On the Capacity of 2-Dimensional Channels

Yonglong Li* and Paul H. Siegel†
Center for Memory and Recording Research*†, Department of Electrical and Computer Engineering†
University of California, San Diego
Email:{yol143, psiegel}@ucsd.edu

***Abstract*—For a 2-dimensional ($2D$) Gaussian inter-symbol interference (ISI) channel with discrete input and a $2D$ discrete memoryless channel with a special class of irreducible constraints, we show that the information capacity is equal to the stationary capacity. As a byproduct, these capacities are shown to be equal to the operational capacity.**

I. Introduction

In storage systems such as holographic memory [18] and non-volatile memory (e.g., NAND flash memory [4], 3D NAND flash memory [15] and 3D XPoint memory [10]), data are stored on $2D$ or $3D$ devices. For holographic memory, the storage channel is modeled as a $2D$ Gaussian ISI channel due to interference from neighboring bits. In [4], the authors propose a $2D$ communication channel to model multilevel NAND flash memory. To mitigate interference in holographic memory and NAND flash memory, high dimensional constrained coding schemes have been proposed [1], [2], [20]. In [22], [23], [16], [12], the authors derived some bounds on the noiseless capacity of some $2D$ constraints. In this paper, we study various concepts of capacity of $2D$ Gaussian ISI channels and input-constrained $2D$ discrete memoryless channels.

In Section II, we consider the $2D$ Gaussian ISI channel model:

$$Y_{k,l} = X_{k,l} + \sum_{(i,j)\in U} h_{i,j} X_{i+k,j+l} + W_{k,l}, \qquad (1)$$

where $U = \{(i,j) : |i| \le k_0, |j| \le l_0\}$, $h_{i,j}$ are real numbers, $W_{(k,l)}$ are independent Gaussian random variables with mean 0 and variance 1 and $\{W_{k,l}\}$ is independent of $\{X_{k,l}\}$. The input random field $\{X_{k,l}\}$ takes real values in a finite set $\mathcal{X}$.

For a $2D$ channel, we can define several notions of capacity. The *operational capacity* C, roughly speaking, is the largest rate under which reliable communication is possible. The *information capacity* or *Shannon capacity* $C_{Shannon}$ is defined in [9, p. 256] as

$$C_{Shannon} = \lim_{m,n\to\infty} C_{m,n}, \qquad (2)$$

where $C_{m,n} = \frac{1}{(m+1)(n+1)} \sup_{p(x_{0,0}^{m,n})} I(X_{0,0}^{m,n}; Y_{0,0}^{m,n})$

and X and Y are input and output random fields, respectively. Finally, the *stationary capacity* C_S is defined as

$$C_S = \sup_X I(X;Y),$$

where the supremum is taken over all the stationary random fields and

$$I(X;Y) = \lim_{m,n\to\infty} \frac{1}{(m+1)(n+1)} I(X_{0,0}^{m,n}; Y_{0,0}^{m,n}).$$

For such $2D$ channels, many tools available in $1D$ information theory are not applicable. First, a Shannon-McMillan-Breiman (SMB) theorem for continuous-valued $2D$ stationary ergodic random fields is not established yet, although a SMB theorem for discrete-valued stationary ergodic random fields was proved in [19]. Second, Ordentlich and Roth in [17] showed that the $2D$ maximum-likelihood detection problem is NP hard. However, some results relating to 2D capacity are known. In [6], Chen and Siegel derived bounds on the symmetric information rate of a $2D$ Gaussian ISI channel. In [21], the authors mapped the $2D$ Gaussian ISI channels to a graphical model and then applied the generalized belief propagation method to estimate the a posteriori probability and information rate. However, in contrast to the $1D$ case, for $2D$ channels with memory and states it is not known whether a stationary ergodic random field will achieve the information capacity, as was pointed out as an open problem in [7]. In this work, for a special class of $2D$ channels, we will show that the operational capacity C and information capacity $C_{Shannon}$ are equal and both can be achieved by a stationary ergodic random field, i.e., $C = C_S = C_{Shannon}$.

In Section III, we consider a $2D$ memoryless channel W with input alphabet $\mathcal{A}$ and output alphabet $\mathcal{Y}$, where $\mathcal{A}$ and $\mathcal{Y}$ are both finite sets. For a finite subset $S \subset \mathbb{Z}^2$, an element $u \in \mathcal{A}^S$ is called a configuration. For $x \in \mathcal{A}^{\mathbb{Z}^2}$, $x|_S$ is the restriction of x to S. Given a set of finite configurations $\mathcal{F}$, we define a $2D$ constraint $X(\mathcal{F}) = \{x \in \mathcal{A}^{\mathbb{Z}^2}$: for any finite subset $S \subset \mathbb{Z}^2, x|_S \notin \mathcal{F}\}$ that forbids $\mathcal{F}$. The language of $X(\mathcal{F})$ is $\mathcal{L}(X(\mathcal{F})) = \cup_U X_U(\mathcal{F})$, where $X_U(\mathcal{F}) = \{x \in \mathcal{A}^U$: for any finite subset $S \subset U, x|_S \notin \mathcal{F}\}$ and the union is taken over all finite $U \subset \mathbb{Z}^2$. For $U \subset \mathbb{Z}^2$, let $\partial_d U = \{(i,j) : (i,j) \in U$ and $|i-k| + |j-l| \le d+1$ for some $(k,l) \notin U\}$. A constraint $X(\mathcal{F})$ is said to have a safe symbol if there is some $a \in \mathcal{A}$ and some positive integer d such that for any configurations $u \in X_U(\mathcal{F})$, if $u_{i,j} = a$ for $(i,j) \in \partial_d U$, then $uv \in \mathcal{L}(X(\mathcal{F}))$ for any $v \in \mathcal{L}(X(\mathcal{F}))$. We call such an a and d a safe symbol and safe distance, respectively. For example for the $2D$-$(1,\infty)$-run-length-limited constraint $X(\mathcal{F})$ with

$$\mathcal{F} = \{11, \begin{matrix}1\\1\end{matrix}\},$$

$a = 0$ is the safe symbol with safe distance $d = 1$. The constraint $X(\mathcal{F})$ with

$$\mathcal{F} = \{101, \begin{matrix}1\\0\\1\end{matrix}\}$$

is relevant to mitigation of inter-cell interference in NAND flash memory [20]. For this constraint, $a = 0$ is the safe symbol

978-1-5386-4780-6/18/$31.00©2018 IEEE

with safe distance $d = 2$.

In [11] and [13], the authors characterized asymptotics of the capacity of a $1D$ binary symmetric channel with the input supported on an irreducible finite-type constraint. In [14], Li and Han derived the asymptotics of the capacity of a $1D$ erasure channel with input supported on an irreducible finite-type constraint. For a $2D$ memoryless channel with input constraint $X(\mathcal{F})$ with a safe symbol a and safe distance d, we consider the corresponding notions of capacity and prove a result analogous to the result above for $2D$ ISI channels. Namely, we show that the operational capacity C and information capacity $C_{Shannon}$ are equal and both can be achieved by a stationary ergodic random field supported on $X(\mathcal{F})$.

Throughout the paper, $X \sim \mu$ means that μ is the probability measure induced by the random field X and we use the logarithm with base e. Also both entropy and differential entropy of some random variable X are denoted by $H(X)$. For $U \subset \mathbb{Z}^2$, X_U means $\{X_{i,j} : (i,j) \in U\}$ and is denoted by $X_{m,n}^{M,N}$ when $U = \{(i,j) : m \leq i \leq M, n \leq j \leq N\}$.

II. $2D$ Gaussian ISI Channels

In this section we consider the $2D$ Gaussian ISI channel (1) and show that its operational capacity C and information capacity $C_{Shannon}$ are equal and both can be achieved by a stationary and ergodic random field.

Remark II.1. *In the definition of $C_{Shannon}$, $X_{i,j} = x_{i,j}$ is fixed for $(i,j) \in \{(i,j) : -k_0 \leq i \leq m + k_0, -l_0 \leq j \leq n + l_0\} - \{(i,j) : 0 \leq i \leq m, 0 \leq j \leq n\}$ and it can be verified that $C_{Shannon}$ is independent of the choice of $x_{i,j}$.*

Theorem II.2. *For the $2D$ Gaussian ISI channel (1),*

$$C = C_S = C_{Shannon}.$$

Proof. **Proof of $C_S \leq C$.** This follows from a standard "achievability part" proof: For any rate $R < C_S$ and $\varepsilon > 0$, choose a stationary ergodic input $\{X_{m,n}\}$ such that $R < I(X;Y) - \varepsilon$. First choose N such that

$$\lim_{m\to\infty} \frac{I(X_{0,0}^{m,N}; Y_{0,0}^{m,N})}{mN} < I(X;Y) - \varepsilon/2.$$

Then it follows from the generalized Shannon-McMillan-Breiman Theorem in [5] that $\{(X_{m,n}, Y_{m,n}) : m \in \mathbb{Z}, 0 \leq n \leq N\}$ satisfies the asymptotic equipartition property. Then the proof of the achievability can be completed by going through the usual random coding argument [3, p. 200].

Proof of $C \leq C_{Shannon}$. This follows from a standard "converse part" proof [3, p. 207].

Proof of $C_{Shannon} \leq C_S$. First we find a capacity-achieving distribution for $C_{Shannon}$ on a large $2D$ grid, then we construct a block independent process with mutual information close to $C_{Shannon}$. Then using a classical argument in probability [8], we construct a stationary and ergodic random field from this block independent random field that achieves $C_{Shannon}$. The proof is similar to the one in [8], so we just outline the main steps.

Step 0. Let $\alpha = \sup_{x\in\mathcal{X}} |x|$. First, for any $\varepsilon > 0$, choose sufficiently large M and N such that

$$\frac{(Ml_0 + Nk_0)\log 2\pi e\left(2 + \sum_{(i,j)\in U} h_{i,j}^2\right)\left((|U|+1)\alpha^2 + 1\right)}{(M+1)(N+1)} + \frac{2(Ml_0 + Nk_0)\log|\mathcal{X}|}{(M+1)(N+1)} \leq \frac{\varepsilon}{2} \quad (3)$$

and

$$\sup_{X_{0,0}^{M,N}} \frac{I(X_{0,0}^{M,N}; Y_{0,0}^{M,N})}{(M+1)(N+1)} > C_{Shannon} - \frac{\varepsilon}{3}, \quad (4)$$

and then choose $X_{0,0}^{M,N} \sim p(x_{0,0}^{M,N})$ such that

$$\frac{I(X_{0,0}^{M,N}; Y_{0,0}^{M,N})}{(M+1)(N+1)} \geq C_{Shannon} - \frac{\varepsilon}{2}. \quad (5)$$

Step 1. Now, let $\hat{X} = \{\hat{X}_{i,j}\}$ be the "independent block" random field defined as follows:

(i) $\hat{X}_{k_1(M+1),k_2(N+1)}^{k_1(M+1)+M,k_2(N+1)+N}$ are i.i.d. for $k_1, k_2 \in \mathbb{Z}$;

(ii) $\hat{X}_{0,0}^{M,N}$ has the same distribution as $X_{0,0}^{M,N}$.

Let $\hat{Y}$ be the output obtained by passing $\hat{X}$ through the channel (1). Let ν be independent of $\{\hat{X}_{i,j}, W_{i,j}\}$ and uniformly distributed over $\{(i', j') : 0 \leq i' \leq M, 0 \leq j' \leq N\}$, and let $\bar{X}_{i,j} = \hat{X}_{\nu+(i,j)}$. It can be verified that $\{\bar{X}_{i,j}\}$ is a stationary and ergodic random field.

Step 2. Let $\{\bar{Y}_{i,j}\}$ be the output obtained by passing the stationary random field $\{\bar{X}_{i,j}\}$ through the channel (1). Letting

$$I(\bar{X};\bar{Y}) = \lim_{k_1,k_2\to\infty} \frac{I(\bar{X}_{0,0}^{k_1(M+1)+M,k_2(N+1)+N}; \bar{Y}_{0,0}^{k_1(M+1)+M,k_2(N+1)+N})}{(k_1+1)(k_2+1)(M+1)(N+1)},$$

we will show that

$$I(\bar{X};\bar{Y}) - \frac{I(X_{0,0}^{M,N}; Y_{0,0}^{M,N})}{(M+1)(N+1)} \geq -\frac{\varepsilon}{2}, \quad (6)$$

which, together with (5) and the arbitrariness of ε, will imply $C_S \geq C_{Shannon}$.

For $0 \leq i \leq M, 0 \leq j \leq N$, Let $\hat{X}_{(i,j),m,n} \triangleq \hat{X}_{m+i,n+j}$ and $\hat{Y}_{(i,j)} = \{\hat{Y}_{(i,j),m,n}\}$ be the output random field obtained by passing the random field $\hat{X}_{(i,j)} = \{\hat{X}_{(i,j)m,n}\}$ through the channel (1). Now one checks that

$$\begin{aligned}
& p_{\bar{X}_{0,0}^{m,n}}(x_{0,0}^{m,n}) f_{\bar{Y}_{0,0}^{m,n}|\bar{X}_{0,0}^{m,n}}(y_{0,0}^{m,n}|x_{0,0}^{m,n}) \\
&= \sum_{i=0}^{M}\sum_{j=0}^{N} \frac{P(\bar{X}_{0,0}^{m,n} = x_{0,0}^{m,n}|\nu = (i,j)) f_{\bar{Y}_{0,0}^{m,n}|\bar{X}_{0,0}^{m,n}}(y_{0,0}^{m,n}|x_{0,0}^{m,n})}{(M+1)(N+1)} \\
&= \sum_{i=0}^{M}\sum_{j=0}^{N} \frac{P(\hat{X}_{i,j}^{m+i,n+j} = x_{0,0}^{m,n}) f_{\bar{Y}_{0,0}^{m,n}|\bar{X}_{0,0}^{m,n}}(y_{0,0}^{m,n}|x_{0,0}^{m,n})}{(M+1)(N+1)} \\
&= \sum_{i=0}^{M}\sum_{j=0}^{N} \frac{P(\hat{X}_{(i,j)0,0}^{m,n} = x_{0,0}^{m,n}) f_{\hat{Y}_{0,0}^{m,n}|\hat{X}_{0,0}^{m,n}}(y_{0,0}^{m,n}|x_{0,0}^{m,n})}{(M+1)(N+1)}.
\end{aligned}$$

Then it follows from Lemma 2 in [8] that

$$I(\bar{X};\bar{Y}) = \frac{1}{(M+1)(N+1)} \sum_{i=0}^{M}\sum_{j=0}^{N} I(\hat{X}_{(i,j)}; \hat{Y}_{(i,j)}),$$

where

$$I(\hat{X}_{(i,j)};\hat{Y}_{(i,j)}) = \lim_{k_1,k_2\to\infty} \frac{I(\hat{X}_{(i,j)0,0}^{k_1(M+1)+M,k_2(N+1)+N};\hat{Y}_{(i,j)0,0}^{k_1(M+1)+M,k_2(N+1)+N})}{(k_1+1)(k_2+1)(M+1)(N+1)}.$$

To prove (6), it suffices to establish that for any (i,j),

$$I(\hat{X}_{(i,j)};\hat{Y}_{(i,j)}) \geq \frac{I(X_{0,0}^{M,N};Y_{0,0}^{M,N})}{(M+1)(N+1)} - \frac{\varepsilon}{2}. \quad (7)$$

The proof of (7) is similar for all values of (i,j), so in the following we only show it holds true for $(i,j)=(0,0)$. Here, we note that when $(i,j)=(0,0)$,

$$I(\hat{X}_{(0,0)};\hat{Y}_{(0,0)}) = I(\hat{X};\hat{Y}) = \lim_{k_1,k_2\to\infty} \frac{I(\hat{X}_{0,0}^{k_1(M+1)+M,k_2(N+1)+N};\hat{Y}_{0,0}^{k_1(M+1)+M,k_2(N+1)+N})}{(k_1+1)(k_2+1)(M+1)(N+1)}.$$

Let

$$G_{s,t} = \left\{(u,v): \begin{array}{c} s(M+1)\leq u \leq (s+1)(M+1)-1, \\ t(N+1)\leq v\leq (t+1)(N+1)-1 \end{array}\right\}$$

and

$$\Gamma_{s,t} = \bigcup_{s_1<s \text{ or } s_1=s,t_1<t} G_{s_1,t_1}.$$

Intuitively, $\Gamma_{s,t}$ is the "block history" of $G_{s,t}$ (see $\Gamma_{4,3}$ in Fig. 1).

Using the chain rule for mutual information, we have,

$$\sum_{s=0}^{k_1}\sum_{t=0}^{k_2} I(\hat{X}_{G_{s,t}};\hat{Y}_{G_{s,t}}|\hat{X}_{\Gamma_{s,t}},\hat{Y}_{\Gamma_{s,t}}) \leq I(\hat{X}_{0,0}^{k_1(M+1)+M,k_2(N+1)+N};\hat{Y}_{0,0}^{k_1(M+1)+M,k_2(N+1)+N}), \quad (8)$$

which means that, to prove (7), it suffices to show that

$$\frac{I(\hat{X}_{s(M+1),t(N+1)}^{s(M+1)+M,t(N+1)+N};\hat{Y}_{s(M+1),t(N+1)}^{s(M+1)+M,t(N+1)+N}|\hat{X}_{\Gamma_{s,t}},\hat{Y}_{\Gamma_{s,t}})}{(M+1)(N+1)} \geq \frac{1}{(M+1)(N+1)}I(X_{0,0}^{M,N};Y_{0,0}^{M,N}) - \frac{\varepsilon}{2}.$$

Without loss of generality, we prove this holds true for $(s,t)=(1,1)$. Let

$$U_{M,N} = \left\{(u,v): \begin{array}{c} M+1+k_0\leq u\leq 2M+1-k_0, \\ N+1+l_0\leq v\leq 2N+1-l_0 \end{array}\right\}$$

and

$$U_0 = G_{1,1} - U_{M,N}.$$

Geometrically, U_0 is the difference of the rectangle $G_{1,1}$ and $U_{M,N}$ (see Fig. 2). In the following, we will first apply the chain rule to decompose $I(\hat{X}_{G_{1,1}};\hat{Y}_{G_{1,1}}|\hat{X}_{\Gamma_{1,1}},\hat{Y}_{\Gamma_{1,1}})$ into three parts. Then we will show that each part is close to the counterpart of $I(X_{0,0}^{M,N};Y_{0,0}^{M,N})$.

Note that

$$\begin{aligned}
&I(\hat{X}_{G_{1,1}};\hat{Y}_{G_{1,1}}|\hat{X}_{\Gamma_{1,1}},\hat{Y}_{\Gamma_{1,1}}) \\
&= I(\hat{X}_{G_{1,1}};\hat{Y}_{U_0}|\hat{X}_{\Gamma_{1,1}},\hat{Y}_{\Gamma_{1,1}}) + I(\hat{X}_{G_{1,1}};\hat{Y}_{U_{M,N}}|\hat{X}_{\Gamma_{1,1}},\hat{Y}_{\Gamma_{1,1}\cup U_0}) \\
&= I(\hat{X}_{G_{1,1}};\hat{Y}_{U_0}|\hat{X}_{\Gamma_{1,1}},\hat{Y}_{\Gamma_{1,1}}) + I(\hat{X}_{U_0};\hat{Y}_{U_{M,N}}|\hat{X}_{\Gamma_{1,1}},\hat{Y}_{\Gamma_{1,1}\cup U_0}) \\
&\quad + I(\hat{X}_{U_{M,N}};\hat{Y}_{U_{M,N}}|\hat{X}_{\Gamma_{1,1}\cup U_0},\hat{Y}_{\Gamma_{1,1}\cup U_0}) \\
&= I(\hat{X}_{G_{1,1}};\hat{Y}_{U_0}|\hat{X}_{\Gamma_{1,1}},\hat{Y}_{\Gamma_{1,1}}) + I(\hat{X}_{U_0};\hat{Y}_{U_{M,N}}|\hat{X}_{\Gamma_{1,1}},\hat{Y}_{\Gamma_{1,1}\cup U_0}) \\
&\quad + I(\hat{X}_{U_{M,N}};\hat{Y}_{U_{M,N}}|\hat{X}_{U_0}), \quad (9)
\end{aligned}$$

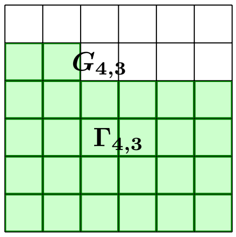

Fig.1

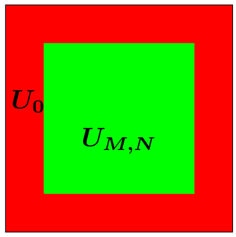

Fig.2

where in (9) we use the fact that given $\hat{X}_{U_0}$, $(\hat{X}_{U_{M,N}},\hat{Y}_{U_{M,N}})$ is independent of $(\hat{X}_{\Gamma_{1,1}},\hat{Y}_{\Gamma_{1,1}\cup U_0})$.

Let

$$V_0 = \{(u,v): 0\leq u\leq M, 0\leq v\leq N\} - \left\{(u,v): \begin{array}{c} k_0\leq u\leq M-k_0, \\ l_0\leq v\leq N-l_0 \end{array}\right\}.$$

Since

$$I(\hat{X}_{0,0}^{M,N};\hat{Y}_{0,0}^{M,N}) = I(\hat{X}_{V_0};\hat{Y}_{0,0}^{M,N}) + I(\hat{X}_{k_0,l_0}^{M-k_0,N-l_0};\hat{Y}_{V_0}|\hat{X}_{U_0}) + I(\hat{X}_{k_0,l_0}^{M-k_0,N-l_0};\hat{Y}_{k_0,l_0}^{M-k_0,N-l_0}|\hat{X}_{U_0}),$$

we have that

$$\begin{aligned}
&|I(\hat{X}_{G_{1,1}};\hat{Y}_{G_{1,1}}|\hat{X}_{\Gamma_{1,1}},\hat{Y}_{\Gamma_{1,1}}) - I(\hat{X}_{0,0}^{M,N};\hat{Y}_{0,0}^{M,N})| \\
&\overset{(a)}{\leq} I(\hat{X}_{V_0};\hat{Y}_{0,0}^{M,N}) + I(\hat{X}_{G_{1,1}};\hat{Y}_{U_0}|\hat{X}_{\Gamma_{1,1}},\hat{Y}_{\Gamma_{1,1}}) \\
&+ I(\hat{X}_{U_0};\hat{Y}_{U_{M,N}}|\hat{X}_{\Gamma_{1,1}},\hat{Y}_{\Gamma_{1,1}\cup U_0}) + I(\hat{X}_{k_0,l_0}^{M-k_0,N-l_0};\hat{Y}_{V_0}|\hat{X}_{U_0}) \\
&\overset{(b)}{\leq} (|U_0|+|V_0|)\log|\mathcal{X}| + I(\hat{X}_{G_{1,1}};\hat{Y}_{U_0}|\hat{X}_{\Gamma_{1,1}},\hat{Y}_{\Gamma_{1,1}}) \\
&\quad + I(\hat{X}_{k_0,l_0}^{M-k_0,N-l_0};\hat{Y}_{V_0}|\hat{X}_{U_0}), \quad (10)
\end{aligned}$$

where in (a) we use the fact that

$$p_{X_{0,0}^{M,N}}(\cdot) = p_{\hat{X}_{M+1,N+1}^{2M+1,2N+1}}(\cdot)$$

and in (b) we use the inequalities

$$I(\hat{X}_{U_0};\hat{Y}_{M+1+k_0,N+1+l_0}^{2M+1-k_0,2N+1-l_0}|\hat{X}_{\Gamma_{1,1}},\hat{Y}_{\Gamma_{1,1}\cup U_0}) \leq H(\hat{X}_{U_0}) \leq |U_0|\log|\mathcal{X}|$$

and

$$I(\hat{X}_{V_0};\hat{Y}_{k_0,l_0}^{M-k_0,N-l_0}) \leq H(\hat{X}_{V_0}) \leq |V_0|\log|\mathcal{X}|.$$

Recall that $\alpha = \sup_{x\in\mathcal{X}}|x|$. Then it follows from Hölder's inequality that

$$\mathbf{E}[Y_{i,j}^2] \leq (2+\sum_{(i,j)\in U} h_{i,j}^2)((|U|+1)\alpha^2 + \mathbf{E}[W_{i,j^2}]).$$

Then we have

$$\begin{aligned}
H(\hat{Y}_{U_0}|\hat{X}_{\Gamma_{1,1}},\hat{Y}_{\Gamma_{1,1}}) &\leq H(\hat{Y}_{U_0}) = \sum_{(i,j)\in U_0} H(\hat{Y}_{i,j}) \\
&\leq \frac{|U_0|}{2}\log 2\pi e((2+\sum_{(i,j)\in U} h_{i,j}^2)((|U|+1)\alpha^2+1)), \quad (11)
\end{aligned}$$

where (11) follows from the fact that Gaussian distribution maximizes the entropy given the second moment.

It then follows from (11) and

$$H(\hat{Y}_{U_0}|\hat{X}_{M+1,N+1}^{2M+1,2N+1},\hat{X}_{\Gamma_{1,1}},\hat{Y}_{\Gamma_{1,1}}) = H(W_{U_0}) > 0,$$

that

$$I(\hat{X}_{M+1,N+1}^{2M+1,2N+1};\hat{Y}_{U_0}|\hat{X}_{\Gamma_{1,1}},\hat{Y}_{\Gamma_{1,1}})$$
$$\leq H(\hat{Y}_{U_0}|\hat{X}_{\Gamma_{1,1}},\hat{Y}_{\Gamma_{1,1}})$$
$$\leq \frac{|U_0|}{2}\log 2\pi e\left(2+\sum_{(i,j)\in U}h_{i,j}^2\right)((|U|+1)\alpha^2+1). \quad (12)$$

Now, with (5) and the fact that

$$|U_0|=|V_0|\leq 2(Ml_0+Nk_0),$$

we conclude that

$$\frac{I(\hat{X}_{M+1,N+1}^{2M+1,2N+1};\hat{Y}_{M+1,N+1}^{2M+1,2N+1}|\hat{X}_{\Gamma_{1,1}},\hat{Y}_{\Gamma_{1,1}})}{(M+1)(N+1)}$$
$$\geq \frac{I(X_{0,0}^{M,N};Y_{0,0}^{M,N})}{(M+1)(N+1)}-\frac{(|U_0|+|V_0|)\log|\mathcal{X}|}{(M+1)(N+1)}$$
$$-\frac{(|U_0|+|V_0|)\log 2\pi e\left(2+\sum_{(i,j)\in U}h_{i,j}^2\right)((|U|+1)\alpha^2+1)}{2(M+1)(N+1)}$$
$$\geq \frac{I(X_{0,0}^{M,N};Y_{0,0}^{M,N})}{(M+1)(N+1)}-\frac{2(Ml_0+Nk_0)\log|\mathcal{X}|}{(M+1)(N+1)}$$
$$-\frac{(Ml_0+Nk_0)\log 2\pi e\left(2+\sum_{(i,j)\in U}h_{i,j}^2\right)((|U|+1)\alpha^2+1)}{(M+1)(N+1)}$$
$$\geq \frac{I(X_{0,0}^{M,N};Y_{0,0}^{M,N})}{(M+1)(N+1)}-\frac{\varepsilon}{2}, \quad (13)$$

as desired. □

III. 2D Memoryless Channels with Input Constraints

In this section, we are concerned with $2D$ memoryless channels with the input constraint $X(\mathcal{F})$ having a safe symbol a and safe distance d. Let $X_{m,n}(\mathcal{F})=\{u\in\mathcal{A}^{R_{m,n}}:$ for any finite subset $S\subset R_{m,n}, u|_S\notin\mathcal{F}\}$, where $R_{m,n}=\{(i,j):0\leq i\leq m, 0\leq j\leq n, i,j\in\mathbb{Z}\}$. Let $\mathcal{M}_{m,n}$ be the set of probability mass functions over $X_{m,n}(\mathcal{F})$ and $\mathcal{M}$ the set of probability measures on $X(\mathcal{F})$. The following theorem says that the capacity of input-constrained $2D$ memoryless channels can be achieved by stationary ergodic random fields.

Theorem III.1. *Let $X(\mathcal{F})$ be a constraint with a safe symbol. For a $2D$ memoryless channel with input constraint $X(\mathcal{F})$,*

$$C=C_{Shannon}=C_S, \quad (14)$$

where

$$C_{Shannon}=\lim_{m,n\to\infty}\frac{1}{(m+1)(n+1)}\sup_{p\in\mathcal{M}_{m,n}}I(X_{0,0}^{m,n};Y_{0,0}^{m,n})$$

and

$$C_S=\sup_{X\sim\mu:\mu\in\mathcal{M}}I(X;Y).$$

Proof. We will apply similar arguments as in the proof of Theorem II.2 to prove (14). The main difference is the construction of the capacity-achieving random field. Due to the limitation of the constraint $X(\mathcal{F})$, we cannot put independent blocks together in a straightforward way. However, for two adjacent blocks with sufficiently large distance, we can, using the property of safe symbol of the constraint $X(\mathcal{F})$, put extra symbols between independent blocks to preserve the constraint $X(\mathcal{F})$. Then using the same idea as in Theorem II.2, we construct a stationary and ergodic random field from this random field that achieves $C_{Shannon}$. In the following, we outline the main steps.

Step 0. First of all, for any $\varepsilon>0$, choose M and N such that

$$\left(1-\frac{(M+1)(N+1)}{(M+d+1)(N+d+1)}\right)\log|\mathcal{Y}|$$
$$+\frac{(\log|\mathcal{A}|+\log|\mathcal{Y}|)(M+N)d}{(M+d+1)(N+d+1)}\leq\frac{\varepsilon}{2}$$

and then choose $X_{0,0}^{M,N}\sim p(x_{0,0}^{M,N})$ such that

$$\frac{I(X_{0,0}^{M,N};Y_{0,0}^{M,N})}{(M+1)(N+1)}\geq C_{Shannon}-\frac{\varepsilon}{2}. \quad (17)$$

Step 1. Let $R(k_1,k_2)$ be the rectangle consisting of (i,j) such that

$$k_1(M+d+1)\leq i\leq(k_1+1)(M+d+1)-1,$$
$$k_2(N+d+1)\leq j\leq(k_2+1)(N+d+1)-1.$$

Let $R_1(k_1,k_2)$ be the complement of $\{(i,j):k_1(M+d+1)\leq i\leq k_1(M+d+1)+M, k_2(N+d+1)\leq j\leq k_2(N+d+1)+N\}$ in $R(k_1,k_2)$. More precisely, $R_1(k_1,k_2)$ is the set of (i,j) such that either

$$k_1(M+d+1)\leq i\leq(k_1+1)(M+d+1)-1$$
$$k_2(N+d+1)+N+1\leq j\leq k_2(N+d+1)+N+d,$$

or

$$k_1(M+d+1)+M+1\leq i\leq k_1(M+d+1)+M+d,$$
$$k_2(N+d+1)\leq j\leq(k_2+1)(N+d+1)-1.$$

Now, let $\hat{X}=\{\hat{X}_{i,j}\}$ be the "block" random field defined as:

(i) $\hat{X}_{k_1(M+d+1),k_2(N+d+1)}^{k_1(M+d+1)+M,k_2(N+d+1)+N}$ are i.i.d. for $k_1,k_2\in\mathbb{Z}$;
(ii) $\hat{X}_{i,j}=a$ for $(i,j)\in R_1(k_1,k_2)$.
(ii) $\hat{X}_{0,0}^{M,N}$ has the same distribution as $X_{0,0}^{M,N}$.

From the definition of a and d, it follows that $\{\hat{X}_{i,j}\}$ is a random field supported on the constraint $X(\mathcal{F})$. One can easily checks that $\hat{X}_{k_1(M+d+1),k_2(N+d+1)}^{k_1(M+d+1)+M+d,k_2(N+d+1)+N+d}$ are i.i.d. Now let $\hat{Y}$ be the output obtained by passing $\hat{X}$ through the channel W. Due to the memoryless property of channel W, $\hat{Y}_{k_1(M+d+1),k_2(N+d+1)}^{k_1(M+d+1)+M+d,k_2(N+d+1)+N+d}$ are also i.i.d. Let ν be independent of $\{\hat{X}_{i,j}\}$ and uniformly distributed over $\{(i,j):0\leq i\leq M+d, 0\leq j\leq N+d\}$, and let $\bar{X}_{i,j}=\hat{X}_{\nu+(i,j)}$. It can be verified that $\{\bar{X}_{i,j}\}$ is a stationary and ergodic random field.

Step 2. Let $\{\bar{Y}_{i,j}\}$ be the output obtained by passing the stationary random field $\{\bar{X}_{i,j}\}$ through the channel W. We will show that

$$I(\bar{X};\bar{Y})-\frac{I(X_{0,0}^{M+d,N+d};Y_{0,0}^{M+d,N+d})}{(M+d+1)(N+d+1)}\geq-\frac{\varepsilon}{2}, \quad (18)$$

which, together with (17) and the arbitrariness of ε, will imply $C_S\geq C_{Shannon}$.

For $0\leq i\leq M+d$ and $0\leq j\leq N+d$, let $\hat{X}_{(i,j),m,n}=\hat{X}_{m+i,n+j}$ and $\hat{Y}_{(i,j)}=\{\hat{Y}_{(i,j),m,n}\}$ denote the output random field obtained by passing the random field

$$I(\hat{X}_{(0,0)};\hat{Y}_{(0,0)}) = I(\hat{X};\hat{Y}) = \lim_{k_1,k_2\to\infty} \frac{I(\hat{X}_{0,0}^{(k_1+1)(M+d+1)-1,(k_2+1)(N+d+1)-1};\hat{Y}_{0,0}^{(k_1+1)(M+d+1)-1,(k_2+1)(N+d+1)-1})}{(k_1+1)(k_2+1)(M+1)(N+1)}. \quad (15)$$

$$I(\hat{X}_{0,0}^{(k_1+1)(M+d+1)-1,(k_2+1)(N+d+1)-1};\hat{Y}_{0,0}^{(k_1+1)(M+d+1)-1,(k_2+1)(N+d+1)-1}) = \sum_{s=1}^{k_1}\sum_{t=1}^{k_2} I(\hat{X}_{R(k_1,k_2)};\hat{Y}_{R(k_1,k_2)}) \quad (16)$$

$\hat{X}_{(i,j)} = \{\hat{X}_{(i,j)m,n}\}$ through the channel W. Then going through the same argument as in **Step 2** of the proof of Theorem II.2, to prove (18), it suffices to show that

$$I(\hat{X}_{(i,j)};\hat{Y}_{(i,j)}) \geq \frac{I(X_{0,0}^{M+d,N+d};Y_{0,0}^{M+d,N+d})}{(M+d+1)(N+d+1)} - \frac{\varepsilon}{2}. \quad (19)$$

The proof of (19) is similar for all values of (i,j), so in the following we only show it holds true for $(i,j)=(0,0)$. Note that when $(i,j)=(0,0)$, (15) holds.

Using the chain rule for mutual information and the fact that $(\hat{X}_{R(k_1,k_2)},\hat{Y}_{R(k_1,k_2)})$ are independent, we have (16), which means that, to prove (19), it suffices to show that

$$\frac{I(\hat{X}_{R(s,t)};\hat{Y}_{R(s,t)})}{(M+d+1)(N+d+1)} \geq \frac{I(X_{0,0}^{M,N};Y_{0,0}^{M,N})}{(M+1)(N+1)} - \frac{\varepsilon}{2}.$$

Without loss of generality, we prove this holds true for $(s,t)=(1,1)$. Using the chain rule for mutual information, we have

$$\begin{aligned} I(\hat{X}_{R(1,1)};\hat{Y}_{R(1,1)}) &= I(\hat{X}_{R(1,1)-R_1(1,1)};\hat{Y}_{R(1,1)-R_1(1,1)}) \\ &\quad + I(\hat{X}_{R_1(1,1)};\hat{Y}_{R(1,1)}|\hat{X}_{R(1,1)-R_1(1,1)}) \\ &\quad + I(\hat{X}_{R(1,1)-R_1(1,1)};\hat{Y}_{R_1(1,1)}|\hat{Y}_{R(1,1)-R_1(1,1)}) \\ &\geq I(X_{0,0}^{M,N};Y_{0,0}^{M,N}) - (\log|\mathcal{A}|+\log|\mathcal{Y}|)|R_1(1,1)|, \quad (20) \end{aligned}$$

where in (20) we use the fact that

$$p_{X_{0,0}^{M,N}}(\cdot) = p_{\hat{X}_{R(1,1)-R_1(1,1)}}(\cdot)$$

and the two inequalities

$$\begin{aligned} &I(\hat{X}_{R_1(1,1)};\hat{Y}_{R(1,1)}|\hat{X}_{R(1,1)-R_1(1,1)}) \leq H(\hat{X}_{R(1,1)}) \\ &\leq |R_1(1,1)|\log|\mathcal{A}| \end{aligned}$$

and

$$\begin{aligned} &I(\hat{X}_{R(1,1)-R_1(1,1)};\hat{Y}_{R_1(1,1)}|Y_{R(1,1)-R_1(1,1)}) \leq H(\hat{Y}_{R(1,1)}) \\ &\leq |R_1(1,1)|\log|\mathcal{Y}|. \end{aligned}$$

Since $|R_1(1,1)| \leq (M+N)d$, it then follows that

$$\begin{aligned} &\frac{I(\hat{X}_{R(s,t)};\hat{Y}_{R(s,t)})}{(M+d+1)(N+d+1)} \\ &\geq \frac{I(X_{0,0}^{M,N};Y_{0,0}^{M,N})}{(M+1)(N+1)} - \frac{(\log|\mathcal{A}|+\log|\mathcal{Y}|)|R_1(1,1)|}{(M+d+1)(N+d+1)} \\ &\quad - \left(1-\frac{(M+1)(N+1)}{(M+d+1)(N+d+1)}\right)\log|\mathcal{Y}| \\ &\geq \frac{I(X_{0,0}^{M,N};Y_{0,0}^{M,N})}{(M+1)(N+1)} - \frac{(\log|\mathcal{A}|+\log|\mathcal{Y}|)(M+N)d}{(M+d+1)(N+d+1)} \\ &\quad - \left(1-\frac{(M+1)(N+1)}{(M+d+1)(N+d+1)}\right)\log|\mathcal{Y}| \\ &\geq \frac{I(X_{0,0}^{M,N};Y_{0,0}^{M,N})}{(M+1)(N+1)} - \frac{\varepsilon}{2}, \end{aligned}$$

as desired. □

IV. Acknowledgement

This work was supported by NSF Grants CCF-1415109 and CCF-1619053.

References

[1] J. J. Ashley, M. Blaum and B. H. Marcus, "Report on coding techniques for holographic storage," *IBM Res. Rep.*, RJ 10013, 1996.

[2] J. J. Ashley and B. H. Marcus, "Two-dimensional low-pass filtering codes," *IEEE Trans. Commun.*, vol. 46, no.6, pp. 724–727, Jun. 1998.

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

[4] M. Asadi, X. Huang, A. Kavcic, and N. Santhanam, "Optimal detector for multilevel NAND flash memory channels with intercell interference," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 5, pp. 825–835, May 2014.

[5] A. R. Barron, "The strong ergodic theorem for densities: generalized Shannon-McMillan-Breiman theorem," *The Annals of Probability*, vol. 13, no. 4, pp. 1292–1303, Nov. 1985.

[6] Jiangxin 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.

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

[8] A. Feinstein, "On the coding theorem and its converse for finite-memory channels," *Il Nuovo Cimento Series 10*, vol. 13, no. 2, pp. 560–575, 1959.

[9] R. M. Gray, *Entropy and Information Theory*. Springer US, 2011.

[10] F. T. Hady, A. Foong, B. Veal and D. Williams, "Platform storage performance with 3D XPoint technology," in *Proceedings of the IEEE*, vol. 105, no. 9, pp. 1822–1833, Sept. 2017.

[11] G. Han and B. Marcus, "Asymptotics of input-constrained binary symmetric channel capacity," *Ann. Appl. Probab.*, vol. 19, no. 3, pp. 1063–1091, 2009.

[12] A. Kato and K. Zeger, "On the capacity of two-dimensional run-length constrained channels," *IEEE Trans. Inf. Theory*, vol. 45, no. 7, pp. 1527–1540, Jul. 1999.

[13] P. Jacquet and W. Szpankowski, "Noisy constrained capacity for BSC channels," *IEEE Trans. Inf. Theory*, vol. 56, no. 11, pp. 5412–5423, Nov. 2010.

[14] Y. Li and G. Han, "Asymptotics of input-constrained erasure channel capacity," *IEEE Trans. Inf. Theory*, vol. 64, no. 1, pp. 148–162, Jan. 2018.

[15] R. Micheloni and L. Crippa, "3D Stacked NAND Flash Memories," *3D Flash Memories* , pp. 63–83, Springer Netherlands, 2016

[16] Z. Nagy and K. Zeger, "Capacity bounds for the 3-dimensional runlength limited channel," *IEEE Trans. Inf. Theory*, vol. 46, no. 5, pp. 1030–1033, May 2000.

[17] E. Ordentlich and R. M. Roth, "Two-dimensional maximum-likelihood sequence detection is NP hard," *IEEE Trans. Inf. Theory*, vol. 57, no. 12, pp. 7661–7670, Dec. 2011.

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

[19] D. Ornstein and B. Weiss, "The Shannon-McMillan-Breiman theorem for a class of amenable groups," *Israel J. Math.*, vol. 44, iss. 1, pp. 53–60, Apr. 1983.

[20] M. Qin, E. Yaakobi, and P. Siegel, "Constrained codes that mitigate inter-cell interference in read/write cycles for flash memories," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 5, pp. 836–846, May 2014.

[21] O. Shental, N. Shental, S. Shamai, I. Kanter, A. J. Weiss and Y. Weiss, "Discrete-input two-dimensional gaussian channels with memory: estimation and information rates via graphical models and statistical mechanics," *IEEE Trans. Inf. Theory*, vol. 54, no. 4, pp. 1500–1513, Apr. 2008.

[22] P. H. Siegel and J. K. Wolf, "Bit-stuffing bounds on the capacity of 2-dimensional constrained arrays," in *Proc. IEEE Int. Symp. Information Theory*, Cambridge, MA, Aug. 1998, p. 323.

[23] W. Weeks and R. E. Blahut, "The capacity and coding gain of certain checkerboard codes," *IEEE Trans. Inform. Theory*, vol. 44, no. 3, pp. 1193–1203, May 1998.