(19)日本国特許庁(JP)　(12)公表特許公報(A)　(11)公表番号
特表2022-552885
(P2022-552885A)
(43)公表日　令和4年12月20日(2022.12.20)

(51)国際特許分類		FI			テーマコード（参考）
Ｇ０６Ｆ　12/00　(2006.01)		Ｇ０６Ｆ	12/00	５６０Ｆ	５Ｂ１６０
Ｇ０６Ｆ　12/0804(2016.01)		Ｇ０６Ｆ	12/0804	１００	５Ｂ２０５

審査請求　未請求　予備審査請求　未請求　（全20頁）

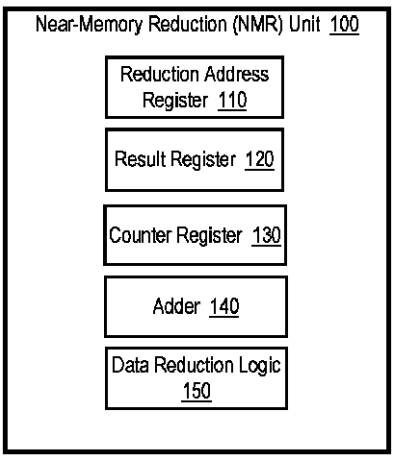
(21)出願番号　特願2022-523534(P2022-523534)
(86)(22)出願日　令和2年7月14日(2020.7.14)
(85)翻訳文提出日　令和4年6月8日(2022.6.8)
(86)国際出願番号　PCT/US2020/041910
(87)国際公開番号　WO2021/080656
(87)国際公開日　令和3年4月29日(2021.4.29)
(31)優先権主張番号　16/658,733
(32)優先日　令和1年10月21日(2019.10.21)
(33)優先権主張国・地域又は機関
米国(US)
(81)指定国・地域　AP(BW,GH,GM,KE,LR,LS,MW,MZ,NA,RW,SD,SL,ST,SZ,TZ,UG,ZM,ZW),EA(AM,AZ,BY,KG,KZ,RU,TJ,TM),EP(AL,AT,BE,BG,CH,CY,CZ,DE,DK,EE,ES,FI,FR,GB,GR,HR,HU,IE,IS,IT,LT,LU,LV,MC,
最終頁に続く

(71)出願人　591016172
アドバンスト・マイクロ・ディバイシズ・インコーポレイテッド
ＡＤＶＡＮＣＥＤ ＭＩＣＲＯ ＤＥＶＩＣＥＳ ＩＮＣＯＲＰＯＲＡＴＥＤ
アメリカ合衆国 ９５０５４ カリフォルニア州、 サンタ クララ、オーガスティン ドライブ ２４８５
(74)代理人　100108833
弁理士　早川 裕司
(74)代理人　100111615
弁理士　佐野 良太
(74)代理人　100162156
弁理士　村雨 圭介
(72)発明者　ヌワン ジャヤセーナ
最終頁に続く

(54)【発明の名称】　ニアメモリデータ削減

(57)【要約】
オフチップ又はオフダイメモリへのストア動作中にニアメモリデータ削減を実施するためのアプローチが提供される。ニアメモリ削減（ＮＭＲ）ユニットは、指定されたアドレス範囲への書き込み操作中にニアメモリデータ削減を提供する。ＮＭＲユニットは、削減されるアドレスの範囲で構成されており、ストア動作がアドレスの範囲内のアドレスを指定すると、ＮＲＭユニットは、ストア動作によって指定されたデータ値を累積削減結果に加算することによってデータ削減を実行する。一実施形態によれば、ＮＲＭユニットは、データ削減がいつ完了したのかを決定するために使用される累積削減結果に対する更新回数のカウントを維持する。
【選択図】図１

FIG. 1



【特許請求の範囲】
【請求項１】
ニアメモリ削減ユニットであって、
削減アドレス範囲を格納するように構成された削減アドレスレジスタと、
データ削減結果を格納するように構成された結果レジスタと、
データ削減ロジックと、を備え、
前記データ削減ロジックは、
オフダイメモリにデータを格納するストア動作の宛先アドレスが、前記削減アドレスレジスタに格納された削減アドレス範囲内にあると判別したことに応じて、
前記ストア動作のデータ値を削減操作によって処理して、更新されたデータ削減結果を生成することと、
前記更新されたデータ削減結果を前記結果レジスタに格納することと、
を行うように構成されている、
ニアメモリ削減ユニット。
【請求項２】
前記データ削減ロジックは、前記削減アドレス範囲を、前記ニアメモリ削減ユニットの前記削減アドレス範囲をプログラムする操作によって指定された値に初期化するように構成されている、
請求項１のニアメモリ削減ユニット。
【請求項３】
前記削減アドレス範囲は、連続するメモリアドレスの範囲である、
請求項１のニアメモリ削減ユニット。
【請求項４】
前記削減アドレス範囲は、メモリアドレスのストライド範囲である、
請求項１のニアメモリ削減ユニット。
【請求項５】
前記結果レジスタに格納された前記データ削減結果は、前記削減アドレスレジスタに格納された前記削減アドレス範囲の変化又は初期化操作に応じて、指定された値に初期化される、
請求項１のニアメモリ削減ユニット。
【請求項６】
前記結果レジスタの更新の現在のカウントを格納するように構成されたカウンタレジスタをさらに備え、前記データ削減ロジックは、データを前記オフダイメモリに格納する前記ストア動作の前記宛先アドレスが、前記削減アドレスレジスタに格納された前記削減アドレス範囲内にあると判別したことに応じて、前記結果レジスタに格納された前記更新の現在のカウントをインクリメントするように構成されている、
請求項１のニアメモリ削減ユニット。
【請求項７】
前記データ削減ロジックは、
データをオフダイメモリに格納する前記ストア動作の前記データ値を処理して、処理済みのデータ値を生成することと、
前記処理済みのデータ値を前記削減操作によって処理して、更新された削減結果を生成することと、
を行うように構成されている、
請求項１のニアメモリ削減ユニット。
【請求項８】
前記データ値を処理することは、前記データ値を関数で処理すること、又は、前記データ値に定数を適用すること、のうち１つ以上を含む、
請求項７のニアメモリ削減ユニット。
【請求項９】

前記オフダイメモリはメインメモリである、
請求項１のニアメモリ削減ユニット。
【請求項１０】
オフダイメモリにデータを格納する前記ストア動作によって指定された前記データがキャッシュに格納され、
前記ストア動作は、前記ストア動作によって指定された前記データを、前記キャッシュからメインメモリに格納するために実行される、
請求項１のニアメモリ削減ユニット。
【請求項１１】
前記ストア動作は第１のストア動作であり、前記データ削減ロジックは、前記第１のストア動作の前記宛先アドレスと同じ宛先アドレスを有する第２のストア動作を受信したことに応じて、エラービットを設定するように構成されている、
請求項１のニアメモリ削減ユニット。
【請求項１２】
前記ニアメモリ削減ユニットは、メモリコントローラ、メモリモジュールに割り当てられるスタンドアロンエンティティ、又は、スタックメモリパッケージのダイのうち１つ以上に実装されている、
請求項１のニアメモリ削減ユニット。
【請求項１３】
前記削減操作は、前記ストア動作の前記データ値を、前記結果レジスタに格納された累積データ削減結果に加算する加算演算である、
請求項１のニアメモリ削減ユニット。
【請求項１４】
データ削減を実行するための方法であって、
１つ以上の非最終メモリストア動作と、最終メモリストア動作と、を含む複数のメモリストア動作の場合であって、前記複数のメモリストア動作の各メモリストア動作が、前記複数のメモリストア動作を受信するメモリインターフェースのビット幅未満であるビット数によってそれぞれ表される複数のデータ要素値を含む、場合に、
前記複数のメモリストア動作のメモリストア動作毎に、各データ要素値に対してデータ削減を実行して、対応するデータ削減値を生成することと、
前記最終メモリストア動作の後に、前記対応するデータ削減値に対してツリーベースのペアワイズデータ削減を実行して、前記複数のメモリストア動作の最終データ削減結果を生成することと、を含む、
方法。
【請求項１５】
前記データ削減結果は、前記複数のメモリストア動作で指定された前記対応するデータ削減値の全ての合計である、
請求項１４の方法。
【請求項１６】
前記複数のデータ要素値内のデータ要素値は、同じビット数で表される、
請求項１４の方法。
【請求項１７】
前記ツリーベースのペアワイズデータ削減は、データ削減値のペアに対してデータ削減を実行して、複数のペアデータ削減値を生成することを含む、
請求項１４の方法。
【請求項１８】
前記ツリーベースのペアワイズデータ削減は、前記複数のペアデータ削減値のうちペアデータ削減値のペアに対してデータ削減を実行して、前記最終データ削減結果を生成することを含む、
請求項１７の方法。

【請求項１９】
前記データ削減値は、ニアメモリ削減ユニットに格納される、
請求項１７の方法。
【請求項２０】
前記方法は、メモリコントローラによって実行される、
請求項１７の方法。
【発明の詳細な説明】
【背景技術】
【０００１】
本項に記載されるアプローチは、追求され得るアプローチであるが、必ずしも以前に着想又は追求されたアプローチではない。したがって、特に断りのない限り、本項に記載される何れのアプローチが、本項に含まれているという理由だけで、従来技術と見なされると想定されるべきではない。さらに、本項に記載される何れのアプローチも、本項に含まれているという理由だけで、よく理解され、日常的で、又は、従来通りであると見なされるべきではない。
【０００２】
機械学習等のある種のコンピューティングプロセスは、大きなデータセットが生成されるため、これを削減する必要がある。データ削減自体は、メモリから読み取られたデータのバイトに対して少ない演算しか行わないという演算量対バイト率の低い加算演算を含む。しかしながら、キャッシュ等の利用可能なオンチップ又はオンダイメモリよりも大きいデータセットは、メインメモリに格納する必要があり、メインメモリからオンチップ又はオンダイメモリにデータを取り出してデータ削減を実行するにはコストがかかる。データ移動コストは、データ削減の計算コストが低い割に法外に高く、データ移動によりキャッシュ汚染が生じる可能性がある。同じ問題は、並列コンピューティングのアプローチにも当てはまる。
【０００３】
この問題に対する技術的な解決策として、データ削減操作を、削減されるデータを生成する操作にピギーバック（piggybacking）させることを伴う。これにより、削減されるデータが生成される際にデータ削減を実行することが可能になるが、不利な点がある。先ず、行列の乗算及び畳み込み等の汎用的な計算は、多くの場合、様々な状況で再利用できるように実装されており、ＣＰＵ／ＧＰＵキャッシュ、ＧＰＵレジスタ、ＬＤＳ／共有メモリ等の利用可能なオンチップレジスタを最適に利用し得るため、データ削減操作を含むことは、性能に悪影響を与え、さらに亜種の生成及び維持にもつながり、その再利用性が阻害される可能性がある。さらに、データセットを生成する操作は、同じ出力先に対して複数の書き込みを行う等の特殊な技術を使用することがあり、１つの書き込みでデータ削減操作を含めることが不可能な場合がある。これは、複数の書き込みからの出力を同じ記憶位置にマージするためにＬ２アトミック（L2 atomics）を利用するＧＰＵで一般的に行われる。したがって、計算操作によって生成される大量のデータを削減するためのより優れたアプローチが必要とされている。
【０００４】
実施形態は、添付図面の図に限定ではなく例として示されており、図において、同様の符号は同様の要素を指す。
【図面の簡単な説明】
【０００５】
【図１】本明細書に記載されるニアメモリデータ削減を実施するためのニアメモリ削減（ＮＭＲ）ユニットを示すブロック図である。
【図２】ＮＭＲユニットを使用してデータを削減するためのアプローチを示すフロー図である。
【図３Ａ】２つのメモリモジュールという観点からデータを削減するためのアプローチの例示的な実装を示す図である。

【図３Ｂ】２つのメモリモジュールという観点からデータを削減するためのアプローチの例示的な実装を示す図である。
【図３Ｃ】２つのメモリモジュールという観点からデータを削減するためのアプローチの例示的な実装を示す図である。
【図３Ｄ】２つのメモリモジュールという観点からデータを削減するためのアプローチの例示的な実装を示す図である。
【図３Ｅ】２つのメモリモジュールという観点からデータを削減するためのアプローチの例示的な実装を示す図である。
【図４】各メモリストア動作後にツリーベースのペアワイズ（pair-wise）データ削減を使用するデータ削減を実行するための例示的なアプローチを示すブロック図である。
【図５】「レーン内（in-lane）」データ削減を実行するためのアプローチを示すブロック図である。
【図６】「レーン内」データ削減を実行するためのアプローチを示すフロー図である。
【発明を実施するための形態】
【０００６】
以下の説明では、説明のために、実施形態の十分な理解を提供するために、多数の具体的な詳細が記載されている。しかしながら、当業者には、実施形態が、これらの具体的な詳細がなくても実施されることが明らかであろう。他の例では、実施形態を不必要に不明瞭にすることを回避するために、周知の構造及びデバイスがブロック図の形態で示されている。
Ｉ．概要
ＩＩ．アーキテクチャ
ＩＩＩ．ニアメモリデータ削減
Ａ．概要
Ｂ．スレッドのための共有リソース割り当ての変更
Ｃ．有用性メトリック
Ｄ．リソース割り当ての揺らぎ（oscillation）への対処
ＩＶ．操作例
Ｖ．レーン内データ削減
【０００７】
Ｉ．概要
オフチップ又はオフダイメモリへのストア動作（格納操作）中にニアメモリデータ削減を実施するためのアプローチが提供される。ニアメモリ削減（ＮＭＲ）ユニットは、指定されたアドレス範囲への書き込み動作中にニアメモリデータ削減を提供する。ＮＭＲユニットは、削減されるアドレスの範囲で構成されており、ストア動作がアドレスの範囲内のアドレスを指定すると、ＮＲＭユニットは、ストア動作によって指定されたデータ値を累積削減結果に加算することによって、データ削減を実行する。一実施形態では、ＮＲＭユニットは、データ削減がいつ完了したのかを決定するために使用される累積削減結果に対する更新回数のカウントを保持する。
【０００８】
このアプローチは、データ削減を実行するためにオフチップメモリからオンチップメモリにデータを読み戻すコストを回避し、マイクロプロセッサ、スタックメモリのロジックダイの演算ユニット、又は、メモリコントローラが容易にアクセスできるレジスタに累積削減結果を維持する。さらに、アプローチは、キャッシュの汚染を低減し、データが準備完了になると、削減されるデータを生成する演算処理と同時にデータ削減を実行することを可能にする。
【０００９】
ＩＩ．アーキテクチャ
図１は、本明細書で説明する、ニアメモリ削減を実施するためのニアメモリ削減（ＮＭＲ）ユニット１００を示すブロック図である。ＮＭＲユニット１００は、削減アドレスレ

ジスタ１１０と、結果レジスタ１２０と、オプションのカウンタレジスタ１３０と、算術論理ユニット（ＡＬＵ）１４０と、データ削減ロジック１５０と、を含む。実施形態は、図に示され、ＡＬＵ１４０を含むＮＭＲ１００という観点から本明細書に記載されているが、いくつかの実施形態では、ＮＭＲ１００は、任意のタイプのデータ削減操作（例えば、加算、最小値若しくは最大値の決定、ＡＮＤ、ＯＲ、ＸＯＲ等の算術演算及び／又は論理演算）を実行するためのデジタル回路を含む。ＮＭＲユニット１００は、特定の実施形態に応じて変化する、より少ない又は追加の要素を含み、実施形態は、任意の特定の要素に限定されない。実施形態によれば、本明細書に記載され、単一のレジスタとして図に示される様々なレジスタは、特定の実施形態に応じて、複数のレジスタによって実装される。

【００１０】

削減アドレスレジスタ１１０は、データ削減が実行されるアドレス範囲を格納する。以下により詳細に説明するように、格納されたアドレス範囲は、ストア動作が削減されるかどうかを決定するためにデータ削減ロジック１５０によって使用される物理アドレス又は論理アドレスである。結果レジスタ１２０は、累積データ削減結果を格納する。結果レジスタ１２０は、初期化コマンドに応じて、例えばゼロ等の指定された値に初期化される。或いは、結果レジスタ１２０は、削減アドレスレジスタ１１０に新たなアドレス範囲が格納されることに応じて、指定された値に初期化される。

【００１１】

カウンタレジスタ１３０は、結果レジスタ１２０が最後に初期化されてからの結果レジスタ１２０の更新の現在のカウントを格納する。更新の現在のカウントは、例えば、データ削減がいつ完了したのかを決定するためにメモリコントローラによって使用される。加算器１４０は、単一の加算器又は複数の加算器として実装され、結果レジスタ１２０に格納された累積データ削減結果にストア動作のデータ値を加算する。

【００１２】

データ削減ロジック１５０は、ＮＭＲユニット１００によって実行されるデータ削減を管理する。これは、削減されるストア動作を特定すること、データ削減を実行すること、及び、カウンタレジスタ１３０内の更新の現在のカウントをオプションでインクリメント（増分）することを含む。削減されるストア動作は、削減アドレスレジスタ１１０に維持されるアドレス範囲内のアドレスを指定する。ストア動作を削減するために、ＮＭＲユニット１００は、ストア動作によって指定されたデータ値を、結果レジスタ１２０に格納された累積データ削減結果に加算させる。また、データ削減ロジック１５０は、カウンタレジスタ１３０に格納されている更新の現在のカウントをオプションでインクリメントする。データ削減ロジック１５０は、コンピュータハードウェア、コンピュータソフトウェア、又は、コンピュータハードウェアとコンピュータソフトウェアとの任意の組み合わせによって実装される。

【００１３】

一実施形態によれば、プリミティブ（つまり、プリミティブ命令又は低レベル命令）のセットは、プロセスがＮＭＲユニット１００の機能を呼び出すことを可能にするために提供される。例示的なプリミティブは以下を含むが、これに限定されない。

ｐｒｏｇｒａｍ＿ｒｅｄｕｃｔｉｏｎ＿ａｄｄｒｅｓｓ＿ｒａｎｇｅ（ａｄｄｒｅｓｓ＿ｂｅｇｉｎ，ａｄｄｒｅｓｓ＿ｅｎｄ）

ｒｅａｄ＿ｒｅｄｕｃｔｉｏｎ＿ｒｅｓｕｌｔ（）

【００１４】

ｐｒｏｇｒａｍ＿ｒｅｄｕｃｔｉｏｎ＿ａｄｄｒｅｓｓ＿ｒａｎｇｅ（ａｄｄｒｅｓｓ＿ｂｅｇｉｎ，ａｄｄｒｅｓｓ＿ｅｎｄ）プリミティブは、データ削減が実行されるアドレス範囲で削減アドレスレジスタ１１０をプログラムすることによって、データ削減のためにＮＭＲユニット１００を準備する。一実施形態によれば、アドレス範囲は、開始アドレス及びｌｅｎｇｔｈ＿ｃｏｕｎｔによって識別され、終了アドレスは、ａｄｄｒｅｓｓ＿ｂｅｇｉｎ＋ｌｅｎｇｔｈ＿ｃｏｕｎｔとして計算される。アドレス範囲は、連続する

アドレスに限定されず、一実施形態によれば、ｐｒｏｇｒａｍ＿ｒｅｄｕｃｔｉｏｎ＿ａｄｄｒｅｓｓ＿ｒａｎｇｅプリミティブは、削減されるアドレス間のギャップを指定するストライド引数（stride argument）をサポートする。或いは、ｐｒｏｇｒａｍ＿ｒｅｄｕｃｔｉｏｎ＿ａｄｄｒｅｓｓ＿ｒａｎｇｅプリミティブは、複数の連続しない、非ストライドのアドレス範囲に亘るデータ削減を可能にするために、複数のアドレス範囲引数をサポートする。一実施形態によれば、ｐｒｏｇｒａｍ＿ｒｅｄｕｃｔｉｏｎ＿ａｄｄｒｅｓｓ＿ｒａｎｇｅプリミティブにより、結果レジスタ１２０内の累積データ削減結果及び／又はカウンタレジスタ１３０の現在のカウントが、ゼロ等の指定された値に初期化される。指定された値は、ｐｒｏｇｒａｍ＿ｒｅｄｕｃｔｉｏｎ＿ａｄｄｒｅｓｓ＿ｒａｎｇｅプリミティブの引数として含まれるか、データ削減ロジック１５０内で別に構成される。或いは、ｉｎｉｔｉａｌｉｚｅ＿ｎｍｒ（ｖａｌｕｅ）等の別のプリミティブが、結果レジスタ１２０及び／又はカウンタレジスタ１３０を初期化するために提供され、このｖａｌｕｅ引数は初期化値である。ｒｅａｄ＿ｒｅｄｕｃｔｉｏｎ＿ｒｅｓｕｌｔ（）プリミティブにより、累積データ削減結果が結果レジスタ１２０から読み取られる。

【００１５】

実施形態は、アドレス範囲及びＮＭＲユニット１００のレジスタに格納された累積データ削減結果という観点から図に示され、本明細書に記載されているが、実施形態は、アドレス範囲及び累積データ削減結果が、通常のロード動作及びストア動作を使用してアクセス可能な記憶位置にマッピングされることを含む。

【００１６】

一実施形態によれば、ＮＭＲユニット１００は、対応するメモリモジュールに割り当てられるスタンドアロンエンティティとして実装される。或いは、ＮＭＲユニット１００に関してここで説明する様々な要素及び機能は、メモリコントローラ、又は、ハイブリッドメモリキューブ（ＨＭＣ）パッケージ若しくは高帯域幅メモリ（ＨＢＭ）パッケージ等のスタックメモリパッケージのダイに実装される。一実施形態によれば、ＮＭＲユニット１００は、３Ｄスタックメモリ若しくは従来のＤＲＡＭのメモリ層のメモリアレイ／バンク近くに、又は、メモリコントローラ若しくは最後のレベルのキャッシュの近く若しくはその内部に実装される。任意の数のＮＭＲユニット１００が、同時ニアメモリ削減を可能にするために実装される。ＮＭＲユニット１００は、例えば、不揮発性デュアルインラインメモリモジュール（ＮＶＤＩＭＭ）及びＮＶＤＩＭＭ－Ｐ、ＳＲＡＭスクラッチパッドメモリ等を含む、ニアメモリコンピューティングに対応可能な任意のタイプのメモリと共に使用される。

【００１７】

ＩＩＩ．ニアメモリデータ削減

図２は、ＮＲＭユニット１００を使用してデータを削減するためのアプローチを示すフロー図２００である。この例では、データ削減が加算演算を使用して実行されるが、実施形態はこの例に限定されず、実施形態は、他のタイプのデータ削減操作を使用することを含む。ステップ２０２では、初期化が実行され、特定の実装に応じて変化するいくつかの異なるタスクを含む。一実施形態によれば、ＮＭＲユニット１００、より具体的には、結果レジスタ１２０及び／又はカウンタレジスタ１３０は、ゼロ等の指定された値に初期化される。他の実施形態では、ゼロ以外の値を使用する。一実施形態によれば、ＮＭＲユニット１００の初期化は、初期化プリミティブ又は本明細書に上述したように、削減されるアドレス範囲を指定するｐｒｏｇｒａｍ＿ｒｅｄｕｃｔｉｏｎ＿ａｄｄｒｅｓｓ＿ｒａｎｇｅによって呼び出される。削減されるアドレス範囲は、メモリモジュールによってサポートされているアドレス範囲全体、又は代替的に、メモリモジュールによってサポートされているアドレス範囲のサブセットである。一実施形態によれば、初期化は、例えば削減されるデータの符号付き／符号なし整数、単精度／倍精度浮動小数点等のように、ビット幅及びデータタイプ等の追加の制御情報を初期化することを含む。これにより、ＮＭＲユニット１００は、受信したストア動作の情報を正しく解釈することができる。

【００１８】

初期化が完了した後、次にステップ２０４では、少なくともアドレス及び値を指定するストア動作（格納操作）が受信され、そのアドレスへのメモリ書き込みが実行される。ストア動作は、特定の実施形態に応じて、他の情報を指定する。一実施形態によれば、メモリ書き込みは、ニアメモリ削減が実行されるメモリにキャッシュをバイパス又はフロースルーする方式で発行される。例えば、一実施形態では、ストア動作は、引数を含むか、キャッシュをバイパスするように特別に設計される。他の実施形態では、メモリに書き込まれる値は、キャッシュラインの何れの部分が書き込まれたかを示す追加のメタデータと共にキャッシュに取り込まれるが、削減される値は、ＮＭＲユニット１００がデータ削減においてデータ値を含むことを確実にするために、キャッシュからオフダイメモリにフラッシュされる。メタデータが追跡され、キャッシュラインの何れのサブセット（複数可）が修正されたかを示す削減エンジンに通信され、データ削減に組み込まれる必要がある。この情報は、部分的な書き込みをサポートするアーキテクチャ（例えば、エラー訂正コード（ＥＣＣ）のない特定のグラフィックスプロセッシングユニット（ＧＰＵ）等）において、バイトイネーブルビットの形態で利用できる。ストア動作は、ＮＭＲユニット１００によって、例えばメモリコントローラから受信される。したがって、本明細書に記載されるアプローチは、キャッシュを使用するストア動作と、キャッシュをバイパスしてオフダイメモリに直接データを格納するストア動作と、に適用できる。

【００１９】

ステップ２０６では、ストア動作が削減アドレス範囲内にあるかどうかの判別が行われる。一実施形態によれば、ストア動作によって指定されたアドレスは、削減アドレスレジスタ１１０に格納されたアドレス範囲と比較される。例えば、ストア動作によって指定されたアドレスのうち、削減アドレスレジスタ１１０内の下位アドレス以上であり、削減アドレスレジスタ１１０内の上位アドレス以下であるアドレスは、削減アドレス範囲内にある。

【００２０】

ステップ２０８において、ストア動作が削減アドレス範囲内にあると判別された場合、ステップ２１０において、ストア動作によって指定されたデータ値が、累積削減結果に加算される。例えば、データ削減ロジック１５０は、加算器１４０を介して、ストア動作で指定されたデータ値を、結果レジスタ１２０に格納された累積データ削減結果に加算させる。或いは、ストア動作で指定されたデータ値は、関数によって処理されるか、例えば定数によってスケーリングされて、結果レジスタ１２０内の累積データ削減結果に加算される処理済みデータ値を生成する。

【００２１】

カウンタレジスタ１３０に格納された現在のカウントは、オプションでインクリメントされる。例えば、整数値の形態の現在のカウントは、データ削減が実行される毎に、つまり、ストア動作が削減アドレス範囲内のアドレスを指定した場合に、１つずつインクリメントされる。一実施形態によれば、カウンタレジスタ１３０内の現在のカウントは、例えば、指定された反復回数の後に、データ削減を停止するために使用される。例えば、アルゴリズムは、現在のカウントを使用して、データ削減を、指定されたアドレス範囲で指定された反復回数に制限し、計算リソースの消費を低減することによって、近似的なデータ削減を提供する。

【００２２】

プロセスは、ステップ２１２の後で、又は、ステップ２０８でストア動作が削減アドレス範囲内にないと判別したことに応じて、ステップ２１４で完了する。

【００２３】

上述したプロセスは、データ削減が完了するまで任意の回数繰り返される。データ削減の完了は、特定の実施形態に応じて変わる異なる方法で決定される。例えば、データ削減ロジック１５０は、指定された反復回数だけデータ削減を実行する。別の例として、ｐｒｏｇｒａｍ＿ｒｅｄｕｃｔｉｏｎ＿ａｄｄｒｅｓｓ＿ｒａｎｇｅプリミティブは、実行されるデータ削減の反復回数を指定する引数を含む。これにより、アプリケーションプログ

ラムは、実行されるデータ削減の特定の数を指定することができる。さらなる例として、メモリコントローラ又は処理ユニットは、対象のメモリモジュールのカウンタレジスタ１３０に格納されている現在のカウントを読み取り、指定された数のデータ削減が完了するとデータ削減を停止する。例えば、メモリコントローラは、複数のＮＲＭ１００のカウンタレジスタ１３０から現在のカウントを取り出し、次に、現在のカウントの全てが指定された数を満たすと、データ削減が完了したことをマイクロプロセッサに通知する。次に、マイクロプロセッサは、各ＮＭＲユニット１００の結果レジスタ１２０から累積削減結果を読み取り、これらの値を最終削減結果に減少させる。或いは、最終的な削減結果は、スタックメモリのロジックダイ内又はメモリコントローラの近くの算術ユニットによって生成される。

【００２４】

カウンタレジスタ１３０に格納されている現在のカウントを読み取る代わりに、メモリコントローラは、削減されるアドレス範囲を知っているので、ストア動作の数を追跡して、いつデータ削減が完了したのかを決定する。一実施形態によれば、データ削減の完了により、以下により詳細に説明するように、ツリー削減がトリガされる。削減されるデータがキャッシュされる実施形態では、ニアメモリ削減の完了は、キャッシュ内のデータがデータ削減結果に含まれていることを確実にするために、キャッシュされたデータのオフチップメモリへのフラッシュをトリガすることを含む。

【００２５】

データ削減サイクルが完了したと見なされる前に、ＮＭＲユニット１００が、以前のストア動作からの値を既に含むアドレスを指定するストア動作を受信する状況が発生する場合がある。一実施形態によれば、ＮＭＲユニット１００は、同じアドレスへの複数の書き込みを検出し、例えば、メモリコントローラによってポーリングされ、マイクロプロセッサに提供され得るエラービットを設定する等して、エラーをフラグ付けする。或いは、新しい値は、結果レジスタ１２０に格納された累積データ削減結果に加算され、これは、エラー耐性がある又は近似アルゴリズムに基づく計算には十分である。

【００２６】

ＩＶ．操作例

図３Ａ～図３Ｅは、図中でメモリモジュール－０及びメモリモジュール－１として識別される２つのメモリモジュールという観点から、データを削減するためのアプローチの例示的な実施形態を示している。図３Ａ～図３Ｅは、ステップの時系列的なシーケンスを示している。この例では、２つのメモリモジュールは、インタリーブされたアドレス、つまり、第１のメモリモジュールのアドレス０及び２と、第２のメモリモジュールのアドレス１及び３とを有し、これは、メモリモジュールレベルの並列処理を提供するために一般的に行われる。メモリモジュールの各々は、特定の実施形態に応じて、任意の数の他のアドレスを有する。２つのメモリモジュールの各々は、対応するＮＭＲユニット１００を有する。

【００２７】

図３Ａで開始し、２つのメモリモジュールが初期化され、これは、各メモリモジュールの削減値をゼロに設定することと、削減されるアドレス範囲（この例では、アドレス０～３）を指定することと、を含む。本明細書で上述したように、ｐｒｏｇｒａｍ＿ｒｅｄｕｃｔｉｏｎ＿ａｄｄｒｅｓｓ＿ｒａｎｇｅプリミティブは、削減されるアドレス範囲を指定し、結果レジスタ１２０及びカウンタレジスタ１３０を初期化するために使用される。或いは、別の初期化コマンドが、結果レジスタ１２０及びカウンタレジスタ１３０を初期化するために使用される。削減されるアドレス範囲は、削減アドレスレジスタ１１０に格納される。

【００２８】

図３Ｂでは、値１０がアドレス０に格納されることを指定する第１のストア動作が受信される。第１のメモリモジュールを管理するメモリコントローラは、値１０をメモリモジュール－０のアドレス位置０に格納させる。第１のメモリモジュールのＮＭＲユニット１

００のデータ削減ロジック１５０は、ストア動作で指定されたアドレスが、削減アドレスレジスタ１１０に格納された削減アドレス範囲内にあるかどうかを判別する。ストア動作で指定されたアドレス０が、削減アドレスレジスタ１１０の削減アドレス範囲内にあるので、データ削減ロジック１５０は、ストア動作で指定された値１０を、加算器１４０を介して、結果レジスタ１２０に格納された累積削減結果に加算させる。図３Ｂにおいて、これは、第１のメモリモジュール、つまりメモリモジュール－０の削減された値である、削減値－０が１０に更新されることによって示されている。第１のメモリモジュールのＮＭＲユニット１００のデータ削減ロジック１５０は、カウンタレジスタ１３０に格納されている現在のカウントをインクリメントする。第２のメモリモジュールであるメモリモジュール－１については、アドレス０を処理しないので、変更がない。

【００２９】

ストア動作によって指定されたアドレスが、削減アドレスレジスタ１１０に格納された削減アドレス範囲外である場合、ストア動作によって指定された値は、第１のメモリモジュールに格納されるが、その値は、結果レジスタ１２０に格納された累積削減結果に加算されないであろう。例えば、ストア動作によって、値７がアドレス５０に格納されることが指定されたと仮定する。第１のメモリモジュールがアドレス５０で構成されていたと仮定すると、値７は、アドレス５０に格納されが、ストア動作で指定されたアドレス５０は、削減アドレスレジスタ１１０に指定された削減アドレス範囲内にないため、結果レジスタ１２０に格納された累積削減結果に加算されないであろう。

【００３０】

図３Ｃでは、値９９がアドレス１に格納されることを指定する第２のストア動作が受信される。第２のメモリモジュールを管理するメモリコントローラは、値９９をアドレス位置１に格納させる。第２のメモリモジュールのＮＭＲユニット１００のデータ削減ロジック１５０は、ストア動作で指定されたアドレスが、削減アドレスレジスタ１１０に格納された削減アドレス範囲内にあるかどうかを判別する。ストア動作で指定されたアドレス１は、削減アドレスレジスタ１１０に格納された削減アドレス範囲０～３内にあるので、データ削減ロジック１５０は、ストア動作で指定された値９９を、加算器１４０を介して、結果レジスタ１２０に格納された累積削減結果に加算させる。図３Ｃでは、これは、第２のメモリモジュール、つまりメモリモジュール－１の削減された値である削減値－１が９９に更新されることによって示されている。第２のメモリモジュールのＮＭＲユニット１００のデータ削減ロジック１５０は、カウンタレジスタ１３０に格納されている現在のカウントをインクリメントする。第１のメモリモジュールであるメモリモジュール－０については、アドレス１を処理しないので、変更がない。

【００３１】

例を続けると、図３Ｄにおいて、値４４がアドレス２に格納されることを指定する第３のストア動作が受信される。第１のメモリモジュールを管理するメモリコントローラは、値４４をアドレス位置２に格納させる。第１のメモリモジュールのＮＭＲユニット１００のデータ削減ロジック１５０は、ストア動作で指定されたアドレスが、削減アドレスレジスタ１１０に格納された削減アドレス範囲内にあるかどうかを判別する。ストア動作で指定されたアドレス２は、削減アドレスレジスタ１１０に格納された削減アドレス範囲０～３内にあるので、データ削減ロジック１５０は、ストア動作で指定された値４４を、加算器１４０を介して、結果レジスタ１２０に格納された累積削減結果に加算させる。図３Ｄにおいて、これは、削減値－０が、以前の値１０に新しい値４４を加えた合計である５４に更新されることによって示されている。第１のメモリモジュールのＮＭＲユニット１００のデータ削減ロジック１５０は、カウンタレジスタ１３０に格納されている現在のカウントをインクリメントする。第２のメモリモジュールであるメモリモジュール－１については、アドレス２を処理しないので、変更がない。

【００３２】

図３Ｅに示す最終操作では、値２２がアドレス３に格納されることを指定する第４のストア動作が受信される。第２のメモリモジュールを管理するメモリコントローラは、値２

２をアドレス位置３に格納させる。第２のメモリモジュールのＮＭＲユニット１００のデータ削減ロジック１５０は、ストア動作で指定されたアドレスが、削減アドレスレジスタ１１０に格納された削減アドレス範囲内にあるかどうかを判別する。ストア動作で指定されたアドレス３は、削減アドレスレジスタ１１０に格納された削減アドレス範囲０～３内にあるので、データ削減ロジック１５０は、ストア動作で指定された値２２を、加算器１４０を介して、結果レジスタ１２０に格納された累積削減結果に加算させる。図３Ｅにおいて、これは、削減値－１が、以前の値９９に新しい値２２を加えた合計である１２１に更新されることによって示されている。第２のメモリモジュールのＮＭＲユニット１００のデータ削減ロジック１５０は、カウンタレジスタ１３０に格納されている現在のカウントをインクリメントする。第１のメモリモジュールであるメモリモジュール－０については、アドレス３を処理しないので、変更がない。

【００３３】

これらの例によって提供される技術的な解決策は、１）単一の記憶位置ではなくメモリアドレスの範囲にわたってメモリ削減を提供すること、及び、２）メモリ内アトミック（in-memory atomics）が（メモリ帯域幅削減なしに）記憶位置に読み取り－修正－書き込み操作を実行するのに対し、この解決策は、累積削減結果を各ＮＭＲユニット１００の結果レジスタ１２０に格納し、データ削減を実行するために、メインメモリからキャッシュ等のオンチップメモリにデータを読み戻す必要性が回避されることによって、メモリ内アトミックと異なる。

【００３４】

Ｖ．レーン内データ削減

メモリインターフェースの幅が、格納されているデータ要素の幅を超えることは少なくない。例えば、メモリインターフェースの論理幅は、２５６ビット又は５１２ビットであるが、格納されているデータ要素の幅は、８ビット、１６ビット、３２ビット又は６４ビットとなる場合がある。したがって、２５６ビット幅のインターフェースを備えたメモリモジュールの場合、メモリストア動作毎に８つの３２ビットデータ要素が格納される。

【００３５】

ハードウェア内の複数のデータ要素にわたってデータ削減を同時に実行するための１つのアプローチは、ツリーベースのペアワイズデータ削減、つまり「クロスレーン」データ削減を使用する。図４は、メモリインターフェースの論理幅が格納されているデータ要素の幅を超える場合に、各メモリストア動作後にツリーベースのペアワイズデータ削減を実行するための例示的なアプローチを示すブロック図である。この例では、メモリインターフェース幅は２５６ビットであり、各データ要素は３２ビットで表され、メモリストア動作毎に最大８つの３２ビットデータ要素を提供する。図４に示すインターフェースデータ要素の幅は例示的な値であり、実施形態はこれらの例に限定されない。また、この例では、データ削減が加算を使用して実行されるが、実施形態は、加算を使用してデータ削減を実行することに限定されず、他のタイプの算術演算及び論理演算を含む。

【００３６】

図４の例では、データ転送プロセスは３つのメモリストア動作を含み、完全なツリーベースのペアワイズデータ削減が各メモリストア動作後に実行される。より具体的には、第１のメモリストア動作（メモリ格納操作）後に、ツリーの第１のレベルで、３２ビットデータ要素の４つのペアに対して４つの加算演算が実行され、ペア合計が生成される。ツリーの次のレベルでは、初期の加算セットから４つのペアの合計に対して２つの加算演算が実行され、残りの２つのペア合計が生成される。ツリーの最終レベルでは、残りの２つのペアの合計が加算されて、第１のメモリストア動作における８つの３２ビットデータ要素全ての完全データ削減、つまり合計である結果１が生成される。

【００３７】

このプロセスは、第２のメモリストア動作で８つの３２ビットデータ要素に対して繰り返され、データ削減結果が結果１に加算されて結果２が生成される。第３（及び最終）のメモリストア動作の８つの３２ビットデータ要素も同様に削減され、結果が結果２に加算

され、３つのメモリストア動作の２４の３２ビットデータ要素全ての合計である最終結果が生成される。このアプローチは、３つのストア動作に少なくとも３つの加算サイクルを必要とし、各加算サイクルは、データ削減を完了するために、７つの追加の加算及び３つのステップを必要とし、これは計算コストがかかる。

【００３８】

一実施形態によれば、図４のクロスレーンデータ削減アプローチとは対照的に、レーン内データ削減アプローチは、データ転送における全てのメモリストア動作にわたって、メモリインターフェースの３２ビット幅の各部分内で加算を実行する。データ転送の最終メモリストア動作の後に、完全なツリーベースのペアワイズデータ削減が実行される。

【００３９】

図５は、「レーン内」データ削減を実行するためのアプローチを示すブロック図である。データは、２つの非最終メモリストア動作と、１つの最終メモリストア動作と、を含む３つのメモリストア動作を使用してメモリに格納される。一実施形態によれば、データの各レーンは、メモリストア動作の全てにわたって別々に削減される。図５に示す例では、各メモリストア動作中に、データ削減が各３２ビットデータ要素、つまりレーンに対して実行され、８つの３２ビットデータ削減値が提供される。加算演算を使用したデータ削減の場合、各３２ビットデータ要素は、その対応する３２ビットデータ要素、つまりレーンの累積値に加算され、８つの３２ビット累積データ値が提供される。最終メモリストア動作の後に、完全なツリーベースのペアワイズデータ削減が８つの３２ビットデータ要素にわたって実行され、最終データ削減結果が生成される。加算演算を使用したデータ削減の場合、最終データ削減結果は、３つのメモリストア動作からの３２ビットデータ要素の全ての合計である。

【００４０】

図５に示すように、第１のメモリストア動作（メモリ格納操作）が、第１の３２ビットデータ要素に値１７を指定したと仮定する。値１７は、第１の３２ビットデータ要素、つまり第１のレーンの累積値に加算される。この例では、第１の３２ビットデータ値の初期の累積値はゼロであると推定されているので、第１のレーンの累積値は、第１のメモリストア動作後には１７である。これは、他の３２ビットデータ値、つまりレーンの各々について繰り返される。累積データ値は、レジスタ、キャッシュ、メインメモリ、又は、本明細書において上述したＮＭＲユニット１００に保持される。

【００４１】

第２のメモリストア動作は、第１の３２ビットデータ要素、つまり第１のレーンに値９を指定する。値９は、第１のレーンの累積値１７に加算され、第１の３２ビットデータ要素の新しい累積値２６が生成される。これは、他の３２ビットデータ値、つまりレーンの各々について繰り返される。

【００４２】

第３（最終）のメモリストア動作は、第１の３２ビットデータ要素に値１１を指定する。値１１が累積値２６に加算され、第１の３２ビットデータ要素の新しい累積値３７が生成される。これは、他の３２ビットデータ値、つまりレーンの各々について繰り返される。第３のメモリストア動作後、他の８つの３２ビットレーンの各々も、３つのメモリストア動作の各々からの対応する３２ビットのデータ要素の合計である累積値を有する。次に、完全なツリーベースのペアワイズデータ削減が８つの累積した３２ビットデータ値に対して実行され、図５に示す最終結果が生成される。

【００４３】

このアプローチは、各メモリストア動作の後に完全なツリーベースのペアワイズデータ削減を実行する代わりに、各メモリストア動作の後に単一のセットの並列加算が実行されるので、図４に示すアプローチよりも計算コストが低い。図５に示す例では、８つの並列加算のセットが、各メモリストア動作後に実行される。加算の総数は、図４に示すアプローチと同じであるが、必要なステップがより少ない。図４の例では、完全なツリーベースのペアワイズデータ削減の各々に３つ、結果２及び最終結果を生成するために２つの、合

計２３の加算及び１１のステップが必要である。対照的に、図５のレーン内データ削減アプローチは、２３の加算を使用するが、必要なステップは５つのみである。

【００４４】

実施形態は、等しい幅のレーン、つまりレーンの全てが３２ビット幅であるという観点から、図に示され、本明細書に記載されているが、実施形態は、これらの例に限定されない。一実施形態によれば、レーンは異なる幅を有し、メモリコントローラ又はレーン内データ削減を実行している他の要素は、メモリストア動作内で別々のデータ要素を決定することができ、データ要素の位置は、メモリストア動作の全てにわたって一貫している。レーン内データ削減の使用は、例えば、メモリストア動作のパラメータを介して選択可能であり、及び／又は、メモリコントローラは、レーン内データ削減を実行するように構成されている。

【００４５】

図６は、「レーン内」データ削減を実行するためのアプローチを示すフロー図６００である。ステップ６０２において、初期化が実行され、特定の実施形態に応じて変化するいくつかの異なるタスクを含む。一実施形態によれば、累積データ値を格納する記憶位置（例えば、レジスタ、ＮＲＭユニット１００、他の記憶位置）は、ゼロ等の指定された値に初期化される。他の実施形態は、特定の実施形態に応じて、ゼロ以外の値を使用する。

【００４６】

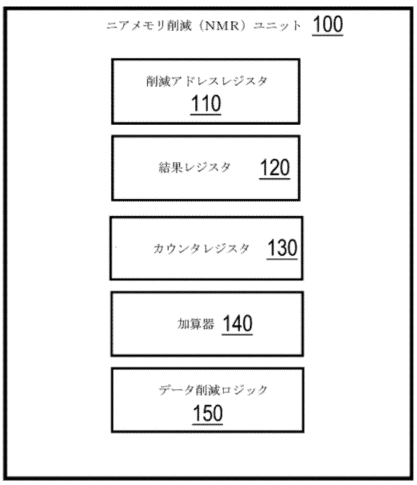
初期化完了後に、ステップ６０４において、メモリインターフェースの幅よりも小さいビット幅を有する複数のデータ要素を含むメモリストア動作（メモリ格納操作）が受信される。ステップ６０６において、データ要素の累積値が更新される。例えば、メモリコントローラは、メモリストア動作で指定されたデータ要素の値を、対応する各レーンの累積値に加算させる。

【００４７】

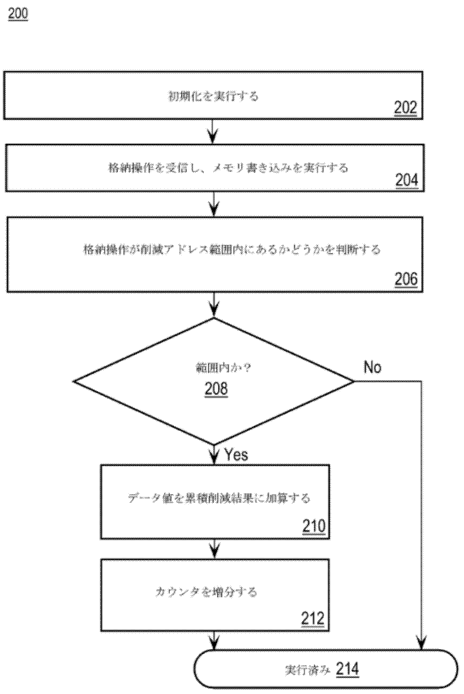
ステップ６０８において、メモリストア動作が最終メモリストア動作であるかどうかが判別される。メモリストア動作が最終メモリストア動作ではない場合、制御はステップ６０４に戻る。現在のメモリストア動作が最終メモリストア動作である場合、ステップ６１０において、上述したように、完全なツリーベースのペアワイズデータ削減が実行されて最終データ削減結果が生成され、プロセスはステップ６１２で完了する。

【図面】

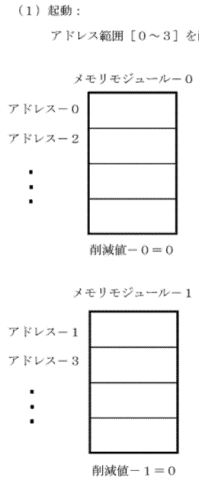
【図１】



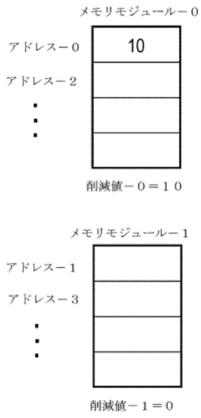
【図２】



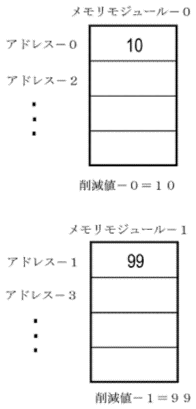
【図３Ａ】



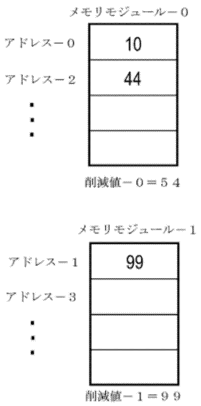
【図３Ｂ】



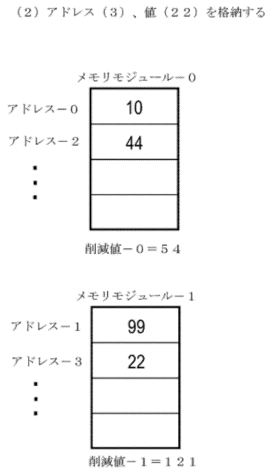
【図３Ｃ】



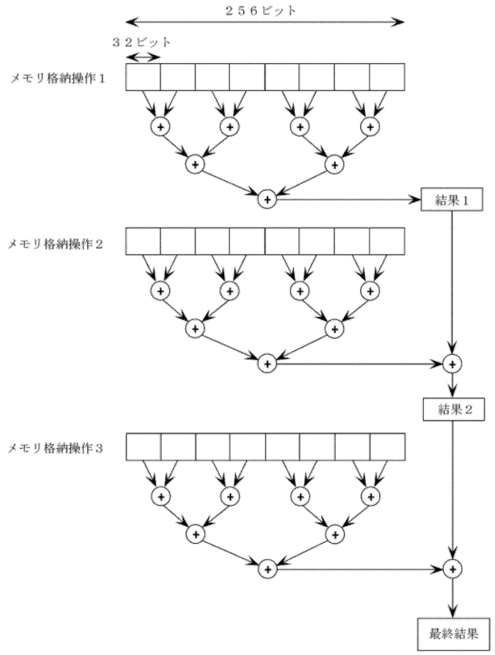
【図３Ｄ】



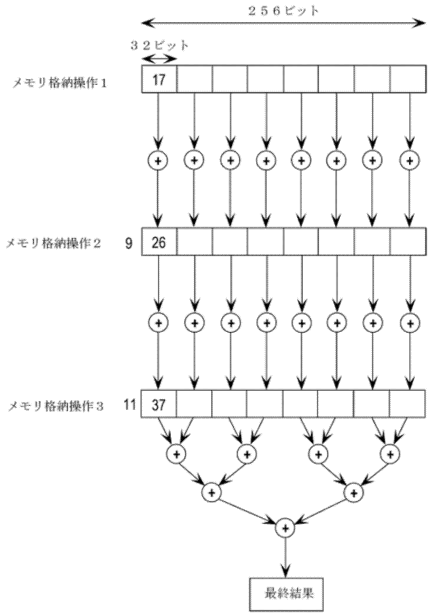
【図３Ｅ】


（２）アドレス（３）、値（２２）を格納する
メモリモジュール－０
アドレス－０
10
アドレス－２
44
削減値－０＝５４
メモリモジュール－１
アドレス－１
99
アドレス－３
22
削減値－１＝１２１

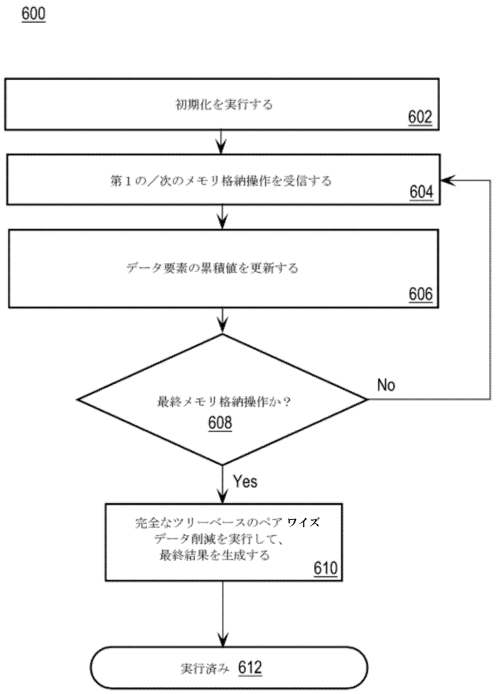
【図４】


２５６ビット
３２ビット
メモリ格納操作１
結果１
メモリ格納操作２
結果２
メモリ格納操作３
最終結果

【図５】


２５６ビット
３２ビット
メモリ格納操作１
17
メモリ格納操作２
9
26
メモリ格納操作３
11
37
最終結果

【図６】


600
初期化を実行する
602
第１の／次のメモリ格納操作を受信する
604
データ要素の累積値を更新する
606
最終メモリ格納操作か？
608
No
Yes
完全なツリーベースのペア ワイズ
データ削減を実行して、
最終結果を生成する
610
実行済み 612

【手続補正書】
【提出日】令和４年６月２２日(２０２２.６.２２)
【手続補正１】
【補正対象書類名】特許請求の範囲
【補正対象項目名】全文
【補正方法】変更
【補正の内容】
【特許請求の範囲】
【請求項１】
ニアメモリ削減ユニットであって、
削減アドレス範囲を格納するように構成された削減アドレスレジスタと、
データ削減結果を格納するように構成された結果レジスタと、
データ削減ロジックと、を備え、
前記データ削減ロジックは、
オフダイメモリにデータを格納するストア動作の宛先アドレスが、前記削減アドレスレジスタに格納された削減アドレス範囲内にあると判別したことに応じて、
前記ストア動作のデータ値を削減操作によって処理して、更新されたデータ削減結果を生成することと、
前記更新されたデータ削減結果を前記結果レジスタに格納することと、
を行うように構成されている、
ニアメモリ削減ユニット。
【請求項２】
前記データ削減ロジックは、前記削減アドレス範囲を、前記ニアメモリ削減ユニットの前記削減アドレス範囲をプログラムする操作によって指定された値に初期化するように構成されている、
請求項１のニアメモリ削減ユニット。
【請求項３】
前記削減アドレス範囲は、連続するメモリアドレスの範囲である、
請求項１のニアメモリ削減ユニット。
【請求項４】
前記削減アドレス範囲は、メモリアドレスのストライド範囲である、
請求項１のニアメモリ削減ユニット。
【請求項５】
前記結果レジスタに格納された前記データ削減結果は、前記削減アドレスレジスタに格納された前記削減アドレス範囲の変化又は初期化操作に応じて、指定された値に初期化される、
請求項１のニアメモリ削減ユニット。
【請求項６】
前記結果レジスタの更新の現在のカウントを格納するように構成されたカウンタレジスタをさらに備え、前記データ削減ロジックは、データを前記オフダイメモリに格納する前記ストア動作の前記宛先アドレスが、前記削減アドレスレジスタに格納された前記削減アドレス範囲内にあると判別したことに応じて、前記結果レジスタに格納された前記更新の現在のカウントをインクリメントするように構成されている、
請求項１のニアメモリ削減ユニット。
【請求項７】
前記データ削減ロジックは、
データをオフダイメモリに格納する前記ストア動作の前記データ値を処理して、処理済みのデータ値を生成することと、
前記処理済みのデータ値を前記削減操作によって処理して、更新された削減結果を生成することと、

を行うように構成されている、
請求項１のニアメモリ削減ユニット。
【請求項８】
前記データ値を処理することは、前記データ値を関数で処理すること、又は、前記データ値に定数を適用すること、のうち１つ以上を含む、
請求項７のニアメモリ削減ユニット。
【請求項９】
前記オフダイメモリはメインメモリである、
請求項１のニアメモリ削減ユニット。
【請求項１０】
オフダイメモリにデータを格納する前記ストア動作によって指定された前記データがキャッシュに格納され、
前記ストア動作は、前記ストア動作によって指定された前記データを、前記キャッシュからメインメモリに格納するために実行される、
請求項１のニアメモリ削減ユニット。
【請求項１１】
前記ストア動作は第１のストア動作であり、前記データ削減ロジックは、前記第１のストア動作の前記宛先アドレスと同じ宛先アドレスを有する第２のストア動作を受信したことに応じて、エラービットを設定するように構成されている、
請求項１のニアメモリ削減ユニット。
【請求項１２】
前記ニアメモリ削減ユニットは、メモリコントローラ、メモリモジュールに割り当てられるスタンドアロンエンティティ、又は、スタックメモリパッケージのダイのうち１つ以上に実装されている、
請求項１のニアメモリ削減ユニット。
【請求項１３】
前記削減操作は、前記ストア動作の前記データ値を、前記結果レジスタに格納された累積データ削減結果に加算する加算演算である、
請求項１のニアメモリ削減ユニット。

【国際調査報告】

INTERNATIONAL SEARCH REPORT

International application No
PCT/US2020/041910

A. CLASSIFICATION OF SUBJECT MATTER
INV. G06F12/02 G06F12/0875
ADD. G06N3/063

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 G06N

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 practicable, search terms used)
EPO-Internal

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 2016/098200 A1 (GUZ ZVI [US] ET AL) 7 April 2016 (2016-04-07) paragraphs [0004], [0027], [0048], [0052] - [0057], [0072], [0075] figures 1B,7	1-20
A	US 2018/365438 A1 (BHATTACHARYYA BINATA [US] ET AL) 20 December 2018 (2018-12-20) paragraphs [0035], [0049] figures 2,4	1-20
A	US 2019/311266 A1 (LIU SHAOLI [CN] ET AL) 10 October 2019 (2019-10-10) paragraph [0053] - paragraph [0054] figure 8	1-20

☐ Further documents are listed in the continuation of Box C.

☒ 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 application or patent 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

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
24 September 2020	07/10/2020

Name and mailing address of the ISA/	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Fax: (+31-70) 340-3016	Braccini, Guido

Form PCT/ISA/210 (second sheet) (April 2005)

INTERNATIONAL SEARCH REPORT

Information on patent family members

International application No
PCT/US2020/041910

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 2016098200 A1	07-04-2016	CN 105487812 A	13-04-2016
		KR 20160039544 A	11-04-2016
		TW 201626207 A	16-07-2016
		US 2016098200 A1	07-04-2016
US 2018365438 A1	20-12-2018	CN 107851170 A	27-03-2018
		EP 3326107 A1	30-05-2018
		TW 201717030 A	16-05-2017
		US 2017024573 A1	26-01-2017
		US 2018365438 A1	20-12-2018
		WO 2017014858 A1	26-01-2017
US 2019311266 A1	10-10-2019	CN 108320018 A	24-07-2018
		CN 108334944 A	27-07-2018
		CN 111126590 A	08-05-2020
		CN 111160547 A	15-05-2020
		EP 3561732 A1	30-10-2019
		US 2019311266 A1	10-10-2019
		WO 2018113790 A1	28-06-2018

Form PCT/ISA/210 (patent family annex) (April 2005)

フロントページの続き

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

アメリカ合衆国 ９５０５４ カリフォルニア州、サンタ クララ、オーガスティン ドライブ ２４８５

(72)発明者 シャイジーン アガ

アメリカ合衆国 ９５０５４ カリフォルニア州、サンタ クララ、オーガスティン ドライブ ２４８５

Ｆターム（参考） 5B160 AA12 AA13 CB03 MM20

5B205 MM01 NN89 NN95