On Achievable Rates of Multistage Decoding on Two-Dimensional ISI Channels

Joseph B. Soriaga, Paul H. Siegel, and Jack K. Wolf
Center for Magnetic Recording Research (CMRR)
University of California, San Diego
9500 Gilman Drive, La Jolla, CA, USA, 92093-0401
Email: {jsoriaga,psiegel,jwolf}@ucsd.edu

Marcus Marrow
Link_A_Media Devices Corp.
2840 San Tomas Expressway, Suite 200
Santa Clara, CA 95051, USA
Email: mmarrow@link-a-media.com

***Abstract*— The achievable information rates for multilevel coding (MLC) systems with multistage decoding (MSD) are examined on two-dimensional binary-input intersymbol interference (ISI) channels. One MSD scheme employs trellis-based detection, while another involves zero-forcing equalization and linear noise prediction. Information rates are determined by examining the output statistics at each stage of MSD. The first scheme is shown to achieve rates very close to known information-theoretic limits. Systems with low-density parity-check codes are then optimized to approach these rates.**

I. Introduction

Two-dimensional (2D) intersymbol-interference (ISI) channels with a binary input constraint arise as models for emerging storage technologies such as multi-track optical recording [1] and holographic recording [2]. The input and output relationship for a discrete-time 2D ISI channel can be described by

$$y_{i,j} = \sum_{s,t} h_{s,t} x_{i-s,j-t} + n_{i,j}, \tag{1}$$

where $\{x_{i,j}\}$ is the channel input 2D-sequence (or page), $\{y_{i,j}\}$ is the channel output, and $\{n_{i,j}\}$ is white Gaussian noise with zero mean and variance σ^2 . The impulse response of the channel $\{h_{i,j}\}$ is assumed to have finite span, and the indices i and j take integer values.

Although this is a natural generalization of the one-dimensional (1D) ISI channel, many important tools are not readily applicable; e.g., the BCJR algorithm [3] for 2D ISI does not necessarily have bounded per-symbol computational complexity [4]. One must rely on suboptimal detection strategies, e.g., Marrow and Wolf [4] or O'Sullivan and Singla [5]. The difficulty in applying the BCJR algorithm also precludes the Monte Carlo method of [6], [7] for calculating achievable information rates, though Chen and Siegel [8] provide related Monte Carlo approaches, while Zhang, Duman, and Kurtas [9] extend the 1D methods of [6], [7] along multiple tracks.

In this paper, we determine achievable rates for 2D ISI channels by considering the application of multilevel coding (MLC) with multistage decoding (MSD) as in [7]. In MLC, several codewords are interleaved, and then sequentially decoded with MSD. Each decoded codeword is accounted for at subsequent decoding stages. Two specific MSD schemes are proposed. One involves the iterative multi-strip (IMS) detection algorithm of [4] (Section III), and the other uses zero-forcing equalization and linear noise prediction (Section IV). As in [7], [10], achievable information rates are determined by analyzing each stage of MSD, and then LDPC component codes are optimized for each stage. This approach is outlined in Section II, and numerical results are given in Section V. When MSD with IMS detection is used, we observe achievable rates near the symmetric information rate (i.e., the mutual information rate for independent and identically distributed equiprobable inputs) of 2D channels, as confirmed by the recent bounds of Chen and Siegel [8]. Additional remarks are provided in Section VI.

II. Multilevel Coding and Multistage Decoding

The structure for MLC and MSD is shown in Fig. 1. For an m-level encoder, a block of data bits $\mathbf{u}$ is partitioned into m subblocks of varying sizes, and each of these is separately mapped to a codeword of length N. For the ℓth code of rate $R_m^{(\ell)}$, the codeword is denoted by $\mathbf{c}^{(\ell)} = (c_1^{(\ell)}, \ldots, c_N^{(\ell)})$. All of the mN code bits are then interleaved into a 2D channel input page, where code bits from the ℓth codeword are placed at positions in the set

$$\mathcal{I}^{(\ell)} = \left\{(i,j) \in \mathbb{Z}^2 \,:\, i + j \equiv (\ell - 1) \bmod m\right\}.$$

This interleaving pattern is illustrated in Fig. 1(b) for the case where $m = 3$. The overall system has a rate $\bar{R}_m = \sum_{\ell=1}^{m} R_m^{(\ell)}/m$.

The multistage decoder, depicted in Fig. 1(c), recovers the codewords sequentially, at each stage using the decisions from previously decoded interleaves. More precisely, for each stage $\ell = 1, \ldots, m$, the decoding comprises two steps:

1) A detector processes the entire received page $\{y_{i,j}\}$ and obtains soft-decisions $\mathbf{z}^{(\ell)} = (z_1^{(\ell)}, \ldots, z_N^{(\ell)})$ corresponding to the code bits in the ℓth interleave. If $\ell > 1$, the codeword decisions $\hat{\mathbf{c}}^{(1)}, \hat{\mathbf{c}}^{(2)}, \ldots, \hat{\mathbf{c}}^{(\ell-1)}$ are also supplied as *a priori* information to the detector.
2) The component decoder determines the ℓth interleaved codeword $\hat{\mathbf{c}}^{(\ell)}$ from $\mathbf{z}^{(\ell)}$.

For $m = 1$, there is a single outer code, and the channel detector is given no *a priori* information. Due to the interleaving

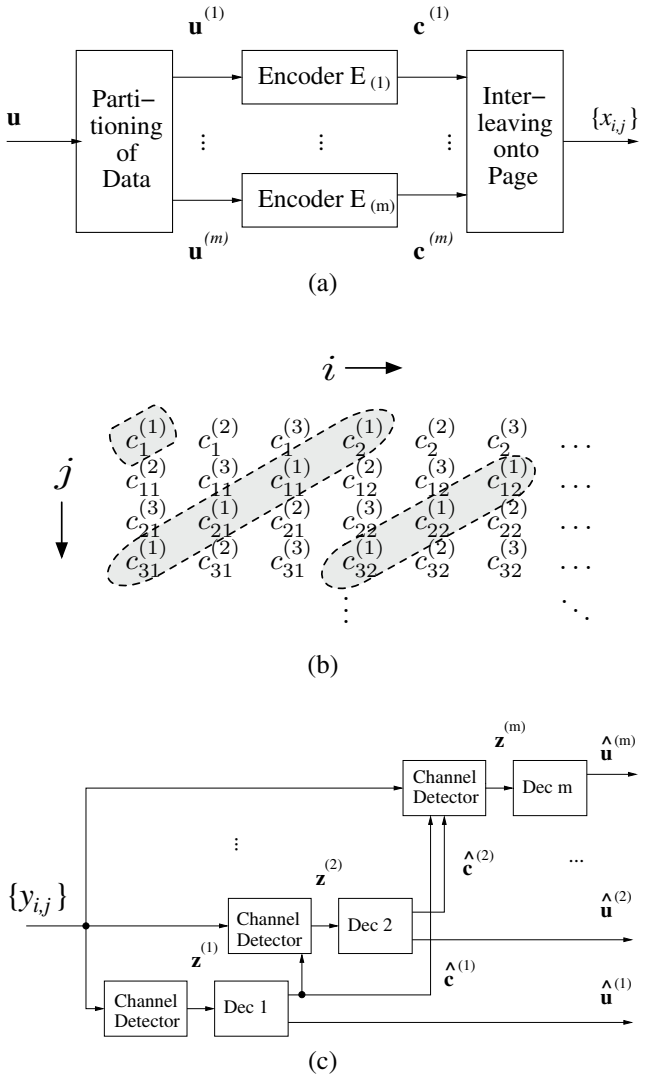

Fig. 1. (a) Multilevel encoder, (b) example interleaving pattern for $m = 3$ on a page of width 30, and (c) multistage decoder.

structure shown in Fig. 1(b), if each of the previous stages decoded correctly, then diagonal bands of the input page are known exactly at subsequent decoding stages.

Generally, a set of achievable information rates for a given m-level system can be determined as follows. At the output of each stage ℓ, one first finds the marginal conditional output distribution for the corresponding interleave, $f(z_j^{(\ell)}|c_j^{(\ell)})$, under the assumption that previous stages decoded correctly, and that all channel inputs are independent and identically distributed equiprobable random variables. From this, the information rate $R_m^{(\ell)} = I(C_j^{(\ell)}; Z_j^{(\ell)})$ can be calculated, and continuing the analysis through all stages yields an overall system rate $\bar{R}_m$. Also, using the marginal distributions at each stage of MSD, one can design LDPC component code ensembles to approach $R_m^{(\ell)}$ by applying methods based on those in [11]. This idea of MLC/MSD system optimization was presented earlier in [10] for 1D ISI channels. In the next two sections, we present two specific implementations of MSD for 2D ISI channels.

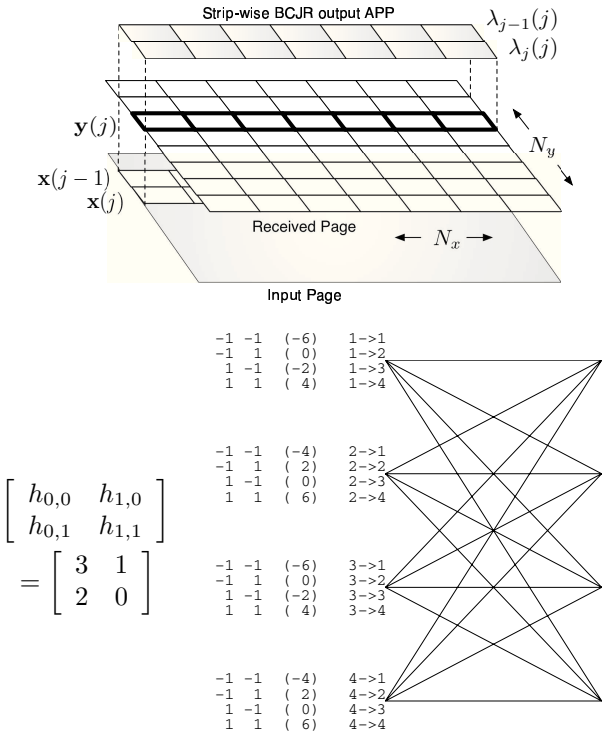

Fig. 2. Illustration of strip-wise BCJR detection and an example trellis.

III. MSD with Iterative Multi-strip Detection

The first MSD scheme we consider involves the IMS detection algorithm of Marrow and Wolf [4]. This soft-decision algorithm has been shown to achieve near-optimal bit detection on 2D ISI channels, while having finite computational and storage complexity per symbol.

The IMS algorithm relies on applying the BCJR algorithm to each row of the received page $\{y_{i,j}\}$. Let us assume that $0 \le i \le N_x$ and $0 \le j \le N_y$ for $\{y_{i,j}\}$, and that the impulse response $\{h_{i,j}\}$ is nonzero only if $0 \le i \le L_x$ and $0 \le j \le L_y$. Any finite-span response can be brought into this form with appropriate delays. In this case, each channel output row $\mathbf{y}(j) = (y_{1,j}, y_{2,j}, \ldots, y_{N_x,j})$ depends upon $L_y + 1$ rows of inputs, $\mathbf{x}(j - L_y)$, $\mathbf{x}(j - L_y + 1), \ldots, \mathbf{x}(j)$. The relationship between the multiple input strips and the output strip can be described exactly with a trellis which has $L_y + 1$ binary inputs per edge, and 1 output per edge. An example of such a trellis is shown in Fig. 2, with labels given for the inputs (and outputs in parenthesis). Consequently, the log *a posteriori* probability ratio (LAPPR) for the inputs given $\mathbf{y}(j)$, i.e.,

$$\lambda_{i,j-t}(j) = \log \frac{P\left(x_{i,j-t} = 1 \,|\, \mathbf{y}(j)\right)}{P\left(x_{i,j-t} = -1 \,|\, \mathbf{y}(j)\right)}, \qquad (2)$$

for $t = 0, \ldots, L_y$ and $i = 0, \ldots, N_x$, can be calculated using the BCJR algorithm on this trellis [3].

Notice from (2) that each decoded strip $\mathbf{y}(j)$ corresponds to $L_y + 1$ input strips, as also depicted in Fig. 2. So when adjacent strips are decoded, overlapping sets of input LAPPRs result.

Altogether a given input strip is addressed by L_y+1 separate strip-wise detectors. Therefore, the IMS algorithm passes these overlapping messages between all strip-wise detectors, effectively providing *extrinsic* information for each detector based on its neighbors. The complete algorithm is summarized below.

Let $a_{i,j} = \log P(x_{i,j}=1)/P(x_{i,j}=-1)$ be the *a priori* information for the channel inputs, and denote the jth row by $\mathbf{a}(j) = (a_{1,j}, a_{2,j}, \ldots, a_{N_x,j})$.

1) *Initialization.* For each row j, let the extrinsic information for the jth strip-wise BCJR detector be

$$\alpha_{j-t}(j) = (\alpha_{1,j-t}(j), \alpha_{2,j-t}(j), \ldots, \alpha_{N_x,j-t}(j)),$$

and set $\alpha_{j-t}(j) = \mathbf{a}(j-t)$, for each $t = 0, \ldots, L_y$.

2) *IMS iterations.* Repeat the following for a fixed number of times, as specified by the user.
   a) *Strip-wise detection.* For each row of outputs $\mathbf{y}(j)$, apply the strip-wise BCJR algorithm to obtain the set of LAPPRs,

$$\lambda_{j-t}(j) = (\lambda_{1,j-t}(j), \lambda_{2,j-t}(j), \ldots, \lambda_{N_x,j-t}(j)),$$

   for $t = 0, \ldots, L_y$, where $\lambda_{j-t}(j)$ is as defined in (2). The extrinsic information $\alpha_{j-t}(j)$ serves as *a priori* input for the strip-wise BCJR algorithm.
   b) *Extrinsic information update.* For each row j and each $t = 0, \ldots, L_y$, use the LAPPRs to calculate $\tilde{\lambda}_{j-t}(j) = \lambda_{j-t}(j) - \alpha_{j-t}(j)$. Then, again for each row j and each $t = 0, \ldots, L_y$, update the extrinsic information for the jth strip-wise detector,

$$\alpha_{j-t}(j) = \mathbf{a}(j-t) + \left(\sum_{k=j-t}^{j-t+L_y} \tilde{\lambda}_{j-t}(k) \right) - \tilde{\lambda}_{j-t}(j).$$

3) *Soft decision.* For each row j, calculate the soft decisions $\mathbf{z}(j) = (z_{1,j}, z_{2,j}, \ldots, z_{N_x,j})$ according to

$$\mathbf{z}(j) = \mathbf{a}(j) + \sum_{k=j}^{j+L_y} \tilde{\lambda}_j(k).$$

Since the IMS algorithm can utilize *a priori* information, it can be readily incorporated into MSD. Given a known set of decisions $\{\hat{x}_{i,j} : i,j \in I^{(\ell)}\}$ from a previous stage ℓ, the *a priori* information $a_{i,j}$ in Step 1 is set to $+\infty$ if $\hat{x}_{i,j} = 1$, and to $-\infty$ otherwise. For this MSD scheme, there is no closed form solution for the achievable rates for each stage. So simulations are used to generate output sequences at each stage ℓ of MSD, and the distribution $f(z_j^{(\ell)}|c_j^{(\ell)})$ is approximated with histograms. Achievable rates are then calculated using these distributions. It can be shown that these densities are symmetric, i.e., $f(z_j^{(\ell)}|c_j^{(\ell)}) = f(-z_j^{(\ell)}| - c_j^{(\ell)})$, and it is also observed empirically that they are exponential symmetric, i.e., $f(z_j^{(\ell)}|1) = e^{z_j^{(\ell)}} f(z_j^{(\ell)}| - 1)$. Numerical results for the achievable rates of this scheme are plotted in Fig. 4.

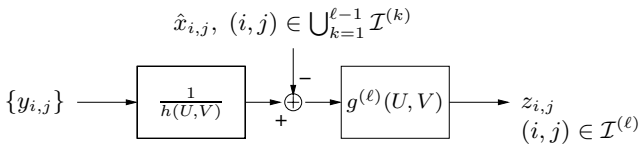

Fig. 3. Detector block for MSD with zero-forcing equalization and noise prediction.

IV. MSD with Zero-Forcing Equalization and Linear Noise Prediction

We now consider a low-complexity alternative to MSD with IMS detection, which further permits a closed form calculation of the densities and achievable rates at each stage.

Recall that a simple method for equalizing a 2D ISI channel is *zero-forcing* equalization; i.e., if we introduce the (U,V)-transform of $\{h_{i,j}\}$ as

$$h(U,V) = \sum_{i,j} h_{i,j} U^i V^j,$$

then the ideal zero-forcing equalizer has a transform equal to $g(U,V) = 1/h(U,V)$. This inverse response may not be stable in the sense that $\sum |g_{i,j}| < \infty$ [12], and so this equalization technique is not applicable to all ISI channels. The zero-forcing equalizer can be very closely approximated using the pseudo-inverse method of Weeks [13], and for certain responses, e.g., $\{h_{i,j}\}$ where $h_{i,j} = 0$ whenever $i < 0$ or $j < 0$, it can be implemented exactly using feedback.

One adverse consequence of zero-forcing equalization is that it colors the noise and enhances the noise power. However, MSD can exploit this noise coloration through linear prediction, because once channel input bits are known from previous stage decisions, then one can solve for the corresponding noise values on those interleaves. The resulting structure for the detector block at each stage of MSD (see Fig. 1(c)) is shown in Fig. 3, and the filter components for each stage are described below. An example is provided in Fig. 5.

Definition 1: $\mathcal{D}_L^{(\ell)}$ denotes the indices for the nonzero coefficients of the ℓ-stage noise prediction filter with a maximum delay of L (excluding the coordinate (0,0)). Since this filter only operates on values from the current and previously decoded interleaves, we can formally define this set as

$$\mathcal{D}_L^{(\ell)} = -\left\{ (s,t) \in \bigcup_{k=1}^{\ell-1} \mathcal{I}^{(k)} - (\ell-1, 0), \quad |s|, |t| \le L \right\}.$$

Using this, we can now define the optimal prediction filter at each stage. Let $\{\tilde{y}_{i,j}\}$ denote the output from the zero-forcing equalizer in Fig. 3, and let $\{\tilde{n}_{i,j}\}$ be the corresponding colored noise such that $\tilde{y}_{i,j} = x_{i,j} + \tilde{n}_{i,j}$.

Lemma 1: Suppose $\mathcal{D}_L^{(\ell)} = \{(s_1,t_1), \ldots, (s_n,t_n)\}$. Let the colored-noise autocorrelation function be $\tilde{R}(s,t) = E(\tilde{n}_{i,j}\tilde{n}_{i-s,j-t})$, let $\tilde{R}$ equal

$$\begin{bmatrix} \tilde{R}(0,0) & \cdots & \tilde{R}(s_1-s_n, t_1-t_n) \\ \tilde{R}(s_2-s_1, t_2-t_1) & & \tilde{R}(s_2-s_n, t_2-t_n) \\ \tilde{R}(s_3-s_1, t_3-t_1) & & \tilde{R}(s_3-s_n, t_3-t_n) \\ \vdots & \ddots & \vdots \\ \tilde{R}(s_n-s_1, t_n-t_1) & \cdots & \tilde{R}(0,0) \end{bmatrix},$$

and let $\mathbf{v} = \left(\tilde{R}(s_1,t_1),\ldots,\tilde{R}(s_n,t_n)\right)^T$. For the ℓ-th stage noise prediction filter, set $g_{0,0}^{(\ell)} = 1$. The optimal values for the remaining coefficients $\{g_{s,t}^{(\ell)},\ (s,t) \in \mathcal{D}_L^{(\ell)}\}$ are then

$$\mathbf{g}^{(\ell)} = \left[g_{s_1,t_1}^{(\ell)},\ldots,g_{s_n,t_n}^{(\ell)}\right]^T = -\tilde{R}^{-1}\mathbf{v},$$

and the noise variance after the ℓ-th stage prediction filter is $E\left(\tilde{n}_{i,j}^{(\ell)}\right)^2 = \tilde{R}(0,0) - \mathbf{v}^T\tilde{R}^{-1}\mathbf{v}$.

Proof: For $g_{0,0}^{(\ell)} = 1$, the output from the ℓ-th stage prediction filter is

$$\tilde{y}_{i,j}^{(\ell)} = \tilde{y}_{i,j} + \sum_{(s,t)\in\mathcal{D}_L^{(\ell)}} g_{s,t}^{(\ell)}\tilde{n}_{i-s,j-t} = x_{i,j} + \tilde{n}_{i,j}^{(\ell)}. \quad (3)$$

The resulting noise variance may then be expressed as

$$E\left(\tilde{n}_{i,j}^{(\ell)}\right)^2 = E\left(\tilde{n}_{i,j} + \sum_{(s,t)\in\mathcal{D}_L^{(\ell)}} g_{s,t}^{(\ell)}\tilde{n}_{i-s,j-t}\right)^2. \quad (4)$$

From the orthogonality principle (or equivalently by taking derivatives), it follows that the remaining filter coefficients which minimize this variance satisfy

$$E\left[\left(\tilde{n}_{i,j} + \sum_{(s,t)\in\mathcal{D}_L^{(\ell)}} g_{s,t}^{(\ell)}\tilde{n}_{i-s,j-t}\right)\tilde{n}_{i-s',j-t'}\right] = 0 \quad (5)$$

for all $(s',t') \in \mathcal{D}_L^{(\ell)}$. This can be simplified to,

$$-E\left(\tilde{n}_{i,j}\tilde{n}_{i-s',j-t'}\right) = \sum_{(s,t)\in\mathcal{D}_L^{(\ell)}} g_{s,t}^{(\ell)}E(\tilde{n}_{i-s,j-t}\tilde{n}_{i-s',j-t'})$$

or

$$-\tilde{R}(s',t') = \sum g_{s,t}^{(\ell)}\tilde{R}(s'-s,t'-t).$$

Furthermore, this linear system of equations can be expressed in matrix form as $-\mathbf{v} = \tilde{R}\mathbf{g}^{(\ell)}$, and thus the optimal coefficients are $\mathbf{g}^{(\ell)} = -\tilde{R}^{-1}\mathbf{v}$.

Noting (4), we find that

$$\begin{aligned} E\left(\tilde{n}_{i,j}^{(\ell)}\right)^2 &= E\left(\tilde{n}_{i,j}^2\right) + 2\sum_{(s,t)\in\mathcal{D}_L^{(\ell)}} g_{s,t}^{(\ell)}E\left(\tilde{n}_{i,j}\tilde{n}_{i-s,j-t}\right) + \\ &\quad \sum_{(s,t),(s',t')\in\mathcal{D}_L^{(\ell)}} g_{s',t'}^{(\ell)}g_{s,t}^{(\ell)}E\left(\tilde{n}_{i-s',j-t'}\tilde{n}_{i-s,j-t}\right) \\ &= \tilde{R}(0,0) + 2\mathbf{v}^T\mathbf{g}^{(\ell)} + (\mathbf{g}^{(\ell)})^T\tilde{R}\mathbf{g}^{(\ell)} \\ &= \tilde{R}(0,0) - 2\mathbf{v}^T\tilde{R}^{-1}\mathbf{v} + \mathbf{v}^T\tilde{R}^{-1}\mathbf{v} \\ &= \tilde{R}(0,0) - \mathbf{v}^T\tilde{R}^{-1}\mathbf{v}. \end{aligned}$$

which completes the proof of the lemma. ■

Since the outputs at each stage are conditionally Gaussian, we also have the next corollary.

Corollary 1: $R_m^{(\ell)} = I(X;X+Z^{(\ell)})$, where $P(X=1) = P(X=-1) = 1/2$, and $Z^{(\ell)}$ is Gaussian with zero mean and $\mathrm{Var}(Z^{(\ell)}) = E\left(\tilde{n}_{i,j}^{(\ell)}\right)^2$.

Incidentally, these results do not require that $h_{i,j} = 0$ whenever $i < 0$ or $j < 0$. However, for channels which do satisfy this condition, we can characterize the asymptotic behavior of the achievable system rate below.

Theorem 1: $\lim_{m\to\infty}\bar{R}_m = I(X;X+Z^*)$, where $P(X=1) = P(X=-1) = 1/2$, and Z^* is Gaussian with zero mean and $\mathrm{Var}(Z^*) = \sigma^2/h_{0,0}^2$.

The proof of this theorem relies on the following lemma.

Lemma 2: Assume that $h_{s,t} = 0$ whenever $s < 0$ or $t < 0$, and let the set of active indices be $\mathcal{J}$ and $|h_{0,0}| > 0$. If at the ℓth stage $\mathcal{J} \subseteq \mathcal{D}_L^{(\ell)} \cup \{(0,0)\}$, and $s,t \geq 0$ for any $(s,t) \in \mathcal{D}_L^{(\ell)}$, then the optimal filter coefficients are $g_{s,t}^{(\ell)} = h_{s,t}/h_{0,0}$ for $(s,t) \in \mathcal{J}$, and $g_{s,t}^{(\ell)} = 0$ otherwise. The noise variance at the output of this filter is $\sigma^2/h_{0,0}^2$.

Proof: For these particular channel responses, the colored noise term $\tilde{n}_{i,j}$ at the output of the zero-forcing equalizer can be described in terms of $n_{i,j}$ using feedback, i.e.,

$$\tilde{n}_{i,j} = \frac{1}{h_{0,0}}\left(n_{i,j} - \sum_{(s,t)\in\mathcal{J}\setminus\{(0,0)\}} h_{s,t}\tilde{n}_{i-s,j-t}\right), \quad (6)$$

with i increasing along the horizontal axis, and then j increasing along the vertical axis. Alternatively, the impulse response of the zero-forcing equalizer $\{g_{s,t}\}$ is such that $g_{s,t} = 0$ if $s < 0$ or $t < 0$, and so we can write

$$\tilde{n}_{i,j} = \sum_{s,t\geq 0} g_{s,t}n_{i-s,j-t}. \quad (7)$$

By noting the relation in (6), direct substitution of the coefficients $\{g_{s,t}^{(\ell)}\}$ into equation (3) yields $\tilde{n}_{i,j}^{(\ell)} = n_{i,j}/h_{0,0}$. Moreover, it follows that $E\left[n_{i,j}\tilde{n}_{i-s,j-t}\right] = 0$ for all $(s,t) \in \mathcal{D}_L^{(\ell)}$, because $n_{i,j}$ never occurs in the summation (7) when shifted by (s,t), where $s,t \geq 0$ and $(s,t) \neq (0,0)$. Therefore, the choice of filter coefficients $\{g_{s,t}^{(\ell)}\}$ is optimal since it satisfies (5), which proves the lemma. ■

The proof of the theorem is omitted, but it essentially relies on the fact that the fraction of stages that satisfy Lemma 2 goes to 1 as $m \to \infty$.

V. Numerical Results

For MSD with IMS detection, we evaluate the achievable rates for $m = 1,2,3$ on the channel considered by Chen and Siegel [8]. The results are plotted in Fig. 4, where it can be observed that a two-stage system has an achievable rate within 0.5 dB of the symmetric information rate (SIR) lower bound, and a three-stage system effectively achieves the bound. As mentioned in Section II, LDPC codes are optimized with respect to the output densities at each stage, and these results are also shown in the figure.

To illustrate the second MSD technique, in Fig. 5 we present the optimal noise prediction filters (with $L = 1$, and with center taps marked by a '†') for a four-stage system on a channel for which the zero-forcing equalizer is stable. The noise variance at the output of the zero-forcing equalizer is about $2.76\sigma^2$, or 4.42 dB larger than the original noise variance σ^2. The achievable rates on this channel for $m = 1,2,3$ are

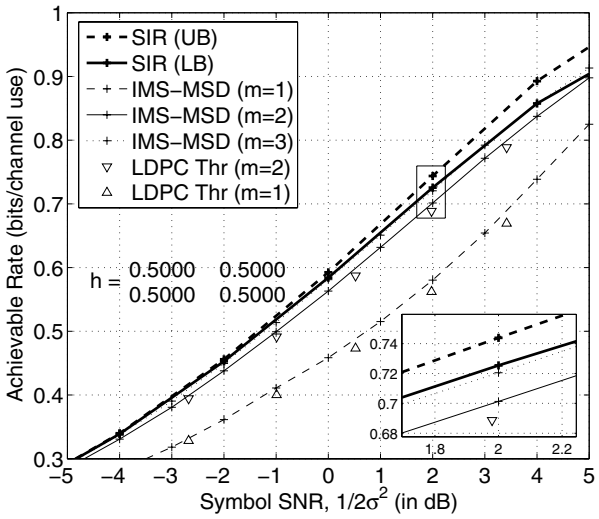

Fig. 4. Achievable information rates and LDPC code thresholds for MSD with IMS detection.

$$\{h_{i,j}\} = \begin{bmatrix} 0.8^{\dagger} & 0.4 \\ 0.4 & 0.2 \end{bmatrix}$$

ℓ	$\mathbf{g}^{(\ell)}$	Variance/σ^2
2	$\begin{bmatrix} 1^{\dagger} & 0.3992 \\ 0.3992 & \end{bmatrix}$	1.66 (2.21 dB)
3	$\begin{bmatrix} -0.1449 & & \\ & 1^{\dagger} & 0.4805 \\ & 0.4805 & 0.2394 \end{bmatrix}$	1.51 (1.78 dB)
4	$\begin{bmatrix} 0.1728 & 0.3475 & \\ 0.3475 & 1^{\dagger} & 0.3475 \\ & 0.3475 & 0.1728 \end{bmatrix}$	1.09 (0.36 dB)

Fig. 5. Example of optimal noise prediction filters for MSD.

shown in Fig. 6, along with thresholds for optimized LDPC code ensembles.

VI. Further Remarks

One important implication here is that IMS detection can be used to nearly achieve the SIR, when incorporated into MSD. It would be interesting to see if turbo-equalization systems can also be designed to approach the SIR using IMS detection. Naturally, other detection methods such as [5] can also be used with MSD.

VII. Acknowledgements

The authors are grateful to Jiangxin Chen for the symmetric information rate bounds from [8], and to Henry Pfister for program code to help with LDPC code optimization. This research was supported by the National Institute of Standards and Technology in cooperation with InPhase Technologies.

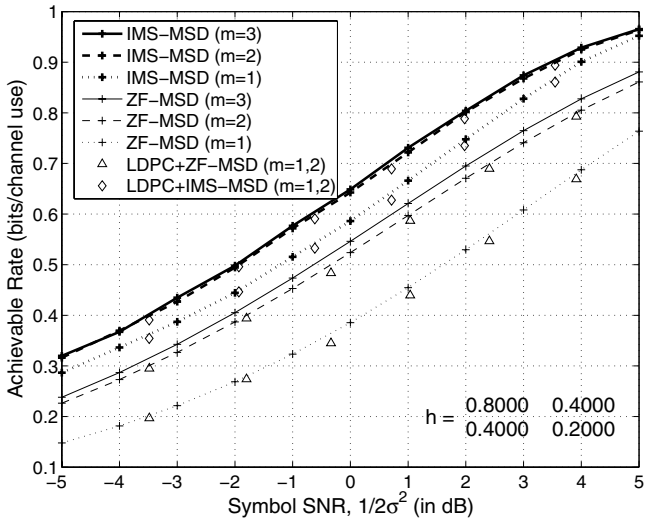

Fig. 6. Comparison of MSD with IMS detection versus MSD with zero-forcing equalization and noise prediction.

References

[1] W. Coene, "Two-dimensional optical storage," in *Technical Digest of Optical Data Storage (ODS) Conference*, (Vancouver, Canada), pp. 90–92, 2003.

[2] J. Ashley, M.-P. Bernal, G. W. Burr, H. Coufal, H. Guenther, J. A. Hoffnagle, C. M. Jefferson, B. Marcus, R. M. Macfarlane, R. M. Shelby, and G. T. Sincerbox, "Holographic data storage," *IBM Journal of Research and Development*, vol. 44, pp. 341–368, May 2000.

[3] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Transactions on Information Theory*, vol. 20, pp. 284–287, March 1974.

[4] M. Marrow and J. K. Wolf, "Detection of 2-dimensional signals in the presence of isi and noise," in *International Symposium on Information Theory and its Applications*, (Parma, Italy), pp. 891–894, October 2004.

[5] J. A. O'Sullivan and N. Singla, "Ordered subsets message-passing," in *Proceedings of IEEE International Symposium on Information Theory*, (Yokohama, Japan), p. 349, June 29-July 4, 2003.

[6] D. Arnold and H. Loeliger, "On the information rate of binary-input channels with memory," in *Proceedings IEEE International Conference on Communications*, (Helsinki, Finland), pp. 2692–2695, June 2001.

[7] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rates of finite state ISI channels," in *Proceedings IEEE Global Telecommunications Conference*, (San Antonio, Texas, USA), pp. 2992–2996, November 2001.

[8] J. Chen and P. Siegel, "On the symmetric information rate of two-dimensional finite state ISI channels," in *Proceedings of the IEEE Information Theory Workshop*, (Paris, France), pp. 320–323, March 2003.

[9] Z. Zhang, T. M. Duman, and E. M. Kurtas, "Achievable information rates and coding for MIMO systems over ISI and frequency-selective fading channels," *IEEE Transactions on Communications*, vol. 52, pp. 1698–1710, October 2004.

[10] J. B. Soriaga, H. D. Pfister, and P. H. Siegel, "On approaching the capacity of finite-state intersymbol interference channels," in *Information, Coding, and Mathematics: Proceedings of Workshop Honoring Prof. Bob McEliece on His 60th Birthday* (M. Blaum, P. G. Farrell, and H. C. A. Van Tilborg, eds.), pp. 365–378, Boston, USA: Kluwer Academic Publishers, 2002.

[11] T. J. Richardson, M. A. Shokrollahi, and R. L. Urbanke, "Design of capacity-approaching irregular low-density parity-check codes," *IEEE Transactions on Information Theory*, vol. 27, pp. 619–637, February 2001.

[12] D. E. Dudgeon and R. M. Mersereau, *Multidimensional Signal Processing*. Englewood Cliffs, New Jersey, USA: Prentice-Hall, Inc., 1984.

[13] W. Weeks IV, *Full-Surface Data Storage*. PhD thesis, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA, 2000.