C-019

HDD 向けキャッシュ性能シミュレーションによる I/O 性能設計の一手法

An I/O Performance Design by the Cache Performance Simulator for HDD

細木 浩二† 長野 岳彦† 石川 誠† 井上 和則†† Hung Vu†††
Koji Hosogi† Takehiko Nagano† Makoto Ishikawa† Kazunori Inoue†† Hung Vu†††

1． はじめに

HDD(Hard Disk Drive)などの I/O 性能向上アプローチの一つはキャッシュメモリであり，その性能を示す一般的な指標はキャッシュヒット率である．一方 HDD は，ディスク回転などの物理的制約と内部構成の特徴により，キャッシュミス時の応答時間が一様でなく，ヒット率以外の性能指標が要求される． そこで本稿は，HDD 性能シミュレータを用いた HDD 向けキャッシュの性能設計手法に関して述べる．まず，本シミュレータ構成について述べ，次に静的評価と動的評価方法について述べる．次に，キャッシュミス要因の分類による性能改善指針，および，HDD 性能に係わるサーボとディスクの性能評価指標について述べる．最後に本手法のシミュレーション結果およびその評価をまとめる．

2．HDD 向けキャッシュメモリ

図 1 に HDD 構成の概略を示す．HDD の基本機能は，ホストから入力される I/O コマンドに応じたディスクのリード・ライトである．HDD には，コマンドを保持するキュー，ディスクやサーボを制御する各コントローラを有し，これらを総合的に管理するコントローラで構成する．ディスク上のデータ参照は，ヘッドを参照対象トラックまでシークし，対象セクタまでの回転待ちを伴って，対象データを参照する[1]．ここで参照データ量が小さい場合，ヘッド到達時間(シーク時間+回転待ち時間)は，データ参照時間に対して十分に大きな値となる．加えて，ヘッド到達時間は機械的動作を伴うため応答性が一様ではない．

一般的に，ヘッド到達時間短縮に向け，ランダム参照時は RPO(Rotational Position Optimization)[1]を，シーケンシャル参照時はキャッシュメモリへのプリフェッチを用いる．一般的なプロセッサに搭載されるキャッシュでは，その性能評価指標としてキャッシュヒット率が用いられる．これは，ミスヒット時のペナルティがほぼ固定値であり，キャッシュヒット率のみにより，大枠の性能を評価できるためである．

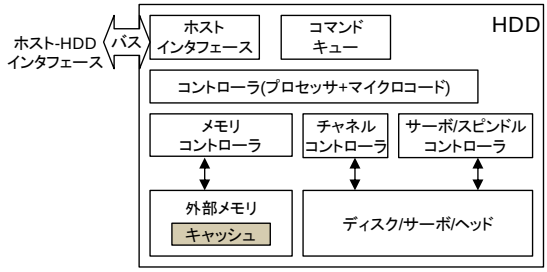

図 1　HDD 構成の概略

† (株)日立製作所 横浜研究所, Hitachi Ltd., Yokohama Lab.,
†† (株) HGST ジャパン, HGST Japan Ltd.,　††† HGST, Ltd.

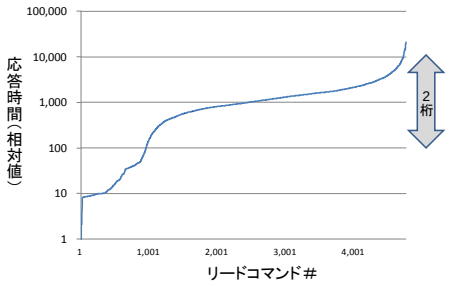

図 2 リードコマンド応答時間(相対値)

図 2 は実 HDD における，リードコマンド応答時間(相対値)を昇順に並べ替えたものである(リードコマンド主体，サンプル数 4,762，値レンジ 1～20,904)．図が示すように，応答時間は 4 桁のレンジにばらついている．キャッシュヒット時の応答時間最大値を 100 と仮定した場合，キャッシュミス時の応答時間ばらつきは 100~20,904(2 桁レンジ)となり，ミスヒット時のペナルティが一様でないことを示す．本原因は，機械的動作を伴うヘッド到達時間に加え，可変長コマンド，多段キュー構成と RPO によるリオーダリングの高い自由度などが起因する．従って HDD の性能評価では，キャッシュヒット率に加え，ばらつきを加味した評価指標を用いて評価しなければならない．

3. キャッシュ性能シミュレーション

3.1 HDD 性能評価プラットフォーム

HDD 性能並びにキャッシュ性能を評価するため，HDD キャッシュ性能シミュレータ開発した(図 2)．本シミュレータは，設計上流段階でのキャッシュアルゴリズム評価を目的とし， HDD 性能に係わる機能のみをモデル化したシミュレータである．

入力は，ディスクやキャッシュの構成パラメータ 5 種，キャッシュ制御パラメータ 11 種ならびにホストアクセスパタンであり，実ワークロードをシミュレーション可能とする．対象とするワークロードは，PC アプリケーションやベンチマークなど，実時間 1,000 秒以上のワークロードである．シミュレータ出力は，各コマンドの応答性などのコマンド・ログ，キャッシュ挙動ログおよびディスク挙動ログである．評価指標解析部は，これら独立したログを総合評価するための機能を含み，性能評価指標として可視化する．

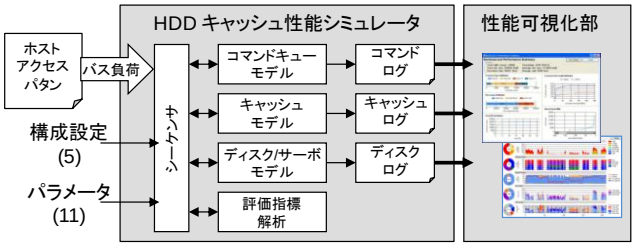

図 2　HDD 性能評価プラットフォーム

3.2　チャンク分割による部分統計の定量化

評価指標の定量化方法は大きく 2 種に分類できる．一方は図 2 のような静的(もしくは統計的)定量化であり，他方は時系列などの動的定量化である．特に対象とする PC 系ワークロードは，入力コマンドの特徴が動的に変化するため，動的定量化が有効となる．そこで，チャンク分割を用いた部分統計を用いた．これは複数コマンドを 1 つのチャンクとし，平均値などの統計値をチャンク毎に定量化することで，複数モデルの動的比較を可能とする．本提案では，表 1 に示す 3 つのチャンク分割モードを定義した．本チャンク分割により，単位時間当たりのスループットや，単位コマンドあたりのキャッシュヒット率など，目的に沿った比較を可能とする．

表 1 チャンク分割モード

チャンク分割モード	内容
Δ時間	単位時間内コマンドをチャンク化
Δ容量	コマンド長の和が単位容量内のコマンドをチャンク化
Δイベント数	単位コマンド数をチャンク化

3.3 キャッシュ・ミスヒットの分類

先にキャッシュ・ミスヒット時の応答時間ばらつきは 2 桁レンジであると述べた．従って設計上流段階では，キャッシュヒット率向上に加え，ミスペナルティの大きなミス要因に対して，性能向上を図る必要がある．そこで，まずキャッシュミス要因を分類し，各ミス要因に対する改善指針を定義した(表 2).

表 2 キャッシュ・ミスヒット分類

ミス原因	内容
ニアミス	現キャッシュの近傍に存在するキャッシュミス．ロングプリフェッチ長により改善可.
パージミス	既にパージされたキャッシュにヒットしたキャッシュミス．キャッシュ再配置アルゴリズムの改編により改善可.
ピュアミス	ランダムなどの過去発行コマンドとは独立したキャッシュミス．改善困難.

3.4 サーボおよびディスクの性能定量化

HDD では機械的制御を伴うサーボとディスクの挙動が性能に大きく係わる．ここではディスク稼働率に着目した．例えばプリフェッチ量を大きくした場合，ディスク稼動率は上昇し，後発のディスク参照は遅延するため，稼働率に応じたプリフェッチ量制御が必要となる．そこで，ディスク稼働率を表現するため，ディスク参照時のヘッド到達時間，データ転送時間及びディスク・アイドル時間を定義し，これらを評価指標の 1 つとした．

4．評価

表 3 の評価条件にて，異なるキャッシュパラメータ(ショート/ロング・プリフェッチ長)でのシミュレーション結果を比較する(表 4 及び図 3)．ショート時のヒット率は 39.2%であり，キャッシュミス要因の大部分はニアミスである．これに対しロング時はヒット率 72.4%に向上し，ショート時と比較し，総シーク回数半減，平均応答時間半減，ディスク・アイドル時間増加であり，表 2 に示したニアミス時の改善指針と同様の結果となった．一方，最大応答時間はロング時が悪化している．この原因はロング・プリフェッチによるディスク稼働率の局所的な上昇である．結果，全般的にはニアミスがフルヒットに代わり性能向上しているが，ワースト応答時間は悪化した (図 3)．このように性能評価では，キャッシュヒット率に加え，複数の評価指標を加味した総合的な判断が必要となる．

なおシミュレーション速度は，実 HDD に対し約 130 倍高速であり，パラメータ探索を含む上流性能設計において，問題ない速度にて評価可能である見通しを得た．

表 3 評価条件

項目	評価条件
実行環境	プロセッサ：Intel® Core™ i7-2600 3.4GHz * 4-CPUs, 主記憶 4.0GB
ベンチマーク	PCMark®7 [2]
パラメータ(2 種)	ショート／ロング・プリフェッチ
分割モード	Δ時間モード (Δ時間=1[sec])

表 4 プリフェッチ量(ショート /ロング)の比較

プリフェッチ長	ショート	ロング	凡例
ヒット率[%]	39.2	72.4	-
シーク数 [num]	6,908	3,209	-
ディスク・アイドル率[%]	55.8	69.1	-
ヒット率 [%]	ニアミス	フルヒット	ピュアミス, リリースドミス, ニアミス, アクティブヒット, フルヒット
シーク数 [num/sec]	100	30	ライトシーク, リードシーク (ニアミス), リードシーク
平均応答時間 [us]	約 5[ms]	約2.5[ms]	平均応答時間
ディスク稼働率 [%]	50, アイドル50%	30, アイドル70%	アイドル, ライトデータ, ライトシーク, リードデータ, リードシーク
最大応答時間 [us]			最大応答時間

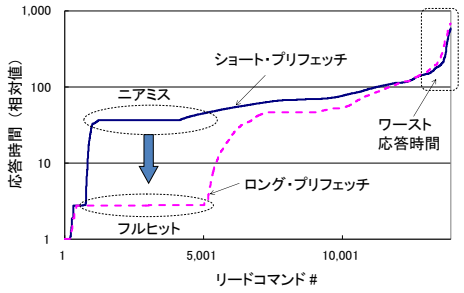

図 3 リードコマンド応答時間(相対値) の比較

5．おわりに

本性能設計手法では，HDD キャッシュの評価指標として，キャッシュヒット率に加え，ディスク・アイドル時間などを定義した．また，チャンク分割による部分統計により時系列での性能比較が可能であり，上流性能設計にて，有効な一手法となる見通しを得た．

参考文献

[1] B.Jocob et al., “Memory Systems : Cache, DRAM, Disk”, Morgan Kaufman Publisher Inc., 2008

[2] FutureMark, http://www.futuremark.com/products/pcmark7/