Channel Models for Multi-Level Cell Flash Memories Based on Empirical Error Analysis

Veeresh Taranalli, *Student Member, IEEE,* Hironori Uchikawa, *Member, IEEE,* Paul H. Siegel, *Fellow, IEEE*

***Abstract*—We propose binary discrete parametric channel models for multi-level cell (MLC) flash memories that provide accurate ECC performance estimation by modeling the empirically observed error characteristics under program/erase (P/E) cycling stress. Through a detailed empirical error characterization of 1X-nm and 2Y-nm MLC flash memory chips from two different vendors, we observe and characterize the *overdispersion* phenomenon in the number of bit errors per ECC frame. A well studied channel model such as the binary asymmetric channel (BAC) model is unable to provide accurate ECC performance estimation. Hence we propose a channel model based on the beta-binomial probability distribution (2-BBM channel model) which is a good fit for the overdispersed empirical error characteristics and show through statistical tests and simulation results for BCH, LDPC and polar codes, that the 2-BBM channel model provides accurate ECC performance estimation in MLC flash memories.**

***Index Terms*—Flash memory, multi-level cell, channel model, error correcting codes, P/E cycling.**

I. Introduction

CHANNEL modeling for NAND flash memories is a developing research area with applications to better signal processing and coding techniques. A channel model for a flash memory can be viewed as a simplified representation of the underlying physical mechanisms which induce errors in stored data. For NAND flash memories, the major error mechanisms are program disturb and cell wear that occur during program/erase cycling, charge loss that occurs during data retention and inter-cell interference (ICI) [1]–[3]. The main applications of a flash memory channel model are improved design, decoding and performance evaluation of error-correcting codes (ECCs) and error-mitigating codes. Other applications include information theoretic studies that provide an analysis of the capacity of flash memories [4], as well as insights for the development of new coding techniques. In this paper, we focus on the development of parametric channel models for multi-level cell (MLC) flash memories based on empirical error characterization, that enable accurate ECC frame error rate (FER) performance estimation/prediction.

This work was supported by National Science Foundation (NSF) Grants CCF-1116739, CCF-1405119 and the Center for Memory and Recording Research, UC San Diego. The material in this paper was presented in part at the IEEE International Conference on Communications, London, UK, June 8-12, 2015 and the Annual Non-Volatile Memories Workshop (NVMW), San Diego, USA, March 6-8, 2016.

V. Taranalli, and P. H. Siegel are with the University of California, San Diego, La Jolla, CA 92093-0401, USA (e-mail: vtaranalli, psiegel@ucsd.edu).

H. Uchikawa is with Toshiba Corporation, Japan (e-mail: hironori.uchikawa@toshiba.co.jp).

A. Overview of the Problem

Efficient evaluation of ECC FER performance is important for storage system design and optimization. One approach to ECC FER performance estimation is to experimentally collect error data for use in Monte-Carlo simulations of the ECC decoder, but this can be impractical because of the large amount of error data required when estimating low frame error rates. Another approach is to analytically predict the performance of a code based upon a measured average raw bit error rate. While this is feasible for algebraic codes with bounded distance decoders, it is difficult for low density parity check (LDPC) codes and polar codes that use probabilistic decoders based upon message passing or successive cancellation. Moreover, the implicit assumption of independent, symmetric bit errors may not be justified.

Previously proposed [5], [6] parametric channel models for MLC flash memories were obtained by using well known probability distributions to model the empirical cell threshold voltage distributions. In [5], a Gaussian distribution, and in [6], a Normal-Laplace mixture model were shown to be a good fit for the experimentally observed cell threshold voltage distributions in MLC flash memories. Such models can be used to reliably predict/estimate the experimentally observed raw bit error rate (RBER) of the flash memory. However in this paper, we show through empirical error characterization that the RBER is not necessarily a good indicator of the ECC FER performance and this is due to the overdispersion phenomenon in the number of bit errors per frame in MLC flash memories. Overdispersion refers to the greater variability in empirical data compared to a statistical model for e.g., the binomial distribution typically used to model count data. Therefore, a memoryless channel model such as the binary asymmetric channel (BAC) model provides an optimistic estimate of the ECC FER performance when compared to the actual ECC FER performance estimate obtained from empirical data.

B. Summary of Contributions

We present a detailed empirical characterization of errors in MLC flash memories at the bit, cell and page granularity levels for 1X-nm and 2Y-nm feature size MLC flash memory chips from two different vendors referred to as vendor-A and vendor-B respectively. We study the asymmetry of bit errors in the lower and upper pages of MLC flash memories with a focus on the *number of bit errors per frame* parameter. We observe that the empirical probability distributions of the number of bit errors per frame parameter are *overdispersed*

when compared to a binomial distribution typically used to model count data.

Based on the empirical error analysis, we study the per-page binary asymmetric channel (BAC) model referred to as the 2-BAC model for MLC flash memories. Using statistical analysis, we show that the 2-BAC model does not provide a good fit for the empirical error data and hence is inadequate for accurate ECC frame error rate (FER) performance estimation. Therefore, we propose a channel model based on the beta-binomial probability distribution referred to as the 2-Beta-Binomial (2-BBM) channel model. We show that it is a good fit for the observed overdispersed empirical error data and performs well for ECC FER performance estimation. We also propose normal and Poisson approximation based channel models for MLC flash memories.

Through quantitative evaluation of the proposed channel models using the statistical Kolmogorov-Smirnov (K-S) Two Sample goodness of fit test and using Monte-Carlo simulation results of FER performance for BCH, LDPC and polar codes, we show that the 2-Beta-Binomial channel model is an accurate channel model to represent the overdispersed nature of bit errors in MLC flash memories.

C. Organization of the Paper

The rest of the paper is organized as follows. Section II presents a brief introduction to flash memories with a focus on the structure of MLC flash memories. In Section III we describe the P/E cycling experiment procedure. Section IV provides a detailed empirical characterization of errors in MLC flash memories, the results of which are utilized for design and evaluation of the proposed channel models. Section V describes the proposed channel models for MLC flash memories and provides statistical analysis results. In Section VI, quantitative results for statistical goodness of fit tests and BCH, LDPC and polar code FER performance are presented to evaluate the proposed channel models. Section VII provides the concluding remarks.

II. Flash Memory Structure

The fundamental data storing unit in NAND flash memories is a floating-gate transistor commonly referred to as a cell. A cell can be programmed to hold different levels of charge and these charge levels represent the data bits stored in a cell. The most commonly used cells in today's flash memories are capable of holding 2, 4 and 8 distinct charge levels (1, 2, 3 bits/cell respectively) and are referred to as single-level cell (SLC), multi-level cell (MLC) and three-level cell (TLC) respectively. These flash memory cells are organized into a rectangular array interconnected through horizontal wordlines (WL) and vertical bitlines (BL) to form a flash memory "block" [1]. A collection of such blocks makes up the flash memory chip. A schematic of the block structure of MLC flash memories is shown in Fig. 1.

The two bits belonging to a MLC flash memory cell are separately mapped to logical units of programming, called pages. A page is also the smallest unit for program and read operations whereas a block is the smallest unit for the erase

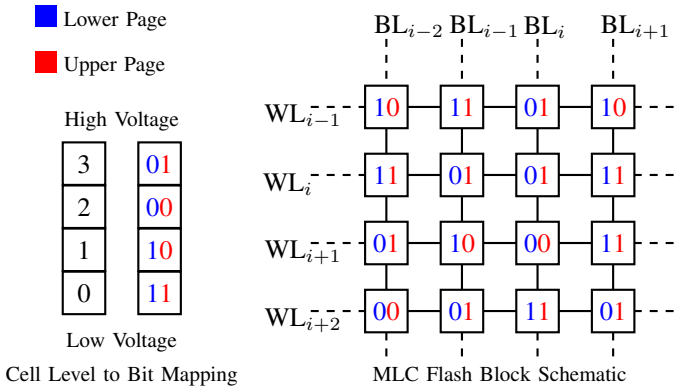

Fig. 1. Cell level to bit mapping and block schematic in MLC flash memories. In the block schematic, the rectangles depict the MLC flash memory cells connected to horizontal wordlines (WL) and vertical bitlines (BL).

operation. The most significant bit (MSB) is mapped to the lower page while the least significant bit (LSB) is mapped to the upper page. The lower page bit of a cell always precedes the corresponding upper page bit in the programming order. We represent the four charge levels in MLC flash memory as 0, 1, 2, 3 in the increasing order of charge levels respectively. The corresponding 2-bit patterns written to the lower (MSB) and upper (LSB) pages are '11', '10', '00' and '01' respectively as shown in Fig. 1.

III. Experiment Procedure

To characterize and quantify the number and types of errors observed, we perform program/erase (P/E) cycling of the MLC flash memory chip under test which consists of repeated application of the following steps:

1) Erase MLC flash memory blocks under test.
2) Program MLC flash memory pages (of blocks under test) with pseudo-random (PR) data generated using a Mersenne-Twister pseudo-random number generator. The pseudo-random number generator is initialized with a randomly generated seed for every page in every P/E cycle.
3) Starting with the first cycle, perform a read operation on the MLC flash memory block(s) at intervals of every 100^{th} cycle. Record bit errors and their locations in the block.

We arbitrarily choose 4 contiguous blocks in an MLC flash memory chip for our experiments. The MLC flash memory blocks are P/E cycled up to 10,000 P/E cycles and the experiments are performed at room temperature in a continuous manner with no extra wait time between the erase/program/read operations.

IV. Characterization of Errors in MLC Flash Memories

The first step in the error characterization of a flash memory chip is to study its raw bit error rate (BER) performance when all the pages in all the blocks under test are programmed with pseudo-random data. This closely resembles the most common use in practice, where random data are stored and retrieved.

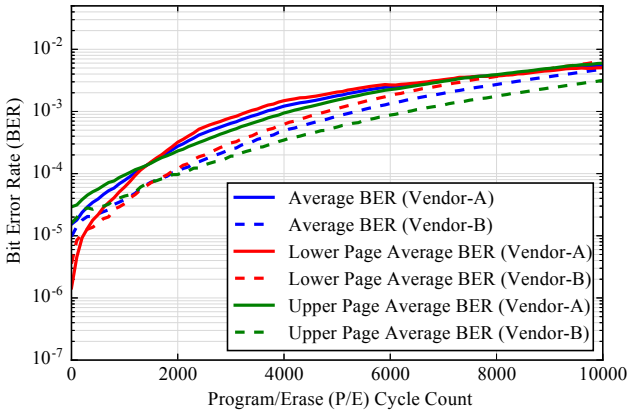

Fig. 2. Measured average raw bit error rates over 4 blocks of vendor-A and vendor-B chips.

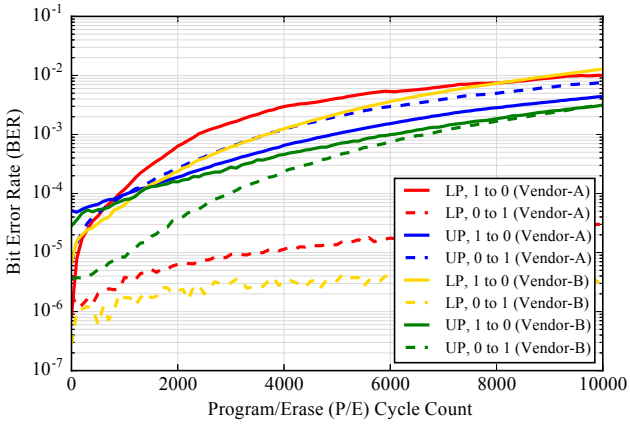

Fig. 3. Average raw bit error rates corresponding to specific bit errors in the lower pages (LP) and upper pages (UP) over 4 blocks of vendor-A and vendor-B chips.

Fig. 2 shows the average raw BER across the P/E cycles when all pages in each block are programmed for both the vendor-A and vendor-B flash memory chips. The raw BER is averaged over 4 blocks tested. Fig. 2 also shows the average raw BER separately for the lower and upper pages of the MLC flash memory. Although the lower page is expected to have a smaller BER compared to the upper page [7], we observe that this is only the case up to a certain number of P/E cycles in the beginning and as the P/E cycle count increases, the lower page begins to show a larger number of errors than the upper page. This observation is consistent across both the vendor-A and vendor-B flash memory chips. Using empirical data from 20 blocks of the same flash memory chip, we have also observed consistent measured average raw BER estimates across all the P/E cycles.

We also record the specific cell (symbol) errors corresponding to all the bit errors observed. Table I shows the frequencies of all possible cell errors as a percentage of the total number of cell errors observed across all the blocks in all the P/E cycles. The corresponding average cell error probabilities across all P/E cycles are $\sim 4.16 \times 10^{-3}$ and $\sim 2.71 \times 10^{-3}$ for vendor-A and vendor-B chips respectively. We observe that the level 1

TABLE I

FREQUENCY OF CELL (SYMBOL) ERRORS MEASURED AS A PERCENTAGE OF TOTAL NUMBER OF CELL ERRORS OBSERVED ACROSS ALL P/E CYCLES WHEN ALL 4 BLOCKS ARE PROGRAMMED WITH PSEUDO-RANDOM DATA.

Vendor-A				
Write Cell	**Read Cell Values**			
Values	**11**	**10**	**00**	**01**
11	0.00	17.25	0.08	2.57
10	0.19	0.00	48.19	0.74
00	0.00	0.14	0.00	30.61
01	0.00	0.03	0.20	0.00

Vendor-B				
Write Cell	**Read Cell Values**			
Values	**11**	**10**	**00**	**01**
11	0.00	18.39	0.03	4.01
10	0.07	0.00	62.22	1.84
00	0.00	0.06	0.00	13.39
01	0.00	0.00	0.00	0.00

to 2 cell error "10 (1) $\rightarrow$ 00 (2)" is the most dominant for both vendor-A and vendor-B chips. This observation explains why the lower page average raw BER is worse than the upper page average raw BER as shown earlier in Fig. 2. We also note that the three adjacent level cell errors "10 (1) $\rightarrow$ 00 (2)", "11 (0) $\rightarrow$ 10 (1)" and "00 (2) $\rightarrow$ 01 (3)" are the most frequent and together make up about 96% and 94% of all the cell errors observed for the vendor-A and vendor-B chips respectively. Such knowledge about dominant cell errors can be very useful in utilizing ECC redundancy more effectively. This was demonstrated in [8], where the authors designed two BCH codes with different error correction capabilities for the lower and upper pages of an MLC flash memory and proposed a stagewise combined decoding algorithm for both pages. Their scheme gave better results than using a single BCH code independently for all pages.

A. Asymmetry of Bit Errors in MLC Flash Memories

Fig. 3 shows the asymmetry of bit errors in MLC flash memories. We present the average raw BERs corresponding to the specific types of bit errors i.e., $0 \rightarrow 1$ and $1 \rightarrow 0$ bit errors, in the lower and upper pages of both vendor-A and vendor-B MLC flash memory chips. While there is a high degree of asymmetry in the lower page bit errors throughout the P/E cycle range, the degree of asymmetry in the upper page bit errors is much lower. This agrees well with the observations in Table I, where the dominant cell errors imply a large proportion of $1 \rightarrow 0$ bit errors in the lower page and comparable proportions of $0 \rightarrow 1$ and $1 \rightarrow 0$ bit errors in the upper page. This asymmetry in bit errors in both the lower and upper pages also reflects the dominance of data dependent inter-cell interference (ICI) errors i.e., the middle cells in the cell level data patterns 303, 313 and 323 across wordlines are highly susceptible to errors [9].

B. Characterization of Number of Bit Errors per Frame

As we want to develop parametric channel models for MLC flash memories which provide an accurate representation of the empirically observed bit errors and enable accurate ECC FER performance estimation, we study the distribution of the number of bit errors per frame parameter. This is the key factor in determining the FER performance of an ECC with a specified error correction capability of t number of bit errors per frame.

From the error data collected during P/E cycling experiments, we obtain the sample counts of the number of bit errors per frame for $0 \to 1$ and $1 \to 0$ bit errors in both the lower and upper pages by choosing a fixed frame length of $N = 8192$ bits. This choice of the frame length is representative of the large ECC frame lengths used in practice, while still being small enough to ensure sufficient empirical data can be collected easily. Commonly used ECC frame lengths range from 8192 to 32768 bits and multiple ECC frames are written to a single flash memory page in practice. The sample mean and variance statistics of the number of bit errors per frame are computed using the sample counts and are shown in Table II for both vendor-A and vendor-B chips. We also plot two dimensional (2D) maps showing the number of bit errors for every frame in a single block of MLC flash memory at 8,000 P/E cycles in Fig. 4. The 2D maps are obtained by stacking horizontally, the bit error counts in frames belonging to a page, and then stacking vertically all the pages belonging to a single block. From Table II and Fig. 4, we clearly observe that the variance in the number of bit errors per frame is much larger than the mean i.e., the experiment data is overdispersed with respect to a binomial distribution, Binomial(n, p), typically used to model count data whose mean and variance are approximately equal when p is small.

TABLE II
SAMPLE MEAN AND VARIANCE OF THE NUMBER OF BIT ERRORS PER FRAME OBTAINED FROM EMPIRICAL DATA FOR LOWER AND UPPER PAGES ACROSS P/E CYCLES WHEN ALL 4 BLOCKS ARE PROGRAMMED WITH PSEUDO-RANDOM DATA. FRAME LENGTH $N = 8192$.

P/E Cycles	Vendor-A				Vendor-B			
	Lower Page		Upper Page		Lower Page		Upper Page	
	Mean	Variance	Mean	Variance	Mean	Variance	Mean	Variance
2000	2.63	3.08	1.90	2.17	0.98	1.05	0.79	0.86
4000	12.21	18.70	7.76	9.84	5.10	6.97	2.84	3.66
6000	21.90	46.71	18.43	30.06	14.85	29.64	7.18	10.23
8000	30.55	75.89	32.01	66.43	30.03	84.81	14.46	24.37
10000	41.37	111.35	48.88	125.99	52.61	216.95	26.06	51.30

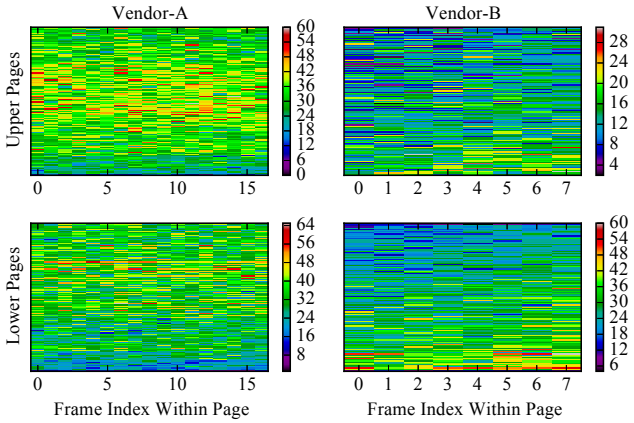

Fig. 4. Two dimensional maps of bit error counts in frames of lower and upper pages in a single block of MLC flash memory chips from vendor-A and vendor-B at 8,000 P/E cycles.

V. CHANNEL MODELS FOR MLC FLASH MEMORIES

In this section, first we study the suitability of well known discrete memoryless channel (DMC) models such as the 4-ary DMC, the BSC and the BAC, to represent the bit errors observed in the MLC flash memory channel. Among the DMC models, a per page BAC (2-BAC) model appears to align well with our empirical error characterization results. However we show through analysis as well as empirical results that the per page BAC model is unable to fit the empirical distribution of the number of bit errors per frame and is not a good model for ECC FER performance estimation. This is due to the interdependence of mean and variance statistics of the number of bit errors per frame for a BAC where the number of $0 \to 1$ and $1 \to 0$ errors are modeled as binomial distributions. The binomial distribution is a single parameter (degree of freedom) distribution, hence its mean and variance cannot be chosen independently. Thus the binomial distribution is unable to accurately model the overdispersed empirical error data as described in the previous section. A natural next choice is to consider the normal approximation to the binomial distribution which provides two parameters (degrees of freedom) for modeling the observed mean and variance statistics independently. However we observe that the normal approximation based channel model does not accurately fit the shape of the empirical data distribution. Another commonly used probability distribution to model overdispersed data with respect to a binomial distribution is the beta-binomial distribution [10], [11]. Hence we propose a discrete channel model based on the beta-binomial distribution for the lower and upper pages referred to as the 2-BBM channel model. We show that this model fits the empirical distribution of the number of bit errors per frame and provides accurate ECC FER performance estimation. We also present simple approximations of the 2-BAC model based on the normal and Poisson probability distributions. Although these approximations are able to fit the empirical distribution of the number of bit errors per frame better than the 2-BAC model, they are not as good a fit as the proposed 2-BBM channel model.

A. Definitions and Notation

Let K represent the total number of bit errors in a frame of length N bits. Let K_m be the total number of bit errors in a frame of N bits which consists of m zeros and $N - m$ ones. The relationship between probability distributions of K and K_m is given by

$$\Pr(K = k) = \sum_{m=0}^{N} \frac{\binom{N}{m}}{2^N} \Pr(K_m = k) \tag{1}$$

where $\frac{\binom{N}{m}}{2^N}$ represents the probability of observing exactly m zeros in a frame of N bits. K_m can be represented as the sum of the number of $0 \to 1$ and $1 \to 0$ bit errors as

$$K_m = K_m^{(0)} + K_{N-m}^{(1)} \tag{2}$$

where $K_m^{(0)}$ and $K_{N-m}^{(1)}$ denote the number of $0 \to 1$ and $1 \to 0$ bit errors respectively. K can also be represented as the sum of the total number of $0 \to 1$ and $1 \to 0$ bit errors as

$$K = K^{(0)} + K^{(1)} \quad \text{where,} \tag{3}$$

$$\Pr(K^{(u)} = k) = \sum_{m=k}^{N} \frac{\binom{N}{l}}{2^N} \Pr(K_l^{(u)} = k) \tag{4}$$

Note that $u \in \{0, 1\}$ where $l = m + (N - 2m)u$. We use $\mathrm{E}[X]$ and $\mathrm{Var}[X]$ to denote the expected value (mean) and the variance of a random variable X respectively. We use $X \mid Y$ to denote "X given Y".

B. Candidate Discrete Memoryless Channel (DMC) Models

The primary error mechanism in MLC flash memories is at the cell level and hence the 4-ary DMC model with 4 inputs and 4 outputs can naturally account for all the cell level errors. This 4-ary DMC model requires 16 parameters (only 12 independent parameters) which are the cell level transition probabilities and these parameters can be easily estimated from experiment data such as that shown in Table I. However the 4-ary DMC model is not useful in practice as the logical unit of progam/read operations in current MLC flash memory applications is a binary page. Hence any practically applicable channel model would have to treat the errors in the lower and upper pages of the MLC flash memory independently, even though it is clear that the errors occur at the cell level and hence the lower and upper page bit errors are not independent. A simpler more commonly used DMC model is the 2-BSC model where two independent BSCs are used to represent the bit errors occuring in the lower and upper pages. The advantage of using the BSC model for each page independently is that it is simple and well studied, with a variety of error correction coding (ECC) techniques available for transmission over the BSC. However, based on our error characterization results in Section IV, the bit errors in MLC flash memories during P/E cycling are mostly asymmetric in nature. Therefore, the BSC is clearly not an accurate model to represent the bit errors in MLC flash memories. A numerical comparison of estimated capacities of the 4-ary DMC model and the 2-BSC model was presented in [9], where it was observed that the 4-ary DMC model provides a significant capacity gain compared to the 2-BSC model for MLC flash memories.

C. The 2-Binary Asymmetric Channel (2-BAC) Model

Based on the asymmetry of bit errors observed in MLC flash memories (Section IV), we propose a per page BAC model called the 2-BAC model where two independent BAC models are used to represent the bit errors occuring in the lower and upper pages. The 2-BAC model is a parametric model with 4 parameters which are the probabilities of $0 \to 1$ and $1 \to 0$ errors in lower and upper page BACs, $p_0^{(l)}$, $p_1^{(l)}$ and $p_0^{(u)}$, $p_1^{(u)}$. For a theoretical evaluation, we mainly compare the mean and variance statistics of the number of bit errors per frame corresponding to a BAC model with the empirically observed sample mean and variances shown in Table II. We consider a BAC as shown in Fig. 5, where p is the probability of $0 \to 1$ error and q is the probability of $1 \to 0$ error. Next, we derive closed form expressions for the mean, $\mathrm{E}[K]$, and the variance, $\mathrm{Var}[K]$, of the number of bit errors per frame corresponding to a BAC model. For the BAC model, $K_m^{(0)}$ and $K_{N-m}^{(1)}$ are distributed according to the binomial probability distribution and are independent i.e.,

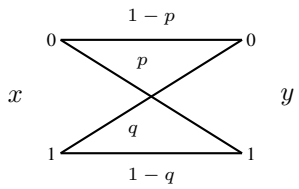

Fig. 5. Binary asymmetric channel

$$K_m^{(0)} \sim \text{Binomial}(m, p) \tag{5}$$

$$K_{N-m}^{(1)} \sim \text{Binomial}(N - m, q) \tag{6}$$

$$K_m^{(0)} \perp\!\!\!\perp K_{N-m}^{(1)} \tag{7}$$

The mean and the variance of $K_m^{(0)}$ are given by

$$\mathrm{E}[K_m^{(0)}] = mp \tag{8}$$

$$\mathrm{Var}[K_m^{(0)}] = mp(1 - p) \tag{9}$$

and those of $K_{N-m}^{(1)}$ are given by

$$\mathrm{E}[K_{N-m}^{(1)}] = (N - m)q \tag{10}$$

$$\mathrm{Var}[K_{N-m}^{(1)}] = (N - m)q(1 - q). \tag{11}$$

Proposition 1: The mean and the variance of K for a BAC model are given by

$$\mathrm{E}[K] = \frac{N}{2}(p + q) \tag{12}$$

$$\mathrm{Var}[K] = \frac{N}{2}\Big((p + q) - pq - \frac{1}{2}(p^2 + q^2)\Big). \tag{13}$$

Proof: See Appendix A. ∎

The parameters of the BAC model p and q are estimated as the average $0 \to 1$ and $1 \to 0$ bit error rates obtained from experimental data corresponding to a particular P/E cycle point in the flash memory lifetime. An algorithmic description of the BAC model is presented in Algorithm 1.

Algorithm 1 BAC Model Implementation

Input: Input frame $\mathbf{x}$ of length N, BAC model parameters (p, q).

Output: Data frame with errors $\mathbf{y}$.

1: **for** $x_i \in \mathbf{x}$ **do**
2: Generate random sample $u \sim$ Uniform$[0, 1]$.
3: **if** $x_i = 0$ **then** $t = p$ **else** $t = q$.
4: **if** $u \leq t$ **then** $e_i = 1$ **else** $e_i = 0$.
5: $y_i = x_i \oplus e_i$.

Using the results of Proposition 1, we compute $\mathrm{E}[K]$ and $\mathrm{Var}[K]$ for a BAC model as follows. For example, at 8,000 P/E cycles for the upper page BAC model for vendor-A chip,

we have $p = 4.97 \times 10^{-3}$ and $q = 2.84 \times 10^{-3}$ and assuming $N = 8192$, we get $\mathrm{E}[K] = 32.01$ and $\mathrm{Var}[K] = 32.02$. Comparing $\mathrm{E}[K]$ and $\mathrm{Var}[K]$ to the sample mean and variance of K recorded using experimental data as shown in Table II, we observe that the BAC model is unable to account for the large observed sample variance. For small values of p and q, from Proposition 1, we have $\mathrm{Var}[K] \approx \mathrm{E}[K]$. Therefore, the BAC model is not a good fit for the observed empirical probability distribution of K as shown in Fig. 8 and Fig. 9 for vendor-A and vendor-B flash memory chips, respectively. As the $\mathrm{Var}[K]$ is much less than the observed sample variance, the 2-BAC model for MLC flash memory is expected to provide a more optimistic estimate of the ECC FER performance when compared to the actual performance. We discuss this in more detail in Section VI. However, note that the 2-BAC model does provide an accurate estimate of the average raw BER which is given by $\frac{\mathrm{E}[K]}{N}$. This shows that the ability to accurately estimate/predict the average raw BER is not the sole criterion for a good MLC flash memory channel model.

D. The 2-Beta-Binomial (2-BBM) Channel Model

As mentioned in Section IV, the empirically observed sample mean and variance estimates show that the number of bit errors per frame data is overdispersed with respect to the binomial distribution. This is the major reason for the poor fit of the 2-BAC model discussed in the previous subsection. To account for the overdispersion, we propose a channel model for MLC flash memories based on the beta-binomial probability distribution called the 2-Beta-Binomial (2-BBM) channel model.

The beta-binomial probability distribution was first proposed in [10] as the probability distribution for counts resulting from a binomial distribution if the probability of success varies according to the beta distribution between sets of trials. Using empirical data, it was also shown in [10] that the beta-binomial probability distribution is a good fit for overdispersed binomial data. Lindsey et al. [11] studied the beta-binomial probability distribution based model in fitting overdispersed human sex ratio in families data and it was found to be a good fit. Stapper et al. [12] developed a yield prediction model for semiconductor memory chips by modeling the overdispersed distribution of number of faults per chip using the gamma-Poisson distribution which is closely related to the beta-binomial distribution.

For the beta-binomial channel model, we model the variables $K_m^{(0)}$ and $K_{N-m}^{(1)}$ as being distributed according to the beta-binomial distribution i.e.,

$$\begin{aligned} p &\sim \text{Beta}(a, b) \\ K_m^{(0)} \mid p &\sim \text{Binomial}(m, p) \\ K_m^{(0)} &\sim \text{Beta-Binomial}(m, a, b) \end{aligned} \tag{14}$$

$$\begin{aligned} q &\sim \text{Beta}(c, d) \\ K_{N-m}^{(1)} \mid q &\sim \text{Binomial}(N-m, q) \\ K_{N-m}^{(1)} &\sim \text{Beta-Binomial}(N-m, c, d) \end{aligned} \tag{15}$$

$$K_m^{(0)} \perp\!\!\!\perp K_{N-m}^{(1)} \tag{16}$$

where (a, b) and (c, d) correspond to the parameters of a beta probability distribution defined as

$$f(\theta; \alpha, \beta) = \frac{\theta^{\alpha-1}(1-\theta)^{\beta-1}}{B(\alpha, \beta)} \quad 0 \le \theta \le 1 \tag{17}$$

$$B(\alpha, \beta) = \int_0^1 \theta^{\alpha-1}(1-\theta)^{\beta-1} \mathrm{d}\theta \tag{18}$$

where $B(\alpha, \beta)$ represents the beta function. Thus the Beta-Binomial (BBM) channel model is derived from a BAC model where the bit error probabilities p and q are random variables which vary from frame to frame and are distributed according to the beta distribution. The BBM channel model is a 4-parameter model (compared to the 2-parameter BAC) and hence the 2-BBM channel model for MLC flash memories will be an 8-parameter model. The beta-binomial probability distributions of $K_m^{(0)}$ and $K_{N-m}^{(1)}$ are given by

$$\Pr(K_m^{(0)} = k) = \binom{m}{k} \frac{B(a+k, b+m-k)}{B(a, b)} \tag{19}$$

$$\Pr(K_{N-m}^{(1)} = k) = \binom{N-m}{k} \frac{B(c+k, d+N-m-k)}{B(c, d)}. \tag{20}$$

The mean and the variance of $K_m^{(0)}$ and $K_{N-m}^{(1)}$ are given by

$$\mathrm{E}[K_m^{(0)}] = \frac{ma}{a+b} \tag{21}$$

$$\mathrm{Var}[K_m^{(0)}] = \frac{mab(a+b+m)}{(a+b)^2(a+b+1)} \tag{22}$$

$$\mathrm{E}[K_{N-m}^{(1)}] = \frac{(N-m)c}{c+d} \tag{23}$$

$$\mathrm{Var}[K_{N-m}^{(1)}] = \frac{(N-m)cd(c+d+N-m)}{(c+d)^2(c+d+1)}. \tag{24}$$

Proposition 2: The mean and the variance of K for a BBM channel model are given by

$$\mathrm{E}[K] = \frac{N}{2}\left(\frac{a}{a+b} + \frac{c}{c+d}\right) \tag{25}$$

$$\begin{aligned} \mathrm{Var}[K] = {} & \frac{N}{4}\left(\frac{a(a+b)(a+2b+1)+Nab}{(a+b)^2(a+b+1)}\right) + \\ & \frac{N}{4}\left(\frac{c(c+d)(c+2d+1)+Ncd}{(c+d)^2(c+d+1)}\right) - \\ & \frac{N}{4}\left(\frac{2ac}{(a+b)(c+d)}\right). \end{aligned} \tag{26}$$

Proof: See Appendix B. ∎

Proposition 3: The mean and the second moment of $K^{(0)}$ and $K^{(1)}$ for a BBM channel model are given by

$$\mathrm{E}[K^{(0)}] = \frac{N}{2}\left(\frac{a}{a+b}\right) \tag{27}$$

$$\mathrm{E}[(K^{(0)})^2] = \frac{N}{4}\left(\frac{a(a+2b+1)+Na(a+1)}{(a+b)(a+b+1)}\right) \tag{28}$$

$$\mathrm{E}[K^{(1)}] = \frac{N}{2}\left(\frac{c}{c+d}\right) \tag{29}$$

$$\mathrm{E}[(K^{(1)})^2] = \frac{N}{4}\left(\frac{c(c+2d+1)+Nc(c+1)}{(c+d)(c+d+1)}\right). \tag{30}$$

Proof: See Appendix C. ∎

The parameters a, b, c, d of the BBM channel model are estimated from the sample moments of $K^{(0)}$ and $K^{(1)}$ using the method of moments [10]. From P/E cycling experiment data, we obtain the sample mean and sample second moment estimates of the random variables $K^{(0)}$ and $K^{(1)}$ which represent the total number of $0 \to 1$ and $1 \to 0$ bit errors per frame. Let μ_1, μ_2 represent the first and second moment estimates of $K^{(0)}$ and μ_3, μ_4 represent the first and second moment estimates of $K^{(1)}$. Solving the equations in Proposition 3 for a, b, c, d, we have the parameter estimates

$$\hat{a} = \frac{\mu_1^2(N+1) - 2\mu_1\mu_2}{N(\mu_2-\mu_1) - \mu_1^2(N-1)} \qquad \hat{b} = \hat{a}\left(\frac{N}{2\mu_1} - 1\right) \tag{31}$$

$$\hat{c} = \frac{\mu_3^2(N+1) - 2\mu_3\mu_4}{N(\mu_4-\mu_3) - \mu_3^2(N-1)} \qquad \hat{d} = \hat{c}\left(\frac{N}{2\mu_3} - 1\right). \tag{32}$$

An algorithmic description of the BBM channel model is presented in Algorithm 2.

Algorithm 2 BBM Channel Model Implementation

Input: Input frame $\mathbf{x}$ of length N, BBM channel model parameters (a, b, c, d).

Output: Data frame with errors $\mathbf{y}$.

1: Generate two independent random samples, $p \sim \text{Beta}(a, b)$ and $q \sim \text{Beta}(c, d)$.
2: $\mathbf{y} = \text{BAC}(\mathbf{x}, p, q)$ [Use Algorithm 1].

TABLE III
UPPER PAGE BBM CHANNEL MODEL PARAMETER ESTIMATES FOR VENDOR-A AND VENDOR-B CHIPS. $N = 8192$.

P/E Cycles	Vendor-A				Vendor-B			
	a	b	c	d	a	b	c	d
2000	12.72	46368.34	8.05	42569.08	10.82	302596.64	6.86	43747.02
4000	25.95	20940.98	15.46	23556.92	11.39	48028.59	6.00	13142.88
6000	22.67	7596.71	18.16	11890.14	15.58	20535.47	7.16	7193.92
8000	20.72	4143.52	22.28	7821.13	15.28	9068.43	7.58	4092.87
10000	21.36	2819.03	26.12	5890.35	13.36	4142.23	9.28	2938.88

For evaluation of the BBM channel model, we compute $\mathrm{E}[K]$ and $\mathrm{Var}[K]$ using Proposition 2. Corresponding to the example used for evaluating the BAC model, the parameter estimates of the upper page BBM channel model for vendor-A

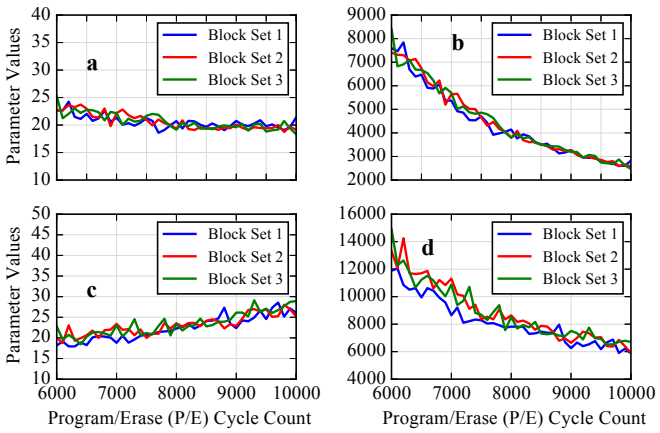

Fig. 6. Variation of parameter estimates for the upper page BBM channel model ((a, b) for $0 \to 1$ error, (c, d) for $1 \to 0$ error) for 3 different 4-block sets for vendor-A chip. $N = 8192$.

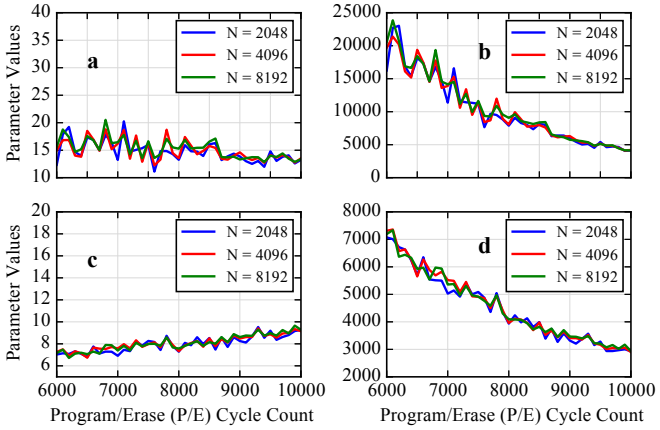

Fig. 7. Variation of parameter estimates for the upper page BBM channel model ((a, b) for $0 \to 1$ error, (c, d) for $1 \to 0$ error) for 3 different frame lengths for vendor-B chip.

are as shown in Table III and using these parameter estimates, we obtain $\mathrm{E}[K] = 32.01$ and $\mathrm{Var}[K] = 57.88$ for $N = 8192$ at 8,000 P/E cycles. Comparing with the results from Table II, we observe that the $\mathrm{Var}[K]$ obtained using the BBM channel model is still lower than the sample variance; however, it is clear that the BBM channel model is vastly better at modeling the overdispersed number of bit errors per frame empirical data than the BAC model. This will be even more evident based on the ECC FER performance estimation results presented in Section VI.

We also observe remarkable consistency in the parameter estimates of the BBM channel model across different blocks of the same MLC flash memory chip. Fig. 6 shows the empirical parameter estimates corresponding to the upper page BBM channel models for vendor-A chip using data collected from 3 different sets of 4 contiguous blocks of the MLC flash memory chip. Fig. 7 shows the empirical parameter estimates corresponding to the upper page BBM channel models for vendor-B chip obtained using different frame sizes. Although not shown (due to lack of space), we also observe similar

consistency in the lower page parameter estimates for both the vendor chips using different sets of blocks on the same chip and different frame sizes. We also note that the estimates for lower page parameters a and b will be noisy because the $0 \to 1$ bit error rate in the lower page is extremely small. This consistency suggests that we may be able to model every flash memory chip with just 8 parameters of the 2-BBM channel model for accurate ECC FER performance estimation.

E. Normal and Poisson Approximation Channel Models

To model the overdispersed number of bit errors per frame empirical data, an alternative approach from a statistical viewpoint is to consider approximations to the binomial probability distribution which retain the general shape of the binomial distribution and whose mean and variance can be controlled independently. We propose two such channel models for MLC flash memories based on the normal and Poisson probability distributions called the 2-Normal Approximation to the BAC (2-NA-BAC) model and the 2-Poisson Approximation to the BAC (2-PA-BAC) model respectively. Similar to the 2-BAC and 2-BBM channel models, the 2-NA-BAC (resp., 2-PA-BAC) model consists of two independent NA-BAC (resp., PA-BAC) models for the lower and upper pages of MLC flash memories. The design goal for the NA-BAC and PA-BAC models is to ensure a match between the mean and variance statistics of the data from the model and the observed sample mean and sample variance. Based on this, we define rules for the normal and Poisson approximation as follows.

Let μ_0 and σ_0^2 denote the sample mean and sample variance of $K^{(0)}$ and μ_1 and σ_1^2 denote the sample mean and sample variance of $K^{(1)}$. Let $\mathcal{N}(\mu, \sigma^2)$ denote a normal distribution with mean μ and variance σ^2 and let $\mathcal{P}(\lambda)$ denote a Poisson distribution with rate parameter λ. Let g_0 and g_1 represent the sampled number of $0 \to 1$ and $1 \to 0$ bit errors per frame.

Definition 1: The normal approximation rules for the NA-BAC model are given by

$$\begin{aligned} g_0 &= [\hat{g}_0] \text{ where } \hat{g}_0 \ \sim \ \mathcal{N}\big(\mu_0, \sigma_0^2\big) \\ g_1 &= [\hat{g}_1] \text{ where } \hat{g}_1 \ \sim \ \mathcal{N}\big(\mu_1, \sigma_1^2\big). \end{aligned} \tag{33}$$

where $[\cdot]$ denotes the round to nearest integer operator.

Definition 2: The Poisson approximation rules for the PA-BAC model are given by

$$\begin{aligned} g_0 &= \hat{g}_0 - (\sigma_0^2 - \mu_0) \text{ where } \hat{g}_0 \ \sim \ \mathcal{P}\big(\sigma_0^2\big) \\ g_1 &= \hat{g}_1 - (\sigma_1^2 - \mu_1) \text{ where } \hat{g}_1 \ \sim \ \mathcal{P}\big(\sigma_1^2\big). \end{aligned} \tag{34}$$

Based on these rules, an algorithmic description of the NA-BAC and PA-BAC models is presented in Algorithm 3. The normal probability distribution is a continuous distribution with infinite support whereas the variables $K^{(0)}$ and $K^{(1)}$ being modeled have finite support and are discrete (integers). Hence we require the round to nearest integer function in Definition 1. The Poisson probability distribution is a discrete distribution with an infinite support set. Using goodness of fit tests in Section VI, we show that the 2-NA-BAC and 2-PA-BAC models are a better fit than the 2-BAC model for the observed empirical data. However, the 2-NA-BAC and the 2-PA-BAC models are not as good a fit as the 2-BBM model to describe the bit errors in MLC flash memories.

Algorithm 3 NA-BAC and PA-BAC Model Implementation

Input: Input frame $\mathbf{x}$ of length N, sample $(\mathrm{E}[K^{(0)}], \mathrm{Var}[K^{(0)}])$, and sample $(\mathrm{E}[K^{(1)}], \mathrm{Var}[K^{(1)}])$.

Output: Data frame with errors $\mathbf{y}$.

1: Generate integers g_0, g_1 according to the Normal or Poisson approximation rules.
2: $\mathcal{T}_0 = \{i \mid x_i = 0\}$, $\mathcal{T}_1 = \{i \mid x_i = 1\}$.
3: Pick subsets $\mathcal{E}_0$ of size g_0 and $\mathcal{E}_1$ of size g_1 uniformly at random from $\mathcal{T}_0$ and $\mathcal{T}_1$, respectively.
4: Create a binary error vector $\mathbf{e}$ of length N such that $e_i = 1$ if $i \in \mathcal{E}_0 \cup \mathcal{E}_1$.
5: $\mathbf{y} = \mathbf{x} \oplus \mathbf{e}$.

VI. Simulation Results and Evaluation of Channel Models

In this section, we provide a quantitative evaluation of the proposed channel models for MLC flash memories. For this we consider two viewpoints. The first one is a purely statistical viewpoint where we perform the Kolmogorov-Smirnov (K-S) Two Sample test [13] to evaluate the goodness of fit of the proposed channel models when compared with the empirical data. Next, we evaluate the proposed channel models for their application in ECC FER performance estimation. We emphasize the results of this latter evaluation when compared to the former, as accurate ECC FER performance estimation has been the main driving factor in the design of the proposed channel models.

A. Statistical Goodness of Fit Tests

The Kolmogorov-Smirnov (K-S) Two Sample test is a commonly used statistical test for determining if two sets of data samples are drawn from the same probability distribution. The K-S test is a very general test in that it makes no assumptions about the underlying probability distributions of the input data samples and is a non-parametric test [13]. This makes it suitable for our purpose as we have a varied set of underlying probability distributions of the number of bit errors per frame corresponding to the proposed channel models. The BAC and BBM model distributions do not match any well known probability distributions exactly although, they are close to the binomial distribution, and the NA-BAC and PA-BAC model distributions are approximately normal and Poisson respectively.

We perform K-S Two Sample tests comparing the number of bit errors per frame data samples from the proposed channel models to the empirical data obtained from P/E cycling experiments. The empirical data sample sizes, i.e., number of frames for each page, are 8704 for vendor-A and 4096 for vendor-B, respectively. For the BAC, BBM, NA-BAC and PA-BAC models, we simulate 10000 frames. The beta random variates to simulate the BBM channel model and the K-S Two Sample test statistic values are computed using the SciPy library [14]. The test statistic values are shown in Tables IV

and V for $8,000$ and $4,000$ P/E cycles, respectively. The null hypothesis is that the data samples from a proposed channel model and empirical data belong to the same underlying probability distribution. The test statistic is indicative of the difference in underlying probability distributions of the two input data samples. From Table IV, we see that the test statistic values are consistently low for the BBM channel model, thus indicating that it provides the best fit to the empirical data among all the proposed channel models. The p-values recorded (not shown) for all the K-S Two Sample tests in Tables IV and V are smaller than 0.01 indicating that the test statistic values are estimated with a significant level of confidence. The K-S Two Sample test compares the cumulative distribution functions (CDF) obtained from input data samples to compute the test statistic. Fig. 8 and Fig. 9 provide a visual comparison of these CDFs corresponding to vendor-A and vendor-B chips.

TABLE IV
TEST STATISTIC VALUES FROM K-S TWO SAMPLE TESTS COMPARING THE LOWER AND UPPER PAGE BAC, BBM, NA-BAC, PA-BAC MODELS WITH EMPIRICAL DATA AT 8,000 P/E CYCLES. FRAME LENGTH $N = 8192$.

K-S Two Sample Tests	Vendor-A		Vendor-B	
	Lower Page	Upper Page	Lower Page	Upper Page
BAC v/s Experiment	0.0979	0.0744	0.1278	0.0669
BBM v/s Experiment	0.0386	0.0357	0.0190	0.0135
NA-BAC v/s Experiment	0.0430	0.0715	0.0373	0.0659
PA-BAC v/s Experiment	0.0268	0.0777	0.0337	0.1008

TABLE V
TEST STATISTIC VALUES FROM K-S TWO SAMPLE TESTS COMPARING THE LOWER AND UPPER PAGE BAC, BBM, NA-BAC, PA-BAC MODELS WITH EMPIRICAL DATA AT 4,000 P/E CYCLES. FRAME LENGTH $N = 8192$.

K-S Two Sample Tests	Vendor-A		Vendor-B	
	Lower Page	Upper Page	Lower Page	Upper Page
BAC vs. Experiment	0.0498	0.0291	0.0436	0.0422
BBM vs. Experiment	0.0268	0.0153	0.0137	0.0053
NA-BAC vs. Experiment	0.0575	0.0973	0.0632	0.1191
PA-BAC vs. Experiment	0.0642	0.0703	0.0223	0.1953

B. ECC FER Performance Estimation

We evaluate the proposed channel models for their accuracy in ECC FER performance estimation using binary BCH, LDPC, and polar codes. The choice of these ECCs reflects the fact that BCH and LDPC codes are already being used in practical flash memory applications, while polar codes are a promising candidate for the future. The baseline ECC FER performance estimates are obtained from the empirical error data collected from MLC flash memory chips during P/E cycling experiments. As pseudo-random data was written to the flash memory chips during P/E cycling experiments, for ECC decoding we assume an all-zero codeword as the transmitted codeword with the error vector obtained from the empirical error data. This assumption is valid because all the ECCs considered are linear codes. To estimate the ECC FER

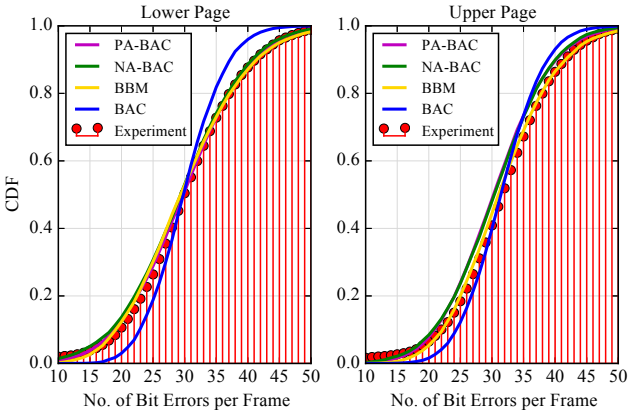

Fig. 8. Comparison of CDFs for number of bit errors per frame observed from empirical data and from the BAC, BBM, NA-BAC, PA-BAC models at 8,000 P/E cycles for vendor-A chip.

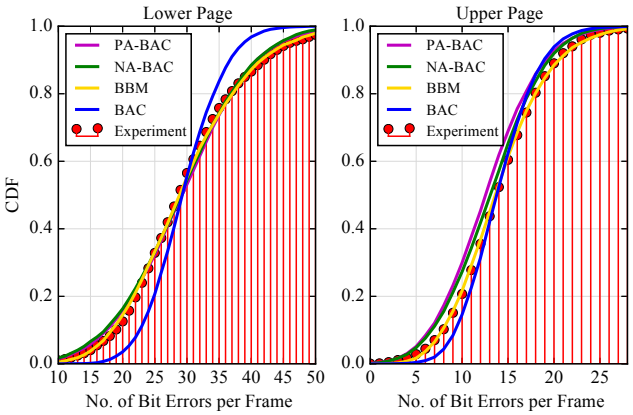

Fig. 9. Comparison of CDFs for number of bit errors per frame observed from empirical data and from the BAC, BBM, NA-BAC, PA-BAC models at 8,000 P/E cycles for vendor-B chip.

performance using the proposed channel models, Monte-Carlo simulations are used where pseudo-random codewords of the ECC are generated and transmitted through the appropriate channel model and the received codeword is decoded. At least 400 frame errors are recorded for FER estimation.

The FER performance of a ($N = 8191$, $k = 7683$, $t = 39$) BCH code using empirical data and the proposed channel models is shown in Fig. 10. Fig. 11 shows the FER performance of a ($N = 8192$, $k = 7683$) regular quasi-cyclic LDPC (QC-LDPC) code with $d_c = 64$ and $d_v = 4$, where d_c and d_v refer to the check node and variable node degrees, respectively, in the parity check matrix. The parity check matrix of the QC-LDPC code is constructed using size 128×128 circulant permutation matrices and the design rate is specified as 0.9375. To ensure the required variable node degree d_v, exactly d_v permutations of the circulant matrix are stacked vertically along the rows of the parity check matrix for every set of columns. Zero matrices of size 128×128 are used to fill up any remaining rows. This is done using the progressive edge growth (PEG) algorithm [15] to avoid short cycles. Note that although the specified design rate corresponds to

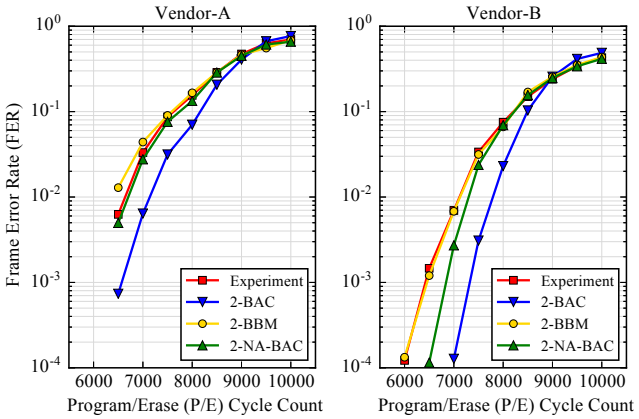

Fig. 10. Comparison of FER performance of a ($N = 8191$, $k = 7683$, $t = 39$) BCH code using empirical error data and error data from simulation using the 2-BAC, 2-BBM, 2-NA-BAC channel models for vendor-A and vendor-B chips.

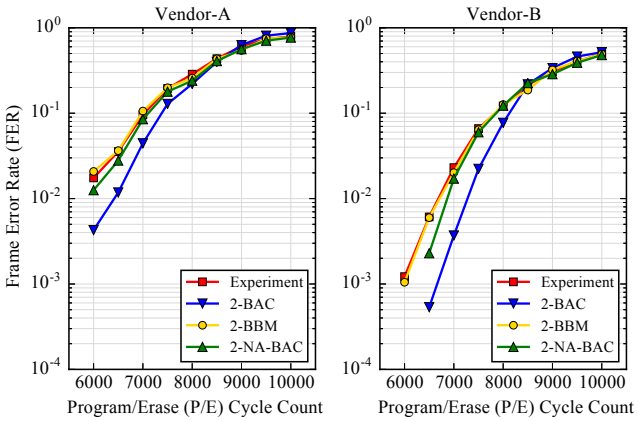

Fig. 11. Comparison of FER performance of a ($N = 8192$, $k = 7683$) regular QC-LDPC code using empirical error data and error data from simulation using the 2-BAC, 2-BBM, 2-NA-BAC channel models for vendor-A and vendor-B chips.

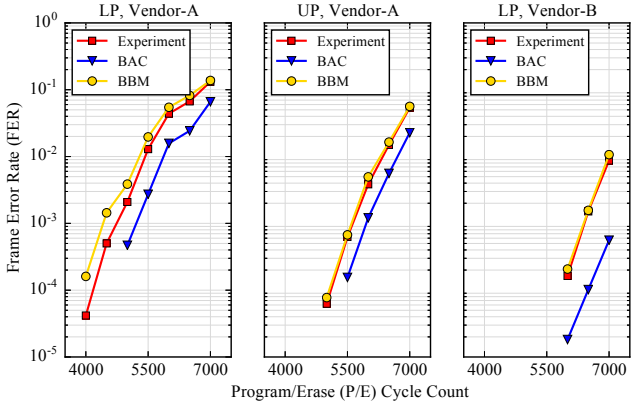

Fig. 12. Comparison of FER performance of a ($N = 8192$, $k = 7683$) regular QC-LDPC code using empirical error data and error data from simulation using the BAC and BBM channel models for both lower and upper pages of vendor-A chip and the lower page of vendor-B chip.

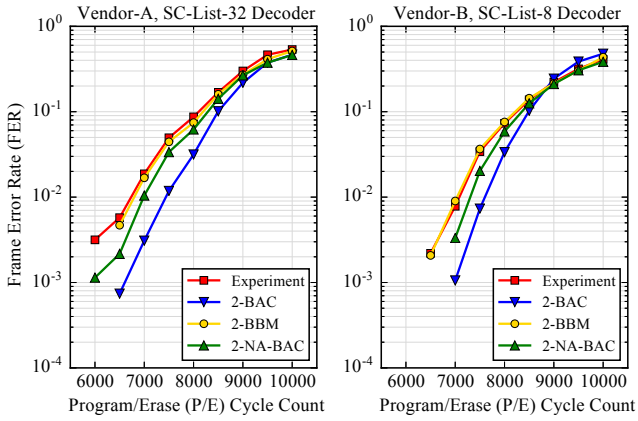

Fig. 13. Comparison of FER performance of a ($N = 8192$, $k = 7684$) polar code optimized for BSC(0.001) using empirical error data and error data from simulation using the 2-BAC, 2-BBM, 2-NA-BAC channel models for vendor-A and vendor-B chips. The SC-List decoder is used with a list size $= 32$ for vendor-A and list size $= 8$ for vendor-B chip.

a code dimension of 7680, we get $k = 7683$ due to three dependent parity checks in the final parity check matrix thus obtained. A sum-product belief propagation decoder with a maximum of 50 iterations and early termination is used to decode the QC-LDPC code. Fig. 12 also shows additional results comparing the FER performance of the QC-LDPC code obtained using empirical data and simulation data from the BAC, BBM channel models, separately for the lower and upper pages of vendor-A chip and the lower page of vendor-B chip. The lowest FER performance estimates from empirical data were obtained by P/E cycling 44 and 24 blocks of vendor-A and vendor-B chips, respectively. A total of 6 and 4 frame errors were observed to obtain the lowest FER performance estimates from empirical data for the lower and upper pages of vendor-A chip, respectively. For the lower page of vendor-B chip, 4 frame errors were observed to estimate the lowest FER performance from empirical data. Note that the results for the upper page of vendor-B chip are not shown as we did not observe any frame errors in the empirical data. We also note that a different vendor-B chip was used to obtain the additional results shown in Fig. 12 when compared to the rest of the paper. Fig. 13 shows the comparison of FER performance of a ($N = 8192$, $k = 7684$) polar code using empirical data and the proposed channel models. The polar code is optimized for a binary symmetric channel (BSC) with bit error probability $p = 0.001$ using the construction technique proposed in [16]. The successive cancellation list (SC-List) decoder proposed in [17] is used for decoding the polar code.

For all the ECCs considered and using data from both vendor chips, we observe that the 2-BAC model provides an optimistic estimate of the FER performance when compared to the empirically observed FER performance. This is mainly due to the inability of the 2-BAC model to capture the high variance in the number of bit errors per frame observed empirically. The gap in ECC FER performance estimates using the 2-BAC model and the empirical data is increasing as the FER decreases, and it is about an order of magnitude for vendor-A chip at $6,500$ P/E cycles and greater than an order of magnitude for vendor-B chip at $7,000$ P/E cycles for

the BCH code as shown in Fig. 10. This gap in ECC FER performance estimates at low FERs is bad for determining the correct endurance (life-time) of a flash memory chip. From the results shown in Fig. 12 for the QC-LDPC code, we observe that the BBM channel model estimates the FER performance accurately even at lower FERs around 10^{-4}, for the upper page of vendor-A chip and the lower page of vendor-B chip. The FER performance estimates obtained using the BBM channel model are better than those obtained using the BAC channel model for the lower page of vendor-A chip, however we observe a small mismatch in the BBM channel model FER performance estimates at lower FERs when compared to the empirical FER estimates. This mismatch is due to the inability of the BBM channel model to fit the larger proportions of frames with small number of bit errors observed in the lower tail of the empirical error histograms for the lower page of vendor-A chip. This appears to be a vendor-specific effect, as this kind of effect was not observed in the empirical error histograms corresponding to the lower page of vendor-B chip. Overall, the 2-BBM model is able to match the empirical ECC FER performance estimates accurately, while the estimates obtained using the 2-NA-BAC model lie between those of the 2-BAC and the 2-BBM models. The ECC FER performance estimates using the 2-PA-BAC model are the same as those using the 2-NA-BAC model and are omitted. From these results it is clear that the proposed 2-BBM channel model is able to accurately describe the nature of the number of bit errors per frame in MLC flash memories and hence provides accurate estimates of the ECC FER performance.

VII. Conclusion

We studied the feasibility of using well known discrete memoryless channel models to model the MLC flash memory channel. Based on empirical error analysis and ECC FER performance estimation for BCH, LDPC, and polar codes, we observe that the 2-BAC model with parameter estimates derived from empirical error data suffices to produce an accurate estimate of the average raw bit error rate, but it provides an incorrect optimistic estimate of the ECC FER performance when compared to the empirically observed ECC FER performance. This is mainly due to the overdispersed nature of the number of bit errors per frame in empirical data which is not modeled well by the 2-BAC model. We proposed the 2-Beta-Binomial (2-BBM) channel model based on the beta-binomial probability distribution and using statistical analysis, goodness of fit tests and ECC FER performance results showed that the 2-BBM channel model accurately describes the nature of the number of bit errors per frame in MLC flash memories. We also note that the BBM channel model can be shown to be equivalent to an urn based channel model [18] and hence has memory associated with it. Although not presented in this paper, our preliminary experiment results for combined data retention plus P/E cycling stress show the evidence of overdispersion in error statistics and the suitability of the proposed 2-BBM channel model. We leave a detailed examination of this as future work. Although the proposed channel models are for MLC flash memories, the proposed empirical design approach is generic and can easily be extended for three-level cell (TLC) flash memories.

Appendix A
Proof of Proposition 1

To compute $\mathrm{Var}[K]$, we compute its mean $\mathrm{E}[K]$ and the second moment $\mathrm{E}[K^2]$. Based on (1), both these moments of K can be computed from the moments of K_m as

$$\mathrm{E}[K] = \sum_{m=0}^{N} \frac{\binom{N}{m}}{2^N} \mathrm{E}[K_m] \tag{35}$$

$$\mathrm{E}[K^2] = \sum_{m=0}^{N} \frac{\binom{N}{m}}{2^N} \mathrm{E}[K_m^2]. \tag{36}$$

From (2) and (7), we have

$$\begin{aligned}\mathrm{E}[K_m] &= \mathrm{E}[K_m^{(0)}] + \mathrm{E}[K_{N-m}^{(1)}] \\ &= mp + (N-m)q \end{aligned} \tag{37}$$

$$\begin{aligned}\mathrm{Var}[K_m] &= \mathrm{Var}[K_m^{(0)}] + \mathrm{Var}[K_{N-m}^{(1)}] \\ &= mp(1-p) + (N-m)q(1-q). \end{aligned} \tag{38}$$

Therefore, $\mathrm{E}[K_m^2]$ is given by

$$\begin{aligned}\mathrm{E}[K_m^2] &= \mathrm{Var}[K_m] + (\mathrm{E}[K_m])^2 \\ &= mp + (N-m)q + m(m-1)p^2 + 2m(N-m)pq \\ &\quad + (N-m)(N-m-1)q^2. \end{aligned} \tag{39}$$

Hence $\mathrm{E}[K]$ and $\mathrm{E}[K^2]$ are given by

$$\begin{aligned}\mathrm{E}[K] &= \sum_{m=0}^{N} \frac{\binom{N}{m}}{2^N} \mathrm{E}[K_m] \\ &= \frac{N}{2}(p+q) \end{aligned} \tag{40}$$

$$\begin{aligned}\mathrm{E}[K^2] &= \sum_{m=0}^{N} \frac{\binom{N}{m}}{2^N} \mathrm{E}[K_m^2] \\ &= \frac{N}{2}(p+q) + \Big(\frac{N^2-N}{2}\Big)pq \\ &\quad + \Big(\frac{N^2-N}{4}\Big)(p^2+q^2). \end{aligned} \tag{41}$$

Note that we have used the combinatorial identities

$$\sum_{m=0}^{N} \binom{N}{m} m = N2^{N-1} \tag{42}$$

$$\sum_{m=0}^{N} \binom{N}{m} m^2 = (N+N^2)2^{N-2}. \tag{43}$$

Therefore we can obtain $\mathrm{Var}[K]$ from (40) and (41) as

$$\begin{aligned}\mathrm{Var}[K] &= \mathrm{E}[K^2] - (\mathrm{E}[K])^2 \\ &= \frac{N}{2}\Big((p+q) - pq - \frac{1}{2}(p^2+q^2)\Big). \end{aligned} \tag{44}$$

■

APPENDIX B
PROOF OF PROPOSITION 2

We take the same approach as the proof of Proposition 1. From (2) and (16), we have

$$\mathrm{E}[K_m] = \left(\frac{ma}{a+b}\right) + \left(\frac{(N-m)c}{c+d}\right) \tag{45}$$

$$\mathrm{Var}[K_m] = \left(\frac{mab(a+b+m)}{(a+b)^2(a+b+1)}\right) + \left(\frac{(N-m)cd(c+d+N-m)}{(c+d)^2(c+d+1)}\right). \tag{46}$$

Therefore, $\mathrm{E}[K_m^2]$ is given by

$$\begin{aligned}\mathrm{E}[K_m^2] &= \mathrm{Var}[K_m] + (\mathrm{E}[K_m])^2 \\ &= \mathrm{Var}[K_m^{(0)}] + (\mathrm{E}[K_m^{(0)}])^2 + \mathrm{Var}[K_{N-m}^{(1)}] + \\ &\quad (\mathrm{E}[K_{N-m}^{(1)}])^2 + 2\,\mathrm{E}[K_m^{(0)}]\,\mathrm{E}[K_{N-m}^{(1)}].\end{aligned} \tag{47}$$

Substituting using (21) - (24) and simplifying, we have

$$\begin{aligned}\mathrm{E}[K_m^2] = &\left(\frac{ma(m(a+1)+b)}{(a+b)(a+b+1)}\right) + \\ &\left(\frac{(N-m)c((N-m)(c+1)+d)}{(c+d)(c+d+1)}\right) + \\ &\left(\frac{2m(N-m)ac}{(a+b)(c+d)}\right).\end{aligned} \tag{48}$$

Hence $\mathrm{E}[K]$ and $\mathrm{E}[K^2]$ are given by

$$\begin{aligned}\mathrm{E}[K] &= \sum_{m=0}^{N} \frac{\binom{N}{m}}{2^N}\ \mathrm{E}[K_m] \\ &= \frac{N}{2}\left(\frac{a}{a+b} + \frac{c}{c+d}\right)\end{aligned} \tag{49}$$

$$\begin{aligned}\mathrm{E}[K^2] &= \sum_{m=0}^{N} \frac{\binom{N}{m}}{2^N}\ \mathrm{E}[K_m^2] \\ &= \frac{N}{4}\left(\frac{(N+1)a(a+1)+2Nab}{(a+b)(a+b+1)}\right) + \\ &\quad \frac{N}{4}\left(\frac{(N+1)c(c+1)+2Ncd}{(c+d)(c+d+1)}\right) + \\ &\quad \frac{N(N-1)}{4}\left(\frac{2ac}{(a+b)(c+d)}\right).\end{aligned} \tag{50}$$

We have used the combinatorial identities (42) and (43). From (49) and (50), $\mathrm{Var}[K]$ is easily obtained as

$$\begin{aligned}\mathrm{Var}[K] &= \mathrm{E}[K^2] - (\mathrm{E}[K])^2 \\ &= \frac{N}{4}\left(\frac{a(a+b)(a+2b+1)+Nab}{(a+b)^2(a+b+1)}\right) + \\ &\quad \frac{N}{4}\left(\frac{c(c+d)(c+2d+1)+Ncd}{(c+d)^2(c+d+1)}\right) - \\ &\quad \frac{N}{4}\left(\frac{2ac}{(a+b)(c+d)}\right).\end{aligned} \tag{51}$$

■

APPENDIX C
PROOF OF PROPOSITION 3

This proof proceeds along similar lines as the proof of Proposition 2. From (4) and (19),

$$\Pr(K^{(0)} = k) = \sum_{m=k}^{N} \frac{\binom{N}{m}}{2^N}\binom{m}{k}\frac{B(a+k, b+m-k)}{B(a,b)}$$

$$\begin{aligned}\mathrm{E}[K^{(0)}] &= \sum_{k=0}^{N} k\Pr(K^{(0)}=k) \\ &= \frac{1}{2^N}\sum_{k=0}^{N}\sum_{m=k}^{N} k\binom{N}{m}\binom{m}{k}\frac{B(a+k,b+m-k)}{B(a,b)} \\ &= \frac{1}{2^N}\sum_{m=0}^{N}\binom{N}{m}\sum_{k=0}^{m} k\binom{m}{k}\frac{B(a+k,b+m-k)}{B(a,b)} \\ &= \frac{1}{2^N}\sum_{m=0}^{N}\binom{N}{m}\,\mathrm{E}[K_m^{(0)}] \\ &= \frac{N}{2}\left(\frac{a}{a+b}\right)\end{aligned} \tag{52}$$

$$\begin{aligned}\mathrm{E}[(K^{(0)})^2] &= \sum_{k=0}^{N} k^2\Pr(K^{(0)}=k) \\ &= \frac{1}{2^N}\sum_{k=0}^{N}\sum_{m=k}^{N} k^2\binom{N}{m}\binom{m}{k}\frac{B(a+k,b+m-k)}{B(a,b)} \\ &= \frac{1}{2^N}\sum_{m=0}^{N}\binom{N}{m}\sum_{k=0}^{m} k^2\binom{m}{k}\frac{B(a+k,b+m-k)}{B(a,b)} \\ &= \frac{1}{2^N}\sum_{m=0}^{N}\binom{N}{m}\,\mathrm{E}[(K_m^{(0)})^2] \\ &= \frac{N}{4}\left(\frac{a(a+2b+1)+Na(a+1)}{(a+b)(a+b+1)}\right)\end{aligned} \tag{53}$$

$$\begin{aligned}\mathrm{Var}[K^{(0)}] &= \mathrm{E}[(K^{(0)})^2] - (\mathrm{E}[K^{(0)}])^2 \\ &= \frac{N}{4}\left(\frac{a(a+b)(a+2b+1)+Nab}{(a+b)^2(a+b+1)}\right).\end{aligned} \tag{54}$$

We have used the combinatorial identities (42) and (43) and also the fact that the second moment of a beta-binomial random variable $K_m^{(0)} \sim$ Beta-Binomial(m,a,b) is given by $\frac{ma(m(a+1)+b)}{(a+b)(a+b+1)}$. The expressions for $\mathrm{E}[K^{(1)}]$ and $\mathrm{Var}[K^{(1)}]$ can be derived similarly. ■

REFERENCES

[1] R. Bez, E. Camerlenghi, A. Modelli, and A. Visconti, "Introduction to flash memory," *Proceedings of the IEEE*, vol. 91, no. 4, pp. 489–502, April 2003.

[2] J. Cooke, "The inconvenient truths about NAND flash memory," in *Micron MEMCON 7*, 2007.

[3] J. D. Lee, S. H. Hur, and J. D. Choi, "Effects of floating-gate interference on NAND flash memory cell operation," *IEEE Electron Device Lett.*, vol. 23, no. 5, pp. 264–266, May 2002.
[4] X. Huang, A. Kavcic, X. Ma, G. Dong, and T. Zhang, "Multilevel flash memories: Channel modeling, capacities and optimal coding rates," *International Journal on Advances in Systems and Measurements*, vol. 6, no. 3 & 4, pp. 364–373, 2013.
[5] Y. Cai, E. F. Haratsch, O. Mutlu, and K. Mai, "Threshold voltage distribution in MLC NAND flash memory: Characterization, analysis, and modeling," in *Proc. of the Conference on Design, Automation and Test in Europe (DATE)*, 2013, pp. 1285–1290.
[6] T. Parnell, N. Papandreou, T. Mittelholzer, and H. Pozidis, "Modelling of the threshold voltage distributions of sub-20 nm NAND flash memory," in *Proc. IEEE Global Communications Conference (GLOBECOM)*, Dec 2014, pp. 2351–2356.
[7] E. Yaakobi, L. Grupp, P. H. Siegel, S. Swanson, and J. K. Wolf, "Characterization and error-correcting codes for TLC flash memories," in *Proc. International Conference on Computing, Networking and Communications (ICNC)*, January 2012, pp. 486–491.
[8] E. Yaakobi, J. Ma, L. Grupp, P. H. Siegel, S. Swanson, and J. K. Wolf, "Error characterization and coding schemes for flash memories," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM) Workshops*, December 2010, pp. 1856–1860.
[9] V. Taranalli, H. Uchikawa, and P. H. Siegel, "Error analysis and inter-cell interference mitigation in multi-level cell flash memories," in *Proc. IEEE International Conference on Communications (ICC)*, London, UK, June 2015, pp. 271–276.
[10] J. G. Skellam, "A probability distribution derived from the binomial distribution by regarding the probability of success as variable between the sets of trials," *Journal of the Royal Statistical Society. Series B (Methodological)*, vol. 10, no. 2, pp. 257–261, 1948.
[11] J. K. Lindsey and P. M. E. Altham, "Analysis of the human sex ratio by using overdispersion models," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 47, no. 1, pp. 149–157, 1998.
[12] C. H. Stapper, A. N. McLarent, and M. Dreckmann, "Yield model for productivity optimization of VLSI memory chips with redundancy and partially good product," *IBM Journal of Research and Development*, vol. 24, no. 3, pp. 398–409, May 1980.
[13] F. J. Massey, "The Kolmogorov-Smirnov test for goodness of fit," *Journal of the American Statistical Association*, vol. 46, no. 253, pp. 68–78, 1951.
[14] E. Jones, T. Oliphant, P. Peterson *et al.*, "SciPy: Open source scientific tools for Python," 2001–. [Online]. Available: http://www.scipy.org/
[15] D. M. Arnold, E. Eleftheriou, and X. Y. Hu, "Progressive edge-growth tanner graphs," in *Proc. IEEE Global Telecommunications Conference (GLOBECOM)*, vol. 2, 2001, pp. 995–1001.
[16] I. Tal and A. Vardy, "How to construct polar codes," *IEEE Transactions on Information Theory*, vol. 59, no. 10, pp. 6562–6582, October 2013.
[17] I. Tal and A. Vardy, "List decoding of polar codes," *IEEE Transactions on Information Theory*, vol. 61, no. 5, pp. 2213–2226, May 2015.
[18] F. Alajaji and T. Fuja, "A communication channel modeled on contagion," *IEEE Transactions on Information Theory*, vol. 40, no. 6, pp. 2035–2041, Nov 1994.

Veeresh Taranalli (S'06-) is a Ph.D. candidate in Electrical Engineering at the University of California, San Diego, CA, USA. He received his M.S. degree in Electrical Engineering from the University of California, San Diego, CA, USA in 2013; and his B.Tech degree in Electronics and Communication Engineering from the National Institute of Technology Karnataka, Surathkal, India, in 2009. His research interests include error characterization, channel modeling and study of modern coding techniques for NAND flash memory applications. He received the 2015 Shannon Memorial Fellowship awarded by the University of California San Diego.

Hironori Uchikawa received the B.E. and M.E. degrees in electrical engineering from Yokohama National University in 2001 and 2003 respectively. He received the Ph.D. degree in electrical engineering from Tokyo Institute of Technology in 2012. Since 2003, he has been with Toshiba Corporation. From March 2013 to December 2014, he was a Visiting Scholar in the Center for Memory Recording Research at University of California, San Diego. His research interests include coding theory, information theory, communication theory and their application to storage systems. He received the 2008 Young Researcher's Award of the Institute of Electronics, Information and Communication Engineers of Japan.

Paul H. Siegel (M'82-SM'90-F'97) received the S.B. and Ph.D. degrees in mathematics from Massachusetts Institute of Technology (MIT), Cambridge, MA, USA, in 1975 and 1979, respectively. He held a Chaim Weizmann Postdoctoral Fellowship with the Courant Institute, New York University, New York, NY, USA. He was with the IBM Research Division, San Jose, CA, USA, from 1980 to 1995. He joined the Faculty with the University of California, San Diego, CA, USA, in July 1995, where he is currently a Professor of Electrical and Computer Engineering in the Jacobs School of Engineering. He is affiliated with the Center for Memory and Recording Research where he holds an Endowed Chair and served as Director from 2000 to 2011. His research interests include information theory and communications, particularly coding and modulation techniques, with applications to digital data storage and transmission. He was a Member of the Board of Governors of the IEEE Information Theory Society from 1991 to 1996 and again from 2009 to 2014. He served as Co-Guest Editor of the May 1991 Special Issue on "Coding for Storage Devices" of the IEEE TRANSACTIONS ON INFORMATION THEORY. He served the same Transactions as Associate Editor for Coding Techniques from 1992 to 1995, and as Editor-in-Chief from July 2001 to July 2004. He was also Co-Guest Editor of the May/September 2001 two-part issue on "The Turbo Principle: From Theory to Practice" and the February 2016 issue on "Recent Advances in Capacity Approaching Codes" of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS. He is a member of the National Academy of Engineering. He was the 2015 Padovani Lecturer of the IEEE Information Theory Society. He was the recipient of the 2007 Best Paper Award in Signal Processing and Coding for Data Storage from the Data Storage Technical Committee of the IEEE Communications Society. He was the corecipient of the 1992 IEEE Information Theory Society Paper Award and the 1993 IEEE Communications Society Leonard G. Abraham Prize Paper Award.