(19)日本国特許庁(JP)　(12)公開特許公報(A)　(11)特許出願公開番号
特開2020-77075
(P2020-77075A)
(43)公開日　令和2年5月21日(2020.5.21)

(51)Int.Cl.		F I			テーマコード（参考）
G06F 12/06	(2006.01)	G06F	12/06	522D	5B160
G06F 12/1027	(2016.01)	G06F	12/1027	120	5B205

審査請求 未請求　請求項の数 14　ＯＬ　（全 21 頁）

(21)出願番号　特願2018-208545 (P2018-208545)
(22)出願日　平成30年11月6日 (2018.11.6)

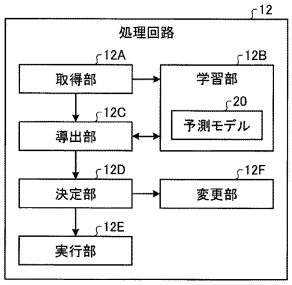
（出願人による申告）平成２８年度　国立研究開発法人新エネルギー・産業技術総合開発機構「ＩｏＴ推進のための横断技術開発プロジェクト／高速大容量ストレージデバイス・システムの研究開発」委託研究、産業技術力強化法第１９条の適用を受ける特許出願

(71)出願人　000003078
株式会社東芝
東京都港区芝浦一丁目１番１号
(74)代理人　110002147
特許業務法人酒井国際特許事務所
(72)発明者　城田　祐介
東京都港区芝浦一丁目１番１号　株式会社東芝内
(72)発明者　金井　達徳
東京都港区芝浦一丁目１番１号　株式会社東芝内

Ｆターム(参考)　5B160 MM03
5B205 MM51　VV03

(54)【発明の名称】情報処理装置、情報処理方法、およびプログラム

(57)【要約】
【課題】複数種類のメモリの使い分けに用いる情報を効率よく提供する。
【解決手段】情報処理装置１０は、取得部１２Ａと、導出部１２Ｃと、決定部１２Ｄと、を備える。取得部１２Ａは、処理回路１２の動作統計情報４２Ａを取得する。導出部１２Ｃは、動作統計情報４２Ａから処理回路１２のメモリアクセス特性４２Ｂを導出するための予測モデル２０に基づいて、取得した動作統計情報４２Ａからメモリアクセス特性４２Ｂを導出する。決定部１２Ｄは、導出したメモリアクセス特性４２Ｂに基づいて、第１記憶部１４Ａより処理回路１２によるアクセス速度が遅い第２記憶部１４Ｂのデータを第１記憶部１４Ａへ転送し、第１記憶部１４Ａ内の該データにアクセスする第１アクセス方式、または、第２記憶部１４Ｂ内のデータにアクセスする第２アクセス方式、の何れかのアクセス方式を決定する。
【選択図】図３



【特許請求の範囲】
【請求項１】
処理回路の動作統計情報を取得する取得部と、
前記動作統計情報から前記処理回路のメモリアクセス特性を導出するための予測モデルに基づいて、取得した前記動作統計情報から前記メモリアクセス特性を導出する導出部と、
導出した前記メモリアクセス特性に基づいて、第１記憶部より前記処理回路によるアクセス速度が遅い第２記憶部のデータを前記第１記憶部へ転送し、前記第１記憶部内の該データにアクセスする第１アクセス方式、または、前記第２記憶部内のデータにアクセスする第２アクセス方式、の何れかのアクセス方式を決定する決定部と、
を備える情報処理装置。
【請求項２】
決定された前記アクセス方式に応じて、前記データの前記第２記憶部から前記第１記憶部への転送および前記第１記憶部内の該データへのアクセス、または、前記第２記憶部内のデータへのアクセス、を実行する実行部、
を備える請求項１に記載の情報処理装置。
【請求項３】
前記動作統計情報は、
前記処理回路が実行中のアプリケーションに割当てられた物理メモリサイズ、および、ＴＬＢ（Ｔｒａｎｓｌａｔｉｏｎ　Ｌｏｏｋａｓｉｄｅ　Ｂｕｆｆｅｒ）ミスに関する動作統計情報、の少なくとも一方を含む、
請求項１または請求項２に記載の情報処理装置。
【請求項４】
前記メモリアクセス特性は、
前記処理回路が単位期間あたりに使用したメモリサイズを示す、
請求項１～請求項３の何れか１項に記載の情報処理装置。
【請求項５】
前記決定部は、
導出した前記メモリアクセス特性が第１閾値より大きい場合、前記第２アクセス方式を決定し、該メモリアクセス特性が前記第１閾値以下の場合、前記第１アクセス方式を決定する、
請求項４に記載の情報処理装置。
【請求項６】
第１閾値は、前記処理回路が利用可能な、前記第１記憶部のサイズ以上の値である、請求項５に記載の情報処理装置。
【請求項７】
前記決定部は、
取得した前記動作統計情報に関する１または複数のアプリケーションの各々に割当てられた物理メモリサイズの合計値に対する、前記メモリアクセス特性の比率が、第２閾値より大きい場合、前記第２アクセス方式を決定し、前記第２閾値以下の場合、前記第１アクセス方式を決定する、
請求項１～請求項３の何れか１項に記載の情報処理装置。
【請求項８】
前記第２閾値は、前記合計値のＮ分の１である（Ｎは、２以上の整数）、請求項７に記載の情報処理装置。
【請求項９】
前記第１アクセス方式を決定した場合、
前記第１記憶部の利用可能なメモリサイズを変更する変更部、
を備える、請求項１～請求項８の何れか１項に記載の情報処理装置。
【請求項１０】

前記動作統計情報と前記メモリアクセス特性との対応を示す教師データを複数含む教師データセットを用いて、前記予測モデルを学習する学習部、
を備える、請求項１～請求項９の何れか１項に記載の情報処理装置。
【請求項１１】
前記教師データは、アプリケーションの命令単位ごとの、前記動作統計情報と前記メモリアクセス特性との対応を示す、
請求項１０に記載の情報処理装置。
【請求項１２】
前記学習部は、学習用のアプリケーションを少なくとも２回実行し、一方の回の該アプリケーションの実行によって前記動作統計情報を取得し、他方の回の該アプリケーションの実行によって前記メモリアクセス特性を取得し、該アプリケーションの命令単位ごとに、取得した前記動作統計情報と取得した前記メモリアクセス特性との対応を示す前記教師データを生成する、
請求項１１に記載の情報処理装置。
【請求項１３】
処理回路の動作統計情報を取得するステップと、
前記動作統計情報から前記処理回路のメモリアクセス特性を導出するための予測モデルに基づいて、取得した前記動作統計情報から前記メモリアクセス特性を導出するステップと、
導出した前記メモリアクセス特性に基づいて、第１記憶部より前記処理回路によるアクセス速度が遅い第２記憶部のデータを前記第１記憶部へ転送し、前記第１記憶部内の該データにアクセスする第１アクセス方式、または、前記第２記憶部内のデータにアクセスする第２アクセス方式、の何れかのアクセス方式を決定するステップと、
を含む情報処理方法。
【請求項１４】
処理回路の動作統計情報を取得するステップと、
前記動作統計情報から前記処理回路のメモリアクセス特性を導出するための予測モデルに基づいて、取得した前記動作統計情報から前記メモリアクセス特性を導出するステップと、
導出した前記メモリアクセス特性に基づいて、第１記憶部より前記処理回路によるアクセス速度が遅い第２記憶部のデータを前記第１記憶部へ転送し、前記第１記憶部内の該データにアクセスする第１アクセス方式、または、前記第２記憶部内のデータにアクセスする第２アクセス方式、の何れかのアクセス方式を決定するステップと、
をコンピュータに実行させるためのプログラム。
【発明の詳細な説明】
【技術分野】
【０００１】
本発明の実施形態は、情報処理装置、情報処理方法、およびプログラムに関する。
【背景技術】
【０００２】
ＭＲＡＭ（Ｍａｇｎｅｔｏｒｅｓｉｓｔｉｖｅ Ｒａｎｄｏｍ Ａｃｃｅｓｓ Ｍｅｍｏｒｙ）、ＲｅＲＡＭ（Ｒｅｓｉｓｔｉｖｅ ＲＡＭ）、ＰＣＭ（Ｐｈａｓｅ－Ｃｈａｎｇｅ Ｍｅｍｏｒｙ）などの各種のストレージクラスメモリ（ＳＣＭ）が開発されている。ＳＣＭは、ＤＲＡＭ（Ｄｙｎａｍｉｃ Ｒａｎｄｏｍ Ａｃｃｅｓｓ Ｍｅｍｏｒｙ）に比べてアクセス速度は遅いが、集積度が高い。一方、ＤＲＡＭは、ＳＣＭに比べて集積度は低いが、アクセス速度が速い。このため、複数種類のメモリを搭載したシステムの場合、これらのメモリを使い分けて用いる必要がある。
【０００３】
しかし、従来では、複数の種類のメモリの使い分けに用いる情報が管理されておらず、また、この情報を収集する手段を有していなかった。このため、従来では、複数種類のメ

モリの使い分けに用いる情報を効率よく提供することは困難であった。
【先行技術文献】
【非特許文献】
【０００４】
【非特許文献１】Ｒ．Ｆ．Ｆｒｅｉｔａｓ ａｎｄ Ｗ．Ｗ．Ｗｉｌｃｋｅ，"Ｓｔｏｒａｇｅ－ｃｌａｓｓ Ｍｅｍｏｒｙ：Ｔｈｅ Ｎｅｘｔ Ｓｔｏｒａｇｅ Ｓｙｓｔｅｍ Ｔｅｃｈｎｏｌｏｇｙ"，ＩＢＭ Ｊｏｕｒｎａｌ ｏｆ Ｒｅｓｅａｒｃｈ ａｎｄ Ｄｅｖｅｌｏｐｍｅｎｔ Ｖｏｌ．５２ Ｎｏ．４，ｐｐ．４３９－４４７，２００８．
【発明の概要】
【発明が解決しようとする課題】
【０００５】
本発明が解決しようとする課題は、複数種類のメモリの使い分けに用いる情報を効率よく提供することができる、情報処理装置、情報処理方法、およびプログラムを提供することである。
【課題を解決するための手段】
【０００６】
実施形態の情報処理装置は、取得部と、導出部と、決定部と、を備える。取得部は、処理回路の動作統計情報を取得する。導出部は、前記動作統計情報から前記処理回路のメモリアクセス特性を導出するための予測モデルに基づいて、取得した前記動作統計情報から前記メモリアクセス特性を導出する。決定部は、導出した前記メモリアクセス特性に基づいて、第１記憶部より前記処理回路によるアクセス速度が遅い第２記憶部のデータを前記第１記憶部へ転送し、前記第１記憶部内の該データにアクセスする第１アクセス方式、または、前記第２記憶部内のデータにアクセスする第２アクセス方式、の何れかのアクセス方式を決定する。
【図面の簡単な説明】
【０００７】
【図１】情報処理装置の構成の一例を示す模式図。
【図２】物理アドレス空間の模式図。
【図３】処理回路の機能ブロック図。
【図４】予測モデルの学習の説明図。
【図５】動作統計情報とメモリアクセス特性との関係の説明図。
【図６Ａ】動作統計情報の模式図。
【図６Ｂ】メモリアクセス特性の模式図。
【図７】導出部および決定部の処理の説明図。
【図８】アクセス方式の決定の説明図。
【図９】情報処理の手順のフローチャート。
【図１０】情報処理の手順のフローチャート。
【発明を実施するための形態】
【０００８】
以下に添付図面を参照して、本実施の形態の詳細を説明する。
【０００９】
図１は、本実施の形態の情報処理装置１０の構成の一例を示す模式図である。情報処理装置１０は、処理回路１２と、キャッシュメモリ１６と、管理装置１８と、を備える。情報処理装置１０のメモリバスには、記憶部１４が接続されている。
【００１０】
処理回路１２とキャッシュメモリ１６、処理回路１２と管理装置１８、および、キャッシュメモリ１６と管理装置１８、の各々は、データや信号を授受可能に接続されている。処理回路１２および管理装置１８と記憶部１４とは、データや信号を授受可能に接続されている。

【００１１】
処理回路１２は、１または複数のプロセッサを有する。プロセッサは、例えば、ＣＰＵ（Ｃｅｎｔｒａｌ　Ｐｒｏｃｅｓｓｉｎｇ　Ｕｎｉｔ）である。プロセッサは、１または複数のＣＰＵコアを含んでいてもよい。処理回路１２は、１または複数のアプリケーションプログラムの実行に応じて、管理装置１８を介して、記憶部１４からのデータ読出しや、記憶部１４へのデータ書込みを行う。
【００１２】
なお、以下では、アプリケーションプログラムを、単に、アプリケーション、と称して説明する場合がある。また、記憶部１４からのデータ読出および記憶部１４へのデータ書込みを総称して説明する場合には、記憶部１４へのアクセス、と称して説明する場合がある。
【００１３】
処理回路１２および管理装置１８は、記憶部１４に記憶されているデータをキャッシュメモリ１６に一時的に記憶し、処理に用いる。
【００１４】
記憶部１４は、処理回路１２による作業領域として用いられるメインメモリである。本実施の形態の情報処理装置１０は、複数種類の記憶部１４を備える。すなわち、本実施の形態の情報処理装置１０は、複数種類の記憶部１４を、メインメモリとして用いる。
【００１５】
複数種類の記憶部１４は、処理回路１２によるアクセス速度が互いに異なる。なお、以下では、処理回路１２によるアクセス速度を、単にアクセス速度と称して説明する場合がある。また、アクセス速度は、アクセス遅延とも呼ばれることもある。アクセス速度が速いというのは、アクセス遅延時間が短いことである。
【００１６】
本実施の形態では、情報処理装置１０は、アクセス速度の異なる複数種類の記憶部１４として、第１記憶部１４Ａと、第２記憶部１４Ｂと、を備える。なお、情報処理装置１０は、３種類以上の記憶部１４を備えた構成であってもよい。
【００１７】
第１記憶部１４Ａは、第２記憶部１４Ｂに比べてアクセス速度が速い。また、本実施の形態では、第１記憶部１４Ａは、第２記憶部１４Ｂより集積度が低い。
【００１８】
第１記憶部１４Ａは、例えば、揮発性メモリである。具体的には、第１記憶部１４Ａは、ＤＲＡＭ（Ｄｙｎａｍｉｃ　Ｒａｎｄｏｍ　Ａｃｃｅｓ　Ｍｅｍｏｒｙ）である。なお、第１記憶部１４Ａは、ＤＲＡＭと同程度に高速アクセスが可能な、ＭＲＡＭ（Ｍａｇｎｅｔｏｒｅｓｉｓｔｉｖｅ　Ｒａｎｄｏｍ　Ａｃｃｅｓ　Ｍｅｍｏｒｙ）等の不揮発メモリであってもよい。
【００１９】
一方、第２記憶部１４Ｂは、第１記憶部１４Ａに比べてアクセス速度が遅い。また、本実施の形態では、第２記憶部１４Ｂは、第１記憶部１４Ａより集積度が高い。すなわち、第２記憶部１４Ｂは、第１記憶部１４Ａより容量が大きい。
【００２０】
第２記憶部１４Ｂは、例えば、不揮発性メモリである。具体的には、第２記憶部１４Ｂは、ＤＲＡＭより大容量な大容量高速不揮発メモリ（Ｎｏｎ－ｖｏｌａｔｉｌｅ　Ｍｅｍｏｒｙ）である。
【００２１】
更に具体的には、第２記憶部１４Ｂは、ＭＲＡＭ、ＰＣＭ（Ｐｈａｓｅ　Ｃｈａｎｇｅ　Ｍｅｍｏｒｙ）、ＰＲＡＭ（Ｐｈａｓｅ　Ｒａｎｄｏｍ　Ａｃｃｅｓｓ　Ｍｅｍｏｒｙ）、ＰＣＲＡＭ（Ｐｈａｓｅ　Ｃｈａｎｇｅ　Ｒａｎｄｏｍ　Ａｃｃｅｓｓ　Ｍｅｍｏｒｙ）、ＲｅＲＡＭ（Ｒｅｓｉｓｔａｎｃｅ　Ｃｈａｎｇｅ　Ｒａｎｄｏｍ　Ａｃｃｅｓｓ　Ｍｅｍｏｒｙ）、ＦｅＲＡＭ（Ｆｅｒｒｏｅｌｅｃｔｒｉｃ　Ｒａｎｄｏｍ　Ａｃｃｅ

ｓｓ Ｍｅｍｏｒｙ）、３ＤＸＰｏｉｎｔまたはＭｅｍｒｉｓｔｏｒなどである。

【００２２】

また、第２記憶部１４Ｂは、いわゆるストレージクラスメモリ（ＳＣＭ）と呼ばれるメモリであってもよい。また、第２記憶部１４Ｂは、複数の半導体装置を１つの基板または筐体等に設けたモジュールであってもよい。

【００２３】

本実施の形態では、第１記憶部１４ＡがＤＲＡＭであり、第２記憶部１４ＢがＳＣＭである場合を、一例として説明する。なお、第１記憶部１４Ａのアクセス速度が第２記憶部１４Ｂより速ければよく、これらの組合せは、第１記憶部１４ＡがＤＲＡＭであり第２記憶部１４ＢがＳＣＭである形態に限定されない。例えば、第１記憶部１４ＡがＭＲＡＭであり、第２記憶部１４ＢがＲｅＲＡＭであってもよい。

【００２４】

なお、第１記憶部１４Ａおよび第２記憶部１４Ｂを総称して説明する場合には、単に、記憶部１４と称して説明する。

【００２５】

記憶部１４は、複数の第１領域を含む。第１領域は、複数の第２領域を含む。言い換えると、本実施の形態では、処理回路１２および管理装置１８が、第１記憶部１４Ａおよび第２記憶部１４Ｂを、第１領域毎に管理すると共に、第１領域内の第２領域ごとに管理する。

【００２６】

図２は、処理回路１２から見た物理アドレス空間を示す模式図である。

【００２７】

図２に示すように、第１記憶部１４Ａおよび第２記憶部１４Ｂは、それぞれ、複数の第１領域を含む。

【００２８】

第１領域は、例えば、処理回路１２によるデータの管理単位、あるいは、処理回路１２上で動作するオペレーティングシステムによるデータの管理単位（例えば、ページ）である。ページは例えば４ＫＢなどである。言い換えると、第１領域は、第１記憶部１４Ａと第２記憶部１４Ｂとの間でデータを転送するときの、転送単位である。なお、第１領域は、処理回路１２によるデータの管理単位の所定数倍の単位などであってもよい。本実施の形態では、第１領域が、ページに相当する場合を、一例として説明する。

【００２９】

第２領域は、第１領域より小さい領域である。例えば、第２領域は、処理回路１２による記憶部１４に対するアクセスに伴うデータの書き換え単位である。すなわち、第２領域は、処理回路１２によるデータのアクセス単位である。具体的には、第２領域は、キャッシュラインと呼ばれる単位である。キャッシュラインは、キャッシュメモリ１６に対するデータの書き換え単位に対応する。すなわち、処理回路１２からのメモリアクセス要求を受けた管理装置１８は、キャッシュラインの単位で、第１記憶部１４Ａまたは第２記憶部１４Ｂにアクセスする。

【００３０】

キャッシュラインは、例えば、６４バイトである。なお、第２領域は、キャッシュラインよりも小さい単位（例えば、バイト単位）であってもよい。また、第２領域は、キャッシュラインのサイズの所定数倍の単位などであってもよい。

【００３１】

本実施の形態では、処理回路１２および管理装置１８は、図２に示す物理アドレス空間１５にマッピングされた第１記憶部１４Ａおよび第２記憶部１４Ｂ内の領域を、第１領域のサイズ（例えば、ページサイズ）に区切って管理する。そして、処理回路１２および管理装置１８は、ページテーブルを用いて論理アドレスから物理アドレスに変換することで、仮想記憶を実現する。

【００３２】

図１に戻り、説明を続ける。管理装置１８は、処理回路１２による、複数種類の記憶部１４（第１記憶部１４Ａ、第２記憶部１４Ｂ）に対するアクセスを管理する。管理装置１８は、メモリ管理ユニット（ＭＭＵ：Ｍｅｍｏｒｙ　Ｍａｎａｇｅｍｅｎｔ　Ｕｎｉｔ）などと称される場合がある。管理装置１８はメモリコントローラなどであってもよい。

【００３３】

管理装置１８は、処理回路１２から受付けたメモリアクセス要求を処理する。メモリアクセス要求は、処理回路１２から記憶部１４に対するアクセス要求である。メモリアクセス要求は、記憶部１４へのデータ書込み、または記憶部１４からのデータ読出しを示す。メモリアクセス要求は、アクセス対象の記憶部１４の第１領域のアドレス情報および第２領域のアドレス情報を含む。これらのアドレス情報は、論理アドレスによって表される。

【００３４】

管理装置１８は、処理回路１２から受付けたメモリアクセス要求によって示されるアクセス対象のデータがキャッシュメモリ１６に格納されていない場合、記憶部１４へアクセスする。この場合、管理装置１８は、処理回路１２から受付けたメモリアクセス要求によって示される、アクセス対象の記憶部１４における第１領域内の第２領域にアクセスする。そして、管理装置１８は、アクセスした該第２領域に対して、メモリアクセス要求によって示される処理（書込みや読出し）を実行する。

【００３５】

具体的には、処理回路１２から受付けたメモリアクセス要求が特定の第２領域への書込みを示す場合がある。この場合、管理装置１８は、メモリアクセス要求に示される、アクセス対象の記憶部１４における、アクセス対象の第１領域内の第２領域に、メモリアクセス要求に示されるデータを書込む。また、処理回路１２から受付けたメモリアクセス要求が特定の第２領域からのデータ読出しを示す場合がある。この場合、管理装置１８は、メモリアクセス要求に示される、アクセス対象の記憶部１４における、アクセス対象の第１領域内の第２領域からデータを読出し、キャッシュメモリ１６に格納するとともに、処理回路１２へ出力する。

【００３６】

次に、処理回路１２について詳細を説明する。上述したように、処理回路１２は、１または複数のアプリケーションの実行に応じて、管理装置１８を介して、記憶部１４へのアクセスを行う。

【００３７】

図３は、処理回路１２の機能ブロック図の一例である。処理回路１２は、取得部１２Ａと、学習部１２Ｂと、導出部１２Ｃと、決定部１２Ｄと、実行部１２Ｅと、変更部１２Ｆと、を備える。

【００３８】

取得部１２Ａ、学習部１２Ｂ、導出部１２Ｃ、決定部１２Ｄ、実行部１２Ｅ、および変更部１２Ｆの少なくとも１つは、ＣＰＵなどのプロセッサにプログラムを実行させること、すなわちソフトウェアにより実現してもよい。また、取得部１２Ａ、学習部１２Ｂ、導出部１２Ｃ、決定部１２Ｄ、実行部１２Ｅ、および変更部１２Ｆの少なくとも１つは、専用のＩＣ（Ｉｎｔｅｇｒａｔｅｄ　Ｃｉｒｃｕｉｔ）などのハードウェアにより実現してもよい。また、取得部１２Ａ、学習部１２Ｂ、導出部１２Ｃ、決定部１２Ｄ、実行部１２Ｅ、および変更部１２Ｆの少なくとも１つは、ソフトウェアおよびハードウェアを併用して実現してもよい。複数のプロセッサを用いる場合、各プロセッサは、これらの取得部１２Ａ、学習部１２Ｂ、導出部１２Ｃ、決定部１２Ｄ、実行部１２Ｅ、および変更部１２Ｆのうち１つを実現してもよいし、各部のうち２以上を実現してもよい。

【００３９】

取得部１２Ａは、処理回路１２の動作統計情報を取得する。

【００４０】

動作統計情報とは、処理回路１２の動作に関する情報の統計値である。詳細には、動作統計情報は、処理回路１２が１または複数のアプリケーションを実行時の、動作に関する

情報の統計値である。動作に関する情報の統計値とは、単位期間Ｔあたりの、動作に関する情報を示す。単位期間Ｔは、予め設定すればよい。動作統計情報は、管理装置１８やキャッシュメモリ１６や情報処理装置１０の動作に関する情報の統計値であってもよい。動作統計情報は、例えば、プロセッサが備える、ハードウェアイベントを測定する性能カウンタにより収集される。動作統計情報は、例えば、ＯＳが管理する、情報処理装置の状態やＯＳ内部の状態を示す情報（例えば、ＯＳ内部のイベント発生回数の統計情報）全般であってもよい。

【００４１】

具体的には、動作統計情報は、性能カウンタなどで収集される、単位期間Ｔあたりの、ＴＬＢ（Ｔｒａｎｓｌａｔｉｏｎ Ｌｏｏｋａｓｉｄｅ Ｂｕｆｆｅｒ）ミスの回数、ＴＬＢミスに関する動作統計情報、キャッシュメモリの各階層（Ｌ１キャッシュ、Ｌ２キャッシュ、Ｌ３キャッシュ、ＬＬＣ（Ｌａｓｔ Ｌｅｖｅｌ Ｃａｃｈｅ）など）のキャッシュミスの回数、キャッシュミスのミスに関する動作統計情報、記憶部１４への書込回数、記憶部１４からの読出回数、ＳＴＬＢ（Ｓｅｃｏｎｄａｒｙ ｌｅｖｅｌ ＴＬＢ）ミスの回数、ＳＴＬＢミスに関する動作統計情報のなどのハードウェアイベントのうち、１または複数によって表される。動作統計情報には、さらに、ＯＳが管理する、当該実行期間に実行されているアプリケーションに当該実行期間中に割当てられている物理メモリサイズ（つまりアプリケーションの実行中のいずれかのタイミングでアクセスされる可能性があるメモリのサイズ）なども含まれていてもよい。なお、動作統計情報は、これらに限定されない。

【００４２】

取得部１２Ａは、公知の方法で、処理回路１２の動作統計情報を取得すればよい。例えば、取得部１２Ａは、処理回路１２に設けられた性能カウンタから、単位期間Ｔごとの動作統計情報を順次取得すればよい。性能カウンタは、例えば、Ｉｎｔｅｌプロセッサのパフォーマンスモニタリングカウンタ（Ｐｅｒｆｏｒｍａｎｃｅ Ｍｏｎｉｔｏｒｉｎｇ Ｃｏｕｎｔｅｒ）などであるが、これに限定されない。取得部１２Ａは、性能カウンタと一体的に構成してもよい。また、取得部１２Ａと性能カウンタとを、別体として構成してもよい。本実施の形態では、取得部１２Ａと性能カウンタとを、別体として構成する形態を一例として説明する。

【００４３】

学習部１２Ｂは、教師データを複数含む教師データセットを用いて、予測モデル２０を学習する。

【００４４】

予測モデル２０は、動作統計情報からメモリアクセス特性を導出するためのモデルである。予測モデル２０は、学習によって生成される、学習モデルである。

【００４５】

メモリアクセス特性とは、第１記憶部１４Ａおよび第２記憶部１４Ｂに対する、処理回路１２によるアクセスの特性を示す。

【００４６】

例えば、メモリアクセス特性は、処理回路１２がアプリケーション実行中に単位期間Ｔあたりに使用したメモリサイズによって表される。詳細には、メモリアクセス特性は、例えば、処理回路１２がアプリケーション実行中に単位期間Ｔあたりに記憶部１４やキャッシュメモリ１６上のデータに対してロード命令とストア命令を発行した場合、そのデータの合計サイズである。具体的には、処理回路１２がアプリケーション実行中において、単位期間ＴあたりＮページに対してアクセスしたと想定する（Ｎは１以上の整数）。この場合、１ページの容量が４Ｋバイトとした場合、メモリアクセス特性は、"Ｎ"に、１ページの容量である"４Ｋバイト"を乗算した結果（Ｎ×４Ｋ）によって表される。これは、一般に、ワーキングセットサイズなどとも呼ばれることもある。なお、メモリサイズは、処理回路１２が記憶部１４あるいは記憶部１４のデータをキャッシュするキャッシュメモリ１６に対してアクセスしたデータを格納するページのページ数（第１領域の数）で表し

てもよい。

【００４７】

学習部１２Ｂは、動作統計情報を入力としメモリアクセス特性を出力とする予測モデル２０を、教師データセットを用いて学習する。

【００４８】

図４は、予測モデル２０の学習の説明図である。教師データセット４０は、複数の教師データ４２を含む。教師データ４２は、単位期間Ｔごとに生成される。教師データ４２は、動作統計情報４２Ａと、メモリアクセス特性４２Ｂと、を含む、なお、１つの教師データ４２には、１または複数の動作統計情報４２Ａが含まれる。また、１つの教師データ４２には、１または複数の動作統計情報４２Ａに対応する１つの正解情報として、１つのメモリアクセス特性４２Ｂが含まれる。

【００４９】

処理回路１２は、予め、教師データセット４０を用意する。例えば、処理回路１２は、１または複数の学習用のアプリケーション３０（例えば、アプリケーション３０Ａ、アプリケーション３０Ｂ、アプリケーション３０Ｃ）を実行する。そして、処理回路１２は、アプリケーション３０を実行中における１または複数の動作統計情報４２Ａとメモリアクセス特性４２Ｂとの組からなる教師データ４２を、単位期間Ｔごとに生成する。この処理により、処理回路１２は、予め、複数の教師データ４２を含む教師データセット４０を用意する。

【００５０】

図５は、学習用のアプリケーション３０の実行期間ＴＡにおける、動作統計情報４２Ａとメモリアクセス特性４２Ｂとの関係の一例を示す説明図である。図５に示すグラフの縦軸は、動作統計情報またはメモリアクセス特性を示す。図５に示すグラフの横軸は、時間を示す。なお、図５には、動作統計情報４２Ａが実行中のアプリケーションに割当てられた物理メモリサイズを示し、メモリアクセス特性４２Ｂが、処理回路１２がアプリケーション実行中に単位期間Ｔあたりに使用したメモリサイズを示す場合を、一例として示した。

【００５１】

処理回路１２が、学習用のアプリケーション３０を実行したときの、動作統計情報４２Ａおよびメモリアクセス特性４２Ｂの時間経過に伴う推移が、図５に示す推移を示したと想定する。この場合、学習部１２Ｂは、単位期間Ｔごとの、動作統計情報４２Ａとメモリアクセス特性４２Ｂとの対応を、単位期間Ｔごとの教師データ４２として生成すればよい。なお、隣接するタイミングの教師データ４２の単位期間Ｔは、一部が重複するタイミングであってもよく、また、互いに非重複のタイミングであってもよい。

【００５２】

学習部１２Ｂは、性能カウンタから取得部１２Ａを介して、単位期間Ｔごとの動作統計情報４２Ａを取得することで、単位期間Ｔの教師データ４２の生成に用いればよい。

【００５３】

また、学習部１２Ｂは、教師データ４２に用いる単位期間Ｔごとのメモリアクセス特性４２Ｂを、以下の方法により取得すればよい。

【００５４】

詳細には、学習部１２Ｂは、単位期間Ｔの最初のタイミング（例えば、ｔ１）で、処理回路１２に予めインストールされたＯＳ（オペレーティングシステム）が管理するページテーブルの一部をリセットする。詳細には、学習部１２Ｂは、ページテーブルの全てのページのアクセス済フラグを"０"として"アクセス未"とすることで、ページテーブルをリセットする。次に、学習部１２Ｂは、該単位期間Ｔの区間の終了タイミング（例えば、ｔ２）に、該ページテーブルにおけるアクセス済フラグ（"１"となっているフラグ）を計数する。この計数処理により、学習部１２Ｂは、単位期間Ｔに処理回路１２によってアクセスされたページ数を求める。そして、学習部１２Ｂは、該ページ数（Ｎ）に、１ページの容量である"４Ｋバイト"を乗算した結果（Ｎ×４Ｋ）を、該単位期間Ｔのメモリア

クセス特性４２Ｂとして取得する。

【００５５】

そして、学習部１２Ｂは、単位期間Ｔごとに、上記処理を実行する。そして、学習部１２Ｂは、単位期間Ｔごとに取得した、動作統計情報４２Ａとメモリアクセス特性４２Ｂとの対応を示す教師データ４２を生成する。

【００５６】

ここで、上述したように、学習部１２Ｂは、性能カウンタから取得部１２Ａを介して動作統計情報４２Ａを取得する。このため、学習部１２Ｂは、学習用のアプリケーション３０の実行中に、リアルタイムで動作統計情報４２Ａを取得可能である。一方、メモリアクセス特性４２Ｂの取得には、学習部１２Ｂは、ページテーブルのリセット、アクセス済フラグの計数、メモリアクセス特性４２Ｂの計算、といった処理を、単位期間Ｔごとに実行する必要がある。このため、学習部１２Ｂは、学習用のアプリケーション３０の実行中に、リアルタイムでメモリアクセス特性４２Ｂを取得することは困難な場合がある。また、ページテーブルのリセット、アクセス済フラグの計数、メモリアクセス特性４２Ｂの計算、といった処理自体が、動作統計情報４２Ａに影響を与えてしまう（アプリケーションのみを実行した場合の動作統計情報４２Ａに対して大きく変化してしまう）可能性もあるため、それを回避するのが望ましい。

【００５７】

そこで、本実施の形態では、処理回路１２は、学習用のアプリケーション３０を２回実行する。そして、学習部１２Ｂは、１回目および２回目の内の一方の回のアプリケーション３０の実行時に動作統計情報４２Ａを取得し、他方の回のアプリケーション３０の実行時にメモリアクセス特性４２Ｂを取得する。そして、学習部１２Ｂは、単位期間Ｔに対応する、アプリケーション３０の命令単位ごとに、動作統計情報４２Ａとメモリアクセス特性４２Ｂとの対応を示す教師データ４２を生成する。

【００５８】

図６Ａは、学習用のアプリケーション３０を実行したときの、動作統計情報４２Ａの一例を示す模式図である。図６Ａには、動作統計情報４２Ａとして、ＴＬＢミスに関する動作統計情報４２Ａ１と、実行中のアプリケーションに割当てられた物理メモリサイズ４２Ａ２と、を一例として示した。なお、図６Ａに示すグラフの縦軸は、動作統計情報を示し、横軸は、時間を示す。

【００５９】

図６Ａに示すように、処理回路１２が、学習用のアプリケーション３０を実行したときの、時間経過に伴う動作統計情報４２Ａの推移が、図６Ａに示す推移を示したと想定する。そして、学習部１２Ｂは、単位期間Ｔを、該アプリケーション３０が命令単位Ｓを実行する期間と定める。命令単位Ｓは、例えば、１０万回である。なお、命令単位Ｓの命令回数は、予め設定すればよく、１０万回に限定されない。

【００６０】

この場合、学習部１２Ｂは、アプリケーション３０の命令単位Ｓごとの動作統計情報４２Ａを、性能カウンタから取得することで、命令単位Ｓごと（すなわち、単位期間Ｔごと）の動作統計情報４２Ａを取得する。例えば、学習用のアプリケーション３０を実行した期間ＴＡを、過去から未来に向かって、命令単位Ｓごとに、フェーズＰ１、フェーズＰ２、フェーズＰ３、フェーズＰ４、フェーズＰ５に分割して管理すると想定する。この場合、学習部１２Ｂは、各フェーズ（フェーズＰ１～フェーズＰ５）の各々ごとの、動作統計情報４２Ａを取得する。

【００６１】

次に、学習部１２Ｂは、同じ学習用のアプリケーション３０を再度実行し、命令単位Ｓごとのメモリアクセス特性４２Ｂを取得する。

【００６２】

図６Ｂは、教師データ４２用の動作統計情報４２Ａの取得時と同じ学習用のアプリケーション３０を実行したときの、メモリアクセス特性４２Ｂの一例を示す模式図である。図

６Ｂには、メモリアクセス特性４２Ｂとして、処理回路１２が単位期間Ｔあたりに使用したメモリサイズを一例として示した。なお、図６Ｂに示すグラフの縦軸は、メモリアクセス特性を示し、横軸は、時間を示す。

【００６３】

学習部１２Ｂは、単位期間Ｔに相当する命令単位Ｓごとに、ページテーブルのリセット、アクセス済フラグの計数、メモリアクセス特性４２Ｂの計算、を実行する。この処理により、学習部１２Ｂは、命令単位Ｓごとの、メモリアクセス特性４２Ｂを取得する。このため、学習部１２Ｂは、フェーズ（フェーズＰ１～フェーズＰ５）の各々の命令単位Ｓごとの、メモリアクセス特性４２Ｂを取得する。

【００６４】

そして、学習部１２Ｂは、学習用のアプリケーション３０の命令単位Ｓごとの、動作統計情報４２Ａとメモリアクセス特性４２Ｂとの対応を示す、教師データ４２を生成すればよい。

【００６５】

ここで、学習部１２Ｂが、命令単位Ｓごとに、ページテーブルのリセット、アクセス済フラグの計数、メモリアクセス特性４２Ｂの計算、を実行するため、これらの実行に要する時間は、単位期間Ｔより長い期間Ｔ'となる場合がある。

【００６６】

しかし、本実施の形態では、学習部１２Ｂが、動作統計情報４２Ａをリアルタイムで取得するときの単位期間Ｔに相当する命令単位Ｓを基準として、該命令単位Ｓごとのメモリアクセス特性４２Ｂを取得する。このため、学習部１２Ｂは、処理回路１２が実際にアプリケーション３０を実行したときの、単位期間Ｔごとの動作統計情報４２Ａとメモリアクセス特性４２Ｂとの対応を示す教師データ４２を、精度良く生成することができる。

【００６７】

図４に戻り説明を続ける。そして、学習部１２Ｂは、複数の教師データ４２を含む教師データセット４０を用いて、動作統計情報４２Ａからメモリアクセス特性４２Ｂを導出するための予測モデル２０を学習する。

【００６８】

学習部１２Ｂは、公知の学習アルゴリズムを用いて、予測モデル２０を学習すればよい。学習アルゴリズムは、例えば、線形回帰、ｋ近傍法（ＫＮＮ：Ｋ－Ｎｅａｒｅｓｔ　Ｎｅｉｇｈｂｏｒ　ａｌｇｏｒｉｔｈｍ）、サポートベクターマシン、Ｒａｎｄｏｍ　Ｆｏｒｅｓｔなどであるが、これらに限定されない。

【００６９】

学習部１２Ｂは、予め予め定めたタイミング毎に、教師データセット４０に含まれる複数の教師データ４２を用いて、予測モデル２０を学習すればよい。また、例えば、学習部１２Ｂは、新たな教師データ４２が教師データセット４０に登録されるごとに、予測モデル２０を学習してもよい。教師データ４２登録は、任意のタイミングで実行すればよい。

【００７０】

また、学習部１２Ｂは、新たな教師データ４２を含む教師データセット４０を用いて新たな予測モデル２０を学習した場合には、学習部１２Ｂに登録されている予測モデル２０を、新たに学習した予測モデル２０に更新すればよい。すなわち、学習部１２Ｂには、１つの予測モデル２０が格納された状態となる。

【００７１】

図３に戻り説明を続ける。次に、導出部１２Ｃについて説明する。

【００７２】

導出部１２Ｃは、学習部１２Ｂが学習した予測モデル２０に基づいて、取得部１２Ａで取得した動作統計情報４２Ａからメモリアクセス特性４２Ｂを導出する。導出部１２Ｃがメモリアクセス特性４２Ｂの導出に用いる動作統計情報４２Ａは、学習部１２Ｂによる予測モデル２０の学習時とは異なり、処理回路１２が学習用以外の実際のアプリケーション３０を実行したときの動作統計情報４２Ａである。導出部１２Ｃは、この動作統計情報４

２Ａと、予測モデル２０と、を用いて、メモリアクセス特性４２Ｂを導出する。

【００７３】

図７は、導出部１２Ｃおよび決定部１２Ｄによる処理の一例の説明図である。

【００７４】

例えば、導出部１２Ｃは、最適化対象のアプリケーション３２を処理回路１２が実行しているときの動作統計情報４２Ａを、取得部１２Ａから取得する。そして、導出部１２Ｃは、取得した動作統計情報４２Ａを予測モデル２０へ入力することで、予測モデル２０からの出力として、メモリアクセス特性４２Ｂを得る。

【００７５】

すなわち、導出部１２Ｃは、予測モデル２０を用いて、取得部１２Ａで取得した動作統計情報４２Ａに対する、メモリアクセス特性４２Ｂの予測値を得る。

【００７６】

決定部１２Ｄは、導出部１２Ｃが導出したメモリアクセス特性４２Ｂに基づいて、アクセス方式を決定する。

【００７７】

アクセス方式は、処理回路１２による記憶部１４に対するアクセス方式を示す。本実施の形態では、アクセス方式は、第１アクセス方式、または、第２アクセス方式を示す。

【００７８】

第１アクセス方式は、第２記憶部１４Ｂのデータを第１記憶部１４Ａへ転送し、該第１記憶部１４Ａ内の該データにアクセスする、アクセス方式である。転送およびアクセスするデータは、該アクセス方式の決定に用いた動作統計情報４２Ａの取得時に処理回路１２がアクセスしていたデータである。言い換えると、転送およびアクセスするデータは、該アクセス方式の決定に用いた動作統計情報４２Ａによって示される動作実行時に処理回路１２がアクセスしていたデータである。

【００７９】

本実施の形態では、転送とは、コピーを意味する。上述したように、処理回路１２は、ページ単位（第１領域の単位）でデータ転送を行う。また、処理回路１２は、キャッシュライン（第２領域）の単位で、記憶部１４へアクセスする。

【００８０】

このため、第１アクセス方式は、該アクセス方式の決定に用いた動作統計情報４２Ａによって示される動作実行時に処理回路１２がアクセスしていたデータを含むページ（第１領域）内のデータを、第２記憶部１４Ｂから第１記憶部１４Ａへ転送し、転送後の第１記憶部１４Ａの該データへアクセスすることを示す。

【００８１】

第２アクセス方式は、第２記憶部１４Ｂ内のデータにアクセスする方式を示す。本実施の形態では、処理回路１２は、原則、第２記憶部１４Ｂ内のデータにアクセスする。そして、処理回路１２は、特定の条件を満たした場合にのみ、第２記憶部１４Ｂから第１記憶部１４Ａへデータを転送し、第１記憶部１４Ａ内のデータへアクセスする。このため、第２アクセス方式は、データを第２記憶部１４Ｂに配置したまま、該第２記憶部１４Ｂにダイレクトにアクセスすることを示す。

【００８２】

例えば、決定部１２Ｄは、導出部１２Ｃが導出したメモリアクセス特性４２Ｂが第１閾値より大きい場合、第２アクセス方式を決定する。また、決定部１２Ｄは、該メモリアクセス特性４２Ｂが第１閾値以下の場合、第１アクセス方式を決定する。

【００８３】

図８は、アクセス方式の決定の説明図である。図８の横軸は時間を示し、縦軸は動作統計情報４２Ａを示す。

【００８４】

例えば、導出部１２Ｃが、アプリケーション３２の実行中に取得部１２Ａで取得した動作統計情報４２Ａを予測モデル２０へ入力することで、図８に示すメモリアクセス特性４

２Ｂを導出したと想定する。

【００８５】

なお、動作統計情報４２Ａが、処理回路１２で実行中のアプリケーションに割当てられた物理メモリサイズを示すと想定する。また、メモリアクセス特性４２Ｂが、アプリケーション３２実行中、単位期間Ｔあたりに処理回路１２（アプリケーション３２）が使用したメモリサイズを示すと想定する。

【００８６】

この場合、決定部１２Ｄは、導出したメモリアクセス特性４２Ｂが第１閾値より大きい場合、第２アクセス方式を決定する。また、決定部１２Ｄは、導出したメモリアクセス特性４２Ｂが第１閾値以下である場合、第１アクセス方式を決定する。

【００８７】

図８に示すように、決定部１２Ｄは、アプリケーション３２の実行期間ＴＡの内の前半の期間Ａのように、メモリアクセス特性４２Ｂが第１閾値以下の場合、第１アクセス方式を決定する。

【００８８】

すなわち、決定部１２Ｄは、メモリアクセス特性４２Ｂが第１閾値以下である状態を、処理回路１２によるメモリアクセスが集中し、アクセスのローカリティの高い状態であると推定する。そして、決定部１２Ｄは、メモリアクセス特性４２Ｂが第１閾値以下の場合、第１アクセス方式を決定する。

【００８９】

このため、決定部１２Ｄは、処理回路１２がアクセスのローカリティの高いデータにアクセスする場合、すなわち、記憶部１４におけるメモリアクセスされる場所が集中している場合、データを第２記憶部１４Ｂから第１記憶部１４Ａへ転送し、処理回路１２が第１記憶部１４Ａ上のデータをキャッシュライン単位でアクセスするように、アクセス方式を決定することができる。

【００９０】

一方、決定部１２Ｄは、アプリケーション３２の実行期間ＴＡの後半の期間Ｂのように、メモリアクセス特性４２Ｂが第１閾値を超える場合、第２アクセス方式を決定する。

【００９１】

図８に示すように、決定部１２Ｄは、アプリケーション３２の実行期間ＴＡの内の後半の期間Ｂのように、メモリアクセス特性４２Ｂが第１閾値を超える場合、第２アクセス方式を決定する。

【００９２】

すなわち、決定部１２Ｄは、メモリアクセス特性４２Ｂが第１閾値を超える状態を、処理回路１２によるメモリアクセスが分散し、アクセスのローカリティが低く、使用中のメモリサイズが大きい状態であると推定する。そして、決定部１２Ｄは、メモリアクセス特性４２Ｂが第１閾値を超える場合、第２アクセス方式を決定する。

【００９３】

メモリアクセス特性４２Ｂが第１閾値を超える場合、第２アクセス方式を決定することで、決定部１２Ｄは、処理回路１２によるメモリアクセスの高速化を図ることができる。

【００９４】

ここで、メモリアクセス特性４２Ｂが第１閾値を超える場合には、第２記憶部１４Ｂのデータを第１記憶部１４Ａへ転送しても、第１記憶部１４Ａの空き容量不足により、すぐに第１記憶部１４Ａから第２記憶部１４Ｂへデータが転送されてしまう。すなわち、メモリアクセス特性４２Ｂが第１閾値を超える場合に、第２記憶部１４Ｂのデータを第１記憶部１４Ａへ転送すると、第１記憶部１４Ａと第２記憶部１４Ｂとの間のページ単位のデータ入れ替えが頻発してしまい、処理回路１２の性能低下を引き起こす可能性がある。

【００９５】

そこで、本実施の形態では、決定部１２Ｄは、メモリアクセス特性４２Ｂが第１閾値を超える場合には、第２アクセス方式を決定する。このため、決定部１２Ｄは、処理回路１

２によるメモリアクセスの高速化を図ることができる。
【００９６】
なお、第１閾値には、任意の値を予め定めればよい。例えば、第１閾値には、処理回路１２が利用可能な、第１記憶部１４Ａのサイズやこれに近い値であればよい。
【００９７】
処理回路１２が利用可能なサイズとは、具体的には、処理回路１２が実行するアプリケーション３２で利用可能な、第１記憶部１４Ａのサイズや、情報処理装置１０で利用可能な記憶部１４のサイズなどである。なお、第１閾値は、これらの利用可能なサイズに対して所定の割合大きい値であってもよい。また、第１閾値は、これらの利用可能なサイズに対して所定の割合小さい値であってもよい。
【００９８】
なお、第１閾値を、処理回路１２が実行するアプリケーション３２で利用可能な第１記憶部１４Ａのサイズより大きい値、または、情報処理装置１０で利用可能な第１記憶部１４Ａのサイズより大きい値とすると、以下の効果が得られる。例えば、第１記憶部は高速なＳＣＭを想定しているため、第１アクセス方式が決定された場合における、第１記憶部１４Ａと第２記憶部１４Ｂとの間のデータの転送は、高速に行う事が可能である。つまり、第２記憶部１４Ｂから第１記憶部１４Ａへのデータ転送あるいは第１記憶部１４Ａから第２記憶部１４Ｂへのデータ転送を積極的に行って頻繁にデータを入れ替えても速度低下は僅かで済み高速に処理することができる。このため、利用可能な第１記憶部１４Ａを最大限活用するという観点では、第１閾値を、利用可能な第１記憶部１４Ａのサイズより大きい値とすることで、利用可能な第１記憶部１４Ａを最大限活用して、より大きなメモリサイズを利用するアプリケーションが実行可能になる。また、利用可能な第１記憶部１４Ａを削減するという観点では、第１閾値を、利用可能な第１記憶部１４Ａのサイズより大きい値とすることで、実際に利用可能な第１記憶部１４Ａのサイズより小さいサイズで、使用メモリサイズが小さく（つまり消費電力が低く）速度低下も抑えた効率の良い処理を実行することができる。
【００９９】
図３に戻り説明を続ける。次に、実行部１２Ｅについて説明する。実行部１２Ｅは、決定部１２Ｄで決定されたアクセス方式に応じて、データの第２記憶部１４Ｂから第１記憶部１４Ａへの転送および第１記憶部１４Ａ内の該データへのアクセス、または、第２記憶部１４Ｂ内のデータへのアクセス、を実行する。
【０１００】
すなわち、決定部１２Ｄが第１アクセス方式を決定した場合、実行部１２Ｅは、該アクセス方式の決定に用いた動作統計情報４２Ａによって示される動作実行時に処理回路１２がアクセスしていたデータを含むページ（第１領域）を、第２記憶部１４Ｂから第１記憶部１４Ａへ転送し、第１記憶部１４Ａにおける転送した該ページ内の該データへのアクセスを実行する。
【０１０１】
なお、実行部１２Ｅによる、第１記憶部１４Ａから第２記憶部１４Ｂへのデータ転送のタイミングは、限定されない。例えば、実行部１２Ｅは、決定部１２Ｄが第１アクセス方式を決定した直後、実行部１２Ｅが次回該データへアクセスする時、予め定めた条件を満たしたタイミング、の何れであってもよい。予め定めた条件を満たすタイミングは、例えば、処理回路１２による記憶部１４へのメモリアクセスが所定値以下の期間などである。
【０１０２】
一方、決定部１２Ｄが第２アクセス方式を決定した場合、実行部１２Ｅは、該アクセス方式の決定に用いた動作統計情報４２Ａによって示される動作実行時に処理回路１２がアクセスしていた、第２記憶部１４Ｂ内のデータへのアクセスを継続して実行する。
【０１０３】
次に、変更部１２Ｆについて説明する。決定部１２Ｄが第１アクセス方式を決定した場合、変更部１２Ｆは、第１記憶部１４Ａの、利用可能なメモリサイズを変更する。

【０１０４】
詳細には、変更部１２Ｆは、第１記憶部１４Ａの利用可能なメモリサイズを、第１アクセス方式の決定に用いたメモリアクセス特性４２Ｂによって示される、処理回路１２がアプリケーション３２の実行中に単位期間Ｔあたりに使用したメモリサイズに変更する。また、変更部１２Ｆは、第１記憶部１４Ａの利用可能なメモリサイズを、該使用した該メモリサイズより一定の割合大きいサイズ、または、該使用した該メモリサイズより一定の割合小さいサイズに変更する。そして、変更部１２Ｆは、利用可能なメモリサイズを変更した後の第１記憶部１４Ａにおける、利用不可能な領域を、パワーオフまたはセルフリフレッシュモードなどの低消費電力モードに設定すればよい。
【０１０５】
なお、処理回路１２は、変更部１２Ｆを備えない構成であってもよい。すなわち、処理回路１２は、第１記憶部１４Ａにおける利用可能なメモリサイズの変更を行わない形態であってもよい。
【０１０６】
次に、本実施の形態の処理回路１２が実行する情報処理の手順の一例を説明する。
【０１０７】
図９は、本実施の形態の処理回路１２が実行する情報処理の手順の一例を示す、フローチャートである。なお、図９に示す情報処理の手順の実行前に、学習部１２Ｂが、予測モデル２０を学習済であるものとして説明する。また、図９には、変更部１２Ｆが第１記憶部１４Ａの利用可能なメモリサイズの変更処理を実行しない形態を、一例として示した。
【０１０８】
まず、取得部１２Ａが、単位期間Ｔにおける、処理回路１２の動作統計情報４２Ａを取得する（ステップＳ１００）。
【０１０９】
次に、導出部１２Ｃが、学習部１２Ｂが学習した予測モデル２０に、ステップＳ１００で取得した動作統計情報４２Ａを入力することで、メモリアクセス特性４２Ｂを導出する（ステップＳ１０２）。
【０１１０】
次に、決定部１２Ｄが、ステップＳ１０２で導出したメモリアクセス特性４２Ｂが第１閾値より大きいか否かを判断する（ステップＳ１０４）。
【０１１１】
メモリアクセス特性４２Ｂが第１閾値より大きいと判断した場合（ステップＳ１０４：Ｙｅｓ）、ステップＳ１０６へ進む。
【０１１２】
ステップＳ１０６では、決定部１２Ｄは、第２アクセス方式を決定する（ステップＳ１０６）。次に、実行部１２Ｅは、ステップＳ１００で取得した動作統計情報４２Ａを示す処理を実行中の処理回路１２がアクセス中の、第２記憶部１４Ｂのデータを、該第２記憶部１４Ｂに配置したまま、該第２記憶部１４Ｂにアクセスする（ステップＳ１０８）。
【０１１３】
次に、処理回路１２は、情報処理を終了するか否かを判断する（ステップＳ１１０）。例えば、処理回路１２は、ステップＳ１００で取得した動作統計情報４２Ａを示す処理を実行中のアプリケーション３２の終了指示を受付けたか否かを判別することで、ステップＳ１１０の判断を行う。ステップＳ１１０で肯定判断すると（ステップＳ１１０：Ｙｅｓ）、本ルーチンを終了する。一方、ステップＳ１１０で否定判断すると（ステップＳ１１０：Ｎｏ）、上記ステップＳ１００へ戻る。
【０１１４】
一方、上記ステップＳ１０４で、ステップＳ１０２で導出したメモリアクセス特性４２Ｂが第１閾値以下であると判断すると（ステップＳ１０４：Ｎｏ）、ステップＳ１１２へ進む。
【０１１５】

ステップＳ１１２では、決定部１２Ｄは、第１アクセス方式を決定する（ステップＰＳ１１２）。次に、実行部１２Ｅは、ステップＳ１００で取得した動作統計情報４２Ａによって示される動作実行時に処理回路１２がアクセスしていた、第２記憶部１４Ｂの第１領域内（ページ内）のデータを、該第２記憶部１４Ｂから第１記憶部１４Ａへ転送する（ステップＳ１１４）。

【０１１６】

次に、実行部１２Ｅは、ページテーブルにおける、ステップＳ１１４で転送した第１領域の論理アドレスに対応する物理アドレスを、ステップＳ１１４で転送した転送先の第１記憶部１４Ａの格納先を示す物理アドレスに更新する（ステップＳ１１６）。このため、処理回路１２は、該データにアクセスする場合には、第１記憶部１４Ａにアクセスすることで、該データにアクセスすることが可能となる。

【０１１７】

そして、実行部１２Ｅは、ステップＳ１１４で第１記憶部１４Ａへ転送されたデータにアクセスする（ステップＳ１１８）。そして、上記ステップＳ１１０へ進む。

【０１１８】

以上説明したように、本実施の形態の情報処理装置１０は、取得部１２Ａと、導出部１２Ｃと、決定部１２Ｄと、を備える。取得部１２Ａは、処理回路１２の動作統計情報４２Ａを取得する。導出部１２Ｃは、動作統計情報４２Ａから処理回路１２のメモリアクセス特性４２Ｂを導出するための予測モデル２０に基づいて、取得した動作統計情報４２Ａからメモリアクセス特性４２Ｂを導出する。決定部１２Ｄは、導出したメモリアクセス特性４２Ｂに基づいて、第１記憶部１４Ａより処理回路１２によるアクセス速度が遅い第２記憶部１４Ｂのデータを第１記憶部１４Ａへ転送し、第１記憶部１４Ａ内の該データにアクセスする第１アクセス方式、または、第２記憶部１４Ｂ内のデータにアクセスする第２アクセス方式、の何れかのアクセス方式を決定する。

【０１１９】

このように、本実施の形態の情報処理装置１０は、予測モデル２０に基づいて、第１アクセス方式または第１アクセス方式の何れかのアクセス方式を決定する。

【０１２０】

ここで、従来では、複数種類のメモリ（記憶部１４）の使い分けに用いる情報を効率よく提供することは困難であった。

【０１２１】

ＳＣＭなどの第２記憶部１４Ｂは、ＤＲＡＭなどの第１記憶部１４Ａより大容量であるがアクセス速度が遅い。このため、処理対象のデータの特性に合わせて、データを第１記憶部１４Ａと第２記憶部１４Ｂとに分散して格納してアクセスすると、処理回路１２は効率よくデータ処理を実行することができる。すなわち、処理回路１２によるアクセスのローカリティが低く、メモリアクセスされる場所が広域に分散されることでアクセスするデータサイズが大きい場合には、データを第２記憶部１４Ｂに配置したままとし、処理回路１２が第２記憶部１４Ｂにダイレクトにアクセスすることが好ましい。また、メモリアクセスされる場所が集中することで、アクセスのローカリティの高いデータに処理回路１２がアクセスする場合には、データを第２記憶部１４Ｂから第１記憶部１４Ａへ転送（コピー）し、処理回路１２は第１記憶部１４Ａ上のデータをキャッシュライン単位でアクセスすることが好ましい。

【０１２２】

しかし、従来では、複数種類の記憶部１４の使い分けに用いる情報、すなわち、処理回路１２のメモリアクセス特性４２Ｂを、効率よく得ることが困難であった。

【０１２３】

一方、本実施の形態の情報処理装置１０では、予測モデル２０に基づいて、取得した動作統計情報４２Ａからメモリアクセス特性４２Ｂを導出し、第１アクセス方式または第１アクセス方式の何れかのアクセス方式を決定する。

【０１２４】

従って、本実施の形態の情報処理装置１０は、複数種類のメモリの使い分けに用いる情報を効率よく提供することができる。

【０１２５】

（変形例１）

なお、上記実施の形態では、情報処理の手順の説明時に、変更部１２Ｆが変更処理を実行しない形態を、一例として示した。

【０１２６】

しかし、情報処理の手順の実行時に、変更部１２Ｆによる変更処理を実行してもよい。

【０１２７】

この場合、例えば、図９に示すステップＳ１１２の第アクセス方式を決定した後に、変更部１２Ｆが、第１記憶部１４Ａの利用可能なメモリサイズを変更する変更処理を実行すればよい。そして、その次に、上記ステップＳ１１４～ステップＳ１１８の処理を実行すればよい。

【０１２８】

なお、変更部１２Ｆによる変更処理のタイミングは、このタイミングに限定されない。例えば、変更部１２Ｆは、第１アクセス方式が決定され、第２記憶部１４Ｂの第１領域内のデータが第１記憶部１４Ａへ転送された後に、変更処理を実行してもよい。また、変更部１２Ｆは、第１アクセス方式が決定され、第２記憶部１４Ｂの第１領域内のデータが第１記憶部１４Ａへ転送され、更に、実行部１２Ｅが第１記憶部１４Ａのデータにアクセスした後に、変更処理を実行してもよい。

【０１２９】

（変形例２）

なお、上記実施の形態では、決定部１２Ｄは、導出部１２Ｃが導出したメモリアクセス特性４２Ｂが第１閾値より大きい場合、第２アクセス方式を決定する形態を説明した。また、決定部１２Ｄは、該メモリアクセス特性４２Ｂが第１閾値以下の場合、第１アクセス方式を決定する形態を説明した。

【０１３０】

しかし、決定部１２Ｄは、他の方法により、第１アクセス方式または第２アクセス方式を決定してもよい。

【０１３１】

例えば、決定部１２Ｄは、メモリアクセス特性４２Ｂの比率が、第２閾値より大きい場合、第２アクセス方式を決定する。メモリアクセス特性４２Ｂの比率とは、取得部１２Ａが取得した動作統計情報４２Ａによって示される動作実行時に実行中の、１または複数のアプリケーション３２の各々に割当てられた、物理メモリサイズの合計値に対する、メモリアクセス特性４２Ｂの比率（割合）を示す。

【０１３２】

具体的には、決定部１２Ｄは、実行中のアプリケーション３２の各々に割当てられた物理メモリサイズの合計値に対する、導出部１２Ｃで導出されたメモリアクセス特性４２Ｂとしての使用中のメモリサイズの比率が、第２閾値以下の場合、第１アクセス方式を決定する。

【０１３３】

該比率が第２閾値以下の状態とは、アプリケーション３２で利用する可能性のあるメモリの一部の領域に対して、メモリアクセスが集中している状態を示す。このため、この場合、決定部１２Ｄは、処理回路１２がアクセスのローカリティの高いデータにアクセスしていると判断し、第２記憶部１４Ｂ上のデータをページ単位（第１領域単位）で第１記憶部１４Ａへ転送し、アクセスする第１アクセス方式を決定する。

【０１３４】

一方、決定部１２Ｄは、該比率が第２閾値を超える場合、第２アクセス方式を決定する。

【０１３５】

該比率が第２閾値を超える状態とは、アプリケーション３２で利用する可能性のあるメモリ領域全体に対する処理回路１２によるアクセスが、該メモリ領域全体に分散されている状況であることを示す。このため、該比率が第２閾値を超える状態の場合、アクセスのローカリティが低く、使用中のメモリサイズが大きい状態である。このため、この場合、決定部１２Ｄは、第２アクセス方式を決定する。

【０１３６】

なお、第２閾値は、予め任意の値を定めればよい。例えば、第２閾値は、上記１または複数のアプリケーション３２の各々に割当てられた物理メモリサイズの合計値のＮ分の１（Ｎは、２以上の整数）。例えば、第２閾値は、上記合計値の１／３、１／５、１／７、１／１０などの値などである。基本的な考え方としては、利用するＤＲＡＭに対してそれに見合う高速化が得られる場合は第１アクセス方式を選択したい。あるアプリケーションに対して、そのアプリケーションに割り当てられた物理メモリサイズを１０としたときに、ＤＲＡＭを１あるいは２あるいは３程度利用することでＤＲＡＭ利用による消費電力増加やコストを抑えつつＤＲＡＭ利用による局所性の高い処理の高速化が得られるのが好ましいためである。

【０１３７】

次に、決定部１２Ｄが比率を用いてアクセス方式を決定する場合に、処理回路１２が実行する情報処理の手順の一例を説明する。

【０１３８】

図１０は、本変形例の処理回路１２が実行する情報処理の手順の一例を示す、フローチャートである。

【０１３９】

まず、処理回路１２は、上記実施の形態のステップＳ１００～ステップＳ１０２（図９参照）と同様にして、ステップＳ２００～ステップＳ２０２の処理を実行する。

【０１４０】

詳細には、取得部１２Ａが、単位期間Ｔにおける、処理回路１２の動作統計情報４２Ａを取得する（ステップＳ２００）。次に、導出部１２Ｃが、学習部１２Ｂが学習した予測モデル２０に、ステップＳ２００で取得した動作統計情報４２Ａを入力することで、メモリアクセス特性４２Ｂを導出する（ステップＳ２０２）。

【０１４１】

次に、決定部１２Ｄが、ステップＳ２０２で導出したメモリアクセス特性４２Ｂの比率が、第２閾値より大きいか否かを判断する（ステップＳ２０４）。

【０１４２】

メモリアクセス特性４２Ｂの比率が第２閾値より大きいと判断した場合（ステップＳ２０４：Ｙｅｓ）、上記実施の形態のステップＳ１０６～ステップＳ１１０（図９参照）と同様にして、ステップＳ２０６～ステップＳ２０８～ステップＳ２１０の処理を実行する。

【０１４３】

一方、メモリアクセス特性４２Ｂの比率が第２閾値以下と判断した場合（ステップＳ２０４：Ｎｏ）、上記実施の形態のステップＳ１１２～ステップＳ１１８（図９参照）と同様にして、ステップＳ２１２～ステップＳ２１８の処理を実行する。そして、ステップＳ２１０で肯定判断すると（ステップＳ２１０：Ｙｅｓ）、本ルーチンを終了する。

【０１４４】

以上説明したように、決定部１２Ｄは、導出部１２Ｃで導出したメモリアクセス特性４２Ｂの比率が、第２閾値より大きいか否かを判断することで、アクセス方式を決定してもよい。この場合についても、上記実施の形態と同様の効果が得られる。

【０１４５】

なお、上記実施の形態および変形例では、情報処理装置１０が、処理回路１２と、キャッシュメモリ１６と、管理装置１８と、を備える形態を一例として説明した（図１参照）。しかし、情報処理装置１０は、処理回路１２と、キャッシュメモリ１６と、管理装置１

８と、記憶部１４と、を備えた構成であってもよい。また、処理回路１２が、キャッシュメモリ１６および管理装置１８の少なくとも一方を含む構成であってもよい。また、管理装置１８が、記憶部１４およびキャッシュメモリ１６を備えた構成であってもよい。

【０１４６】

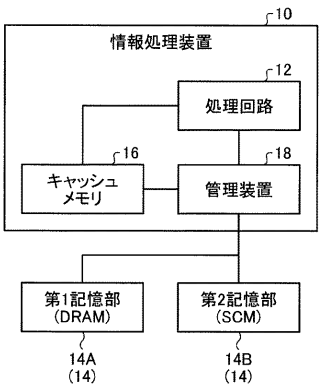
以上、本発明の実施の形態および変形例を説明したが、これらの実施の形態および変形例は、例として提示したものであり、発明の範囲を限定することは意図していない。これら新規な実施の形態および変形例は、その他の様々な形態で実施されることが可能であり、発明の要旨を逸脱しない範囲で、種々の省略、置き換え、変更を行うことができる。これらの実施の形態やその変形例は、発明の範囲や要旨に含まれるとともに、請求の範囲に記載された発明とその均等の範囲に含まれる。

【符号の説明】

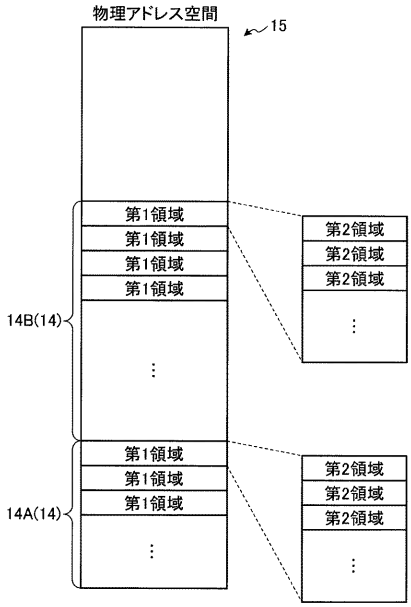
【０１４７】

１０　情報処理装置
１２　処理回路
１２Ａ　取得部
１２Ｂ　学習部
１２Ｃ　導出部
１２Ｄ　決定部
１２Ｅ　実行部
１２Ｆ　変更部
１４　記憶部
１４Ａ　第１記憶部
１４Ｂ　第２記憶部
２０　予測モデル

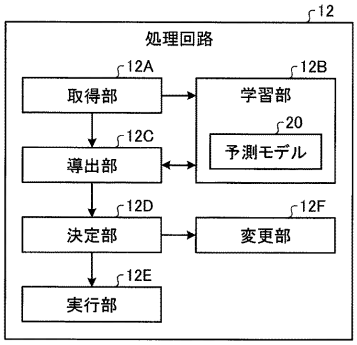
【図１】



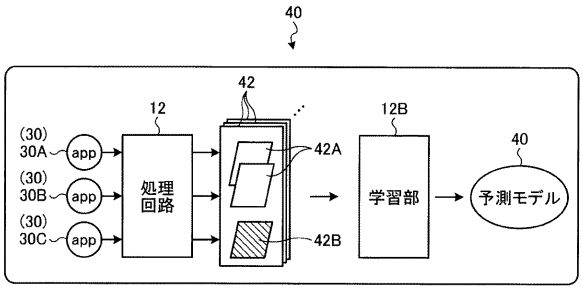
【図２】



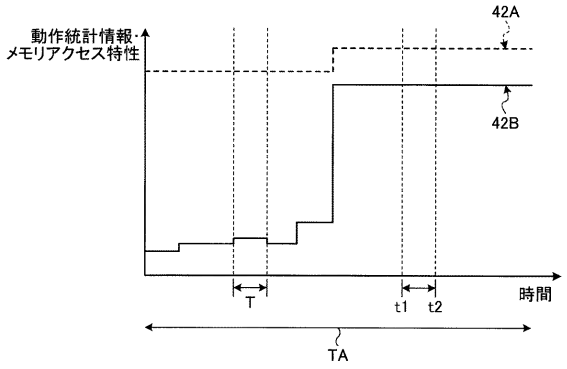
【図３】



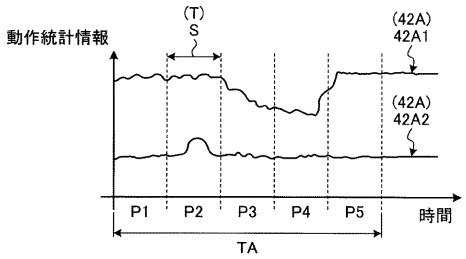
【図４】



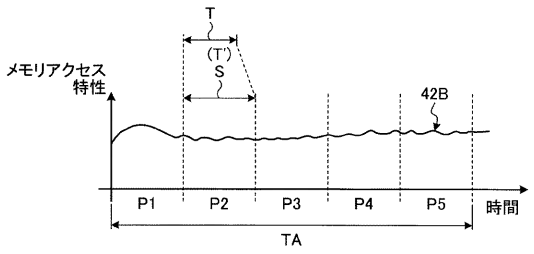
【図５】



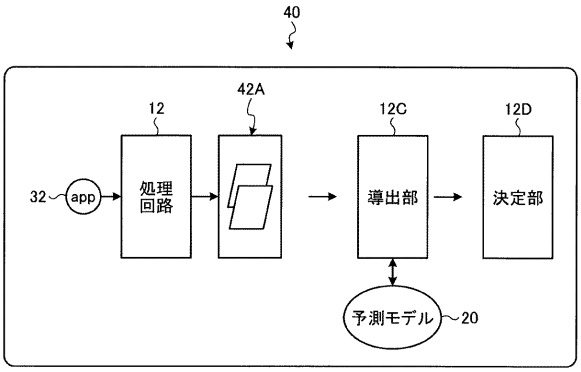
【図６Ａ】



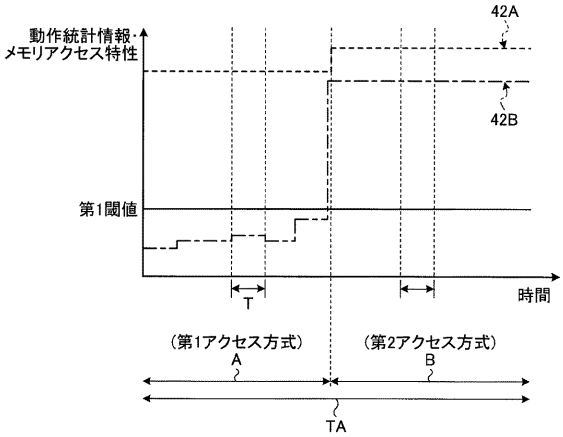
【図６Ｂ】



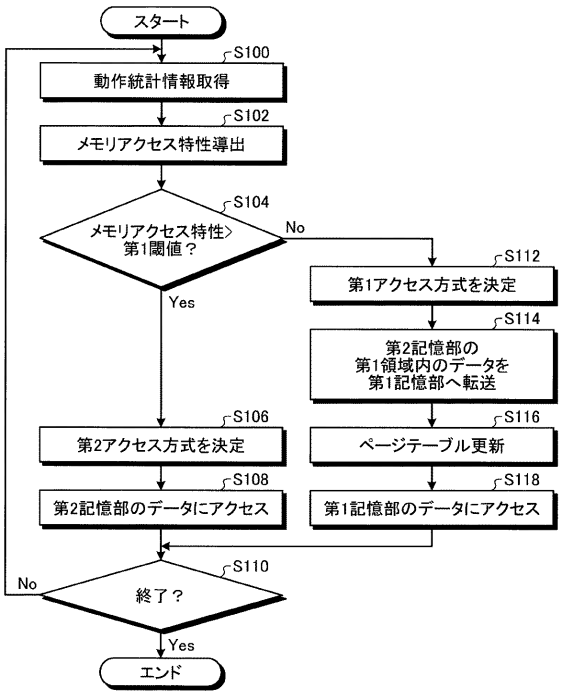
【図７】


40
42A
12
12C
12D
32
app
処理回路
導出部
決定部
予測モデル
20

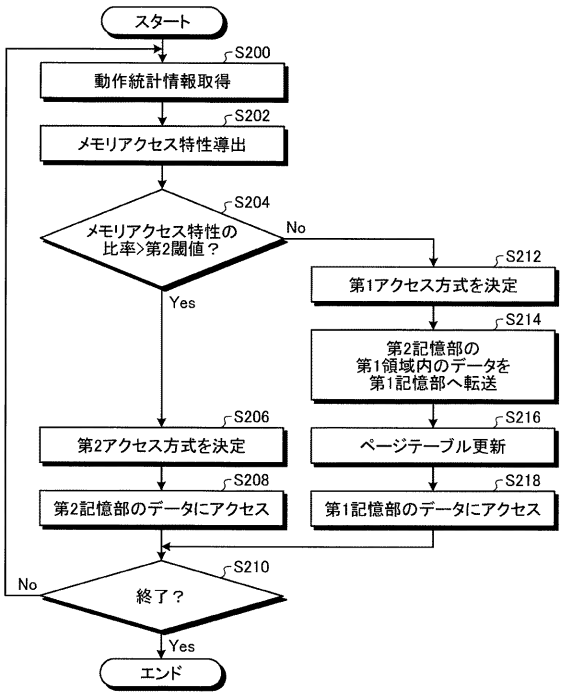
【図８】


42A
動作統計情報・メモリアクセス特性
42B
第1閾値
時間
T
（第1アクセス方式）
A
（第2アクセス方式）
B
TA

【図９】


スタート
S100
動作統計情報取得
S102
メモリアクセス特性導出
S104
メモリアクセス特性>第1閾値？
No
Yes
S112
第1アクセス方式を決定
S114
第2記憶部の第1領域内のデータを第1記憶部へ転送
S106
第2アクセス方式を決定
S116
ページテーブル更新
S108
第2記憶部のデータにアクセス
S118
第1記憶部のデータにアクセス
S110
終了？
No
Yes
エンド

【図１０】


スタート
S200
動作統計情報取得
S202
メモリアクセス特性導出
S204
メモリアクセス特性の比率>第2閾値？
No
Yes
S212
第1アクセス方式を決定
S214
第2記憶部の第1領域内のデータを第1記憶部へ転送
S206
第2アクセス方式を決定
S216
ページテーブル更新
S208
第2記憶部のデータにアクセス
S218
第1記憶部のデータにアクセス
S210
終了？
No
Yes
エンド