

(19) 日本国特許庁(JP)

(12) 特 許 公 報(B2)

(11) 特許番号

特許第6229347号
(P6229347)

(45) 発行日 平成29年11月15日(2017.11.15)

(24) 登録日 平成29年10月27日(2017.10.27)

(51) Int.Cl.

F I

G 0 6 F 3 / 0 6 (2006.01)

G 0 6 F 3 / 0 6 3 0 1 F

請求項の数 7 (全 16 頁)

(21) 出願番号	特願2013-154585 (P2013-154585)	(73) 特許権者	000005223
(22) 出願日	平成25年7月25日(2013.7.25)		富士通株式会社
(65) 公開番号	特開2015-26183 (P2015-26183A)		神奈川県川崎市中原区上小田中4丁目1番1号
(43) 公開日	平成27年2月5日(2015.2.5)	(74) 代理人	100092152
審査請求日	平成28年4月5日(2016.4.5)		弁理士 服部 毅巖
		(72) 発明者	鈴木 友憲
			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	遠藤 篤
			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	桜井 茂行

最終頁に続く

(54) 【発明の名称】 ストレージ制御装置、ストレージ制御プログラム、およびストレージ制御方法

(57) 【特許請求の範囲】

【請求項 1】

論理ボリュームにおけるストレージ装置の構成、および前記ストレージ装置の性能にもとづいて前記論理ボリュームに対して単位時間あたりに発行可能な I / O 発行可能数を可変設定する設定部と、

前記論理ボリュームに対して受け付けた I / O 要求を要求蓄積部に蓄積し、前記要求蓄積部が蓄積する前記 I / O 要求のうちから、単位時間あたりの I / O 発行数を設定された前記 I / O 発行可能数の範囲に制限して前記論理ボリュームに前記 I / O 要求の発行をおこなう発行制御部と、

を備えることを特徴とするストレージ制御装置。

10

【請求項 2】

前記設定部は、前記ストレージ装置の構成、および前記ストレージ装置の性能ごとに要求される性能情報にもとづいて、前記論理ボリュームごとに前記性能情報から導出される I / O 発行可能数を前記 I / O 発行可能数として設定することを特徴とする請求項 1 記載のストレージ制御装置。

【請求項 3】

前記設定部は、前記発行制御部が所定時間の間に発行した前記 I / O 要求の転送単位に応じて前記 I / O 発行可能数を変更することを特徴とする請求項 1 または請求項 2 記載のストレージ制御装置。

【請求項 4】

20

前記発行制御部は、前記所定時間の間に発行した前記 I / O 要求の転送単位を統計情報として保持し、前記統計情報にもとづいて前記転送単位を選択することを特徴とする請求項 3 記載のストレージ制御装置。

【請求項 5】

前記設定部は、前記論理ボリュームの生成時に前記性能情報にもとづいて転送単位ごとに対応する I / O 発行可能数を算出することを特徴とする請求項 2 記載のストレージ制御装置。

【請求項 6】

コンピュータに、

論理ボリュームにおけるストレージ装置の構成、および前記ストレージ装置の性能にもとづいて前記論理ボリュームに対して単位時間あたりに発行可能な I / O 発行可能数を可変設定し、

10

前記論理ボリュームに対して受け付けた I / O 要求を要求蓄積部に蓄積し、前記要求蓄積部が蓄積する前記 I / O 要求のうちから、単位時間あたりの I / O 発行数を設定された前記 I / O 発行可能数の範囲に制限して前記論理ボリュームに前記 I / O 要求の発行をおこなう、

処理を実行させることを特徴とするストレージ制御プログラム。

【請求項 7】

コンピュータが、

論理ボリュームにおけるストレージ装置の構成、および前記ストレージ装置の性能にもとづいて前記論理ボリュームに対して単位時間あたりに発行可能な I / O 発行可能数を可変設定し、

20

前記論理ボリュームに対して受け付けた I / O 要求を要求蓄積部に蓄積し、前記要求蓄積部が蓄積する前記 I / O 要求のうちから、単位時間あたりの I / O 発行数を設定された前記 I / O 発行可能数の範囲に制限して前記論理ボリュームに前記 I / O 要求の発行をおこなう、

処理を実行することを特徴とするストレージ制御方法。

【発明の詳細な説明】

【技術分野】

30

【0001】

本発明は、ストレージ制御装置、ストレージ制御プログラム、およびストレージ制御方法に関する。

【背景技術】

【0002】

複数の H D D (Hard Disk Drive) を 1 つの論理ボリュームとして扱うストレージ制御装置として R A I D (Redundant Array of Inexpensive Disks) 装置が知られている。R A I D 装置は、複数の H D D を搭載したドライブエンクロージャ (以下、D E (Drive Enclosure)) を備え、D E 内の H D D から R A I D を構成する。また、R A I D 装置は、D E 内の H D D の交換や追加、あるいは D E の交換や追加により H D D 構成を更新可能にする。

40

【0003】

R A I D 装置は、H D D 構成に変更が生じて安定した性能値 (たとえば、スループット、I O P S (Input/Output Per Second)) を提供することが求められる。そのため、R A I D 装置は、所定の性能を有することを検証された H D D を搭載している。H D D に求められる所定の性能は、その下限だけでなく上限も規定した基準の範囲内に適合することが求められる。

【0004】

R A I D 装置ベンダと H D D ベンダは、H D D に求められる性能の上限が H D D 内部の処理に委ねられることから、ファームウェアの変更と検証を繰り返してファームウェアの

50

最適化を図っている。

【先行技術文献】

【特許文献】

【0005】

【特許文献1】特開平8-69359号公報

【特許文献2】特開平8-63298号公報

【発明の概要】

【発明が解決しようとする課題】

【0006】

しかしながら、市場や生産拠点のグローバル化に伴うサプライチェーンの最適化を図る上で、HDDのマルチベンダ化が進展している。このようなマルチベンダ化の流れは、自然災害や政情不安などの不安定要因に対抗するため一層進展する傾向にある。

10

【0007】

また、ユーザのニーズにより細やかに対応するため、HDDの多機種化（サイズ、回転数など）に伴い様々な性能特性を有するHDDが存在する。

RAID装置ベンダとHDDベンダによるファームウェアの最適化を図るための作業負担は、ますます増大する傾向にある。

【0008】

1つの側面では、本発明は、ストレージ装置の性能最適化に伴う作業負担を軽減可能なストレージ制御装置、ストレージ制御プログラム、およびストレージ制御方法を提供することを目的とする。

20

【課題を解決するための手段】

【0009】

上記目的を達成するために、以下に示すような、ストレージ制御装置が提供される。ストレージ制御装置は、設定部と、発行制御部を備える。設定部は、論理ボリュームにおけるストレージ装置の構成、およびストレージ装置の性能にもとづいて論理ボリュームに対して単位時間あたりに発行可能なI/O発行可能数を可変設定する。発行制御部は、論理ボリュームに対して受け付けたI/O要求を要求蓄積部に蓄積し、要求蓄積部が蓄積するI/O要求のうちから、単位時間あたりのI/O発行数を設定されたI/O発行可能数の範囲に制限して論理ボリュームにI/O要求の発行をおこなう。

30

【0010】

また、上記目的を達成するために、以下に示すような、ストレージ制御プログラム、およびストレージ制御方法が提供される。ストレージ制御プログラムは、以下に示すような処理をコンピュータに実行させ、ストレージ制御方法は、以下に示すような処理をコンピュータが実行する。コンピュータは、論理ボリュームにおけるストレージ装置の構成、およびストレージ装置の性能にもとづいて論理ボリュームに対して単位時間あたりに発行可能なI/O発行可能数を可変設定し、論理ボリュームに対して受け付けたI/O要求を要求蓄積部に蓄積し要求蓄積部が蓄積するI/O要求のうちから、単位時間あたりのI/O発行数を設定されたI/O発行可能数の範囲に制限して論理ボリュームにI/O要求の発行をおこなう処理を実行する。

40

【発明の効果】

【0011】

1態様によれば、ストレージ制御装置、ストレージ制御プログラム、およびストレージ制御方法において、ストレージ装置の性能最適化に伴う作業負担を軽減する。

【図面の簡単な説明】

【0012】

【図1】第1の実施形態のストレージ制御装置の構成例を示す図である。

【図2】第2の実施形態のストレージシステムの構成例を示す図である。

【図3】異なるHDD種別におけるRAID性能の測定例を示す図である。

【図4】第2の実施形態のコントローラモジュールの構成例を示す図である。

50

【図 5】第 2 の実施形態の R A I D 装置のハードウェア構成例を示す図である。

【図 6】第 2 の実施形態の R A I D グループ追加処理のフローチャートを示す図である。

【図 7】第 2 の実施形態の性能情報テーブルの一例を示す図である。

【図 8】第 2 の実施形態の I / O 発行数管理テーブルの一例を示す図である。

【図 9】第 2 の実施形態の R A I D グループ I / O 受付処理のフローチャートを示す図である。

【図 10】第 2 の実施形態の統計情報管理テーブルの一例を示す図である。

【図 11】第 2 の実施形態の R A I D グループ I / O 発行数管理処理のフローチャートを示す図である。

【発明を実施するための形態】

10

【0013】

以下、図面を参照して実施の形態を詳細に説明する。

[第 1 の実施形態]

まず、第 1 の実施形態のストレージ制御装置について図 1 を用いて説明する。図 1 は、第 1 の実施形態のストレージ制御装置の構成例を示す図である。

【0014】

ストレージ制御装置 1 は、ホストコンピュータ 9 から論理ボリュームへの I / O 要求 7 を受け付ける。ストレージ制御装置 1 は、複数のストレージ装置 5 (5 a , 5 b , ... , 5 n) の組合せから論理ボリュームを生成する。論理ボリュームは、たとえば、R A I D がある。

20

【0015】

ストレージ装置 5 は、たとえば、H D D であるが、フラッシュメモリドライブなどの S S D (Solid State Drive) などであってもよい。また、ストレージ装置 5 は、ストレージ制御装置 1 が内蔵するものであってもよいし、ストレージ制御装置 1 に外付けされるものであってもよい。

【0016】

ストレージ制御装置 1 は、設定部 2 と、発行制御部 3 と、要求蓄積部 4 を含む。設定部 2 は、論理ボリュームに対して単位時間あたりに発行可能な I / O 発行可能数 6 を設定する。設定部 2 は、論理ボリュームにおけるストレージ装置 5 の構成、およびストレージ装置 5 の性能にもとづいて、I / O 発行可能数を設定する。

30

【0017】

これにより、ストレージ制御装置 1 は、論理ボリュームを構成するストレージ装置 5 の性能を反映した I / O 発行可能数を設定することができる。

発行制御部 3 は、論理ボリュームに対して受け付けた I / O 要求 7 を要求蓄積部 4 に蓄積する。発行制御部 3 は、要求蓄積部 4 が蓄積する I / O 要求 7 から論理ボリュームに I / O 要求の発行をおこなう。このとき、発行制御部 3 は、単位時間あたりの I / O 発行数を I / O 発行可能数 6 の範囲に制限する。

【0018】

これにより、ストレージ制御装置 1 は、ストレージ装置 5 が過大な応答性能を有する場合があっても、論理ボリュームが過大な応答性能を有することを抑制する。このようなストレージ制御装置 1 は、論理ボリュームを構成するストレージ装置 5 の組合せに変更が生じて安定した性能値を発揮することができる。そのため、ストレージ制御装置 1 は、ストレージ装置 5 の性能最適化に伴う作業負担を軽減することができる。

40

【0019】

[第 2 の実施形態]

次に、第 2 の実施形態のストレージ装置の接続例について図 2 を用いて説明する。図 2 は、第 2 の実施形態のストレージシステムの構成例を示す図である。

【0020】

ストレージシステム 10 は、1 以上のホストコンピュータ 11 と、ホストコンピュータ 11 と通信可能に接続する 1 以上の R A I D 装置 12 とを含んで構成される。R A I D 装

50

置 1 2 は、複数の H D D 1 9 を 1 つの論理ボリュームとして扱うストレージ装置の 1 つである。R A I D 装置 1 2 は、R A I D への I / O 要求をホストコンピュータ 1 1 から受け付ける。R A I D 装置 1 2 は、コントローラモジュール（以下、C M (Controller Module)）1 3 と、ドライブエンクロージャ（D E）1 4 とを含む。

【 0 0 2 1 】

D E 1 4 は、複数の H D D 1 9 を収容する筐体である。D E 1 4 は、複数の H D D 1 9 への電源供給を担うとともに、H D D 1 9 と C M 1 3 とを接続するインタフェースを有する。R A I D 装置 1 2 は、D E 1 4 が収容する複数の H D D 1 9 の組合せにより R A I D を構成する。なお、D E 1 4 は、R A I D 装置 1 2 が内蔵するものであってもよいし、R A I D 装置 1 2 に外付けされるものであってもよい。また、図示する R A I D 装置 1 2 は、1 つの D E 1 4 を備えるが 2 以上の D E 1 4 を備えてもよい。

10

【 0 0 2 2 】

C M 1 3 は、ストレージ制御装置の一形態であり、ホストコンピュータ 1 1 からの I / O 要求（たとえば、W r i t e 要求、R e a d 要求など）を受け付けて、D E 1 4 が備える複数の H D D 1 9 へのアクセスを制御する。C M 1 3 は、ホスト I / O 制御部 1 5 と、R A I D 制御部 1 6 と、ディスク I / O 制御部 1 7 を含む。なお、図示する R A I D 装置 1 2 は、1 つの C M 1 3 を備えるが 2 以上の C M 1 3 を備える冗長構成としてもよい。

【 0 0 2 3 】

ホスト I / O 制御部 1 5 は、ホストコンピュータ 1 1 と C M 1 3 とのインタフェース制御をおこなう。C M 1 3 は、ホスト I / O 制御部 1 5 を介してホストコンピュータ 1 1 との間で所要のデータの送受信をおこなう。R A I D 制御部 1 6 は、R A I D グループごとの I / O 発行数を管理する。ディスク I / O 制御部 1 7 は、D E 1 4 と C M 1 3 とのインタフェース制御をおこなう。C M 1 3 は、ディスク I / O 制御部 1 7 を介して D E 1 4 との間で所要のデータの送受信をおこなう。

20

【 0 0 2 4 】

ここで、異なる H D D 種別における R A I D 性能の測定例について図 3 を用いて説明する。図 3 は、異なる H D D 種別における R A I D 性能の測定例を示す図である。

H D D __ A と H D D __ B とは、R A I D 装置 1 2 において相互に交換可能な H D D である。所定条件下でアクセス性能を測定したとき、H D D __ A は P 1 (M B / s)、H D D __ B は P 2 (M B / s) であり、H D D __ A と H D D __ B の間に d 1 (M B / s) の性能差がある。

30

【 0 0 2 5 】

H D D __ A は、要求性能下限 P L と要求性能上限 P H を満足するが、H D D __ B は、要求性能下限 P L を満足するものの要求性能上限 P H を満足していない。たとえば、H D D __ B を含む H D D の組合せで、ユーザが要求する性能を満足する R A I D を構成していたときに、H D D __ B を H D D __ A に交換するとユーザが要求する性能を満たさなくなってしまう場合がある。

【 0 0 2 6 】

このように、H D D ごとの I / O 発行数に上限値を設けたとしても I / O 要求の発行後の処理は、H D D 側に任せられるため、R A I D 性能は、H D D 内部の処理方法に委ねられる。そのため、従来であれば、H D D __ A と H D D __ B が互換性を有して運用できるように、設計段階で H D D のファームウェア性能の最適化を図っている。

40

【 0 0 2 7 】

たとえば、R A I D 装置ベンダは、新規の H D D の性能測定を実施し、測定結果にもとづいて要求性能値を算出して H D D ベンダに通知する。H D D ベンダは、通知された要求性能値にもとづいて H D D のファームウェアを更新して R A I D 装置ベンダに提供する。R A I D 装置ベンダと H D D ベンダは、要求性能値が満たされるまで上記作業を反復する。

【 0 0 2 8 】

このような H D D の性能最適化に伴う作業負担を軽減するための処理を実行する R A I

50

D制御部について、図4を用いて説明する。図4は、第2の実施形態のコントローラモジュールの構成例を示す図である。

【0029】

CM(コントローラモジュール)13は、前述した通り、ホストI/O制御部15と、RAID制御部16と、ディスクI/O制御部17を含む。ホストI/O制御部15は、I/O受付部20と、I/O完了受付部21を含む。I/O受付部20は、ホストコンピュータ11からI/O要求を受け付けてRAID制御部16にI/O要求を通知する。I/O完了受付部21は、RAID制御部16からI/O完了応答を受け付けてホストコンピュータ11にI/O完了応答を通知する。

【0030】

RAID制御部16は、I/O受付部22と、I/O発行可能数算出部25と、I/O完了受付部28を含む。I/O受付部22は、ホストI/O制御部15からI/O要求を受け付けて、RAIDグループごとに時間当たりのI/O発行数を制限してディスクI/O制御部17にI/O要求を通知する。I/O受付部22は、I/O発行数管理部23と、管理時間監視部24を含む。

【0031】

I/O発行数管理部23は、I/O受付部22が受け付けたI/O要求をI/O発行待ちキューに蓄積し、時間当たりのI/O発行数をI/O発行可能数27の範囲に制限しながらディスクI/O制御部17にI/O要求の発行(I/O発行)をおこなう。管理時間監視部24は、I/O発行数を制限する際の単位時間を管理時間として監視する。

【0032】

I/O発行可能数算出部25は、あらかじめ設定される性能情報テーブル26を保持し、性能情報テーブル26にもとづいてRAIDグループごとにI/O発行可能数27を算出し、I/O発行可能数27を保持する。性能情報テーブル26は、RAIDの構成条件別の要求性能を示す。性能情報テーブル26の詳細は、後で図7を用いて説明する。I/O完了受付部28は、ディスクI/O制御部17からI/O完了応答を受け付けてホストI/O制御部15にI/O完了応答を通知する。

【0033】

ディスクI/O制御部17は、I/O受付部29と、I/O完了受付部30を含む。I/O受付部29は、RAID制御部16からI/O要求を受け付けてRAIDを構成するHDD19にI/O要求を通知する。HDD19は、I/O要求を受け付け、I/O要求に対応するコマンド処理をおこなう。HDD19は、コマンド処理の実行後にI/O完了受付部30にI/O完了応答を通知する。I/O完了受付部30は、HDD19からI/O完了応答を受け付けてRAID制御部16にI/O完了応答を通知する。

【0034】

次に、第2の実施形態のRAID装置のハードウェア構成について図5を用いて説明する。図5は、第2の実施形態のRAID装置のハードウェア構成例を示す図である。

RAID装置12は、チャンネルアダプタ31と、コントローラモジュール13と、DE14を備える。RAID装置12は、チャンネルアダプタ31を介してホストコンピュータ11と接続する。コントローラモジュール13は、チャンネルアダプタ31を複数(たとえば、2つ)備え、複数のチャンネルアダプタ31による冗長構成のもと複数系統でホストコンピュータ11と接続する。コントローラモジュール13は、ディスクアダプタ34を介して、DE14が収容するHDD19と接続する。

【0035】

コントローラモジュール13は、プロセッサ32、メモリ33、ディスクアダプタ34を備え、図示しないバスを介して接続されている。プロセッサ32は、コントローラモジュール13全体を制御し、RAID制御を含むストレージ制御をおこなう。プロセッサ32は、マルチプロセッサであってもよい。プロセッサ32は、たとえばCPU(Central Processing Unit)、MPU(Micro Processing Unit)、DSP(Digital Signal Processor)、ASIC(Application Specific Integrated Circuit)、またはPLD(Progr

10

20

30

40

50

amable Logic Device)である。またプロセッサ32は、CPU、MPU、DSP、ASIC、PLDのうちの2以上の要素の組合せであってもよい。

【0036】

メモリ33は、たとえば、RAM(Random Access Memory)や不揮発性メモリを含む。メモリ33は、HDD19からデータを読み出したときにデータを保持するほか、I/O要求を一時的に蓄積するI/O発行待ちキューやHDD19にデータを書き込むときのバッファとなる。また、メモリ33は、ユーザデータや制御情報を格納する。ディスクアダプタ34は、HDD19とのインタフェース制御(アクセス制御)をおこなう。たとえば、RAMは、コントローラモジュール13の主記憶装置として使用される。RAMには、プロセッサ32に実行させるオペレーティングシステム(Operating System)のプログラムやファームウェア、アプリケーションプログラムの少なくとも一部が一時的に格納される。また、RAMには、プロセッサ32による処理に必要な各種データが格納される。また、RAMは、各種データの格納に用いるメモリと別体にキャッシュメモリを含むものであってもよい。

10

【0037】

不揮発性メモリは、RAID装置12の電源遮断時においても記憶内容を保持する。不揮発性メモリは、たとえば、EEPROM(Electrically Erasable and Programmable Read Only Memory)やフラッシュメモリなどの半導体記憶装置や、HDDなどである。不揮発性メモリには、オペレーティングシステムのプログラムやファームウェア、アプリケーションプログラム、および各種データが格納される。

20

【0038】

以上のようなハードウェア構成によって、第2の実施形態のRAID装置12、コントローラモジュール13の処理機能を実現することができる。なお、第1の実施形態に示したストレージ制御装置1も、図示したRAID装置12、コントローラモジュール13と同様のハードウェアにより実現することができる。

【0039】

次に、第2の実施形態のRAID制御部16が実行するRAIDグループ追加処理について図6を用いて説明する。図6は、第2の実施形態のRAIDグループ追加処理のフローチャートを示す図である。RAIDグループ追加処理は、新規のRAIDグループが追加されたときにRAID制御部16により、より詳しくはI/O発行可能数算出部25により実行される。RAIDグループ追加処理は、新規に追加されたRAIDグループについてI/O発行可能数を決定し、決定したI/O発行可能数のもとでI/O制御を開始する処理である。

30

【0040】

[ステップS11] RAID制御部16は、新規に追加されたRAIDグループのRAIDレベルを取得する。RAIDレベルは、たとえば、RAID1、RAID5、RAID6、RAID1+0、RAID5+0などがあるが、その他のRAIDレベルであってもよい。

【0041】

[ステップS12] RAID制御部16は、ディスク種別を取得する。ディスク種別は、HDDの回転数である。HDDの回転数は、たとえば、15krpm(revolution per minute)、10krpm、7.2krpmなどがある。なお、ディスク種別は、ディスクの性能を区別する指標であれば、HDDの回転数に限らず、HDDのサイズや記録容量、バッファ容量、最大読み出し速度や最大書き込み速度、MTBF(平均故障間隔)などであってもよい。また、ディスクの性能を区別する指標は、1つに限らず複数であってもよい。また、ディスクは、HDDに限らず、SSDなどであってもよい。

40

【0042】

[ステップS13] RAID制御部16は、ディスク実構成数を取得する。ディスク実構成数は、新規に追加されたRAIDグループに割り当てられたディスクの数である。

[ステップS14] RAID制御部16は、性能情報テーブルを参照する。性能情報テ

50

ーブルは、あらかじめRAIDレベル、ディスク種別に用意される。たとえば、性能情報テーブルは、メモリ33に保持されるが、ホストコンピュータ11など外部装置から取得されるものであってもよい。

【0043】

ここで、性能情報テーブルについて図7を用いて説明する。図7は、第2の実施形態の性能情報テーブルの一例を示す図である。性能情報テーブル100は、あらかじめ用意されている性能情報テーブルの一例である。性能情報テーブル100は、RAIDレベル、ディスク種別(回転数)、基本構成ディスク数、転送単位(kbyte)、および要求性能(MB/s)を項目として有する。RAIDレベル、ディスク種別(回転数)は、あらかじめ用意されている性能情報テーブルのうちから対応する性能情報テーブルを参照するためのインデックスとなる項目である。性能情報テーブル100は、新規に追加されたRAIDグループが、RAIDレベル「RAID5」、ディスク種別「15krpm」であるときに参照される。基本構成ディスク数は、RAIDレベルに応じたディスク数であり、RAIDレベル「RAID5」の基本構成ディスク数は、「5」である。

【0044】

転送単位は、RAIDグループに設定される転送単位である。たとえば、転送単位は、基本転送単位Nを基準にして7段階に分けられている。まず、第1段階が「 $N \times 8$ より大きい範囲」、第2段階が「 $N \times 8$ から $N \times 4$ の範囲」、第3段階が「 $N \times 4$ から $N + 1$ の範囲」、第4段階が「N」、第5段階が「 $N - 1$ から $N / 4$ の範囲」、第6段階が「 $N / 4$ から $N / 8$ の範囲」、第7段階が「 $N / 8$ より小さい範囲」である。なお、転送単位は、たとえば、5段階や10段階など任意の段階に設定してもよい。また、各段階の刻み幅も任意の刻み幅に設定してもよい。

【0045】

なお、転送単位の初期値は、第4段階のN(たとえば、メインフレーム用途で「8」kbyte、バックアップ用途で「1000」kbyteなど)が設定される。転送単位は、RAIDグループの稼働状態に応じて所要のタイミングで更新される。要求性能は、転送単位ごとに要求される性能である。たとえば、転送単位Nのとき、要求性能は、「200」MB/sである。

【0046】

[ステップS15] RAID制御部16は、RAIDグループにおける単位時間当たりのI/O処理数を算出する。RAID制御部16は、性能情報テーブル100の要求性能を転送単位で除することにより単位時間当たりのI/O処理数を算出する。たとえば、要求性能「200」MB/s、転送単位Nのとき、単位時間当たりのI/O処理数は、「 $200 / N$ 」である。

【0047】

[ステップS16] RAID制御部16は、RAIDグループにおける構成ディスク比率を算出する。RAID制御部16は、ステップS13で取得したディスク実構成数を性能情報テーブル100の基本構成ディスク数で除することにより構成ディスク比率を算出する。たとえば、ディスク実構成数「5」、基本構成ディスク数「5」のとき、構成ディスク比率は、「1」であり、ディスク実構成数「6」、基本構成ディスク数「5」のとき、構成ディスク比率は、「1.2」である。構成ディスク比率は、「1」以上の値をとり、値が大きいほどI/O発行可能数を大きくすることができる。

【0048】

[ステップS17] RAID制御部16は、管理時間当たりのI/O発行可能数を算出する。RAID制御部16は、ステップS15で算出した単位時間当たりのI/O処理数と、ステップS16で算出した構成ディスク比率と、管理時間の積により管理時間当たりのI/O発行可能数を算出する。たとえば、単位時間当たりのI/O処理数「 $200 / N$ 」、構成ディスク比率「1」、管理時間「1(秒)」のとき、管理時間当たりのI/O発行可能数は、「 $200 / N (= (200 / N) \times 1 \times 1)$ 」である。なお、管理時間は、あらかじめ設定される値である。

【 0 0 4 9 】

[ステップ S 1 8] R A I D 制御部 1 6 は、 I / O 発行数管理テーブルを生成する。 I / O 発行数管理テーブルは、 R A I D グループごとに生成される。

ここで、 I / O 発行数管理テーブルについて図 8 を用いて説明する。図 8 は、第 2 の実施形態の I / O 発行数管理テーブルの一例を示す図である。 I / O 発行数管理テーブル 1 1 0 は、 R A I D 制御部 1 6 が R A I D グループごとに生成する I / O 発行数管理テーブルの一例である。 I / O 発行数管理テーブル 1 1 0 は、 R A I D グループ識別情報、転送単位 (k b y t e)、および I / O 発行可能数を項目として有する。 R A I D グループ識別情報は、 R A I D グループを一意に特定可能な識別情報である。 I / O 発行数管理テーブル 1 1 0 は、 R A I D グループ「 # 1 」の I / O 発行数管理テーブルである。転送単位 (k b y t e) は、所定の転送単位を示す。 I / O 発行数管理テーブル 1 1 0 の転送単位は、性能情報テーブル 1 0 0 の転送単位と同じである。 I / O 発行可能数は、ステップ S 1 7 で算出した値を、転送単位「 N 」に対応する I / O 発行可能数「 M 」に設定する。その他の転送単位の I / O 発行可能数は、転送単位「 N 」に対応する I / O 発行可能数「 M 」に所定の計数を乗じた値が設定される。たとえば、転送単位「 $N \times 4$ から $N + 1$ の範囲」に対応する I / O 発行可能数は、「 $k 3 \times M$ 」である。所定の計数「 k 1 」, 「 k 2 」, 「 k 3 」, 「 k 4 」, 「 k 5 」, 「 k 6 」は、転送単位ごとにあらかじめ設定される。

10

【 0 0 5 0 】

R A I D 制御部 1 6 は、生成した I / O 発行数管理テーブルをメモリ 3 3 に保持する。したがって、 R A I D 制御部 1 6 は、 R A I D グループごとに I / O 発行数管理テーブルをメモリ 3 3 に保持する。

20

【 0 0 5 1 】

[ステップ S 1 9] R A I D 制御部 1 6 は、 I / O 発行数管理テーブルを参照して、設定されている転送単位に対応する I / O 発行可能数を取得して設定する。たとえば、 R A I D 制御部 1 6 は、 I / O 発行数管理テーブル 1 1 0 を参照して、初期値として設定されている転送単位「 N 」に対応する I / O 発行可能数「 M 」を取得し、 R A I D グループ「 # 1 」の I / O 発行可能数に「 M 」を設定する。

【 0 0 5 2 】

[ステップ S 2 0] R A I D 制御部 1 6 は、新規に追加された R A I D グループの I / O 制御を開始し、 R A I D グループ追加処理を終了する。

30

次に、第 2 の実施形態の R A I D 制御部 1 6 が実行する R A I D グループ I / O 受付処理について図 9 を用いて説明する。図 9 は、第 2 の実施形態の R A I D グループ I / O 受付処理のフローチャートを示す図である。 R A I D グループ I / O 受付処理は、 R A I D 制御部 1 6 がホスト I / O 制御部 1 5 から I / O 要求を受け付けることにより実行される。 R A I D グループ I / O 受付処理は、ホスト I / O 制御部 1 5 から受け付けた I / O 要求を I / O 発行待ちキューにキューイングするとともに、 R A I D グループ追加処理で設定した I / O 発行可能数を更新する処理である。

【 0 0 5 3 】

[ステップ S 2 1] R A I D 制御部 1 6 (より詳しくは I / O 発行可能数算出部 2 5) は、 I / O 受付部 2 2 がホスト I / O 制御部 1 5 から受け付けた I / O 要求を、 I / O 要求に対応する R A I D グループごとの I / O 発行待ちキューにキューイングする。

40

【 0 0 5 4 】

[ステップ S 2 2] R A I D 制御部 1 6 は、 R A I D グループごとの I / O 要求の統計情報を更新する。統計情報は、所定の監視時間当たりの転送単位使用頻度を含み、統計情報管理テーブルに記録される。統計情報管理テーブルは、メモリ 3 3 に保持される。ここで、統計情報管理テーブルについて図 1 0 を用いて説明する。図 1 0 は、第 2 の実施形態の統計情報管理テーブルの一例を示す図である。

【 0 0 5 5 】

統計情報管理テーブル 1 2 0 は、 R A I D 制御部 1 6 が R A I D グループごとに管理する統計情報管理テーブルの一例である。統計情報管理テーブル 1 2 0 は、 R A I D グループ

50

ブ識別情報、転送単位 (k b y t e)、および受付 I / O 数を項目として有する。R A I D グループ識別情報は、R A I D グループを一意に特定可能な識別情報である。

【 0 0 5 6 】

統計情報管理テーブル 1 2 0 は、R A I D グループ「 # 1 」の I / O 発行数管理テーブルである。転送単位 (k b y t e) は、所定の転送単位を示す。統計情報管理テーブル 1 2 0 の転送単位は、性能情報テーブル 1 0 0 の転送単位、および I / O 発行数管理テーブル 1 1 0 の転送単位と同じである。受付 I / O 数は、所定の監視時間当たりの転送単位使用頻度である。より詳しくは、受付 I / O 数は、I / O 受付部 2 2 が受け付けた I / O 要求のデータサイズに対応する転送単位ごとの I / O 要求の受付数である。たとえば、転送単位「 N 」の受付 I / O 数「 3 0 」は、所定の監視時間の間に、データサイズ「 N 」の I / O 要求を 3 0 回受け付けたことを示す。

10

【 0 0 5 7 】

なお、所定の監視時間は、たとえば、6 0 分など統計的に有意な情報を収集できる時間を任意に設定することができる。監視時間は、R A I D グループで共通の値としてもよいし、R A I D グループごとに異なる値としてもよい。また、監視時間は、あらかじめ設定する固定値としてもよいし、時間帯や時間当たりの I / O 要求数に応じた可変値としてもよい。

【 0 0 5 8 】

[ステップ S 2 3] R A I D 制御部 1 6 は、I / O 発行可能数の更新タイミングであるか否かを判定する。更新タイミングは、あらかじめ設定した更新周期 (たとえば、1 分など) であってもよいし、所定のトリガ検出 (たとえば、統計情報管理テーブル 1 2 0 の受付 I / O 数のいずれかが所定値を超えたなど) であってもよい。R A I D 制御部 1 6 は、I / O 発行可能数の更新タイミングである場合にステップ S 2 4 にすすみ、I / O 発行可能数の更新タイミングでない場合に R A I D グループ I / O 受付処理を終了する。

20

【 0 0 5 9 】

[ステップ S 2 4] R A I D 制御部 1 6 は、統計情報管理テーブルを参照して受付 I / O 数のボリュームゾーンを選択する。たとえば、R A I D 制御部 1 6 は、統計情報管理テーブル 1 2 0 を参照して、受付 I / O 数のボリュームゾーンとして受付 I / O 数が最大の「 1 0 0 0 」となる転送単位「 N / 4 から N / 8 の範囲」を選択する。なお、ボリュームゾーンとして選択する転送単位は、受付 I / O 数が最大となる場合に限らず、平均値や中央値など所定の選択基準に従い選択されるものであってもよい。

30

【 0 0 6 0 】

[ステップ S 2 5] R A I D 制御部 1 6 は、I / O 発行数管理テーブルを参照して、選択した転送単位に対応する I / O 発行可能数を取得して設定する。たとえば、R A I D 制御部 1 6 は、転送単位「 N / 4 から N / 8 の範囲」を選択した場合、I / O 発行数管理テーブル 1 1 0 を参照して、転送単位「 N / 4 から N / 8 の範囲」に対応する I / O 発行可能数「 k 5 × M 」を取得し、R A I D グループ「 # 1 」の I / O 発行可能数に「 k 5 × M 」を設定する。R A I D 制御部 1 6 は、I / O 発行可能数を設定した後、R A I D グループ I / O 受付処理を終了する。

【 0 0 6 1 】

このように、I / O 発行可能数は、受け付けた I / O 要求に応じて適宜更新される。

40

次に、第 2 の実施形態の R A I D 制御部 1 6 が実行する R A I D グループ I / O 発行数管理処理について図 1 1 を用いて説明する。図 1 1 は、第 2 の実施形態の R A I D グループ I / O 発行数管理処理のフローチャートを示す図である。R A I D グループ I / O 発行数管理処理は、R A I D グループの I / O 制御の開始を受けて実行される。R A I D グループ I / O 受付処理は、I / O 発行待ちキューにキューイングされている I / O 要求を I / O 発行可能数の範囲に制限しながら R A I D グループに I / O 要求を発行する処理である。

【 0 0 6 2 】

[ステップ S 3 1] R A I D 制御部 1 6 (より詳しくは I / O 発行数管理部 2 3) は、

50

I/O発行待ちキューにキューイングされているI/O要求の数(受付I/O発行数)が「0」より大きいかな否かを判定する。すなわち、RAID制御部16は、I/O発行待ちキューにキューイングされているI/O要求があるかな否かを判定する。RAID制御部16は、I/O発行待ちキューにキューイングされているI/O要求がある場合にステップS32にすすみ、I/O要求がない場合にI/O発行待ちキューへのキューイングを待ち受ける。

【0063】

[ステップS32] RAID制御部16は、受付I/O発行数がRAIDグループに設定されているI/O発行可能数を超えているかな否かを判定する。RAID制御部16は、受付I/O発行数がI/O発行可能数を超えている場合にステップS33にすすみ、受付I/O発行数がI/O発行可能数を超えていない場合にステップS34にすすむ。

10

【0064】

[ステップS33] RAID制御部16は、I/O発行待ちキューにキューイングされているI/O要求のうちからI/O発行可能数だけI/O要求をディスクI/O制御部17に向けて発行する。

【0065】

[ステップS34] RAID制御部16は、I/O発行待ちキューにキューイングされているI/O要求をディスクI/O制御部17に向けて発行する。

[ステップS35] RAID制御部16は、I/O要求の発行後の管理時間の経過を待つ。管理時間は、RAIDグループ追加処理のステップS17でI/O発行可能数を算出した際に用いた管理時間である。RAID制御部16は、I/O要求の発行後の管理時間を経過している場合にステップS31にすすみ、I/O要求の発行後の管理時間を経過していない場合に管理時間の経過を待つ。

20

【0066】

このようにして、RAID制御部16は、RAIDグループに向けて発行するI/O要求の数を、管理時間当たりI/O発行可能数の範囲に制限することができる。したがって、RAIDグループを構成する一部のHDD19の処理性能が過大であっても、RAIDグループのI/O処理数が過大になることがない。

【0067】

したがって、RAID装置12は、性能の上限が多様なHDD19を搭載しても、RAIDグループを構成したときに所定の性能値を得ることができる。このようなRAID装置12は、搭載するHDD19のマルチベンダ化の推進の障害となることがない。また、このようなRAID装置12は、搭載するHDD19の性能最適化に伴う、RAID装置ベンダおよびHDDベンダの作業負担を軽減することができる。

30

【0068】

また、このようなRAID装置12は、RAIDグループを構成するHDD19を故障などにより交換したような場合に、RAIDグループの応答性能が著しく変化することがない。したがって、RAID装置12は、メンテナンスの前後においてユーザに対して安定した性能を発揮するRAIDグループを提供することができる。

【0069】

また、RAID装置12は、新規にRAIDグループを構成したときに、新規に構成したRAIDグループとすでにあるRAIDグループとの間で性能差が過大になることがない。したがって、RAID装置12は、ユーザに対して安定した性能を発揮するRAIDグループを提供することができる。

40

【0070】

なお、上記の処理機能は、コンピュータによって実現することができる。その場合、ストレージ制御装置1、RAID装置12、コントローラモジュール13などが有すべき機能の処理内容を記述したプログラムが提供される。そのプログラムをコンピュータで実行することにより、上記処理機能がコンピュータ上で実現される。処理内容を記述したプログラムは、コンピュータで読み取り可能な記録媒体に記録しておくことができる。コンピ

50

ュータで読み取り可能な記録媒体としては、磁気記憶装置、光ディスク、光磁気記録媒体、半導体メモリなどがある。磁気記憶装置には、ハードディスク装置（HDD）、フレキシブルディスク（FD）、磁気テープなどがある。光ディスクには、DVD、DVD-RAM、CD-ROM/RWなどがある。光磁気記録媒体には、MO（Magneto-Optical disk）などがある。

【0071】

プログラムを流通させる場合には、たとえば、そのプログラムが記録されたDVD、CD-ROMなどの可搬型記録媒体が販売される。また、プログラムをサーバコンピュータの記憶装置に格納しておき、ネットワークを介して、サーバコンピュータから他のコンピュータにそのプログラムを転送することもできる。

10

【0072】

プログラムを実行するコンピュータは、たとえば、可搬型記録媒体に記録されたプログラムもしくはサーバコンピュータから転送されたプログラムを、自己の記憶装置に格納する。そして、コンピュータは、自己の記憶装置からプログラムを読み取り、プログラムに従った処理を実行する。なお、コンピュータは、可搬型記録媒体から直接プログラムを読み取り、そのプログラムに従った処理を実行することもできる。また、コンピュータは、ネットワークを介して接続されたサーバコンピュータからプログラムが転送されるごとに、逐次、受け取ったプログラムに従った処理を実行することもできる。

【0073】

また、上記の処理機能の少なくとも一部を、DSP、ASIC、PLDなどの電子回路で実現することもできる。

20

【符号の説明】

【0074】

- 1 ストレージ制御装置
- 2 設定部
- 3 発行制御部
- 4 要求蓄積部
- 5 ストレージ装置
- 6 I/O発行可能数
- 7 I/O要求
- 9 ホストコンピュータ
- 10 ストレージシステム
- 11 ホストコンピュータ
- 12 RAID装置
- 13 コントローラモジュール（CM）
- 14 ドライブエンクロージャ（DE）
- 15 ホストI/O制御部
- 16 RAID制御部
- 17 ディスクI/O制御部
- 19 HDD
- 20, 22, 29 I/O受付部
- 21, 28, 30 I/O完了受付部
- 23 I/O発行数管理部
- 24 管理時間監視部
- 25 I/O発行可能数算出部
- 26 性能情報テーブル
- 27 I/O発行可能数
- 31 チャンネルアダプタ
- 32 プロセッサ
- 33 メモリ

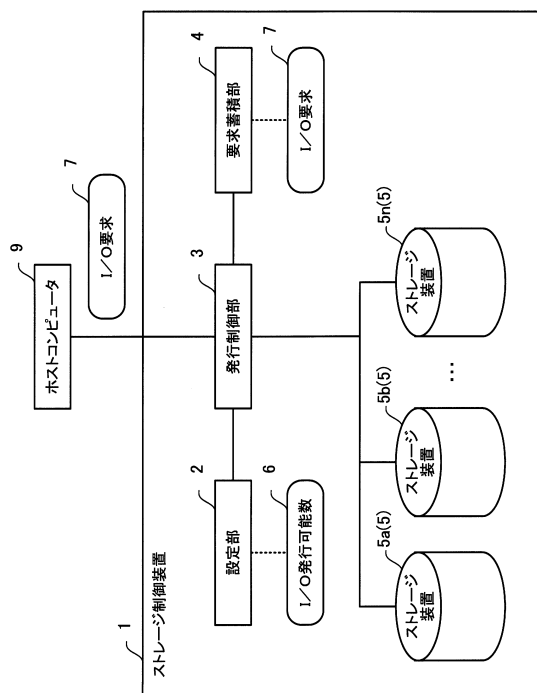
30

40

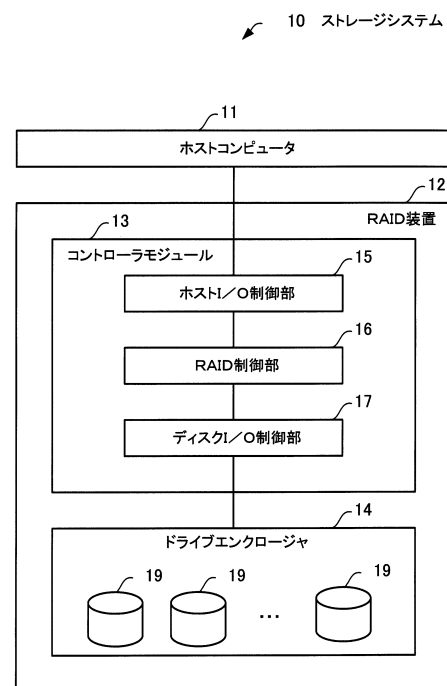
50

- 3 4 ディスクアダプタ
 1 0 0 性能情報テーブル
 1 1 0 I/O発行数管理テーブル
 1 2 0 統計情報管理テーブル

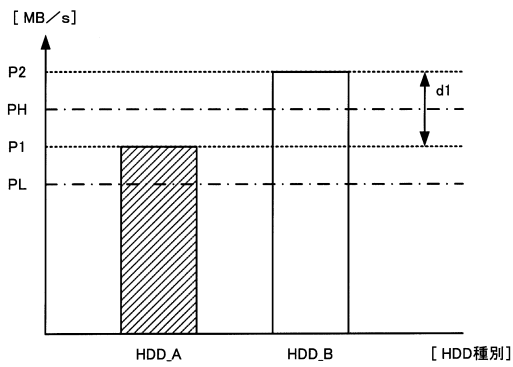
【図 1】



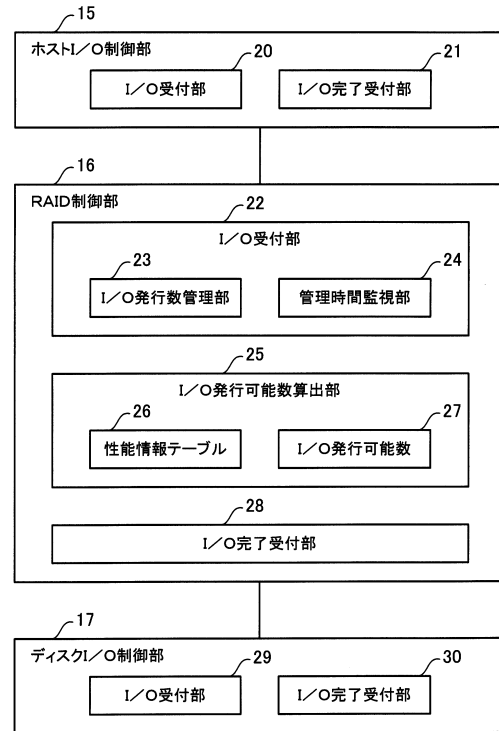
【図 2】



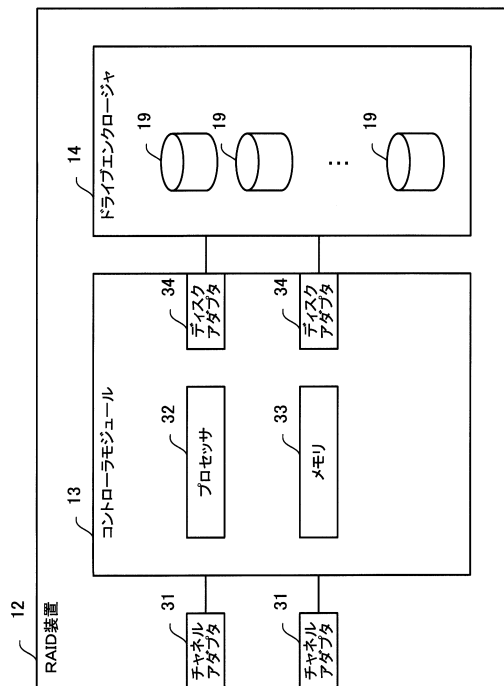
【図 3】



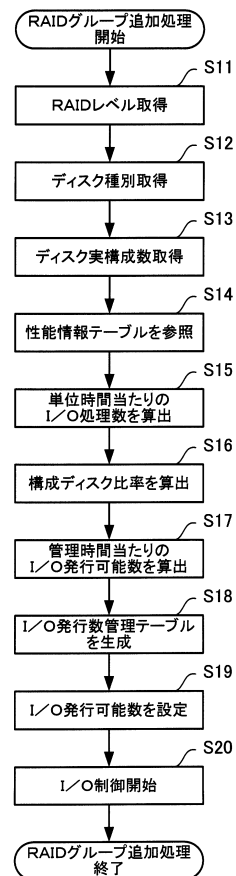
【図 4】



【図 5】



【図 6】



【図 7】

100 性能情報テーブル

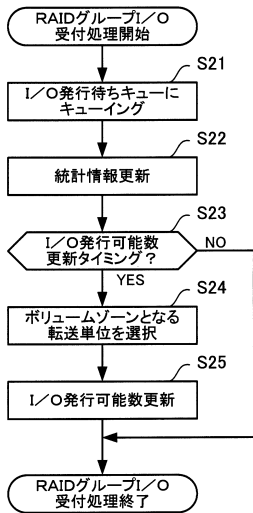
RAIDレベル	ディスク種別 (回転数)	基本構成 ディスク数	転送単位 (kbyte)	要求性能 (MB/s)
RAID5	15krpm	5	$>(N \times 8)$	400
			$(N \times 8) \sim (N \times 4)$	350
			$(N \times 4) \sim (N + 1)$	300
			N	200
			$(N - 1) \sim (N / 4)$	100
			$(N / 4) \sim (N / 8)$	10
			$<(N / 8)$	8

【図 8】

110 I/O発行数管理テーブル

RAIDグループ 識別情報	転送単位 (kbyte)	I/O発行可能数
#1	$>(N \times 8)$	$k1 \times M$
	$(N \times 8) \sim (N \times 4)$	$k2 \times M$
	$(N \times 4) \sim (N + 1)$	$k3 \times M$
	N	M
	$(N - 1) \sim (N / 4)$	$k4 \times M$
	$(N / 4) \sim (N / 8)$	$k5 \times M$
	$<(N / 8)$	$k6 \times M$

【図 9】

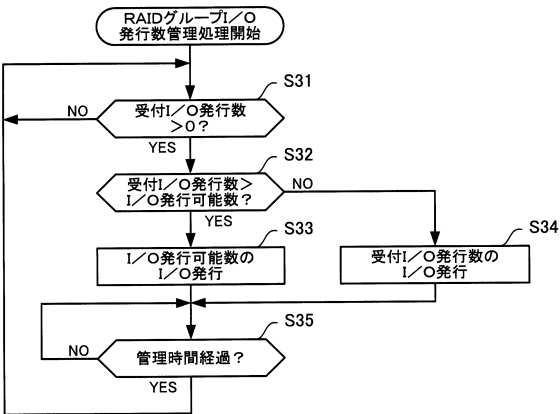


【図 10】

120 統計情報管理テーブル

RAIDグループ 識別情報	転送単位 (kbyte)	受付I/O数
#1	$>(N \times 8)$	0
	$(N \times 8) \sim (N \times 4)$	10
	$(N \times 4) \sim (N + 1)$	5
	N	30
	$(N - 1) \sim (N / 4)$	5
	$(N / 4) \sim (N / 8)$	1000
	$<(N / 8)$	0

【図 11】



フロントページの続き

(56)参考文献 特開平 1 1 - 1 6 7 4 7 0 (J P , A)
米国特許第 6 3 6 6 9 8 1 (U S , B 1)
特開 2 0 0 9 - 2 9 4 9 4 9 (J P , A)
米国特許出願公開第 2 0 0 9 / 0 3 0 7 5 3 4 (U S , A 1)
特開 2 0 0 5 - 3 3 2 2 3 6 (J P , A)
特開 2 0 0 8 - 1 8 6 2 1 1 (J P , A)
米国特許出願公開第 2 0 1 5 / 0 2 6 1 4 6 2 (U S , A 1)
米国特許第 7 5 5 2 2 8 0 (U S , B 1)
米国特許出願公開第 2 0 1 3 / 0 2 1 2 3 4 9 (U S , A 1)

(58)調査した分野(Int.Cl. , D B 名)

G 0 6 F 3 / 0 6 - 3 / 0 8
G 0 6 F 1 3 / 1 0 - 1 3 / 1 4