データベースシステムにおける プロアクティブなディスクアレイ省電力化手法に関する一考察

平井　遥† 星野　喬† 合田　和生†† 喜連川　優††

† 東京大学 大学院情報理工学系研究科
†† 東京大学 生産技術研究所

あらまし 近年のコンピュータ社会では扱われるデータ量が爆発的に増大し、データセンタなどにおいて多数のディスクドライブが格納された大規模ディスクアレイが使用されるようになっている。これに伴いディスクアレイの消費する電力も無視できないものとなっており、これを削減することが課題と考えられている。本論文では、ディスクアレイの消費電力のモデル化を行うとともに、データベースシステムを対象としたプロアクティブな省電力化手法について有効性を検証する。

キーワード 省電力化, ストレージ, 問い合わせ処理

A Study on Proactive Methods of Disk Array Power Saving for Database Systems

Haruka HIRAI†, Takashi HOSHINO†, Kazuo GODA††, and Masaru KITSUREGAWA††

† Graduate School of Information Science and Technology, the University of Tokyo
†† Institute of Industrial Science, the University of Tokyo

Abstract The volume of information treated in the computer society is increasing explosively. The disk arrays which comprises a massive number of disk drives is being deployed into data centers and their power consumption is not negligible. Power saving for disk storage is important. In this paper, we discuss our power consumption model of the disk array, and propose a proactive power reduction method for database systems.

Key words Energy saving, Storage, Query processing

1. はじめに

近年のコンピュータ社会では扱われる情報量が爆発的に増大しており、これは年率約 2 倍の割合で増加していると言われている。それに伴ってデータセンタなどでストレージに使用される記憶とその管理に係わるコストも増加しており、システム全体の消費電力量に対してストレージの消費電力量は無視できないものとなっている。

実際のデータセンタにおけるストレージの消費電力量については以下のような報告が行われている。典型的なデータセンタにおいては全体の 27 %の電力がストレージによって消費されていると言われている [1]。さらに巨大データに対してトランザクション処理を行うような超高性能 OLTP システムにおいては全体の 71 %の電力がストレージによって消費されていると言われている [2]。したがって、データセンタなどにおいてシステム全体の消費電力の削減を考えなくてはならない場合には、ストレージの消費電力を削減することが重要であると考えられる。

そこで著者らはディスクアレイの省電力化手法について研究を行った。まず、ディスクアレイ内部の各構成部位の消費電力を計測することにより、消費電力の内訳について調査を行った。次に、ディスクアレイの消費電力を実測に基づいてモデル化し、構築したモデルからディスクアレイの消費電力シミュレータを構築した。さらにデータベースシステムとの連携によるディスクアレイの省電力化手法を提案し、構築した消費電力シミュレータを用いて検証を行った。提案手法は、省電力化指向の問合せ処理とディスクアレイ全体の省電力化制御によって既存手法よりも高効率の省電力化を図る手法である。検証の結果、提案手法は TPCH の Q17 相当の問合せを行う場合に最大で約 49.5 %の省電力化効果があることが分かった。

本論文の構成は以下の通りである。まず、2. ではディスクアレイの各構成部位の消費電力について調査を行ったので、その方法と結果を示す。次に 3. ではディスクアレイの消費電力モデルと消費電力シミュレータの構築についてその方法と結果を示す。また 4. ではデータベースシステムとの連携によるディスク

アレイの省電力化手法について提案を行い、5. で提案手法の検証を行う。そして 6. で関連研究について簡単に述べ、最後に 7. でまとめと今後の課題について述べる。

2. ディスクアレイの各構成部位の消費電力の計測

2.1 目　　的

実際にシステムが稼動している最中のディスクアレイの具体的な消費電力量や各構成部位ごとの消費電力の内訳などについては詳細な調査は行われていない。そこで著者らは、まずディスクアレイに DBMS を用いてアクセスを行い、その際の消費電力を実際に電力計を用いて計測することによって、

- システム稼動中にディスクアレイがどの程度の電力を消費しているか
- その際、ディスクアレイの各構成部位がそれぞれどの程度の電力を消費しているか

ということについて調査を行った。

2.2 ディスクアレイの電力計測システム

計測のために著者らは図 1 のような電力計測システムを構築した。まず、データベースサーバとディスクアレイとを SCSI によって接続した。これにより、データベースサーバ側から DBMS や I/O ベンチマークを用いてディスクアレイに対して負荷をかけられるようにした。次に、ディスクアレイ全体とディスクアレイ内部の各構成部位ごとの消費電力を計測できるように 3 台の電力計を用いて、ディスクアレイ内部に図に示したような配線をおこなった。これにより、ディスクアレイに対して負荷をかけた際の各構成部位ごとの消費電力をリアルタイムで計測できるようにした。また、データベースサーバ側の I/O トレーサによってディスクアレイに対して行った I/O トレースのログを取ることができるようにした。

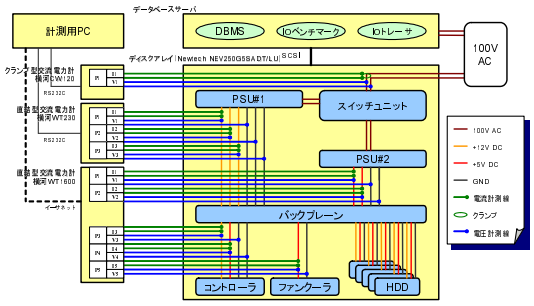

図 1　ディスクアレイの電力計測システム

2.3 実 験 環 境

次に計測環境について述べる。データベースサーバ用の PC としては DELL PRECISION 390(CentOS 3.9) を使用した。また、DBMS にはオープンソースのソフトウェアである MySQL を用い、代表的なデータベースベンチマークである TPCH の Q1 と Q8 による問い合わせを行った。測定対象とするディスクアレイには Newtech 製の EvolutionII Desktop NEV250G5SADT/LU を使用し、以下の各構成部位について問合せを行っている際の消費電力を計測した。

(1) パワーサプライ（PSU）

100V 交流電源によって供給される電力を直流に変換し、ディスクアレイ内部の他の各部位へと供給する部位。使用したディスクアレイでは PSU によって変換された電力はいったんバックプレーンへと送られ、そこから各部位へと供給される仕組みになっている。

(2) ファンクーラ（FAN）

ディスクアレイには専用のファンクーラが装着されている。使用したディスクアレイのファンクーラには一般 PC 向けの同種のケースファンと比較して、回転数がやや高めのものが使用されている。

(3) コントローラ（CTL）

RAID を制御するための回路部分。ディスクアレイへの I/O の有無により消費する電力が変化する。詳細については 3. で詳しく述べる。

(4) ディスクドライブ（HDD）

使用したディスクアレイには HITACHI 製のハードディスクドライブ Deskstar T7K500 が 5 台搭載されている。ディスクアレイの構造上、直接電力を測定することが難しかったため、PSU で供給される電力から他の部位で消費される電力を引いた残りの電力を HDD の消費電力とした。

2.4 計測結果と考察

計測の結果は図 2、図 3 のようになった。図 2 は Q1 を実行した際の時間経過に伴う消費電力の変化をグラフ化したもの、図 3 は同じく Q8 を実行した際のものである。また、表 1 にはクエリを実行していない状態、Q1 を実行している状態、Q8 を実行している状態のそれぞれの部位ごとの消費電力の平均値をまとめた。ディスクアレイ全体の消費電力はクエリを実行していない状態では約 51.5 ワット、Q1 と Q8 を実行している状態ではそれぞれ約 53.31 ワット、58.62 ワットとなった。各構成部位ごとの電力については、パワーサプライが約 10 ワット、ファンクーラが約 3 ワット、コントローラが約 8.5 ワット、そして残りの約 30 ワットをディスクドライブが消費していることが分かった。

この計測によってまず、計測に使用したディスクアレイにおいてはディスクアレイ全体が消費している電力のうち、ディスクドライブが約 60 %程度の電力を消費していることが分かった。したがって、ディスクアレイの消費電力を削減するためにはやはりディスクドライブの省電力化を考えることが有効であると考えられる。しかし、その一方でディスクドライブ以外の部位によって全体の約 40 %程度の電力が消費されており、ディスクドライブ以外の部位の消費電力も決して無視することはできないことが分かった。そのため、ディスクドライブの省電力化を行うことに加えてこれらの部位の省電力化を行うことによって、より高効率の省電力化を行うことが出来ると考えられる。4. で提案するディスクアレイ全体の電力制御による省電力化はこの実験結果に基づくものである。

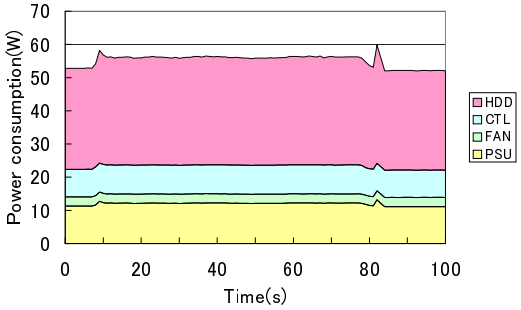

図 2　ディスクアレイの消費電力内訳 (Q1)

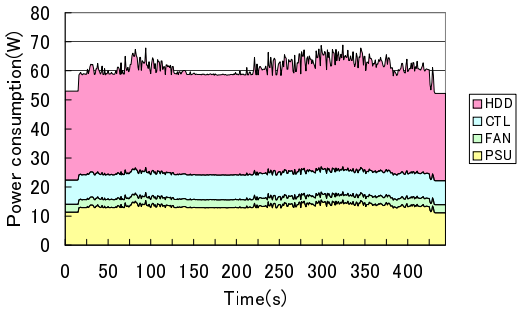

図 3　ディスクアレイの消費電力内訳 (Q8)

表 1　ディスクアレイの各構成部位ごとの消費電力

	total	PSU	FAN	CTL	HDD
Idle	51.50	11.33	2.76	8.29	30.50
Q1(Avg)	53.31	12.17	2.77	8.70	32.45
Q8(Avg)	58.62	13.59	2.76	8.57	36.46

3.　ディスクアレイの消費電力モデルの構築

3.1　目　　的

ディスクアレイの省電力化手法の検証のためにはディスクアレイの消費電力のシミュレーションを行う必要がある。そのため、シミュレーションのためのディスクアレイの消費電力モデルが必要となる。しかし、著者の知る範囲ではディスクアレイの構成部位のうち、ディスクドライブについての消費電力モデルはいくつか存在するものの、その他の部位についての消費電力モデルは現在のところ存在していない。そこで、著者らは 2.2 に示した計測システムを用いてディスクアレイの消費電力の計測を行い、実測に基づいてディスクアレイの各構成部位について消費電力モデルの構築を行った。

3.2　モデル化の方法

モデル化はディスクアレイにさまざまな負荷を与え、その際の各構成部位ごとの消費電力の実測データを解析することによって行った。モデル化のために行った実験の手順は以下のようなものである。まず、2.2 に示した計測システムを用いて、マイクロ I/O ベンチマークによってディスクアレイに対してブロックサイズを 16KB としてシーケンシャルアクセスとランダムアクセスを行った。そしてその際に 2.3 に示した各構成部位ごとの消費電力を計測した。この実測データから各構成部位ごとに負荷と消費電力の変化量の関係を解析し、それぞれの特徴に基づいてモデル化を行った。以降ではそれぞれの部位ごとのモデル化の結果について述べる。

3.3　ディスクドライブの消費電力モデル

ディスクドライブの消費電力のモデル化は [7] に示す方式に基づいて行った。この方式ではディスクドライブが取り得る各状態ごとに分類してモデル化を行い、特にディスクアクセスを行っている状態であるアクティブ状態の消費電力を負荷の大きさを表す指標を用いて高精度で予測することに成功している。そこで本論文でもディスクドライブの状態ごとに分類してモデル化を行い、アクティブ状態については負荷の大きさの指標としてディスクドライブビジー率を用いることによってモデル化を行うことにした。

以降ではディスクドライブの取り得る状態について説明した後、各状態のモデル化の結果について説明する。

(1) アクティブ状態

ディスクドライブが入出力処理を行っている状態。スピンドルモータは最高回転数で回転しており、ヘッドはディスク上に存在する。

(2) アイドル状態

ディスクドライブは入出力処理を行ってはいないが、即座に入出力処理を開始することができる状態である。スピンドルモータは最高回転数で回転しており、ヘッドはディスク上に存在する。

(3) スタンバイ状態

ディスクドライブが入出力処理を行っておらず、ヘッドはランプへと退避されており、スピンドルモータが完全に停止している状態である。入出力処理を検知するための電力を除けばほぼ電力を必要としないため必要消費電力はかなり少ないが、再びアクセスを行う際にはスピンアップを必要とするため相応の時間コストと電力コストを必要とする。

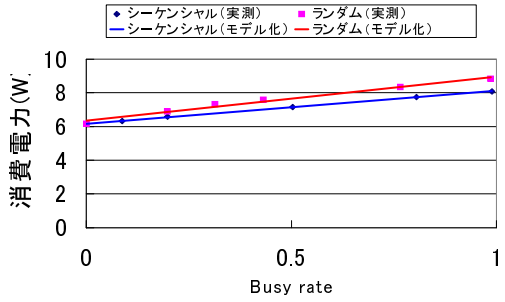

図 4　ディスクドライブビジー率と消費電力の関係

まず、アクティブ状態の消費電力モデルについて説明する。この場合は負荷の大きさや種類によって消費電力が変化する。そこでディスクドライブビジー率と消費電力の関係について計測したところ図 4 のようなグラフが得られた。そこでこの実測データから近似を行ったところ、消費電力の算出方法として以下のような式が得られた。

シーケンシャルアクセス : $P = 1.96 \cdot \beta + 6.17$

ランダムアクセス : $P = 2.63 \cdot \beta + 6.35$

(β : $busy\,rate[\%]$, P : $power\,consumption[W]$)

次にアイドル状態とスタンバイ状態の消費電力モデルについて説明する。この場合には消費電力は定常値を取るため、計測を行って表 2 に示す値を得た。また、状態間遷移の際の時間コストと電力コストについても計測を行い、表 3 に示す値を得た。

表 2　アイドル・スタンバイ状態の消費電力（ディスクドライブ）

状態	アイドル	スタンバイ
消費電力 [W]	6.1	1.7

表 3　ディスクドライブの状態間遷移コスト

	時間コスト [s]	電力コスト [J]
アクティブ→アイドル	0	0
アイドル→アクティブ	0	0
アイドル→スタンバイ	～0	～0
スタンバイ→アイドル	3.0	120

3.4　コントローラの消費電力モデル

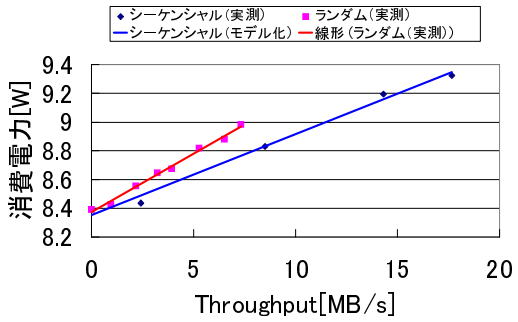

図 5　スループットとコントローラの消費電力の関係

次にコントローラの消費電力モデルについてモデル化の結果を説明する。コントローラはマイクロ I/O ベンチマークによる電力計測の結果、ディスクドライブと同じく、アクティブ状態、アイドル状態、スタンバイ状態の 3 状態を遷移することが分かった。以降では各状態ごとの消費電力モデルについて説明する。

まずアクティブ状態の消費電力モデルについて説明する。この場合はディスクドライブと同様に負荷の大きさや種類によって消費電力が変化する。計測の結果、コントローラの場合にはスループットと消費電力の間に図 5 に示すような関係を得ることが出来た。そこでこの実測データから近似を行ったところ、消費電力の算出方法として以下のような式が得られた。

シーケンシャルアクセス : $P = 0.0563 \cdot \theta + 8.35$

ランダムアクセス : $P = 0.0814 \cdot \theta + 8.37$

(θ : $throughput[MB/s]$, P : $power\,consumption[W]$)

次にアイドル状態の消費電力モデルについて説明する。この場合にはディスクドライブと同じく消費電力は定常値を取るため、計測を行って 8.39 ワットの電力を消費するということが分かった。また、同様にしてスタンバイ状態の消費電力は 0 ワットであるということが分かった。また、状態間遷移の際の時間コストと電力コストについて計測を行い、表 4 に示す値を得た。なおこの表の値のうち、アイドルとスタンバイの間の状態遷移コストはディスクアレイ全体の電源を ON,OFF した場合の値である。

表 4　コントローラの状態間遷移コスト

	時間コスト [s]	電力コスト [J]
アクティブ→アイドル	0	0
アイドル→アクティブ	0	0
アイドル→スタンバイ	2.0	15.7
スタンバイ→アイドル	26.0	193.2

3.5　ファンクーラとパワーサプライの消費電力モデル

最後にファンクーラとパワーサプライの消費電力モデルについて説明する。

まず、ファンクーラは計測の結果、常に 2.76[W] を消費していることが分かった。したがってこの値をファンクーラの定常的な消費電力値とした。

次にパワーサプライについては計測の結果、全体消費電力の一定割合を消費していることが分かった。そこでこの値を算出して、ディスクドライブ、コントローラおよびファンクーラの消費電力の合計値の 0.186 倍をパワーサプライの消費電力値とした。

3.6　消費電力シミュレータの作成と消費電力モデルの検証

次に、構築した消費電力モデルを用いてディスクアレイの消費電力シミュレータを作成した。このシミュレータはディスクアレイへの I/O トレースを入力としてディスクアレイの消費電力のシミュレーションを行うものである。ディスクアレイへの I/O トレースからディスクドライブビジー率やスループットの情報を得ることにより、それらを消費電力モデルに適用することによって予測電力値を算出することができるのである。

この消費電力シミュレータによるディスクアレイの消費電力の予測精度を評価することによって消費電力モデルの検証を行った。検証は MySQL を用いて TPCH のクエリ Q1～Q10 までの問い合わせを行った際の消費電力量について、その際の I/O トレースを入力としてシミュレータで算出した予測値と実測値とを比較することによって行った。

検証の結果は図 6 のようになった。図は Q1～Q10 のそれぞれについて実測電力値を 1 に正規化して、実測電力値と予測電力値を比較したものである。また、それぞれの場合についてディスクアレイの各構成部位ごとの内訳も表している。図に示したように最も誤差の大きい Q6 でもその誤差は 7.6 %となっており、すべてのクエリにおいて誤差を 10 %未満に抑えて消費電力量を予測できていることが分かる。したがって、構築した消費電力モデルとディスクアレイの I/O トレースを入力としたシミュレータによってディスクアレイの消費電力を高精度で予測できるということが検証できた。

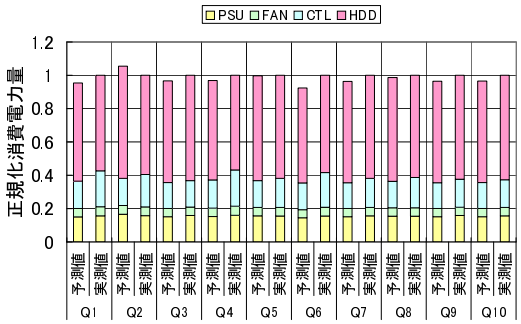

図 6　消費電力モデルの検証結果

4.　データベースシステムとの連携によるディスクアレイ省電力化手法

4.1　従来の省電力化手法と能動的なディスクドライブ省電力化手法

従来の省電力化手法はいずれもストレージ内で得られる情報のみを活用した受動的な省電力化制御を行うものである。これは I/O の応答時間やアクセス頻度などストレージ内で得られる情報とデータアクセスには空間的・時間的局所性があるということを利用して省電力化を図るものである。たとえば、長時間アクセスが行われていないディスクドライブを、今後もアクセスが行われないとみなしてスピンダウン制御を行うことで省電力化を図る手法や、アクセス頻度の高いデータを一部のボリュームへと集中化させることでその他のボリュームの省電力化を図る手法などが存在する [3], [4]。しかし、これらの手法はストレージ内で得られる情報のみからの予測による省電力化制御を行うものであるため、予測と異なる挙動が起きた場合には充分な省電力化効果が得られないという問題点がある。

これに対してデータベースシステムの問合せ実行計画を利用することで能動的なディスクドライブの省電力化制御を行う手法が提案されている [7]。この手法ではストレージ内で得られる情報に加えてデータベースサーバが有する問合せ実行計画の情報を活用することによって、確実で高効率な省電力化制御を行っている。問合せ実行計画から入出力予定情報を抽出し、ディスクドライブの直接的な制御を行うことによって確実な省電力化制御を行うことが出来るのである。この手法は論文 [7] において単体のディスクドライブのモデルを用いた解析によって TPCH の Q8 などに対して消費電力量を約 35 %〜50 %削減することができるということが示されている。

比較のために従来手法による省電力化と能動的なディスクドライブ省電力化手法による省電力化のイメージを図 7 に示す。従来手法は図に示したようにストレージ内で得られる情報のみからストレージが単体で省電力化制御を行っている。それに対して、能動的なディスクドライブ省電力化手法ではストレージに対してデータベースサーバが情報を提供することによって、能動的にストレージの省電力化を行うのである。これによって能動的なディスクドライブ省電力化手法は従来の省電力化手法よりも高効率な省電力化を行うことが出来るのである。

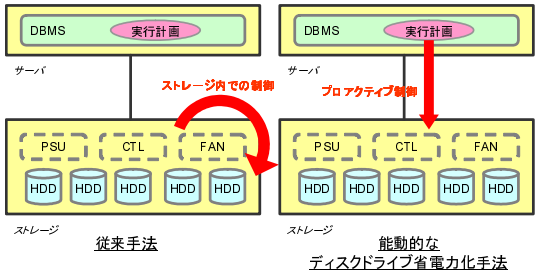

図 7　従来手法と能動的な省電力化手法の省電力化のイメージ図

4.2　既存手法の問題点

次に本節では従来の省電力化手法および能動的な省電力化手法について二つの問題点を挙げる。

まず、一つ目の問題点として能動的なディスクドライブ省電力化手法の問題点について述べる。4.1 で述べたように能動的なディスクドライブ省電力化手法は従来の手法に比べて確実で高効率な省電力化を行うことが出来る手法である。しかし、能動的なディスクドライブ省電力化手法ではデータベースサーバが立案する問合せ実行計画にしたがって省電力化制御を行うため、問合せ処理の問合せ方法によっては充分な省電力化が行えないという問題点がある。たとえば、問合せ処理が各ボリュームに対して常にアクセスを行うようなものであれば、問合せ実行計画を活用しても省電力化制御を一切行うことが出来ない。これは、データベースサーバがストレージの省電力化を考慮せずに性能指向の問合せ処理を行っていることに原因があると考えられる。

また、二つ目の問題点として従来の省電力化手法および能動的な省電力化手法では、省電力化制御を行う際にディスクドライブの省電力化制御のみを行っているということが挙げられる。ストレージの物理構成によってはディスクドライブ以外の消費電力がストレージ全体の消費電力に対して大きな割合を占めることもある。本研究で使用したディスクアレイにおいても 2.4 で述べたようにディスクドライブ以外の構成部位の消費電力が全体の約 40 %程度を占めており、これらの部位が消費する電力も決して無視することはできない。そのため、ストレージに対して省電力化制御を行う際にはディスクドライブ以外の消費電力についても考慮しなければならない場合もあると考えられる。

4.3　データベースシステムとの連携によるディスクアレイ省電力化手法の提案

そこでこれらの問題点を解決する手法としてデータベースシステムとの連携によるディスクアレイ省電力化手法を提案する。この手法はデータベースサーバがストレージと深い連携を取ることによって、既存手法よりも一層の省電力化を行うことを目指すものである。

提案手法では 4.2 で示した問題点を解決するために以下の二つの試みを行う。

まず、最初の試みとして提案手法では従来の性能指向の問合せ処理とは異なる省電力化指向の問合せ処理を行う。これは問合せの実行時間が多少長くなってしまうことを許容する代わり

にストレージの省電力化を考慮した問合せ処理を行うというものである。省電力化指向の問合せ処理としてたとえば、ストレージの物理構成を意識して各ボリュームへのアクセスを時間的に集中化させることによって、非アクセス状態の時間を長期化させることなどが考えられる。このようにすることによって非アクセス状態のボリュームの効率的な省電力化制御を行うことが出来るのである。この省電力化指向の問合せ処理の具体的な方法については、5.2 において説明する。このように省電力化指向の問合せ処理を行うことによって、その問合せ実行計画の情報を利用して効率的な省電力化を行うことが出来ると思われる。

次に二つめの試みとして提案手法ではディスクアレイの省電力化制御を行う際には、ディスクドライブだけではなくその他の構成部位に対しても省電力化制御を行う。規模の大きなデータベース環境においては一つのボリュームを単体もしくは複数のディスクアレイに対して割り当てることも少なくない。この環境においては、ディスクドライブのみに対して省電力化制御を行ったとしてもディスクアレイのその他の構成部位によって消費される電力は削減することが出来ない。このような場合に提案手法はディスクアレイ全体に対して省電力化制御を行うことによって、より高効率の省電力化を行うことが出来ると思われる。

4.1 と同じく比較のために提案手法による省電力化のイメージを既存手法のイメージとともに図 8 に示す。4.1 で述べたように能動的なディスクドライブ省電力化手法ではストレージに対してデータベースサーバが問合せ実行計画の情報を提供することで、省電力化を行っている。それに対して、提案手法ではストレージが物理構成情報などをサーバ側に提供し、その情報を考慮したうえでデータベースサーバが省電力化指向の問合せ実行計画を立案する。そして、その問合せ実行計画の情報をストレージに提供することによって、より効率的な省電力化を行うのである。

また、従来の省電力化手法と能動的なディスクドライブ省電力化手法では図に示したようにディスクアレイ中のディスクドライブのみに対して省電力化制御を行っている。それに対して、提案手法ではディスクドライブ以外の構成部位についても省電力化制御を行う。このディスクアレイ全体に対する省電力化制御によってすべての構成部位の電力を削減することが出来るため、効率的な省電力化を行うことができるのである。

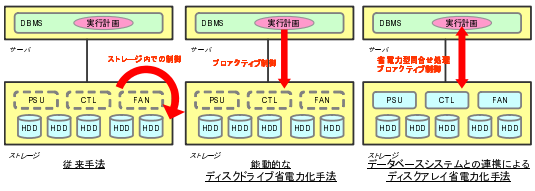

図 8　従来手法と能動的な省電力化手法および提案手法の省電力化のイメージ図

5.　提案手法の評価

5.1　評 価 方 法

本章では提案手法の評価を行う。評価は TPCH の Q17 相当の問合せ処理を MySQL 環境において実行した場合のケーススタディによって行った。この場合に、既存手法と提案手法による問合せ処理を行い、その際の消費電力量と実行時間の比較を行った。これらの結果から省電力化効果と実行時間の増加率を検討することによって、提案手法の有効性を検証した。

実験環境としては 2. で使用した環境を用いた。この環境において実際に行われるようなデータ配置を想定して TPCH の各テーブルの物理的配置を行い、これに対して問合せ処理を行った。その際の I/O トレースを入力として 3. で構築した消費電力モデルに基づくシミュレータによって、既存手法と提案手法の消費電力量をそれぞれ算出した。また、問合せ処理の実行時間については実際に計測した。

評価はデータベースの物理構成として小規模ディスクアレイ環境を想定した場合と中規模ディスクアレイ環境を想定した場合について行った。各環境における具体的なテーブルの物理的配置方法については 5.3,5.4 においてそれぞれの評価結果とともに詳しく述べる。

5.2　省電力化指向の問合せ処理

```
select
        part.p_partkey,
        part.p_name,
        lineitem.l_orderkey
from
        part,
        lineitem
where
        p_container = 'MED BOX';
```

図 9　評価に使用したクエリ

本節では評価に使用した Q17 相当の問合せを行う場合について、省電力化指向の問合せ処理を行う場合の具体的な処理方法を従来の性能指向の問合せ処理を行う場合と比較して説明する。評価に使用したクエリを図 9 に示す。このクエリによる問合せ処理を MySQL 環境において実行すると、TPCH の構成テーブルである part テーブルと lineitem テーブルのネステッドループ結合が行われる。この結合を行う際に、省電力化指向の問合せ処理ではこれらのテーブルのストレージ上での物理的な配置方法を意識することによって省電力化を図る。

はじめにこれらのテーブルの特徴から想定される、それぞれのテーブルのストレージ上での配置方法について説明する。まず、part テーブルは比較的サイズの小さなテーブルであり、実際のデータベース環境においても単一のボリュームに配置されていることが多いと想定される。それに対して lineitem テーブルは非常にサイズの大きなテーブルであり、実際のデータベース環境においてもしばしば複数のボリュームへと分散して配置されていることが想定される。

次にこれらのテーブルが想定した配置方法によってストレージ上で配置されている場合のそれぞれの問合せ処理方法を説明する。図 10 は lineitem テーブルが三つのボリューム上に

lineitem1〜3 として配置されている場合について、性能指向の問合せ処理と省電力化指向の問合せ処理による問合せの様子を示したものである。

まず、性能指向の問合せ処理では図に示したように part テーブルと lineitem テーブル全体の結合を一括して行う。この場合、lineitem1〜3 の全体に対してアクセスが行われるため、すべてのボリュームに対して散発的にアクセスが行われることになる。そのため、図のグラフに示したように問合せ処理中にいずれのボリュームも省電力化制御を行うことは出来ない。これは各テーブルの物理構成を意識せずにテーブル単位で結合処理を行ったためであるといえる。

それに対して省電力化指向の問合せ処理では図に示したように part テーブルと lineitem テーブルの結合を各テーブルの物理構成を意識して逐次的に行うことで省電力化を図っている。part テーブルと lineitem テーブルの結合結果は part と lineitem1 の結合結果、part と lineitem2 の結合結果、part と lineitem3 の結合結果のそれぞれを合わせたものに等しい。そこで、省電力化指向の問合せ処理ではまずはじめに part と lineitem1 の結合を行う。この際には、図のグラフに示したように lineitem2 と lineitem3 が含まれるボリュームに対して省電力化制御を行うことが出来る。そして、次に part と lineitem2 の結合を行う。この際には、lineitem1 と lineitem3 が含まれるボリュームに対して省電力化制御を行うことが出来る。最後に同様にして part と lineitem3 の結合を行う。省電力化指向の問合せ処理ではこのように散発的だった各ボリュームへのアクセスを時間的に集中化させることによって、それぞれのボリュームのアイドル時間を長大化させ、省電力化制御を行うことで省電力化を行う。これは各テーブルの物理構成を意識してボリューム単位で結合処理を行うことによって達成される。しかし、この処理方法では part テーブルを何度も読み込むことになるため、実行時間自体は増加してしまうものと思われる。

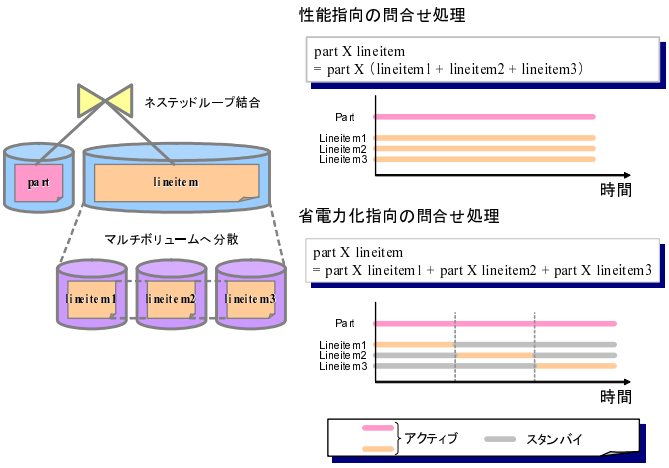

図 10　性能指向の問合せ処理と省電力化指向の問合せ処理

5.3　小規模ディスクアレイ環境での評価

本節では小規模ディスクアレイでのデータベース空間構成を想定した場合の評価結果を示す。

想定環境でのデータベースの物理構成について説明する。想定した環境では各ボリュームを単一のディスクアレイの各ディスクドライブに対してそれぞれ割り当てた。使用したディスクアレイは 5 台のディスクドライブによって構成されていたため、TPCH のスケールファクタを 1 としてこのうちの 1 台に part テーブルを配置し、残りの 4 台に lineitem テーブルを分割して配置した。この環境において提案手法として省電力化指向の問合せ処理とディスクドライブの省電力化制御を行った場合の省電力化効果を評価した。

評価の結果は図 11 のようになった。図は既存手法と提案手法の消費電力量と実行時間を表したものである。また、消費電力量に関しては各構成部位ごとと遷移コストの内訳、実行時間に関しては処理時間と遷移コストの内訳も表している。提案手法は既存手法に対して 26.7 %の省電力化効果がある一方で、実行時間の増加率は 1.8 %にとどまっていることが分かった。

この評価によって省電力化指向の問合せ処理とディスクドライブの省電力化制御によって実行時間の増加割合に対して充分な省電力化効果が得られることが分かった。しかし、ディスクドライブの消費電力が削減されたことによってその他の部位の消費電力がより支配的になることが分かった。

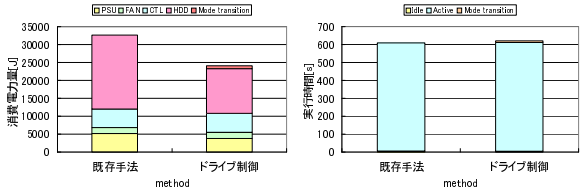

図 11　小規模ディスクアレイ環境での評価結果

5.4　中規模ディスクアレイ環境での評価

次に 5.3 よりも規模の大きい中規模ディスクアレイでのデータベース空間構成を想定した場合の評価の結果を示す。

想定環境でのデータベースの物理構成について説明する。この環境では評価に使用したディスクアレイ 5 台からなるデータベース空間を想定し、各ボリュームを 1 台のディスクアレイに対してそれぞれ割り当てた。TPCH のスケールファクタを 5 として 1 台に part テーブルを配置し、残りの 4 台に lineitem テーブルを分割して配置した。この環境において提案手法として省電力化指向の問合せ処理とディスクドライブのみの省電力化制御を行った場合、およびディスクアレイ全体の省電力化制御を行った場合の省電力化効果を評価した。

評価の結果は図 12 のようになった。図は既存手法とディスクドライブのみの制御、ディスクアレイ全体の制御を行った場合の消費電力量と実行時間を表したものである。また、消費電力量に関してはディスクアレイごとと遷移コストの内訳、実行時間に関しては処理時間と遷移コストの内訳も表している。まず、ディスクドライブのみの制御を行った場合には 31.0 %、ディスクアレイ全体の制御を行った場合には 49.5 %の省電力化効果があることが分かった。一方で実行時間はディスクドライブのみの制御の場合はほぼ増加せず、ディスクアレイ全体の制御を行った場合にも 1.2 %の増加にとどまるということが分かった。

この評価によって省電力化指向の問合せ処理とともにディス

クアレイ全体に対して省電力化制御を行うことで、ディスクドライブのみの制御を行った場合よりも、さらに高効率な省電力化を行えるということが分かった。

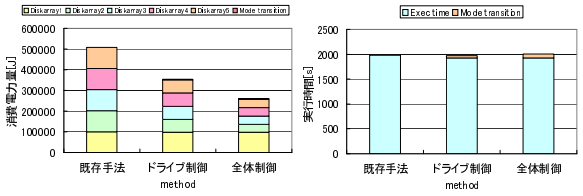

図 12　中規模ディスクアレイ環境での評価結果

6. 関連研究

6.1 PARAID

PARAID(Power-Aware RAID) [5] は RAID の構成をベースとして余っているディスク・スペースを有効に使うことによって省電力化を図る手法である。従来の RAID の構成の特徴として、すべてのディスクへのアクセス頻度がある程度均等になるようにデータが配置されていることがあげられる。これはディスク・アクセスを均等化することによって高バンド幅を得るためであり、性能面では非常に有効であるといえる。しかし、消費電力の面ではアクセス頻度が少ないときでもすべてのディスクにアクセスする機会が多くなってしまい、ディスクを standby 状態にしにくいことからあまりよい構成であるとはいえない。

一方でデータセンタ環境での使用ディスク・スペースは全体の 3 割〜6 割程度にとどまっているという報告がされている。これはデータセンタ環境などでは高バンド幅を得るために必要以上に多くのディスク・ドライブを使用しているためであると考えられる。

PARAID はこの余ったディスク・スペースにデータを冗長に配置することによって使用ディスク数を減少させて電力を削減している。また、この使用ディスク数を必要なバンド幅に伴って変更することにより、性能の保証も行っている。

6.2 Hibernator

Hibernator [6] は動的ディスク回転速度制御とマイグレーションによるデータ配置の動的最適化を行うことによる省電力化手法である。

Hibernator では省電力化のためにまず粗い粒度の調整として動的ディスク回転速度制御を行う。Hibernator は同一回転速度で回転しているディスクのグループ（Tier）によって構成される。Hibernator ではよりアクセス頻度の高いディスクを高速で回転する Tier に所属させ、アクセス頻度の低いディスクを低速で回転する Tier に所属させることによってそれぞれのディスクを最適な回転速度で回転させる。

さらに Hibernator では動的ディスク回転速度制御よりも細かい粒度の調整としてマイグレーションによるデータ配置の動的最適化を行っている。これは Tier 間でデータ・ブロックのマイグレーションを行うものである。アクセス頻度の高いデータ・ブロックをより高速で回転している Tier にマイグレートすることによってデータ配置を動的に最適化するのである。

7. おわりに

本稿では近年の情報量の爆発的増大に伴って、ストレージの省電力化が重要になってきていることについて述べた。一方でストレージの実際の消費電力量に関しては調査が十分ではないことを踏まえてディスクアレイ内部の消費電力について計測を行った。また、ディスクアレイの消費電力モデルを構築し、その有効性について検証した。さらにディスクアレイの省電力化を図る手法としてデータベースシステムとの連携によるディスクアレイ省電力化手法の提案と構築したモデルに基づく検証を行った。

今後の課題としては省電力化指向の問合せ処理のための最適なデータ配置を考えることが挙げられる。最適なデータ配置は問合せごとに異なってくると思われるため、様々な問合せに対して有効なデータ配置を考える必要があると思われる。これに加えて、実際のデータベース環境においては時間とともに負荷の傾向が変わることも想定されるため、動的なデータマイグレーションによるデータ配置の動的最適化を行うことなども考えられる。また、データベースシステムとの連携によるディスクアレイ省電力化手法として本稿に示した以外にも様々な連携方法があると考えられるため、それらを考案・評価していくことも課題として挙げられる。

謝辞　本研究の一部は，文部科学省リーディングプロジェクト e-Society 基盤ソフトウェアの総合開発「先進的なストレージ技術」の助成により行われた．協力企業である株式会社日立製作所より多くの有益なコメントを頂戴した．感謝する次第である．

文　献

[1] *Q. Zhu, F.M. David, C. Devaraj, Z. Li, Y. Zhou, P. Cao, Reducing Energy Consumption of Disk Storage Using Power-Aware Cache Management, Proceedings of the 10th International Symposium on High Performance Computer Architecture, 2004.*

[2] *Dell PowerEdge 6650 executive summary. http://www.tpc.org/results/individual results/Dell/dell 6650 010603 es.pdf, March 31 2003.*

[3] *D. Colarelli and D. Grunwald. Massive Arrays of Idle Disks for Storage Archive. In Proc. Int'l Conf. on Supercomputing, pp. 1-11, 2002.*

[4] *E. V. Carrera, E. Pinheiro, and R. Bianchini. Conserving Disk Energy in Network Servers. In Proc. Int'l Conf. on Supercomputing, pp. 86-97, 2003.*

[5] *Charles Weddle, Mathew Oldham, Jin Qian, and An-I Andy Wang, Peter Reiher, Geoff Kuenning. PARAID: A Gear-Shifting Power-Aware RAID. USENIX FAST 2007.*

[6] *Qingbo Zhu, Zhifeng Chen, Lin Tan, Yuanyuan Zhou, Kimberly Keeton, John Wilkes," Hibernator: Helping Disk Arrays Sleep through the Winter ", SOSP '05, October 23-26, 2005.*

[7] 上野 裕也, 合田 和生, 喜連川 優, データベースシステムの問合せ実行計画を利用したディスクアレイ省電力化に関する一考察, 日本データベース学会論文誌 *(DBSJ Letters) Vol.6 No.1, 2007*