# Optimal Detector for Multilevel NAND Flash Memory Channels with Intercell Interference

Meysam Asadi, *Student Member, IEEE,* Xiujie Huang, *Member, IEEE,*
Aleksandar Kavcic, *Senior Member, IEEE*, and Narayana (Prasad) Santhanam, *Member, IEEE*

***Abstract*—In this paper we derive the optimal detector for multilevel cell (MLC) flash memory channels with intercell interference (ICI). We start with the MLC channel model proposed by Dong *et al.* and just slightly alter the model to guarantee mathematical tractability of the optimal detectors (maximum likelihood and maximum a-posteriori sequence and symbol detectors). The optimal detector is obtained by computing branch metrics using Fourier transforms of analytically computable characteristic functions (corresponding to likelihood functions). We derive the detectors for both simple one-dimensional (1D) channel models and more realistic page-orientated two-dimensional (2D) channel models. Simulation results show that the hard-output bit error rate (BER) performance matches some previously known detectors, but that the soft-output detector outperforms previously known detectors by $0.35$ dB.**

***Index Terms*—Fast Fourier transform (FFT), intercell interference (ICI), maximum *a posteriori* (MAP) detector, multilevel cell (MLC), NAND flash memory.**

## I. Introduction

FLASH-based Solid State Drives (SSDs) are used in diverse consumer electronics applications. The NAND flash memory usage grows because of its low cost and high density resulting from the continuous improvements in scaling technology that shrinks the sizes of CMOS transistors and multilevel cell (MLC) technology that stores more than 1 bit per cell. However, as the storage density in flash memories continues to grow, various other factors such as energy consumption, intercell interference (ICI) and program/erase (PE) endurance, continue to degrade the overall system performance.

In this paper, we are mostly concerned with the detector design for a MLC NAND flash memory in the presence of ICI. First, the MLC technology narrows the width of threshold voltage for each level and reduces the margins between adjacent levels (voltages) in a cell, which results in degradation of reliability. Second, the scaling technology continues to increase the cell density, which dramatically enhances ICI and requires ever more complex detectors. These two issues complicate detection and encoding/decoding in flash memories.

To guarantee the reliability, on-chip error correcting techniques are widely employed in MLC NAND flash memory products [1]. However, as the market continues to demand higher densities and more levels per cell, simple error-correcting codes (ECCs) (for example, BCH codes with hard decoding) cease to be adequate. Hence, stronger ECCs with soft (decision) decoding, for example, low-density-parity-check (LDPC) codes, are desired in the next-generation flash memories. Consequently, this requires precise flash memory channel modeling and the knowledge of the exact statistics of the model. Existing literature, such as [2], [3], [4], derived so-called soft (decision) information by using uniform and/or non-uniform channel output quantization methods. To the best of our knowledge, there exists no open literature providing an exact soft-output detector without channel output quantization in MLC NAND flash memories.

To compensate for ICI, two similar methods, one called *post-compensation* in [5] and the other called *coupling canceller* in [6], were presented to subtract estimates of ICI from the noisy observation of the channel output (i.e., the sensed voltage of each cell). These two methods can be considered as (hard) detection schemes for the MLC flash memory.

The MLC flash memory system can be represented as a concatenation of 4 components: ECC encoder, channel, detector and ECC decoder (as depicted in Fig. 1). In this paper, we focus on the detector (the shaded block in Fig. 1). Motivated by [5], we focus on designing the detector for MLC flash memory in order to improve the hard decision bit-error-rate if possible. In addition, we would like to improve the soft decision quality of the detector if possible. Furthermore, designing the optimal soft and hard detector provides a benchmark which all other (sub-optimal) detector would be compared to. The optimal detector design also helps to derive closed form expression for the optimal decision making strategy in order to gain insight (such as sufficient statistics) and understand the interplay between channel parameters. Finally, we can use the attained insights to guide the derivation of novel low-complexity suboptimal detector.

We first provide channel models, including the one-dimensional (1D) model with *causal output memory* and the two-dimensional (2D) anti-causal model of the MLC flash memory, as shown in Sec. III. Second, in Sec. IV, we present a mathematically tractable Viterbi-like *maximum a*

Manuscript received May 15, 2013; revised October 1, 2013 and December 10, 2013. This work was supported by LSI Corporation and by NSF Grants CCF-1018984 and EECS-1029081. This work was also supported by NSF of China under Grants 61133014 and 61272413. Parts of this work were performed while A. Kavcic was a visiting scholar at the Institute of Network Coding at the Chinese University of Hong Kong.

X. Huang is with the Department of Computer Science, Jinan University, Guangzhou, 510632 China (e-mail: t_xiujie@jnu.edu.cn).

M. Asadi, A. Kavcic, and N. Santhanam are with the Department of Electrical Engineering, University of Hawaii at Manoa, Honolulu, 96822 HI USA (e-mail: masadi@hawaii.edu; kavcic@hawaii.edu, nsanthan@hawaii.edu).

Digital Object Identifier 10.1109/JSAC.2014.140503.

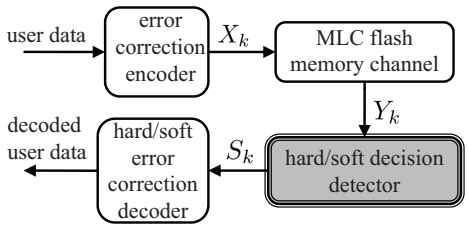


Fig. 1. A simple MLC flash memory system block diagram.

*posteriori* (MAP) sequence detector for the 1D causal model with output memory. The exact statistics of the channel model necessary for implementing the MAP detector can be obtained by using the fast Fourier transform (FFT). Third, we introduce a simplified Gaussian approximation (GA) sequence detector at the expense of reduced performance, which is shown in Sec. V. Both the MAP detector and the GA detector can be employed in the 2D anti-causal flash memory channel. Fourth, in Sec. VI, we extend the channel model and detector design to more general scenarios including those with signal-dependent noise, input intersymbol interference, and 2D Markov channel inputs. Fifth, in Sec. VII, we utilize simulation results to show that the MAP detector outperforms the existing detectors in the literature. Finally, we conclude this work in Sec. VIII.

## II. MLC NAND Flash Basics

### A. Structure

A NAND flash memory consists of lots of cells. Each cell is a transistor with an extra polysilicon strip between the control gate and the device channel, called the floating gate. By applying a voltage to the floating gate, a charge is maintained/stored in a cell. In order to store data in the cell of a MLC flash memory, a certain voltage (that falls into one of multiple required voltage ranges) is applied to the cell. All memory cells are hierarchically organized in arrays, blocks and page partitions, as shown in Fig. 2. The smallest unit that can be simultaneously accessed for programming (writing) or reading is a page; while the smallest unit that can be erased is a block.

An approach, known as incremental step pulse program (ISPP), also called program-and-verify technique with a staircase in [7, 8] and iterative programming in [9], is an iterative technique that can verify the amount of voltage carried at each cell after each programming step. The ISPP approach provides a series of verification pulses right after each program pulse. Consequently, the threshold voltage deviation of a programmed cell tends to behave like a uniform random variable [8]. As the programming of a cell is a one-way operation and because it is not possible to erase a specific cell separately from other cells in a block, it is necessary to erase a memory cell before being able to program it. The distribution of the threshold voltage of an erased memory cell tends to be Gaussian [10].

One architecture (called "even/odd bit-line structure") to program (write) the data is to separate all the cells into those at even bit-lines and those at odd bit-lines. During the process of programming, the cells at even bit-lines along a word-line are written at the same time instant, and then the cells at odd bit-lines along this word-line are written at the next time instant. The other architecture (called "all-bit-line structure") to program the data is to write all cells along a word-line simultaneously without distinguishing between even and odd cells. The even/odd bit-line structure has the advantage that circuitry could be shared and reused, while the all-bit-line structure has the advantage that the ICI is lower (as pictorially explained further down in Sec. III).

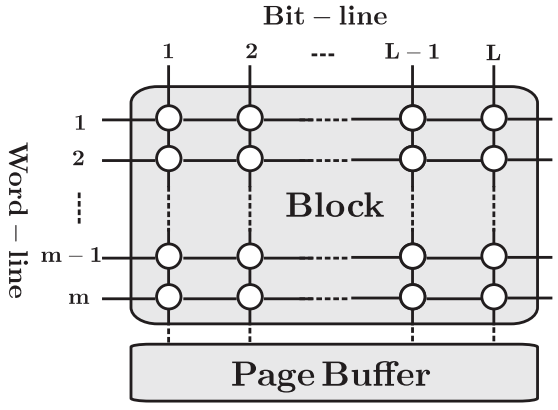


Fig. 2. NAND Flash memory structure [7].

### B. Degradation Sources

As we discussed in the Introduction, two major sources of performance degradation that affect the threshold voltage in each memory cell are PE cycling and ICI. The PE cycling process distorts the final threshold voltage of a transistor in two different ways. The first major distortion of PE cycling is due to the trapping and detrapping ability of the interface at the transistor gate, which leads to fluctuation of the final threshold voltage of the cell. This fluctuation is usually modeled by a Gaussian distribution with parameters dependent on the input voltage at the floating gate and the number of times that a cell has been programmed and erased [8]. The second distortion arises when electrons are trapped in the cell's interface area, which causes degradation of the threshold voltage. This effect is exacerbated as the device undergoes many PE cycles [8].

ICI is a degradation that grows with density. As cells are packed closer to each other, the influence of threshold voltages from neighboring cells increases. In other words, due to the parasitic capacitance coupling effects among the neighboring cells, the change in the threshold voltage on one cell during the programming (charging), affects the final voltages of all the other cells (especially those cells that were already programmed) [11]. This disturbance is usually modeled by a (truncated) Gaussian distribution whose parameters depend on the distance between cells [2, 5].

## III. NAND Flash Channel Model

In this section, we provide a stochastic model that includes all the important features of a multilevel NAND flash memory as illustrated in Fig. 3. We use a channel model that is mathematically almost identical to the channel model in [2], but we do make certain small changes (as described below)

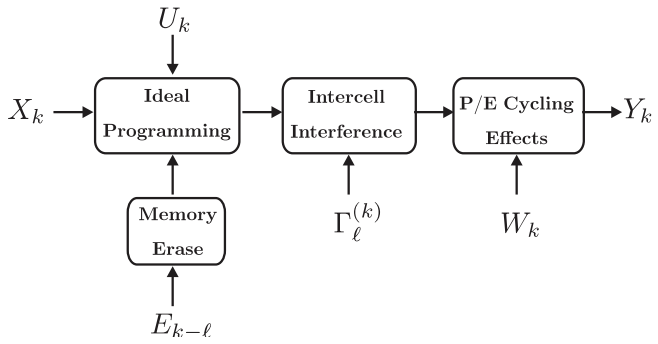


Fig. 3. NAND Flash Memory Channel Model.

to achieve complete mathematical tractability of the optimal detector design. These changes only marginally effect the model and the detector performance.

We start our exposition of the channel model by first presenting a simple *one-dimensional (1D) causal* channel model. Note that an actual flash memory channel is not one-dimensional, but rather two-dimensional (2D) because the channel is a page-oriented channel. Also, note that the actual flash memory channel is not causal, but rather anti-causal, because ICI is an anti-causal effect as only those cells that are programmed *after* the victim cell actually affect the victim cell. Nonetheless, the 1D causal channel model is very useful because it allows us to formulate the optimal detector in the universally accepted manner, namely for a 1D *causal* channel. Extrapolating the detector to cover 2D anti-causal channels is then fairly straightforward.

### A. 1D Causal Channel with Memory

Let $k \in \mathbb{Z}$ stand for discrete time. The channel input, denoted by $X_k$, is the intended stored voltage amount in the $k$-th cell. The channel output denoted by $Y_k$ is the channel output voltage corresponding to the input value $X_k$. According to MLC technology, we assume that the channel input random variable $X_k$ takes value from a finite alphabet $\mathcal{X} = \{v_0, v_1, \cdots, v_{m-1}\}$ with $|\mathcal{X}| = m < \infty$. We assume that the channel input and channel output have the relation[1]:

$$Y_k = X_k + \sum_{\ell=1}^{L} \Gamma_\ell^{(k)} \left(Y_{k-\ell} - E_{k-\ell}\right) + W_k + U_k \qquad (1)$$

where[2],

a) $E_k$ is the erase-state noise at the $k$-th cell, modeled as a Gaussian random variable with mean $\mu_e$ and variance $\sigma_e^2$, that is, $E_k \sim \mathcal{N}(\mu_e, \sigma_e^2)$.

b) $\Gamma_\ell^{(k)}$ is a fading-like coefficient that models *causal* ICI from the $(k-\ell)$-th cell towards the $k$-th cell (victim cell). We assume $\Gamma_\ell^{(k)}$ also to be a Gaussian random variable, $\Gamma_\ell^{(k)} \sim \mathcal{N}(\gamma_\ell, g_\ell)$.

c) $L$ is the output memory, which implies that the current channel output $Y_k$ is affected by its $L$ neighbors $Y_{k-1}, Y_{k-2}, \cdots Y_{k-L}$.

d) $U_k$ denotes the programming noise, resulting from using the ISPP method of programming the $k$-th cell of a certain word-line. This noise is modeled as a zero mean uniform random variable with width $\Delta$, that is, $U_k \sim \mathcal{U}\left(-\Delta/2\,,\, \Delta/2\right)$.

e) $W_k$ is observation noise due to the PE cycling, and is distributed as a zero mean Gaussian random variable with variance $\sigma_w^2$, that is, $W_k \sim \mathcal{N}(0, \sigma_w^2)$.

We assume that all random variables $\Gamma_\ell^{(k)}, E_{k-\ell}, W_k$ and $U_k$ are mutually independent for all $k$ and all $\ell$.

**Remark 1.** *We assume that the PE cycling/aging effect is incorporated into the model through the knowledge of $\sigma_w^2$. That is, $\sigma_w^2$ may depend on the device age.*

The major differences in the properties of random variables of this 1D channel model and the proposed channel model in [2, 5] are as follows:

i) We assume that $\Gamma_\ell^{(k)}$ are Gaussian, whereas Dong *et al.* [2, 5] assumed the corresponding variables to be distributed as truncated Gaussians. Note that these two distributions almost have the same behavior in the common support range, but the Gaussian distribution is easier to track analytically.

ii) We assume that the observation noise $U_k + W_k$ is a mixture of uniform and Gaussian noises. The pdf of this mixture is actually very similar (though not identical) to the pdf shown in [5, Fig. 4].

**Remark 2.** *The channel model (1) is actually a general 1D causal ICI channel which belongs to the class of channels with memory. Note that the memory of an ICI channel depends on the values of channel outputs $Y_k$. Although the output process $Y_k$ in this channel (for a stationary i.i.d. input process $X_k$) is a Markov process, this channel does not belong to well-know class of Markov finite-state channels [12, 13].*

Since the ICI channel has memory, it is obvious that a capacity-achieving process $X_k$ may also need to have memory. For this reason, we assume that the process $X_k$ is a Markov process[3] of order $M$. That is,

$$P_{X_k|X^{k-1}}(x_k|x^{k-1}) = P_{X_k|X_{k-M}^{k-1}}(x_k|x_{k-M}^{k-1}). \qquad (2)$$

The optimal detector is either a trellis-based Viterbi detector [15] (if we are interested in maximum likelihood (ML) or MAP sequence detection) or a trellis-based BCJR detector [16] (if we are interested in ML or MAP symbol detection). In both sequence or symbol detection cases, the trellis state at time $k$ is defined as $(X_{k-M+1}^k)$. Therefore, the number of trellis states in one trellis section is $|\mathcal{X}|^M = m^M$.

[1]The model in (1) is an autoregressive model, but only if the channel input $X_k$ is memoryless (uncorrelated). However, if the channel input has memory (for example, if the channel input is a Markov process of order $M > 0$), this is not an autoregressive model.

[2]The channel model in (1) does not take the same form as equation (1) in [5], however, they are actually identical. In Dong et al. [5], the authors chose to write the channel model in terms of the "voltage shift". In our equation (1), the "voltage shift" is equal to the difference $Y_{k-\ell} - E_{k-\ell}$, but instead of calling explicit attention to the "voltage shift", we call explicit attention to the channel outputs $Y_k, Y_{k-1}, \ldots, Y_{k-L}$. We think that, for the purpose of detector design, it is much better to explicitly call attention to the channel outputs because the detector specifically operates on the channel outputs.

[3]This paper is not concerned with finding the actual optimal Markov-memory input process $X_k$; The interested reader is referred to [14].

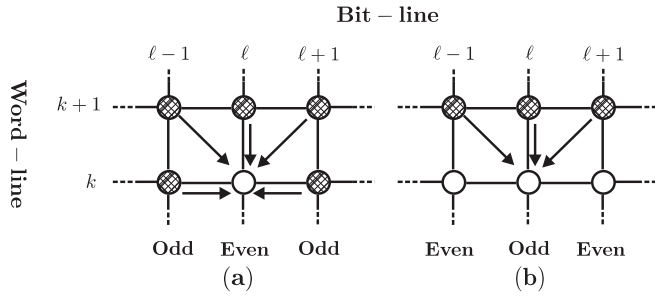


Fig. 4. Even/odd bit-line structure used in a NAND flash memory. (a) A cell on the even bit-line is affected by the ICI of 5 neighbors. (b) A cell on the odd bit-line is affected by 3 neighbors.

Although it seems reasonable to use Markov input process to mitigate the ICI influence in MLC flash memory channels, it is a very difficult task to find the best Markov process. In this difficult optimization problem, one should not only find the optimal number of voltage levels ($m$) and the best values of these levels $\{v_0, v_1, \cdots, v_{m-1}\}$, but also find the Markov distribution that maximizes the channel mutual information rate [17]. To the best of our knowledge, this problem is still open and there exists no literature to address the aforementioned problem. Note that finding the best Markov input process is beyond the scope of this paper. However, the results in this paper are still valid even for the case that the input distribution is Markov.

If $X_k$ is a memoryless (i.i.d) process (i.e., if $M = 0$), even though $Y_k$ has memory, a trellis-based detector is not needed (because $|\mathcal{X}|^M = 1$), and optimal detection (both soft and hard) can be executed on a symbol-by-symbol basis. That is, we can make the optimal decision on the random variable $X_k$ without postulating neighboring channel input realizations $x_{k-M}^{k-1}$ at all, and by considering only the received channel output realizations $y_{k-L}^{k}$.

### *B. Page-Oriented Memories (2D)*

In two-dimensional (2D) page-oriented memories with ICI, a single (victim) cell is only affected by a finite anticausal neighborhood of near-by cells (which are programmed *after* the victim cell). As discussed in section II-A, either the even/odd bit-line or the all-bit-line structure is used in most modern NAND flash memories. Since the amount of interference in the even/odd bit-line structure is higher than that in the all-bit-line structure, we consider only the even/odd bit-line structure using the full-sequence programming strategy. Note that our results can also be applied to the all-bit-line structure.

For the even/odd bit-line structure, cells in even bit-lines along a given word-line, referred to as even cells, are programmed first at one time instant, and then cells in odd bit-lines, referred to as odd cells, are programmed at a later time instant. Hence, the ICI neighborhoods are also dependent on whether an even cell or an odd cell is programmed in the programming cycle. Let $(k, \ell)$ denote the location of a memory cell, which means that the cell is located at the $k$-th word-line and the $\ell$-th bit-line. We denote by $\mathcal{O}_{(k,\ell)}$ the indices of the anticausal neighborhood for the *odd* cell and by $\mathcal{E}_{(k,\ell)}$ the indices of the anticausal ICI neighborhood for the *even* cell (illustrated in Fig. 4). That is,

$$\mathcal{O}_{(k,\ell)} \triangleq \{(k+1,\ell-1),(k+1,\ell),(k+1,\ell+1)\} \tag{3}$$

and

$$\mathcal{E}_{(k,\ell)} \triangleq \{(k,\ell-1),(k,\ell+1)\} \cup \mathcal{O}_{(k,\ell)}. \tag{4}$$

Therefore, we introduce an appropriate ICI channel model for the $(k, \ell)$-th victim cell as

$$Y_{(k,\ell)} = X_{(k,\ell)} + \sum_{\substack{(a,b)\in \\ \mathcal{S}_{(k,\ell)}}} \Gamma_{(a,b)}^{(k,\ell)} \big(Y_{(a,b)} - E_{(a,b)}\big) + W_{(k,\ell)} + U_{(k,\ell)} \tag{5}$$

where, the set $\mathcal{S}_{(k,\ell)}$ is either $\mathcal{O}_{(k,\ell)}$ or $\mathcal{E}_{(k,\ell)}$ depending on the location of the cell. If $X_{(k,\ell)}$ is a process with 2D memory, an optimal detector is not known (a 2D equivalent of a Viterbi detector is not available), and must be appropriately approximated using adequate (possibly interleaved) one dimensional (1D) detectors [18]. If $X_{(k,\ell)}$ is an i.i.d. process, optimal detection (both soft and hard) can be executed on a symbol-by-symbol basis.

## IV. Viterbi-like 1D Sequence Detection

We denote the sequence of random variables $(X_1, X_2, \cdots, X_n)$ of length $n$ by $X_1^n$. The realization sequence $(x_1, x_2, \cdots, x_n)$ is denoted by $x_1^n$. The set of all possible realizations of the random sequence $X_1^n$ is denoted by $\mathcal{X}^n$.

The MAP sequence detector of the state sequence $x_1^n$ is the sequence $\hat{x}_1^n$ that maximizes the joint conditional pdf, i.e.,

$$\hat{x}_1^n = \arg \max_{x_1^n \in \mathcal{X}^n} f(x_1^n, y_1^n | x_{1-M}^0, y_{1-L}^0), \tag{6}$$

where $M$ and $L$ are the order of the Markov input process and the output memory, respectively.

As shorthand, denote $f(\underline{x}, \underline{y}|\text{i.c.})$ as the conditional pdf of the right hand side of (6), where i.c. stands for the initial condition $(x_{1-M}^0, y_{1-L}^0)$. We start by factoring the pdf in (6) as

$$\begin{aligned} f(\underline{x}, \underline{y}|\text{i.c.}) &= f(x_1^n, y_1^n | x_{1-M}^0, y_{1-L}^0) \\ &= P(x_1^n | x_{1-M}^0, y_{1-L}^0) f(y_1^n | x_{1-M}^n, y_{1-L}^0) \\ &= \prod_{k=1}^{n} P(x_k | x_{k-M}^{k-1}) f(y_k | x_k, y_{k-L}^{k-1}). \end{aligned}$$

Consequently, the MAP detected sequence is equal to

$$\hat{x}_1^n = \arg \min_{x_1^n \in \mathcal{X}^n} \sum_{k=1}^{n} \underbrace{\big[-\ln\big(P\big(x_k | x_{k-M}^{k-1}\big) f\big(y_k | x_k, y_{k-L}^{k-1}\big)\big)\big]}_{\Lambda_{\text{MAP}}(x_{k-M}^k, y_{k-L}^k)}. \tag{7}$$

The term inside the summation in (7) is called a *branch metric* and is denoted by $\Lambda_{\text{MAP}}(x_{k-M}^k, y_{k-L}^k)$. It is clear that evaluating the branch metric $\Lambda_{\text{MAP}}(\cdot,\cdot)$ requires evaluating the conditional pdf $f(y_k|x_k, y_{k-L}^{k-1})$ or some function there of. Obviously, the branch metric depends on $L+1$ real valued variables $y_k, \cdots, y_{k-L}$. So, it is desired to extract a sufficient statistics from $y_{k-L}^k$ that will allow efficient computations of branch metrics. We next derive the desired sufficient statistics.

### A. Calculation of the Characteristic Function

Computing $f(y_k|x_k, y_{k-L}^{k-1})$ analytically is intractable. Instead, we calculate the conditional characteristic function of $Y_k$ under the assumptions that $X_k = x_k$ and $Y_{k-L}^{k-1} = y_{k-L}^{k-1}$ are given. The conditional pdf $f(y_k|x_{k-M}^k, y_{k-L}^{k-1})$ for each realization $y_k$ is then derived by taking the Fourier transform of the characteristic function.

We rewrite the channel model as

$$Y_k = \underbrace{X_k + W_k}_{R} + U_k + \sum_{\ell=1}^{L} \underbrace{\Gamma_\ell^{(k)}(Y_{k-\ell} - E_{k-\ell})}_{Z_\ell}. \tag{8}$$

Next, we compute the conditional characteristic function of $R$ and $Z_\ell$ under the assumptions that $X_k = x_k$ and $Y_{k-L}^{k-1} = y_{k-L}^{k-1}$ are given. Note that if $X_k = x_k$ is given, $R$ is Gaussian $\mathcal{N}(\mu_R, \sigma_R^2)$ where

$$\begin{aligned} \mu_R &= \mathbb{E}[R|X_k = x_k] = x_k \\ \sigma_R^2 &= \mathrm{Var}[R|X_k = x_k] = \sigma_w^2 \end{aligned}$$

Hence, the conditional characteristic function of $R$ is

$$\begin{aligned} G_{R|X_k}(t) &= \mathbb{E}\left[e^{iRt}|X_k = x_k\right] \\ &= \exp\left(-\frac{1}{2}\sigma_R^2 t^2 + i\mu_R t\right). \end{aligned} \tag{9}$$

Similarly, the conditional characteristic function of $Z_\ell$, when $Y_{k-\ell} = y_{k-\ell}$ and $X_k = x_k$ are given, is derived in the Appendix and denoted by $G_{Z_\ell|Y_{k-\ell}}(t)$. Finally, combining $G_{Z_\ell|Y_{k-\ell}}(t)$ and (9), and utilizing the conditional independence of $R$ and $Z_\ell$ (given $y_{k-\ell}^{k-1}$ and $x_k$), we get

$$\begin{aligned} &G_{Y_k|X_k, Y_{k-L}^{k-1}}(t) \\ &= G_{R|X_k}(t)\, G_{U_k}(t) \prod_{\ell=1}^{L} G_{Z_\ell|Y_{k-\ell}}(t) \\ &= \frac{\mathrm{sinc}\,(t\Delta/2)}{\sqrt{\prod_{\ell=1}^{L}(1 + g_\ell \sigma_e^2 t^2)}} \exp\left(-\frac{1}{2}\sigma_R^2 t^2 + i\mu_R t + \Phi(t)\right) \end{aligned} \tag{10}$$

where

$$\Phi(t) = \sum_{\ell=1}^{L} \frac{-t^2\left[(y_{k-\ell} - \mu_e)^2 g_\ell + \gamma_\ell^2\, \sigma_e^2\right] + 2it\,(y_{k-\ell} - \mu_e)\gamma_\ell}{2\left(1 + g_\ell \sigma_e^2 t^2\right)}. \tag{11}$$

### B. FFT Implementation

Since the pdf is the Fourier transform of the characteristic function, the conditional probability $f(y_k|x_k, y_{k-L}^{k-1})$ can be obtained as

$$f(y_k|x_k, y_{k-L}^{k-1}) = \int_{-\infty}^{\infty} G_{Y_k|X_k, Y_{k-L}^{k-1}}(t) e^{-iy_k t} dt. \tag{12}$$

Hence, it is possible to numerically compute the branch metric $\Lambda_{\mathrm{MAP}}(x_{k-M}^k, y_{k-L}^k)$ in (7) for each branch in the Viterbi trellis using the fast Fourier transform (FFT). For each trellis section, we only need to compute one FFT. In other words, the FFT is the same for all branches of the trellis section, but the actual branch metric values are obtained by sampling the FFT at different points as illustrated in Fig. 6 below.

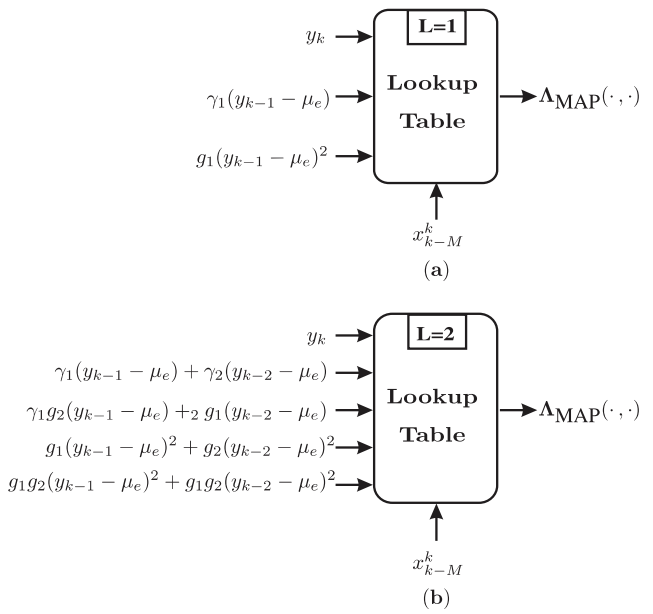


Fig. 5. Branch metric $\Lambda_{\mathrm{MAP}}$ for cases (a) $L = 1$ and (b) $L = 2$.

### C. Sufficient Statistics

A look at (10) and (11) reveals that the channel outputs $y_{k-L}^k$ need to be processed (in some way) in order to formulate the branch metrics. The processing complexity depends on the order $L$.

**Example 1.** *If $L = 1$, then (10) and (11) reveal that a set of sufficient statistics necessary for the computation of branch metrics is*

*a)* $y_k$
*b)* $\gamma_1(y_{k-1} - \mu_e)$
*c)* $g_1(y_{k-1} - \mu_e)^2$

*A way to obtain $\Lambda_{\mathrm{MAP}}$ in this case could be using the lookup table in Fig. 5 (a).*

**Example 2.** *If $L = 2$, then (11) reveals that the sufficient statistics are obtained by finite impulse response (FIR) filters. The sufficient statistics are:*

*a)* $y_k$, *and the following FIR filter outputs*
*b)* $\gamma_1(y_{k-1} - \mu_e) + \gamma_2(y_{k-2} - \mu_e)$
*c)* $\gamma_1 g_2(y_{k-1} - \mu_e) + \gamma_2 g_1(y_{k-2} - \mu_e)$
*d)* $g_1(y_{k-1} - \mu_e)^2 + g_2(y_{k-2} - \mu_e)^2$
*e)* $g_1 g_2(y_{k-1} - \mu_e)^2 + g_1 g_2(y_{k-2} - \mu_e)^2$

*Consequently, the branch metrics $\Lambda_{\mathrm{MAP}}$ can be computed using a lookup table in Fig. 5 (b).*

The complexity of implementing the lookup table grows linearly with $L$ because we need $2L + 1$ sufficient statistics. Quantizing each sufficient statistic to, say, a 7 bit precision requires $7(2L + 1)$ binary inputs to the lookup table. An alternative is to use FFT to compute the pdf $f(y_k|x_k, y_{k-L}^{k-1})$ from the characteristic function $G_{Y_k|X_k, Y_{k-L}^{k-1}}(t)$. Thereby, we can use the sufficient statistics to compute an equivalent form of the characteristic function (10) with

$$\Phi(t) = \frac{\sum_{\ell=1}^{2L} C_\ell(y_{k-L}^{k-1}) t^\ell + q(t)}{2\prod_{\ell=1}^{L}(1 + g_\ell \sigma_e^2 t^2)}$$

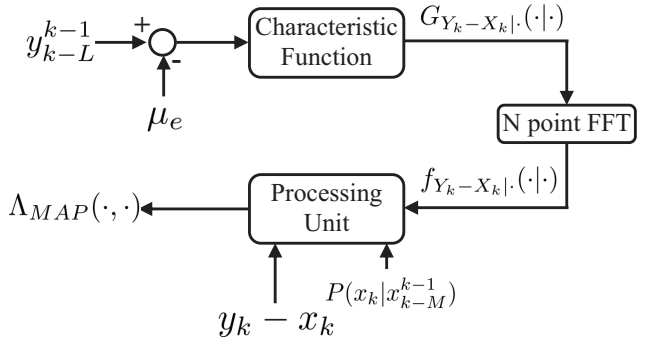


Fig. 6. Branch metric computation using the FFT.

where

$$C_\ell(y_{k-L}^{k-1}) = \begin{cases} \sum_{j=1}^{L} \alpha_j^{(\ell)} \cdot (y_{k-j} - \mu_e), & \text{if } \ell \text{ is odd} \\ \sum_{j=1}^{L} \beta_j^{(\ell)} \cdot (y_{k-j} - \mu_e)^2, & \text{if } \ell \text{ is even.} \end{cases} \tag{13}$$

and $q(t)$ is a polynomial (whose coefficients are independent of $y_{k-j}^{k-1}$). It is clear from (13) that the sufficient statistics for this computation are outputs of two types of FIR filters where:

a) inputs are signals $(y_k - \mu_e)$, if $\ell$ is odd.
b) inputs are signals $(y_k - \mu_e)^2$, if $\ell$ is even.

This approach is illustrated in Fig. 6.

Lookup tables may be complicated to implement because of the need to quantize all the sufficient statistics. However, the purpose of this section is not to suggest that a lookup table is always a practical approach. Rather, it is to reveal what the sufficient statistics are. Knowing the sufficient statistics is important because it gives us an analytic insight into what to compute at the receiver end. For example, when $L = 2$, Fig. 5 (b) shows that there are 5 sufficient statistics (the 5 inputs to the lookup table from the left). This can actually guide the development of suboptimal detectors. For example, it is easy to verify that if we ignore the 3rd, 4th and 5th inputs to the lookup table, we get the suboptimal post-compensation detector of Dong *et al.* [5]. In the same spirit, we will reveal in Section V how to derive another suboptimal detector (based on a Gaussian approximation) that ignores the 3rd and 5th inputs in Fig. 5 (b), has a closed-form expression for the branch metric and is particularly easy to implement.

## V. Gaussian Approximation (GA) Detector

At the expense of reduced performance, as an alternative to the optimal procedure given in Sec. IV, we give a simplified procedure for computing an approximation of $\Lambda_{\text{MAP}}(\cdot,\cdot)$ based on a Gaussian approximation. We rewrite the channel model as

$$\begin{aligned} Y_k &= X_k + \sum_{\ell=1}^{L} \Gamma_\ell^{(k)} (Y_{k-\ell} - E_{k-\ell}) + W_k + U_k \\ &= V_k + U_k. \end{aligned} \tag{14}$$

According to (1), $V_k$ is obtained as the summation of several random variables. Assume that we can approximate $f(v_k|x_k, y_{k-L}^{k-1})$ by a Gaussian pdf as follows:

$$f(v_k|x_k, y_{k-L}^{k-1}) \sim \mathcal{N}(\mu_G(k), \sigma_G^2(k)), \tag{15}$$

where

$$\begin{aligned} \mu_G(k) &= \mathbb{E}\big[V_k|Y_{k-\ell}^{k-1} = y_{k-\ell}^{k-1}, X_k = x_k\big] \\ &= x_k + \sum_{\ell=1}^{L} \gamma_\ell (y_{k-\ell} - \mu_e) \end{aligned} \tag{16}$$

$$\begin{aligned} \sigma_G^2(k) &= \mathrm{Var}\big[V_k|Y_{k-\ell}^{k-1} = y_{k-\ell}^{k-1}, X_k = x_k\big] \\ &= \sum_{\ell=1}^{L} \big(g_\ell\big(\sigma_e^2 + (y_{k-\ell} - \mu_e)^2\big) + \sigma_e^2\gamma_\ell^2\big) + \sigma_w^2. \end{aligned} \tag{17}$$

The new approximate conditional distribution $f^{(G)}(y_k|x_k, y_{k-L}^{k-1})$, is obtained by convolving the Gaussian distribution $\mathcal{N}(\mu_G(k), \sigma_G^2(k))$ and the uniform distribution $\mathcal{U}(-\Delta/2\,,\,\Delta/2)$. That is,

$$\begin{aligned} &f^{(G)}(y_k|x_k, y_{k-L}^{k-1}) \\ &= \int_{y_k - \frac{\Delta}{2}}^{y_k + \frac{\Delta}{2}} \frac{1}{\sqrt{2\pi}\sigma_G\Delta}\, e^{-\frac{(v_k - \mu_G)^2}{2\sigma_G^2}}\, dv_k \\ &= \frac{1}{\Delta}\left[Q\left(\frac{y_k - \mu_G - \frac{\Delta}{2}}{\sigma_G}\right) - Q\left(\frac{y_k - \mu_G + \frac{\Delta}{2}}{\sigma_G}\right)\right] \end{aligned}$$

where the standard $Q$-function is defined as $Q(\zeta) = \frac{1}{\sqrt{2\pi}} \int_\zeta^\infty \exp(-\frac{\eta^2}{2})\, d\eta$.

So, clearly by examining (14)-(17), we conclude that to compute the branch metrics $\Lambda^{(G)}(\cdot,\cdot)$, we need the following subset of the sufficient statistics

$$\begin{aligned} &y_k \\ \theta_k &= \sum_{\ell=1}^{L} \gamma_\ell\, (y_{k-\ell} - \mu_e) \\ \nu_k &= \sum_{\ell=1}^{L} g_\ell\, (y_{k-\ell} - \mu_e)^2, \end{aligned}$$

which actually coincide with the three sufficient statistics a), b) and d) in Example 2 (Fig. 5 (b)). Hence, the computation of $\Lambda^{(G)}(x_{k-M}^k, y_{k-L}^k)$ is equivalent to computing $\Lambda^{(G)}(x_{k-M}^k, y_k, \theta_k, \nu_k)$. Note, again, that $\theta_k$ and $\nu_k$ are obtained by FIR-filtering $(y_{k-\ell} - \mu_e)$ and $(y_{k-\ell} - \mu_e)^2$, respectively. Thereby, the entire set of sufficient statistics can be replaced by a new vector $[y_k, \theta_k, \nu_k]$ of only three components (even if $L > 1$). Furthermore, the actual computation of $\Lambda^{(G)}(x_{k-M}^k, y_k, \theta_k, \nu_k)$ does not require generating lookup tables or FFTs, but can be implemented using sample DSP components such as multipliers and adders. Fig. 7 illustrates the branch metric computation module of the suboptimal GA detector.

It is interesting to note that some prior-art detectors, such as the hard-decision detectors in [5, 19], and the soft-decision detectors in [2, 20] can be obtained by further approximations of the Gaussian-approximation detector given in this section. In particular, the hard decision detector (post-compensation detector) in [5] can be obtained as a symbol-by-symbol detector (when $X_k$ is i.i.d.) by computing the decision variables $y_k - \mu_G(k)$ and heuristically determining the decision thresholds to achieve minimum probability of symbol error. The detector in [19] is very similar to [5] in the way that

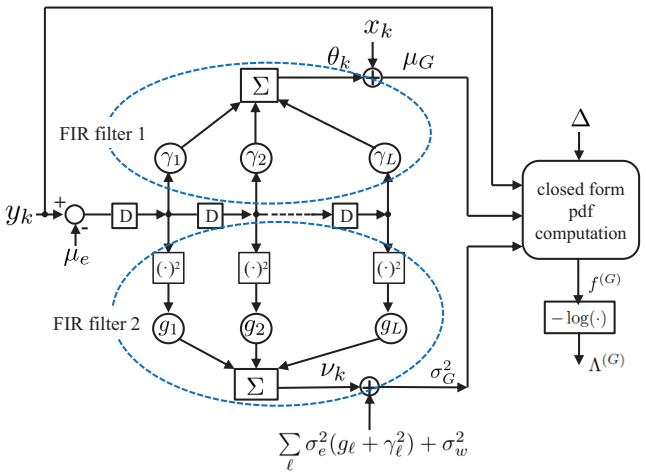


Fig. 7. The branch metric computation module of the GA detector using FIR filters.

decision variables are determined, but in [19] a heuristically-derived Viterbi detector (obtained by exhaustive target training) is used to refine the detection process. Similarly, the soft-output detector in [2, 20] can be obtained if instead of $\sigma_G^2(k)$ in (17), we use $\sigma_e^2\gamma_1^2$. Consequently, the detectors in [2, 20] are suboptimal versions of the detector proposed here (particularly, if used as precursors to a soft-in-soft-out decoder for LDPC codes) as shown explicitly in Sec. VII.

Finally, we note that in [5], a separate predistortion detector and a separate post-compensation detector were proposed, showing that each has some advantages. The detector in this paper combines the strengths of both strategies *simultaneously*. This can be achieved by shaping the channel input process $X_k$ into a Markov process (2) such as in [21, 22] (which is akin to predistortion in [5]) and subsequently using a Viterbi/BCJR trellis detector as in Secs. IV and V to detect the Markov input process $X_k$ (akin to the post-compensation detector in [5]).

## VI. Extensions

In this section, we briefly explain how to extend the channel model and detector design when the channel suffers from additional degradations.

### A. Signal-Dependent Noise

If the noise in the channel is signal-dependent, the channel model and the detector must be appropriately altered. Several studies [8, 17] have shown that different levels $v_0, v_1, \cdots, v_{m-1}$ give rise to different statistics of the channel noise $W_k$ and $U_k$ (see model in (1)). For example, it is possible that the random variables $W_k$ and $U_k$ depend on the realization of the channel input random variable $X_k = v_j$. In that case, we model $W_k$ and $U_k$ to be signal-dependent,

$$\begin{aligned} f_{W_k|X_k}(\cdot|X_k=v_j) &\sim \mathcal{N}(0,\sigma_w^2(j)) \\ f_{U_k|X_k}(\cdot|X_k=v_j) &\sim \mathcal{U}\left(\frac{-\Delta(j)}{2},\frac{\Delta(j)}{2}\right) \end{aligned}$$

In other words, the parameters $\sigma_w^2$ and $\Delta$ are appropriately substituted by $\sigma_w^2(j)$ and $\Delta(j)$, depending on the (postulated) realization of the random variable $X_k = v_j$.

Similarly, we may also extend the model of the random fading-like coefficient $\Gamma_\ell^{(k)}$ to be signal-dependent. This means that instead of assuming $\mathbb{E}\left[\Gamma_\ell^{(k)}\right] = \gamma_\ell$ and $\mathrm{Var}\left[\Gamma_\ell^{(k)}\right] = g_\ell$, we would assume a signal-dependent model $\mathbb{E}\left[\Gamma_\ell^{(k)}|X_{k-\ell}=v_j\right] = \gamma_\ell(j)$ and $\mathrm{Var}\left(\Gamma_\ell^{(k)}|X_{k-\ell}=v_j\right) = g_\ell(j)$. These changes would appropriately alter the Viterbi/BCJR detectors.

### B. Input Intersymbol Interference

The channel model in (1) assumed only output ICI, but no input intersymbol interference (ISI). We can alter the model in (1) to account for input ISI as follows:

$$Y_k = \sum_{m=0}^{M} A_m^{(k)} X_{k-m} + \sum_{\ell=1}^{L} \Gamma_\ell^{(k)}(Y_{k-\ell}-E_{k-\ell}) + W_k + U_k \quad (18)$$

where $A_m^{(k)}$ are either constant coefficients or random variables, say $A_m^{(k)} \sim \mathcal{N}(a(m),\sigma_a^2(m))$. In either case, the optimal detector design is still a Viterbi-like or a BCJR-like detector whose branch metrics $\Lambda_{\mathrm{MAP}}(x_{k-M}^k, y_{k-L}^k)$ can be determined using the FFT of the appropriate characteristic function, or an appropriate Gaussian approximation. The model in (18) can also be extended to be signal-dependent and/or 2D. We omit the details.

### C. 2D Channels

When the 2D channel model in (5) is appropriate, and the channel input is i.i.d, the optimal detector is a simple symbol-by-symbol detector. However, since the channel does have 2D-memory, it is reasonable to expect that the information-theoretically optimal channel input process $X_{(k,\ell)}$ will not be i.i.d. In this case we have two complications for which exact solutions are not known, and we likely need to resort to ad-hoc and/or heuristic approaches:

a) Optimizing the input distribution (even under the 2D Markov input assumption) to maximize the information rate of a 2D channel with memory is not known. To date, only a limited number of computational methods to evaluate information rates of 2D channels with memory are known [23–26], but to the best of our knowledge, no 2D information-rate optimization techniques are available. One approach may be to heuristically adapt 1D techniques to optimize lower-bounds on 2D information rates as in [27], but this is certainly subject the further research.

b) Even if an appropriate 2D Markov process $X_{(k,\ell)}$ could be constructed to guarantee a nearly optimal information rate, the optimal detector for a 2D channel is not available (because there exists no equivalents of the Viterbi/BCJR detectors in 2D). A plausible solution is to apply 1D methods in some heuristic fashion (such as, for example, interleaving vertical and horizontal detectors [18, 28–30], or combining 1D horizontal detectors with 1D vertical decision feedback [31–33]), or to design entirely new 2D detectors [34–36] (which is, of course, subject to further research).

TABLE I
PARAMETERS OF THE 4-LEVEL FLASH MEMORY

| $i$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $i$th level $v_i$ | 1.1 | 2.7 | 3.3 | 3.9 |
| $\Delta(i)$ | 0 | 0.3 | 0.3 | 0.3 |
| $\sigma_w(i)$ | $0.35\sigma$ | $0.03\sigma$ | $0.03\sigma$ | $0.03\sigma$ |

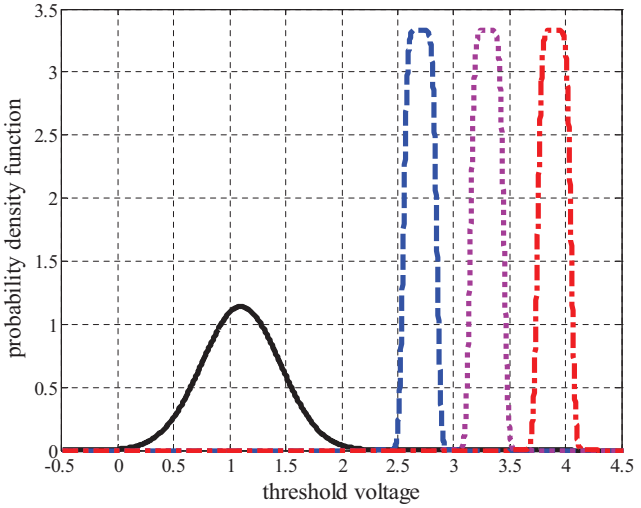


Fig. 8. The pdf of each level's voltage for the 4-level flash memory without ICI.

From points (a) and (b) above, it is clear that to achieve an information-theoretically optimal transmission/reception strategy further research on 2D capacity computing techniques and 2D detection/interleaving techniques is needed.

## VII. Simulation Results and Discussion

In this section, we give simulation results to show the performances of the detectors presented in this paper when using an even/odd bit-line structure. We use a 4-level flash memory channel, where the channel input $X_k$ is an i.i.d. process with parameters $\Pr(X_k = v_j) = 0.25$ for any of the 4 levels $v_0$, $v_1$, $v_2$ or $v_3$. The parameters of the 4-level flash memory (2D channel) with signal-dependent noise are given in Table I. With the parameters as in Table I, and using $\sigma = 1$, Fig. 8 depicts the pdf of each level's voltage when no ICI occurs.

We next assume that the random coupling ratios $\Gamma^{(k,\ell)}_{(a,b)}$ (see (5)) have the following Gaussian distributions

$$\begin{aligned} &\Gamma^{(k,\ell)}_{(k,\ell-1)} \sim \mathcal{N}(\gamma_h, g_h), \quad &&\Gamma^{(k,\ell)}_{(k,\ell+1)} \sim \mathcal{N}(\gamma_h, g_h), \\ &\Gamma^{(k,\ell)}_{(k+1,\ell-1)} \sim \mathcal{N}(\gamma_d, g_d), \quad &&\Gamma^{(k,\ell)}_{(k+1,\ell+1)} \sim \mathcal{N}(\gamma_d, g_d) \\ &\Gamma^{(k,\ell)}_{(k+1,\ell)} \sim \mathcal{N}(\gamma_v, g_v), \end{aligned} \tag{19}$$

where the subscripts $h$, $v$ and $d$ mean horizontal, vertical and diagonal interference, respectively. We also assume that[4] $\gamma_h : \gamma_v : \gamma_d = 0.1 : 0.08 : 0.006$ and $g_i = 0.09\gamma_i^2$ for $i \in \{h, v, d\}$ as introduced in [5] and the references therein. Let $s$ be the *intercell coupling strength factor*. Then $\gamma_h = 0.1s$, $\gamma_v = 0.08s$ and $\gamma_d = 0.006s$.

[4]These notations denote the relative magnitudes of horizontal, vertical and diagonal capacitance couplings.

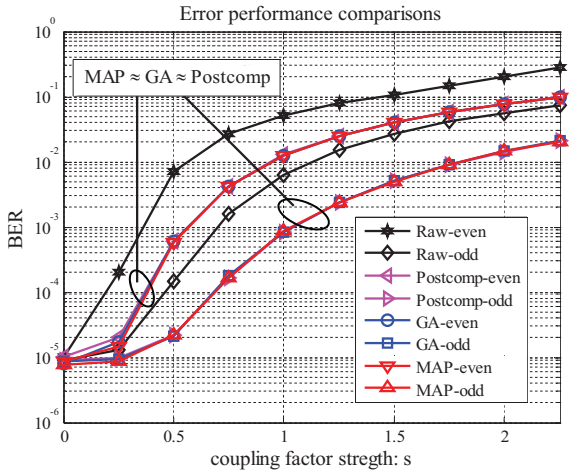


Fig. 9. BER comparisons for different detectors when the coupling factor strength is varying and $\sigma = 1$.

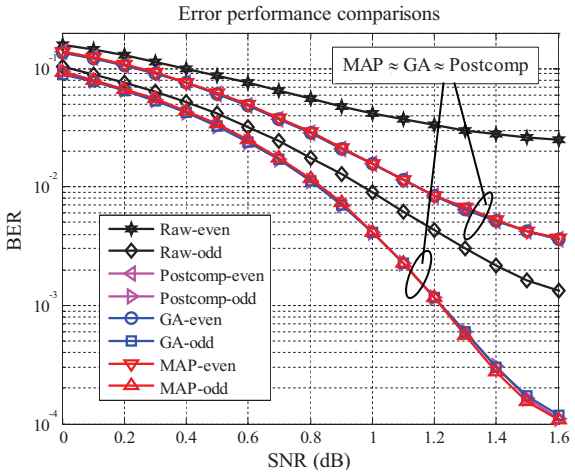


Fig. 10. BER comparisons for different detectors when the SNR is varying and the coupling factor strength is fixed at $s = 0.75$.

In the first simulation scenario, we fix $\sigma = 1$ (see Table I) and we let the coupling strength factor $s$ vary from 0 to 2. The bit-error-rate (BER) performances of the MAP detector and the GA detector are shown in Fig. 9. In Fig. 9, we also show the BER performances of the post-compensation detector [5] and the raw detector [5].

In the second simulation scenario, we fix $s = 0.75$, and vary the parameter $\sigma$ (see Table I). By varying $\sigma$, we effectively vary the signal-to-noise ratio (SNR), defined as

$$\text{SNR} \triangleq \frac{1}{\sum_i \Pr(X_k = v_i)\sigma_w^2(i)}. \tag{20}$$

The BER curves for varying SNRs are shown in Fig. 10, depicting the performances of the MAP detector, the GA detector, the post-compensation detector [5] and the raw detector [5].

Figs. 9 and 10 reveal that if the BER is the figure of merit, neither the MAP detector nor the GA detector outperforms the post-compensation detector (originally disclosed in [5]). Hence, to get a better sense of the quality of each detector, we must compare the qualities of their *soft* outputs. Here, we measure the quality of a detector's soft output as follows. Let $X_k$ be the i.i.d. equiprobable channel input, meaning that

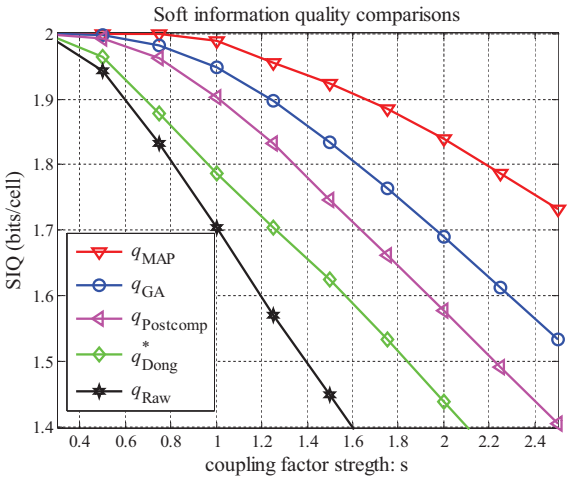


Fig. 11. SIQ comparisons for different detectors when the coupling factor strength is varying and $\sigma = 1$.

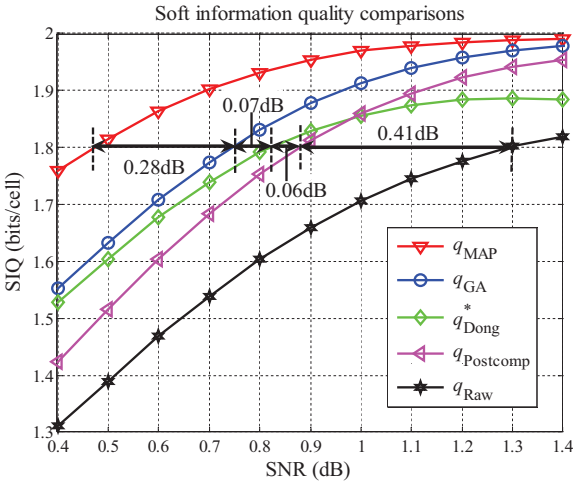


Fig. 12. SIQ comparisons for different detectors when the SNR is varying and the coupling factor strength is fixed at $s = 0.75$.

$P(X_k = v_j) = \frac{1}{m}$. In the case of a soft decision detector, the detector output $S_k$ is a vector defined as[5]

$$S_k = \begin{bmatrix} P(X_k = v_0 | Y_1^n = y_1^n) \\ P(X_k = v_1 | Y_1^n = y_1^n) \\ \vdots \\ P(X_k = v_{m-1} | Y_1^n = y_1^n) \end{bmatrix}$$

and in the case of a hard detector, the detector output $S_k$ is a scalar estimate of the channel input

$$S_k = \hat{X}_k \in \{v_0, v_1, \cdots, v_{m-1}\}.$$

We define the *soft information quality* (SIQ) of a detector as

$$q = \frac{1}{2}\Big[ I(X_k; S_k)|_{\text{even } k} + I(X_k; S_k)|_{\text{odd } k} \Big]. \qquad (21)$$

As explained in [37], SIQ is the capacity of random linear block codes. Therefore, this quantity is proved to be the highest information rate achievable by a random low-density parity-check (LDPC) error correction code. Furthermore, the SIQ allows us to compare performances of codes without going through the complicated task of simulating the actual codes. For example, if SIQ of detector $A$ is $0.5$ dB better than SIQ of detector $B$, then a random LDPC code using outputs from detector $A$ will outperform the same random LDPC code using outputs from detector $B$ by $0.5$ dB. In other words, if we use detector $A$, we can afford to use a $0.5$ dB weaker code and achieve the same overall system performance.

The mutual information terms in (21) can be readily computed numerically using Monte-Carlo simulations for any detector (also for a hard-decision detector). For the special case of a MAP detector, the soft-information quality $q_{\text{MAP}}$ has an alternative interpretation, i.e., $q_{\text{MAP}}$ is equal to the so-called *BCJR-once bound* (see [37] for details). Fig. 11 shows the soft information qualities of the MAP detector and the GA detector when the coupling strength factor $s$ varies for fixed SNR, while Fig. 12 shows the soft information quality curves when the SNR varies for fixed $s = 0.75$. Also shown in Figs. 11 and 12 are soft information qualities of the post-compensation detector [5] and the raw detector [5]. Finally, the figures also show an upper bound on the soft information quality of the soft-output detector presented in [2], denoted by $q^{*}_{\text{Dong}}$. At SIQ $= 1.8$ bits per cell (which corresponds to a code rate of $0.9$ user bits per channel bit), the MAP detector outperforms known detectors by $0.35$ dB, as shown in Fig. 12.

[5]Obviously, in a multilevel flash memory channel, the variable $S_k$ is a collection of $m$ likelihood values - one likelihood value for each level.

TABLE II
COMPLEXITY COMPARISON FOR DIFFERENT DETECTORS

| Detector | Read Process | Metric Computation |
|---|---|---|
| No ICI (threshold detector) | $O(1)$ | $O(1)$ |
| Post-compensation [5] | $O(L)$ | $O(L)$ |
| Gaussian Approximation | $O(L)$ | $O(L)$ |
| MAP(FFT-based) | $O(L)$ | $O(L)+$ $O(NlogN)$ |

As we show in Fig. 12, the performance of MAP detector is significantly better than current detectors. Table II compares the computational penalty of each detector. In Table II, the "read process" penalty stands for the required number of reads per written symbol, and the "metric computation" penalty is the computational complexity of computing the branch metrics per written symbol. Note that the variable $N$ in the MAP detector is the number of quantization points in computing the FFT (i.e, $N$ is the support length of the FFT)[6].

## VIII. Conclusion

We derived the optimal detector structure for multilevel cell (MLC) NAND flash memory channels. The optimal detector is attainable using FFTs of analytically computable characteristic functions. Alternatively, at a small performance loss, a Gaussian-approximation detector is attainable using two FIR filters, i) the first operating on channel outputs, and ii) the second operating on the squares of the channel outputs. We derived the optimal detectors here for both 1D and 2D page-oriented channels and demonstrated their superior performances (particularly if executed as soft-output detectors) through simulations.

[6]In all our simulations, we used $N = 512$.

## APPENDIX A
## CHARACTERISTIC FUNCTION $G_{Z_\ell|Y_{k-\ell}}(t)$

Under the assumption that $Y_{k-\ell} = y_{k-\ell}$ and $X_k = x_k$ are given, $Z_\ell$ is the product of two Gaussian random variables, which can be rewritten as

$$Z_\ell \quad = \quad \Gamma\,\Omega = \Gamma_\ell^{(k)}\,(Y_{k-\ell} - E_{k-\ell}) \tag{22}$$

where $\Gamma \sim \mathcal{N}(\gamma_\ell, g_\ell)$ and $\Omega \sim \mathcal{N}(y_{k-\ell} - \mu_e, \sigma_e^2)$. Then, the characteristic function for the product of two normal random variables $\Gamma\,\Omega$, denoted by $G_{Z_\ell|Y_{k-\ell}}(t)$, is computed as

$$\begin{aligned}
&G_{Z_\ell|Y_{k-\ell}}(t)\\
&= \mathbb{E}[e^{i\Gamma\,\Omega t}|Y_{k-\ell} = y_{k-\ell}]\\
&= \mathbb{E}\left[\mathbb{E}\left[e^{i\Gamma\,\Omega t}|\Omega, Y_{k-\ell} = y_{k-\ell}\right]\right]\\
&= \mathbb{E}\left[e^{i\gamma_\ell\Omega t - \frac{1}{2}g_\ell\Omega^2 t^2}|Y_{k-\ell} = y_{k-\ell}\right]\\
&= \frac{1}{\sqrt{2\pi\sigma_e^2}}\int_{-\infty}^{\infty} e^{(i\gamma_\ell\omega t - \frac{1}{2}g_\ell\omega^2 t^2)} e^{-\frac{(\omega+\mu_e-y_{k-\ell})^2}{2\sigma_e^2}}\,d\omega\\
&= \frac{\exp\left(\frac{-t^2\left((y_{k-\ell}-\mu_e)^2 g_\ell + \gamma_\ell^2\,\sigma_e^2\right) + 2it\,(y_{k-\ell}-\mu_e)\gamma_\ell}{2\left(1+g_\ell\sigma_e^2 t^2\right)}\right)}{\sqrt{1+g_\ell\sigma_e^2 t^2}}.
\end{aligned}$$

## ACKNOWLEDGMENT

We would like to thank Dr. Guiqiang Dong and Prof. Tong Zhang for teaching us the NAND flash memory channel model. We also thank Dr. Bruce Wilson and Dr. Erich Haratsch of LSI Corporation for steering this research through monthly conference calls.

## REFERENCES

[1] S. Gregori, A. Cabrini, O. Khouri, and G. Torelli, "On-chip error correcting techniques for new-generation flash memories," *Proc. IEEE*, vol. 91, no. 4, pp. 602–616, Apr. 2003.

[2] G. Dong, N. Xie, and T. Zhang, "On the use of soft-decision error-correction codes in NAND flash memory," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 58, no. 2, pp. 429–439, Feb. 2011.

[3] J. Kim and W. Sung, "Low-energy error correction of NAND flash memory through soft-decision decoding," *EURASIP J. Advances in Signal Processing 2012*, 2012:195.

[4] J. Wang, T. Courtade, H. Shankar, and R. D. Wesel, "Soft information for LDPC decoding in flash: mutual-information optimized quantization," in *Proc. IEEE GLOBECOM 2011*, Houston, Texas, USA, Dec. 2011.

[5] G. Dong, S. Li, and T. Zhang, "Using data postcompensation and predistortion to tolerate cell-to-cell interference in MLC NAND flash memory," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 10, pp. 2718–2728, Oct. 2010.

[6] D. Park and J. Lee, "Floating-gate coupling canceller for multi-level cell NAND flash," *IEEE Trans. Magn.*, vol. 47, no. 3, pp. 624–628, Mar. 2011.

[7] B. Ricco, G. Torelli, M. Lanzoni, A. Manstretta, H. E. Maes, D. Montanari, and A. Modelli, "Nonvolatile multilevel memories for digital applications," *Proc. IEEE*, vol. 86, no. 12, pp. 2399–2420, Dec. 1998.

[8] R. Bez, E. Camerlenghi, A. Modelli, and A. Visconti, "Introduction to flash memory," *Proc. IEEE*, vol. 91, no. 4, pp. 489–502, Apr. 2003.

[9] A. Jiang, H. Li, and J. Bruck, "On the capacity and programming of flash memories," *IEEE Trans. Inf. Theory*, vol. 58, no. 3, pp. 1549–1564, Mar. 2012.

[10] K. Takeuchi, T. Tanaka, and H. Nakamura, "A double-level-$V_{th}$ select gate array architecture for multilevel NAND flash memories," *IEEE J. Solid-State Circuits*, vol. 31, no. 4, pp. 602–609, Apr. 1996.

[11] 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.

[12] R. G. Gallager, *Information Theory and Reliable Communication*. New York: John Wiley & Sons, Inc, 1968.

[13] J. Chen and P. H. Siegel, "Markov processes asymptotically achieve the capacity of finite-state intersymbol interference channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 3, pp. 1295–1303, Mar. 2008.

[14] P. O. Vontobel, A. Kavčić, D. M. Arnold, and H.-A. Loeliger, "A generalization of the Blahut-Arimoto algorithm to finite-state channels," *IEEE Trans. Inf. Theory*, vol. 54, no. 5, pp. 1887–1918, May 2008.

[15] G. D. Forney, Jr., "The Viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268–278, Mar. 1973.

[16] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inf. Theory*, vol. IT-20, no. 2, pp. 284–287, Mar. 1974.

[17] X. Huang, A. Kavcic, X. Ma, G. Dong, and T. Zhang, "Optimization of achievable information rates and number of levels in multilevel flash memories," in *ICN 2013 : The Twelfth International Conference on Networks*, Seville, Spain, Jan. 27-Feb. 1 2013, pp. 125–131.

[18] M. Marrow, "Detection and modeling of 2-dimensional signals," Ph.D. thesis, University of California, San Diego, 2004.

[19] D. Park and J. Lee, "Coupling canceller maximum-likelihood (CCML) detection for multi-level cell NAND flash memory," *IEEE Trans. Consumer Electron.*, vol. 57, no. 1, pp. 160–163, Feb. 2011.

[20] X. Wang, G. Dong, L. Pan, and R. Zhou, "Error correction codes and signal processing in flash memory," in *Flash Memories*, I. S. Stievano, Ed. InTech, 2011, pp. 57–82.

[21] A. Kavčić, X. Ma, and N. Varnica, "Matched information rate codes for partial response channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 3, pp. 973–989, Mar. 2005.

[22] J. B. Soriaga, H. D. Pfister, and P. H. Siegel, "Determining and approaching achievable rates of binary intersymbol interference channels using multistage decoding," *IEEE Trans. Inf. Theory*, vol. 53, no. 4, pp. 1416–1429, Apr. 2007.

[23] J. Chen and P. H. Siegel, "On the symmetric information rate of two-dimensional finite state ISI channels," *IEEE Trans. Inf. Theory*, vol. 52, no. 1, pp. 227–236, Jan. 2006.

[24] O. Shental, N. Shental, and S. S. (Shitz), "On the achievable information rates of finite-state input two-dimensional channels with memory," in *Proc. IEEE. Int. Symp. Inform. Theory*, Adelaide, Australia, Sept. 2005, pp. 2354–2358.

[25] H.-A. Loeliger and M. Molkaraie, "Simulation-based estimation of the partition function and the information rate of two-dimensional models," in *Proc. IEEE Int. Symp. Inform. Theory*, Toronto, Canada, July 6-11 2008, pp. 1113–1117.

[26] M. Molkaraie and H.-A. Loeliger, "Monte Carlo algorithms for the partition function and information rates of two-dimensional channels," *IEEE Trans. Inf. Theory*, vol. 59, no. 1, pp. 495–503, Jan. 2013.

[27] A. Kavčić, X. Huang, B. Vasic, W. Ryan, and M. F. Erden, "Channel modeling and capacity bounds for two-dimensional magnetic recording," *IEEE Trans. Magn.*, vol. 46, no. 3, pp. 812–818, Mar. 2010.

[28] X. Chen and K. M. Chugg, "Near-optimal data detection for two-dimensional ISI/AWGN channels using concatenated modeling and iterative algorithms," in *Proc. Int. Conf. Commun.*, vol. 2, Atlanta, Georgia, Jun. 1998, pp. 952–956.

[29] Y. Wu, J. A. O'Sullivan, N. Singla, and R. S. Indeck, "Interative detection and decoding for separable two-dimensional intersymbol interference," *IEEE Trans. Magn.*, vol. 39, no. 4, pp. 2115–2120, July 2003.

[30] M. Marrow and J. K. Wolf, "Iterative detection of 2-dimensional ISI channels," in *Proc. ITW 2003*, Paris, France, Mar. 31-Apr. 4 2003, pp. 131–134.

[31] L. Huang, G. Mathew, and T. C. Chong, "Reduced complexity Viterbi detection for two-dimensional optical recording," *IEEE Trans. Consum. Electron.*, vol. 51, no. 1, pp. 123–129, Feb. 2005.

[32] S. Nabavi and B. V. K. V. Kumar, "Detection methods for holographic data storage," in *2006 Optical Data Storage Meeting*, Montreal, Canada, Apr. 2006, pp. 156–158, Paper TuC3.

[33] M. Keskinoz and B. V. K. V. Kumar, "Efficient modeling and iterative magnitude-squared decision feedback equalization (dfe) for volume holographic storage channel," in *Proc. Int. Conf. Commun.*, vol. 9, Helsinki, Finland, June 11-14 2001, pp. 2696–2700.

[34] O. Shental, A. J. Weiss, N. Shental, and Y. Weiss, "Generalized belief propagation receiver for near-optimal detection of two-dimensional channels with memory," in *Proc. IEEE Inform. Theory Workshop*, San Antonio, Texas, Oct. 2004, pp. 225–229.

[35] N. Singla, J. A. O'Sullivan, R. S. Indeck, and Y. Wu, "Iterative decoding and equalization for 2-D recording channels," *IEEE Trans. Magn.*, vol. 38, no. 5, pp. 2328–2330, Sep. 2002.

[36] A. H. J. Immink and *et al.*, "Signal processing and coding for two-dimensional optical storage," in *Proc. GLOBECOM 2003*, vol. 7, San Francisco, California, Dec. 2003, pp. 3904–3908.

[37] A. Kavčić, X. Ma, and M. Mitzenmacher, "Binary intersymbol interference channels: Gallager codes, density evolution, and code performance bounds," *IEEE Trans. Inf. Theory*, vol. 49, no. 7, pp. 1636–1652, July 2003.

**Meysam Asadi** received his B.Sc. degree in Electrical Engineering from Razi University of Kermanshah, Iran, in 2004, followed by a M.Sc. degree at AmirKabir University, Tehran, Iran, in 2007, in Electrical Engineering and Computer Sciences. He is currently a PhD candidate in the Department of Electrical Engineering at University of Hawaii, Honolulu, USA. His research interests cover estimating channels with memory, estimation in slow mixing Markov processes and detector design for storage systems.

**Xiujie Huang** (S'10-M'12) received the M.Sc. degree in mathematics and the Ph.D. degree in communication and information system from Sun Yat-sen University, Guangzhou, China, in 2006 and 2012, respectively. She was a Postdoctoral Fellow in the Department of Electrical Engineering, University of Hawaii, Honolulu, USA, during 2012-2013. Since November 2013 she has been with Jinan University, Guangzhou, China. Her research interests include Shannon information theory, network information theory and their applications in digital communication and storage systems. Her current researches focus on the capacity (region) determination for the interference and relay channels, and the detection and code design for the flash memory.

**Aleksandar Kavcic** received the Dipl. Ing. degree in Electrical Engineering from Ruhr-University, Bochum, Germany in 1993, and the Ph.D. degree in Electrical and Computer Engineering from Carnegie Mellon University in 1998. Since 2007 he has been with the University of Hawaii, Honolulu where he is presently Professor of Electrical Engineering. Prior to 2007, he was in the Division of Engineering and Applied Sciences at Harvard University. He also held-short term visiting and advisory positions at City University of Hong Kong, Chinese University of Hong Kong, Seagate Technology, Read-Rite Corporation, Quantum Corporation and Link-A-Media Devices. Prof. Kavcic received the IBM Partnership Award in 1999 and the NSF CAREER Award in 2000. He is a co-recipient, with X. Ma and N. Varnica, of the 2005 IEEE Best Paper Award in Signal Processing and Coding for Data Storage. He served on the Editorial Board of the IEEE Transactions on Information Theory as Associate Editor for Detection and Estimation from 2001 to 2004, as Guest Editor of the IEEE Signal Processing Magazine in 2003-2004, and as Guest Editor of the IEEE Journal on Selected Areas in Communications in 2008-2009. From 2005 until 2007, he was the Chair of the Data Storage Technical Committee of the IEEE Communications Society, and in 2014 was the general co-chair of the IEEE Symposium on Information Theory in Honolulu, Hawaii.

**Narayana Prasad Santhanam** is an Assistant Professor at the University of Hawaii since 2009. He obtained his B.Tech from the Indian Institute of Technology, Chennai (then Madras) in 2000; MS and PhD from the University of California, San Diego in 2003 and 2006 respectively. From 2007-2008, he held a postdoctoral position at the University of California, Berkeley. His research interests lie in the intersection of information theory and statistics, with a focus on the undersampled/high dimensional regime and including applications to finance, biology, communication and estimation theory. He is the recipient of the 2006 Information Theory Best Paper award from the IEEE Information Theory Society along with A. Orlitsky and J. Zhang; as well as the organizer of several workshops on high dimensional statistics and "big data" problems over the last five years.