

RC-015

多レベルセルフラッシュメモリに対する多元LDPCの構成と評価

Construction and Evaluation of Non-Binary LDPC Codes for Multilevel Cell Flash Memories

前田遊*
Yuu Maeda金子晴彦*
Haruhiko Kaneko

1 はじめに

フラッシュメモリは、高記録密度を有する不揮発性半導体メモリであり、デジタルカメラ、USBメモリなど多岐にわたる分野で使用されている。また、メモリ素子の記憶密度向上によりハードディスクドライブ(HDD)の代替として、フラッシュメモリを用いたソリッドステートドライブ(SSD)が開発されている。フラッシュメモリはメモリセルの浮遊ゲートに蓄える電荷の量によってデータの値を保持する。従来の単一レベルセル(SLC)においては2値データを保持するのに対し、多レベルセル(MLC)は $Q = 2^b$ 値のデータ、すなわち **b** ビットのデータを保持することが可能である。これによりMLCはSLCよりも高い記憶密度を持つ。一方、MLCはSLCと比較して誤りの発生確率が高い。これは浮遊ゲートに蓄えられた電荷量や制御ゲート電圧のわずかな変化により記憶データを誤った値として読み出してしまふためである。従来、SLCフラッシュメモリや **$b = 2$** のMLCフラッシュメモリに対しては、誤り制御符号としてHamming符号や **t** ビット誤り訂正(t は3から7程度) BCH符号が適用されている。しかし、セルのレベル数 **Q** が増加すると、それに伴い閾値電圧の間隔が小さくなり誤りの確率が急激に上昇することが予想される。よってMLCにおける誤りの特性を考慮した、より高い誤り制御能力を有する符号の設計が必要である。

本稿では、MLCフラッシュメモリにおいて、メモリセルの閾値電圧が正規分布に従うと仮定して、MLCにおける誤りを多元低密度パリティ検査(LDPC:Low-Density Parity Check)符号を用いて訂正する手法を示す。また、シミュレーションにより復号後のビット誤り率(BER)の評価を行う。

本稿の構成は以下のとおりである。2章でフラッシュメモリセルの構造とMLCにおける誤りについて述べる。3章で多元LDPC符号の構成及び復号について述べ、4章で評価を示す。5章で結論と今後の課題について述べる。

2 フラッシュメモリセルの構造と誤り

2.1 メモリセルの構造

フラッシュメモリのメモリセルは、図1に示すようにMOSトランジスタにデータを記憶させるための浮遊ゲートを追加した構造を有する。メモリセルへの書き込みは制御ゲートに高電圧を加えて浮遊ゲートに電荷を挿入することにより行い、消去は基板に高電圧を加え浮遊ゲートから電荷を除去することにより行う。この動作を図2に示す。

メモリセルからの読み出しは、浮遊ゲート内の電荷量に応じて制御ゲート電圧 V_{CG} とドレインソース電流 I_D の関係が変化することを用いて行う。浮遊ゲート内に電荷がない場合には V_{CG} が小さくても I_D が大きくなるのに対し、電荷がある場合には低い V_{CG} に対してほぼ0になることを用いて読み出し電圧 V_R を設定し、保持されているデータを読み出す。

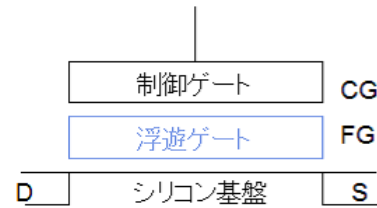


図1 メモリセルの構造

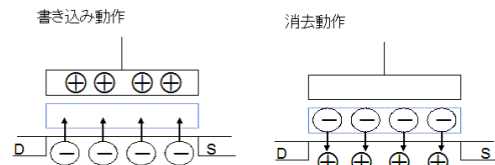


図2 セルの書き込み・消去動作

2.2 多レベルセルにおける非対称誤り

MLCにおいては浮遊ゲートの電荷量を **2^b** 段階に制御することによって **b** ビットの情報表現する。浮遊ゲート内の電荷量が少ない場合、制御ゲート電圧 V_{CG} が低いときでもドレインソース電流 I_D が大きくなるのに対し、浮遊ゲート内の電荷量が多い場合、電圧 V_{CG} を大きくしなければ電流 I_D が流れない。このことから V_{CG} の値を変えて電流を識別することにより、浮遊ゲート内の電荷量を推定することができる。例として4レベルのセルにおける制御ゲート電圧とドレインソース電流の関係を図3に示す。

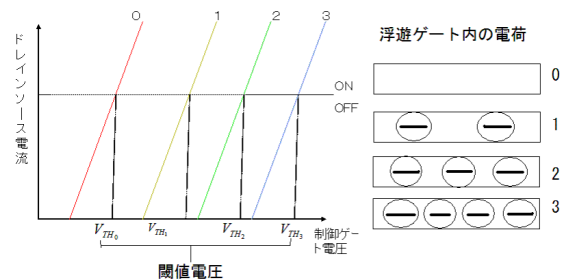


図3 4レベルセルにおける制御ゲート電圧とドレインソース電流の関係

以下では、浮遊ゲートが $Q = 2^b$ 段階の電荷レベルをとるとし、電荷レベルを整数値 $\{0, 1, \dots, Q - 1\}$ によ

*東京工業大学 大学院情報理工学研究所, Graduate School of Information Science and Engineering, Tokyo Institute of Technology

り表現する。浮遊ゲートの電荷レベルが i であるときドレインソース電流 I_D がある電流 I より大きくなるような最小のゲート電圧を電荷レベル i に対する閾値電圧とよび V_{TH_i} であらわす。多レベルセルでは閾値電圧 ($V_{TH_0}, V_{TH_1}, \dots, V_{TH_{Q-1}}$) に対して $V_{TH_i} < V_{R_i} < V_{TH_{i+1}}$ となるように読み出し電圧 V_{R_i} を定める。これに従いそれぞれの読み出し電圧に対してドレインソース電流が流れるか否かを判定することでセルの値を読み出すことができる。

しかし、閾値電圧 V_{TH_i} は書き込み電荷量の誤差、セルの劣化などによって固定されずに変動するのが一般的であり、その分布は正規分布により近似する場合が多い [2]。本稿では閾値電圧 V_{TH_i} の確率密度関数が以下の式で与えられるとする。

$$P_i(x) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}}$$

ここで、 μ_i および σ_i はそれぞれ V_{TH_i} の平均および標準偏差である。

2.3 通信路行列

多レベルセルにおける誤り確率を次の通信路行列によって表現する。

$$P = \begin{bmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,Q-1} \\ p_{1,0} & p_{1,1} & \cdots & p_{1,Q-1} \\ \vdots & \vdots & \ddots & \vdots \\ p_{Q-1,0} & p_{Q-1,1} & \cdots & p_{Q-1,Q-1} \end{bmatrix}$$

ここで、 $p_{i,j}$ は電荷レベル i で情報を書き込んだとき、電荷レベル j として読み出される確率を表す。この確率 $p_{i,j}$ は以下の式により与えられる。

$$p_{i,j} = \int_{V_{R_{j-1}}}^{V_{R_j}} P_i(x) dx$$

ただし、 $P_i(x)$ は閾値電圧を表す正規分布、 $V_{R_{-1}} = -\infty, V_{R_{Q-1}} = +\infty$ とする。

例 $Q = 4$ レベルにおける通信路行列の例を以下に示す。

$$P = [p_{i,j}]_{4 \times 4} = \begin{bmatrix} p_{0,0} & 1 \times 10^{-16} & 5 \times 10^{-82} & 2 \times 10^{-199} \\ 8 \times 10^{-17} & p_{1,1} & 1 \times 10^{-16} & 3 \times 10^{-134} \\ 2 \times 10^{-134} & 1 \times 10^{-16} & p_{2,2} & 9 \times 10^{-17} \\ 8 \times 10^{-278} & 1 \times 10^{-106} & 1 \times 10^{-16} & p_{3,3} \end{bmatrix}$$

ただし、閾値電圧分布において $\sigma_0 = 0.15, \sigma_1 = 0.10, \sigma_2 = 0.12, \sigma_3 = 0.12, \mu_0 = -2.50, \mu_1 = -0.45, \mu_2 = 1.19, \mu_3 = 3.00$ とし、読み出し電圧を $V_{R_0} = -1.27, V_{R_1} = 0.37, V_{R_2} = 2.01$ と仮定したものである。また、 $p_{i,i} = 1 - \sum_{0 \leq j \leq 3, i \neq j} p_{i,j}$ である。

このように、多レベルセルにおいては隣接する電荷レベルに誤りやすく、離れた電荷レベルには誤りにくいという特性を有する非対称誤りが発生する。例として電荷レベル i のセルの値を誤識別する分布を示したものを図 4 に示す。

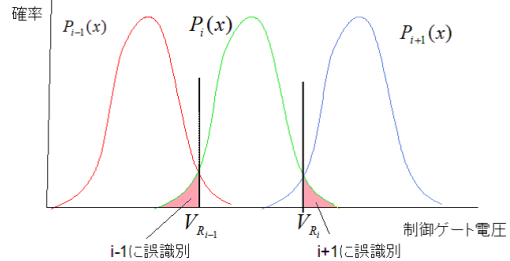


図 4 電荷レベル i を誤識別する確率

3 MLCフラッシュメモリに対する多元LDPC符号

3.1 LDPC符号

LDPC符号は、非常に疎な検査行列によって定義される線形符号である。非常に疎な行列とは、非零の要素が非常に少ない行列のことである。以下に定義を述べる。

定義 1 M 行 N 列の Q 元検査行列 H において、すべての列ハミング重みが ω_c であり、全ての行ハミング重みも ω_r であるとする。更に $\omega_c \ll N$ であるとき、その検査行列 H によって定義される符号をレギュラーLDPC符号と呼ぶ。これに対し、各行、各列の重みが一定でないLDPC符号のことをイレギュラーLDPC符号と呼ぶ。□

適切な列重み分布・行重み分布を持つイレギュラーLDPC符号は、レギュラーLDPC符号よりも優れた復号誤り特性を有することが知られている [1]。

3.2 多元LDPC符号の検査行列の構成

検査行列を構成する準備として、線形符号に対するタナーグラフの定義を示す。

定義 2 タナーグラフは線形符号の検査行列に対して定義される 2 部グラフである。いま、 $M \times N$ 行列である検査行列 H が与えられているとする。タナーグラフを構成するノードは 2 つのグループからなり、ひとつは M 個のノードからなるチェックノード、もうひとつは N 個のノードからなるメッセージノードである。タナーグラフでは検査行列 H の i 行 j 列目要素が非零ならば j 番目のメッセージノードと i 番目のチェックノードが辺で結ばれる。□

例として、二元の検査行列

$$H = \begin{bmatrix} 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

に対応するタナーグラフを図 5 に示す。

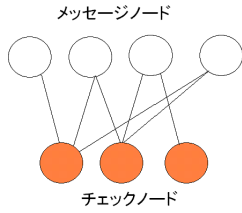


図5 タナーグラフの例

多元LDPC符号の検査行列は乱数を使って計算機により構成する。LDPC符号の検査行列を乱数を用いて生成するとき、符号長が短い場合(数百シンボル以下)は符号の復号特性に差が生じるが、符号長が長い場合にはランダムに構成しても高い確率で同程度の復号性能が得られるため、乱数を用いても問題はないとされている[3]。多元LDPC符号の検査行列は以下の条件を満たすように構成する。

- 可能な限り行重みを一定にする。
- 任意の二列の同じ行に非零の要素が存在する場合、これらの値が同じにならないようにする。
- 検査行列の定めるタナーグラフにおいて、可能な限り長さ4のループが生じないようにする。

ここでループとは、タナーグラフのあるノードから始まりそのノードで終わる周回路を意味する。タナーグラフから長さ4のループを排除することにより、次に示す復号アルゴリズムにおいて、事後確率をより正確に計算できる[5]。

3.3 多元LDPC符号の復号のアルゴリズム

Q レベルセルフフラッシュメモリに対して $q = Q$ 元LDPC符号を適用する。 q 元LDPCの復号アルゴリズムとして sum-product アルゴリズムを使用する。

q 元 $M \times N$ 行列 $H = [H_{mn}]$ を多元LDPC符号の検査行列とし、集合 $[1, N] = \{1, 2, \dots, N\}$ の部分集合 $A(m), B(n)$ を次のように定義する。

$$A(m) = \{n : H_{mn} \text{が非零}\} \quad B(n) = \{m : H_{mn} \text{が非零}\}$$

すなわち $A(m)$ は検査行列 H の m 行目において、非零の値が入っている列インデックスの集合であり、 $B(n)$ は検査行列 H の n 列目において非零の値が入っている行インデックスの集合を指す。また、受信語を $y = (y_1, y_2, \dots, y_N)$ とし、通信路として、2.3節に示す通信路行列で定義される通信路を仮定する。多元LDPC符号に対する sum-product アルゴリズムの基本的な考え方を以下に示す。

ステップ1(初期化) $H_{mn} \neq 0$ を満たすすべての組 (m, n) に対して $v_{mn}(i) = 1/q : (i = 0, 1, \dots, q-1)$ と初期設定をする。これらの v_{mn} を事前値と呼ぶ。また、ループ変数を $l = 1$ とし、ループ最大回数を変数 l_{max} に設定する。

ステップ2(行処理) $m = 1, 2, \dots, M$ の順に $H_{mn} \neq 0$ を満たす全ての組 (m, n) に対して、次の更新式を利用して $c_{mn}(i) : (i = 0, 1, \dots, q-1)$ を更新する。

$$\Omega_{mn}(i) = \left\{ t_j \in GF(q) \mid \sum_{j \in A(m) \setminus n} t_j h_{mj} = -i \right\}$$

$$c_{m,n}(i) = K \sum_{\Omega_{mn}(i)} \left(\prod_{n' \in A(m) \setminus n} v_{mn'}(t_{n'}) P(y_{n'} | t_{n'}) \right)$$

ここで、 \setminus は差集合を示し、定数 K は、 $\sum_{k=0}^{q-1} c_{mn}(k) = 1$ が成り立つように定められるものとする。これらの c_{mn} は外部値と呼ばれる。ステップ2で外部値が更新される様子を図6に示す。

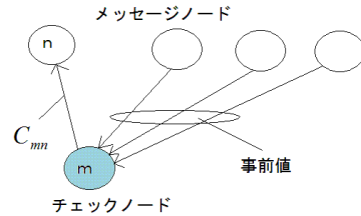


図6 外部値の更新

ステップ3(列処理) $n = 1, 2, \dots, N$ の順に $H_{mn} \neq 0$ を満たす全ての組 (m, n) に対して、次の更新式を利用して $v_{mn}(i) : (i = 0, 1, \dots, q-1)$ を更新する。

$$v_{mn}(i) = K' \prod_{m' \in B(n) \setminus n} c_{m'n}(i)$$

ここで K' は $\sum_{k=0}^{q-1} v_{mn}(k) = 1$ が成り立つように定められた定数とする。ステップ3で事前値が更新される様子を図7に示す。

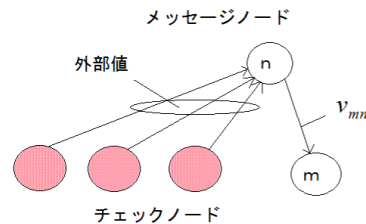


図7 事前値の更新

ステップ4(一時推定語の計算) $n = 1, 2, \dots, N$ について q 個のシンボル全てに対して、

$$R_n(i) = P(y_n | x_n = i) \prod_{m' \in B(n)} c_{m'n}(i)$$

$$(i = 0, 1, 2, \dots, q-1)$$

を計算し、その中で最も確率の高い $R_n(i)$ を与えるような i を r_n とする。ここで得られる (r_1, r_2, \dots, r_N) を一時推定語と呼ぶ。

ステップ5(パリティ検査) 一時推定語が符号語になっているかを確認する。すなわち、 (r_1, \dots, r_n) が $GF(q)$ 上で

$$(r_1, \dots, r_n)H^t = 0$$

を満たせば一時推定語を復号語として出力し、アルゴリズムを終了する。

ステップ6(ループ回数のカウント) もし $l \leq l_{max}$ ならば l をインクリメントしてステップ2にもどる。 $l > l_{max}$ ならば一時推定語を復号語として出力し、アルゴリズムを終了する。

符号のタナーグラフにループが含まれないのであれば、十分な回数の反復計算後に $R_n(a) (a \in \{0, 1, \dots, q-1\}, n \in [1, N])$ は事後確率分布 $P(x_n = a|y)$ に収束する。しかし符号長が有限長である場合、タナーグラフには必ずループが存在することから、sum-product アルゴリズムは事後確率分布の近似計算となる。また、このアルゴリズムは計算が複雑なため、復号に要する時間計算量は大きくなる。

4 評価と考察

本節では次式で定義するビット誤り率 (BER) の評価を行う。

$$BER = \frac{\text{誤りビット数}}{\text{総ビット数}}$$

4.1 ビット誤り率の評価

表1に示す閾値分布を有する8レベルセルフラッシュメモリ、及び表2に示す閾値分布を有する16レベルセルフラッシュメモリを仮定して、多元LDPC符号を適用した場合のBERを、シミュレーションにより評価した。表3にLDPC符号の評価のパラメータを示す。

表1 8レベルセルの閾値分布

	V_{TH_0}	V_{TH_1}	V_{TH_2}	V_{TH_3}
分布の平均 μ_i	-3.0000	-2.0945	-1.2795	-0.4645
	V_{TH_4}	V_{TH_5}	V_{TH_6}	V_{TH_7}
分布の平均 μ_i	0.3505	1.1655	1.9805	3.0000

表2 16レベルセルの閾値分布

	V_{TH_0}	V_{TH_1}	V_{TH_2}	V_{TH_3}
分布の平均 μ_i	-3.0000	-2.5810	-2.1870	-1.7930
	V_{TH_4}	V_{TH_5}	V_{TH_6}	V_{TH_7}
分布の平均 μ_i	-1.3990	-1.0050	-0.6110	-0.2170
	V_{TH_8}	V_{TH_9}	$V_{TH_{10}}$	$V_{TH_{11}}$
分布の平均 μ_i	0.1770	0.5710	0.9650	1.3590
	$V_{TH_{12}}$	$V_{TH_{13}}$	$V_{TH_{14}}$	$V_{TH_{15}}$
分布の平均 μ_i	1.7530	2.1470	2.5410	3.0000

シミュレーション条件は以下のとおりである。

- 符号語を100,000語生成し、復号語の誤り率を求める。

表3 多元LDPC符号の評価のパラメータ

セルのレベル数	8	16	
符号長	8000シンボル	4000シンボル	
符号化率	1/2	5/8	3/4
閾値の標準偏差 σ	0.10	0.15	0.20
	0.25	0.30	0.40
	0.50	0.60	

- q レベルセルに対して q 元LDPC符号を適用する。
- 多元LDPC符号として、列重み2,3のレギュラーLDPC符号と、列重みに2と3が1:1の割合で含まれる(列重み2.5)イレギュラーLDPC符号を用いる。
- sum-product アルゴリズムにおけるループの最大回数を $l_{max} = 200$ とする。

また、一般的な Q レベルセルフラッシュメモリにおける閾値分布を考慮し [2]、 V_{TH_i} の標準偏差 σ_i は以下のような関係を有するものと仮定する。

$$\sigma_i = \begin{cases} \sigma & (i \in \{1, 2, \dots, Q-2\}) \\ 1.2\sigma & (i = 0) \\ 1.5\sigma & (i = Q-1) \end{cases}$$

8レベルセルに対して符号化率1/2, 5/8, 及び3/4の多元LDPC符号を適用した場合のBERをそれぞれ図8, 9, 及び10に示す。また、16レベルセルに対して符号化率1/2, 5/8, 及び3/4の多元LDPC符号を適用した場合のBERをそれぞれ図11, 12, 及び13に示す。グラフにおいて、グラフの左側が途切れているものは100,000語を復号しても誤りが発生しなかったことを意味する。また、符号長4000の符号を実線、符号長8000の符号を点線で示す。8レベルセルにおいては、2元LDPC符号の復号結果も比較している。

2元LDPC符号と多元LDPC符号のBERを比較すると、 σ が小さい場合においては大きな差がないが、 σ が大きい場合は多元LDPC符号を使用した方が2元LDPC符号を使用したものよりBERが低くなる。このことから、多レベルセルのレベルが増え更にBERが高くなった状況では、多元LDPC符号は2元LDPC符号よりも低いBERを与えることが予測される。

また、列重み3以外の符号では σ が低下しても、一定よりBERが減少しにくくなるエラーフロアが生じている。

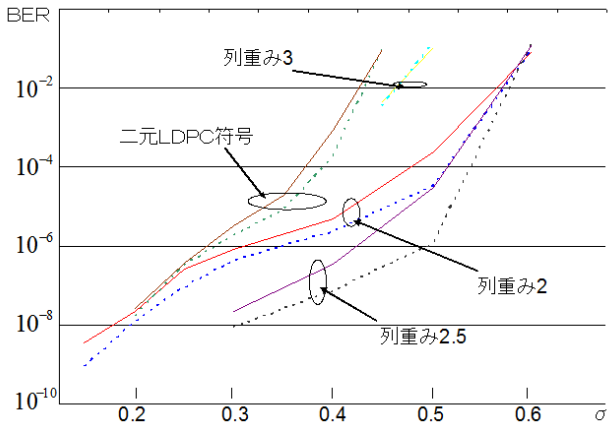


図 8 8 レベル、符号化率 1/2 の場合のBER

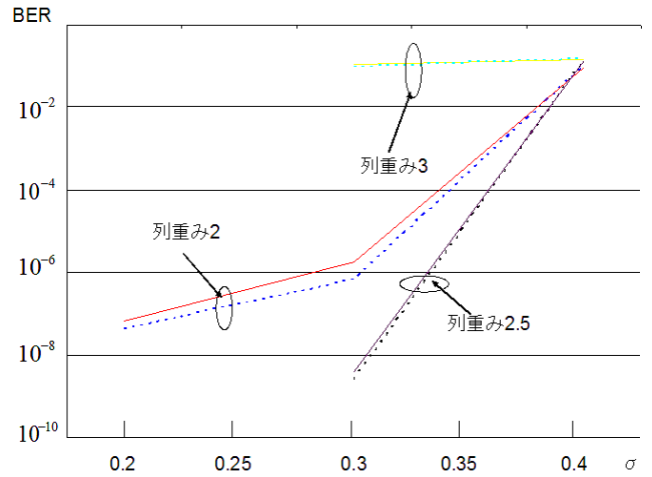


図 11 16 レベル、符号化率 1/2 の場合のBER

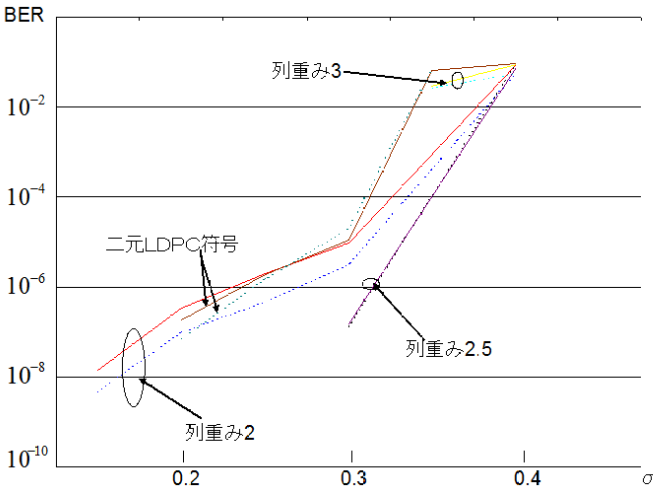


図 9 8 レベル、符号化率 5/8 の場合のBER

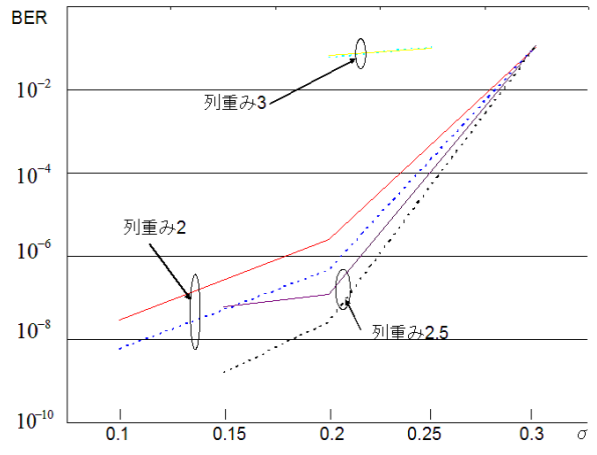


図 12 16 レベル、符号化率 5/8 の場合のBER

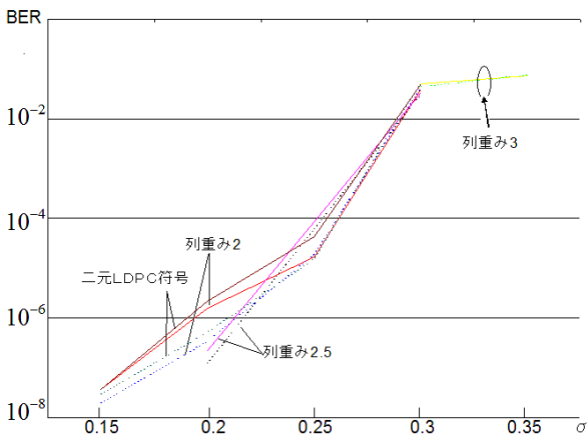


図 10 8 レベル、符号化率 3/4 の場合のBER

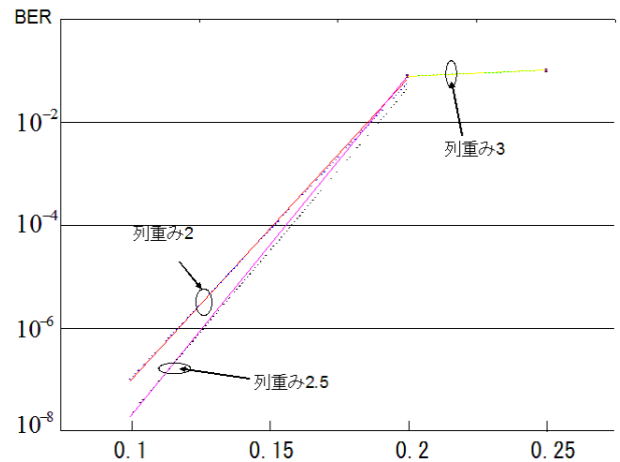


図 13 16 レベル、符号化率 3/4 の場合のBER

4.2 列重みとBERの関係

評価結果から、列重みが2の符号よりも、列重み2.5の符号のほうがより低いBERを示していることがわかる。列重みが3のものについては、 σ が低い場合では100,000語では誤りが発生せずに列重みが2のものと2.5のものよりも高い復号性能を見せたが、 σ が上がり、誤りの発生率が上昇し復号できないものが存在し始めるとBERが急激に上昇し、他の二つよりもBERが高くなる。8レベルセルにおいては重み3のLDPC符号のBERが上昇するのは $\sigma = 0.3$ 以上の場合であるが、16レベルセルにおいては符号化率が $3/8, 3/4$ のものについては $\sigma = 0.2$ でBERが高くなっている。このことから16レベルよりレベル数が増えたMLCは、列重みが3の符号より、列重み2.5のイレギュラーLDPC符号がよい性能を示すということが考えられる。

列重みとBERの関係を詳しく調べるため、列重み2,3及び4のレギュラーLDPC符号を用いた場合のBERとWERを表4に、重み2及び3の列が3:1,1:1及び1:3で存在するイレギュラーLDPC符号を用いた場合のBERとWERを表5に示す。ただし、WERとはワードの誤り率を示し、表4においては $\sigma = 0.4$ 、表5においては $\sigma = 0.5$ 、符号化率はどちらも1/2である。表5に示す比率 $x:y$ は、重み2の列の比率が x 、重み3の列の比率が y であることを示す。

その結果、列重み2及び3のレギュラーLDPC符号に比べ、列重み4のレギュラーLDPC符号が低いBER及びWERを与えることはなく、列重み比率1:1のイレギュラーLDPC符号に比べ、1:3及び3:1のイレギュラーLDPC符号が、低いBER及びWERを示すことはなかった。これは、列重みが大きくなるとLDPC符号のタナグラフに含まれる短いループの数が増えsum-product アルゴリズムによる事後確率分布の近似計算が適当な値に収束しないためだと考えられる。列重み2の場合にBERが高くなる理由としては、短いループの数は最も少ないが、それと同時に事前値および外部値の更新に使用する情報の量も少なくなるために十分な反復計算を行ったとしても、事後確率分布が適当な値に収束しないためと考えられる。

表4 列重み2,3及び4のレギュラーLDPC符号の誤り率($\sigma = 0.4$)

列重み	BER	WER
2	7.50×10^{-6}	1.29×10^{-2}
3	0	0
4	1.12×10^{-1}	9.14×10^{-1}

表5 イレギュラーLDPC符号の誤り率($\sigma = 0.5$)

列重み2と3の比率	BER	WER
3:1	3.76×10^{-5}	7.65×10^{-2}
1:1	4.58×10^{-5}	4.29×10^{-3}
1:3	8.37×10^{-2}	6.27×10^{-1}

5 おわりに

本稿では、多レベルセルフラッシュメモリの電荷レベル数が現在より増加することにより、誤り確率が急激に上昇すると予測されることから、多レベルセルフラッシュメモリに対し強力な誤り制御符号である多元LDPC符号を適用する手法を示し、シミュレーションにより誤り率の評価を行った。

その結果、多くのパラメータにおいて、列重み2.5の多元イレギュラーLDPC符号が最も低いBERを与えることを示した。また、列重み3の多元レギュラーLDPC符号では、エラーフロア現象が生じにくいことを示した。例えば、8レベルセルにおいて、符号化率が1/2、閾値分布の標準偏差が $\sigma = 0.3$ のとき、列重み2,3及び2.5のLDPC符号を適用した場合のBERはそれぞれ 4.06×10^{-7} , 0及び 7.91×10^{-9} であったのに対し、 $\sigma = 0.5$ においてはそれぞれ 3.11×10^{-5} , 1.08×10^{-1} 及び 1.06×10^{-6} であった。また、 $\sigma = 0.4$ の8レベルセルにおいて、符号化率1/2の多元LDPC符号及び2元LDPC符号のBERはそれぞれ 2.12×10^{-6} 及び 1.73×10^{-4} であり、多元LDPC符号が有効であることを示した。

今後の課題として、32レベル程度までの誤り特性の評価と、それぞれのレベルについて、接続符号の外符号の提案・評価を行う必要がある。また、既存の多元誤り訂正符号との性能比較や、多元LDPC符号の復号時間が長い場合、復号に要する時間を短くする手法の構築が必要である。

謝辞

本研究を行うに当たり、有益なご助言をいただきました藤原英二氏に心から感謝申し上げます。

参考文献

- [1] D. J. C. MacKay, "Good Error-Correcting Codes Based On Very Sparse Matrices," *IEEE Trans. Inform. Theory*, Vol.47, No.2, pp.498-519, 2001
- [2] S. Lee, et al., "A 3.3V 4Gb Four-Level NAND Flash Memory with 90nm CMOS Technology," *Proc. 2004 IEEE International Solid-State Circuits Conference*, p.2.7, 2004.
- [3] 和田山 正, "低密度パリティ検査符号とその復号法について," 電子情報通信学会技術研究報告, MR2001-83, Dec. 2001,
- [4] Alan R. Olson, Denis J. Langlois "Solid State Drives Data Reliability and Lifetime," *Imation*, 2008.
- [5] J. Moreira, et al., "Essentials of Error-Control Coding," *Wiley-interscience*, 2006.
- [6] E. Fujiwara, "Code Design for Dependable Systems, Theory and Practical Applications," *Wiley-interscience*, 2006.