EURASIP Journal on Advances in Signal Processing
a SpringerOpen Journal

RESEARCH Open Access

Low-energy error correction of NAND Flash memory through soft-decision decoding

Jonghong Kim and Wonyong Sung*

Abstract

The raw bit error rate of NAND Flash memory increases as the semiconductor geometry shrinks for high density, which makes it very necessary to employ a very strong error correction circuit. The soft-decision-based error correction algorithms, such as low-density parity-check (LDPC) codes, can enhance the error correction capability without increasing the number of parity bits. However, soft-decision error correction schemes need multiple precision data, which obviously increases the energy consumption in NAND Flash memory for more sensing operations as well as more data output. We examine the energy consumption of a NAND Flash memory system with an LDPC code-based soft-decision error correction algorithm. The energy consumed at multiple-precision NAND Flash memory as well as the LDPC decoder is considered. The output precision employed is 1.0, 1.4, 1.7, and 2.0 bits per data. In addition, we also propose an LDPC decoder-assisted precision selection method that needs virtually no overhead. The experiment was conducted with 32-nm 128-Gbit 2-bit multi-level cell NAND Flash memory and a 65-nm LDPC decoding VLSI.

Keywords: NAND Flash memory, LDPC, Low-density parity-check codes, Multi-precision sensing operation, Soft-decision decoding, Low energy

Introduction

NAND Flash memory is widely used for handheld devices and notebook PCs because of its high density and low power consumption. As the semiconductor geometry shrinks, the error performance of NAND Flash memory becomes worse, thus it is greatly needed to increase the reliability by using memory signal processing and forward-error correction (FEC) methods. Among various FEC codes, Bose-Chaudhuri-Hocquenghem (BCH) and Reed-Solomon (RS) codes have widely been used for NAND Flash error correction [1-3]. However, because of severe performance degradation of recent NAND Flash memory devices, more advanced FEC codes are needed.

Low-density parity-check (LDPC) codes [4] show excellent error correcting performance close to the Shannon-limit when decoded with the belief-propagation (BP) algorithm [5] using soft-decision information. LDPC codes have successfully been applied to many communication systems such as DVB-S2 [6], IEEE 802.3an [7], and IEEE 802.16e [8]. However, despite of good characteristics of LDPC codes, their application to NAND Flash memory is not straightforward because multiple precision output data are needed for exploiting the advantages of LDPC algorithms that show high performance with soft-decision decoding. Moreover, multiple sensing operations and delivering multiple precision data also increase the energy consumption of NAND Flash memory.

In this article, we analyze the energy consumption of a NAND Flash memory error correction system that adopts LDPC soft-decision decoding. The energy consumption of NAND Flash memory as well as that of the LDPC decoder is all considered. A VLSI circuit-based decoder for a rate-0.96 (68254, 65536) LDPC code is used for error performance and energy estimation. Especially, the effect of energy consumption when increasing the precision of NAND Flash memory is analyzed. The LDPC decoder tends to consume more energy when the precision of NAND Flash memory output is very low, such as 1.0 bit per data; however, increasing the precision also demands more energy in NAND Flash memory for sensing and data transfer. As a result, the optimum precision is closely related to the signal quality of NAND Flash memory. We analyze this relation quantitatively, and also

*Correspondence: wysung@snu.ac.kr
Department of Electrical Engineering and Computer Science, Seoul National University, Gwanak-gu, Seoul 151-744 Korea

propose a method that can find the optimum precision using the iteration count of an LDPC decoder.

The rest of this article is organized as follows. "Energy consumption of multi-bit data read in NAND Flash memory" section explains the read operation of NAND Flash memory and its energy consumption. In "Soft-decision error correcting performance in NAND Flash memory" section, the performance of LDPC decoding with multi-precision output data in NAND Flash memory is presented. "Hardware performance of (68254, 65536) LDPC decoder" section describes the energy consumption of a rate-0.96 (68254, 65536) LDPC decoder with a 65-nm technology. In "Low-energy error correction scheme for NAND Flash memory" section, we analyze the total energy consumption of NAND Flash memory with LDPC code based soft-decision decoding and also propose an LDPC decoder-assisted precision selection method. Finally, this article ends with conclusion section.

Energy consumption of multi-bit data read in NAND Flash memory

NAND Flash memory overview

A NAND Flash memory device contains thousands of cell blocks that can independently be erased. Each cell block consists of rows and columns of cells. The cells in the same row are controlled by the same word-line, and can be read or programmed simultaneously. The number of columns determines the page size, and the typical page size of the current generation of NAND Flash memory is 64 kbits (8 kbytes) besides the parity data. Each Flash memory cell is a floating gate NMOS transistor, in which the gate stores charges to control the threshold voltage of the transistor. Because of the process variation, program inaccuracy, charge leakage, and noise, the threshold voltage of NAND Flash memory has a Gaussian-like distribution. Today's NAND Flash memory adopts the multi-level cell (MLC) technology that has more than one bit per memory cell to increase the density. The organization of a 128-Gbit NAND Flash memory device with 2-bit MLC technology is shown in Table 1 [9].

Voltage sensing scheme for multi-precision output

In 2-bit MLC NAND Flash memory, each memory cell has one of four different threshold voltages that have Gaussian-like distributions as illustrated in Figure 1. The left-most distribution is the erased state (symbol *11*), while the remaining distributions correspond to three different programmed states (symbol *01*, *00*, and *10*, respectively).

Table 1 The features of 34-nm 2-bit MLC NAND Flash memory [9]

Capacity	128 Gbits
MLC tech.	2 bits/cell
Device size	8,192 blocks
Block size	256 pages
Page size	8,192 + 448 bytes

In conventional Flash memory with hard-decision data output, three sensing reference voltages (SRVs), namely, $V_{r.1}$, $V_{r.2}$, and $V_{r.3}$, are needed to fully resolve the four threshold voltage distributions. Note that $V_{r.1}$ resolves the boundary between symbols *11* and *01*, while $V_{r.2}$ is for the boundary of symbols *01* and *00*, and $V_{r.3}$ is for symbols *00* and *10*. Since a pair of LSB and MSB pages is mapped into a word-line and the bits are gray coded, $V_{r.1}$ and $V_{r.3}$ are required to read MSB pages, while only $V_{r.2}$ is needed for LSB pages. The LSB sensing operation (SO) with $V_{r.2}$ is referred to SO_1, and the MSB sensing operation with $V_{r.1}$ and $V_{r.3}$ is represented by SO_2.

For soft-decision error correction, each page should be sensed with an increased number of reference voltages. Especially, it is needed to increase the resolution in the overlapping regions, where most of bit errors are occurred, as shown in Figure 1. The simplest form of multi-bit sensing is to provide an erasure region at each symbol boundary. In this case, we need six SRVs and can obtain seven different threshold values. The lowest voltage region can be considered a strong *11* symbol, and the next lowest region is a value between *11* and *01*. Figure 1 shows four different sensing schemes, including the conventional sensing for hard-decision data output. Increasing the number of sensing operations at each symbol boundary can provide more accurate reliability information, which, however, increases the latency and energy consumption in NAND Flash memory.

Let the number of SRVs be N_s, the sensed threshold voltage belongs to one of $N_s + 1$ regions, and N_b ($= \log_2(N_s + 1)$) bits are needed to represent the threshold voltage. Hence, each bit of a page is represented by $N_b/2$ bits for 2-bit MLC NAND Flash memory. The memory sensing operations with 3, 6, 9, and 15 SRVs yield 1-, 1.4 ($= 0.5 \times \log_2(7)$)-, 1.7 ($= 0.5 \times \log_2(10)$)-, and 2 ($= 0.5 \times \log_2(16)$)-bit soft-decision bits, respectively. For example, in the 2-bit soft-decision memory sensing scheme, there exist $N_s = 15$ SRVs and 4 bits are enough to represent the 16 threshold levels for both LSB and MSB data.

Since conventional NAND Flash memory devices do not provide multi-precision data output, obtaining the soft-decision data from conventional memory requires multiple hard-decision sensing and data output operations. Note that conventional NAND Flash memory devices provide command sequences that can change the SRVs. Figure 2 illustrates the voltage sensing scheme for 1.7-bit soft-decision data output with conventional hard-decision Flash memory, where $V_{r.i}$'s are SRVs for $1 \leq i \leq 9$.

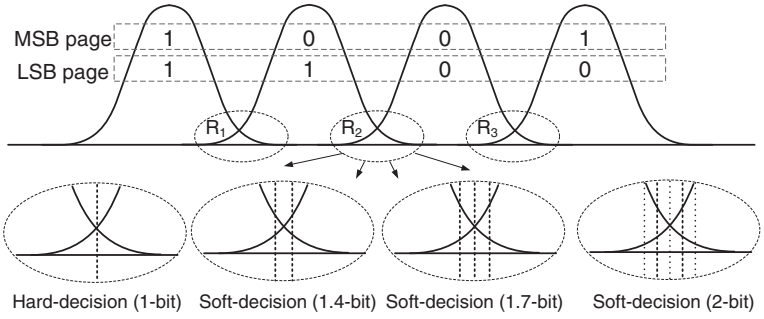

Figure 1 Voltage sensing schemes for soft-decision data output.

With hard-decision sensing SO_1 using $V_{r.5}$ and SO_2 using $V_{r.4}$ and $V_{r.6}$ around the overlapping region R_2, an LSB page is read with four levels as shown in Figure 2a. In this case, two data output operations are performed. Meanwhile, because an MSB page has two overlapping regions, R_1 and R_3, three SO_2's using $V_{r.i}$'s, where $i \in \{1, 2, 3, 7, 8, 9\}$, are needed. In addition, one SO_1 using $V_{r.5}$ is also performed to distinguish the region below $V_{r.1}$ and that above $V_{r.9}$ as illustrated in Figure 2b. As a result, in order to read an MSB page with eight levels, one SO_1 and three SO_2 are demanded, which results in four times many data output operations when compared to the conventional hard-decision mode. Finally, Table 2 summarizes the number of sensing operations for the 1-bit hard-decision and the 1.4-, 1.7-, and 2-bit soft-decision data output. Note that the sensing results are mapped to log-likelihood ratio (LLR) values by using a look-up table in the Flash memory controller.

LSB and MSB concurrent access scheme for low-energy soft-decision data output

As explained, the soft-decision scheme with conventional memory demands multiple hard-decision sensing and data transfer operations to increase the resolution in the overlapping region. Moreover, an additional LSB sensing operation is needed to access an MSB page as shown in

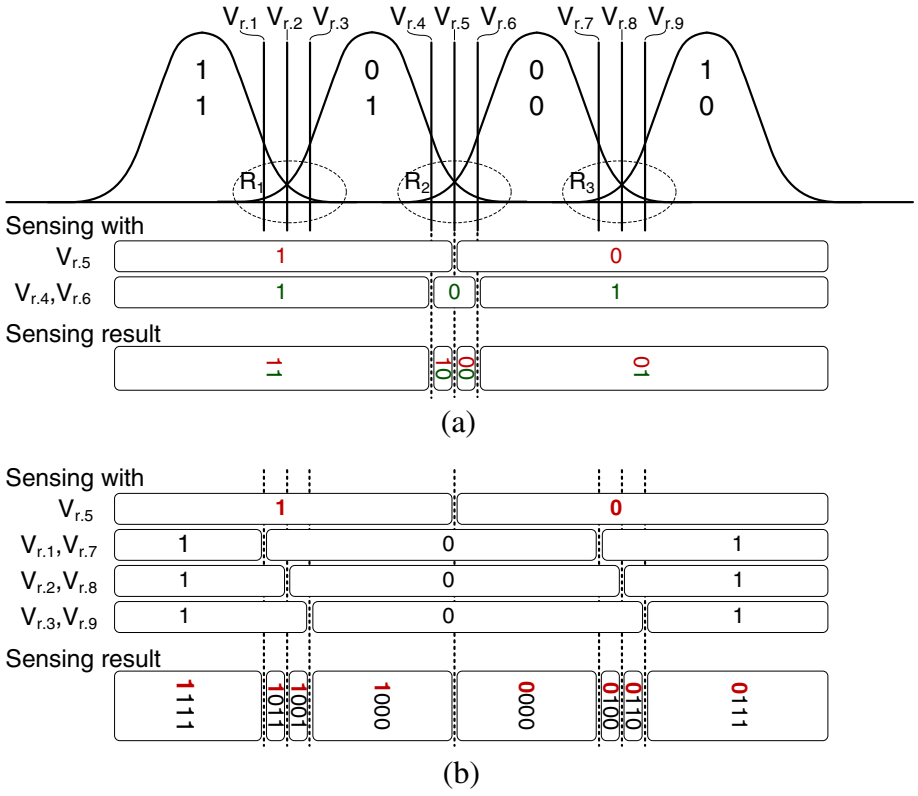

Figure 2 Voltage sensing scheme of 1.7-bit soft-decision data output for (a) LSB and (b) MSB pages.

Table 2 The number of sensing and data output (DO) operations for hard- and soft-decision sensing with conventional NAND Flash memory

	LSB pages			MSB pages		
	SO_1	SO_2	DO	SO_1	SO_2	DO
1.0-bit	1	0	1	0	1	1
1.4-bit	0	1	1	1	2	3
1.7-bit	1	1	2	1	3	4
2.0-bit	1	2	3	1	5	6

Figure 2. This scheme incurs a high amount of data output operations when high precision data are needed. In order to reduce the energy consumption of soft-decision data output, we consider a method that senses the LSB and MSB data simultaneously with multiple SRVs.

In this scheme, N_s memory sensing operations are performed for a row of transistors in the NAND Flash array, and all the sensing results are stored to the page register in $\lceil N_b \rceil$ bits, where $N_b = \log_2(N_s + 1)$. Assuming that up to 2-bit precision is used for each data, $N_b = 4$ bits are needed to represent all soft-decision sensing results. Of course, this scheme needs increased hardware of $4 \times N_{\text{pagebits}}$ data registers to store the soft-decision sensing results as shown in Figure 3, while the conventional NAND Flash memory has only N_{pagebits} data registers, where N_{pagebits} is the number of bits in each page.

When compared to the soft-decision sensing using conventional NAND Flash memory, this concurrent access scheme greatly reduces the number of data transfer operations, only $\lceil N_b \rceil$ bits for both LSB and MSB data, because the data are composed within a memory device. Thus, this method reduces not only the data output latency, but also the energy consumption for off-chip data transfer. Therefore, we only consider the LSB and MSB concurrent access scheme in this article.

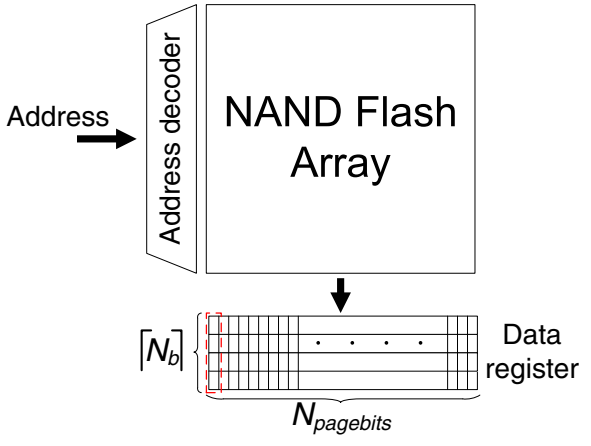

Figure 3 NAND Flash memory with internal multi-precision data composition.

Energy consumption of read operations in NAND Flash memory

The read operation of NAND Flash memory involves *address decoding*, *NAND Flash array access*, and *data output*. The conventional NAND Flash memory supports various types of read operations such as *read page* and *read page cache*, where the *read page* mode accesses only one page, while the *read page cache* mode reads the next sequential pages in a block consecutively. The timing diagram of the read page mode is illustrated in Figure 4, where t_{clk}, t_R, and t_{rc} denote the clock period, NAND Flash array access time per voltage sensing operation, and read cycle time, respectively. The array access time, t_R, includes the threshold voltage sensing operation time as well as the data transfer time from NAND Flash array to either the data or cache register.

In this section, we analyze the energy consumption of reading 2-bit MLC NAND Flash memory. We estimate the energy consumption based on the electrical specifications listed in the data book from Micron technology [9]. We model the energy consumption of reading NAND Flash memory as the sum of the energy for array access (E_{ac}) and that for data output (E_{do}), where

$$E_{\text{ac}} = V_{\text{cc}} I_{\text{cc}} t_R N_s, \tag{1}$$

$$E_{\text{do}} = V_{\text{ccq}} I_{\text{io}} t_{\text{do}}. \tag{2}$$

Note that we only concern the active energy and ignore the idle energy. V_{cc} and V_{ccq} are the core and the I/O supply voltages, while I_{cc} and I_{io} represent the core and the I/O supply currents, respectively. Finally, the data output time is represented by t_{do}, which is determined by the number of bytes to output and the period of data output clock, as a result $t_{\text{do}} = t_{\text{rc}} \times \lceil N_b \rceil \times N_{\text{pagebits}}/8$.

Since the read operation is performed simultaneously for both LSB and MSB data, the energy consumption of LSB and MSB pages is considered as follows. Let E_{LSB} and E_{MSB} be the read energy for an LSB page and an MSB page, respectively. In 2-bit MLC, reading an MSB page uses two times many SRVs than that of an LSB page access, hence the energy consumption of the array access operations for an LSB page and an MSB page can be modeled as $E_{\text{ac}}/3$ and $E_{\text{ac}} \times 2/3$, respectively. Because two pages of data are delivered simultaneously in the LSB and MSB concurrent access scheme, the data output energy of each page is modeled as $E_{\text{do}}/2$. Therefore, the energy consumption of each page can be represented as follows:

$$E_{\text{LSB}} = \frac{E_{\text{ac}}}{3} + \frac{E_{\text{do}}}{2}, \tag{3}$$

$$E_{\text{MSB}} = \frac{2}{3} E_{\text{ac}} + \frac{E_{\text{do}}}{2}. \tag{4}$$

Table 3 shows the voltage, current, and timing parameters noted in the 34-nm 2-bit MLC NAND Flash data book from Micron technology [9].

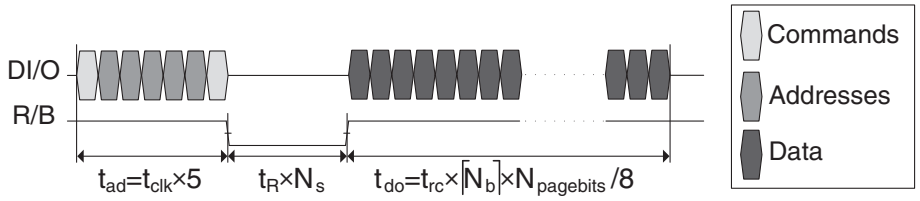

Figure 4 Timing diagram of read page mode.

Table 4 shows the estimated energy consumption and the latency of read operation for different output precision cases. Since the data output operation takes a long time due to the limited number of I/O ports, the operating condition that needs the smallest t_{rc} in the synchronous mode shows the minimum energy consumption. In this simulation, NAND Flash memory that operates at 100 MHz and V_{ccq} of 1.8 V in the synchronous mode consumes the minimum read energy. Since the energy consumption of the *read page* mode is almost similar to that of the *read page cache* mode, we only consider the read page mode of the above operating condition ($t_{clk} = 10$ ns, $V_{ccq} = 1.8$ V, and synchronous mode).

As summarized in Table 4, the 1.4-, 1.7-, and 2-bit data output of an LSB page consume 1.7, 2.4, and 3.2 times more energy, respectively, when compared to the 1-bit hard-decision data output. MSB pages consume approximately 1.5 times more energy than LSB pages.

Soft-decision error correcting performance in NAND Flash memory

In this section, we employ the MLC NAND Flash memory channel modeled in [10,11], where random telegraph noise [12], the incremental step pulse programming [13], cell-to-cell interference [14], and non-uniform quantization [15] are considered. In particular, in order to support soft-decision LDPC decoding, we adopt the LLR computation method proposed in [16], in which the four threshold voltage distributions are assumed as Gaussian distributions and the partial cumulative distribution functions of the Gaussian distribution are used to compute quantized LLRs. Thus, the LLR computation method only requires the means and the variances of the distributions obtained by performing channel estimation. Note that the LLR computation can be implemented by using a look-up table.

For the error correction in NAND Flash memory, we employ a rate-0.96 (68254, 65536) shortened Euclidean geometry (EG) LDPC code whose message size matches the page size of the 128-Gbit 2-bit MLC NAND Flash memory. The EG-LDPC codes [17] are a class of finite-geometry codes and show very low error-floor performance [18] as well as fast convergence speed [17], which are important properties for application to NAND Flash error correction.

In this study, we estimate the error performances of the NAND Flash memory channel with LDPC and BCH decoders. We assume that the erased state (symbol *11*) has a Gaussian distribution whose mean and standard deviation are 1.0 and 0.32 V, respectively, and the target programming voltages for the symbol *01*, *00*, and *10* are 2.6, 3.2, and 3.8 V, respectively. In order to generate the NAND Flash memory channel with different bit-error rates (BERs), we change the cell-to-cell coupling coefficient factor (CCF) [15,16]. The CCF primarily affects the variances of the threshold voltage distributions. Increasing the CCF results in high raw BER (RBER) because the variance of Flash memory signal becomes larger.

The error performances of a rate-0.96 (68254, 65536) EG-LDPC code and two BCH codes over the NAND

Table 3 The voltage, current, and timing parameters of 2-bit MLC NAND Flash memory

	Asynchronous	Synchronous	Unit
t_{clk}	20, 25, 30, 35, 50, 100	10, 12, 15, 20, 30, 50	ns
V_{cc}	3.3	3.3	V
V_{ccq}	1.8, 3.3	1.8, 3.3	V
I_{cc}	25	25	mA
I_{io}	8	20	mA
t_{ad}	150–450	168–288	ns
t_R	12.5	12.5	μs/sensing
t_{rc}	t_{clk}	$0.5 \times t_{clk}$	ns

Table 4 The energy consumption of a read operation for LSB and MSB pages

	E_{ac} (nJ/byte)	E_{do} (nJ/byte)	E_r (nJ/byte)
LSB pages			
1.0-bit	0.12	0.18	0.30
1.4-bit	0.24	0.27	0.51
1.7-bit	0.36	0.36	0.72
2.0-bit	0.60	0.36	0.96
MSB pages			
1.0-bit	0.24	0.18	0.42
1.4-bit	0.48	0.27	0.72
1.7-bit	0.72	0.36	1.08
2.0-bit	1.19	0.36	1.55

Flash memory channel are plotted in Figure 5a for LSB pages, where the min-sum (MS) algorithm [19] is used for low-complexity LDPC decoding. The performance of BP-based LDPC decoding is also shown for comparison. The simulation of the LDPC code is performed in floating-point arithmetic. The x-axis represents RBER and the numbers in parentheses are the corresponding signal-to-noise ratio (SNR) values, which are computed assuming a 4-pulse amplitude modulation channel with additive white Gaussian noise. The BP algorithm with infinite-bit soft-decision information yields the best error correcting performance, and the MS decoding with 1.7- and 2-bit soft-decision data output also shows good error performance fairly close the BP decoding. The (68256, 65536, 160) BCH code, which has the same code rate of 0.96, shows a worse performance than the LDPC decoder with

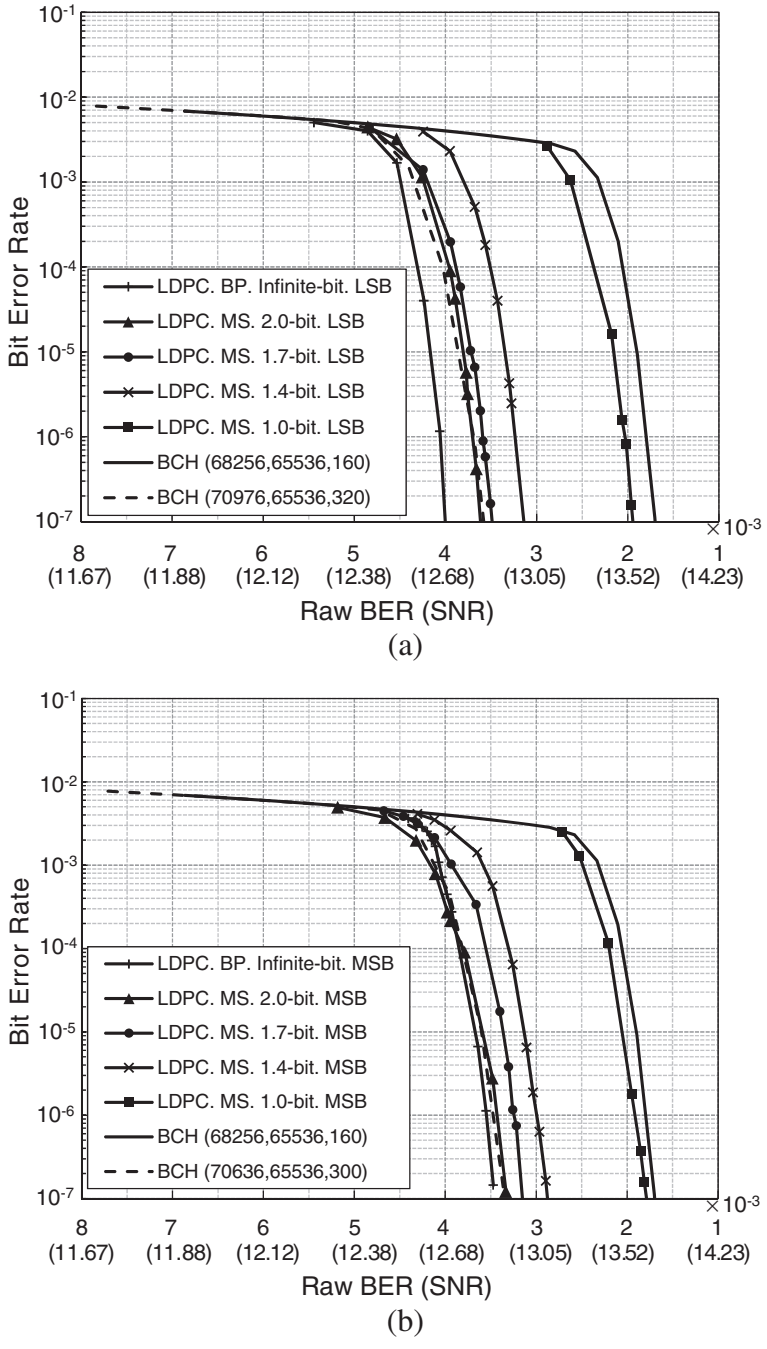

Figure 5 Error-performance of a (68254, 65536) EG-LDPC code and two BCH codes for (a) LSB and (b) MSB pages.

1-bit (hard-decision) data. In order to make the error performance of the BCH code close to that of LDPC code with 2-bit MS decoding, the error-correcting capability t is increased from $t = 160$ to $t = 320$, which corresponds to the code rate of 0.92 and requires more hardware resources. The comparison of soft-decision LDPC and hard-decision BCH codes clearly shows the advantage of the soft-decision decoding.

Figure 5b shows the error performances of the LDPC code and two BCH codes for MSB pages. The overall performance of the LDPC code for MSB pages is slightly worse than that for LSB pages. In this case, a BCH code with the error-correcting capability of $t = 300$ is required to achieve the comparable performance of the LDPC code with 2-bit soft-decision MS decoding.

In Figure 5a, we can find that even hard-decision-based decoding works when the RBER is lower than 1.95×10^{-3}. However, when the RBER is between 1.95×10^{-3} and 3.15×10^{-3}, the hard-decision-based decoding does not work and only soft-decision decoding can remove most of the errors. When the RBER is greater than 3.62×10^{-3}, even 2-bit MS decoding cannot correct the data properly. From this observation, we can divide the RBER values into five regions as shown in Table 5. Although a NAND Flash memory system requires error-free decoding with BER less than 10^{-15}, here we set the target BER to 10^{-7} because the simulation of the LDPC code takes much time to observe the minimum requirement. Note again that EG-LDPC codes show very low error-floor performance and have fast convergence speed. Finally, Table 5 summarizes the results for LSB and MSB pages. Here, we can find that the 1.4-bit precision enhances the error correcting performance very much when compared to 1-bit hard-decision decoding. However, further increasing the precision brings diminishing returns. As a result, the Region II is quite wider than Region III or IV.

Hardware performance of (68254, 65536) LDPC decoder

In order to assess the energy consumption of LDPC decoding, we have implemented the (68254, 65536) EG-LDPC decoder employing the normalized *a posteriori* probability (APP)-based algorithm and layered decoding that lead to simplified functional units and halved decoding iterations, respectively. In addition, a conditional variable node update scheme is employed to improve the error performance and reduce circuit switching activities in the node processing units [20]. The decoding throughput is increased by employing 5-stage pipelined 8-way parallel architecture, while the chip size is much reduced by adopting memory optimization techniques [20]. Because the error performance of fixed-point LDPC decoding with the normalized APP-based algorithm is very close to that of the floating-point decoding, the (68254, 65536) LDPC decoder yields almost the same performance as shown in Figure 5 for the NAND Flash memory channel.

Table 5 The operating regions according to memory output precision

	RBER ($\times 10^{-3}$)		Memory output precision
	LSB pages	MSB pages	needed for 10^{-7} BER
Region I (R_1)	$\sim$ 1.95	$\sim$ 1.79	1-, 1.4- 1.7-, and 2-bit
Region II (R_2)	1.95–3.15	1.79–2.90	1.4- 1.7-, and 2-bit
Region III (R_3)	3.15–3.50	2.90–3.15	1.7- and 2-bit
Region IV (R_4)	3.50–3.62	3.15–3.33	2-bit
Region V (R_5)	3.62+	3.33+	–

The LDPC decoder was synthesized, placed, and routed in 0.13-μm CMOS technology using Synopsys tools, then parasitic resistances and capacitances were extracted to estimate the energy consumption accurately. Randomly generated information bits were encoded and Gaussian noise was added to make test vectors. Then, the power consumption, iteration count, and decoding latency were estimated by using Synopsys PrimeTime. From the simulation results, we obtained the average energy consumption as a first-order function of the iteration count. Finally, the energy consumption of the LDPC decoder was computed using the average iteration counts found by simulations for each memory output precision and RBER. In order to consider the implementation with a recent process technology, the decoding energy of the LDPC decoder is scaled down to a 65-nm technology. The core supply voltages of 130 and 65 nm nodes are 1.2 and 1.0 V, respectively. In addition, the maximum clock frequencies are assumed to be the same, 131 MHz, for both processes. Considering the process technologies and the supply voltages, the energy consumption is scaled down by a factor of 2.88 ($= \left[(65/130\,\text{nm}) \times (1.0/1.2\,\text{V})^2\right]^{-1}$) for the 65-nm technology node according to [21].

The energy consumption of the (68254, 65536) LDPC VLSI with the 65-nm technology for hard-decision and soft-decision data is shown in Figure 6, where the input to the LDPC decoder are LLR values. The clock frequency was set to 131 MHz. Here, we set the maximum iteration limit as eight and the number of quantization bits in the LDPC decoder as seven including two bits for the fractional part. Since the implemented LDPC decoder shows very fast convergence speed, the decoding energy consumption decreases rapidly at low RBER (high SNR). For the low RBER region below 10^{-3}, all decoders demand mostly one decoding iterations, thus resulting in the minimum energy consumption of 0.7 nJ/byte. For the region exceeding the RBER of 10^{-3}, decoding with multi-precision data consumes less energy than that with the hard-decision data because of the decreased number of iterations. In addition, in the region below the RBER of

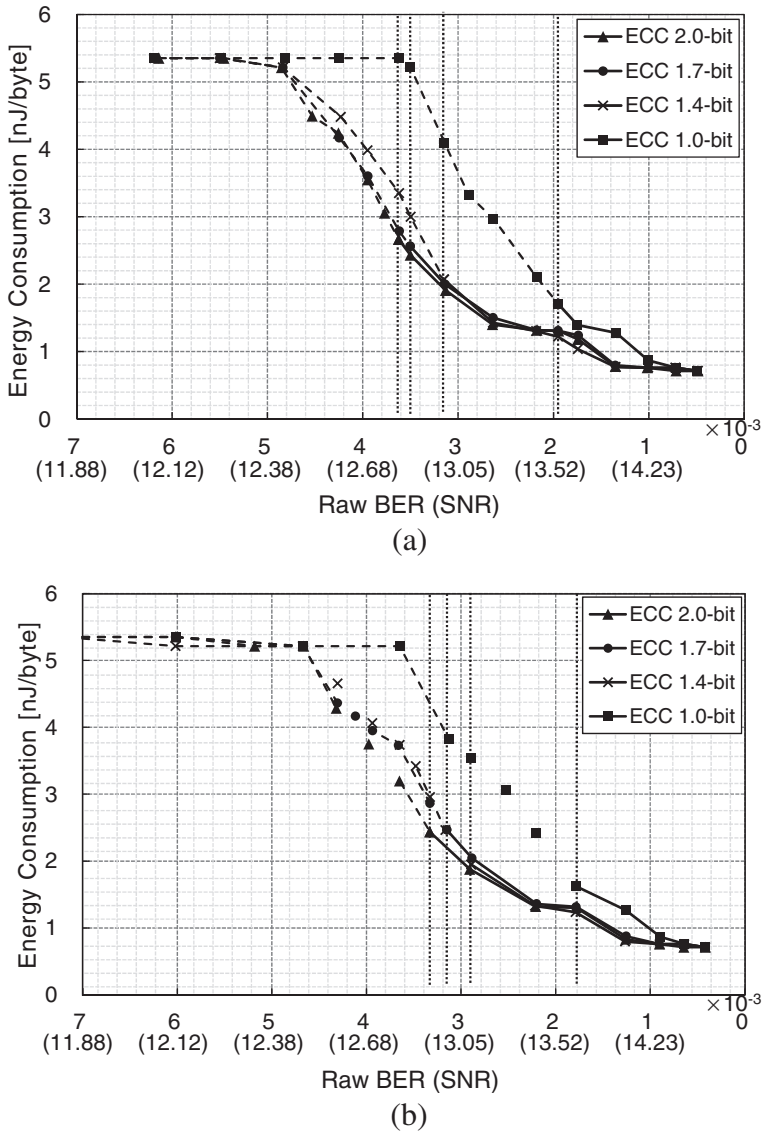

Figure 6 The energy consumption of the (68254, 65536) LDPC decoder (65-nm VLSI) over NAND Flash memory channel for (a) LSB and (b) MSB pages.

3×10^{-3}, all soft-decision decoding shows similar energy consumption.

At the high RBER region where only 2-bit soft-decision decoding is allowed to use, we can find that the average energy consumption of the LDPC decoder is 1.6 to 8.4 times higher than that of a read operation in MLC NAND Flash memory. However, in the low RBER (high SNR) region, in which all kinds of precision can be used, the LDPC decoder consumes only 0.5 to 2.3 times of the energy needed for the read operation in MLC NAND Flash memory. Therefore, we can consider that the total energy consumption is significantly affected by the LPDC decoder in the high RBER region, but is more influenced by the read operation of NAND Flash memory in the low RBER region.

Low-energy error correction scheme for NAND Flash memory

Optimum output precision for low-energy decoding

The total energy consumption of NAND Flash memory access can be obtained by adding that for memory access and that for error correction. We observe that high output precision increases the energy for memory access, while it can reduce the LDPC decoding energy.

Figure 7 shows the total energy consumption of NAND Flash memory with the LDPC decoder for LSB and MSB

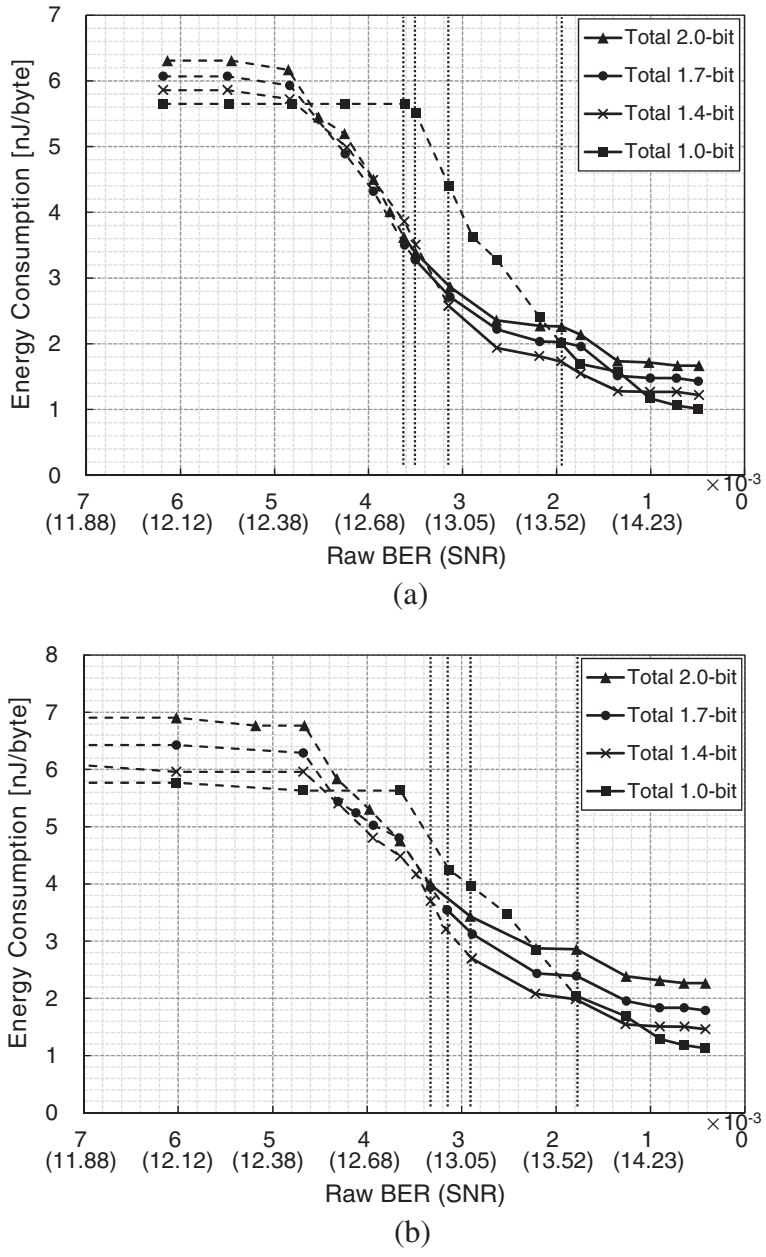

Figure 7 The total energy consumption for (a) LSB and (b) MSB pages, where the LDPC decoder was scaled down to 65-nm technology node.

pages, where NAND Flash memory operates at 100 MHz and V_{ccq} of 1.8 V in the synchronous data output mode. The vertical dotted lines divide the operating regions according to Table 5.

In the region I, where all hard- and soft-decision decoding operate, the 1-bit hard-decision decoding shows the smallest energy consumption when the RBER is very low, while the 1.4-bit soft-decision decoding consumes less energy than the hard-decision decoding as RBER increases. In the region II, the 1.4-bit memory output precision results in the lowest energy consumption, while in the region III, the 1.7-bit precision leads to the lowest consumption. Finally, in the region IV, there is no other choice except the 2-bit soft-decision decoding.

In summary, for each operating region, decoding with the smallest output bits allowed consumes the least energy among possible decoding schemes, especially for decoding MSB pages. Although the 2-bit soft-decision decoding shows the best error correcting performance over all RBER regions, it consumes up to two times more energy than the hard-decision decoding at the low RBER (high SNR) region because of the additional

memory sensing operations. Therefore, depending on the channel condition, appropriate memory output precision should be chosen to minimize the total energy consumption.

We also studied the trend of total energy consumption when considering both program-and-erase (PE) cycling and data retention. The NAND Flash memory channel estimation proposed in [22] was used to decide the SRVs and the smallest output precision was chosen among the possible decoding schemes. Figure 8 shows the total energy consumption for MSB pages. The number of PE cycles and retention time vary from 1 to 5K times and from 1 to 9K hours, respectively. The coupling coefficients of the x and xy directions are set to 0.1034 and 0.006721, respectively, in order to consider 20-nm Flash memory technology [23,24]. We can find that the total energy consumption is very strongly affected by the PE cycling. When the number of PE cycles is less than or equal to 1K, the total energy consumption shows the least amount, which is around 1 nJ/byte regardless of the retention time. However, the total energy consumption also increases with the retention time when the number of PE cycles is larger than 1K.

Iteration count-based precision selection

The presented experimental results show that optimum precision selection is very important for low-energy soft-decision-based decoding of NAND Flash memory. One straightforward idea is to conduct *failure-based* precision selection. In this method, the precision is increased when the decoding is failed. For example, the decoding begins with 1-bit (hard-decision) precision, and if it fails, the decoding is retried with an improved precision. Although this method is very simple and there is no need of storing the precision information, this can consume a large amount of energy when the decoding fails because LDPC decoders iterate many cycles. Of course, the failure-based scheme also incurs additional time-delay for retrying the decoding with an updated precision.

Another approach is to estimate the signal quality of NAND Flash memory periodically with channel estimation algorithms [22]. By sensing the signal with multiple threshold voltages, we can estimate the mean and the variance of each symbol. This method, however, demands extra time and energy for signal quality estimation. Considering that the signal quality deteriorates when the number of PE cycles and the retention time increase, the overhead of periodic estimation can be quite high, especially for a large capacity solid-state drives.

We propose a precision selection method that utilizes the iteration count of the LDPC decoder. In this explanation, we use the precision of 1.0, 1.4, and 2.0 bits because the optimum operating range of the 1.7-bit precision is quite narrow. As shown in Figure 9, when the RBER is very low, such as less than 1.0×10^{-3}, the average iteration count is around one even with 1-bit precision decoding. Thus, employing the 1-bit precision is the best for low energy decoding in this region. However, as the RBER

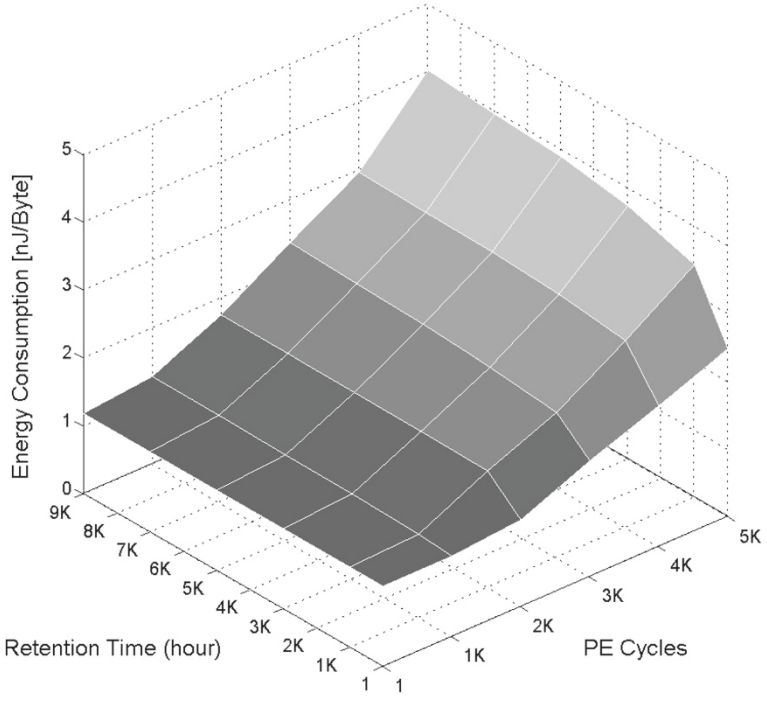

Figure 8 The total energy consumption for MSB pages with the number of PE cycles and retention time.

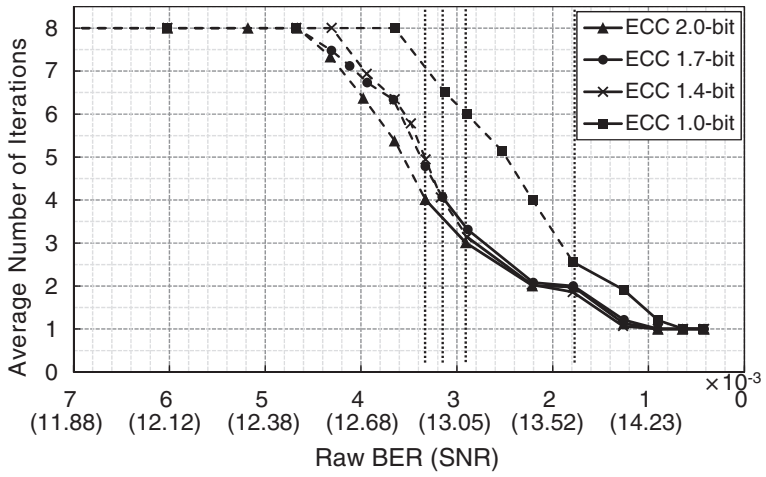

Figure 9 Average number of decoding iterations of the (68254, 65536) LDPC decoder for MSB pages.

grows and when it becomes approximately between 1.0×10^{-3} and 1.79×10^{-3}, the decoding with 1-bit precision for Flash memory output demands an increased number of iterations. Thus, we need to increase the precision to 1.4-bit for low energy when the iteration count with 1-bit precision is repeatedly two or greater. Of course, the opposite path is also needed. If the iteration count is repeatedly only one with 1.4-bit precision, then it is needed to lower the precision into 1-bit. A similar scenario happens when the RBER is close to 3.0×10^{-3}. At this region, the decoding with 1.4-bit demands the iteration count of three or more. This means that it is the time to increase the iteration count to 2-bit. Of course, when the iteration count with 2-bit decoding is repeatedly equal to or less than two, we need to decrease the precision to 1.4-bit. Since we increase the precision before the decoding failure, we can avoid the energy loss and delay.

The iteration count-based precision selection can also be applied to adapt the reference voltages. When the bit error pattern shows an asymmetric result, which means that the number of errors from 1 to 0 is significantly higher or lower than that from 0 to 1, we need to adjust the sensing reference voltages and the direction is easily determined by the error statistics. The channel estimation is performed only when the iteration count with 2-bit precision is repeatedly four or greater.

Concluding remarks

We studied the optimum output precision of NAND Flash memory for low-energy soft-decision-based error correction. The energy consumed at NAND Flash memory as well as the LDPC decoder is considered. This study shows that the optimum precision of Flash memory data for soft-decision LDPC decoding depends on the signal quality, which implies that knowing the SNR of NAND Flash memory is quite important for low-energy error correction. When the SNR is relatively high, the conventional 1-bit (hard-decision) decoding leads to the lowest energy consumption because of minimum sensing and output energy consumed at NAND Flash memory; however, as the SNR decreases the optimum number of bits for low energy needs to be increased. We find that the precision of 1.4-bit for each output, which represents providing an erasure region at each signal boundary, leads to minimum energy decoding at a broad range of signal quality. We also propose an adaptive, feedback-based, precision selection scheme that needs virtually no overhead.

Competing interests
The authors declare that they have no competing interests.

Acknowledgements
This work was supported in part by the Brain Korea 21 Project and the National Research Foundation of Korea(NRF) grants funded by the Korea government(MEST) (No. 2011-0027502 and 2012R1A2A2A06047297).

Received: 17 April 2012 Accepted: 17 August 2012 Published: 5 September 2012

References

1. W Liu, J Rho, W Sung, in *Proceedings of the IEEE Workshop Signal Processing Systems (SiPS'2006)*. Low-power high-throughput BCH error correction VLSI design for multi-level cell NAND Flash memories, (Alberta, Canada, 2–4 October 2006). pp. 303–308
2. R Micheloni, R Ravasio, A Marelli, E Alice, V Altieri, A Bovino, L Crippa, ED Martino, L D'Onofrio, A Gambardella, E Grillea, G Guerra, D Kim, C Missiroli, I Motta, A Prisco, G Ragone, M Romano, M Sangalli, P Sauro, M Scotti, S Won, in *IEEE ISSCC Digest of Technical Papers*. A 4-Gb 2b/cell NAND flash memory with embedded 5b BCH ECC for 36 MB/s system read throughput (San Francisco, CA, 6–9 February 2006). pp. 497–506
3. B Chen, X Zhang, Z Wang, in *Proceedings of the IEEE Workshop Signal Processing Systems (SiPS'2008)*. Error correction for multi-level NAND flash memory using Reed-Solomon codes, (Hainan Island, China, 8–10 October 2008). pp. 94–99
4. RG Gallager, Low density parity check codes, IRE Trans Inf. Theory. **IT-8**(1), 21–28 (1962)

5. DJC MacKay, Good error-correcting codes based on very sparse matrices, IEEE Trans. Inf. Theory. **45**(2), 399–432 (1999)
6. Digital Video Broadcasting (DVB): Second Generation System for Broadcasting, Interactive Services, News Gathering and Other Broadband Satellite Applications. *ETSI Standard, ETS 302 307* (March 2005)
7. IEEE Standard for Information Technology-Telecommunications and Information Exchange between Systems-Local and Metropolitan Area Networks-Specific Requirements Part 3, Carrier Sense Multiple Access with Collision Detection (CSMA/CD) Access Method and Physical Layer Specifications. *IEEE Standard IEEE 802.3an* (June 2006)
8. IEEE Standard for Local and Metropolitan Area Networks Part 16: Air Interface for Fixed and Mobile Broadband Wireless Access Systems. *IEEE Standard IEEE 802.16e* (February 2006)
9. *64Gb, 128Gb, 256Gb, 512Gb asynchronous/synchronous NAND features*. (Micron Technology Inc., 2009). http://www.micron.com/products/nand-flash/mlc-nand. Accessed 5 December 2012
10. G Dong, S Li, T Zhang, Using data postcompensation and predistortion to tolerate cell-to-cell interference in MLC NAND flash memory, IEEE Trans Circuits Syst. I: Regular Papers. **57**(10), 2718–2728 (2010)
11. G Dong, Y Pan, N Xie, C Varanasi, T Zhang, Estimating information-theoretical NAND flash memory storage capacity and its implication to memory system design space exploration, IEEE Trans. VLSI Syst. **20**(9), 1705–1714 (2012)
12. C Monzio, M Ghidotti, A Lacaita, A Spinelli, A Visconti, Random telegraph noise effect on the programmed threshold-voltage distribution of flash memories, IEEE Electron. Dev. Lett. **30**(9), 984–986 (2009)
13. BH Suh, YH Um, JK Kim, YJ Choi, YN Koh, SC Lee, SS Kwon, BS Choi, JS Yum, JH Choi, JR Kim, HK Lim, A 3.3V 32Mb NAND flash memory with incremental step pulse programming scheme, IEEE J. Solid State Circuits. **30**(11), 1149–1156 (1995)
14. JD Lee, SH Hur, JD Choi, Effects of floating-gate interference on NAND flash memory cell operation, IEEE Electron. Dev. Lett. **23**(5), 264–266 (2002)
15. G Dong, N Xie, T Zhang, On the use of soft-decision error-correction codes in NAND flash memory, IEEE Trans. Circuit Syst. I: Regular Papers. **58**(2), 429–439 (2011)
16. J Kim, D Lee, W Sung, in *Proceedings of IEEE International Conference on Communications (ICC'2012), Workshop on Emerging Data Storage Technologies*. Performance of rate 0.96 (68254, 65536) EG-LDPC code for NAND Flash memory error correction, (Ottawa, Canada, 10–15 June 2012)
17. Y Kou, S Lin, MPC Fossorier, Low-density parity-check codes based on finite geometries: a rediscovery and new results, IEEE Trans. Inf. Theory. **47**(7), 2711–2736 (2001)
18. Q Huang, Q Diao, S Lin, K Abdel-Ghaffar, Cyclic and quasi-cyclic LDPC codes on constrained parity-check matrices and their trapping sets, IEEE Trans. Inf. Theory. **58**(5), 2648–2671 (2012)
19. J Chen, A Dholakia, E Eleftheriou, MPC Fossorier, XY Hu, Reduced-complexity decoding of LDPC codes, IEEE Trans. Commun. **53**(8), 1288–1299 (2005)
20. J Kim, W Sung, A rate-0.96 LDPC decoding VLSI for soft-decision error correction of NAND flash memory, IEEE Trans. VLSI Syst (2012). (submitted, in review)
21. S Borkar, Design challenges of technology scaling, IEEE Micro. **19**(4), 23–29 (1999)
22. D Lee, W Sung, Estimation of NAND flash memory threshold voltage distribution for optimum soft-decision error correction, IEEE Trans. Signal Process (2012). (accepted for publication)
23. K Prall, in *Proceedings of the 22nd IEEE Non-Volatile Semiconductor Memory Workshop(SiPS'2007)*. Scaling non-volatile memory below 30nm, (Monterey, CA, 26–30 August 2007), pp. 5–10
24. P Poliakov, P Blomme, MM Corbalan, JV Houdt, W Dehaene, Cross-cell interference variability aware model of fully planar NAND Flash memory including line edge roughness, Microelectron Reliab. **51**(5), 919–924 (2011)

doi:10.1186/1687-6180-2012-195
Cite this article as: Kim and Sung: **Low-energy error correction of NAND Flash memory through soft-decision decoding.** *EURASIP Journal on Advances in Signal Processing* 2012 **2012**:195.