

(19) 日本国特許庁 (JP)

(12) 特 許 公 報 (B2)

(11) 特許番号

特許第6149595号  
(P6149595)

(45) 発行日 平成29年6月21日 (2017.6.21)

(24) 登録日 平成29年6月2日 (2017.6.2)

(51) Int.Cl.

F I

G 0 6 F 12/08 (2016.01)

G 0 6 F 12/08 5 2 3 B

G 0 6 F 12/08 5 4 3 B

G 0 6 F 12/08 5 6 5

請求項の数 14 (全 41 頁)

(21) 出願番号 特願2013-168069 (P2013-168069)  
 (22) 出願日 平成25年8月13日 (2013.8.13)  
 (65) 公開番号 特開2015-36873 (P2015-36873A)  
 (43) 公開日 平成27年2月23日 (2015.2.23)  
 審査請求日 平成28年5月10日 (2016.5.10)

(73) 特許権者 000005223  
 富士通株式会社  
 神奈川県川崎市中原区上小田中4丁目1番  
 1号  
 (74) 代理人 100094525  
 弁理士 土井 健二  
 (74) 代理人 100094514  
 弁理士 林 恒徳  
 (72) 発明者 藤井 正俊  
 神奈川県川崎市中原区上小田中4丁目1番  
 1号 富士通株式会社内  
 (72) 発明者 日野原 久  
 神奈川県川崎市中原区上小田中4丁目1番  
 1号 富士通株式会社内

最終頁に続く

(54) 【発明の名称】 キャッシュメモリ制御プログラム、キャッシュメモリを内蔵するプロセッサ及びキャッシュメモリ制御方法

(57) 【特許請求の範囲】

【請求項 1】

キャッシュメモリを共用キャッシュ領域と専用キャッシュ領域とに分割して制御するキャッシュメモリ制御工程をプロセッサに実行させるプロセッサ読み取り可能なキャッシュメモリ制御プログラムであって、

前記キャッシュメモリ制御工程は、

前記プロセッサによるプロセスの実行中に発行される専用キャッシュ領域の割当を要求する専用領域獲得要求に応答して、前記専用領域獲得要求が対象とする対象処理について前記専用キャッシュ領域を割り当てた場合のキャッシュヒット率と前記共用キャッシュ領域を割り当てた場合のキャッシュヒット率との差と、メモリアクセス頻度とに基づくキャッシュ有効利用度に応じて、前記キャッシュ有効利用度が基準値より高い場合専用キャッシュ領域を割当て、低い場合共用キャッシュ領域を割り当てるキャッシュ領域割当工程と

前記プロセッサによるプロセスの実行中に発行される割り当て済み専用キャッシュ領域の開放を要求する専用領域開放要求に応答して、前記割り当てられた専用キャッシュ領域の割当を解放する専用キャッシュ領域開放工程とを有し、

前記専用キャッシュ領域が割り当てられた対象処理中に発行されるメモリアクセス命令が、前記プロセッサにより前記割り当てた専用キャッシュ領域を利用して実行されるキャッシュメモリ制御プログラム。

【請求項 2】

請求項 1 において、

前記キャッシュ有効利用度は、前記専用キャッシュ領域を割り当てた場合のキャッシュヒット率と前記共用キャッシュ領域を割り当てた場合のキャッシュヒット率との差に、前記メモリアクセス頻度を乗算し利用する専用キャッシュ領域のサイズで除算した値であるキャッシュメモリ制御プログラム。

【請求項 3】

請求項 2 において、

キャッシュ領域割当工程では、前記キャッシュ有効利用度に代えて、前記キャッシュ有効利用度に更に前記対象処理の優先度を乗じた実キャッシュ有効利用度に応じて専用キャッシュ領域または共用キャッシュ領域を割り当てるキャッシュメモリ制御プログラム。

10

【請求項 4】

請求項 1 において、

前記専用キャッシュ領域を割り当てた場合のキャッシュヒット率は、予め、割り当てる専用キャッシュ領域のサイズを可変させて前記対象処理を前記プロセッサに実行させたときに取得した専用キャッシュ領域サイズ対キャッシュヒット率の特性データであるキャッシュメモリ制御プログラム。

【請求項 5】

請求項 1 において、

前記キャッシュ領域割当工程において、前記専用領域獲得要求に応答して、前記専用領域獲得要求が対象とする対象処理についてのキャッシュミス率に前記メモリアクセス頻度を乗算したキャッシュ攪乱力が基準値より高い場合に、前記共用キャッシュ領域と専用キャッシュ領域とは異なる攪乱処理キャッシュ領域を割り当てるキャッシュメモリ制御プログラム。

20

【請求項 6】

請求項 2 または 3 において、

前記キャッシュ領域割当工程において、新たに発行された発行中専用領域獲得要求の前記キャッシュ有効利用度と、前記専用キャッシュ領域が割り当て済みの割当済み専用領域獲得要求の前記キャッシュ有効利用度とを比較し、前記発行中専用領域獲得要求のキャッシュ有効利用度のほうが高い場合に、前記割当済み専用領域獲得要求に対して割り当て済みの専用キャッシュ領域を開放して前記共用キャッシュ領域を割り当て、前記発行中専用領域獲得要求に前記専用キャッシュ領域を割り当てるキャッシュ制御プログラム。

30

【請求項 7】

請求項 2 または 3 において、

前記キャッシュ領域割当工程において、新たに発行された発行中専用領域獲得要求の前記キャッシュ有効利用度と前記専用キャッシュ領域サイズを乗算したキャッシュ有効利用数と、前記専用キャッシュ領域が割り当て済みの割当済み専用領域獲得要求の単位領域の増加に対するキャッシュヒット率の増加率である増加率に、前記メモリアクセス頻度を乗算したキャッシュ有効利用数の増加率を、前記発行中専用領域獲得要求の専用キャッシュ領域サイズ分だけ累積した累積増加数とを比較し、

40

前記発行中専用領域獲得要求の前記キャッシュ有効利用数のほうが多い場合に、前記割当済み専用領域獲得要求に対して割り当て済みの専用キャッシュ領域を開放して前記共用キャッシュ領域を割り当て、前記発行中専用領域獲得要求に前記専用キャッシュ領域を割り当てるキャッシュ制御プログラム。

【請求項 8】

請求項 2 または 3 において、

前記キャッシュメモリ制御工程は、更に、

前記プロセス実行中に、前記専用キャッシュ領域が割り当て済みの割当済み専用領域獲得要求が対象とする対象処理について、前記割当済み専用キャッシュ領域のキャッシュヒット率とアクセス頻度を調査し、前記割当済み専用領域獲得要求の単位領域の増加に対す

50

るキャッシュヒット率の増加数である増加率に、前記メモリアクセス頻度を乗算したキャッシュ有効利用数の増加率のリストを有する割当済み専用領域リストと、前記専用キャッシュ領域が未だ割り当てられていない未割当専用領域獲得要求の前記キャッシュ有効利用数の増加率のリストを有する未割当専用領域リストとを作成するリスト作成工程と、

前記プロセス実行中に定期的に行われ、前記割当済み専用領域リストの最小値より前記未割当専用領域リストの最大値が入れ替え処理に要するオーバーヘッド値以上大きい場合に、対応する前記割当済み専用領域獲得要求に対して割当済みの専用キャッシュ領域を開放して前記共用キャッシュ領域を割り当て、対応する前記未割当専用領域獲得要求に前記専用キャッシュ領域を割り当てる定期的専用領域入れ替え工程とを有するキャッシュ制御プログラム。

10

【請求項 9】

請求項 8 において、

前記定期的専用領域入れ替え工程において、前記未割当専用領域リストの最大値における専用キャッシュ領域サイズを増加させながら、前記未割当専用領域リストの最大値のキャッシュ有効利用数の増加率が、前記割当済み専用領域リストの最小値から順に前記増加させる専用キャッシュ領域サイズに対応するリスト数までの割当済み専用領域のキャッシュ有効利用数の増加率の累計値より、オーバーヘッド値以上多いか否かを判定し、前記判定が多い場合に、前記未割当専用領域リストの最大値に対応する未割当専用領域獲得要求に前記増加させた専用キャッシュ領域サイズの最大のサイズ分だけ前記専用キャッシュ領域を割り当て、前記割当済み専用領域リストの前記リスト数まで前記割当済みの専用キャッシュ領域を開放して前記共用キャッシュ領域を割り当てるキャッシュ制御プログラム。

20

【請求項 10】

請求項 2 または 3 において、

前記キャッシュメモリ制御工程は、更に、

前記プロセス実行中に定期的に行われる定期的専用領域入れ替え工程を有し、

前記定期的専用領域入れ替え工程において、

現実の加重メモリアクセス比に基づく共用キャッシュ領域の理想サイズと、現実の共用キャッシュ領域のサイズとの差分が、オーバーヘッド値を超えて高い場合に、前記共用キャッシュ領域が理想のサイズ比に近づくように、前記共用キャッシュ領域を未だ専用領域が割り当てられていない未割当専用領域獲得要求に専用キャッシュ領域として割り当てるか、または、専用領域が割当済みの割当済み専用領域獲得要求に割り当てられた割当済み専用キャッシュ領域を開放するかにより、前記専用キャッシュ領域の入れ替え処理を行うキャッシュ制御プログラム。

30

【請求項 11】

請求項 10 において、

前記キャッシュメモリ制御工程は、更に、

前記プロセス実行中に、前記割当済み専用領域獲得要求が対象とする対象処理について、前記割当済み専用キャッシュ領域のキャッシュヒット率とアクセス頻度を調査し、前記割当済み専用領域獲得要求の単位領域増加に対するキャッシュヒット率の増加数である増加率に前記メモリアクセス頻度を乗算したキャッシュ有効利用数の増加率のリストを有する割当済み専用領域リストと、前記未割当専用領域獲得要求の前記キャッシュ有効利用数の増加率のリストを有する未割当専用領域リストとを作成するリスト作成工程を有し、

40

前記定期的専用領域入れ替え工程において、

前記割当済み専用キャッシュ領域を開放する場合、前記割当済み専用領域リスト内の前記増加率が少ない順に、前記割当済み専用キャッシュ領域を開放して前記共用キャッシュ領域を割り当て、

前記専用キャッシュ領域を割り当てる場合、前記未割当専用領域リスト内の前記増加率が多い順に、前記専用キャッシュ領域を割り当てる、

キャッシュ制御プログラム。

【請求項 12】

50

キャッシュメモリを共用キャッシュ領域と専用キャッシュ領域とに分割して制御するキャッシュメモリ制御工程をプロセッサに実行させるプロセッサ読み取り可能なキャッシュメモリ制御プログラムであって、

前記キャッシュメモリ制御工程は、

前記プロセッサによるプロセスの実行中に発行される専用キャッシュ領域の割当を要求する専用領域獲得要求に応答して、前記専用領域獲得要求が対象とする対象処理についてのキャッシュミス率にメモリアクセス頻度を乗算したキャッシュ攪乱力が基準値より高い場合に、他のキャッシュ領域とはキャッシュデータ管理が隔離された専用キャッシュ領域を割り当てるキャッシュ領域割当工程と、

前記プロセッサによるプロセスの実行中に発行される割り当て済み専用キャッシュ領域の開放を要求する専用領域開放要求に応答して、前記割り当てられた専用キャッシュ領域の割当を開放する専用キャッシュ領域開放工程とを有し、

前記専用キャッシュ領域が割り当てられた対象処理中に発行されるメモリアクセス命令が、前記プロセッサにより前記割り当てた専用キャッシュ領域を利用して実行されるキャッシュメモリ制御プログラム。

#### 【請求項 13】

命令を実行してメモリアクセスを行うCPUと、

前記メモリアクセスが行われるときに参照されるキャッシュメモリと、

前記キャッシュメモリを共用キャッシュ領域と専用キャッシュ領域とに分割して制御するキャッシュメモリ制御ユニットとを有し、

前記キャッシュメモリ制御ユニットは、

前記CPUによるプロセスの実行中に発行される専用キャッシュ領域の割当を要求する専用領域獲得要求に応答して、前記専用領域獲得要求が対象とする対象処理について前記専用キャッシュ領域を割り当てた場合のキャッシュヒット率と前記共用キャッシュ領域を割り当てた場合のキャッシュヒット率との差と、メモリアクセス頻度とに基づくキャッシュ有効利用度に応じて、前記キャッシュ有効利用度が基準値より高い場合専用キャッシュ領域を割当て、低い場合共用キャッシュ領域を割り当てるキャッシュ領域割当ユニットと

、前記CPUによるプロセスの実行中に発行される割り当て済み専用キャッシュ領域の開放を要求する専用領域開放要求に応答して、前記割り当てられた専用キャッシュ領域の割当を解放する専用キャッシュ領域開放ユニットとを有し、

前記専用キャッシュ領域が割り当てられた対象処理中に発行される前記メモリアクセスが、前記割り当てた専用キャッシュ領域を利用して実行されるプロセッサ。

#### 【請求項 14】

キャッシュメモリを共用キャッシュ領域と専用キャッシュ領域とに分割して制御するキャッシュメモリ制御方法であって、

プロセッサによるプロセスの実行中に発行される専用キャッシュ領域の割当を要求する専用領域獲得要求に応答して、前記専用領域獲得要求が対象とする対象処理について前記専用キャッシュ領域を割り当てた場合のキャッシュヒット率と前記共用キャッシュ領域を割り当てた場合のキャッシュヒット率との差と、メモリアクセス頻度とに基づくキャッシュ有効利用度に応じて、前記キャッシュ有効利用度が基準値より高い場合専用キャッシュ領域を割当て、低い場合共用キャッシュ領域を割り当てるキャッシュ領域割当工程と、

前記プロセッサによるプロセスの実行中に発行される割り当て済み専用キャッシュ領域の開放を要求する専用領域開放要求に応答して、前記割り当てられた専用キャッシュ領域の割当を解放する専用キャッシュ領域開放工程とを有し、

前記専用キャッシュ領域が割り当てられた対象処理中に発行されるメモリアクセス命令が、前記プロセッサにより前記割り当てた専用キャッシュ領域を利用して実行されるキャッシュメモリ制御方法。

#### 【発明の詳細な説明】

#### 【技術分野】

10

20

30

40

50

## 【 0 0 0 1 】

本発明は、キャッシュメモリ制御プログラム、キャッシュメモリを内蔵するプロセッサ及びキャッシュメモリ制御方法に関する。

## 【 背景技術 】

## 【 0 0 0 2 】

近年のサーバシステムは、プロセッサの高性能化が進んでいる。しかし、CPUによるメモリアクセス速度が遅く、その点が性能向上のネックになっている。プロセッサはCPUに加えてキャッシュメモリを有し、CPUによるメモリアクセスが発生すると、まずキャッシュメモリがチェックされて、キャッシュヒットすればキャッシュメモリにアクセスが行われる。そして、キャッシュミスした場合にメインメモリにアクセスが行われる。キャッシュメモリへのアクセスは、メインメモリへのアクセスと比べると100～300倍高速である。したがって、CPUによるメモリアクセスの性能を向上させるためには、できるだけキャッシュヒットさせてメインメモリへのアクセスを発生させないことが重要である。

10

## 【 0 0 0 3 】

一般に、キャッシュヒット率は、キャッシュサイズを増加すると向上する。しかし、キャッシュヒット率は、キャッシュサイズが比較的小さい場合は、キャッシュサイズの増大に依存して大きく増加するが、ある程度までキャッシュサイズが大きくなると、キャッシュサイズを増加してもキャッシュヒット率はそれほど向上しない。

## 【 0 0 0 4 】

一方で、マルチタスク化に伴い、プロセッサは複数のプロセスを同時にまたは交互に実行する。そのため、キャッシュメモリには複数のプロセスのデータが記憶され、第1のプロセスで利用していたキャッシュメモリの領域が、第1のプロセスとは異なる第2のプロセスで利用されることになり、第1のプロセスのデータがキャッシュメモリから押し出されてしまう。そのため、その後第1のプロセスが再開したときに、キャッシュミスヒットが頻発する。従って、全てのプロセスが同じキャッシュメモリを共用するアーキテクチャでは、キャッシュヒット率向上にはキャッシュメモリの制御に何らかの改良が必要である。

20

## 【 0 0 0 5 】

その一例として、特定の処理に対して専用のキャッシュ領域を割り当てることが提案されている。例えば、以下の特許文献などである。これにより、特定の処理のメモリプロセスが専用キャッシュ領域に限定されるので、共用キャッシュ領域でのキャッシュミスヒットの発生を抑制することができる。

30

## 【 先行技術文献 】

## 【 特許文献 】

## 【 0 0 0 6 】

【 特許文献 1 】 特開 2 0 0 4 - 1 3 3 9 3 1 号公報

【 特許文献 2 】 特表 2 0 0 9 - 5 1 8 7 5 4 号公報

【 特許文献 3 】 特開 2 0 0 2 - 7 2 1 3 号公報

【 特許文献 4 】 特開 2 0 0 6 - 4 2 0 3 号公報

【 特許文献 5 】 特開平 0 2 - 1 8 6 4 4 号公報

40

【 特許文献 6 】 特開 2 0 0 9 - 2 1 1 6 4 9 号公報

【 特許文献 7 】 特開 2 0 0 9 - 1 6 2 4 5 0 号公報

## 【 発明の概要 】

## 【 発明が解決しようとする課題 】

## 【 0 0 0 7 】

しかしながら、上記のキャッシュメモリの制御方法では、キャッシュメモリの領域を特定の処理に対して専用キャッシュ領域を固定的に割り当てただけである。また、キャッシュメモリの領域を、共用キャッシュ領域と特定用途の専用キャッシュ領域とに固定的に割り当てただけである。そのため、事前にキャッシュメモリの領域を固定的に分割して割り当てることになり、常に発生するわけではない割り込み処理や大量データのIO処理のため

50

に専用キャッシュ領域が割り当てられ、全てのキャッシュメモリを有効活用することには  
ならない。さらに、一度しかデータを参照しない処理や、固定的にキャッシュメモリにデ  
ータを格納する処理など、様々な処理に対して専用キャッシュ領域を割り当てる必要が生  
じ、それら全てに対してそれぞれ専用キャッシュ領域を固定的に割り当てるのは、無駄で  
あり非現実的といえる。

【0008】

また、アプリケーションプログラムは、各処理に対してキャッシュメモリ内のどの専用  
キャッシュ領域を利用するかを事前に記述される必要があり、プログラム開発の工数を増  
加させる要因にもなる。特殊なアプリケーションプログラムだけを超高速に実行するスー  
パーコンピュータなどでは、キャッシュメモリ内に固定的に専用キャッシュ領域を割り当  
てること、キャッシュヒット率を向上させることができるかもしれない。しかし、前述  
のマルチタスク化されたサーバのプロセッサは、様々なアプリケーションプログラムを実  
行する必要があり、しかも、どのアプリケーションプログラムが実行されるかを予測する  
ことは困難であり、固定的に専用キャッシュ領域を割り当てる方法は現実的ではない。

【0009】

そこで、1つの側面では、本発明の目的は、キャッシュヒット率を向上させるキャッシ  
ュメモリ制御プログラム、キャッシュメモリを内蔵するプロセッサ及びキャッシュメモリ  
制御方法を提供することにある。

【課題を解決するための手段】

【0010】

実施の形態の第1の側面は、キャッシュメモリを共用キャッシュ領域と専用キャッシュ  
領域とに分割して制御するキャッシュメモリ制御工程をプロセッサに実行させるプロセッ  
サ読み取り可能なキャッシュメモリ制御プログラムであって、

前記キャッシュメモリ制御工程は、

前記プロセッサによるプロセスの実行中に発行される専用キャッシュ領域の割当を要求  
する専用領域獲得要求に応答して、前記専用領域獲得要求が対象とする対象処理について  
前記専用キャッシュ領域を割り当てた場合のキャッシュヒット率と前記共用キャッシュ領  
域を割り当てた場合のキャッシュヒット率との差と、メモリアクセス頻度とに基づくキャ  
ッシュ有効利用度に応じて、前記キャッシュ有効利用度が高い程専用キャッシュ領域を割  
当て、低い程共用キャッシュ領域を割り当てるキャッシュ領域割当工程と、

前記プロセッサによるプロセスの実行中に発行される割り当て済み専用キャッシュ領域  
の開放を要求する専用領域開放要求に応答して、前記割り当てられた専用キャッシュ領域  
の割当を解放する専用キャッシュ領域開放工程とを有し、

前記専用キャッシュ領域が割り当てられた対象処理中に発行されるメモリアクセス命令  
が、前記プロセッサにより前記割り当てた専用キャッシュ領域を利用して実行される。

【発明の効果】

【0011】

第1の側面によれば、キャッシュメモリの利用効率が上昇する。

【図面の簡単な説明】

【0012】

【図1】本実施の形態におけるサーバの構成図である。

【図2】本実施の形態におけるサーバのハードウェアとソフトウェアの構成図である。

【図3】本実施の形態におけるキャッシュ領域の動的割当の概略を示すフローチャート図  
である。

【図4】コンパイラ21がアプリケーションプログラムのソースコードを機械語にコンパ  
イルする処理を示す図である。

【図5】上記の特性調査で得られたキャッシュサイズ対キャッシュヒット率の特性例で  
ある。

【図6】キャッシュ領域の構成と専用領域獲得要求についての割当について示す図である

【図 7】変換テーブルの一例を示す図である。

【図 8】図 3 に示した専用領域獲得要求に対する判定処理S16について第 1 の例のフローチャート図である。

【図 9】専用領域獲得要求に専用キャッシュ領域を割り当ててする場合を説明する図である。

【図 10】図 3 に示した専用領域獲得要求に対する判定処理S16について第 2 の例のフローチャート図である。

【図 11】専用領域獲得要求と割当済専用領域獲得要求との間で専用キャッシュ領域を入れ替える処理を説明する図である。

【図 12】推定キャッシュヒット率を示す図である。

【図 13】割当済専用領域リストLIST-FREEと未割当専用領域リストLIST-DEMANDとを示す図である。 10

【図 14】定期割当見直し処理S18の第 1 の例のフローチャート図である。

【図 15】定期割当見直し処理S18の第 1 の例を説明するための図である。

【図 16】定期割当見直し処理S18の第 2 の例のフローチャート図である。

【図 17】定期割当見直し処理S18の第 2 の例を説明するための図である。

【図 18】定期割当見直し処理S18の第 3 の例のフローチャート図である。

【図 19】定期割当見直し処理S18の第 3 の例を説明するための図である。

【図 20】図 3 の工程S2とS3のサーバシステム内での処理を示す図である。

【図 21】図 20 のサーバシステム内の処理に対応するフローチャート図である。

【図 22】図 3 の工程S13, S16, S19のサーバシステム内での処理を示す図である。 20

【図 23】図 22 のサーバシステム内の処理に対応するフローチャート図である。

【図 24】本専用領域獲得要求判定処理S16 ( 2 ) のフローチャート図である。

【図 25】アプリケーションプログラムのプロセス 22 が切り替わるときのセクタID変換テーブルの入れ替え処理を示す図である。

【図 26】第 2 の実施の形態におけるOS割込処理プログラム ( OS割込処理プロセス203 ) の構成例を示す図である。

【図 27】第 2 の実施の形態におけるメモリアクセスを説明する図である。

【図 28】第 2 の実施の形態における割込・例外処理発生時の処理を示す図である。

【図 29】第 2 の実施の形態における割込・例外処理発生時の処理を示す図である。

【図 30】第 2 の実施の形態における割込・例外処理発生時の処理のフローチャート図である。 30

【発明を実施するための形態】

【 0 0 1 3 】

以下は、本実施の形態の目次である。

【 0 0 1 4 】

[ 第 1 の実施の形態 ]

[ 本実施の形態のキャッシュ領域の動的割当の概略 ]

[ 専用領域獲得要求での専用キャッシュ領域の割当S15, S16 ]

[ 専用キャッシュ領域を利用するメモリアクセスS13 ]

[ 専用領域開放要求での専用キャッシュ領域の開放 ] 40

[ 本実施の形態のキャッシュ領域の動的割当処理の詳細 ]

[ キャッシュ領域の割当判定の指標 ]

[ 実キャッシュヒット数 ]

[ メモリアクセスの局所性 ]

[ キャッシュ有効利用度, 実キャッシュ有効利用度, キャッシュ攪乱力 ]

[ 専用領域獲得要求判定処理S16 ( 1 ) ]

[ 専用領域獲得要求判定処理S16 ( 2 ) ]

[ 定期割当見直し処理S18 ]

[ 定期割当見直し処理S18 ( 1 ) ]

[ 定期割当見直し処理S18 ( 2 ) ] 50

[ 定期割当見直し処理S18 ( 3 ) ]

[ プロセス切り替え時のセクタID変換テーブルの入れ替え処理 ]

[ 第 2 の実施の形態 ]

以下、上記の目次にしたがって本実施の形態を説明する。

【 0 0 1 5 】

[ 第 1 の実施の形態 ]

図 1 は、本実施の形態におけるサーバの構成図である。サーバは、プロセッサユニット 1 0 と、メインメモリ 1 6 と、キーボードや表示装置などの入出力部 1 7 と、ハードディスクなどの大容量のストレージ 1 8 とを有する。プロセッサユニット 1 0 は、機械語命令を実行する CPU 1 2 と、キャッシュメモリとキャッシュメモリの制御部を有するキャッシュユニット 1 3 と、CPU からメインメモリ 1 6 へのアクセスの制御を行うバスインターフェース 1 4 とを有する。

10

【 0 0 1 6 】

ストレージ 1 8 には、オペレーションシステム ( OS ) と、コンパレータなどのミドルウェア ( M-Ware ) と、アプリケーションプログラム ( APL ) などが格納されている。これらのプログラムは、サーバ起動時にメインメモリ 1 6 内にロードされ、CPU 1 2 によって実行される。また、メインメモリ 1 6 は、アプリケーションプログラム APL に含まれるプロセスの実行に必要なデータを記憶する。また、CPU 1 2 によるキャッシュメモリへのアクセス速度と比較すると、メインメモリ 1 6 へのアクセス速度は極端に遅い。一方、キャッシュメモリの容量は、メインメモリ 1 6 の容量に比較すると非常に小さい。

20

【 0 0 1 7 】

図 2 は、本実施の形態におけるサーバのハードウェアとソフトウェアの構成図である。ハードウェア 1 は、図 1 に示した構成を有し、図 2 には、特に、メインメモリ 1 6 と、キャッシュシステム 1 3 と、CPU 1 2 とが示されている。

【 0 0 1 8 】

キャッシュシステム 1 3 は、SRAM など高速の半導体メモリなどで構成されるキャッシュメモリ 1 3 0 と、そのキャッシュメモリ 1 3 0 へアクセスするときに参照されるセクタID変換テーブル 1 3 1、専用キャッシュ領域のサイズ設定などが行われるキャッシュセクタ情報設定部 1 3 2、キャッシュメモリへアクセスするときに使用されるキャッシュアドレス変換部 1 3 3 などとを有する。

30

【 0 0 1 9 】

本実施の形態において、CPU12 は、通常の機械語命令に加えて、ローカルセクタIDが付加された機械語命令を実行する。後述するとおり、ローカルセクタIDは、その機械語命令が利用するキャッシュメモリの領域を識別するために使用される。

【 0 0 2 0 】

OS20 は、アプリケーションプログラムの様々なプロセスにキャッシュメモリの領域 ( キャッシュセクタ ) を動的に割り当てる場合に、セクタID変換テーブル 1 3 1 にローカルセクタIDに対して割り当てるキャッシュセクタIDをひもづける。そして、セクタID変換テーブル 1 3 1 は、キャッシュメモリにアクセスするときに参照され、機械語命令に付加されたローカルセクタIDをキャッシュメモリ内の使用領域を示すキャッシュセクタIDに変換する。このような方法により、OS20 は、プログラムの様々なプロセスやプロセス内の処理に対して、専用キャッシュ領域を動的に割り当てて、そのプロセスやプロセス内の処理に割り当てた専用キャッシュ領域を使用させることができる。

40

【 0 0 2 1 】

キャッシュシステム 1 3 内のキャッシュセクタ情報設定部 1 3 2 は、キャッシュメモリの領域 ( キャッシュセクタ ) に対してどの程度のキャッシュサイズを割り当てるかが設定される。後述するとおり、OS20 は、専用キャッシュ領域を割り当てる場合に、そのキャッシュサイズをキャッシュセクタ情報設定部 1 3 2 に設定する。キャッシュアドレス変換部 1 3 3 は、CPU12 によるメモリアクセスのアドレスをキャッシュメモリのアドレスに変換する。

50



## 【 0 0 2 2 】

図2のソフトウェア2は、オペレーションシステム(OS)20と、アプリケーションプログラムを機械語に変換するコンパイラなどのミドルウェア21と、複数のアプリケーションプログラムを有するアプリケーションプログラム群22とを有する。

## 【 0 0 2 3 】

OS内の関数制御管理部201は、アプリケーションプログラムの各プロセス内で呼び出される関数の管理を行う。一般に、関数制御管理部201は、関数が呼び出されるとその関数で利用されるデータをスタックデータとして管理し、また、関数が読み戻されるとスタックデータを開放する。

## 【 0 0 2 4 】

OS20は、マルチタスク(またはマルチプロセス)を切り替えながら各プロセスの処理を実行する。プロセス管理部202は、この各プロセスの切替制御を行う。本実施の形態では、特に、プロセス管理部202は、プロセスが切り替えられるたびに、キャッシュシステム13内のセクタID変換テーブル131を切り替える。これにより、各プロセスに対応して適切なキャッシュ領域の動的な割当を行うことができる。具体的には後述する。

## 【 0 0 2 5 】

割り込み処理プロセス203は、IO割り込みや内部プログラムからの例外的な割り込みなどが発生したときに、実行されるOS割り込み処理プログラムである。この割り込み処理プロセス内には、専用領域獲得要求と専用領域解放要求が含まれている。具体的には後述する。

## 【 0 0 2 6 】

本実施の形態では、OS12が、アプリケーションプログラムのプロセス内で発行される専用領域獲得要求に応答して、キャッシュメモリ内の例えば共用キャッシュ領域と、専用キャッシュ領域と、攪乱処理キャッシュ領域などの割当を、動的に実行する。この動的な割り当ては、専用領域獲得要求の対応する処理(以下対応処理)のメモリアクセスの特性や専用領域獲得要求時のキャッシュメモリのアクセス状況などに基づいて行われる。キャッシュ割当管理部204は、これらのキャッシュ領域の動的な割当に関する制御を行う。

## 【 0 0 2 7 】

上記の共用キャッシュ領域(または共用領域)とは、複数のプロセスや対応処理が共有して利用するキャッシュメモリの領域であり、あるプロセスや対応処理が共用キャッシュ領域を利用することで、他のプロセスや対応処理によって共用キャッシュ領域に記憶されたデータが追い出されることがある。

## 【 0 0 2 8 】

一方、専用キャッシュ領域(または専用領域)とは、ある専用領域獲得要求に対して割り当てられるキャッシュメモリの領域であり、その専用領域獲得要求の対象処理内のメモリアクセスは、それに割り当てられた専用キャッシュ領域を使用する。したがって、他のプロセスや対応処理のメモリアクセスにより専用キャッシュ領域内のデータが追い出されることはないし、専用キャッシュ領域へのアクセスにより、共用キャッシュ領域や他の専用キャッシュ領域のデータが追い出されることもない。

## 【 0 0 2 9 】

また、攪乱処理キャッシュ領域とは、対応する処理のキャッシュミスヒット率が非常に高い対応処理に対する専用領域獲得要求に対して割り当てられるキャッシュメモリの領域であり、一種の専用キャッシュ領域である。攪乱処理キャッシュ領域では、共用キャッシュ領域や他の専用キャッシュ領域から隔離して、そのキャッシュ制御が行われる。そのため、攪乱処理キャッシュ領域が割り当てられた対応処理の高いキャッシュミスヒット率により、共用キャッシュ領域や他の専用キャッシュ領域内のデータが追い出されてその後のキャッシュミスヒット率が高くなることが回避される。攪乱処理キャッシュ領域は、複数の専用領域獲得要求に割り当てられることがある。攪乱処理キャッシュ領域を設けることで、攪乱処理による高いキャッシュミスヒット率により共用キャッシュ領域や専用キャッシュ領域のデータが追い出されることが抑制され、キャッシュヒット率が極端に低下する

10

20

30

40

50

ことが抑制される。

【 0 0 3 0 】

キャッシュ割当管理部 2 0 4 が実行する処理工程には、例えば、次のようなものがある。第 1 に、開発環境下において、各プロセスや専用領域獲得要求が対象とする対象処理などについて、メモリアクセス頻度や、専用キャッシュ領域を割り当てられた場合のキャッシュヒット率などのメモリアクセステ性の調査を行う特性調査工程と、第 2 に、プロセス切り替え時においてキャッシュシステム 1 3 内のセクタID変換テーブル 1 3 1 を入れ替える工程と、第 3 に、共用キャッシュ領域と、専用キャッシュ領域と、攪乱処理キャッシュ領域間のキャッシュ領域入れ替え判定工程と、第 4 に、キャッシュ領域入れ替え判定結果に基づくセクタID変換テーブルの更新及びキャッシュセクタ情報設定部 1 3 2 の割当キャッシュ領域のサイズ設定の更新工程などである。

10

【 0 0 3 1 】

ミドルウェア 2 1 のコンパイラは、ソースコードで記述されたアプリケーションプログラムを実行形式の機械語（オブジェクトコード）に変換する。このコンパイラは、一般的なコンパイル処理に加えて、各プロセスにおいて適宜発行される専用キャッシュ領域の割当を要求する専用領域獲得要求の機械語命令と、その割り当てられた専用領域を開放する専用領域開放要求の機械語命令を追加する。また、コンパイラは、発行する各専用領域獲得要求には、プロセス内でユニークなローカルセクタIDを付加する。そして、コンパイラは、獲得した専用キャッシュ領域を利用する機械語命令には、そのローカルセクタIDを付加する。これにより、専用領域獲得要求に対して、その対象処理で利用するキャッシュ領域に専用キャッシュ領域が動的に割り当てられ、キャッシュメモリ領域全体を有効に利用することができる。すなわち、後述するキャッシュ有効利用度や実キャッシュ有効利用度を高くすることができる。

20

【 0 0 3 2 】

アプリケーションプログラム（APL） 2 2 は、アプリケーションプログラムの実行形式である機械語にコンパイルされている。これにより、CPU12は、このアプリケーションプログラムの機械語を実行する。

【 0 0 3 3 】

[ 本実施の形態のキャッシュ領域の動的割当の概略 ]

図 3 は、本実施の形態におけるキャッシュ領域の動的割当の概略を示すフローチャート図である。まず、サーバにアプリケーションプログラムAPLをインストールした開発環境でシステムを稼働させる（S1）。まず、コンパイラをプロセッサ（CPU）に実行させて、アプリケーションプログラムAPLをコンパイルする（S2）。このコンパイル工程S2におけるキャッシュ領域の動的割当に関連する部分について以下説明する。

30

【 0 0 3 4 】

図 4 は、コンパイラ 2 1 がアプリケーションプログラムのソースコードを機械語にコンパイルする処理を示す図である。コンパイラは、アプリケーションプログラムのソースコードAPL\_SCを、その実行形式である機械語（オブジェクトコード）APL\_OCに変換する。一般的なコンパイラの機能は、オブジェクトコードを命令部とオペランドを有する機械語の命令に変換することである。

40

【 0 0 3 5 】

本実施の形態では、コンパイラは、さらに、アプリケーションプログラム内の複数のプロセス内に、キャッシュメモリの専用キャッシュ領域への割当を要求する専用領域獲得要求 3 0 と、それに対応する専用領域開放要求 3 4 とを適宜挿入する。この専用領域獲得要求 3 0 は、図 4 に示されるとおり、例えば、プロセスIDと、ローカルセクタIDと、専用領域獲得要求 3 0 の対象処理の優先度を有する機械語命令である。各プロセスにおいて、キャッシュ領域を占有して処理を行いたい部分で、この専用領域獲得要求 3 0 を発行してOS 20に専用キャッシュ領域の割当を要求する。

【 0 0 3 6 】

プロセスIDは、OS20が管理している稼働プロセスPR1の識別情報である。ローカルセク

50

タIDは、この専用領域獲得要求30の対象処理32を識別するための、プロセスPR1内におけるローカルな識別情報であり、割り当てられた専用キャッシュ領域を特定する識別情報である。

【0037】

また、優先度は、対応処理32の優先度を示す値であり、例えば、0 - 7の8段階で表す。ただし、OS内部の処理を優先させる必要があるため、ユーザプログラムには例えば0 - 5までしか優先度が割り当てられないなどの制約をかけることが行われる。

【0038】

一方、専用領域開放要求34は、これと対になる専用領域獲得要求30と同じプロセスIDとローカルセクタIDとを有する。これにより、OS20が割り当てた専用キャッシュ領域が開放され、その対応処理には共用キャッシュ領域が代わりに割り当てられる。

10

【0039】

さらに、コンパイラは、専用領域獲得要求30と専用領域開放要求34との間の命令に、その命令が利用する専用キャッシュ領域を識別するローカルセクタIDを付与して、ローカルセクタID付き命令を生成する。したがって、専用領域獲得要求が対象とする対象処理32内の命令は、ローカルセクタIDに基づいて、割り当てられた専用キャッシュ領域を利用してメモリアクセスを実行する。ただし、専用領域獲得要求30と専用領域開放要求34との間の全ての命令が、専用キャッシュ領域を利用する必要があるわけではない。そのような命令には、ローカルセクタIDを例えば「0」に指定することで、専用キャッシュ領域ではなく通常の命令と同様に共用キャッシュ領域を利用してメモリアクセスを実行させることができる。

20

【0040】

プロセスPR1内には、上記の専用領域獲得要求30が複数発行されることがある。そして、専用領域獲得要求30に対応する専用領域解放要求34は、それぞれ、専用領域を開放するに適切な位置で発行される。

【0041】

図3に戻り、開発環境下で、専用領域獲得要求の特性調査を実行する(S3)。本実施の形態では、キャッシュメモリ内の一部の領域を専用キャッシュ領域として特定のプロセスまたはプロセス内の一部の処理に動的に割り当てる。動的に割り当てる目的は、適切に専用キャッシュ領域を割り当てることによりキャッシュメモリの単位時間当たりのキャッシュヒット数の総計を向上させ、キャッシュミスヒット数をできるだけ抑制することにある。そのため、専用領域獲得要求に対して専用キャッシュ領域を割り当てるべきか否かを判断するための情報が必要になる。

30

【0042】

各プロセスの専用領域獲得要求が対象とする処理には、様々なタイプがある。例えば、次のようなものである。

「プロセス全体が専用領域獲得要求の対象処理であり、そのプロセス全体で利用する専用キャッシュ領域を要求するもの」

「専用領域獲得要求の対象処理がプロセス内の一部の処理であるもの」

「専用領域獲得要求の対象処理がメモリの広範囲にアクセスしてキャッシュ領域内のデータを追い出して攪乱するもの」

40

「専用領域獲得要求の対象処理がキャッシュ領域にデータを固定的に記憶させ、キャッシュミスヒットがほとんど発生しないもの」

「対象処理が超短時間で終了するもの」

「対象処理が専用キャッシュ領域利用時と共用キャッシュ領域利用時とでキャッシュヒット率が大きく異なるもの」

これらを事前に調査しておくことで、本番稼働時において専用領域獲得要求に対して専用キャッシュ領域を割り当てるべきかを適切に判断することができる。

【0043】

そこで、工程S3において、開発環境下でコンパイルされた機械語のアプリケーションブ

50

ログラム内の専用領域獲得要求に専用キャッシュ領域の割当を許可した場合に、どの程度キャッシュヒット率が向上するかの特性調査を実行する。具体的には、特性調査対象の専用領域獲得要求だけに専用キャッシュ領域を割り当て、且つ、その専用キャッシュ領域のサイズを0から徐々に大きくしながら、それぞれの場合の(1)短時間当たりのメモリアクセス回数(メモリアクセス頻度)と、(2)専用キャッシュ領域内のキャッシュヒット率とを取得する。

#### 【0044】

図5は、上記の特性調査で得られたキャッシュサイズ対キャッシュヒット率の特性例である。横軸のキャッシュサイズ $x$ を0から増大させた場合のそれぞれのキャッシュサイズでのキャッシュヒット率 $H(x)$ が縦軸に対応されている。一般的に、このキャッシュサイズ対キャッシュヒット率の特性は、キャッシュサイズが小さい範囲ではキャッシュサイズの増加はそれほどキャッシュヒット率の上昇に貢献しないが、あるキャッシュサイズの範囲ではキャッシュサイズの増加でキャッシュヒット率が急激に上昇し、そして、あるキャッシュサイズより大きいサイズではキャッシュヒット率の上昇はあまりない。

10

#### 【0045】

したがって、図5中の原点から延びる直線(一点鎖線)がキャッシュヒット率特性 $H(x)$ に接する点のキャッシュサイズ $x_1$ が、キャッシュサイズの増大がキャッシュヒット率の向上に最も寄与する基準点となる。

#### 【0046】

なお、図5のキャッシュヒット率を調査する段階で、同時に単位時間当たりのメモリアクセス回数(メモリアクセス頻度)とキャッシュヒット数とキャッシュミスヒット数が取得される。

20

#### 【0047】

図20は、図3の工程S2とS3のサーバシステム内での処理を示す図である。また、図21は、図20のサーバシステム内の処理に対応するフローチャート図である。これらの図を参照して、図3の工程S2,S3の処理を説明する。

#### 【0048】

まず、サーバシステムの入力部からアプリケーションプログラムのソースコードが入力される(S201)。そして、コンパイラ21が、このアプリケーションプログラムのソースコードをコンパイルしてオブジェクトコード(機械語)に変換する(S202)。この機械語命令への変換については、図4で説明したとおりである。

30

#### 【0049】

次に、開発環境下でサーバシステムの入力部からシステム稼働が起動される(S301)。これにตอบสนองして、OS20内のプロセス管理部202がプロセス立ち上げ処理を実施してシステム稼働処理を行う(S302)。

#### 【0050】

この開発環境下でのプロセス稼働中に、OS20内のキャッシュ割当管理部204は、特性調査対象のアプリケーションプログラムに含まれている各専用領域獲得要求に対して、その要求を許可し、割り当てた専用キャッシュ領域を使用しながら命令を実行して、メモリアクセス頻度とキャッシュヒット率を計測する。この場合、キャッシュ割当管理部204は、専用キャッシュ領域のサイズを0から最大値まで変化させながらそれぞれのサイズでのメモリアクセス頻度とキャッシュヒット率を計測する(S303)。これにより、図5に示した特性グラフを生成する。

40

#### 【0051】

特性調査の開始に当たり、キャッシュ割当管理部204は、特性調査対象の専用領域獲得要求に対応したプロセスのセクタID変換テーブル131を読み出して、調査対象のローカルセクタIDに、割り当てる専用キャッシュ領域のキャッシュセクタIDを対応付け、その専用キャッシュ領域に設定したサイズを割り当てる(S304)。また、調査対象以外の専用領域獲得要求に対応したローカルセクタIDに対しては、共用キャッシュ領域を示すキャッシュセクタID=0を対応付ける。その結果、キャッシュメモリ130には、調査対象の処理

50

のローカルセクタIDに対してのみ指定されたサイズの専用キャッシュ領域が生成される (S305)。

【 0 0 5 2 】

プロセス管理部 2 0 2 は、特性調査の開始の準備が整ったところで、調査対象プロセスの起動を行い (S306)、アプリケーションプログラム 2 2 内の調査対象プロセスが実行される (S307)。この調査対象プロセスの実行中は、割り当てられた専用キャッシュ領域を使用してメモリアクセスが実行される。

【 0 0 5 3 】

全プロセスの全専用領域獲得要求の特性調査が終了するまで、処理S303-S307が繰り返し実行される。そして、全てが終了すると、キャッシュ割当管理部 2 0 4 は、調査対象プロセスのメモリアクセス頻度とキャッシュヒット率を記録して、特性調査を終了する (S308)。

10

【 0 0 5 4 】

図 3 に戻り、本番環境下でシステムを稼働すると (S10)、次のような処理が実行される。OS12は、バックグラウンドで共用キャッシュ領域と専用領域獲得要求で獲得された専用キャッシュ領域のキャッシュヒット率を調査する (S11)。この調査で、共用キャッシュ領域と専用キャッシュ領域でのメモリアクセス頻度なども取得される。具体的には、各領域へのアクセス関数のカウンタや、キャッシュヒット回数のカウンタなどを利用して、調査する。

【 0 0 5 5 】

20

OS20は、上記の調査結果に基づいて、共用キャッシュ領域と専用キャッシュ領域の適正比率や、専用領域獲得要求毎の実キャッシュ有効利用度などを算出する。共用・専用領域の適正比率は、後で詳述するが、両領域へのメモリアクセス数の比率により算出される。また、実キャッシュ有効利用度は、専用キャッシュ領域を利用している場合と共用キャッシュ領域を利用している場合のキャッシュヒット率の差分に、メモリアクセス頻度を乗算し、さらに、優先度を乗算して求められる。後に詳述する。

【 0 0 5 6 】

そして、CPU12は、図 4 で説明したローカルセクタID付きの命令を実行する (S13)。この命令の実行については後で詳述する。

【 0 0 5 7 】

30

さらに、OS20は、キャッシュ領域の動的割当処理を行う (S14)。キャッシュメモリ内の領域の動的割当処理では、第 1 に、専用領域獲得要求が発行されたときに (S15)、OS20内のキャッシュ割当管理部 2 0 4 が専用領域獲得要求について専用キャッシュ領域を割り当てるべきか否かの判定処理を行い、専用キャッシュ領域か、共用キャッシュ領域か、または攪乱処理キャッシュ領域かを割り当てる (S16)。

【 0 0 5 8 】

第 2 に、定期的な割当見直しタイミングで (S17)、OS20内のキャッシュ割当管理部 2 0 4 が、定期割当見直し処理を実行する (S18)。この定期割当見直し処理には、( 1 ) 割当済み専用領域獲得要求の割当済み専用キャッシュ領域と未割当専用領域獲得要求が要求する未割当専用キャッシュ領域とを入れ替えるべきか否かの判定処理、( 2 ) 共用キャッシュ領域の一部を未割当専用領域獲得要求に専用キャッシュ領域として割り当てるべきか否かの判定処理、( 3 ) その逆の、割当済み専用キャッシュ領域を開放して共用キャッシュ領域に変更すべきか否かの判定処理などが含まれる。そして、定期割当見直し処理S18では、キャッシュ割当管理部 2 0 4 が、判定結果に応じて、専用キャッシュ領域の置き換えや、専用キャッシュ領域の割当や開放を行う。

40

【 0 0 5 9 】

第 3 に、専用領域開放要求が発行されたときに (S19)、キャッシュ割当管理部 2 0 4 が、要求された専用キャッシュ領域を開放して共用キャッシュ領域の一部にする開放要求処理を行う (S20)。

【 0 0 6 0 】

50

上記の工程S11-S20が，システム稼働が終了するまで繰り返される。

【 0 0 6 1 】

図 6 は，キャッシュ領域の構成と専用領域獲得要求についての割当について示す図である。キャッシュメモリ 1 3 0 は，共用キャッシュ領域130\_1と，動的割当領域である専用キャッシュ領域130\_2とを有する。専用キャッシュ領域130\_2には，専用領域獲得要求に対して割り当てられる複数の専用キャッシュ領域と攪乱処理キャッシュ領域とが含まれる。つまり，攪乱処理キャッシュ領域は，一種の専用キャッシュ領域である。

【 0 0 6 2 】

キャッシュメモリ 1 3 0 は，共用キャッシュ領域と専用キャッシュ領域にセクタ分割されていて，各領域にはセクタIDが 0 から順に割り振られている。共用キャッシュ領域はセクタID=0である。共用キャッシュ領域は，複数のプロセスやプロセス内の対象処理が共通で使用するキャッシュ領域である。一方，専用キャッシュ領域は，判定により許可されたサイズを有し，割り当てられた専用領域獲得要求の対象プロセスが占有して使用するキャッシュ領域である。さらに，専用キャッシュ領域の一形態である攪乱処理キャッシュ領域は，専用領域獲得要求の対象処理のメモリアクセスが広範囲のアドレスに対して行われ，共用キャッシュ領域に割り当てるとその共用キャッシュ領域内の他のデータを追い出して攪乱してしまう場合に割り当てられる。したがって，複数の専用領域獲得要求に対して攪乱処理キャッシュ領域が割り当てられる場合がある。

【 0 0 6 3 】

専用領域獲得要求は，前述したとおり，アプリケーションプログラムAPL内のプロセス毎にあるいはプロセス内の対象処理毎に発行され，判定結果に応じて，専用領域獲得要求に専用キャッシュ領域が割り当てられる。したがって，必ずしも専用キャッシュ領域が割り当てられるとは限らず，共用キャッシュ領域が割り当てられて未割当専用領域獲得要求として専用キャッシュ領域の割当まで待たされる場合がある。キャッシュメモリ 1 3 0 は，サイズや分割可能数に上限があるからである。

【 0 0 6 4 】

図 6 に示した例では，プロセス 1 の専用領域獲得要求要求 0 ，プロセス 2 の専用領域獲得要求 0 ， 3 は，判定の結果共用キャッシュ領域130\_1が割り当てられている。また，プロセス 2 の専用領域獲得要求 1 ， 2 ， 4 には，それぞれ専用キャッシュ領域であるセクタID = 1 ， 2 ， 3 が割り当てられている。また，プロセス 3 の専用領域獲得要求 0 に対しては，判定の結果，攪乱処理キャッシュ領域であるセクタID = 4 が割り当てられている。

【 0 0 6 5 】

図 2 2 は，図 3 の工程S13，S16，S19のサーバシステム内での処理を示す図である。また，図 2 3 は，図 2 2 のサーバシステム内の処理に対応するフローチャート図である。これらの図を参照して，専用領域獲得要求での専用キャッシュ領域の割当と，その専用キャッシュ領域を利用するメモリアクセスと，専用領域開放要求での専用キャッシュ領域の開放とを説明する。

【 0 0 6 6 】

[ 専用領域獲得要求での専用キャッシュ領域の割当S15,S16 ]

まず，アプリケーションプログラム 2 2 のプロセス内で，専用領域獲得要求が発行される (S400)。この専用領域獲得要求にはローカルセクタIDが付与されている。これにตอบสนองして，OS20内のキャッシュ割当管理部 2 0 4 は，専用キャッシュ領域の割当を許可するかどうかの判定を実行する (S401)。この判定処理については，後述する。そして，割当が許可されると，キャッシュ割当管理部 2 0 4 は，割り当てる専用キャッシュ領域のキャッシュサイズをキャッシュセクタ情報設定部 1 3 2 に設定するとともに，キャッシュシステム 1 3 内のセクタID変換テーブル 1 3 1 に，ローカルセクタIDに対応するキャッシュセクタIDの欄に，割り当てた領域のセクタIDを設定する (S402)。

【 0 0 6 7 】

図 7 は，変換テーブルの一例を示す図である。図 7 中には，プロセスID = 2 ，ローカ

10

20

30

40

50

ルセクタID = 2の専用領域獲得要求30に対して、プロセスID = 2の変換テーブル131内のローカルセクタID = 2に対してキャッシュセクタID = 2の専用キャッシュ領域のセクタIDが記入されている。

【0068】

[専用キャッシュ領域を利用するメモリアクセスS13]

アプリケーションプログラム22のプロセスが実行される場合、使用する専用キャッシュ領域を識別するローカルセクタID付きの命令(機械語命令)が発行される(S403)。そして、CPU12は、そのローカルセクタID付きの命令を解釈して実行する(S403-1)。この実行に際し、CPU12は、セクタID変換テーブル131を参照してローカルセクタIDに対応するキャッシュセクタIDを取得する(S403-2)。すなわち、図7中にS403-2として示す通りである。

10

【0069】

そして、キャッシュアドレス変換部133が、メモリアクセスアドレスをキャッシュメモリ内のアドレスに変換して、キャッシュメモリ130内のキャッシュセクタID = 2の専用領域獲得要求キャッシュ領域にキャッシュアクセスする(S403-3)。ここで、キャッシュミスヒットが発生すると、バスインターフェース14を介してメインメモリ16にアクセスする(S403-4)。キャッシュヒットの場合は、専用キャッシュ領域内のデータが読み出されまたは専用キャッシュ領域にデータが書き込まれる。

【0070】

以上のように、専用キャッシュ領域が割り当てられている場合は、プロセス内の命令が割り当てられている専用キャッシュ領域を使用してメモリアクセスする。

20

【0071】

[専用領域開放要求での専用キャッシュ領域の開放]

次に、アプリケーションプログラム22のプロセスで、専用領域開放要求が発行されると(S404)、キャッシュ割当管理部204は、専用キャッシュ領域の開放処理を実行する(S405)。具体的には、セクタID変換テーブル内のローカルセクタIDに対応するキャッシュセクタIDを共用キャッシュ領域のID=0に変更する(S406)。すなわち、図7中にS406として示す通りである。また、キャッシュセクタ情報設定部132内の解放するキャッシュセクタIDのサイズが0に変更される。

【0072】

30

[本実施の形態のキャッシュ領域の動的割当処理の詳細]

以下、キャッシュ領域の動的割当処理の詳細について説明する。動的割当処理は、図3で示されるとおり、専用領域獲得要求に対する判定処理S16と、定期的な割当見直しタイミングでの判定処理S18とがある。それらについて順に説明する前に、キャッシュ領域の割当判定の指標を説明する。

【0073】

[キャッシュ領域の割当判定の指標]

[実キャッシュヒット数]

キャッシュメモリを共用キャッシュ領域と複数の専用キャッシュ領域(攪乱処理キャッシュ領域を含む)とに分割する目的は、キャッシュメモリ全体の単位時間当たりのキャッシュヒット数を最大化することにある。この単位時間当たりのキャッシュヒット数は、次の式により求めることができる。

40

単位時間当たりのキャッシュヒット数 = キャッシュヒット率 \* メモリアクセス頻度 (1)

したがって、専用領域獲得要求に専用キャッシュ領域を割り当てた場合に、専用領域獲得要求の対象処理のキャッシュヒット数が高ければ、その要求に専用キャッシュ領域を割り当てることで全体のキャッシュヒット数を高めることができる。そして、キャッシュヒット率が高いがメモリアクセス頻度が低い対象処理に対して専用キャッシュ領域を割り当てるよりも、キャッシュヒット率が多少低くてもメモリアクセス頻度が高い対象処理に専用キャッシュ領域を割り当てた方が、式(1)を大きくすることができ、メインメモリに

50

アクセスする時間を短縮する回数が多くなり、全体としてメモリアクセスが短縮できる。

【0074】

一方、アプリケーションプログラムの処理には優先度があり、特定のアプリケーションプログラムや特定のプロセスを特に高速化したいなどの要求がある。この要求が高いほど処理の優先度が高い。したがって、実際には以下の実キャッシュヒット数を最大化することがキャッシュ領域の動的割当の目的になる。

実キャッシュヒット数 = (処理のキャッシュヒット数 \* 処理優先度) (2)

すなわち、上記式(2)において、は、専用領域獲得要求の対象とする処理のキャッシュヒット数にその処理の優先度を乗算したものを、全処理分累積することを意味する。

【0075】

[メモリアクセスの局所性]

次に、専用領域獲得要求の対象の処理には、メモリアクセスの局所性が高いものと低いものがある。メモリアクセスの局所性は、メモリアクセスのアドレスの領域が狭い範囲に集中していれば局所性は高く、広い範囲に分散していれば局所性は低くなる。対象処理のメモリアクセスの局所性が低いと、キャッシュヒット率は低くなる。ただし、キャッシュサイズを大きくすることで局所性が高まり、キャッシュヒット率が高くなる場合がある。

【0076】

図5に示されたキャッシュサイズに対するキャッシュヒット率の特性曲線の一例をみれば、上記の局所性とキャッシュサイズとの関係が理解できる。この例では、キャッシュサイズが小さい場合は、メモリアクセスのアドレス範囲に比較してキャッシュサイズが小さいため、キャッシュヒット率は向上していない。しかし、キャッシュサイズを増大させることで、メモリアクセスのアドレス範囲に比較してキャッシュサイズが追いつき、キャッシュヒット率が向上している。そして、キャッシュサイズを基準値×よりさらに大きくしてもキャッシュヒット率は向上していない。

【0077】

さらに、メモリアクセスの局所性が極端に高い場合は、キャッシュサイズを増大しても、キャッシュヒット率を向上させることができず、キャッシュミスヒット率が高止まりする。このようなキャッシュミスヒット率が高い処理を、共用キャッシュ領域に割り当てると、キャッシュミスヒットが頻発して他の処理のデータを追い出してしまうことになる。このような処理を攪乱処理と称する。

【0078】

したがって、このような攪乱処理に対しては、共用キャッシュ領域を割り当てるよりも専用のキャッシュ領域を割り当てることが、共有キャッシュ領域のキャッシュヒット率向上に有益である。ただし、このような攪乱処理は、専用キャッシュ領域を割り当てられたとしてもキャッシュミスヒットを頻発するので、式(2)の実キャッシュヒット率は上がらない。そこで、攪乱処理には、攪乱処理用のキャッシュ領域を割り当て、しかも、複数の攪乱処理に、共通の攪乱処理用キャッシュ領域を割り当てて、それらの処理を他のキャッシュ領域から隔離することが有効である。

【0079】

[キャッシュ有効利用度、実キャッシュ有効利用度、キャッシュ攪乱力]

次に、専用領域獲得要求に対して専用のキャッシュ領域を割り当てるか否かの判断の指標として、キャッシュ有効利用度、実キャッシュ有効利用度、キャッシュ攪乱力について説明する。

【0080】

キャッシュ有効利用度は、専用のキャッシュ領域を割り当てた場合のキャッシュヒット数よりも、共用キャッシュ領域を割り当てた場合のキャッシュヒット数がどのくらい向上するかを示す度合いである。すなわち、キャッシュ有効利用度は次の式(3)のとおりである。

キャッシュ有効利用度 = { (専用領域時キャッシュヒット率 - 共用領域時キャッシュヒット率) \* メモリアクセス頻度 } / 使用専用キャッシュ領域のサイズ (3)

10

20

30

40

50



つまり、キャッシュ有効利用度（度合い）とは、共用キャッシュ領域を割り当てた場合よりも専用キャッシュ領域を割り当てた場合のほうがどれくらいキャッシュヒット率が增大するかを示すキャッシュヒット率の差分に、メモリアクセス頻度（単位時間当たりのメモリアクセス数）を乗算したキャッシュヒット数の増大を、単位キャッシュ領域当たりで示す指標である。

#### 【 0 0 8 1 】

この場合、専用領域時キャッシュヒット率とメモリアクセス頻度は、事前に行う特性調査結果から取得することができ、共用領域時キャッシュヒット率は、本番稼働時に取得することができる。また、既に専用キャッシュ領域が割り当てられている場合は、その割当済み専用キャッシュ領域について、キャッシュヒット率を取得することが可能である。したがって、割当済み専用キャッシュ領域については、割り当てられたキャッシュサイズに対する実際のキャッシュヒット率を取得し、事前の特性調査での同じキャッシュサイズに対するキャッシュヒット率との比率を、事前特性調査で得た全キャッシュサイズに対するキャッシュヒット率に乘算して、推定キャッシュヒット率を求めることができる。この推定キャッシュヒット率については、後で再度詳述する。

#### 【 0 0 8 2 】

次に、アプリケーションプログラムのプロセスには前述した優先度が存在し、優先度が高い処理ほど、優先的にキャッシュヒット数を向上させることが必要になる。そこで、以下の式（４）に示すとおり、上記のキャッシュ有効利用度に、対応する処理の優先度を乗算することで、実キャッシュ有効利用度を算出することができる。

実キャッシュ有効利用度 = キャッシュ有効利用度 \* 対応処理優先度 （４）

式（３）（４）において、専用領域時キャッシュヒット率を図５の $H(x)$ に、共用領域時キャッシュヒット率を $BASEH$ に、メモリアクセス頻度を $MACCESS$ に、使用専用キャッシュ領域のサイズを $x$ に、対応処理優先度を $PRIORITY$ に置き換えると、式（４）の実キャッシュ有効利用度は、次の通りになる。

実キャッシュ有効利用度 =  $[ \{ H(x) - BASEH \} * MACCESS * PRIORITY ] / x$  （４）

この実キャッシュ有効利用度についても、割当済み専用キャッシュ領域に対しては上記の推定キャッシュヒット率を利用することも可能である。

#### 【 0 0 8 3 】

式（３）のキャッシュ有効利用度と、それに優先度を乗じた式（４）の実キャッシュ有効利用度とは、主に、専用領域獲得要求に対して専用領域を割り当てるべきか否かの判断指標として利用される。

#### 【 0 0 8 4 】

次に、前述のメモリアクセスの局所性が極端に高いほど、他の処理と同じ共用キャッシュ領域に割り当てられると他の処理のデータを追い出す力が強くなる。この追い出す力をキャッシュ攪乱力とする。このキャッシュ攪乱力は、以下の式（５）に示したとおり、キャッシュミスヒット率（１ - キャッシュヒット率）にメモリアクセス頻度を乗算して算出することができる。

キャッシュ攪乱力 =  $( 1 - \text{キャッシュヒット率} ) * \text{メモリアクセス頻度}$  （５）

キャッシュ攪乱力は、キャッシュ有効利用度とは相反する値となる。そして、このキャッシュ攪乱力は、開発環境下での特性調査でキャッシュヒット率とメモリアクセス頻度が取得されるので、予め処理毎にまたはプロセス毎に算出しておくことができる。このキャッシュ攪乱力がある基準値以上の処理は、キャッシュ攪乱処理と見なして、共用キャッシュ領域や他の専用キャッシュ領域から隔離して、攪乱処理キャッシュ領域に割り当てることで、他の処理のキャッシュヒット率の低下を抑制することができる。

#### 【 0 0 8 5 】

キャッシュ攪乱力は、専用領域獲得要求に対して、専用領域を割り当てるべきか判定するときに、攪乱処理用の専用領域に割り当てるべきか否かの判断指標として利用される。

#### 【 0 0 8 6 】

[ 専用領域獲得要求判定処理S16（１） ]

図 8 は、図 3 に示した専用領域獲得要求に対する判定処理S16について第 1 の例のフローチャート図である。この第 1 の判定処理では、OS20内のキャッシュ割当管理部 2 0 4 が、専用領域獲得要求の対象処理のキャッシュ攪乱力（式（ 5 ））を算出して判定し（S30）、キャッシュ攪乱力がある第 1 の基準値より高い場合は（S31のYES）、その専用領域獲得要求に攪乱処理キャッシュ領域を割り当てる（S32）。

【 0 0 8 7 】

さらに、キャッシュ割当管理部 2 0 4 は、キャッシュ攪乱力が第 1 の基準値以下の場合（S31のNO）、その専用領域獲得要求の対象処理の実キャッシュ有効利用度（式（ 4 ））を算出して判定し（S33）、実キャッシュ有効利用度が第 2 の基準値より高い場合は（S34のYES）、その専用領域獲得要求に専用キャッシュ領域を割り当て（S35）、実キャッシュ有効利用度が第 2 の基準値以下の場合（S34のNO）、その専用領域獲得要求に共用キャッシュ領域を割り当てる（S36）。

【 0 0 8 8 】

図 9 は、専用領域獲得要求に専用キャッシュ領域を割り当てる場合を説明する図である。図 9 中（ 1 ）のキャッシュメモリ 1 3 0 の分割状態において、専用領域獲得要求 3 0 が発行されたとする。この場合、専用領域獲得要求の対象処理の実キャッシュ有効度を式（ 4 ）に基づいて算出する。そして、その実キャッシュ有効度が第 2 の基準値より高ければ、図 9 中（ 2 ）のキャッシュメモリ 1 3 0 のように、その専用領域獲得要求 3 0 に対してキャッシュセクタ ID = 2 の専用キャッシュ領域を割り当てる。その結果、この割り当てられた専用キャッシュ領域は、それ以降、割当済み専用領域と称し、割り当てられた専用領域獲得要求は割当済み専用領域獲得要求と称することとする。

【 0 0 8 9 】

さらに、割り当てる専用キャッシュ領域のサイズについて説明する。図 5 に示したとおり、処理のキャッシュヒット率は、キャッシュサイズを増大することで急激に上昇し、ある基準キャッシュサイズ× 1 で飽和する。そこで、実キャッシュ有効利用度は、図 5 のキャッシュサイズ対キャッシュヒット率の特性曲線の基準キャッシュサイズ× 1 でのキャッシュヒット率を元に算出することが望ましい。そして、そのようにして算出した実キャッシュ有効利用度が第 2 の基準値より大きい場合には、その専用領域獲得要求に対して割り当てる専用キャッシュ領域のサイズは、基準キャッシュサイズ× 1 に設定するのが望ましい。

【 0 0 9 0 】

このように、専用キャッシュ領域のキャッシュサイズを自動的に基準キャッシュサイズ× 1 に割り当てるようにすることで、最適なキャッシュサイズを割り当てることが可能になる。

【 0 0 9 1 】

図 9 によれば、実キャッシュ有効利用度が第 2 の基準値より高い専用領域獲得要求に次々と専用キャッシュ領域を割り当てていくと、キャッシュメモリ 1 3 0 が専用キャッシュ領域であふれることになる。そこで、後述する定期割当見直し処理（ 3 ）により、共用キャッシュ領域と専用キャッシュ領域とのバランスを適正化することで、キャッシュメモリが専用キャッシュ領域であふれることを回避することができる。

【 0 0 9 2 】

[ 専用領域獲得要求判定処理S16（ 2 ） ]

図 1 0 は、図 3 に示した専用領域獲得要求に対する判定処理S16について第 2 の例のフローチャート図である。この第 2 の判定処理では、OS20内のキャッシュ割当管理部 2 0 4 が、専用領域獲得要求の対象処理のキャッシュ攪乱力を算出して第 1 の基準値より大きい場合に攪乱処理キャッシュ領域を割り当てる（S30,S31,S32）ことは同じである。そして、キャッシュ割当管理部 2 0 4 は、キャッシュ攪乱力が第 1 の基準値以下の場合に、専用領域獲得要求の対象処理と割当済み専用領域の実キャッシュ有効利用度（望ましくは実キャッシュ有効利用値の増加率）を比較判定して（S37）、入れ替えることがより効率的である場合は（S37のYES）、専用領域獲得要求に専用領域を割り当て、割当済み専用領域を開放し

10

20

30

40

50

て、両者を入れ替える (S39)。逆に、入れ替えるのが効率的ではない場合は (S37のNO)、専用領域獲得要求に共用キャッシュ領域を割り当てる (S40)。

#### 【0093】

図11は、専用領域獲得要求と割当済専用領域獲得要求との間で専用キャッシュ領域を入れ替える処理を説明する図である。図11中(1)のキャッシュメモリ130のような分割状態において、専用領域獲得要求30が発行されると、その対象処理のキャッシュ攪乱力が第1の基準値以下の場合、割当済み専用領域と入れ替えて新たな専用領域を割り当てることで、キャッシュメモリ全体のキャッシュヒット数を増大できるか否かを判定する。

#### 【0094】

図11に示した例では、キャッシュセクタID=2の割当済専用領域の実キャッシュ有効利用度よりも、専用領域獲得要求の対象処理の実キャッシュ有効利用度のほうが、入れ替えに要するオーバーヘッド値を超えて高いため、割当済専用領域を開放して共用キャッシュ領域に移動し、代わりに、専用領域獲得要求に新たな専用キャッシュ領域を割り当てている。(付記9)

そこで、上記の比較判定S37について説明する。

#### 【0095】

まず、専用キャッシュ領域が既に割り当てられている場合と、未だ割り当てられていない場合とでキャッシュヒット率の取り扱いを異ならせる。割当済専用領域に対しては、推定キャッシュヒット率を採用し、未割当専用領域に対しては、事前調査で取得したキャッシュヒット率を採用する。

#### 【0096】

図12は、推定キャッシュヒット率を示す図である。図中、横軸はキャッシュサイズ、縦軸はキャッシュヒット率である。実線のH(x)は、事前調査で取得したキャッシュサイズに対するキャッシュヒット率の特性曲線である。それに対して、割当済み専用キャッシュ領域の場合は、割り当てられたキャッシュサイズ×2に対するキャッシュヒット率を、キャッシュシステム内のカウンタなどから取得することができる。

#### 【0097】

その場合は、キャッシュサイズ×2での現在のキャッシュヒット率SH(x2)と、事前調査時のキャッシュヒット率H(x2)との比率SH(x2)/H(x2)を、事前調査時の全キャッシュサイズに対するキャッシュヒット率H(x)に乘じることで、推定キャッシュヒット率SH(x)を得ることができる。即ち、次の式(6)の通りである。

$$SH(x) = \{ SH(x2)/H(x2) \} * H(x) \quad (6)$$

この推定キャッシュヒット率SH(x)を利用すれば、上記の式(4)から、割当済み専用キャッシュ領域の実キャッシュ有効利用度は、次の式(7)のように導くことができる。

$$CCn = [ \{ SH(x) - BASEH \} * MACCESSn * PRIORITYn ] / x \quad (7)$$

ここで、nは、専用領域獲得要求の対象処理を示す。したがって、式(7)はある一つの対象処理に対する推定キャッシュヒット率SH(x)を利用した実キャッシュ有効利用度である。そして、未割当の専用領域獲得要求に対する実キャッシュ有効利用度は、前述の式(4)のとおりであり、以下のようになる。

$$\text{実キャッシュ有効利用度}n = [ \{ H(x) - BASEH \} * MACCESSn * PRIORITYn ] / x \quad (4)$$

専用領域獲得要求にはあるサイズxの専用キャッシュ領域が割り当てられる。したがって、割当済み専用キャッシュ領域との入れ替えを判定するためには、単位領域当たりの実キャッシュ有効利用度の増加値(実キャッシュ有効利用度の増加率)を考慮することが有効である。これを利用することで、割当済専用キャッシュ領域のサイズを+1するか-1するかにより実キャッシュ有効利用度がどの程度変化するかを考慮することができる。

#### 【0098】

上記の実キャッシュ有効利用度の増加率CHHn(x)は、割り当てる専用キャッシュ領域のサイズをx-1からxに増加したときの実キャッシュ有効利用度の増加量であり、式(7)から次の式(8)のようになる。

10

20

30

40

50

$$CHHn(x) = \{ SH(x) - SH(x-1) \} * MACCESSn * PRIORITYn \quad (8)$$

式(8)からわかるとおり、式(7)のように領域サイズxで除算していないので、CHHn(x)は、実キャッシュ有効利用数の増加率である。以下、式(8)を実キャッシュ有効利用数の増加率と称する。

【0099】

専用領域獲得要求に専用領域が未割当の場合は、推定キャッシュヒット率SH(x)を利用できないので、事前調査でのキャッシュヒット率H(x)を利用して、上記式(8)の実キャッシュ有効数の増加率CHHn(x)は、次の式(9)のようになる。

$$CHHn(x) = \{ H(x) - H(x-1) \} * MACCESSn * PRIORITYn \quad (9)$$

本実施の形態の判定工程S37では、式(8)の実キャッシュ有効利用数の増加率を割当済み専用キャッシュ領域について算出した割当済み専用領域リストLIST-FREEを利用する。以下、この割当済み専用領域リストLIST-FREEについて説明する。

【0100】

図13は、割当済み専用領域リストLIST-FREEと未割当専用領域リストLIST-DEMANDとを示す図である。図13の右側が割当済み専用領域リストLIST-FREEを示す。キャッシュメモリ130において、割当済みの専用領域(キャッシュセクタID=1,2)を、共用キャッシュ領域(キャッシュセクタID=0)に入れ替える場合に、実キャッシュ有効利用数の増加率がどうなるかを式(8)によって算出しておき、その増加率が少ない順に割当済み専用領域リストLIST-FREEに登録しておく。増加率が少ないとは、割当済み専用領域を共用領域に単位領域だけ入れ替えた場合の実キャッシュヒット数の減少が少ないことを意味し、現在のキャッシュメモリの有効利用に寄与していないことを意味する。

【0101】

以上のように、図13の右側の割当済み専用領域リストLIST-FREEには、増加率が小さい順にその増加率(LIST-FREE.CHH)が並べられている。

【0102】

図13の左側が未割当専用領域リストLIST-DEMANDを示す。キャッシュメモリ130において、専用領域獲得要求に対して専用領域が未割当の場合には共用領域が割り当てられていて、さらに、専用領域獲得要求に対して一部のサイズだけ専用領域が割当られている場合には残りのサイズは共用領域に割り当てられているので、それらの専用領域に割り当てられていない未割当専用領域が、この未割当専用領域リストLIST-DEMANDに含まれる。

【0103】

そして、それらの未割当専用領域を共用領域から専用領域に入れ替えた場合に、実キャッシュ有効利用数の増加率がどうなるかを式(9)によって算出しておき、その増加率が多い順に未割当専用領域リストLIST-DEMANDに登録しておく。増加率が多いとは、未割当専用領域を共用領域から専用領域に入れ替えた場合の実キャッシュヒット数の増加が多いことを意味し、入れ替えることによりキャッシュメモリの有効利用に大きく寄与することを意味する。

【0104】

以上のように、図13の左側の未割当専用領域リストLIST-DEMANDには、増加率が多い順にその増加率(LIST-DEMAND.CHH)が並べられている。

【0105】

そこで、本実施の形態の判定工程S37では、専用領域獲得要求30に対して専用領域を割り当てた場合の実キャッシュ有効利用数の増加と、割当済み専用領域リストLIST-FREEに登録されている割当済み専用領域の実キャッシュ有効利用数の減少の差分が、入れ替えに要するオーバーヘッド値を超える場合は、入れ替えして専用領域獲得要求に専用領域を割り当てるべきと判断する。

【0106】

この判断は、次の式(10)を満たすか否かにより行われる。

$$y * CCn(y) - (m-1-y)LIST-FREE.m.CHH > OVH2 * \{ (m-1-y)LIST-FREE.m.PRIORITY \} \quad (10)$$

10

20

30

40

50

ここで、 $y$  は専用領域のサイズ  $y$  を、 $m$  はリスト内の番号を示し、LIST-FREEm.CHH は実キャッシュ有効利用数の増加率を示し、OVH2は入れ替えに要するオーバーヘッド値を示し、LIST-FREEm.PRIORITYはその処理の優先度を示す。つまり、オーバーヘッド値OVH2が優先度で重み付けされている。

#### 【0107】

OS20内のキャッシュ割当部204は、 $y=1$ から式(10)を判定していき、式(10)の条件を満たす最大のサイズ  $y$  を、専用領域獲得要求に割り当てる専用領域のサイズとする。すなわち、図13に示した専用領域獲得要求30と割当済専用領域リストLIST-FREEとの間の矢印で示した入れ替えを行う。その結果、サイズ  $y$  に対応する割当済専用領域が共用領域に開放される。どのサイズ  $y$  でも式(10)を満たさない場合は、専用領域獲得要求には専用領域を割り当てない。

10

#### 【0108】

図12に示されるとおり、キャッシュヒット率の特性は、キャッシュサイズをある程度大きくすると急激に増加する傾向にある。したがって、新たに割り当てる専用領域のサイズ  $y$  を1から大きくすることで、式(10)の条件を満たすサイズ  $y$  が見つかることがある。特に、サイズ  $y$  が基準値  $\times 1$  の付近で式(10)の条件が満たされることがある。

#### 【0109】

式(10)の条件を満たすキャッシュサイズ  $y$  と、専用領域獲得要求30が要求するキャッシュサイズとの関係は、例えば、次の通りである。条件を満たすキャッシュサイズ  $y$  が要求するキャッシュサイズ以下であれば、キャッシュサイズ  $y$  から要求するキャッシュサイズまでのサイズを割り当てることでキャッシュメモリの効率を向上することができる。一方、キャッシュサイズ  $y$  が要求するキャッシュサイズを超えている場合は、式(10)の条件は満たされないと判定して専用領域を割り当てないのが好ましい。

20

#### 【0110】

図24は、本専用領域獲得要求判定処理S16(2)のフローチャート図である。図10と同じ処理には同じ番号を与えている。まず、キャッシュ割当管理部204は、キャッシュ攪乱力が第1の基準値を超えるか判定し(S30,31)、超えていれば、専用の攪乱処理キャッシュ領域を割り当てる(S32)。その結果、セクタID変換テーブル131で、専用領域獲得要求のローカルセクタIDに、攪乱処理キャッシュ領域のキャッシュセクタIDが紐付けられる(S32-1)。

30

#### 【0111】

キャッシュ割当管理部204は、キャッシュ攪乱力が第1の基準値を超えていなければ、専用キャッシュ領域を割り当てるべきか否かの判定を、式(10)を満たすサイズ  $y$  が存在するか否かで行う(S37,38)。キャッシュ割当管理部204は、条件を満たさない場合は、専用領域獲得要求に対して共用領域を割り当てて(S40)、条件を満たす場合は、専用領域獲得要求に対してサイズ  $y$  の専用領域を割り当てることをハードウェアに指示する(S39)。

#### 【0112】

ハードウェアでは、キャッシュ割当管理部204が、キャッシュセクタ情報設定部132に、新規に割り当てる専用キャッシュ領域とそのサイズ  $y$  を設定する(S39-1)。また、入れ替え対象の割当済専用領域のキャッシュセクタIDの削除やサイズ変更も同時に行う(S39-1)。そして、キャッシュ割当管理部204が、セクタID変換テーブル131において、専用領域獲得要求のローカルセクタIDに、割り当てた専用キャッシュ領域のキャッシュセクタIDが紐付けられる(S39-2)。

40

#### 【0113】

##### [ 定期割当見直し処理S18 ]

次に、図3の定期割当見直し処理S18について3つの例を説明する。専用領域獲得要求に回答して、OS20内のキャッシュ割当管理部204は、専用キャッシュ領域を割り当てるべきか否か、割当済専用領域と入れ替えすべきか否かを判定した。

#### 【0114】

50

これに対して、キャッシュ割当管理部204は、定期的に、例えば2秒毎に、専用領域獲得要求に割り当てた割当済専用領域と、専用領域獲得要求に未だ割り当てていない未割当専用領域とを入れ替えるべきか否かの判定(例1, 例2)と、共用領域と全専用領域とのサイズの比率がアクセス頻度の比率に基づく理想値とずれている場合に、共用領域を専用領域として未割当の専用領域獲得要求に新たに割り当てるか、または、割当済専用領域を開放して共用領域に入れ替えるかの判定(例3)とを行う。これらの判定において、前述の割当済専用領域リストLIST-FREEと未割当専用領域リストLIST-DEMANDを利用する。

#### 【0115】

[ 定期割当見直し処理S18(1) ]

図14は、定期割当見直し処理S18の第1の例のフローチャート図である。図15は、定期割当見直し処理S18の第1の例を説明するための図である。図14において、定期的割当見直し処理S18では、キャッシュ割当管理部204は、未割当専用領域と割当済専用領域の入れ替え判定を行う(S50)。この判定処理では、図13で説明した未割当専用領域リストLIST-DEMAND内の未割当専用領域の実キャッシュ有効利用数の増加率CHHn(x)と、割当済専用領域リストLIST-FREE内の割当済専用領域の実キャッシュ有効利用数の増加率CHHn(x)とを比較する(S51)。比較対象は、例えば、図15に矢印で示されるとおり、未割当専用領域リストLIST-DEMAND内の最上位にある最大の増加率と、割当済専用領域リストLIST-FREE内の最上位にある最小の増加率とを比較し、さらに、下位にある増加率どうしを比較する。

#### 【0116】

そして、以下の式(11)の条件を満たす場合は、入れ替えで効率がアップすると見なすことができる(S52)。

$$\text{LIST-DEMAND}_m.\text{CHH} - \text{LIST-FREE}_m.\text{CHH} > \text{OVH2} * \text{LIST-FREE}_m.\text{PRIORITY} \quad (11)$$

ここで、mはリストの順番を示し、OVH2は入れ替えに要するオーバーヘッド値を示し、LIST-FREE<sub>m</sub>.PRIORITYは割当済専用領域リストLIST-FREE内の処理の優先度を示す。つまり、比較対象の未割当専用領域の増加率から割当済専用領域の増加率への差分が、入れ替えに伴うオーバーヘッド値OVH2を優先度で重み付けした値を超える場合は、入れ替えで効率がアップすると見なされる。

#### 【0117】

式(11)の条件が、両リストの先頭(m=1)から順番に判定され、この条件が満たされなくなる順番まで繰り返し判定される。ただし、式(11)には優先度LIST-FREE<sub>m</sub>.PRIORITYが算入されているので、右辺の値は優先度に応じて変動する。したがって、未割当専用領域リストLIST-DEMAND内の最上位にある最大の増加率CHHと比較される対象は、割当済専用領域リストLIST-FREE内の最上位の増加率CHHだけでなく、最上位から順に下位の増加率CHHも含まれる。式(11)を満たす組み合わせが見つければ、それらが入れ替え対象になる。そして、未割当専用領域リストLIST-DEMAND内の次の増加率CHHについて、上記と同じ比較が行われ、式(11)を満たす組み合わせが検出される。

#### 【0118】

入れ替えで効率がアップすると見なされる場合は、キャッシュ割当管理部204は、未割当専用領域と割当済専用領域とを入れ替える処理を行う(S53)。具体的には、キャッシュセクタ情報設定部132に未割当専用領域に新たに専用領域を割り当てさせ、割り当てるサイズを登録させ、セクタID変換テーブル内の未割当専用領域獲得要求のローカルセクタIDに新たに割り当てる専用領域のキャッシュセクタIDを紐づける。逆に、キャッシュセクタ情報設定部132に割当済専用領域の登録を抹消させ、セクタID変換テーブル内の割当済専用領域獲得要求のローカルセクタIDに紐づけられているキャッシュセクタIDを共用領域のIDに変更する。

#### 【0119】

そして、入れ替え処理を行うと、入れ替えられた専用領域の増加率は、未割当専用領域リストLIST-DEMANDから割当済専用領域リストLIST-FREEに移動させ、同時に、割当済専用領域リストLIST-FREEから未割当専用領域リストLIST-DEMANDに移動させる。さらに、各リ

スト内で増加率CHHの値に基づいて、ソートし直しておく。

【 0 1 2 0 】

入れ替えで効率がアップすると見なされない場合は、キャッシュ割当管理部 2 0 4 は、未割当専用領域と割当済専用領域とを入れ替える処理は行わない (S54)。

【 0 1 2 1 】

[ 定期割当見直し処理S18 ( 2 ) ]

図 1 6 は、定期割当見直し処理S18の第 2 の例のフローチャート図である。図 1 7 は、定期割当見直し処理S18の第 2 の例を説明するための図である。図 1 6 において、定期的割当見直し処理S18では、キャッシュ割当管理部 2 0 4 は、未割当専用領域と割当済専用領域の入れ替え判定を行う (S60)。この判定処理では、図 1 7 に示したとおり、未割当専用領域リストLIST-DEMAND内の先頭の未割当専用領域の実キャッシュ有効利用数の増加率CHH<sub>1</sub>(x)の専用キャッシュ領域サイズxを増加させながら、割当済専用領域リストLIST-FREE内の割当済専用領域の実キャッシュ有効利用数の増加率CHH<sub>m</sub>(x)を先頭から増加させながら、両者の累積値の比較をする (S61)。

【 0 1 2 2 】

この判断S61は、次の式 ( 1 2 ) を満たすか否かにより行われる。

$$(x=1-y)LIST-DEMAND_1.CHH(x) - (m=1-y)LIST-FREE_m.CHH > OVH2 * \{ (m=1-y)LIST-FREE_m.PRIORITY \} \quad (12)$$

ここで、y は専用領域のサイズ y を、m はリスト内の番号を示し、LIST-DEMAND<sub>1</sub>.CHH、LIST-FREE<sub>m</sub>.CHH は実キャッシュ有効利用数の増加率を示し、OVH2は入れ替えに要するオーバーヘッド値を示し、LIST-FREE<sub>m</sub>.PRIORITYはその処理の優先度を示す。ここでも、オーバーヘッド値OVH2が優先度で重み付けされている。

【 0 1 2 3 】

OS20内のキャッシュ割当部 2 0 4 は、式 ( 1 2 ) の判定をy=1から増加しながら行い、式 ( 1 2 ) の条件を満たす最大のサイズyを検出する。すなわち、実キャッシュ有効利用数の増加率CHH<sub>1</sub>(x)は、単位サイズ増加させたときの実キャッシュ有効利用数の増加量である。そして、図 1 2 に示すとおり、専用キャッシュ領域のサイズxを増加させるほどキャッシュヒット率が上昇する傾向にあるので、未割当専用領域の実キャッシュ有効利用数の増加率CHH<sub>1</sub>(x)が最大化できる専用キャッシュ領域のサイズxに対して、同じサイズの割当済専用領域の増加率と比較して、入れ替え対象にする。式 ( 1 0 ) と同じ考え方である。

【 0 1 2 4 】

上記式 ( 1 2 ) を満たして入れ替えで効率がアップすると見なされる場合は、キャッシュ割当管理部 2 0 4 は、未割当専用領域と割当済専用領域とを入れ替える処理を行う (S63)。入れ替えで効率がアップすると見なされない場合は、キャッシュ割当管理部 2 0 4 は、入れ替え処理を行わない (S64)。

【 0 1 2 5 】

未割当専用領域獲得要求 3 0 が要求するキャッシュサイズと、式 ( 1 2 ) を満たすキャッシュサイズyとの関係は、専用領域獲得要求判定処理 ( 2 ) について図 1 3 で説明したのと同様である。つまり、式 ( 1 2 ) を満たすキャッシュサイズyが要求するキャッシュサイズ以下であれば、サイズyと要求サイズとの間のサイズを割り当てる。また、キャッシュサイズyが要求キャッシュサイズを超える場合は、式 ( 1 2 ) が満たされないとして、入れ替え処理は行わない。

【 0 1 2 6 】

[ 定期割当見直し処理S18 ( 3 ) ]

図 1 8 は、定期割当見直し処理S18の第 3 の例のフローチャート図である。図 1 9 は、定期割当見直し処理S18の第 3 の例を説明するための図である。定期割当見直し処理S18の第 3 の例では、現実の加重メモリアクセス比に基づく共用キャッシュ領域の理想サイズと、現実の共用キャッシュ領域のサイズとの間に乖離が発生している場合は、共用キャッシュ領域が理想のサイズ比に近づくように、共用キャッシュ領域を未割当専用領域獲得要求

に専用キャッシュ領域として割り当てるか、割当済専用キャッシュ領域を開放して共用キャッシュ領域に割り当てるかの入れ替え処理を行う。

【 0 1 2 7 】

図 19 に示した左右のキャッシュメモリ 130 は、実線が実際のサイズ比を示し、破線が理想のサイズ比を示している。左側のキャッシュメモリ 130 は、共用キャッシュ領域が理想サイズよりも大きくなっているため、その分を専用領域として未割当専用領域獲得要求に割り当てることが望ましい。また、逆に、右側のキャッシュメモリ 130 は、共用キャッシュ領域が理想サイズよりも小さくなっているため、その分を割当済専用領域を開放して共用領域に変更することが望ましい。

【 0 1 2 8 】

共用キャッシュ領域の理想サイズは、次の式 ( 1 3 ) ( 1 4 ) により求められる。

$$\text{MACCESSX}_n = (m=1-n) \text{MACCESS}_m * \text{PRIORITY}_m \quad (13)$$

$$\text{RSIZE}_{\text{SHR}} = \text{CEIL}\{(\text{MACCESSX}_{\text{SHR}}/\text{MACCESSX}_{\text{ALL}}) * \text{SIZE}_{\text{ALL}}\} \quad (14)$$

ここで、 $\text{MACCESSX}_n$  は  $n$  個の対象処理の加重メモリアクセス量の合計値を示し、 $\text{MACCESS}_m$  は対象処理  $m$  のメモリアクセス量を示し、 $\text{PRIORITY}_m$  は処理  $m$  の優先度を示す。さらに、 $\text{RSIZE}_{\text{SHR}}$  は共用キャッシュ領域 (SHR) の理想的なサイズを示し、 $\text{MACCESSX}_{\text{SHR}}$ 、 $\text{MACCESSX}_{\text{ALL}}$  は共有キャッシュ領域 (SHR) へのアクセス量、全キャッシュ領域へのアクセス量、全キャッシュ領域のサイズを示す。また、CEIL は切り上げを意味する。

【 0 1 2 9 】

したがって、式 ( 1 3 ) により  $n$  個の対象処理の加重メモリアクセス量の合計値が求められるので、共用キャッシュ領域を使用する対象処理全ての加重メモリアクセス量は、式 ( 1 3 ) により算出できる。同様に、全専用キャッシュ領域または全キャッシュ領域を使用する対象処理全ての加重メモリアクセス量も、同様に式 ( 1 3 ) により算出できる。

【 0 1 3 0 】

そして、式 ( 1 4 ) は、共用キャッシュ領域の理想サイズ  $\text{RSIZE}_{\text{SHR}}$  を、全キャッシュ領域サイズ ( $\text{SIZE}_{\text{ALL}}$ ) を共用領域と専用領域への加重メモリアクセス量で按分して ( $\text{MACCESSX}_{\text{SHR}}/\text{MACCESSX}_{\text{ALL}}$ ) 求めている。したがって、共用キャッシュ領域の理想サイズ  $\text{RSIZE}_{\text{SHR}}$  は、共用領域と専用領域との加重メモリアクセス数の比率に基づいて求められる。

【 0 1 3 1 】

そして、前述した、共用キャッシュ領域の理想サイズと、現実のサイズとを比較して、その差分が入れ替えに要するオーバーヘッド値  $\text{OVH1}$  を超えるか否かの判断が行われる。その判断は、以下の式 ( 1 5 ) が満たされるか否かによって行われる。

$$\text{ABS}\{\text{MACCESSX}_{\text{SHR}} - \text{MACCESSX}_{\text{ALL}} * (\text{SIZE}_{\text{SHR}}/\text{SIZE}_{\text{ALL}})\} > \text{OVH1} \quad (15)$$

ここで、ABS は絶対値を示している。

【 0 1 3 2 】

式 ( 1 5 ) は、プロセス実行中における共用キャッシュ領域に割り当てられた対象処理でのメモリアクセス頻度を各対象処理の優先度で重み付けした第 1 の加重アクセス頻度  $\text{MACCESSX}_{\text{SHR}}$  と、各対象処理のメモリアクセス頻度を各対象処理の優先度で重み付けした加重アクセス頻度の累計値  $\text{MACCESSX}_{\text{ALL}}$  を共用キャッシュ領域と前記専用キャッシュ領域のサイズ比 ( $\text{SIZE}_{\text{SHR}}/\text{SIZE}_{\text{ALL}}$ ) で按分した共用キャッシュ領域の第 2 の加重アクセス頻度  $\text{MACCESSX}_{\text{ALL}} * (\text{SIZE}_{\text{SHR}}/\text{SIZE}_{\text{ALL}})$  との差分が、オーバーヘッド値  $\text{OVH1}$  より高い場合の条件式である。この条件式が満たされる場合に、その差分に対応したサイズだけ、割当済み専用キャッシュ領域を共用キャッシュ領域に入れ替えるか、または共用キャッシュ領域を未割当専用キャッシュ領域に入れ替える。

【 0 1 3 3 】

上記式 ( 1 5 ) は、式全体に  $\text{SIZE}_{\text{ALL}}/\text{MACCESSX}_{\text{ALL}}$  を乗算すると、次の式 ( 1 6 ) になる。

$$\text{ABS}\{(\text{MACCESSX}_{\text{SHR}}/\text{MACCESSX}_{\text{ALL}}) * \text{SIZE}_{\text{ALL}} - \text{SIZE}_{\text{SHR}}\} > \text{OVH1} * (\text{SIZE}_{\text{ALL}}/\text{MACCESSX}_{\text{ALL}}) \quad (16)$$

$$\text{ABS}\{(\text{RSIZE}_{\text{SHR}} - \text{SIZE}_{\text{SHR}})\} > \text{OVH1} * (\text{SIZE}_{\text{ALL}}/\text{MACCESSX}_{\text{ALL}}) \quad (16)$$

10

20

30

40

50



すなわち，現実の加重アクセス頻度比による共用キャッシュ領域の理想的なサイズ $RSIZE_{S_{HR}}$ と，現実のサイズ $SIZE_{SHR}$ との差分がオーバーヘッド値を超えるか否かを判定することと同じである。

#### 【 0 1 3 4 】

そこで，上記式（ 1 5 ）または（ 1 6 ）が満たされる場合は，共用キャッシュ領域のサイズが理想的なサイズに近づくように，割当済み専用キャッシュ領域を共用キャッシュ領域に入れ替えるか，逆に，共用キャッシュ領域を未割当専用キャッシュ領域に入れ替える。

#### 【 0 1 3 5 】

そのために，本実施の形態では，図 1 9 の左側に示すように，共用領域の理想サイズと比較して現実のサイズが大きい場合は，専用領域候補リストのサイズ（WAY数）を増加し，定期的見直しでは未割当専用領域獲得要求に対して専用領域を割り当てることができることを記録する。逆に，図 1 9 の右側に示すように，共用領域の理想サイズと比較して現実のサイズが小さい場合は，専用領域借りリストのサイズ（WAY数）を増加し，定期的見直しでは割当済み専用領域を共用領域を割り当てることができることを記録する。

#### 【 0 1 3 6 】

そして，定期的見直しでは，図 1 9 の左側に示すように，専用領域候補リストのサイズ（WAY数）分だけ，未割当専用領域リストLIST-DEMANDの先頭から専用領域を割り当てる。あるいは，図 1 9 の右側に示すように，専用領域借りリストのサイズ（WAY数）分だけ，割当済み専用領域リストLIST-FREEの先頭から専用領域を開放する。

#### 【 0 1 3 7 】

図 1 8 のフローチャートにしたがって再度説明すると，キャッシュ割当管理部 2 0 4 は，共用領域の理想サイズと現実サイズとの差分に基づいて，共用領域と専用領域の入れ替えの必要性を判定する（S70）。具体的には，専用領域候補リストのサイズが正であれば（S71），未割当専用領域獲得要求に専用領域を割り当てる（S72）。一方，専用領域借りリストのサイズが正であれば（S73），割当済み専用領域を開放する（S74）。

#### 【 0 1 3 8 】

キャッシュ割当管理部 2 0 4 は，共用領域の理想サイズと現実サイズとの差分がオーバーヘッド値を超えてない場合は，図 1 4 ， 1 6 で説明した定期割当見直し処理（ 1 ）（ 2 ）を実行する（S75）。

#### 【 0 1 3 9 】

上記の第 3 の例では，専用領域候補リストに記録されたキャッシュサイズ（Way数）で割り当てることが望ましい。

#### 【 0 1 4 0 】

〔 プロセス切り替え時のセクタID変換テーブルの入れ替え処理 〕

図 2 5 は，アプリケーションプログラムのプロセス 2 2 が切り替わるときのセクタID変換テーブルの入れ替え処理を示す図である。OS内のプロセス管理部 2 0 2 は，切り替え元の稼働プロセス22Aがプロセスを停止するときに，そのプロセスの停止処理を行い（S81），同時に，キャッシュ割当管理部 2 0 4 にセクタID変換テーブルの切替制御を指示し（S83），切り替え元の稼働プロセス22AのセクタID変換テーブルを，切り替え先の稼働プロセス22BのセクタID変換テーブルに入れ替えさせる（S84）。その後，プロセス管理部 2 0 2 は，切り替え先の稼働プロセス22Bのプロセス稼働を開始させる（S86）。

#### 【 0 1 4 1 】

このように，プロセスが切り替わるときにプロセス管理部 2 0 2 がキャッシュ割当管理部 2 0 4 にセクタID変換テーブルを入れ替えさせ，切り替え先のプロセスがそのセクタID変換テーブルのキャッシュメモリ分割状態に基づいてメモリアクセスを行う。

#### 【 0 1 4 2 】

第 1 の実施の形態では，図 8 の専用領域獲得要求に対する判定（ 1 ）では，専用領域獲得要求の対象処理の実キャッシュ有効利用度が第 2 の基準値を超えていれば，専用領域を割り当てた。この判定において，式（ 4 ）（ 7 ）において優先度を乗算していないキャッ

シュ有効利用度を指標にして専用領域の割当の有無を判断してもよい。

【 0 1 4 3 】

第 1 の実施の形態では、図 1 0 - 1 3 の専用領域獲得要求に対する判定 ( 2 ) では、実キャッシュ有効利用数の増加率を判断指標にして、専用領域獲得要求に対して専用領域を割り当てるか否かを判断した。この判定においても、式 ( 8 ) ( 9 ) において優先度を乗算していないキャッシュ有効利用数の増加率を指標にして専用領域の割当の有無を判断してもよい。

【 0 1 4 4 】

第 1 の実施の形態において、定期割当見直し ( 1 ) ( 2 ) ( 3 ) においても、同様に、優先度を考慮しない指標を利用して判断してもよい。

10

【 0 1 4 5 】

[ 第 2 の実施の形態 ]

第 2 の実施の形態は、通常のアプリケーションプログラムのプロセス稼働中に、割込・例外処理が発生した場合のキャッシュメモリの動的割込に関する。割込・例外処理は、優先度の高い処理やプログラムであり、CPU内部割込として、数値オーバーフローなどによるプログラム割込、入出力処理要求などのスーパーバイザーコール割込などがあり、CPU外部割込として、入出力要求の完了通知などの入出力割込、タイマドリブンによる割込のタイマ割込、コンソール操作による割込のコンソール割込などがある。

【 0 1 4 6 】

これらの割込・例外処理が発生すると、OSは、稼働中のプログラムや処理を中断して、OSが有するOS割込処理を実行する。このOS割込処理は、それまでの稼働中のプログラムや処理が実行するメモリアクセスのアドレスとは全く異なるアドレスにメモリアクセスを実行する。そのため、それまで稼働中だったプログラムや処理の共用キャッシュ領域内のデータは、OS割込処理によるメモリアクセスの結果、追い出されてしまうことが多い。つまり、OS割込処理は、共用キャッシュ領域内のデータをかき乱す攪乱力が高い処理である。

20

【 0 1 4 7 】

そこで、第 2 の実施の形態では、OS20に含まれるOS割込処理内に、特定のローカルセクタIDを割り当てた割込・例外処理専用領域獲得要求と、専用領域開放要求とを挿入しておく、さらに、OS割込処理内の命令にその特定のローカルセクタIDを割り当てる。それにより、第 1 の実施の形態で説明したように、OS割込処理には攪乱処理用の専用キャッシュ領域が割り当てられ、その攪乱処理用の専用キャッシュ領域を使用してOS割込処理内のメモリアクセスが実行される。

30

【 0 1 4 8 】

図 2 6 は、第 2 の実施の形態におけるOS割込処理プログラム ( OS割込処理プロセス203 ) の構成例を示す図である。OS設計において、OS割込処理プログラム内に割込・例外処理専用領域獲得要求30Eと、その専用領域開放要求34Eが挿入される。この割込・例外処理専用領域獲得要求30Eと、専用領域開放要求34Eには、対象のプロセスのプロセスIDと、特定のローカルセクタID=1が設定されている。さらに、OS割込処理プログラム内の命令には、その特定のローカルセクタID=1が設定されている。

【 0 1 4 9 】

さらに、第 2 の実施の形態においても、開発環境で割込・例外処理を発生させて、その特性調査が行われる。この特性調査により、割込・例外処理は、キャッシュミスヒット率が高く、前述の式 ( 5 ) に示した攪乱力が高い処理となっている。

40

【 0 1 5 0 】

図 2 7 は、第 2 の実施の形態におけるメモリアクセスを説明する図である。図 2 6 には、通常プロセス稼働中 ( S90 ) でのCPU12によるメモリアクセスと、割込処理時 ( S91 ) でのキャッシュメモリの動的割込状態とメモリアクセスとが示されている。

【 0 1 5 1 】

通常プロセス稼働中S90では、CPU12は、キャッシュメモリ 1 3 0 内の共用キャッシュ領域 1 3 0 - 1 を使用してメモリアクセスを実行する。したがって、共用キャッシュ領域 1

50

30 - 1内には、メインメモリ16内の稼働プロセス使用領域内の一部のデータが書き込まれている。

#### 【0152】

一方、割込・例外処理が発生すると、図26に示したOS割込処理プロセス内の割込・例外処理専用領域獲得要求30Eが発行され、図8のフローチャートに示したように、キャッシュ割当管理部204が、特性調査結果に基づいて攪乱力が高いと判定し、キャッシュメモリ130内に攪乱処理用の専用キャッシュ領域130-2を割り当てる。この割当によりセクタID変換テーブルには、ローカルセクタID=1に対して割り当てた専用キャッシュ領域のキャッシュセクタIDが紐づけられる。その結果、図26で示したOS割込処理プロセス内の命令がローカルセクタID=1に基づいて専用キャッシュ領域130-2を使用してメモリアクセスを実行する。その結果、キャッシュメモリ130の共用キャッシュ領域130-1内の稼働プロセスデータがOS割込処理によるキャッシュミスヒットで追い出されることが回避される。

10

#### 【0153】

図28、図29は、第2の実施の形態における割込・例外処理発生時の処理を示す図である。図30は、第2の実施の形態における割込・例外処理発生時の処理のフローチャート図である。これらを参照して、割込・例外処理発生時の処理について説明する。

#### 【0154】

通常プロセス稼働中は、OS20内のプロセス管理部202が通常プロセスの動作を管理している(S90)。そして、例えば入出力部17からハードウェア割込・例外処理が発生する(S91)。この割込・例外処理は、プロセス管理部202に通知される。これに応答して、プロセス管理部202は、割込処理への切り替え処理を実行する(S92)。

20

#### 【0155】

すなわち、OS内の割込処理プロセス203が、割込処理プロセスに挿入されている割込・例外処理の専用領域獲得要求30Eを発行する(S93)。これに応答して、キャッシュ割当管理部204が、専用領域の割当の有無を判定する(S94)。この判定で、キャッシュ割当管理部204は、対応する処理がキャッシュ攪乱力が高いことを検出し、攪乱処理用の専用キャッシュ領域を割り当てる(S94)。具体的には、キャッシュ割当管理部204は、OS割込処理プロセスのセクタID変換テーブル131において、ローカルセクタID=1に対して割り当てる専用キャッシュ領域のキャッシュセクタIDを紐づけて、さらに、キャッシュセクタ設定部132に専用キャッシュ領域のサイズを設定する。

30

#### 【0156】

そして、割込処理プロセス203は、OS割込処理の機械語の命令を発行し、CPU12に実行させる(S95)。このOS割込処理の機械語の命令には、ローカルセクタID=1が設定されているので、CPU12は、セクタID変換テーブル131を介して攪乱処理用の専用キャッシュ領域のキャッシュセクタIDを取得し、そのキャッシュセクタIDの攪乱処理用の専用キャッシュ領域130-2を使用してメモリアクセスを実行する(S95-1)。このメモリアクセスでキャッシュミスヒットが発生すると、メインメモリ16に対してアクセスが実行される(S95-2)。これにより、先に稼働中だったプロセスの共用キャッシュ領域130-1内のデータは、OS割込処理のメモリアクセスの影響を受けることがなく、キャッシュメモリから追い出されることはない。

40

#### 【0157】

割込処理プロセス203は、OS割込処理が完了すると、OS割込処理専用領域開放要求34Eを発行する(S96)。これに応答して、キャッシュ割当管理部204は、セクタID変換テーブル131のローカルセクタID=1に紐づけられていたキャッシュセクタIDをID=0(共用キャッシュ領域)に書き換えて、攪乱処理用の専用キャッシュ領域を開放する(S97)。その後、プロセス管理部202が、元の稼働プロセスへの切替を実行して、稼働中だったプロセスの再稼働を行う(S99)。

#### 【0158】

以上の通り、第2の実施の形態によれば、割込・例外処理が発生した場合、図8のフロ

50

ーチャートにしたがってキャッシュ攪乱力が高いと評価されて、攪乱処理用の専用キャッシュ領域が割り当てられ、割込前の稼働中のプロセスのデータが、共用キャッシュ領域から追い出されることが回避される。

【0159】

以上の実施の形態をまとめると、次の付記のとおりである。

【0160】

(付記1)

キャッシュメモリを共用キャッシュ領域と専用キャッシュ領域とに分割して制御するキャッシュメモリ制御工程をプロセッサに実行させるプロセッサ読み取り可能なキャッシュメモリ制御プログラムであって、

前記キャッシュメモリ制御工程は、

前記プロセッサによるプロセスの実行中に発行される専用キャッシュ領域の割当を要求する専用領域獲得要求に応答して、前記専用領域獲得要求が対象とする対象処理について前記専用キャッシュ領域を割り当てた場合のキャッシュヒット率と前記共用キャッシュ領域を割り当てた場合のキャッシュヒット率との差と、メモリアクセス頻度とに基づくキャッシュ有効利用度に応じて、前記キャッシュ有効利用度が高い程専用キャッシュ領域を割当て、低い程共用キャッシュ領域を割り当てるキャッシュ領域割当工程と、

前記プロセッサによるプロセスの実行中に発行される割り当て済み専用キャッシュ領域の開放を要求する専用領域開放要求に応答して、前記割り当てられた専用キャッシュ領域の割当を解放する専用キャッシュ領域開放工程とを有し、

前記専用キャッシュ領域が割り当てられた対象処理中に発行されるメモリアクセス命令が、前記プロセッサにより前記割り当てた専用キャッシュ領域を利用して実行されるキャッシュメモリ制御プログラム。

【0161】

(付記2)

付記1において、

前記キャッシュ有効利用度は、前記専用キャッシュ領域を割り当てた場合のキャッシュヒット率と前記共用キャッシュ領域を割り当てた場合のキャッシュヒット率との差に、前記メモリアクセス頻度を乗算し利用する専用キャッシュ領域のサイズで除算した値であるキャッシュメモリ制御プログラム。

【0162】

(付記3)

付記2において、

キャッシュ領域割当工程では、前記キャッシュ有効利用度に代えて、前記キャッシュ有効利用度以上に前記対象処理の優先度を乗じた実キャッシュ有効利用度に応じて専用キャッシュ領域または共用キャッシュ領域を割り当てるキャッシュメモリ制御プログラム。

【0163】

(付記4)

付記1において、

前記専用キャッシュ領域を割り当てた場合のキャッシュヒット率は、予め、割り当てる専用キャッシュ領域のサイズを可変させて前記対象処理を前記プロセッサに実行させたときに取得した専用キャッシュ領域サイズ対キャッシュヒット率の特性データであるキャッシュメモリ制御プログラム。

【0164】

(付記5)

付記1において、

前記共用キャッシュ領域を割り当てた場合のキャッシュヒット率は、前記専用領域獲得要求が発行された対象処理が実行中における前記共用キャッシュ領域のキャッシュヒット率であるキャッシュメモリ制御プログラム。

【0165】

## ( 付記 6 )

付記 1 において、

前記キャッシュ領域割当工程において、前記専用領域獲得要求に応答して、前記専用領域獲得要求が対象とする対象処理についてのキャッシュミス率に前記メモリアクセス頻度を乗算したキャッシュ攪乱力が基準値より高い場合に、前記共用キャッシュ領域と専用キャッシュ領域とは異なる攪乱処理キャッシュ領域を割り当てるキャッシュメモリ制御プログラム。

## 【 0 1 6 6 】

## ( 付記 7 )

付記 6 において、

前記キャッシュ領域割当工程において、前記攪乱処理キャッシュ領域を複数の前記専用領域獲得要求に共通に割り当てるキャッシュメモリ制御プログラム。

## 【 0 1 6 7 】

## ( 付記 8 )

付記 1 において、

前記キャッシュ領域割当工程において、前記専用キャッシュ領域の割当または共用キャッシュ領域の割当にしたがって、前記対象処理に対する前記専用キャッシュ領域と共用キャッシュ領域の割当情報を有する割当情報テーブルを更新し、

前記キャッシュメモリ制御工程は、更に、

前記プロセスが変更されるときに、変更されるプロセスの前記割当情報テーブルを変更後のプロセスの前記割当情報テーブルに切り替える割当情報テーブル切り替え工程を有するキャッシュメモリ制御プログラム。

## 【 0 1 6 8 】

## ( 付記 9 )

付記 2 または 3 において、

前記キャッシュ領域割当工程において、新たに発行された発行中専用領域獲得要求の前記キャッシュ有効利用度と、前記専用キャッシュ領域が割り当て済みの割当済み専用領域獲得要求の前記キャッシュ有効利用度とを比較し、前記発行中専用領域獲得要求のキャッシュ有効利用度のほうが高い場合に、前記割当済み専用領域獲得要求に対して割り当て済みの専用キャッシュ領域を開放して前記共用キャッシュ領域を割り当て、前記発行中専用領域獲得要求に前記専用キャッシュ領域を割り当てるキャッシュ制御プログラム。

## 【 0 1 6 9 】

## ( 付記 1 0 )

付記 2 または 3 において、

前記キャッシュ領域割当工程において、新たに発行された発行中専用領域獲得要求の前記キャッシュ有効利用度に前記専用キャッシュ領域サイズを乗算したキャッシュ有効利用数と、前記専用キャッシュ領域が割り当て済みの割当済み専用領域獲得要求の単位領域の増加に対するキャッシュヒット率の増加率である増加率に、前記メモリアクセス頻度を乗算したキャッシュ有効利用数の増加率を、前記発行中専用領域獲得要求の専用キャッシュ領域サイズ分だけ累積した累積増加数とを比較し、

前記発行中専用領域獲得要求の前記キャッシュ有効利用数のほうが多い場合に、前記割当済み専用領域獲得要求に対して割り当て済みの専用キャッシュ領域を開放して前記共用キャッシュ領域を割り当て、前記発行中専用領域獲得要求に前記専用キャッシュ領域を割り当てるキャッシュ制御プログラム。

## 【 0 1 7 0 】

## ( 付記 1 1 )

付記 1 0 において、

前記キャッシュメモリ制御工程は、更に、

前記プロセス実行中に前記割当済み専用領域獲得要求が対象とする対象処理について前

10

20

30

40

50

記割当済み専用キャッシュ領域のキャッシュヒット率とアクセス頻度を調査し，前記キャッシュ有効利用数の増加率のリストを有する割当済み専用領域リスト（LIST-FREE）を作成するリスト作成工程を有し，

前記キャッシュ領域割当工程において，

前記割当済み専用領域リスト内の増加数率が小さい順に累積して前記累積増加数を算出するキャッシュ制御プログラム。

【 0 1 7 1 】

（付記 1 2 ）

付記 2 または 3 において，

前記キャッシュメモリ制御工程は，更に，

前記プロセス実行中に，前記専用キャッシュ領域が割り当て済みの割当済み専用領域獲得要求が対象とする対象処理について，前記割当済み専用キャッシュ領域のキャッシュヒット率とアクセス頻度を調査し，前記割当済み専用領域獲得要求の単位領域の増加に対するキャッシュヒット率の増加数である増加率に，前記メモリアクセス頻度を乗算したキャッシュ有効利用数の増加率のリストを有する割当済み専用領域リスト（LIST-FREE）と，前記専用キャッシュ領域が未だ割り当てられていない未割当専用領域獲得要求の前記キャッシュ有効利用数の増加率のリストを有する未割当専用領域リスト（LIST-DEMAND）とを作成するリスト作成工程と，

前記プロセス実行中に定期的に実行され，前記割当済み専用領域リストの最小値より前記未割当専用領域リストの最大値が入れ替え処理に要するオーバーヘッド値以上大きい場合に，対応する前記割当済み専用領域獲得要求に対して割当済みの専用キャッシュ領域を開放して前記共用キャッシュ領域を割り当て，対応する前記未割当専用領域獲得要求に前記専用キャッシュ領域を割り当てる定期的専用領域入れ替え工程とを有するキャッシュ制御プログラム。

【 0 1 7 2 】

（付記 1 3 ）

付記 1 2 において，

前記定期的専用領域入れ替え工程において，前記未割当専用領域リストの最大値に前記専用キャッシュ領域サイズを乗算したキャッシュ有効利用数が，前記割当済み専用領域リストの最小値から順に前記専用キャッシュ領域サイズに対応するリスト数までの割当済み専用領域のキャッシュ有効利用数の増加率の累計値より，オーバーヘッド値以上多い場合に，前記未割当専用領域リストの最大値に対応する未割当専用領域獲得要求に前記専用キャッシュ領域サイズ分だけ前記専用キャッシュ領域を割り当て，前記割当済み専用領域リストの前記リスト数まで前記割当済みの専用キャッシュ領域を開放して前記共用キャッシュ領域を割り当てるキャッシュ制御プログラム。

【 0 1 7 3 】

（付記 1 4 ）

付記 2 または 3 において，

前記キャッシュメモリ制御工程は，更に，

前記プロセス実行中に定期的に実行される定期的専用領域入れ替え工程を有し，

前記定期的専用領域入れ替え工程において，

現実の加重メモリアクセス比に基づく共用キャッシュ領域の理想サイズと，現実の共用キャッシュ領域のサイズとの差分が，オーバーヘッド値を超えて高い場合に，前記共用キャッシュ領域が理想のサイズ比に近づくように，前記共用キャッシュ領域を未だ専用領域が割り当てられていない未割当専用領域獲得要求に専用キャッシュ領域として割り当てるか，または，専用領域が割当済みの割当済み専用領域獲得要求に割り当てられた割当済み専用キャッシュ領域を開放するかにより，前記専用キャッシュ領域の入れ替え処理を行うキャッシュ制御プログラム。

【 0 1 7 4 】

（付記 1 5 ）

付記 14 において、

前記キャッシュメモリ制御工程は、更に、

前記プロセス実行中に、前記割当済み専用領域獲得要求が対象とする対象処理について、前記割当済み専用キャッシュ領域のキャッシュヒット率とアクセス頻度を調査し、前記割当済み専用領域獲得要求の単位領域増加に対するキャッシュヒット率の増加数である増加率に前記メモリアクセス頻度を乗算したキャッシュ有効利用数の増加率のリストを有する割当済み専用領域リスト (LIST-FREE) と、前記未割当専用領域獲得要求の前記キャッシュ有効利用数の増加率のリストを有する未割当専用領域リスト (LIST-DEMAND) とを作成するリスト作成工程を有し、

前記定期的専用領域入れ替え工程において、

前記専用領域借リストのサイズに対しては、前記割当済み専用領域リスト (LIST - FREE) 内の前記増加率が少ない順に、前記割当済み専用キャッシュ領域を開放して前記共用キャッシュ領域を割り当て、

前記専用領域候補リストのサイズに対しては、前記未割当専用領域リスト (LIST-DEMAND) 内の前記増加率が多い順に、前記専用キャッシュ領域を割り当てる、

キャッシュ制御プログラム。

【 0 1 7 5 】

( 付記 1 6 )

キャッシュメモリを共用キャッシュ領域と専用キャッシュ領域とに分割して制御するキャッシュメモリ制御工程をプロセッサに実行させるプロセッサ読み取り可能なキャッシュメモリ制御プログラムであって、

前記キャッシュメモリ制御工程は、

前記プロセッサによるプロセスの実行中に発行される専用キャッシュ領域の割当を要求する専用領域獲得要求に回答して、前記専用領域獲得要求が対象とする対象処理についてのキャッシュミス率にメモリアクセス頻度を乗算したキャッシュ攪乱力が基準値より高い場合に、他のキャッシュ領域とはキャッシュデータ管理が隔離された専用キャッシュ領域を割り当てるキャッシュ領域割当工程と、

前記プロセッサによるプロセスの実行中に発行される割り当て済み専用キャッシュ領域の開放を要求する専用領域開放要求に回答して、前記割り当てられた専用キャッシュ領域の割当を開放する専用キャッシュ領域開放工程とを有し、

前記専用キャッシュ領域が割り当てられた対象処理中に発行されるメモリアクセス命令が、前記プロセッサに前記割り当てた専用キャッシュ領域を利用して実行されるキャッシュメモリ制御プログラム。

【 0 1 7 6 】

( 付記 1 7 )

命令を実行してメモリアクセスを行う CPU と、

前記メモリアクセスが行われるときに参照されるキャッシュメモリと、

前記キャッシュメモリを共用キャッシュ領域と専用キャッシュ領域とに分割して制御するキャッシュメモリ制御ユニットとを有し、

前記キャッシュメモリ制御ユニットは、

前記 CPU によるプロセスの実行中に発行される専用キャッシュ領域の割当を要求する専用領域獲得要求に回答して、前記専用領域獲得要求が対象とする対象処理について前記専用キャッシュ領域を割り当てた場合のキャッシュヒット率と前記共用キャッシュ領域を割り当てた場合のキャッシュヒット率との差と、メモリアクセス頻度とに基づくキャッシュ有効利用度に応じて、前記キャッシュ有効利用度が高い程専用キャッシュ領域を割当て、低い程共用キャッシュ領域を割り当てるキャッシュ領域割当ユニットと、

前記 CPU によるプロセスの実行中に発行される割り当て済み専用キャッシュ領域の開放を要求する専用領域開放要求に回答して、前記割り当てられた専用キャッシュ領域の割当を解放する専用キャッシュ領域開放ユニットとを有し、

前記専用キャッシュ領域が割り当てられた対象処理中に発行される前記メモリアクセス

10

20

30

40

50

が、前記割り当てた専用キャッシュ領域を利用して実行されるプロセッサ。

【0177】

(付記18)

キャッシュメモリを共用キャッシュ領域と専用キャッシュ領域とに分割して制御するキャッシュメモリ制御方法であって、

前記プロセッサによるプロセスの実行中に発行される専用キャッシュ領域の割当を要求する専用領域獲得要求に応答して、前記専用領域獲得要求が対象とする対象処理について前記専用キャッシュ領域を割り当てた場合のキャッシュヒット率と前記共用キャッシュ領域を割り当てた場合のキャッシュヒット率との差と、メモリアクセス頻度とに基づくキャッシュ有効利用度に応じて、前記キャッシュ有効利用度が高い程専用キャッシュ領域を割

10

当て、低い程共用キャッシュ領域を割り当てるキャッシュ領域割当工程と、  
前記プロセッサによるプロセスの実行中に発行される割り当て済み専用キャッシュ領域の開放を要求する専用領域開放要求に応答して、前記割り当てられた専用キャッシュ領域の割当を解放する専用キャッシュ領域開放工程とを有し、

前記専用キャッシュ領域が割り当てられた対象処理中に発行されるメモリアクセス命令が、前記プロセッサにより前記割り当てた専用キャッシュ領域を利用して実行されるキャッシュメモリ制御方法。

【符号の説明】

【0178】

10：プロセッサユニット

20

12：CPU

13：キャッシュユニット

16：メインメモリ

20：OS

21：コンパイラ

22：アプリケーションプログラム

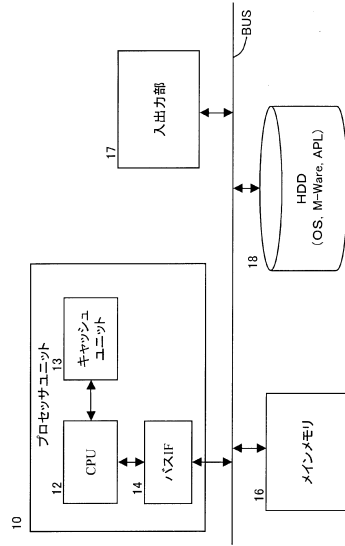
130：キャッシュメモリ

130\_\_1：共用キャッシュ領域

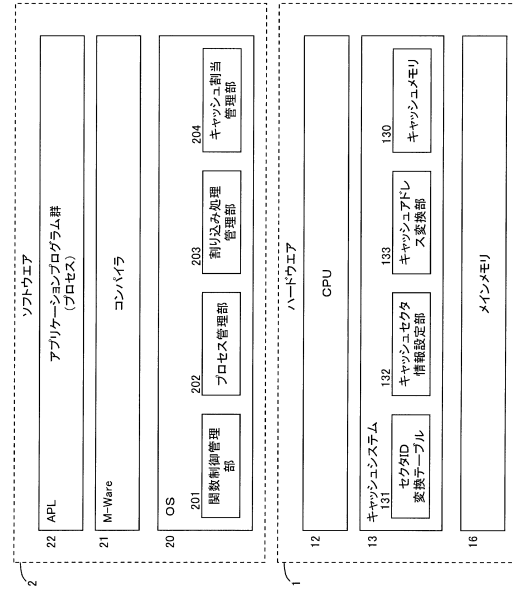
130\_\_2：専用キャッシュ領域



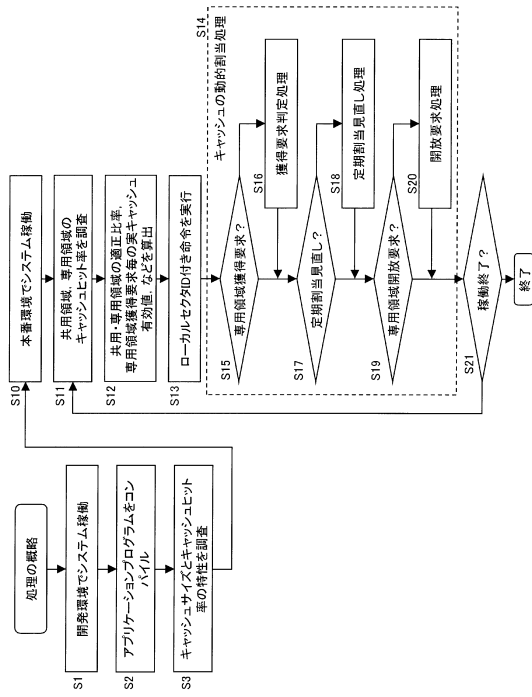
【図 1】



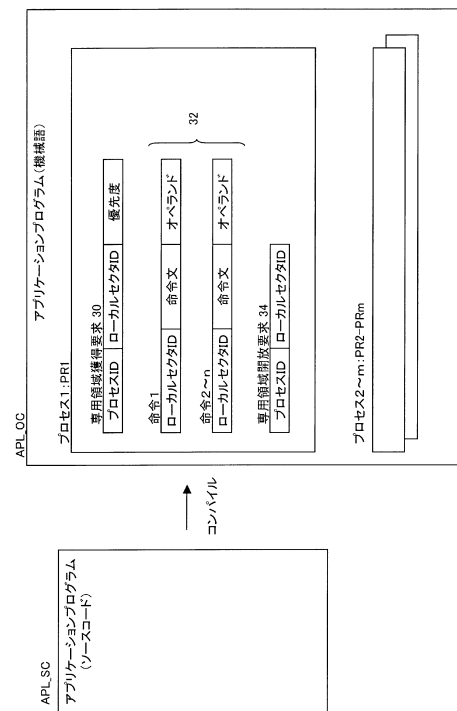
【図 2】



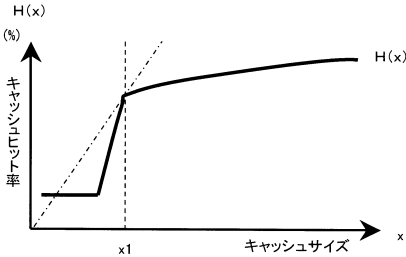
【図 3】



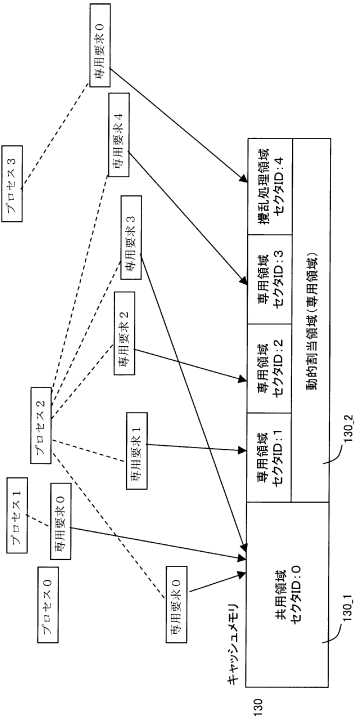
【図 4】



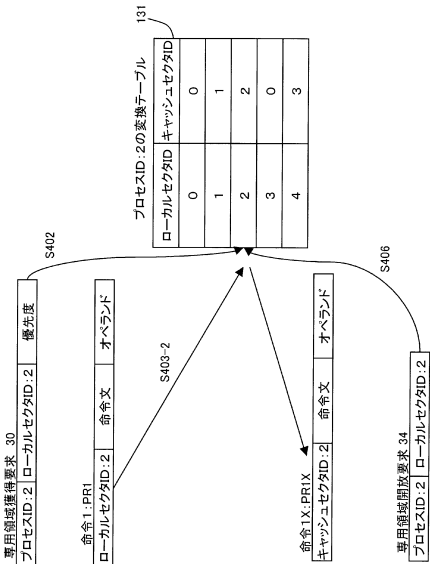
【図 5】



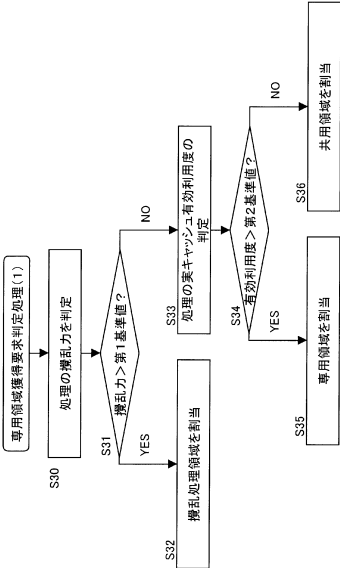
【図 6】



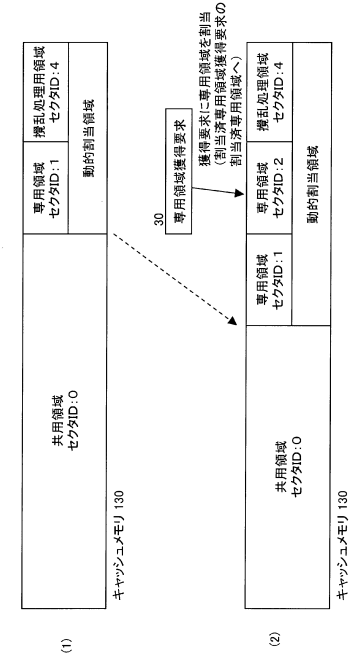
【図 7】



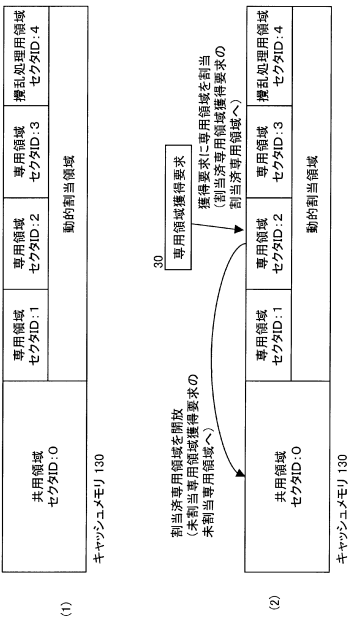
【図 8】



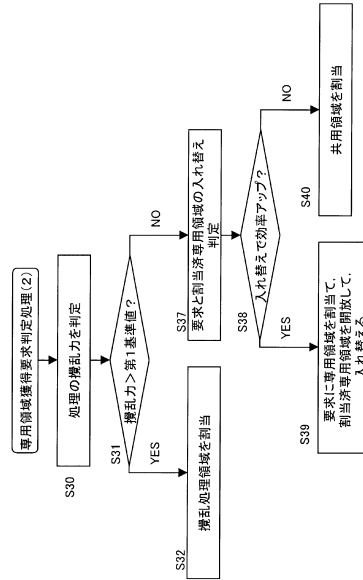
【図 9】



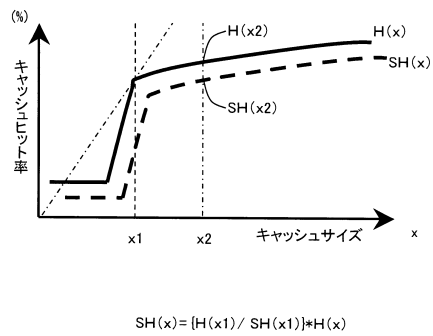
【図 1 1】



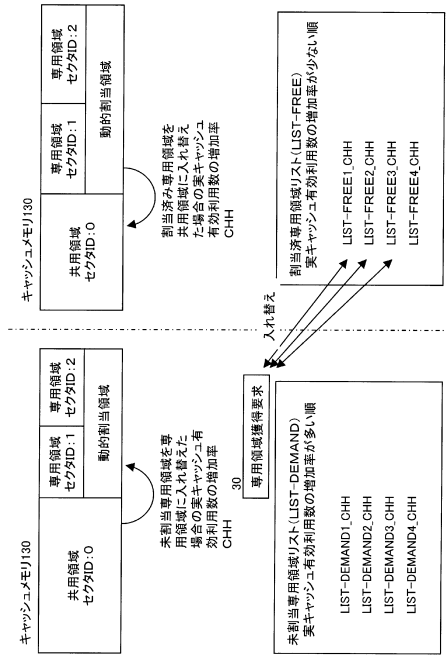
【図 1 0】



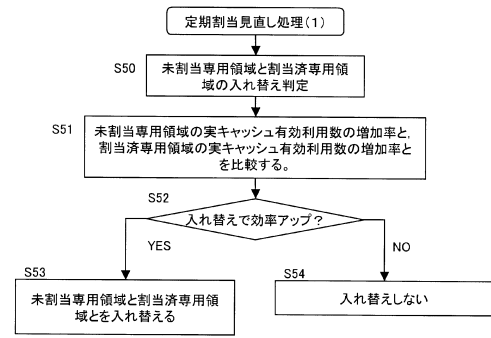
【図 1 2】



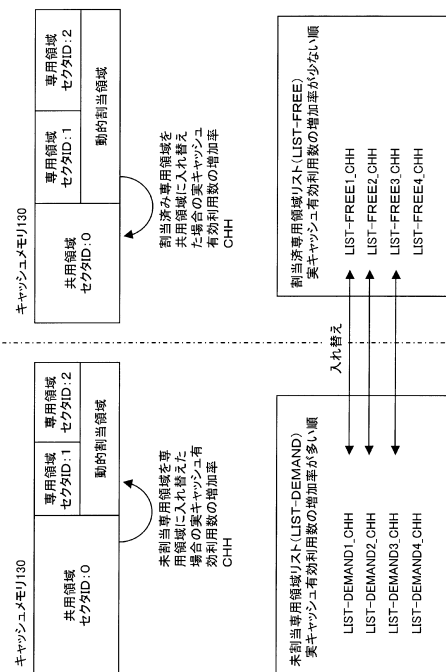
【図 13】



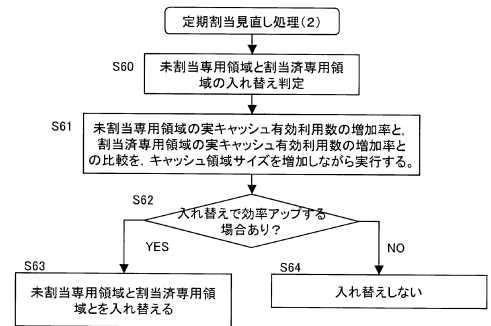
【図 14】



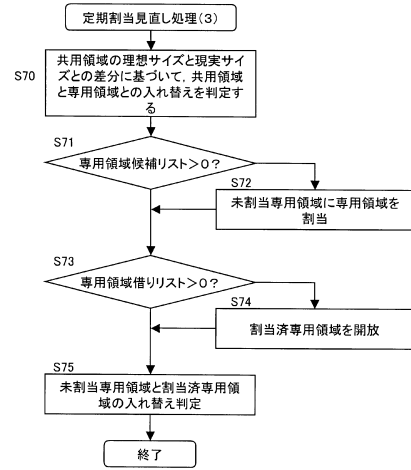
【図 15】



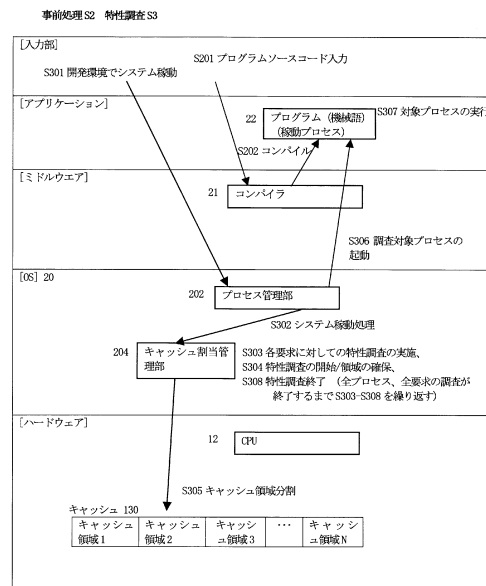
【図 16】



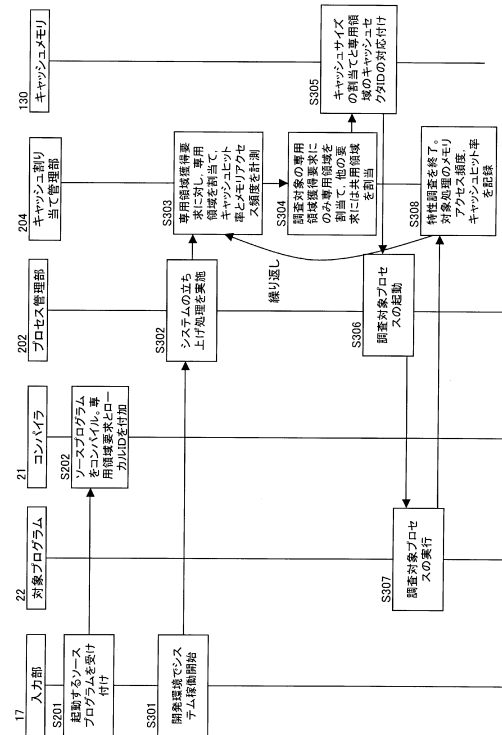
【 図 1 8 】



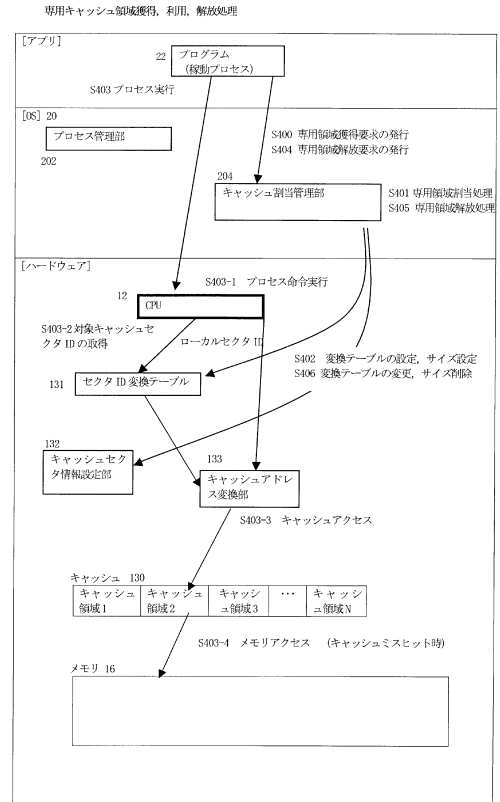
【 図 2 0 】



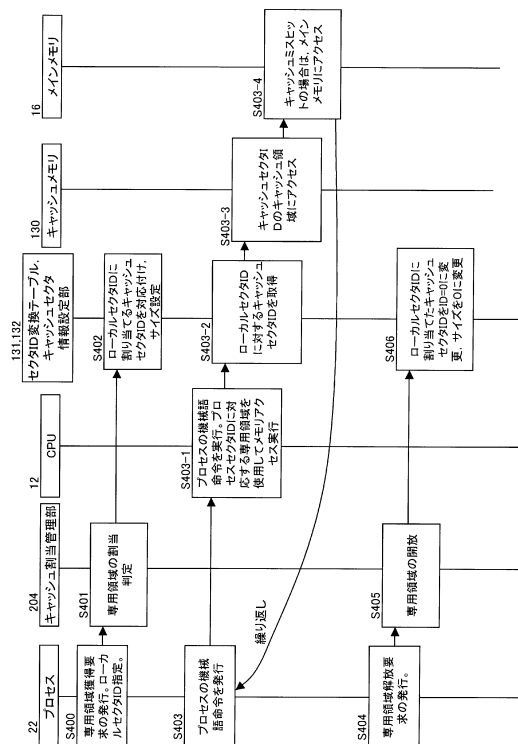
【 図 2 1 】



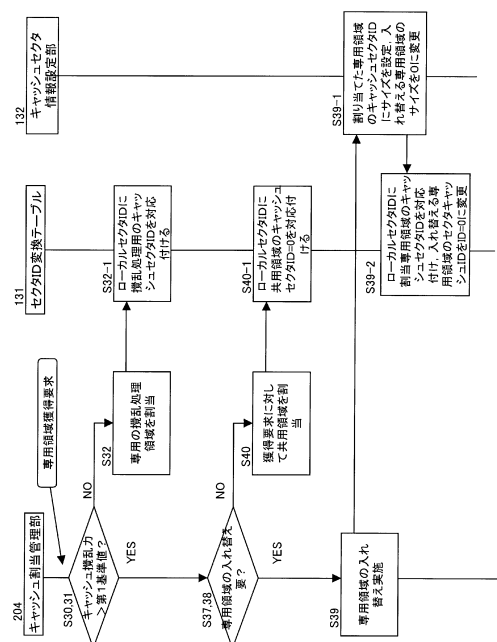
【 図 2 2 】



【 図 2 3 】

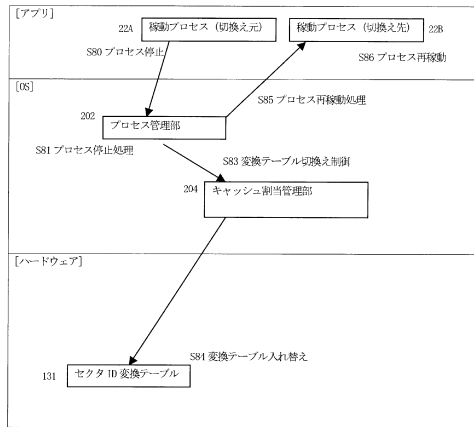


【 図 2 4 】

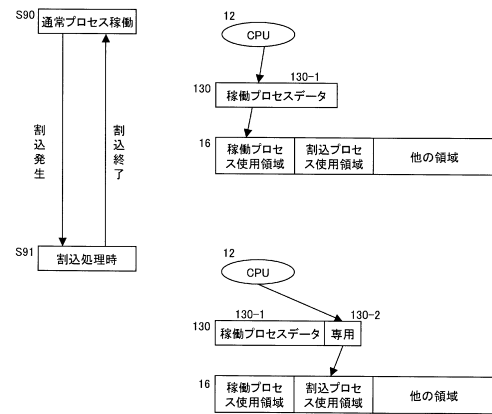


【図 25】

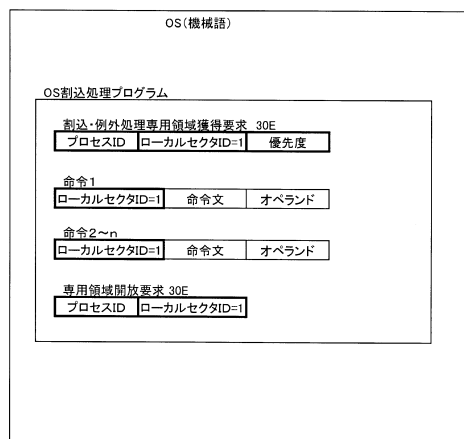
プロセス切換え時のセクタ ID 変換テーブル入れ替え処理



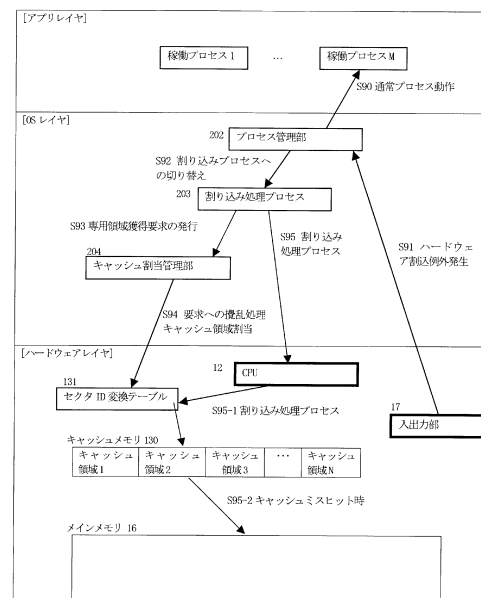
【図 26】



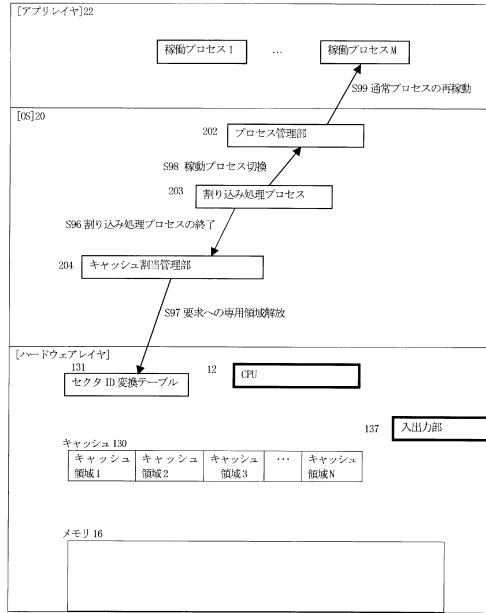
【図 27】



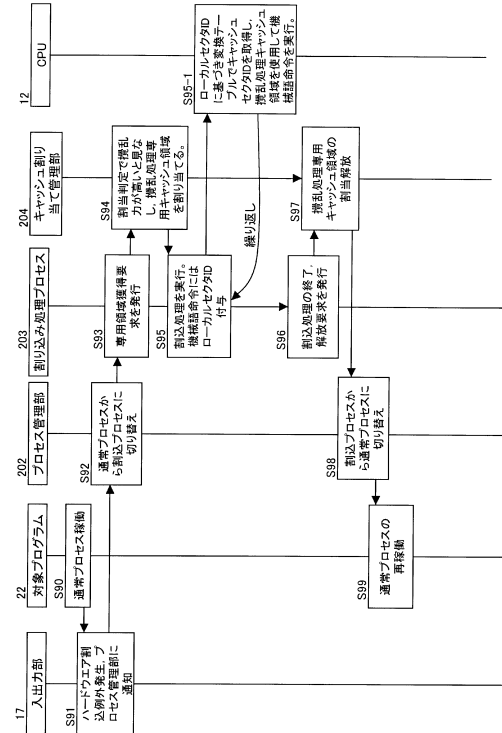
【図 28】



【図 29】



【図 30】





---

フロントページの続き

(72)発明者 弓場 康弘  
神奈川県川崎市中原区上小田中4丁目1番1号 富士通株式会社内

審査官 酒井 恭信

(56)参考文献 特開2008-097572(JP,A)  
米国特許出願公開第2008/0065833(US,A1)  
特開2009-245047(JP,A)  
米国特許出願公開第2009/0248922(US,A1)  
特開2002-342163(JP,A)  
国際公開第2010/093003(WO,A1)

(58)調査した分野(Int.Cl., DB名)  
G06F 12/08