Correspondence

On the Symmetric Information Rate of Two-Dimensional Finite-State ISI Channels

Jiangxin Chen and Paul H. Siegel, *Fellow, IEEE*

***Abstract*—We derive a pair of bounds (upper and lower) on the symmetric information rate of a two-dimensional finite-state intersymbol interference (ISI) channel model. For channels with small impulse response support, they can be estimated via a modified forward recursion of the Bahl–Cocke–Jelinek–Raviv (BCJR) algorithm. The convergence of the bounds is also analyzed. To relax the constraint on the size of the impulse response, a new upper bound is proposed which allows the tradeoff of the computational complexity and the tightness of the bound. These bounds are further extended to d-dimensional $(d > 2)$ ISI channels.**

***Index Terms*—Bahl–Cocke–Jelinek–Raviv (BCJR) algorithm, hidden Markov field, information rate, Peano–Hilbert curve, two-dimensional intersymbol interference (ISI) channel.**

I. Introduction

One approach to achieving higher information storage density is the use of page-oriented data recording technologies, such as holographic memory [1]. Instead of recording the data in one-dimensional tracks, these technologies store the data on two-dimensional surfaces. A commonly used channel model for such a two-dimensional recording channel is the two-dimensional finite-state intersymbol interference (ISI) channel with additive white Gaussian noise (AWGN), as described by

$$y[i,j] = \sum_{k=0}^{n_1-1} \sum_{l=0}^{n_2-1} h[k,l]x[i-k,j-l] + n[i,j] \tag{1}$$

where $x[i,j]$ is the channel input with finite alphabet of cardinality $|\mathcal{X}|$, $y[i,j]$ is the channel output, and $n[i,j]$ is independent and identically distributed (i.i.d.) zero-mean Gaussian noise with variance $\sigma^2 = N_0/2$.

As in the case of one-dimensional channels, the capacity of this two-dimensional channel is defined as the maximum mutual information rate $I(\mathcal{X},\mathcal{Y})$ over all input distributions, where $\mathcal{X} = \{x[i,j]\}$ and $\mathcal{Y} = \{y[i,j]\}$. When the input is i.i.d., equiprobable, the mutual information rate is called the symmetric information rate (SIR). The capacity and the SIR provide useful measures of the storage density that can, in principle, be achieved with page-oriented technologies, and also serve as performance benchmarks for channel coding and detection methods.

Various bounds on the capacity and the SIR have been developed for certain one-dimensional ISI channels; see, e.g., [2]–[5]. Recently, several authors independently proposed a new Monte Carlo approach to calculating a convergent sequence of lower bounds on these information rates [6]–[8]. The method requires the calculation of the joint probability of a long sample realization of the channel output. The forward recursion of the sum-product Bahl–Cocke–Jelinek–Raviv (BCJR) algorithm [10], applied to the combined source–channel trellis, can be used to reduce the overall computational complexity of the calculations. This approach was further generalized and applied to multitrack recording systems [9].

In this correspondence, we investigate the extension of this approach to two-dimensional finite-state ISI channels. The main difficulty that prevents us from extending the one-dimensional approach directly is that there is no direct counterpart of the BCJR algorithm that can simplify the calculations for a large sample output array in the two-dimensional setting. To overcome this problem, we derive upper and lower bounds on the entropy rate of a large output array based upon conditional entropies of smaller output arrays. The convergence of these bounds is investigated. By modifying the one-dimensional Monte Carlo technique to the calculation of these conditional entropies, we are able to compute fairly tight upper and lower bounds on the SIR for two-dimensional ISI channels with small impulse response support. To further reduce the computational complexity for the channels with larger impulse response support, we derive an alternative upper bound which uses an auxiliary one-dimensional ISI channel so that the one-dimensional Monte Carlo technique can be applied directly. It allows us to make the tradeoff between the auxiliary one-dimensional channel memory length (which determines the complexity) and the tightness of the bound. Although these bounds are derived in the context of two-dimensional ISI channels, they can be easily extended to higher dimensional ISI channels which may be used to model the inter-page interference in holographic memories [12].

The paper is organized as follows. In Section II, we derive the upper and lower bounds on the SIR of two-dimensional finite-state ISI channels. We then present a lower bound on the convergence rate and describe the Monte Carlo approach to computing these bounds. Section III introduces the alternative upper bound which makes use of an auxiliary one-dimensional ISI channel. These bounds are extended to d-dimensional $(d > 2)$ ISI channels in Section IV. Section V provides numerical results for some two-dimensional ISI channels, and Section VI concludes the paper.

II. Bounds on the SIR of Two-Dimensional ISI Channels

A. SIR Bounds Based on Conditional Entropies

In this two-dimensional ISI channel model, the input array $\mathcal{X}$ is a two-dimensional discrete random process consisting of i.i.d. random variables. The corresponding output array $\mathcal{Y}$ is a two-dimensional continuous random process. It is easy to see that both are stationary random processes. Let $L_{i,j}^{i+m-1,j+n-1}$ be the set of positions forming an $m \times n$ rectangular block whose upper left corner is (i,j) [1] and whose lower right corner is $(i+m-1, j+n-1)$. That is,

$$L_{i,j}^{i+m-1,j+n-1} = \{(s,t) \in Z^2 : i \leq s \leq i+m-1, j \leq t \leq j+n-1\}.$$

The output array indexed on the set $L_{i,j}^{i+m-1,j+n-1}$ is $Y\{L_{i,j}^{i+m-1,j+n-1}\}$ and is abbreviated as $Y_{i,j}^{i+m-1,j+n-1}$. The mutual information rate between $\mathcal{X}$ and $\mathcal{Y}$ is

$$I(\mathcal{X};\mathcal{Y}) = H(\mathcal{Y}) - H(\mathcal{Y}|\mathcal{X}) = H(\mathcal{Y}) - H(\mathcal{N}) \tag{2}$$

[1]Throughout the correspondence, (i,j) denotes the position in the ith row and the jth column.

Manuscript received March 10, 2004; revised July 15, 2005. This work was supported in part by the Center for Magnetic Recording Research at the University of California, San Diego, and by the National Science Foundation under Grant CCR-0219582. The material in this correspondence was presented in part at the IEEE Information Theory Workshop, Paris, France, April 2003.

J. Chen is with Prediction Company, UBS, Santa Fe, NM 87505 USA (e-mail: simple_address@yahoo.com).

P. H. Siegel is with the Department of Electrical and Computer Engineering, and the Center for Magnetic Recording Research, University of California, San Diego, La Jolla, CA 92093 USA (e-mail: psiegel@ucsd.edu).

Communicated by R. W. Yeung, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2005.860467

where

$$H(\mathcal{Y}) = \lim_{m,n\to\infty} \frac{1}{mn} H(Y_{1,1}^{m,n})$$

is the entropy rate of the output process and $H(\mathcal{N}) = \frac{1}{2}\log(\pi e N_0)$ is the noise entropy rate. Equation (2) comes from the fact that

$$\begin{aligned} H(\mathcal{Y}|\mathcal{X}) &= \lim_{m,n\to\infty} \frac{1}{mn} H\left(Y_{1,1}^{m,n} \left| X_{1,1}^{m,n}\right.\right) \\ &= \lim_{m,n\to\infty} \frac{1}{mn} \left[H\left(Y_{1,1}^{m,n} \left| X_{-n_1+2,-n_2+2}^{m,n}\right.\right) \right. \\ &\quad \left. + I\left(X_{-n_1+2,-n_2+2}^{0,n}, X_{1,-n_2+2}^{m,0}; Y_{1,1}^{m,n} \left| X_{1,1}^{m,n}\right.\right)\right] \\ &= \lim_{m,n\to\infty} \frac{1}{mn} H\left(Y_{1,1}^{m,n} \left| X_{-n_1+2,-n_2+2}^{m,n}\right.\right) \\ &= \lim_{m,n\to\infty} \frac{1}{mn} H\left(N_{1,1}^{m,n}\right) \\ &= H(\mathcal{N}) = \frac{1}{2}\log(\pi e N_0) \end{aligned} \tag{3}$$

since

$$\begin{aligned} 0 &\le \lim_{m,n\to\infty} \frac{1}{mn} I\left(X_{-n_1+2,-n_2+2}^{0,n}, X_{1,-n_2+2}^{m,0}; Y_{1,1}^{m,n} \left| X_{1,1}^{m,n}\right.\right) \\ &\le \lim_{m,n\to\infty} \frac{(n_1-1)(n_2+n-1)+(n_2-1)m}{mn} \log|\mathcal{X}| = 0. \end{aligned}$$

In the rest of the correspondence, we will no longer distinguish between bounding the channel output entropy rate and bounding the SIR since the two rates differ only by a constant. The problem is now reduced to the evaluation of the entropy rate of the two-dimensional output process $\mathcal{Y}$, if it exists. For a stationary one-dimensional random process $\mathcal{Z} = \{Z[i]\}$, the limit

$$H(\mathcal{Z}) = \lim_{n\to\infty} \frac{1}{n} H(Z[1], \ldots, Z[n])$$

exists due to the subadditivity of the entropy $H(Z[1], \ldots, Z[n])$. Similarly, since the subadditive property holds for the entropy $H(Y_{1,1}^{m,n})$ in each dimension, the limit $\lim_{m,n\to\infty} \frac{1}{mn} H(Y_{1,1}^{m,n})$ also exists when $\mathcal{Y}$ is stationary [11].

Although we could estimate $H(\mathcal{Y})$ by calculating the sample entropy rate of a very large array, similar to the approach used in [6]–[8], the huge computational complexity makes it impractical. As pointed out earlier, there is no counterpart of the BCJR algorithm for two-dimensional processes to help reduce the complexity. Instead, we will use conditional entropies based upon a smaller array to derive upper and lower bounds on $H(\mathcal{Y})$, and subsequently, the SIR. Conditional entropies have been used to bound the entropy rate of one-dimensional hidden Markov processes (see, e.g., [13]). We will adapt the approach to accommodate the nature of the two-dimensional processes.

Before the development of the bounds on the SIR, we first define the ordering $\prec$ of the elements in a two-dimensional array Y. Unlike the one-dimensional case in which there is a natural ordering that coincides with the progression of time, there are many equally viable ways to order the elements in a two-dimensional array. In this correspondence, we consider two simple orderings, row-by-row ordering $\prec_R$ and column-by-column ordering $\prec_C$. Given two positions (t_1, t_2) and (t_1', t_2'), in the row-by-row ordering, we have $(t_1, t_2) \prec_R (t_1', t_2')$ if $t_1 < t_1'$ or if $t_1 = t_1'$ and $t_2 < t_2'$. Similarly, in the column-by-column ordering, we have $(t_1, t_2) \prec_C (t_1', t_2')$ if $t_2 < t_2'$ or if $t_2 = t_2'$ and $t_1 < t_1'$. To simplify the notation, we will use R and C to denote row-by-row ordering and column-by-column ordering, respectively. Therefore, we have $\prec \in \{R, C\}$.

Let $\underline{l} = [l_1, \ldots, l_4]$ be a nonnegative vector (i.e., $l_i \ge 0, \forall i$). We denote by $\mathrm{Past}_{\prec,\underline{l}}(i,j)$ the subset of positions inside $L_{i-l_1,j-l_3}^{i+l_2,j+l_4}$ that precede position (i,j) according to the ordering $\prec$. Fig. 1 depicts $\mathrm{Past}_{C,[3,2,6,0]}(i,j)$ and $\mathrm{Past}_{R,[6,0,4,2]}(i,j)$. The channel outputs indexed by $\mathrm{Past}_{\prec,\underline{l}}(i,j)$ are denoted by $Y\{\mathrm{Past}_{\prec,\underline{l}}(i,j)\}$. The boundary of $\mathrm{Past}_{\prec,\underline{l}}(i,j)$ is denoted by $\mathrm{Strip}_{\prec,\underline{l}}(i,j)$. That is,

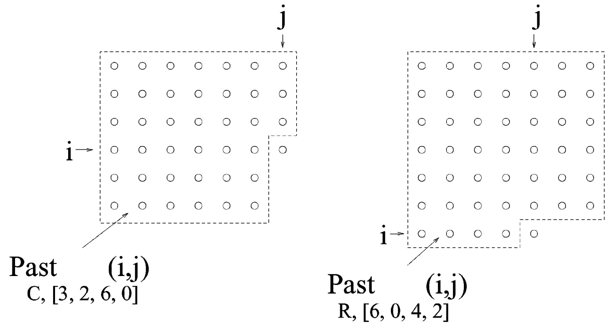

Fig. 1. The region $\mathrm{Past}_{C,[3,2,6,0]}(i,j)$ and $\mathrm{Past}_{R,[6,0,4,2]}(i,j)$.

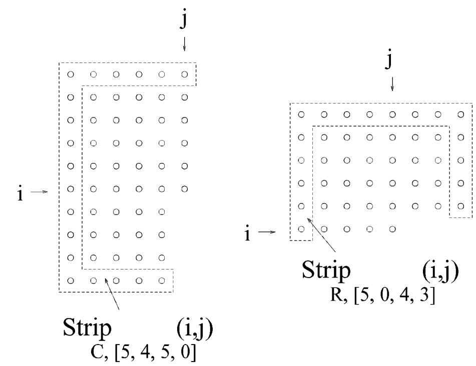

Fig. 2. The region $\mathrm{Strip}_{C,[5,4,5,0]}(i,j)$ and $\mathrm{Strip}_{R,[5,0,4,3]}(i,j)$.

$$\mathrm{Strip}_{\prec,\underline{l}}(i,j) = \begin{cases} L_{i-l_1,j-l_3}^{i,j-l_3} \cup L_{i-l_1,j-l_3}^{i-l_1,j+l_4} \cup L_{i-l_1,j+l_4}^{i-1,j+l_4}, & \text{if } \prec = R \\ L_{i-l_1,j-l_3}^{i+l_2,j-l_3} \cup L_{i-l_1,j-l_3}^{i-l_1,j} \cup L_{i+l_2,j-l_3}^{i+l_2,j-1}, & \text{if } \prec = C. \end{cases} \tag{4}$$

Fig. 2 illustrates $\mathrm{Strip}_{C,[5,4,5,0]}(i,j)$ and $\mathrm{Strip}_{R,[5,0,4,3]}(i,j)$ as examples. It is obvious that the value of l_2 is irrelevant to $\mathrm{Past}_{R,\underline{l}}(i,j)$ and $\mathrm{Strip}_{R,\underline{l}}(i,j)$, as is l_4 to $\mathrm{Past}_{C,\underline{l}}(i,j)$ and $\mathrm{Strip}_{C,\underline{l}}(i,j)$.

In order to derive the bounds based on the conditional entropy, we first introduce a lemma which links the entropy rate defined above and the conditional entropy of the stationary two-dimensional random process. This result was first derived by Katznelson and Weiss [14] for the more general Z^d lattice. Anastassiou and Sakrison [15] obtained a similar result for stationary two-dimensional fields on Z^2. For the time being, we state this result in the context of the Z^2 lattice.

Lemma 2.1: For a stationary two-dimensional random field $\mathcal{Y}$ on Z^2, the entropy rate $H(\mathcal{Y})$ satisfies the following equality:

$$H(\mathcal{Y}) = H(Y[i,j] | Y\{\mathrm{Past}_{\prec,\underline{\infty}}(i,j)\})$$

where $\underline{\infty} = [\infty, \infty, \infty, \infty]$, meaning that for any given $\epsilon > 0$, there exist $M_i > 0, i = 1, \ldots, 4$, such that when $l_i > M_i$ for every $1 \le i \le 4$

$$|H(\mathcal{Y}) - H(Y[i,j] | Y\{\mathrm{Past}_{\prec,\underline{l}}(i,j)\})| < \epsilon.$$

This result holds for both orderings. The lemma also applies to random fields on lattices that can be mapped to Z^2, such as the hexagonal lattice [16]. In Section V, we will give an example of a two-dimensional ISI channel on a hexagonal lattice.

Applying a bounding technique similar to that used in the setting of a one-dimensional hidden Markov process [13], we obtain the following bounds.

Theorem 2.1: For a stationary two-dimensional random field $\mathcal{Y}$, an upper bound on its entropy rate is

$$H(\mathcal{Y}) \leq \min_{\prec \in \{R,C\}} H^{U1}_{\prec,\underline{l}}$$

where

$$H^{U1}_{\prec,\underline{l}} = H(Y[i,j]|Y\{\mathrm{Past}_{\prec,\underline{l}}(i,j)\})$$

and the l_i's are finite nonnegative integers.

For a stationary two-dimensional hidden Markov field $\mathcal{Y}$, a lower bound on its entropy rate is

$$H(\mathcal{Y}) \geq \max_{\prec \in \{R,C\}} H^{L1}_{\prec,\underline{l}}$$

where

$$H^{L1}_{\prec,\underline{l}} = H(Y[i,j]|Y\{\mathrm{Past}_{\prec,\underline{l}}(i,j)\}, S\{\mathrm{Strip}_{\prec,\underline{l}}(i,j)\})$$

and $S\{\mathrm{Strip}_{\prec,\underline{l}}(i,j)\}$ is the state information for $Y\{\mathrm{Strip}_{\prec,\underline{l}}(i,j)\}$, meaning the subset of $\{X[i,j]\}$ that is related to $Y\{\mathrm{Strip}_{\prec,\underline{l}}(i,j)\}$ via the transfer function $h[i,j]$.

Proof: The proof is very similar to its one-dimensional counterpart (see, e.g., [13]). Note that $\mathrm{Past}_{\prec,\underline{l}}(i,j) \subset \mathrm{Past}_{\prec,\underline{\infty}}(i,j)$ for every $\prec$. Applying Lemma 2.1 and the fact that conditioning reduces entropy, we have

$$H(\mathcal{Y}) \leq H(Y[i,j]|Y\{\mathrm{Past}_{\prec,\underline{l}}(i,j)\}) = H^{U1}_{\prec,\underline{l}}.$$

Since this inequality holds for both orders, we can conclude that

$$H(\mathcal{Y}) \leq \min_{\prec} H^{U1}_{\prec,\underline{l}}.$$

The lower bound is based on the observation that given $S\{\mathrm{Strip}_{\prec,\underline{l}}(i,j)\}$, the channel output $Y[i,j]$ is independent of the elements in

$$Y\{\mathrm{Past}_{\prec,\underline{\infty}}(i,j)\} \setminus Y\{\mathrm{Past}_{\prec,\underline{l}}(i,j)\}$$

for every $\prec$. This Markov property combined with Lemma 2.1 allows us to follow a similar procedure as in [13, pp. 69-70] to derive the lower bound

$$\begin{aligned} H^{L1}_{\prec,\underline{l}} &= H(Y[i,j]|Y\{\mathrm{Past}_{\prec,\underline{\infty}}(i,j)\}, S\{\mathrm{Strip}_{\prec,\underline{l}}(i,j)\}) \\ &\leq H(Y[i,j]|Y\{\mathrm{Past}_{\prec,\underline{\infty}}(i,j)\}) = H(\mathcal{Y}). \end{aligned}$$ □

Although the conditional entropies used in the bounds above are on a rectangular array, similar bounds can be derived using conditional entropies on an array of a different geometrical shape. However, as will be shown in Section II-C, a rectangular array simplifies the numerical estimation of the bounds. We also note that the bounds derived here are very similar to those we presented in [17]. The new bounds are in fact tighter, and their derivation is simpler.

Given the bounds above and the obvious fact that $I(\mathcal{X},\mathcal{Y}) \leq \log|\mathcal{X}|$, we obtain the following bounds for the mutual information rate:

$$\begin{aligned} \max_{\prec \in \{R,C\}} \left\{H^{L1}_{\prec,\underline{l}}\right\} - \frac{1}{2}\log(\pi e N_0) &\leq I(\mathcal{X};\mathcal{Y}) \\ &\leq \min\left\{\min_{\prec \in \{R,C\}} H^{U1}_{\prec,\underline{l}} - \frac{1}{2}\log(\pi e N_0), \log|\mathcal{X}|\right\}. \end{aligned} \quad (5)$$

B. Convergence of the Bounds

Given the bounds on $H(\mathcal{Y})$ developed above, we would like to know if they converge to the true entropy rate as the size of the array increases. If they do, we would also be interested to know the rate of convergence. The convergence rate will determine the size of the array needed to achieve tight upper and lower bounds in our numerical estimation.

It is well known that the entropy rate of a stationary one-dimensional hidden Markov process can be bounded by the conditional entropies of a finite sequence. They converge to the true entropy rate [13]. Birch [18] also showed that if the transition probabilities of the underlying Markov process are all positive, these bounds converge exponentially fast as the length of the sequence increases. In this section, we analyze the convergence of the above bounds for two-dimensional hidden Markov fields. We will show that they converge to the true entropy rate and a lower bound on the convergence rate is provided.

We first prove a lemma on the monotonicity of the upper and lower bounds.

Lemma 2.2: For a stationary two-dimensional hidden Markov field $\mathcal{Y}$, the upper bound $H^{U1}_{\prec,\underline{l}}$ on its entropy rate is monotonically nonincreasing as the size of the array defined by the vector $\underline{l}$ increases; the lower bound $H^{L1}_{\prec,\underline{l}}$ is monotonically nondecreasing as the size of the array increases.

Proof: It is easy to see that if $\underline{l} \leq \underline{l}'$ (that is, $l_i \leq l'_i, \forall i$), $\mathrm{Past}_{\prec,\underline{l}}(i,j) \subseteq \mathrm{Past}_{\prec,\underline{l}'}(i,j)$. Therefore,

$$H(Y[i,j]|Y\{\mathrm{Past}_{\prec,\underline{l}}(i,j)\}) \geq H(Y[i,j]|Y\{\mathrm{Past}_{\prec,\underline{l}'}(i,j)\})$$

since conditioning reduces entropy. Thus, the upper bound is nonincreasing as $\underline{l}$ increases.

To prove the monotonic property of the lower bound, we first examine

$$H^{L1}_{R,\underline{l}} = H(Y[i,j]|Y\{\mathrm{Past}_{R,\underline{l}}(i,j)\}, S\{\mathrm{Strip}_{R,\underline{l}}(i,j)\}).$$

Let $\underline{l}_1 = [l_1 + \Delta_1, l_2, l_3, l_4]$, where Δ_1 is a positive integer. Then we have

$$\begin{aligned} &H\Big(Y[i,j]|Y\{\mathrm{Past}_{R,\underline{l}}(i,j)\}, S\{\mathrm{Strip}_{R,\underline{l}}(i,j)\}\Big) \\ &\quad = H\Big(Y[i,j]|Y\{\mathrm{Past}_{R,\underline{l}_1}(i,j)\}, \\ &\qquad\qquad S\{\mathrm{Strip}_{R,\underline{l}}(i,j)\}, S\{\mathrm{Strip}_{R,\underline{l}_1}(i,j)\}\Big) \\ &\quad \leq H\Big(Y[i,j]|Y\{\mathrm{Past}_{R,\underline{l}_1}(i,j)\}, S\{St_{R,\underline{l}_1}(i,j)\}\Big). \end{aligned} \quad (6)$$

The equality comes from the Markov property, namely, that $Y[i,j]$ is independent of

$$S\{\mathrm{Strip}_{R,\underline{l}_1}(i,j)\} \setminus S\{\mathrm{Strip}_{R,\underline{l}}(i,j)\}$$

and

$$Y\{\mathrm{Past}_{R,\underline{l}_1}(i,j)\} \setminus Y\{\mathrm{Past}_{R,\underline{l}}(i,j)\}$$

given $S\{\mathrm{Strip}_{R,\underline{l}}(i,j)\}$, and the inequality comes from the fact that conditioning reduces entropy. Following the same line of reasoning, we can show that

$$\begin{aligned} &H\Big(Y[i,j]|Y\{\mathrm{Past}_{R,\underline{l}}(i,j)\}, S\{\mathrm{Strip}_{R,\underline{l}}(i,j)\}\Big) \\ &\quad \leq H\left(Y[i,j]|Y\{\mathrm{Past}_{R,\underline{l}}(i,j)\}, S\left\{\mathrm{Strip}_{R,\underline{l}_m}(i,j)\right\}\right), \\ &\qquad\qquad \text{for } m = 3,4 \end{aligned} \quad (7)$$

where $\underline{l}_3 = [l_1, l_2, l_3 + \Delta_3, l_4], \underline{l}_4 = [l_1, l_2, l_3, l_4 + \Delta_4]$, and Δ_3, Δ_4 are positive integers. The same conclusion can be reached for the lower bound when the column-by-column ordering is used. □

Next we use Lemma 2.2 to derive a lower bound on the convergence rate of the bounds.

Theorem 2.2: For a stationary two-dimensional hidden Markov field $\mathcal{Y}$, a lower bound on the convergence rate of $H^{U1}_{\prec,\underline{l}}$ and $H^{L1}_{\prec,\underline{l}}$ is $O(1/l_{\min})$, where $l_{\min} = \min\{l_1, l_3, l_4\}$ for $\prec = R$ and $l_{\min} = \min\{l_1, l_2, l_3\}$ for $\prec = C$.

Proof: We use the bounds obtained with the ordering $\prec = R$ as the example. We order the elements in the set $Y\{(i,j) \cup \text{Past}_{R,\underline{l}}(i,j)\}$ row by row beginning from the upper left corner $(i - l_1, j - l_3)$. We then convert this two-dimensional set to a one-dimensional sequence Y_{S_m} of $(l_3 + l_4 + 1)l_1 + l_3 + 1$ elements, with Y_{S_1} corresponding to $Y[i - l_1, j - l_3]$ and $Y_{S_{(l_3+l_4+1)l_1+l_3+1}}$ corresponding to $Y[i,j]$. Now the chain rule can be applied to the joint entropy

$$H(Y[i,j], Y\{\text{Past}_{R,\underline{l}}(i,j)\}) = \sum_{m=1}^{(l_3+l_4+1)l_1+l_3+1} H\left(Y_{S_m} \middle| Y_{S_1}^{S_{m-1}}\right). \quad (8)$$

Following the same line of reasoning as in Section II-A, we can derive an alternative upper and lower bound on the entropy rate $H(\mathcal{Y})$

$$\frac{\sum_{m=1}^{(l_3+l_4+1)l_1+l_3+1} H\left(Y_{S_m} \middle| Y_{S_1}^{S_{m-1}}, S\{\text{Strip}_{R,\underline{l}}(i,j)\}\right)}{(l_3+l_4+1)l_1+l_3+1} \le H(\mathcal{Y}) \le \frac{\sum_{m=1}^{(l_3+l_4+1)l_1+l_3+1} H\left(Y_{S_m} \middle| Y_{S_1}^{S_{m-1}}\right)}{(l_3+l_4+1)l_1+l_3+1}.$$

We further note that

$$\begin{aligned}
&\frac{\sum_{m=1}^{(l_3+l_4+1)l_1+l_3+1} H\left(Y_{S_m} \middle| Y_{S_1}^{S_{m-1}}\right)}{(l_3+l_4+1)l_1+l_3+1} - \frac{\sum_{m=1}^{(l_3+l_4+1)l_1+l_3+1} H\left(Y_{S_m} \middle| Y_{S_1}^{S_{m-1}}, S\{\text{Strip}_{R,\underline{l}}(i,j)\}\right)}{(l_3+l_4+1)l_1+l_3+1} \\
&= \frac{1}{(l_3+l_4+1)l_1+l_3+1} \times I(S\{\text{Strip}_{R,\underline{l}}(i,j)\}; Y[i,j], Y\{\text{Past}_{R,\underline{l}}(i,j)\}) \\
&\le \frac{[n_1(n_2+l_3+l_4) + n_2(2l_1-1)]\log|\mathcal{X}|}{(l_3+l_4+1)l_1+l_3+1} \sim O\left(\frac{1}{l_n}\right), \quad (9)
\end{aligned}$$

where $l_n = \min\{l_1, l_3 + l_4\}$. The inequality comes from the fact that the mutual information is upper-bounded by the source entropy. As $\underline{l} \to \underline{\infty}$, the difference between the bounds goes to zero. Utilizing Lemma 2.2, we can show that

$$H(Y[i,j] | Y\{\text{Past}_{R,\underline{l}'}(i,j)\}) \le H\left(Y_{S_m} \middle| Y_{S_1}^{S_{m-1}}\right)$$

and

$$H(Y[i,j] | Y\{\text{Past}_{R,\underline{l}'}(i,j)\}, S\{\text{Strip}_{R,\underline{l}'}(i,j)\}) \ge H\left(Y_{S_m} \middle| Y_{S_1}^{S_{m-1}}, S\left\{\text{Strip}_{R,\underline{l}}(i,j)\right\}\right)$$

for every m satisfying $1 \le m \le (l_3 + l_4 + 1)l_1 + l_3 + 1$, where $\underline{l}' = [l_1, l_2, l'_3, l'_4]$ with $l'_3 > l_3 + l_4$ and $l'_4 > l_3 + l_4$. Therefore, the convergence rate of $H^{U1}_{\prec,\underline{l}'}$ and $H^{L1}_{\prec,\underline{l}'}$ is no slower than $O(\frac{1}{l'_{\min}})$. □

For certain channel models, we conjecture that these bounds actually converge exponentially fast. In Section V, we will see that for the 2×2 channel transfer function used in the simulation, we can obtain fairly tight upper and lower bounds with a small rectangular array.

C. Computing the SIR Bounds

1) Monte Carlo Approach: The bounds derived in Section II-A convert the problem of estimating the entropy of a large two-dimensional array to the problem of estimating conditional entropies over a smaller array. As suggested in [6], the conditional entropy

$$h(\mathcal{A}|\mathcal{B}) = -E[\log P(a|b)]$$

can be estimated by $-\frac{1}{N}\sum_{k=1}^{N} \log P(a^{(k)}|b^{(k)})$, where $a^{(k)}, b^{(k)}$ are the kth realizations of A and B, respectively. By the law of large numbers, the estimate converges to $h(\mathcal{A}|\mathcal{B})$ with probability 1 as $N \to \infty$. Therefore, we simulate the channel N times, each time with i.i.d. channel inputs $\{x[i,j]\}$, generating the corresponding channel outputs $\{y[i,j]\}$ on a small rectangular array. For each of these two-dimensional realizations, we calculate the conditional probabilities needed to estimate the conditional entropies above. The calculation of the conditional probability $P(A|B)$ is, in turn, converted to the calculation of the joint probability $P(A,B)$ and the probability $P(B)$.

Without loss of generality, we will consider the estimation of $H(Y[i,j] | Y\{\text{Past}_{C,\underline{l}}(i,j)\})$ and

$$H\Big(Y[i,j] | Y\{\text{Past}_{C,\underline{l}}(i,j)\}, S\{\text{Strip}_{C,\underline{l}}(i,j)\}\Big)$$

only in the remainder of this section. If the ordering is $\prec = R$, one can consider column vectors instead and follow the same procedure. We also assume that the inputs are i.i.d., equiprobable binary symbols.

In order to compute the joint probability of the two-dimensional arrays, we adapt the technique proposed in [6]–[8] for one-dimensional sequences, treating each row vector $\underline{Y}_i^-$ as a variable and calculating the joint probability $P(\underline{Y}^-_{i-l_1}, \underline{Y}^-_{i-l_1+1}, \ldots, \underline{Y}^-_{i+l_2})$ row by row, using the forward recursion of the BCJR algorithm.

Since each row vector is considered as a single variable, the number of states for each variable increases exponentially as the number of columns in the array increases. Therefore, this numerical scheme can only be used to calculate the probability of two-dimensional arrays with relatively few columns. This limitation, in turn, requires that the two-dimensional channel impulse response have a small region of support.

2) Computing the Upper Bound: For the upper bound on $H(\mathcal{Y})$, we compute the conditional probability of $P(y[i,j] | y\{\text{Past}_{C,\underline{l}}(i,j)\})$ by calculating $P(y[i,j], y\{\text{Past}_{C,\underline{l}}(i,j)\})$ and $P(y\{\text{Past}_{C,\underline{l}}(i,j)\})$, and then taking the quotient of the two.

For $r < i$, the state information that corresponds to the rth row $\underline{Y}_r^-$ is

$$S_r = X^{r,j}_{r-n_1+2, j-l_3-n_2+1}.$$

Applying the forward recursion of the BCJR algorithm, we have

$$\begin{aligned}
P\left(y^-_{(i-l_1)\to r}, S_r = q\right) &= \sum_{q' \Rightarrow q} P\left(y^-_{(i-l_1)\to(r-1)}, S_{r-1} = q'\right) P\left(\underline{y}^-_r | S_r = q, S_{r-1} = q'\right) \\
&\quad \cdot P(S_r = q | S_{r-1} = q') \quad (10)
\end{aligned}$$

where $y^-_{(i-l_1)\to r} = \{\underline{y}^-_{i-l_1}, \ldots, \underline{y}^-_r\}$, and $q' \Rightarrow q$ denotes the event that state $S_{r-1} = q'$ can transition to $S_r = q$ with nonzero probability. When $q' \Rightarrow q$, the conditional probability

$$P(S_r = q | S_{r-1} = q') = \frac{1}{2^{l_3+n_2}}$$

since the inputs are i.i.d., equiprobable binary symbols. The probability $P(\underline{y}_r^-|S_r = q, S_{r-1} = q')$ admits a joint Gaussian distribution

$$P(\underline{y}_r^-|S_r = q, S_{r-1} = q') = \frac{1}{(2\pi\sigma^2)^{(l_3+1)/2}} \cdot \exp\left\{ -\frac{1}{2\sigma^2} \sum_{m=j-l_3}^{j} \left(y[r,m] - \sum_{p=0}^{n_1-1} \sum_{s=0}^{n_2-1} h[p,s]x[r-p,m-s] \right)^2 \right\}.$$

The initial condition of the recursion is

$$\begin{aligned} P\left(y^-_{(i-l_1)\to(i-l_1)}, S_{i-l_1} = q\right) &= \sum_{q'\Rightarrow q} P\left(\underline{y}^-_{i-l_1}|S_{i-l_1} = q, S_{i-l_1-1} = q'\right) \cdot P(S_{i-l_1} = q|S_{i-l_1-1} = q')P(S_{i-l_1-1} = q') \\ &= \frac{1}{2^{(l_3+n_2)(n_1-1)}} \frac{1}{(2\pi\sigma^2)^{\frac{l_3+1}{2}}} \cdot \exp\left\{ -\frac{1}{2\sigma^2} \sum_{m=j-l_3}^{j} \left(y[i-l_1,m] - \sum_{p=0}^{n_1-1} \sum_{s=0}^{n_2-1} h[p,s]x[i-l_1-p,m-s] \right)^2 \right\} \end{aligned}$$

where

$$P(S_{i-l_1-1} = q') = \frac{1}{2^{(l_3+n_2)(n_1-1)}}.$$

The difference between the calculation of $P(y[i,j], y\{\text{Past}_{C,\underline{l}}(i,j)\})$ and that of $P(y\{\text{Past}_{C,\underline{l}}(i,j)\})$ arises from the difference between the sets of the outputs in the ith row. For $P(y[i,j], y\{\text{Past}_{C,\underline{l}}(i,j)\})$, the ith row consists of the elements $Y_{i,j-l_3}^{i,j}$ while $Y[i,j]$ is excluded from that row in the calculation of $P(y\{\text{Past}_{C,\underline{l}}(i,j)\})$. We only need to make a minor modification of the recursive formula (10) to accommodate this difference. When $r > i$, each row only has l_3 elements $Y_{r,j-l_3}^{r,j-1}$. Equation (10) can be modified in the same way.

Finally, we obtain the joint probability $P(y[i,j], y\{\text{Past}_{C,\underline{l}}(i,j)\})$ and the probability $P(y\{\text{Past}_{C,\underline{l}}(i,j)\})$, namely

$$P(y[i,j], y\{\text{Past}_{C,\underline{l}}(i,j)\}) = \sum_q P(y[i,j], y\{\text{Past}_{C,\underline{l}}(i,j)\}, S_{i+l_2} = q)$$

and

$$P(y\{\text{Past}_{C,\underline{l}}(i,j)\}) = \sum_q P(y\{\text{Past}_{C,\underline{l}}(i,j)\}, S_{i+l_2} = q).$$

To increase numerical accuracy, a normalization scheme proposed in [6] may be applied.

3) Computing the Lower Bound: The calculation of the lower bound

$$H\left(Y[i,j]|Y\{\text{Past}_{C,\underline{l}}(i,j)\}, S\{\text{Strip}_{C,\underline{l}}(i,j)\}\right)$$

is similar to that of the upper bound. The state information $S\{\text{Strip}_{C,\underline{l}}(i,j)\}$ is

$$S\{\text{Strip}_{C,\underline{l}}(i,j)\} = X\left\{ L_{i-l_1-n_1+1,j-l_3-n_2+1}^{i-l_1,j} \cup L_{i-l_1+1,j-l_3-n_2+1}^{i+l_2-n_1,j-l_3} \cup L_{i+l_2-n_1+1,j-l_3-n_2+1}^{i+l_2,j-1} \right\}.$$

We again calculate the joint probability row by row with the following initial condition:

$$P(S_{i-l_1-1} = q) = \begin{cases} 1, & \text{if } q = X_{i-l_1-n_1+1,j-l_3-n_2+1}^{i-l_1-1,j} \\ 0, & \text{else.} \end{cases} \tag{11}$$

The recursion is similar to (10) except that we need to take care of the additional state information contained in $S\{\text{Strip}_{C,\underline{l}}(i,j)\}$. That is, each row vector in the calculation of the forward recursion includes both the channel output in that row and the additional state information. For every row r in the array, there is additional state information from $X_{r,i-l_1-n_2+1}^{r,i-l_1}$. For rows $i-l_1$ to $i-l_1+n_1-1$ and rows $i+l_2-n_1+1$ to $i+l_2$, the effects of $X_{i-l_1-n_1+1,j-l_3-n_2+1}^{i-l_1,j}$ and $X_{i+l_2-n_1+1,j-l_3-n_2+1}^{i+l_2,j-1}$ also need to be considered, respectively. Besides, we still need to pay special attention to the difference of the two sets

$$\{y[i,j], y\{\text{Past}_{C,\underline{l}}(i,j)\}, s\{\text{Strip}_{C,\underline{l}}(i,j)\}\}$$

and

$$\{y\{\text{Past}_{C,\underline{l}}(i,j)\}, s\{\text{Strip}_{C,\underline{l}}(i,j)\}\}$$

beginning from the ith row when calculating

$$P(y[i,j], y\{\text{Past}_{C,\underline{l}}(i,j)\}, s\{\text{Strip}_{C,\underline{l}}(i,j)\})$$

and

$$P(y\{\text{Past}_{C,\underline{l}}(i,j)\}, s\{\text{Strip}_{C,\underline{l}}(i,j)\}).$$

Although there are some subtle details in the lower bound estimation, the modification of the BCJR recursion formula is rather straightforward.

Since the bounds in Section II-A are given for both orders, we can calculate two upper bounds and two lower bounds to make the overall bounds tighter, if the computation in each order is feasible. In some cases, we can only calculate the bound for a specific ordering due to complexity constraints. For example, we can only calculate the bounds for the column-by-column ordering if n_1 is small, but n_2 is large. We also note that if the impulse response is symmetric, i.e., $h[i,j] = h[j,i]$, it is sufficient to calculate only one of the two conditional entropies in (5) for the upper and the lower bounds.

III. An Alternative Upper Bound

As shown in the preceding section, the modified BCJR algorithm allows us to calculate the probability of an output array whose size can be extended in one dimension, but not both, since the modified algorithm is still one-dimensional in nature. Thus, it only allows us to compute the symmetric information rate for a channel with a small support region. This constraint motivates us to search for other approaches to further reduce the computational complexity, albeit at the cost of weakening the bound.

In this section, we present an alternative upper bound which relies upon an auxiliary ISI channel in the evaluation of the upper bound on the channel output entropy rate. We have freedom in selecting the memory length and the dimensionality of this auxiliary channel, parameters that determine the complexity of this approach. This upper bound is an extension of the upper bound proposed in [19] for the evaluation of the mutual information rate of one-dimensional ISI channels. The main differences between the upper bound here and that in [19] are the following. First, the upper bound here is for the output entropy rate of a two-dimensional ISI channel, while [19] focuses on the upper-bounding of the mutual information rate of a one-dimensional channel. We therefore have more freedom in choosing the auxiliary channel model, which can be either one-dimensional or two-dimensional. Second, we use the conditional probability of a finite sequence (in the case of a one-dimensional auxiliary channel model) or finite array (in the case of a two-dimensional auxiliary channel model) instead of the joint probability of a very long sequence or large array.

Theorem 3.1: For a stationary two-dimensional random field $\mathcal{Y}$, an upper bound on its entropy rate is

$$H(\mathcal{Y}) \leq H^{U2}_{\prec,\underline{l}}$$

where

$$H^{U2}_{\prec,\underline{l}} = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} -p(y[i,j], y\{\text{Past}_{\prec,\underline{l}}(i,j)\}) \cdot \log q(y[i,j] | y\{\text{Past}_{\prec,\underline{l}}(i,j)\}) d\underline{y}$$

and $q(y[i,j]|y\{\text{Past}_{\prec,\underline{l}}(i,j)\})$ is an arbitrary conditional probability distribution.

Proof: The proof combines the proof technique in [19] and Theorem 2.1. It is clear that

$$\begin{aligned} H^{U1}_{\prec,\underline{l}} - H^{U2}_{\prec,\underline{l}} &= -\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} p(y[i,j], y\{\text{Past}_{\prec,\underline{l}}(i,j)\}) \\ &\quad \cdot \log \frac{p(y[i,j]|y\{\text{Past}_{\prec,\underline{l}}(i,j)\})}{q(y[i,j]|y\{\text{Past}_{\prec,\underline{l}}(i,j)\})} d\underline{y} \\ &= -D(p(Y[i,j]|Y\{\text{Past}_{\prec,\underline{l}}(i,j)\}) \| \\ &\quad q(Y[i,j]|Y\{\text{Past}_{\prec,\underline{l}}(i,j)\})) \leq 0. \end{aligned} \tag{12}$$

Combining (12) and Theorem 2.1, we can obtain the upper bound. □

The proof shows that to get a tight upper bound, we need to choose a large array size (i.e., the elements in $\underline{l}$ should be large), and $q(Y[i,j]|Y\{\text{Past}_{\prec,\underline{l}}(i,j)\})$ should approximate $p(Y[i,j]|Y\{\text{Past}_{\prec,\underline{l}}(i,j)\})$ as closely as possible.

Theorem 3.1 converts the problem of calculating the conditional probability $p(Y[i,j]|Y\{\text{Past}_{\prec,\underline{l}}(i,j)\})$ of a two-dimensional channel output array to that of calculating the conditional probability $q(Y[i,j]|Y\{\text{Past}_{\prec,\underline{l}}(i,j)\})$ of an array generated by an arbitrary source. It gives us the freedom to explore various approaches in selecting the source. The most promising one proposed by Arnold *et al.* [19] introduces the concept of an auxiliary channel. The source is assumed to be the output of an auxiliary one-dimensional ISI channel. If the divergence of the corresponding probability distribution and the distribution of the original channel output is small, it will be a good bound. If we apply the auxiliary channel approach, the computation procedure is as follows.

1. Choose the auxiliary ISI channel impulse response.
2. Generate an output array $\{y[i,j], y\{\text{Past}_{\prec,\underline{l}}(i,j)\}\}$ from the original two-dimensional ISI channel model with i.i.d channel input. Calculate $-\log q(y[i,j]|y\{\text{Past}_{\prec,\underline{l}}(i,j)\})$ assuming that the array was generated from the auxiliary channel with the same kind of i.i.d. input.
3. Repeat the same process N times and calculate the average of the N realizations of $-\log q(y[i,j]|y\{\text{Past}_{\prec,\underline{l}}(i,j)\})$.

We have two classes of auxiliary channels to choose from—one-dimensional auxiliary channels and two-dimensional auxiliary channels. If we choose a two-dimensional auxiliary channel, the support of the auxiliary channel impulse response and the size of the overall output array must be small in order to limit the computational complexity. Having specified these quantities, we can then choose the impulse response coefficients and proceed according to the procedure specified above. Since we use a different channel transfer function to approximate the output of the original channel, it is not clear how to choose the auxiliary channel impulse response appropriately. One approach is to use existing optimization techniques such as differential evolution [20] to find the optimal auxiliary channel impulse response. Of course, this incurs additional computational cost to find the upper bound. To ensure a fair comparison, we normalize both the original and the auxiliary channel transfer functions to have unit energy; thus,

$$\sum_{j=0}^{n_1-1} \sum_{k=0}^{n_2-1} h^2[j,k] = 1.$$

Since the complexity of this approach still depends upon the output array size, it is only applicable to channels with support regions that are slightly larger than those that can be handled directly with the techniques of Section II. With only a few columns in the output array, it is very difficult to carry out the calculations for channels with large support regions. In order to remove the constraint on the number of columns in the output array, we instead consider the use of one-dimensional auxiliary channels. One immediate problem associated with this approach is how to convert a two-dimensional array to a one-dimensional sequence in such a way that the critical statistical properties of the array are captured by the sequence. To solve this problem, we propose the use of the Peano–Hilbert plane-filling curve [21]. It has the property that if a square is mapped to an interval, its subsquares are mapped to the subintervals of that interval. Lempel and Ziv [22] showed that a compression scheme for two-dimensional arrays which uses the Peano–Hilbert curve combined with a compression scheme for one-dimensional sequences [23] is asymptotically optimal. Their result indicates that this dimensional transformation does preserve the statistical properties that are essential for the purpose of entropy estimation.

The Peano–Hilbert curve exists only for squares with size $2^k \times 2^k$. Unless we set $\underline{l} = [2^k - 1, 0, 2^k - 1, 0]$, this type of curve cannot be directly applied in our calculation. However, for more general rectangular arrays, one can construct related curves known as pseudo-Peano–Hilbert curves (see, e.g., [24]). Although these curves may lose some statistical information, we expect them to provide reasonably accurate entropy estimates, as has been observed empirically. Fig. 3 shows an example of using a pseudo-Peano–Hilbert curve to convert a two-dimensional array $\{Y[i,j], Y\{\text{Past}_{C,[7,8,7,l_4]}(i,j)\}\}$ into a one-dimensional sequence. One important requirement for such a curve is that the output array element $Y[i,j]$ should be at the end of the curve. It is also desirable to have the curve differ from a standard Peano–Hilbert curve only near the beginning, as it will then lead to a conditional probability $q(Y[i,j]|Y\{\text{Past}_{\prec,\underline{l}}(i,j)\})$ that better approximates the actual conditional probability of the array. In the figure, only the beginning portion of the curve differs from a Peano–Hilbert curve, and the affected portion lies relatively far away from the array element $Y[i,j]$.

The auxiliary channel approach, as illustrated above with a pseudo-Peano–Hilbert curve, makes it possible to explore additional trade-offs between complexity and accuracy in estimating entropy rate upper bounds for two-dimensional ISI channels, irrespective of the size of the impulse response support region. Although this bound is likely to be looser than that described in Section II for any given array size, the pseudo-Peano–Hilbert curve enables us to compute the bound using a much larger array. Since the alternative upper bound becomes tighter when the array size is increased, it is possible that for certain two-dimensional ISI channels we can actually determine, for a specified computational complexity, a tighter bound with this approach than with the method of Section II by using a much larger array size. The alternative approach is especially helpful when the ISI channel has a large impulse response support region; in such a case, we simply cannot compute the upper bound described in Section II due to the exponential growth in complexity as a function of the support region size.

Finally, we mention two reasons for basing the upper bound on a conditional probability, rather than on a joint probability over a very large two-dimensional array, which would seem to be the natural direct extension of the approach in [19]. First, it is not clear if the limit

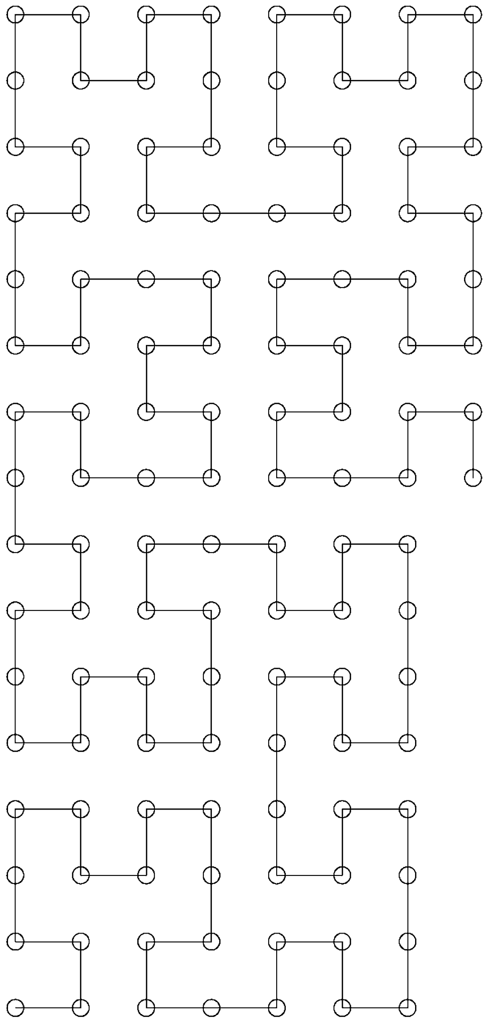

Fig. 3. Pseudo-Peano–Hilbert curve for the SIR upper bound calculation.

exists for such a two-dimensional joint probability. Arnold *et al.* listed several sufficient conditions which guarantee the convergence of their approach [25]. Unfortunately, our system does not satisfy these conditions. Second, even if the limit does exist, the joint probability approach may not provide much gain compared to the bound $H^{U2}_{\prec,\underline{l}}$. As will be seen in Section V, the memory length of the one-dimensional auxiliary channel will be a limiting factor on the tightness of the bound. The accuracy of the joint probability bound would suffer a similar sort of limitation.

One can also develop an alternative lower bound based on the auxiliary channel approach in [19]. However, it cannot be numerically estimated by the Monte Carlo approach, as we now discuss.

The auxiliary channel approach in [19] is applicable only to lower bounds on the channel mutual information rate. Therefore, we must first convert our lower bound on the channel output entropy rate to a lower bound on the mutual information rate. We note that

$$\begin{aligned}H(Y[i,j]|Y\{\mathrm{Past}_{\prec,\underline{l}}(i,j)\}, S\{\mathrm{Strip}_{\prec,\underline{l}}(i,j)\}, S\{(i,j)\})\\ = H(Y[i,j]|S\{(i,j)\}) = H(\mathcal{N})\end{aligned}$$

where $S\{(i,j)\}$ is the state information for $Y[i,j]$. The first equality comes from the Markov property, which implies that, given $S\{(i,j)\}$, the output element $Y[i,j]$ is independent of any other channel output or state information. Thus, a lower bound on the mutual information rate of the two-dimensional ISI channel is

$$I(\mathcal{X};\mathcal{Y}) \geq I^{L1}_{\prec,\underline{l}}$$

where

$$I^{L1}_{\prec,\underline{l}} = I(S\{(i,j)\}; Y[i,j]|Y\{\mathrm{Past}_{\prec,\underline{l}}(i,j)\}, S\{\mathrm{Strip}_{\prec,\underline{l}}(i,j)\}).$$

Applying the technique in [19], we can show that

$$I^{L2}_{\prec,\underline{l}} \leq I^{L1}_{\prec,\underline{l}} \leq I(\mathcal{X};\mathcal{Y})$$

where

$$\begin{aligned}I^{L2}_{\prec,\underline{l}} = &\sum_{s\{(i,j)\}, s\{\mathrm{Strip}_{\prec,\underline{l}}(i,j)\}} \int_{-\infty}^{\infty}\cdots\int_{-\infty}^{\infty}\\ &p(y[i,j], y\{\mathrm{Past}_{\prec,\underline{l}}(i,j)\} s\{(i,j)\}, s\{\mathrm{Strip}_{\prec,\underline{l}}(i,j)\})\\ &\cdot \log \frac{q(y[i,j]|s\{(i,j)\})}{q(y[i,j]|y\{\mathrm{Past}_{\prec,\underline{l}}(i,j)\}, s\{\mathrm{Strip}_{\prec,\underline{l}}(i,j)\})} d\underline{y}\\ &- D(p(S\{(i,j)\}|Y\{\mathrm{Past}_{\prec,\underline{l}}(i,j)\}, S\{\mathrm{Strip}_{\prec,\underline{l}}(i,j)\})\|\\ &q(S\{(i,j)\}|Y\{\mathrm{Past}_{\prec,\underline{l}}(i,j)\}, S\{\mathrm{Strip}_{\prec,\underline{l}}(i,j)\}))\end{aligned}$$

and the distribution $q(\cdot)$ comes from the auxiliary channel. Other than the additional conditioning, the main difference between this bound and the one in [19] is the second term. This new term arises from the conditioning introduced in the original lower bound in Section II-A. Without this conditioning, the two distributions $p(S\{(i,j)\})$ and $q(S\{(i,j)\})$ are the same, and the second term is equal to 0. With the conditioning, the conditional state distributions also depend on the underlying channel models and are generally different. Thus, the second term becomes positive and needs to be evaluated numerically. While the distribution $q(S\{(i,j)\}|Y\{\mathrm{Past}_{\prec,\underline{l}}(i,j)\}, S\{\mathrm{Strip}_{\prec,\underline{l}}(i,j)\})$ can be evaluated by choosing an appropriate auxiliary channel to reduce the complexity, the evaluation of $p(S\{(i,j)\}|Y\{\mathrm{Past}_{\prec,\underline{l}}(i,j)\}, S\{\mathrm{Strip}_{\prec,\underline{l}}(i,j)\})$ is much more complex and cannot be simplified by use of the auxiliary channel. This problem once again illustrates the difficulties inherent in analyzing two-dimensional channels.

IV. Extension to d-Dimensional ISI Channels

The bounds we have developed so far apply to two-dimensional ISI channels. These results are in fact more general and can be extended to d-dimensional ISI channels where $d \geq 2$, as we now briefly describe.

Given a d-dimensional space Z^d, we consider a generalized lexicographical order $\prec_\kappa$ which we refer to as the permuted lexicographical order with permutation vector κ. The length-d vector $\kappa = [k_1, \ldots, k_d]$ represents a permutation of the elements $\{1, 2, \ldots, d\}$. The ordering of two points $(t_1, \ldots, t_d)$ and $(t'_1, \ldots, t'_d)$ with respect to $\prec_\kappa$ is induced by the standard d-dimensional lexicographical ordering of the points $(t_{k_1}, \ldots, t_{k_d})$ and $(t'_{k_1}, \ldots, t'_{k_d})$. Specifically, $(t_1, \ldots, t_d) \prec_\kappa (t'_1, \ldots, t'_d)$ if $(t_{k_1}, \ldots, t_{k_d})$ precedes $(t'_{k_1}, \ldots, t'_{k_d})$ in the standard d-dimensional lexicographical order. In total, there are $d!$ possible orders. When $d = 2$, there are two possible permutation vectors $\kappa_R = [1, 2]$ and $\kappa_C = [2, 1]$. It is easy to verify that the permuted lexicographical order with permutation vector $\kappa_R = [1, 2]$ corresponds to the row-by-row order, and the permuted lexicographical order with permutation vector $\kappa_C = [2, 1]$ corresponds to the column-by-column order.

Let $\underline{l}$ be a nonnegative vector of length $2d$. We denote by $\mathrm{Past}_{\prec_\kappa,\underline{l}}(i_1, \ldots, i_d)$ the subset of positions inside the d-dimensional subspace defined by the 2^d vertices $(t_1, \ldots, t_d)$ that precede

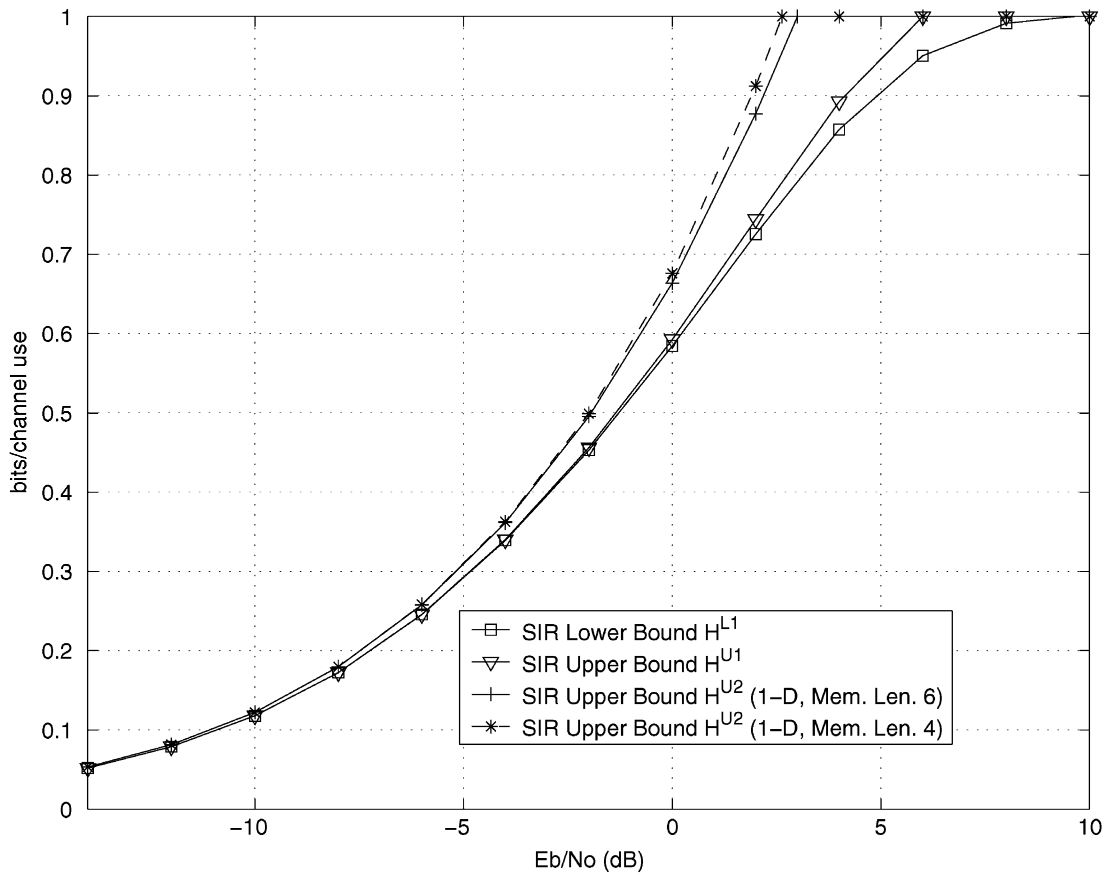

Fig. 4. Upper and lower bounds on the SIR of the two-dimensional ISI channel h_1.

$(i_1, \ldots, i_d)$ according to the order $\prec_\kappa$, where $t_k = i_k - l_{2k-1}$ or $t_k = i_k + l_{2k}$.

With this modified notation, we can see that Lemma 2.1 is applicable to Z^d, as was shown in [14]. Therefore, Theorem 2.1 follows immediately. The only modification necessary relates to the definition of $\mathrm{Strip}_{\prec_\kappa, \underline{l}}$ — rather than consist of one-dimensional rows and columns on the boundary of a two-dimensional array; it instead consists of $(d-1)$-dimensional layers on the boundary of the d-dimensional array $\mathrm{Past}_{\prec_\kappa, \underline{l}}(i_1, \ldots, i_d)$. The convergence results remain valid and the lower bound of the convergence rate is still $O(1/l_{\min})$ where $l_{\min}$ is the minimum of the l_i's.

When $d > 2$, the computational complexity of the bounds in Theorem 2.1 becomes enormous as we need to treat a $(d-1)$-dimensional layer as a single state in the application of the BCJR algorithm. In this case, the alternative upper bound derived in Section III becomes much more feasible to compute. The Peano–Hilbert curve and pseudo-Peano–Hilbert curves can be extended to Z^d [21], [24], so we can still obtain an upper bound with a one-dimensional auxiliary ISI channel.

V. Numerical Results

In this section, we present numerical results for the upper and lower bounds on the SIR of a two-dimensional ISI channel. We begin with a channel impulse response of size 2×2, namely

$$h_1 = \begin{bmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{bmatrix}.$$

The coefficients of the impulse response are chosen to cause a significant amount of ISI. In a channel with little ISI, we expect our bounds to be tight. In choosing this impulse response, our intention is to test the tightness of the bounds in a "worst case" scenario. The channel inputs are i.i.d., equiprobable Bernoulli samples over the input alphabet $\{1, -1\}$; thus, each channel input symbol has energy $E_b = 1$. The noise variance is $N_0/2$. For each E_b/N_0 point, the corresponding conditional entropies are computed from 100000 output array realizations.

In particular, for the SIR upper bound, we calculate

$$\min\left\{H^{U1}_{C,[7,7,3,0]} - \frac{1}{2}\log(\pi e N_0), 1\right\}$$

using the approach described in Section II-C2. The SIR lower bound computation uses the method of Section II-C3 to determine $H^{L1}_{C,[7,7,3,0]} - \frac{1}{2}\log(\pi e N_0)$. In these calculations, we use the conditional entropy corresponding to the output in the middle of the last column, because the state information is expected to have the least impact on the result, resulting in a lower bound closer to the upper bound.

Fig. 4 shows the numerical results. It can be seen that, despite the use of a relatively small sample array, the upper and lower bounds are still very close, particularly at low signal-to-noise ratios. The maximum difference is approximately 0.05 bits per channel use. The discrepancy at higher values of E_b/N_0 reflects the larger impact of the ISI on the conditional entropies and, therefore, on the resulting tightness of the bounds. The small difference between the bounds based upon such a small array suggests that the actual convergence rate of the bounds for this channel may be faster than $O(1/l_{\min})$. We conjecture that they may converge exponentially fast, as in the case described in [18].

In the same figure, we also show the numerical results for $H^{U2}_{C,[7,8,7,0]}$ when applied to this channel model. The results are obtained with one-

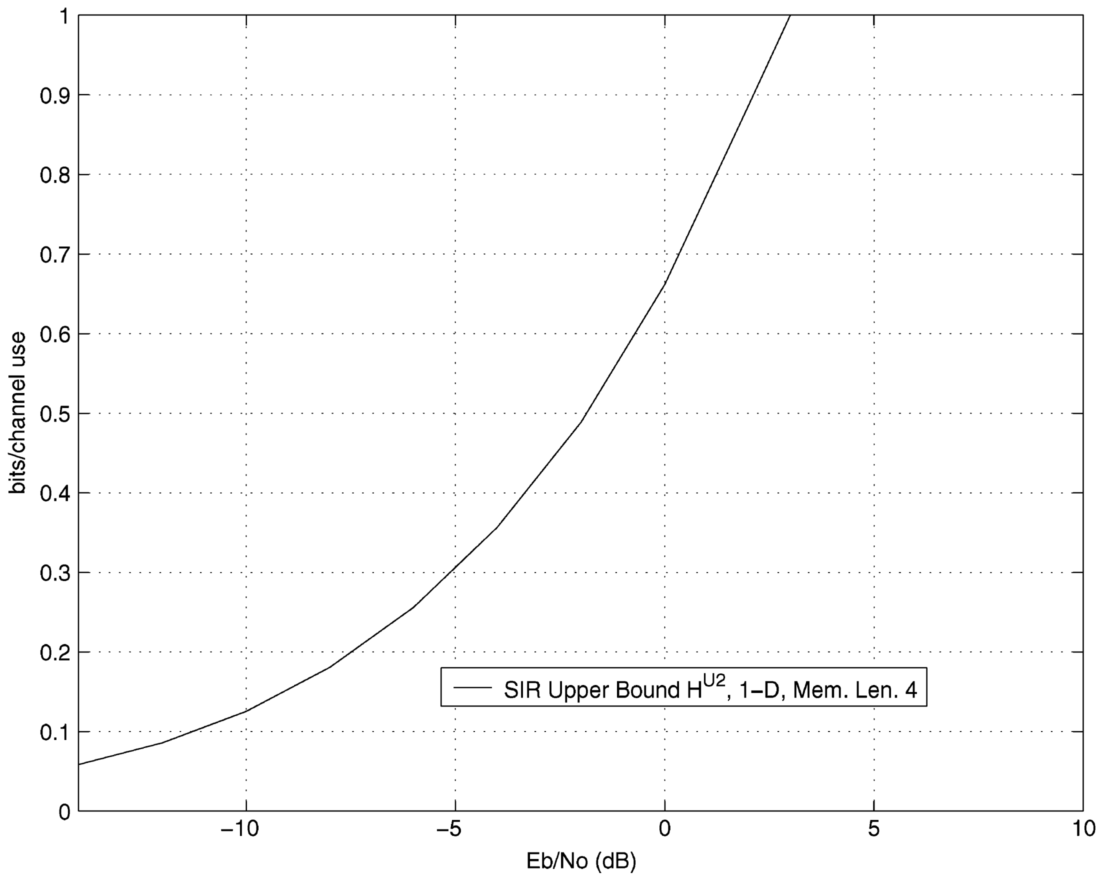

Fig. 5. Upper bound on the SIR of the two-dimensional ISI channel h_2.

dimensional auxiliary ISI channels of memory length 4 and 6 respectively, whose transfer functions are found via differential evolution. The array size and the corresponding pseudo-Peano-Hilbert curve are shown in Fig. 3. Although we almost double the size of the array, the alternative upper bound is still looser than the other upper bound. Furthermore, increasing the memory length only brings minor improvement at relatively high signal-to-noise ratios. This indicates that we need a one-dimensional auxiliary channel with much longer memory to closely approximate the output of the target two-dimensional ISI channel, even though the size of the two-dimensional impulse response is only 2×2. In this case, the increased array size cannot compensate for the looseness of $H^{U2}_{C,[7,8,7,0]}$, underscoring the limitations of using one-dimensional ISI channels to analyze two-dimensional systems. We conclude that it is preferable to estimate the upper bound in Section 2 directly for this channel, as long as the complexity is acceptable.

However, for a two-dimensional channel whose transfer function size is too large to allow numerical estimation of $H^{U1}_{\prec,\underline{l}}$, the bound $H^{U2}_{\prec,\underline{l}}$ does provide a way to estimate an upper bound on the SIR, although the resulting bound may be loose. To illustrate this point, we estimate the upper bound $H^{U2}_{C,[7,8,7,0]}$ on the SIR of a two-dimensional channel with a 3×3 transfer function given by

$$h_2 = \frac{1}{\sqrt{10}} \begin{bmatrix} 0 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 0 \end{bmatrix}.$$

This channel model is equivalent to one that arises in the context of a two-dimensional optical storage system that records data on a hexagonal lattice [26]. In that system, the readback signal corresponding to a specified position may contain interference from the six immediate neighbors of the position. As noted earlier, the hexagonal lattice can be mapped to a square lattice, and the transfer function above is equivalent under this mapping to that of the system based upon the hexagonal lattice. Fig. 5 presents the numerical upper bound obtained with a one-dimensional auxiliary ISI channel of memory length 4. Similar to the previous figure, we obtain a useful upper bound (i.e., a bound less than 1 bit/channel use) up to $E_b/N_0 = 3$ dB.

VI. Concluding Remarks

We develop upper and lower bounds on the SIR of two-dimensional finite-state ISI channels. These bounds are expressed in terms of conditional entropies of channel output arrays. Their convergence properties are analyzed. We also describe a method that can be used to compute these conditional entropies when the array size is not too large. To relax this constraint, we propose an alternative upper bound which uses an auxiliary channel to explore more fully the possible tradeoffs between the tightness of bounds and their computational complexity.

We then present numerical results for a channel with 2×2 impulse response support, demonstrating that we can obtain fairly tight bounds on the SIR using conditional entropies on an array which is only a few times larger than the support of the impulse response. With the same channel model, we show that the alternative upper bound $H^{U2}_{\prec,\underline{l}}$ is not as tight as $H^{U1}_{\prec,\underline{l}}$ when the memory length of the one-dimensional auxiliary channel is small. However, we also show that the alternative approach does provide a feasible way to numerically estimate bounds on the symmetric information rate for channels with large impulse response support regions. We conclude that, despite the usefulness of the upper and lower bounds derived here, further research is required to develop improved, computable bounds on the SIR for general two-dimensional ISI channels.

REFERENCES

[1] D. Psaltis and F. Mok, "Holographic memories," *Scient. Amer.*, pp. 70–76, Nov. 1995.

[2] W. Hirt, "Capacity and Information Rates of Discrete-Time Channels with Memory," Ph.D. dissertation (dissertation ETH 8671), Swiss Federal Inst. Tech. (ETH), Zurich, Switzerland, 1988.

[3] W. Hirt and J. L. Massey, "Capacity of the discrete-time Gaussian channel with intersymbol interference," *IEEE Trans. Inf. Theory*, vol. 34, no. 3, pp. 380–388, May 1988.

[4] S. Shamai (Shitz), L. H. Ozarow, and A. D. Wyner, "Information rates for a discrete-time Gaussian channel with intersymbol interference and stationary inputs," *IEEE Trans. Inf. Theory*, vol. 37, no. 6, pp. 1527–1539, Nov. 1991.

[5] S. Shamai (Shitz) and R. Laroia, "The intersymbol interference channel: Lower bounds on capacity and channel precoding loss," *IEEE Trans. Inf. Theory*, vol. 42, no. 5, pp. 1388–1404, Sep. 1996.

[6] D. Arnold and H. Loeliger, "On the information rate of binary-input channels with memory," in *Proc. IEEE Int. Conf. Communications*, vol. 9, Helsinki, Finland, Jun. 2001, pp. 2692–2695.

[7] V. Sharma and S. K. Singh, "Entropy and channel capacity in the regenerative setup with applications to Markov channels," in *Proc. IEEE Int. Symp. Information Theory*, Washington, DC, Jun. 2001, p. 283.

[8] H. D. Pfister, J. B. Soriaga, and P. H. Siegel, "On the achievable information rate of finite state ISI channels," in *Proc. GLOBECOM 2001*, vol. 5, San Antonio, TX, Nov. 2001, pp. 2992–2996.

[9] Z. Zhang, T. M. Duman, and E. Kurtas, "On information rates of single track and multi-track magnetic recording channels in intertrack interference," in *Proc. IEEE Int. Symp. Information Theory*, Lausanne, Switzerland, Jun./Jul. 2002, p. 163.

[10] 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.

[11] H. O. Georgii, *Gibbs Measures and Phase Transitions*. Berlin, Germany: DeGruyter, 1988, Studies in Mathematics 9.

[12] J. J. Ashley, M. Blaum, and B. H. Marcus, "Report on Coding Techniques for Holographic Storage," IBM Res. Rep., RJ-10013, Mar. 1996.

[13] T. M. Cover and J. A. Thomas, *Elements of Information Theory*: Wiley, 1991.

[14] Y. Katznelson and B. Weiss, "Commuting measure-preserving transformations," *Israel J. Math.*, vol. 12, pp. 161–173, 1972.

[15] D. Anastassiou and D. J. Sakrison, "Some results regarding the entropy rates of random fields," *IEEE Trans. Inf. Theory*, vol. IT-28, no. 2, pp. 340–343, Mar. 1982.

[16] Z. Kukorelly and K. Zeger, "The capacity of some hexagonal (d, k)-constraints," in *Proc. IEEE Int. Symp. Information Theory*, Washington, DC, Jun. 2001, p. 64.

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

[18] J. J. Birch, "Approximations for the entropy for functions of Markov chains," *Ann. Math. Statist.*, vol. 33, no. 3, pp. 930–938, Sept. 1962.

[19] D. Arnold, H. Loeliger, and P. O. Vontobel, "Computation of information rates from finite state source/channel models," in *Proc. 40th Annu. Allerton Conf. Communications, Control, and Computing*, Monticello, IL, Oct. 2002, pp. 457–466.

[20] K. Price and R. Storn, "Differential evolution—A simple and efficient heuristic for global optimization over continuous spaces," *J. Global Optimiz.*, vol. 11, pp. 341–359, Dec. 1997.

[21] H. Sagan, *Space-Filling Curves*. New York: Springer-Verlag, 1991.

[22] A. Lempel and J. Ziv, "Compression of two-dimensional data," *IEEE Trans. Inf. Theory*, vol. IT-32, no. 1, pp. 2–8, Jan. 1986.

[23] J. Ziv and A. Lempel, "Compression of individual sequences via variable-rate coding," *IEEE Trans. Inf. Theory*, vol. IT-24, no. 5, pp. 530–536, Sep. 1978.

[24] Y. Bandoh and S. Kamata, "An address generator for an N-dimensional pseudo-Hilbert scan in a hyper-rectangular parallelepiped region," in *Proc. Int. Conf. Image Processing*, vol. 1, Vancouver, BC, Canada, Sep. 2000, pp. 737–740.

[25] D. Arnold, H. Loeliger, P. O. Vontobel, A. Kavčić, and W. Zeng, "Simulation-based computation of information rates for channels with memory," *IEEE Trans. Inf. Theory*, submitted for publication.

[26] A. H. J. Immink, W. M. J. Coene, A. M. van der Lee, C. Busch, A. P. Hekstra, J. W. M. Bergmans, J. Riani, S. J. L. v. Beneden, and T. Conway, "Signal processing and coding for two-dimensional optical storage," in *Proc. GLOBECOM 2003*, vol. 7, San Francisco, CA, Dec. 2003, pp. 3904–3908.

Piecewise Linear Conditional Information Inequality

František Matúš

***Abstract*—A new information inequality of non-Shannon type is proved for three discrete random variables under conditional independence constraints, using the framework of entropy functions and polymatroids. Tightness of the inequality is described via quasi-groups.**

***Index Terms*—Conditional independence, entropy function, information inequality, polymatroid, quasi-group, Shannon entropy.**

I. INTRODUCTION

The *entropy function* of a random vector $(\xi_i)_{i\in N}$ indexed by a finite set N assigns to each subset I of N the Shannon entropy of the subvector $(\xi_i)_{i\in I}$. The notion depends on the base of logarithms occurring in the entropy; our preference is given to the natural ones.

By basic properties of the Shannon entropy, an entropy function g satisfies $g(\emptyset) = 0$, is nondecreasing, thus $g(I) \le g(J)$ for $I \subseteq J$, and submodular, $g(I) + g(J) \ge g(I \cup J) + g(I \cap J)$ for $I, J \subseteq N$. A pair (N, g) where g is a real function on the power set $\mathcal{P}(N)$ of N that satisfies these three requirements is called a *polymatroid* with the ground set N and rank function g [3], [2], [8], making usually no difference between a polymatroid and its rank function. Thus, real-valued entropy functions are polymatroids [1]. Let us call a polymatroid *entropic* if it is equal to the entropy function of a random vector taking a finite number of values.

Theorem 1 below solves the problem which of the polymatroids $ag_1 + bg_2, a, b \ge 0$, is entropic for two very special polymatroids g_1, g_2 with a common three-element ground set.

A real function ϕ on $\mathbb{R}^{\mathcal{P}(N)}$, the Euclidean space containing entropy functions, generates an *information inequality* if $\phi(g) \ge 0$ holds for the entropic polymatroids. If the nonnegativity takes place even for all polymatroids then the inequality is of *Shannon type*. First non-Shannon type information inequalities appeared in [12], see also [10], [7]. If $\phi(g) \ge 0$ holds only for the entropic polymatroids satisfying additional constraints, usually on conditional independences among subvectors, then the inequality is *conditional*. For non-Shannon type conditional information inequalities see [11], [5].

One implication of Theorem 1 interprets as a non-Shannon type, conditional, piecewise linear information inequality, see Remarks 1 and 2. Tightness is discussed in Remark 3.

Manuscript received July 5, 2005; revised September 22, 2005. This work was supported by the Grant Agency of Academy of Sciences of the Czech Republic under Grant A1075104, and by the Grant Agency of the Czech Republic under Grant 201/01/1482.

The author is with the Institute of Information Theory and Automation, Academy of Sciences of the Czech Republic, 182 08 Prague, Czech Republic (e-mail: matus@utia.cas.cz).

Communicated by R. W. Yeung, Associate Editor for Shannon Theory.
Digital Object Identifier 10.1109/TIT.2005.860438