(19)日本国特許庁(JP)

(12)特許公報(B2)

(11)特許番号
特許第7187635号
(P7187635)

(45)発行日　令和4年12月12日(2022.12.12)

(24)登録日　令和4年12月2日(2022.12.2)

(51)国際特許分類
G06F 17/16 (2006.01)

FI
G06F 17/16 P
G06F 17/16 Q
G06F 17/16 A

請求項の数 15 外国語出願 （全21頁）

(21)出願番号　特願2021-147894(P2021-147894)
(22)出願日　令和3年9月10日(2021.9.10)
(62)分割の表示　特願2019-85959(P2019-85959)の分割
原出願日　平成28年12月26日(2016.12.26)
(65)公開番号　特開2022-781(P2022-781A)
(43)公開日　令和4年1月4日(2022.1.4)
審査請求日　令和3年10月8日(2021.10.8)
(31)優先権主張番号　15/016,420
(32)優先日　平成28年2月5日(2016.2.5)
(33)優先権主張国・地域又は機関
米国(US)

(73)特許権者　502208397
グーグル エルエルシー
Ｇｏｏｇｌｅ ＬＬＣ
アメリカ合衆国 カリフォルニア州 ９４０４３ マウンテン ビュー アンフィシアター パークウェイ １６００
１６００ Ａｍｐｈｉｔｈｅａｔｒｅ Ｐａｒｋｗａｙ ９４０４３ Ｍｏｕｎｔａｉｎ Ｖｉｅｗ， ＣＡ Ｕ．Ｓ．Ａ．
(74)代理人　110001195
弁理士法人深見特許事務所
(72)発明者　ラビ・ナラヤナスワミ
アメリカ合衆国、９４０４３ カリフォルニア州、マウンテン・ビュー、アンフィシアター・パークウェイ、１６００

最終頁に続く

(54)【発明の名称】　疎要素を密行列に変換するためのシステムおよび方法

(57)【特許請求の範囲】

【請求項１】

疎要素を密行列に変換するためのシステムであって、

複数の疎要素アクセスユニットを備え、前記複数の疎要素アクセスユニットの各疎要素アクセスユニットは、

それぞれの制御信号を受信し、

前記それぞれの制御信号に基づいて、前記疎要素アクセスユニットに対応するデータ片に格納された疎要素にアクセスし、

前記データ片から得られる前記疎要素に適用される変換に基づいて出力密行列を生成し、

前記出力密行列を前記システムのノードネットワークに与えるよう構成されている、システム。

【請求項２】

前記それぞれの制御信号に対応する命令を受けるよう構成される疎密変換ユニットをさらに備え、

前記複数の疎要素アクセスユニットは、前記疎密変換ユニットに位置する、請求項１に記載のシステム。

【請求項３】

前記疎密変換ユニットは、行次元および列次元を含む多次元疎密変換ユニットである、請求項２に記載のシステム。

【請求項４】

前記複数の疎要素アクセスユニットは、前記多次元疎密変換ユニットのそれぞれの次元に沿って配置されている、請求項３に記載のシステム。
【請求項５】
前記複数の疎要素アクセスユニットの各々は、
前記疎要素に前記変換を適用して前記出力密行列を生成するように構成されたそれぞれの第１のユニットを含む、請求項２に記載のシステム。
【請求項６】
前記第１のユニットは連結ユニットであり、
前記変換は連結演算に基づく、請求項５に記載のシステム。
【請求項７】
前記第１のユニットは圧縮／伸長ユニットであり、
前記変換は、前記疎要素を圧縮する演算に基づく、請求項５に記載のシステム。
【請求項８】
複数の疎要素アクセスユニットを含むシステムを用いて疎要素を密行列に変換するための方法であって、
疎要素アクセスユニットがそれぞれの制御信号を受信することと、
前記それぞれの制御信号に基づいて、前記疎要素アクセスユニットが、前記疎要素アクセスユニットに対応するデータ片に格納された疎要素にアクセスすることと、
前記データ片から得られる前記疎要素に前記疎要素アクセスユニットによって適用される変換に基づいて、出力密行列を生成することと、
前記出力密行列を前記システムのノードネットワークに与えることとを含む、方法。
【請求項９】
前記システムは、命令を受けるように構成された疎密変換ユニットを含み、
前記複数の疎要素アクセスユニットは、前記疎密変換ユニットに位置しており、
前記方法は、前記疎密変換ユニットが、前記複数の疎要素アクセスユニットの各々についてそれぞれの制御信号に対応する命令を受けることを含む、請求項８に記載の方法。
【請求項１０】
前記疎密変換ユニットは、行次元および列次元を含む多次元疎密変換ユニットである、請求項９に記載の方法。
【請求項１１】
前記複数の疎要素アクセスユニットは、前記多次元疎密変換ユニットのそれぞれの次元に沿って配置される、請求項１０に記載の方法。
【請求項１２】
前記複数の疎要素アクセスユニットの各々は、前記疎要素を変換するように構成されたそれぞれの第１のユニットを含み、
前記方法は、前記第１のユニットが前記疎要素に前記変換を適用して前記出力密行列を生成することを含む、請求項９に記載の方法。
【請求項１３】
前記第１のユニットは連結ユニットであり、
前記変換を適用することは、
　前記疎要素に連結演算を適用することと、
　前記連結演算に基づいて前記疎要素を連結して前記出力密行列を生成することとを含む、請求項１２に記載の方法。
【請求項１４】
前記第１のユニットは圧縮／伸長ユニットであり、
前記変換を適用することは、
前記疎要素に圧縮演算を適用することと、
前記圧縮演算に基づいて前記疎要素を圧縮して前記出力密行列を生成することとを含む、請求項１２に記載の方法。
【請求項１５】

システムの疎要素アクセスユニットによって疎要素を密行列に変換するために用いられる命令を格納したコンピュータプログラムであって、前記命令は、請求項８～１４のいずれか１項に記載の方法を引き起こすようプロセッサによって実行可能である、コンピュータプログラム。

【発明の詳細な説明】

【技術分野】

【０００１】

発明の詳細な記載

背景

この明細書は、一般に、回路系を用いて行列を処理することに関する。

【発明の概要】

【発明が解決しようとする課題】

【０００２】

概要

この明細書に記載される主題の１つの革新的な局面によれば、行列プロセッサを用いて、疎から密への、または密から疎への行列変換を実行することができる。一般に、高性能計算システムは、行列を処理するために線形代数ルーチンを用い得る。いくつかの例においては、行列のサイズは１つのデータストレージにはまるには大きすぎるかもしれず、行列の異なる部分は、分散型データストレージシステムの異なる位置に疎に格納され得る。行列をロードするために、計算システムの中央処理ユニットは、別の回路系に行列の異なる部分にアクセスするよう命令し得る。この回路系は、ネットワークトポロジーに従って構成された複数のメモリコントローラを含んでもよく、疎データは、予め定められるルールの組に基いて、区分され格納されてもよい。各メモリコントローラは、予め定められるルールの組に基いて疎データを集めて、疎データ上において同時計算を実行し、および、中央処理ユニットがその後の処理を実行するために、ともに連結することができる密行列を生成してもよい。

【課題を解決するための手段】

【０００３】

一般に、この明細書に記載される主題の１つの革新的な局面は、疎要素を密行列に変換するためのシステムにおいて実施することができる。このシステムは、第１の密行列と関連付けられる疎要素をフェッチするよう構成された疎要素アクセスユニットの第１の群と、第１の密行列とは異なる第２の密行列と関連付けられる疎要素をフェッチするよう構成された疎要素アクセスユニットの第２の群とを備える。システムは、第１の密行列と関連付けられる疎要素および第２の密行列と関連付けられる疎要素を含む疎要素に基いた出力行列に対する要求を受取り、疎要素アクセスユニットの第１の群によってフェッチされる第１の密行列と関連付けられる疎要素を得、疎要素アクセスユニットの第２の群によってフェッチされる第２の密行列と関連付けられる疎要素を得、第１の密行列と関連付けられる疎要素および第２の密行列と関連付けられる疎要素を変換して、第１の密行列と関連付けられる疎要素および第２の密行列と関連付けられる疎要素を含む出力密行列を生成するよう構成される。

【０００４】

これらおよび他の実現例は、各々、任意で以下の特徴の１つ以上を含むことができる。たとえば、疎要素アクセスユニットの第１の群は第１の疎要素アクセスユニットおよび第２の疎要素アクセスユニットを含んでもよい。第１の疎要素アクセスユニットは、第１の密行列と関連付けられる疎要素の第１の部分集合をフェッチするよう構成されてもよい。第２の疎要素アクセスユニットは、第１の密行列と関連付けられる疎要素の第２の異なる部分集合をフェッチするよう構成されてもよい。

【０００５】

第１の疎要素アクセスユニットは、第１の密行列と関連付けられる疎要素および第２の密行列と関連付けられる疎要素を含む複数個の疎要素に対する要求を受取り、要求を第２

の疎要素アクセスユニットに送信するよう構成される。第１の疎要素アクセスユニットは、複数個の疎要素のうちの特定の疎要素のアイデンティティが、第１の密行列と関連付けられる疎要素の第１の部分集合のうちの１つのアイデンティティと一致する、と判断してもよい。複数個の疎要素のうちの特定の疎要素のアイデンティティが、第１の密行列と関連付けられる疎要素の第１の部分集合のうちの１つのアイデンティティと一致する、と判断することに応じて、第１の疎要素アクセスユニットは、特定の疎要素を含む第１の密行列と関連付けられる疎要素の第１の部分集合をフェッチするよう構成されてもよい。

【０００６】

第１の疎要素アクセスユニットは、第１のデータ片から第１の密行列と関連付けられる疎要素の第１の部分集合をフェッチするよう構成されてもよく、第２の疎要素アクセスユニットは、第２の異なるデータ片から第１の密行列と関連付けられる疎要素の第２の異なる部分集合をフェッチするよう構成されてもよい。第１の疎要素アクセスユニットは、第１の密行列と関連付けられる疎要素の第１の部分集合を変換して第３の密行列を生成するよう構成されてもよく、第２の疎要素アクセスユニットは、第３の密行列を受け、第２の密行列と関連付けられる疎要素の第２の部分集合を変換して第４の密行列を生成し、第３の密行列を第４の密行列とともに変換して、第１の密行列と関連付けられる疎要素の第１の部分集合および第１の密行列と関連付けられる疎要素の第２の部分集合を含む第５の密行列を生成するよう構成されてもよい。

【０００７】

疎要素アクセスユニットの第１の群および疎要素アクセスユニットの第２の群は二次元のメッシュ構成で配列されてもよい。疎要素アクセスユニットの第１の群および疎要素アクセスユニットの第２の群は二次元の円環面構成で配列されてもよい。第１の密行列と関連付けられる疎要素および第２の密行列と関連付けられる疎要素は多次元の行列であってもよく、出力密行列はベクトルであってもよい。

【０００８】

この明細書において記載される主題は、以下の利点の１つ以上を実現するように特定の実施の形態において実現することができる。ネットワークトポロジーに従ってメモリコントローラユニットを接続することは、予め定められるルールの組に従う疎データの格納の区分化を可能にする。中央処理ユニットから別の回路系に疎密データロードタスクをシフトすることは、中央処理ユニットの計算帯域幅を増大し、システムの処理費を低減する。特殊化された回路系を用いることによって、疎データをフェッチするために密な線形代数に対して特殊化されるプロセッサの使用を回避することができる。分散型システムにおいて多数のメモリを同時に用いることによって、分散型システムにおいて利用可能な和集合帯域幅は、直列化を必要とし、集合の帯域幅上において単一のメモリキャップを有する単一のメモリバンクに対する帯域幅よりも高い。

【０００９】

この局面および他の局面の他の実現例は、計算機記憶装置上でエンコードされる、方法のアクションを実行するように構成される、対応のシステム、装置およびコンピュータプログラムを含む。１つ以上のコンピュータのシステムは、システムにインストールされ、動作でシステムにアクションを実行させるソフトウェア、ファームウェア、ハードウェアまたはそれらの組合せによってそのように構成することができる。１つ以上のコンピュータプログラムは、データ処理装置によって実行されたとき、装置にアクションを実行させる命令を有することによって、そのように構成することができる。

【００１０】

この明細書に記載される主題の１つ以上の実現例の詳細は、添付の図面および以下の記載において述べられる。主題の他の潜在的な特徴、局面および利点は、記載、図面および特許請求の範囲から明らかになる。

【図面の簡単な説明】

【００１１】

【図１】例示の計算システムのブロック図である。

【図２Ａ】例示の疎密変換ユニットを示す図である。
【図２Ｂ】例示の疎密変換ユニットを示す図である。
【図２Ｃ】例示の疎密変換ユニットを示す図である。
【図２Ｄ】例示の疎密変換ユニットを示す図である。
【図３Ａ】例示の疎要素アクセスユニットを示す図である。
【図３Ｂ】例示の疎要素アクセスユニットを示す図である。
【図４】密行列を生成するためのプロセスの例を示すフローチャート図である。
【図５】疎要素を密行列に変換するためのプロセスの例を示すフローチャート図である。
【発明を実施するための形態】
【００１２】
さまざまな図面における同様の参照番号および指定は同様の要素を示す。
詳細な記載
一般に、データは行列の形式において表すことができ、計算システムは線形代数アルゴリズムを用いてデータを操作し得る。行列は一次元のベクトルまたは多次元行列であり得る。行列は、データベーステーブルまたは変数などのようなデータ構造によって表されてもよい。しかしながら、行列のサイズが大きすぎると、１つのデータストレージに行列全体を格納することは可能ではないかもしれない。密行列は複数の疎要素に変換され得、各疎要素は異なるデータストレージに格納され得る。密行列の疎要素は行列であってもよく、行列のうちの小さな部分行列（たとえば単一値要素、行、列、または部分行列）のみが非零値を有する。計算システムが密行列にアクセスすることを必要とするときに、中央処理ユニット（ＣＰＵ）は、データストレージの各々に到達するスレッドを開始して、格納された疎要素をフェッチしてもよく、そして、疎密変換を適用して密行列を戻す。しかしながら、それが疎要素すべてをフェッチするのにかかる時間の量は長いかもしれず、ＣＰＵの計算帯域幅は結果として十分に利用されないかもしれない。いくつかの場合では、計算システムはいくつかの密行列の疎要素にアクセスして新たな密行列を形成する必要があるかもしれず、それらの密行列は等しい次元を有さないかもしれない。異なる密行列の疎要素をフェッチするようデータストレージの各々に到達するスレッドと関連付けられるＣＰＵアイドル時間は、異なる待ち時間に遭遇し得、さらに、計算装置の性能に望ましくない態様で影響を与えるかもしれない。いくつかの場合では、計算システムはいくつかの密行列の疎要素にアクセスして新たな密行列を形成する必要があるかもしれず、それらの疎要素は等しい次元を有さないかもしれない。異なる密行列の疎要素をフェッチするようデータストレージの各々に到達するスレッドと関連付けられるＣＰＵアイドル時間は、異なる待ち時間に遭遇し得、さらに、計算装置の性能に望ましくない態様で影響を与えるかもしれない。ＣＰＵから分離しているハードウェア疎密変換ユニットは、プロセッサの計算帯域幅をＣＰＵ動作から独立した疎要素の収集および疎要素の密行列への変換によって、増大させ得る。
【００１３】
図１は、１つ以上の密行列から疎要素を変換して密行列を生成するための例示の計算システム１００のブロック図を示す。計算システム１００は、処理ユニット１０２、疎密変換ユニット１０４およびデータ片１０６ａ～１０６ｋを含み、ｋは１以上の整数である。一般に、処理ユニット１０２は、目標密行列へのアクセスための命令を処理し、目標密行列を生成するために疎密変換ユニット１０４に命令１１０を送信する。疎密変換ユニット１０４はデータ片１０６ａ～１０６ｋの１つ以上から対応する疎要素１０８ａ～１０８ｎにアクセスし、ｎは１つ以上の整数である。疎密変換ユニット１０４は対応する疎要素１０８ａ～１０８ｎを用いて目標密行列１１２を生成し、目標密行列１１２をその後の処理のために処理ユニット１０２に与える。たとえば、疎要素１０８ａ～１０８ｎは異なるサイズを有する二次元の行列であり得、疎密変換ユニット１０４は疎要素１０８ａ～１０８ｎの各々をベクトルに変換することおよびｎ個のベクトルを単一のベクトルに連結することによって目標密行列１１２を生成し得る。
【００１４】

いくつかの実現例では、処理ユニット１０２は、目標密行列の更新ための命令を処理し、更新された密行列を疎密変換ユニット１０４に送信してもよい。疎密変換ユニット１０４は、更新された密行列を対応する疎要素に変換し、したがって、データ片１０６ａ～１０６ｋに格納された１つ以上の疎要素を更新してもよい。

【００１５】

処理ユニット１０２は計算システム１００内で実行のために命令を処理するように構成される。処理ユニット１０２は１つ以上のプロセッサを含んでもよい。いくつかの実現例では、処理ユニット１０２は疎密変換ユニット１０４によって生成された目標密行列１１２を処理するように構成される。他のいくつかの実現例では、処理ユニット１０２は、疎密変換ユニット１０４に目標密行列１１２を生成することを要求するよう構成されてもよく、別の処理ユニットが目標密行列１１２を処理するように構成されてもよい。データ片１０６ａ～１０６ｋは疎要素１０８ａ～１０８ｎを含むデータを格納する。いくつかの実現例では、データ片１０６ａ～１０６ｋは単数または複数の揮発性記憶装置であってもよい。他のいくつかの実現例では、データ片１０６ａ～１０６ｋは単数または複数の不揮発性記憶装置であってもよい。データ片１０６ａ～１０６ｋは、さらに、ストレージエリアネットワークまたは他の構成における装置などのような別のコンピュータ読取可能媒体の形式であってもよい。データ片１０６ａ～１０６ｋは電気的接続、光接続または無線接続を用いて、疎密変換ユニット１０４に結合されてもよい。いくつかの実現例では、データ片１０６ａ～１０６ｋは疎密変換ユニット１０４の一部であってもよい。

【００１６】

疎密変換ユニット１０４は疎要素に基いて密行列を判断するように構成される。いくつかの実現例では、疎密変換ユニット１０４は密行列に基いて疎要素の位置を判断するように構成されてもよい。いくつかの実現例では、図２Ａ～図２Ｄを参照してより詳細に以下に記載されるように、疎密変換ユニット１０４は複数の相互接続された疎要素アクセスユニットを含んでもよい。

【００１７】

図２Ａは例示の疎密変換ユニット２００を示す。疎密変換ユニット２００は疎密変換ユニット１０４に対応してもよい。疎密変換ユニット２００は、Ｍ×Ｎ個の疎要素アクセスユニット$X_{1,1}$～$X_{M,N}$を含み、それらは物理的にまたは論理的にＭ個の行およびＮ個の列に配列され、ＭおよびＮは１以上の整数である。いくつかの実現例では、疎密変換ユニット２００は、データを処理するように構成されたさらなる回路系を含んでもよい。一般に、疎密変換ユニット２００は、密行列に対する要求を受け、疎要素アクセスユニット$X_{1,1}$～$X_{M,N}$によってアクセス可能な対応する疎要素に基いて密行列を判断するよう構成される。一般に、各疎要素アクセスユニットは、指定される疎要素の組にアクセスするよう構成され、図３Ａ～図３Ｂを参照してより詳細に以下に記載される。いくつかの実現例では、疎要素アクセスユニットは単一命令・多重データ（ＳＩＭＤ）処理装置であってもよい。

【００１８】

いくつかの実現例では、疎要素アクセスユニット$X_{1,1}$～$X_{M,N}$は、二次元のメッシュ構成に物理的にまたは論理的に配列されてもよい。たとえば疎要素アクセスユニット$X_{1,1}$は、疎要素アクセスユニット$X_{1,2}$および$X_{2,1}$に直接結合される。別の例として、疎要素アクセスユニット$X_{2,2}$は、疎要素アクセスユニット$X_{2,1}$、$X_{3,1}$、$X_{2,3}$および$X_{1,2}$に直接結合される。２つの疎要素アクセスユニット間の結合は、電気的接続、光接続、無線接続または任意の他の好適な接続であってもよい。

【００１９】

他のいくつかの実現例では、疎要素アクセスユニット$X_{1,1}$～$X_{M,N}$は、二次元の円環面構成に物理的にまたは論理的に配列されてもよい。たとえば疎要素アクセスユニット$X_{1,1}$は、疎要素アクセスユニット$X_{1,2}$、$X_{2,1}$、$X_{1,N}$および$X_{M,1}$に直接結合される。別の例として、疎要素アクセスユニット$X_{M,N}$は、疎要素アクセスユニット$X_{M,N-1}$、$X_{M-1,N}$、$X_{M,1}$および$X_{1,N}$に直接結合される。

【００２０】
いくつかの実現例では、疎密変換ユニット２００は、予め定められる条件の組に従って密行列から変換される疎要素を区分するよう構成されてもよい。疎要素アクセスユニット$X_{1,1}$～$X_{M,N}$の各行は、特定の密行列から変換される疎要素にアクセスするよう区分されてもよい。たとえば、疎密変換ユニット２００は、コンピュータモデルの１，０００個の異なるデータベーステーブルに対応する密行列から変換される疎要素にアクセスするよう構成されてもよい。データベーステーブルの１つ以上は異なるサイズを有してもよい。疎要素アクセスユニットの第１番目の行２０２は、データベーステーブル１番～データベーステーブル１００番から変換される疎要素にアクセスするよう構成されてもよく、疎要素アクセスユニットの第２番目の行２０４は、データベーステーブル１０１番～データベーステーブル３００番から変換される疎要素にアクセスするよう構成されてもよく、疎要素アクセスユニットのＭ番目の行２０６は、データベーステーブル７５１番～データベーステーブル１，０００番から変換される疎要素にアクセスするよう構成されてもよい。いくつかの実現例では、区分は、疎密変換ユニット２００を用いて、プロセッサが疎要素にアクセスする前に、ハードウェア命令によって構成されてもよい。
【００２１】
疎要素アクセスユニット$X_{1,1}$～$X_{M,N}$の各列は、特定の密行列から変換される疎要素の部分集合にアクセスするよう区分されてもよい。たとえば、データベーステーブル１番に対応する密行列は１，０００個の疎要素に変換されてもよく、１，０００個の疎要素は上に記載されるような第１番目の行２０２によってアクセス可能である。疎要素アクセスユニット$X_{1,1}$は、データベーステーブル１番の疎要素１番～２００番にアクセスするよう構成されてもよく、疎要素アクセスユニット$X_{1,2}$は、データベーステーブル１番の疎要素２０１番～５００番にアクセスするよう構成されてもよい。別の例として、データベーステーブル２番に対応する密行列は５００個の疎要素に変換されてもよく、５００の疎要素は上に記載されるような第１番目の行２０２によってアクセス可能である。疎要素アクセスユニット$X_{1,1}$は、データベーステーブル２番の疎要素１番～５０番にアクセスするよう構成されてもよく、疎要素アクセスユニット$X_{1,2}$は、データベーステーブル２番の疎要素５１番～２００番にアクセスするよう構成されてもよい。別の例として、データベーステーブル１，０００番に対応する密行列は１０，０００個の疎要素に変換されてもよく、１０，０００個の疎要素は上に記載されるような第Ｍ番目の行２０６によってアクセス可能である。疎要素アクセスユニット$X_{M,1}$は、データベーステーブル１，０００番の疎要素１番～２，０００番にアクセスするよう構成されてもよく、疎要素アクセスユニット$X_{M,N}$は、データベーステーブル１，０００番の疎要素９，０００番～１０，０００番にアクセスするよう構成されてもよい。
【００２２】
図２Ｂは、疎密変換ユニット２００が、疎要素アクセスユニットの二次元のメッシュネットワークを用いて、どのように疎要素を要求し得るかの例を示す。例として、処理ユニットは、疎密変換ユニット２００に対して、データベーステーブル１番の疎要素１番～５０番、データベーステーブル２番の疎要素１００番～２００番、およびデータベーステーブル１，０００番の疎要素９，０５０番～９，０６０番を用いて生成される密な一次元のベクトルを要求する命令を実行してもよい。疎密変換ユニット２００は、処理ユニットから要求を受けた後、疎密変換ユニット２００は、疎要素アクセスユニット$X_{1,1}$に命令して、疎要素に対する要求をメッシュネットワークにおける他の疎要素アクセスユニットに同報通信させてもよい。疎要素アクセスユニット$X_{1,1}$は、疎要素アクセスユニット$X_{1,2}$に要求２２２を、および疎要素アクセスユニット$X_{2,1}$に要求２２４を同報通信してもよい。要求２２２を受けた後、疎要素アクセスユニット$X_{1,2}$は、要求２２６を疎要素アクセスユニット$X_{1,3}$に同報通信してもよい。いくつかの実現例では、疎要素アクセスユニットは、ルーティングスキームに基いて別の疎要素アクセスユニットに要求を同報通信するよう構成されてもよい。たとえば、疎要素アクセスユニット$X_{1,2}$は疎要素アクセスユニット$X_{2,2}$に要求を同報通信するよう構成されなくてもよく、なぜならば、疎要素

アクセスユニット$X_{2,2}$は疎要素アクセスユニット$X_{2,1}$から同報通信を受けるよう構成されるからである。ルーティングスキームは静的であってもよく、または動的に生成されてもよい。たとえば、ルーティングスキームはルックアップテーブルであってもよい。いくつかの実現例では、疎要素アクセスユニットは、要求２２４を要求２２４に基いて別の疎要素アクセスユニットに同報通信するよう構成されてもよい。たとえば、要求２２４は要求された疎要素の識別を含んでもよく（たとえばデータベーステーブル１番、疎要素１番～５０番）、疎要素アクセスユニット$X_{1,2}$は、要求２２４を疎要素アクセスユニット$X_{2,2}$および／または疎要素アクセスユニット$X_{1,3}$に同報通信するべきであるかどうかを、識別に基づいて判断してもよい。同報通信プロセスは、メッシュネットワークを介して伝搬し、疎要素アクセスユニット$X_{M,N}$は疎要素アクセスユニット$X_{M,N-1}$から要求２３０を受ける。

【００２３】

図２Ｃは、疎密変換ユニット２００が、疎要素アクセスユニットの二次元のメッシュネットワークを用いて、要求される密行列をどのように生成し得るかの例を示す。いくつかの実現例では、或る疎要素アクセスユニットが同報通信された要求を受けた後、その疎要素アクセスユニットは、それは要求される疎要素のいずれかにアクセスするよう構成されるかどうかを判断するよう構成される。たとえば疎要素アクセスユニット$X_{1,1}$は、それは、データベーステーブル１番の疎要素１番～５０番にアクセスするよう構成されるが、データベーステーブル２番の疎要素１００番～２００番またはデータベーステーブル１，０００番の疎要素９，０５０番～９，０６０番にアクセスするようには構成されない、と判断してもよい。それがデータベーステーブル１番の疎要素１番～５０番にアクセスするよう構成されると判断することに応じて、疎要素アクセスユニット$X_{1,1}$は、データベーステーブル１番の疎要素１番～５０番を、これらの疎要素が格納されているデータ片からフェッチし、これらの疎要素に基いて密行列２４２を生成してもよい。

【００２４】

別の例として、疎要素アクセスユニット$X_{2,1}$は、それが、データベーステーブル１番の疎要素１番～５０番、データベーステーブル２番の疎要素１００番～２００番、またはデータベーステーブル１，０００番の疎要素９，０５０番～９，０６０番のいずれにもアクセスするよう構成されない、と判断してもよい。それが要求される疎要素のいずれにもアクセスするよう構成されないと判断することに応じて、疎要素アクセスユニット$X_{2,1}$はさらなるアクションを実行しなくてもよい。

【００２５】

別の例として、疎要素アクセスユニット$X_{1,2}$は、それはデータベーステーブル２番の疎要素１００番～２００番にアクセスするよう構成されるが、データベーステーブル１番の疎要素１番～５０番またはデータベーステーブル１，０００番の疎要素９，０５０番～９，０６０番にアクセスするようには構成されない、と判断してもよい。それがデータベーステーブル２番の疎要素１００番～２００番にアクセスするよう構成されると判断することに応じて、疎要素アクセスユニット$X_{1,2}$は、これらの疎要素が格納されているデータ片からこれらの疎要素をフェッチし、これらの疎要素に基いて密行列２４４を生成してもよい。いくつかの実現例では、ある疎要素アクセスユニットが、密行列を生成した後、その疎要素アクセスユニットは、同報通信された要求の送信側にその密行列を転送するよう構成されてもよい。ここでは、疎要素アクセスユニット$X_{1,2}$は密行列２４４を疎要素アクセスユニット$X_{1,1}$に転送する。

【００２６】

別の例として、疎要素アクセスユニット$X_{M,N}$は、それはデータベーステーブル１，０００番の疎要素９，０５０番～９，０６０番にアクセスするよう構成されるが、データベーステーブル１番の疎要素１番～５０番またはデータベーステーブル２番の疎要素１００番～２００番にアクセスするようには構成されない、と判断してもよい。それがデータベーステーブル１，０００番の疎要素９，０５０番～９，０６０番にアクセスするよう構成されると判断することに応じて、疎要素アクセスユニット$X_{M,N}$は、これらの疎要素が格

納されているデータ片からこれらの疎要素をフェッチし、これらの疎要素に基いて密行列２４６を生成してもよい。いくつかの実現例では、ある疎要素アクセスユニットが、密行列を生成した後、その疎要素アクセスユニットは、同報通信された要求の送信側にその密行列を転送するよう構成されてもよい。ここでは、疎要素アクセスユニット$X_{M,N}$は密行列２４６を疎要素アクセスユニット$X_{M,N-1}$に転送する。次のサイクルで、疎要素アクセスユニット$X_{M,N-1}$は、密行列２４６を疎要素アクセスユニット$X_{M,N-1}$に転送するよう構成される。このプロセスは、疎要素アクセスユニット$X_{2,1}$が密行列２４６を疎要素アクセスユニット$X_{1,1}$に転送するまで継続する。

【００２７】

いくつかの実現例では、疎密変換ユニット２００は、疎要素アクセスユニットによって生成された密行列を変換し、プロセッサユニットのための密行列を生成するように構成される。ここで、疎密変換ユニット２００は、密行列２４２、２４４および２４６を、プロセッサユニットのための密行列に変換する。たとえば、密行列２４２は１００×１０の次元を有してもよく、密行列２４４は２０×１００の次元を有してもよく、密行列２４６は３×３の次元を有してもよい。疎密変換ユニット２００は、密行列２４２、２４４および２４６を、１×３００９の次元でベクトルに変換してもよい。有利なことに、密行列（たとえばデータベーステーブル）に従う行の区分化は、疎密変換ユニット２００が、生成された密行列が列Ｎから列１に伝搬した後に、要求された疎要素をすべて得ることを可能にする。列の区分化は、疎要素アクセスユニットのわずか１つを用いてあまりにも多数の疎要素にアクセスすることによって引起される帯域幅ボトルネックを低減する。

【００２８】

図２Ｄは、疎密変換ユニット２００が、疎要素アクセスユニットの二次元のメッシュネットワークを用いて、密行列に基いて疎要素をどのように更新し得るかの例を示す。例として、処理ユニットは、疎密変換ユニット２００に対して、データベーステーブル１番の疎要素１番～５０番およびデータベーステーブル１，０００番の疎要素９，０５０番～９，０６０番を用いて生成される密な一次元のベクトルを用いて、格納された疎要素を更新するよう要求する命令を実行してもよい。疎密変換ユニット２００は、処理ユニットから要求を受けた後、疎密変換ユニット２００は疎要素アクセスユニット$X_{1,1}$に対して、疎要素更新要求をメッシュネットワークにおける他の疎要素アクセスユニットに同報通信するよう命令してもよく、疎要素更新要求は、処理ユニットによって与えられる密な一次元のベクトルを含んでもよい。いくつかの実現例では、疎要素アクセスユニット$X_{1,1}$は、それが密な一次元のベクトルに含まれる疎要素にアクセスするよう割当てられるかどうかを判断してもよい。それが密な一次元のベクトルに含まれる疎要素にアクセスするよう割当てられると判断することに応じて、疎要素アクセスユニット$X_{1,1}$は、データ片に格納される疎要素を更新してもよい。ここで、疎要素アクセスユニット$X_{1,1}$は、それがデータベーステーブル１番の疎要素１番～５０番に割当てられると判断し、疎要素アクセスユニット$X_{1,1}$は、データ片におけるこれらの疎要素を更新するよう命令を実行する。

【００２９】

疎要素アクセスユニット$X_{1,1}$は、疎要素アクセスユニット$X_{1,2}$に疎要素更新要求２５２を、および疎要素アクセスユニット$X_{2,1}$に疎要素更新要求２５４を同報通信してもよい。疎要素更新要求２５２を受けた後、疎要素アクセスユニット$X_{1,2}$は、それは、密な一次元のベクトルに含まれる疎要素にアクセスするよう割当てられない、と判断してもよい。疎要素アクセスユニット$X_{1,2}$は、要求２５６を疎要素アクセスユニット$X_{1,3}$に同報通信する。同報通信プロセスは、メッシュネットワークを介して伝搬し、疎要素アクセスユニット$X_{M,N}$は疎要素アクセスユニット$X_{M,N-1}$から要求２６０を受ける。ここで、疎要素アクセスユニット$X_{M,N}$は、それがデータベーステーブル１，０００番の疎要素９，０５０番～９，０６０番に割当てられると判断し、疎要素アクセスユニット$X_{M,N}$はデータ片におけるこれらの疎要素を更新するよう命令を実行する。

【００３０】

図３Ａは例示の疎要素アクセスユニット３００を示す。疎要素アクセスユニット３００

は、疎要素アクセスユニット$X_{1,1}$～$X_{M,N}$の任意の１つであってもよい。一般に、疎要素アクセスユニット３００は、ノードネットワーク３２０から、１つ以上のデータ片に格納された疎要素をフェッチし、フェッチされた疎要素を密行列に変換する要求３４２を受けるよう構成される。いくつかの実現例では、処理ユニット３１６が、ノードネットワーク３２０における疎要素アクセスユニットに対して、疎要素を用いて生成された密行列を求める要求を送信する。疎要素アクセスユニットは、疎要素アクセスユニット３００に要求３４２を同報通信してもよい。同報通信された要求３４２のルーティングは図２Ｂにおける記載に類似してもよい。疎要素アクセスユニット３００は、要求識別ユニット３０２、データフェッチユニット３０４、疎低減ユニット３０６、連結ユニット３０８、圧縮／伸長ユニット３１０、および分割ユニット３１２を含む。ノードネットワーク３２０は二次元のメッシュネットワークであってもよい。処理ユニット３１６は処理ユニット１０２と類似してもよい。

【００３１】

一般に、要求識別ユニット３０２は、１つ以上のデータ片３３０に格納された疎要素をフェッチするよう要求３４２を受け、疎要素アクセスユニット３００は要求３４２によって示された疎要素にアクセスするよう割当てられるかどうかを判断するよう構成される。いくつかの実現例では、要求識別ユニット３０２は、疎要素アクセスユニット３００が要求３４２によって示される疎要素にアクセスするよう割当てられるかどうかを、ルックアップテーブルを用いることによって判断してもよい。特定の要求された疎要素の識別（たとえば、データベーステーブル１番の１番）がルックアップテーブルに含まれている場合には、要求識別ユニット３０２は、特定の要求された疎要素をフェッチするよう、信号３４４をデータフェッチユニット３０４に送信してもよい。特定の要求された疎要素の識別（たとえば、データベーステーブル１番の１番）がルックアップテーブルに含まれない場合には、要求識別ユニット３０２は受取った要求を破棄してもよい。いくつかの実現例では、要求識別ユニット３０２は、受取った要求をノードネットワーク３２０上における別の疎要素アクセスユニットに同報通信するよう構成されてもよい。

【００３２】

データフェッチユニット３０４は、信号３４４を受信することに応じて、データ片３３０から１つ以上の要求された疎要素をフェッチするよう構成される。いくつかの実現例では、データフェッチユニット３０４は１つ以上のプロセッサ３２２ａ～３２２ｋを含み、ｋは整数である。プロセッサ３２２ａ～３２２ｋは、ベクトル処理ユニット（ＶＰＵ）、アレイ処理ユニットまたは任意の好適な処理ユニットであってもよい。いくつかの実現例では、プロセッサ３２２ａ～３２２ｋは、データ片３３０近くに配置されて、プロセッサ３２２ａ～３２２ｋとデータ片３３０との間のレイテンシを低減するようにする。疎要素アクセスユニット３００がフェッチするよう割当てられる、要求された疎要素の数に基いて、データフェッチユニット３０４は、プロセッサ３２２ａ～３２２ｋ間に分散されるべき１つ以上の要求を発生させるよう構成されてもよい。いくつかの実現例では、プロセッサ３２２ａ～３２２ｋの各々は、疎要素の識別に基いて特定の疎要素に割当てられてもよく、データフェッチユニット３０４は、プロセッサ３２２ａ～３２２ｋに対する１つ以上の要求を疎要素の識別に基いて発生させるよう構成されてもよい。いくつかの実現例では、データフェッチユニット３０４はルックアップテーブルを用いることによってプロセッサ割当を判断してもよい。いくつかの実現例では、データフェッチユニット３０４は、プロセッサ３２２ａ～３２２ｋのために複数のバッチを生成してもよく、各バッチは要求された疎要素の部分集合に対する要求である。プロセッサ３２２ａ～３２２ｋは、割当てられた疎要素をデータ片３３０から独立してフェッチし、フェッチされた疎要素を疎低減ユニット３０６に転送するよう構成される。

【００３３】

疎低減ユニット３０６はフェッチされた疎要素３４６の次元を低減するように構成される。たとえば、プロセッサ３２２ａ～３２２ｋの各々は、１００×１の次元を有する疎要素を生成してもよい。疎低減ユニット３０６は、１００×ｋの次元を有する、フェッチさ

れた疎要素３４６を受け、フェッチされた疎要素３４６の次元を論理演算、算術演算または両方の組合せによって１００×１に低減することによって疎低減要素３４８を生成してもよい。疎低減ユニット３０６は疎低減要素３４８を連結ユニット３０８に出力するように構成される。

【００３４】

連結ユニット３０８は、疎低減要素３４８を再配列および連結して、連結された要素３５０を生成するように構成される。たとえば、疎要素アクセスユニット$X_{1,1}$は、データベーステーブル１番の疎要素１番～２００番にアクセスするよう構成されてもよい。プロセッサ３２２ａは、フェッチされた疎要素１０番を、フェッチされた疎要素５番を返すように構成されるプロセッサ３２２ｂよりも早く、疎低減ユニット３０６に返すかもしれない。連結ユニット３０８は、その後受取られる疎要素５番を、より早く受取られた疎要素１０番の前に順序づけられるように再配列し、疎要素１番～２００番を連結された要素３５０として連結するよう構成される。

【００３５】

圧縮／伸長ユニット３１０は、連結された要素３５０を圧縮して、ノードネットワーク３２０のための密行列３５２を生成するよう構成される。たとえば、圧縮／伸長ユニット３１０は、連結された要素３５０における零値を圧縮して、ノードネットワーク３２０の帯域幅を改善するよう構成されてもよい。いくつかの実現例では、圧縮／伸長ユニット３１０は、受取られた密行列を伸長してもよい。たとえば、疎要素アクセスユニット３００は、ノードネットワーク３２０を介して近隣の疎要素アクセスユニットから密行列を受けてもよい。疎要素アクセスユニット３００は受取られた密行列を伸長してもよく、伸長された密行列を連結された要素３５０と連結して、更新された連結された要素を形成してもよく、それらは圧縮され、次いでノードネットワーク３２０に出力されることができる。

【００３６】

図３Ｂは、疎要素アクセスユニット３００がノードネットワーク３２０から受取られる密行列に基いて疎要素をどのように更新し得るかの例を示す。例として、処理ユニットは、疎密変換ユニットに対して、データベーステーブル１番の疎要素１番～５０番およびデータベーステーブル１，０００番の疎要素９，０５０番～９，０６０番を用いて生成される密な一次元のベクトルを用いて、格納された疎要素を更新するよう要求する命令を実行してもよい。疎密変換ユニットは、処理ユニットから要求を受けた後、疎密変換ユニットは要求３６２を送信して、疎要素アクセスユニット３００に対して、それが密な一次元のベクトルに含まれる疎要素にアクセスするよう割当てられるかどうかを判断するよう命令してもよい。要求識別ユニット３０２は、疎要素アクセスユニット３００が密な一次元のベクトルに含まれる疎要素にアクセスするよう割当てられるかどうかを判断するよう構成される。疎要素アクセスユニット３００が密な一次元のベクトルに含まれる疎要素にアクセスするよう割当てられると判断することに応じて、要求識別ユニット３０２は、データ片において格納された疎要素を更新するよう、指示３６４を分割ユニット３１２に送信してもよい。

【００３７】

分割ユニット３１２は、受取られた密行列を、データ片３３０においてデータフェッチユニット３０４によって更新することができる疎要素に変換するように構成される。たとえば、分割ユニット３１２は、密な一次元のベクトルを複数の疎要素に変換し、データフェッチユニット３０４に対して、疎要素アクセスユニット３００がフェッチするよう割当てられるデータ片３３０において格納された疎要素を更新するよう命令するよう構成されてもよい。

【００３８】

図４は、密行列を生成するためのプロセス４００の例を示すフローチャートである。プロセス４００は、疎密変換ユニット１０４または疎密変換ユニット２００などのようなシステムによって実行されてもよい。システムは、疎要素アクセスユニットの第１の群および疎要素アクセスユニットの第２の群を含んでもよい。たとえば、図２Ａを参照して、疎

密変換ユニット２００は、Ｍ×Ｎ個の疎要素アクセスユニット$X_{1,1}$～$X_{M,N}$を含み、それらは物理的にまたは論理的にＭ個の行およびＮ個の列に配列される。疎要素アクセスユニット$X_{1,1}$～$X_{M,N}$の各行は、特定の密行列から変換される疎要素にアクセスするよう区分されてもよい。いくつかの実現例では、疎要素アクセスユニットの第１の群は第１の疎要素アクセスユニットおよび第２の疎要素アクセスユニットを含んでもよい。たとえば、疎密変換ユニット２００の第１番目の行は疎要素アクセスユニット$X_{1,1}$および、$X_{1,2}$を含んでもよい。いくつかの実現例では、疎要素アクセスユニットの第１の群および疎要素アクセスユニットの第２の群は二次元のメッシュ構成で配列されてもよい。いくつかの実現例では、疎要素アクセスユニットの第１の群および疎要素アクセスユニットの第２の群は二次元の円環面構成で配列されてもよい。

【００３９】

システムは、第１の密行列と関連付けられる疎要素および第２の密行列と関連付けられる疎要素を含む疎要素に基いた出力行列に対する要求を受ける（４０２）。たとえば、図２Ｂを参照して、処理ユニットは、疎密変換ユニット２００に対して、データベーステーブル１番の疎要素１番～５０番、データベーステーブル２番の疎要素１００番～２００番およびデータベーステーブル１，０００番の疎要素９，０５０番～９，０６０番を用いて生成される密な一次元のベクトルを要求する命令を実行してもよい。

【００４０】

いくつかの実現例では、第１の疎要素アクセスユニットは、第１の密行列と関連付けられる疎要素および第２の密行列と関連付けられる疎要素を含む複数個の疎要素に対する要求を受取ってもよい。第１の疎要素アクセスユニットは、第２の疎要素アクセスユニットに要求を送信してもよい。たとえば、図２Ｂを参照して、疎密変換ユニット２００は、処理ユニットから要求を受けた後、疎密変換ユニット２００は、疎要素アクセスユニット$X_{1、1}$に命令して、疎要素に対する要求をメッシュネットワークにおける他の疎要素アクセスユニットに同報通信させてもよい。疎要素アクセスユニット$X_{1、1}$は、要求２２２を疎要素アクセスユニット$X_{1、2}$に同報通信してもよい。

【００４１】

システムは、疎要素アクセスユニットの第１の群によってフェッチされる第１の密行列と関連付けられる疎要素を得る（４０４）。いくつかの実現例では、第１の疎要素アクセスユニットは、複数個の疎要素のうちの特定の疎要素のアイデンティティが、第１の密行列と関連付けられる疎要素の第１の部分集合のうちの１つのアイデンティティと一致する、と判断してもよい。たとえば、図２Ｃを参照して、疎要素アクセスユニット$X_{1,1}$は、データベーステーブル１番の疎要素１番～２００番にアクセスするよう構成されてもよい。疎要素アクセスユニット$X_{1,1}$は、それはデータベーステーブル１番の疎要素１番～５０番にアクセスするよう構成されるが、データベーステーブル２番の疎要素１００番～２００番またはデータベーステーブル１，０００番の疎要素９，０５０番～９，０６０番にアクセスするようには構成されない、と判断してもよい。複数個の疎要素のうちの特定の疎要素のアイデンティティが、第１の密行列と関連付けられる疎要素の第１の部分集合のうちの１つのアイデンティティと一致する、と判断することに応じて、第１の疎要素アクセスユニットは、特定の疎要素を含む第１の密行列と関連付けられる疎要素の第１の部分集合をフェッチしてもよい。たとえば、それがデータベーステーブル１番の疎要素１番～５０番にアクセスするよう構成されると判断することに応じて、疎要素アクセスユニット$X_{1,1}$は、データベーステーブル１番の疎要素１番～５０番を、これらの疎要素が格納されているデータ片からフェッチしてもよい。

【００４２】

第２の疎要素アクセスユニットは、第１の密行列と関連付けられる疎要素の第２の異なる部分集合をフェッチしてもよい。たとえば、図２Ｃを参照して、疎要素アクセスユニット$X_{1,2}$は、データベーステーブル２番の疎要素５１番～２００番にアクセスするよう構成されてもよい。それがデータベーステーブル２番の疎要素１００番～２００番にアクセスするよう構成されると判断することに応じて、疎要素アクセスユニット$X_{1,2}$は、

これらの疎要素が格納されているデータ片からこれらの疎要素をフェッチしてもよい。

【００４３】

システムは、疎要素アクセスユニットの第２の群によってフェッチされる第２の密行列と関連付けられる疎要素を得る（４０６）。たとえば、図２Ｃを参照して、第２の群疎要素アクセスユニットは、Ｍ×Ｎ個の疎要素アクセスユニットのＭ番目の行であってもよく、疎要素アクセスユニット$X_{M,N}$は、データベーステーブル１，０００番の疎要素９，０００番～１０，０００番にアクセスするよう構成されてもよい。それがデータベーステーブル１，０００番の疎要素９，０５０番～９，０６０番にアクセスするよう構成されると判断することに応じて、疎要素アクセスユニット$X_{M,N}$は、これらの疎要素が格納されているデータ片からこれらの疎要素をフェッチし、これらの疎要素に基いて密行列２４６を生成してもよい。

【００４４】

いくつかの実現例では、第１の疎要素アクセスユニットは、第１のデータ片から第１の密行列と関連付けられる疎要素の第１の部分集合をフェッチしてもよく、第２の疎要素アクセスユニットは、第２の異なるデータ片から第１の密行列と関連付けられる疎要素の第２の異なる部分集合をフェッチしてもよい。たとえば、図１を参照して、第１の疎要素アクセスユニットは、データ片１０６ａから第１の密行列と関連付けられる疎要素の第１の部分集合をフェッチしてもよく、第２の疎要素アクセスユニットは、データ片１０６ｂから第１の密行列と関連付けられる疎要素の第２の異なる部分集合をフェッチしてもよい。

【００４５】

システムは、第１の密行列と関連付けられる疎要素および第２の密行列と関連付けられる疎要素を変換して、第１の密行列と関連付けられる疎要素および第２の密行列と関連付けられる疎要素を含む出力密行列を生成する（４０８）。たとえば、図２Ｃを参照して、疎密変換ユニット２００は、密行列２４２、２４４および２４６を、プロセッサユニットのための密行列に変換してもよい。

【００４６】

いくつかの実現例では、第１の密行列と関連付けられる疎要素および第２の密行列と関連付けられる疎要素は多次元の行列であってもよく、出力密行列はベクトルであってもよい。たとえば、密行列２４２は１００×１０の次元を有してもよく、密行列２４４は２０×１００の次元を有してもよく、密行列２４６は３×３の次元を有してもよい。疎密変換ユニット２００は、密行列２４２、２４４および２４６を、１×３００９の次元でベクトルに変換してもよい。

【００４７】

図５は、密行列を生成するためのプロセス５００の例を示すフローチャートである。プロセス５００は、疎密変換ユニット１０４または疎要素アクセスユニット３００などのようなシステムによって実行されてもよい。

【００４８】

システムは特定の疎要素の部分集合にアクセスすることに対する指示を受ける（５０２）。たとえば図３Ａを参照して、データフェッチユニット３０４は、データ片３３０から１つ以上の要求された疎要素をフェッチするための信号３４４を受信するよう構成されてもよい。いくつかの実現例では、１つ以上のデータ片において格納される特定の疎要素に対する要求が、ノードネットワーク上で受けられてもよい。たとえば図３Ａを参照して、要求識別ユニット３０２は、データ片３３０において格納された疎要素をフェッチするようノードネットワーク３２０上で要求３４２を受けるよう構成されてもよい。システムは、データフェッチユニットは特定の疎要素の部分集合を扱うよう割当てられる、と判断してもよい。たとえば、要求識別ユニット３０２は、疎要素アクセスユニット３００は要求３４２によって示される疎要素にアクセスするよう割当てられるかどうかを判断するよう構成されてもよい。データフェッチユニットは特定の疎要素の部分集合を扱うよう割当てられる、と判断することに応じて、その指示は特定の疎要素の部分集合にアクセスすることに対して生成されてもよい。たとえば、特定の要求された疎要素の識別（たとえば、デ

ータベーステーブル１番の１番）がルックアップテーブルに含まれている場合には、要求識別ユニット３０２は、特定の要求された疎要素をフェッチするよう、信号３４４をデータフェッチユニット３０４に送信してもよい。

【００４９】

システムは、特定の疎要素の部分集合の識別に基いて、特定の疎要素の部分集合をフェッチするためのプロセッサ指定を判断する（５０４）。たとえば図３Ａを参照して、データフェッチユニット３０４は１つ以上のプロセッサ３２２ａ～３２２ｋを含む。プロセッサ３２２ａ～３２２ｋの各々は、疎要素の識別に基いて特定の疎要素に割当てられてもよく、データフェッチユニット３０４は、プロセッサ３２２ａ～３２２ｋに対する１つ以上の要求を疎要素の識別に基いて発生させるよう構成されてもよい。いくつかの実現例では、システムは、システムが特定の疎要素の部分集合を扱うよう割当てられると判断してもよく、システムはルックアップテーブルに基いて特定の疎要素の部分集合を扱うよう割当てられると判断することを含む。たとえば、データフェッチユニット３０４はルックアップテーブルを用いることによってプロセッサ割当を判断してもよい。

【００５０】

システムは、指定に基いて、および複数個のプロセッサのうちの第１のプロセッサによって、特定の疎要素の部分集合の第１の疎要素をフェッチする（５０６）。たとえば図３Ａを参照して、データフェッチユニット３０４は、信号３４４に含まれる疎要素をフェッチするようプロセッサ３２２ａに命令してもよい。

【００５１】

システムは、指定に基いて、および複数個のプロセッサのうちの第２のプロセッサによって、特定の疎要素の部分集合の第２の疎要素をフェッチする（５０８）。たとえば図３Ａを参照して、データフェッチユニット３０４は、信号３４４に含まれる異なる疎要素をフェッチするようプロセッサ３２２ｂに命令してもよい。

【００５２】

いくつかの実現例では、第１のプロセッサから第１の疎要素を含む第１の行列を受取ってもよく、第１の行列は第１の次元を有してもよい。システムは、第１の疎要素を含む第２の行列を生成してもよく、第２の行列は、第１の次元よりも小さい第２の次元を有する。たとえば、疎低減ユニット３０６はフェッチされた疎要素３４６の次元を低減するように構成されてもよい。プロセッサ３２２ａ～３２２ｋの各々は、１００×１の次元を有する疎要素を生成してもよい。疎低減ユニット３０６は、１００×ｋの次元を有する、フェッチされた疎要素３４６を受け、フェッチされた疎要素３４６の次元を論理演算、算術演算または両方の組合せによって１００×１に低減することによって疎低減要素３４８を生成してもよい。システムは出力密行列を生成してもよく、出力密行列は第２の行列に基いて生成されてもよい。たとえば、連結ユニット３０８は、疎低減要素３４８を再配列および連結して、連結された要素３５０を生成するように構成されてもよい。

【００５３】

いくつかの実現例では、第１の疎要素は第１の時間の点において受取られてもよく、第２の疎要素は第２の異なる時間の点において受取られてもよい。システムは、出力密行列のために第１の疎要素および第２の疎要素の順序を判断してもよい。たとえば図３Ａを参照して、プロセッサ３２２ａは、フェッチされた疎要素１０番を、フェッチされた疎要素５番を返すように構成されるプロセッサ３２２ｂよりも早く、疎低減ユニット３０６に返すかもしれない。連結ユニット３０８は、その後受取られる疎要素５番を、より早く受取られた疎要素１０番の前に順序づけられるように再配列し、疎要素１番～２００番を連結された要素３５０として連結するよう構成される。

【００５４】

システムは、少なくとも第１の疎要素および第２の疎要素に適用される変換に基いて出力密行列を生成する（５１０）。いくつかの実現例では、システムは、出力密行列を圧縮して、圧縮された出力密行列を生成してもよい。システムは、圧縮された出力密行列をノードネットワークに与えてもよい。たとえば、圧縮／伸長ユニット３１０は、連結された

要素３５０を圧縮して、ノードネットワーク３２０のための密行列３５２を生成するよう構成されてもよい。

【００５５】

いくつかの実現例では、システムは、ノードネットワーク上で送信される密行列を表す第１の密行列を受取ってもよく、第１の密行列、第１の疎要素および第２の疎要素に基いて出力密行列を生成してもよい。たとえば、疎要素アクセスユニット３００は、ノードネットワーク３２０を介して近隣の疎要素アクセスユニットから密行列を受けてもよい。疎要素アクセスユニット３００は受取られた密行列を伸長してもよく、伸長された密行列を連結された要素３５０と連結して、更新された連結された要素を形成してもよく、それらは圧縮され、次いでノードネットワーク３２０に出力されることができる。

【００５６】

いくつかの実現例では、特定の疎要素のうちの１つ以上の疎要素は、多次元の行列であり、出力密行列はベクトルである。主題の実施形態および本明細書に記載される機能的動作は、デジタル電子回路において、有形に実施されたコンピュータソフトウェアもしくはファームウェアにおいて、本明細書において開示された構造およびそれらの構造等価物を含むコンピュータソフトウェアにおいて、または、それらの１つ以上の組合せにおいて実現され得る。本明細書に記載される主題の実施形態は、１つ以上のコンピュータプログラムとして、すなわち、データ処理装置による実行のために、または、データ処理装置の動作を制御するために有形の非一時的なプログラム担体上でエンコードされたコンピュータプログラム命令の１つ以上のモジュールとして実現され得る。代替的に、または加えて、プログラム命令は、データ処理装置による実行に対して好適な受信側装置への送信のために情報をエンコードするよう生成される、たとえばマシンにより生成された電気信号、光信号、または電磁気信号などの、人為的に生成された伝播される信号上でエンコードすることができる。コンピュータ記憶媒体は、コンピュータ読取可能記憶装置、コンピュータ読取可能記憶基板、ランダムもしくはシリアルアクセスメモリデバイス、または、それらの１つ以上の組合せであり得る。

【００５７】

「データ処理装置」という用語は、例としてプログラマブルプロセッサ、コンピュータ、または複数のプロセッサもしくはコンピュータを含む、データを処理するためのすべての種類の装置、デバイスおよびマシンを包含する。当該装置は、たとえばＦＰＧＡ（フィールドプログラマブルゲートアレイ）またはＡＳＩＣ（特定用途向け集積回路）といった特定目的論理回路を含み得る。当該装置は、ハードウェアに加えて、たとえばプロセッサファームウェア、プロトコルスタック、データベース管理システム、オペレーティングシステム、または、それらの１つ以上の組合せを構成するコードといった、当該コンピュータプログラムについて実行環境を作成するコードをさらに含み得る。

【００５８】

（プログラム、ソフトウェア、ソフトウェアアプリケーション、モジュール、ソフトウェアモジュール、スクリプトまたはコードとも称され、または記載され得る）コンピュータプログラムは、コンパイル型もしくはインタープリタ型言語、または宣言型もしくは手続き型言語を含む任意の形態のプログラミング言語で記述され得、スタンドアロンプログラムとして、または、モジュール、コンポーネント、サブルーチン、もしくは、コンピューティング環境で使用するのに好適な他のユニットとして任意の形態で展開され得る。コンピュータプログラムは、ファイルシステムにおけるファイルに対応し得るが、対応する必要があるわけではない。プログラムは、当該プログラムに専用である単一のファイルにおいて、または、複数の連携ファイル（coordinated files）（たとえばコードの１つ以上のモジュール、サブプログラムまたは部分を格納するファイル）において、他のプログラムまたはデータ（たとえばマークアップ言語ドキュメントに格納される１つ以上のスクリプト）を保持するファイルの一部に格納され得る。コンピュータプログラムは、１つの場所に位置するかもしくは複数の場所にわたって分散され通信ネットワークによって相互接続される１つのコンピュータまたは複数のコンピュータ上で実行されるように展開され

得る。

【００５９】

本明細書に記載されるプロセスおよび論理フローは、入力データ上で動作し出力を生成することにより機能を実行するよう１つ以上のプログラマブルプロセッサが１つ以上のコンピュータプログラムを実行することによって実行され得る。プロセスおよび論理フローは、たとえばＦＰＧＡ（フィールドプログラマブルゲートアレイ）、ＡＳＩＣ（特定用途向け集積回路）といった特殊目的論理回路として、またはＧＰＧＰＵ（汎用グラフィック処理装置）として実現され得る。

【００６０】

コンピュータプログラムの実行に好適であるプロセッサは、例として、汎用マイクロプロセッサもしくは特殊目的マイクロプロセッサもしくはその両方または任意の種類の中央処理ユニットに基づき得る。一般に、中央処理ユニットは、リードオンリメモリもしくはランダムアクセスメモリまたはその両方から命令およびデータを受取る。コンピュータの必須の要素は、命令を実行するための中央処理ユニットと、命令およびデータを格納するための１つ以上のメモリデバイスとである。一般に、コンピュータはさらに、たとえば磁気ディスク、光磁気ディスクまたは光ディスクといった、データを格納するための１つ以上の大容量記憶装置を含むか、当該１つ以上の大容量記憶装置からデータを受取るかもしくは当該１つ以上の大容量記憶装置にデータを転送するよう動作可能に結合されるか、またはその両方を行う。しかしながら、コンピュータはそのような装置を有する必要はない。さらに、コンピュータはたとえば、携帯電話、携帯情報端末（ＰＤＡ）、モバイルオーディオまたはビデオプレーヤ、ゲームコンソール、全地球測位システム（ＧＰＳ）受信機、またはポータブル記憶装置（たとえばユニバーサルシリアルバス（ＵＳＢ）フラッシュドライブ）といった別のデバイスに埋め込まれ得る。

【００６１】

コンピュータプログラム命令およびデータを格納するのに好適であるコンピュータ読取可能媒体は、例として、たとえばＥＰＲＯＭ、ＥＥＰＲＯＭおよびフラッシュメモリ素子といった半導体メモリデバイスと、たとえば内部ハードディスクまたはリムーバブルディスクといった磁気ディスクと、光磁気ディスクと、ＣＤ－ＲＯＭおよびＤＶＤ－ＲＯＭディスクとを含むすべての形態の不揮発性メモリ、媒体およびメモリデバイスを含む。プロセッサおよびメモリは、特殊目的論理回路によって補足され得るか、または特殊目的論理回路に組み込まれ得る。

【００６２】

ユーザとのインタラクションを提供するために、本明細書に記載される主題の実施形態は、たとえばＣＲＴ（陰極線管）またはＬＣＤ（液晶ディスプレイ）モニタといったユーザに対して情報を表示するための表示デバイスと、たとえばマウス、トラックボールといったユーザがコンピュータに入力を提供可能であるキーボードおよびポインティングデバイスとを有するコンピュータ上で実現され得る。他の種類のデバイスが同様に、ユーザとのインタラクションを提供するために使用され得；たとえば、ユーザに提供されるフィードバックは、たとえば視覚フィードバック、聴覚フィードバックまたは触覚フィードバックといった任意の形態の感覚フィードバックであり得；ユーザからの入力は、音響入力、音声入力、または触覚入力を含む任意の形態で受取られ得る。加えて、コンピュータは、ユーザが使用するデバイスにドキュメントを送信しユーザが使用するデバイスからドキュメントを受信することによって、たとえば、ウェブブラウザから受信された要求に応答してユーザのクライアントデバイス上のウェブブラウザにウェブページを送信することによって、ユーザと対話し得る。

【００６３】

本明細書に記載される主題の実施形態は、たとえばデータサーバとしてバックエンドコンポーネントを含む計算システムにおいて実現され得るか、たとえばアプリケーションサーバといったミドルウェアコンポーネントを含む計算システムにおいて実現され得るか、たとえば本明細書に記載される主題の実現例とユーザが対話することが可能であるグラフ

ィカルユーザーインターフェイスもしくはウェブブラウザを有するクライアントコンピュータといったフロントエンドコンポーネントを含む計算システムにおいて実現され得るか、または１つ以上のそのようなバックエンドコンポーネント、ミドルウェアコンポーネントもしくはフロントエンドコンポーネントの任意の組合せの計算システムにおいて実現され得る。システムのコンポーネントは、たとえば通信ネットワークといったデジタルデータ通信の任意の形態または媒体によって相互接続され得る。通信ネットワークの例は、ローカルエリアネットワーク（「ＬＡＮ」）およびワイドエリアネットワーク（「ＷＡＮ」）、たとえばインターネットを含む。

【００６４】

計算システムはクライアントおよびサーバを含むことができる。クライアントとサーバとは一般に互いから遠隔にあり、典型的には通信ネットワークを通じて対話する。クライアントとサーバとの関係は、それぞれのコンピュータ上で実行されるとともに互いに対してクライアント－サーバ関係を有するコンピュータプログラムによって発生する。

【００６５】

本明細書は多くの特定の実現例の詳細を含んでいるが、これらは如何なる発明の範囲または請求され得るものの範囲に対する限定としても解釈されるべきではなく、特定の発明の特定の実施形態に特有の特徴であり得る記載として解釈されるべきである。別個の実施形態の文脈で本明細書において記載されるある特徴は、単一の実施形態において組合せでも実現され得る。反対に、単一の実施形態の文脈において記載されるさまざまな特徴は、複数の実施形態において別々に、または任意の好適な部分的組合わせでも実現され得る。さらに、特徴は、ある組合せにおいて作用すると上で記載されているとともにそのように最初は請求されている場合があるが、請求される組合せのうちの１つ以上の特徴はいくつかの場合には当該組合せから削除され得、請求される組合せは、部分的組合わせまたは部分的組合わせの変形例に関し得る。

【００６６】

同様に、動作が図においては特定の順に示されているが、そのような動作は、望ましい結果を達成するために、示された当該特定の順もしくは連続した順で実行される必要があると理解されるべきではなく、または、すべての示された動作が実行される必要があると理解されるべきではない。ある状況においては、マルチタスキングおよび並列処理が有利であり得る。さらに、上に記載された実施形態におけるさまざまなシステムモジュールおよびコンポーネントの分離は、すべての実施形態においてそのような分離を必要とすると理解されるべきでなく、記載されたプログラムコンポーネントおよびシステムは一般に、単一のソフトウェアプロダクトに一緒に統合され得るか、または、複数のソフトウェアプロダクトへとパッケージ化され得るということが理解されるべきである。

【００６７】

主題の特定の実施形態が記載された。他の実施形態は以下の請求の範囲内にある。たとえば、請求項において記載されるアクションは、異なる順で実行され得、それでも望ましい結果を達成し得る。一例として、添付の図において示されるプロセスは、望ましい結果を達成するために、示された特定の順または連続する順であることを必ずしも必要としない。ある実現例においては、マルチタスキングおよび並列処理が有利であり得る。

【符号の説明】

【００６８】

１００　計算システム、　１０２　処理ユニット、１０４　疎密変換ユニット、１０６ａ～１０６ｋ　データ片、２００　疎密変換ユニット、３００　疎要素アクセスユニット、３０２　要求識別ユニット、３０４　データフェッチユニット、３０６　疎低減ユニット、３０８　連結ユニット、３１０　圧縮／伸長ユニット、３１２　分割ユニット、３３０　データ片、３２０　ノードネットワーク、３２２ａ～３２２ｋ　プロセッサ。

【図面】

【図１】

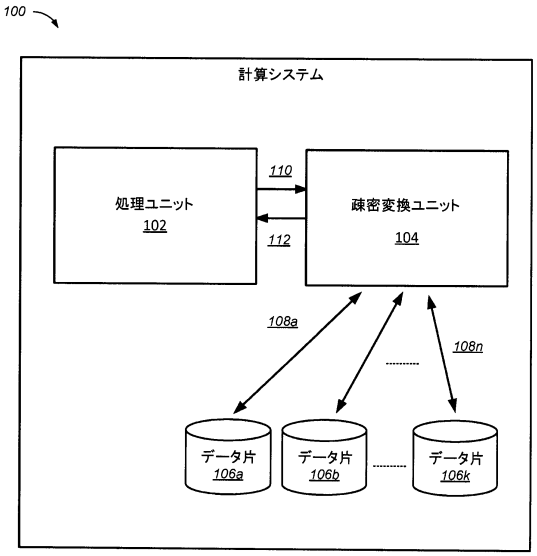

FIG. 1

【図２Ａ】

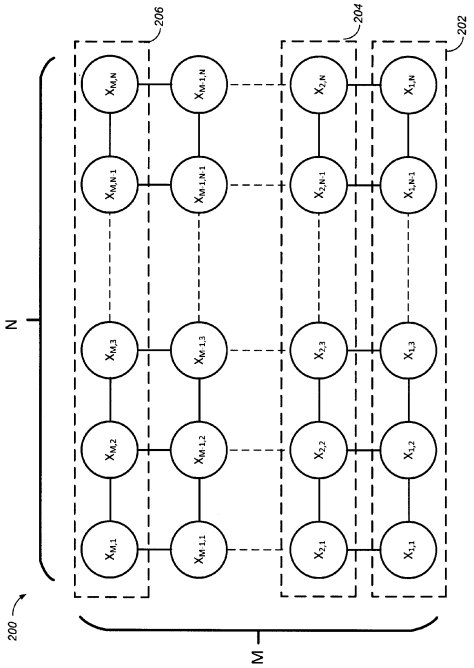

FIG. 2A

【図２Ｂ】

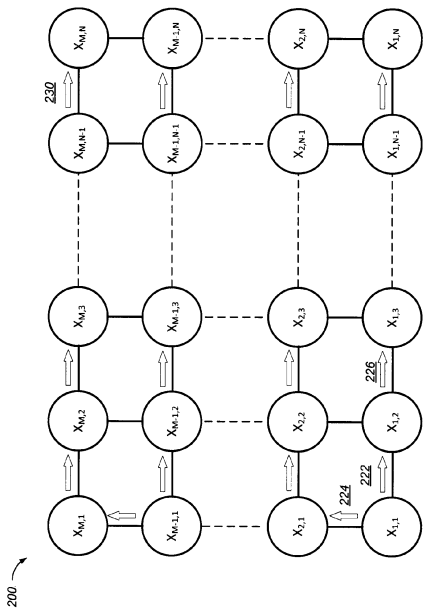

FIG. 2B

【図２Ｃ】

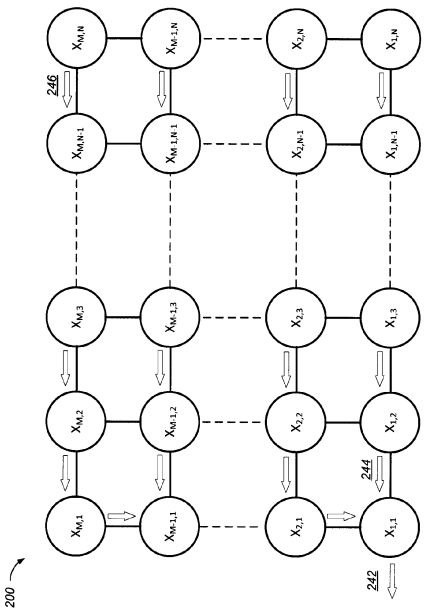

FIG. 2C

【図２Ｄ】

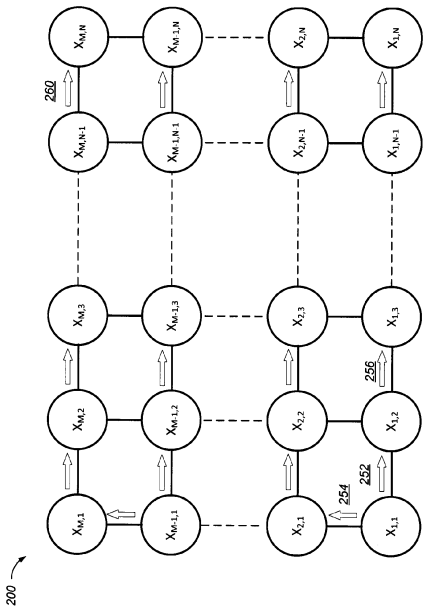

FIG. 2D

【図３Ａ】

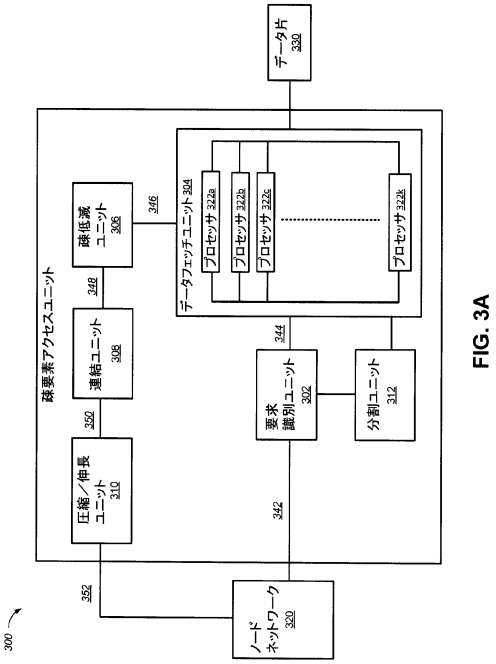

FIG. 3A

【図３Ｂ】

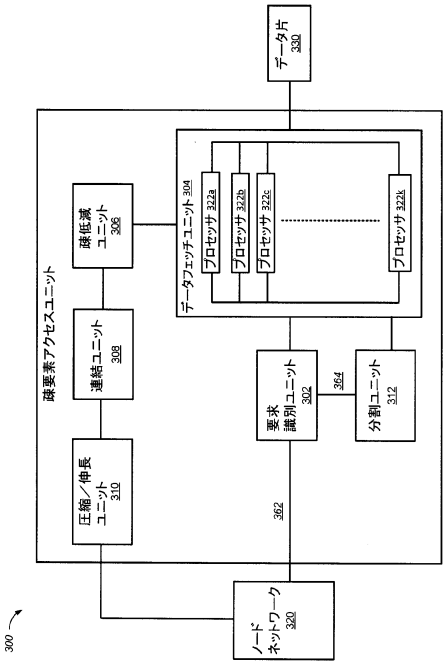

FIG. 3B

【図４】

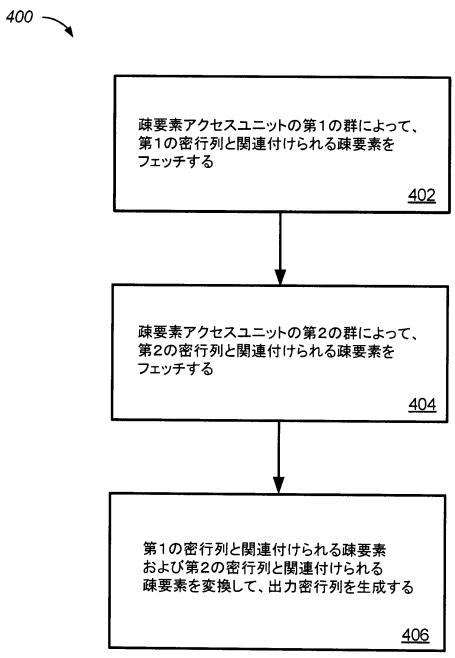

FIG. 4

【図５】

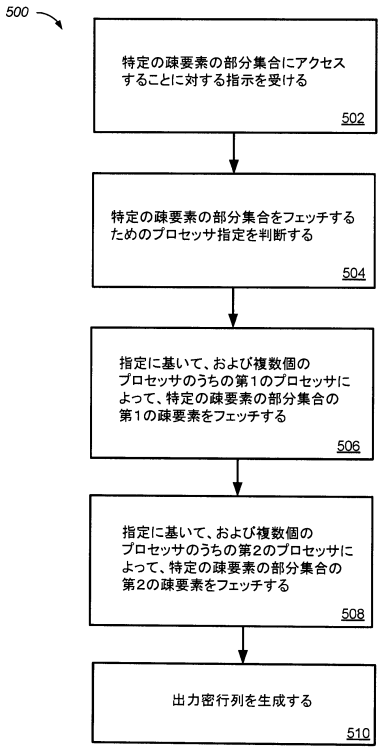
500
特定の疎要素の部分集合にアクセスすることに対する指示を受ける
502
特定の疎要素の部分集合をフェッチするためのプロセッサ指定を判断する
504
指定に基いて、および複数個のプロセッサのうちの第１のプロセッサによって、特定の疎要素の部分集合の第１の疎要素をフェッチする
506
指定に基いて、および複数個のプロセッサのうちの第２のプロセッサによって、特定の疎要素の部分集合の第２の疎要素をフェッチする
508
出力密行列を生成する
510

FIG. 5

フロントページの続き

(72)発明者　ラフル・ナガラジャン
アメリカ合衆国、９４０４３　カリフォルニア州、マウンテン・ビュー、アンフィシアター・パークウェイ、１６００
(72)発明者　ウ・ドン・ヒョク
アメリカ合衆国、９４０４３　カリフォルニア州、マウンテン・ビュー、アンフィシアター・パークウェイ、１６００
(72)発明者　クリストファー・ダニエル・リアリー
アメリカ合衆国、９４０４３　カリフォルニア州、マウンテン・ビュー、アンフィシアター・パークウェイ、１６００

審査官　坂庭 剛史
(56)参考文献　米国特許出願公開第２０１５／００６７００９（ＵＳ，Ａ１）
特開２０１４－１９９５４５（ＪＰ，Ａ）
特開平０６－０５２１２５（ＪＰ，Ａ）
(58)調査した分野　(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　１７／１６