(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7400169号
(P7400169)

(45)発行日　令和5年12月19日(2023.12.19)

(24)登録日　令和5年12月11日(2023.12.11)

(51)国際特許分類		FI		
G06F	*9/50 (2006.01)*	G06F	9/50	150A
G06F	*9/38 (2018.01)*	G06F	9/38	310J
G06F	*9/48 (2006.01)*	G06F	9/38	370C
		G06F	9/38	370X
		G06F	9/48	300G

請求項の数　14　外国語出願　（全40頁）

(21)出願番号　特願2020-104328(P2020-104328)
(22)出願日　令和2年6月17日(2020.6.17)
(65)公開番号　特開2021-34020(P2021-34020A)
(43)公開日　令和3年3月1日(2021.3.1)
審査請求日　令和4年6月13日(2022.6.13)
(31)優先権主張番号　16/542,012
(32)優先日　令和1年8月15日(2019.8.15)
(33)優先権主張国・地域又は機関
米国(US)

(73)特許権者　591003943
インテル・コーポレーション
アメリカ合衆国 ９５０５４ カリフォルニア州・サンタクララ・ミッション カレッジ ブーレバード・２２００
(74)代理人　110000877
弁理士法人ＲＹＵＫＡ国際特許事務所
(72)発明者　マイケル ベハー
アメリカ合衆国 ９５０５４ カリフォルニア州・サンタクララ・ミッション カレッジ ブーレバード・２２００ インテル・コーポレーション内
(72)発明者　モシェ マオル
アメリカ合衆国 ９５０５４ カリフォルニア州・サンタクララ・ミッション カ

最終頁に続く

(54)【発明の名称】　ワークロードのスタティックマッピングの順不同にパイプライン化された実行を可能にする方法及び装置

(57)【特許請求の範囲】

【請求項１】

第１のローカルクレジットマネージャーを有する第１のコンピューティングユニットであって、前記第１のコンピューティングユニットは、第１のバッファにデータを書き込むタスクを処理する、前記第１のコンピューティングユニットと、

第２のローカルクレジットマネージャーを有する第２のコンピューティングユニットであって、前記第２のコンピューティングユニットは、第２のバッファからデータを読み出すタスクを処理する、前記第２のコンピューティングユニットと、

前記第１のコンピューティングユニットと前記第２のコンピューティングユニットに結合された少なくとも１つのファブリックと、

前記少なくとも１つのファブリックに結合された中央クレジットマネージャーとを備え、

前記中央クレジットマネージャーは、

前記第１のローカルクレジットマネージャーへの第１のクレジットの送信を生じさせることであって、前記第１のクレジットは、前記第１のバッファに格納される第２のデータを生成すべく前記第１のコンピューティングユニットにより処理される第１のデータに対応する、前記送信を生じさせることと、

前記第１のコンピューティングユニットの前記第１のローカルクレジットマネージャーから前記第１のクレジットにアクセスすることと、

前記第２のコンピューティングユニットに対するクレジットのカウントを減らすこととを行う

装置。
【請求項２】
前記中央クレジットマネージャーは、前記第１のコンピューティングユニットが前記第１のデータを処理することに応じて、前記第１のコンピューティングユニットの前記第１のローカルクレジットマネージャーから前記第１のクレジットにアクセスする
請求項１に記載の装置。
【請求項３】
前記中央クレジットマネージャーは、前記第２のバッファにおける前記第２のデータの利用可能性に応じて、前記第２のコンピューティングユニットに対する前記クレジットのカウントを減らす
請求項１又は２に記載の装置。
【請求項４】
前記第２のコンピューティングユニットに対する前記クレジットのカウントは、クレジットの第１のカウントであり、
前記中央クレジットマネージャーは、
前記第１のコンピューティングユニットに対するクレジットの第２のカウントを初期化し、
前記第２のコンピューティングユニットに対するクレジットの前記第１のカウントを初期化する
請求項１から３のいずれか一項に記載の装置。
【請求項５】
前記中央クレジットマネージャーは、前記第１のデータが前記第１のコンピューティングユニットに割り当てられるタスクに関連付けられることに基づいて、前記第１のローカルクレジットマネージャーへの前記第１のクレジットの送信を生じさせる
請求項１から４のいずれか一項に記載の装置。
【請求項６】
コンピューティングシステムにおける少なくとも一つのプロセッサによって命令が実行されたとき、前記コンピューティングシステムにおける中央クレジットマネージャーが、第１のコンピューティングユニットの第１のローカルクレジットマネージャーへ、第１のクレジットを送信する段階であって、前記第１のコンピューティングユニットは、第１のバッファにデータを書き込むタスクを処理し、前記第１のクレジットは、前記第１のバッファに格納される第２のデータを生成すべく前記第１のコンピューティングユニットにより処理される第１のデータに対応する、段階と、
前記少なくとも一つのプロセッサによって命令が実行されたとき、前記中央クレジットマネージャーが、前記第１のコンピューティングユニットの前記第１のローカルクレジットマネージャーから前記第１のクレジットにアクセスする段階と、
前記少なくとも一つのプロセッサによって命令が実行されたとき、前記中央クレジットマネージャーが、第２のローカルクレジットマネージャーを有する第２のコンピューティングユニットに対するクレジットのカウントを減らす段階であって、前記第２のコンピューティングユニットは、第２のバッファからデータを読み出すタスクを処理する、段階と、を備える
方法。
【請求項７】
前記第１のコンピューティングユニットが前記第１のデータを処理することに応じて、前記中央クレジットマネージャーが、前記第１のコンピューティングユニットの前記第１のローカルクレジットマネージャーから前記第１のクレジットにアクセスする段階を更に備える
請求項６に記載の方法。
【請求項８】
前記第２のバッファにおける前記第２のデータの利用可能性に応じて、前記中央クレジ

ットマネージャーが、前記第２のコンピューティングユニットに対する前記クレジットのカウントを減らす段階を更に備える
　請求項６又は７に記載の方法。
【請求項９】
　前記第２のコンピューティングユニットに対する前記クレジットのカウントは、クレジットの第１のカウントであり、
　前記方法は、
前記中央クレジットマネージャーが、前記第１のコンピューティングユニットに対するクレジットの第２のカウント、及び前記第２のコンピューティングユニットに対するクレジットの前記第１のカウントを初期化する段階を更に備える
　請求項６から８のいずれか一項に記載の方法。
【請求項１０】
　前記第１のデータが前記第１のコンピューティングユニットに割り当てられるタスクに関連付けられることに基づいて、前記中央クレジットマネージャーが、前記第１のローカルクレジットマネージャーへ、前記第１のクレジットを送信する段階を更に備える
　請求項６から９のいずれか一項に記載の方法。
【請求項１１】
　メモリと、複数の命令と、請求項６から１０のいずれか一項に記載の方法を実行するべく、前記複数の命令を実行する前記少なくとも一つのプロセッサとを備える装置。
【請求項１２】
　命令を備え、前記命令は、実行されると、請求項６から１０のいずれか一項に記載の方法を前記少なくとも一つのプロセッサに実行させる
　非一時的コンピュータ可読媒体。
【請求項１３】
　請求項６から１０のいずれか一項に記載の方法を実行するための手段を備える装置。
【請求項１４】
　請求項６から１０のいずれか一項に記載の方法を前記少なくとも一つのプロセッサに実行させるコンピュータプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
　この開示は概して、処理に関し、より詳細にはワークロードのスタティックマッピングの順不同にパイプライン化された実行を可能にする方法及び装置に関する。
【背景技術】
【０００２】
　コンピュータハードウェア製造者は、コンピュータプラットフォームの様々なコンポーネントに用いられるハードウェアコンポーネントを開発する。例えば、コンピュータハードウェア製造者は、マザーボード、マザーボード用のチップセット、中央処理ユニット（ＣＰＵ）、ハードディスクドライブ（ＨＤＤ）、ソリッドステートドライブ（ＳＳＤ）及び、他のコンピュータコンポーネントを開発する。更に、コンピュータハードウェア製造者は、アクセラレータとして知られる、ワークロードの処理を加速する処理要素を開発する。例えば、アクセラレータは、ＣＰＵ、グラフィック処理ユニット（ＧＰＵ）、ビジョン処理ユニット（ＶＰＵ）及び／又はフィールドプログラマブルゲートアレイ（ＦＰＧＡ）などであり得る。
【図面の簡単な説明】
【０００３】
【図１】異種システムのアクセラレータで実行されるワークロードを表したグラフの図である。
【０００４】
【図２】パイプライン及びバッファを実装した異種システムのアクセラレータで実行され

るワークロードを表したグラフの図である。
【０００５】
【図３】本開示の教示に従って構築された例示的なコンピューティングシステムを示すブロック図である。
【０００６】
【図４】例示的な１又は複数のスケジューラを含む例示的なコンピューティングシステムを示すブロック図である。
【０００７】
【図５】図３及び４の１又は複数のスケジューラを実装し得る例示的なスケジューラのブロック図である。
【０００８】
【図６】図５のバッファクレジット記録装置のさらなる詳細を示す例示的なスケジューラのブロック図である。
【０００９】
【図７】パイプライン及びバッファを実装した異種システムのアクセラレータで実行するワークロードを表す例示的なグラフの図である。
【００１０】
【図８】図５のスケジューラ及び／又は図６のスケジューラを実装するために実行できる機械可読命令によって実装され得る処理を表すフローチャートである。
【００１１】
【図９】図５のスケジューラ及び／又は図６のスケジューラの１又は複数のインスタンス化を実装するための図８の命令を実行するよう構築された例示的なプロセッサプラットフォームのブロック図である。
【００１２】
図は縮尺通りではない。概して、同じもの又は一部のようなものを指すべく、図面及び付随する記述説明全体で同じ参照が用いられるであろう。接続についての言及（例えば、取り付け、結合、接続及び結合）は広く解釈されるべきであり、そうでないと示していない限り、要素の集合の間の中間部材及び要素の間の相対的な移動を含んでよい。従って、接続についての言及は、２つの要素が直接接続されたり互いに固定された関係であることを必ずしも推論されるものではない。
【００１３】
「第１」、「第２」、「第３」等の記述子は、別個に称される複数の要素又はコンポーネントを識別する場合に本明細書で用いられる。用いられるそれらの文脈に基づいて特定又は理解されるのでない限り、そのような記述子は、優先性、物理的順序若しくはリストの配置、又は、時間的な順序のいかなる意味を負わせることを意図しておらず、開示された例の理解の簡略化のために、複数の要素又はコンポーネントを別個に参照するためのラベルとして、単に用いられている。いくつかの例において、ある要素を指すのに「第１」という記述子が詳細な説明で用いられる一方で、同じ要素が請求項で「第２」又は「第３」のような異なる記述子で称されてよい。このような例において、そのような記述子は単に、複数の要素又はコンポーネントの参照を簡略化するために用いられていると理解されるべきである。
【発明を実施するための形態】
【００１４】
多くのコンピュータハードウェア製造者は、アクセラレータとして知られる、ワークロードの処理を加速する処理要素を開発する。例えば、アクセラレータは、中央処理ユニット（ＣＰＵ）、グラフィック処理ユニット（ＧＰＵ）、ビジョン処理ユニット（ＶＰＵ）及び／又はフィールドプログラマブルゲートアレイ（ＦＰＧＡ）であり得る。さらに、アクセラレータは、ワークロードの任意のタイプを処理可能であるが、ワークロードの特定のタイプを最適化するように設計される。例えば、ＣＰＵ及びＦＰＧＡはより汎用の処理を扱うよう設計され得るが、ＧＰＵは、ビデオ、ゲーム及び／又は他の物理及び数学に基

づく計算の処理を向上するよう設計され得るとともに、ＶＰＵは、マシンビジョンタスクの処理を向上するよう設計され得る。

【００１５】

更に、いくつかのアクセラレータは、人工知能（ＡＩ）アプリケーションの処理を特に向上するよう設計される。ＶＰＵはＡＩアクセラレータの特定のタイプであるが、多くの異なるＡＩアクセラレータが用いられ得る。実際、多くのＡＩアクセラレータは、特定用途向け集積回路（ＡＳＩＣ）によって実装され得る。このようなＡＳＩＣベースのＡＩアクセラレータは、機械学習（ＭＬ）、深層学習（ＤＬ）、及び／又は、サポートベクタマシン（ＳＶＭ）、ニューラルネットワーク（ＮＮ）、リカレントニューラルネットワーク（ＲＮＮ）、畳み込みニューラルネットワーク（ＣＮＮ）、ロングショートタームメモリ（ＬＳＴＭ）、ゲートリカレントユニット（ＧＲＵ）等を含む他の人工機械駆動ロジックのようなＡＩの特定のタイプに関したタスクの処理を向上するよう設計され得る。

【００１６】

コンピュータハードウェア製造者は、１つより多いタイプの処理要素を含む異種システムもまた開発している。例えば、コンピュータハードウェア製造者は、ＣＰＵのような汎用の処理要素と、ＦＰＧＡのような汎用アクセラレータ、及び／又は、ＧＰＵ、ＶＰＵ及び／又は他のＡＩアクセラレータのようなより調整されたアクセラレータのいずれかとの両方を組み合わせてよい。このような異種システムは、システムオンチップ（ＳｏＣ）として実装され得る。

【００１７】

開発者が異種システム上で機能、アルゴリズム、プログラム、アプリケーション及び／又は他のコードを動作させることを望む場合、開発者及び／又はソフトウェアは、コンパイル時に、機能、アルゴリズム、プログラム、アプリケーション及び／又は他のコードのためのスケジュールを生成する。一旦スケジュールが生成されると、スケジュールは、実行可能ファイルを生成するために（アヘッドオブタイム又はジャストインタイムのいずれかのパラダイムで）、機能、アルゴリズム、プログラム、アプリケーション及び／又は他のコードの仕様と組み合わせられる。さらに、機能、アルゴリズム、プログラム、アプリケーション及び／又は他のコードはノードを含むグラフとして表されてよく、ここで、グラフはワークロードを示し、各ノードはそのワークロードの特定のタスクを示す。さらに、グラフ内の異なるノード間の接続は、特定のノードが実行されるために必要なデータ入力及び／又は出力を示し、グラフの頂点はグラフのノード間のデータ依存性を示す。

【００１８】

実行可能ファイルは多数の異なる実行可能なセクションを含み、ここで、各実行可能なセクションは特定の処理要素（例えば、ＣＰＵ、ＧＰＵ、ＶＰＵ及び／又はＦＰＧＡ）によって実行可能である。実行可能ファイルの各実行可能なセクションは実行可能なサブセクションをさらに含んでよく、ここで、各実行可能サブセクションは特定の処理要素の計算ビルディングブロック（ＣＢＢ）によって実行可能である。更に又は代替的に、本明細書で開示されるいくつかの例において、開発者及び／又はソフトウェア開発用ソフトウェアは、実行ファイルの実行の成功を判断する基準を定め得る（例えば成功基準）。例えば、このような成功基準は、異種システム及び／又は特定の処理要素の利用の閾値と満たし及び／又はそうでなければ満足するよう実行ファイルを実行することに対応してよい。他の例において、成功基準は、閾値量の時間に実行ファイルを実行することに対応してよい。しかしながら、異種システム及び／又は特定の処理要素でどのように実行ファイルを実行するかを判断する場合に任意の適切な成功機能が用いられてよい。このように、成功基準は、開発者、ソフトウェア及び／又は人工知能システムが成功基準を満たすよう最適化されたスケジュールを含む実行ファイルを生成するのに有益であり得る。

【００１９】

図１は、異種システムのアクセラレータで実行されるワークロードのグラフ１００を表す図である。グラフ１００は、第１ワークロードノード１０２（ＷＮ［０］）、第２ワークロードノード１０４（ＷＮ［１］）、第３ワークロードノード１０６（ＷＮ［２］）、

第４ワークロードノード１０８（ＷＮ［３］）及び第５ワークロードノード１１０（ＷＮ［４］）を含む。図１において、アクセラレータは、スタティックソフトウェアスケジュールでグラフ１００によって表されるワークロードを行っている。スタティックソフトウェアスケジューリングは、アクセラレータの計算ビルディングブロック（ＣＢＢ）上でグラフ１００の異なるワークロードノードを実行するための予め定義された態様を決定することを含む。例えば、スタティックソフトウェアスケジュールは、第１ワークロードノード１０２（ＷＮ［０］）を第１ＣＢＢ１１２に、第２ワークロードノード１０４（ＷＮ［１］）を第２ＣＢＢ１１４に、第３ワークロードノード１０６（ＷＮ［２］）を第３ＣＢＢ１１６に、第４ワークロードノード１０８（ＷＮ［３］）を第４ＣＢＢ１１８に、第５ワークロードノード１１０（ＷＮ［４］）を第２ＣＢＢ１１４に割り当てる。

【００２０】

図１において、スタティックソフトウェアスケジュールは、第４ＣＢＢ１１８で実行する第４ワークロードノード１０８（ＷＮ［３］）と並列して第１ワークロードノード１０２（ＷＮ［０］）が第１ＣＢＢ１１２で実行するという枠組みを作っている。図１において、第１ＣＢＢ１１２が第１ワークロードノード１０２（ＷＮ［０］）を実行するより速く、第４ＣＢＢ１１８は第４ワークロードノード１０８（ＷＮ［３］）を実行する。スタティックソフトウェアスケジュールが、第２ＣＢＢ１１４が第５ワークロードノード１１０（ＷＮ［４］）を実行する前に第２ＣＢＢ１１４が第２ワークロードノード１０４（ＷＮ［１］）を実行するという枠組みを作っているように、第１ＣＢＢ１１２が第１ワークロードノード１０２（ＷＮ［０］）の実行を完了するまで、第２ＣＢＢ１１４はアイドル状態である。さらに、次のワークロードノードの実行前にワークロードノードが全て実行されるまで待つことは、著しいメモリオーバーヘッドを必要とする。というのは、ＣＢＢが２番目のワークロードノード（例えば第２ワークロードノード１０４（ＷＮ［１］））を実行し得る前に、ＣＢＢによる１番目のワークロードノード（例えば第１ワークロードノード１０２（ＷＮ［０］）の実行で生成されたデータをアクセラレータに格納することが必要とされるからである。

【００２１】

図２は、パイプライン及びバッファを実装している異種システムのアクセラレータで実行するワークロードを表すグラフ２００の図である。グラフ２００は、第１ワークロードノード１０２（ＷＮ［０］）、第２ワークロードノード１０４（ＷＮ［１］）、第３ワークロードノード１０６（ＷＮ［２］）、第４ワークロードノード１０８（ＷＮ［３］）及び第５ワークロードノード１１０（ＷＮ［４］）を含む。図２において、アクセラレータは、スタティックソフトウェアスケジュールでグラフ２００によって表されるワークロードを行っている。図２のスタティックソフトウェアスケジュールは、パイプラインを実装するとともに第１バッファ２０２、第２バッファ２０４及び第３バッファ２０６を含むアクセラレータのＣＢＢでのグラフ２００の異なるワークロードノードに対する実行スケジュールの枠組みを作っている。更に、スタティックソフトウェアスケジュールは、第１ワークロードノード１０２（ＷＮ［０］）を第１ＣＢＢ１１２に、第２ワークロードノード１０４（ＷＮ［１］）を第２ＣＢＢ１１４に、第３ワークロードノード１０６（ＷＮ［２］）を第３ＣＢＢ１１６に、第４ワークロードノード１０８（ＷＮ［３］）を第４ＣＢＢ１１８に、第５ワークロードノード１１０（ＷＮ［４］）を第２ＣＢＢ１１４に割り当てる。第１バッファ２０２は第１ＣＢＢ１１２及び第２ＣＢＢ１１４と結合し、第２バッファ２０４は第２ＣＢＢ１１４及び第３ＣＢＢ１１６と結合し、第３バッファ２０６は第２ＣＢＢ１１４及び第４ＣＢＢ１１８と結合する。

【００２２】

バッファ２０２、２０４及び２０６によって、スタティックソフトウェアスケジュールが、ある時間間隔内でワークロードノードの全体を実行するよりむしろ各ＣＢＢがワークロードノードの一部（例えばタイル）をその時間間隔内で処理する枠組みを作るのが可能となる。同様に、スタティックソフトウェアスケジュールは、ワークロードのそのような一部が利用可能となった場合に、他のＣＢＢ（例えばコンシューマー）によって生成され

たデータを処理しているＣＢＢがワークロードノードの一部（例えばタイル）を実行し得る枠組みを作り得る。しかしながら、ワークロードノードを実行しているＣＢＢは利用可能なデータを処理して新たなデータをメモリに書き込むので、ＣＢＢで所与のワークロードノードを実行するためには、ランタイムにおいて閾値量のデータが利用可能でなければならず、ランタイムにおいて結果を書き込むメモリ内の閾値量のスペースがなければならない。バッファは基本的なスタティックソフトウェアスケジューリングによってメモリのオーバーヘッドを減少させるが、それはランタイムにおいてデータ利用可能性及び／又は依存性に高く依存するので、バッファでスタティックソフトウェアスケジュールの枠組みを作ることはますます難しい。さらに、アクセラレータ全体の負荷はアクセラレータ上の各ＣＢＢの処理速度に影響し得るので、所与のアクセラレータのＣＢＢを効果的に利用するスタティックソフトウェアスケジュールを開発するのは難しい。

【００２３】

本明細書で開示された例は、ワークロードのスタティックマッピングの順不同にパイプライン化された実行を可能にする方法及び装置を含む。スタティックソフトウェアスケジューリングとは対照的に、本明細書で開示された例は、予め定められたスタティックソフトウェアスケジュールには依存しない。むしろ、本明細書で開示された例は、アクセラレータ及び／又は他の処理要素上の利用可能なデータ及び利用可能なメモリに基づいて、所与のＣＢＢに割り当てられているワークロードノードのどれを行うかを決定する。さらに、各ＣＢＢは、クレジットの第１の数で表される、第１バッファで利用可能な所与のワークロードに関連づけられたデータの量、及び、クレジットの第２の数で表される、第２バッファで利用可能なスペースの量を追跡する。これは、所与のＣＢＢでのワークロードノードのダイナミックランタイムスケジューリングを可能にする。

【００２４】

ワークロードノードごとに、クレジットの第１の数が第１閾値を満たしかつクレジットの第２の数が第２閾値を満たす場合に、ＣＢＢはワークロードノードを実行し得る。これは、ワークロード全体の所与のグラフから独立した順不同にパイプライン化された実行を可能にする。本明細書で開示された例は、アクセラレータの１又は複数の計算ビルディングブロックにワークロードのスタティックマッピングの順不同にパイプライン化された実行を可能にする装置を提供する。例示的な装置は、クレジットの第１の数をメモリ内へ読み込むインターフェースと、クレジットの第１の数をバッファのメモリ利用可能性に関連付けられたクレジットの閾値数と比較する比較器と、クレジットの第１の数がクレジットの閾値数を満たす場合に、１又は複数の計算ビルディングブロックの最初の一つで実行されるワークロードのワークロードノードを選択するディスパッチャとを含む。

【００２５】

図３は、本開示の教示に従い構築される例示的なコンピューティングシステム３００を示すブロック図である。図３の例において、コンピューティングシステム３００は、例示的なシステムメモリ３０２及び例示的な異種システム３０４を含む。例示的な異種システム３０４は、例示的なホストプロセッサ３０６、例示的な第１通信バス３０８、例示的な第１アクセラレータ３１０ａ、例示的な第２アクセラレータ３１０ｂ及び例示的な第３アクセラレータ３１０ｃを含む。例示的な第１アクセラレータ３１０ａ、例示的な第２アクセラレータ３１０ｂ及び例示的な第３アクセラレータ３１０ｃの各々は、いくつかはアクセラレータの演算に対して汎用的で、いくつかはそれぞれのアクセラレータの演算に対して特化した様々のＣＢＢを含む。

【００２６】

図３の例において、システムメモリ３０２は異種システム３０４に結合される。システムメモリ３０２はメモリである。図３において、システムメモリ３０２は、ホストプロセッサ３０６、第１アクセラレータ３１０ａ、第２アクセラレータ３１０ｂ及び第３アクセラレータ３１０ｃのうちの少なくとも１つの間での共有ストレージである。図３の例において、システムメモリ３０２はコンピューティングシステム３００に位置する物理ストレージである。しかしながら、他の例において、システムメモリ３０２はコンピューティン

グシステム３００の外部にあってよく及び／又はそうでなければ離れていてよい。さらなる例において、システムメモリ３０２は仮想記憶装置であってよい。図３の例において、システムメモリ３０２は、永続ストレージ（例えば読み出し専用メモリ（ＲＯＭ）、プログラマブルＲＯＭ（ＰＲＯＭ）、消去可能ＰＲＯＭ（ＥＰＲＯＭ）、電気的消去可能ＰＲＯＭ（ＥＥＰＲＯＭ等））である。他の例において、システムメモリ３０２は、永続基本入出力システム（ＢＩＯＳ）又はフラッシュストレージであってよい。さらなる例において、システムメモリ３０２は揮発性メモリであってよい。

【００２７】

図３において、異種システム３０４はシステムメモリ３０２と結合される。図３の例において、異種システム３０４は、ホストプロセッサ３０６、及び／又は、第１アクセラレータ３１０ａ、第２アクセラレータ３１０ｂ又は第３アクセラレータ３１０ｃの１又は複数でワークロードを実行することによって、ワークロードを処理する。図３において、異種システム３０４はＳｏＣである。代替的に、異種システム３０４はいかなるその他のタイプのコンピューティング又はハードウェアシステムであってよい。

【００２８】

図３の例において、ホストプロセッサ３０６は、コンピュータ又はコンピューティングデバイス（例えばコンピューティングシステム３００）に関連付けられた演算の完了の実行、遂行及び／又は促進するための命令（例えば機械可読命令）を実行する処理要素である。図３の例において、ホストプロセッサ３０６は、異種システム３０４にとって基本の処理要素であり、かつ、少なくとも１つのコアを含む。代替的に、ホストプロセッサ３０６は、（例えば１つより多いＣＰＵが用いられる例において）共同した一次的な処理要素であってよいが、他の例において、ホストプロセッサ３０６は、二次的な処理要素であってよい。

【００２９】

図３に図示の例において、第１アクセラレータ３１０ａ、第２アクセラレータ３１０ｂ及び／又は第３アクセラレータ３１０ｃの１又は複数は、ハードウェアアクセラレーションのようなコンピューティングタスクのための異種システム３０４で実行するプログラムによって利用されてよい処理要素である。例えば、第１アクセラレータ３１０ａは、ＡＩに対するマシンビジョンタスク（例えばＶＰＵ）を処理する処理速度及び全体性能を向上するよう設計され及び／又はそうでなければ構成され若しくは構築された処理リソースを含む処理要素である。

【００３０】

本明細書で開示された例において、ホストプロセッサ３０６、第１アクセラレータ３１０ａ、第２アクセラレータ３１０ｂ及び第３アクセラレータ３１０ｃの各々は、コンピューティングシステム３００及び／又はシステムメモリ３０２の他の要素と通信する。例えば、ホストプロセッサ３０６、第１アクセラレータ３１０ａ、第２アクセラレータ３１０ｂ、第３アクセラレータ３１０ｃ及び／又はシステムメモリ３０２は第１通信バス３０８で通信する。本明細書で開示されたいくつかの例において、ホストプロセッサ３０６、第１アクセラレータ３１０ａ、第２アクセラレータ３１０ｂ、第３アクセラレータ３１０ｃ及び／又はシステムメモリ３０２は、任意の適切な有線及び／又は無線通信システムで通信してよい。更に、本明細書で開示されたいくつかの例において、ホストプロセッサ３０６、第１アクセラレータ３１０ａ、第２アクセラレータ３１０ｂ、第３アクセラレータ３１０ｃ及び／又はシステムメモリ３０２の各々は、任意の適切な有線及び／又は無線通信システムで、コンピューティングシステム３００の外部の任意のコンポーネントと通信してよい。

【００３１】

図３の例において、第１アクセラレータ３１０ａは、例示的な畳み込みエンジン３１２、例示的なＲＮＮエンジン３１４、例示的なメモリ３１６、例示的なメモリ管理ユニット（ＭＭＵ）３１８、例示的なＤＳＰ３２０、例示的なコントローラ３２２及び例示的なダイレクトメモリアクセス（ＤＭＡ）ユニット３２４を含む。更に、例示的な畳み込みエン

ジン３１２、例示的なＲＮＮエンジン３１４、例示的なＤＭＡユニット３２４、例示的なＤＳＰ３２０及び例示的なコントローラ３２２及びの各々は、例示的な第１スケジューラ３２６、例示的な第２スケジューラ３２８、例示的な第３スケジューラ３３０、例示的な第４スケジューラ３３２及び例示的な第５スケジューラ３３４をそれぞれ含む。例示的なＤＳＰ３２０及び例示的なコントローラ３２２の各々は更に、例示的な第１カーネルライブラリ３３６及び例示的な第２カーネルライブラリ３３８を含む。

【００３２】

図３に図示の例において、畳み込みエンジン３１２は畳み込みに関連したタスクの処理を向上させるよう構成されたデバイスである。さらに、畳み込みエンジン３１２は、視覚イメージの解析に関連するタスク及び／又はＣＮＮに関連する他のタスクの処理を向上させる。図３において、ＲＮＮエンジン３１４はＲＮＮに関連するタスクの処理を向上するよう構成されたデバイスである。更に、ＲＮＮエンジン３１４は、セグメント化されていない繋がった手書き認識、音声認識の解析に関連するタスク及び／又はＲＮＮに関連する他のタスクの処理を向上させる。

【００３３】

図３の例において、メモリ３１６は、畳み込みエンジン３１２、ＲＮＮエンジン３１４、ＭＭＵ３１８、ＤＳＰ３２０、コントローラ３２２及びＤＭＡユニット３２４のうちの少なくとも１つの間の共有ストレージである。図３の例において、メモリ３１６は第１アクセラレータ３１０ａに位置する物理ストレージである。しかしながら、他の例において、メモリ３１６は、第１アクセラレータ３１０ａの外部にあってよく及び／又はそうでなければ離れていてよい。さらなる例において、メモリ３１６は仮想記憶装置であってよい。図３の例において、メモリ３１６は、永続ストレージ（例えばＲＯＭ、ＰＲＯＭ、ＥＰＲＯＭ、ＥＥＰＲＯＭ等）である。他の例において、メモリ３１６は永続ＢＩＯＳ又はフラッシュストレージであってよい。さらなる例において、メモリ３１６は揮発性メモリであってよい。

【００３４】

図３に図示の例において、例示的なＭＭＵ３１８は、メモリ３１６及び／又はシステムメモリ３０２のアドレスへの参照を含むデバイスである。ＭＭＵ３１８は更に、畳み込みエンジン３１２、ＲＮＮエンジン３１４、ＤＳＰ３２０及び／又はコントローラ３２２の１又は複数によって用いられる仮想的メモリアドレスを、メモリ３１６及び／又はシステムメモリ３０２内の物理アドレスへ変換する。

【００３５】

図３の例において、ＤＳＰ３２０は、デジタル信号の処理を向上させるデバイスである。例えば、ＤＳＰ３２０は、カメラ及び／又はコンピュータビジョンに関する他のセンサからのデータのような、連続的な実世界の信号を測定、フィルタ及び／又は圧縮する処理を促進する。図３において、コントローラ３２２は第１アクセラレータ３１０ａの制御ユニットとして実装される。例えば、コントローラ３２２は、第１アクセラレータ３１０ａの演算を管理する。いくつかの例において、コントローラ３２２はクレジットマネージャーを実装する。さらに、コントローラ３２２は、畳み込みエンジン３１２、ＲＮＮエンジン３１４、メモリ３１６、ＭＭＵ３１８及び／又はＤＳＰ３２０の１又は複数に、ホストプロセッサ３０６から受信した機械可読命令にどのように応答するかを命令し得る。

【００３６】

図３に図示の例において、ＤＭＡユニット３２４は、畳み込みエンジン３１２、ＲＮＮエンジン３１４、ＤＳＰ３２０及びコントローラ３２２のうちの少なくとも１つに、ホストプロセッサ３０６から独立してシステムメモリ３０２にアクセスすることを可能にするデバイスである。例えば、ＤＭＡユニット３２４は、アナログ又はデジタル回路、ロジック回路、プログラマブルプロセッサ、プログラマブルコントローラ、グラフィック処理ユニット（ＧＰＵ）、デジタルシグナルプロセッサ（ＤＳＰ）、特定用途向け集積回路（ＡＳＩＣ）、プログラマブル論理デバイス（ＰＬＤ）及び／又はフィールドプログラマブル論理デバイス（ＦＰＬＤ）の１又は複数によって実装され得る。

【００３７】
図３の例において、第１スケジューラ３２６、第２スケジューラ３２８、第３スケジューラ３３０、第４スケジューラ３３２及び第５スケジューラ３３４の各々は、畳み込みエンジン３１２、ＲＮＮエンジン３１４、ＤＭＡユニット３２４、ＤＳＰ３２０及びコントローラ３２２がそれぞれ、オフロードされていた及び／又はそうでなければ第１アクセラレータ３１０ａに送信されているワークロードの一部をいつ実行するかを決定するデバイスである。更に、第１カーネルライブラリ３３６及び第２カーネルライブラリ３３８の各々は、１又は複数のカーネルを含むデータ構造である。第１カーネルライブラリ３３６及び第２カーネルライブラリ３３８のカーネルは、例えば、ＤＳＰ３２０及びコントローラ３２２のそれぞれで高スループットのためにコンパイルされたルーチンである。カーネルは、例えば、コンピューティングシステム３００で行われる実行ファイルの実行可能なサブセクションに対応する。
【００３８】
本明細書で開示された例において、畳み込みエンジン３１２、ＲＮＮエンジン３１４、メモリ３１６、ＭＭＵ３１８、ＤＳＰ３２０、コントローラ３２２及びＤＭＡユニット３２４の各々は、第１アクセラレータ３１０ａの他の要素と通信する。例えば、畳み込みエンジン３１２、ＲＮＮエンジン３１４、メモリ３１６、ＭＭＵ３１８、ＤＳＰ３２０、コントローラ３２２及びＤＭＡユニット３２４は、例示的な第２通信バス３４０で通信する。いくつかの例において、第２通信バス３４０は、コンフィギュレーションアンドコントロール（ＣｎＣ）ファブリック及びデータファブリックにより実装されてよい。本明細書で開示されたいくつかの例において、畳み込みエンジン３１２、ＲＮＮエンジン３１４、メモリ３１６、ＭＭＵ３１８、ＤＳＰ３２０、コントローラ３２２及びＤＭＡユニット３２４は、任意の適切な有線及び／又は無線通信システムで通信してよい。更に、本明細書に開示されたいくつかの例において、畳み込みエンジン３１２、ＲＮＮエンジン３１４、メモリ３１６、ＭＭＵ３１８、ＤＳＰ３２０、コントローラ３２２及びＤＭＡユニット３２４の各々は、任意の適切な有線及び／又は無線通信システムで第１アクセラレータ３１０ａの外部の任意のコンポーネントと通信してよい。
【００３９】
前に言及したように、例示的な第１アクセラレータ３１０ａ、例示的な第２アクセラレータ３１０ｂ及び例示的な第３アクセラレータ３１０ｃの各々は、いくつかはアクセラレータの演算に対して汎用的で、いくつかはそれぞれのアクセラレータの演算に対して特化した様々のＣＢＢを含む。例えば、第１アクセラレータ３１０ａ、第２アクセラレータ３１０ｂ及び第３アクセラレータ３１０ｃの各々は、メモリ、ＭＭＵ、コントローラ、及び、ＣＢＢの各々に対するそれぞれのスケジューラのような汎用ＣＢＢを含む。
【００４０】
図３の例において、第１アクセラレータ３１０ａはＶＰＵを実装し、かつ、畳み込みエンジン３１２、ＲＮＮエンジン３１４及びＤＳＰ３２０（例えば第１アクセラレータ３１０ａの演算に特化したＣＢＢ）を含み、第２アクセラレータ３１０ｂ及び第３アクセラレータ３１０ｃは第２アクセラレータ３１０ｂ及び／又は第３アクセラレータ３１０ｃに特化した追加的又は代替的なＣＢＢを含んでよい。例えば、もし第２アクセラレータ３１０ｂがＧＰＵを実装していれば、第２アクセラレータ３１０ｂの演算に特化したＣＢＢは、スレッドディスパッチャ、グラフィックテクノロジーインターフェース及び／又はコンピュータグラフィック及び／又は画像処理を処理する処理速度及び全体性能を向上するのに好ましい任意のその他のＣＢＢを含み得る。さらに、もし第３アクセラレータ３１０ｃがＦＰＧＡを実装していれば、第３アクセラレータ３１０ｃの演算に特化したＣＢＢは、１又は複数の算術ロジックユニット（ＡＬＵ）及び／又は汎用の計算を処理する処理速度及び全体性能を向上するのに好ましい任意のその他のＣＢＢを含み得る。
【００４１】
図３の異種システム３０４は、ホストプロセッサ３０６、第１アクセラレータ３１０ａ、第２アクセラレータ３１０ｂ及び第３アクセラレータ３１０ｃを含むが、いくつかの例

において、異種システム３０４は、特定用途向け命令セットプロセッサ（ＡＳＩＰ）、物理演算ユニット（ＰＰＵ）、指定されたＤＳＰ、画像プロセッサ、コプロセッサ、浮動小数点ユニット、ネットワークプロセッサ、マルチコア及びフロントエンドプロセッサを含む、任意の数の処理要素（例えばホストプロセッサ及び／又はアクセラレータ）を含んでよい。

【００４２】

さらに、図３の例において、畳み込みエンジン３１２、ＲＮＮエンジン３１４、メモリ３１６、ＭＭＵ３１８、ＤＳＰ３２０、コントローラ３２２、ＤＭＡユニット３２４、第１スケジューラ３２６、第２スケジューラ３２８、第３スケジューラ３３０、第４スケジューラ３３２、第５スケジューラ３３４、第１カーネルライブラリ３３６及び第２カーネルライブラリ３３８は第１アクセラレータ３１０ａ上に実装されるが、畳み込みエンジン３１２、ＲＮＮエンジン３１４、メモリ３１６、ＭＭＵ３１８、ＤＳＰ３２０、コントローラ３２２、ＤＭＡユニット３２４、第１スケジューラ３２６、第２スケジューラ３２８、第３スケジューラ３３０、第４スケジューラ３３２、第５スケジューラ３３４、第１カーネルライブラリ３３６及び第２カーネルライブラリ３３８の１又は複数は、ホストプロセッサ３０６、第２アクセラレータ３１０ｂ及び／又は第３アクセラレータ３１０ｃに実装され得る。

【００４３】

図４は、例示的な１又は複数のスケジューラを含む例示的なコンピューティングシステム４００を示すブロック図である。いくつかの例において、コンピューティングシステム４００は、図３のコンピューティングシステム３００に対応し得る。図４の例において、コンピューティングシステム４００は、例示的な入力４０２、例示的なコンパイラ４０４及び例示的なアクセラレータ４０６を含む。いくつかの例において、アクセラレータ４０６は、図３の第１アクセラレータ３１０ａに対応し得る。図４において、入力４０２はコンパイラ４０４に結合される。入力４０２は、アクセラレータ４０６で実行されるべきワークロードである。いくつかの例において、コンパイラ４０４は、図３のホストプロセッサ３０６及び／又は外部デバイスに対応し得る。

【００４４】

図４の例において、入力４０２は、例えば、機能、アルゴリズム、プログラム、アプリケーション、及び／又は、アクセラレータ４０６によって実行される他のコードである。いくつかの例において、入力４０２は、機能、アルゴリズム、プログラム、アプリケーション及び／又は他のコードのグラフ記述であってよい。追加的又は代替的な例において、入力４０２は深層学習及び／又はコンピュータビジョンのようなＡＩ処理に関するワークロードである。

【００４５】

図４に図示の例において、コンパイラ４０４は入力４０２及びアクセラレータ４０６に結合される。コンパイラ４０４は入力４０２を受信し、入力４０２をアクセラレータ４０６によって実行される１又は複数の実行ファイル内へコンパイルする。例えば、コンパイラ４０４は、入力４０２を受信し、ワークロード（例えば入力４０２）の様々なワークロードノードをアクセラレータ４０６の様々なＣＢＢに割り当てるグラフコンパイラである。更に、コンパイラ４０４は、アクセラレータ４０６のメモリ内の１又は複数のバッファに対してメモリを割り振る。

【００４６】

図４の例において、アクセラレータ４０６はコンパイラ４０４に結合されており、例示的なクレジットマネージャー４０８、例示的なＣｎＣファブリック４１０、例示的なデータファブリック４１１、例示的な畳み込みエンジン４１２、例示的なＤＭＡユニット４１４、例示的なＲＮＮエンジン４１６、例示的なＤＳＰ４１８、例示的なメモリ４２０及び例示的なＭＭＵ４２２を含む。更に、例示的な畳み込みエンジン４１２、例示的なＤＭＡユニット４１４、例示的なＲＮＮエンジン４１６及び例示的なＤＳＰ４１８の各々は、例示的な第１スケジューラ４２４、例示的な第２スケジューラ４２６、例示的な第３スケジ

ューラ４２８及び例示的な第４スケジューラ４３０をそれぞれ含む。さらに、例示的なＤＳＰ４１８は例示的なカーネルライブラリ４３２を含む。いくつかの例において、第１スケジューラ４２４は図３の第１スケジューラ３２６に対応し得る。追加的又は代替的な例において、第２スケジューラ４２６は図３の第３スケジューラ３３０に対応し得る。さらなる例において、第３スケジューラ４２８は図３の第２スケジューラ３２８に対応し得る。いくつかの例において、第４スケジューラ４３０は図４の第４スケジューラ３３２に対応し得る。

【００４７】

図４に図示の例において、クレジットマネージャー４０８はコンパイラ４０４及びＣｎＣファブリック４１０に結合される。クレジットマネージャー４０８は、畳み込みエンジン４１２、ＤＭＡユニット４１４、ＲＮＮエンジン４１６及び／又はＤＳＰ４１８の１又は複数に関連付けられたクレジットを管理するデバイスである。いくつかの例において、クレジットマネージャー４０８は、クレジットマネージャーコントローラとしてコントローラにより実装され得る。クレジットは、メモリ４２０内で利用可能なワークロードノードに関連付けられたデータ、及び／又は、ワークロードノードの出力に対してメモリ４２０内で利用可能なスペースの量を表す。例えば、クレジットマネージャー４０８は、コンパイラ４０４から受信した１又は複数の実行ファイルに基づいて、メモリ４２０を、所与のワークロードのワークロードノードごとに関連付けられた１又は複数のバッファに区分し得る。もしワークロードノードがバッファにデータを書き込むよう構成されていれば、ワークロードノードはプロデューサーであり、もしワークロードノードがバッファからデータを読み出すよう構成されていれば、ワークロードノードはコンシューマーである。

【００４８】

図４の例において、クレジットマネージャー４０８は更に、畳み込みエンジン４１２、ＤＭＡユニット４１４、ＲＮＮエンジン４１６及び／又はＤＳＰ４１８の１又は複数にクレジットを送信及び／又はクレジット受信するよう構成される。いくつかの例において、クレジットマネージャー４０８は、アクセラレータ４０６の制御ユニットとして実装される。例えば、クレジットマネージャー４０８はアクセラレータ４０６の演算を管理し得る。さらに、クレジットマネージャー４０８は、畳み込みエンジン４１２、ＤＭＡユニット４１４、ＲＮＮエンジン４１６及び／又はＤＳＰ４１８の１又は複数に、実行ファイル及び／又はコンパイラ４０４から受信した他の機械可読命令に対してどのように応答するかを命令し得る。

【００４９】

図４の例において、ＣｎＣファブリック４１０は、クレジットマネージャー４０８、畳み込みエンジン４１２、ＤＭＡユニット４１４、ＲＮＮエンジン４１６及びＤＳＰ４１８に結合される。ＣｎＣファブリック４１０は、クレジットマネージャー４０８、畳み込みエンジン４１２、ＤＭＡユニット４１４、ＲＮＮエンジン４１６及び／又はＤＳＰ４１８の１又は複数が、クレジットマネージャー４０８、畳み込みエンジン４１２、ＤＭＡユニット４１４、ＲＮＮエンジン４１６及び／又はＤＳＰ４１８の１又は複数にクレジットを送信及び／又はそれらからクレジットを受信することを可能にする少なくとも１つのロジック回路と電気的に相互接続されるネットワークである。いくつかの例において、ＣｎＣファブリック４１０は、図３の第２通信バス３４０に対応し得る。

【００５０】

図４の例において、データファブリック４１１は、畳み込みエンジン４１２、ＤＭＡユニット４１４、ＲＮＮエンジン４１６、ＤＳＰ４１８、メモリ４２０及びＭＭＵ４２２に結合される。データファブリック４１１は、クレジットマネージャー４０８、畳み込みエンジン４１２、ＲＮＮエンジン４１６、ＤＳＰ４１８、メモリ４２０及び／又はＭＭＵ４２２の１又は複数が、クレジットマネージャー４０８、畳み込みエンジン４１２、ＲＮＮエンジン４１６、ＤＳＰ４１８、メモリ４２０及び／又はＭＭＵ４２２の１又は複数にデータを送信及び／又はそれらからデータを受信することを可能にする少なくとも１つのロジック回路と電気的に相互接続するネットワークである。いくつかの例において、データ

ファブリック４１１は図３の第２通信バス３４０に対応し得る。
【００５１】
図４に図示の例において、畳み込みエンジン４１２はＣｎＣファブリック４１０及びデータファブリック４１１に結合される。畳み込みエンジン４１２は、畳み込みに関連するタスクの処理を向上するよう構成されたデバイスである。さらに、畳み込みエンジン４１２は、視覚イメージの解析に関連付けられたタスク及び／又はＣＮＮに関連付けられた他のタスクの処理を向上させる。いくつかの例において、畳み込みエンジン４１２は図３の畳み込みエンジン３１２に対応し得る。
【００５２】
図４に図示の例において、ＤＭＡユニット４１４はＣｎＣファブリック４１０及びデータファブリック４１１に結合される。ＤＭＡユニット４１４は、畳み込みエンジン４１２、ＲＮＮエンジン４１６又はＤＳＰ４１８のうちの少なくとも１つが、対応するプロセッサ（例えばホストプロセッサ３０６）から独立して、アクセラレータ４０６から離れたメモリ（例えばシステムメモリ３０２）にアクセスすることを可能にするデバイスである。いくつかの例において、ＤＭＡユニット４１４は図３のＤＭＡユニット３２４に対応し得る。例えば、ＤＭＡユニット４１４は、アナログ又はデジタル回路、ロジック回路、プログラマブルプロセッサ、プログラマブルコントローラ、ＧＰＵ、ＤＳＰ、ＡＳＩＣ、ＰＬＤ及び／又はＦＰＬＤの１又は複数によって実装され得る。
【００５３】
図４において、ＲＮＮエンジン４１６はＣｎＣファブリック４１０及びデータファブリック４１１に結合される。ＲＮＮエンジン４１６は、ＲＮＮに関連するタスクの処理を向上するよう構成されたデバイスである。更に、ＲＮＮエンジン４１６は、セグメント化されていない繋がった手書き認識、音声認識の解析に関連付けられたタスク及び／又はＲＮＮに関連付けられた他のタスクの処理を向上させる。いくつかの例において、ＲＮＮエンジン４１６は図３のＲＮＮエンジン３１４に対応し得る。
【００５４】
図４の例において、ＤＳＰ４１８はＣｎＣファブリック４１０及びデータファブリック４１１に結合される。ＤＳＰ４１８はデジタル信号の処理を向上させるデバイスである。例えば、ＤＳＰ４１８は、カメラ及び／又はコンピュータビジョンに関する他のセンサからのデータのような、連続的な実世界の信号を測定、フィルタ及び／又は圧縮する処理を促進する。いくつかの例において、ＤＳＰ４１８は図３のＤＳＰ３２０に対応し得る。
【００５５】
図４の例において、メモリ４２０はデータファブリック４１１に結合される。メモリ４２０は、畳み込みエンジン４１２、ＤＭＡユニット４１４、ＲＮＮエンジン４１６及びＤＳＰ４１８のうちの少なくとも１つの間での共有ストレージである。いくつかの例において、メモリ４２０は図３のメモリ３１６に対応し得る。メモリ４２０は、クレジットマネージャー４０８から受信した実行ファイルに関連付けられたワークロードの１又は複数のワークロードノードに関連付けられた１又は複数のバッファに区分化され得る。図４の例において、メモリ４２０はアクセラレータ４０６に位置する物理ストレージである。しかしながら、他の例において、メモリ４２０はアクセラレータ４０６の外部にあってよく及び／又はそうでなければ離れていてよい。さらなる例において、メモリ４２０は仮想記憶装置であってよい。図４の例において、メモリ４２０は永続ストレージ（例えばＲＯＭ、ＰＲＯＭ、ＥＰＲＯＭ、ＥＥＰＲＯＭ等）である。他の例において、メモリ４２０は永続ＢＩＯＳ又はフラッシュストレージであってよい。さらなる例において、メモリ４２０は揮発性メモリであってよい。
【００５６】
図４に図示の例において、例示的なＭＭＵ４２２はデータファブリック４１１に結合される。ＭＭＵ４２２は、メモリ４２０及び／又はアクセラレータ４０６から離れたメモリのアドレスへの参照を含むデバイスである。ＭＭＵ４２２は更に、畳み込みエンジン４１２、ＤＭＡユニット４１４、ＲＮＮエンジン４１６及び／又はＤＳＰ４１８の１又は複数

に利用される仮想的メモリアドレスを、メモリ４２０及び／又はアクセラレータ４０６から離れたメモリ内の物理アドレスに変換する。いくつかの例において、ＭＭＵ４２２は図３のＭＭＵ３１８に対応し得る。

【００５７】

図４の例において、第１スケジューラ４２４、第２スケジューラ４２６、第３スケジューラ４２８及び第４スケジューラ４３０の各々は、クレジットマネージャー４０８及び／又はアクセラレータ４０６の追加的なＣＢＢによって、畳み込みエンジン４１２、ＤＭＡユニット４１４、ＲＮＮエンジン４１６及びＤＳＰ４１８にそれぞれ割り当てられているワークロードの一部（例えばワークロードノード）を、畳み込みエンジン４１２、ＤＭＡユニット４１４、ＲＮＮエンジン４１６及びＤＳＰ４１８がそれぞれいつ実行するかを決定するデバイスである。タスク及び／又は所与のワークロードノードの他の演算に応じて、ワークロードノードはプロデューサー又はコンシューマーであり得る。プロデューサーワークロードノードは他のワークロードノードで利用されるデータを生成し、他方、コンシューマーワークロードノードは他のワークロードノードで生成されたデータを消費及び／又はそうでなければ処理する。

【００５８】

図４に図示の例において、カーネルライブラリ４３２は１又は複数のカーネルを含むデータ構造である。いくつかの例において、カーネルライブラリ４３２は図３の第１カーネルライブラリ３３６に対応し得る。カーネルライブラリ４３２のカーネルは、例えば、ＤＳＰ４１８で高スループットとなるためにコンパイルされたルーチンである。カーネルは、例えば、アクセラレータ４０６上で動作する実行ファイルの実行可能なサブセクションに対応する。図４の例において、アクセラレータ４０６はＶＰＵを実装し、クレジットマネージャー４０８、ＣｎＣファブリック４１０、データファブリック４１１、畳み込みエンジン４１２、ＤＭＡユニット４１４、ＲＮＮエンジン４１６、ＤＳＰ４１８、メモリ４２０及びＭＭＵ４２２を含むが、アクセラレータ４０６は図４に図示されたこれらに追加的又は代替的なＣＢＢを含んでよい。

【００５９】

図４の例で、演算において、第１スケジューラ４２４は、畳み込みエンジン４１２に割り当てられたワークロードノードに対するワークロードノードへの入力バッファ及びワークロードノードからの出力バッファに対応したクレジットを読み込む。例えば、入力バッファはワークロードノードがそこからデータを読み出すよう構成されたバッファである一方、出力バッファはワークロードノードがそこからデータを書き込むよう構成されたバッファである。いくつかの例において、第１ワークロードノードの入力バッファは第２ワークロードノードの出力バッファであり得る。さらに、第１スケジューラ４２４はクレジットマネージャー４０８からクレジットを受信及び／又はそうでなければ取得する。

【００６０】

図４の例で、演算において、第１スケジューラ４２４は、畳み込みエンジン４１２に割り当てられたワークロードノードを選択し、選択されたワークロードノードへの入力バッファに格納されているデータを演算するために、第１スケジューラ４２４がクレジットの閾値量を受信しているか否かを決定する。例えば、第１スケジューラ４２４は、入力バッファに対するプロデューサーワークロードノードから受信したクレジット数を、入力バッファに対するクレジットの閾値数と比較する。もし第１スケジューラ４２４はクレジットの閾値量を受信していなければ、第１スケジューラ４２４は畳み込みエンジン４１２に割り当てられた他のワークロードノードの処理を繰り返す。

【００６１】

図４に図示した例において、演算において、もし第１スケジューラ４２４が、選択されたワークロードノードへの入力バッファに格納されているデータを演算するためにクレジットの閾値量を受信していれば、第１スケジューラ４２４は、第１スケジューラ４２４が選択されたワークロードノードに対して出力バッファにデータを書き込むためにクレジットの閾値量を受信しているか否かを判断する。例えば、第１スケジューラ４２４は、出力

バッファに対するコンシューマーワークロードノードから受信したクレジット数を、選択されたワークロードノードのための出力バッファに対するクレジットの閾値数と比較する。もし第１スケジューラ４２４がクレジットの閾値量を受信していなければ、第１スケジューラ４２４は畳み込みエンジン４１２に割り当てられた他のワークロードノードの処理を繰り返す。もし第１スケジューラ４２４が出力バッファにデータを書き込むためにクレジットの閾値量を受信していれば、第１スケジューラ４２４は選択されたワークロードノードの実行が準備できたことを示す。次に、第１スケジューラ４２４は畳み込みエンジン４１２に割り当てられた追加的なワークロードノードに対してこの処理を繰り返す。

【００６２】

図４の例で、演算において、畳み込みエンジン４１２に割り当てられたワークロードノードが解析された後に、第１スケジューラ４２４は実行の準備ができたワークロードノードをスケジューリングする。第１スケジューラ４２４は次に、スケジュールに従ってワークロードノードをディスパッチする。ディスパッチされたワークロードノードが畳み込みエンジン４１２によって実行された後に、第１スケジューラ４２４は、入力バッファ及び／又は出力バッファに対応するクレジットをクレジットマネージャー４０８に送信する。第１スケジューラ４２４は実行されるスケジュール内に追加的なワークロードノードがあるかどうかを判断する。もしスケジュール内に追加的なワークロードノードがあるなら、第１スケジューラ４２４は、畳み込みエンジン４１２で実行されるスケジュール内の次のワークロードノードを生じさせる。

【００６３】

図５は図３及び４の１又は複数のスケジューラを実装し得る例示的なスケジューラ５００のブロック図である。例えば、スケジューラ５００は、図３の第１スケジューラ３２６、第２スケジューラ３２８、第３スケジューラ３３０、第４スケジューラ３３２及び／又は第５スケジューラ３３４、及び／又は、図４の第１スケジューラ４２４、第２スケジューラ４２６、第３スケジューラ４２８及び／又は第４スケジューラ４３０、及び／又は、図６のスケジューラ６００、及び／又は、図７の第１スケジューラ７２２、第２スケジューラ７２４、第３スケジューラ７２６及び／又は第４スケジューラ７２８の例示的な実装である。

【００６４】

図５の例において、スケジューラ５００は、例示的なワークロードインターフェース５０２、例示的なバッファクレジット格納装置５０４、例示的なクレジット比較器５０６、例示的なワークロードノードディスパッチャ５０８及び例示的な通信バス５１０を含む。スケジューラ５００は、スケジューラ５００が関連付けられるところＣＢＢに割り当てられているワークロードの一部（例えばワークロードノード）を、スケジューラ５００が関連付けられるところのＣＢＢがいつ実行するかを決定するデバイスである。

【００６５】

図５に図示の例において、ワークロードインターフェース５０２は、スケジューラ５００、バッファクレジット記録装置５０４、クレジット比較器５０６及び／又はワークロードノードディスパッチャ５０８の外部の他のデバイスと通信するよう構成されたデバイスである。

例えば、ワークロードインターフェース５０２は、スケジューラ５００が関連付けられるところのＣＢＢによって実行されるワークロードノードを受信及び／又はそうでなければ取得し得る。更に又は代替的に、ワークロードインターフェース５０２は他のスケジューラ、他のＣＢＢ及び／又は他のデバイスにクレジットを送信及び／又はそれらから受信し得る。さらに、ワークロードインターフェース５０２は、ワークロードノードへの入力バッファ及び／又はワークロードノードからの出力バッファに対応するクレジットを、バッファクレジット記録装置５０４内へ及び／又はそこから読み込み得る。

【００６６】

いくつかの例において、例示的なワークロードインターフェース５０２は例示的なインターフェースする手段を実装する。インターフェース手段は、図８の少なくともブロック

８０２、８１８及び８２２によって実装されるような実行可能命令によって実装される。例えば、図８のブロック８０２、８１８及び８２２の実行可能命令は、図９の例に示される例示的なプロセッサ９１０及び／又は例示的なアクセラレータ９１２のような少なくとも１つのプロセッサで実行されてよい。他の例において、インターフェース手段は、ハードウェアロジック、ハードウェア実装ステートマシン、論理回路、及び／又は、ハードウェア、ソフトウェア及び／又はファームウェアの他の任意の組み合わせによって実装される。

【００６７】

図５に図示される例において、バッファクレジット記録装置５０４は、ワークロードインターフェース５０２、クレジット比較器５０６及び／又はワークロードノードディスパッチャ５０８のうちの少なくとも１つの間での共有ストレージである。バッファクレジット記録装置５０４はスケジューラ５００に位置する物理ストレージである。しかしながら、他の例において、バッファクレジット記録装置５０４はスケジューラ５００の外部にあってよく及び／又はそうでなければそれから離れていてよい。さらなる例において、バッファクレジット記録装置５０４は仮想記憶装置であってよい。図５の例において、バッファクレジット記録装置５０４は永続ストレージ（例えばＲＯＭ、ＰＲＯＭ、ＥＰＲＯＭ、ＥＥＰＲＯＭ等）である。他の例において、バッファクレジット記録装置５０４は永続ＢＩＯＳ又はフラッシュストレージであってよい。さらなる例において、バッファクレジット記録装置５０４は揮発性メモリであってよい。

【００６８】

図５の例において、バッファクレジット記録装置５０４は、スケジューラ５００が関連付けられるところのＣＢＢに割り当てられたワークロードノードに関連付けられたワークロードノードへの入力バッファ及び／又はワークロードノードからの出力バッファに対応したクレジットを格納することに関連付けられたメモリである。例えば、バッファクレジット記録装置５０４は、スケジューラ５００が関連付けられるところのＣＢＢに割り当てられたワークロードノードごとに対するフィールド、及び、スケジューラ５００が関連付けられるところのＣＢＢに割り当てられたワークロードノードに関連付けられたワークロードノードへの各入力バッファ及び／又はワークロードノードからの各出力バッファに対するフィールド、を含むデータ構造として実装され得る。

【００６９】

図５の図示の例において、バッファクレジット記録装置５０４は更に又は代替的に、スケジューラ５００が関連付けられるところのＣＢＢに割り当てられているワークロードノード、及び／又は、ワークロードノードへの入力バッファ及び／又はワークロードノードからの出力バッファに対応するクレジットの閾値量を格納し得る。さらに、バッファクレジット記録装置５０４は、各ワークロードノードへの入力バッファ及び／又は各ワークロードノードからの出力バッファに対するクレジットの閾値数に関連付けられたフィールドを含む。

【００７０】

図５の例において、ワークロードノードがプロデューサーである（例えばワークロードノードが他のワークロードノードによって利用されるデータを生成する）場合に、クレジットの閾値数は、スケジューラ５００が関連付けられるところのＣＢＢがプロデューサーワークロードノードを実行し得る前に満たされるべき出力バッファのスペースの閾値量（例えばメモリ４２０の区分化されたスペース）に対応する。更に、ワークロードノードがコンシューマーである（例えばワークロードノードが他のワークロードノードによって生成されたデータを処理する）場合に、クレジットの閾値数は、スケジューラ５００が関連付けられるところのＣＢＢがコンシューマーワークロードノードを実行し得る前に満たされるべき入力バッファのデータの閾値量（例えばメモリ４２０の区分化されたスペース）に対応する。

【００７１】

いくつかの例において、例示的なバッファクレジット格納装置５０４は例示的な格納す

る手段を実装する。格納手段は、図８において実装されたもののような実行可能命令によって実装され得る。例えば、実行可能命令は図９の例に示された例示的なプロセッサ９１０及び／又は例示的なアクセラレータ９１２のような少なくとも１つのプロセッサで実行されてよい。他の例において、格納手段は、ハードウェアロジック、ハードウェア実装ステートマシン、論理回路、及び／又は、ハードウェア、ソフトウェア及び／又はファームウェアの他の任意の組み合わせによって実装される。

【００７２】

図５に図示された例において、クレジット比較器５０６は、スケジューラ５００が関連付けられるところのＣＢＢに割り当てられたワークロードノードへの入力バッファ及び／又はワークロードノードからの出力バッファに対応したクレジットの閾値数が受信されているか否かを判断するよう構成されたデバイスである。クレジット比較器５０６は、スケジューラ５００が関連付けられるところのＣＢＢに割り当てられたワークロードノードを選択するよう構成される。

【００７３】

図５の例において、クレジット比較器５０６は更に、選択されたワークロードノードに対して入力バッファに格納されたデータを演算するために、スケジューラ５００がクレジットの閾値量を受信したか否かを判断するよう構成される。例えば、クレジット比較器５０６は、外部デバイス（例えばクレジットマネージャー４０８、コントローラ３２２等）から受信したクレジット数に関連付けられたバッファクレジット記録装置５０４内のフィールドを、選択されたワークロードノードへの入力バッファに対するクレジットの閾値数に関連付けられたバッファクレジット記録装置５０４内のフィールドと比較する。もしスケジューラ５００がクレジットの閾値量を受信していないならば、クレジット比較器５０６はスケジューラ５００が関連付けられるところのＣＢＢに割り当てられた他のワークロードノードの処理を繰り返す。

【００７４】

図５に図示した例において、もしスケジューラ５００が入力バッファに格納されたデータを演算するためにクレジットの閾値量を受信しているならば、クレジット比較器５０６は選択されたワークロードノードに対する出力バッファにデータを書き込むためにクレジットの閾値量をスケジューラ５００が受信したか否かを判断する。例えば、クレジット比較器５０６は、選択されたワークロードノードに対する出力バッファに対する外部デバイス（例えばクレジットマネージャー４０８、コントローラ３２２等）から受信したクレジット数に関連付けられたバッファクレジット記録装置５０４内のフィールドを、出力バッファに対するクレジットの閾値数に関連付けられたバッファクレジット記録装置５０４内のフィールドと比較する。

【００７５】

図５の例において、もしスケジューラ５００がクレジットの閾値量を受信していないならば、クレジット比較器５０６はスケジューラ５００に関連付けられるところのＣＢＢに割り当てられた他のワークロードノードの処理を繰り返す。もしスケジューラ５００が出力バッファにデータを書き込むためにクレジットの閾値量を受信しているならば、クレジット比較器５０６は選択されたワークロードノードが実行する準備があることを示す。次に、クレジット比較器５０６は、スケジューラ５００が関連付けられるところのＣＢＢに割り当てられた追加的なワークロードノードに対するこの処理を繰り返す。

【００７６】

いくつかの例において、例示的なクレジット比較器５０６は例示的な比較する手段を実装する。比較手段は、少なくとも図８のブロック８０４、８０６、８０８、８１０及び８１２によって実装されるような実行可能命令によって実装される。例えば、図８のブロック８０４、８０６、８０８、８１０及び８１２の実行可能命令は、図９の例に示された例示的なプロセッサ９１０及び／又は例示的なアクセラレータ９１２のような少なくとも１つのプロセッサで実行されてよい。他の例において、比較手段は、ハードウェアロジック、ハードウェア実装ステートマシン、論理回路、及び／又は、ハードウェア、ソフトウェ

ア及び／又はファームウェアの他の任意の組み合わせによって実装される。
【００７７】
図５の例において、ワークロードノードディスパッチャ５０８は、スケジューラ５００に関連付けられるところのＣＢＢで実行されるべくスケジューラ５００に関連付けられるところのＣＢＢに割り当てられた１又は複数のワークロードノードをスケジューリングするデバイスである。例えば、スケジューラ５００に関連付けられるところのＣＢＢに割り当てられたワークロードノードが解析された後に、ワークロードノードディスパッチャ５０８は実行する準備ができたワークロードノードをスケジューリングする。例えば、ワークロードノードディスパッチャ５０８は、実行する準備ができたワークロードノードをラウンドロビンスケジュールのようなスケジューリングアルゴリズムに基づいてスケジューリングする。ワークロードノードディスパッチャ５０８は次に、スケジュールに従ってワークロードノードをディスパッチする。他の例において、ワークロードノードディスパッチャ５０８は、実行する準備ができたワークロードノードをスケジューリングする任意の他の適切な任意のアルゴリズムを利用し得る。
【００７８】
図５に図示された例において、ディスパッチされたワークロードノードがスケジューラ５００が関連付けられるところのＣＢＢによって実行されるにつれ、ワークロードインターフェース５０２は、ワークロードインターフェース５０２がクレジットをそこから受信するところの外部デバイス（例えばクレジットマネージャー４０８、コントローラ３２２等）への入力バッファに関連付けられたクレジットを送信する。ワークロードノードディスパッチャ５０８は更に、実行されるスケジュール内に追加的なワークロードノードがあるかどうかを判断する。もしスケジュール内に追加的なワークロードノードがあるならば、ワークロードノードディスパッチャ５０８はスケジュール内の次のワークロードノードをディスパッチする。
【００７９】
いくつかの例において、例示的なワークロードノードディスパッチャ５０８は、例示的なディスパッチする手段を実装する。ディスパッチ手段は、少なくとも図８のブロック８１４、８１６及び８２０によって実装されるような実行可能命令によって実装される。例えば、図８のブロック８１４、８１６及び８２０の実行可能命令は、図９の例に示される例示的なプロセッサ９１０及び／又は例示的なアクセラレータ９１２のような少なくとも１つのプロセッサで実行されてよい。他の例において、ディスパッチ手段は、ハードウェアロジック、ハードウェア実装ステートマシン、論理回路、及び／又は、ハードウェア、ソフトウェア及び／又はファームウェアの他の任意の組み合わせによって実装される。
【００８０】
本明細書で開示された例において、ワークロードインターフェース５０２、バッファクレジット記録装置５０４、クレジット比較器５０６及びワークロードノードディスパッチャ５０８の各々は、スケジューラ５００の他の要素と通信する。例えば、ワークロードインターフェース５０２、バッファクレジット記録装置５０４、クレジット比較器５０６及びワークロードノードディスパッチャ５０８は例示的な通信バス５１０で通信する。本明細書で開示されるいくつかの例において、ワークロードインターフェース５０２、バッファクレジット記録装置５０４、クレジット比較器５０６及びワークロードノードディスパッチャ５０８は、任意の適切な有線及び／又は無線通信システムで通信してよい。更に、本明細書で開示されたいくつかの例において、ワークロードインターフェース５０２、バッファクレジット記録装置５０４、クレジット比較器５０６及びワークロードノードディスパッチャ５０８の各々は、任意の適切な有線及び／又は無線通信システムでスケジューラ５００の外部の任意のコンポーネントと通信してよい。
【００８１】
図６は、図５のバッファクレジット記録装置５０４のさらなる詳細を示す例示的なスケジューラ６００のブロック図である。スケジューラ６００は、図３の第１スケジューラ３２６、第２スケジューラ３２８、第３スケジューラ３３０、第４スケジューラ３３２及び

／又は第５スケジューラ３３４、及び／又は、図４の第１スケジューラ４２４、第２スケジューラ４２６、第３スケジューラ４２８及び／又は第４スケジューラ４３０、及び／又は、図５のスケジューラ５００、及び／又は、図７の第１スケジューラ７２２、第２スケジューラ７２４、第３スケジューラ７２６及び／又は第４スケジューラ７２８の例示的な実装である。

【００８２】

図６の例において、スケジューラ６００は、例示的なワークロードインターフェース５０２、例示的なバッファクレジット格納装置５０４、例示的なクレジット比較器５０６及び例示的なワークロードノードディスパッチャ５０８を含む。スケジューラ６００は、スケジューラ６００が関連付けられるところのＣＢＢが、スケジューラ６００が関連付けられるところのＣＢＢに割り当てられているワークロードの一部（例えばワークロードノード）をいつ実行するかを決定するデバイスである。

【００８３】

図６に図示の例において、ワークロードインターフェース５０２は、スケジューラ６００の外部の１又は複数のデバイス、バッファクレジット記録装置５０４及びワークロードノードディスパッチャ５０８に結合される。ワークロードインターフェース５０２は、スケジューラ６００の外部の他のデバイス、バッファクレジット記録装置５０４及び／又はワークロードノードディスパッチャ５０８と通信するよう構成されているデバイスである。例えば、ワークロードインターフェース５０２は、スケジューラ６００が関連付けられるところのＣＢＢによって実行されるワークロードノードを受信及び／又はそうでなければ取得し得る。更に又は代替的に、ワークロードインターフェース５０２は、クレジットをスケジューラ６００の外部の１又は複数のデバイスへ送信及び／又はそこから受信し得る。さらに、ワークロードインターフェース５０２は、ワークロードノードへの入力バッファ及び／又はワークロードノードからの出力バッファに対応するクレジットを、バッファクレジット記録装置５０４内へ及び／又はから読み込み得る。

【００８４】

図６に図示した例において、バッファクレジット記録装置５０４は、ワークロードインターフェース５０２、クレジット比較器５０６及び／又はワークロードノードディスパッチャ５０８のうちの少なくとも１つの間での共有ストレージである。バッファクレジット記録装置５０４はスケジューラ５００に位置する物理ストレージである。しかしながら、他の例において、バッファクレジット記録装置５０４はスケジューラ５００の外部にあってよく及び／又はそうでなければそれから離れていてもよい。さらなる例において、バッファクレジット記録装置５０４は仮想記憶装置であってよい。図５の例において、バッファクレジット記録装置５０４は、永続ストレージ（例えばＲＯＭ、ＰＲＯＭ、ＥＰＲＯＭ、ＥＥＰＲＯＭ等）である。他の例において、バッファクレジット記録装置５０４は永続ＢＩＯＳ又はフラッシュストレージであってよい。さらなる例において、バッファクレジット記録装置５０４は揮発性メモリであってよい。

【００８５】

図６の例において、バッファクレジット記録装置５０４は、第１ワークロードノードＷＮ［０］、第２ワークロードノードＷＮ［１］及び第ｎワークロードノードＷＮ［ｎ］に対応する行を含むデータ構造である。バッファクレジット記録装置５０４は更に、第１コンシューマー（例えばコンシューマー［０］）に対する入力バッファ、第ｌコンシューマー（例えばコンシューマー［ｌ］）に対する入力バッファ、第１プロデューサー（例えばプロデューサー［０］）に対する出力バッファ、及び、第ｍプロデューサー（例えばプロデューサー［ｍ］）に対する出力バッファに対応する列を含む。バッファクレジット記録装置５０４はさらに、各ワークロードノードへの入力バッファ及び／又は各ワークロードノードからの出力バッファのクレジットの閾値数に対応する列を含む。

【００８６】

図６に図示の例において、第１ワークロードノードＷＮ［０］、第２ワークロードノードＷＮ［１］及び第ｎワークロードノードＷＮ［ｎ］の各々は、スケジューラ６００が関

連付けられるところのＣＢＢに割り当てられる。バッファクレジット記録装置５０４において、第１ワークロードノードＷＮ［０］、第２ワークロードノードＷＮ［１］及び第ｎワークロードノードＷＮ［ｎ］に対応する行と、第１コンシューマー（例えばコンシューマー［０］）に対する入力バッファ、第ｌコンシューマー（例えばコンシューマー［ｌ］）に対する入力バッファ、第１プロデューサー（例えばプロデューサー［０］）に対する出力バッファ、第ｍプロデューサー（例えばプロデューサー［ｍ］）に対する出力バッファに対応する列との間の交差は、そのバッファに対する１又は複数の外部デバイスから受信するクレジット数に対応するフィールドを示す。さらに、各ワークロードノードへの入力バッファ及び／又は各ワークロードノードからの出力バッファに対するクレジットの閾値数に対応する列は、スケジューラ６００が関連付けられるところのＣＢＢがそれぞれのワークロードノードを演算し得る前にバッファに対して満たされるべきクレジットの閾値数を示す。

【００８７】

図６の例において、バッファクレジット記録装置５０４における、第１ワークロードノードＷＮ［０］、第２ワークロードノードＷＮ［１］及び第ｎワークロードノードＷＮ［ｎ］に対応する行と、第１コンシューマー（例えばコンシューマー［０］）に対する入力バッファ及び第ｌコンシューマー（例えばコンシューマー［ｌ］）に対する入力バッファに対応する列との間の交差のフィールドは、外部デバイス（例えばクレジットマネージャー４０８、コントローラ３２２等）によってゼロの値に初期化される。更に、バッファクレジット記録装置５０４における、第１ワークロードノードＷＮ［０］、第２ワークロードノードＷＮ［１］及び第ｎワークロードノードＷＮ［ｎ］に対応する行と、第１プロデューサー（例えばプロデューサー［０］）に対する出力バッファ及び第ｍプロデューサー（例えばプロデューサー［ｍ］）に対する出力バッファに対応する列との間の交差のフィールドは、外部デバイス（例えばクレジットマネージャー４０８、コントローラ３２２等）によって関連するバッファ内に区分化されたメモリの量に対応した値に初期化される。さらに、入力バッファ及び／又は出力バッファに対するクレジットの閾値数に対応する列は、外部デバイス（例えばクレジットマネージャー４０８、コントローラ３２２、ホストプロセッサ３０６で実行するソフトウェア等）によって初期化される。

【００８８】

図６に図示の例において、クレジット比較器５０６は、バッファクレジット記録装置５０４及びワークロードノードディスパッチャ５０８に結合される。クレジット比較器５０６は、スケジューラ６００が関連付けられるところのＣＢＢに割り当てられたワークロードノードへの入力バッファ及び／又はワークロードノードからの出力バッファに対応するクレジットの閾値数を受信しているか否かを判断するよう構成されたデバイスである。図６の例において、ワークロードノードディスパッチャ５０８は、ワークロードインターフェース５０２、バッファクレジット記録装置５０４、クレジット比較器５０６及びスケジューラ６００の外部の１又は複数のデバイスに結合される。ワークロードノードディスパッチャ５０８は、例えば、スケジューラ６００が関連付けられるところのＣＢＢで実行されるべく、スケジューラ６００が関連付けられるところのＣＢＢに割り当てられた１又は複数のワークロードノードをスケジューリングするデバイスである。

【００８９】

図６の例で、演算において、ワークロードインターフェース５０２が外部デバイス（例えばクレジットマネージャー４０８、コントローラ３２２等）からワークロードノードを受信及び／又はそうでなければ取得した場合に、ワークロードインターフェース５０２はワークロードノードをワークロードノードに対応するバッファクレジット記録装置５０４のそれぞれのフィールド内へ読み込む。さらに、クレジット比較器５０６は、スケジューラ６００が関連付けられるところのＣＢＢに割り当てられたワークロードノードを選択する。

【００９０】

図６に図示の例において、クレジット比較器５０６は、選択されたワークロードノード

に対する入力バッファに格納されたデータを演算するためにクレジットの閾値量をスケジューラ６００が受信しているか否かを判断する。例えば、クレジット比較器５０６は、外部デバイス（例えばクレジットマネージャー４０８、コントローラ３２２等）から受信したクレジット数に関連付けられたバッファクレジット記録装置５０４内のフィールドを、選択されたワークロードノードへの入力バッファに対するクレジットの閾値数に関連付けられたバッファクレジット記録装置５０４内のフィールドと比較する。クレジットの閾値数は、スケジューラ６００が関連付けられるところのＣＢＢがコンシューマーワークロードノードを実行し得る前に満たされるべき入力バッファのデータの閾値量（例えばメモリ４２０の区分化されたスペース）に対応する。もしスケジューラ６００がクレジットの閾値量を受信していないならば、クレジット比較器５０６は、スケジューラ６００が関連付けられるところのＣＢＢに割り当てられた他のワークロードノードの処理を繰り返す。

【００９１】

図６に図示した例において、もしスケジューラ６００が入力バッファに格納されたデータを演算するためにクレジットの閾値量を受信しているならば、クレジット比較器５０６は、スケジューラ６００が選択されたワークロードノードに対する出力バッファへデータを書き込むためにクレジットの閾値量を受信しているか否かを判断する。例えば、クレジット比較器５０６は、選択されたワークロードノードに対する出力バッファに対する外部デバイス（例えばクレジットマネージャー４０８、コントローラ３２２等）から受信したクレジット数に関連付けられたバッファクレジット記録装置５０４内のフィールドを、出力バッファに対するクレジットの閾値数に関連付けられたバッファクレジット記録装置５０４内のフィールドと比較する。クレジットの閾値数は、スケジューラ６００が関連付けられるところのＣＢＢがプロデューサーワークロードノードを実行し得る前に満たされるべき出力バッファのスペースの閾値量（例えばメモリの区分化されたスペース）に対応し得る。

【００９２】

図６の例において、もしスケジューラ６００がクレジットの閾値量を受信していないならば、クレジット比較器５０６はスケジューラ６００が関連付けられるところのＣＢＢに割り当てられた他のワークロードノードの処理を繰り返す。もしスケジューラ６００が出力バッファへデータを書き込ためにクレジットの閾値量を受信しているならば、クレジット比較器５０６は、選択されたワークロードノードが実行する準備ができていることを示す。次に、クレジット比較器５０６は、スケジューラ６００が関連付けられるところのＣＢＢに割り当てられた追加的なワークロードノードに対するこの処理を繰り返す。

【００９３】

図６の例において、ワークロードノードディスパッチャ５０８は、スケジューラ６００が関連付けられるところのＣＢＢで実行されるべく、スケジューラ６００が関連付けられるところのＣＢＢに割り当てられた１又は複数のワークロードノードをスケジューリングするデバイスである。例えば、スケジューラ６００が関連付けられるところのＣＢＢに割り当てられたワークロードノードが解析された後に、ワークロードノードディスパッチャ５０８は、実行の準備ができたワークロードノードをスケジューリングする。例えば、ワークロードノードディスパッチャ５０８は、実行の準備ができたワークロードノードを、ラウンドロビンスケジュールのようなスケジューリングアルゴリズムに基づいてスケジューリングする。ワークロードノードディスパッチャ５０８は次に、スケジュールに従ってワークロードノードをディスパッチする。他の例において、ワークロードノードディスパッチャ５０８は実行の準備ができたワークロードノードをスケジューリングする任意の他の適切な任意のアルゴリズムを利用し得る。

【００９４】

図６に図示した例において、ディスパッチされたワークロードノードがスケジューラ６００が関連付けられるところのＣＢＢによって実行されるにつれ、ワークロードインターフェース５０２は、入力バッファに関連付けられたクレジットをワークロードインターフェース５０２がそこからクレジットを受信したところの外部デバイス（例えばクレジット

マネージャー４０８、コントローラ３２２等）へ送信する。ワークロードノードディスパッチャ５０８は更に、実行されるべきスケジュール内に追加的なワークロードノードがあるかどうかを判断する。もしスケジュール内に追加的なワークロードノードがあるならば、ワークロードノードディスパッチャ５０８はスケジュール内の次のワークロードノードをディスパッチする。

【００９５】

図７は、パイプライン及びバッファを実装している異種システムのアクセラレータで実行するワークロードを表す例示的なグラフ７００の図である。例えば、アクセラレータは、図３の第１アクセラレータ３１０ａであり、異種システムは異種システム３０４である。例示的なグラフ７００は、例示的な第１ワークロードノード７０２（ＷＮ［０］）、例示的な第２ワークロードノード７０４（ＷＮ［１］）、例示的な第３ワークロードノード７０６（ＷＮ［２］）、例示的な第４ワークロードノード７０８（ＷＮ［３］）及び例示的な第５ワークロードノード７１０（ＷＮ［４］）を含む。図７の例において、アクセラレータは、ワークロードノードを様々なＣＢＢに割り当てる例示的なクレジットマネージャー７１２からのスケジュールに基づいたグラフ７００によって表されるワークロードを実行するよう構成されている。例えば、クレジットマネージャー７１２及び／又は他のコントローラは、第１ワークロードノード７０２（ＷＮ［０］）を例示的な第１ＣＢＢ７１４に、第２ワークロードノード７０４（ＷＮ［１］）を例示的な第２ＣＢＢ７１６に、第３ワークロードノード７０６（ＷＮ［２］）を例示的な第３ＣＢＢ７１８に、第４ワークロードノード７０８（ＷＮ［３］）を例示的な第４ＣＢＢ７２０に、及び、第５ワークロードノード７１０（ＷＮ［４］）を例示的な第２ＣＢＢ７１６に割り当てる。

【００９６】

図７の例において、例示的な第１ＣＢＢ７１４、例示的な第２ＣＢＢ７１６、例示的な第３ＣＢＢ７１８及び例示的な第４ＣＢＢ７２０の各々は、例示的な第１スケジューラ７２２、例示的な第２スケジューラ７２４、例示的な第３スケジューラ７２６及び例示的な第４スケジューラ７２８を含む。第１スケジューラ７２２、第２スケジューラ７２４、第３スケジューラ７２６及び第４スケジューラ７２８の各々は、図５のスケジューラ５００及び／又は図６のスケジューラ６００によって実装され得る。

【００９７】

図７に図示の例において、第１ワークロードノード７０２（ＷＮ［０］）及び第２ワークロードノード７０４（ＷＮ［１］）は例示的な第１バッファ７３０に関連付けられる。第１バッファ７３０は、第１ワークロードノード７０２（ＷＮ［０］）の出力バッファ及び第２ワークロードノード７０４（ＷＮ［１］）の入力バッファである。第２ワークロードノード７０４（ＷＮ［１］）及び第３ワークロードノード７０６（ＷＮ［２］）は例示的な第２バッファ７３２に関連付けられる。第２バッファ７３２は第２ワークロードノード７０４（ＷＮ［１］）の出力バッファ及び第３ワークロードノード７０６（ＷＮ［２］）の入力バッファである。第４ワークロードノード７０８（ＷＮ［３］）及び第５ワークロードノード７１０（ＷＮ［４］）は例示的な第３バッファ７３４に関連付けられる。第３バッファ７３４は第４ワークロードノード７０８（ＷＮ［３］）の出力バッファ及び第５ワークロードノード７１０（ＷＮ［４］）の入力バッファである。第１バッファ７３０、第２バッファ７３２及び第３バッファ７３４の各々は循環バッファによって実装され得る。図７の例において、第１バッファ７３０、第２バッファ７３２及び第３バッファ７３４の各々は、アクセラレータのメモリの５つの区分を含み、それらの各々はデータのタイルを格納し得る。

【００９８】

図７に図示した例において、第１ワークロードノード７０２（ＷＮ［０］）はプロデューサーワークロードノードであり、クレジットマネージャー７１２は第１バッファ７３０に対する５つのクレジットで第１スケジューラ７２２を初期化する。同様に、第２ワークロードノード７０４（ＷＮ［１］）はプロデューサーワークロードノードなので、クレジットマネージャー７１２は第２バッファ７３２に対する５つのクレジットで第２スケジュ

ーラ７２４を初期化する。更に、第４ワークロードノード７０８（ＷＮ［３］）はプロデューサーワークロードノードであり、クレジットマネージャー７１２は第３バッファ７３４に対する５つのクレジットで第４スケジューラ７２８を初期化する。

【００９９】

第１スケジューラ７２２、第２スケジューラ７２４及び第４スケジューラ７２８の各々に提供された５つのクレジットは、第１バッファ７３０、第２バッファ７３２及び第３バッファ７３４のサイズの表現である。更に、第２ワークロードノード７０４（ＷＮ［１］）はまたコンシューマーワークロードノードでもあり、クレジットマネージャー７１２は第１バッファ７３０に対するゼロクレジットで第２スケジューラ７２４を初期化する。さらに、第３ワークロードノード７０６（ＷＮ［２］）はコンシューマーワークロードノードなので、クレジットマネージャー７１２は第２バッファ７３２に対するゼロクレジットで第３スケジューラ７２６を初期化する。さらに、第５ワークロードノード７１０（ＷＮ［４］）はコンシューマーワークロードノードであり、クレジットマネージャー７１２は第３バッファ７３４に対するゼロクレジットで第２スケジューラ７２４を初期化する。

【０１００】

図７の例において、第１スケジューラ７２２は第１ワークロードノード７０２（ＷＮ［０］）への入力バッファ及びそこからの出力バッファの両方に対するクレジットの閾値数を受信しているので、第１スケジューラ７２２は第１ＣＢＢ７１４で実行する第１ワークロードノード７０２（ＷＮ［０］）をディスパッチする。更に、第４スケジューラ７２８は、第４ワークロードノード７０８（ＷＮ［３］）への入力バッファ及びそこからの出力バッファの両方に対するクレジットの閾値数を受信しているので、第４スケジューラ７２８は第４ＣＢＢ７２０で実行する第４ワークロードノード７０８（ＷＮ［３］）をディスパッチする。第１ワークロードノード７０２（ＷＮ［０］）が第１ＣＢＢ７１４で実行されるにつれ、第１ＣＢＢ７１４はデータを第１バッファ７３０に送信する。同様に、第４ワークロードノード７０８（ＷＮ［３］）が第４ＣＢＢ７２０で実行されるにつれ、第４ＣＢＢ７２０はデータを第３バッファ７３４に送信する。

【０１０１】

図７に図示された例において、第１ＣＢＢ７１４及び第４ＣＢＢ７２０の各々がそれぞれ第１ワークロードノード７０２（ＷＮ［０］）及び第４ワークロードノード７０８（ＷＮ［３］）に関連付けられたデータのタイルを送信するにつれ、第１スケジューラ７２２及び第４スケジューラ７２８は、それぞれ、第１ＣＢＢ７１４及び第４ＣＢＢ７２０から第１バッファ７３０及び第３バッファ７３４へ送信されたデータのタイルごとにクレジットマネージャー７１２へクレジットを送信する。クレジットマネージャー７１２は、第１スケジューラ７２２から受信したクレジットを第２スケジューラ７２４へ、第４スケジューラ７２８から受信したクレジットを第２スケジューラ７２４へ送信する。第４ＣＢＢ７２０が第４ワークロードノード７０８（ＷＮ［３］）を実行するにつれ、第４ＣＢＢ７２０は第３バッファ７３４に格納するデータの２つのタイルを生成する。同様に、第１ＣＢＢ７１４が第１ワークロードノード７０２（ＷＮ［０］）を実行するにつれ、第１ＣＢＢ７１４は第１バッファ７３０に格納するデータの５つのタイルを生成する。

【０１０２】

図７の例において、第１ＣＢＢ７１４が第１ワークロードノード７０２（ＷＮ［０］）を実行するよりも迅速に、第４ＣＢＢ７２０は第４ワークロードノード７０８（ＷＮ［３］）を実行する。第２バッファ７３２には利用可能なメモリはあるが、第２ワークロードノード７０４（ＷＮ［１］）が依存しているデータが準備される前に第５ワークロードノード７１０（ＷＮ［４］）が依存しているデータが準備されるので、第２スケジューラ７２４は、第２ワークロードノード７０４（ＷＮ［１］）とは対照的に、第５ワークロードノード７１０（ＷＮ［４］）を第２ＣＢＢ７１６で実行するものとして選択する。

【０１０３】

図７に図示の例において、第５ワークロードノード７１０（ＷＮ［４］）が第２ＣＢＢ７１６で実行され、第２ＣＢＢ７１６が第３バッファ７３４に格納されたデータのタイル

を消費するにつれ、第２スケジューラ７２４は、第３バッファ７３４に関連付けられたクレジットを、第３バッファ７３４からの第２ＣＢＢ７１６で消費されたデータのタイルごとに、クレジットマネージャー７１２へ返送する。次に、第１バッファ７３０及び第２バッファ７３２に対するクレジットの閾値量が満たされると、第２スケジューラ７２４は第２ＣＢＢ７１６で実行する第２ワークロードノード７０４（ＷＮ［１］）をディスパッチする。第２ＣＢＢ７１６が第２ワークロードノード７０４（ＷＮ［１］）に関連付けられたデータのタイルを生成し、第２バッファ７３２にデータを出力するにつれ、第２スケジューラ７２４は、第２ＣＢＢ７１６から第２バッファ７３２へ送信されたデータのタイルごとに、第２バッファ７３２に関連付けられたクレジットをクレジットマネージャー７１２に送信する。

【０１０４】

図７の例において、第２スケジューラ７２４から第２バッファ７３２に関連付けられたクレジットを受信すると、クレジットマネージャー７１２は、第２バッファ７３２に関連付けられたクレジットを第３スケジューラ７２６に送信する。第３スケジューラ７２６が第２バッファ７３２に関連付けられたクレジットの閾値量を受信した場合に、第３スケジューラ７２６は、第３ＣＢＢ７１８で実行する第３ワークロードノード７０６（ＷＮ［２］）をディスパッチする。第３ＣＢＢ７１８が第３ワークロードノード７０６（ＷＮ［２］）を実行し、第３ＣＢＢ７１８が第２バッファ７３２に格納されたデータのタイルを消費するにつれ、第３スケジューラ７２６は、第２バッファ７３２に関連付けられたクレジットを、第２バッファ７３２からの第３ＣＢＢ７１８で消費されたデータのタイルごとに、クレジットマネージャー７１２へ返送する。

【０１０５】

追加的又は代替的な例において、第１ＣＢＢ７１４は図４の畳み込みエンジン４１２に対応し得、第１スケジューラ７２２は図４の第１スケジューラ４２４に対応し得る。いくつかの例において、第２ＣＢＢ７１６は図４のＲＮＮエンジン４１６に対応し得、第２スケジューラ７２４は図４の第３スケジューラ４２８に対応し得る。さらなる例において、第３ＣＢＢ７１８は図４のＤＭＡユニット４１４に対応し得、第３スケジューラ７２６は図４の第２スケジューラ４２６に対応し得る。いくつかの例において、第４ＣＢＢ７２０は図４のＤＳＰ４１８に対応し得、第４スケジューラ７２８は図４の第４スケジューラ４３０に対応し得る。

【０１０６】

図３の第１スケジューラ３２６、第２スケジューラ３２８、第３スケジューラ３３０、第４スケジューラ３３２及び／又は第５スケジューラ３３４、及び／又は、図４の第１スケジューラ４２４、第２スケジューラ４２６、第３スケジューラ４２８及び／又は第４スケジューラ４３０、及び／又は、図７の第１スケジューラ７２２、第２スケジューラ７２４、第３スケジューラ７２６及び／又は第４スケジューラ７２８の実装の例示的な態様は図５及び／又は図６に図示されているが、図５及び／又は図６に図示された要素、処理及び／又はデバイスの１又は複数は、組み合わせたり、分割されたり、再構成されたり、省略されたり、除去されたり、及び／又は、その他の方式によって実装されてよい。さらに、図５の例示的なワークロードインターフェース５０２、例示的なバッファクレジット格納装置５０４、例示的なクレジット比較器５０６、例示的なワークロードノードディスパッチャ５０８、例示的な通信バス５１０及び／又は、より一般的に、例示的なスケジューラ５００及び／又は図６の例示的なスケジューラ６００は、ハードウェア、ソフトウェア、ファームウェア、及び／又は、ハードウェア、ソフトウェア及び／又はファームウェアの任意の組み合わせによって実装されてよい。従って、例えば、図５の例示的なワークロードインターフェース５０２、例示的なバッファクレジット格納装置５０４、例示的なクレジット比較器５０６、例示的なワークロードノードディスパッチャ５０８、例示的な通信バス５１０及び／又は、より一般的に、例示的なスケジューラ５００及び／又は図６の例示的なスケジューラ６００のいずれも、アナログ又はデジタル回路、ロジック回路、プログラマブルプロセッサ、プログラマブルコントローラ、グラフィック処理ユニット（Ｇ

ＰＵ）、デジタルシグナルプロセッサ（ＤＳＰ）、特定用途向け集積回路（ＡＳＩＣ）、プログラマブル論理デバイス（ＰＬＤ）及び／又はフィールドプログラマブル論理デバイス（ＦＰＬＤ）の１又は複数によって実装され得る。本特許の任意の装置又はシステムの請求項を純粋なソフトウェア及び／又はファームウェア実装を包含するように読む場合に、図５の例示的なワークロードインターフェース５０２、例示的なバッファクレジット格納装置５０４、例示的なクレジット比較器５０６、例示的なワークロードノードディスパッチャ５０８、例示的な通信バス５１０、及び／又は、より一般的に、例示的なスケジューラ５００及び／又は図６の例示的なスケジューラ６００のうちの少なくとも１つは、ソフトウェア及び／又はファームウェアを含む、メモリ、デジタルバーサタイルディスク（ＤＶＤ）、コンパクトディスク（ＣＤ）、ブルーレイディスク等のような、非一時的コンピュータ可読記憶装置デバイス、又は、ストレージディスクを含むように本明細書では明示的に定義される。更にまた、図５の例示的なスケジューラ５００及び／又は図６の例示的なスケジューラ６００は、図５及び／又は図６に図示されている、要素、処理及び／又はデバイスに追加し、又は、代わりに、１又は複数のそれらを含んでよく、及び／又は、図示された要素、処理及びデバイスのいずれか一つより多く又は全てを含んでよい。本明細書で用いられるように、「通信」及びそれらの変形を含む語句は、直接的な通信及び／又は１又は複数の中間媒介コンポーネントを介した間接的な通信を包含し、直接物理的な（例えば有線）通信及び／又は一定の通信を必要としておらず、むしろ、周期的な間隔、スケジュールされた間隔、非周期的な間隔及び／又は一回的なイベントでの選択的な通信を更に含む。

【０１０７】

図５のスケジューラ５００及び／又は図６のスケジューラ６００を実装するための、例示的なハードウェアロジック、機械可読命令、ハードウェア実装ステートマシン及び／又はそれらの任意の組み合わせを表すフローチャートが図８に示される。機械可読命令は、図９に関して以下で説明される例示的なプロセッサプラットフォーム９００に示されるプロセッサ９１０及び／又はアクセラレータ９１２のようなコンピュータプロセッサによる実行のための実行可能プログラム又は実行可能プログラムの一部の１又は複数であってよい。プログラムは、プロセッサ９１０及び／又はアクセラレータ９１２に関連付けられた、ＣＤ－ＲＯＭ、フロッピーディスク、ハードドライブ、ＤＶＤ、ブルーレイディスクのような非一時的コンピュータ可読記憶媒体又はメモリに格納されたソフトウェア内に具現されてよいが、プログラム全体及び／又はそれらの一部は代替的にプロセッサ９１０及び／又はアクセラレータ９１２以外のデバイスで実行され得、及び／又は、ファームウェア又は専用のハードウェアに具現され得る。さらに、例示的なプログラムが図８に図示されたフローチャートに言及して記載されるが、図５の例示的なスケジューラ５００及び／又は図６スケジューラ６００を実装する多くの他の方法が代替的に用いられてよい。例えば、ブロックの実行の順序が変更されてよく、及び／又は、記載されたいくつかのブロックは変更、除去又は組み合わせされてよい。更に又は代替的に、ブロックのいずれか又は全部は、ソフトウェア又はファームウェアを実行することなく対応する演算を実行するよう構築された１又は複数のハードウェア回路（例えば別個及び／又は統合されたアナログ及び／又はデジタル回路、ＦＰＧＡ、ＡＳＩＣ、比較器、オペレーショナルアンプ（オペアンプ）、ロジック回路等）によって実装されてよい。

【０１０８】

本明細書に記載された機械可読命令は、圧縮されたフォーマット、暗号化されたフォーマット、細分化されたフォーマット、コンパイルされたフォーマット、実行可能なフォーマット、パッケージされたフォーマット等の１又は複数で格納されてよい。本明細書で記載される機械可読命令は、機械実行可能命令を創造、製造及び／又は生成するのに利用されてよいデータ（例えば命令の一部、コード、コードの表現等）として格納されてよい。例えば、機械可読命令は細分化されて、ストレージデバイス及び／又はコンピューティングデバイス（例えばサーバ）の１又は複数に格納されてよい。機械可読命令は、コンピューティングデバイス及び／又は、他の機械でそれらを直接的に可読、変換可能及び／又は

実行可能にするために、インストール、変更、適合、更新、組み合わせ、補完、構成、復号化、解凍、アンパッキング、分配、再割り当て、コンパイル等の１又は複数が必要であってよい。例えば、機械可読命令は、複数部分に格納されてよく、それらは、個別に圧縮され、暗号化されかつ別個のコンピューティングデバイスに格納され、それらの部分は、復号化され、解凍され及び組み合わせされた場合に本明細書で記載されるようなプログラムに実装される実行可能命令のセットを形成する。

【０１０９】

別の例において、機械可読命令は、コンピュータによってそれらが読み出しえるが、特定のコンピューティングデバイス又は他のデバイスで命令を実行するために、ライブラリ（例えばダイナミックリンクライブラリ（ＤＬＬ））、ソフトウェア開発キット（ＳＤＫ）、アプリケーションプログラミングインターフェース（ＡＰＩ）等の追加を必要してよい状態で格納されてよい。別の例において、機械可読命令は、機械可読命令及び／又は対応するプログラムが全部又は一部実行され得る前に構成されること（例えば、格納された設定、データ入力、記録されたネットワークアドレス等）が必要でもよい。従って、開示された機械可読命令及び／又は対応するプログラムは、格納され又はそうでなければ残り若しくは送信中の場合の、機械可読命令及び／又はプログラムの特定のフォーマット又は状態に関わらず、そのような機械可読命令及び／又はプログラムを包含するよう意図される。

【０１１０】

本明細書で記載される機械可読命令は、過去、存在又は未来命令言語、スクリプト言語、プログラミング言語等のいずれかで表され得る。例えば、機械可読命令は以下の言語の任意のものを使用して表され得る、すなわち、Ｃ、Ｃ＋＋、Ｊａｖａ（登録商標）、Ｃ＃、ＰＥＲＬ、ＰＹＴＨＯＮ、ＪａｖａＳｃｒｉｐｔ（登録商標）、ハイパーテキストマークアップ言語（ＨＴＭＬ）、構造化照会言語（ＳＱＬ）、Ｓｗｉｆｔ等。

【０１１１】

上で言及したように、図８の例示的な処理は、ハードディスクドライブ、フラッシュメモリ、読み出し専用メモリ、コンパクトディスク、デジタルバーサタイルディスク、キャッシュ、ランダムアクセスメモリ、及び／又は、任意の期間（例えば拡張された時間の間、恒久的に、短い期間の間、一時的なバッファリングの間及び／又は情報のキャッシングの間）情報が格納されるその他のストレージデバイス及び／又はストレージディスクのような、非一時的コンピュータ及び／又は機械可読媒体に格納された実行可能命令（例えばコンピュータ及び／又は機械可読命令）を使用することを実装されてよい。本明細書で用いられるように、非一時的コンピュータ可読媒体という用語は、コンピュータ可読記憶装置デバイス及び／又はストレージディスクの任意のタイプを含むように、かつ、伝播する信号を除外しかつ送信媒体を除外するよう明示的に定義される。

【０１１２】

「含み」及び「備え」（及びそれらの全ての型及び時制）は本明細書において非限定的な用語に用いられている。従って、請求項がプリアンブルとして又は請求項内の任意の種類の記述において「含む」又は「備える」の任意の型（例えば、備える、含む、備え、含み、有し等）が用いるときにはいつでも、対応する請求項又は記述の範囲の外部に入ることなく、追加的な要素、用語等が存在してよいと理解されるべきである。本明細書で用いられている、「少なくとも」という語句は、例えば、請求項のプリアンブルの遷移用語として用いられる場合、それは「備え」及び「含み」が非限定的な用語であるのと同じ態様において非限定的である。「及び／又は」という用語は例えば、Ａ、Ｂ及び／又はＣのような型で用いられる場合、（１）Ａ単独、（２）Ｂ単独、（３）Ｃ単独、（４）ＡとＢ，（５）ＡとＣ、（６）ＢとＣ及び（７）ＡとＢとＣのような、Ａ，Ｂ，Ｃの任意の組み合わせ又はサブセットを指す。本明細書において、構造、コンポーネント、項目、オブジェクト及び／又は物を説明する文脈の中で用いられると、「Ａ及びＢのうちの少なくとも１つ」という語句は、（１）少なくとも１つのＡ、（２）少なくとも１つのＢ及び（３）少なくとも１つＡ及び少なくとも１つのＢのいずれかを含む実装を指すことを意図している

。同様に、本明細書において、構造、コンポーネント、項目、オブジェクト及び／又は物を説明する文脈の中で用いられると、「Ａ又はＢのうちの少なくとも１つ」という語句は、（１）少なくとも１つのＡ、（２）少なくとも１つのＢ及び（３）少なくとも１つＡ及び少なくとも１つのＢのいずれかを含む実装を指すことを意図している。本明細書において、処理、命令、動作、活動及び／又は段階の遂行又は実行を説明する文脈の中で用いられると、「Ａ及びＢのうちの少なくとも１つ」という語句は、（１）少なくとも１つのＡ、（２）少なくとも１つのＢ及び（３）少なくとも１つＡ及び少なくとも１つのＢのいずれかを含む実装を指すことを意図している。同様に、本明細書において、処理、命令、動作、活動及び／又は段階の遂行又は実行を説明する文脈の中で用いられると、「Ａ又はＢのうちの少なくとも１つ」という語句は、（１）少なくとも１つのＡ、（２）少なくとも１つのＢ及び（３）少なくとも１つＡ及び少なくとも１つのＢのいずれかを含む実装を指すことを意図している。

【０１１３】

本明細書で用いられると、単数の参照（例えば「ａ」、「ａｎ」、「第１」、「第２」等）は複数を除外していない。「ａ」又は「ａｎ」エンティティという用語は、本明細書で用いられると、そのエンティティの１又は複数を指す。「ａ」（又は「ａｎ」）、「１又は複数」、「少なくとも１つの」という用語は本明細書では同じ意味で用いられ得る。さらに、個別に列挙されているが、複数の手段、要素又は方法動作は、例えば単一のユニット又はプロセッサによって実装されてよい。更に、別個の特徴が異なる例又はクレイムに含まれていてよいが、これらはおそらく組み合わせられてよく、異なる例又はクレイムに含まれていることは特徴の組み合わせが実現可能でない及び／又は有利でないことを暗示するものではない。

【０１１４】

図８は、図５のスケジューラ５００及び／又は図６のスケジューラ６００を実装するために実行されてよい機械可読命令によって実装され得る処理８００を表すフローチャートである。処理８００はブロック８０２で開始し、そこにおいて、ワークロードインターフェース５０２は、スケジューラ５００及び／又はスケジューラ６００が関連付けられるところのＣＢＢに割り当てられたワークロードノードへの入力バッファ及び／又はそこからの出力バッファに対応するクレジットを、バッファクレジット記録装置５０４内へ読み込む。

【０１１５】

図８に図示される例において、処理８００はブロック８０４で継続し、そこにおいて、クレジット比較器５０６はスケジューラ５００及び／又はスケジューラ６００が関連付けられるところのＣＢＢに割り当てられたワークロードノードを選択する。ブロック８０６において、クレジット比較器５０６は、選択されたワークロードノードに対する入力バッファに格納されたデータを演算するためにスケジューラ５００及び／又はスケジューラ６００がクレジットの閾値量を受信しているか否かを判断する。例えば、クレジット比較器５０６は、外部デバイス（例えばクレジットマネージャー４０８、コントローラ３２２等）から受信したクレジット数に関連付けられた配列又は他のデータ構造内のフィールドを、選択されたワークロードノードへの入力バッファに対するクレジットの閾値数に関連付けられた配列又は他のデータ構造内のフィールドと比較する。もしスケジューラ５００及び／又はスケジューラ６００が選択されたワークロードノードに対する入力バッファに格納されているデータを演算するためにクレジットの閾値量を受信していない、とクレジット比較器５０６が判断したら（ブロック８０６：ＮＯ）、処理８００はブロック８１２に進む。

【０１１６】

図８の例において、もしスケジューラ５００及び／又はスケジューラ６００が入力バッファに格納されているデータを演算するためにクレジットの閾値量を受信している、とクレジット比較器５０６が判断したら（ブロック８０６：ＹＥＳ）、処理８００はブロック８０８に進む。ブロック８０８において、クレジット比較器５０６は、選択されたワーク

ロードノードに対する出力バッファにデータを書き込むためにスケジューラ５００及び／又はスケジューラ６００がクレジットの閾値量を受信しているか否かを判断する。例えば、クレジット比較器５０６は、選択されたワークロードノードに対する出力バッファに対する外部デバイス（例えばクレジットマネージャー４０８、コントローラ３２２等）から受信したクレジット数に関連付けられた配列又は他のデータ構造内のフィールドを、出力バッファに対するクレジットの閾値数に関連付けられた配列又は他のデータ構造内のフィールドと比較する。もしスケジューラ５００及び／又はスケジューラ６００がクレジットの閾値量を受信していない、とクレジット比較器５０６が判断したら（ブロック８０８：ＮＯ）、処理８００はブロック８１２に進む。もし出力バッファにデータを書き込むためにスケジューラ５００及び／又はスケジューラ６００がクレジットの閾値量を受信している、とクレジット比較器５０６が判断したら（ブロック８０８：ＹＥＳ）、クレジット比較器５０６は、選択されたワークロードノードの実行の準備ができたことをブロック８１０において示す。

【０１１７】

図８に図示した例において、ブロック８１２で、クレジット比較器５０６は処理すべき追加的なワークロードノードがあるか否かを判断する。もしクレジット比較器５０６が処理すべき追加的なワークロードノードがあると判断したなら（ブロック８１２：ＹＥＳ）、クレジット比較器５０６は追加的なワークロードノードを選択して、処理８００はブロック８０６に進む。もしクレジット比較器５０６が処理すべき追加的なワークロードノードがないと判断したなら（ブロック８１２：ＮＯ）、処理８００はブロック８１４に進む。

【０１１８】

図８に図示の例において、ブロック８１４で、ワークロードノードディスパッチャ５０８は、実行の準備ができたワークロードノードをスケジューリングする。ブロック８１６において、ワークロードノードディスパッチャ５０８はスケジュールに従ってワークロードノードをディスパッチする。ブロック８１８において、ディスパッチされたワークロードノードがスケジューラ５００及び／又はスケジューラ６００が関連付けられるところのＣＢＢによって実行されるにつれ、ワークロードインターフェース５０２は、入力バッファに関連付けられたクレジットを、そこからワークロードインターフェース５０２がクレジットを受信したところの外部デバイス（例えばクレジットマネージャー４０８、コントローラ３２２等）へ送信する。

【０１１９】

図８に図示した例において、ブロック８２０で、ワークロードノードディスパッチャ５０８は実行すべきスケジュール内に追加的なワークロードノードがあるかどうかを判断する。もしワークロードノードディスパッチャ５０８がスケジュール内に追加的なワークロードノードがあると判断したなら（ブロック８２０：ＹＥＳ）、処理８００はブロック８１６に進む。もしワークロードノードディスパッチャ５０８がスケジュールに実行すべき追加的なワークロードノードがないと判断したなら（ブロック８２０：ＮＯ）、処理８００はブロック８２２に進む。

【０１２０】

図８の例において、ブロック８２２で、ワークロードインターフェース５０２は演算を継続するか否かを判断する。例えば、ワークロードインターフェース５０２で演算を継続するとの判断が生じるであろう条件は、追加的なワークロードノードを受信することを含む。もしワークロードインターフェース５０２が演算を継続することを決定したら（ブロック８２２：ＹＥＳ）、処理８００はブロック８０２に進む。もしワークロードインターフェース５０２が演算を継続しないことを決定したら（ブロック８２２：ＮＯ）、処理８００は終了する。

【０１２１】

図９は、図５のスケジューラ５００及び／又は図６のスケジューラ６００の１又は複数のインスタンス化を実装するために図８の命令を実行するよう構築された例示的なプロセッサプラットフォーム９００のブロック図である。プロセッサプラットフォーム９００は

、例えば、サーバ、パーソナルコンピュータ、ワークステーション、自己学習機械（例えばニューラルネットワーク）、モバイルデバイス（例えばセルフォン、スマートフォン、ｉＰａｄ（登録商標）のようなタブレット）、パーソナルデジタルアシスタント（ＰＤＡ）、インターネット機器、ＤＶＤプレイヤ、ＣＤプレイヤ、デジタルビデオレコーダ、ブルーレイプレイヤ、ゲームコンソール、パーソナルビデオレコーダ、セットトップボックス、ヘッドセット若しくは他のウェアラブルデバイス、又は、任意のその他のタイプのコンピューティングデバイスであり得る。

【０１２２】

図示の例のプロセッサプラットフォーム９００は、プロセッサ９１０及びアクセラレータ９１２を含む。図示の例のプロセッサ９１０はハードウェアである。例えば、プロセッサ９１０は、任意の所望なファミリー若しくは製造者からの集積回路、ロジック回路、マイクロプロセッサ、ＧＰＵ、ＤＳＰ又はコントローラの１又は複数によって実装され得る。ハードウェアプロセッサは、半導体ベース（例えばシリコンベース）デバイスであってよい。更に、アクセラレータ９１２は、例えば、集積回路、ロジック回路、マイクロプロセッサ、ＧＰＵ、ＤＳＰ、ＦＰＧＡ、ＶＰＵ、コントローラ、及び／又は、任意の所望なファミリー若しくは製造者からの他のＣＢＢの１又は複数により実装され得る。図示の例のアクセラレータ９１２はハードウェアである。ハードウェアアクセラレータは、半導体ベース（例えばシリコンベース）デバイスであってよい。この例において、アクセラレータ９１２は、例示的な畳み込みエンジン３１２、例示的なＲＮＮエンジン３１４、例示的なメモリ３１６、例示的なＭＭＵ３１８、例示的なＤＳＰ３２０、例示的なコントローラ３２２及び例示的なＤＭＡユニット３２４を実装する。さらに、例示的な畳み込みエンジン３１２、例示的なＲＮＮエンジン３１４、例示的なＤＭＡユニット３２４、例示的なＤＳＰ３２０及び例示的なコントローラ３２２の各々は、例示的な第１スケジューラ３２６、例示的な第２スケジューラ３２８、例示的な第３スケジューラ３３０、例示的な第４スケジューラ３３２及び例示的な第５スケジューラ３３４をそれぞれ含む。図９の例において、例示的な第１スケジューラ３２６、例示的な第２スケジューラ３２８、例示的な第３スケジューラ３３０、例示的な第４スケジューラ３３２及び例示的な第５スケジューラ３３４の各々は、例示的なワークロードインターフェース５０２、例示的なバッファクレジット格納装置５０４、例示的なクレジット比較器５０６、例示的なワークロードノードディスパッチャ５０８、及び／又は、より一般的に、スケジューラ５００を含む。

【０１２３】

追加的又は代替的な例において、プロセッサ９１０は、例示的な畳み込みエンジン３１２、例示的なＲＮＮエンジン３１４、例示的なメモリ３１６、例示的なＭＭＵ３１８、例示的なＤＳＰ３２０、例示的なコントローラ３２２及び例示的なＤＭＡユニット３２４を実装する。さらにこのような追加的又は代替的な例において、例示的な畳み込みエンジン３１２、例示的なＲＮＮエンジン３１４、例示的なＤＭＡユニット３２４、例示的なＤＳＰ３２０及び例示的なコントローラ３２２の各々は、例示的な第１スケジューラ３２６、例示的な第２スケジューラ３２８、例示的な第３スケジューラ３３０、例示的な第４スケジューラ３３２及び例示的な第５スケジューラ３３４をそれぞれ含む。このような追加的又は代替的な例において、例示的な第１スケジューラ３２６、例示的な第２スケジューラ３２８、例示的な第３スケジューラ３３０、例示的な第４スケジューラ３３２及び例示的な第５スケジューラ３３４の各々は、例示的なワークロードインターフェース５０２、例示的なバッファクレジット格納装置５０４、例示的なクレジット比較器５０６、例示的なワークロードノードディスパッチャ５０８、及び／又は、より一般的に、スケジューラ５００を含む。

【０１２４】

図示の例のプロセッサ９１０はローカルメモリ９１１（例えばキャッシュ）を含む。図示の例のプロセッサ９１０は、バス９１８で揮発性メモリ９１４及び不揮発性メモリ９１６を含むメインメモリと通信する。さらには図示の例のアクセラレータ９１２は、ローカルメモリ９１３（例えばキャッシュ）を含む。図示の例のアクセラレータ９１２は、バス

９１８で揮発性メモリ９１４及び不揮発性メモリ９１６を含むメインメモリと通信する。揮発性メモリ９１４は、同期ダイナミックランダムアクセスメモリ（ＳＤＲＡＭ）、ダイナミックランダムアクセスメモリ（ＤＲＡＭ）、ＲＡＭＢＵＳ（登録商標）ダイナミックランダムアクセスメモリ（ＲＤＲＡＭ（登録商標））及び／又は任意のその他のタイプのアクセスメモリデバイスによって実装されてよい。不揮発性メモリ９１６は、フラッシュメモリ及び／又はその他の所望の任意のタイプのメモリデバイスによって実装されてよい。メインメモリ９１４、９１６へのアクセスはメモリコントローラによって制御される。

【０１２５】

図示の例のプロセッサプラットフォーム９００はまたインターフェース回路９２０をも含む。インターフェース回路９２０は、Ｅｔｈｅｒｎｅｔ（登録商標）インターフェース、ユニバーサルシリアルバス（ＵＳＢ）、Ｂｌｕｅｔｏｏｔｈ（登録商標）インターフェース，近距離無線通信（ＮＦＣ）インターフェース、及び／又は、ＰＣＩエクスプレスインターフェースのような任意のタイプのインターフェース規格によって実装されてよい。

【０１２６】

図示の例において、１又は複数入力デバイス９２２はインターフェース回路９２０に接続される。入力デバイス９２２は、ユーザにデータ及び／又はコマンドをプロセッサ９１０及び／又はアクセラレータ９１２内へ入力させるのを可能にする。入力デバイスは、例えば、オーディオセンサ、マイク、カメラ（静止画又は動画）、キーボード、ボタン、マウス、タッチスクリーン、トラックパッド、トラックボール、ｉｓｏｐｏｉｎｔ及び／又は音声認識システムによって実装され得る。

【０１２７】

１又は複数出力デバイス９２４もまた、図示の例のインターフェース回路９２０に接続される。出力デバイス９２４は、例えば、ディスプレイデバイス（例えば発光ダイオード（ＬＥＤ）、有機発光ダイオード（ＯＬＥＤ）、液晶ディスプレイ（ＬＣＤ）、カソードレイ管ディスプレイ（ＣＲＴ），面内スイッチング（ＩＰＳ）ディスプレイ、タッチスクリーン等）、触覚出力デバイス、プリンタ及び／又はスピーカによって実装され得る。図示の例のインターフェース回路９２０は従って、典型的には、グラフィックドライバカード、グラフィックドライバチップ及び／又はグラフィックドライバプロセッサを含む。

【０１２８】

図示の例のインターフェース回路９２０はまた、送信機、受信機、トランシーバ、モデム、住宅ゲートウェイ、無線アクセスポイント、及び／又は、ネットワーク９２６を介した外部の機械（例えば任意の種類のコンピューティングデバイス）とのデータの交換を促進するネットワークインターフェースのような通信デバイスを含む。通信は、例えば、Ｅｔｈｅｒｎｅｔ（登録商標）接続、デジタル加入者ライン（ＤＳＬ）接続、電話線接続、同軸ケーブルシステム、衛星システム、ラインオブサイト無線システム、セルラ電話システム等を介し得る。

【０１２９】

図示の例のプロセッサプラットフォーム９００はまた、ソフトウェア及び／又はデータを格納するための１又は複数の大容量ストレージデバイス９２８を含む。このような大容量ストレージデバイス９２８の例は、フロッピーディスクドライブ、ハードドライブディスク、コンパクトディスクドライブ、ブルーレイディスクドライブ、独立ディスクの冗長アレイ（ＲＡＩＤ）システム、デジタルバーサタイルディスク（ＤＶＤ）ドライブを含む。

【０１３０】

図８の機械実行可能命令９３２は、大容量ストレージデバイス９２８内、揮発性メモリ９１４内、不揮発性メモリ９１６内、及び／又は、ＣＤ又はＤＶＤのようなリムーバブル非一時的コンピュータ可読記憶媒体上に格納されてよい。

【０１３１】

上記から、ワークロードのスタティックマッピングの順不同にパイプライン化された実行が可能に、例示的な方法、装置及び製造物が開示されていることが理解されるであろう。さらに、ワークロードノードが依存するところのデータが利用可能であり、かつ、ワー

クロードノードの実行によって生成された出力を格納するのに利用可能な十分なメモリがある場合に、計算ビルディングブロックがワークロードノードを実行することが可能であるように、例示的な方法、装置及び製造物が開示されている。更に、本明細書で開示された例は、スケジュール及び／又は他の順序から独立してワークロードノードが割り当てられるところの計算ビルディングブロックによってワークロードノードが実行されることを可能にする。開示された方法、装置及び製造物は、処理デバイスの利用を増加することによってコンピューティングデバイスの使用の効率性を向上する。さらに、本明細書で開示された例示的な方法、装置及び製造物は、ワークロードを処理及び／又はそうでなければ実行するために処理デバイスによって用いられる計算サイクルの数を減少させる。従って、開示された方法、装置及び製造物は、コンピュータの機能を１又は複数改善するよう方向付けられている。

【０１３２】

ワークロードのスタティックマッピングの順不同にパイプライン化された実行を可能にする例示的な方法、装置、システム及び製造物が本明細書に開示されている。さらなる例及びそれらの組み合わせは以下のものを含む：例１は、クレジットの第１の数をメモリ内へ読み込むインターフェースと、クレジットの第１の数をバッファ内のメモリ利用可能性に関連付けられたクレジットの閾値数と比較する比較器と、クレジットの第１の数がクレジットの閾値数を満たす場合に、１又は複数の計算ビルディングブロックの最初の一つで実行されるワークロードのワークロードノードを選択するディスパッチャとを備える装置を含む。

【０１３３】

例２は、例１の装置を含み、インターフェースは、インターフェースがクレジットマネージャーからクレジットの第１の数を受信した場合にクレジットの第１の数をメモリ内へ読み込み、ワークロードノードに関連付けられたデータの１又は複数のタイルが１又は複数の計算ビルディングブロックの最初の一つからバッファへ送信されるにつれ、バッファに送信された各タイルに対してクレジットをクレジットマネージャーに送信するものである。

【０１３４】

例３は例１の装置を含み、バッファはワークロードノードに関連付けられた出力バッファであり、クレジットの第１の数は出力バッファに対応しており、クレジットの閾値数は出力バッファ内のメモリの閾値量に対応する。

【０１３５】

例４は例１の装置を含み、バッファはワークロードノードに関連付けられた入力バッファであり、クレジットの第１の数は入力バッファに対応しており、クレジットの閾値数は入力バッファ内のデータの閾値量に対応する。

【０１３６】

例５は例１の装置を含み、バッファは第１バッファであり、クレジットの閾値数はクレジットの第１閾値数であり、比較器はクレジットの第２の数を第２バッファ内のメモリ利用可能性に関連付けられたクレジットの第２閾値数と比較するものであり、ディスパッチャは、（１）クレジットの第１の数がクレジットの第１閾値数を満たしし、（２）クレジットの第２の数がクレジットの第２閾値数を満たす場合に、１又は複数の計算ビルディングブロックの最初の一つで実行されるワークロードノードを選択するものである。

【０１３７】

例６は例５の装置を含み、第２バッファはワークロードノードに関連付けられた入力バッファであり、クレジットの第２の数は入力バッファに対応しており、クレジットの第２閾値数は入力バッファ内のデータの閾値量に対応する。

【０１３８】

例７は例１の装置を含み、クレジットの閾値数はクレジットの第１閾値数であり、ワークロードノードは第１ワークロードノードであり、（１）クレジットの第１の数がクレジットの第１閾値数と満たし、（２）クレジットの第２の数がクレジットの第２閾値数を満

たす場合に、ディスパッチャは、１又は複数の計算ビルディングブロックの最初の一つで実行される第１ワークロードノード及び第２ワークロードノードをスケジューリングするものである。

【０１３９】

例８は命令を備える非一時的コンピュータ可読記憶媒体を含み、命令は実行された場合に、少なくとも１つのプロセッサに、クレジットの第１の数をメモリ内へ読み込むこと、クレジットの第１の数をバッファ内のメモリ利用可能性に関連付けられたクレジットの閾値数と比較すること、及び、クレジットの第１の数がクレジットの閾値数を満たす場合に、計算ビルディングブロックで実行されるワークロードのワークロードノードを選択することを少なくとも生じせしめる。

【０１４０】

例９は例８の非一時的コンピュータ可読記憶媒体を含み、命令は実行された場合に、少なくとも１つのプロセッサに、クレジットの第１の数がクレジットマネージャーから受信された場合にクレジットの第１の数をメモリ内へ読み込むこと、及び、ワークロードノードに関連付けられたデータの１又は複数のタイルが計算ビルディングブロックからバッファへ送信されるにつれ、バッファに送信された各タイルに対してクレジットをクレジットマネージャーに送信することを生じせしめる。

【０１４１】

例１０は例８の非一時的コンピュータ可読記憶媒体を含み、バッファはワークロードノードに関連付けられた出力バッファであり、クレジットの第１の数は出力バッファに対応しており、クレジットの閾値数は出力バッファ内のメモリの閾値量に対応する。

【０１４２】

例１１は例８の非一時的コンピュータ可読記憶媒体を含み、バッファはワークロードノードに関連付けられた入力バッファであり、クレジットの第１の数は入力バッファに対応しており、クレジットの閾値数は入力バッファ内のデータの閾値量に対応する。

【０１４３】

例１２は例８の非一時的コンピュータ可読記憶媒体を含み、バッファは第１バッファであり、クレジットの閾値数はクレジットの第１閾値数であり、かつ、命令は実行された場合に、少なくとも１つのプロセッサに、クレジットの第２の数を第２バッファ内のメモリ利用可能性に関連付けられたクレジットの第２閾値数と比較すること、（１）クレジットの第１の数がクレジットの第１閾値数を満たし、（２）クレジットの第２の数がクレジットの第２閾値数を満たす場合に、計算ビルディングブロックで実行されるワークロードノードを選択することを生じせしめる。

【０１４４】

例１３は例１２の非一時的コンピュータ可読記憶媒体を含み、第２バッファはワークロードノードに関連付けられた入力バッファであり、クレジットの第２の数は第２バッファに対応しており、クレジットの第２閾値数は入力バッファ内のデータの閾値量に対応する。

【０１４５】

例１４は例８の非一時的コンピュータ可読記憶媒体を含み、クレジットの閾値数はクレジットの第１閾値数であり、ワークロードノードは第１ワークロードノードであり、命令は実行された場合に、少なくとも１つのプロセッサに、（１）クレジットの第１の数がクレジットの第１閾値数を満たし、（２）クレジットの第２の数がクレジットの第２閾値数を満たす場合に、計算ビルディングブロックで実行される第１ワークロードノード及び第２ワークロードノードをスケジューリングすることを生じせしめる。

【０１４６】

例１５は、インターフェースする手段であって、クレジットの第１の数をメモリ内へ読み込むためのインターフェースする手段と、比較する手段であって、クレジットの第１の数をバッファ内のメモリ利用可能性に関連付けられたクレジットの閾値数と比較するための比較する手段と、ディスパッチする手段であって、クレジットの第１の数がクレジットの閾値数を満たす場合に、１又は複数の計算ビルディングブロックの最初の一つで実行さ

れるワークロードのワークロードノードを選択するためのディスパッチする手段とを備える装置を含む。

【０１４７】

例１６は例１５の装置を含み、インターフェースする手段は、インターフェースする手段がクレジットマネージャーからクレジットの第１の数を受信した場合にクレジットの第１の数をメモリ内へ読み込み、ワークロードノードに関連付けられたデータの１又は複数のタイルが１又は複数の計算ビルディングブロックの最初の一つからバッファへ送信されるにつれ、バッファに送信された各タイルに対してクレジットをクレジットマネージャーに送信するものである。

【０１４８】

例１７は例１５の装置を含み、バッファはワークロードノードに関連付けられた出力バッファであり、クレジットの第１の数は出力バッファに対応しており、クレジットの閾値数は出力バッファ内のメモリの閾値量に対応する。

【０１４９】

例１８は例１５の装置を含み、バッファはワークロードノードに関連付けられた入力バッファであり、クレジットの第１の数は入力バッファに対応しており、クレジットの閾値数は入力バッファ内のデータの閾値量に対応する。

【０１５０】

例１９は例１５の装置を含み、バッファは第１バッファであり、クレジットの閾値数はクレジットの第１閾値数であり、比較する手段はクレジットの第２の数を第２バッファ内のメモリ利用可能性に関連付けられたクレジットの第２閾値数と比較するものであり、ディスパッチする手段は、（１）クレジットの第１の数がクレジットの第１閾値数を満たし、（２）クレジットの第２の数がクレジットの第２閾値数を満たす場合に、１又は複数の計算ビルディングブロックの最初の一つで実行されるワークロードノードを選択するものである。

【０１５１】

例２０は例１９の装置を含み、第２バッファはワークロードノードに関連付けられた入力バッファであり、クレジットの第２の数は入力バッファに対応しており、クレジットの第２閾値数は入力バッファ内のデータの閾値量に対応する。

【０１５２】

例２１は例１５の装置を含み、クレジットの閾値数はクレジットの第１閾値数であり、ワークロードノードは第１ワークロードノードであり、（１）クレジットの第１の数がクレジットの第１閾値数を満たし、（２）クレジットの第２の数がクレジットの第２閾値数を満たす場合に、ディスパッチする手段は、１又は複数の計算ビルディングブロックの最初の一つで実行される第１ワークロードノード及び第２ワークロードノードをスケジューリングするものである。

【０１５３】

例２２は、クレジットの第１の数をメモリ内へ読み込むことと、クレジットの第１の数をバッファ内のメモリ利用可能性に関連付けられたクレジットの閾値数と比較することと、クレジットの第１の数がクレジットの閾値数を満たす場合に、１又は複数の計算ビルディングブロックの最初の一つで実行されるワークロードのワークロードノードを選択することとを備える方法を含む。

【０１５４】

例２３は、例２２の方法を含み、クレジットマネージャーからクレジットの第１の数を受信した場合にクレジットの第１の数をメモリ内へ読み込むことと、ワークロードノードに関連付けられたデータの１又は複数のタイルが１又は複数の計算ビルディングブロックの最初の一つからバッファへ送信されるにつれ、バッファに送信された各タイルに対してクレジットをクレジットマネージャーに送信することとをさらに含む。

【０１５５】

例２４は例２２の方法を含み、バッファはワークロードノードに関連付けられた出力バ

ッファであり、クレジットの第１の数は出力バッファに対応しており、クレジットの閾値数は出力バッファ内のメモリの閾値量に対応する。

【０１５６】

例２５は例２２の方法を含み、バッファはワークロードノードに関連付けられた入力バッファであり、クレジットの第１の数は入力バッファに対応しており、クレジットの閾値数は入力バッファ内のデータの閾値量に対応する。

【０１５７】

特定の例示的な方法、装置及び製造物が本明細書で開示されているが、この特許のカバレッジの範囲はそれに限定されない。反対に、この特許は、この特許の請求項の範囲内に適正に含まれる全ての方法、装置及び製造物を包含する。

【０１５８】

以下の請求項は本明細書においてこの参照により本詳細な説明に組み込まれ、各請求項は本開示の別個の実施形態を独自に代表する。

他の可能な項目

［項目１］

クレジットの第１の数をメモリ内へ読み込むインターフェースと、

クレジットの上記第１の数をバッファ内のメモリ利用可能性に関連付けられたクレジットの閾値数と比較する比較器と、

クレジットの上記第１の数がクレジットの上記閾値数を満たす場合に、上記１又は複数の計算ビルディングブロックの最初の一つで実行される上記ワークロードのワークロードノードを選択するディスパッチャと

を備える装置。

［項目２］

上記インターフェースは、

上記インターフェースがクレジットマネージャーからクレジットの上記第１の数を受信した場合にクレジットの上記第１の数をメモリ内へ読み込み、

上記ワークロードノードに関連付けられたデータの１又は複数のタイルが上記１又は複数の計算ビルディングブロックの上記最初の一つから上記バッファへ送信されるにつれ、上記バッファに送信された各タイルに対してクレジットを上記クレジットマネージャーに送信するものである

項目１の装置。

［項目３］

上記バッファは上記ワークロードノードに関連付けられた出力バッファであり、クレジットの上記第１の数は上記出力バッファに対応しており、クレジットの上記閾値数は上記出力バッファ内のメモリの閾値量に対応する項目１の装置。

［項目４］

上記バッファは上記ワークロードノードに関連付けられた入力バッファであり、クレジットの上記第１の数は上記入力バッファに対応しており、クレジットの上記閾値数は上記入力バッファ内のデータの閾値量に対応する項目１の装置。

［項目５］

上記バッファは第１バッファであり、クレジットの上記閾値数はクレジットの第１閾値数であり、上記比較器はクレジットの第２の数を第２バッファ内のメモリ利用可能性に関連付けられたクレジットの第２閾値数と比較するものであり、上記ディスパッチャは、（１）クレジットの上記第１の数がクレジットの上記第１閾値数を満たし、（２）クレジットの上記第２の数がクレジットの上記第２閾値数を満たす場合に、上記１又は複数の計算ビルディングブロックの上記最初の一つで実行される上記ワークロードノードを選択するものである項目１の装置。

［項目６］

上記第２バッファは上記ワークロードノードに関連付けられた入力バッファであり、クレジットの上記第２の数は上記入力バッファに対応しており、クレジットの上記第２閾値

数は上記入力バッファ内のデータの閾値量に対応する項目５の装置。
［項目７］
クレジットの上記閾値数はクレジットの第１閾値数であり、上記ワークロードノードは第１ワークロードノードであり、（１）クレジットの上記第１の数がクレジットの上記第１閾値数を満たし、（２）クレジットの第２の数がクレジットの第２閾値数を満たす場合に、上記ディスパッチャは、上記１又は複数の計算ビルディングブロックの上記最初の一つで実行される上記第１ワークロードノード及び第２ワークロードノードをスケジューリングするものである項目１の装置。
［項目８］
実行された場合に、少なくとも１つのプロセッサに、
クレジットの第１の数をメモリ内へ読み込むこと、
クレジットの上記第１の数をバッファ内のメモリ利用可能性に関連付けられたクレジットの閾値数と比較すること、及び、
クレジットの上記第１の数がクレジットの上記閾値数を満たす場合に、計算ビルディングブロックで実行される上記ワークロードのワークロードノードを選択すること
を少なくとも生じせしめる命令を備える非一時的コンピュータ可読記憶媒体。
［項目９］
上記命令は実行された場合に、上記少なくとも１つのプロセッサに、
クレジットの上記第１の数がクレジットマネージャーから受信された場合にクレジットの上記第１の数をメモリ内へ読み込むこと、及び、
上記ワークロードノードに関連付けられたデータの１又は複数のタイルが上記計算ビルディングブロックから上記バッファへ送信されるにつれ、上記バッファに送信された各タイルに対してクレジットを上記クレジットマネージャーに送信することを生じせしめる項目８の非一時的コンピュータ可読記憶媒体。
［項目１０］
上記バッファは上記ワークロードノードに関連付けられた出力バッファであり、クレジットの上記第１の数は上記出力バッファに対応しており、クレジットの上記閾値数は上記出力バッファ内のメモリの閾値量に対応する項目８の非一時的コンピュータ可読記憶媒体。
［項目１１］
上記バッファは上記ワークロードノードに関連付けられた入力バッファであり、クレジットの上記第１の数は上記入力バッファに対応しており、クレジットの上記閾値数は上記入力バッファ内のデータの閾値量に対応する項目８の非一時的コンピュータ可読記憶媒体。
［項目１２］
上記バッファは第１バッファであり、クレジットの上記閾値数はクレジットの第１閾値数であり、かつ、上記命令は実行された場合に、上記少なくとも１つのプロセッサに、
クレジットの第２の数を第２バッファ内のメモリ利用可能性に関連付けられたクレジットの第２閾値数と比較すること、
（１）クレジットの上記第１の数がクレジットの上記第１閾値数を満たし、（２）クレジットの上記第２の数がクレジットの上記第２閾値数を満たす場合に、上記計算ビルディングブロックで実行される上記ワークロードノードを選択すること
を生じせしめる項目８の非一時的コンピュータ可読記憶媒体。
［項目１３］
上記第２バッファは上記ワークロードノードに関連付けられた入力バッファであり、クレジットの上記第２の数は上記第２バッファに対応しており、クレジットの第２閾値数は上記入力バッファ内のデータの閾値量に対応する項目１２の非一時的コンピュータ可読記憶媒体。
［項目１４］
クレジットの上記閾値数はクレジットの第１閾値数であり、上記ワークロードノードは第１ワークロードノードであり、命令は実行された場合に、上記少なくとも１つのプロセッサに、（１）クレジットの上記第１の数がクレジットの上記第１閾値数を満たし、（２

）クレジットの第２の数がクレジットの第２閾値数を満たす場合に、上記計算ビルディングブロックで実行される上記第１ワークロードノード及び第２ワークロードノードをスケジューリングすることを生じせしめる項目８の非一時的コンピュータ可読記憶媒体。

［項目１５］

インターフェースする手段であって、クレジットの第１の数をメモリ内へ読み込むための上記インターフェースする手段と、

比較する手段であって、クレジットの上記第１の数をバッファ内のメモリ利用可能性に関連付けられたクレジットの閾値数と比較するための上記比較する手段と、

ディスパッチする手段であって、クレジットの上記第１の数がクレジットの上記閾値数に一致する場合に、上記１又は複数の計算ビルディングブロックの最初の一つで実行される上記ワークロードのワークロードノードを選択するための上記ディスパッチする手段とを備える装置。

［項目１６］

上記インターフェースする手段は、

上記インターフェースする手段がクレジットマネージャーからクレジットの上記第１の数を受信した場合にクレジットの上記第１の数をメモリ内へ読み込み、

上記ワークロードノードに関連付けられたデータの１又は複数のタイルが上記１又は複数の計算ビルディングブロックの上記最初の一つから上記バッファへ送信されるにつれ、上記バッファに送信された各タイルに対してクレジットを上記クレジットマネージャーに送信する

ものである項目１５の装置。

［項目１７］

上記バッファは上記ワークロードノードに関連付けられた出力バッファであり、クレジットの上記第１の数は上記出力バッファに対応しており、クレジットの上記閾値数は上記出力バッファ内のメモリの閾値量に対応する項目１５の装置。［項目１８］上記バッファは上記ワークロードノードに関連付けられた入力バッファであり、クレジットの上記第１の数は上記入力バッファに対応しており、クレジットの上記閾値数は上記入力バッファ内のデータの閾値量に対応する項目１５の装置。

［項目１９］

上記バッファは第１バッファであり、クレジットの上記閾値数はクレジットの第１閾値数であり、上記比較する手段はクレジットの第２の数を第２バッファ内のメモリ利用可能性に関連付けられたクレジットの第２閾値数と比較するものであり、上記ディスパッチする手段は、（１）クレジットの上記第１の数がクレジットの上記第１閾値数を満たし、（２）クレジットの上記第２の数がクレジットの上記第２閾値数を満たす場合に、上記１又は複数の計算ビルディングブロックの上記最初の一つで実行される上記ワークロードノードを選択するものである項目１５の装置。

［項目２０］

上記第２バッファは上記ワークロードノードに関連付けられた入力バッファであり、クレジットの上記第２の数は上記入力バッファに対応しており、クレジットの上記第２閾値数は上記入力バッファ内のデータの閾値量に対応する項目１９の装置。

［項目２１］

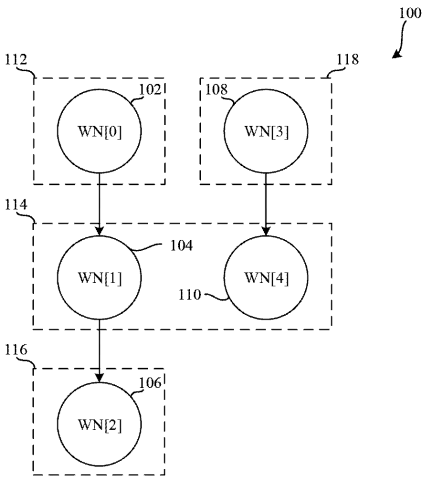
クレジットの上記閾値数はクレジットの第１閾値数であり、上記ワークロードノードは第１ワークロードノードであり、（１）クレジットの上記第１の数がクレジットの上記第１閾値数を満たし、（２）クレジットの第２の数がクレジットの第２閾値数を満たす場合に、上記ディスパッチする手段は、上記１又は複数の計算ビルディングブロックの上記最初の一つで実行される上記第１ワークロードノード及び第２ワークロードノードをスケジューリングするものである項目１５の装置。

［項目２２］

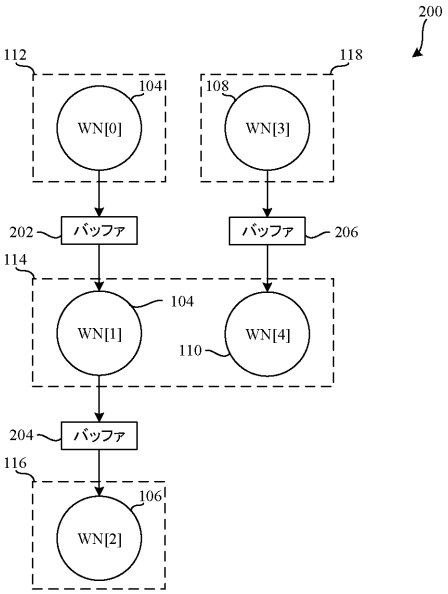
クレジットの第１の数をメモリ内へ読み込むことと、

クレジットの上記第１の数をバッファ内のメモリ利用可能性に関連付けられたクレジッ

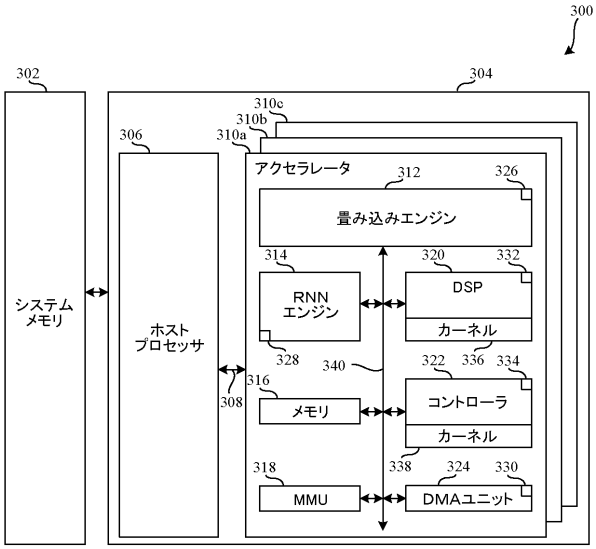
トの閾値数と比較することと、
クレジットの上記第１の数がクレジットの上記閾値数に一致する場合に、上記１又は複数の計算ビルディングブロックの最初の一つで実行される上記ワークロードのワークロードノードを選択することと
を備える方法。
［項目２３］
クレジットマネージャーからクレジットの上記第１の数を受信する場合にクレジットの上記第１の数をメモリ内へ読み込むことと、
上記ワークロードノードに関連付けられたデータの１又は複数のタイルが上記１又は複数の計算ビルディングブロックの上記最初の一つから上記バッファへ送信されるにつれ、上記バッファに送信された各タイルに対してクレジットを上記クレジットマネージャーに送信することと
をさらに含む項目２２の方法。
［項目２４］
上記バッファは上記ワークロードノードに関連付けられた出力バッファであり、クレジットの上記第１の数は上記出力バッファに対応しており、クレジットの上記閾値数は上記出力バッファ内のメモリの閾値量に対応する項目２２の方法。
［項目２５］
上記バッファは上記ワークロードノードに関連付けられた入力バッファであり、クレジットの上記第１の数は上記入力バッファに対応しており、クレジットの上記閾値数は上記入力バッファ内のデータの閾値量に対応する項目２２の方法。
【図面】
【図１】



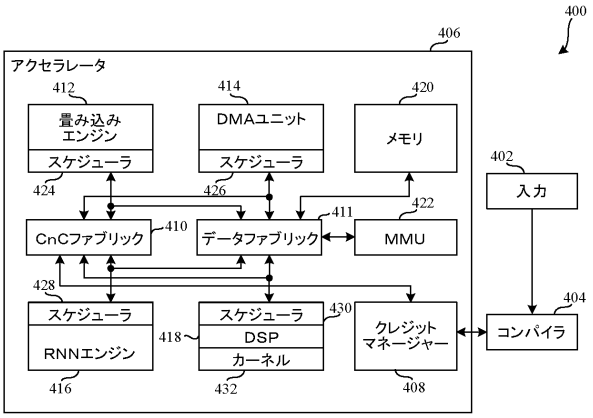
【図２】



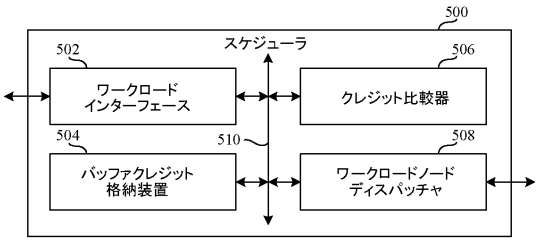
【図３】


300
302
304
310c
310b
310a
306
アクセラレータ
312
326
畳み込みエンジン
314
320
332
RNN
エンジン
DSP
カーネル
328
340
322
336
334
316
308
メモリ
コントローラ
カーネル
318
338
324
330
MMU
DMAユニット
システム
メモリ
ホスト
プロセッサ

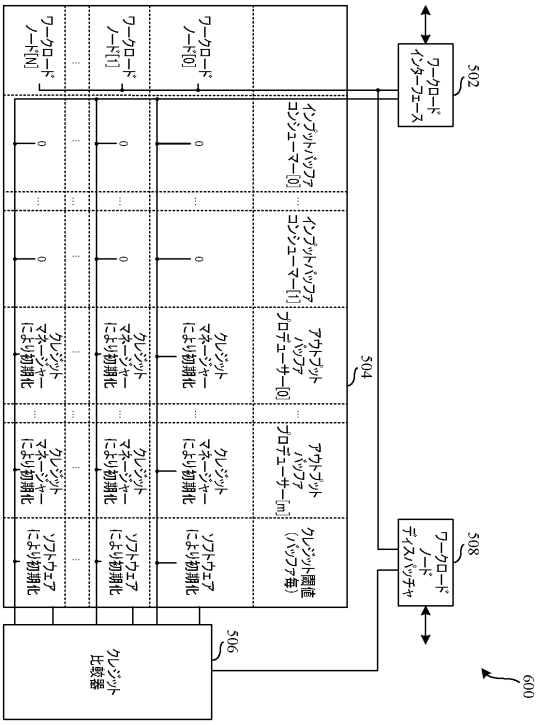
【図４】


400
406
アクセラレータ
412
414
420
畳み込み
エンジン
DMAユニット
メモリ
スケジューラ
スケジューラ
402
424
426
入力
422
411
CnCファブリック
410
データファブリック
MMU
428
404
スケジューラ
スケジューラ
430
クレジット
マネージャー
コンパイラ
418
DSP
RNNエンジン
カーネル
416
432
408

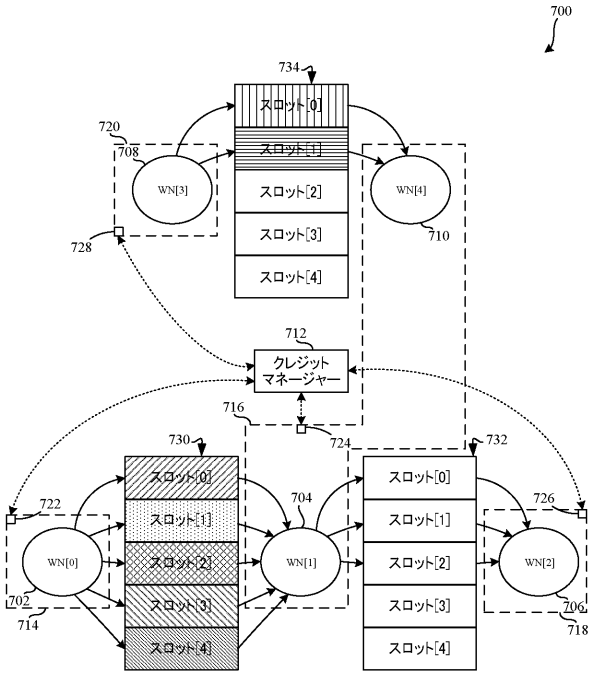
【図５】


500
スケジューラ
502
506
ワークロード
インターフェース
クレジット比較器
504
510
508
バッファクレジット
格納装置
ワークロードノード
ディスパッチャ

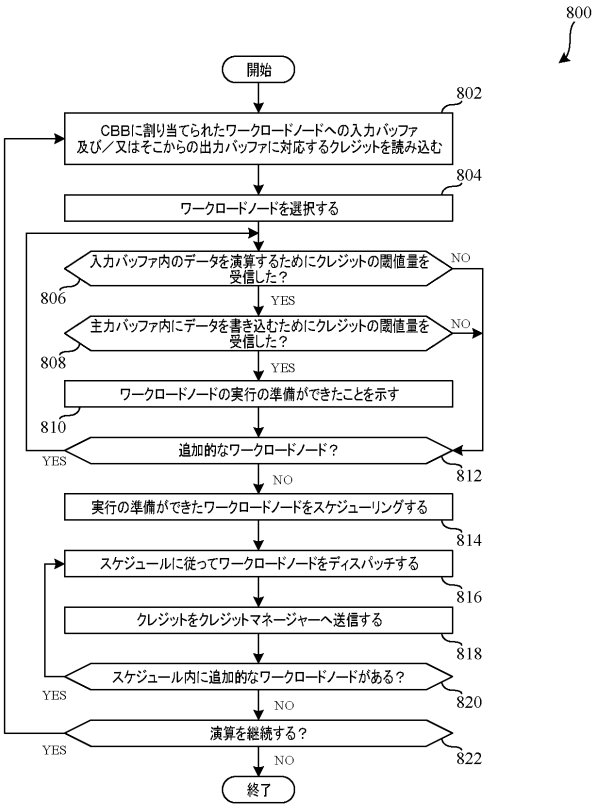
【図６】


ワークロード
ノード[N]
ワークロード
ノード[1]
ワークロード
ノード[0]
ワークロード
インターフェース
502
インプットバッファ
コンシューマー[0]
インプットバッファ
コンシューマー[1]
クレジット
マネージャー
により初期化
アウトプット
バッファ
プロデューサー[0]
アウトプット
バッファ
プロデューサー[m]
504
ソフトウェア
により初期化
クレジット閾値
(バッファ毎)
ワークロード
ノード
ディスパッチャ
508
クレジット
比較器
506
600

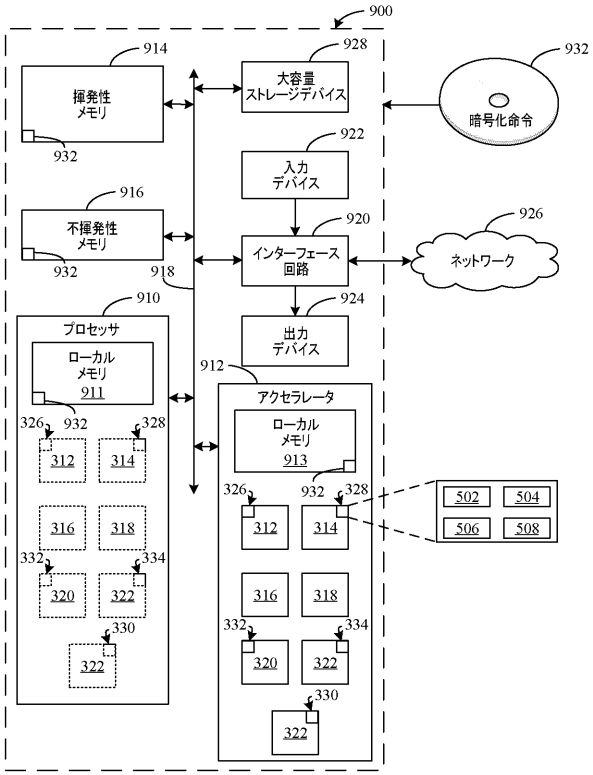
【図７】


700
734
スロット[0]
スロット[1]
スロット[2]
スロット[3]
スロット[4]
720
708
WN[3]
WN[4]
710
728
712
クレジット
マネージャー
716
724
730
732
722
704
726
WN[0]
WN[1]
WN[2]
702
706
714
718

【図８】


800
開始
802
CBBに割り当てられたワークロードノードへの入力バッファ
及び／又はそこからの出力バッファに対応するクレジットを読み込む
804
ワークロードノードを選択する
入力バッファ内のデータを演算するためにクレジットの閾値量を受信した？
NO
806
YES
主力バッファ内にデータを書き込むためにクレジットの閾値量を受信した？
NO
808
YES
ワークロードノードの実行の準備ができたことを示す
810
追加的なワークロードノード？
YES
812
NO
実行の準備ができたワークロードノードをスケジューリングする
814
スケジュールに従ってワークロードノードをディスパッチする
816
クレジットをクレジットマネージャーへ送信する
818
スケジュール内に追加的なワークロードノードがある？
YES
820
NO
演算を継続する？
YES
822
NO
終了

【図９】


900
914
揮発性
メモリ
932
928
大容量
ストレージデバイス
932
暗号化命令
922
入力
デバイス
916
不揮発性
メモリ
932
920
インターフェース
回路
926
ネットワーク
918
910
924
出力
デバイス
プロセッサ
ローカル
メモリ
911
912
アクセラレータ
ローカル
メモリ
913
326
328
312
314
316
318
320
322
330
332
334
502
504
506
508

フロントページの続き

レッジ ブーレバード・２２００ インテル・コーポレーション内
(72)発明者 ロネン ガバイ
アメリカ合衆国 ９５０５４ カリフォルニア州・サンタクララ・ミッション カレッジ ブーレバード・２２００ インテル・コーポレーション内
(72)発明者 ロニ ロスナー
アメリカ合衆国 ９５０５４ カリフォルニア州・サンタクララ・ミッション カレッジ ブーレバード・２２００ インテル・コーポレーション内
(72)発明者 ジギ ウォルター
アメリカ合衆国 ９５０５４ カリフォルニア州・サンタクララ・ミッション カレッジ ブーレバード・２２００ インテル・コーポレーション内
(72)発明者 オレン アガム
アメリカ合衆国 ９５０５４ カリフォルニア州・サンタクララ・ミッション カレッジ ブーレバード・２２００ インテル・コーポレーション内

審査官 坂東 博司
(56)参考文献 特開２００５－０１８６２０（ＪＰ，Ａ）
米国特許出願公開第２００４／０２６８０８３（ＵＳ，Ａ１）
特表２０１７－５２５０４７（ＪＰ，Ａ）
特開平０７－０２１１４４（ＪＰ，Ａ）
特開２０１２－１９０４１５（ＪＰ，Ａ）
米国特許出願公開第２０１２／０２３９８３３（ＵＳ，Ａ１）
米国特許出願公開第２０１６／０１４００７１（ＵＳ，Ａ１）
米国特許出願公開第２０１９／００５０２６１（ＵＳ，Ａ１）
(58)調査した分野 (Int.Cl.，ＤＢ名)
Ｇ０６Ｆ ９／５０
Ｇ０６Ｆ ９／３８
Ｇ０６Ｆ ９／４８