(19)日本国特許庁(JP)　(12)特　許　公　報(B2)　(11)特許番号
特許第6674460号
(P6674460)
(45)発行日　令和2年4月1日(2020.4.1)　(24)登録日　令和2年3月10日(2020.3.10)

(51)Int.Cl.　F I
G06F 12/1009 (2016.01)　G06F 12/1009
G06F 12/1027 (2016.01)　G06F 12/1027
G06F 12/0806 (2016.01)　G06F 12/0806

請求項の数 15　(全 18 頁)

(21)出願番号　特願2017-528906 (P2017-528906)
(86)(22)出願日　平成27年11月20日 (2015.11.20)
(65)公表番号　特表2018-501559 (P2018-501559A)
(43)公表日　平成30年1月18日 (2018.1.18)
(86)国際出願番号　PCT/US2015/061987
(87)国際公開番号　WO2016/089631
(87)国際公開日　平成28年6月9日 (2016.6.9)
審査請求日　平成30年11月5日 (2018.11.5)
(31)優先権主張番号　14/560,290
(32)優先日　平成26年12月4日 (2014.12.4)
(33)優先権主張国・地域又は機関
米国 (US)

(73)特許権者　507364838
クアルコム，インコーポレイテッド
アメリカ合衆国　カリフォルニア　９２１２１　サン　ディエゴ　モアハウス　ドライブ　５７７５
(74)代理人　100108453
弁理士　村山　靖彦
(74)代理人　100163522
弁理士　黒田　晋平
(72)発明者　スティーヴン・アーサー・モロイ
アメリカ合衆国・カリフォルニア・９２１２１・サン・ディエゴ・モアハウス・ドライヴ・５７７５

最終頁に続く

(54)【発明の名称】不均一メモリアーキテクチャにおける改善されたレイテンシのためのシステムおよび方法

(57)【特許請求の範囲】
【請求項１】
不均一メモリアーキテクチャを有するポータブルコンピューティングデバイス内のメモリを割り振るための方法であって、
第1のシステムオンチップ(SoC)上で実行されているプロセスから、仮想メモリページについての要求を受け取るステップであって、前記第1のSoCがチップ間インターフェースを介して第2のSoCに電気的に結合され、前記第1のSoCが第1の高性能バスを介して第1のローカル揮発性メモリデバイスに電気的に結合され、前記第2のSoCが第2の高性能バスを介して第2のローカル揮発性メモリデバイスに電気的に結合される、ステップと、
前記第1および第2のローカル揮発性メモリデバイス上で利用可能な同じ物理アドレスを含んだ空き物理ページペアを決定するステップと、
前記空き物理ページペアを単一の仮想ページアドレスにマップするステップと
を含む、方法。
【請求項２】
前記空き物理ページペアを前記単一の仮想ページアドレスに前記マップするステップが、前記同じ物理アドレスに関連するページテーブルエントリを変更するステップを含む、請求項1に記載の方法。
【請求項３】
前記ページテーブルエントリを前記変更するステップが、前記第1および第2のローカル揮発性メモリデバイス上の前記同じ物理アドレスに記憶されたメモリデータを複製するよ

うにコピー属性を設定するステップを含む、請求項2に記載の方法。
【請求項４】
前記第1および第2のローカル揮発性メモリデバイス上の前記同じ物理アドレスに記憶されたメモリデータを複製するステップ
をさらに含む、請求項1に記載の方法。
【請求項５】
前記空き物理ページペアを前記決定するステップが、
前記利用可能な同じ物理アドレスを識別するためにグローバルディレクトリで物理ページフレームを検索するステップ
を含む、請求項1に記載の方法。
【請求項６】
前記空き物理ページペアを前記決定するステップが、
前記第1および第2のローカル揮発性メモリデバイス用の前記同じ物理アドレスが、異なる仮想アドレスに割り当てられることを決定するステップと、
前記物理ページペアを空けるために、前記異なる仮想アドレスのうちの少なくとも１つを再割り当てするステップと
を含む、請求項1に記載の方法。
【請求項７】
別の仮想メモリページを求めるさらなる要求を受け取るステップと、
さらなる物理ページペアが利用可能ではないと決定するステップと、
前記さらなる要求に応答して、さらなる物理ページを別の仮想ページアドレスにマップするステップと
をさらに含む、請求項1に記載の方法。
【請求項８】
オペレーティングシステムが、以前に割り振られた物理ページをさらなる空き物理ページペアに変換するステップ
をさらに含む、請求項1に記載の方法。
【請求項９】
前記第1および第2のローカル揮発性メモリデバイスのうちの1つまたは複数が、ダイナミックランダムアクセスメモリ(DRAM)デバイスを含む、請求項1に記載の方法。
【請求項１０】
前記ポータブルコンピューティングデバイスが、モバイルフォン、タブレットコンピュータ、ナビゲーションデバイス、および携帯ゲーム機のうちの1つを含む、請求項1に記載の方法。
【請求項１１】
不均一メモリアーキテクチャを有するポータブルコンピューティングデバイス内のメモリを割り振るためのシステムであって、
第1のシステムオンチップ(SoC)上で実行されているプロセスから、仮想メモリページについての要求を受け取るための手段であって、前記第1のSoCがチップ間インターフェースを介して第2のSoCに電気的に結合され、前記第1のSoCが第1の高性能バスを介して第1のローカル揮発性メモリデバイスに電気的に結合され、前記第2のSoCが第2の高性能バスを介して第2のローカル揮発性メモリデバイスに電気的に結合される、手段と、
前記第1および第2のローカル揮発性メモリデバイス上で利用可能な同じ物理アドレスを含んだ空き物理ページペアを決定するための手段と、
前記空き物理ページペアを単一の仮想ページアドレスにマップするための手段と
を含む、システム。
【請求項１２】
前記空き物理ページペアを前記単一の仮想ページアドレスにマップするための前記手段が、前記同じ物理アドレスに関連するページテーブルエントリを変更するための手段をさらに含む、請求項11に記載のシステム。

【請求項１３】
前記ページテーブルエントリを変更するための前記手段が、前記第1および第2のローカル揮発性メモリデバイス上の前記同じ物理アドレスに記憶されたメモリデータを複製するようにコピー属性を設定するための手段を含む、請求項12に記載のシステム。
【請求項１４】
前記第1および第2のローカル揮発性メモリデバイス上の前記同じ物理アドレスに記憶されたメモリデータを複製するための手段
をさらに含む、請求項11に記載のシステム。
【請求項１５】
不均一メモリアーキテクチャでメモリを割り振るために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であって、請求項1から10のいずれか一項に記載の方法を実行するように構成された論理を含むコンピュータプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
不均一メモリアーキテクチャにおける改善されたレイテンシのためのシステムおよび方法に関する。
【背景技術】
【０００２】
ポータブルコンピューティングデバイス(たとえば、セルラー電話、スマートフォン、タブレットコンピュータ、携帯情報端末(PDA)、および携帯ゲーム機)は、ますます拡大する数々の機能およびサービスを提供し続けており、ユーザが情報、リソース、および通信にかつてないレベルでアクセスできるようにしている。これらのサービス拡張とペースを合わせるために、そのようなデバイスは、より強力、かつより複雑になってきた。ここで、ポータブルコンピューティングデバイスには、一般に、単一の基板上に組み込まれた1つまたは複数のチップ構成要素(たとえば、1つまたは複数の中央処理ユニット(CPU)、グラフィックス処理ユニット(GPU)、デジタル信号プロセッサなど)を含む、システムオンチップ(SoC)が含まれる。
【０００３】
集積回路上のトランジスタ密度を上げることがより難しくなるにつれて、2次元モノリシック集積のコストは法外になり、結果的に、ポータブルコンピューティングデバイスにおいてマルチダイまたはマルチSoCの製品の使用が増えることがある。そのようなマルチダイの製品は、ダイナミックランダムアクセスメモリ(DRAM)などの高速ローカルメモリに各々アクセスできる、相互接続された物理ダイを含む場合がある。そのようなアーキテクチャは、一般に、非統合メモリアーキテクチャ(NUMA: Non-unified Memory Architecture)と呼ばれる。しかしながら、NUMA設計は、高性能のバスを介してアクセス可能な近い、すなわちローカルのDRAM、またはより低い性能のチップ間インターフェースを介してアクセス可能な遠いDRAMにあるデータが、どちらのダイ上でもプロセッサによってアクセスされる必要がある状況をもたらす。これは、たとえば、プロセッサが遠いDRAMに向かわなければならないとき、より高いレイテンシを生じる可能性がある。
【発明の概要】
【発明が解決しようとする課題】
【０００４】
したがって、非統合メモリアーキテクチャにおけるすべてのプロセッサに対する低レイテンシのメモリアクセスのシステムおよび方法を提供する必要がある。
【課題を解決するための手段】
【０００５】
不均一メモリアーキテクチャ(Non-uniform Memory Architecture)を有するポータブルコンピューティングデバイス内のメモリを割り振るためのシステム、方法、およびコンピュータプログラムが開示される。1つのそのような方法は、第1のシステムオンチップ(SoC

)上で実行されているプロセスから、仮想メモリページについての要求を受け取るステップを含む。第1のSoCは、チップ間インターフェースを介して第2のSoCに電気的に結合される。第1のSoCは、第1の高性能バスを介して第1のローカル揮発性メモリデバイスに電気的に結合され、第2のSoCは、第2の高性能バスを介して第2のローカル揮発性メモリデバイスに電気的に結合される。第1および第2のローカル揮発性メモリデバイス上で利用可能である同じ物理アドレスを含む、空き物理ページペア(Free Physical Page Pair)が決定される。空き物理ページペアは、単一の仮想ページアドレスにマップされる。

【０００６】

図において、別段に規定されていない限り、同様の参照番号は、様々な図の全体を通して同様の部分を指す。「102A」または「102B」などの文字指定を伴う参照番号について、文字指定は、同じ図に存在する2つの同様の部分または要素を区別する場合がある。参照番号がすべての図において同じ参照番号を有するすべての部分を含むことを意図するとき、参照番号に対する文字指定は省略される場合がある。

【図面の簡単な説明】

【０００７】

【図１】不均一メモリアーキテクチャ(NUMA)を有する複数の相互接続されたシステムオンチップ(SoC)を備えたシステムの一実施形態のブロック図である。

【図２】メモリデータがSoC間で選択的に複製された図1のシステムを示す図である。

【図３】同じ物理アドレスを有する空き物理ページペアを含んだSoC内のページテーブルの一実施形態を示すブロック図である。

【図４】空き物理ページペアを単一の仮想アドレスにマップするためにコピー属性フィールドを含んだページテーブルエントリを実装するためのデータ構造の一実施形態を示すブロック図である。

【図５】図1および図2における適応的NUMAレイテンシ最適化モジュールによって実施される方法の一実施形態を示すフローチャートである。

【図６】物理ページペアを空けるためにオペレーティングシステムによって実施される方法の一実施形態を示すフローチャートである。

【図７】以前に割り振られた単一のページを空き物理ページペアに変換するための方法の一実施形態を示すフローチャートである。

【図８】図1および図2のシステムにおいてページテーブルエントリを実装するためのデータ構造の別の実施形態を示すブロック図である。

【図９ａ】図1および図2におけるノンブロッキングNUMAレイテンシ最適化モジュールによって実施される方法の一実施形態を示すフローチャートである。

【図９ｂ】図1および図2におけるノンブロッキングNUMAレイテンシ最適化モジュールによって実施される方法の一実施形態を示すフローチャートである。

【図１０】図8のページテーブルエントリを実装する例示的な書込みトランザクションを示すフローチャートである。

【図１１】図8のページテーブルエントリを実装する例示的な読取りトランザクションを示すフローチャートである。

【図１２】図8のページテーブルエントリを使用してデータを複製するためのページ変換図を示す機能ブロック図である。

【図１３】システムメモリを拡張するためのRAMカード/ソケットを内蔵する場合があるポータブル通信デバイスの別の実施形態のブロック図である。

【発明を実施するための形態】

【０００８】

「例示的」という語は、本明細書では、「例、事例、または例示として役に立つ」ことを意味するために使用される。本明細書で「例示的」として説明されるいかなる態様も、必ずしも他の態様よりも好ましいまたは有利なものと解釈されるべきではない。

【０００９】

この説明では、「アプリケーション」または「イメージ」という用語は、オブジェクト

コード、スクリプト、バイトコード、マークアップ言語ファイル、およびパッチなどの、実行可能なコンテンツを有するファイルを含む場合もある。加えて、本明細書で参照される「アプリケーション」は、オープンされる必要があり得るドキュメントまたはアクセスされる必要がある他のデータファイルなどの、本質的に実行可能でないファイルを含む場合もある。

【００１０】

「コンテンツ」という用語はまた、オブジェクトコード、スクリプト、バイトコード、マークアップ言語ファイル、およびパッチなどの、実行可能コンテンツを有するファイルを含む場合もある。加えて、本明細書において言及する「コンテンツ」は、開かれる必要があることがある文書、またはアクセスされる必要がある他のデータファイルなど、本来は実行可能ではないファイルを含む場合もある。

【００１１】

「構成要素」、「データベース」、「モジュール」、「システム」などの用語は、本明細書で使用されるとき、ハードウェア、ファームウェア、ハードウェアとソフトウェアの組合せ、ソフトウェア、または実行中のソフトウェアのいずれかであるコンピュータ関連エンティティを指すものとする。たとえば、構成要素は、限定はしないが、プロセッサ上で実行されているプロセス、プロセッサ、オブジェクト、実行可能ファイル、実行のスレッド、プログラム、および/またはコンピュータであってもよい。例として、コンピューティングデバイス上で実行されるアプリケーションとコンピューティングデバイスの両方が構成要素であってもよい。1つまたは複数の構成要素は、プロセスおよび/または実行のスレッド内に存在してもよく、構成要素は、1つのコンピュータ上に局在化され、および/または2つ以上のコンピュータ間で分散されてもよい。加えて、これらの構成要素は、様々なデータ構造が記憶された様々なコンピュータ可読媒体から実行されてもよい。構成要素は、1つまたは複数のデータパケット(たとえば、ローカルシステム、分散システム内の別の構成要素と対話し、かつ/または、信号によってインターネットなどのネットワークにわたって他のシステムと対話する、1つの構成要素からのデータ)を有する信号などに従って、ローカルプロセスおよび/またはリモートプロセスによって通信する場合がある。

【００１２】

「仮想メモリ」という用語は、メモリを参照しているアプリケーションまたはイメージから見た実際の物理メモリの抽象化を指す。仮想メモリアドレスを物理メモリアドレスに変換するために、変換またはマッピングが使用され得る。マッピングは、1対1と同じほど単純である(たとえば、物理アドレスが仮想アドレスに等しい)か、中程度に複雑である(たとえば、物理アドレスが仮想アドレスからの一定のオフセットに等しい)か、または、複雑である(たとえば、すべての4KBページが一意的にマップされる)場合がある。マッピングは、静的である(たとえば、起動時に一度実行される)場合があり、または、動的である(たとえば、メモリが割り振られ、空けられながら、連続的に進展する)場合がある。

【００１３】

本明細書では、「通信デバイス」、「ワイヤレスデバイス」、「ワイヤレス電話」、「ワイヤレス通信デバイス」、および「ワイヤレスハンドセット」という用語は、互換的に使用される。第3世代(「3G」)ワイヤレス技術および第4世代(「4G」)が出現したことによって、利用可能な帯域幅が拡大されたので、より多様なワイヤレス能力を備えた、より多くのポータブルコンピューティングデバイスが可能になった。したがって、ポータブルコンピューティングデバイスには、セルラー電話、ページャ、PDA、スマートフォン、ナビゲーションデバイス、またはワイヤレス接続もしくはワイヤレスリンクを有するハンドヘルドコンピュータが含まれる場合がある。

【００１４】

図1は、不均一メモリアーキテクチャ(NUMA)を有する複数の相互接続された物理ダイ(たとえば、システムオンチップ(SoC)102およびSoC202)を備えたシステム100の一実施形態を示す。システム100は、パーソナルコンピュータと、ワークステーションと、サーバと、セルラー電話、携帯情報端末(PDA)、携帯ゲーム機、パームトップコンピュータ、または

タブレットコンピュータなどのポータブルコンピューティングデバイス(PCD)とを含む、任意のコンピューティングデバイス用に設計された、またはさもなければ任意のコンピューティングデバイスに常駐するマルチダイ製品として実装される場合がある。SoC102および202は、ダイ間インターフェース116を介して電気的に結合される。各SoCは、高性能バスを介して、近い、すなわちローカルの揮発性メモリデバイス(たとえば、ダイナミックランダムアクセスメモリ(DRAM)デバイス)に電気的に結合される。図1の実施形態に示すように、SoC102は、バス105を介してローカルDRAM104に接続され、SoC202は、バス205を介してローカルDRAM204に接続される。バス105および205は、それぞれSoC102およびSoC202によるローカルDRAM104およびローカルDRAM204への、低レイテンシでより高速、高性能のアクセスを可能にする。当技術分野で知られているように、NUMAは、SoC102および202各々が他方のSoCのローカルDRAMにアクセスすることを可能にするが、ダイ間インターフェース116は結果的にレイテンシをより高くし、性能を比較的低くする場合がある。

【００１５】

SoC102および202は、様々なオンチップまたはオンダイ構成要素を含む。オンチップ構成要素は、必要に応じて変わってもよく、システム100は、任意の数のSoCを含んでもよいことを諒解されたい。図1の実施形態では、SoC102は、SoCバス112を介して相互接続された、1つまたは複数のプロセッサ108(たとえば、中央処理ユニット(CPU)、グラフィックス処理ユニット(GPU)、デジタル信号プロセッサ(DSP)など)と、DRAMコントローラ106と、チップ間インターフェースコントローラ114とを備える。SoC202は、SoCバス212を介して相互接続された、1つまたは複数のプロセッサ208と、DRAMコントローラ206と、チップ間インターフェースコントローラ214とを備える。SoC102およびSoC202は、(それぞれバス105およびバス205を介して)SoCローカルDRAMから、またはダイ間インターフェース116を介して他のSoCに接続された遠いDRAMから、メモリリソースを要求する1つまたは複数のメモリクライアントを含む場合がある。DRAMコントローラ106およびコントローラ206は、それぞれDRAM104およびDRAM204との間を行き来するデータのフローを管理する。チップ間インターフェースコントローラ114および214は、SoC102とSoC202との間のデータのフローを管理する。

【００１６】

各SoCは、オペレーティングシステム(O/S110および210)を備えることができ、オペレーティングシステムは、たとえば、システムメモリマネージャ200を介して仮想メモリ管理をサポートする。システムメモリマネージャ200は、ハードウェアおよび/またはソフトウェアをともに使用して実装され得る様々なメモリ管理技法を制御するように構成される。当技術分野で知られているように、システムメモリマネージャ200は、仮想アドレスと呼ばれる、プログラムによって使用されるメモリアドレスを、コンピュータメモリ内の物理アドレスにマップする。O/S110および210は、仮想アドレス空間と、物理メモリ(たとえば、DRAM104および204)を仮想メモリへ割り当てることとを管理する。メモリ管理ユニット(MMU)などのアドレス変換ハードウェアが、仮想アドレスを物理アドレスに変換する。

【００１７】

図2に関しては、O/S110、O/S210、およびシステムメモリマネージャ200は、メモリアクセス、タスク、作業負荷が複数のプロセッサにわたって管理されるNUMAをサポートするように構成され得ることを諒解されたい。下記でより詳細に述べるように、システムメモリマネージャ200は、システム100の不均一メモリアーキテクチャにおいて複数のプロセッサにわたって改善されたレイテンシでのメモリアクセスを可能にするための様々なモジュールを備える場合がある。適応的NUMAレイテンシ最適化モジュール201について、図3～図7に関して以下で説明し、ノンブロッキングNUMAレイテンシ最適化モジュールについて、図8～図12に関して以下で説明する。

【００１８】

図2の例示的な実施形態で説明したように、システムメモリマネージャ200は、各ダイ(たとえば、SoC102および202)がローカルDRAM(すなわち、それに直接接続されたDRAM)にデータのコピーを有するように、メモリデータを選択的に複製するように構成され得る。た

とえば、DRAM204にあるメモリデータ300が、複数のプロセッサによってアクセスされる場合がある。メモリデータ300は、複数のデータ部分302、304、306、および308を含む場合がある。SoC202上にある、プロセッサ220が、データ304および308へのアクセスを要求する場合があり、プロセッサ222が、データ302および306へのアクセスを要求する場合がある。SoC102上にある、プロセッサ120が、データ302および304へのアクセスを要求する場合があり、プロセッサ122が、データ306へのアクセスを要求する場合がある。NUMAはプロセッサ120および122がダイ間インターフェース116を介してメモリデータ300にアクセスすることを可能にするが、SoC102上のプロセッサ120および122に、それらが必要とするメモリへのより高性能でより低レイテンシのアクセスを与えるために、DRAM204にあるメモリデータをDRAM104上に(またはその逆に)選択的に複製することが望ましい場合がある。

【００１９】

システム100においてメモリデータを選択的に複製するために、様々な方法、アルゴリズム、および技法が採用され得ることを諒解されたい。図3～図7に示す実施形態では、システム100は、空き物理ページペア(Free Physical Page Pair)を探すこと、識別すること、および/または管理することによって、メモリデータを複製する。図3に示すように、空き物理ページペア399は、DRAM104内の利用可能な物理アドレスの、DRAM204内の同じ利用可能な物理アドレスとの論理マッピングを含む。例示的な実施形態では、同じ物理アドレスは、それらのページアドレスの同一の下位Nビットを有する2つの物理ページアドレスを指し得ることを諒解されたい。ただし、N=log2(単一メモリチップの容量)。たとえば、2個の1GBメモリチップを備えた2GBシステムにおいて、物理ページペアが、位置(29、28、27…14、13、12)に同一のアドレスビットを有する場合がある。ビット(11、10…1、0)は、それらがすでに、たとえば4KBページの内部にある可能性があるので、比較されなくてもよいことを諒解されたい。論理マッピングは、ページテーブル350および360により提供され得る。(DRAM104に対応する)ページテーブル350は、SoC102上で実行している仮想メモリマネージャ402によって管理され得る。(DRAM204に対応する)ページテーブル360は、SoC202上で実行している仮想メモリマネージャ404によって管理され得る。ページテーブル350および360は、物理アドレス402～432の範囲へのインデックスを含む。一例として、ページテーブル350内の物理アドレス402aおよびページテーブル360内の物理アドレス402bは、それらが同じ物理アドレスを有するので、物理ページペアを表す。空いている、または利用可能な物理ページペア399は、DRAM104とDRAM204の両方においてメモリ割振りに利用可能である物理ページペア(すなわち、402a/b、404a/b、406a/bなど)を指す。図3では、空き物理ページペアが、灰色に表示されたボックスにおいて識別される。この点について、「a」という文字で参照される物理アドレス(たとえば、406a、408aなど)は、SoC102/DRAM104に対応し、「b」という文字で参照される物理アドレス(たとえば、406b、408bなど)は、SoC202/DRAM204に対応し、同じ番号を付けられた「a/b」ペアが、物理ページペア399を含んでいる。

【００２０】

複数のプロセッサ(たとえば、SoC102上のプロセッサ120および122、ならびにSoC202上のプロセッサ220および222)にわたるアクセスのためにメモリデータを選択的に複製するために、図4に示すように、変更されたページテーブルエントリ400が与えられてよい。変更されたページテーブルエントリ400は、物理アドレスへの物理ページインデックス454を記憶するためのフィールド454、ならびにコピービット値452を記憶するためのコピー属性(Copy Attribute)フィールド450を備える。単一の物理ページを対応する仮想ページアドレスにマップするためのデフォルト動作では、「0」または「false」というコピービット値が使用されてよい。改善されたレイテンシが求められるとき、かつ空き物理ページペア399が利用可能である限り、コピービット値は「1」または「true」に設定されてよく、これによりシステム100は、空き物理ページペア399を単一の仮想ページアドレスに論理的にマップすることができる。コピー属性フィールド450は、各SoCダイがローカルDRAMにデータのコピーを有するように、メモリデータを選択的に複製するために使用され得ることを諒解されたい。

【００２１】
図5は、空き物理ページペア399を使用してNUMAにおいてメモリを割り振るための方法500を示す。方法500は、0/S110、0/S210、および/またはシステムメモリマネージャ200によって実施されてよい。ブロック502において、第1のSoC102上で実行されているプロセスから仮想メモリページについての要求が受け取られる場合がある。ブロック504において、システムは、空き物理ページペア399があるかどうかを決定する。空き物理ページペア399が利用可能ではない場合(決定ブロック506)、単一の物理ページに仮想ページアドレスが論理的にマップされ得る。しかしながら、利用可能である場合、空き物理ページペア399は同じ仮想ページアドレスに論理的にマップされ得る(ブロック508)。上記で説明したように、論理的マッピングは、ページテーブルエントリ350を変更することによって実行されてよい。たとえば、ブロック510において、コピー属性フィールド450は、コピービット値452を「1」または「true」という値に設定することによって変更されてよく、これにより、DRAM104および204上の同じ物理アドレスに記憶されたメモリデータを複製する。
【００２２】
システムメモリマネージャ200は、0/S110および0/S210が空き物理ページペア399を探すおよび/または管理するバックグラウンドプロセスを実行できるように構成されてよい。図6は、さらなる物理ページペアを空けるための方法600を示す。すべてのメモリ物理ページのグローバルディレクトリが設けられてもよい。ブロック602において、オペレーティングシステムは、グローバルディレクトリを検索してよい。ブロック604において、オペレーティングシステムは、ページテーブル350および360が、異なる仮想アドレスに割り当てられた一致する物理アドレスを有するいずれかの物理ページを識別するかどうかを決定してよい。一致が存在しない場合(決定ブロック606)、フローは、続いて潜在的な一致をチェックするためにブロック602に戻ってよい。一致が見つかる場合、オペレーティングシステムは、ブロック608において、物理ページのうちの1つの競合する仮想アドレスを再割り当てすることによって、空き物理ページペア399を作成してよい。関連のある競合する仮想アドレスを移動した後に、オリジナル物理アドレスを有する残りの物理ページは、今では新しい空き物理ページペア399として利用可能である。
【００２３】
図7は、以前に割り振られた単一のページをペアにされたページに変換することによってメモリ複製の性能を向上させるための別の技法を示す。方法700は、メモリ圧迫が少なく、デフォルトの複製しないモードから上記の複製するモードに切り替えることが望ましい状況を決定するために使用されてよい。決定ブロック702において、オペレーティングシステムは、空きページの総数が最小しきい値を超えるかどうかを決定してよい。しきい値を超えない場合、決定ブロック702が所定の間隔で繰り返されてよい。しきい値を超える場合、ブロック704において、オペレーティングシステムは、グローバルディレクトリですべてのメモリ物理ページフレームを検索してよい。オペレーティングシステムは、コピービット値452が有効ではない(値=「0」または「false」)ページが存在するが、一致するページペアは空いているかどうかを決定してよい。決定ブロック706において、一致が探し当てられる場合、オペレーティングシステムはページを、ページペアの他方にコピーし、コピービット値を「1」または「true」に設定してよい。
【００２４】
図8～図12は、同じ物理アドレスを共有しない空き物理ページペアに基づくノンブロッキングの、無名の(Anonymous)割振りを可能にする複製方式の別の実施形態を示す。この複製方式は、システム100においてノンブロッキングNUMAレイテンシモジュール203(図1)によって実施されてよい。一致する物理アドレスを顧慮せずにメモリデータを選択的に複製するには、図8に示すように、変更されたページテーブルエントリ800が与えられてよい。変更されたページテーブルエントリ800は、コピー属性フィールド450と、第1のDRAM104と関連する第1の物理アドレスへの物理ページインデックス#1 1406を記憶するためのフィールド802と、レプリカアドレスを記憶するための新しいフィールド804とを含む。レプリカアドレスは、第2のDRAM204と関連する第2の物理アドレス1408への物理ページインデッ

クス#2を含む。この点について、ページテーブルエントリ800は、単一の仮想アドレスを、各ダイから1つの、任意の(たとえば、同じである必要はない)物理アドレスを有する物理ページペアにマップすることをサポートしてよい。ページテーブルエントリ800は、両方の物理アドレスへの変換を可能にする。図12は、単一の仮想アドレスが、ページインデックス1402(第13ビット以上)、およびページオフセット1404(下位12ビット)でどのように構成されるか、また1つのページインデックス1402が、ページテーブルフィールド802および804にそれぞれ基づいて、物理ページインデックス#1 1406および物理ページインデックス#2 1408にどのようにマップするかを示す。ページオフセット1404は、変更されず、各4KBページ内のワードにアクセスするために使用される。

【００２５】

図9は、任意の物理ページペアを使用してNUMAでメモリを割り振るための方法900を示す。ブロック902において、第1のSoC102上で実行されているプロセスから仮想メモリページについての要求が受け取られる場合がある。システムは、複製を可能にするための十分なメモリがあるかどうかを決定するための様々なしきい値を実装してよい。複製がページごとに行われる場合があることを諒解されたい。一実施形態では、決定ブロック904において、システムは、利用可能な物理ページの数が、DRAM104を使用するSoC102についての最小しきい値を超えるかどうかを決定してよい。「はい」の場合、ブロック906において、SoC102についてのメモリ充足値が「true」に設定されてよい。「いいえ」の場合、ブロック908において、SoC102についてのメモリ充足値が「false」に設定されてよい。決定ブロック910において、システムは、利用可能な物理ページの数が、DRAM204を使用するSoC202についての最小しきい値を超えるかどうかを決定してよい。「はい」の場合、ブロック912において、SoC202についてのメモリ充足値が「true」に設定されてよい。「いいえ」の場合、ブロック914において、SoC202についてのメモリ充足値が「false」に設定されてよい。メモリ充足値に基づいて、ブロック916においてシステムは、実行する適切な割振りアクションを決定してよい。図9bに示すように、SoC102とSoC202の両方において十分なメモリが利用可能である(すなわち、両方の値=「true」)場合、オペレーティングシステムは、DRAM104を使用するSoC102、およびDRAM204を使用するSoC202からのページを割り振り、コピービット値を「ture」または「1」に設定して、コピービット値がこのように有効であり、レプリカアドレスもページテーブルエントリ800に追加され得るとき、複製を可能にし得る。SoC102とSoC202の両方ではなく、SoC102またはSoC202のいずれかで十分なメモリが利用可能である場合、オペレーティングシステムは、どちらのSoCが十分なメモリを有する(すなわち、値=「ture」)かに応じて、DRAM104を使用するSoC102、またはDRAM204を使用するSoC202からの単一のページを割り振ってよい。SoC102とSoC202の両方に、十分なメモリがない(すなわち、両方の値=「false」)場合、オペレーティングシステムは、割振りをすることができず、例外をトリガすることになる。割振り失敗に対する例外処理は、既存の方法と変わらないものであり、より優先度の低い、まれにアクセスされるプロセスを、それらに割り振られているメモリを空けるために、終了する実行プログラムまたはサービスを呼び出す。メモリが共有され得るNUMAでは、単一のページは、DRAM104を使用するSoC102、またはDRAM204を使用するSoC202のいずれかから割り振られる場合があることを諒解されたい。

【００２６】

図10は、変更されたページテーブルエントリ800を含んだメモリ書込みトランザクションを実行するための方法1000の一実施形態を示す。方法1000は、ソフトウェアおよび/またはハードウェアによって実施される場合があることを諒解されたい。ハードウェア実施形態では、方法は、たとえば、メモリ管理ユニット(MMU)においてトランスレーションルックアサイドバッファ(TLB:Translation Look Aside Buffer)によって実施される場合がある。ブロック1000において、TLBによってメモリ書込みトランザクションが受け取られる。ブロック1004において、TLBは、ページテーブルエントリ800へのルックアップを実行する。ブロック1006において、コピービット値452が読み取られる。コピービット値が「true」である場合(決定ブロック1008)、レプリカアドレスが読み取られ(ブロック1010)、

キャッシュハードウェアがオリジナル物理アドレスとレプリカ物理アドレスの両方へデータをフラッシュする。コピービット値が「false」である場合、キャッシュハードウェアは、オリジナル物理アドレスのみへデータをフラッシュする。

【００２７】

図11は、変更されたページテーブルエントリ800を含んだメモリ読取りトランザクションを実行するための方法1100の一実施形態を示す。方法1100は、ソフトウェアおよび/またはハードウェアによって実施される場合があることを諒解されたい。ハードウェア実施形態では、方法は、たとえば、メモリ管理ユニット(MMU)においてトランスレーションルックアサイドバッファ(TLB)によって実施される場合がある。ブロック1100において、TLBによってメモリ読取りトランザクションが受け取られる。ブロック1104において、TLBは、ページテーブルエントリ800へのルックアップを実行する。ブロック1106において、コピービット値452が読み取られる。コピービット値が「true」である場合(決定ブロック1108)、レプリカアドレスが読み取られ(ブロック1110)、キャッシュフィルがレプリカアドレスまたはオリジナルアドレスのいずれかから行われる(ブロック1112)。コピービット値が「false」である場合、オリジナルアドレスからキャッシュフィルが行われる。

【００２８】

上述のように、システム100は任意の望ましいコンピューティングシステムに組み込むことができる。図13は、SoC102と、SoC202とを備えた例示的なポータブルコンピューティングデバイス(PCD)1300を示す。この実施形態では、SoC102およびSoC202は、マルチコアCPU1302を含む場合がある。マルチコアCPU1302は、第0のコア1310と、第1のコア1312と、第Nのコア1314とを含む場合がある。コアのうちの1つは、たとえば、グラフィックス処理ユニット(GPU)を含み、他のコアのうちの1つまたは複数はCPUを含む場合がある。

【００２９】

ディスプレイコントローラ328およびタッチスクリーンコントローラ330は、CPU602に結合されてもよい。一方、SoC102および202の外部にあるタッチスクリーンディスプレイ606は、ディスプレイコントローラ328およびタッチスクリーンコントローラ330に結合されてもよい。

【００３０】

図13は、ビデオエンコーダ334、たとえば、位相反転線(PAL)エンコーダ、順次式カラーメモリ(SECAM)エンコーダ、または全米テレビジョン方式委員会(NTSC)エンコーダが、マルチコアCPU1302に結合されることをさらに示す。さらに、ビデオ増幅器336がビデオエンコーダ334とタッチスクリーンディスプレイ1306とに結合される。また、ビデオポート338がビデオ増幅器336に結合される。図13に示すように、ユニバーサルシリアルバス(USB)コントローラ340が、マルチコアCPU602に結合される。また、USBポート342が、USBコントローラ340に結合される。メモリ104および204ならびに加入者識別モジュール(SIM)カード346が、マルチコアCPU1302に結合される場合もある。

【００３１】

さらに、図13に示すように、デジタルカメラ348は、マルチコアCPU1302に結合される場合がある。例示的な態様では、デジタルカメラ348は、電荷結合デバイス(CCD)カメラまたは相補型金属酸化物半導体(CMOS)カメラである。

【００３２】

図13にさらに示すように、ステレオオーディオコーダ-デコーダ(コーデック)350が、マルチコアCPU802に結合されてもよい。さらに、オーディオ増幅器352が、ステレオオーディオコーデック350に結合されてもよい。例示的な態様では、第1のステレオスピーカ354および第2のステレオスピーカ356が、オーディオ増幅器352に結合される。図13は、マイクロフォン増幅器358もステレオオーディオコーデック350に結合される場合があることを示している。加えて、マイクロフォン360が、マイクロフォン増幅器358に結合される場合がある。特定の態様では、周波数変調(FM)無線チューナー362がステレオオーディオコーデック350に結合される場合がある。また、FMアンテナ364が、FM無線チューナー362に結合される。さらに、ステレオヘッドフォン366が、ステレオオーディオコーデック350に結

合される場合がある。

【００３３】

図13は、無線周波数(RF)トランシーバ368がマルチコアCPU1302に結合される場合があることをさらに示す。RFスイッチ370が、RFトランシーバ368とRFアンテナ372とに結合される場合がある。キーパッド204が、マルチコアCPU602に結合される場合がある。また、マイクロフォンを備えたモノヘッドセット376が、マルチコアCPU1302に結合される場合がある。さらに、バイブレータデバイス378がマルチコアCPU1302に結合される場合がある。

【００３４】

図13はまた、電源380がSoC102およびSoC202に結合される場合があることを示す。特定の態様では、電源380は、電力を必要とするPCD1300の種々の構成要素に電力を供給する直流(DC)電源である。さらに、特定の態様では、電源は、充電式DCバッテリー、または交流(AC)電源に接続されたAC/DC変換器から得られるDC電源である。

【００３５】

図13は、PCD1300が、データネットワーク、たとえば、ローカルエリアネットワーク、パーソナルエリアネットワーク、または任意の他のネットワークにアクセスするために使用される場合があるネットワークカード388を含む場合もあることをさらに示す。ネットワークカード388は、ブルートゥース(登録商標)ネットワークカード、WiFiネットワークカード、パーソナルエリアネットワーク(PAN)カード、パーソナルエリアネットワーク超低電力技術(PeANUT)ネットワークカード、テレビジョン/ケーブル/衛星チューナー、または当技術分野でよく知られている任意の他のネットワークカードとすることができる。さらに、ネットワークカード388は、チップに組み込まれる場合があり、すなわち、ネットワークカード388は、チップ内のフルソリューションである場合があり、別個のネットワークカード388でなくてもよい。

【００３６】

図13を参照すると、メモリ104、RAMカード105、タッチスクリーンディスプレイ606、ビデオポート338、USBポート342、カメラ348、第1のステレオスピーカ354、第2のステレオスピーカ356、マイクロフォン360、FMアンテナ364、ステレオヘッドフォン366、RFスイッチ370、RFアンテナ372、キーパッド374、モノヘッドセット376、バイブレータ378、および電源380は、オンチップシステム102の外部に存在する場合があることを諒解されたい。

【００３７】

本明細書で説明した方法ステップのうちの1つまたは複数は、上記のモジュールなどのコンピュータプログラム命令としてメモリに記憶される場合があることを諒解されたい。これらの命令は、本明細書で説明した方法を実行するために、対応するモジュールと組み合わせてまたは協働して、任意の適切なプロセッサによって実行される場合がある。

【００３８】

本明細書で説明したプロセスまたはプロセスフローにおける特定のステップは、当然、説明したように本発明が機能するために他のステップに先行する。しかしながら、そのような順序またはシーケンスが本発明の機能を変えない場合、本発明は、説明したステップの順序に限定されない。すなわち、本発明の範囲および趣旨から逸脱することなく、いくつかのステップは、他のステップの前に実行されるか、後に実行されるか、または他のステップと並行して(実質的に同時に)実行されてよいことを認識されたい。場合によっては、本発明から逸脱することなく、いくつかのステップが省略されてもよく、または実行されなくてもよい。さらに、「その後」、「次いで」、「次に」などの語は、ステップの順番を限定することを意図していない。これらの言葉は単に、例示的な方法の説明に読者を導くために使用される。

【００３９】

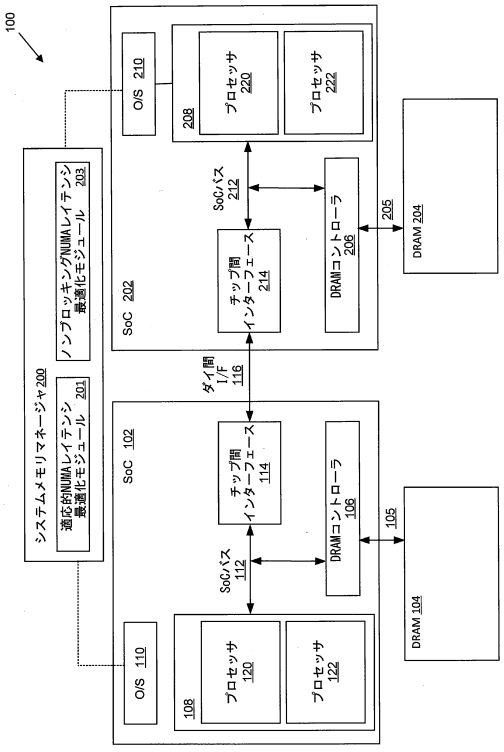
さらに、プログラミングに関する当業者は、たとえば、本明細書におけるフローチャートおよび関連する説明に基づいて、難なく、開示した発明を実装するコンピュータコードを書くことができるか、または実装するのに適したハードウェアおよび/もしくは回路を特定することができる。

【００４０】
したがって、プログラムコード命令または詳細なハードウェアデバイスの特定のセットの開示は、本発明をどのように作製し使用するのかを十分に理解するために必要であるとは見なされない。特許請求されるコンピュータ実施プロセスの発明性のある機能は、上記の説明において、かつ様々なプロセスフローを示す場合がある図面とともに、より詳細に説明される。
【００４１】
1つまたは複数の例示的な態様では、説明する機能は、ハードウェア、ソフトウェア、ファームウェア、またはそれらの任意の組合せで実装されてもよい。ソフトウェアに実装される場合、機能は、1つもしくは複数の命令もしくはコードとして、コンピュータ可読媒体上に記憶される場合もあり、またはコンピュータ可読媒体上に送信される場合もある。コンピュータ可読媒体は、ある場所から別の場所へのコンピュータプログラムの転送を容易にする任意の媒体を含む、コンピュータ記憶媒体とコンピュータ通信媒体との両方を含む。記憶媒体は、コンピュータによってアクセスされる場合がある任意の利用可能な媒体であってもよい。限定ではなく例として、そのようなコンピュータ可読媒体は、RAM、ROM、EEPROM、NANDフラッシュ、NORフラッシュ、M-RAM、P-RAM、R-RAM、CD-ROMもしくは他の光ディスクストレージ、磁気ディスクストレージもしくは他の磁気記憶デバイス、または命令もしくはデータ構造の形態で所望のプログラムコードを搬送もしくは記憶するために使用され、コンピュータによってアクセスされてもよい任意の他の媒体を備えてもよい。
【００４２】
また、いかなる接続も、厳密にはコンピュータ可読媒体と呼ばれる。たとえば、ソフトウェアが、同軸ケーブル、光ファイバーケーブル、ツイストペア、デジタル加入者回線(「DSL」)、または赤外線、無線、およびマイクロ波などのワイヤレス技術を使用してウェブサイト、サーバ、または他のリモートソースから送信される場合、同軸ケーブル、光ファイバーケーブル、ツイストペア、DSL、または赤外線、無線、およびマイクロ波などのワイヤレス技術は、媒体の定義に含まれる。
【００４３】
ディスク(disk)およびディスク(disc)は、本明細書で使用するときに、コンパクトディスク(disc)(「CD」)、レーザーディスク(登録商標)(disc)、光ディスク(disc)、デジタル多用途ディスク(disc)(「DVD」)、フロッピーディスク(disk)およびブルーレイディスク(disc)を含み、ディスク(disk)は通常、データを磁気的に再生するが、ディスク(disc)は、レーザーを用いてデータを光学的に再生する。上記の組合せも、コンピュータ可読媒体の範囲に含まれるべきである。
【００４４】
本発明の趣旨および範囲から逸脱することなく、本発明が関係する代替的な実施形態が、当業者には明らかになるであろう。したがって、選択された態様が図示され詳細に説明されてきたが、以下の特許請求の範囲によって定義されるように、本発明の趣旨および範囲から逸脱することなく、各態様において様々な置換および改変が行われてよいことが理解されよう。
【符号の説明】
【００４５】
100 システム
102 システムオンチップ(SoC)
104 ダイナミックランダムアクセスメモリ(DRAM)
105 バス
106 DRAMコントローラ
108 プロセッサ
110 オペレーティングシステム
112 SoCバス

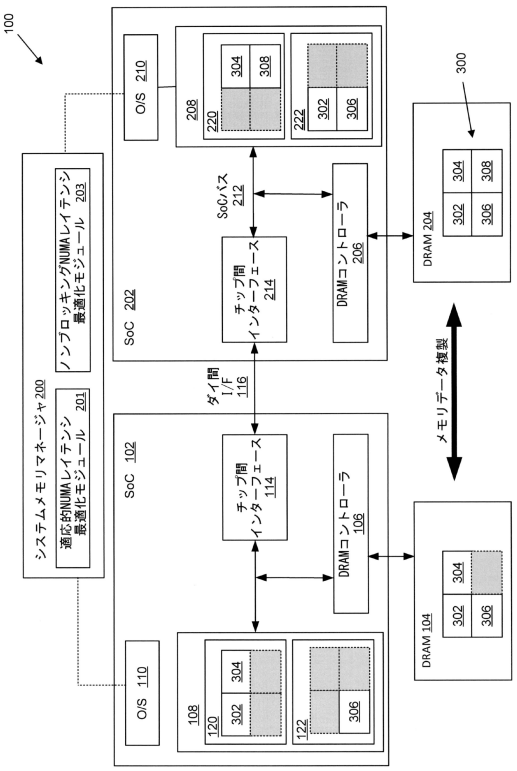
114 チップ間インターフェースコントローラ
116 ダイ間インターフェース
120 プロセッサ
122 プロセッサ
200 システムメモリマネージャ
201 適応的NUMAレイテンシ最適化モジュール
202 SoC
203 ノンブロッキングNUMAレイテンシ最適化モジュール
204 DRAM
205 バス
206 DRAMコントローラ
208 プロセッサ
210 オペレーティングシステム
212 SoCバス
214 チップ間インターフェースコントローラ
220 プロセッサ
222 プロセッサ
300 メモリデータ
302 データ
304 データ
306 データ
308 データ
328 ディスプレイコントローラ
330 タッチスクリーンコントローラ
334 ビデオエンコーダ
336 ビデオ増幅器
338 ビデオポート
340 ユニバーサルシリアルバス(USB)コントローラ
342 USBポート
346 加入者識別モジュール(SIM)カード
348 デジタルカメラ
350 ステレオ/オーディオコーデック
352 オーディオ増幅器
354 ステレオスピーカ
356 ステレオスピーカ
358 マイクロフォン増幅器
360 マイクロフォン
362 周波数変調(FM)無線チューナー
364 FMアンテナ
366 ステレオヘッドフォン
368 無線周波数(RF)トランシーバ
370 RFスイッチ
372 RFアンテナ
374 キーパッド
376 モノヘッドセット
378 バイブレータデバイス
380 電源
388 ネットワークカード
399 空き物理ページペア
402 仮想メモリマネージャ

404　仮想メモリマネージャ
450　コピー属性フィールド
452　コピービット値
454　物理ページインデックス
1300　PCD
1302　マルチコアCPU
1306　ディスプレイ/タッチスクリーン
1310　第0のコア
1312　第1のコア
1314　第Nのコア

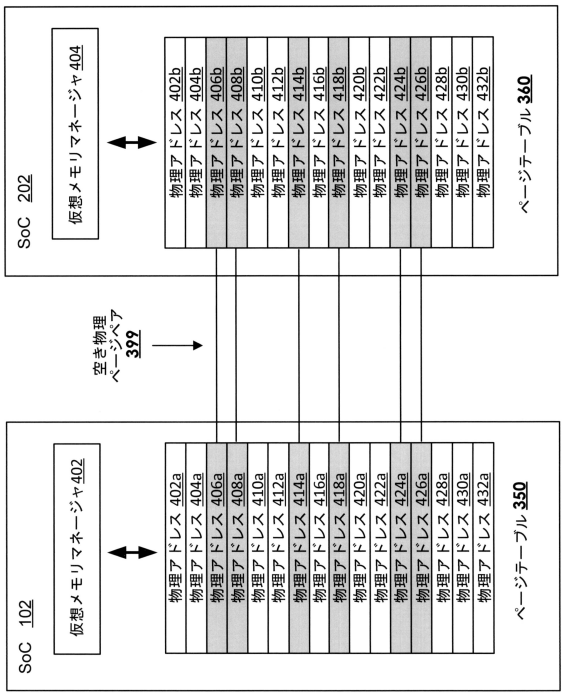
【図１】



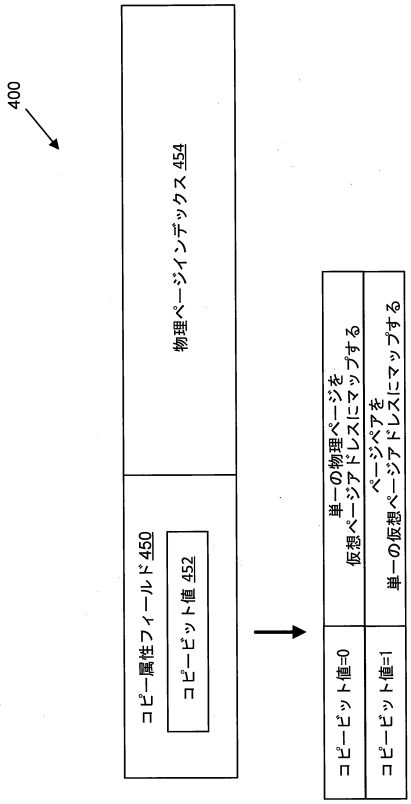
【図２】



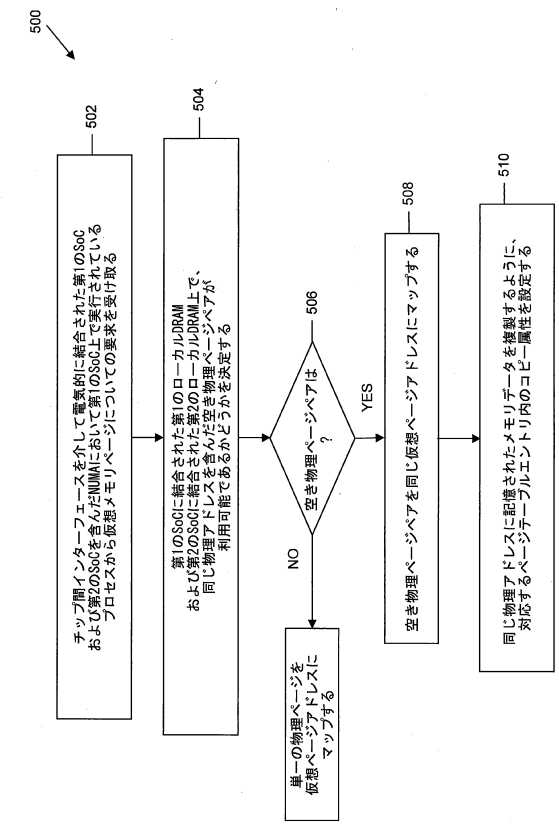
【図３】


SoC 202
仮想メモリマネージャ404
物理アドレス 402b
物理アドレス 404b
物理アドレス 406b
物理アドレス 408b
物理アドレス 410b
物理アドレス 412b
物理アドレス 414b
物理アドレス 416b
物理アドレス 418b
物理アドレス 420b
物理アドレス 422b
物理アドレス 424b
物理アドレス 426b
物理アドレス 428b
物理アドレス 430b
物理アドレス 432b
ページテーブル 360
空き物理ページペア 399
SoC 102
仮想メモリマネージャ402
物理アドレス 402a
物理アドレス 404a
物理アドレス 406a
物理アドレス 408a
物理アドレス 410a
物理アドレス 412a
物理アドレス 414a
物理アドレス 416a
物理アドレス 418a
物理アドレス 420a
物理アドレス 422a
物理アドレス 424a
物理アドレス 426a
物理アドレス 428a
物理アドレス 430a
物理アドレス 432a
ページテーブル 350

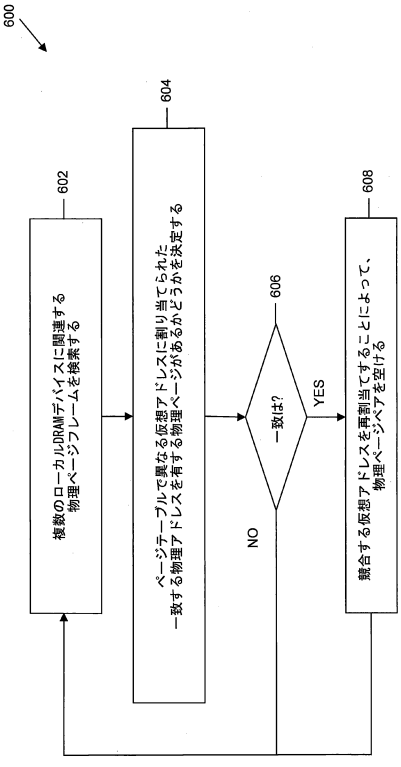
【図４】


400
コピー属性フィールド 450
コピービット値 452
物理ページインデックス 454
コピービット値=0
単一の物理ページを仮想ページアドレスにマップする
コピービット値=1
ページペアを単一の仮想ページアドレスにマップする

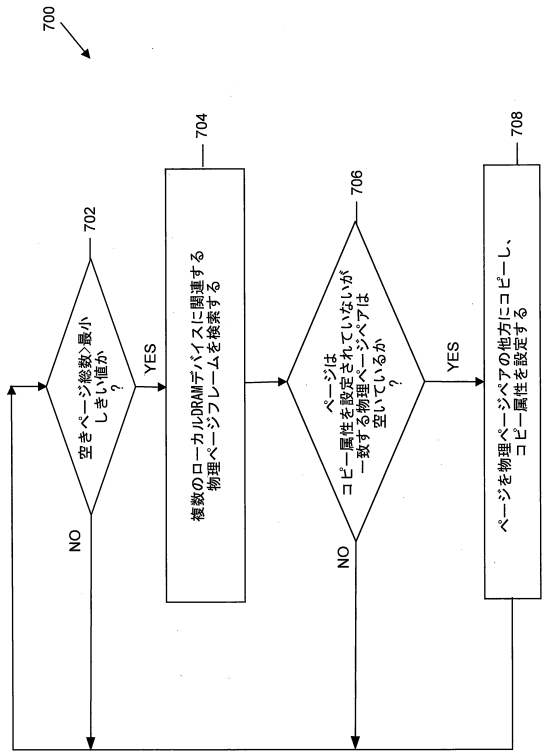
【図５】


500
チップ間インターフェースを介して電気的に結合された第1のSoCおよび第2のSoCを含んだNUMAにおいて第1のSoC上で実行されているプロセスから仮想メモリページについての要求を受け取る
502
第1のSoCに結合された第1のローカルDRAMおよび第2のSoCに結合された第2のローカルDRAM上で、同じ物理アドレスを含んだ空き物理ページペアが利用可能であるかどうかを決定する
504
空き物理ページペアは？
506
NO
YES
単一の物理ページを仮想ページアドレスにマップする
空き物理ページペアを同じ仮想ページアドレスにマップする
508
同じ物理アドレスに記憶されたメモリデータを複製するように、対応するページテーブルエントリ内のコピー属性を設定する
510

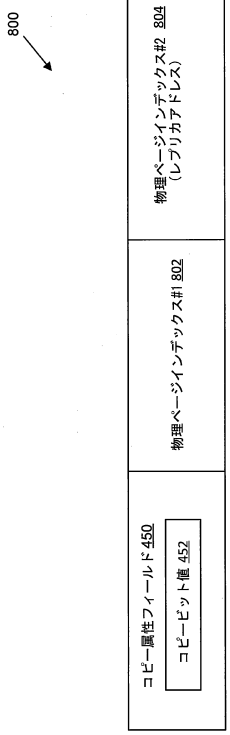
【図６】


600
複数のローカルDRAMデバイスに関連する物理ページフレームを検索する
602
ページテーブルで異なる仮想アドレスに割り当てられた一致する物理アドレスを有する物理ページがあるかどうかを決定する
604
一致は?
606
NO
YES
競合する仮想アドレスを再割当てすることによって、物理ページペアを空ける
608

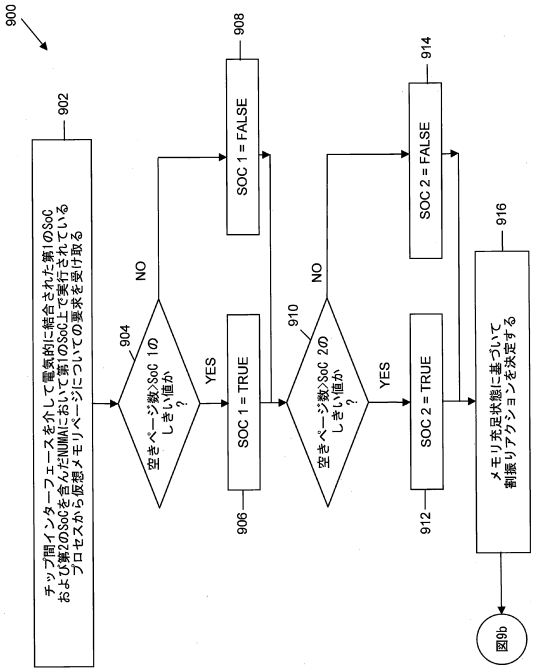
【図７】


700
702
空きページ総数>最小しきい値か？
YES
NO
704
複数のローカルDRAMデバイスに関連する物理ページフレームを検索する
706
ページはコピー属性を設定されていないが一致する物理ページペアは空いているか？
YES
NO
708
ページを物理ページペアの他方にコピーし、コピー属性を設定する

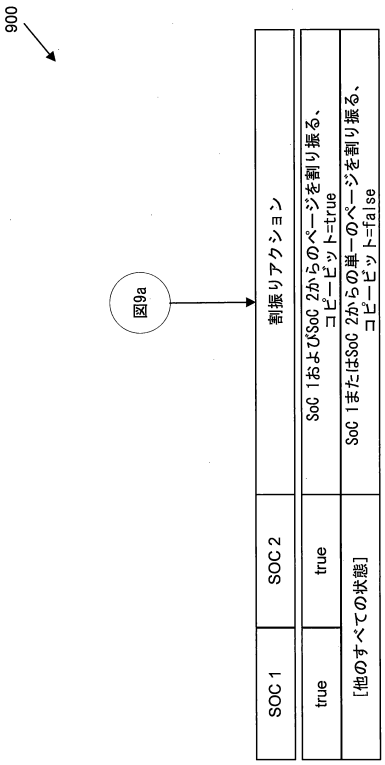
【図８】


800
コピー属性フィールド 450
コピービット値 452
物理ページインデックス#1 802
物理ページインデックス#2 804
（レプリカアドレス）

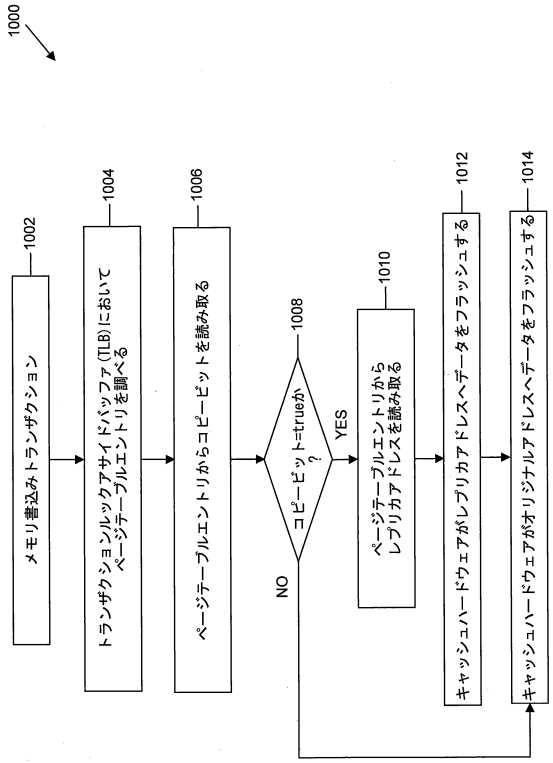
【図９ａ】


900
902
チップ間インターフェースを介して電気的に結合された第1のSoCおよび第2のSoCを含んだNUMAにおいて第1のSoC上で実行されているプロセスから仮想メモリページについての要求を受け取る
904
空きページ数>SoC 1のしきい値か？
YES
NO
906
SOC 1 = TRUE
908
SOC 1 = FALSE
910
空きページ数>SoC 2のしきい値か？
YES
NO
912
SOC 2 = TRUE
914
SOC 2 = FALSE
916
メモリ充足状態に基づいて割振りアクションを決定する
図9b

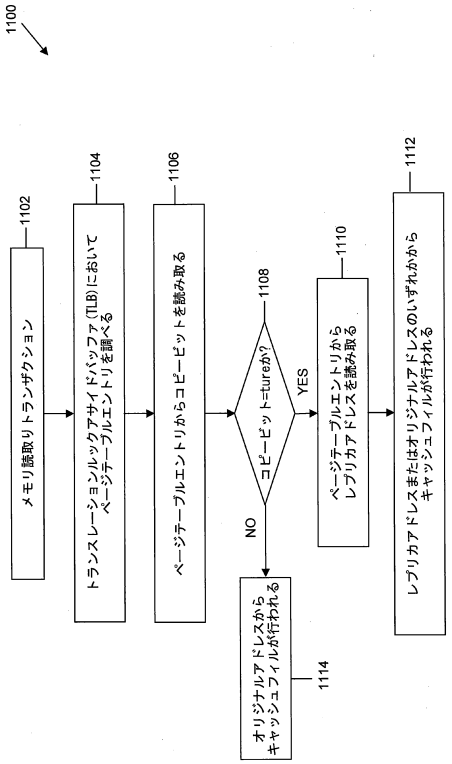
【図９ｂ】


900
図9a
割振りアクション
SOC 1
SOC 2
true
true
SoC 1およびSoC 2からのページを割り振る、コピービット=true
[他のすべての状態]
SoC 1またはSoC 2からの単一のページを割り振る、コピービット=false

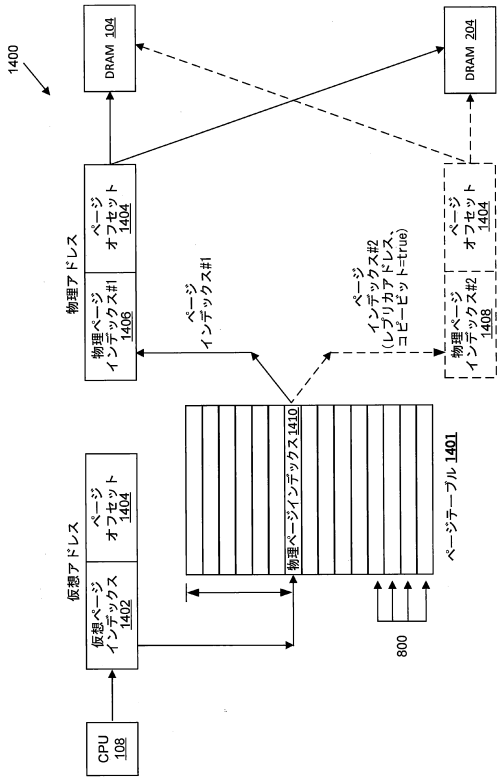
【図１０】



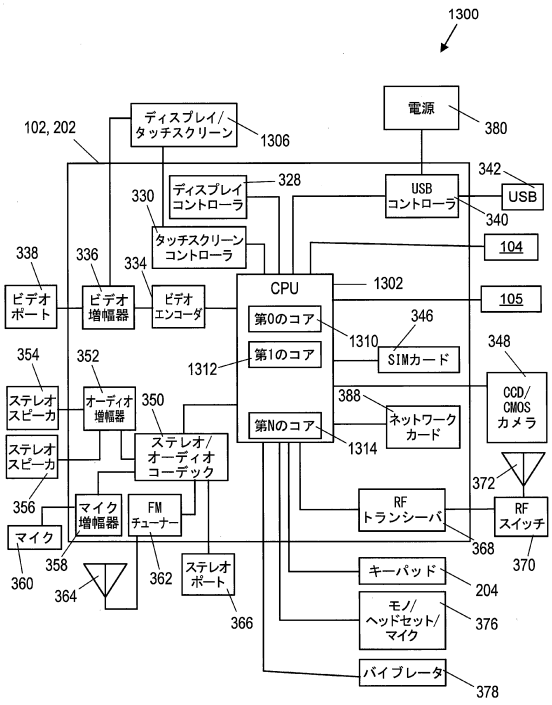
【図１１】



【図１２】



【図１３】



フロントページの続き

(72)発明者　デクスター・タミオ・チュン
アメリカ合衆国・カリフォルニア・９２１２１・サン・ディエゴ・モアハウス・ドライヴ・５７７５

審査官　中村　康司

(56)参考文献　特開２０１１－０６５６５０（ＪＰ，Ａ）
米国特許出願公開第２００９／０１５３８９７（ＵＳ，Ａ１）

(58)調査した分野(Int.Cl.，ＤＢ名)
Ｇ０６Ｆ　１２／１０
Ｇ０６Ｆ　１２／０８