要旨

近年 DRAM と NAND 型フラッシュメモリとのアクセス性能差を埋めるメモリとしてストレージクラスメモリ（storage class memory，SCM）が研究，開発されてきた．磁気抵抗型メモリ（magnetoresistive RAM，MRAM），抵抗変化型メモリ（resistive RAM，ReRAM），相変化メモリ（phase change RAM，PRAM）が SCM の主な候補である．SCM は DRAM と同じくバイトアクセス可能で，DRAM と同等の 0.1 μsec 未満のアクセス性能を持つ．一方で NAND 型フラッシュメモリと同じく不揮発であり，Single-level cell（SLC，1 bit/cell）NAND 型フラッシュメモリに近い 1-10 μsec 程度のアクセス性能を持つ SCM も存在する．本論文では，高速なアクセス性能を持つ SCM をメモリタイプ SCM（memory-type SCM，M-SCM）と呼び，NAND 型フラッシュメモリに近い容量を達成するであろう SCM をストレージタイプ SCM（storage-type SCM，S-SCM）と呼ぶ．NAND 型フラッシュメモリもまた多値化技術により大容量化が進んだが，その書き換え性能は低下する．本論文では不揮発性半導体メモリを複数種用いてデータを管理，保存するヘテロジニアスストレージシステムを提案する．データを保存する不揮発性の記憶媒体として，M-SCM，S-SCM および Multiple-level cell（MLC，2 bit/cell），Triple-level cell（TLC，3 bit/cell）NAND 型フラッシュメモリを用いる．

本論文では三種以上の不揮発性半導体メモリを用い，二種類のヘテロジニアスストレージを提案する．また用いる不揮発性半導体メモリの特性に適したデータ管理アルゴリズムを提案する．SCM および MLC NAND 型フラッシュメモリを用いたハイブリッドストレージと比較して，第一の SCM，MLC および TLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージは，SCM に頻繁にアクセスされるデータを保存し，MLC NAND 型フラッシュメモリに滞留するアクセス頻度の低いデータを TLC NAND 型フラッシュメモリに保存することで MLC および TLC NAND 型フラッシュメモリの寿命を延ばす．第二の M-SCM，S-SCM および NAND 型フラッシュメモリを用いたヘテロジニアスストレージは，SCM を二種類用いて極端にアクセス頻度の高いデータを M-SCM に，ややアクセス頻度の高いデータを S-SCM に保存することを特徴とする．これらの不揮発性半導体メモリの組み合わせを用いたストレージは性能が向上し，消費エネルギーが削減される．

一方でストレージアプリケーションは，読み出し・書き込み量の多寡，平均データアクセス頻度，平均データサイズなどの特性が異なる．したがってストレージアプリケーションの特性に対して，ヘテロジニアスストレージの適切な不揮発性半導体メモリの組み合わせが異

なる．不揮発性半導体メモリの特性および容量を変え，ベンチマークとするストレージアプリケーションを用いた評価を行なった．その結果，高速な M-SCM を大容量用いるほどヘテロジニアスストレージの性能が向上することが明らかとなった．しかし単位容量当たりの M-SCM のコストは，NAND 型フラッシュメモリと比較して約 10 倍と予想される．そのため本論文では，MLC NAND 型フラッシュメモリのみを用いたストレージのコストと比較して，ヘテロジニアスストレージは 1.5 倍のコスト増が許容できると仮定した．その結果，読み出し・書き込み量の多寡および平均データサイズ（ランダム・シーケンシャル）と比較して，ストレージアプリケーションの平均データアクセス頻度（ホット・コールド）が，ヘテロジニアスストレージの最適な構成の決定に重要であることを明らかにした．一部のデータが頻繁に書き換えられるストレージアプリケーションに対しては，S-SCM を大容量用いることで性能を向上できることを明らかにした．あるいは M-SCM を極小容量用いて書き込み性能を向上できることを明らかにした．また一部のデータが頻繁に読み出し・書き込みされるストレージアプリケーションに対しては, M-SCM を大容量用いることで性能を向上できることを明らかにした. 一方で, 上書きおよび読み出しが頻繁に行われないアプリケーションに対しては, 小容量で高速な M-SCM を書き込みバッファとして機能させることが良いことを明らかにした．

続いて，エラー訂正符号（error-correcting code，ECC）を用いて SCM および MLC NAND 型フラッシュメモリを用いたハイブリッドストレージの信頼性を向上させる．しかし ECC の復号によりストレージ性能は低下するトレードオフがある．不揮発性半導体メモリの種類によって信頼性は異なりまた許容書き換え回数も異なる．一部のデータが頻繁にアクセスされ，平均データサイズが小さいホット・ランダムアプリケーションに対して，大容量の SCM は性能を大きく向上させるが，SCM に適用する ECC 強度を弱くしなければ求められる性能を維持できない．これと比較して NAND 型フラッシュメモリは，大容量の SCM を用いるハイブリッドストレージでは多量のデータが SCM で処理されるため，復号時間の長い LDPC 符号を NAND 型フラッシュメモリに適用することを可能にすることを明らかにした．

最後にストレージアプリケーションに対して必要な SCM の容量を自律調整する手法を提案する．頻繁にアクセスされるデータを SCM に保存することでストレージ性能は向上するが，適切な SCM 容量はストレージアプリケーションの特性によって異なる．また，複数のストレージアプリケーションが動作するデータセンターにおいて，それぞれのストレージアプリケーションに対して適切な SCM 容量を手動で決定することは実用的ではない．そこで M-SCM および MLC NAND 型フラッシュメモリを用いたハイブリッドストレージにおいて，

MLC NAND 型フラッシュメモリに保存されたデータのアクセス頻度を管理することで調整すべき SCM の容量を判断することを提案する. その結果頻繁にアクセスされるデータが SCM に保存され，ストレージ性能を低下させることなく SCM の容量を調整できることを示した.

以上により，一部のデータが頻繁に書き込み，読み出されるストレージアプリケーションに対して高速な M-SCM が有効であることを明らかにした．また，やや頻繁に書き込み，読み出しが行われるアプリケーションに対しては，大容量で NAND 型フラッシュメモリより高速な S-SCM が必要であることを示した．一方で，上書きおよび読み出しが頻繁に行われないアプリケーションに対しては，小容量で高速な M-SCM を書き込みバッファとして機能させると良いことを明らかにした.

目次

第1章 序論

1.1 研究の背景

近年 Internet of Things（IoT）や機械学習技術の興隆によりデータセンターでは，種類，速度, 量など特徴の異なる多種多様なデータが実行される[1][2]. 自動運転車, インダストリ 4.0, セキュリティシステムのような IoT アプリケーションではネットワークのエッジで収集した温度や画像などのデータをデータセンターで処理する．データセンターに集められたデータはたとえば機械学習技術を用いて多量のデータから類似するパターンを発見，モデル化し，将来の予測を行う．これらのデータは大きさやアクセスパターンなどの特徴がそれぞれ異なるため，ストレージに必要な特性も異なる．データセンターでは現在，高速な NAND 型フラッシュメモリを用いたソリッドステートドライブ（solid-state drive，SSD）などのストレージがハードディスクドライブ（hard-disk drive，HDD）を置き換えつつある．さらに HDD は書き込み・読み取り部に機械部品を用いるため衝撃に弱く故障が予測できない．一方で NAND 型フラッシュメモリは不揮発性半導体メモリであるため機械部品は使われず，データ書き換えやデータ保持によりビット反転が発生し，データの書き込み・読み出しエラーの原因となる．

図 1.1 にメモリおよびストレージ階層を示す[3][4][5][6][7][8][9]．階層の上から下へメモリのアクセス性能は低下する．その一方でメモリダイ（メモリチップ）当たりの容量は増加するため，ビットコストは低減する．上層の static random access memory（static RAM，SRAM）および dynamic RAM（DRAM）は揮発メモリであり電源消失によりデータが失われるが，高速に書き換え可能である．SRAM はデータ記憶にフリップフロップ回路を用い，6 トランジスタで構成されるため回路面積が必要でビットコストが高い．そのため小容量で十分なプロセッサのキャッシュメモリなどで用いられる．DRAM のメモリセルはトランジスタとキャパシタによって構成され，キャパシタの電荷を保存することでデータ保持を行う．リーク電流によりキャパシタ内の電荷が失われるとデータが失われるため，常に電源供給が必要で，さらにデータ消失を防ぐために定期的なデータリフレッシュを行う．常に電源供給が必要なことおよび定期的なデータリフレッシュを行うことから, DRAM は SRAM と比較して消費電力

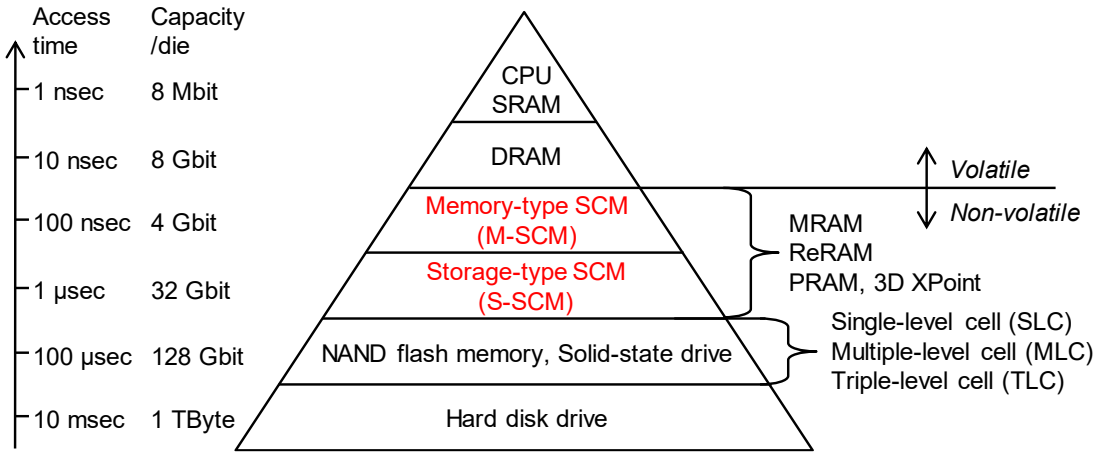


図 1.1 メモリおよびストレージ階層 [3][4][5][6][7][8][9]

が高い．しかしながら微細化によってビットコストが低減し続けたため，DRAM はキャッシュよりも大容量が必要な主記憶（メインメモリ）として用いられる．さらにまた DRAM を不揮発化するために電池を用い，瞬断に対応した battery back-uped（BBU）DRAM がある．

メモリおよびストレージ階層の下層に位置するストレージクラスメモリ（storage class memory，SCM）および NAND 型フラッシュメモリは不揮発であり，電源を消失してもデータは失われないため恒常的な電源供給は不要である．代表的な不揮発性半導体メモリである NAND 型フラッシュメモリは，メモリおよびストレージ階層（図 1.1）の最下層に位置する．NAND 型フラッシュメモリのセルは P 型半導体の上に構成された MOSFET であり，浮遊ゲートに電子を注入することでデータを保存する．浮遊ゲートは電気的に浮遊しているため電源を失ってもデータを保存できる．NAND 型フラッシュメモリは微細化および多値化技術により大容量化・ビットコスト低減が進められてきた．一方で多値化技術により Multiple-level cell（MLC）NAND 型フラッシュメモリ，Triple-level cell（TLC）NAND 型フラッシュメモリが実現した[10][11]．多値化技術は NAND 型フラッシュメモリを大容量化した一方で，そのアクセス性能および信頼性は低下する[12]．NAND 型フラッシュメモリはその大容量・低ビットコストによって携帯可能な USB メモリおよびメモリカード，個人向けコンピュータの二次記憶としての SSD や，さらにアクセス性能および大容量が求められるデータセンターのストレージとして用いられる．しかし DRAM と NAND 型フラッシュメモリとの間には約 10^4 倍のアクセス性能差がある．

このため DRAM と NAND 型フラッシュメモリとのアクセス性能差を埋めるメモリとして

SCM が研究, 開発されてきた[13][14]. 磁気抵抗メモリ（magnetoresistive RAM, MRAM）[15][16], 抵抗変化型メモリ（resistive RAM，ReRAM）[6][17][18]，相変化メモリ（phase change RAM, PRAM）[7][19][20]が SCM の主な候補である．SCM は DRAM と同じくバイトアクセス可能で，DRAM と同等のアクセス性能を持つ．一方で NAND 型フラッシュメモリと同じく不揮発であり，SLC NAND 型フラッシュメモリに近いアクセス性能を持つ．近年ではさらに SCM はその特性によりメモリタイプ SCM（memory-type SCM, M-SCM）およびストレージタイプ SCM（storage-type SCM，S-SCM）に細分化されている[21]．本論文では MRAM は高速なアクセス性能を持つため M-SCM と呼び，ReRAM および PRAM は大容量であるため S-SCM と呼ぶ．しかし SCM は開発途上であり，そのデータ保持方式などから NAND 型フラッシュメモリと同等に大容量化することは難しく SCM 単体でストレージとして使うにはコストが高い．さらにここに述べたように性能，信頼性，コストすべての面で優れたユニバーサルメモリは現在存在しない．

1.2 本研究の目的

本研究では第 1.1 節で述べたさまざまな不揮発性半導体メモリを組み合わせ，異種の不揮発性半導体メモリで構成されるヘテロジニアスストレージを提案する．ストレージアプリケーションは書き込み・読み出しの多寡，データアクセス頻度，データアクセスサイズなどの点でそれぞれ特性が異なる．また不揮発性半導体メモリはアクセス速度，信頼性，書き換え耐久性（許容書き換え回数），コストなどの点でそれぞれ異なる特性を持つ．そのため特性の異なるストレージアプリケーションに最適な不揮発性半導体メモリの構成および要件を提示する．従来研究として SCM および NAND 型フラッシュメモリを用いたハイブリッドストレージが提案されている[22][23][24]．NAND 型フラッシュメモリの低速な書き込み速度を隠ぺいするため，SCM は不揮発性キャッシュメモリあるいは小容量ストレージとして用いる．また SCM を用いた SSD [25]は NAND 型フラッシュメモリを用いた SSD の不揮発性キャッシュとして期待されている．これらの従来技術では二種類の不揮発性半導体メモリを用いてストレージを構成するのに対し，本論文で論じるヘテロジニアスストレージは三種類以上の不揮発性半導体メモリを用いてストレージを構成する．

不揮発性半導体メモリは HDD と異なり突然の機械部品の故障は起きないが，データの書き換えやデータ保持などにより徐々にエラーが発生する．不揮発性半導体メモリの種類によって信頼性は異なりまた書き換え耐久性（許容書き換え回数）も異なる．ストレージコントローラ内のエラー訂正回路が不揮発性半導体メモリに生じたエラーを訂正する．エラー訂正

回路はエラー訂正符号（error-correcting code，ECC）の種類によって異なる．ECC は書き込みたいデータにパリティを付加してデータを符号化してメモリに書き込む．またメモリからの読み出し時にデータ復号しエラー訂正を行う．一般に書き込み時の符号化と比較して読み出し時の復号に時間がかかる．さらに，エラー訂正能力の高い ECC はエラー訂正能力の低い ECC と比較して，ECC の動作に長い時間を要する．本論文では，SCM および NAND 型フラッシュメモリを用いたストレージについて，エラー訂正能力の異なる ECC を適用し高信頼化を図る．一方で ECC によるストレージ性能の低下を評価し，SCM および NAND 型フラッシュメモリに適用可能な ECC の強度を見積もる．

さらに異種の不揮発性半導体メモリを用いたストレージの不揮発性半導体メモリ構成の自動最適化を行う. 頻繁にアクセスされるデータを含むアプリケーションほど多量の SCM を必要とする一方で, 頻繁にアクセスされるデータが少ないアプリケーションは多量の SCM を用いてもストレージ性能に寄与しない．データセンターではさまざまなストレージアプリケーションが動作するため，それぞれのアプリケーションに対し適切な不揮発性半導体メモリの容量を手動で決定することは不可能である．また SCM のビットコストは NAND 型フラッシュメモリのそれと比較して高価であるため, アプリケーションが必要な場合にのみ SCM を用いたい．そのためストレージコントローラがストレージアプリケーションの特性を把握し，随時必要な SCM 容量を調整することが必要である．

以上により，不揮発性半導体メモリの特性およびストレージアプリケーションの特性を考慮し，高速，高信頼，低コストなヘテロジニアスストレージを構築することを目的とする．

1.3 本論文の構成

本論文は全 7 章から成る．図 1.2 に本論文の構成を示す．第 1 章では研究の背景と目的について述べた．ストレージクラスメモリ（storage class memory，SCM）と呼ばれる新規の不揮発性半導体メモリの登場によってメモリおよびストレージ階層が多層化しつつある．またストレージアプリケーションの特性はさまざまに異なるため，アプリケーションの特性に応じたストレージ構成が必要であることを述べた．さらに不揮発性半導体メモリに発生するエラーを訂正するためにエラー訂正符号（error-correcting code，ECC）が不可欠であるが，ECC によりストレージ性能が低下するため不揮発性半導体メモリに適用可能な ECC 強度を見積もることを述べた．またストレージアプリケーションの特性によって必要な SCM 容量は異なり，SCM のビットコストは NAND 型フラッシュメモリと比較して高価であるため，SCM 容量の自動最適化が必要であることを述べた．

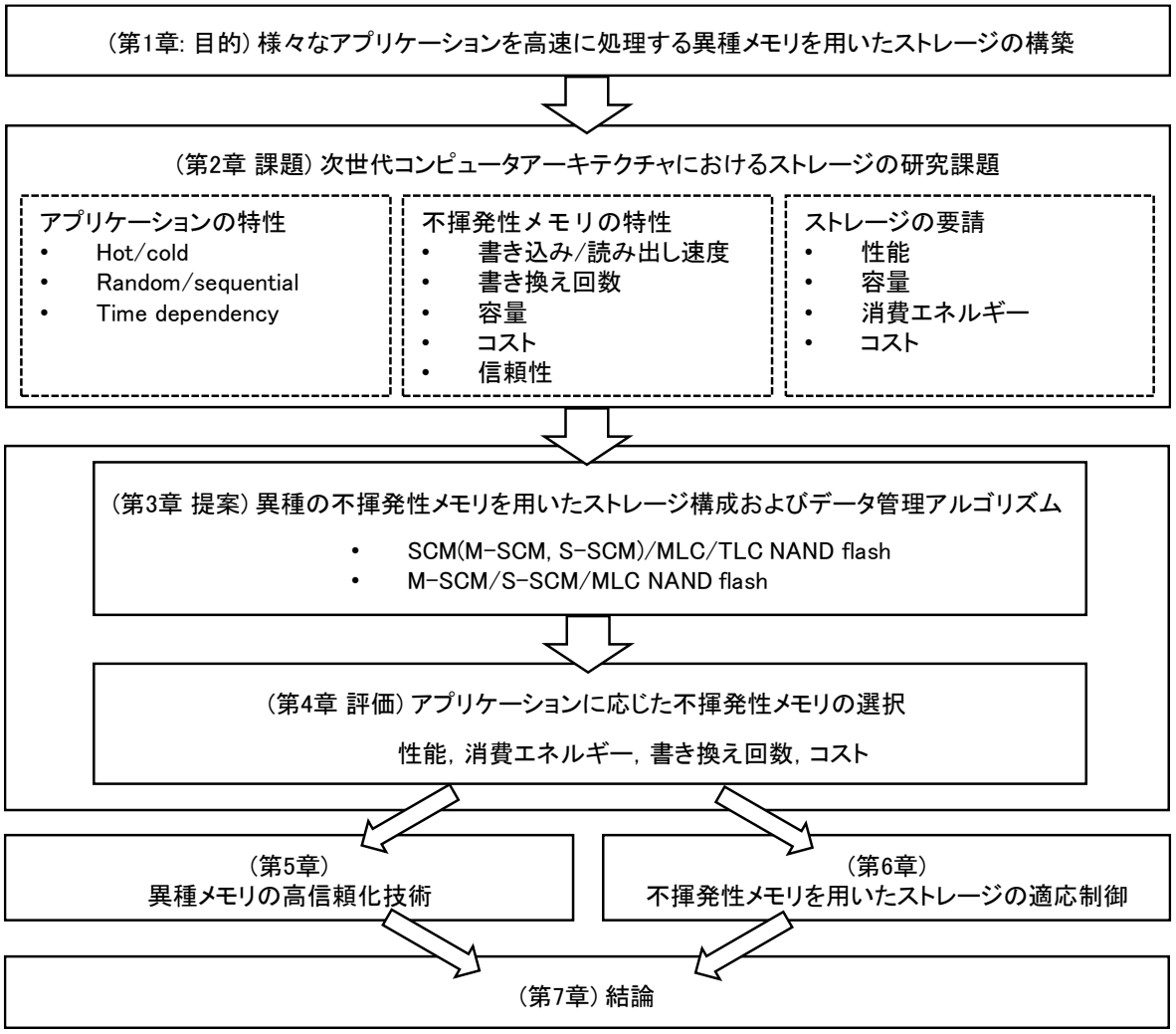


図 1.2 本論文の構成

第 2 章では不揮発性半導体メモリを用いたストレージの従来研究および問題点について述べる．まず初めに不揮発性半導体メモリである SCM と NAND 型フラッシュメモリの動作について述べる．SCM はその特性によってメモリタイプおよびストレージタイプに分類でき，一方で NAND 型フラッシュメモリはセル当たりに保存するビット数によって SLC，MLC，TLC に分類できることを示す．従来研究として SCM 一種および NAND 型フラッシュメモリ一種を用いたハイブリッドストレージストレージの構成，および SCM を NAND 型フラッシュメモリの不揮発性キャッシュあるいは小容量ストレージとして用いるデータマネジメント手法について説明する．さらに不揮発性半導体メモリを用いた次世代のコンピュータアーキテクチャにおけるストレージの課題を論じる．第 1 章で述べたようにストレージアプリケーションの特性がさまざまに異なるためストレージアプリケーション内のデータの特徴によって，異種の不揮発性半導体メモリを用いたヘテロジニアスストレージの不揮発性半導体メモ

リの構成の最適化が必要となる．またヘテロジニアスストレージの SCM と NAND 型フラッシュメモリとではアクセス頻度が異なり，不揮発性半導体メモリの種類によってエラー発生頻度や許容書き換え回数が異なるため，それぞれに異なる強度の ECC を適用することが必要となる．さらにストレージアプリケーションの特性によって最適な SCM 容量は異るが，データセンター事業者やユーザが手動でさまざまな種類のストレージアプリケーションに必要な SCM 容量を設定することは困難であるため，自動で SCM 容量を最適化する手法が必要であることを述べる．

第 3 章では異種の不揮発性半導体メモリを用いたヘテロジニアスストレージ構成を提案する．従来研究と異なり，三種以上の不揮発性半導体メモリを用いて構成したストレージをヘテロジニアスストレージと呼び本論文で扱う．ヘテロジニアスストレージとして，1）SCM，MLC および TLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージおよび 2）M-SCM，S-SCM および NAND 型フラッシュメモリを用いたヘテロジニアスストレージを提案する．第一の SCM，MLC および TLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージは，SCM および MLC NAND 型フラッシュメモリを用いたハイブリッドストレージと比較して, MLC NAND 型フラッシュメモリに滞留するアクセス頻度の低いデータを TLC NAND 型フラッシュメモリに保存することで MLC NAND 型フラッシュメモリの書き換え回数を削減することを目的とする. さらに SCM の導入で上昇する総ストレージコストをビットコストの低い TLC NAND 型フラッシュメモリでバランスできる．第二の M-SCM，S-SCM および NAND 型フラッシュメモリを用いたヘテロジニアスストレージは，SCM を二種類用いて極端にアクセス頻度の高いデータを M-SCM に，ややアクセス頻度の高いデータを S-SCM に保存することを特徴とする．M-SCM は S-SCM と比較して高速だがビットコストが高いため，ごく少量のデータが頻繁にアクセスされるストレージアプリケーションに適していると考える．

第 4 章では, 第 3 章で提案した二種類のヘテロジニアスストレージの評価を行なう. SystemC ベースのストレージエミュレータに，不揮発性半導体メモリの動作およびデータマネジメントアルゴリズムを実装した．不揮発性半導体メモリの容量比や書き込み・読み出し時間などのアクセス性能を変化させ，ヘテロジニアスストレージのアクセス性能，消費エネルギー，不揮発性半導体メモリの書き換え回数の点から評価し比較する．代表的なストレージアプリケーション毎にヘテロジニアスストレージの最適な不揮発性半導体メモリ構成を示す.

第 5 章では異種メモリの高信頼化技術について述べる．ここでは SCM および NAND 型フ

ラッシュメモリを用いたハイブリッドストレージに強度の異なる ECC を適用する．NAND 型フラッシュメモリには従来から Bose-Chaudhuri-Hocquenghem（BCH）符号が用いられている．BCH 符号は高速でランダムエラーを訂正できる ECC であるため，ストレージとして用いる SCM にも適用する．また NAND 型フラッシュメモリの微細化および多値化が進むにつれ，BCH 符号より訂正能力の高い low-density parity-check（LDPC）符号も適用されつつある．これらの ECC を不揮発性半導体メモリに用いるとストレージ性能が低下するため，ハイブリッドストレージで用いる SCM および NAND 型フラッシュメモリに適用できる ECC の強度を示す．

第 6 章ではストレージアプリケーションの特性に応じた SCM 容量の自律調整手法を述べる．SCM と NAND 型フラッシュメモリを用いたハイブリッドストレージは，ストレージアプリケーションの性質によって必要となる SCM 容量は異なる．さまざまな特性を持つストレージアプリケーションに最適な SCM 容量をそれぞれ手動で決定することは難しい．SCM から NAND 型フラッシュメモリへ移されたデータの中で頻繁にアクセスされるデータをゴースト least recently used（ghost LRU）リストを用いて検出し拡大すべき SCM 容量を計算する手法について述べる．また SCM 内で頻繁にアクセスされるデータを NAND 型フラッシュメモリへ移動することを防ぐアルゴリズムについても述べ，ハイブリッドストレージ性能を評価する．

第 7 章に本論文の結論と今後の研究について展望を述べる．本論文ではストレージアプリケーションの特性に応じた異種の不揮発性半導体メモリを用いたヘテロジニアスストレージの最適な構成を論じた．極頻繁にアクセスされるデータを含むストレージアプリケーションには高速な M-SCM が必要である．一方で頻繁にアクセスされないデータを含むストレージアプリケーションは TLC NAND 型フラッシュメモリが必要であることがわかる．また SCM および NAND 型フラッシュメモリを用いたストレージにおいて，SCM 容量が多いと SCM へのアクセス頻度が増すために SCM に適用する ECC の強度は一定以下に抑えねばならず，しかし NAND 型フラッシュメモリへのアクセス頻度は減少するために NAND 型フラッシュメモリに適用する ECC の強度を強めることができる．さらに SCM 容量の自律調整手法を用いることで，ストレージ動作期間中のストレージコストを抑えることができることを示す．最後に 3 次元積層された NAND 型フラッシュメモリを用いた場合のヘテロジニアスストレージのデータマネジメントアルゴリズム最適化や，SCM 容量調整手法に機械学習を用いることが今後の研究課題であることを述べる．

参考文献

[1] M. Fink, “Toward a memory-centric architecture,” in *Flash Memory Summit*, 2017.

[2] EMC Digital Universe with Research & Analysis by IDC, “The Digital Universe of Opportunities: Rich Data and the Increasing Value of the Internet of Things,” Apr. 2014, http://www.emc.com/leadership/digital-universe/2014iview/business-imperatives.htm.

[3] M. Yabuuchi, K. Nii, S. Tanaka, Y. Shinozaki, Y. Yamamoto, T. Hasegawa, H. Shinkawata, and S. Kamohara, “A 65 nm 1.0 V 1.84 ns Silicon-on-Thin-Box (SOTB) embedded SRAM with 13.72 nW/Mbit standby power for smart IoT,” in *IEEE Symposium on VLSI Circuits Digest of Technical Papers*, Jun. 2017, pp. 220-221.

[4] K. Song, S. Lee, D. Kim, Y. Shim, S. Park, B. Ko, D. Hong, Y. Joo, W. Lee, Y. Cho, W. Shin, J. Yun, H. Lee, J. Lee, E. Lee, N. Jang. J. Yang, H.-K. Jung, J. Cho, H. Kim, and J. Kim, “A 1.1 V 2y-nm 4.35 Gb/s/pin 8 Gb LPDDR4 mobile device with bandwidth improvement techniques,” *IEEE Journal of Solid-State Circuits (JSSC)*, vol. 50, no. 8, pp. 1945-1959, Aug. 2015.

[5] S.-W. Chung, T. Kishi, J. W. Park, M. Yoshikawa, K. S. Park, T. Nagase, K. Sunouchi, H. Kanaya, G. C. Kim, K. Noma, M. S. Lee, A. Yamamoto, K. M. Rho, K. Tsuchida, S. J. Chung, J. Y. Li, H. S. Chun, H. Oyamatsu, and S. J. Hong, “4Gbit density STT-MRAM using perpendicular MTJ realized with compact cell structure,” in *IEEE International Electron Devices Meeting (IEDM) Technical Digest*, Dec. 2016, pp. 27.1.1-27.1.4.

[6] T.-Y. Liu, T. H. Yan, R. Scheuerlein, Y. Chen, J. K. Lee, G. Balakrishnan, G. Yee, H. Zhang, A. Yap, J. Ouyang, T. Sasaki, A. Al-Shamma, C. Chen, M. Gupta, G. Hilton, A. Kathuria, V. Lai, M. Matsumoto, A. Nigam, A. Pai, J. Pakhale, C. H. Siau, X. Wu, Y. Yin, N. Nagel, Y. Tanaka, M. Higashitani, T. Minvielle, C. Gorla, T. Tsukamoto, T. Yamaguchi, M. Okajima, T. Okamura, S. Takase, H. Inoue, and L. Fasoli, “A 130.7-mm2, 2-layer 32Gb ReRAM memory device in 24-nm technology,” *IEEE Journal of Solid-State Circuits (JSSC)*, vol. 49, no. 1, pp. 140-153, Jan. 2014.

[7] Micron 3D XPoint Technology, https://www.micron.com/about/emerging-technologies/3d-xpoint-technology.

[8] M. Helm, J.-K. Park, A. Ghalam, J. Guo, C. W. Ha, C. Hu, H. Kim, K. Kavalipurapu, E. Lee, A. Mohammadzadeh, D. Nguyen, V. Patel, T. Pekny, B. Saiki, D. Song, J. Tsai, V. Viajedor, L. Vu, T. Wong, J. H. Yun, R. Ghodsi, A. D’Alessandro, D. Di Cicco, V. Moschiano, “A 128Gb MLC NAND-flash device using 16nm planer cell,” in *IEEE International Solid-State Circuits Conference (ISSCC) Digest of Technical Papers*, Feb. 2014, pp. 326-327.

[9] G. Naso, L. Botticchio, M. Castelli, C. Cerafogli, M. Cichocki, P. Conenna, A. D’Alessandro, L.

De Santis, D. Di Cicco, W. Di Francesco, M. L. Gallese, G. Gallo, M. Incarnati, C. Lattaro, A, Macerola, G. Marotta, V. Moschiano, D. Orlandi, F. Paolini, S. Perugini, L. Pilolli, P. Pistilli. G. Rizzo, F. Rori, M. Rossini, G. Santin, E. Sirizotti, A. Smaniotto, U. Siciliani, M. Tiburzi, R. Meyer, A. Goda, B. Filipiak, T. Vali, M. Helm, and R. Ghodsi, "A 128Gb 3b/cell NAND flash design using 20nm planer-cell technology," in *IEEE International Solid-State Circuits Conference (ISSCC) Digest of Technical Papers*, Feb. 2013, pp. 218-219.

[10] M. Bauer, R. Alexis, G. Atwood, B. Baltar, A. Fazio, K. Frary, M. Hensel, M. Ishac, J. Javanifard, M. Landgraf, D. Leak, K. Loe, D. Mills, P. Ruby, R. Rozman, S. Sweha, S. Talreja, and K. Wojciechowski, "A multilevel-cell 32Mb flash memory," in *IEEE International Solid-State Circuits Conference (ISSCC) Digest of Technical Papers*, Feb. 1995, pp. 132-133.

[11] K. Takeuchi, T. Tanaka, and T. Tanzawa, "A multi-page cell architecture for high-speed programming multi-level NAND flash memories," in *IEEE Symposium on VLSI Circuits Digest of Technical Papers*, Jun. 1997, pp. 67-68.

[12] HP, "Solid state drive technology: Differences between SLC, MLC and TLC NAND," http://h10032.www1.hp.com/ctg/Manual/c03757461.pdf, May 2013, Rev. 1.

[13] R. F. Freitas and W. W. Wilcke, "Storage-class memory: The next storage system technology," *IBM Journal of Research and Development*, vol. 52, no. 4/5, pp. 439-447, Jul. 2008.

[14] G. W. Burr, B. N. Kurdi, J. C. Scott, C. H. Lam, K. Gopalakrishnan, and R. S. Shenoy, "Overview of candidate device technologies for storage-class memory," *IBM Journal of Research and Development*, vol. 52, no. 4/5, pp. 449-464, Jul. 2008.

[15] K. Tsuchida, T. Inaba, K. Fujita, Y. Ueda, T. Shimizu, Y. Asao, T. Kajiyama, M. Iwayama, K. Sugiura, S. Ikegawa, T. Kishi, T. Kai, M. Amano, N. Shimomura, H. Yoda, and Y. Watanabe, "A 64Mb MRAM with clamped reference and adequate-reference schemes," in *IEEE International Solid-State Circuits Conference (ISSCC) Digest of Technical Papers*, Feb. 2010, pp. 258-259.

[16] S.-W. Chung, T. Kishi, J. W. Park, M. Yoshikawa, K. S. Park, T. Nagase, K. Sunouchi, H. Kanaya, G. C. Kim, K. Noma, M. S. Lee, A. Yamamoto, K. M. Rho, K. Tsuchida, S. J. Chung, J. Y. Li, H. S. Chun, H. Oyamatsu, and S. J. Hong, "4Gbit density STT-MRAM using perpendicular MTJ realized with compact cell structure," in *IEEE International Electron Devices Meeting (IEDM) Technical Digest*, Dec. 2016, pp. 27.1.1-27.1.4.

[17] A. Kawahara, R. Azuma, Y. Ikeda, K. Kawai, Y. Katoh, K. Tanabe, T. Nakamura, Y. Sumimoto, N. Yamada, N. Nakai, S. Sakamoto, Y. Hayakawa, K. Tsuji, S. Yoneda, A. Himeno, K. Origasa, K. Shimakawa, T. Takagi, T. Mikawa, and K. Aono, "An 8Mb multi-layered cross-point ReRAM

macro with 43 MB/s write throughput," *IEEE Journal of Solid-State Circuits (JSSC)*, vol. 48, no. 1, pp. 178-185, Oct. 2013.

[18] K. Kawai, A. Kawahara, R. Yasuhara, S. Muraoka, Z. Wei, R. Azuma, K. Tanabe, and K. Shimakawa, "Highly-reliable TaOx ReRAM technology using automatic forming circuit," in *Proceedings of IEEE International Conference on IC Design and Technology (ICICDT)*, May 2014, pp. 100-103.

[19] K.-J. Lee, B.-H. Cho, W.-Y. Cho, S. Kang, B.-G. Choi, H.-R. Oh, C.-S. Lee, H.-J. Kim, J.-M. Park, Q. Wang, M.-H. Park, Y.-H. Ro, J.-Y. Choi, K.-S. Kim, Y.-R. Kim, I.-C. Shin, K.-W. Lim, H.-K. Cho, C.-H. Choi, W.-R. Chung, D.-E. Kim, Y.-J. Yoon, K.-S. Yu, G.-T. Jeong, H.-S. Jeong, C.-K. Kwak, C.-H. Kim, K. Kim, "A 90nm 1.8 V 512 Mb diode-switch PRAM with 266 MB/s read throughput," *IEEE Journal of Solid-State Circuits (JSSC)*, vol. 43, no. 1, pp. 150-162, Jan. 2008.

[20] Y. Choi, I. Song, M. Park, H. Chung, S. Chang, B. Cho, J. Kim, Y. Oh, D. Kwon, J. Sunwoo, J. Shin, Y. Rho, C. Lee, M. G. Kang, J. Lee, Y. Kwon, S. Kim, J. Kim, Y. Lee, Q. Wang, S. Cha, S. Ahn, H. Horii, J. Lee, K. Kim, H. Joo, K. Lee, Y. Lee, J. Yoo, and G. Jeong, "A 20nm 1.8V 8Gb PRAM with 40MB/s program bandwidth," in *IEEE International Solid-State Circuits Conference (ISSCC) Digest of Technical Papers*, Feb. 2012, pp. 46-47.

[21] IBM Almaden Research Center, "Storage class memory: Towards a disruptively low-cost solid-state non-volatile memory," http://researcher.watson.ibm.com/researcher/files/us-gwburr/Almaden_SCM_overview_Jan2013.pdf, Jan. 2013.

[22] H. Fujii, K. Miyaji, K. Johguchi, K. Higuchi, C. Sun, and K. Takeuchi, "x11 performance increase, x6.9 endurance enhancement, 93% energy reduction of 3D TSV-integrated hybrid ReRAM/MLC NAND SSDs by data fragmentation suppression,"in *IEEE Symposium on VLSI Circuits Digest of Technical Papers*, Jun. 2012, pp. 134-135.

[23] C. Sun, K. Miyaji, K. Johguchi, and K. Takeuchi, "A high performance and energy-efficient cold data eviction algorithm for 3D-TSV hybrid ReRAM/MLC NAND SSD," *IEEE Transactions on Circuits and Systems-I (TCAS-I)*, vol. 61, no. 2, pp. 382-392, Feb. 2014.

[24] S. Okamoto, C. Sun, S. Hachiya, T. Yamada, Y. Saito, T. O. Iwasaki, and K. Takeuchi, "Application driven SCM and NAND flash hybrid SSD design for data-centric computing system," in *Proceedings of IEEE International Memory Workshop (IMW)*, May 2015, pp. 157-160.

[25] Intel Optane Technology, http://www.intel.com/content/www/us/en/architecture-and-technology/intel-optane-technology.html.

第2章 次世代コンピュータアーキテクチャにおけるストレージの研究課題

2.1 はじめに

本章では不揮発性半導体メモリを用いたストレージの従来研究および次世代のコンピュータアーキテクチャにおけるストレージの研究課題について述べる．まず不揮発性半導体メモリであるストレージクラスメモリ（storage class memory，SCM）と NAND 型フラッシュメモリの動作および特性について述べる．SCM はその特性によってメモリタイプおよびストレージタイプに分類でき，一方で NAND 型フラッシュメモリはセル当たりに保存するビット数によって異なる特性を持つことを示す．続いて，新しい不揮発性半導体メモリである SCM を用いた次世代のコンピュータアーキテクチャについて述べる．SCM を用いたストレージに関するこれまでの研究状況について述べ，次世代のコンピュータアーキテクチャにおけるストレージの研究課題を論じる．

2.2 次世代コンピュータアーキテクチャ

SCM の出現により次世代のコンピュータアーキテクチャは大きく変わろうとしている．図 2.1 に現在のノイマン型コンピュータアーキテクチャを示す．ノイマン型コンピュータの基本ハードウエア構成は，プロセッサ（central processing unit，CPU），主記憶（メインメモリ），入出力（Input/Output，I/O）および周辺装置からなる．周辺装置は二次記憶（ストレージ），ディスプレイ，キーボード，マウスなどを含む．ノイマン型コンピュータの主な特徴として処理（プログラム）と情報（データ）の分離がある．データの処理はプロセッサで行い，データの保存は半導体メモリで行う[1][2]. データは逐次メインメモリからキャッシュへ送られ CPU で処理される．半導体メモリは図 1.1 のように階層構造を構成する．CPU 内のレジスタがもっとも高速で，static random access memory（static RAM，SRAM）はメインメモリのキャッシュとして用いられる．SRAM キャッシュもまた L1，L2，L3 と階層構造を成し，L1 キャッシュは高速で小容量，L2 および L3 キャッシュは低速で大容量という特徴がある．頻繁にアクセスするデータは高速な L1 キャッシュに，頻繁にアクセスしないデータは L2 あるいは L3 キ

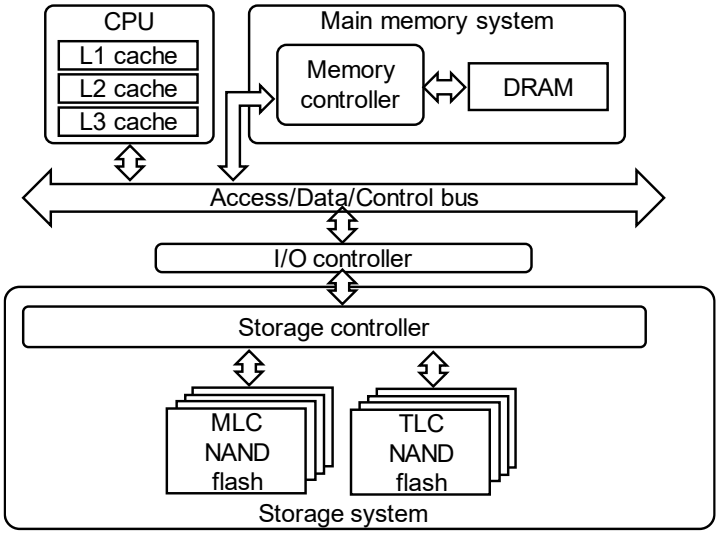


図 2.1 現在のノイマン型コンピュータアーキテクチャ

ャッシュに保存する．メインメモリとして高速で比較的大容量な dynamic RAM（DRAM）が用いられる．DRAM はキャパシタに電荷を保存することでデータを保持するため，電源を消失するとデータを失う．またリーク電流により電荷が失われるとデータのエラーが発生するため，定期的なデータリフレッシュを行う必要がある．DRAM は SRAM キャッシュと比較して十分大容量であるが，DRAM 内の頻繁にアクセスされないデータは二次記憶に移動し保存する．二次記憶として従来ハードディスクドライブ（hard-disk drive，HDD）が用いられてきたが機械部品による突然の故障が発生する．そのため HDD は不揮発性半導体メモリである NAND 型フラッシュメモリを用いたソリッドステートドライブ（solid-state drive，SSD）に置き換わりつつある．このようにノイマン型コンピュータでは，CPU とメモリ間のデータ転送時の衝突やアクセス性能差がコンピュータアーキテクチャ全体の性能のボトルネックとなる．これをフォンノイマン・ボトルネック（von Neumann bottleneck）と呼ぶ．データ転送時の衝突を回避し高速化するために CPU は投機実行などを行う．一方で半導体メモリの観点から見ると図 1.1 で示したように，SRAM と DRAM との間には約 10^1 倍，DRAM と NAND 型フラッシュメモリとの間には約 10^4 倍のアクセス性能差がある．

フォンノイマン・ボトルネックとなる DRAM と NAND 型フラッシュメモリとのアクセス性能差を埋めるため，ストレージクラスメモリ（storage class memory，SCM）と呼ばれる不揮発性半導体メモリが研究, 開発されてきた[3][4]. SCM は magnetoresistive RAM（MRAM）[5][6], resistive RAM（ReRAM）[7][8][9]，phase change RAM（PRAM）[10][11][12]など次世代の不揮

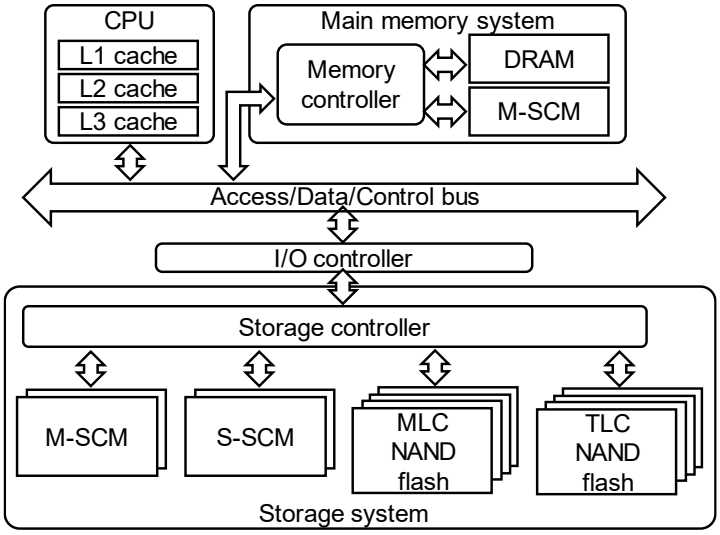


図 2.2 M-SCM および S-SCM を用いた将来のコンピュータアーキテクチャ

発性半導体メモリの総称である．SCM は DRAM より大容量で NAND 型フラッシュメモリより高いアクセス性能を持つことを特徴とする．SCM を用いた次世代のコンピュータアーキテクチャを図 2.2 に示す．SCM はメインメモリシステムおよびストレージシステムにおいて用いられる．図 1.1 に示したように SCM はメモリタイプ（memory-type SCM，M-SCM）とストレージタイプ（storage-type SCM，S-SCM）とに分類できる．M-SCM は DRAM により近い特性を持つためメインメモリシステムで用いられる. メインメモリシステムでは例えば, DRAM とハイブリッド化し低いアクセス頻度のデータを M-SCM に保存する研究が行われている [13]. またストレージシステムでは M-SCM および S-SCM を用いることを本論文で提案する．M-SCM および S-SCM はフォンノイマン・ボトルネックである DRAM と NAND 型フラッシュメモリ間の性能差を埋めるために用いる．さらに NAND 型フラッシュメモリも，NAND 型フラッシュメモリセル当たりに保存するビット数によって, Single-level cell（SLC，1 bit/cell），Multiple-level cell（MLC，2 bit/cell），Triple-level cell（TLC，3 bit/cell）が存在しそれぞれメモリ特性が異なる．これらの半導体メモリはその特性に一長一短があるため，高速，大容量，不揮発なユニバーサルメモリは存在しない．そのためこれらの異種の不揮発性半導体メモリを複数種用いて構成し，本論文でヘテロジニアスストレージと呼ぶ.

2.3 不揮発性半導体メモリ

本論文で提案するヘテロジニアスストレージを構成する不揮発性半導体メモリである，SCM および NAND 型フラッシュメモリについてその動作およびその特性を述べる.

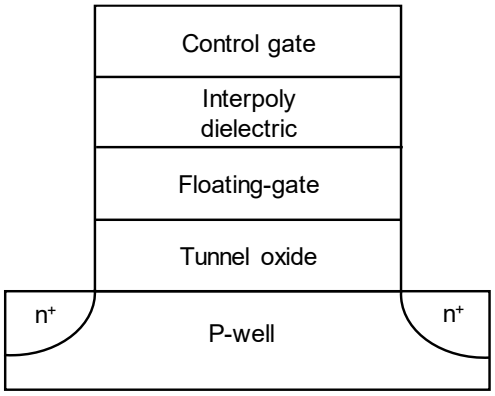


図 2.3 フローティングゲート型 NAND 型フラッシュメモリセル [14]

2.3.1 NAND 型フラッシュメモリ

図 2.3 に NAND 型フラッシュメモリセルを示す. P 型半導体基板（P-well）上に酸化膜（tunnel oxide）と制御ゲート（control gate，CG）を構成する MOS トランジスタと比較して，NAND 型フラッシュメモリセルは酸化膜と制御ゲートとの間に浮遊ゲート（floating gate，FG）層を構成する点が異なる[14]. 浮遊ゲートは絶縁膜により電気的に隔離されており，電源を与えることなく浮遊ゲート内に保存された電子を保持することができるため不揮発である．NAND 型フラッシュメモリセルの書き込みおよび読み出しは，制御ゲートと P 型半導体基板間に電圧を加えることで行う．データ書き込み時は P 型半導体基板に 0V および制御ゲートに 20V を加えることで，Fowler-Nordheim トンネリングにより電子は NAND 型フラッシュメモリセルの浮遊ゲートに注入される．このとき論理的な"0"状態にしきい値電圧が上昇する[15]．一方でデータ消去時，トンネル電圧を書き込み時とは反対に P 型半導体基板に 20V および制御ゲートに 0V を加えることで，電子を浮遊ゲートから排出する．これはセルのしきい値電圧を低下させ，論理的な"1"状態にする．つまりデータの書き込み/消去（write/erase）を行うと，NAND 型フラッシュメモリセルの P 型半導体基板と酸化膜との間で電子の移動が起こる．何度もデータの書き込み/消去を行うと酸化膜が劣化するため NAND 型フラッシュメモリの書き換え回数（Write/erase cycle）は制限されている．

図 2.4 に NAND 型フラッシュメモリの回路を示す[16]．NAND 型フラッシュメモリの回路は，ビットライン（bit-line，BL）とワードライン（word-line，WL）の交点に NAND 型フラッシュメモリセルが格子状に並ぶ．ビットラインはそれぞれの NAND 型フラッシュメモリセルのソースとドレインに接続される．ワードラインはまた，それぞれの NAND 型フラッシュメモリセルの制御ゲートに接続される．ワードラインおよびビットラインの電圧を制御するこ

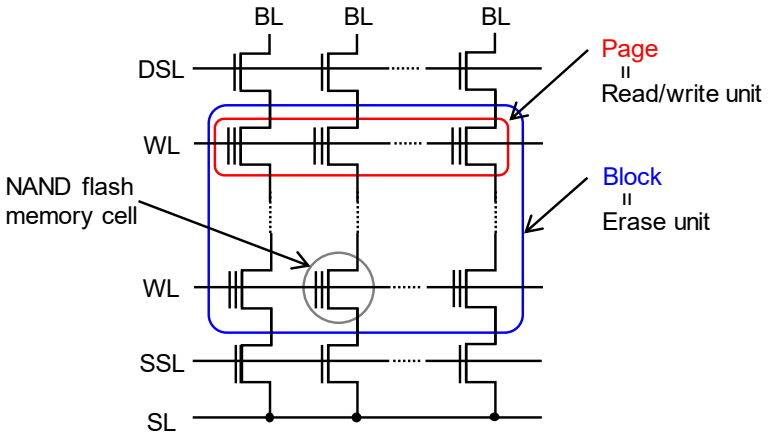


図 2.4 NAND 型フラッシュメモリ回路 [16]

とで，NAND 型フラッシュメモリセルの書き込み，読み出し，消去を行う．NAND 型フラッシュメモリ回路にはページおよびブロックの単位がある．ページは同一のワードラインを共有する NAND 型フラッシュメモリセルの集まりで，ワードライン電圧を制御することで同時に読み出し，書き込みができる．ブロックは同一のビットラインを共有する NAND 型フラッシュメモリセルの集まりである．NAND 型フラッシュメモリセルのデータ消去は P 型半導体基板に 20V 加えることで達成されるため，同一のビットラインを共有する複数のページからなるブロック単位でのみ消去できる．このように NAND 型フラッシュメモリは読み出し・書き込み動作はページ単位で，消去動作はブロック単位で行われる．NAND 型フラッシュメモリセルの書き込み・消去単位の非対称性はチップ面積を削減しコスト低減する目的であり，現在ではさまざまな用途に NAND 型フラッシュメモリが用いられている．

NAND 型フラッシュメモリの書き込み・消去単位の非対称性のため，同一ページ内で上書き動作ができない[17][18][19]．そのため NAND 型フラッシュメモリのページの状態を有効（valid），空（free），無効（invalid）と表しコントローラが管理する．有効ページは現存のデータを保存したページである．空のページはデータが無い状態であり，データを書き込むことができる．無効ページはすでにデータが書き込まれているが，データは古いあるいは無効な状態である．書き込み・消去単位の非対称性により，ブロック全体を消去しなければ無効ページにデータを書き込むことができない．NAND 型フラッシュメモリに上書きするとき，古いページのデータを読み出し，新しいデータと統合した後，同じあるいは別のブロックの空きページに書き込まれる（図 2.5）．その後古いページを無効化し，ブロックを消去する準備を行う．NAND 型フラッシュメモリのコントローラのフラッシュトランスレーションレイ

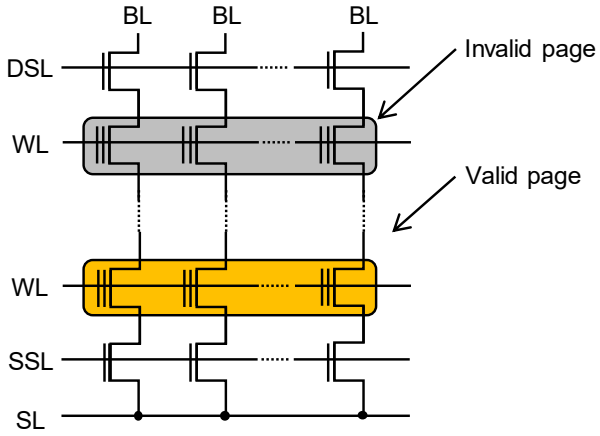


図 2.5 NAND 型フラッシュメモリのページ管理

ヤ（flash translation layer，FTL）の機能の一つは，これらの複雑なページの状態を管理することである．

ページの上書きによって無効ページが増えると，書き込みに使える空ページの数が減少する．そのため無効ページを消去し，空ページをつくるガベージコレクション（garbage collection，GC）を行う必要がある．実際の NAND 型フラッシュメモリでは，空ページの数があらかじめ決めたしきい値より少なくなると GC が行われる．図 2.6 に GC 動作を示す．初めに消去予定のブロックの有効ページをコントローラに読み出しエラー訂正を行う．消去予定ブロック内の有効ページ数を N_{valid} とする．次にエラー訂正を行った有効ページのデータを，消去予定のブロックとは異なるブロックに書き込む．最後に消去予定のブロックを消去する．これらの GC 動作に要する時間は，有効ページを消去予定のブロックから別のブロックにコピーする時間と，ブロックを消去する時間との和となる．消去するブロックの有効ページ数が多い場合，GC に要する時間は 100 msec を超える[20][21]．また消去するブロックの有効ページ数が多いと write amplification が起きる[22]．Write amplification とは NAND 型フラッシュメモリ内部の上書きや GC 動作によって，ホストからの書き込みと比較して NAND 型フラッシュメモリへの書き込み量が増える現象である．Write amplification が起きると NAND 型フラッシュメモリの書き換え回数が増加し，制限のある NAND 型フラッシュメモリの許容書き換え回数を圧迫する．このように GC 動作は NAND 型フラッシュメモリの性能を低下させるボトルネックとなる．

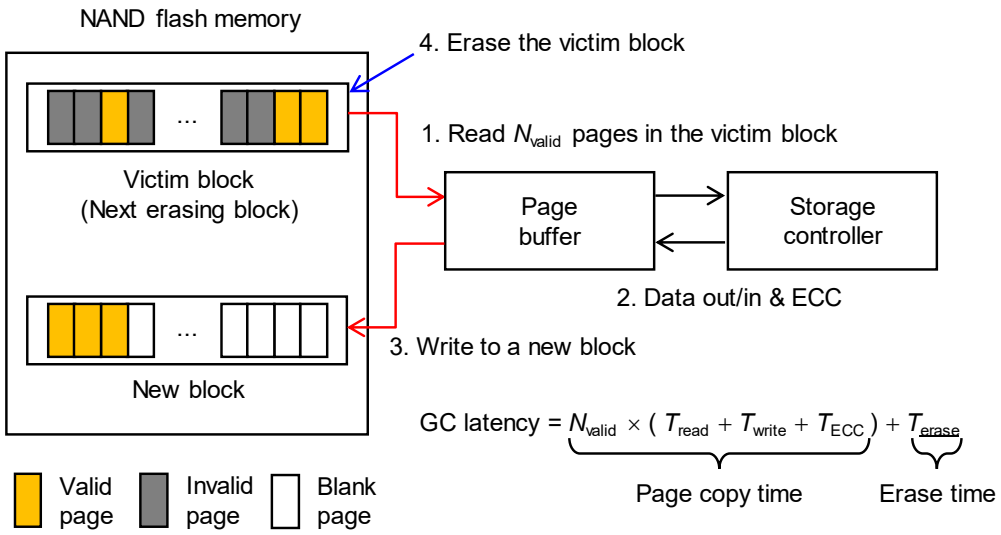


図 2.6 NAND 型フラッシュメモリのガベージコレクション動作 [20]

NAND 型フラッシュメモリは 2 次元のスケーリング[23]，多値化技術[24][25]および 3 次元積層技術[26]によって大容量化しビットコストを低下させている．多値化技術は NAND 型フラッシュメモリのセル当たりに 2 bit 以上を保存する．初期の NAND 型フラッシュメモリは，NAND 型フラッシュメモリのセル当たりに 1 bit を保存する Single-level cell（SLC，1 bit/cell）であった．次に NAND 型フラッシュメモリのセル当たりに 2 bit を保存する Multiple-level cell（MLC，2 bit/cell）が開発され，現在でも多くのストレージ製品に用いられている．近年では NAND 型フラッシュメモリのセル当たりに 3 bit を保存する Triple-level cell（TLC，3 bit/cell）が大容量を目的とする製品に用いられる．また NAND 型フラッシュメモリのセル当たりに 4 bit を保存する Quadruple-level cell（QLC，4 bit/cell）製品の開発も予定されている[27]．図 2.7 に SLC，MLC および TLC NAND 型フラッシュメモリのしきい値（V_{TH}）分布を示す．SLC NAND 型フラッシュメモリは消去状態“1”および書き込み状態“0”と，2（= 2^1）状態のしきい値電圧を持つ．NAND 型フラッシュメモリセルの浮遊ゲートに注入する電子の量を制御することで，MLC NAND 型フラッシュメモリは 4（= 2^2）状態のしきい値電圧を持つ．同様に TLC NAND 型フラッシュメモリは 8（= 2^3）状態のしきい値電圧を持つ．そのため，MLC および TLC NAND 型フラッシュメモリは SLC NAND 型フラッシュメモリと比較して，それぞれ 2 倍，3 倍のビット密度を持つ．しかし図 2.7 から明らかなように，多値化技術を用いた MLC および TLC NAND 型フラッシュメモリは SLC NAND 型フラッシュメモリと比較してしきい値電圧の間隔が狭くなる. そのためしきい値分布を詳細に制御するための時間がかかり, MLC

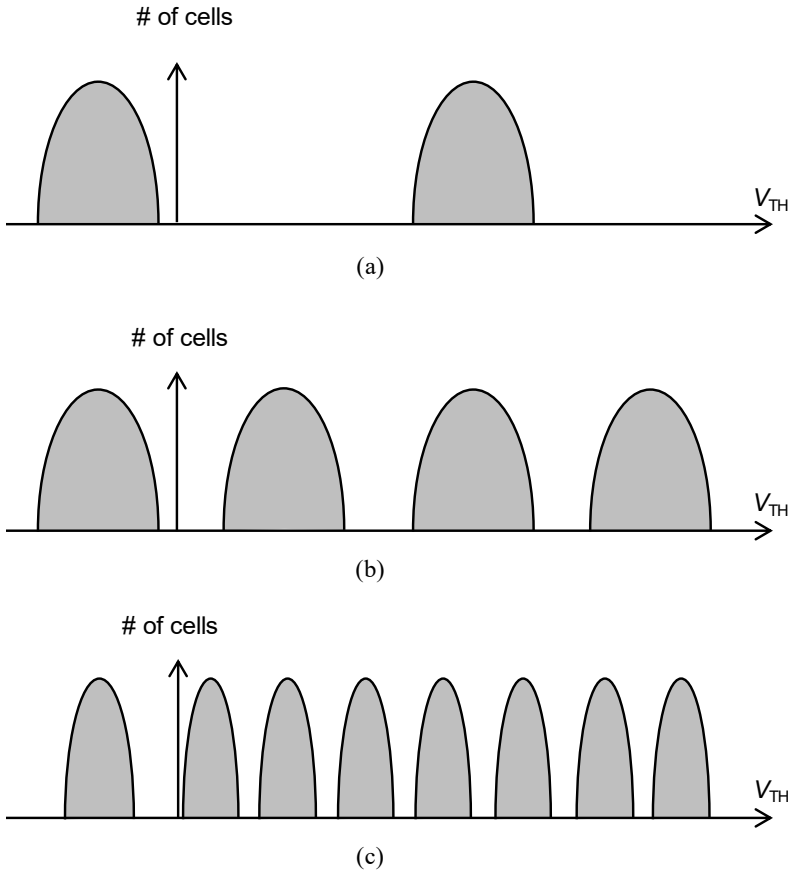


図 2.7 多値化 NAND 型フラッシュメモリのしきい値分布．（a）SLC，（b）MLC，（c）TLC NAND 型フラッシュメモリ

および TLC NAND 型フラッシュメモリは SLC NAND 型フラッシュメモリと比較して読み出し，書き込み，消去時間が長くなる．NAND 型フラッシュメモリは時間経過などにより，トンネル酸化膜中のトラップサイトを経由して電子のトンネリングが発生し，さらにトンネル酸化膜中にトラップされた電子がデトラップされる．また隣接セルに書き込む際，電子が誤って注入されるなどしてビット反転が起き，異なるしきい値電圧に変化するエラーが起きる．特に TLC NAND 型フラッシュメモリはしきい値電圧の間隔が狭いためビット反転が起きやすい．

図 2.8，図 2.9 に MLC および TLC NAND 型フラッシュメモリを実測した（a）ページ読み出し時間，（b）ページ書き込み時間，（c）ブロック消去時間，（d）ビットエラーレート（bit error rate，BER）を示す[28]．NAND 型フラッシュメモリの書き換え（W/E cycle）に対して（a）ページ読み出し時間はほぼ一定である．しかし NAND 型フラッシュメモリの書き換えによってトンネル酸化膜に捕獲された電子が増加する[29]．その結果しきい値電圧が上昇するため，

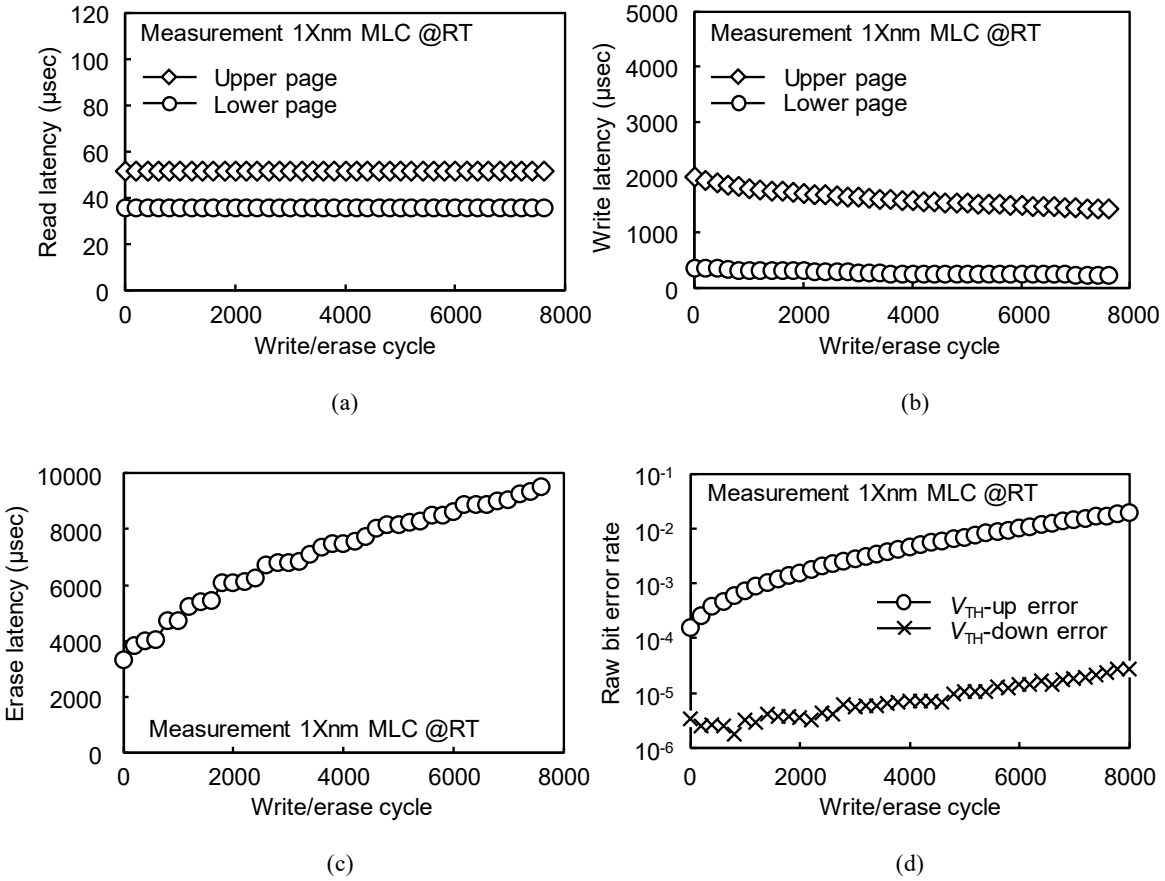


図 2.8 MLC NAND 型フラッシュメモリの（a）ページ読み出し時間，（b）ページ書き込み時間，（c）ブロック消去時間，（d）Bit error rate [28]

W/E cycle の増加によって（b）ページ書き込み時間は減少し，（c）ブロック消去時間は増加する．また W/E cycle の増加によってトンネル酸化膜が劣化し（d）BER が増加する．次に MLC および TLC NAND 型フラッシュメモリの動作時間を比較する．TLC NAND 型フラッシュメモリは MLC NAND 型フラッシュメモリと比較してしきい値状態の数が多いため，しきい値分布を詳細に制御するための時間がかかり，TLC NAND 型フラッシュメモリは MLC NAND 型フラッシュメモリと比較して（a）ページ読み出し，（b）ページ書き込み，（c）ブロック消去時間が長くなる．また（d）BER のようにしきい値が上がるエラー（V_{TH}-up error）としきい値が下がるエラー（V_{TH}-down error）を区別して計測した．特に TLC NAND 型フラッシュメモリはしきい値分布が狭いため，書き換えによってしきい値が下がるエラーが増加する．NAND 型フラッシュメモリはメモリコストが市場を拡大する推進力であるため，多値化技術による欠点があるにも関わらずセル当たりにより多くのビットが保存される．

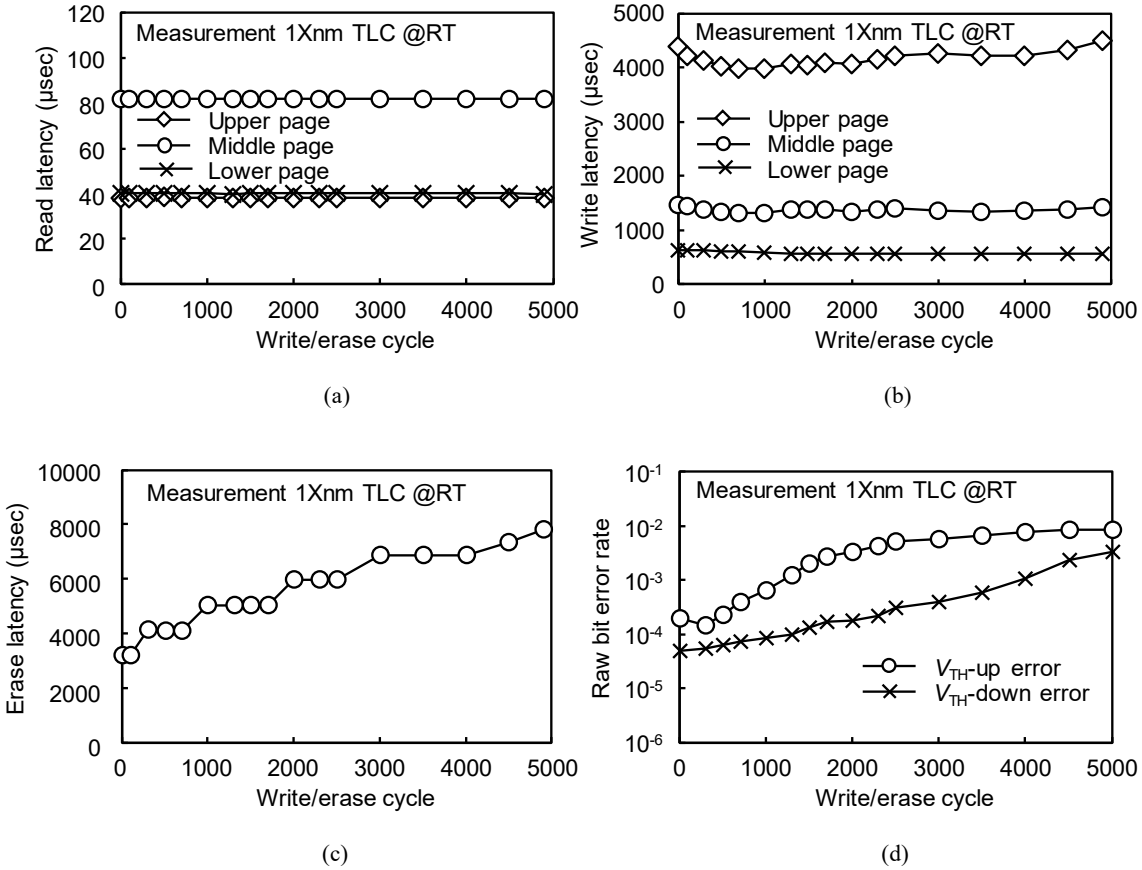


図 2.9 TLC NAND 型フラッシュメモリの（a）ページ読み出し時間，（b）ページ書き込み時間，（c）ブロック消去時間，（d）Bit error rate [28]

2.3.2 ストレージクラスメモリ（SCM）

図 1.1 に示したメモリおよびストレージ階層において，ストレージクラスメモリ（storage class memory，SCM）は DRAM と NAND 型フラッシュメモリとの間に位置する．SCM は DRAM と NAND 型フラッシュメモリとの間の特性を有する不揮発性半導体メモリのことを指す総称である．磁気抵抗型メモリ（magnetoresistive RAM，MRAM）[5][6]，抵抗変化型メモリ（resistive RAM，ReRAM）[7][8][9]，相変化メモリ（phase change RAM，PRAM）[10][11][12] が SCM の候補である．以下に SCM の動作を示す．

図 2.10 にスピン注入型磁気抵抗型メモリ（spin transfer torque MRAM，STT-MRAM）の書き込み原理を示す．STT-MRAM は強磁性体に挟まれた絶縁層から構成され，これを磁気トンネル接合（magnetic tunnel junction，MTJ）素子と呼ぶ．強磁性体に電圧を加えることで free layer の磁化が reference layer の磁化と同一極性あるいは反極性に変化しトンネル電流が流れる．Free layer の磁化が reference layer の磁化と同一極性のとき高抵抗状態で論理的な“1”となり，

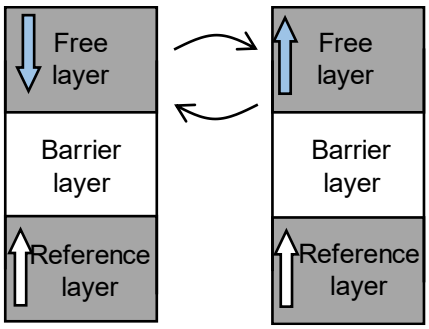


図 2.10 磁気抵抗型メモリ（MRAM）の動作 [5]

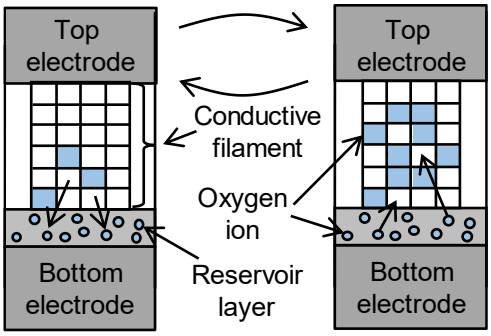


図 2.11 抵抗変化型メモリ（ReRAM）の動作 [9]

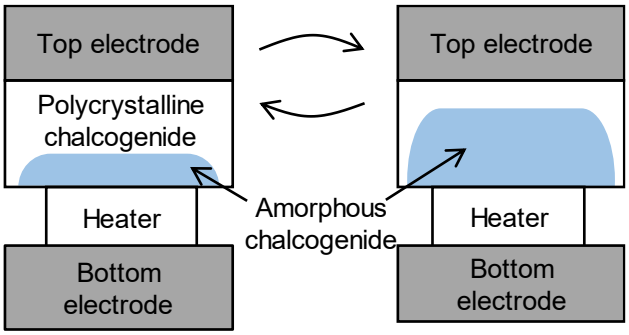


図 2.12 相変化メモリ（PRAM）の動作 [11]

反極性のとき低抵抗状態となり論理的な“0”を表す.

図 2.11 に ReRAM の書き込み原理を示す．ReRAM は一般に遷移金属酸化物層と上下の電極（top/bottom electrode）から構成される．図 2.11 に示す ReRAM の遷移金属酸化物層は，導電性フィラメント（conductive filament）を形成する層と酸素イオンを貯める層（reservoir layer）

を持つ[7]．下部電極に電圧を加えると酸素イオンが下部の reservoir layer へ移動し，導電性フィラメントには酸素欠陥が発生する．このとき導電性フィラメントは電流が流れやすい低抵抗状態（“1”）となる．また上部電極に電圧を加えると reservoir layer から導電性フィラメントへ酸素イオンが拡散する．酸素イオンによって導電性フィラメント内の酸素欠陥が減少し，電流が流れにくい高抵抗状態（“0”）となる．

図 2.12 に PRAM の書き込み原理を示す．PRAM はカルコゲナイド材料を上下電極で挟んだ構成である．電極に電圧を加え大電流が流れると，ジュール熱によりカルコゲナイド材料の温度が上昇し結晶状態となる．このとき電流が流れやすい低抵抗状態（“1”）と呼ぶ．また高電圧を加えると大電流が流れカルコゲナイド材料は融解する．続いて電圧を下げるとカルコゲナイド材料の温度は急激に低下しアモルファス状態となる．このとき電流が流れにくい高抵抗状態（“0”）となる．このように PRAM はカルコゲナイド材料を結晶化および融解することにより状態を変化させるため，MRAM，ReRAM と比較して消費電力が大きく並列化の点で不利である．

これら将来の不揮発性半導体メモリである SCM は，NAND 型フラッシュメモリより短い読み出し・書き込み時間を持ち，DRAM より大容量であるため，DRAM と NAND 型フラッシュメモリ間の性能差を埋めるメモリとして期待されている．SCM は NAND 型フラッシュメモリと同様に不揮発性である．さらに NAND 型フラッシュメモリの最小アクセス単位がページであるのに対し，SCM はバイトアクセスが可能である．また，NAND 型フラッシュメモリがアクセス単位の非対称性により GC 動作が必要であるのに対し，ストレージで用いる SCM は HDD と同じくセクタ（ブロックともいう，512 Byte）で行うため消去動作は不要で同一セクタで上書き可能である．しかしほぼ無制限の書き換え回数を許容する DRAM と比較して，SCM は上記で述べた動作原理により書き換え回数に制限があるため，完全なユニバーサルメモリではない．

SCM はそれぞれ書き込み原理が異なるため，MRAM，ReRAM，PRAM は異なる特性を持つ．SCM のアクセス速度と容量に応じて，SCM はメモリタイプ SCM（memory-type SCM，M-SCM）とストレージタイプ SCM（storage-type SCM，S-SCM）に分類できる[30]．本研究では，MRAM は DRAM 並みに高速化が実現可能で高い書き換え耐久性を持つため M-SCM として分類する．一方で ReRAM や PRAM は，SLC NAND 型フラッシュメモリと同等の大容量を実現し低コストとなる可能性があるため S-SCM に分類する．表 2.1 および表 2.2 に本研究で仮定した不揮発性半導体メモリの特性をまとめる[31]. M-SCM は S-SCM と比較して約 10^1-

表 2.1 不揮発性半導体メモリの特性 [31]

Memory device	SCM	MLC NAND flash	TLC NAND flash
Read latency	表2.2に示す	52 µsec/page (U) 36 µsec/page (L)	80 µsec/page (U) 100 µsec/page (M) 80 µsec/page (L)
Write latency		2000 µsec/page (U) 370 µsec/page (L)	4400 µsec/page (U) 1500 µsec/page (M) 640 µsec/page (L)
Erase latency	Not required	3300 µsec/block	3200 µsec/block
I/O frequency	1066 MHz	400 MHz	
V_{DD} (Core, I/O)	1.8 V, 1.2 V	3.3 V, 1.8 V	
I (read, write, erase)	20 mA, 40 mA, -	45 mA, 45 mA, 45 mA	
Minimum access unit	Sector (512 Byte)	Page (16 KByte)	Block (4.03 MByte)
Acceptable endurance	表2.2に示す	10^4	10^3
Bit cost		1	2/3

U: Upper page, M: Middle page, L: Lower page of NAND flash

表 2.2 ストレージクラスメモリ（SCM）の特性 [31]

SCM scenario	1 (M-SCM)	2 (S-SCM)	3 (S-SCM)
Read latency	0.1 µsec	1 µsec	10 µsec
Write latency	0.1 µsec	1 µsec	10 µsec
Acceptable endurance	10^{12}	10^8	10^6
Bit cost	10	6	4

10^2 倍短い読み出し・書き込み時間を持ち，M-SCM は S-SCM より約 10^6-10^2 倍高い書き換え耐久性を持つと仮定した．DRAM のビットコストが MLC NAND 型フラッシュメモリの約 12 倍であることから[32], M-SCM および S-SCM のビットコストはそれぞれ MLC NAND 型フラッシュメモリの 10 倍，6 倍，4 倍と仮定した[31]．M-SCM のアクセス時間は DRAM 並みに短いが, その書き込みエネルギーは DRAM より高い. 一方で, S-SCM の書き換え回数は MLC NAND 型フラッシュメモリより約 10^2-10^4 倍多い．S-SCM は NAND 型フラッシュメモリより高コストだが，M-SCM と比較すると費用対効果が高い.

2.4 SCM および NAND 型フラッシュメモリを用いたストレージ

ストレージクラスメモリ（SCM）の出現により，図 2.2 のように次世代のコンピュータア

ーキテクチャが大きく変わろうとしている．現在，メインメモリには DRAM，二次記憶（ストレージ）には NAND 型フラッシュメモリあるいは HDD が主に用いられている．SCM はメインメモリシステムおよびストレージシステムの性能を向上させると期待される．メインメモリシステムでは，SCM はシステムの消費エネルギーを削減するのに役立つ．SRAM のリーク電力を無くすために，MRAM と SRAM をハイブリッド化した L2 キャッシュ[33]，MRAM と DRAM をハイブリッド化した last level cache（LLC）が提案されている[34]．さらに，PRAM および DRAM をハイブリッド化したメインメモリは，PDRAM と呼ばれ，待機電力を削減する[35]．また，SCM に対応した Linux ベースのファイルシステムとして，Diret Access（DAX）[36]がある．SCM から読み出したデータをカーネル空間に DMA 転送し，カーネル空間から直接 CPU へデータ転送する．ファイルシステム経由だが，カーネル空間からユーザ空間へコピーしない点を直接アクセスできるという．また，バイト単位でも SCM へアクセス可能である．

2.4.1 SCM および NAND 型フラッシュメモリを用いたハイブリッドストレージ

一方で，ストレージ性能を向上させるため，SCM と NAND 型フラッシュメモリを用いたハイブリッドストレージシステムの研究が行われている．SCM は小さいサイズのデータを保存し，NAND 型フラッシュメモリの write amplification を減少する．例えばデータの種類によって書き込む不揮発性半導体メモリを変える．NAND 型フラッシュメモリがユーザデータだけを保存するのに対し，SCM はメタデータ，論物変換テーブル，ECC のパリティビットを保存するのに用いられる．たとえば Sun ら[37]は PRAM と NAND 型フラッシュメモリでデータを分けるハイブリッドストレージを提案している．その中で，PRAM は頻繁に更新されるログデータを保存するために用いられ，この結果，NAND 型フラッシュメモリの GC オーバーヘッドが減少する．Fujii ら[21]は ReRAM と NAND 型フラッシュメモリを用いたハイブリッドストレージにおいて, ホットかつランダムなデータを ReRAM に保存することを提案した．この結果 NAND 型フラッシュメモリのみを用いたストレージと比較して，書き込みだけを行うアプリケーションに対して 10 倍以上性能を向上させることを示した．

また，SCM は不揮発性キャッシュあるいは小容量のストレージとして用いられる．SCM と NAND 型フラッシュメモリとを階層化して構成したハイブリッドストレージは，NAND 型フラッシュメモリを用いたストレージの性能を向上させる有望な手法である．いくつかのハイブリッドストレージの構成およびデータマネジメントアルゴリズムが提案されている[38][39][40]．これらのアルゴリズムでは，頻繁にアクセスされるホットデータ（hot data）は

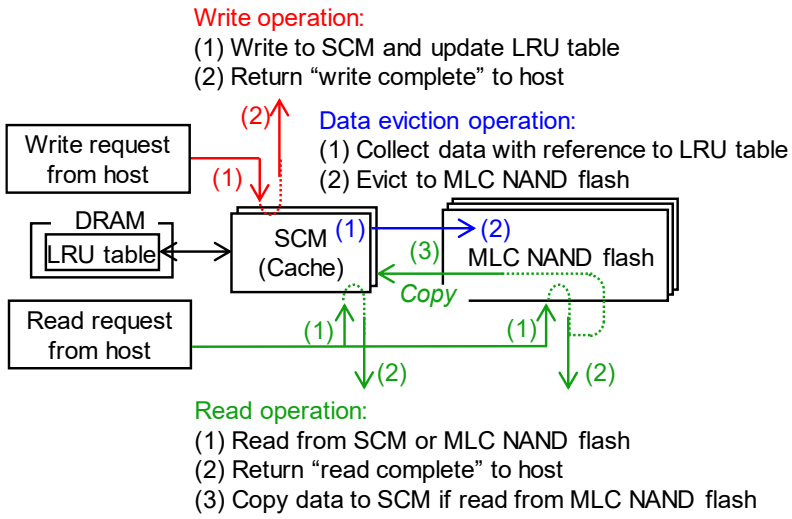


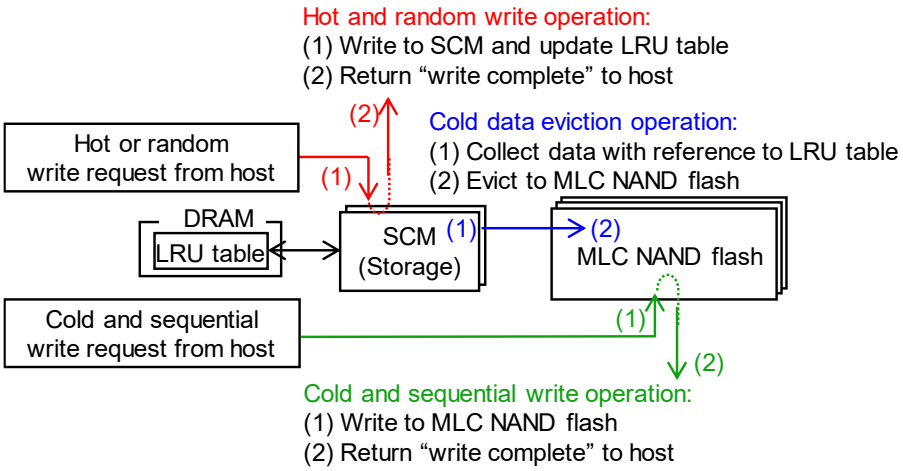
図 2.13 不揮発性半導体メモリ向けライトバックキャッシュ（Non-Volatile Memory Write-Back, NV-WB）アルゴリズム [39]

SCM に保存される．従来の揮発性半導体メモリを用いたキャッシュアルゴリズムでは，キャッシュミスおよびデータの一貫性を保証するために，ホットデータはキャッシュメモリおよび不揮発性ストレージの両方に保存されている．しかし，以下で説明するデータマネジメントアルゴリズムは SCM の不揮発性のために，電源遮断時にデータを退避させる必要がなく，SCM あるいは NAND 型フラッシュメモリだけに保存することもできる．

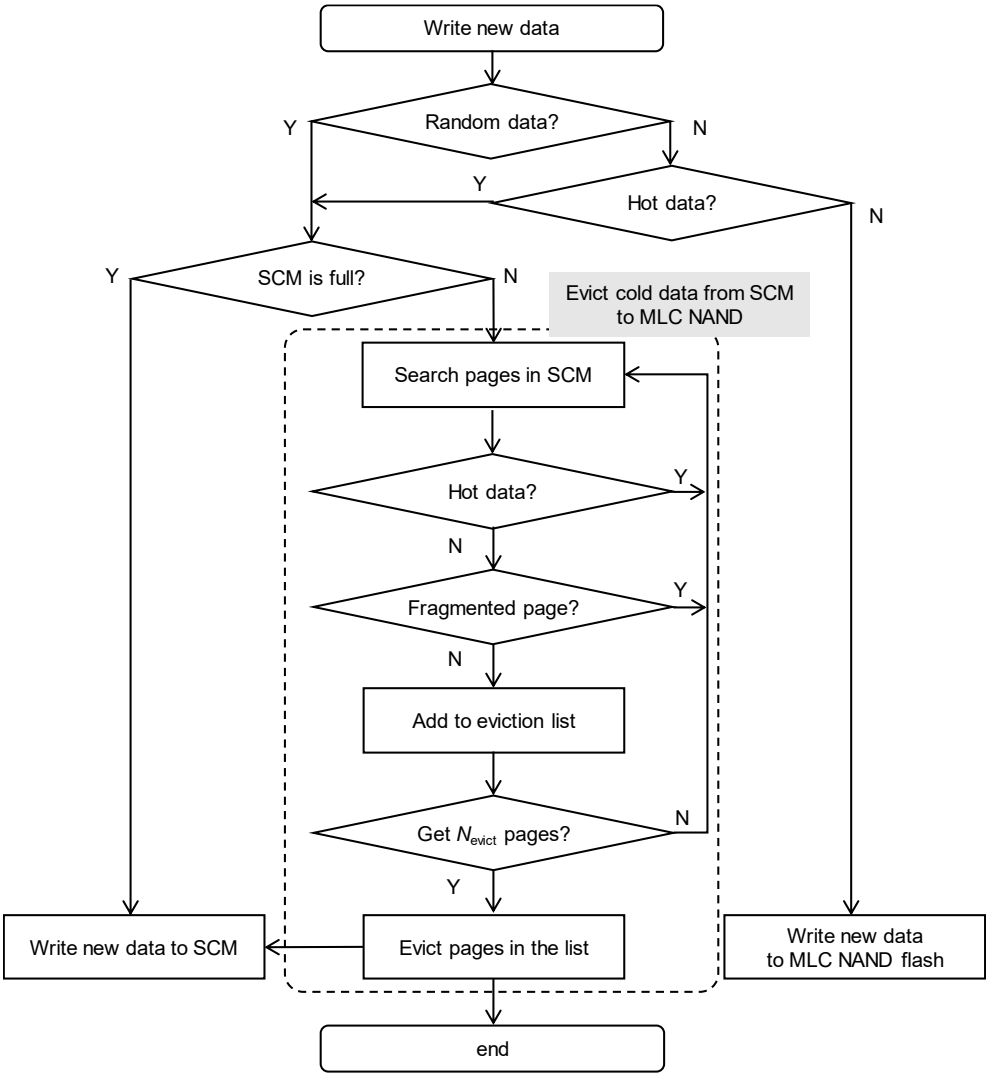
図 2.13 は SCM と NAND 型フラッシュメモリを用いたハイブリッドストレージに用いる，不揮発性半導体メモリ向けライトバック（Non-Volatile Memory Write-Back，NV-WB）キャッシュデータマネジメントアルゴリズムを示す[39][40]. DRAM がその揮発性のために定期的なデータフラッシュ動作を必要とする．しかし SCM の不揮発性のために，NV-WB キャッシュは定期的なデータフラッシュ動作が不要である．加えて揮発性の DRAM と異なり，NV-WB キャッシュは突然の電源障害に対して安全である．NV-WB キャッシュアルゴリズムでは，すべてのデータは初めにキャッシュメモリとしての SCM に書き込まれる．NV-WB キャッシュアルゴリズムは，すべてのデータを SCM に書き込むため高い性能を示す．したがって高速な SCM がハイブリッドストレージの応答時間を短縮する．SCM がデータでいっぱいになると，least recently used（LRU）順に SCM 内のデータを NAND 型フラッシュメモリに evict する．Evict とは，上位のメモリから下位のメモリへ不要なデータを移動することである．LRU リストは，一番最近にアクセスされた（most recently used，MRU）データから，最も古くアクセスされたデータ（LRU）データの順番を記録する．本研究では SCM 容量が残り 20%になると

evict 動作を発動することとした[39]．SCM からデータを読み出すとき，SCM と NAND 型フラッシュメモリとの間でデータの移動は発生しない．一方 NAND 型フラッシュメモリからデータを読み出すとき，そのデータは近い将来再びアクセスされる可能性があるため SCM へコピーされクリーンデータ（clean data）となる．これに対し，ダーティデータ（dirty data）は SCM への上書きで発生する．SCM と NAND 型フラッシュメモリとの間でデータの一貫性が保たれていれば，そのデータはクリーンであるという．そうでないとき，そのデータはダーティであるという．クリーンデータが SCM から NAND 型フラッシュメモリに evict されるとき，そのデータはすでに NAND 型フラッシュメモリに保存されているため書き込み動作は不要である．しかし，ダーティデータを evict するとき，SCM のデータを NAND 型フラッシュメモリに書き込む必要がある．このようなデータマネジメントにより，頻繁にアクセスされるホットデータ，めったにアクセスされないコールドデータは SCM および NAND 型フラッシュメモリにそれぞれ保存される．NV-WB キャッシュの一つの問題は，頻繁にアクセスされないコールドデータおよびデータサイズの大きいシーケンシャルデータもすべて初めに SCM に保存されるため，ハイブリッドストレージの性能を低下させることである．また SCM 容量が小さい場合に読み出しの多いアプリケーションに対して，SCM と NAND 型フラッシュメモリ間でデータが循環する問題がある[40]．SCM でキャッシュヒットする前にデータが SCM から NAND 型フラッシュメモリへ evicts され，さらに NAND 型フラッシュメモリから読み出しが頻繁に発生する場合データが循環する．

ホットあるいはランダムデータとコールドかつシーケンシャルデータを分離するため，コールドデータエビクション（Cold Data Eviction，CDE）アルゴリズムが提案された[38]．NV-WB キャッシュではたくさんのコールド・シーケンシャルデータが SCM に書き込まれると，頻繁にアクセスされるホットデータが NAND 型フラッシュメモリに evict され SCM でのヒット率が低下する．それに加えて，ランダムデータは NAND 型フラッシュメモリのデータ断片化を起こし，write amplification を発生させる．したがって，CDE アルゴリズムでは，SCM はホットあるいはランダムデータを保存し，NAND 型フラッシュメモリはコールドかつシーケンシャルデータを保存する．SCM と NAND 型フラッシュメモリで重複したデータを保存しないため，SCM はストレージとして用いられる．図 2.14（a）に示すように，ホストからのホットあるいはランダムデータは SCM に書き込まれ，コールドかつシーケンシャルデータは直接 NAND 型フラッシュメモリに書き込まれる．図 2.14（b）に CDE アルゴリズムの詳細を示す．固定したサイズの LRU リストを用い，データのアクセス履歴を記録する．LRU に存在するデータはホットデータと判断される．書き込みリクエストのサイズによって，ランダムあ



(a)



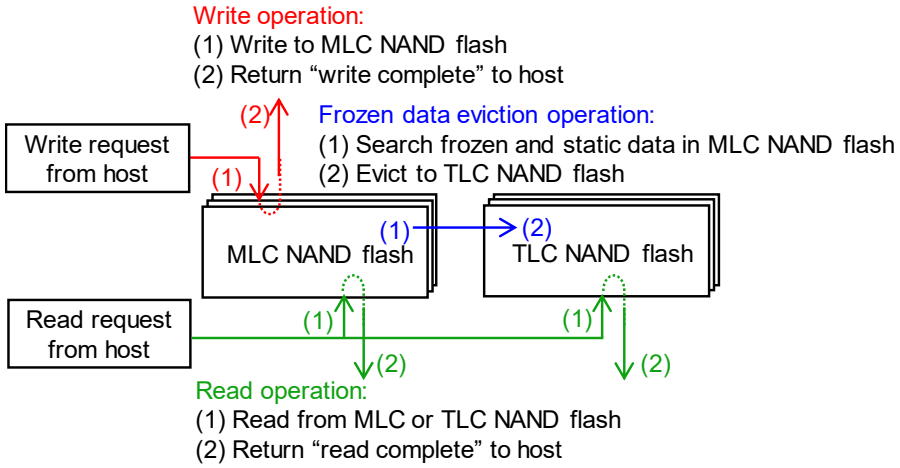
(b)

図 2.14 コールドデータエビクション（Cold Data Eviction，CDE）アルゴリズムの（a）概要，（b）フローチャート [38]

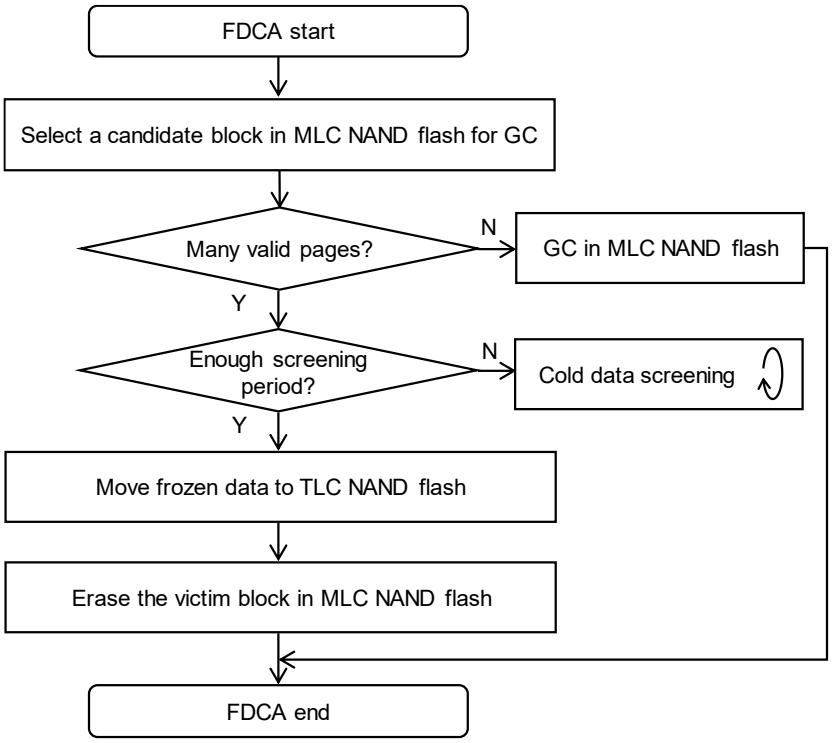
るいはシーケンシャルと判断される．本研究で定義した NAND 型フラッシュメモリのページサイズ（16 KByte）と比較して，以下の図 2.16 で示すアプリケーション分類と同様に，書き込みデータのサイズが NAND 型フラッシュメモリのページサイズの半分より小さいとランダムであると判断される．SCM がデータでいっぱいになると，LRU リストを参照して，コールドかつシーケンシャルデータは SCM から NAND 型フラッシュメモリに evict される．ホストからのリクエストが読み出しの場合，データが保存されたメモリから読み出され，メモリ間のデータの移動は発生しない．したがって，CDE アルゴリズムは書き込みの多いアプリケーションに適している[41].

2.4.2 MLC および TLC NAND 型フラッシュメモリを用いたハイブリッドストレージ

さらにまた，MLC NAND 型フラッシュメモリおよび TLC NAND 型フラッシュメモリを用いたハイブリッドストレージおよびそのデータマネジメントアルゴリズムが提案されている[42]．参考文献[43][44]によると，2D TLC NAND 型フラッシュメモリはセル間干渉を削減するため，その書き込み単位は複数ページをまとめたブロックである．一方で，SLC および MLC NAND 型フラッシュメモリの書き込み単位はページである．つまり NAND 型フラッシュメモリの 1 ページを上書きするとき，SLC および MLC NAND 型フラッシュメモリは古いデータを持つページを読み出し，新しいデータと統合して別のページに書き込む．しかし TLC NAND 型フラッシュメモリの 1 ページを上書きするとき，古いデータを持つページを含む 1 ブロックを読み出し新しいデータと統合して，上書きの必要のないページも別のブロックに書き込む必要がある．このように書き込み単位の大きい TLC NAND 型フラッシュメモリは SLC および MLC NAND 型フラッシュメモリと比較して write amplification が多いため，TLC NAND 型フラッシュメモリの性能は低下し書き換え回数が圧迫される．従来，SRAM が書き込みバッファとしてデータを保存し，TLC NAND 型フラッシュメモリに書き込まれるまで待つ[43]．しかし，第 1 章で述べたように SRAM は高速だが高価で小容量である．MLC および TLC NAND 型フラッシュメモリを用いたハイブリッドストレージにおいて，MLC NAND 型フラッシュメモリは低コストと大容量をバランスした TLC NAND 型フラッシュメモリの書き込みバッファを実現する．TLC NAND 型フラッシュメモリへのアクセスを削減するため，図 2.15（a）に示すラウンドロビン・フローズンデータコレクションアルゴリズム（Round-Robin Frozen Data Collection Algorithm，RR-FDCA）を用いて，TLC NAND 型フラッシュメモリはほとんど上書きされないフローズンデータを保存する．RR-FDCA においてホストからのすべての書き込みリクエストは，書き込みバッファとしての MLC NAND 型フラッシュメモリに書き込まれる．フローズンデータは MLC NAND 型フラッシュメモリの GC 中に，MLC NAND



(a)



(b)

図 2.15 ラウンドロビン・フローズンデータコレクションアルゴリズム（Round-Robin Frozen Data Collection Algorithm，RR-FDCA）の（a）概要，（b）フローチャート [38]

型フラッシュメモリから TLC NAND 型フラッシュメモリにコピーされる．図 2.15（b）にアルゴリズムの詳細を示す. フローズンデータの収集は, MLC NAND 型フラッシュメモリの GC 動作中に行われる．消去するブロックは，NAND 型フラッシュメモリで初めに書き込まれた最も古いブロックがラウンドロビン方式で選択される．その結果，NAND 型フラッシュのウ

エアレベリングが実行される．NAND 型フラッシュメモリのブロックが GC 動作で消去されるとき, もしブロック内のページが上書きされていれば, 有効ページは無効ページに代わる．このため RR-FDCA アルゴリズムでは消去するブロックの有効ページは，MLC NAND 型フラッシュメモリの GC 動作が一巡する間に上書きされなかったフローズンデータであると考える．MLC NAND 型フラッシュメモリのよりフローズンなデータを収集するためブロックの古さを示すカウンタを用いる．MLC NAND 型フラッシュメモリのより大きなカウントを持つブロックは, GC 中に消去するブロックとして選択される．GC 対象として選択した MLC NAND 型フラッシュメモリブロック内の有効ページであるフローズンデータを TLC NAND 型フラッシュメモリにコピーする．その後, MLC NAND 型フラッシュメモリのブロックを消去する．RR-FDCA により MLC NAND 型フラッシュメモリは TLC NAND 型フラッシュメモリに evict することで，フローズンデータを保存することから解放される．その結果，MLC NAND 型フラッシュメモリはホットデータとコールドデータを混在して保存するブロックが減少し，MLC NAND 型フラッシュメモリの GC 時にコピーすべき有効ページ数が減少する．したがって，MLC NAND 型フラッシュメモリのみを用いたストレージと比較して，MLC および TLC NAND 型フラッシュメモリを用いたハイブリッドストレージは，TLC NAND 型フラッシュメモリは自身の書き換え回数を小さく保ったまま，MLC NAND 型フラッシュメモリの寿命を延ばす．RR-FDCA のアルゴリズムを用いると，TLC NAND 型フラッシュメモリはフローズンデータ，MLC NAND 型フラッシュメモリはホットあるいはやや頻繁にアクセスされるワーム（warm）データを保存する．さらに，TLC NAND 型フラッシュメモリはめったに上書きされないため，書き換え回数は小さく保たれる．MLC および TLC NAND 型フラッシュメモリを用いたハイブリッドストレージは, MLC NAND 型フラッシュメモリのみを用いたストレージと比較して高い性能を達成する．しかし SCM を用いたハイブリッドストレージと比較すると，その性能は TLC NAND 型フラッシュメモリの低いアクセス性能で制限される．

2.5 次世代コンピュータアーキテクチャにおけるストレージの課題

第 2.4 節で説明した SCM，MLC NAND 型フラッシュメモリ，TLC NAND 型フラッシュメモリを用いたハイブリッドストレージは，上位の高速な不揮発性半導体メモリにアクセスの多いホットあるいはワームデータを保存し，下位の低速な不揮発性半導体メモリにアクセスの少ないコールドあるいはフローズンデータを保存する．表 2.3 に，第 2.4 節で述べたストレージの利点および問題点を示す[41]．SCM と MLC NAND 型フラッシュメモリを用いたハイブリッドストレージは，頻繁にアクセスされるデータを SCM へ書き込むことで高性能になる．NV-WB キャッシュアルゴリズムを用いる場合[38][39]，すべてのデータは初めに高速な

表 2.3 SCM および NAND 型フラッシュメモリを用いたストレージの利点と問題点 [41]

Storage architecture	Data management algorithm	Objective	Pros	Cons
SCM and MLC NAND flash	NV-WB cache [38, 39]	SCM accelerates performance as simple write-back cache	・Fast write latency than MLC NAND flash	・Cost increase by SCM ・Application dependent performance ・Complex data management algorithm between SCM and MLC/TLC NAND flash
	CDE [37]	Reduce cold or sequential data access to SCM	・Increase performance of write-intensive apps	
MLC and TLC NAND flash	RR-FDCA [40]	MLC works as write buffer of TLC NAND flash	・Lower cost by TLC NAND flash ・MLC NAND flash endurance improvement	・Limited performance improvement

SCM に書き込むため，ホストへの応答速度（レイテンシ）が速くなる．また CDE アルゴリズム[37]では，ホストからのホットあるいはランダムデータを選択して SCM に書き込むため，コールドかつシーケンシャルデータを SCM に書き込む必要がない．しかし CDE アルゴリズムは読み出し動作による不揮発性半導体メモリ間のデータ移動が生じないため，NV-WB キャッシュアルゴリズムと比較して書き込みの多いアプリケーションの性能を向上させる．これらの SCM と MLC NAND 型フラッシュメモリを用いたハイブリッドストレージは，ホット・ランダムなデータを SCM で処理するため MLC NAND 型フラッシュメモリの書き換え回数が削減し，MLC NAND 型フラッシュメモリを長寿命化することができる．一方 SCM を用いることによる問題は，SCM の高いビットコストである．MLC NAND 型フラッシュメモリ容量に対して M-SCM を 10%追加する場合，総ストレージコストは約 2 倍となる．また，SCM および MLC NAND 型フラッシュメモリ間での複雑なデータ管理が必要となる．MLC および TLC NAND 型フラッシュメモリを用いたハイブリッドストレージは，SCM を用いる場合と比較して低コストで実現できる．しかし SCM と比較して MLC および TLC NAND 型フラッシュメモリの性能は劣るため，ストレージ性能の向上は制限される．また，これらの複数種の不揮発性半導体メモリを用いたストレージは，ストレージアプリケーションの特性に大きく依存する．

本研究ではストレージアプリケーションとして，Microsoft Research Cambridge の 1 週間のブロック I/O トレースを用いる[45]．MSR ストレージアプリケーションは，プロキシデータベースサーバ，プロジェクト用ディレクトリなどから取得した 1 週間の読み出し・書き込みリクエストを含む．参考文献[39]では，読み出し・書き込み量の多寡，平均データアクセス頻度，平均データサイズを用いて，図 2.16 のようにストレージアプリケーションを 8 分類している．第一に，アプリケーションの読み出し・書き込みの多さは，読み出しデータ量および書き込みデータ量を用いて判断される．書き込みデータ量が読み出しデータ量より多い場合，

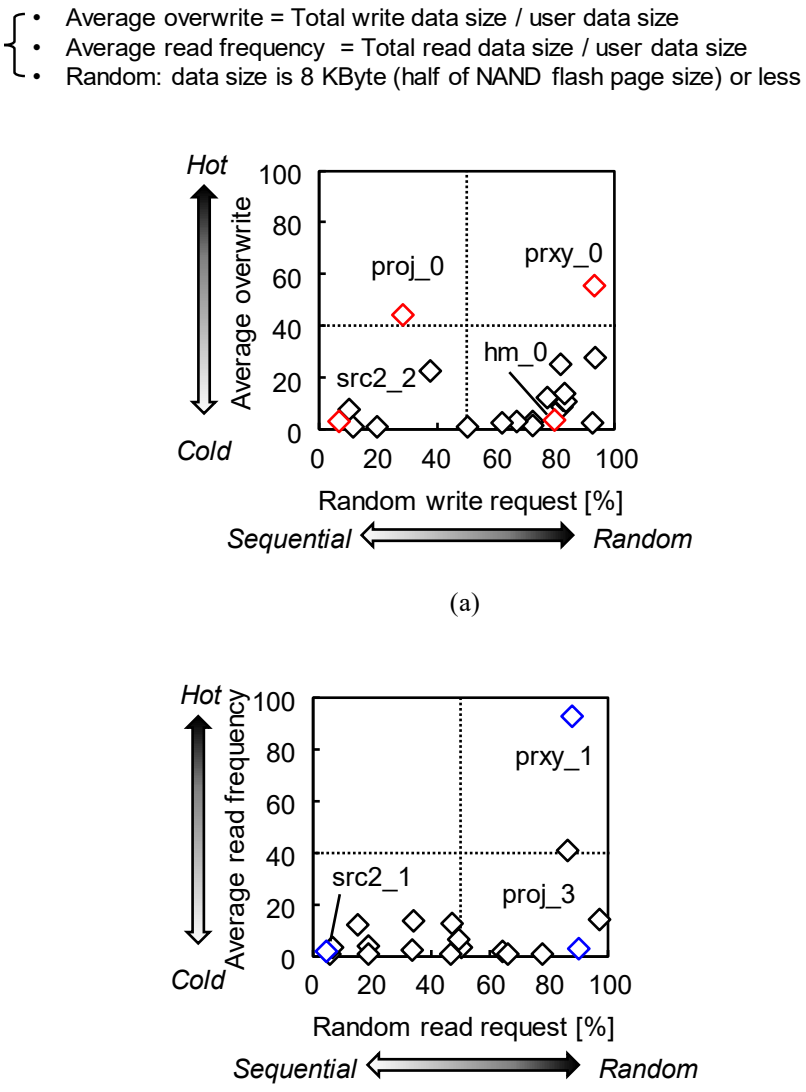
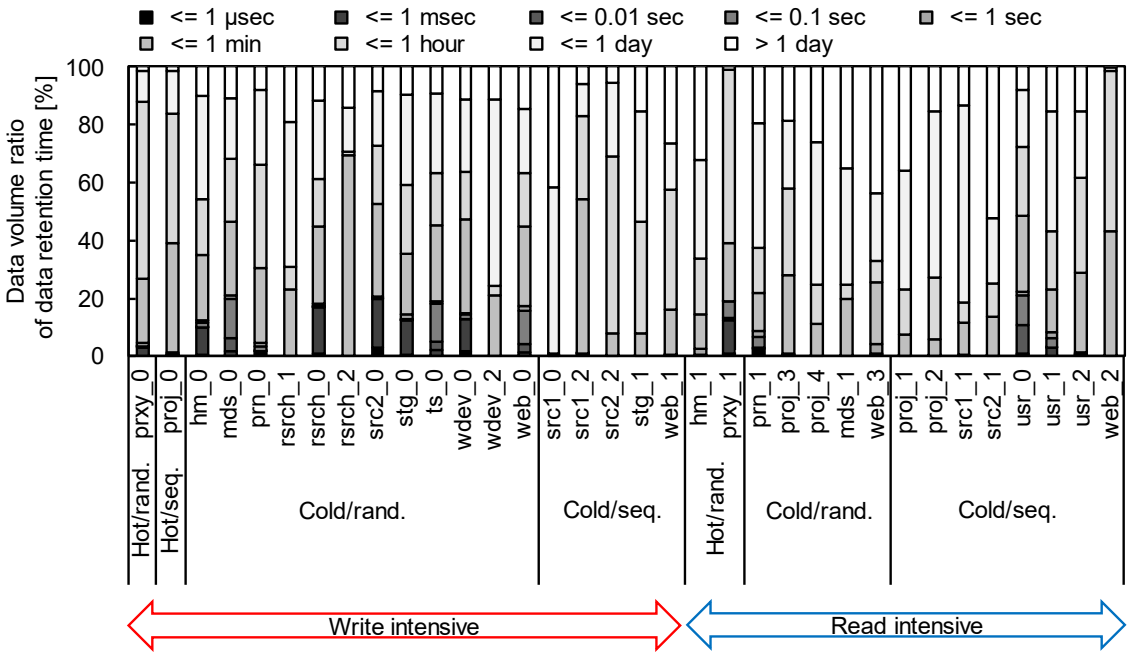
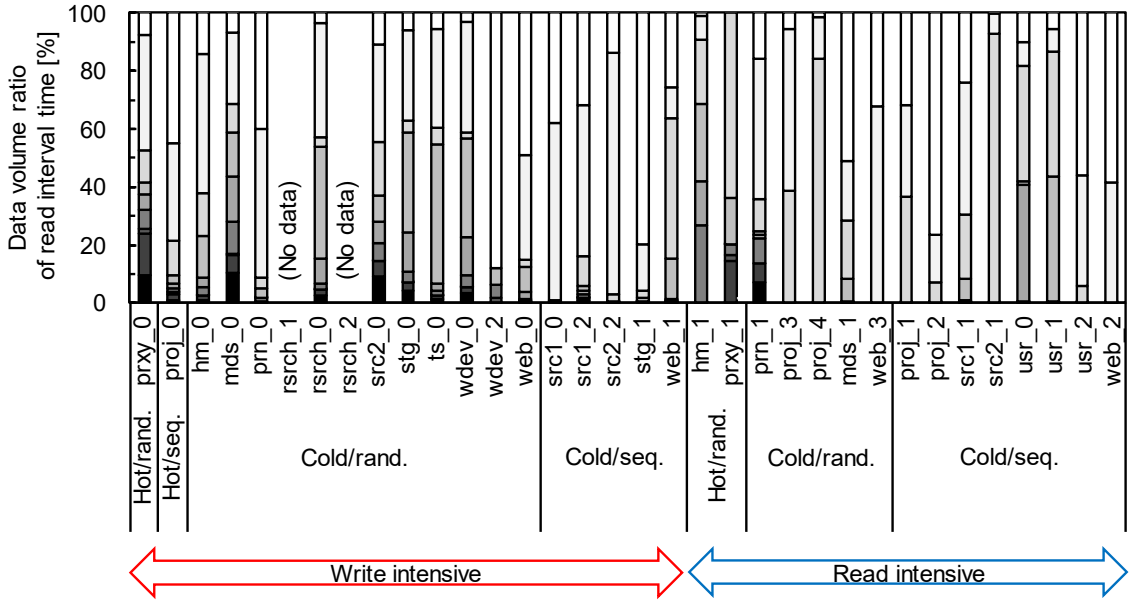


図 2.16 ストレージアプリケーション[45]の分類．（a）書き込み多いアプリケーション，（b）読み出しの多いアプリケーション [39]

そのストレージアプリケーションは書き込みの多い（write-intensive）アプリケーションと定義する[図 2.16（a）]．反対に読み出しデータ量が書き込みデータ量より多い場合，そのストレージアプリケーションは読み出しの多い（read-intensive）アプリケーションと定義する[図 2.16（b）]．第二にストレージアプリケーションの各リクエストは，ホストからの論理アドレス（logical block address，LBA）へのアクセス頻度によって，ホット（hot）またはコールド（cold）に分類する．アプリケーションの論理アドレスが平均して頻繁にアクセスされる場合をホットアプリケーションとし，そうでない場合をコールドアプリケーションとする．第三に，ストレージアプリケーションの平均アクセスサイズを用いて，ランダムあるいはシーケ



(a)



(b)

図 2.17 ストレージアプリケーション[45]のアクセス時間間隔の解析．（a）書き込み時間間隔（データリテンション時間），（b）読み出し時間間隔

ンシャルであると判断する．具体的には，アクセス（読み出しおよび書き込み）リクエストの平均サイズが NAND 型フラッシュメモリのページサイズの半分である 8 KByte と比較して，小さい場合をランダム（random）と定義する．反対に平均アクセスサイズが 8 KByte より大きい場合をシーケンシャル（sequential）とする．一例として，prxy_0 アプリケーションは，書き込みが多くホット・ランダムであると分類される．また，参考文献[46]で解析してい

るように，MSR ストレージアプリケーションによりデータアクセスの時間間隔が異なる．図 2.17 に本研究で用いるストレージアプリケーションのアクセス時間間隔を示す．先の図 2.16 に分類したホットなアプリケーションほどアクセス時間間隔が短いことがわかる．特に prxy_0 アプリケーションは 88%のデータが 1 時間以内に上書きされる．

したがって，このように複雑な特性を持つストレージアプリケーションに最適化したストレージを構築する必要がある．これを実現するために，本論文において異種の不揮発性半導体メモリを用いたヘテロジニアスストレージを提案する．ヘテロジニアスストレージの記憶には第 2.3 節で説明した新たな不揮発性半導体メモリである M-SCM，S-SCM，および現在ストレージシステムで用いられている MLC NAND 型フラッシュメモリおよび TLC NAND 型フラッシュメモリを用いる．これらの不揮発性半導体メモリはアクセス時間，書き換え耐久性（許容書き換え回数)，容量，ビットコストなどが異なる．そのため，それぞれの不揮発性半導体メモリの特性を最大限利用するためのデータマネジメントアルゴリズムやメモリ容量比を，ストレージアプリケーションに対して最適化する必要がある．また，各種 SCM および NAND 型フラッシュメモリは動作原理が異なるため，書き換えや読み出しに対するエラー発生頻度が異なる．エラー訂正符号（error-correcting code，ECC）を用いて不揮発性半導体メモリに発生したエラーを訂正するが，ストレージアクセス性能を低下させる．またデータマネジメントアルゴリズムを用いることによって，SCM と NAND 型フラッシュメモリのアクセス頻度も異なる. さらに SCM の容量が多くなるほど SCM へのアクセス頻度が高まり, NAND 型フラッシュメモリへのアクセス頻度は低減される．ECC によりストレージの信頼性を高める一方で，ストレージのアクセス性能を低下させないことが必要である．さらにストレージアプリケーションにより，ストレージを高速化するために必要な SCM 容量が異なる．しかしデータセンターで動作するさまざまなストレージアプリケーションに対し，手動で最適な SCM 容量を調整することは不可能である．また SCM は NAND 型フラッシュメモリと比較して高いビットコストを持つ．ストレージアプリケーションが必要な場合にのみ SCM を用いる，SCM 容量の自律調整が必要である．

2.6 まとめ

本章では不揮発性半導体メモリを用いたストレージの従来研究および問題点について述べた．初めに SCM の登場により次世代のコンピュータアーキテクチャが変わることを述べた．SCM はメモリシステムおよびストレージシステムで用いられる可能性があり，本論文では SCM と NAND 型フラッシュメモリを用いたストレージシステムの研究を行う．次に本研究

のストレージシステムの記憶として用いる SCM と NAND 型フラッシュメモリの動作を述べた．SCM はその特性によってメモリタイプおよびストレージタイプに分類でき，一方で NAND 型フラッシュメモリはセル当たりに保存するビット数によって SLC，MLC，TLC に分類できることを示した．これらの不揮発性半導体メモリはアクセス時間，書き換え耐久性などがそれぞれ異なることを述べた．従来研究として SCM 一種および NAND 型フラッシュメモリ一種を用いたハイブリッドストレージストレージの構成，および SCM を NAND 型フラッシュメモリの不揮発性キャッシュあるいは小容量ストレージとして用いるデータマネジメント手法について述べた．さらに不揮発性半導体メモリを用いた次世代のコンピュータアーキテクチャにおけるストレージの課題を論じた．ストレージアプリケーションの特性がさまざまに異なるためストレージアプリケーション内のデータの特徴によって，異種の不揮発性半導体メモリを用いたヘテロジニアスストレージの不揮発性半導体メモリの構成の最適化が必要となる．またヘテロジニアスストレージの M-SCM あるいは S-SCM と NAND 型フラッシュメモリとではアクセス頻度が異なり，不揮発性半導体メモリの種類によってエラー発生頻度や許容書き換え回数が異なるため，それぞれに異なる強度の ECC を適用することが必要となる．さらにストレージアプリケーションの特性によって最適な SCM 容量は異なり，データセンター事業者やユーザが手動でさまざまな種類のストレージアプリケーションに必要な SCM 容量を設定することは困難であるため，自動で SCM 容量を最適化する手法が必要であることを述べた．

参考文献

[1] 柴山 潔，“コンピュータアーキテクチャの基礎”初版第 9 刷，2010 年，近代科学社.

[2] H. Ando，“コンピュータアーキテクチャ技術入門”初版第 1 刷，2014 年，技術評論社.

[3] R. F. Freitas and W. W. Wilcke, “Storage-class memory: The next storage system technology,” *IBM Journal of Research and Development*, vol. 52, no. 4/5, pp. 439-447, Jul. 2008.

[4] G. W. Burr, B. N. Kurdi, J. C. Scott, C. H. Lam, K. Gopalakrishnan, and R. S. Shenoy, “Overview of candidate device technologies for storage-class memory,” *IBM Journal of Research and Development*, vol. 52, no. 4/5, pp. 449-464, Jul. 2008.

[5] K. Tsuchida, T. Inaba, K. Fujita, Y. Ueda, T. Shimizu, Y. Asao, T. Kajiyama, M. Iwayama, K. Sugiura, S. Ikegawa, T. Kishi, T. Kai, M. Amano, N. Shimomura, H. Yoda, and Y. Watanabe, “A 64Mb MRAM with clamped reference and adequate-reference schemes,” in I*EEE International Solid-State Circuits Conference (ISSCC) Digest of Technical Papers*, Feb. 2010, pp. 258-259.

[6] S.-W. Chung, T. Kishi, J. W. Park, M. Yoshikawa, K. S. Park, T. Nagase, K. Sunouchi, H. Kanaya,

G. C. Kim, K. Noma, M. S. Lee, A. Yamamoto, K. M. Rho, K. Tsuchida, S. J. Chung, J. Y. Li, H. S. Chun, H. Oyamatsu, and S. J. Hong, “4Gbit density STT-MRAM using perpendicular MTJ realized with compact cell structure,” in *IEEE International Electron Devices Meeting (IEDM) Technical Digest*, Dec. 2016, pp. 27.1.1-27.1.4.

[7] A. Kawahara, R. Azuma, Y. Ikeda, K. Kawai, Y. Katoh, K. Tanabe, T. Nakamura, Y. Sumimoto, N. Yamada, N. Nakai, S. Sakamoto, Y. Hayakawa, K. Tsuji, S. Yoneda, A. Himeno, K. Origasa, K. Shimakawa, T. Takagi, T. Mikawa, and K. Aono, “An 8Mb multi-layered cross-point ReRAM macro with 43 MB/s write throughput,” *IEEE Journal of Solid-State Circuits (JSSC)*, vol. 48, no. 1, pp. 178-185, Oct. 2013.

[8] T.-Y. Liu, T. H. Yan, R. Scheuerlein, Y. Chen, J. K. Lee, G. Balakrishnan, G. Yee, H. Zhang, A. Yap, J. Ouyang, T. Sasaki, A. Al-Shamma, C. Chen, M. Gupta, G. Hilton, A. Kathuria, V. Lai, M. Matsumoto, A. Nigam, A. Pai, J. Pakhale, C. H. Siau, X. Wu, Y. Yin, N. Nagel, Y. Tanaka, M. Higashitani, T. Minvielle, C. Gorla, T. Tsukamoto, T. Yamaguchi, M. Okajima, T. Okamura, S. Takase, H. Inoue, and L. Fasoli, “A 130.7-mm2, 2-layer 32Gb ReRAM memory device in 24-nm technology,” *IEEE Journal of Solid-State Circuits (JSSC)*, vol. 49, no. 1, pp. 140-153, Jan. 2014.

[9] K. Kawai, A. Kawahara, R. Yasuhara, S. Muraoka, Z. Wei, R. Azuma, K. Tanabe, and K. Shimakawa, “Highly-reliable TaOx ReRAM technology using automatic forming circuit,” in *Proceedings of IEEE International Conference on IC Design and Technology (ICICDT)*, May 2014, pp. 100-103.

[10] K.-J. Lee, B.-H. Cho, W.-Y. Cho, S. Kang, B.-G. Choi, H.-R. Oh, C.-S. Lee, H.-J. Kim, J.-M. Park, Q. Wang, M.-H. Park, Y.-H. Ro, J.-Y. Choi, K.-S. Kim, Y.-R. Kim, I.-C. Shin, K.-W. Lim, H.-K. Cho, C.-H. Choi, W.-R. Chung, D.-E. Kim, Y.-J. Yoon, K.-S. Yu, G.-T. Jeong, H.-S. Jeong, C.-K. Kwak, C.-H. Kim, K. Kim, “A 90nm 1.8 V 512 Mb diode-switch PRAM with 266 MB/s read throughput,” *IEEE Journal of Solid-State Circuits (JSSC)*, vol. 43, no. 1, pp. 150-162, Jan. 2008.

[11] Y. Choi, I. Song, M. Park, H. Chung, S. Chang, B. Cho, J. Kim, Y. Oh, D. Kwon, J. Sunwoo, J. Shin, Y. Rho, C. Lee, M. G. Kang, J. Lee, Y. Kwon, S. Kim, J. Kim, Y. Lee, Q. Wang, S. Cha, S. Ahn, H. Horii, J. Lee, K. Kim, H. Joo, K. Lee, Y. Lee, J. Yoo, and G. Jeong, “A 20nm 1.8V 8Gb PRAM with 40MB/s program bandwidth,” in *IEEE International Solid-State Circuits Conference (ISSCC) Diegest of Technical Papers*, Feb. 2012, pp. 46-47.

[12] Micron 3D XPoint Technology, https://www.micron.com/about/emerging-technologies/3d-xpoint-technology.

[13] T. Hirofuchi and R. Takano, “RAMinate: Hypervisor-based virtualization for hybrid main memory

systems," in *Proceedings of ACM Symposium on Cloud Computing (SoCC)*, Oct. 2016, pp. 112-125.

[14] F. Masuoka, M. Momodomi, Y. Iwata, and R. Shirota, "New ultra high density EPROM and flash EEPROM with NAND structure cell," in *IEEE International Electron Devices Meeting (IEDM) Technical Digest*, Dec. 1987, pp. 552-555.

[15] R. H. Fowler and L. Nordheim, "Electron emission in intense electric fields," *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, vol. 119, no. 781, pp. 173-181, May 1928.

[16] K. Takeuchi, "Novel co-design of NAND flash memory and NAND flash controller circuits for sub-30 nm low-power high-speed solid-state drives (SSD)," *IEEE Journal of Solid-State Circuits (JSSC)*, vol. 44, no. 4, pp. 1227-1234, Apr. 2009.

[17] P. Pavan, R. Bez, P. Olivo, and E. Zanoni, "Flash memory cells - an overview," *Proceedings of the IEEE*, vol. 85, no. 8. pp. 1248-1271, Aug. 1997.

[18] R. Bez, E. Camerlinghi, A. Modelli, and A. Visconti, "Introduction to flash memory," *Proceedings of the IEEE*, vol. 91, no. 4, pp. 489-502, Apr. 2003.

[19] H. Niijima, "Design of a solid-state file using flash EEPROM," *IBM Journal of Research and Development*, vol. 39, no. 5, pp. 531-545, Sep. 1995.

[20] K. Takeuchi, Y. Kameda, S. Fujimura, H. Otake, K. Hosono, H. Shiga, Y. Watanabe, T. Futatsuyama, Y. Shindo, M. Kojima, M. Iwai, M. Shirakawa, M. Ichige, K. Hatakeyama, S. Tanaka, T. Kamei, J.-Y. Fu, A. Cernea, Y. Li, M. Higashitani, G. Hemink, S. Sato, K. Oowada, S.-C. Lee, N. Hayashida, J. Wan, J. Lutze, S. Tsao, M. Mofidi, K, Sakurai, N. Tokiwa, H. Waki, Y. Nozawa, K. Kanazawa, and S. Ohshima, "A 56-nm CMOS 99-mm2 8-Gb multi-level NAND flash memory with 10-MB/s program throughput," *IEEE Journal of Solid-State Circuits (JSSC)*, vol. 42, no. 1, pp. 219-232, Jan. 2007.

[21] H. Fujii, K. Miyaji, K. Johguchi, K. Higuchi, C. Sun, and K. Takeuchi., "x11 performance increase, x6.9 endurance enhancement, 93% energy reduction of 3D TSV-integrated hybrid ReRAM/MLC NAND SSDs by data fragmentation suppression," in *IEEE Symposium on VLSI Circuits Digest of Technical Papers*, Jun. 2012, pp.134-135.

[22] X.-Y. Hu, E. Eleftheriou, R. Haas, I. Iliadis, and R. Pletka, "Write amplification analysis in flash-based solid state drives," in *Proceedings of ACM International Systems and Storage Conference (SYSTOR)*, May 2009, pp. 191-202.

[23] Y. Koh, "NAND flash scaling beyond 20nm," in *Proceedings of IEEE International Memory Workshop (IMW)*, May 2009, pp. 3-5.

[24] M. Bauer, R. Alexis, G. Atwood, B. Baltar, A. Fazio, K. Frary, M. Hensel, M. Ishac, J. Javanifard, M. Landgraf, D. Leak, K. Loe, D. Mills, P. Ruby, R. Rozman, S. Sweha, S. Talreja, and K. Wojciechowski, "A multilevel-cell 32Mb flash memory," in *IEEE International Solid-State Circuits Conference (ISSCC) Digest of Technical Papers*, Feb. 1995, pp. 132-133.

[25] K. Takeuchi, T. Tanaka, and T. Tanzawa, "A multi-level cell architecture for high-speed programming multi-level NAND flash memories," in *IEEE Symposium on VLSI Circuits Digest of Technical Papers*, Jun. 1997, pp. 67-68.

[26] Y. Fukuzumi, R. Katsumata, M. Kito, M. Kido, M. Sato, H. Tanaka, Y. Nagata, Y. Matsuoka, Y. Iwata, H. Aochi, and A. Nitayama, "Optimal integration and characteristics of vertical array devices for ultra-high density, bit-cost scalable flash memory," in *IEEE International Electron Devices Meeting (IEDM) Technical Digest*, Dec. 2007, pp. 449-452.

[27] S. Oshima and S. Fingerhut, "Flash memory is going places we have never been before," in *Flash Memory Summit*, Aug. 2017.

[28] C. Matsui, Y. Yamaga, Y. Sugiyama, and K. Takeuchi, "8.9-times performance improvement by tri-hybrid storage system with SCM and MLC/TLC NAND flash memory," in *Extended Abstracts of International Conference on Solid State Devices and Materials (SSDM)*, Sep. 2016, pp. 105-106.

[29] A. Faryushin, K. Seol, J. Na, S. Hur, J. Choi, and K. Kim, "The new program/erase cycling degradation mechanism of NAND flash memory devices," in *IEEE International Electron Devices Meeting (IEDM) Technical Digest*, Dec. 2009, pp. 823-826.

[30] IBM Almaden Research Center, "Storage class memory: Towards a disruptively low-cost solid-state non-volatile memory," http://researcher.watson.ibm.com/researcher/files/us-gwburr/Almaden_SCM_overview_Jan2013.pdf, Jan. 2013.

[31] C. Matsui, T. Yamada, Y. Sugiyama, Y. Yamaga, and K. Takeuchi, "Optimal memory configuration analysis in tri-hybrid solid-state drives with storage class memory and multi-level cell/triple-level cell NAND flash memory," *Japanese Journal of Applied Physics (JJAP)*, vol. 56, no. 4S, pp. 04CE02-1 - 04CE02-9, Apr. 2017.

[32] DRAMexchange, http://www.dramexchange.com.

[33] X. Wu, J. Li, L. Zhang, E. Speight, and Y. Xie, "Power and performance of read-write aware hybrid cache with non-volatile memories," in *Proceedings of Design, Automation and Test in Europe Conference and Exhibition (DATE)*, Apr. 2009, pp. 737-742.

[34] K. Abe, H. Noguchi, E. Kitagawa, N. Shimomura, J. Ito, and S. Fujita, "Novel hybrid DRAM/MRAM design for reducing power of high performance mobile CPU," in *IEEE*

International Electron Devices Meeting (IEDM) Technical Digest, Dec. 2012, pp. 10.5.1-10.5.4.

[35] G. Dhiman, R. Ayoub, and T. Rosing, “PDRAM: A hybrid PRAM and DRAM main memory system,” in *Proceedings of Design Automation Conference (DAC)*, Jul. 2009, pp. 664-669.

[36] Direct Access for files, https://www.kernel.org/doc/Documentation/filesystems/dax.txt.

[37] G. Sun, Y. Joo, Y. Chen, D. Niu, Y. Xie, Y. Chen, and H. Li, “A hybrid solid-state storage architecture for the performance, energy consumption, and lifetime improvement,” in *Proceedings of International Symposium on High-Performance Computer Architecture (HPCA)*, Jan. 2010, pp. 141-152.

[38] C. Sun, K. Miyaji, K. Johguchi, and K. Takeuchi, “A high performance and energy-efficient cold data eviction algorithm for 3D-TSV hybrid ReRAM/MLC NAND SSD,” *IEEE Transactions on Circuits and Systems-I (TCAS-I)*, vol. 61, no. 2, pp. 382-392, Feb. 2014.

[39] S. Okamoto, C. Sun, S. Hachiya, T. Yamada, Y. Saito, T. O. Iwasaki, and K. Takeuchi, “Application driven SCM and NAND flash hybrid SSD design for data-centric computation system,” in *Proceedings of IEEE International Memory Workshop (IMW)*, May 2015, pp. 157-160.

[40] T. Yamada, C. Matsui, and K. Takeuchi, “Optimal combinations of SCM characteristics and non-volatile cache algorithm for high-performance SCM/NAND flash hybrid SSD,” in *Proceedings of IEEE Silicon Nanoelectronics Workshop (SNW)*, Jun. 2016, pp. 88-89, poster presentation.

[41] C. Matsui, C. Sun, and K. Takeuchi, “Design of hybrid SSDs with storage class memory and NAND flash memory,” *Proceedings of the IEEE*, vol. 105, no. 9, pp. 1812-1821, Sep. 2017.

[42] S. Hachiya, K. Johguchi, K. Miyaji, and K. Takeuchi, “TLC/MLC NAND flash mix-and-match design with exchangeable storage array,” in *Extended Abstracts of International Conference on Solid State Devices and Materials (SSDM)*, Sep. 2013, pp. 894-895.

[43] S.-H. Shin, D.-K. Shim, J.-Y. Jeong, O.-S. Kwon, S.-Y. Yoon, M.-H. Choi, T.-Y. Kim, H.-W. Park, H.-J. Yoon, Y.-S. Song, Y.-H. Choi, S.-W. Shim, Y.-L. Ahn, K-T. Park, J.-M. Han, K.-H. Kyung, and Y.-H. Jun, “A new 3-bit programming algorithm using SLC-to-TLC migration for 8MB/s high performance TLC NAND flash memory,” in *IEEE Symposium on VLSI Circuits Digest of Technical Papers*, Jun. 2012, pp. 132-133.

[44] I. J. Chang and J.-S. Yang, “Bit-error rate improvement of TLC NAND flash using state re-ordering,” *IEICE Electronics Express (ELEX)*, vol. 9, no. 34, pp. 1775-1779, Dec. 2012.

[45] MSR Cambridge Traces, http://iotta.snia.org/traces/388.

[46] 新屋敷 裕太，飯澤 健，小沢 年弘，荒堀 喜貴，横田 治夫，“アクセス時間に基づいた字アクセス予想によるデバイスミックスストレージシステムの制御手法”，第 8 回データ工

学と情報マネジメントに関するフォーラム（DEIM Forum），2016，pp. D8-2-1 - D8-2-6.

第3章 異種の不揮発性メモリを用いたストレージ構成およびデータ管理アルゴリズム

3.1 はじめに

本章では複雑な特性を持つストレージアプリケーションを処理するため，異種の不揮発性半導体メモリを用いたヘテロジニアスストレージを提案する．第 2 章で述べた従来研究と異なり，三種以上の不揮発性半導体メモリを用いて構成したストレージをヘテロジニアスストレージと呼ぶ．ヘテロジニアスストレージの記憶として M-SCM，S-SCM，MLC NAND 型フラッシュメモリ，TLC NAND 型フラッシュメモリを用い，1）SCM，MLC および TLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージ，および 2）M-SCM，S-SCM および NAND 型フラッシュメモリを用いたヘテロジニアスストレージを提案する．提案する二種類のヘテロジニアスストレージは，不揮発性半導体メモリ特性の長所を用いる一方で，それぞれの不揮発性半導体メモリの短所を隠ぺいする．

3.2 不揮発性メモリの読み出し，書き込み時間

本研究において M-SCM および S-SCM を用いる理由はそれらのアクセス性能の高さであり，TLC NAND 型フラッシュメモリを用いる理由はその低いビットコストである．第 2 章で述べたように，TLC NAND 型フラッシュメモリは MLC NAND 型フラッシュメモリと比較して読み出し・書き込み時間が長いが，そのビットコストは約 2/3 である．従来の SCM と MLC NAND 型フラッシュメモリを用いたハイブリッドストレージでは，SCM によるストレージコストの増加が問題であった．表 2.1 および表 2.2 で定めたように，M-SCM，S-SCM（scenario 2），MLC NAND 型フラッシュメモリ，TLC NAND 型フラッシュメモリのビットコスト比は，10 : 6 : 1 : 2/3 である．M-SCM，S-SCM，MLC および TLC NAND 型フラッシュメモリを用いる場合の総ストレージコストの計算式を式（3.1）に示す．

$$\text{Total storage cost} = \sum_{\text{M-SCM,S-SCM,MLC flash,TLC flash}} (\text{memory capacity ratio} \times \text{bit cost ratio}) \tag{3.1}$$

本論文において，総ストレージコストはヘテロジニアスストレージの不揮発性半導体メモリの容量とそれらのビットコストで求められるとする．式（3.1）から SCM によるストレージコストの増加は，低いビットコストの TLC NAND 型フラッシュメモリを用いることで相殺できることがわかる．

SCM は NAND 型フラッシュメモリと同様に不揮発性半導体メモリであるが，第 2 章で述べたようにこれらの特性はさまざまな点で異なる．SCM はセクタ（あるいはブロック，512 Byte）単位で読み出し・書き込み動作が可能であるため，SCM に保存された古いデータを上書きすることが可能である．一方で MLC NAND 型フラッシュメモリはページ（32 セクタ，16 KByte）単位で読み出し・書き込みを行い，ブロック（256 page，4.0 MByte）単位で消去する．MLC NAND 型フラッシュメモリは 128 ワードラインを持ち，1 ワードライン当たり Upper page，Lower page を持つので，1 ブロック当たりのページ数は 256 であると仮定する．また TLC NAND 型フラッシュメモリはページ（16 KByte）単位で読み出し，ブロック（258 page，4.03 MByte）単位で書き込み・消去を行う．TLC NAND 型フラッシュメモリは 86 ワードラインを持ち，1 ワードライン当たり Upper page，Middle page，Lower page を持つため，1 ブロック当たりのページ数は 258 であると仮定する．このため MLC NAND 型フラッシュメモリの 1 ページ内のデータを上書きするとき，上書きするデータを有するページを読み出し，上書きするデータとともにコントローラで統合して新しいページに書き込む必要がある．さらに第 2.4.2 節で述べたように TLC NAND 型フラッシュメモリはセル間干渉を防ぐためブロック書き込みをする必要がある．TLC NAND 型フラッシュメモリの 1 ページ内のデータを上書きするとき，上書きするデータを有するページおよびそのページが存在するブロックの他の有効ページも読み出し，別の新しいブロックに書き込み必要がある．悪い条件として 1 セクタのみを上書きすることを考えると，表 2.1 より M-SCM の場合は 0.1 μsec，S-SCM（scenario 2）の場合は 1 μsec だけ時間を要する．一方で MLC および TLC NAND 型フラッシュメモリの読み出し・書き込み時間として表 2.2 に示す Upper/Middle/Lower page から求めた平均読み出し・書き込み時間を用いると，MLC NAND 型フラッシュメモリの 1 セクタを上書きするのに要する時間は

$$
\begin{aligned}
&1\ \text{sector overwrite time of MLC NAND flash} \\
&= 1\ \text{page} \times (\text{average page read time} + \text{average page write time}) \\
&= 1\ \text{page} \times (44\ \mu sec + 1185\ \mu sec) \\
&= 1.3\ msec
\end{aligned}
\tag{3.2}
$$

となり，TLC NAND 型フラッシュメモリの 1 セクタを上書きするのに要する時間は

$$\begin{aligned} &1\ \text{sector overwrite time of TLC NAND flash} \\ &= \#\ \text{of page in block} \times (\text{average page read time} + \text{average page write time}) \\ &= 258\ \text{page} \times (87\ \mu sec + 2180\ \mu sec) \\ &= 585\ msec \end{aligned} \tag{3.3}$$

と計算することができる．さらに悪いことに，NAND 型フラッシュメモリは無効ページを消去し空ページを確保するガベージコレクション（garbage collection，GC）動作を要する．例としてブロック内の有効ページ数が 100 のとき，MLC および TLC NAND 型フラッシュメモリの GC に必要な時間は次の式（3.4）および式（3.5）のように計算できる．

$$\begin{aligned} &\text{GC time of MLC NAND flash} \\ &= \#\ \text{of valid pages} \times (\text{average page read time} + \text{average page write time}) \\ &\qquad + \text{block erase time} \\ &= 100\ \text{page} \times (44\ \mu sec + 1185\ \mu sec) + 3300\ \mu sec \\ &= 126\ msec \end{aligned} \tag{3.4}$$

$$\begin{aligned} &\text{GC time of TLC NAND flash} \\ &= \#\ \text{of valid pages} \times (\text{average page read time} + \text{average page write time}) \\ &\qquad + \text{block erase time} \\ &= 100\ \text{page} \times (87\ \mu sec + 2180\ \mu sec) + 3200\ \mu sec \\ &= 230\ msec \end{aligned} \tag{3.5}$$

実際の TLC NAND 型フラッシュメモリにおいては，SRAM キャッシュや TLC NAND 型フラッシュメモリを SLC モードで使うなどして，ブロック書き込みに要する時間を削減する努力をしている[1]. 一方で 1 セクタを読み出すとき, 表 2.1 より M-SCM は 0.1 μsec, S-SCM（scenario 2）は 1.0 μsec を要する．また NAND 型フラッシュメモリ内の 1 セクタを読み出すときそのセクタを含む 1 ページ全体を読み出すことが必要となり，MLC および TLC NAND 型フラッシュメモリでそれぞれ平均して 44 μsec，87 μsec を要する．SCM と NAND 型フラッシュメモリの書き込み時間の差と比較して，読み出し時間の差は十分小さいことがわかる．ただし本論文では，TLC NAND 型フラッシュメモリの書き込み単位はページであると仮定し第 4 章の評価を行なった.

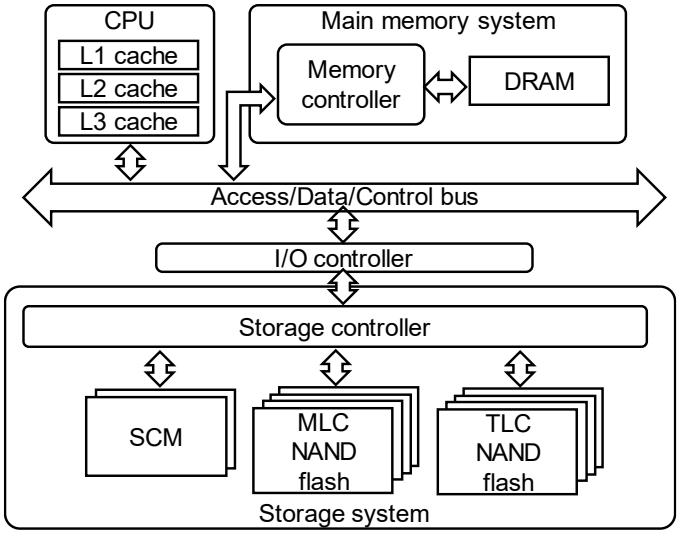


図 3.1 SCM，MLC および TLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージ [2]

3.3 SCM，MLC および TLC NAND 型フラッシュメモリを用いたストレージ

図 3.1 に SCM，MLC および TLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージのアーキテクチャを示す[2]．SCM は用途に応じて M-SCM あるいは S-SCM のいずれかを用い，さらに MLC および TLC NAND 型フラッシュメモリを用いる．第 3.2 節で議論したように SCM と MLC および TLC NAND 型フラッシュメモリは書き込み特性が特に異なる．また SCM（M-SCM あるいは S-SCM）はそのビットコストの高さから，データセンターでの実用上，MLC および TLC NAND 型フラッシュメモリと比較すると小容量しか用いることができないと考えられる．そのため本研究で提案する SCM，MLC および TLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージでは書き込み動作に適したデータマネジメントアルゴリズムを用いる. 第 2.4.1 項で述べたコールドデータエビクション（Cold Data Eviction, CDE）[3]およびラウンドロビン・フローズンデータコレクションアルゴリズム（Round-Robin Frozen Data Collection Algorithm，RR-FDCA）[4]はそれぞれ，不揮発性半導体メモリの特性を考慮した書き込み動作に適したデータマネジメントアルゴリズムである．CDE は SCM および MLC NAND 型フラッシュメモリを用いたハイブリッドストレージにおいて，書き込みデータの特性から SCM あるいは MLC NAND 型フラッシュメモリに書き込むことを判断する．また RR-FDCA は MLC および TLC NAND 型フラッシュメモリを用いたハイブリッドストレージにおいて，MLC NAND 型フラッシュメモリの上書きされないデータを TLC NAND 型フ

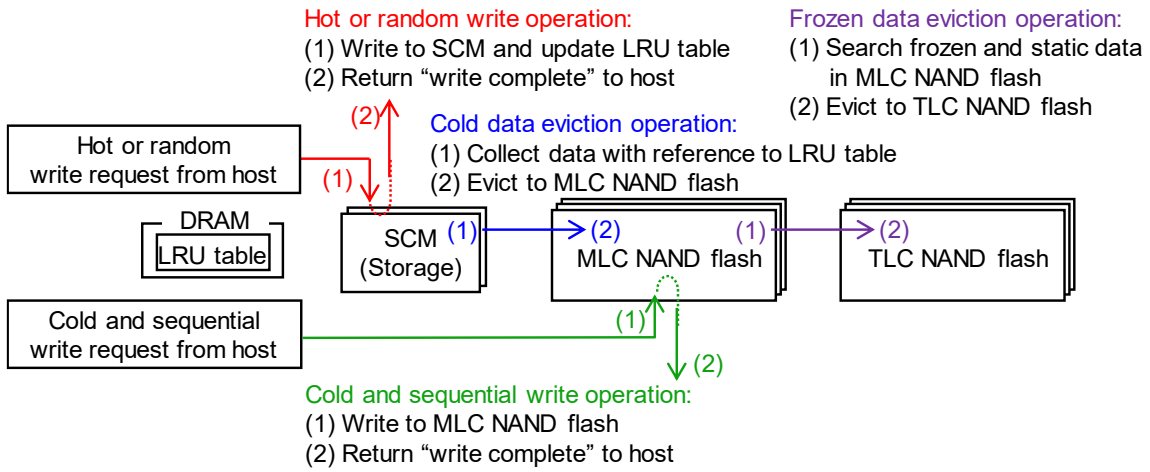


図 3.2 SCM，MLC および TLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージに適用するコールドアンドフローズンデータエビクション（Cold and Frozen Data Eviction，CFDE）アルゴリズム [2]

ラッシュメモリに書き移す．この 2 つのデータマネジメントアルゴリズムを組み合わせることで，SCM には頻繁にアクセスされるホットデータ（hot data），MLC NAND 型フラッシュメモリにはアクセス頻度の低いコールドデータ（cold data），さらに TLC NAND 型フラッシュメモリにはほとんどアクセスの無いフローズンデータ（frozen data）をそれぞれ保存する．SCM はストレージシステムの性能を向上させ，TLC NAND 型フラッシュメモリは SCM によるコスト上昇を均衡させストレージ容量を増やす．図 1.1 のメモリおよびストレージ階層と同様に，MLC NAND 型フラッシュメモリは SCM と TLC NAND 型フラッシュメモリとの間の中間的なストレージとして機能する．

以下に SCM，MLC および TLC NAND 型フラッシュメモリを用いたヘテロジアスストレージに適用するデータマネジメントアルゴリズムの詳細を述べる．このアルゴリズムをコールドアンドフローズンデータエビクション（Cold and Frozen Data Eviction，CFDE）と呼び，動作を図 3.2 に示す．提案の SCM，MLC および TLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージでは，SCM と MLC NAND 型フラッシュメモリ間は CDE アルゴリズム，MLC および TLC NAND 型フラッシュメモリ間は RR-FDCA を用いる．まずホストから書き込みリクエストを受け取ると，CDE アルゴリズムに従ってデータを書き込む．CDE アルゴリズムでは固定長の least recently used（LRU）テーブルを用いて SCM 内のデータの書き込み順を管理する．本研究では LRU リストのサイズは従来研究の結果[3]より SCM 容量の 80%とした．LRU リストは SCM に保存された論理ページアドレス（logical page address，LPA）の

順番を記録する．LPA は式（3.6）のように，論理ブロックアドレス（logical block address, LBA）から求められる．

$$\mathrm{LPA} = \frac{\mathrm{LBA}}{\mathrm{NAND\ \ flash\ \ page\ \ size}\quad (=16\ \ KByte)} \tag{3.6}$$

CDE アルゴリズムでは LRU リストに LPA のアクセス順序を記録するほか，頻繁にアクセスされるホットデータを判断するの役割も持つ．

図 2.14（b）に示したように，ホストからの新しい書き込みデータは初めにデータの大きさを判断する．書き込みリクエストのデータサイズが NAND 型フラッシュメモリのページサイズの半分である 8 KByte と比較して小さい場合ランダム（random）と判断する．反対に書き込みリクエストのデータサイズが 8 KByte より大きい場合をシーケンシャル（sequential）とする．シーケンシャルと判断されたデータで LRU リストにあるデータはホットと判断される．したがって，ホットあるいはランダムなデータは SCM へ（図 3.2 Hot or random write operation），コールドかつシーケンシャルなデータは MLC NAND 型フラッシュメモリへ（図 3.2 Cold and sequential write operation）書き込む．SCM へデータを書き込むべきデータを判断した後，SCM 内の容量がすでにデータでいっぱいか否かを判断する．本研究では従来研究[3]と同様に，SCM 容量のうち空の容量が 20%以上あれば新しいデータを書き込むのに十分であるとした．SCM 内の空の容量が 20%以上あるとき，ホストからのホットあるいはランダムなデータは SCM へすぐに書き込む．しかし SCM 内の空の容量が 20%未満の場合，新しいデータを書き込むのに不十分であるとして，SCM から MLC NAND 型フラッシュメモリへ頻繁にアクセスされないコールドデータを evict する必要がある．まず LRU リストを用いて SCM 内のデータがコールドであるか判断する．LRU リストに存在しないコールドデータは続いて，物理ページ内で断片化（fragmentation）しているかを判断する．MLC NAND 型フラッシュメモリの物理ページサイズ（16 KByte）の 60%より少ないデータがある場合，そのページは断片化しているという[3]．参考文献[3]と同様に SCM 容量およびストレージアプリケーション特性に依らず，SCM 内で 4,000 ページ分のコールドかつシーケンシャルデータが集まると，それらのデータを SCM から MLC NAND 型フラッシュメモリへ evict する（図 3.2 Cold data eviction operation）．また，MLC NAND 型フラッシュメモリに保存されたデータを上書きする場合がある．この場合もまた LRU リストを用いて上書きする古いデータがホットでさらに断片化していれば，NAND 型フラッシュメモリの同一ページに存在する有効なデータと一緒に SCM へ書き込む．一方でホストからの読み出しリクエストを受け取った時，データが保存された SCM あるいは MLC NAND 型フラッシュメモリから読み出す．つまり頻繁に読み出しリ

クエストがあるデータでも MLC NAND 型フラッシュメモリから読み出し，SCM へのデータコピーやデータの移動は行わない．そのため，CDE アルゴリズムは書き込みリクエストに適したアルゴリズムであると言える．

次に MLC NAND 型フラッシュメモリから TLC NAND 型フラッシュメモリへの eviction について述べる（図 3.2 Frozen data eviction operation）．MLC NAND 型フラッシュメモリに多くのデータが書き込まれガベージコレクションが必要になると，RR-FDCA [4]に従って MLC NAND 型フラッシュメモリから TLC NAND 型フラッシュメモリへデータを evict する．MLC NAND 型フラッシュメモリの GC 動作で消去するブロックは，MLC NAND 型フラッシュメモリで初めに書き込まれた最も古いブロックをラウンドロビン方式で選択する．ラウンドロビン方式の GC は言い換えると，NAND 型フラッシュメモリで最も消去回数の少ないブロックを選択することである．NAND 型フラッシュメモリのブロックのウエアレベリングを達成するためにラウンドロビン方式の GC（round-robin GC，RR-GC）を適用した．しかし RR-GC に NAND 型フラッシュメモリの性能を劣化させる問題があることを第 6.3 節で議論する．RR-FDCA では MLC NAND 型フラッシュメモリの GC 対象となるブロックを選択した後，そのブロックが上書きの少ないフローズンであるかを判断する．第 2.3 節で述べたように，NAND 型フラッシュメモリのページに上書きを行うと，古いデータを含むページを読み出し上書きしたいデータと統合して別のページに書き込む．そして上書きされたページは無効ページとして管理される．つまり RR-GC を用いる場合 GC 動作が NAND 型フラッシュメモリ内のブロックを一周する間に，上書きされたページは無効ページとなり，上書きされなかったページは有効ページのまま存在する．このように NAND 型フラッシュメモリではページの有効・無効によって，上書きされたか否かが判断できる．MLC NAND 型フラッシュメモリ内の有効ページ数がブロックサイズの 75%以上[5]ある場合，そのブロックはホットであるとして MLC NAND 型フラッシュメモリ内で GC を行う．つぎに図 2.15（b）に示した RR-GC のコールドデータスクリーニングが 2 回以上行われた MLC NAND 型フラッシュメモリのブロックを選択することで，ほとんど上書きされないフローズンなデータを含む MLC NAND 型フラッシュメモリのブロックを GC 対象とする．続いて GC 対象となったフローズンデータを含む MLC NAND 型フラッシュメモリのブロックの有効ページを TLC NAND 型フラッシュメモリへ書き込む．最後に有効ページを TLC NAND 型フラッシュメモリへ移動した後の MLC NAND 型フラッシュメモリのブロックを消去する．なお，NAND 型フラッシュメモリのページ読み出しを行っても，ページの有効・無効は変化しない．さらに，ホストからの読み出しリクエストを受け取った時，データが保存された MLC あるいは TLC NAND 型フラッシュメモリから

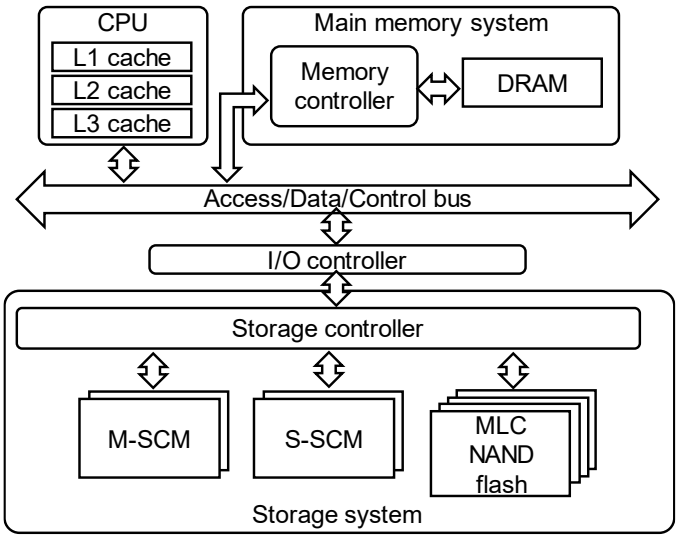


図 3.3 M-SCM，S-SCM および MLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージ [6]

読み出す．つまり頻繁に読み出しリクエストがあるデータでも TLC NAND 型フラッシュメモリから読み出し，MLC NAND 型フラッシュメモリへのデータコピーやデータの移動は行わない．そのため，RR-FDCA もまた書き込みリクエストに適したアルゴリズムであると言える．

3.4 M-SCM，S-SCM および NAND 型フラッシュメモリを用いたストレージ

図 3.3 に M-SCM，S-SCM および MLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージのアーキテクチャを示す[6]．第 3.2 節で議論したように，M-SCM のビットコストは MLC NAND 型フラッシュメモリと比較して約 10 倍であるため，データセンターでの実用上小容量しか用いることができないと考えられる．また表 2.2 に示したように，M-SCM と S-SCM（scenario 2）の読み出し・書き込み時間は約 10 倍の差である．しかし，S-SCM と MLC NAND 型フラッシュメモリの読み出し時間の差は平均で 44 倍，書き込み時間の差は平均で約 1200 倍と大きな差がある．さらにデータ上書きを行う場合は書き込み時間の差は広がる．そのため M-SCM，S-SCM および MLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージでは，M-SCM および S-SCM へのアクセスが多く MLC NAND 型フラッシュメモリへのアクセスが少ないアルゴリズムが必要である．第 2.4 節で議論した不揮発性半導体メモリ向けライトバック（NV-WB）キャッシュを発展させた，二種類の SCM を用いたライトバック（2 Non-Volatile Memory Write-Back，2NV-WB）キャッシュデータマネジメントアルゴリズムを提案する．書き込みの多いアプリケーションに対して性能向上する CDE を基にした

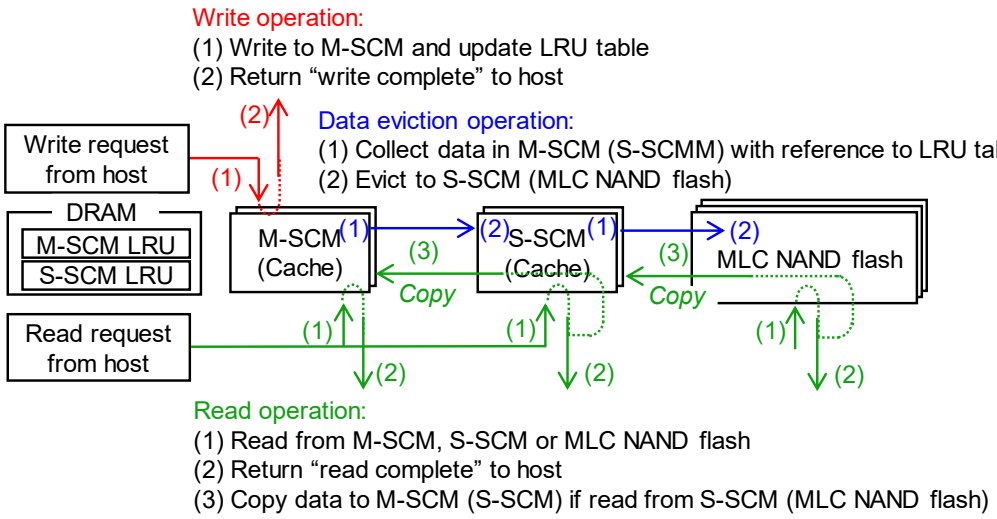


図 3.4 二種類の SCM および MLC NAND 型フラッシュメモリを用いたライトバック（2 Non-Volatile Memory Write-Back，2NV-WB）キャッシュデータマネジメントアルゴリズム [6]

アルゴリズムと異なり，書き込みおよび読み出しの多いアプリケーションに対して性能向上を行うことができると考える．SCM を二種類用いて極端にアクセス頻度の高いデータを M-SCM に，ややアクセス頻度の高いデータを S-SCM に保存することを目的とする．

図 3.4 に二種類の SCM および MLC NAND 型フラッシュメモリを用いたライトバック（2NV-WB）キャッシュデータマネジメントアルゴリズム[6]を示す．一種類の SCM を用いる NV-WB [7]と同様に，M-SCM および S-SCM の不揮発性のため 2NV-WB キャッシュは定期的なデータフラッシュ動作が不要で，突然の電源障害に対して安全である．初めに 2NV-WB キャッシュアルゴリズムの書き込み動作について述べる．すべてのデータは初めに，S-SCM の不揮発性キャッシュメモリとしての M-SCM に書き込まれる（図 3.4 Write operation）．M-SCM 内に書き込まれた順番を管理するために，M-SCM LRU リストを用いる．CDE アルゴリズムで SCM 内のデータアクセス順序を管理しデータのホットあるいはコールドを判断する LRU リストと異なる点は，2NV-WB キャッシュアルゴリズムで用いる M-SCM LRU リストは LPA のほかに 1 bit の clean/dirty flag を必要とすること，また M-SCM 容量全体を表現するリストサイズが必要なことである．2NV-WB キャッシュアルゴリズムでは，M-SCM と S-SCM 間で同じデータを保存し一貫性が保たれている場合をクリーンデータ（clean data）と呼び clean flag “0”を，M-SCM にしかデータが無い場合はダーティデータ（dirty data）と呼び dirty flag “1”をたてる．M-SCM の空き容量が残り 20%未満になると evict 動作を発動する．M-SCM の空き容量が 20%以上になるまで, LRU 順にアクセス頻度の少ないデータを M-SCM から S-SCM へ evict

する（図 3.4 Evict operation）．S-SCM は M-SCM から evict されたデータおよび後述する MLC NAND 型フラッシュメモリから書き戻されたデータを保存する．S-SCM は MLC NAND 型フラッシュメモリの不揮発性キャッシュとして動作する．M-SCM LRU と同様に，S-SCM 内のデータのアクセス順序および clean/dirty flag を管理するために S-SCM LRU リストを用いる．S-SCM と MLC NAND 型フラッシュメモリとの間でデータの一貫性があればクリーンデータ（clean data）であるため clean flag “0”を，M-SCM にしかデータが無い場合はダーティデータ（dirty data）であり dirty flag “1”をたてる．さらに S-SCM の空き容量が減少すると，S-SCM から MLC NAND 型フラッシュメモリへデータを evict する．S-SCM の空き容量が 20%以上になるまで，S-SCM LRU リストに従いアクセス頻度の低いデータを S-SCM から MLC NAND 型フラッシュメモリへ evict する（図 3.4 Evict operation）．MLC NAND 型フラッシュメモリは大容量の不揮発性半導体メモリとして機能し，S-SCM から evict されたデータを保存する．M-SCM および S-SCM と異なり，MLC NAND 型フラッシュメモリのデータアクセス順序を管理するための LRU リストは用いない．

M-SCM LRU リストに基づいた M-SCM から S-SCM への eviction および S-SCM LRU リストに基づいた S-SCM から MLC NAND 型フラッシュメモリへの eviction は，evict すべきデータがクリーンあるいはダーティによって動作が異なる．第 2.4 節で述べた SCM および NAND 型フラッシュメモリを用いたストレージ向けの NV-WB キャッシュアルゴリズムと同様に，クリーンデータの場合は上位の不揮発性半導体メモリ（M-SCM あるいは S-SCM）から下位の不揮発性半導体メモリ（S-SCM あるいは MLC NAND 型フラッシュメモリ）へデータを移動する必要は無い．一方，ダーティデータの場合は上位の不揮発性半導体メモリ（M-SCM あるいは S-SCM）から下位の不揮発性半導体メモリ（S-SCM あるいは MLC NAND 型フラッシュメモリ）へデータを移動する必要がある．第 2.3 節および第 3.2 節で議論したように，M-SCM と S-SCM のアクセス性能差と比較して，S-SCM と MLC NAND 型フラッシュメモリのアクセス性能の差は大きく，さらに MLC NAND 型フラッシュメモリは GC が必要である．そのため，S-SCM から MLC NAND 型フラッシュメモリへ多くのデータが evict されると，ストレージ性能全体を低下させる．このことから S-SCM と MLC NAND 型フラッシュメモリとの間では，できるだけデータの一貫性（クリーン）を保つ方がよいと考える．

続いて上書き動作について述べる．M-SCM に保存されたデータを上書きするとき，その場での上書きが可能である．上書きされるデータ LPA はもっとも最近にアクセスがあったとして，M-SCM LRU リストの先頭に移動する．S-SCM もその場での上書きが可能であるが，今後頻繁に上書きや読み出しがある可能性を考慮してより高速な M-SCM に書き込む. その後，

M-SCM LRU リストの先頭に移動し, 一方で S-SCM の上書きされたデータは消去する. また, MLC NAND 型フラッシュメモリに上書きするときの動作は第 2.4 節で述べた SCM および NAND 型フラッシュメモリを用いたハイブリッドストレージに用いる NV-WB キャッシュアルゴリズムと同じである．MLC NAND 型フラッシュメモリの上書きされるデータを含むページを読み出し，上書きしたいデータと統合して S-SCM に書き込む．このとき上書きしたデータを含む LPA は，S-SCM LRU リストの先頭である MRU の位置に移動する．

続いて 2NV-WB の読み出し動作について述べる（図 3.4 Read operation）．M-SCM に保存されたデータを読み出すときは, M-SCM から読み出し, M-SCM LRU リストの先頭に移動する．S-SCM に保存されたデータを読み出すときは，初めに S-SCM から読み出しストレージコントローラへデータを伝える．次に S-SCM から読み出したデータを M-SCM へコピーする．M-SCM LRU リストの先頭に，S-SCM LRU リストの先頭に，読み出したデータの順序を移動する．また，MLC NAND 型フラッシュメモリから読み出すとき，第 2.4 節で述べた NV-WB キャッシュアルゴリズムと同様, MLC NAND 型フラッシュメモリからデータを読み出しストレージコントローラへ返す．次に MLC NAND 型フラッシュメモリから読み出したデータを S-SCM へコピーする．その後，S-SCM LRU リストの先頭へ読み出したデータの順序を移動する．M-SCM と MLC NAND 型フラッシュメモリは性能差および容量差が大きいため，MLC NAND 型フラッシュメモリから M-SCM へ直接データをコピーしない管理方式を設計した．読み出し動作は上書き動作と似ているが，下位のメモリから上位のメモリへデータをコピーするときに，下位メモリのデータをそのまま保存する点が異なる．下位メモリから上位メモリへデータを移動ではなくコピーするため，上位のメモリおよび下位のメモリで同じデータを持つ場合がある．これがクリーンデータとなる．

このように 2NV-WB では M-SCM と S-SCM 間，および S-SCM と MLC NAND 型フラッシュメモリ間でデータの移動が発生する．しかし M-SCM と MLC NAND 型フラッシュメモリ間のアクセス性能差を考慮し，M-SCM と MLC NAND 型フラッシュメモリ間ではデータの移動はしないとした．しかし，M-SCM および S-SCM の容量は MLC NAND 型フラッシュメモリと比較して小さいと考えられるため，参考文献[8]で明らかにしたように読み出しの多いアプリケーションに対して，M-SCM と S-SCM との間で循環するデータがストレージ性能を低下させる可能性がある．そのために，L2 キャッシュ向けのデータ管理方法[9]を参考に，M-SCM および S-SCM 間のデータ移動を減らす仕組みを取り入れた．2NV-WB キャッシュアルゴリズムにおいては，S-SCM LRU に保存されたデータはそれぞれ（Read_count, Write_count）をペアで持つ．まず，M-SCM から evict されたデータは，（Read_count, Write_count）＝（0, 0）

で S-SCM LRU に登録される．MLC NAND 型フラッシュメモリへの上書きや読み出しによって S-SCM LRU へ登録されたデータもまた（Read_count, Write_count）=（0, 0）で S-SCM LRU に登録される．S-SCM に保存されたデータに読み出しリクエストがあった場合 Read_count をインクリメントする．同様に S-SCM に保存されたデータに上書きリクエストがあった場合 Write_count をインクリメントする．（Read_count, Write_count）を導入しない場合，S-SCM に保存されたデータに一度でも読み出しあるいは上書きリクエストがあれば，そのデータは M-SCM にコピーあるいは書き込まれる．S-SCM LRU に（Read_count, Write_count）を導入することによって，Read_count が Read_threshold あるいは Write_count が Write_threshold を超えた場合だけ，S-SCM から M-SCM へのデータ移動が発生する．S-SCM からの読み出し動作では M-SCM にデータをコピーした後，M-SCM でそのデータに上書きが無ければ M-SCM と S-SCM との間でクリーンとして保たれる．その後 M-SCM から S-SCM への evict するときのオーバーヘッドが少ないため，本論文では Read_threshold = 5，Write_threshold = 10 として第 4 章で性能評価を行なう．

3.5 まとめ

本章では異種の不揮発性半導体メモリを用いたヘテロジニアスストレージ構成およびそのデータマネジメントアルゴリズムを述べた．従来研究と異なり，三種以上の不揮発性半導体メモリを用いて構成したストレージをヘテロジニアスストレージと呼び本論文で扱う．ヘテロジニアスストレージとして，1）SCM，MLC および TLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージおよび 2）M-SCM，S-SCM および NAND 型フラッシュメモリを用いたヘテロジニアスストレージを提案した．第一の SCM，MLC および TLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージは，SCM および MLC NAND 型フラッシュメモリを用いたハイブリッドストレージと比較して，MLC NAND 型フラッシュメモリに滞留するアクセス頻度の低いデータを TLC NAND 型フラッシュメモリに保存することで MLC NAND 型フラッシュメモリの書き換え回数を削減することを目的とする．さらに SCM の導入で上昇する総ストレージコストをビットコストの低い TLC NAND 型フラッシュメモリで均衡することができる．第二の M-SCM，S-SCM および NAND 型フラッシュメモリを用いたヘテロジニアスストレージは，SCM を二種類用いて極端にアクセス頻度の高いデータを M-SCM に，ややアクセス頻度の高いデータを S-SCM に保存することを特徴とする．

参考文献

[1] S.-H. Shin, D.-K. Shim, J.-Y. Jeong, O.-S. Kwon, S.-Y. Yoon, M.-H. Choi, T.-Y. Kim, H.-W. Park, H.-J. Yoon, Y.-S. Song, Y.-H. Choi, S.-W. Shim, Y.-L. Ahn, K-T. Park, J.-M. Han, K.-H. Kyung, and Y.-H. Jun, "A new 3-bit programming algorithm using SLC-to-TLC migration for 8MB/s high performance TLC NAND flash memory," in *IEEE Symposium on VLSI Circuits Digest of Technical Papers*, Jun. 2012, pp. 132-133.

[2] C. Matsui, Y. Yamaga, Y. Sugiyama, and K. Takeuchi, "8.9-times performance improvement by tri-hybrid storage system with SCM and MLC/TLC NAND flash memory," in *Extended Abstracts of International Conference on Solid State Devices and Materials (SSDM)*, Sep. 2016, pp. 105-106.

[3] C. Sun, K. Miyaji, K. Johguchi, and K. Takeuchi, "A high performance and energy-efficient cold data eviction algorithm for 3D-TSV hybrid ReRAM/MLC NAND SSD," *IEEE Transactions on Circuits and Systems-I (TCAS-I)*, vol. 61, no. 2, pp. 382-392, Feb. 2014.

[4] S. Hachiya, K. Johguchi, K. Miyaji, and K. Takeuchi, "TLC/MLC NAND flash mix-and-match design with exchangeable storage array," in *Extended Abstracts of International Conference on Solid State Devices and Materials (SSDM)*, Sep. 2013, pp. 894-895.

[5] S. Hachiya, K. Johguchi, K. Miyaji, and K. Takeuchi, "Hybrid triple-level-cell/multi-level-cell NAND flash storage array with chip exchangeable method," *Japanese Journal of Applied Physics (JJAP)*, vol. 53, no. 4S, pp. 04EE04-1 - 04EE04-10, Apr. 2014.

[6] C. Matsui and K. Takeuchi, "22% higher performance, 2x SCM write endurance heterogeneous storage with dual SCM and NAND flash memory," in *Proceedings of European Solid-State Device Research Conference (ESSDERC)*, Sep. 2017, pp. 6-9.

[7] S. Okamoto, C. Sun, S. Hachiya, T. Yamada, Y. Saito, T. O. Iwasaki, and K. Takeuchi, "Application driven SCM and NAND flash hybrid SSD design for data-centric computation system," in *Proceedings of IEEE International Memory Workshop (IMW)*, May 2015, pp. 157-160.

[8] T. Yamada, C. Matsui, and K. Takeuchi, "Workload-based co-design of non-volatile cache algorithm and storage class memory specifications for storage class memory/NAND flash hybrid SSDs," *IEICE Transactions on Electronics*, vol. E100-C, no. 4, pp. 373-381, Apr. 2017.

[9] X. Wu, L. Li, L. Zhang, E. Speight, and Y. Xie, "Power and performance of read-write aware hybrid cache with non-volatile memories," in *Proceedings of Design, Automation and Test in Europe Conference and Exhibition (DATE)*, Apr. 2009, pp. 737-742.

第4章 アプリケーションに応じた不揮発性メモリの選択

4.1 はじめに

本章では，第 3 章で提案した二種類の異種メモリを用いたヘテロジニアスストレージの性能評価を行なう．SystemC ベースのストレージエミュレータに，不揮発性半導体メモリの動作およびデータマネジメントアルゴリズムを実装する．不揮発性半導体メモリの容量比や書き込み・読み出し時間などのアクセス性能を変化させ，ヘテロジニアスストレージのアクセス性能，消費エネルギー，不揮発性半導体メモリの書き換え回数の点から評価し比較する．代表的なストレージアプリケーション毎にヘテロジニアスストレージの最適な不揮発性半導体メモリ構成を示す.

4.2 評価環境

ヘテロジニアスストレージの性能を評価し，アプリケーション特性に最適なメモリ構成を提示するため，SystemC を用いた transaction-level modeling（TLM）ベースのエミュレータを用いる[1]. TLM ではデータの書き込みや読み出しという転送単位ごとに抽象度の高いモデリングを行う．Register transfer level（RTL）ベースのエミュレータと比較すると，TLM は高速なシミュレーションが可能である．図 4.1 に示すように，ストレージエミュレータは，論理・物理アドレス変換，ウエアレベリング，GC などの flash translation layer（FTL）の機能および不揮発性半導体メモリ間のデータマネジメントアルゴリズムをモデル化している．NAND 型フラッシュメモリの論理アドレス・物理アドレス変換は，ページレベルマッピング[2][3][4]を用いる．ストレージの合計容量は各アプリケーションによって異なり，アプリケーションの最大アクセスアドレスに NAND 型フラッシュメモリのオーバープロビジョニング容量として 25%上乗せした．オーバープロビジョニング容量は NAND 型フラッシュメモリに追加される容量であり，オーバープロビジョニング容量が多いとガベージコレクションの頻度が減少し性能が向上する．ストレージエミュレータの入力はログベースのトレースであり，本論文では図 2.16 で特性ごとに分類した MSR Cambridge のトレース[5]を用いる．特性分類した各カ

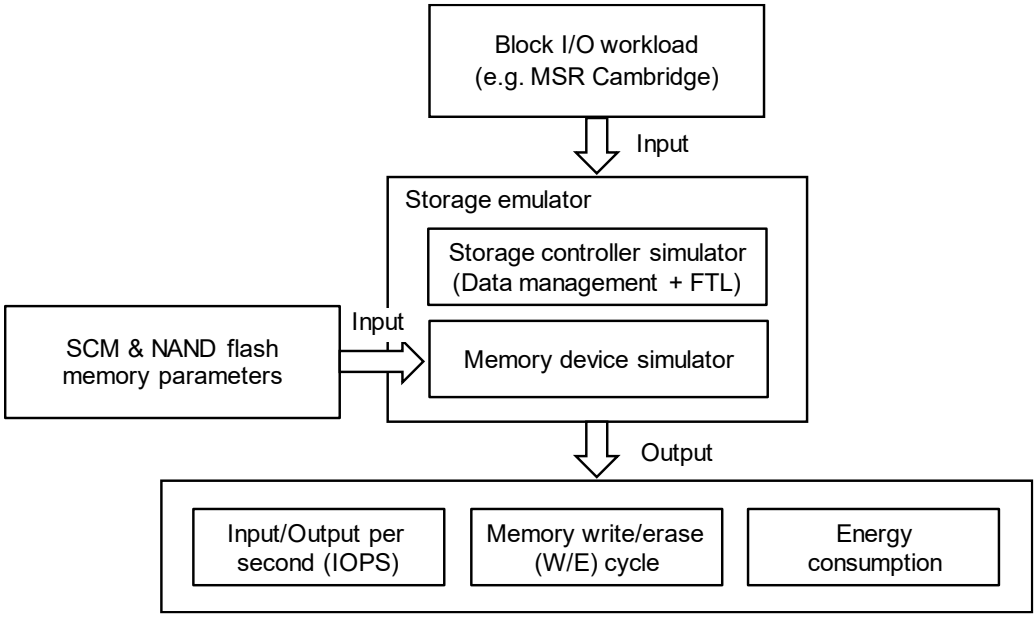


図 4.1 SystemC を用いた TLM ベースのエミュレータ概要 [1][6]

テゴリから 1 アプリケーションずつ選択し，prxy_0，proj_0，hm_0，src2_2，prxy_1，proj_3，src2_1 の 7 種を用いる．また，書き込み・読み出し時間，メモリコア数，I/O 電力のようなメモリに関するパラメータは表 2.1 および表 2.2 に示した不揮発性半導体メモリの特性を入力として用いる．本章では，異なる特性を持つ不揮発性半導体メモリは各 1 チップとし，同じ種類の不揮発性半導体メモリ間で並列動作を行わないこととした．さらにデータマネジメントアルゴリズムを実装し，用いる不揮発性半導体メモリによってデータ管理手法を変更することができる. シミュレーションが終わると, ストレージ性能（Input/Output per second, IOPS）, メモリの書き換え回数，消費エネルギーを出力する[6].

$$\mathrm{IOPS} = \frac{\#\ \text{of read requests} + \#\ \text{of write requests}}{\text{Total consuming time of (read operation + write operation)}} \tag{4.1}$$

$$\text{Total energy consumption} = \sum_{\mathrm{M-SCM,S-SCM,MLC\ flash,TLC\ flash}} \int \left(V_{\mathrm{DD\_}IO} \times I_{IO} + V_{\mathrm{DD\_}memory} \times I_{memory}\right) \mathrm{dt} \tag{4.2}$$

また NAND 型フラッシュメモリの書き換え回数（Write/erase cycle，W/E cycle）は，全 NAND 型フラッシュメモリのブロックの平均値を用いる．一方，M-SCM および S-SCM はそれぞれ全セクタの最大を W/E cycle とする．なお参考文献[7]に基づき M-SCM および S-SCM は，同一セクタに 5 回上書きがあった場合に，別のセクタへ書き込むウエアレベリングを行う．

以下の評価では，MLC NAND 型フラッシュメモリのみを用いたストレージを IOPS および消費エネルギーの基準とする．また，M-SCM，S-SCM，TLC NAND 型フラッシュメモリの書き換え回数は，MLC NAND 型フラッシュメモリの書き換え回数に対する比率とする．

4.3 アプリケーション特性に応じた不揮発性半導体メモリの構成

本節では，ストレージ容量が一定という仮定の下で，さまざまな不揮発性半導体メモリの組み合わせを用い，さまざまな不揮発性半導体メモリを用いたヘテロジニアスストレージの性能，消費エネルギー，メモリ書き換え回数を評価する．表 4.1 に示すようにストレージの構成は，(I) MLC NAND 型フラッシュメモリのみを用いたストレージ，(II) S-SCM および MLC NAND 型フラッシュメモリを用いたハイブリッドストレージ，(III) M-SCM および MLC NAND 型フラッシュメモリを用いたハイブリッドストレージ，(IV) M-SCM，S-SCM および MLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージ，(V) M-SCM，MLC および TLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージ，(VI) S-SCM，MLC および TLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージ，(VII) MLC および TLC NAND 型フラッシュメモリを用いたハイブリッドストレージである．それぞれのストレージ構成に対して不揮発性半導体メモリ間のデータ管理に，次のデータマネジメントアルゴリズムを適用する．(II) S-SCM および MLC NAND 型フラッシュメモリを用いたハイブリッドストレージ，(III) M-SCM および MLC NAND 型フラッシュメモリを用いたハイブリッドストレージには，第 2.4 節で述べた不揮発性半導体メモリ向けライトバック (Non-Volatile Memory Write-Back，NV-WB) キャッシュデータマネジメントアルゴリズムを適用する．(IV) M-SCM，S-SCM および MLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージには，第 3.4 節で述べた二種類の SCM を用いたライトバック (2 Non-Volatile Memory Write-Back，2NV-WB) キャッシュデータマネジメントアルゴリズムを適用する．(V) M-SCM，MLC および TLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージ，(VI) S-SCM，MLC および TLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージには，第 3.3 節で述べたコールドアンドフローズンデータエビクション (Cold and Frozen Data Eviction，CFDE) アルゴリズムを適用する．(VII) MLC および TLC NAND 型フラッシュメモリを用いたハイブリッドストレージには，第 2.4 節で述べたラウンドロビン・フローズンデータコレクションアルゴリズム (Round-Robin Frozen Data Collection Algorithm，RR-FDCA) を適用する．

表 4.1 不揮発性半導体メモリを用いたストレージの構成

Memory combination		M-SCM capacity [%]	S-SCM capacity [%]	MLC flash capacity [%]	TLC flash capacity [%]	Data management	Total memory cost
(I) MLC flash	1	0	0	100	0	-	1.0
(II) S-SCM &MLC flash	2	0	1	99	0	NV-WB cache	1.0
	3	0	3	97	0		1.1
	4	0	5	95	0		1.2
	5	0	7	93	0		1.3
	6	0	10	90	0		1.5
(III) M-SCM &MLC flash	7	1	0	99	0	NV-WB cache	1.1
	8	3	0	97	0		1.3
	9	5	0	95	0		1.4
	10	7	0	93	0		1.6
	11	10	0	90	0		1.9
(IV) M-SCM, S-SCM &MLC flash	12	9	1	90	0	2NV-WB cache	1.9
	13	7	3	90	0		1.8
	14	5	5	90	0		1.7
	15	3	7	90	0		1.6
	16	1	9	90	0		1.5
(V) M-SCM, MLC flash &TLC flash	17	10	0	78.8	11.2	CFDE	1.9
	18	10	0	67.5	22.5		1.8
	19	10	0	56.2	33.8		1.8
	20	10	0	45.0	45.0		1.7
	21	10	0	33.8	56.2		1.7
	22	10	0	22.5	67.5		1.6
(VI) S-SCM, MLC flash &TLC flash	23	0	10	78.8	11.2	CFDE	1.5
	24	0	10	67.5	22.5		1.4
	25	0	10	56.2	33.8		1.4
	26	0	10	45.0	45.0		1.3
	27	0	10	33.8	56.2		1.3
	28	0	10	22.5	67.5		1.3
(VII) MLC flash &TLC flash	29	0	0	87.5	12.5	RR-FDCA	1.0
	30	0	0	75	25		0.92
	31	0	0	62.5	37.5		0.88
	32	0	0	50	50		0.83
	33	0	0	37.5	62.5		0.79
	34	0	0	25	75		0.75

さらに，同じ組み合わせの不揮発性半導体メモリを用いる場合でも，不揮発性半導体メモリの容量比を変更することで，ストレージアプリケーションに適したストレージ構成を明らかにする．図 4.2 のように不揮発性半導体メモリの容量比を決める．ストレージ全体の容量

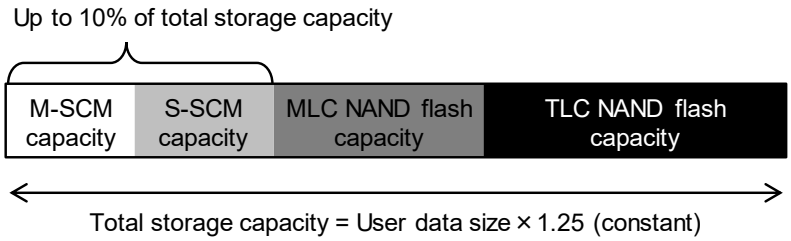


図 4.2 同一ストレージ容量のときの不揮発性半導体メモリ容量比

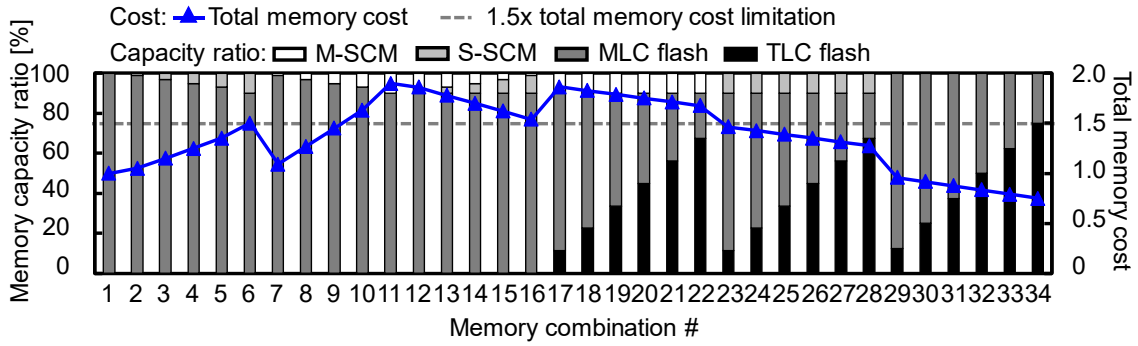


図 4.3 不揮発性半導体メモリ容量比とストレージコストの関係

は，ストレージアプリケーションのユーザデータサイズの 1.25 倍に統一した．本論文ではストレージアプリケーションのユーザデータサイズを，アクセスされる最大の論理アドレス（logical block address，LBA）と定義する．また本章では M-SCM および S-SCM はそれぞれ，表 2.2 に示した scenario 1（read/write latency = 0.1 μsec）および scenario 2（read/write latency = 1 μsec）の SCM であるとする．このとき，それぞれのストレージ構成の総ストレージコストは，式（4.3）で求められる．

$$\text{Total storage cost} = \sum_{\text{M-SCM,S-SCM,MLC flash,TLC flash}} (\text{memory capacity ratio} \times \text{bit cost ratio}) \tag{4.3}$$

メモリ容量比と総ストレージコストとの関係を図 4.3 に示す. 高速な M-SCM あるいは S-SCM の容量が増加するほど総ストレージコストは増加し，一方，TLC NAND 型フラッシュメモリの容量を増やすと総ストレージコストが低減できる．本章ではハイブリッドストレージあるいはヘテロジニアスストレージの総コストは，MLC NAND 型フラッシュメモリのみを用いたストレージと比較して 1.5 倍まで許容できると仮定する．その結果，いくつかの組み合わせはストレージコストの許容範囲外となる．特に M-SCM の容量が多い組み合わせ 10-15 および 17-22 は，総ストレージコストが MLC NAND 型フラッシュメモリのみを用いたストレージと

比較して最大で 1.9 倍となる. 一方で TLC NAND 型フラッシュメモリのビットコストが MLC NAND 型フラッシュメモリのビットコストの 2/3 であることから，TLC NAND 型フラッシュメモリの容量が多い組み合わせ 29-34 は MLC NAND 型フラッシュメモリのみを用いたストレージより低コストとなる.

図 4.4-図 4.10 に異なるストレージアプリケーションに対する，(a) MLC および TLC NAND 型フラッシュメモリの書き換え回数 (W/E cycle)，(b) MLC NAND 型フラッシュメモリの W/E cycle に対して規格化した M-SCM，S-SCM および TLC NAND 型フラッシュメモリの W/E cycle，(c) MLC NAND 型フラッシュメモリを用いたストレージに対して規格化した IOPS 性能，(d) MLC NAND 型フラッシュメモリを用いたストレージに対して規格化した消費エネルギーを示す[8][9].

4.3.1 書き込みが多くホットなストレージアプリケーション

書き込みが多く上書きが多く発生するホットな prxy_0 および proj_0 アプリケーションに対する性能を評価する．はじめに，ヘテロジニアスストレージの MLC および TLC NAND 型フラッシュメモリの書き換え回数（W/E cycle）を解析する．第 3.2 節で議論したように，NAND 型フラッシュメモリへ書き込みが多く行われると，無効ページが発生し，空のブロックを確保するためのガベージコレクション（GC）動作が必要となる．この GC 動作は有効ページのコピーとブロックの消去を行うため時間がかかる．そのため NAND 型フラッシュメモリの W/E cycle を解析することで，ストレージ性能に影響する MLC および TLC NAND 型フラッシュメモリの書き込みがわかる．図 4.4（a)，図 4.5（a）に prxy_0 および proj_0 アプリケーションに対する MLC および TLC NAND 型フラッシュメモリの W/E cycle を示す．MLC NAND 型フラッシュメモリのみを用いる場合（組み合わせ 1）と比較して，M-SCM あるいは S-SCM を用いると（組み合わせ 2-16)，MLC NAND 型フラッシュメモリの W/E cycle が減少していることがわかる. これは M-SCM あるいは S-SCM にホットなデータを書き込むことによって，MLC NAND 型フラッシュメモリへ書き込まれるデータが減少したことを示す. また，M-SCM あるいは S-SCM と MLC および TLC NAND 型フラッシュメモリを用いると（組み合わせ 17-28)，MLC NAND 型フラッシュメモリの容量が減少するに従い，MLC NAND 型フラッシュメモリの W/E cycle が増加する．これは， M-SCM あるいは S-SCM から evict され MLC NAND 型フラッシュメモリに書き込まれたデータによって，容量の少ない MLC NAND 型フラッシュメモリの書き換えが多く発生したことを示す.

次に図 4.4（b)，図 4.5（b）に示す，MLC NAND 型フラッシュメモリの書き換え一回当た

りの M-SCM および S-SCM の W/E cycle を解析する．M-SCM あるいは S-SCM と MLC NAND 型フラッシュメモリを用いたハイブリッドストレージ（組み合わせ 2-6 および 7-11）において，M-SCM および S-SCM の容量比が同じ場合，これらの W/E cycle は同一の数値を示す．M-SCM あるいは S-SCM の容量が少ない場合，ホストからのデータが頻繁に書き込まれるため高い W/E cycle を示す．一方 M-SCM あるいは S-SCM の容量が多くなると W/E cycle は低下する．しかし M-SCM あるいは S-SCM の容量が 10%の場合は，図 4.4（a），図 4.5（a）で示したように MLC NAND 型フラッシュメモリの W/E cycle が十分小さくなるため，相対的な値である図 4.4 (b)，図 4.5 (b) の M-SCM および S-SCM の W/E cycle は高くなる．また M-SCM，S-SCM および MLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージ（組み合わせ 12-16）では，S-SCM と比較して M-SCM の容量が十分多いとき，M-SCM の書き換え回数は少なく，S-SCM の書き換え回数が多くなる．第 3.4 節で述べた二種類の SCM を用いたライトバック（2 Non-Volatile Memory Write-Back，2NV-WB）キャッシュデータマネジメントアルゴリズムでは，S-SCM は二つの役割を持つためである．一つ目は M-SCM から evict されたデータを保存することである．二つ目は MLC NAND 型フラッシュメモリへ読み出しアクセスがあった場合に，読み出されたデータを S-SCM へコピーし次の読み出しアクセスに備えることである．このため，M-SCM の容量が多く S-SCM の容量が少ない場合，S-SCM の書き換え回数が M-SCM の書き換え回数を上回る．これば表 2.2 に示した予想される M-SCM および S-SCM の書き換え耐久性の比率と異なる．したがって組み合わせ 15，16 のように，M-SCM の書き換え回数が S-SCM の書き換え回数より高くなるように，M-SCM および S-SCM の容量を決めるべきであると考える．さらに，M-SCM あるいは S-SCM と MLC および TLC NAND 型フラッシュメモリを用いるヘテロジニアスストレージ（組み合わせ 17-22 および 23-28）では，追加した M-SCM および S-SCM の容量比が同じ場合（たとえば組み合わせ 17 と 23），これらの W/E cycle は同一の数値を示す．第 3.3 節で述べた SCM，MLC および TLC NAND 型フラッシュメモリに用いるコールドアンドフローズンデータエビクション（Cold and Frozen Data Eviction，CFDE）アルゴリズムにおいて，M-SCM あるいは S-SCM には頻繁に書き換えがあり，データサイズが小さいページ内で断片化したデータが書き込まれる．そのため M-SCM あるいは S-SCM の容量が多いほど，その W/E cycle は減少する．さらに，MLC および TLC NAND 型フラッシュメモリを用いる場合（組み合わせ 17-22，23-28，29-34），図 4.4（a），図 4.5（a）でも示したように，TLC NAND 型フラッシュメモリの W/E cycle は MLC NAND 型フラッシュメモリの W/E cycle の約 1/100 以下に抑えられている．

図 4.4（a），図 4.5（a）に示した MLC NAND 型フラッシュメモリの W/E cycle が減少する

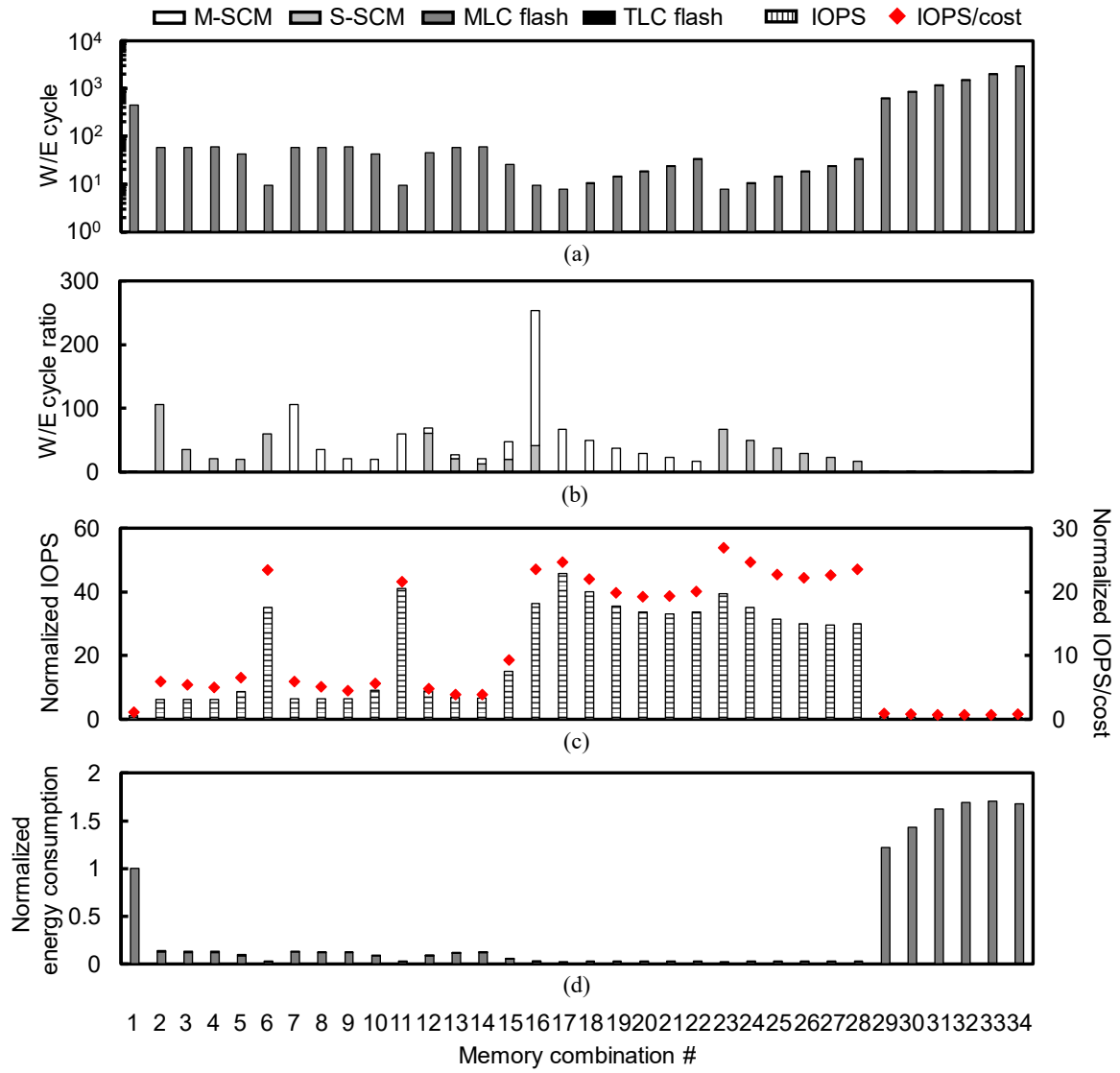


図 4.4 prxy_0（write-hot-random）アプリケーションの（a）MLC および TLC NAND 型フラッシュメモリの書き換え回数，（b）MLC NAND 型フラッシュメモリの書き換え一回当たりの M-SCM，S-SCM，TLC NAND 型フラッシュメモリの書き換え回数，（c）MLC NAND 型フラッシュメモリのみを用いたストレージに対するヘテロジニアスストレージの IOPS 性能，（d）MLC NAND 型フラッシュメモリのみを用いたストレージに対するヘテロジニアスストレージの消費エネルギー

ほど，図 4.4（c），図 4.5（c）に示す IOPS 性能が向上することは明らかである．特に高速な M-SCM の容量が増えるほど，ストレージ性能は向上する．prxy_0 および proj_0 アプリケーションでは，組み合わせ 17 の M-SCM 容量が 10%，MLC および TLC NAND 型フラッシュメモリの容量がそれぞれ 78.8%，11.2%の場合に最も高い IOPS 性能を示す．これは頻繁に上書きされるデータを M-SCM に保存することで，書き込みに要する時間を短くすることができたためである．さらに MLC および TLC NAND 型フラッシュメモリを併用することで，ほと

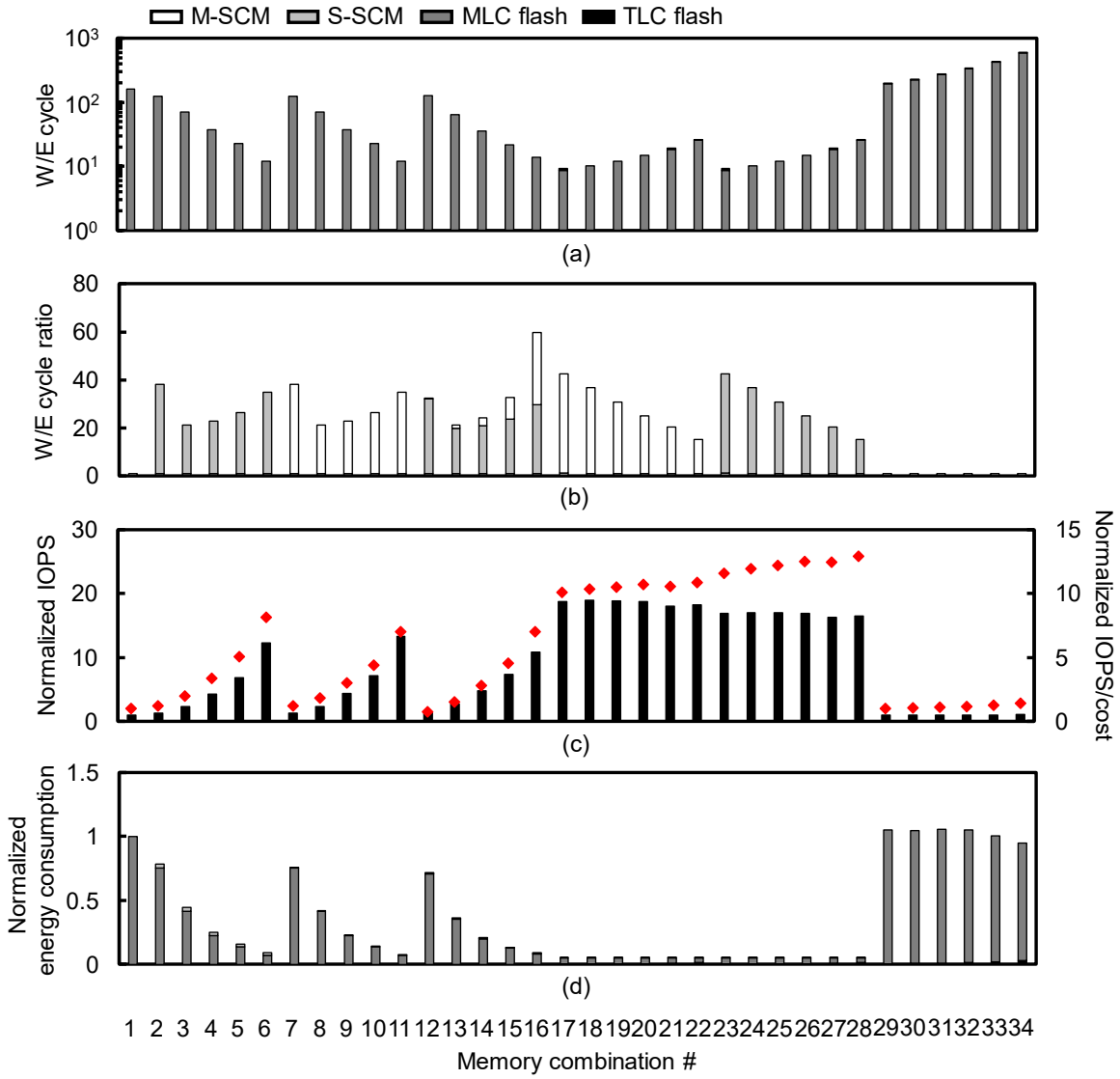


図 4.5 proj_0（write-hot-sequential）アプリケーションの（a）MLC および TLC NAND 型フラッシュメモリの書き換え回数，（b）MLC NAND 型フラッシュメモリの書き換え一回当たりの M-SCM，S-SCM，TLC NAND 型フラッシュメモリの書き換え回数，（c）MLC NAND 型フラッシュメモリのみを用いたストレージに対するヘテロジニアスストレージの IOPS 性能，（d）MLC NAND 型フラッシュメモリのみを用いたストレージに対するヘテロジニアスストレージの消費エネルギー

んど上書きされないコールドあるいはフローズンデータを TLC NAND 型フラッシュメモリへ追い出す．これにより，MLC NAND 型フラッシュメモリのガベージコレクション時，有効ページの少ないブロックを選択し消去することができる．さらにコールドあるいはフローズンデータを TLC NAND 型フラッシュメモリへ追い出すことで，ガベージコレクションの発生頻度を低減することができた．

しかし総ストレージコストを考慮すると，表 4.1 の赤字で示したようにいくつかの不揮発性半導体メモリの組み合わせは，総ストレージコストが MLC NAND 型フラッシュメモリのみを用いたストレージの 1.5 倍を超える．そのため prxy_0 アプリケーションに対しては， M-SCM 1%，S-SCM 9%，MLC NAND 型フラッシュメモリ（組み合わせ 16）あるいは S-SCM 10%，MLC NAND 型フラッシュメモリ 78.8%，TLC NAND 型フラッシュメモリ 11.2%（組み合わせ 23）の場合に高い性能を示す．また proj_0 アプリケーションに対しては，S-SCM 10%，MLC NAND 型フラッシュメモリ 78.8%，TLC NAND 型フラッシュメモリ 11.2%（組み合わせ 23）の場合に高い性能を示す．さらに総ストレージコストを考慮した，図 4.4（c），図 4.5（c）の右縦軸に示す IOPS/cost という指標を導入する．MLC NAND 型フラッシュメモリのみを用いたストレージに対して，IOPS/cost = 1 を示す．つまり MLC NAND 型フラッシュメモリのみを用いたストレージに対して，異種の不揮発性半導体メモリを用いたヘテロジニアスストレージの性能がどれだけ向上したか，また総ストレージコストがどれほど増加したかを示す．したがって性能向上率が高く総ストレージコストが低いほど IOPS/cost は高い数値を示す．IOPS/cost を用いると，prxy_0 アプリケーションに対しては S-SCM 10%，MLC NAND 型フラッシュメモリ 78.8%，TLC NAND 型フラッシュメモリ 11.2%（組み合わせ 23）の場合に最も高い．また proj_0 アプリケーションに対しては，S-SCM 10%，MLC NAND 型フラッシュメモリ 22.5%，TLC NAND 型フラッシュメモリ 67.5%（組み合わせ 28）の場合に最も高い．

図 4.4（d），図 4.5（d）に示すストレージの消費エネルギーは，図 4.4（c），図 4.5（c）に示した IOPS 性能と逆の相関を示す．表 2.1 および表 2.2 に示したように，読み出し・書き込み時間も，M-SCM および S-SCM と比較して MLC および TLC NAND 型フラッシュメモリのほうが長い．そのため，MLC および TLC NAND 型フラッシュメモリへのアクセスが少なく，IOPS 性能の高い不揮発性半導体メモリの組み合わせほど消費エネルギーは低い．

4.3.2 書き込みが多くコールドなストレージアプリケーション

図 4.6（a）および図 4.7（a）に示す hm_0 および src2_2 アプリケーションに対する MLC および TLC NAND 型フラッシュメモリの W/E cycle は，M-SCM あるいは S-SCM を用いることで（組み合わせ 2-16），MLC NAND 型フラッシュメモリの W/E cycle が減少する．しかし M-SCM あるいは S-SCM 容量を増加しても，MLC NAND 型フラッシュメモリの W/E cycle の減少率はほとんど変化しない．これは，頻繁に書き込み・読み出しされるデータが少ないため，M-SCM あるいは S-SCM は単にデータを一時的に書き込むバッファとして機能しているためである．また，M-SCM あるいは S-SCM と MLC および TLC NAND 型フラッシュメモリを用

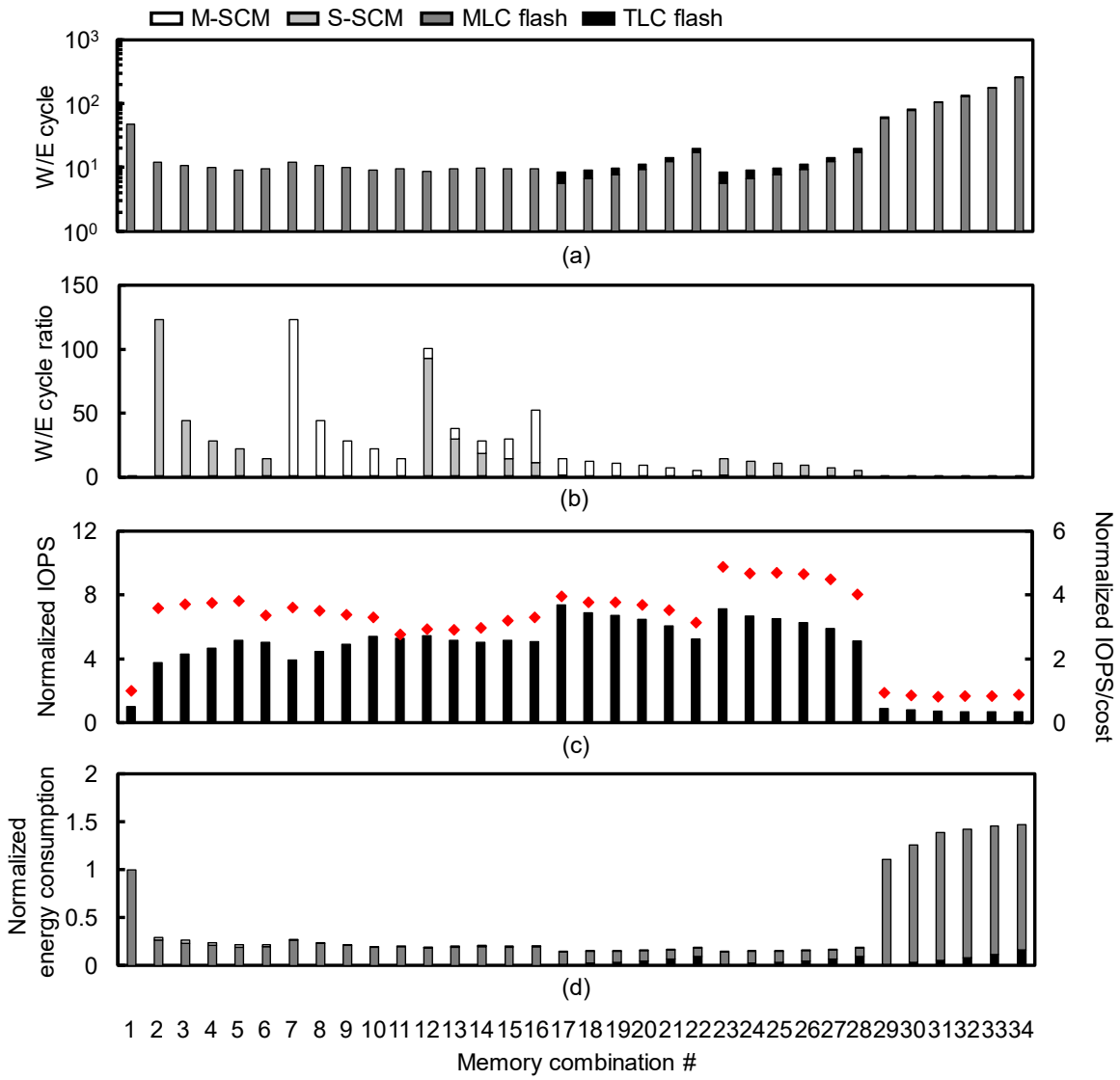


図 4.6 hm_0（write-cold-random）アプリケーションの（a）MLC および TLC NAND 型フラッシュメモリの書き換え回数，（b）MLC NAND 型フラッシュメモリの書き換え一回当たりの M-SCM，S-SCM，TLC NAND 型フラッシュメモリの書き換え回数，（c）MLC NAND 型フラッシュメモリのみを用いたストレージに対するヘテロジニアスストレージの IOPS 性能，（d）MLC NAND 型フラッシュメモリのみを用いたストレージに対するヘテロジニアスストレージの消費エネルギー

いると（組み合わせ 17-28），MLC NAND 型フラッシュメモリの容量が減少するに従い，MLC NAND 型フラッシュメモリの W/E cycle が増加する．これは， M-SCM あるいは S-SCM から evict され MLC NAND 型フラッシュメモリに書き込まれたデータによって，容量の少ない MLC NAND 型フラッシュメモリの書き換えが多く発生したことを示す．さらに，MLC および TLC NAND 型フラッシュメモリを用いる場合（組み合わせ 17-22，23-28，29-34），hm_0 アプリケーションに対しては，TLC NAND 型フラッシュメモリの W/E cycle は MLC NAND 型

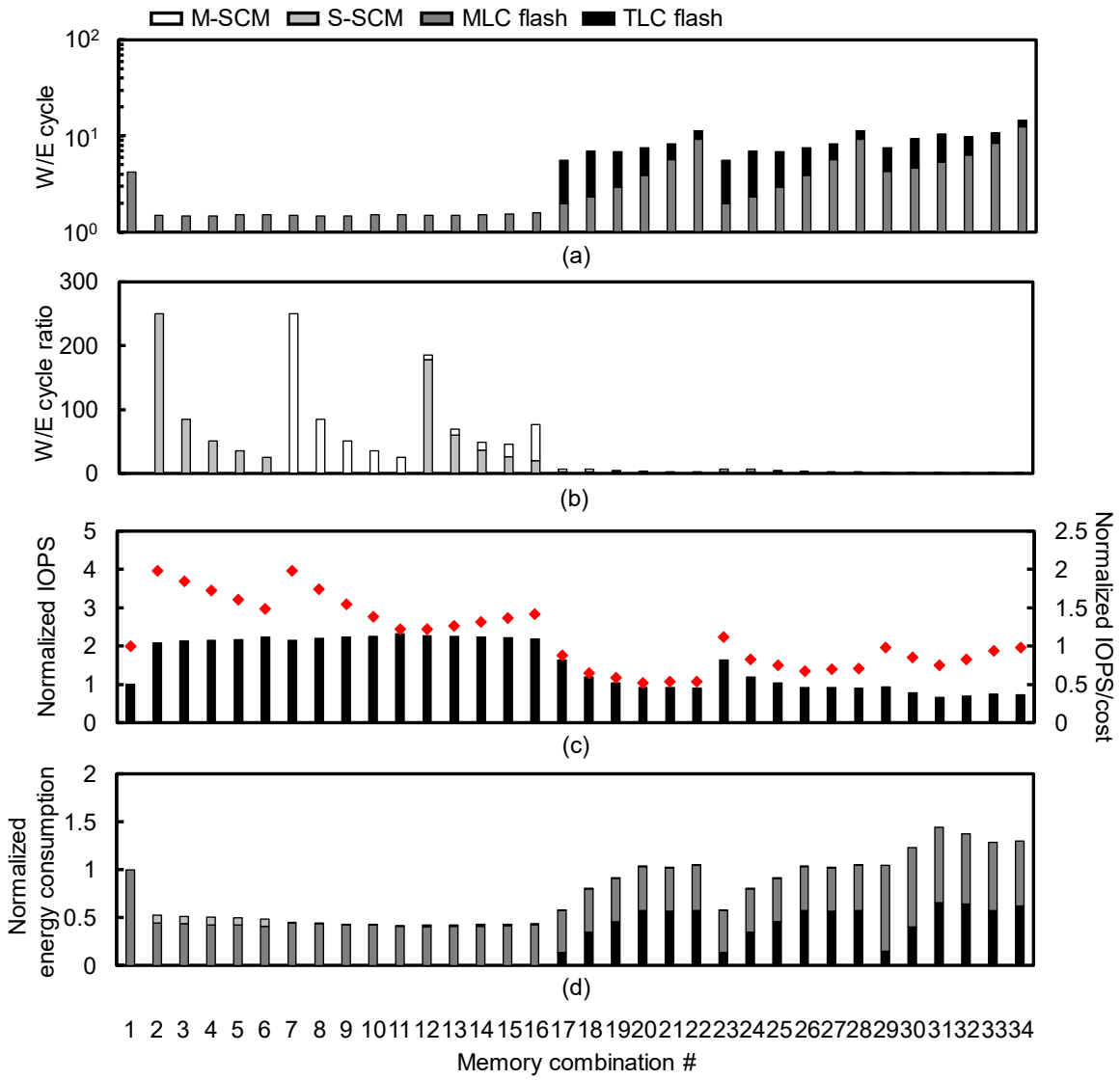


図 4.7 src2_2（write-cold-sequential）アプリケーションの（a）MLC および TLC NAND 型フラッシュメモリの書き換え回数，（b）MLC NAND 型フラッシュメモリの書き換え一回当たりの M-SCM，S-SCM，TLC NAND 型フラッシュメモリの書き換え回数，（c）MLC NAND 型フラッシュメモリのみを用いたストレージに対するヘテロジニアスストレージの IOPS 性能，（d）MLC NAND 型フラッシュメモリのみを用いたストレージに対するヘテロジニアスストレージの消費エネルギー

フラッシュメモリの W/E cycle の約 1/2 以下に抑えられている．一方で src2_2 アプリケーションに対しては，MLC NAND 型フラッシュメモリの容量が小さい場合（組み合わせ 17-18, 23-24），TLC NAND 型フラッシュメモリの W/E cycle は MLC NAND 型フラッシュメモリの W/E cycle の約 2 倍となる．このため，src2_2 アプリケーションに対して TLC NAND 型フラッシュメモリを用いる場合，その容量は十分に MLC NAND 型フラッシュメモリの容量より大きいことが求められる．

次に図 4.6（b），図 4.7（b）に示す，MLC NAND 型フラッシュメモリの書き換え一回当たりの M-SCM および S-SCM の W/E cycle を解析する．M-SCM あるいは S-SCM と MLC NAND 型フラッシュメモリを用いたハイブリッドストレージ（組み合わせ 2-6 および 7-11）において，M-SCM あるいは S-SCM の容量を増やすとそれらの W/E cycle が減少する．これは M-SCM あるいは S-SCM の容量が増えたため，セクタ当たりの平均書き換え回数が減少したことによる．また M-SCM，S-SCM および MLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージ（組み合わせ 12-16）では，第 4.3.1 章で述べた理由により S-SCM と比較して M-SCM の容量が十分多いとき，M-SCM の書き換え回数は少なく，S-SCM の書き換え回数が多くなる．

MLC NAND 型フラッシュメモリの W/E cycle が減少するほど，図 4.6（c），図 4.7（c）に示す IOPS 性能が向上する. hm_0 アプリケーションでは, 組み合わせ 17 の M-SCM 容量が 10%, MLC および TLC NAND 型フラッシュメモリの容量がそれぞれ 78.8%，11.2%の場合に最も高い IOPS 性能を示す．書き換えられるデータと書き換えられないデータとを MLC および TLC NAND 型フラッシュメモリに分けて保存できたと考える．一方 src2_2 アプリケーションに対しては，組み合わせ 11 の M-SCM 10%および MLC NAND 型フラッシュメモリ 90%を用いる場合がもっとも高い IOPS 性能を示す．しかし，M-SCM あるいは S-SCM および MLC NAND 型フラッシュメモリを用いたハイブリッドストレージにおいては，M-SCM あるいは S-SCM の容量に関わらずその性能はほとんど変化しない. ストレージコストも考慮した IOPS/cost の指標で考えると，M-SCM あるいは S-SCM を小容量の 1%だけ用いることが良いとわかる．

図 4.6（d），図 4.7（d）に示すストレージの消費エネルギーは，図 4.6（c），図 4.7（c）に示した IOPS 性能と逆の相関を示す．

4.3.3 読み出しが多くホットなストレージアプリケーション

読み出しが多いホットな prxy_1 アプリケーションに対する性能を評価する．図 4.8（a）に示す prxy_1 アプリケーションに対する MLC および TLC NAND 型フラッシュメモリの W/E cycle は，M-SCM あるいは S-SCM を用いることで（組み合わせ 2-16），MLC NAND 型フラッシュメモリの W/E cycle が減少する．特に M-SCM あるいは S-SCM の容量が 7%以上になると，MLC NAND 型フラッシュメモリの W/E cycle はゼロになる．また，M-SCM あるいは S-SCM と MLC および TLC NAND 型フラッシュメモリを用いると（組み合わせ 17-28），MLC NAND 型フラッシュメモリの容量が減少するに従い, MLC NAND 型フラッシュメモリの W/E cycle が増加する．これは， M-SCM あるいは S-SCM から evict され MLC NAND 型フラッシ

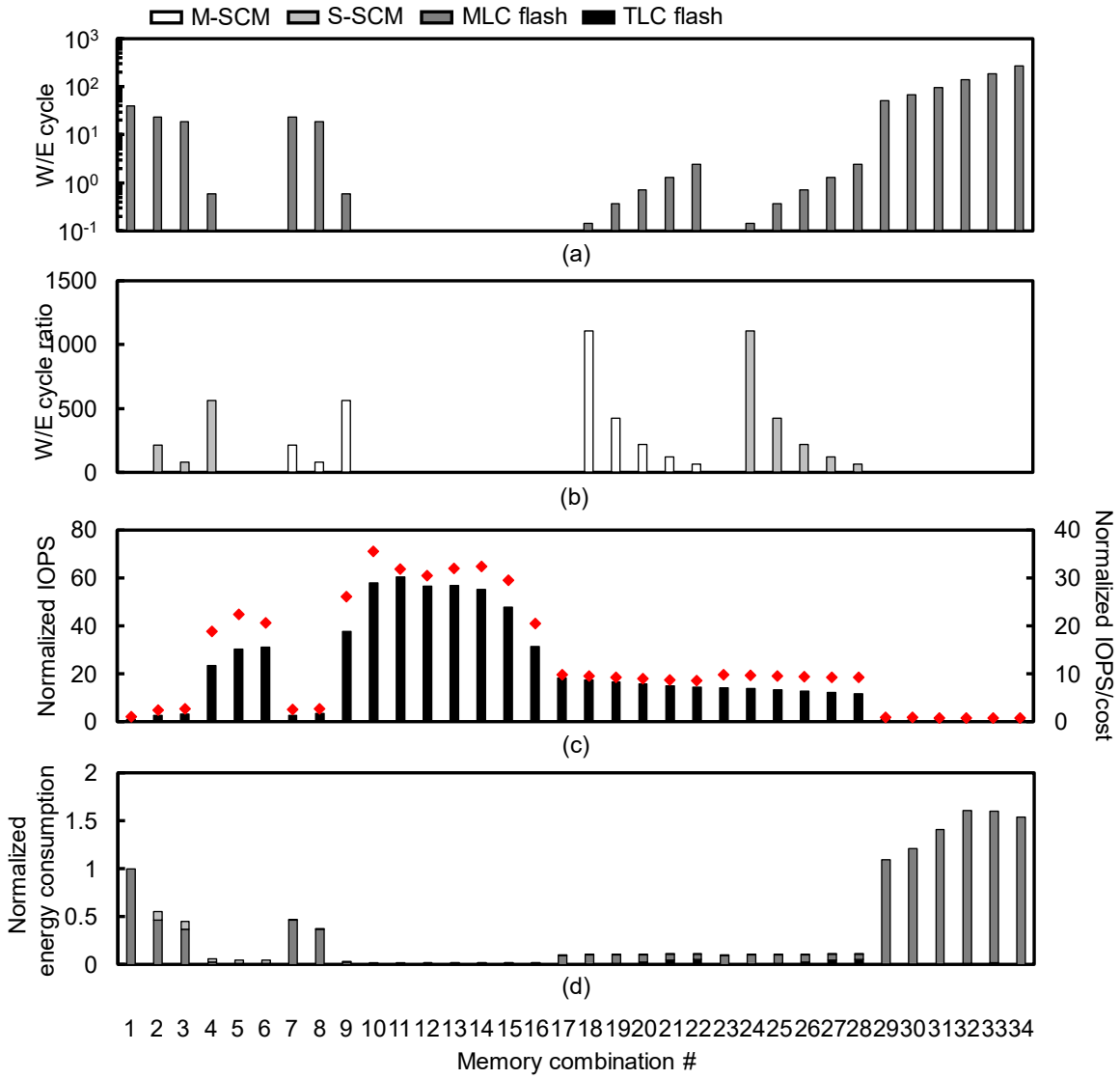


図 4.8 prxy_1（read-hot-random）アプリケーションの（a）MLC および TLC NAND 型フラッシュメモリの書き換え回数，（b）MLC NAND 型フラッシュメモリの書き換え一回当たりの M-SCM，S-SCM，TLC NAND 型フラッシュメモリの書き換え回数，（c）MLC NAND 型フラッシュメモリのみを用いたストレージに対するヘテロジニアスストレージの IOPS 性能，（d）MLC NAND 型フラッシュメモリのみを用いたストレージに対するヘテロジニアスストレージの消費エネルギー

ュメモリに書き込まれたデータによって，容量の少ない MLC NAND 型フラッシュメモリの書き換えが多く発生したことを示す．prxy_1 アプリケーションに対しては，TLC NAND 型フラッシュメモリの W/E cycle は常にゼロであるため（組み合わせ 17-34），TLC NAND 型フラッシュメモリを用いる必要は無いことがわかる．

次に図 4.8（b）に示す，MLC NAND 型フラッシュメモリの書き換え一回当たりの M-SCM

および S-SCM の W/E cycle を解析する．M-SCM あるいは S-SCM と MLC NAND 型フラッシュメモリを用いたハイブリッドストレージ（組み合わせ 2-6 および 7-11）において，M-SCM あるいは S-SCM の容量が少ない場合，MLC NAND 型フラッシュメモリから頻繁に書き戻されるため高い W/E cycle を示す．一方 M-SCM あるいは S-SCM の容量が多くなると W/E cycle は低下する．しかし M-SCM あるいは S-SCM の容量が 5%の場合は，図 4.8（a）で示したように MLC NAND 型フラッシュメモリの W/E cycle が十分小さくなるため，相対的な値である図 4.8（b）の M-SCM および S-SCM の W/E cycle は高くなる．ここで図 4.8（b）で，MLC NAND 型フラッシュメモリの W/E cycle がゼロになる組み合わせ（5-6，9-17，23）については，そのほかのメモリの相対的な W/E cycle を示していない．

MLC NAND 型フラッシュメモリの W/E cycle が減少するほど，図 4.8（c）に示す IOPS 性能が向上することは明らかである．特に高速な M-SCM の容量が増えるほど，ストレージ性能は向上する．prxy_1 アプリケーションでは，組み合わせ 11 の M-SCM 10%および MLC NAND 型フラッシュメモリ 90%を用いる場合がもっとも高い IOPS 性能を示す．これは頻繁に読み出されるデータを M-SCM に保存することで，読み出しに要する時間を短くすることができたためである．図 4.8（a）で見たように prxy_1 アプリケーションでは TLC NAND 型フラッシュメモリへほとんどアクセスしない．そのため TLC NAND 型フラッシュメモリを用いる組み合わせ 17-22，23-28，29-34 より，M-SCM あるいは S-SCM を用いる不揮発性半導体メモリの組み合わせが良いと考える．しかし第 4.3 節で述べたように，ハイブリッドストレージあるいはヘテロジニアスストレージの総コストは，MLC NAND 型フラッシュメモリのみを用いたストレージと比較して 1.5 倍まで許容できると仮定するため，IOPS/cost を用いると，prxy_1 アプリケーションに対しては M-SCM 5%，MLC NAND 型フラッシュメモリ 95%（組み合わせ 9）の場合に最も高い．

図 4.8（d）に示すストレージの消費エネルギーは，図 4.8（c）に示した IOPS 性能と逆の相関を示す．

4.3.4 読み出しが多くコールドなストレージアプリケーション

図 4.9（a）に示す proj_3 アプリケーションに対する MLC および TLC NAND 型フラッシュメモリの W/E cycle は，M-SCM あるいは S-SCM を用いることで（組み合わせ 2-16），MLC NAND 型フラッシュメモリの W/E cycle が減少しゼロになる．また，M-SCM あるいは S-SCM と MLC および TLC NAND 型フラッシュメモリを用いると（組み合わせ 17-28），MLC NAND 型フラッシュメモリの容量が減少するに従い，MLC NAND 型フラッシュメモリの W/E cycle

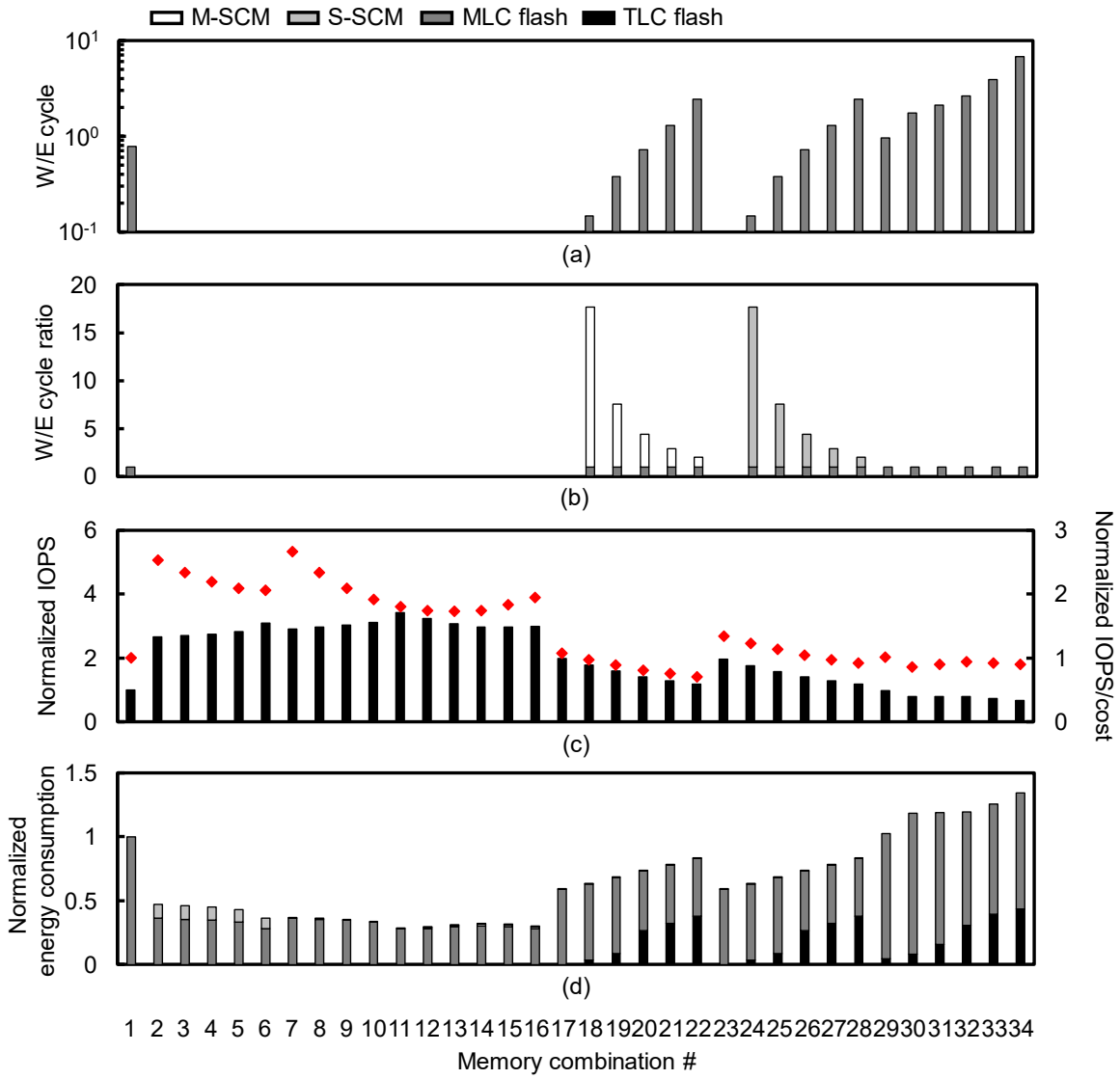


図 4.9 proj_3（write-cold-random）アプリケーションの（a）MLC および TLC NAND 型フラッシュメモリの書き換え回数，（b）MLC NAND 型フラッシュメモリの書き換え一回当たりの M-SCM，S-SCM，TLC NAND 型フラッシュメモリの書き換え回数，（c）MLC NAND 型フラッシュメモリのみを用いたストレージに対するヘテロジニアスストレージの IOPS 性能，（d）MLC NAND 型フラッシュメモリのみを用いたストレージに対するヘテロジニアスストレージの消費エネルギー

が増加する．これは，M-SCM あるいは S-SCM から evict され MLC NAND 型フラッシュメモリに書き込まれたデータによって，容量の少ない MLC NAND 型フラッシュメモリの書き換えが多く発生したことを示す．proj_3 アプリケーションに対しては，TLC NAND 型フラッシュメモリの W/E cycle は常にゼロであるため（組み合わせ 17-34），TLC NAND 型フラッシュメモリは一度書き込まれただけで消去は行われていないことが分かる．

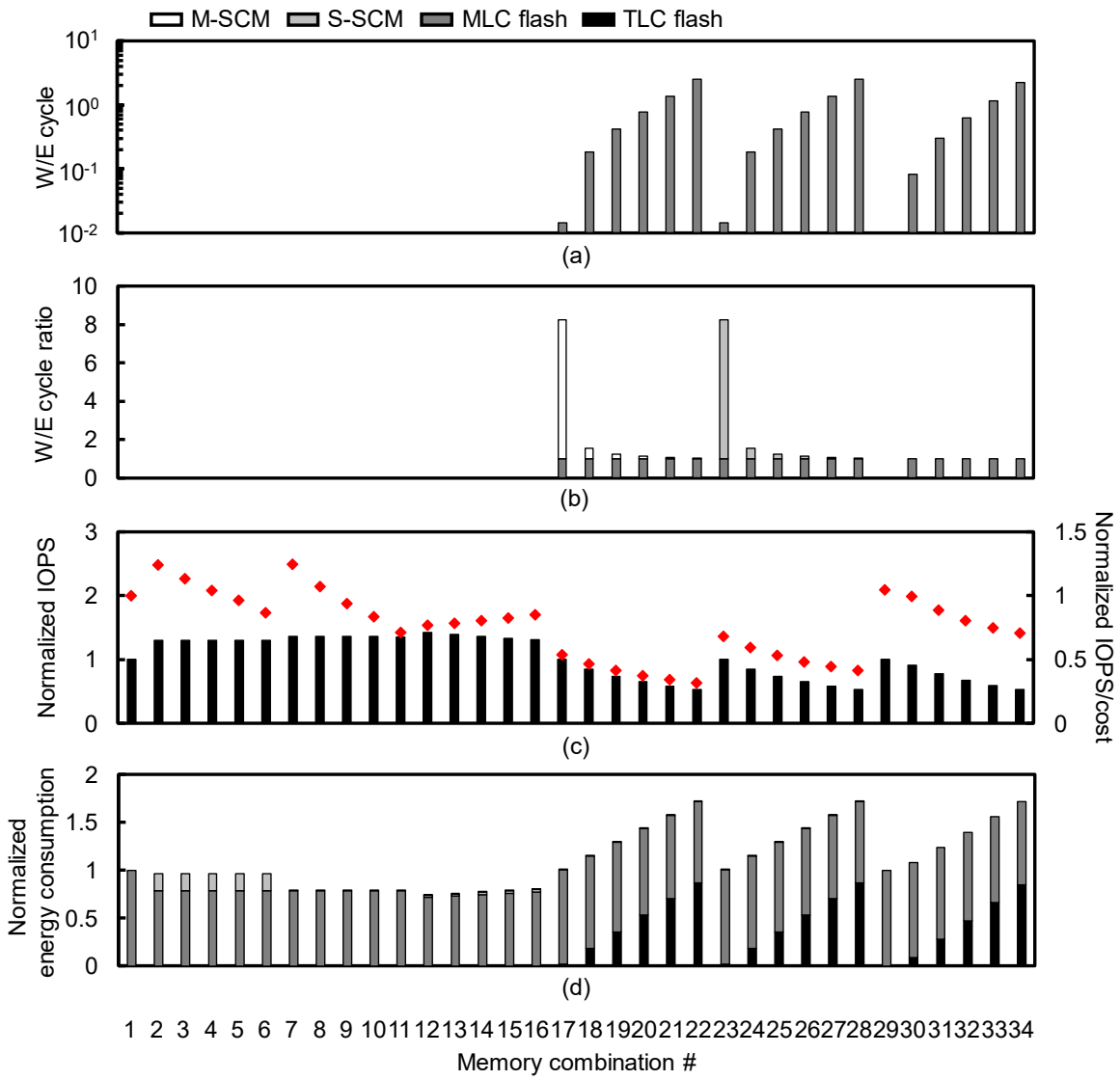


図 4.10 src2_1（read-cold-sequential）アプリケーションの（a）MLC および TLC NAND 型フラッシュメモリの書き換え回数，（b）MLC NAND 型フラッシュメモリの書き換え一回当たりの M-SCM，S-SCM，TLC NAND 型フラッシュメモリの書き換え回数，（c）MLC NAND 型フラッシュメモリのみを用いたストレージに対するヘテロジニアスストレージの IOPS 性能，（d）MLC NAND 型フラッシュメモリのみを用いたストレージに対するヘテロジニアスストレージの消費エネルギー

図 4.10（a）に示す src2_1 アプリケーションは書き込みデータが少ないため，MLC NAND 型フラッシュメモリのみを用いたストレージ（組み合わせ 1）の場合でもその W/E cycle はゼロである．そのため，M-SCM あるいは S-SCM を用いたハイブリッドストレージ（組み合わせ 2-6 および 7-11）においても，MLC NAND 型フラッシュメモリの W/E cycle はゼロのまま変わらない．また，M-SCM あるいは S-SCM と MLC および TLC NAND 型フラッシュメモリ

を用いると（組み合わせ 17-28），MLC NAND 型フラッシュメモリの容量が減少するに従い，MLC NAND 型フラッシュメモリの W/E cycle が増加する．これは，M-SCM あるいは S-SCM から evict され MLC NAND 型フラッシュメモリに書き込まれたデータによって，容量の少ない MLC NAND 型フラッシュメモリの書き換えが多く発生したことを示す．しかし src2_1 アプリケーションに対しては，TLC NAND 型フラッシュメモリの W/E cycle は常にゼロであるため（組み合わせ 17-34），TLC NAND 型フラッシュメモリは一度書き込まれただけで消去は行われていないことが分かる．

次に図 4.9（b），図 4.10（b）に示す，MLC NAND 型フラッシュメモリの書き換え一回当たりの M-SCM および S-SCM の W/E cycle はせいぜい 10-20 回程度であることがわかる（組み合わせ 17-28）．いったん M-SCM あるいは S-SCM に書き込まれたデータは，ふたたびアクセスされることなく MLC NAND 型フラッシュメモリへ evict されている．ここで図 4.8 (b) で，MLC NAND 型フラッシュメモリの W/E cycle がゼロになる組み合わせについては，そのほかのメモリの相対的な W/E cycle を示していない．

proj_3 および src2_1 アプリケーションに対しては，組み合わせ 11 の M-SCM 10%および MLC NAND 型フラッシュメモリ 90%を用いる場合がもっとも高い IOPS 性能を示すが，M-SCM あるいは S-SCM および MLC NAND 型フラッシュメモリを用いたハイブリッドストレージにおいては, M-SCM あるいは S-SCM の容量に関わらずその性能はほとんど変化しない．ストレージコストも考慮した IOPS/cost の指標で考えると，M-SCM あるいは S-SCM を極小容量の 1%だけを書き込みバッファとして用いることが良いとわかる．

4.4 まとめ

本章では，SCM，MLC および TLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージ，M-SCM，S-SCM および MLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージについて，メモリ容量比を変化させアプリケーションに最適なメモリ構成を評価した．高速な M-SCM を大容量用いるほどヘテロジニアスストレージの性能が向上するが，単位容量当たりの M-SCM のコストは，NAND 型フラッシュメモリと比較して約 10 倍と予想されるため，本論文では MLC NAND 型フラッシュメモリのみを用いたストレージのコストと比較して，ヘテロジニアスストレージは 1.5 倍のコスト増が許容できると仮定した．

SCM の特性およびストレージアプリケーション特性に依存してヘテロジニアスストレージの最適な不揮発性半導体メモリの組み合わせが異なることを明らかにした．一部のデータ

が頻繁に書き換えられるストレージアプリケーション（prxy_0，proj_0）に対しては，S-SCM を大容量用いることで性能を向上できることを明らかにした．特に頻繁に書き換えられるデータの多いアプリケーション（prxy_0）については，M-SCM を極小容量用いて書き込み性能を向上できることを明らかにした．また一部のデータが頻繁に読み出し・書き込みされるストレージアプリケーション（prxy_1）に対しては，M-SCM を大容量用いることで性能を向上できることを明らかにした. さらに高速な M-SCM を用いるよりも, 低速大容量な TLC NAND 型フラッシュメモリを用いるべきストレージアプリケーション（hm_0）が存在することを明らかにした．一方で，頻繁に上書きおよび読み出しされないデータの多いアプリケーション（proj_3，src2_2，src2_1）に対しては，小容量で高速な M-SCM を書き込みバッファとして機能させることが良いことを明らかにした.

参考文献

[1] H. Fujii, K. Miyaji, K. Johguchi, K. Higuchi, C. Sun, and K. Takeuchi, “x11 performance increase, x6.9 endurance enhancement, 93% energy reduction of 3D TSV-integrated hybrid ReRAM/MLC NAND SSDs by data fragmentation suppression,” in *IEEE Symposium on VLSI Circuits Digest of Technical Papers*, Jun. 2012, pp.134-135.

[2] J. Kim, J. M. Kim, S. H. Noh, S. L. Min, and Y. Cho, “A space-efficient flash translation layer for compactflash systems,” *IEEE Transactions on Consumer Electronics*, vol. 48, no. 2, pp. 366-375, May 2002.

[3] A. Gupta, Y. Kim, and B. Urgaonkar, “DFTL: A flash translation layer employing demand-based selective caching of page-level address mappings,” in *Proceedings of ACM International Conference on Architectural Support for Programming Languages and Operating Systems (ASPLOS)*, Mar. 2009, pp. 229-240.

[4] Z. Xu, R. Li, and C-Z Xu, “CAST: a page-level FTL with compact address mapping and parallel data blocks,” in *Proceedings of IEEE International Performance Computing and Communications Conference (IPCCC)*, Dec. 2012, pp. 142-151.

[5] MSR Cambridge Traces, http://iotta.snia.org/traces/388.

[6] C. Sun, A. Soga, C. Matsui, A. Arakawa, and K. Takeuchi, “LBA scrambler: A NAND flash aware data management scheme for high-performance solid-state drives,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 24, no. 1, pp. 115-128, Jan. 2016.

[7] T. Onagi, C. Sun, and K. Takeuchi, “Design guidelines of all storage class memory (SCM) SSD and hybrid SCM/NAND flash SSD to balance performance, power, endurance and cost,” in

Extended Abstracts of International Conference on Solid State Devices and Materials (SSDM), Sep. 2014, pp. 106-107.

[8] C. Matsui, T. Yamada, Y. Sugiyama, Y. Yamaga, and K. Takeuchi, “Optimal memory configuration analysis in tri-hybrid solid-state drives with storage class memory and multi-level cell/triple-level cell NAND flash memory,” *Japanese Journal of Applied Physics (JJAP)*, vol. 56, no. 4S, pp. 04CE02-1 - 04CE02-9, Apr. 2017.

[9] C. Matsui and K. Takeuchi, “22% higher performance, 2x SCM write endurance heterogeneous storage with dual SCM and NAND flash memory,” in *Proceedings of European Solid-State Device Research Conference (ESSDERC)*, Sep. 2017, pp. 6-9.

第5章 異種メモリの高信頼化技術

5.1 はじめに

本章では SCM および NAND 型フラッシュメモリを用いたストレージの高信頼化技術について述べる．不揮発性半導体メモリに発生するエラーは，SCM あるいは NAND 型フラッシュメモリによって異なる．ストレージコントローラに実装したエラー訂正回路によって，これらの不揮発性半導体メモリに生じたエラーを訂正する．ここでは SCM および NAND 型フラッシュメモリを用いたハイブリッドストレージに強度の異なるエラー訂正符号（error-correcting code，ECC）を適用する．高速にランダムエラーを訂正する Bose-Chaudhuri-Hocquenghem（BCH）符号を SCM および NAND 型フラッシュメモリに適用する．また NAND 型フラッシュメモリの微細化および多値化が進むにつれ信頼性が低下するため，BCH 符号より訂正能力の高い low-density parity-check（LDPC）符号も適用することが提案されている．しかし，ECC のためのメモリ読み出しおよび符号・復号時間がストレージ性能を低下させる．そのため，SCM と NAND 型フラッシュメモリを用いたハイブリッドストレージの信頼性と性能の関係をシステムレベルで理解する．

5.2 不揮発性メモリに適用するエラー訂正符号

不揮発性半導体メモリに生じるエラーはその種類によって異なる．高信頼化技術を議論する本章では SCM として ReRAM を，NAND 型フラッシュメモリとして MLC NAND 型フラッシュメモリを想定する．MLC NAND 型フラッシュメモリでは，書き込むメモリセルのワードライン（word-line，WL）に高い 20V を印加し，同一 WL 上の書き込まないメモリセルのビットライン（bit-line，BL）には V_{CC} を印加する．このとき制御ゲートと浮遊ゲート間の容量，浮遊ゲートと P 型半導体基板間の容量により，目的のメモリセル以外も書き込まれてしまう．これを容量結合による書き込みディスターブエラー（program-disturb error）と言う．また MLC NAND 型フラッシュメモリのセルを読み出す場合，読み出すメモリセルの WL に参照電圧（V_{REF}）を印加するが，読み出し対象外の WL には V_{READ} を印加する．V_{READ} が印加されたメモリセルは弱い書き込み状態となり間違って書き込まれることがある[1]．これを読み出しデ

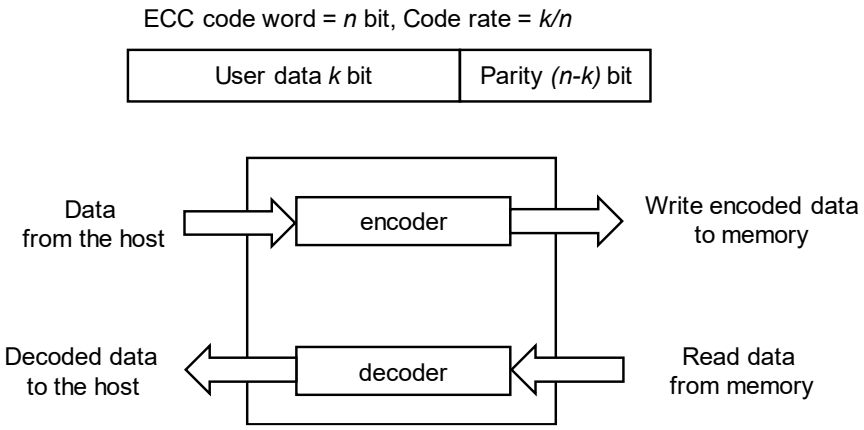


図 5.1 不揮発性半導体メモリの書き込み・読み出し時に適用する ECC 符号化および復号

ィスターブエラー（read-disturb error）と呼ぶ．さらに MLC NAND 型フラッシュメモリは，浮遊ゲートに電子を注入することでデータを記録する．しかし時間が経つにつれて浮遊ゲート内の電子が P 型半導体基板へ抜け落ちる．これをデータ保持エラー（data-retention error）と呼ぶ．また MLC NAND 型フラッシュメモリセルの書き換え回数が多くなると，図 2.8（d）で示したようにセルの酸化膜が劣化しエラーが発生しやすくなる．さらに NAND 型フラッシュメモリセルの微細化および多値化によって，これらのエラー発生率が高くなる．

一方 1T-1R メモリセル構成を持つ ReRAM [2]は NAND 型フラッシュメモリと異なり，書き込みおよび読み出し対象のセルのみに電圧を印加するため，書き込みディスターブエラーおよび読み出しディスターブエラーは発生しにくい．しかし書き換え回数（set/reset cycles）が増えるほどあるいはデータ保持中に，図 2.11 で示した ReRAM の導電性フィラメントの直径が拡大し酸素欠陥の密度が低下するためエラーが発生する[3][4].

不揮発性半導体メモリに生じたエラーは，ストレージコントローラ内に実装されたエラー訂正回路により削減，訂正する．誤り訂正符号（error-correcting code，ECC）をエラー訂正回路に実装する．ECC を用いたエラー訂正の流れを図 5.1 に示す．ホストからの書き込みリクエストデータは，ECC 回路で冗長ビットを付加する符号化（encode）を行い，不揮発性半導体メモリに書き込む．ホストからのデータ読み出しリクエストに対しては，不揮発性半導体メモリから読み出し，ECC 回路で復号（decode）したデータをホストに返す．一般に，ホストからの書き込みリクエストデータをユーザデータ（user data）と言い，符号化によりユーザデータにパリティを足したものを符号（code word）と言う．本論文ではユーザデータサイズ *k* bit，符号長（code length）*n* bit と表す．このときユーザデータに付加したパリティサイズは

（*n-k*）bit となる．また ECC の符号化率（code rate）は *k/n* と求められる．一般に符号化率が低いほど，つまりパリティサイズが大きいほど ECC の訂正能力は高い[5]．しかし符号化率が高いと，不揮発性半導体メモリにユーザデータを書き込むことができる容量が減少する．そのため本論文では符号化率 9/10 を許容すると仮定した．また，符号化は符号器（encoder）内で直ちにパリティビットを生成し付加するため符号化に要する時間は短い．一方，復号は復号器（decoder）内で複数の操作が必要なため，符号より長い時間を要する．したがって符号化と比較して，復号はストレージ性能に大きく影響する，ここで ECC の符号化および復号を含む，不揮発性半導体メモリへの書き込みおよびメモリからの読み出し時間は式（5.1）および式（5.2）で計算できる．

$$\text{Write time} = \text{memory write time} + \text{ECC encoding time} \tag{5.1}$$

$$\text{Read time} = (\text{memory read time} \times \#\text{ of } V_{REF} \times \#\text{ of pages}) + \text{ECC decoding time} \tag{5.2}$$

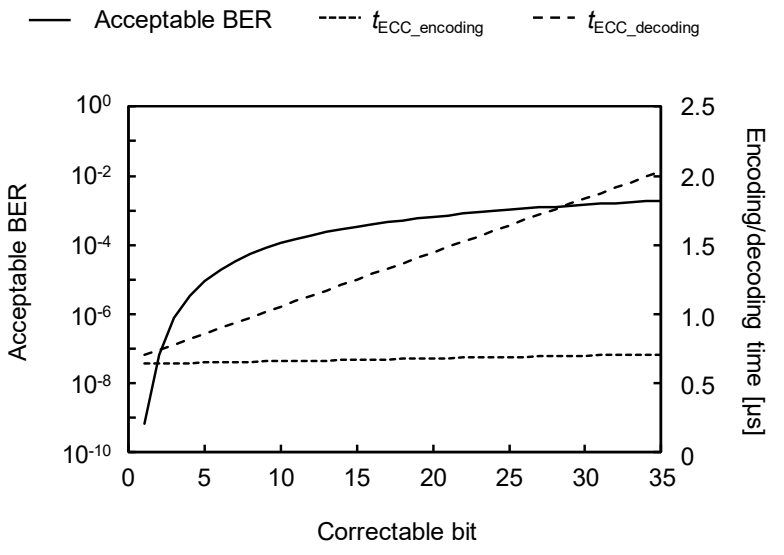
ここで不揮発性半導体メモリの書き込みおよび読み出し時間は表 2.1 および表 2.2 で定義される．式（5.2）に示した読み出すしきい値（V_{REF}）の数およびページ数は以下に示す ECC の種類によって異なる．符号化時間（ECC encoding time）および復号時間（ECC decoding time）もまた，ECC の種類や ECC の強度によって異なる．

NAND 型フラッシュメモリでは，高速にランダムエラーを訂正する Bose–Chaudhuri–Hocquenghem（BCH）符号が広く用いられている．整数 *m* に対し BCH 符号の符号長 *n* およびユーザデータサイズ *k* は式（5.3）および式（5.4）で定義される[6].

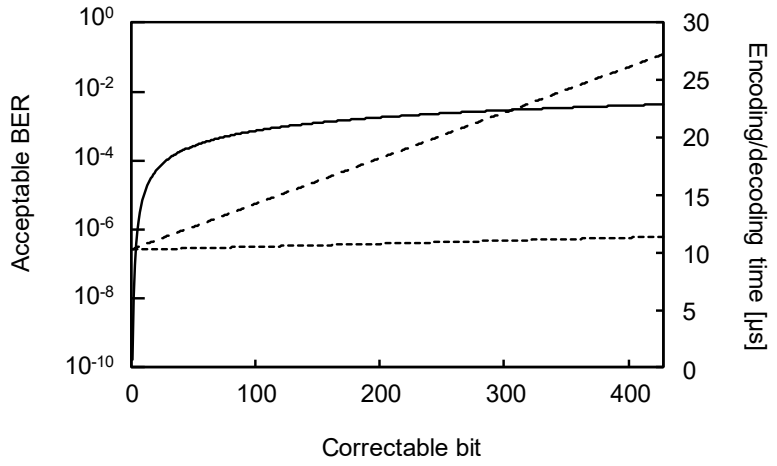
$$n \geq 2^m - 1 \tag{5.3}$$

$$k \geq 2^m - 1 - mt \tag{5.4}$$

m はガロア体の次数であり，このとき符号長 *n* bit のうち *t* bit 以下のエラーを訂正することができる．これらの値を用い BCH 符号は（*n*, *k*, *t*）と表される．BCH 符号は付加するパリティビットを増やし符号長を長くすることで，エラー訂正能力を高めることができる．ユーザデータサイズを長くすることでもまた，同じ符号化率でもエラー訂正能力を高めることができる．図 5.2 に，M-SCM および S-SCM のユーザデータサイズ 512 Byte および MLC NAND 型フラッシュメモリのユーザデータサイズ 8 KByte に対する訂正可能 BER（acceptable BER）を示す．上で述べたように，同じユーザデータサイズであればパリティビットを増やすほど，訂正可能 BER は増加する．また同じ符号化率であれば， SCM のユーザデータサイズ（512 Byte）と比較して，MLC NAND 型フラッシュメモリの大きなユーザデータサイズ（8 KByte）



(a)



(b)

図 5.2 BCH 符号による訂正可能 BER.（a）ユーザデータサイズ 512 Byte,（b）ユーザデータサイズ 8 KByte [13]

が高い訂正能力を示す. SCM のエラーを訂正するために, 13×t bit のパリティが必要となる. 同様に，NAND 型フラッシュメモリのエラーを訂正するために，17×t bit のパリティが必要となる.

BCH 符号はガロア体に基づく生成多項式を用いて符号化を行い，シフトレジスタで実現できるため高速である．一方 BCH 符号の復号は初めに誤りを含むシンドロームを生成（Syndrome generation）し，誤り位置多項式を求める．誤り位置多項式の解法として，Berlekamp-Massey（BM）やユークリッドの互除法がある．このように BCH 符号の復号は符号化と比較して時間を要するため，図 5.3 に示すような並列化により演算の高速する工夫が

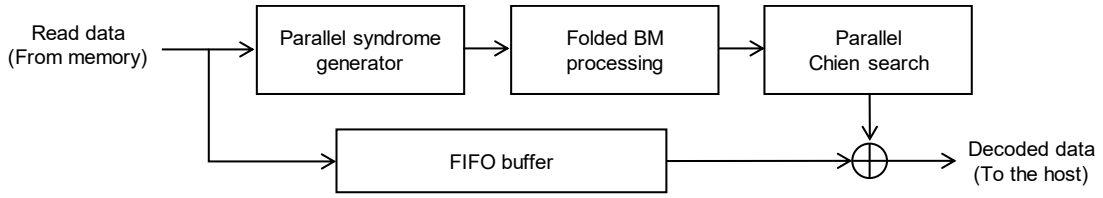


図 5.3 BCH 符号を用いた復号 [7]

行われている[7]. 図 5.3 に示す回路を用いる BCH 符号の復号では，シンドローム生成，BM，チェンサーチ（Chien search）を連続して行う．チェンサーチは誤り位置多項式の係数から，誤り位置を求める．その後，メモリから読み出したデータと，チェンサーチを行い誤りパターンの判明したデータを用いて誤りを訂正する．図 5.3 に示す回路を用いる BCH 符号の復号時間は式（5.5）で表される．

$$\text{BCH ECC decoding time} = \left\{ \frac{mt}{p} + \frac{(t+1)f}{2} + (t-1) + \frac{n}{p} \right\} / F \tag{5.5}$$

ここで，p は並列数，f はフォールディングファクターである．F は ECC 復号回路の動作周波数である．参考文献[7]と同様に，p, f, F はそれぞれ 32, 12, 200 MHz とする．ハミング符号を基にした single error correction, double error detection（SECDED）は ECC 符号長内の 1 ビットしか訂正できないため，SCM にも BCH を適用する．

BCH 符号はハード情報のみを必要とする．BCH 符号による MLC NAND 型フラッシュメモリの復号は，図 5.4 に示すように Lower page，Upper page それぞれで“0”，“1”情報を必要とする．ページ読み出しに要する時間は，Lower page，Upper page を読み出して復号するとき，表 2.1 に示した通りそれぞれ 36 μsec，52 μsec だけ要する．また，BCH 符号による SCM の復号は ReRAM の場合，図 5.5 のように低抵抗状態“1”か高抵抗状態“0”かをメモリセルを読み出して判断できる．そのため ReRAM の 1 セクタ（= 512 Byte）を復号するとき，表 2.2 に示した M-SCM の読み出し時間 0.1 μsec あるいは S-SCM の読み出し時間 1 μsec だけ要する．

また，NAND 型フラッシュメモリは多値化および微細化により BER が高くなるため，BCH 符号より訂正能力の高い low-density parity-check（LDPC）符号を用いることが提案されている[8]. LDPC 符号は NAND 型フラッシュメモリから読み出したデータを用いて対数尤度比（log-likelihood ratio，LLR）を計算し，LLR を用いて復号を行う．LLR はメモリから読み出したデータを y，正解データを x として式（5.6）で定義される．

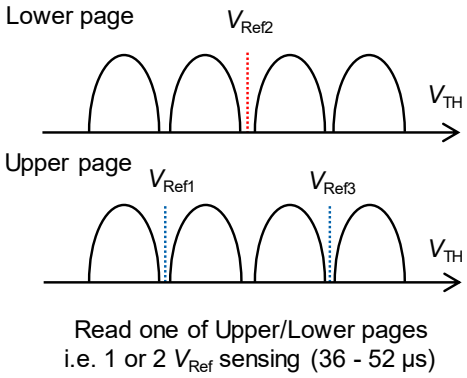


図 5.4 BCH 符号用いた復号時の MLC NAND 型フラッシュメモリの読み出し動作 [10]

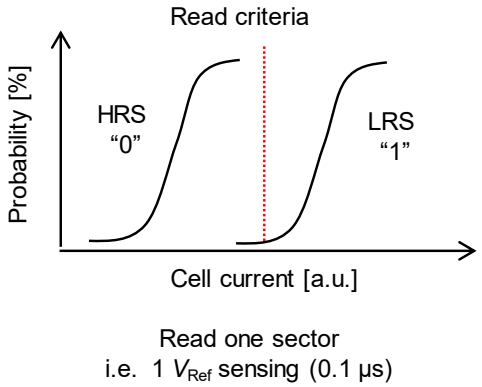


図 5.5 BCH 符号用いた復号時の ReRAM の読み出し動作

$$\mathrm{LLR}(y) = \ln\frac{p(x=0 \mid y)}{p(x=1 \mid y)} \tag{5.6}$$

読み出したデータの BER で表すと LLR は次の式（5.7）および式（5.8）のようになる.

$$\mathrm{LLR}(0) = \ln\frac{1-BER}{BER} \tag{5.7}$$

$$\mathrm{LLR}(1) = \ln\frac{BER}{1-BER} \tag{5.8}$$

図 5.6 のように復号を行うたびにエラーが残っているか確認し，エラーが残っていればエラーがなくなるまで復号を繰り返す. この繰り返し回数を LDPC 符号の復号の繰り返し（iteration）L と言う．Soft-decoding LDPC 符号は，ハード情報である“0”あるいは“1”を得るほか，図 5.7 のようにソフト情報である詳細な V_{TH} 情報を必要とする. ソフト情報を得ることによって LLR をより詳細に計算でき，エラー訂正能力が高くなる．Soft-decoding LDPC 符号は高いエラー訂

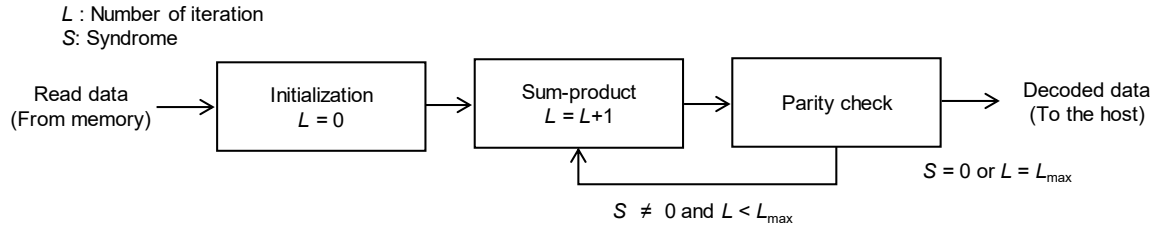
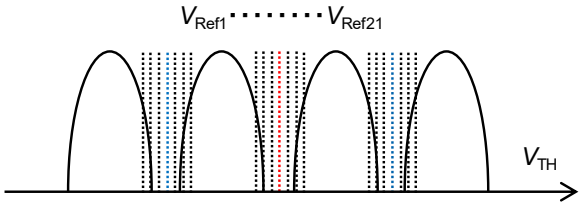


図 5.6 LDPC 符号を用いた復号 [9]



x7 sequential reads
i.e. 21 V_{Ref} sensing (616 µs)

図 5.7 Soft-decoding LDPC 符号用いた復号時の MLC NAND 型フラッシュメモリの読み出し動作 [10]

正能力を持つが，BCH 符号と比較してソフト情報を得るためのメモリ読み出しに時間がかかる．ユーザデータサイズ n = 8 KByte，符号化率 9/10 のとき，LDPC 復号時間は式（5.9）で決まる[9].

$$\text{LDPC ECC decoding time} = 1.58\ \mu sec \times L \tag{5.9}$$

イタレーション回数（L）が増えると復号時間が長くなる．その結果，式（5.2）の Soft-decoding LDPC の読み出し時間は BCH よりはるかに長くなる．従来研究[10]によると， MLC NAND 型フラッシュのみを用いたストレージに Soft-decoding LDPC を用いると，読み出しの多いアプリケーションに対して，BCH と比べて 2 倍性能が悪化する．

このため NAND 型フラッシュメモリの読み出し回数を削減した LDPC 符号が提案されている．図 5.8 に示す Error-prediction（EP-）LDPC without（w/o）upper/lower cells [11]では，NAND 型フラッシュメモリからの読み出しは 1 ワードラインのみとなる．MLC NAND 型フラッシュメモリでは 1 ワードラインに含まれる Upper page および Lower page を読み出す．EP-LDPC w/o upper/lower cells では LLR を計算する際に 1 ワードラインを読み出したデータのほか，メモリ出荷時に測定しコントローラに保存したデータを用いる．テーブルに保存した NAND 型

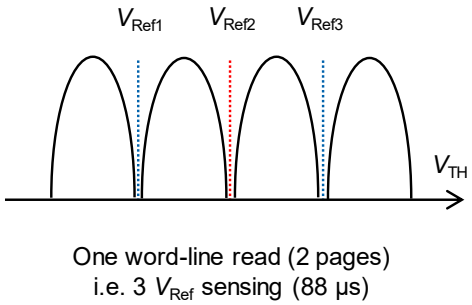


図 5.8 Error-prediction（EP-）LDPC without（w/o）upper/lower cells 符号を用いた復号時の MLC NAND 型フラッシュメモリの読み出し動作 [10]

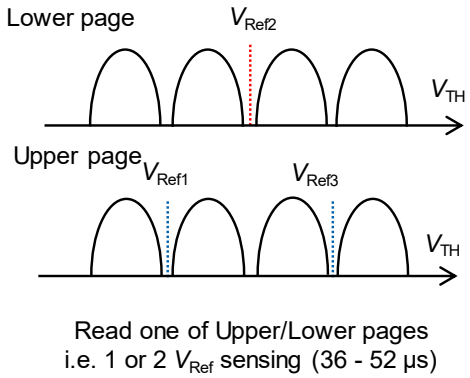
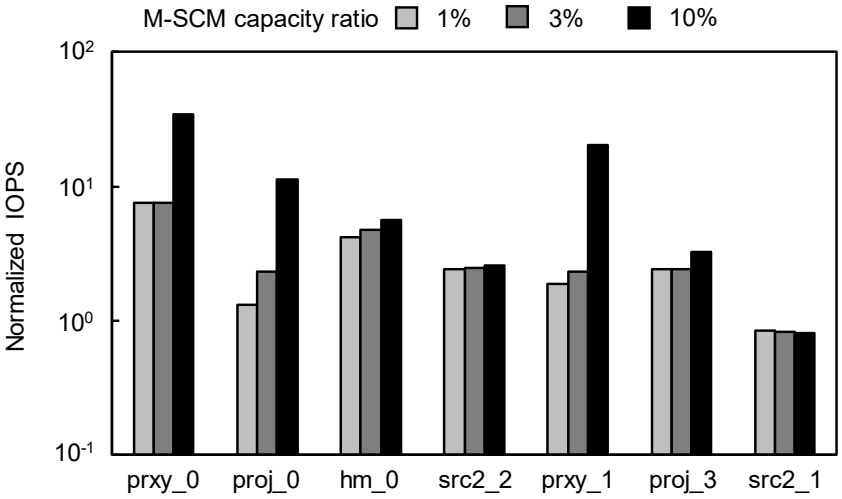


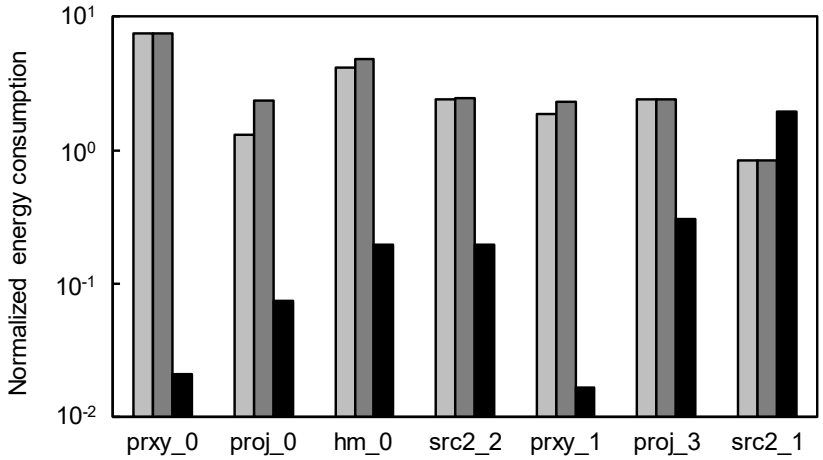
図 5.9 Quick-LDPC 符号を用いた復号時の MLC NAND 型フラッシュメモリの読み出し動作 [10]

フラッシュメモリ書き換え回数，データ保持時間，およびセル間干渉データを用いて BER を予測する．EP-LDPC w/o upper/lower cells の復号に要する時間は Soft-decoding LDPC と同じだが，NAND 型フラッシュメモリセルを読み出す回数が減少するため，式（5.2）のうちメモリ読み出し時間が減少する．

図 5.9 に示す Quick-LDPC [11]では，NAND 型フラッシュメモリからの読み出しは，ホストから読み出しリクエストを受けた対象のページのみとなる．MLC NAND 型フラッシュメモリでは Upper page のみあるいは Lower page のみを読み出す．EP-LDPC w/o upper/lower cells と同様，Quick-LDPC では LLR を計算する際に，読み出した 1 ページのデータのほかメモリ出荷時に測定しコントローラに保存したデータを用いる．テーブルに保存した NAND 型フラッシュメモリ書き換え回数およびデータ保持時間を用いて BER を予測する．Quick-LDPC の復号に要する時間は Soft-decoding LDPC と同じだが，NAND 型フラッシュメモリセルを読み出す回数が減少するため，式（5.2）のうちメモリ読み出し時間が減少する．



(a)



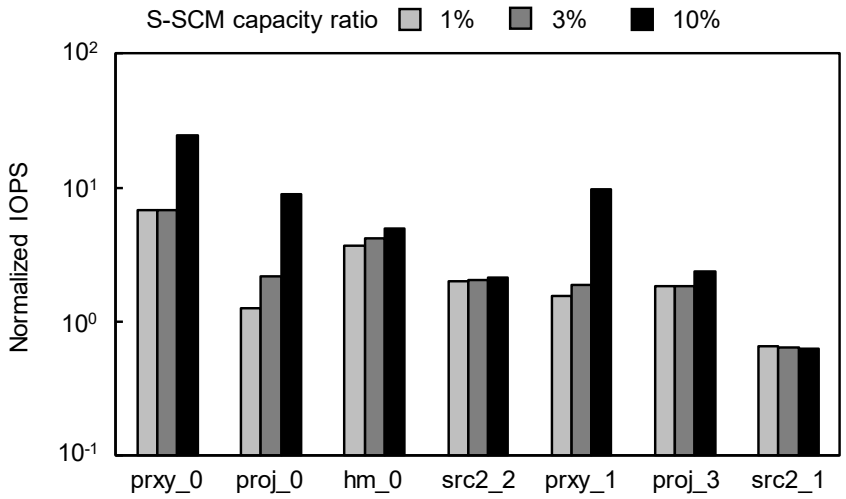
(b)

図 5.10 M-SCM に 1 ビット訂正する BCH ECC を適用したときのハイブリッドストレージ性能．（a）IOPS 性能，（b）消費エネルギー [13]

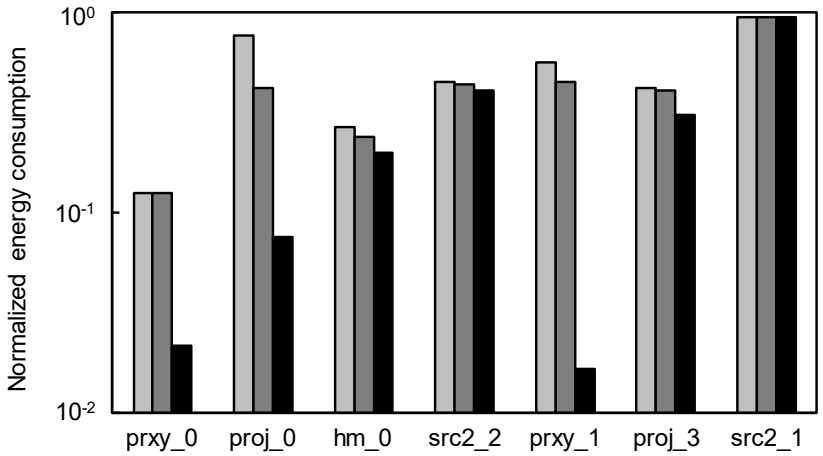
LDPC 符号は BCH 符号と比較してはるかに復号時間が長いため，本論文では SCM には適用しない．

5.3 BCH ECC による SCM の高信頼化

ハイブリッドストレージは，アプリケーションに依存した性能を示す[12]．図 5.10 および図 5.11 は，訂正能力 1 bit の BCH をそれぞれ M-SCM，S-SCM に適用したときの，M-SCM および S-SCM の読み出し・書き込み時間，M-SCM および S-SCM 容量比，アプリケーションを



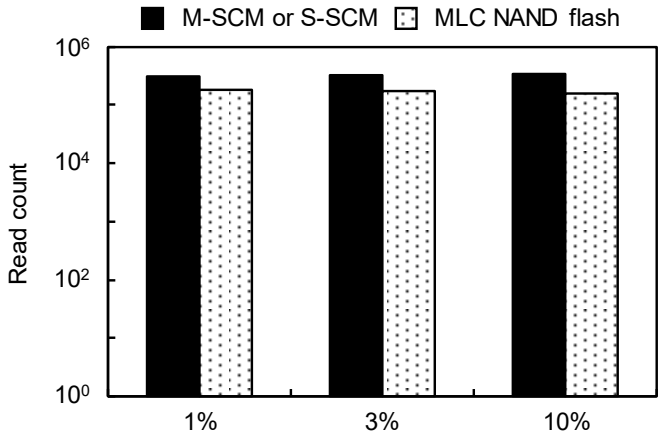
(a)



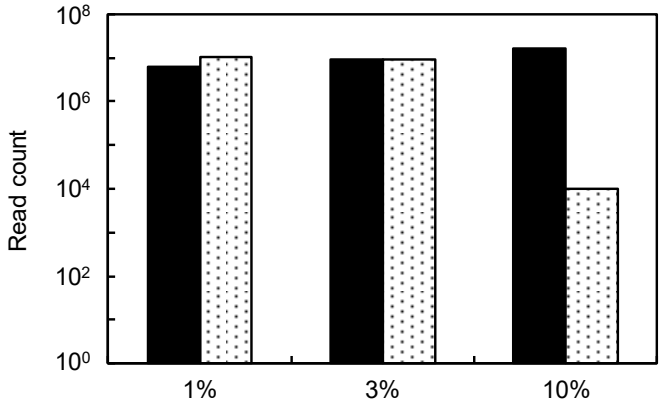
(b)

図 5.11 S-SCM に 1 ビット訂正する BCH ECC を適用したときのハイブリッドストレージ性能．（a）IOPS 性能，（b）消費エネルギー [13]

変化させた（a）IOPS 性能および（b）消費エネルギーを示す[13]．図 2.14 の各カテゴリから一アプリケーションを選択する．M-SCM あるいは S-SCM を大容量（10%）用いるとハイブリッドストレージ性能は向上する．ハイブリッドストレージの IOPS 性能は，NAND 型フラッシュメモリのみを用いたストレージの性能で規格化する．また図 5.10（a）に示すように，0.1 μsec の M-SCM を 1%だけ追加すると，prxy_0（write-hot-random）アプリケーションのハイブリッドストレージ性能は，NAND 型フラッシュメモリのみを用いたストレージと比較して 7.5 倍向上する．M-SCM が 10%追加されると，性能は 35 倍向上する．prxy_0 のホット・ランダムデータは M-SCM に保存されるため，NAND 型フラッシュメモリの GC 頻度が減少



(a)



(b)

図 5.12 M-SCM あるいは S-SCM および MLC NAND 型フラッシュメモリからの読み出し回数．（a）prxy_0，（b）prxy_1

する．prxy_0 と同様に，prxy_1（read-hot-random）のハイブリッドストレージ性能は大容量の M-SCM あるいは S-SCM によって向上する．ホット・ランダムデータが M-SCM あるいは S-SCM から直接読み出されるため，読み出しアクセス時間が減少する．上書き動作が M-SCM あるいは S-SCM で行われるため，大容量の M-SCM あるいは S-SCM は proj_0（write-hot-sequential）のハイブリッドストレージ性能が向上する．コールドなアプリケーション（hm_0，src2_2，proj_3，src2_1）のハイブリッドストレージではデータが多く読み出し・上書きされないため，ストレージ性能は M-SCM あるいは S-SCM 容量によって大幅に向上しない．一方で図 5.10（b）および図 5.11（b）に示すように性能向上に反比例して消費エネルギーは削減する．

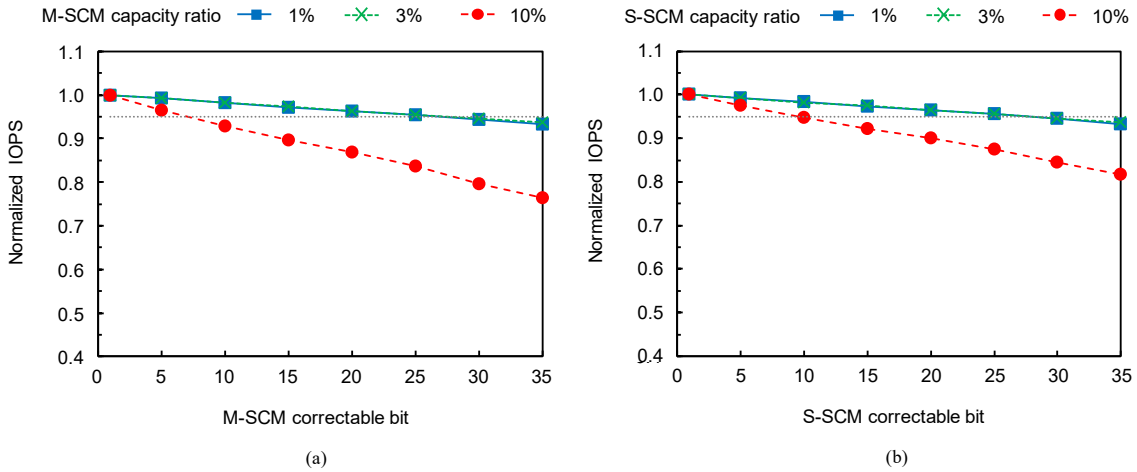


図 5.13 M-SCM あるいは S-SCM に BCH ECC を適用したときの，prxy_0 アプリケーションに対するハイブリッドストレージ性能．（a）M-SCM（scenario 1），（b）S-SCM（scenario 2）[13]

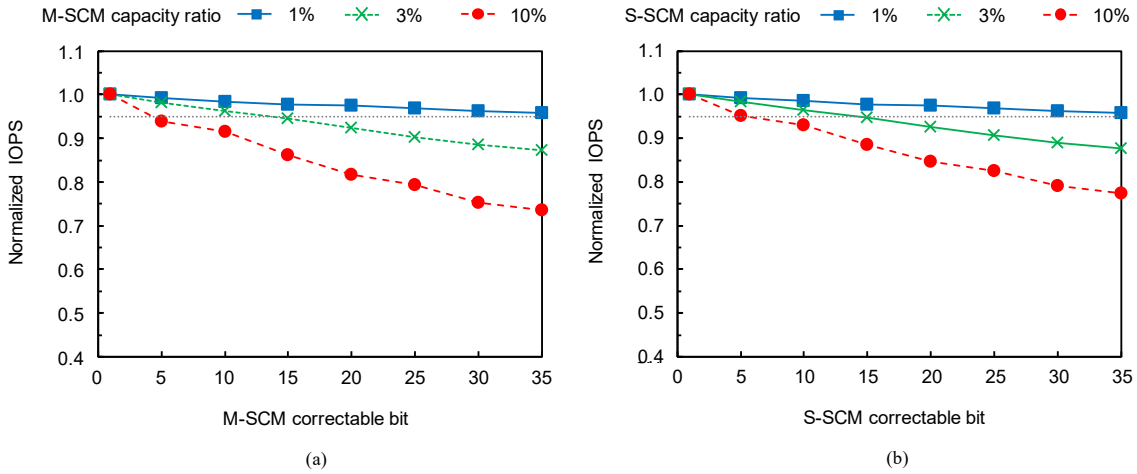


図 5.14 M-SCM あるいは S-SCM に BCH ECC を適用したときの，proj_0 アプリケーションに対するハイブリッドストレージ性能．（a）M-SCM（scenario 1），（b）S-SCM（scenario 2）

本節では，M-SCM あるいは S-SCM および NAND 型フラッシュメモリを用いたハイブリッドストレージの SCM の適切な ECC 強度を評価した．このとき，NAND 型フラッシュメモリの ECC 強度は，符号化率 9/10 の BCH 符号を一律に適用する．ハイブリッドストレージの M-SCM あるいは S-SCM に適用可能な BCH の強度を次のように決めた．図 5.10 および図 5.11 に示した，M-SCM あるいは S-SCM の訂正可能ビット数が 1 の時の IOPS 性能を基準として，性能劣化が 5%以内で最大の BCH 訂正可能ビットとする．図 5.12 に M-SCM あるいは S-SCM 容量を変えた時の，M-SCM あるいは S-SCM および MLC NAND 型フラッシュメモリからの読み出し回数を示す．図 5.12（a）に示す prxy_0 アプリケーションは書き込みの多いアプリケ

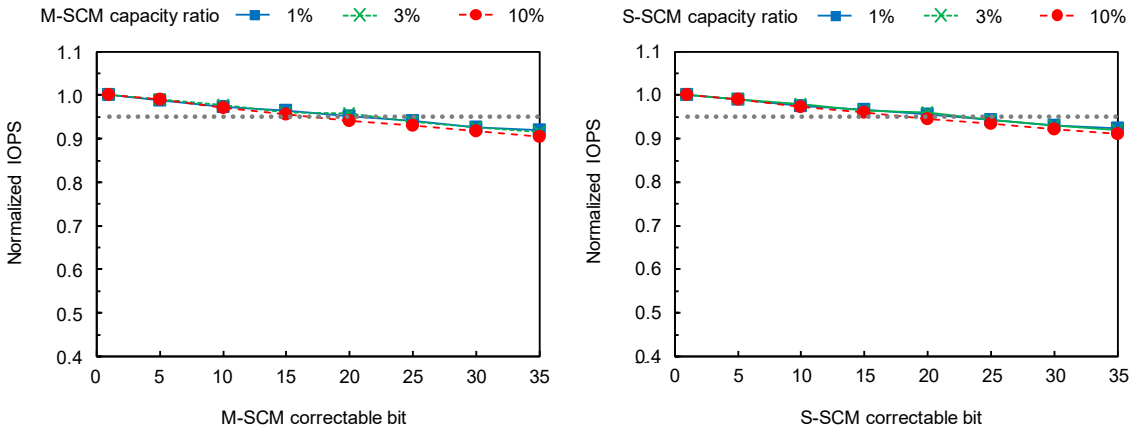


図 5.15 M-SCM あるいは S-SCM に BCH ECC を適用したときの，hm_0 アプリケーションに対するハイブリッドストレージ性能．（a）M-SCM（scenario 1），（b）S-SCM（scenario 2）

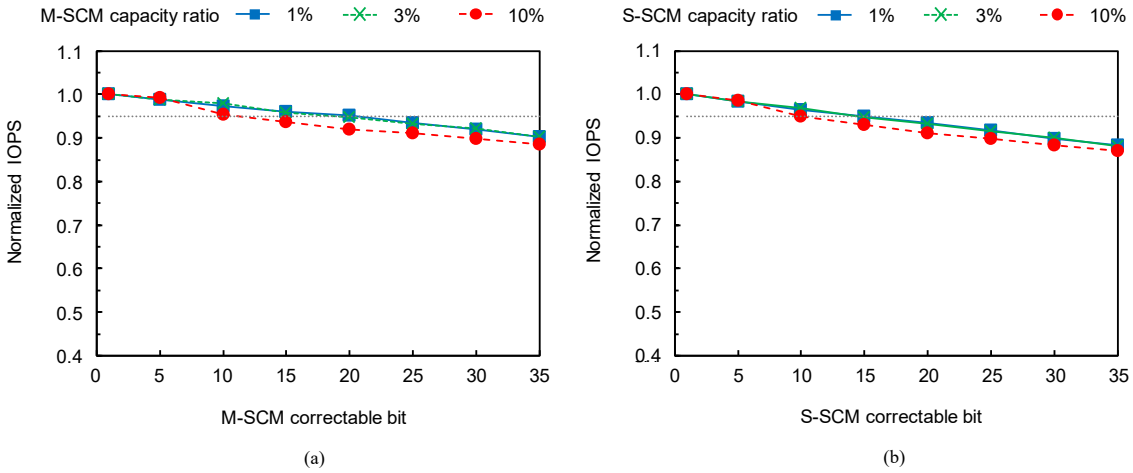


図 5.16 M-SCM あるいは S-SCM に BCH ECC を適用したときの，src2_1 アプリケーションに対するハイブリッドストレージ性能．（a）M-SCM（scenario 1），（b）S-SCM（scenario 2）

ーションであるため，M-SCM あるいは S-SCM 容量を増やしても，M-SCM あるいは S-SCM および MLC NAND 型フラッシュメモリからの読み出し回数はほぼ変わらない．これに対して図 5.12（b）に示す読み出しの多い prxy_1 アプリケーションは，M-SCM あるいは S-SCM 容量を増やすほど M-SCM あるいは S-SCM からの読み出し回数が増加し，MLC NAND 型フラッシュメモリからの読み出し回数は減少する．

図 5.13-図 5.19 に，7 ストレージアプリケーションの M-SCM あるいは S-SCM 訂正可能ビット数による性能劣化率を示す．図 5.13 および図 5.14 に示すように prxy_0 および proj_0 アプリケーションに対して，M-SCM あるいは S-SCM の訂正可能ビット数が増えると，より多

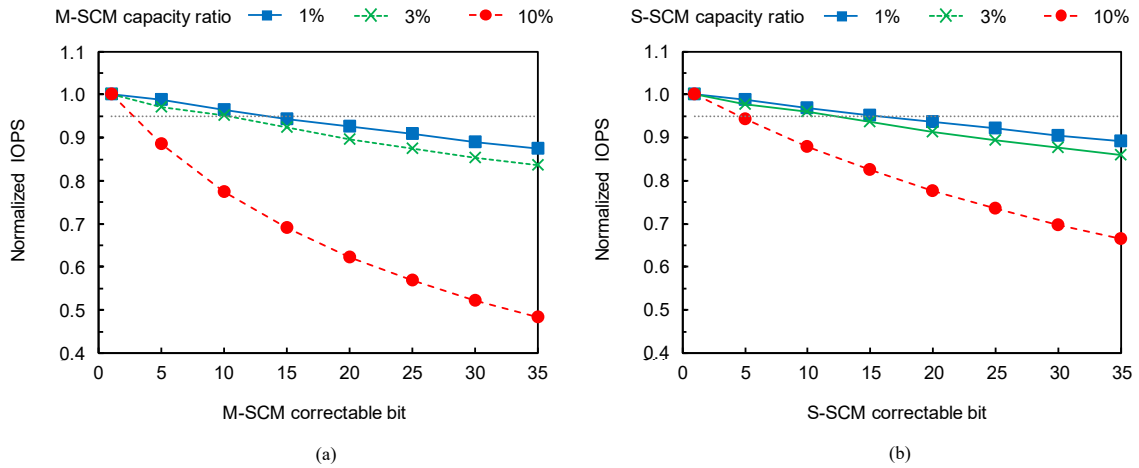


図 5.17 M-SCM あるいは S-SCM に BCH ECC を適用したときの prxy_1 アプリケーションに対するハイブリッドストレージ性能．（a）M-SCM（scenario 1），（b）S-SCM（scenario 2）[13]

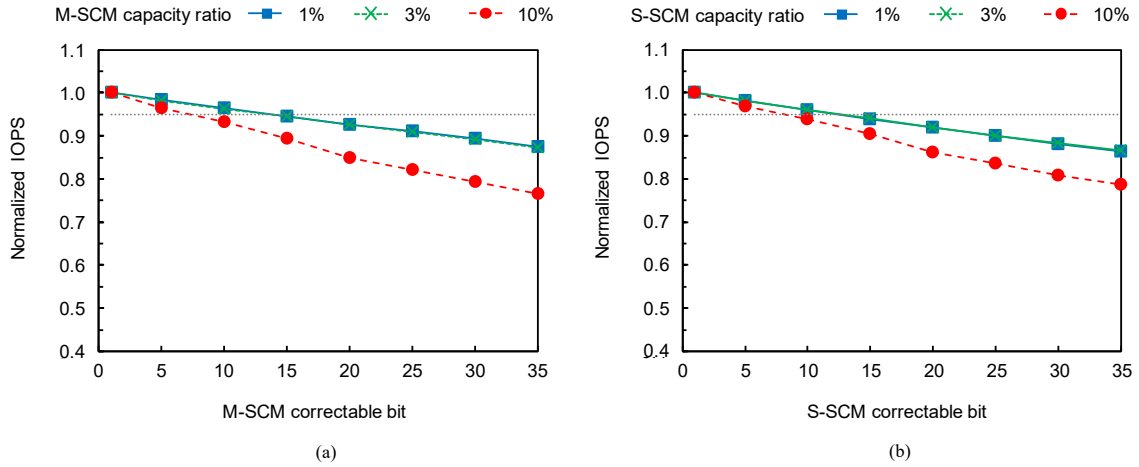
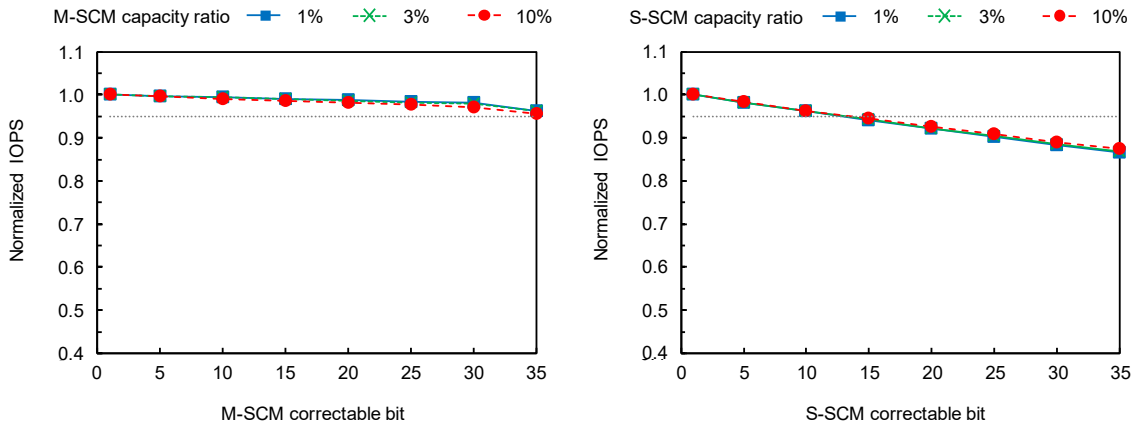


図 5.18 M-SCM あるいは S-SCM に BCH ECC を適用したときの proj_3 アプリケーションに対するハイブリッドストレージ性能．（a）M-SCM（scenario 1），（b）S-SCM（scenario 2）

くのパリティビットが M-SCM あるいは S-SCM のユーザデータに付加され（図 5.1），ユーザデータを書き込める M-SCM あるいは S-SCM 容量が削減する．M-SCM あるいは S-SCM から NAND 型フラッシュメモリへのデータ evict が頻繁に発生し，ハイブリッドストレージの性能が低下する．さらに，M-SCM あるいは S-SCM 容量が大きいと M-SCM あるいは S-SCM へのアクセスが増えるため，性能が急激に劣化する．図 5.13（a）および図 5.13（b）を比較すると，読み出し・書き込み時間が 0.1 μsec の M-SCM（scenario 1）による性能低下は，1 μsec の S-SCM（scenario 2）と比較して大きいことが明らかになった．これは，読み出し・書き込み時間が 0.1 μsec の M-SCM は，BCH の復号時間が性能に大きく影響するからである．しかし，読み出し・書き込み時間が長い S-SCM は，性能への影響は削減される．



(a) (b)

図 5.19 M-SCM あるいは S-SCM に BCH ECC を適用したときの src2_1 アプリケーションに対するハイブリッドストレージ性能．（a）M-SCM（scenario 1），（b）S-SCM（scenario 2）

図 5.15 に示す hm_0 アプリケーションは M-SCM あるいは S-SCM によって性能が向上するが，書き込みの多いアプリケーションであるため，M-SCM あるいは S-SCM に適用す BCH ECC の強度による性能低下率は少ない．

図 5.17 は，prxy_1（read-hot-random）の性能劣化の傾向が，prxy_0（write-hot-random）と同様の傾向を示す．prxy_1 は読み出しの多いアプリケーションであるため，ハイブリッドストレージの性能は，式（5.2）に示す読み出し時間が強く影響する．

図 5.16，図 5.19 に示す src2_2，src2_1 アプリケーションは，図 5.10 および図 5.11 で示したように M-SCM あるいは S-SCM による性能向上の効果が他のアプリケーションと比較して少ない．つまり M-SCM あるいは S-SCM へのアクセスが少なく，これらの高速なメモリが活用されていないためである．このため M-SCM あるいは S-SCM 容量に対して，訂正能力の異なる BCH ECC を適用したときのハイブリッドストレージ性能の低下は似た傾向を示す. 図 5.18 に示す proj_3 アプリケーションも M-SCM あるいは S-SCM による性能向上が低いが，読み出しの多いアプリケーションであるため，高い訂正能力を持つ BCH ECC を適用すると式 (5.5) で示した復号化時間のためにハイブリッドストレージ性能が低下する．

表 5.1 および表 5.2 は，M-SCM および S-SCM に対する訂正可能ビットおよび許容できる BER の概要を示す．より大きい M-SCM あるいは S-SCM 容量は，M-SCM あるいは S-SCM 訂正可能ビットを減少させる．また，S-SCM の長い読み出し・書き込み時間（1 μsec）は IOPS 増加率が平たんになるため，許容できる S-SCM の ECC 強度を低減させる．さらに，IOPS（図

表 5.1 ハイブリッドストレージの M-SCM（scenario 1）に適用可能な BCH ECC 強度 [13]

Application	M-SCM 1%	M-SCM 3%	M-SCM 10%
prxy_0	25 bit	25 bit	5 bit
proj_0	> 35 bit	10 bit	1 bit
hm_0	20 bit	20 bit	15 bit
src2_2	20 bit	15 bit	10 bit
prxy_1	10 bit	10 bit	1 bit
proj_3	10 bit	10 bit	5 bit
src2_1	> 35 bit	> 35 bit	> 35 bit

表 5.2 ハイブリッドストレージの S-SCM（scenario 2）に適用可能な BCH ECC 強度 [13]

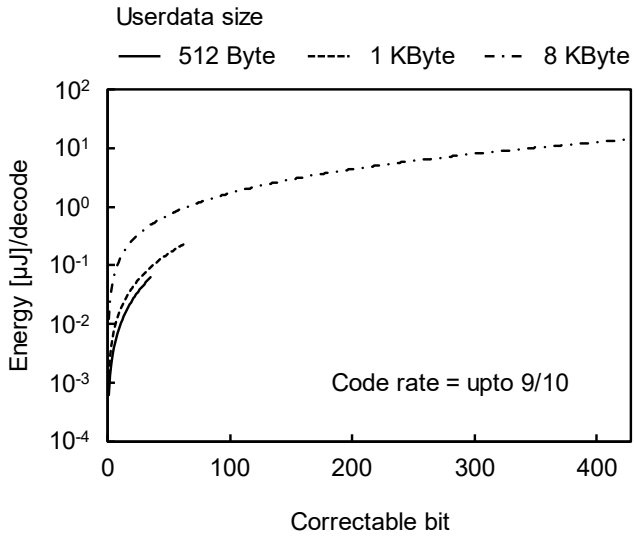
Application	S-SCM 1%	S-SCM 3%	S-SCM 10%
prxy_0	25 bit	25 bit	5 bit
proj_0	>35 bit	10 bit	5 bit
hm_0	20 bit	20 bit	15 bit
src2_2	10 bit	10 bit	10 bit
prxy_1	15 bit	10 bit	1 bit
proj_3	10 bit	10 bit	5 bit
src2_1	10 bit	10 bit	10 bit

5.10，図 5.11）と BCH ECC による IOPS 劣化率（図 5.13-図 5.19）は，M-SCM あるいは S-SCM へのアクセスデータ量（図 5.12）を通じて関係がある．その結果大容量の M-SCM あるいは S-SCM を用いるハイブリッドストレージにでは，M-SCM あるいは S-SCM にアクセスが頻繁に起こるため，訂正ビット数の少ない弱い BCH ECC が適切である．

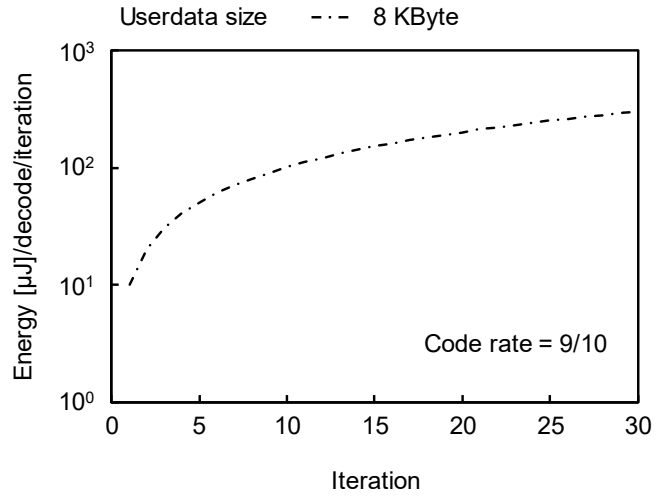
5.4 BCH および LDPC ECC による NAND 型フラッシュメモリの高信頼化

本節では，性能および消費エネルギーの観点で，NAND 型フラッシュメモリに適用できる ECC 強度を解析する[13]．このとき，SCM はユーザデータ 512 Byte に対して一律に符号化率 9/10 の BCH 符号を適用する．

第 5.2 節で説明したように，LDPC 符号の問題はソフト情報を得るための長い読み出し時間である．また長い復号時間のために，LDPC デコーダによる消費エネルギーは BCH よりはるかに大きい. BCH およびハードセンシング LDPC による復号時のエネルギー消費を図 5.20 に示す．BCH の復号時の消費エネルギーは訂正可能ビット数によって増加し，LDPC 復号時の



(a)



(b)

図 5.20 ECC 復号時の消費エネルギー．(a) 最大符号化率 9/10 の BCH 符号，(b) 符号化率 9/10 の LDPC 符号 [13]

消費エネルギーはイタレーション回数によって増加する．NAND 型フラッシュメモリのユーザデータ 8 KByte，符号化率 9/10 とするとき，イタレーション回数 10 回の時の LDPC 復号時の消費エネルギーは BCH の 10 倍以上となる．したがって，LDPC 符号は，ハイブリッドストレージの性能および消費エネルギーに大きな影響を与える．

図 5.21-図 5.27 は 7 ストレージアプリケーションに対し，M-SCM および MLC NAND 型フ

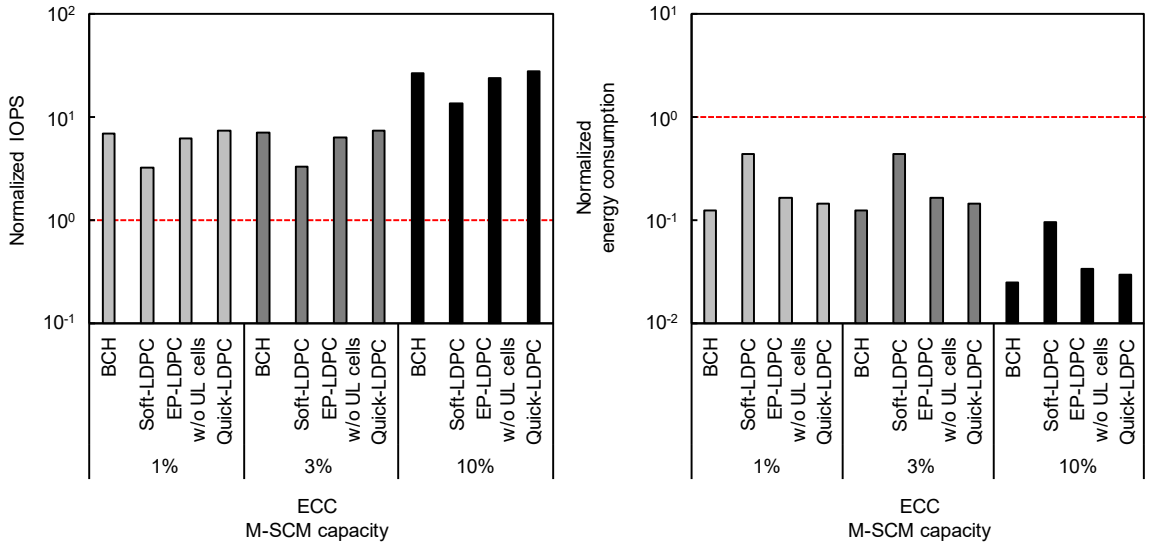


図 5.21 NAND 型フラッシュメモリに BCH あるいは LDPC ECC を適用したときの，prxy_0 アプリケーションに対するハイブリッドストレージの（a）IOPS 性能，（b）消費エネルギー [13]

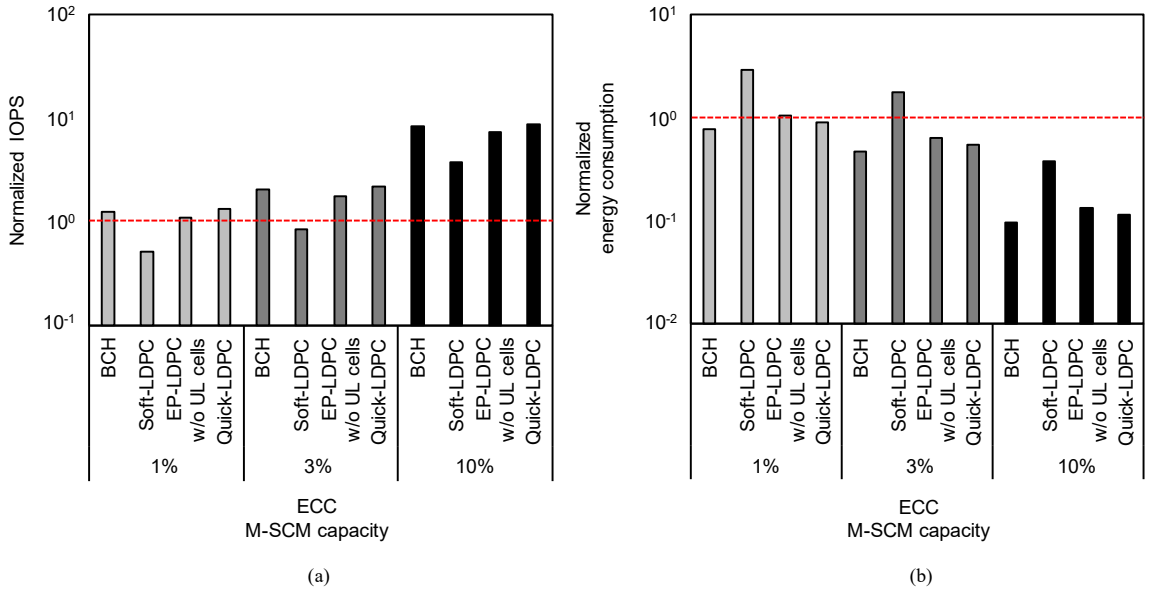


図 5.22 NAND 型フラッシュメモリに BCH あるいは LDPC ECC を適用したときの，proj_0 アプリケーションに対するハイブリッドストレージの（a）IOPS 性能，（b）消費エネルギー

ラッシュメモリを用いたハイブリッドストレージの MLC NAND 型フラッシュメモリに BCH あるいは LDPC 符号を適用したときの IOPS 性能および消費エネルギーを示す．このとき，M-SCM には符号化率 9/10 の BCH 符号を一律に適用した．またハイブリッドストレージの IOPS および消費エネルギーは，MLC NAND 型フラッシュメモリのみを用いたストレージで

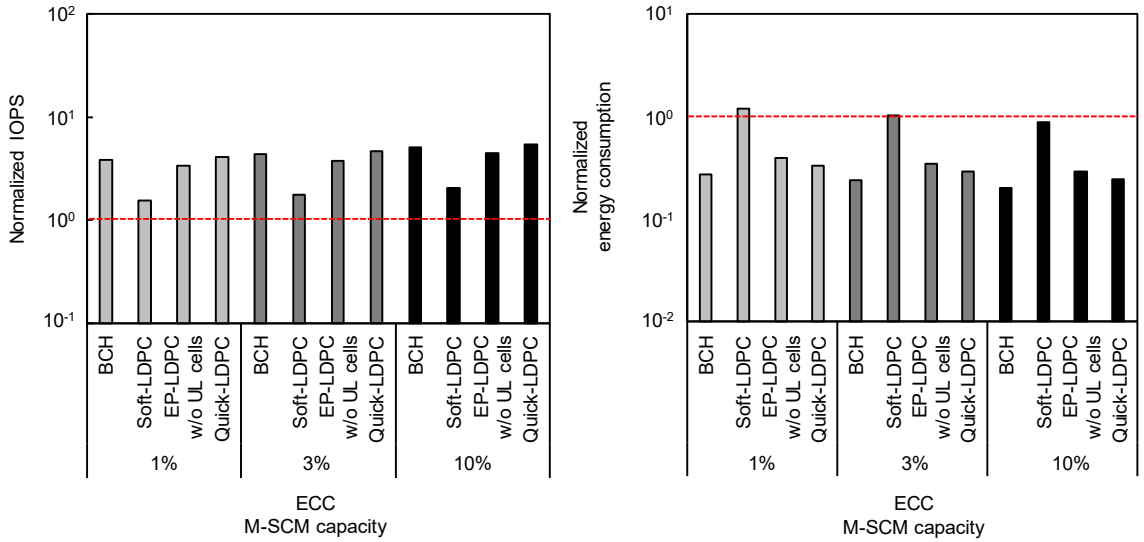


図 5.23 NAND 型フラッシュメモリに BCH あるいは LDPC ECC を適用したときの，hm_0 アプリケーションに対するハイブリッドストレージの（a）IOPS 性能，（b）消費エネルギー

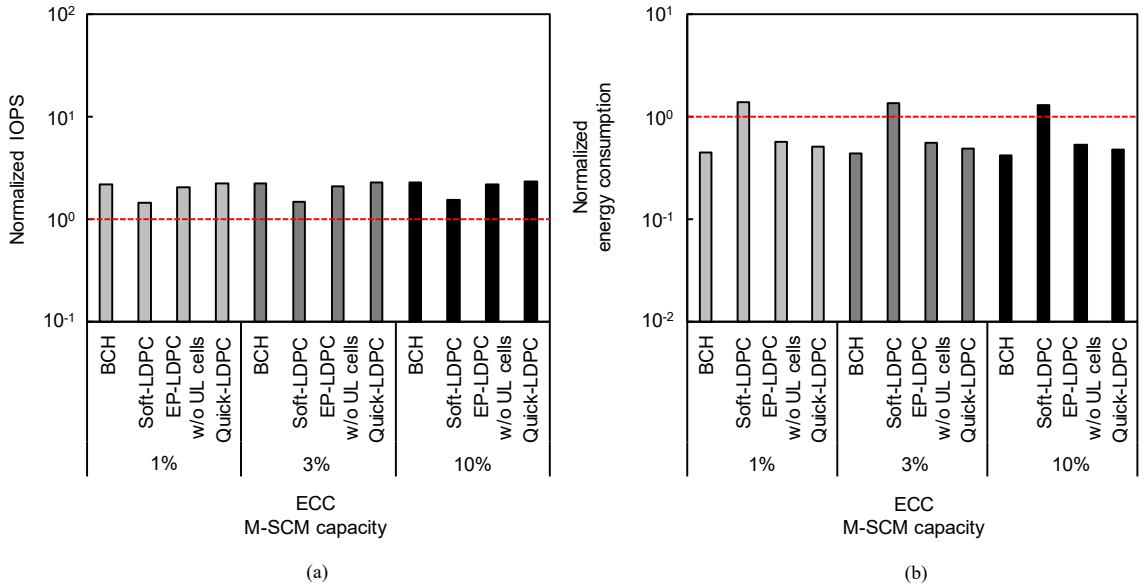


図 5.24 NAND 型フラッシュメモリに BCH あるいは LDPC ECC を適用したときの，src2_2 アプリケーションに対するハイブリッドストレージの（a）IOPS 性能，（b）消費エネルギー

規格化した．BCH 符号と比較すると，MLC NAND 型フラッシュメモリに適用した Soft-decoding LDPC は性能を低下させ，消費エネルギーを増大させる．より少ない読み出し回数を実現した EP-LDPC w/o UP cells および Quick-LDPC などの高度な LDPC 符号は，Soft-decoding LDPC と比較して高い性能と低い消費エネルギーを示す．M-SCM 容量比が大きくなると，MLC NAND 型フラッシュメモリへのアクセス頻度が減少する．したがって，MLC NAND 型

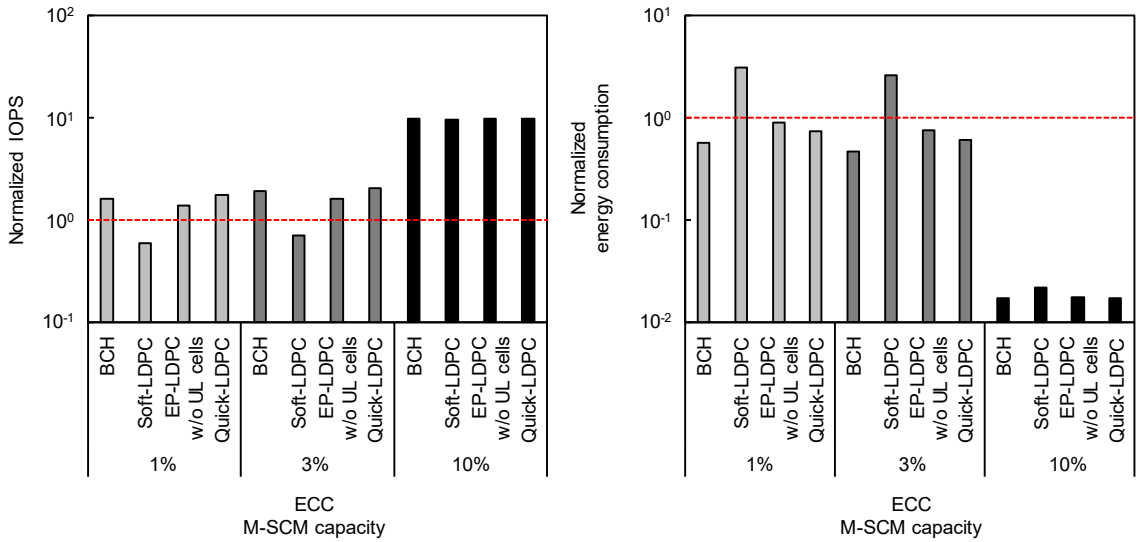


図 5.25 NAND 型フラッシュメモリに BCH あるいは LDPC ECC を適用したときの，prxy_1 アプリケーションに対するハイブリッドストレージの（a）IOPS 性能，（b）消費エネルギー [13]

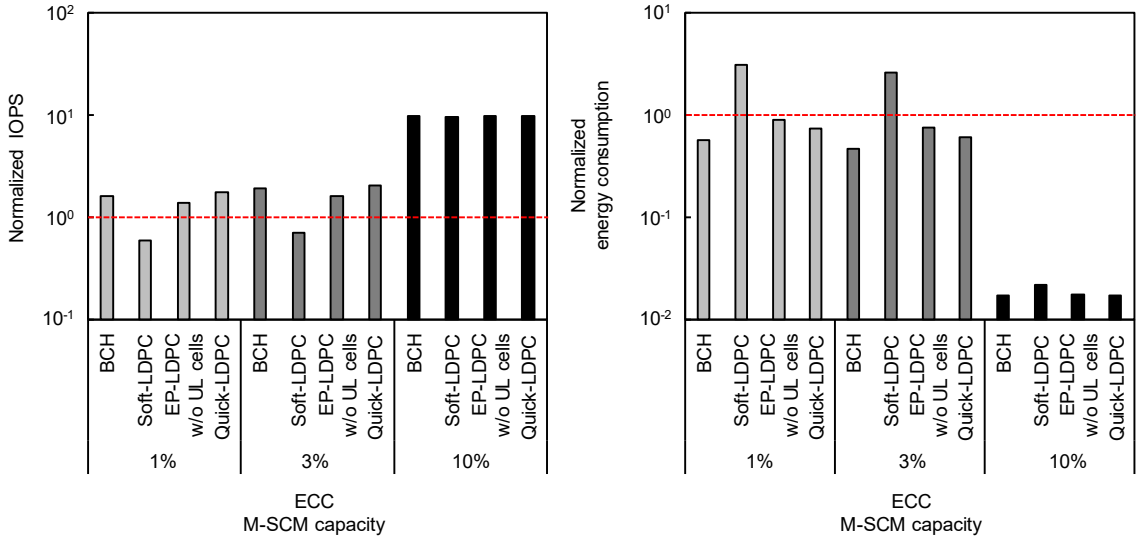


図 5.26 NAND 型フラッシュメモリに BCH あるいは LDPC ECC を適用したときの，proj_3 アプリケーションに対するハイブリッドストレージの（a）IOPS 性能，（b）消費エネルギー

フラッシュメモリに適用した LDPC 符号による性能低下率は，M-SCM を大容量用いることによる性能向上率より小さい．

表 5.3 は，MLC NAND 型フラッシュメモリを用いたストレージと比較してハイブリッドストレージが 2 倍以上の性能を得られるときの，MLC NAND 型フラッシュメモリに適用可能な

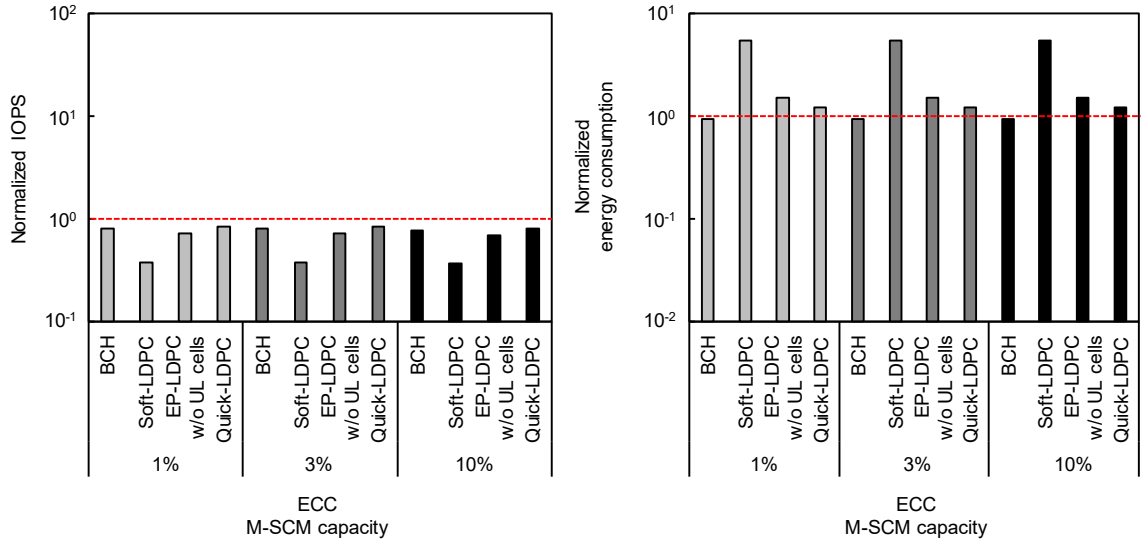


図 5.27 NAND 型フラッシュメモリに BCH あるいは LDPC ECC を適用したときの，src2_1 アプリケーションに対するハイブリッドストレージの（a）IOPS 性能，（b）消費エネルギー

表 5.3 ハイブリッドストレージの MLC NAND 型フラッシュメモリに適用可能な LDPC 符号

Application	S-SCM 1%	S-SCM 3%	S-SCM 10%
prxy_0	Soft-decoding LDPC	Soft-decoding LDPC	Soft-decoding LDPC
proj_0	BCH	Quick-LDPC	Soft-decoding LDPC
hm_0	Quick-LDPC	Quick-LDPC	Soft-decoding LDPC
src2_2	Quick-LDPC	EP-LDPC w/o UL cells	EP-LDPC w/o UL cells
prxy_1	BCH	Quick-LDPC	Soft-decoding LDPC
proj_3	Quick-LDPC	Quick-LDPC	EP-LDPC w/o UL cells
src2_1	BCH	BCH	BCH

高信頼の ECC 符号を示す．図 5.21-図 5.27 より，読み出し・書き込み時間が 0.1 μsec の高速な M-SCM と MLC NAND 型フラッシュメモリとをハイブリッド化することで，ホットなアプリケーション（prxy_0，proj_0，prxy_1）に対する性能は向上する．より大容量の M-SCM 10%を用いることで，MLC NAND 型フラッシュメモリへのアクセス頻度が減少し，高信頼だが低速な Soft-decoding LDPC を MLC NAND 型フラッシュメモリへ適用することが可能となる．一方ホットアプリケーションと比較して，ハイブリッド化によるストレージ性能向上率

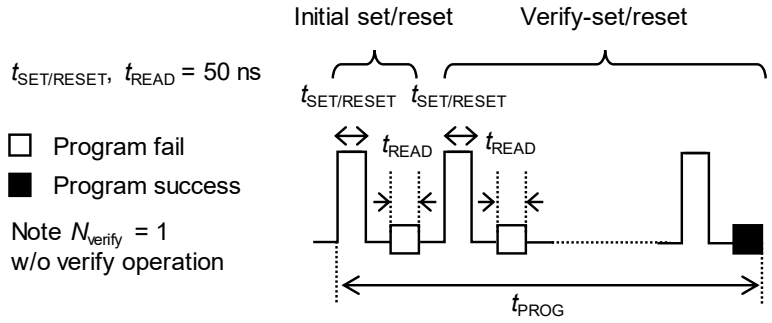


図 5.28 S-SCM の set/reset verify 動作 [15]

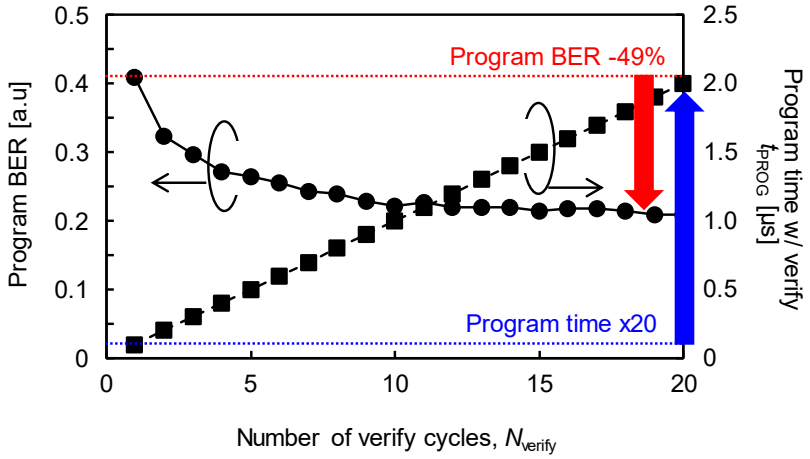


図 5.29 S-SCM の set/reset verify 動作による書き込みエラーの削減と書き込み時間の関係 [14]

が低いコールドなアプリケーション（hm_0，src2_2，proj_3）は，M-SCM を大容量にすることで，訂正能力が高いかつ高速な EP-LDPC 符号あるいは Quick-LDPC 符号を適用できることが明らかとなった．しかしコールドかつ読み出しの多い src2_2 アプリケーションは，図 5.10 で示したように，MLC NAND 型フラッシュメモリに 1 ビット訂正の BCH 符号を用いても，ハイブリッド化による性能向上は 2 倍を満たさない．そのため src2_2 のようなハイブリッド化による性能向上しないアプリケーションに対しては，NAND 型フラッシュメモリのみのストレージに BCH 符号を適用することがよいと考える．

5.5 SCM の ECC 強度と Set/Reset Verify 動作の関係

S-SCM および NAND 型フラッシュメモリを用いたハイブリッドストレージの，S-SCM の ECC および set/reset ベリファイ戦略を提案する[14]．S-SCM のエラーは，set/reset 時のベリファイ動作によっても低減できる[15]．従来研究として，ReRAM の書き換え回数によって all S-

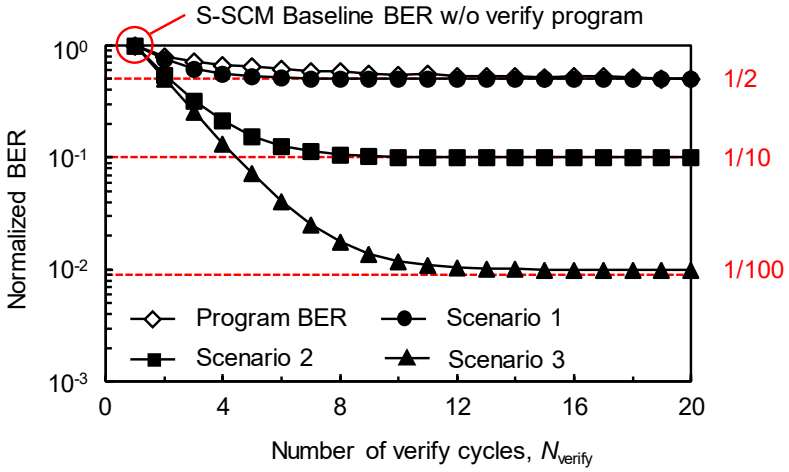


図 5.30 S-SCM の set/reset verify 動作による BER 低減シナリオ [16]

表 5.4 S-SCM の set/reset verify 動作による必要な訂正ビットの削減

Verify cycles	Required correctable bits
0	35 bit (= 9/10 code rate)
5	11 bit
10	7 bit
15	6 bit
20	6 bit

SCM ストレージのエラー訂正手法を set/reset ベリファイから ECC へと変更する提案がされている[4]．図 5.28 のように，S-SCM では書き込みが成功するまで，ベリファイ書き込みとベリファイ読み出しを行う[15]．ベリファイ回数 N_{verify} が増えると S-SCM の書き込み時間は式（5.10）に従い長くなる．

$$\text{SCM } t_{PROG} = \left(t_{SET/RESET} + t_{READ}\right) \times N_{verify} \tag{5.10}$$

図 5.29 に示すように，S-SCM の書き込み時の BER は 20 回ベリファイ動作を行うことで 49%削減するが，書き込み時間が 20 倍長くなる．このように，ECC は S-SCM の読み出し時間，ベリファイ動作は S-SCM の書き込み時間をそれぞれ長くする．アプリケーション特性および S-SCM 容量によって S-SCM へのアクセスデータ量が異なるため，S-SCM の ECC および set/reset ベリファイ戦略を変える必要がある．本節では，S-SCM の読み出し・書き込み時間は 50 nsec とし，ベリファイを行わないときをベリファイ回数 1 とする．

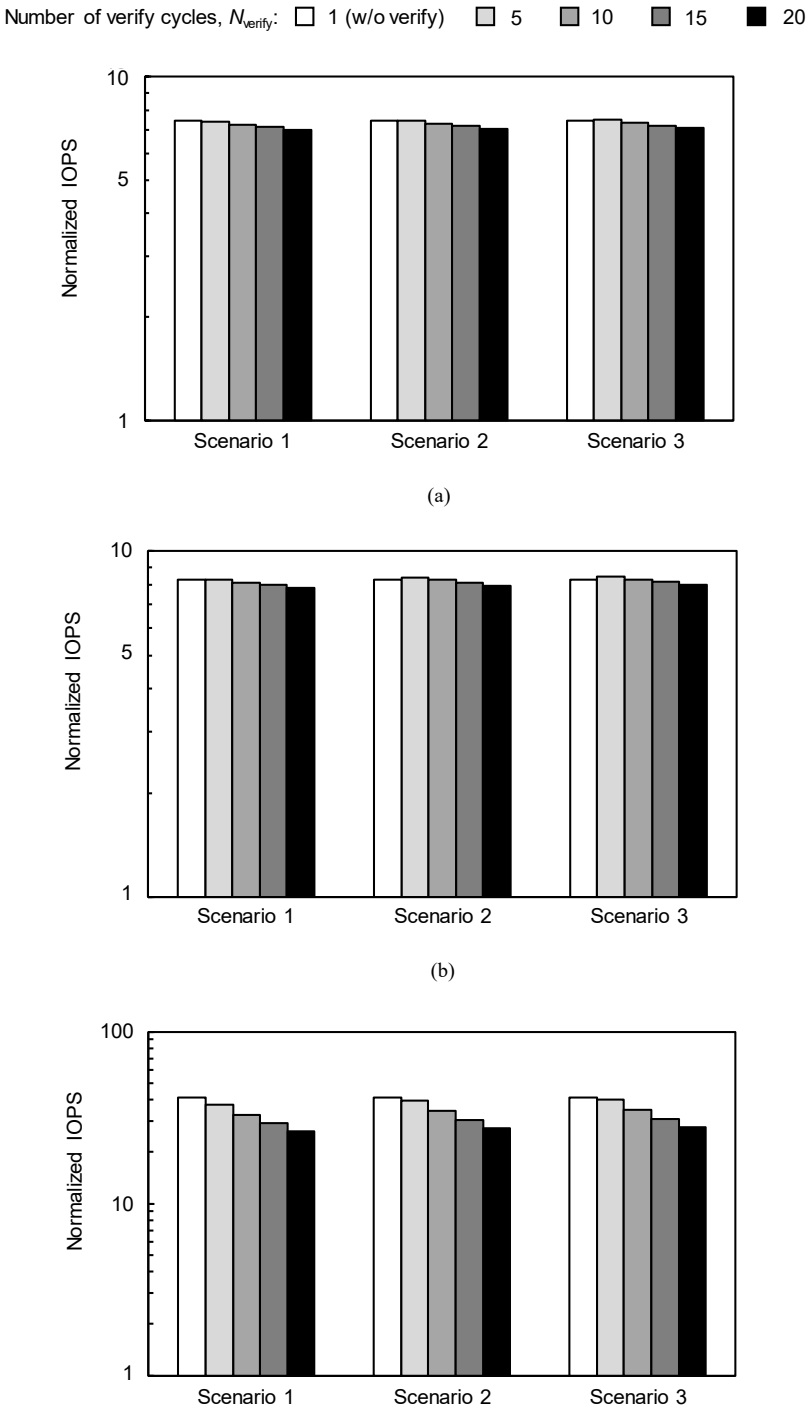


図 5.31 S-SCM の verify 動作動作および ECC による prxy_0 アプリケーションに対するハイブリッドストレージの IOPS 性能．SCM 容量比（a）1%，（b）3%，（c）10% [13]

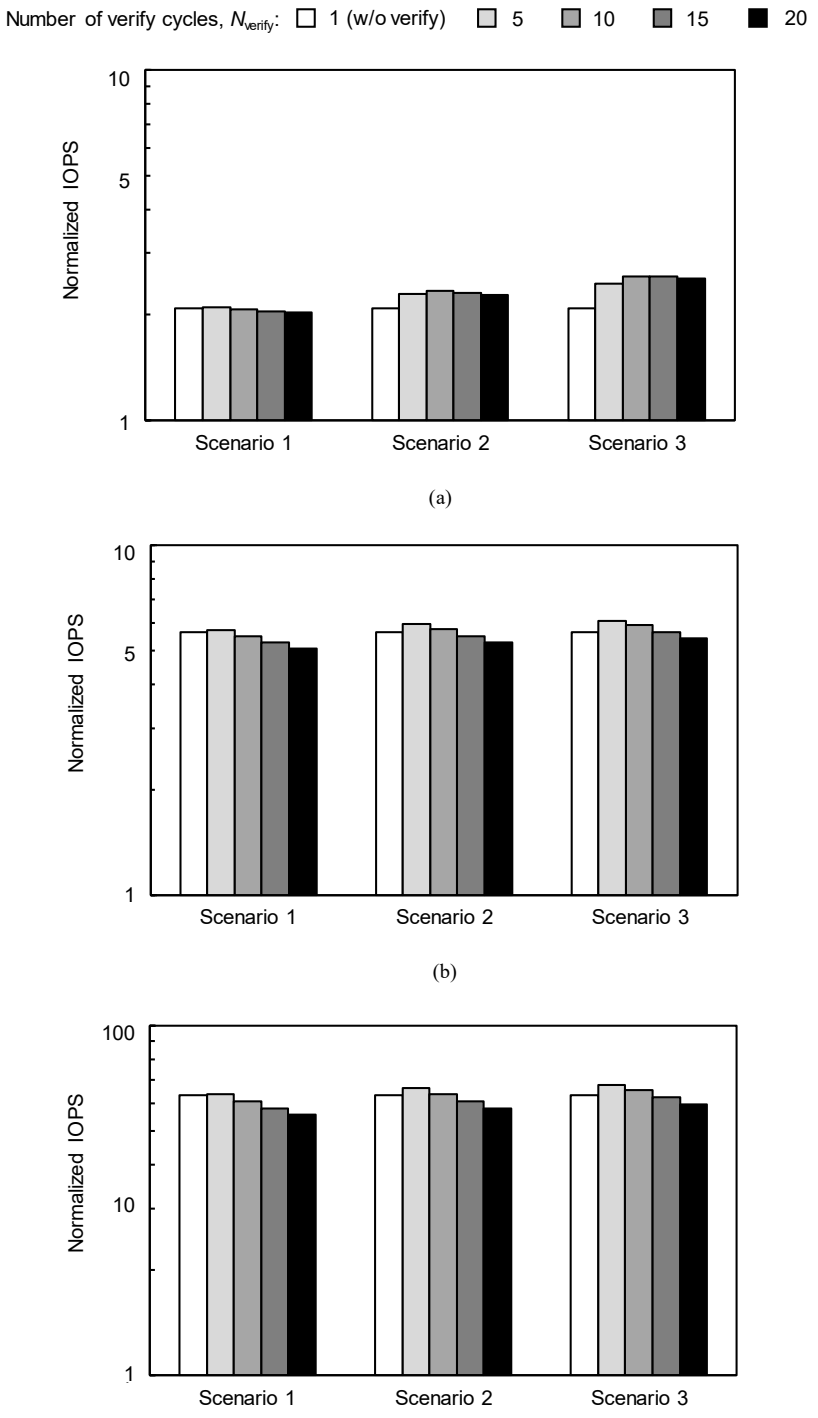


図 5.32 S-SCM の verify 動作動作および ECC による prxy_1 アプリケーションに対するハイブリッドストレージの IOPS 性能．SCM 容量比（a）1%，（b）3%，（c）10% [13]

S-SCM の set/reset 時にベリファイ動作が行われると，書き込み時の BER は図 5.30 の 3 シナリオに従い低減する．その結果表 5.4 に示すように，必要な訂正ビット数が低下する．これらのシナリオでは，ベリファイ動作 1 回につき BER が半分になり，BER は 1/2（scenario 1），1/10（scenario 2），1/100（scenario 3）に収束すると想定する[16].

MLC NAND 型フラッシュメモリでは，ユーザデータ 8 KByte の 428 ビットエラーを訂正するために符号化率 9/10 の BCH ECC が必要となる．S-SCM によるストレージ性能向上が顕著な，書き込みの多い prxy_0 および読み出しの多い prxy_1 アプリケーションを用いて評価した. ハイブリッドストレージのデータマネジメントアルゴリズムとして第 2.4 節で述べた CDE, S-SCM 容量は MLC NAND 型フラッシュメモリの 1%, 3%, 10%とした. S-SCM と MLC NAND 型フラッシュメモリのパリティビットはそれぞれのメモリに保存する．

図 5.31 および図 5.32 に S-SCM および MLC NAND 型フラッシュメモリを用いたハイブリッドストレージの性能を示す．ハイブリッドストレージの性能は，MLC NAND 型フラッシュメモリのみを用いたストレージと比較して向上する．しかしベリファイ動作および ECC がその性能を低下させる．さらに性能低下の傾向はストレージ構成およびアプリケーション特性によって異なる．prxy_0 は書き込みが多いため，BER シナリオに関わらずベリファイ動作は S-SCM の書き込み時間を長くし性能が劣化する．読み出しの多い prxy_1 では，S-SCM のエラー削減戦略は異なる．S-SCM 容量が 1%のハイブリッドストレージに対し，ベリファイ動作を行わない場合と比較して，ベリファイ動作を 5 回行うことでストレージ性能は 23％向上する．これに対し S-SCM 容量が 5%および 10%のとき，ベリファイ動作を行わない場合と比較して, ベリファイ書き込みを用いることによるハイブリッドストレージの性能の向上は 7%である．図 5.29 のようにベリファイ動作を 20 回行うと，S-SCM の書き込み時間は 5 倍長くなる．一方表 5.4 で示したように，ベリファイ動作を 20 回行うことで S-SCM の訂正可能ビットが 35 bit から 6 bit へ削減でき，式（5..2）で表される読み出し時間は 46%削減できることがわかる．このため S-SCM 容量が 3%および 10%の場合に，S-SCM のベリファイ動作を 10 回以上行うと書き込み時間が長くなり，ハイブリッドストレージ性能向上を妨げる要因となる．

5.6 まとめ

本章では M-SCM あるいは S-SCM および MLC NAND 型フラッシュメモリを用いたハイブリッドストレージの信頼性と性能との関係をシステムレベルで理解した. M-SCM あるいは S-SCM のエラーは BCH 符号で訂正し，MLC NAND 型フラッシュメモリのエラーは BCH ある

いは LDPC 符号で訂正する．ホットアプリケーション（prxy_0，proj_0，prxy_1）に対して，大容量の M-SCM あるいは S-SCM は性能を大きく向上させる一方，M-SCM あるいは S-SCM へのデータアクセス頻度が高くなる．このため，M-SCM あるいは S-SCM に訂正能力の高い BCH ECC を適用すると性能の低下が起こる．また ECC は，メモリへの書き込み時に符号に要する時間と比較して，メモリからの読み出し時に復号に要する時間が長い．そのため，読み出しの多い prxy_1 アプリケーションに対して，M-SCM あるいは S-SCM に適用する BCH 符号の訂正能力を高くすることはできない．

これと比較して，大容量の M-SCM あるいは S-SCM を用いるハイブリッドストレージでは多量のデータが M-SCM あるいは S-SCM で処理されるため，復号時間の長い LDPC 符号を MLC NAND 型フラッシュメモリに適用することを可能にする．小容量の M-SCM は MLC NAND 型フラッシュメモリへのアクセス頻度を低減するため, EP-LDPC w/o UP cells や Quick-LDPC のような復号時間の短く信頼性の高い LDPC 符号を適用できる一方で，M-SCM へのアクセスが減るため，訂正能力の高い BCH 符号を M-SCM に適用できる．MLC NAND 型フラッシュメモリのみを用いたストレージと比較して 2 倍の性能を求めるとき, 大容量の M-SCM を用いることで，高信頼だが復号時間の長い Soft-decoding LDPC 符号を MLC NAND 型フラッシュメモリに適用することも可能となる．

3 次元積層（3-dimentional，3D）TLC NAND 型フラッシュメモリは，今後データセンターストレージおいて利用が拡大すると考えられる．参考文献[17]によると 3D TLC NAND 型フラッシュメモリの読み出し時間は，本章で評価に用いた 2 次元積層（2-dimentional，2D）MLC NAND 型フラッシュメモリのそれと比較して約 1.7 倍長い．そのため LDPC 符号の長い復号時間が 3D TLC NAND 型フラッシュメモリの読み出し時間と比較に隠れ，ハイブリッドストレージ性能低下への影響が緩和すると考えられる．また，M-SCM および 3D TLC NAND 型フラッシュメモリを用いたハイブリッドストレージの性能は，M-SCM および”D MLC NAND 型フラッシュメモリを用いたハイブリッドストレージの性能と比較して 2-20%向上することが示されている[17]．したがって，M-SCM あるいは S-SCM および 3D TLC NAND 型フラッシュメモリを用いたハイブリッドストレージの 3D TLC NAND 型フラッシュメモリには，2D MLC NAND 型フラッシュメモリと比較して，M-SCM あるいは S-SCM 容量が少ない場合においても，復号時間が長いが訂正能力が高い LDPC 符号を適用することが可能になると考える．

さらにハイブリッドストレージにおける S-SCM のエラー削減戦略を検討した．システムの

観点から，S-SCM の set/reset ベリファイおよび BCH ECC によるハイブリッドストレージ性能への影響は，ストレージの構成およびアプリケーション特性つまり読み出し書き込みの多寡による．書き込みの多いアプリケーションでは訂正能力の高い BCH ECC でエラー訂正することが必要となる．読み出しの多いアプリケーションに対しては set/reset ベリファイ動作を 5 回程度行い書き込み時の S-SCM のエラーを低減させ，訂正能力が低いが復号時間が短い BCH ECC を用いて残ったエラーを訂正することでストレージシステムの性能を向上することができるを明らかにした．

参考文献

[1] Y. Cai, Y. Luo, S. Ghose, and O. Mutlu, “Read disturb errors in MLC NAND flash memory: Characterization, Mitigation and Recovery,” in *Proceedings of IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)*, Jun. 2015, pp. 438-439.

[2] K. Kawai, A. Kawahara, R. Yasuhara, S. Muraoka, Z. Wei, R. Azuma, K. Tanabe, and K. Shimakawa, “Highly-reliable TaOx ReRAM technology using automatic forming circuit,” in *Proceedings of IEEE International Conference on IC Design and Technology (ICICDT)*, May 2014, pp. 100-103.

[3] C. Y. Chen, A. Fantini, R. Degraeve, A. Redolfi, G. Groeseneken, L. Goux, and G. S. Kar, “Statistical investigation of the impact of program history and oxide-metal interface on OxRRAM retention,” in *IEEE International Electron Devices Meeting (IEDM) Technical Digest*, Dec. 2016, pp. 99-102.

[4] A. Hayakawa, K. Maeda, S. Fukuyama, H. Takishita, R. Yasuhara, S. Mishima, and K. Takeuchi, “Resolving endurance and program time trade-off of 40nm TaOx-based ReRAM by co-optimizing verify cycles, reset voltage and ECC strength,” in *Proceedings of IEEE International Memory Workshop (IMW)*, May 2017, pp. 24-27.

[5] 和田山 正，“誤り訂正技術の基礎”第 1 版第 2 刷，2011 年，森北出版.

[6] 今井 秀樹，“符号理論” 初版第 12 刷，1990 年，電気電子情報通信学会.

[7] Y. Lee, H. Yoo, I, Yoo, and I.-C. Park, “6.4 Gb/s multi-threaded BCH encoder and decoder for multi-channel SSD controllers,” in *IEEE International Solid-State Circuits Conference (ISSCC) Digest of Technical Papers*, Feb. 2012, pp. 426-428.

[8] K. Zhao, W. Zhao, H. Sun, T. Zhang, X. Zhang, and N. Zheng, “LDPC-in-SSD: Making advanced error correction codes work effectively in solid state drives,” in *Proceedings of USENIX Conference on File and Storage Technologies (FAST)*, Feb. 2013, pp. 243-256.

[9] C.-L. Chen, K.-S. Lin, H.-C. Chang, W.-C. Fang, and C.-Y. Lee, “A 11.5 Gbps LDPC decoder based on CP-PEG code construction,” in *Proceedings of European Solid-State Circuits and Conference (ESSCIRC)*, Sep. 2009, pp. 412-415.

[10] Y. Yamaga, C. Matsui, S. Hachiya, and K. Takeuchi, “Application optimized adaptive ECC with advanced LDPCs to resolve trade-off among reliability, performance, and cost of solid-state drives,” in *Proceedings of IEEE International Memory Workshop (IMW)*, May 2016, pp. 129-132.

[11] T. Tokutomi, M. Doi, S. Hachiya, A. Kobayashi, S. Tanakamaru, and K. Takeuchi, “Enterprise-grade 6x fast read and 5x highly reliable SSD with TLC NAND-flash memory for big-data storage,” in *IEEE International Solid-State Circuits Conference (ISSCC) Digest of Technical Papers*, Feb. 2015, pp. 140-141.

[12] S. Okamoto, C. Sun, S. Hachiya, T. Yamada, Y. Saito, T. O. Iwasaki, and K. Takeuchi, “Application driven SCM and NAND flash hybrid SSD design for data-centric computation system,” in *Proceedings of IEEE International Memory Workshop (IMW)*, May 2015, pp. 157-160.

[13] C. Matsui, R. Kinoshita, and K. Takeuchi, “Analysis on applicable ECC strength of SCM and NAND flash in hybrid storage,” *Japanese Journal of Applied Physics (JJAP)* to be published in vol. 57, no. 4S, Apr. 2018.

[14] C. Matsui and K. Takeuchi, “Error-correction & set/reset verify strategy of storage class memory (SCM) for SCM/NAND flash hybrid and all-SCM storage,” in *Extended Abstracts of International Conference on Solid State Devices and Materials (SSDM)*, Sep. 2017, pp. 783-784.

[15] S. Ning, T. O. Iwasaki, and K. Takeuchi, “50 nm Al_xO_y resistive random access memory array program bit error reduction and high temperature operation,” *Japanese Journal of Applied Physics (JJAP)*, vol. 53, no. 4S, pp. 04ED09-1- 04ED09-7, Apr. 2014.

[16] H. Takishita, Y. Adachi, and K. Takeuchi, “ReRAM-based SSD performance considering verify-program cycles and ECC capabilities,” in *Non-Volatile Memoryies Workshop (NVMW)*, Mar. 2018.

[17] M. Fukuchi, Y. Sakaki, C. Matsui, and K. Takeuchi, “20% system-performance gain of 3D charge-trap TLC NAND flash over 2D floating-gate MLC NAND flash for SCM/NAND flash hybrid SSD,” to be presented in *IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2018.

第6章 不揮発性メモリを用いたストレージの適応制御

6.1 はじめに

本章では，ストレージアプリケーションを考慮した SCM 容量自律調整手法を提案する．NAND 型フラッシュメモリの不揮発性キャッシュとして SCM を用いることは，高い性能を得るストレージのために有望な解である．しかし SCM のビットコストは NAND 型フラッシュメモリと比較して高く，最適な SCM 容量はアプリケーションに依存する点が問題である．データセンターストレージで動作するすべてのアプリケーションに対しては従来，最適な SCM 容量は手動で決定すると考えてきた．高性能を達成しながらストレージコストを削減するために，SCM と NAND 型フラッシュメモリからなるハイブリッドメモリプールの SCM 容量を，NAND 型フラッシュメモリ内のワーム（warm）データを観察することで調整する．さらにワームデータを削減し SCM をより効率的に使うため，SCM 内でのデータアクセス頻度を考慮したデータエビクションアルゴリズムも提案する．また，NAND 型フラッシュメモリの動作を高速化するためにガベージコレクション手法を提案する．

6.2 ハイブリッドストレージにおける SCM 容量の適応制御手法

第 6 章第 2 節の内容は，2018 年 5 月に開催される IEEE International Symposium on Circuits and Systems（ISCAS）にて発表予定であり，また学術論文誌に投稿する予定である．

6.3 NAND 型フラッシュメモリの書き込み順序を用いたガベージコレクション手法

NAND 型フラッシュメモリの書き込み性能は，同一ページでの上書き不可および GC により制限される．この問題を解決するために，書き込み順序を用いた GC（Write-Order based GC, WO-GC）手法を提案する[7]．提案手法では，NAND 型フラッシュメモリのブロックの有効ページ数，書き込み順序，消去回数を GC 時の消去ブロック選択時に考慮する．WO-GC の利点の一つはストレージシステム内に時間を計測する時計が不要なことであり，ストレージの電源消失時にも動作可能である．

GC は図 2.6 で示したように，NAND 型フラッシュメモリの空ページが少なくなった場合に空ページをつくる動作であり，消去予定ブロックの有効ページを別のブロックに書き込む．

表 6.2 従来および提案のガベージコレクション手法の概要 [22]

GC algorithm	Score formula	Choose a victim block by	Pros	Cons
Round-robin	N_{erase}	Minimum or FIFO manner	・Even wearout	・Large latency
Greedy	μ	Minimum	・Short latency	・Uneven wearout
Cost-benefit	$\frac{1-\mu}{2\mu} \times age$	Maximum	・Short latency ・Age considered	・Uneven wearout
Cost-age-times (CAT)	$\frac{\mu}{1-\mu} \times \frac{1}{f(age)} \times N_{\text{erase}}$	Minimum	・Short latency ・Age and erase count considered	・Internal timer
Write order based GC (WO-GC)	$\frac{\mu}{1-\mu} \times \frac{1}{\frac{Max.WSN-WSN}{Max.WSN}} \times \frac{N_{\text{erase}}}{Max.N_{\text{erase}}}$	Minimum	・Short latency ・Age and erase count considered ・NO timer	

μ : valid page ratio in the block, which is # of valid pages divided by # of pages in a block
N_{erase}: erase count of the block
age: elapsed time since the block was last modified
WSN: write sequence number since the SSD is power on
Max. WSN and *Max.* N_{erase}: Maximum values of *WSN* and N_{erase} each time GC is triggered

GC アルゴリズムおよび消去ブロック選択手法を設計するとき，次の 3 つのガイドラインを考慮すべきである[8].

第一に，データ保持エラーが蓄積している，最も古いブロックの中から消去ブロックを選択しなければならない.

第二に，有効ページのコピー時間を抑えるために，消去するブロックはできるだけ多くの無効ページを持っていること.

第三に，NAND 型フラッシュメモリの書き換え回数は制限されているため，できるだけ均等にブロックを使用する[9][10].

したがって，GC アルゴリズムは，ブロックの年齢，有効ページ数，消去回数などのパラメータを考慮し，ブロックを選択することが必要となる.

第一のパラメータは，ブロックの年齢（*age*），つまり，最後に書き込まれてからの時間である．時間が経つにつれて，メモリセルは蓄積された電荷の一部あるいはすべてを失う可能性があり，データ保持エラーと呼ばれる．データ保持エラーをリセットするためには，最も古いブロックを消去ブロックとして選択し，消去するのがよい.

第二のパラメータは，ブロックの使用率あるいは有効ページ数（μ）である．有効ページ数が少ない場合, GC 時に消去するブロックから新しいブロックに有効ページをコピーする時間が少なくなる．そのため，最小な有効ページ数を持つブロックを消去するブロックとして選

択するのがよい.

第三のパラメータは，ブロックの消去回数（N_{erase}）である．最小の消去回数をもつブロックを選択することで，NAND 型フラッシュメモリのブロックは均一に使われる．

表 6.2 に NAND 型フラッシュメモリ向けの従来の GC 手法および提案の WO-GC 手法を示す．各アルゴリズムは，3 パラメータ（年齢 *age*，有効ページ割合 μ，消去回数 N_{erase}）の中で考慮するパラメータが異なる．第一の RR アルゴリズムは，最も少ない消去回数（N_{erase}）を持つブロックを選択する[11][12][13]．これは，first-in, first-out と同じである．このアルゴリズムは，NAND 型フラッシュメモリのブロックを均等に使うが，有効ページ数を考慮していない．参考文献[12][13]において，RR アルゴリズムは有効ページ数の多いブロックが頻繁に消去されると報告されている．したがって，RR GC は write amplification が高い傾向を持つ．

第二の greedy アルゴリズムは，有効ページ数（μ）の最も少ないブロックを選択する[13][14][15]．しかし greedy アルゴリズムは，頻繁に上書きされるホットデータに対して write amplification が高い．NAND 型フラッシュメモリでは古いページに直接上書きせず，古いページの上書きされるデータを無効化し，新しいページに書き込む．これにより，greedy アルゴリズムは，有効ページ数が少ない頻繁に上書きされるページを含むブロックを消去対象ブロックとして選択する傾向を持つ．また，このアルゴリズムは，ブロックの消去回数が均一にならない．greedy アルゴリズムの派生として，windowed greedy GC [12][15]，d-choice GC [17] がある．

第三の cost-benefit アルゴリズムは，利益（benefit）対コスト（cost）の割合が最も高いブロックを消去ブロックとして選択する[14][18]．具体的にブロックを再利用するコスト（cost）は，消去するブロック内のすべての有効ページを読み出し，別の空きブロックに書き込む時間である．一方ブロックを消去する利益（benefit）は，2 つのパラメータに基づいて計算される．第一のパラメータはブロックの無効ページの数（1-μ）であり，つまり GC 後に解放されるページ数である．第二のパラメータは，ブロックの年齢（*age*）である．ここでブロックの年齢は，ブロックに書き込み・消去などの変更がなされてからの時間である．ブロックの年齢と有効ページ割合を考慮することにより，cost-benefit アルゴリズムは時間的局所性の高いアプリケーションに対し高い性能を持つ．しかし依然として，ブロックの消去回数は不均一である．

より高度な手法は，有効ページ，年齢，消去回数の 3 パラメータすべてを含む第四の Cost-age-times（CAT）アルゴリズムである[19]．CAT アルゴリズムにおいて消去するブロックを選

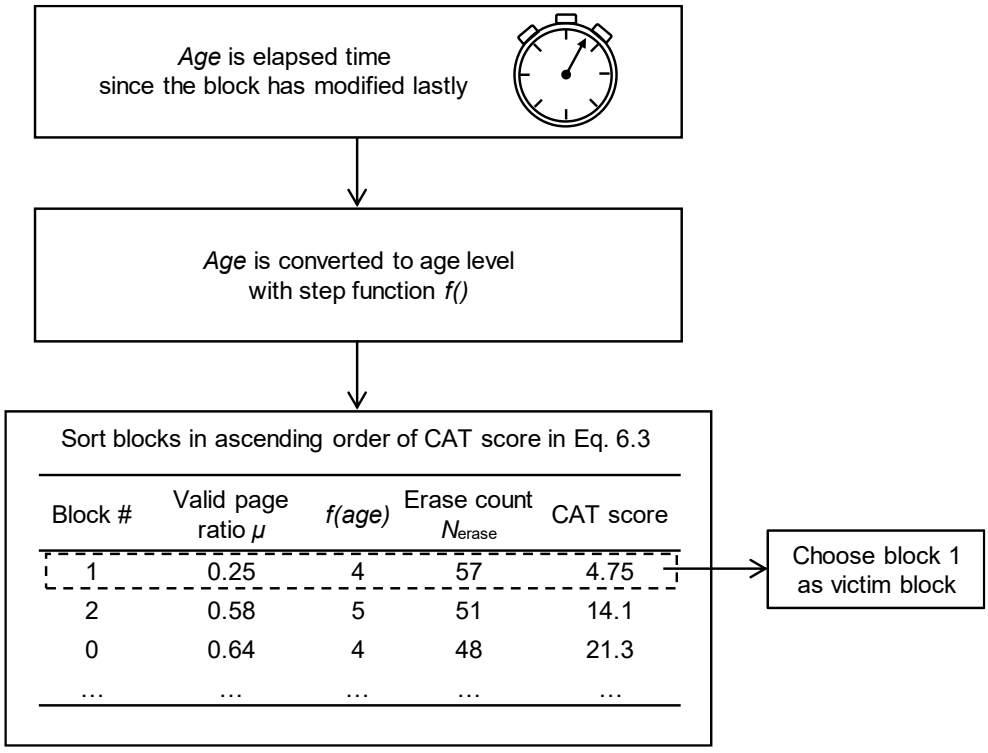


図 6.13 NAND 型フラッシュメモリの GC 対象ブロックを選択する CAT アルゴリズム [7][22]

択するための CAT score を式（6.3）に示す.

$$\text{CAT score} = \frac{\mu}{1-\mu} \times \frac{1}{f(age)} \times N_{erase} \tag{6.3}$$

CAT アルゴリズムでは，最小の CAT score を持つブロックを消去ブロックとして選択する．式（6.3）の第一項はブロック内の有効ページ，第二項はブロックが上書きされてからの時間を参考文献[19]で示されたステップ関数 *f()*による age level への変換，第三項はブロックの消去回数を表す．このようにして，GC の利益・コストおよびブロックのウエアレベリングを考慮するため，前述した 3 つの基本的なガイドラインを満たす．しかし，CAT アルゴリズムは実用的な問題がある．第一の問題は図 6.13 に示すように，CAT アルゴリズムはブロックの年齢をモニタするために，ストレージ内部にタイマーを必要とすることである．ストレージが突然停止した場合，ブロック年齢の情報が消失する．第二の問題は，年齢を変換する関数 *f()* はアプリケーション毎に決めなければならない点である．各ブロックの年齢はタイマーで計測される．そして式（6.3）の年齢の項の重みを削減するために，関数 *f()*を用いて age level に変換する．第三の問題は，式（6.3）の各項の重みはバランスされておらず，異なる範囲の値をとることである．すなわち，有効ページ割合 μ は 0-1，年齢変換関数 *f()*による age level は

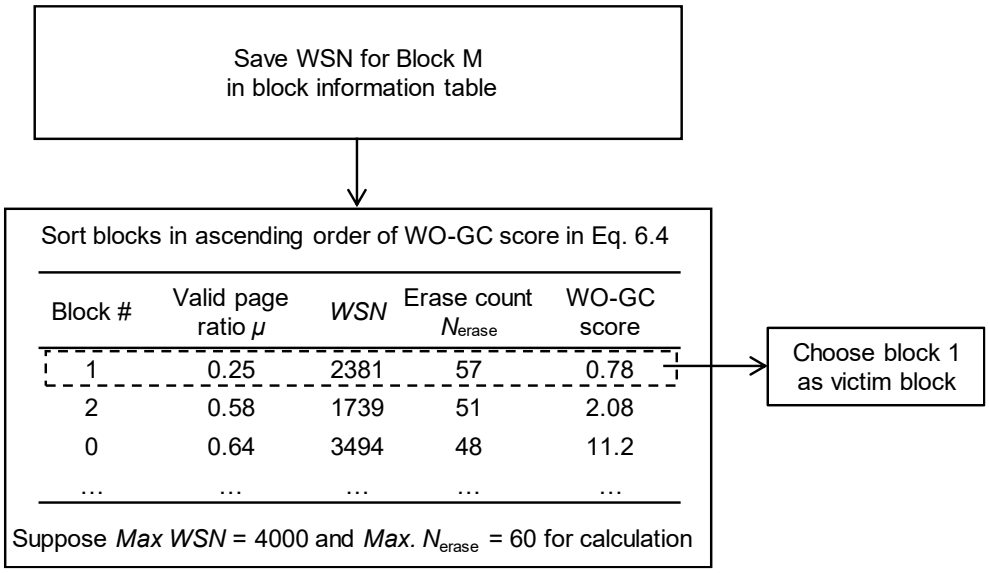


図 6.14 NAND 型フラッシュメモリの GC 対象ブロックを選択する書き込み順序を用いた GC アルゴリズム [7][22]

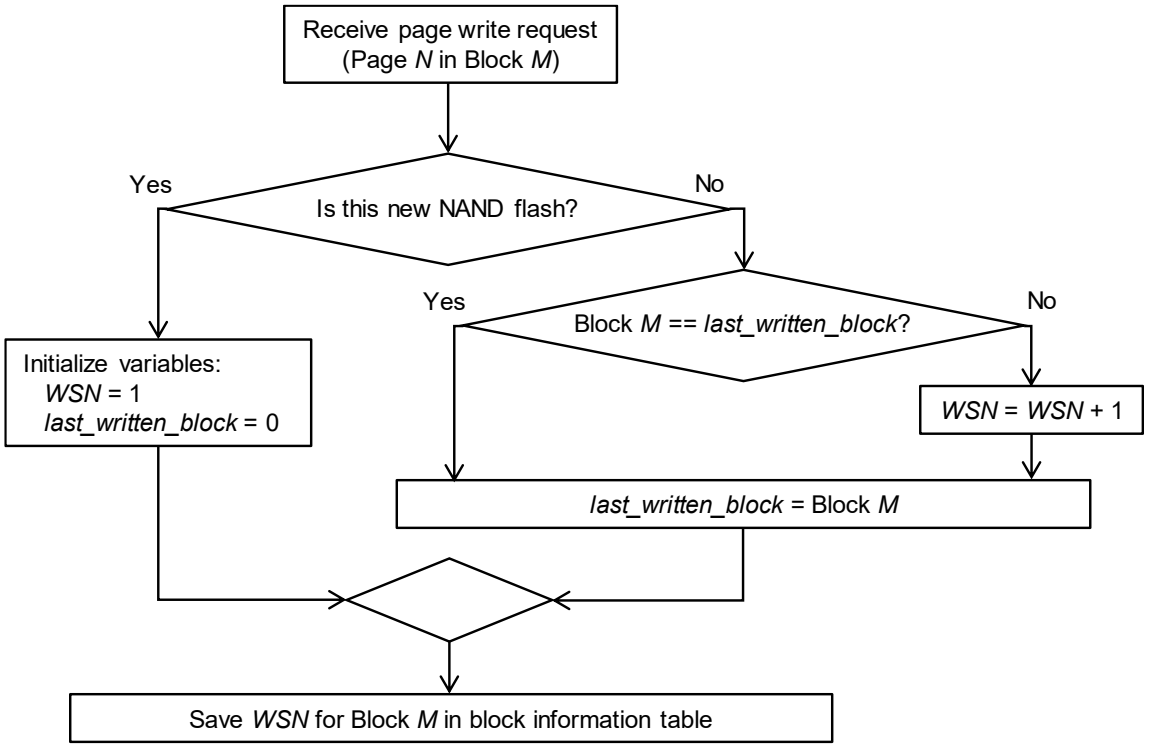


図 6.15 NAND 型フラッシュメモリへの書き込み順決定アルゴリズム [22]

0-7，ブロック消去回数 N_{erase} は 0 から NAND 型フラッシュメモリの最大書き換え回数までの値をとりうる．例えば MLC NAND 型フラッシュメモリは 1 万書き換え回数である．したがって，式（6.3）の第三項に示す消去回数が過度に支配的になる．この場合 NAND 型フラッシュメモリのブロックのウエアレベリングに最適化され，性能面では有効ページ割合が考慮されなくなる．その結果，有効ページ割合の多いブロックを選択することによって，NAND 型フラッシュメモリ性能が低下する．このように CAT アルゴリズムを NAND 型フラッシュメモリに適用するには実用的な問題がある. その結果, ストレージ内部にタイマーを使用せず, 3 パラメータを考慮する GC アルゴリズムが必要となる．

内部にタイマーを必要とする CAT アルゴリズムの代替として，図 6.14 に示す書き込み順序を用いた GC 手法（write order-based GC，WO-GC）を提案した[7].

WO GC において消去するブロックを選択するための WO score を式（6.4）に定義する．そして最小の WO score を持つブロックを消去するブロックとして選択する．

$$\text{WO score} = \frac{\mu}{1-\mu} \times \frac{1}{\frac{Max\ WSN - WSN}{Max\ WSN}} \times \frac{N_{erase}}{Max\ N_{erase}} \tag{6.4}$$

CAT アルゴリズムと同様に，（6.4）は有効ページ割合およびブロック消去回数を用いる．さらに, ブロックの相対的な年齢を測定するために通し番号である write sequence number（*WSN*）と呼ばれる新しいパラメータを導入する．図 6.15 に *WSN* 決定のアルゴリズムを示す．新しいページ書き込みリクエストが来たときに *WSN* をインクリメントする．各ブロックの *WSN* 値はブロック情報テーブル（block information table）に保存される．初めに *WSN* 値はゼロにセットされ, 物理ページへの書き込みリクエストが来るたびに, ブロック *M*, ページ *N* の *WSN* をインクリメントする．DRAM 容量を節約するために，同一ブロックの連続したページに書き込まれるときは，*WSN* をインクリメントしない．ページ書き込み動作が終わると，ブロックの *WSN* はブロック情報テーブルに保存される．NAND 型フラッシュメモリ内の空きブロック数がしきい値を下回ると, 提案の WO-GC がトリガされる. 各ブロックの WO スコアは, 式 (6.4) に基づいて計算され, 最小の値を持つブロックを消去ブロックとして選択する．*WSN* を用いる提案の GC アルゴリズムは実際的であり，ほとんど無視できるオーバーヘッドで NAND 型フラッシュメモリを用いたストレージシステムに実装できる．WO-GC は初めに logical address block（LBA）scrambler [20][21]を適用した NAND 型フラッシュメモリを用いたストレージの性能を向上させるために提案した．LBA scrambler は NAND 型フラッシュメモリのページ内の無効セクタにあらかじめ書き込むことで GC 対象ブロック内の有効ページを

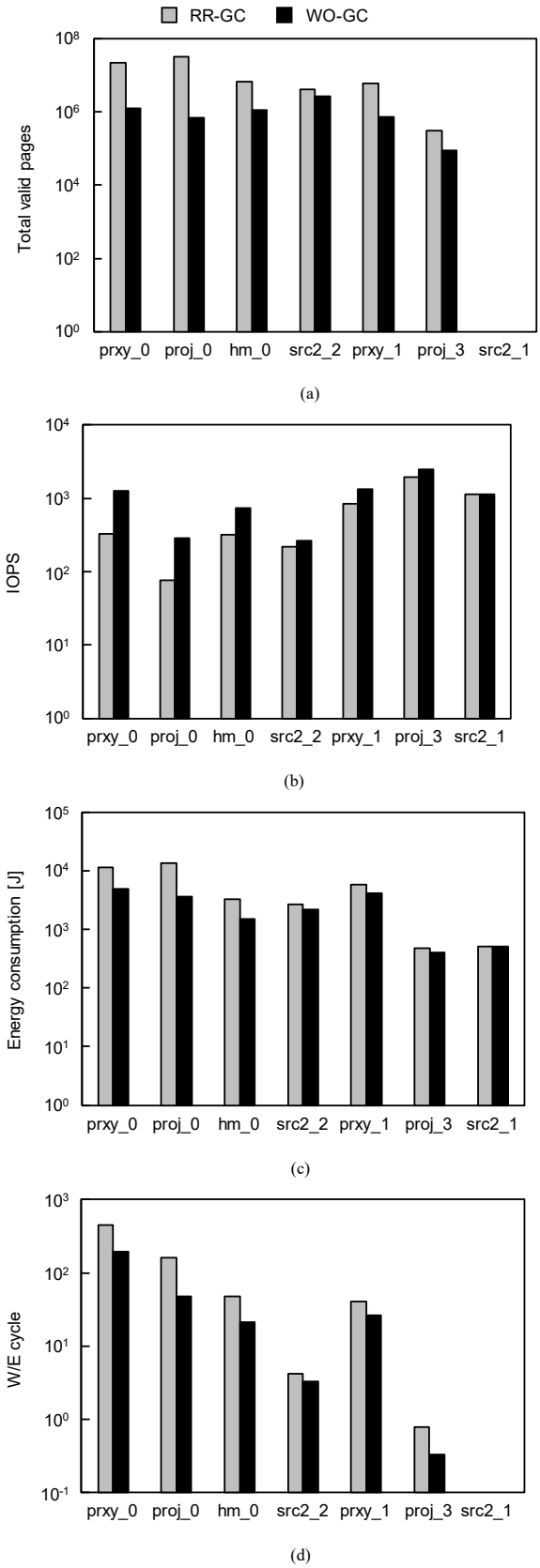


図 6.16 WO-GC による MLC NAND 型フラッシュメモリを用いたストレージの性能．（a）GC 時の有効ページ数の総計，（b）IOPS 性能，（c）消費エネルギー，（d）W/E cycle

削減する手法である．しかし RR-GC アルゴリズムを用いた LBA scrambler は prxy_0（write-hot-random）アプリケーションに対して，有効ページを削減できず性能が悪化した[20][21]．ブロック内有効ページを考慮した WO-GC を適用することで様々なストレージアプリケーションに対して，LBA scrambler が性能向上に寄与することを示した[7][22]．

図 6.16 に MLC NAND 型フラッシュメモリのみを用いたストレージの性能を示す[22]．提案した書き換え回数を用いた GC（WO-GC）と，従来手法としてラウンドロビン方式 GC（RR-GC）を比較した．図 6.16（a）は GC 中に消去するブロックから空のブロックへコピーした有効ページ数の総計を示す．式（6.4）に従い，WO-GC はできるだけ少ない有効ページを持つブロックが GC 対象ブロックとして選択されるため，RR-GC と比較して GC でコピーした有効ページ数の総計が少ない．特に書き込みが多くホットな prxy_0, proj_0 アプリケーションは，WO-GC を用いることでコピーした有効ページ数を 90%以上削減した．ただし，読み出しが多くコールド・シーケンシャルな src2_1 アプリケーションは，RR-GC あるいは WO-GC を用いても，MLC NAND 型フラッシュメモリの GC は一度も発生しなかった．WO-GC によりコピーする有効ページ数が削減したストレージアプリケーションほど，図 6.9（b）に示すストレージの IOPS 性能が向上する．書き込みが多くホットな prxy_0，proj_0 アプリケーションは

約 3.8 倍性能が向上した．また，コピーする有効ページ数が少ないほど消費エネルギーも低減できる．図 6.9（c）に示すように prxy_0，proj_0 アプリケーションは消費エネルギーが 60%以上低減できた．さらに有効ページ数が多い GC 対象ブロックは，ページの書き換えが発生しないフローズンデータであると考えられる．そのため WO-GC を用いて一部の有効ページ数が少ないブロックを GC 対象として選択することで，図 6.9（d）に示すように MLC NAND 型フラッシュメモリの平均書き換え回数（W/E cycle）が削減する．

6.4 まとめ

本章ではストレージアプリケーションの特性に応じた SCM 容量の自律調整手法を述べた．提案手法である Application-aware Autonomous SCM Capacity Adjustment（3ASCA）は SCM から NAND 型フラッシュメモリに evict されたデータをモニタすることで，SCM 容量を調整することを提案した．NAND 型フラッシュメモリでアクセスされた回数をカウントするためにゴースト LRU リストを用いた．prxy_1（read-hot-random）アプリケーションについては，SCM 容量 10%と比較して，IOPS 性能が 4.8%低下するだけで，総コストが 42%低減できた．さらに SCM で頻繁にアクセスされたデータを MLC NAND 型フラッシュメモリへ evict しない SCM-assisted data eviction 手法を提案した．これにより，SCM から NAND 型フラッシュメモ

リへ evict してすぐにまたアクセスされるデータを削減することができた．3ASCA と SCM-assisted data eviction を併用することで，prxy_0（write-hot-random）のストレージ動作中に必要なメモリのコストはさらに 6.4%削減できることを明らかにした．

また，MLC NAND 型フラッシュメモリの書き込み順を用いる，書き込み順序を用いたガベージコレクション（write order-based GC，WO-GC）を提案した．WO-GC を用いることで GC 時にコピーする有効ページ数が削減できる．特に書き込みが多くホットなアプリケーション（prxy_0，proj_0）に対して，MLC NAND 型フラッシュメモリのみを用いたストレージの性能は約 3.8 倍向上できた．ストレージコントローラ内に時計を用いる CAT GC と比較して，WSN を用いる提案の WO-GC アルゴリズムは実用的であり，NAND 型フラッシュメモリを用いたストレージシステムに実装できる．

参考文献

[1] S. Okamoto, C. Sun, S. Hachiya, T. Yamada, Y. Saito, T. O. Iwasaki, and K. Takeuchi, “Application driven SCM and NAND flash hybrid SSD design for data-centric computation system,” in *Proceedings of IEEE International Memory Workshop (IMW)*, May 2015, pp. 157-160.

[2] Y. Sugiyama, T. Yamada, C. Matsui, and K. Takeuchi, “Reconfigurable SCM capacity identification method for SCM/NAND flash hybrid disaggregated storage,” in *Proceedings of IEEE International Memory Workshop (IMW)*, May 2017, pp. 60-63.

[3] DRAMexchange, http://www.dramexchange.com.

[4] T. Yamada, C. Matsui, and K. Takeuchi, “Workload-based co-design of non-volatile cache algorithm and storage class memory specifications for storage class memory/NAND flash hybrid SSDs,” *IEICE Transactions on Electronics*, vol. E100-C, no. 4, pp. 373-381, Apr. 2017.

[5] N. Megiddo and D. S. Modha, “ARC: A self-tuning, low overhead replacement cache,” in *Proceedings of USENIX Conference on File and Storage Technologies (FAST)*, Mar. 2003, pp. 115-130.

[6] C. Matsui and K. Takeuchi, “3ASCA: Application-aware autonomous SCM capacity adjustment for SCM and NAND flash pooled storage”, to be presented in *IEEE International Symposium on Circuits and Systems (ISCAS)*, May 2018.

[7] C. Matsui, A. Arakawa, C. Sun, T. O. Iwasaki, and K. Takeuchi, “3x faster speed solid-state drive with a write-order based garbage collection scheme,” in *Proceedings of IEEE International Memory Workshop (IMW)*, May 2015, pp. 153-156.

[8] H. Niijima, “Design of a solid-state file using flash EEPROM,” *IBM Journal of Research and*

Development, vol. 39, no. 5, pp. 531-545, Sep. 1995.

[9] P. Pavan, R. Bez, P. Olivo, and E. Zanoni, “Flash memory cells - an overview,” *Proceedings of the IEEE*, vol. 85, no. 8. pp. 1248-1271, Aug. 1997.

[10] R. Bez, E. Camerlinghi, A. Modelli, and A. Visconti, “Introduction to flash memory,” *Proceedings of the IEEE*, vol. 91, no. 4, pp. 489-502, Apr. 2003.

[11] D. Shmidt, “TrueFFS: Wear-leveling mechanism,” M-Systems, Technical note (TN-DOC-017), May 2002.

[12] I. Iliadis, “Rectifying pitfalls in the performance evaluation of flash solid-state drives,” *Performance Evaluation*, vol. 79, pp. 235-257, Sep. 2014.

[13] P. Desnoyers, “Analytic models of SSD write performance,” ACM Transactions on Storage, vol. 10, no. 2, pp. 8:1-8:25, Mar. 2014.

[14] A. Kawaguchi, S. Nishioka, and H. Motoda, “A flash-memory based file system,” in *Proceedings of USENIX Technical Conference (TCON)*, Jan. 1995, pp. 155-164.

[15] B. Van Houdt, “Performance of garbage collection algorithms for flash-based solid-state drives with hot/cold data,” *Performance Evaluation*, vol. 70, no. 10, pp. 692-703, Oct. 2013

[16] X.-Y. Hu, E. Eleftheriou, R. Haas, I. Iliadis, and R. Pletka, “Write amplification analysis in flash-based solid state drives,” in *Proceedings of ACM International Systems and Storage Conference (SYSTOR): The Israeli Experimental Systems Conference*, May 2009, pp. 191-202.

[17] B. Van Houdt, “A mean field model for a class of garbage collection algorithms in flash-based solid state drives,” *ACM SIGMETRICS Performance Evaluation Review*, vol. 41, no. 1, pp. 191-202, Jun. 2013.

[18] M. Rosenblum and J. K. Ousterhout, “The design and implementation of a log-structured file system,” *ACM Transactions on Computer Systems (TCOS)*, vol. 10, no. 1, pp. 26-52, Feb. 1992.

[19] M.-L. Chiang and R.-C. Chang, “Cleaning policies in mobile computers using flash memory,” *Journal of Systems and Software*, vol. 48, no. 3, pp. 213-231, Nov. 1999.

[20] A. Soga, C. Sun, and K. Takeuchi, “NAND flash aware data management system for high-speed SSDs by garbage collection overhead suppression,” in *Proceedings of IEEE International Memory Workshop (IMW)*, May 2014, pp. 95-98.

[21] C. Sun, A. Soga, C. Matsui, A. Arakawa, and K. Takeuchi, “LBA scrambler: A NAND flash aware data management scheme for high-performance solid-state drives,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 24, no. 1, pp. 115-128, Jan, 2016.

[22] C. Matsui, A. Arakawa, C. Sun, and K. Takeuchi, “Write order-based garbage collection scheme

for an LBA scrambler integrated SSD," *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 25, no. 2, pp. 510-519, Feb. 2017.

第7章 結論

7.1 結論

本論文では不揮発性半導体メモリを複数種用いてデータを管理，保存するヘテロジニアスストレージシステムを提案した．データを保存する不揮発性の記憶媒体として，M-SCM，S-SCM および MLC，TLC NAND 型フラッシュメモリを用いる．これらの不揮発性半導体メモリの特性およびストレージアプリケーションのデータの特性を考慮し，階層構造を持つ不揮発性半導体メモリに保存する. 頻繁にアクセスされるデータは高速な M-SCM あるいは S-SCM に保存し，頻繁にアクセスされないデータは低速だが大容量な MLC あるいは TLC NAND 型フラッシュメモリに保存する．

三種以上の不揮発性半導体メモリを用い，二種類のヘテロジニアスストレージを提案した．また用いる不揮発性半導体メモリの特性に適したデータ管理アルゴリズムを提案した．第一の SCM, MLC および TLC NAND 型フラッシュメモリを用いたヘテロジニアスストレージは，SCM および MLC NAND 型フラッシュメモリを用いたハイブリッドストレージと比較して，MLC NAND 型フラッシュメモリに滞留するアクセス頻度の低いデータを TLC NAND 型フラッシュメモリに保存することで MLC NAND 型フラッシュメモリの書き換え回数を削減することを実現した．第二の M-SCM，S-SCM および NAND 型フラッシュメモリを用いたヘテロジニアスストレージは, SCM を二種類用いて極端にアクセス頻度の高いデータを M-SCM に，ややアクセス頻度の高いデータを S-SCM に保存することで高い性能を達成した．

読み出し・書き込み量の多寡，平均データアクセス頻度，平均データサイズなどの特性が異なるストレージアプリケーションに対して，適切な不揮発性半導体メモリの組み合わせが異なることを示した．不揮発性半導体メモリの特性および容量を変え，ベンチマークとするストレージアプリケーションを用いた評価を行なった．その結果，読み出し・書き込み量の多寡および平均データサイズ（ランダム・シーケンシャル）と比較して，ストレージアプリケーションの平均データアクセス頻度（ホット・コールド）が，ヘテロジニアスストレージの最適な構成の決定に重要であることを明らかにした．さらに高速な M-SCM を大容量用いるほどヘテロジニアスストレージの性能が向上することが明らかとなった．しかし単位容量

当たりの M-SCM のコストは，NAND 型フラッシュメモリと比較して約 10 倍と予想されるため，本論文では MLC NAND 型フラッシュメモリのみを用いたストレージのコストと比較して，ヘテロジニアスストレージは 1.5 倍のコスト増が許容できると仮定した．その結果，一部のデータが頻繁に書き換えられるストレージアプリケーション（prxy_0, proj_0）に対しては，S-SCM を大容量用いることで性能を向上できることを明らかにした．特に頻繁に書き換えられるデータの多いアプリケーション（prxy_0）については，M-SCM を極小容量用いて書き込み性能を向上できることを明らかにした．また一部のデータが頻繁に読み出し・書き込みされるストレージアプリケーション（prxy_1）に対しては，M-SCM を大容量用いることで性能を向上できることを明らかにした．さらに高速な M-SCM を用いるよりも，低速大容量な TLC NAND 型フラッシュメモリを用いるべきストレージアプリケーション（hm_0）が存在することを明らかにした．一方で，頻繁に上書きおよび読み出しされないデータの多いアプリケーション（proj_3，src2_2，src2_1）に対しては，小容量で高速な M-SCM を書き込みバッファとして機能させることが良いことを明らかにした．

続いて，SCM および MLC NAND 型フラッシュメモリを用いたハイブリッドストレージの信頼性と性能との関係をシステムレベルで理解した．SCM のエラーは BCH 符号で訂正し，NAND 型フラッシュメモリのエラーは BCH あるいは LDPC 符号で訂正する．ホット・ランダムアプリケーション（prxy_0，prxy_1）に対して，大容量の SCM は性能を大きく向上させるが，SCM に適用できる ECC 強度は弱くなることを明らかにした．これと比較して，大容量の SCM を用いるハイブリッドストレージでは多量のデータが SCM で処理されるため，復号時間の長い LDPC 符号を NAND 型フラッシュメモリに適用できることを明らかにした．小容量の SCM は NAND 型フラッシュメモリのアクセス頻度を効果的に低減するため，EP-LDPC w/o upper/lower cells や Quick-LDPC のような復号時間の短い LDPC 符号を適用でき，一方で SCM へのアクセスが減るため強い BCH 符号を SCM に適用できることを明らかにした．

最後にストレージアプリケーションに対して必要な SCM の容量を自動調整する手法を提案した．最適な SCM 容量は従来，手動で決定された．しかし提案手法を用いることで，データセンターで動作するアプリケーション毎に自動で最適な SCM 容量を管理することができることを示した．M-SCM および MLC NAND 型フラッシュメモリを用いたハイブリッドストレージにおいて，MLC NAND 型フラッシュメモリに保存されたデータのアクセス頻度を管理することで調整すべき SCM の容量を判断することを提案した．その結果頻繁にアクセスされるデータが SCM に保存され，ストレージ性能を低下させることなく SCM の容量を調整できることを示した．M-SCM 容量を 10%に固定する場合と比較して容量自律調整手法を用いる

ことで，IOPS 性能が 4.8%低下するだけでストレージ動作中の総コストが最大で 42%低減することを示した．SCM の自律容量調整手法に加えて SCM でのアクセス回数を考慮した eviction を行うことで, ストレージ動作中の総コストがさらに 6.4%削減できることを示した．

7.2 今後の展望

本研究により，異種の不揮発性半導体メモリを用いたヘテロジニアスストレージのアプリケーションに最適な構成を示すことができた．本論文では MSR Cambridge のブロック I/O トレースを提案したヘテロジニアスストレージ性能評価に用いた．今後，ディープラーニングや IoT 機器からの実トレースおよび擬似的に作成した synthetic workload を用い，真にアプリケーションに最適なストレージを示す．ディープラーニングのワークロードは本論文で用いた MSR Cambridge の src2_2 および src2_1 アプリケーションに近い特性を持つため，類似のストレージ性能を示すと考えられる．

また, NAND 型フラッシュメモリのさらなるビットコスト削減のため, 3 次元積層の NAND 型フラッシュメモリが開発されている．平面型の NAND 型フラッシュメモリと比較して，平均書き込み時間は短く，高い書き換え回数を持つ．NAND 型フラッシュメモリの特性の違いによって，これらを組み合わせたヘテロジニアスストレージの性能も異なると予想できる．今後は 3D NAND 型フラッシュメモリを用いたヘテロジニアスストレージに最低なデータマネジメントアルゴリズムを提案することが今後の課題である．

本論文の第 4 章では，一つの半導体不揮発性メモリの組み合わせに対して一つのデータ管理手法を提案した．しかし M-SCM あるいは S-SCM をキャッシュあるいはストレージとして用いるべきかは，アプリケーション特性によって異なると考えられる．そのため一つの半導体不揮発性メモリの組み合わせに対して，本論文で提案した手法とは異なるデータマネジメントアルゴリズムを提案することが必要となる．

本論文では三種の不揮発性半導体メモリを用いたストレージをヘテロジニアスストレージと呼んだが，四種以上の不揮発性半導体メモリを用いることも考えられる．第 4.3 節で議論した prxy_0 アプリケーションは，ひとつのアプリケーション中の異なるアクセス頻度を持つデータを，M-SCM，S-SCM および MLC NAND 型フラッシュメモリにそれぞれ保存することができた．prxy_0 アプリケーションに対してはさらに TLC NAND 型フラッシュメモリを追加することで，三種のメモリの場合に MLC NAND 型フラッシュメモリに保存されていたフローズンデータを分離して保存することができると考える．しかし，その他の特性を持つアプリ

ケーションに対してはデータがメモリ間で循環するため，用いる不揮発性半導体メモリ種の数を増やす利点はないと考える．

MRAM および ReRAM と比較して，PRAM の書き込み時間はその読み出し時間より長い．三種あるいは四種以上の不揮発性半導体メモリを用いたヘテロジニアスストレージにおいて，MRAM および ReRAM はキャッシュあるいはストレージとして同時に用いることができる一方，PRAM はその書き込み速度のために書き込みの少ないアプリケーションに対して用いることがよいと考える．一方本論文で用いた LPDDR2 を模擬した I/O 周波数および BCH ECC は MRAM の書き込み・読み出し速度と比較して長いため，MRAM をストレージとして用いる場合のインターフェースおよび ECC の最適化が必要となる．

さらに，ストレージアプリケーションに適した SCM 容量調整手法について，ストレージアプリケーションの時間的局所性をモニタするなどして得た情報を用いて機械学習の手法により最適化を行うことも考えたい．さらに SCM 容量の減少を行い，および不揮発性半導体メモリのビットコストを考慮したヘテロジニアスストレージおよび SCM 容量調整手法を用いることで，さまざまなストレージアプリケーションに多品種のストレージを開発することが不要になると考える．

研究業績

学術誌発表論文

筆頭

1) C. Matsui, A. Arakawa, C. Sun, and K. Takeuchi, “Write order-based garbage collection scheme for an LBA scrambler integrated SSD,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 25, no. 2, pp. 510-519, Feb. 2017.

2) C. Matsui, T. Yamada, Y. Sugiyama, Y. Yamaga, and K. Takeuchi, “Optimal memory configuration analysis in tri-hybrid solid-state drives with storage class memory and multi-level cell/triple-level cell NAND flash memory,” *Japanese Journal of Applied Physics (JJAP)*, vol. 56, no. 4S, pp. 04CE02-1 - 04CE02-9, Apr. 2017.

3) **(Invited)** C. Matsui, C. Sun, and K. Takeuchi, “Design of hybrid SSDs with storage class memory and NAND flash memory,” *Proceedings of the IEEE*, vol. 105, no. 9, pp. 1812-1821, Sep. 2017.

4) C. Matsui, R. Kinoshita, and K. Takeuchi, “Analysis on applicable ECC strength of SCM and NAND flash in hybrid storage,” *Japanese Journal of Applied Physics (JJAP)*, to be published in vol. 57, no. 4S, Apr. 2018.

共著

5) C. Sun, A. Soga, C. Matsui, A. Arakawa, and K. Takeuchi, “LBA scrambler: A NAND flash aware data management scheme for high-performance solid-state drives,” *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, vol. 24, no. 1, pp. 115-128, Jan. 2016.

6) T. Yamada, C. Matsui, and K. Takeuchi, “Workload-based co-design of non-volatile cache algorithm and storage class memory specifications for storage class memory/NAND flash hybrid SSDs,” *IEICE Transactions on Electronics*, vol. E100-C, no. 4, pp. 373-381, Apr. 2017.

7) H. Takishita, Y. Adachi, C. Matsui, and K. Takeuchi, “Analysis of SCM-based SSD performance in consideration of SCM access unit size, write/read latencies and application request size,” *IEICE Transactions on Electronics*, to be published in vol. E101-C, no. 4, Apr. 2018.

8) Y. Yamaga, C. Matsui, Y. Sakaki, and K. Takeuchi, “Reliability analysis of scaled NAND flash

memory based SSDs with real workload characteristics by using real usage-based precise reliability test," *IEICE Transactions on Electronics*, to be published in vol. E101-C, no. 4, Apr. 2018.

9) Y. Sakaki, T. Yamada, C. Matsui, Y. Yamaga, and K. Takeuchi, "Performance analysis of 3D-triple-level cell and 2D multi-level cell NAND flash hybrid solid-state drives," *Japanese Journal of Applied Physics (JJAP)*, to be published in vol. 57, no. 4S, Apr. 2018.

国際会議発表分

筆頭

1) C. Matsui, A. Arakawa, C. Sun, T. O. Iwasaki, and K. Takeuchi, "3x faster speed solid-state drive with a write-order based garbage collection scheme," in *Proceedings of IEEE International Memory Workshop (IMW)*, Monterey, May 2015, pp. 153-156.

2) C. Matsui, T. Yamada, S. Okamoto, C. Sun, and K. Takeuchi, "Application-dependent SCM/NAND flash hybrid solid-state drive design," in *Flash Memory Summit (FMS)*, Santa Clara, Aug. 2016.

3) C. Matsui, Y. Yamaga, Y. Sugiyama, and K. Takeuchi, "8.9-times performance improvement by tri-hybrid storage system with SCM and MLC/TLC NAND flash memory," in *Extended Abstracts of International Conference on Solid State Devices and Materials (SSDM)*, Tsukuba, Sep. 2016, pp. 105-106.

4) C. Matsui, T. Yamada, Y. Sugiyama, Y. Yamaga, and K. Takeuchi, "Tri-hybrid SSD with storage class memory (SCM) and MLC/TLC NAND flash memories," in *Flash Memory Summit (FMS)*, Santa Clara, Aug. 2017.

5) C. Matsui and K. Takeuchi, "22% higher performance, 2x SCM write endurance heterogeneous storage with dual SCM and NAND flash memory," in *Proceedings of European Solid-State Device Research Conference (ESSDERC)*, Leuven, Sep. 2017, pp. 6-9.

6) C. Matsui and K. Takeuchi, "Error-correction & set/reset verify strategy of storage class memory (SCM) for SCM/NAND flash hybrid and all-SCM storage," in *Extended Abstracts of International Conference on Solid State Devices and Materials (SSDM)*, Sendai, Sep. 2017, pp. 783-784, poster presentation.

7) **(Invited)** C. Matsui and K. Takeuchi, “Heterogeneous storage with storage class memories and NAND flash memory for big and fast data processing,” in *Proceedings of Symposium of Phase Change Oriented Science (PCOS)*, Atami, Nov. 2017, pp. 27-28.

8) C. Matsui and K. Takeuchi, “Application-optimized non-volatile memory combination in heterogeneously-integrated disaggregated storage,” in *Non-Volatile Memories Workshop (NVMW)*, San Diego, Mar. 2018, poster presentation.

9) C. Matsui and K. Takeuchi, “3ASCA: Application-aware autonomous SCM capacity adjustment for SCM and NAND flash pooled storage,” to be presented in *IEEE International Symposium on Circuits and Systems (ISCAS)*, Florence, May 2018.

共著

10) C. Sun, A. Arakawa, A. Soga, C. Matsui, and K. Takeuchi, “Middleware and flash translation layer co-design for the performance boost of solid-state drives,” in *Flash Memory Summit (FMS)*, Santa Clara, Aug. 2015.

11) Y. Sugiyama, T. Yamada, C. Matsui, T. Onagi, and K. Takeuchi, “Application dependency of 3-D integrated hybrid solid-state drive system with through-silicon via technology,” in *Proceedings of International Conference on Electronics Packaging (ICEP)*, Sapporo, Apr. 2016, pp. 79-82.

12) Y. Yamaga, C. Matsui, S. Hachiya, and K. Takeuchi, “Application optimized adaptive ECC with advanced LDPCs to resolve trade-off among reliability, performance, and cost of solid-state drives,” in *Proceedings of IEEE International Memory Workshop (IMW)*, Paris, May 2016, pp. 129-132.

13) T. Yamada, C. Matsui, and K. Takeuchi, “Optimal combinations of SCM characteristics and non-volatile cache algorithm for high-performance SCM/NAND flash hybrid SSD,” in *Proceedings of IEEE Silicon Nanoelectronics Workshop (SNW)*, Honolulu, Jun. 2016, pp. 88-89, poster presentation.

14) T. Yamada, A. Suzuki, Y. Sugiyama, C. Matsui, and K. Takeuchi, “Comprehensive analysis on SCM specifications for high-performance SCM/NAND flash hybrid SSD with through-silicon via,” in *Proceedings of International Conference on Electronics Packaging (ICEP)*, Yamagata, Apr. 2017, pp. 268-271.

15) Y. Yamaga, C. Matsui, Y. Sakaki, A. Kobayashi, and K. Takeuchi, “Real usage-based precise

reliability test by extracting read/write/retention-mixed real-life access of NAND flash memory from system-level SSD emulator," in *Proceedings of IEEE International Reliability Physics Symposium (IRPS)*, Monterey, Apr. 2017, pp. PM-12.1 - PM-12.5, poster presentation.

16) Y. Sugiyama. T. Yamada, C. Matsui, and K. Takeuchi, "Reconfigurable SCM capacity identification method for SCM/NAND flash hybrid disaggregated storage," in *Proceedings of IEEE International Memory Workshop (IMW)*, Monterey, May 2017, pp. 60-63, poster presentation.

17) Y. Sakaki, T. Yamada, C. Matsui, Y. Yamaga, and K. Takeuchi, "23% higher performance of 2D MLC/3D TLC NAND flash hybrid solid-state drive," in *Extended Abstracts of International Conference on Solid State Devices and Materials (SSDM)*, Sendai, Sep. 2017, pp. 183-184.

18) M. Nakanishi, Y. Adachi, C. Matsui, Y. Sugiyama, and K. Takeuchi, "Application-oriented wear-leveling optimization of 3D TSV-integrated storage class memory-based solid-state drives," to be presented in *International Conference on Electronics Packaging and iMAPS All Asia Conference (ICEP-IAAC)*, Kuwana, Apr. 2018.

19) Y. Adachi, C. Matsui, and K. Takeuchi, "Double asymmetric-latency storage class memories (SCMs) for fast-write SCM, fast-read SCM & NAND flash hybrid SSDs," to be presented in *International Symposium on VLSI Design, Automation and Test (VLSI-DAT)*, Hsinchu, Apr. 2018.

20) M. Fukuchi, Y. Sakaki, C. Matsui, and K. Takeuchi, "20% system-performance gain of 3D charge-trap TLC NAND flash over 2D floating-gate MLC NAND flash for SCM/NAND flash hybrid SSD," to be presented in *IEEE International Symposium on Circuits and Systems (ISCAS)*, Florence, May 2018.

国内会議発表分

筆頭

1) **（依頼講演）**松井 千尋，荒川 飛鳥，孫 超，竹内 健，"ガベージコレクション最適化による LBA scrambler を使用した SSD の高速化"，電子情報通信学会 集積回路研究会，信学技報，vol. 116, no. 3, ICD2016-04, pp. 65-69, 2016 年 4 月.

2) 松井 千尋，孫 超，竹内 健，"LBA scrambler を用いた SSD の高速化"，電子情報通信学

会 集積回路研究会，LSI とシステムのワークショップ 2016，2016 年 5 月，ポスター発表.

3) 松井 千尋, 山賀 祐典, 杉山 佑輔, 竹内 健, “半導体ストレージシステムにおける SCM, MLC/TLC NAND フラシュメモリの最適な構成の設計”, 電子情報通信学会 集積回路研究会，デザインガイア 2016 VLSI 設計の新しい大地，信学技報，vol. 116, no. 334, ICD2016-11, pp. 7-10, 2016 年 11 月.

4) 松井 千尋，杉山 佑輔，竹内 健，“ストレージクラスメモリおよび NAND フラシュメモリを用いたハイブリッドストレージのアプリケーション依存性”，情報処理学会 システムと LSI の設計技術研究会，DA シンポジウム，DA シンポジウム 2017 論文集，pp. 109-110，2017 年 8 月，ポスター発表.

共著

5) 山賀 祐典，松井 千尋，竹内 健，“SSD の性能と信頼性を考慮したデータアクセスパターンに適した ECC システム”，情報処理学会 システムソフトウエアとオペレーティングシステム研究会，第 28 回コンピュータ・システム研究会（ComSys2016），2016 年 11 月，ポスター発表.

6) 杉山 佑輔，山田 知明，松井 千尋，小名木 貴裕，竹内 健，“TSV を用いた 3 次元実装ハイブリッド SSD のアプリケーション依存性”，情報処理学会 システムソフトウエアとオペレーティングシステム研究会, 第 28 回コンピュータ・システム研究会（ComSys2016），2016 年 11 月，ポスター発表.

7) 山賀 祐典, 松井 千尋, 竹内 健, “アプリケーションに適用可能なエンタープライズ SSD 向け ECC システム”，第 64 回応用物理学会春季学術講演会，2017 年 3 月，ポスター発表.

8) 木下 怜佳，松井 千尋，山賀 祐典，安達 優，竹内 健，“SCM/NAND 型フラッシュハイブリッド SSD のワークロード特性に応じた SCM のエラー救済手法”，情報処理学会 システムソフトウエアとオペレーティングシステム研究会，第 29 回コンピュータ・システム研究会（ComSys2017），2017 年 12 月，ポスター発表.

9) 山賀 祐典，松井 千尋，榊 佑季哉，竹内 健，“リアルワークロードを用いた NAND 型フラッシュメモリの信頼性評価”，第 65 回応用物理学会春季学術講演会，2018 年 3 月，ポスター発表.

10) 榊 佑季哉，松井 千尋，山賀 祐典，竹内 健，“3D-TLC NAND 型フラッシュメモリを用いたハイブリッド SSD の性能評価”, 第 65 回応用物理学会春季学術講演会, 2018 年 3 月，ポスター発表.

11) 鈴木 敦也，杉山 佑輔，松井 千尋，竹内 健，“TSV を用いた SCM/ NAND 型フラッシュメモリのハイブリッド SSD における SCM の仕様の評価”，第 65 回応用物理学会春季学術講演会，2018 年 3 月，ポスター発表.

12) 木下 怜佳，松井 千尋，杉山 佑輔，安達 優，竹内 健，“2 階層のストレージ・クラス・メモリシステムの性能評価”，電子情報通信学会総合大会 ISS 特別企画「学生ポスターセッション」，2018 年 3 月，ポスター発表.

13) 中西 優，安達 優，松井 千尋，杉山 佑輔，竹内 健，“ストレージ・クラス・メモリで構成した SSD の寿命を考慮した性能評価”，電子情報通信学会総合大会 ISS 特別企画「学生ポスターセッション」，2018 年 3 月，ポスター発表.

14) 福地 守，松井 千尋，榊 佑季哉，竹内 健，“3 次元構造チャージトラップ型メモリで構成されるハイブリッド SSD の特性解析”，電子情報通信学会総合大会 ISS 特別企画「学生ポスターセッション」，2018 年 3 月，ポスター発表.

謝辞

指導教員 竹内 健 教授に心より御礼申し上げます. 博士論文研究を行なう貴重な機会を頂き，日々ご指導くださりありがとうございました．博士論文をまとめるにあたり，ご指導をいただきました山村 清隆 教授，築山 修治 教授，首藤 一幸 准教授に心より御礼申し上げます.

本研究を行なうにあたり，多くの方々からご指導，ご助力，ご支援を賜りました．心より御礼申し上げます．ありがとうございました.

孫 超 博士，Tomoko Ogura Iwasaki さん，蜂谷 尚悟さんに御礼申し上げます．研究内容にかかわる多くの点でご指導いただきました.

竹内研究室 渋谷 弘枝 秘書に御礼申し上げます. 多くの事務手続きのご助力をいただきました.

プロファウンド・デザイン・テクノロジー株式会社 塚本 泰隆 代表取締役に御礼申し上げます．データ管理手法の実装にご助力をいただきました.

株式会社富士通研究所 小川 淳二さん，吉田 英司 博士，風間 哲さん，桑村 慎哉さん，日本電気株式会社 吉川 隆士 博士，Little Wing, LLC 菅 真樹さんに多大なご助言をいただきました．ここに御礼申し上げます.

竹内研究室 修士課程および学部学生のみなさまに御礼申し上げます.

常に支えてくれた家族に，心より御礼申し上げます.

2018 年 1 月 23 日

松井 千尋