Optimized Degree Distributions for Binary and Non-Binary LDPC Codes in Flash Memory

Kasra Vakilinia, Dariush Divsalar*, and Richard D. Wesel
Department of Electrical Engineering, University of California, Los Angeles, Los Angeles, California 90095
*Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California 91109

***Abstract*—This paper uses extrinsic-information-transfer (EXIT)-function analysis employing the reciprocal channel approximation (RCA) to obtain optimal LDPC code degree distributions for initial hard decoding (one-bit quantization of the channel output) and for decoding with the soft information provided by additional reads in both SLC (two-level cell) and MLC (four-level-cell) Flash memory. These results indicate that design for hard decoding can provide irregular degree distributions that have good thresholds across the range of possible decoding precisions. These results also quantify the potential benefit of irregular LDPC codes as compared to regular LDPC codes in the flash setting and compare the RCA-EXIT thresholds of word-line voltages optimized for maximum mutual information (MMI) and word-line voltages that explicitly minimize the RCA-EXIT threshold of a specific LDPC degree distribution. Along the way, this paper illustrates that the MMI optimization of word-line voltages for five reads is a quasi-convex problem for the Gaussian model of SLC Flash. The paper also uses RCA-based EXIT analysis to show that for the same spectral efficiency of 0.9 bits per cell, rate-0.45 non-binary LDPC codes on MLC Flash systems provide thresholds more than 0.5 dB better than rate-0.9 binary LDPC codes on SLC Flash with the same number of reads (i.e. three reads that would provide hard decisions for MLC and limited soft information for SLC). The MLC approach has a potential threshold improvement of more than 1.5 dB over the SLC approach when both systems have access to the full soft information.**

I. Introduction

Single-level-cell (SLC) NAND Flash memory cells have two charge levels. More recent multi-level-cell (MLC) Flash memory cells have four charge levels. The current SLC and MLC Flash systems commonly use binary low-density parity-check (LDPC) codes to protect the information stored in memory cells from distortion introduced by wear-out noise, retention loss, and cell-to-cell interference [1]. Flash memory cells traditionally employ hard decoding. However, multiple reads with distinct (optimized) word-line voltages provide additional soft information that significantly improves the error-correction capability of the LDPC codes used in Flash systems [2], [3]. These additional reads significantly increase decoding time, so they might only be employed when needed to facilitate successful decoding. Thus a single LDPC code might be decoded first with hard-decoded symbols and later with soft information. This is an example of a feedback communication system with multiple decoding attempts where incremental redundancy is transmitted only when needed.

Results in [3] suggest that the best LDPC degree distributions are different for these different decoding scenarios. This paper uses RCA-based EXIT function analysis to investigate the optimization of LDPC degree distributions and word-line voltage optimization in light of the multiple decoding attempts.

The remainder of the paper proceeds as follows: Sec. II provides an overview of Flash with multiple reads and the maximum-mutual-information (MMI) approach for selecting word-line voltages. The contributions of this section are the illustration that the word-line voltage optimization for the five-read Gaussian model of SLC flash is a quasi-convex problem and the consideration of three-read and five-read word-line thresholds from the perspective of multiple decoding attempts.

Sec. III presents RCA-based EXIT-function analysis. Sec. IV applies this technique to optimize degree distributions for LDPC codes for various precision levels for both SLC and MLC Flash. The contributions of these sections include the use of RCA-based EXIT analysis to study the effect of decoding precision on degree-distribution optimization, an explicit quantification of the trade-off associated with selecting a single code for multiple decoding attempts at increasing precision levels, a validation of the use of the MMI-optimized word-line voltages of [3] for RCA-EXIT based degree distribution optimization, and a demonstration of the potential benefit of using MLC even when the information density is in the range traditionally associated with SLC. Sec. V concludes the paper.

II. Multiple-read Flash systems

In this paper we assume i.i.d. Gaussian threshold voltages for each charge level in SLC or MLC Flash memory cells. These models are equivalent to 2-level or 4-level Pulse-Amplitude Modulation with additive white Gaussian ($\mathcal{N}(0,\sigma^2)$) noise (AWGN). Fig. 1 shows the model of the threshold voltage distribution as a mixture of the two identically distributed Gaussian random variables that comprise the SLC model. Since the levels are at $+1$ and -1, the average energy (E_{avg}) of this model is 1.

Similar to Fig. 1, the shifted 4-level MLC threshold voltage distribution is a mixture of four identically distributed Gaussian random variables. The voltage levels are

This material is based upon work supported by the National Science Foundation under Grant Numbers 1162501 and 1161822. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. This research was carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, under NASA JPL Task Plan 82-17473.

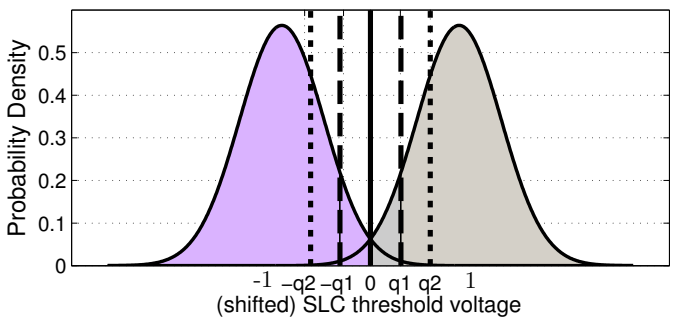

Fig. 1: SLC threshold voltage model

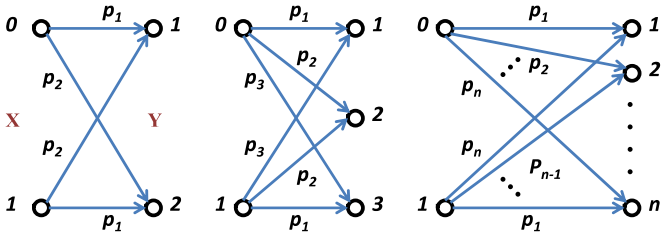

Fig. 2: SLC equivalent discrete read channels

at $\{+\frac{3}{\sqrt{5}}, +\frac{1}{\sqrt{5}}, -\frac{1}{\sqrt{5}}, -\frac{3}{\sqrt{5}}\}$ and correspond to the Gray labeled assignment $\{00, 01, 11, 10\}$ respectively. In this model $E_{avg} = 1$, similar to the SLC model.

A. Progressive Quantization on Read Channels

Each time a cell is read, the result is a single bit indicating whether the threshold voltage is above the word-line voltage at the time of the read. As described in [2], [3], reading the same SLC cell $n-1$ times (for $n \geq 2$) with different word-line voltages effectively produces an equivalent channel with two inputs and n outputs as shown in Fig. 2.

In [2], [3], the word-line voltages for each quantization are chosen by maximizing the mutual information (MI) between the input and output of the equivalent discrete read channels. Let's consider the MI for the three equivalent channels corresponding to a single read, three reads, and five reads, respectively. This set of channels might be seen by three LDPC decoding attempts with progressively more reads.

Define $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-u^2/2} du$, $Q_\sigma^-(x) = Q(\frac{x-1}{\sigma})$, and let $p_{ij} = p_i + p_j$.

For the 1-read SLC in Fig. 2a, the MI can be expressed as

$$I(X;Y) = 1 - H(p_1, p_2)\,, \tag{1}$$

where H is the entropy function. $p_1 = Q_\sigma^-(q)$ and $p_2 = 1 - p_1$. The quantization threshold (q) that maximizes the MI is $q = 0$.

For the 3-read SLC model in Fig. 2c with $n = 4$,

$$I(X;Y) = H(p_{14}, p_{23}) + 1 - H(p_1, p_2, p_3, p_4)\,, \tag{2}$$

where $p_1 = Q_\sigma^-(q_1)$, $p_2 = Q_\sigma^-(0) - p_1$, $p_3 = Q_\sigma^-(-q_1) - p_{12}$, and $p_4 = 1 - Q_\sigma^-(-q_1)$ with q_1 the word-line voltage shown

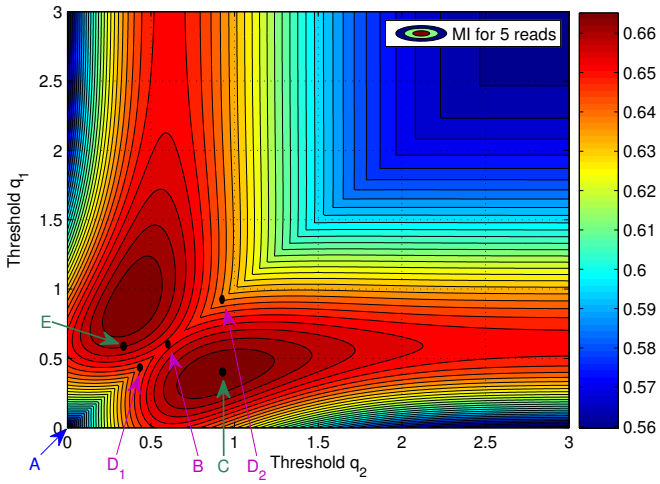

Fig. 3: Mutual information vs. q_1 and q_2 for SLC_5

in Fig 1. As described in [3], for 3-read SLC, the optimal q_1 satisfies $\frac{dI}{dq_1} = 0$. Since the MI is quasi-concave in q_1, the optimal q_1 can be found by a bisection search algorithm.

For the 5-read SLC (SLC_5) model in Fig. 2c with $n = 6$,

$$I(X;Y) = H(p_{16}, p_{25}, p_{34}) + 1 - H(p_1, p_2, \ldots, p_6)\,, \tag{3}$$

where $p_1 = Q_\sigma^-(q_2)$, $p_2 = Q_\sigma^-(q_1) - p_1$, $p_3 = Q_\sigma^-(0) - p_{12}$, $p_4 = Q_\sigma^-(-q_1) - Q_\sigma^-(0)$, $p_5 = Q_\sigma^-(-q_2) - Q_\sigma^-(-q_1)$, and $p_6 = 1 - Q_\sigma^-(-q_2)$ for q_1 and q_2 shown in Fig 1.

Fig. 3 shows the contour plot of mutual information (MI) vs. (q_1,q_2) for the 5-read case. The maximum MI points can be obtained by solving the two partial differential equations $\frac{dI}{dq_1} = 0$ and $\frac{dI}{dq_2} = 0$. Assuming $q_2 \geq q_1$, the MI is quasi-concave in q_1 for a fixed value of q_2 and vice-versa (in each dimension) and can be maximized by bisection search.

Let's use Fig. 3 to explore the effect of multiple decoding attempts with progressive reads on word-line voltage optimization. Note that the MI along the $q_1 = q_2$ line shows the three-read MI performance as a function of q_1. The optimal single-read word-line voltage of zero (Point A) is also optimal as the first read of three reads or five reads. However, The optimal position of q_1 for three reads is $q_1 = 0.61$ (point B), while the optimal position of q_1 for five reads is $q_1 = 0.4$ (with $q_2 = 0.9$, point C). We note that the MI obtained with three reads is 0.6524 when q_1 is optimized for three reads, 0.6439 when q_1 is optimized for five reads with $q_1 = 0.4$ (point D_1), and 0.6414 when q_1 is optimized for five reads with $q_1 = 0.9$ (point D_2). Also, the MI obtained with five reads is 0.6634 when q_1 is optimized for three reads (point E, with $q_1 = 0.61$ and $q_2 = 0.28$) and 0.6687 when q_1 is optimized for five reads (point C), assuming that q_2 is chosen optimally in each case. The MI differences are small in either case, but large enough to have noticeable effects on frame error rate according to [3]. A greedy approach (optimizing q_1 for three-read performance) leads to the smallest degradation and would lower decoding time by increasing the probability of decoding successfully

after three reads, but optimizing q_1 for the five-read scenario will produce the lowest probability of losing a page.

III. RCA-based EXIT Function Analysis

In their density evolution analysis of binary-input AWGN channels, Chung et al. [4] observed that the distribution of the LLR messages at any iteration is approximately Gaussian. They showed that based on a stability condition, the variance of the Gaussian distribution has to be twice its mean. Therefore, the Gaussian distribution could be fully characterized by a single parameter. They also observed that due to a duality property (reciprocal approximation) the LLR messages from the variable to check nodes become additive at check nodes.

The references [5] and [6] applied similar ideas by using the EXtrinsic mutual Information Transfer (EXIT) functions to simplify the DE threshold analysis. EXIT functions in iterative coding schemes characterize how *a priori* input information to a decoder check or variable node converts into *extrinsic* output information from the decoder check or variable node. The EXIT functions track the evolution of the MI between the variable-to-check messages and the variable node true value (I_v^{out}) and the MI between check-to-variable messages and the variable node true value (I_v^{in}).

We now describe the use of EXIT functions with the Gaussian approximation and RCA to compute LDPC decoding thresholds for different number of reads. In the following sections, similar to Fig. 2, p_is represent the transition probabilities. The LLRs are given by $L_{ij} = \log \frac{p_i}{p_j}$. $f_L^{ch}(l)$ represents the apriori channel LLR probability distribution function (p.d.f.). The notation $\oplus$ represents the convolution operator. $\delta(.)$ represents the Dirac delta function.

A. Binary LDPC Thresholds from RCA-based EXIT Functions

By the assumption of uniform input, symmetric channel, and transmission of the all-zero codeword the initial channel LLR ($\log \frac{P(Y|X=0)}{P(Y|X=1)}$) distribution at a variable node for the single-read SLC channel in Fig. 2a is

$$f_L^{ch}(l) = p_1\delta(l - L_{12}) + p_2\delta(l - L_{21})\,, \tag{4}$$

where $L_{ij} = \log \frac{p_i}{p_j}$. This p.d.f. is effectively a p.m.f. with support at $L_{12} = \log(\frac{P(Y=1|X=0)}{P(Y=1|X=1)})$ and $L_{21} = \log(\frac{P(Y=2|X=0)}{P(Y=2|X=1)})$. Since the all-zeros codeword is assumed to be transmitted, the channel LLR takes the values of L_{12} and L_{21} with probabilities p_1 and p_2 respectively. Similar expressions hold for more reads, e.g. with five reads the channel LLR p.d.f. is

$$\begin{aligned} f_L^{ch}(l) = & p_1\delta(l - L_{16}) + p_2\delta(l - L_{25}) + p_3\delta(l - L_{34}) \\ & + p_4\delta(l - L_{43}) + p_5\delta(l - L_{52}) + p_6\delta(l - L_{61})\,, \end{aligned} \tag{5}$$

with the p_is as defined previously for the five-read channel. For the full soft-information (i.e. in the limit of infinitely many reads of the flash cell) BI-AWGN channel, $LLR = \log(\frac{P(Y=y|X=0)}{P(Y=y|X=1)}) = \frac{2}{\sigma^2}y$ where σ^2 is the variance of the Gaussian noise. Therefore, the LLR is normally distributed with a mean of $\frac{2}{\sigma^2}$ and variance of $\frac{4}{\sigma^2}$:

$$f_L^{ch}(l) = \mathcal{N}(2/\sigma^2, 4/\sigma^2)\,. \tag{6}$$

In the first iteration (initialization step) of the threshold calculation algorithm, the extrinsic mutual information from a variable node is calculated by $I_v^{out} = J(f_L^{ch}(l))$ where

$$J(f_L(l)) = 1 - \int_L \log_2(1 + e^{-l}) f_L(l) dl \tag{7}$$

$J(f_L(l))$ is a function of the LLR distribution $f_L(l)$ which gives the mutual information of a binary-input symmetric-output channel with the LLR distribution of $f_L(l)$. For instance, for the 1-read SLC channel we can alternatively calculate the mutual information of (1) as $J(f_L^{ch}(l))$ where $f_L^{ch}(l)$ is given in (4). If the LLR distribution is in form of $\mathcal{N}(\mu, 2\mu)$, $J_{\mathcal{N}}(\mu) = J(\mathcal{N}(\mu, 2\mu))$.

After initialization, (8-13) are performed. This process is repeated until $H(X) - I_v^{out} < \epsilon'$ (e.g. 10^{-7}), where $H(X)$ is the entropy of the input in Fig. 2. This is the convergence or the stopping criterion of the algorithm. If the threshold-calculation algorithm fails at a particular noise level, the noise level is decreased until the maximum noise level (minimum E_b/N_0) at which the algorithm converges is identified.

$$I_c^{in} = 1 - I_v^{out} \tag{8}$$

$$\mu_c^{in} = J_{\mathcal{N}}^{-1}(I_c^{in}) \tag{9}$$

$$I_c^{out} = \sum_{j=1}^{m} \rho_i J_{\mathcal{N}}((j-1)\mu_c^{in}, 2(j-1)\mu_c^{in}) \tag{10}$$

$$I_v^{in} = 1 - I_c^{out} \tag{11}$$

$$\mu_v^{in} = J_{\mathcal{N}}^{-1}(I_v^{in}) \tag{12}$$

$$I_v^{out} = \sum_{i=1}^{n} \lambda_i J(f_L^{ch}(l) \oplus \mathcal{N}((i-1)\mu_v^{in}, 2(i-1)\mu_v^{in}) \tag{13}$$

Eq. (8) follows from the duality property between the extrinsic information from a variable node (I_v^{out}) and the apriori information to a neighboring check node (I_c^{in}). Eq. (9) follows from the monotonicity of J, hence the inverse exists, as well as the assumption of normally distributed LLR messages at each iteration. μ_c^{in} is the mean of the approximately normally distributed LLR messages at check nodes. Eq. (10) follows from the reciprocal approximation of the messages exchanged between variable and check nodes, resulting in additive messages at check nodes. Eq. (11) follows from the duality property between I_v^{in} and I_c^{out}. Eq. (12) follows the monotonicity of J and the assumption of normally distributed LLR messages. μ_v^{in} is the mean of the approximately normally distributed LLR messages at variable nodes. Eq. (13) follows from the additive property of messages at variable nodes due to the independence assumption of apriori LLR messages.

For the SLC channels with one or more reads, such as the single-read and five-read cases of (4) and (5), the convolution of the channel LLR p.d.f. ($f_L^{ch}(l)$) and the LLR p.d.f. of the incoming messages ($\mathcal{N}(\mu_v^{in}, 2\mu_v^{in})$) from the neighboring check nodes to variable node v in (13) results in a mean shift LLR distribution of the messages from the check nodes. For example, for the single-read case, $f_L^{ch}(l) \oplus \mathcal{N}(\mu_v^{in}, 2\mu_v^{in}) = p_1\mathcal{N}(\mu_v^{in} - L_{12}, 2\mu_v^{in}) + p_2\mathcal{N}(\mu_v^{in} - L_{21}, 2\mu_v^{in})$. However,

for BI-AWGN (6) the convolution results in the Gaussian distribution $\mathcal{N}(\mu_v^{in} + 2/\sigma^2, 2\mu_v^{in} + 4/\sigma^2)$.

B. RCA-EXIT for Non-binary LDPC Codes for MLC

We use $GF(4)$ NB-LDPC threshold analysis to find the best code for the MLC Flash models. The J functional for MLC under the assumption of uniform input is

$$J(f_{\boldsymbol{L}}(\boldsymbol{l})) = 2 - \int \log_2(1 + \sum_{i=1}^{3} e^{-l_i}) f_{\boldsymbol{L}}(\boldsymbol{l}) d\boldsymbol{l}\,. \quad (14)$$

The channel LLR p.d.f. is described as follows:

$$f_{\boldsymbol{L}}^{ch}(\boldsymbol{l}) = \frac{1}{4}(\mathcal{N}(\boldsymbol{\mu}_{00}^{ch}, \boldsymbol{\Sigma}_{00}^{ch}) + \mathcal{N}(\boldsymbol{\mu}_{01}^{ch}, \boldsymbol{\Sigma}_{01}^{ch}) \quad (15)$$
$$+\mathcal{N}(\boldsymbol{\mu}_{10}^{ch}, \boldsymbol{\Sigma}_{10}^{ch}) + \mathcal{N}(\boldsymbol{\mu}_{11}^{ch}, \boldsymbol{\Sigma}_{11}^{ch}))$$

$$\boldsymbol{\mu}_{00}^{ch} = \begin{bmatrix} 2/5\sigma^2 \\ 18/5\sigma^2 \\ 8/5\sigma^2 \end{bmatrix} \boldsymbol{\Sigma}_{00}^{ch} = \begin{pmatrix} 4/\sigma^2 & 12/\sigma^2 & 8/\sigma^2 \\ 12/\sigma^2 & 36/\sigma^2 & 24/\sigma^2 \\ 8/\sigma^2 & 24/\sigma^2 & 16/\sigma^2 \end{pmatrix} \quad (16)$$

$$\boldsymbol{\mu}_{01}^{ch} = \begin{bmatrix} 2/5\sigma^2 \\ 2/5\sigma^2 \\ 8/5\sigma^2 \end{bmatrix} \boldsymbol{\Sigma}_{01}^{ch} = \begin{pmatrix} 4/\sigma^2 & -4/\sigma^2 & -8/\sigma^2 \\ -4/\sigma^2 & 4/\sigma^2 & 8/\sigma^2 \\ -8/\sigma^2 & 8/\sigma^2 & 16/\sigma^2 \end{pmatrix}$$

$$\boldsymbol{\mu}_{10}^{ch} = \boldsymbol{\mu}_{00}^{ch} \quad \boldsymbol{\mu}_{11}^{ch} = \boldsymbol{\mu}_{01}^{ch} \quad \boldsymbol{\Sigma}_{10}^{ch} = \boldsymbol{\Sigma}_{00}^{ch} \quad \boldsymbol{\Sigma}_{11}^{ch} = \boldsymbol{\Sigma}_{01}^{ch}$$

The p.d.f. vectors in Eq. (16) are obtained by using the idea in [7] for cosets over $GF(q)$ and the assumption that the non-binary labels are selected at random and uniformly.

IV. LDPC Degree Distributions for Flash

The asymptotic threshold analysis of an LDPC code depends on the degree distribution of its variable and check nodes. $\lambda(x) = \sum_i \lambda_i x^{i-1}$ represents the variable-node degree distribution where λ_i is the fraction of the total number of edges connected to degree-i variable nodes. Similarly, $\rho(x) = \sum_j \rho_j x^{j-1}$ represents the check-node degree distribution where ρ_j is the fraction of all edges that are connected to degree-j check nodes. An LDPC code with variable-node and check-node degree distributions $\lambda(x)$ and $\rho(x)$ has a rate $r = 1 - \frac{\int_0^1 \rho(x)dx}{\int_0^1 \lambda(x)dx}$. For a particular code rate r, the code design optimization consists of finding the variable-node and check-node degree distributions that minimize the E_b/N_0 threshold.

The degree-distribution optimization algorithm for a rate-r code starts with finding the threshold for an initial (e.g. regular) degree distribution that results in a rate-r code. While the rate is kept constant, the parameters of $\lambda(x)$ and $\rho(x)$ are slightly changed and the new threshold is calculated. Once there is no more improvement in threshold by changing the degree distribution, the degree distribution with the lowest threshold is considered to be optimal. Table I shows the degree distributions we obtained using this approach.

A. MMI vs. RCA-EXIT for Word-Line-Voltage Optimization

For a fixed degree distribution, RCA-EXIT analysis can determine the word-line voltages that minimize the E_b/N_0 threshold for a multiple-read Flash channel. This is an alternative approach to MMI word-line voltages. Fig. 4 shows the plot

TABLE I: Optimized degree distributions for the Gaussian model of SLC Flash with 1,2,3, and 5 reads. MLC Flash with 3 reads, and both SLC and MLC with full soft information.

# Reads	Coefficients					
SLC 1 read	λ_2 0.07	λ_3 0.25	λ_7 0.11	λ_8 0.13	λ_{27} 0.44	ρ_{61} 1
SLC 2 reads	λ_2 0.1	λ_3 0.21	λ_7 0.25	λ_{25} 0.44		ρ_{57} 1
SLC 3 reads	λ_2 0.1	λ_3 0.21	λ_6 0.11	λ_7 0.12	λ_{26} 0.46	ρ_{56} 1
SLC 5 reads	λ_2 0.11	λ_3 0.21	λ_5 0.09	λ_8 0.14	λ_{25} 0.45	ρ_{56} 1
SLC soft	λ_2 0.11	λ_3 0.21	λ_6 0.21	λ_8 0.14	λ_{26} 0.47	ρ_{56} 1
MLC 3 reads	λ_2 0.16 ρ_5 0.45	λ_3 0.31 ρ_6 0.16	λ_4 0.1 ρ_8 0.1	λ_7 0.18 ρ_{22} 0.29	λ_8 0.1	λ_{11} 0.15
MLC soft	λ_2 0.15 ρ_5 0.44	λ_3 0.3 ρ_6 0.07	λ_4 0.1 ρ_8 0.1	λ_7 0.09 ρ_9 0.11	λ_9 0.22 ρ_{22} 0.27	λ_{11} 0.14

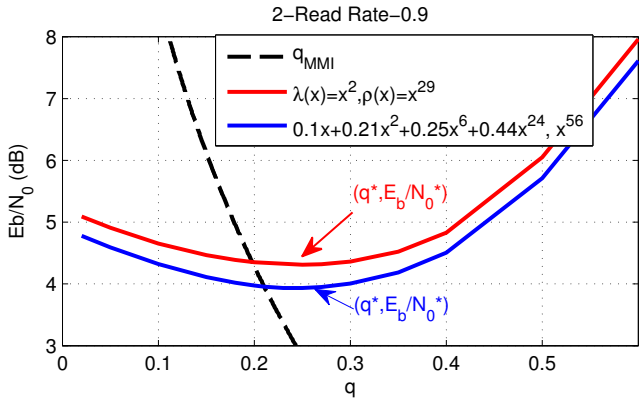

Fig. 4: Threshold E_b/N_0 vs q for a regular and the optimized rate-0.9 binary LDPC codes for 2-read SLC model

of the threshold E_b/N_0 for each word-line voltage $q_1 = q$ for the 2-read SLC model in Fig. 2b for two degree distributions. Also shown is the curve showing the MMI word-line voltage (q_{MMI}) for each E_b/N_0. For a fixed degree distribution and q, we find the minimum required E_b/N_0 such that the threshold calculation algorithm converges as was explained in III-A. The optimal value of q and its corresponding E_b/N_0 are shown in Fig. 4 by the pair $(q^*, E_b/N_0^*)$. E_b/N_0^* is the absolute minimum required E_b/N_0 for the RCA-EXIT to converge for at least one quantization threshold q. The threshold E_b/N_0 at q_{MMI} is less than 1% away from the optimal E_b/N_0^*.

It is difficult to simultaneously optimize both degree distribution and word-line voltage. Thus, even if the final word-line voltage will be selected to explicitly minimize the threshold, the MMI word-line voltage at a given E_b/N_0 is an excellent approximation when optimizing degree distributions. We use the q_{MMI} to optimize the degree distribution until the RCA-EXIT no longer converges at an E_b/N_0 and then adjust the q to obtain the final small improvement in E_b/N_0.

TABLE II: Thresholds for optimized degree distributions on SLC flash with 1, 2, 3, and 5 reads as well as soft information.

Target	1 read	2 reads	3 reads	5 reads	Soft
1 read	**4.752**	3.995	3.728	3.542	3.398
2 reads	4.922	**3.943**	3.658	3.470	3.324
3 reads	4.923	3.958	**3.640**	3.441	3.295
5 reads	4.926	3.973	3.649	**3.437**	3.288
Soft	4.926	3.982	3.662	3.443	**3.275**
Shannon-Limit	4.400	3.733	3.495	3.328	3.198

B. Optimized Degree Distributions for Each Precision Level

Table II shows the thresholds achieved by the various SLC degree distributions with various levels of precision. As expected, for a specified number of reads, the degree distributions optimized for that number of reads has the lowest threshold. However, this table quantifies the relatively small performance loss in threshold that is required for the same code to be used with multiple decoding attempts using increased precision.

The largest performance loss occurs when using a degree distribution optimized for 5 reads on the SLC 1-read channel. Here, the loss is 0.17 dB. Any degree distribution except the one designed for the single-read channel experiences the 0.17 dB loss. In contrast, if the degree distribution optimized for the single-read SLC channel is used on the 5-read channel, only 0.11 dB of loss is incurred. Thus there is reason to consider using the code designed for a single read. However, using the degree distribution optimized for 5 reads would minimize the probability of losing a page at the expense of additional decoder latency for those times when additional reads might have been avoided by using the degree distribution optimized for one read. If it becomes feasible to switch among codes as the Flash cells deteriorate over time, an optimized code for a higher number of reads may replace the previously optimized code with lower amount of precision.

Fig. 5 compares the threshold of regular LDPC degree distributions with the optimized irregular degree distribution $(\lambda(x), \rho(x))$ for each number of reads. There is roughly a gain of 0.4 dB for the optimized codes compared to the regular (x^2, x^{29}) and (x^3, x^{39}) codes.

For the same average energy, the optimized rate-0.45 non-binary codes used in the MLC model have a much lower threshold compared to the rate-0.9 codes for the SLC model. MLC Flash degree distributions provide thresholds more than 0.5 dB lower than degree distributions for rate-0.9 binary LDPC codes on SLC Flash with the same number of reads (i.e. three reads that would provide hard decisions for MLC and limited soft information for SLC). The MLC approach has a potential threshold reduction of about 1.5 dB over the SLC when both systems have access to full soft information. This is consistent with the MI analysis of [8] showing that a spectral efficiency of 0.9 is too high for binary PAM.

V. CONCLUSION

We have used RCA-EXIT analysis to optimize the degree distribution of binary and non-binary LDPC codes with E_b/N_0 thresholds of about 0.1 dB away from the Shannon-Limit for multiple-read models. While regular codes are usually used in Flash systems due to their simple decoder implementation, the optimized irregular binary codes have 0.4 dB lower threshold than the regular codes across different number of reads.

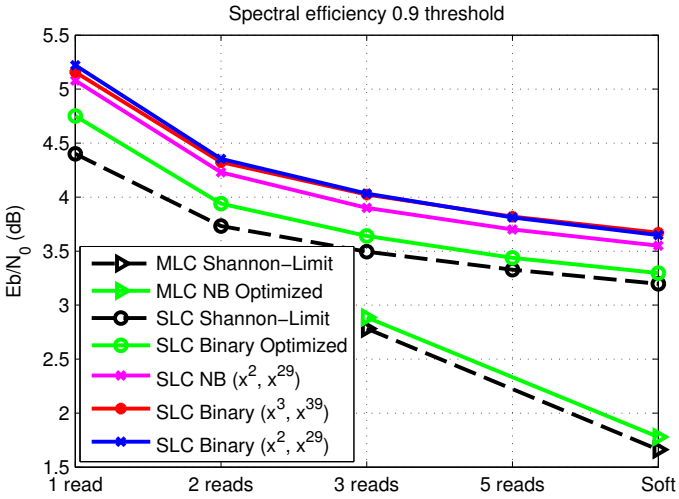

Fig. 5: Threshold E_b/N_0 vs number of reads for spectral efficiency of 0.9 bits per cell.

We have examined and validated the use of MMI word-line voltages in threshold-based degree distribution optimization. For the same spectral efficiency, low-rate non-binary LDPC codes for MLC Flash systems have lower thresholds compared to high-rate LDPC codes used in SLC systems. In addition, it is easier to design low-rate (e.g. rate-0.45) than high-rate (e.g. rate-0.9) LDPC codes.

REFERENCES

[1] T-.Y. Chen. A. R. Williamson and R. D. Wesel, "Increasing Flash memory lifetime by dynamic voltage allocation for constant mutual information," in *Proc. ITA Workshop*, San Diego, CA, 2014.

[2] J. Wang, T. Courtade, H. Shankar, and R. Wesel, "Soft information for LDPC decoding in Flash: Mutual-information optimized quantization," in *GLOBECOM*, Dec 2011, pp. 1–6.

[3] J. Wang, K. Vakilinia, T.-Y. Chen, T. Courtade, G. Dong, T. Zhang, H. Shankar, and R. Wesel, "Enhanced precision through multiple reads for LDPC decoding in Flash memories," *Selected Areas in Communications, IEEE Journal on*, vol. 32, no. 5, pp. 880–891, May 2014.

[4] S.-Y. Chung, T. Richardson, and R. Urbanke, "Analysis of sum-product decoding of low-density parity-check codes using a Gaussian approximation," *Information Theory, IEEE Transactions on*, vol. 47, no. 2, pp. 657–670, Feb 2001.

[5] S. Ten Brink, "Convergence behavior of iteratively decoded parallel concatenated codes," *Communications, IEEE Transactions on*, vol. 49, no. 10, pp. 1727–1737, Oct 2001.

[6] A. Roumy, S. Guemghar, G. Caire, and S. Verdu, "Design methods for irregular repeat-accumulate codes," *Information Theory, IEEE Transactions on*, vol. 50, no. 8, pp. 1711–1727, Aug 2004.

[7] A. Bennatan and D. Burshtein, "Design and analysis of nonbinary LDPC codes for arbitrary discrete-memoryless channels," *Information Theory, IEEE Transactions on*, vol. 52, no. 2, pp. 549–583, Feb 2006.

[8] G. Ungerboeck, "Channel coding with multilevel/phase signals," *Information Theory, IEEE Transactions on*, vol. 28, no. 1, pp. 55–67, Jan 1982.