと同等な)及び大きな帯域幅(埋め込みメモリと同等な)の両方を提供するメモリ及びそのアプリケーションが必要とされている。

30

【課題を解決するための手段】

【0007】

本発明の実施形態は、汎用使用のための処理ユニット内部メモリ(an internal, processing-unit memory)及びそのアプリケーションを提供することによって、上述の必要性を満たす。本発明の実施形態の処理ユニット内部メモリは、それが処理ユニット内に埋め込まれているという理由で、大きな帯域幅を提供する。また、複数の処理ユニットメモリは十分に大きなメモリプールへと組み合わせられ得るので、十分な記憶容量が提供される。

【0008】

例えば、本発明の実施形態はGPUを提供する。GPUは、第1の内部メモリと、第1の内部メモリに結合される実行ユニットと、第1の内部メモリを他の処理ユニットの第2の内部メモリに結合するように構成されるインタフェースと、を含む。ある実施形態においては、GPUはソフトウェアにおいて具現化される。別の実施形態においては、GPUはシステム内に含まれる。システムは、例えば、スーパーコンピュータ、デスクトップコンピュータ、ラップトップコンピュータ、ビデオゲームコンソール、埋め込みデバイス、携帯デバイス(例えば携帯電話、スマートフォン、MP3プレイヤー、カメラ、GPSデバイス等)、又はGPUを含み若しくはGPUを含むように構成される別のシステムを備えていてよい。

40

【0009】

本発明の更なる特徴及び利点の他、本発明の種々の実施形態の構成及び動作は、添付の図面を参照して以下に詳細に説明される。尚、本発明はここに説明される特定の実施形態

50

に限定されない。そのような実施形態は例示の目的のみのためにここに提示されている。追加的な実施形態はここに含まれる教示に基き関連分野を含めた当業者にとって明らかであろう。

【図面の簡単な説明】

【0010】

ここに組み込まれ且つ出願書類の一部をなす添付の図面は本発明を示し、そして明細書と共に、本発明の原理を説明すること及び関連分野を含めた当業者が本発明を作りそして使用するのを可能にすることに更に役立つ。

【0011】

【図1A】図1Aは本発明の実施形態に従い汎用使用のための処理ユニット内部メモリを含む例示的なシステムを示す図（その1）である。

10

【図1B】図1Bは本発明の実施形態に従い汎用使用のための処理ユニット内部メモリを含む例示的なシステムを示す図（その2）である。

【0012】

【図2】図2は本発明の実施形態に従い汎用使用のための内部メモリを有する例示的なGPUの詳細を示す図である。

【0013】

【図3】図3は本発明の実施形態に従い処理要素内に含まれていてよい例示的な積層メモリを示す図である。

【0014】

20

【図4】図4は本発明の実施形態に従い図2のGPUによって実装される例示的な方法を示す図である。

【0015】

本発明の特徴及び利益は、図面と共に以下に記述される詳細な説明からより明らかになり、図面において同様の参照符号は全体を通して対応する要素を識別する。図面において、同様の参照数字は一般的に同一の、機能的に類似の、及び/又は構造的に類似の要素を示す。ある要素が最初に現れる図面は対応する参照番号の一番左の単一又は複数の桁によって示される。

【発明を実施するための形態】

【0016】

30

I. 概説

本発明は、汎用使用のためのGPU内部メモリ及びそのアプリケーションを提供する。以下の詳細な説明において、「1つの実施形態」、「ある実施形態」、「例示的实施形態」等に対する言及は、説明される実施形態が特定の特徴、構造又は特性を含んでいてよいが、全ての実施形態が必ずしも当該特定の特徴、構造又は特性を含む必要がなくてよいことを示している。また、そのような表現は必ずしも同じ実施形態を参照しているとは限らない。更に、特定の特徴、構造又は特性がある実施形態に関連して説明されている場合には、明示的に説明されていようとなかろうと、他の実施形態に関連して当該特定の特徴、構造又は特性を具現化することは当業者の知識の範囲内にあることと言える。

【0017】

40

ある実施形態によると、GPUは、1つ以上の他の処理ユニットによって使用されるように構成される内部メモリ（例えば埋め込みDRAM又は積層DRAM）を含む。GPUはインタフェースを含むと共にプロトコルを実装しており、1つ以上の他のGPUが当該内部メモリにアクセスすることを可能にしている。インタフェースは、各他のGPUに当該内部メモリへの専用のアクセスを提供してよく、あるいは他のGPUに当該内部メモリへの共有アクセスを提供してよい。GPUの内部メモリへのアクセスは、GPUそれ自身又は各他のGPUによって制御されてよい。

【0018】

ある実施形態においては、インタフェース及びプロトコルは、内部メモリが外部メモリと組み合わせられることを可能にし、GPUがアクセス可能なより大きなメモリプールを形

50

成する。外部メモリは他のGPU内に含まれていてよい。ある実施形態においては、例えば、コンピューティングデバイスは複数のGPUを含み、ここでは各GPUは他のGPUと共有されるように構成される内部メモリを含む。この実施形態においては、各GPUの内部メモリは統合化メモリプール(unified memory pool)へと組み合わせられる。メモリプールのサイズは、共有しているGPUの数に対応する。共有しているGPUはいずれもがその記憶要求に対してメモリプールを使用してよい。

【0019】

本発明の実施形態に従う例示的なGPUの更なる詳細が以下に説明される。しかし、これらの詳細を提供するのに先立ちそのようなGPUが実装されるであろう例示的なコンピューティングデバイスを説明することは有用である。

【0020】

II. 例示的なコンピューティングシステム

図1A及び1Bは複数のGPUを有する例示的なコンピューティングシステム100を示しており、各GPUは、本発明の実施形態に従い汎用使用のために構成される内部メモリを含んでいる。外部メモリと比べて、内部メモリは、データへのより大きな帯域幅アクセスを各GPUに提供する。また、各GPUの内部メモリは、各GPUによってアクセス可能な更に大きなメモリプールへと組み合わせられてよく、それにより各GPUに十分な記憶容量を提供することができる。

【0021】

図1Aの実施形態においては、各GPUは他のGPUの内部メモリへの専用のアクセスを与えられている。図1Bの実施形態においては、各GPUは他のGPUの内部メモリへの共有インタフェースを介した共有アクセスを有している。実施形態においては、コンピューティングシステム100は、スーパーコンピュータ、デスクトップコンピュータ、ラップトップコンピュータ、ビデオゲームコンソール、埋め込みデバイス、携帯デバイス(例えば携帯電話、スマートフォン、MP3プレイヤー、カメラ、GPSデバイス等)、又はCPU及び/若しくはGPUを含み若しくはCPU及び/若しくはGPUを含むように構成される何らかの他のデバイスを備えていてよい。

【0022】

図1A及び図1Bを参照すると、コンピューティングデバイス100は、CPU102、第1のGPU110A、及び第2のGPU110Bを含む。CPU102は命令を実行してコンピューティングデバイス100の機能を制御する。GPU110は、データ並列処理タスク(例えばグラフィクス処理タスク及び/又は一般計算タスク)を行うことによってCPU102を支援する。GPU110は典型的には、それらの設計に基づいて、データ並列処理タスクを、CPU102がソフトウェアにおいてそれらを行い得るであろうよりも高速に行うことができる。

【0023】

第1のGPU110A及び第2のGPU110Bは、それら自身の内部メモリ及び実行ユニットを各々が含む。具体的には、第1のGPU110Aは内部メモリ106A及び実行ユニット142Aを含み、また第2のGPU110Bは内部メモリ106B及び実行ユニット142Bを含む。同様にCPU102はキャッシュメモリ130及び実行ユニット132を含む。内部メモリ106(及び随意的にキャッシュメモリ130)は、特定のデータが外部的に記憶されていたとした場合(例えばそのデータがシステムメモリ104内に記憶されていたとした場合)に可能であろうよりも高速なそのデータへのアクセス及び大きな帯域幅を提供するために、GPU110が利用可能である。内部メモリ106は、例えば埋め込みDRAM又は積層DRAMを備えていてよい。

【0024】

内部メモリ106A、106B(及び随意的にキャッシュメモリ130)は、高速で大きな帯域幅のメモリアccessを提供しながら相当の記憶容量(例えば4GBより大きい)を提供するために、より大きなメモリプールへと組み合わせられてよい。従来の外部メモリは十分な記憶容量(例えば4GBより大きい)を提供し得るが、従来の外部メモリは特定

10

20

30

40

50

の高性能な用途に対して不十分な帯域幅をもたらす。同様に、従来の埋め込みメモリはこれらの高性能な用途に対して十分な帯域幅を提供し得るが、従来の埋め込みメモリはこれらの高性能な用途に対して不十分な記憶容量（例えば４ＧＢ未満）をもたらす。従来の外部メモリ及び／又は従来の埋め込みメモリとは異なり、本発明の実施形態は、汎用使用のために他のＧＰＵが利用可能な内部メモリを含むＧＰＵを提供することによって、十分な記憶容量（例えば４ＧＢより大きい）を提供するだけでなく、大きな帯域幅をも提供する。

【００２５】

例えば、高性能なＧＰＵのフレームバッファ（即ちディスプレイデバイス上に表示されるべきデータの完全なフレームを記憶するバッファ）は、相当大きなメモリ（例えば４ギガバイト（ＧＢ）より大きい）への大きな帯域幅のアクセスを必要とするであろう。実施形態においては、第１のＧＰＵ１１０Ａは、内部メモリ１０６Ａ、Ｂ及び随意的にＣＰＵ１０２のキャッシュメモリ１３０を使用して第１のＧＰＵ１１０Ａのフレームバッファを定義してよい。同様に、第２のＧＰＵ１１０Ｂもまた、内部メモリ１０６Ａ、Ｂ及び随意的にＣＰＵ１０２のキャッシュメモリ１３０を使用して第２のＧＰＵ１１０Ｂのフレームバッファを定義してよい。このようにして、従来の外部メモリ又は埋め込みメモリとは異なり、本発明の実施形態に従い定義されるフレームバッファは、相当大きなメモリ（例えば４ＧＢより大きい）への大きな帯域幅のアクセスを提供する。

【００２６】

図１Ａの実施形態においては、既に示唆されたように、各ＧＰＵ１１０は、他の処理ユニットの内部メモリ１０６への専用のアクセスを与えられている。具体的には、第１のインタフェース１０１は、第２のＧＰＵ１１０Ｂの内部メモリ１０６Ｂへの専用アクセスを第１のＧＰＵ１１０Ａに提供し、また第１のＧＰＵ１１０Ａの内部メモリ１０６Ａへの専用アクセスを第２のＧＰＵ１１０Ｂに提供する。データは、そのデータのアドレス範囲に基づいて、内部メモリ１０６Ａ又は内部メモリ１０６Ｂのいずれかに対して書き込まれ又はいずれかからリトリブされる(retrieved)。例えば、内部メモリ１０６Ａは第１のアドレス範囲（例えば第１の予め定められたアドレスＡ未満且つ第２の予め定められたアドレスＢ以上）を割り当てられてよく、また内部メモリ１０６Ｂは第２のアドレス範囲（例えば第１のアドレス範囲内でない全てのアドレス）を与えられてよい。しかし、第１のＧＰＵ１１０Ａ及び第２のＧＰＵ１１０Ｂが各々第１のＧＰＵ１１０Ａの内部メモリ１０６Ａ及び第２のＧＰＵ１１０Ｂの内部メモリ１０６Ｂへのアクセスを有し得ることを前提として、内部メモリ１０６Ａ及び／又は内部メモリ１０６Ｂへデータを書き込み且つこれらからデータをリトリブするための他のスキームが、本発明の精神及び範囲から逸脱することなく実装されてよいことが理解されるはずである。

【００２７】

ある実施形態においては、第１のインタフェース１０１はディスプレイ制御器インタフェースを備えている。ディスプレイ制御器インタフェースは、ディスプレイデバイス１４０にＧＰＵのフレームバッファへのアクセスを提供する。ディスプレイ制御器インタフェースを第１のインタフェース１０１内に組み込むことによって、第１のインタフェース１０１は、従来のＧＰＵ設計に既に含まれている標準ピン上に設けられ得る。

【００２８】

第１のインタフェース１０１に加えて、第２のインタフェース１０３は、第２のＧＰＵ１１０Ｂの内部メモリ１０６Ｂへの専用アクセスをＣＰＵ１０２に提供し、またＣＰＵ１０２のキャッシュメモリ１３０への専用アクセスを第２のＧＰＵ１１０Ｂに提供する。このようにして第２のＧＰＵ１１０Ｂ及びＣＰＵ１０２は、各々第２のＧＰＵ１１０Ｂの内部メモリ１０６Ｂ及びＣＰＵ１０２のキャッシュメモリ１３０へのアクセスを有することができる。同様に、第３のインタフェース１０５は、ＣＰＵ１０２のキャッシュメモリ１３０への専用アクセスを第１のＧＰＵ１１０Ａに提供し、また第１のＧＰＵ１１０Ａの内部メモリ１０６Ａへの専用アクセスをＣＰＵ１０２に提供する。このようにして第１のＧＰＵ１１０Ａ及びＣＰＵ１０２は、各々第１のＧＰＵ１１０Ａの内部メモリ１０６Ａ及び

CPU 102 のキャッシュメモリ 130 へのアクセスを有することができる。

【0029】

図 1 B の実施形態においては、各処理ユニットは、他の処理ユニットの内部メモリへの共有インタフェース 164 を介した共有アクセスを有する。共有インタフェース 164 は、各処理ユニット（例えば第 1 の GPU 110 A、第 2 の GPU 110 B 及び CPU 102）に他の処理ユニットの内部メモリへの大きな帯域幅のアクセスを提供する。データは、そのデータのアドレス範囲に基づいて、内部メモリ 106 A、内部メモリ 106 B 又はキャッシュメモリ 130 に対して書き込まれ又はいずれかからリトリブされる。例えば、内部メモリ 106 A は第 1 のアドレス範囲を割り当てられてよく、内部メモリ 106 B は第 2 のアドレス範囲を与えられてよく、そしてキャッシュメモリ 130 は第 3 のアドレス範囲を割り当てられてよい。しかし、第 1 の GPU 110 A、第 2 の GPU 110 B 及び CPU 102 が各々第 1 の GPU 110 A の内部メモリ 106 A、第 2 の GPU 110 B の内部メモリ 106 B、及び CPU 102 のキャッシュメモリ 130 へのアクセスを有し得ることを前提として、内部メモリ 106 A、内部メモリ 106 B 及び / 又はキャッシュメモリ 130 へデータを書き込み且つこれらからデータをリトリブするための他のスキームが、本発明の精神及び範囲から逸脱することなく実装されてよいことが理解されるはずである。

【0030】

実施形態においては、コンピューティングデバイス 100 はまた、システムメモリ 104、補助メモリ 120、入力・出力 (I/O) インタフェース 116、及び / 又はディスプレイデバイス 140 を含む。システムメモリ 104 は、CPU 102 上で動作中のプログラムによって頻繁にアクセスされる情報を記憶する。システムメモリ 104 は典型的には揮発性メモリを備えており、これはコンピューティングデバイス 100 への電力がオフになったときにシステムメモリ 104 内に記憶されているデータが喪失することを意味している。補助メモリ 120 は、コンピューティングデバイス 100 によって使用されるデータ及び / 又はアプリケーションを記憶する。補助メモリ 120 は、典型的にはシステムメモリ 104 に比べて大きな記憶容量を有しており、また典型的には不揮発性（永続的）メモリを備えており、これはコンピューティングデバイス 100 への電力がオフになったとしても補助メモリ 120 内に記憶されているデータが持続することを意味している。I/O インタフェース 116 は、コンピューティングデバイスシステム 100 が外部デバイス 116（例えば外部ディスプレイデバイス、外部記憶デバイス（例えばビデオゲームカートリッジ、CD、DVD、フラッシュドライブ等）、ネットワークカード、又は何らかの他の種類の外部デバイス）と結合されることを可能にする。ディスプレイデバイス 140 はコンピューティングデバイス 100 のコンテンツを表示する。ディスプレイデバイスは、陰極線管、液晶ディスプレイ (LCD)、プラズマスクリーン、又は現在既知であるか将来開発されるかにかかわらず何らかの他の種類のディスプレイデバイスを備えていてよい。

【0031】

GPU 110 及び CPU 102 は、互いに、そしてシステムメモリ 104、補助メモリ 120 及び I/O インタフェース 116 とバス 114 を介して通信する。バス 114 は、周辺要素インタフェース (peripheral component interface) (PCI) バス、アクセラレーテッドグラフィクスポート (accelerated graphics port) (AGP) バス、PCI エクスプレス (PCI Express) (PCI E) バス、又は現在利用可能であり若しくは将来開発される他の種類のバスを含めてコンピューティングデバイスにおいて用いられる任意の種類のバスであってよい。

【0032】

実施形態においては、コンピューティングデバイス 100 は、GPU 110 の代わりに又は GPU 110 に加えてビデオ処理ユニット (VPU) を含んでいてよい。例えばある実施形態においては、コンピューティングデバイス 100 は GPU 110 A、CPU 102 を含み、そして図 1 A 及び 1 B に示される GPU 110 B に代えて、コンピューティン

グデバイス 100 は VPU を含む。このようにして、CPU 102 は一般的な処理機能を行うことができ、GPU 110A はグラフィクス処理機能を行うことができ、そして VPU はビデオ処理機能を行うことができる。

【0033】

III. 例示的な GPU

図 2 は内部メモリ 106 を有する GPU 110 の例示的な詳細を示している。本発明の実施形態に従い、増大されたメモリフットプリント (footprint) サイズに基づいてグラフィクス処理能力を結合することによって全体的なシステム性能を増大するために、内部メモリ 106 は他の GPU 又は CPU により使用され得る。

【0034】

上述したように、GPU 110 は実行ユニット 142 及び内部メモリ 106 を含む。図 2 を参照すると、実行ユニット 142 は入力論理 202、シェーダコア 204 及び出力論理 206 を含む。内部メモリ 106 はメモリ制御器 210 及びメモリセル 212 を含む。メモリ制御器 210 はメモリセル 212 へのアクセスを制御する。メモリセル 212 はデータを記憶する。

【0035】

ある実施形態においては、内部メモリ 106 は埋め込みダイナミックランダムアクセスメモリ (DRAM) を備える。埋め込み DRAM は、処理ユニットと共に共通のパッケージ内に密閉された (encapsulated) メモリである。別の実施形態においては、内部メモリ 106 は図 3 に示されるような積層 DRAM を備える。積層メモリは、互いの上に 3 次元構造で積層される複数のメモリ要素を含む。

【0036】

内部メモリ 106 は、入力論理 202 及び出力論理 206 の両方を介して実行ユニット 142 に結合される。特に、入力論理 202 は内部メモリ 106 からデータをリトリブすることができ、また出力論理 206 は、データをメモリセル 212 内に記憶されるように内部メモリ 106 へ送ることができる。

【0037】

内部メモリ 106 はまた、第 1 のインタフェース 250 を介して他の GPU の内部メモリに結合されていてよい。内部メモリ 106 を他の GPU の内部メモリに結合することは、実行ユニット 142 が利用可能な総メモリプールを増大させることができる。ある実施形態においては、第 1 のインタフェース 250 は、図 1A のインタフェース 101 によって示されるように、GPU 110 の内部メモリ 106 と他の GPU の内部メモリとの間に専用のアクセスを提供する。この実施形態においては、第 1 のインタフェース 250 は従来の GPU の標準ピン上に設けられたものである。例えば第 1 のインタフェース 250 はディスプレイ制御器インタフェースを備えていてよく、ディスプレイ制御器インタフェースは、内部メモリ 106 に含まれるローカルフレームバッファへのディスプレイデバイスアクセスを提供する。別の実施形態においては、第 1 のインタフェース 250 は、図 1B のインタフェース 164 によって示されるように、GPU 110 の内部メモリ 106 と他の処理ユニットの内部メモリとの間で共有されるアクセスを提供する。

【0038】

内部メモリ 106 はまた、第 2 のインタフェース 260 を介して CPU 102 のキャッシュメモリ 130 に結合されていてよい。その結果、内部メモリ 106 とキャッシュメモリ 130 の組み合わせは、GPU 110 が利用可能なメモリプールを増大させることができる。ある実施形態においては、第 2 のインタフェース 260 は、図 1A の接続 103 又は接続 105 のように、GPU 110 の内部メモリ 106 と CPU 102 のキャッシュメモリ 130 との間での専用接続を提供する。別の実施形態においては、第 2 のインタフェース 260 は、図 1B の接続 164 のように、GPU 110 及び CPU 102 のみによって共有される接続を提供する。更なる実施形態においては、第 2 のインタフェースは、図 1A 及び図 1B のバス 114 のように、GPU 110 を共通のバス上で CPU 102 と結合する。

10

20

30

40

50

【 0 0 3 9 】

IV . G P U 1 1 0 の例示的な動作

図 4 は本発明の実施形態に従い G P U 1 1 0 によって実装される例示的な方法 4 0 0 を示している。方法 4 0 0 は図 2 及び 4 を参照して以下に説明される。

【 0 0 4 0 】

方法 4 0 0 はステップ 4 0 2 で開始し、命令が受信される。ある実施形態においては、G P U 1 1 0 によって実行されるべき命令を入力論理 2 0 2 が受信する。命令は、例えば、システム 1 0 0 の C P U 1 0 2 上で実行中のエンドユーザアプリケーションによって提供されるグラフィクス処理タスク又はデータ並列処理タスクを備えていてよい。

【 0 0 4 1 】

ステップ 4 0 4 では、命令に関連するロケーションが識別される。1 つの例においては、データは受信された命令と共に含まれていてよい。そのようなデータは一般的に即時データと称される。別の例においては、命令はデータのロケーションを提供する。例えば、命令はデータが記憶されているアドレスを含んでいてよい。更なる例においては、命令は、データが記憶されているアドレスを入力論理 2 0 2 が計算するための情報を含む。データは、内部メモリ 1 0 6、内部メモリ 1 0 6 が結合される他の G P U の内部メモリ、又は C P U 1 0 2 のキャッシュメモリ 1 3 0 のいずれかに記憶されてよい。

【 0 0 4 2 】

ステップ 4 0 6 では、データがリトリブされる。データが即時データである場合には、入力論理 2 0 2 は命令から単純に即時データを抽出する。データが内部メモリ 1 0 6 内又は内部メモリ 1 0 6 が結合されるメモリ内に記憶されている場合には、入力論理 2 0 2 はデータにアクセスするための要求をメモリ制御器 2 1 0 へ送る。一方において、データがメモリセル 2 1 2 内に記憶されている場合には、データはリトリブされて入力論理 2 0 2 へ供給される。他方、内部メモリ 1 0 6 に結合される他のメモリ内にデータが記憶されている場合には、入力論理 2 0 2 からの要求が他のメモリにインタフェース 2 5 0 又はインタフェース 2 6 0 を介して転送される。データは次いで、他のメモリからリトリブされて入力論理 2 0 2 へ供給される。

【 0 0 4 3 】

ステップ 4 0 8 では、命令が実行される。シェーダコア 2 0 4 は、ステップ 4 0 6 で入力論理 2 0 2 によって獲得されたデータに基づいて命令を実行する。

【 0 0 4 4 】

ステップ 4 1 0 では、命令実行の結果が出力論理 2 0 6 へ提供される。出力論理 2 0 6 は、判断ステップ 4 1 2 に示されるように、これらの結果に基づいて更なる処理が必要であるかどうかを決定する。出力論理 2 0 6 に提供される結果は、追加の処理が必要であるかどうかを示すために、フラグ又は他のなんらかの印を有していてよい。判断ステップ 4 1 2 において更なる処理が必要であることを出力論理 2 0 6 が決定した場合、出力論理 2 0 6 は結果をシェーダコア 2 0 4 へ転送し戻し、そして方法 4 0 0 のステップ 4 0 8 及び 4 1 0 が繰り返される。一方、判断ステップ 4 1 2 において更なる処理が必要でないと出力論理 2 0 6 が決定した場合には、出力論理 2 0 6 は、ステップ 4 1 4 に示されるように結果を内部メモリ 1 0 6 へ提供する。

【 0 0 4 5 】

結果は次いで、結果が書き込まれるべきアドレスに応じて、内部メモリ 1 0 6 又は内部メモリ 1 0 6 に結合されるメモリへ書き込まれてよい。結果が内部メモリ 1 0 6 へ書き込まれるべきである場合には、メモリ制御器 2 1 0 はメモリセル 2 1 2 内の適切なアドレスへのアクセスを提供し、そして結果はそこに記憶される。一方、内部メモリに結合されるメモリへ結果が書き込まれるべきである場合には、メモリ制御器 2 1 0 はインタフェース 2 5 0 又はインタフェース 2 6 0 を介して結果を他のメモリへ転送し、そして結果は他のメモリのメモリセル内に記憶される。

【 0 0 4 6 】

V . 例示的なソフトウェア実装

10

20

30

40

50

G P U 1 1 0 のハードウェア実装に加えて、そのような G P U はまた、例えばソフトウェア（例えばコンピュータ可読プログラムコード）を記憶するように構成されるコンピュータ可読媒体内に配置されるソフトウェアにおいて具現化されてもよい。コンピュータ可読プログラムコードは、（i）ここに開示されるシステムの機能及び技術（例えば G P U 1 1 0 にタスクを提供すること、G P U 1 1 0 内でタスクをスケジューリングすること、G P U 1 1 0 内でタスクを実行すること、等）、（ii）ここに開示されるシステムの製造及び技術（例えば G P U 1 1 0 の製造）、又は（iii）ここに開示されるシステムの機能及び製造並びに技術の組み合わせ、の実施形態を含めて本発明の実施形態を可能にする。

【 0 0 4 7 】

このことは、例えば、一般的なプログラミング言語（例えば C 又は C + + ）、ベリログ (Verilog) H D L 、 V H D L 、 アルテラ (Altera) H D L (A H D L) 等を含むハードウェア記述言語 (hardware description languages) (H D L) 、あるいは他の利用可能なプログラミング及び / 又は回路図等 (schematic) キャプチャツール (capture tools) (例えば回路キャプチャツール) の使用を通して達成され得る。コンピュータ可読プログラムコードは、半導体、磁気ディスク、光学ディスク（例えば C D - R O M 、 D V D - R O M ）を含む任意の既知のコンピュータ可読媒体内に配置され得る。従って、コンピュータ可読コードは、インターネット及びそれと同等のもの (the Internet and internets) を含む通信ネットワークを介して伝送され得る。上述したシステム及び技術によって達成される機能及び / 又は提供される構造は、コンピュータ可読プログラムコードにおいて具現化されるコア（例えばシェーダコア）内で表現することができ、また集積回路の生産の一部としてハードウェアに変換されてよいことが理解される。

【 0 0 4 8 】

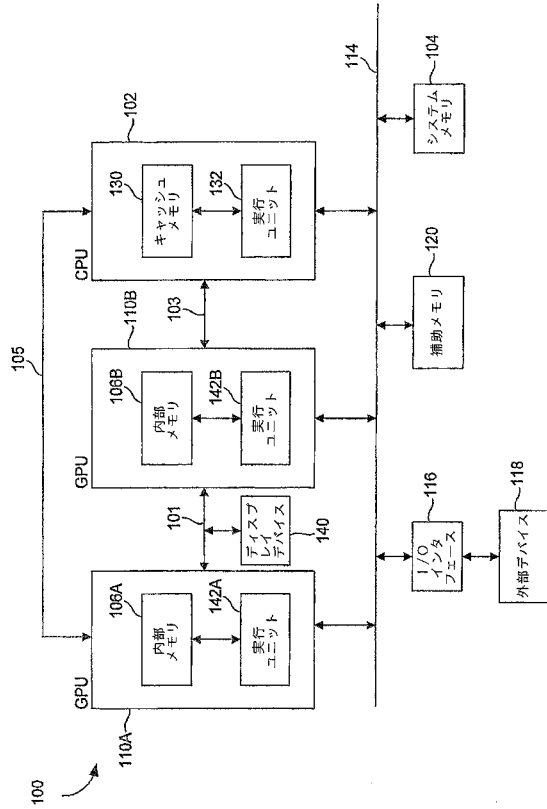
VI . 結論

汎用使用のための G P U 内部メモリ及びそのアプリケーションが上に説明される。概要及び要約の欄ではなく詳細な説明の欄が特許請求の範囲を解釈するために用いられることを意図されていることが理解されるべきである。概要及び要約の欄は、発明者によって検討されているような本発明の 1 つ以上であるが全てではない例示的な実施形態を記述することができ、従って、本発明及び添付の特許請求の範囲を限定することを意図されるものではない。

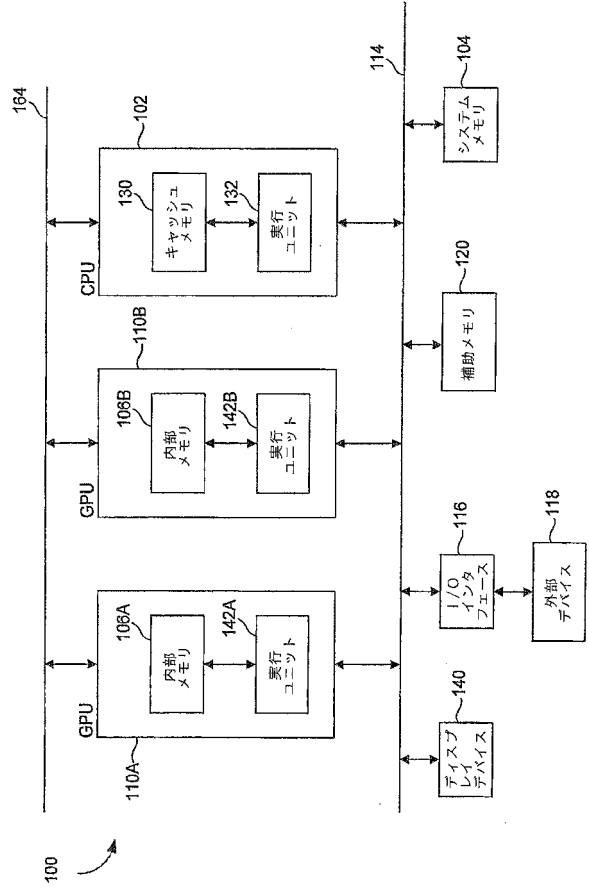
10

20

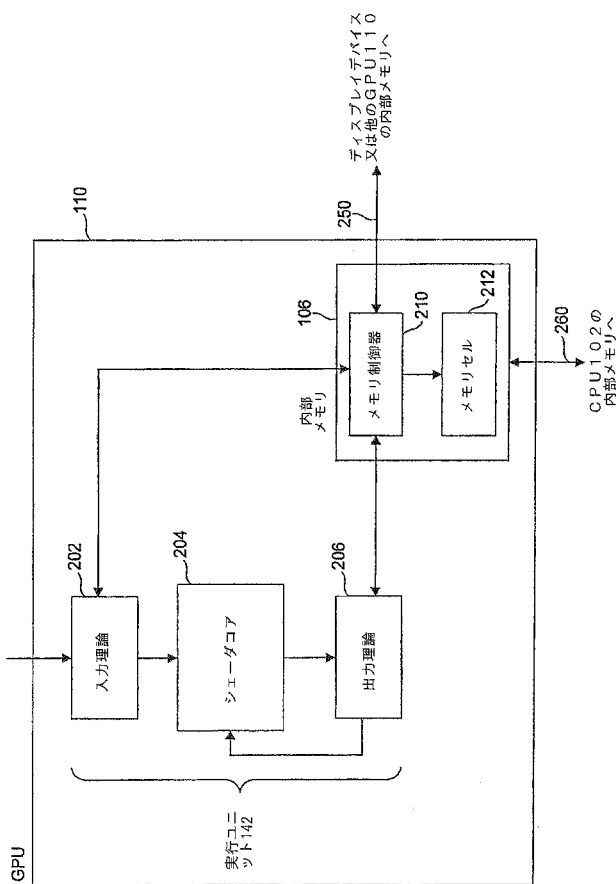
【図 1 A】



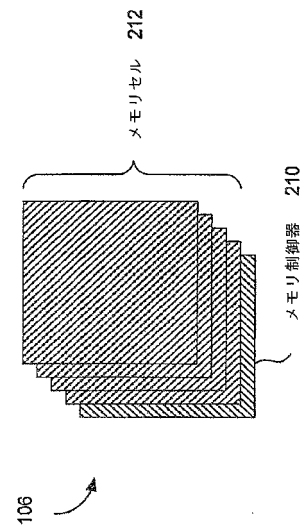
【図 1 B】



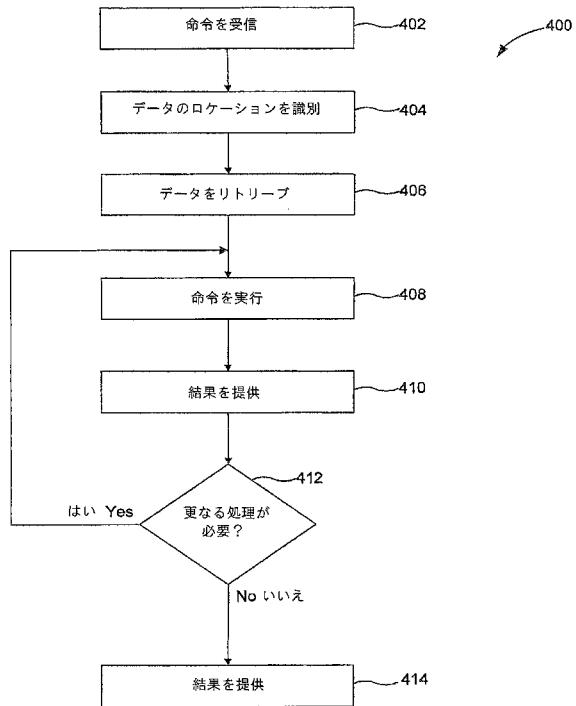
【図 2】



【図 3】



【図 4】



【 国際調査報告 】

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2010/047784

A. CLASSIFICATION OF SUBJECT MATTER INV. G06F9/38 G06F9/46 G06F9/50 ADD.		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED Minimum documentation searched (classification system followed by classification symbols) G06F		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched		
Electronic data base consulted during the International search (name of data base and, where practical, search terms used) EPO-Internal, INSPEC, IBM-TDB, WPI Data		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 2007/294696 A1 (PAPAKIPOS MATTHEW N [US] ET AL) 20 December 2007 (2007-12-20) paragraphs [0031], [0091], [0153], [0383], [0386], [0389], [0438]; figures 7D,10	1-20
Y	EP 1 557 755 A1 (THOMSON LICENSING SA [FR]) 27 July 2005 (2005-07-27) paragraphs [0006], [0034] - [0043]	1-20
A	US 2007/074221 A1 (STENSON RICHARD B [US] ET AL) 29 March 2007 (2007-03-29) paragraphs [0006], [0008], [0022], [0024], [0030]	1-20
A	EP 0 442 041 A2 (NAT SEMICONDUCTOR CORP [US]) 21 August 1991 (1991-08-21) page 5, line 10 - line 24; figure 3	1-20
-/-		
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C. <input checked="" type="checkbox"/> See patent family annex.		
* Special categories of cited documents : *A* document defining the general state of the art which is not considered to be of particular relevance *E* earlier document but published on or after the international filing date *L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) *O* document referring to an oral disclosure, use, exhibition or other means *P* document published prior to the international filing date but later than the priority date claimed *T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention *X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone *Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art *Z* document member of the same patent family		
Date of the actual completion of the international search 15 November 2010		Date of mailing of the international search report 29/11/2010
Name and mailing address of the ISA/ European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016		Authorized officer Thibaudeau, Jean

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2010/047784

C(Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	US 2004/160449 A1 (GOSSALIA ANUJ B [US] ET AL) 19 August 2004 (2004-08-19) paragraph [0040]; figure 2	6, 17

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No

PCT/US2010/047784

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
US 2007294696	A1	20-12-2007	NONE	
EP 1557755	A1	27-07-2005	CN 1645351 A FR 2865291 A1 JP 2005209206 A KR 20050076702 A MX PA05000788 A US 2005172104 A1	27-07-2005 22-07-2005 04-08-2005 26-07-2005 29-08-2005 04-08-2005
US 2007074221	A1	29-03-2007	EP 1934738 A1 JP 2009510612 T US 2009147013 A1 US 2010251245 A1 WO 2007038456 A1	25-06-2008 12-03-2009 11-06-2009 30-09-2010 05-04-2007
EP 0442041	A2	21-08-1991	JP 3227177 A US RE40942 E1	08-10-1991 20-10-2009
US 2004160449	A1	19-08-2004	US 2005168472 A1	04-08-2005

フロントページの続き

(51) Int.Cl.		F I		テーマコード (参考)
G 0 6 F 12/08 (2006.01)		G 0 6 F 12/08	5 5 1 C	
		G 0 6 F 12/08	5 0 1 B	

(81) 指定国 AP(BW, GH, GM, KE, LR, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), EP(AL, AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MK, MT, NL, NO, PL, PT, RO, SE, SI, SK, SM, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA, CH, CL, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE, EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID, IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK, LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW, MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PE, PG, PH, PL, PT, RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TH, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM, ZW

(74) 代理人 100162156

弁理士 村雨 圭介

(72) 発明者 グレグ サドウスキー

アメリカ合衆国 0 2 1 3 9 マサチューセッツ州、ケンブリッジ、ハーバード ストリート
3 2 1、# 3 0 3

(72) 発明者 コンスタンチン アイオールチャ

アメリカ合衆国 9 5 1 2 0 カリフォルニア州、サン ノゼ、ウッディード レイク ドライブ
7 1 8 6

(72) 発明者 ジョン ブラザーズ

アメリカ合衆国 9 4 0 8 5 カリフォルニア州、サニーベール、レークサイド ドライブ 1 2
5 7、アパートメント 1 2 2 6

F ターム (参考) 5B005 KK13 LL15 MM01 NN12

5B013 DD03

5B045 AA01 BB28 BB29 DD10 DD12

5B047 EA06 EA07 EA09 EB15

5B057 CH02 CH11 CH12 CH14 CH16 CH20