Abstract—In this paper, an ordering result is given for Markov channels with respect to mutual information, under the assumption of an independent and identically distributed (i.i.d.) input distribution. For those Markov channels in which the capacity-achieving input distribution is i.i.d., this allows ordering of the channels by capacity. The complexity of analyzing general Markov channels is mitigated by this ordering, since it is possible to immediately determine that a wide class of channels, with different numbers of states, has a smaller mutual information than a given channel.

Index Terms—Finite-state Markov channels, mutual information, partial ordering.

I. INTRODUCTION

In a finite-state Markov channel, the instantaneous value of the channel parameter is selected by the state of a hidden Markov chain. Such channels are useful for modeling systems in which the statistics of the noise process vary with time. As an example, the well-known Gilbert–Elliott (GE) channel [1] consists of a binary-input, binary-output channel, and a two-state Markov chain; the channel behaves as a binary-symmetric channel (BSC) with two crossover probabilities, where the instantaneous crossover probability is chosen by the state of the Markov chain.

For the GE channel, capacity and coding have been studied in many papers, such as in [2]–[4]. Capacity and coding for the more general case, allowing arbitrary input and output alphabets, and larger state spaces, were originally studied in [5]. In contemporary applications, finite-state Markov channels are commonly used to model the effects of mobile fading [6], [7], as well as the characteristics of indoor wireless local-area networks [8], [9].

Because of the requirement to specify the transition probability matrix for the Markov chain, the number of parameters for a Markov channel scales with the square of the number of states. These large parameter spaces complicate the analysis of Markov channels. For example, for a given input distribution, suppose a system designer was interested in finding all GE channels for which the mutual information $I(X;Y)$ was less than 1/2. The equivalent problem for a memoryless BSC is solved by finding the crossover probability $\nu$ corresponding to $I(X;Y) = 1/2$ (which is $\nu = 0,1$), and noting the well-known property of the BSC that all larger values of $\nu$ have smaller $I(X;Y)$. No complete channel ordering exists for the GE channel, so the system designer would have to sample the four-dimensional parameter space to estimate the region in which $I(X;Y) < 1/2$.

In this paper, we address this complexity by showing how to construct Markov channels that have smaller mutual information than an initial channel, thereby forming a partial ordering with respect to mutual information. Using this ordering, if the mutual information using some channel $c$ is known, then $c$ casts a multidimensional “shadow” of neighboring channels where the mutual information is known to be smaller. The results in this paper are related to our previous work in [4], [10], in which general Markov channels were ordered with respect to symbol error probability after the iterative decoding of a low-density parity-check (LDPC) code. We start out with a similar approach (including a proof technique that generalizes one used in [2]), although a completely different method is required to prove its relevance to the present ordering.

It is important to note that, throughout this paper, we operate under the restriction that the input distribution to the channel is independent and identically distributed (i.i.d.). As a result, the ordering is generally with respect to mutual information, rather than with respect to capacity. However, the i.i.d. input distribution is capacity-achieving for many Markov channels [5], including the GE channel [2]; and the analysis of Markov channels under non-i.i.d. input distributions is known to be very difficult.

II. MODEL AND DEFINITIONS

A. Channel Model

In this paper, we adopt notation and definitions for Markov channels similar to those in [10]. In particular, capital letters, like $X$, represent random variables, and lower case letters, like $x$, represent particular values of the corresponding random variables (or constants if there is no corresponding random variable). Boldface letters, like $\mathbf{X}$ and $\mathbf{x}$, represent vector random variables, and particular values of the corresponding vector random variables (or vector constants), respectively. A boldface capital letter may also represent a constant matrix, such as the transition probability matrix $P$. Calligraphic letters, like $\mathcal{X}$, represent sets. Furthermore, for probability mass functions (PMFs), such as $p_X(x)$, the subscript is omitted when it is unambiguous to do so, resulting in $p(x)$.

Consider a channel with inputs selected from an alphabet $\mathcal{X}$, outputs selected from an alphabet $\mathcal{Y}$, and hidden channel states selected from an alphabet $\mathcal{S}$. For notational convenience, throughout this paper we will assume that $\mathcal{X}$ and $\mathcal{Y}$ are discrete sets; nothing changes if they are continuous, except the substitution of probability density functions (PDFs) for the corresponding PMFs. However, $\mathcal{S}$ is always discrete and finite for a
finite-state Markov channel, so we will usually denote the state alphabet as \( S = \{1, 2, \ldots, |S|\} \).

Let \( X \in \mathbb{X}^n \), \( Y \in \mathbb{Y}^n \), and \( S \in \mathcal{S}^{n+1} \) represent random vectors, consisting of channel inputs, channel outputs, and channel states, respectively. A Markov channel is a channel in which the following two conditions hold. First, the channel is memoryless given the channel state \( S \). Second, \( S \) forms a regular Markov chain operating in steady state, which is independent of the channel inputs \( X \). (Note that the second property excludes partial response channels from the discussion.) With these two properties in mind, we can write

\[
p(y, s | x) = p(s_1) \prod_{t=1}^{n} p(y_t | s_t, x_t)p(s_{t+1} | s_t)
\]

(1)

and the channel input–output relationship \( p(y | x) \) is given by marginalizing (1) over \( s \), which can be calculated efficiently by distributing the summation over the \( s_t \) variables. Throughout this paper, we assume that the initial state distribution \( p(s_1) \) is equal to the steady-state distribution.

From (1), the channel is fully parametrized by specifying \( p(s_{t+1} | s_t) \) and \( p(y_t | s_t, x_t) \). The values of \( p(s_{t+1} | s_t) \) are commonly specified in a \(|S| \times |S| \) matrix \( P \), known as the transition probability matrix. If \( S = \{1, 2, \ldots, |S|\} \), then the element of \( P \) on the \( j \)th row and \( i \)th column is given by

\[
P_{i,j} = p_{S_{t+1}S_t}(j | i).
\]

(2)

We assume that \( p(y_t | s_t, x_t) \) is drawn from a given family of channels, where \( s_t \) corresponds to a particular channel parameter for that family. Let \( \eta \) represent the vector of parameters corresponding to each state \( s_t \), where

\[
\eta = [\eta_1, \eta_2, \ldots, \eta_{|S|}]
\]

(3)

and the value of \( s_t \) selects the value within \( \eta \) to be used at time \( t \). For example, if \( p(y_t | s_t, x_t) \) is drawn from a BSC family, then each possible value of \( s_t \) in \( S \) corresponds to an inversion probability, and

\[
p(y_t | s_t, x_t) = \begin{cases} 1 - \eta_{s_t}, & y_t = x_t \\ \eta_{s_t}, & y_t \neq x_t \end{cases}
\]

(4)

Given the family, a Markov channel \( c \) is completely specified by the parameters

\[
c = (P, \eta).
\]

(5)

For a given family of channels, let \( \mathcal{M} \) represent the set of all valid selections of \( c \). That is, for all \( c = (P, \eta) \in \mathcal{M} \):

1) there exists a discrete, finite state alphabet \( S \) such that \( P \) is an \(|S| \times |S| \) matrix, and \( \eta \) has \(|S| \) elements;

2) \( P \) is a valid transition probability matrix for a regular Markov chain; and

3) each element in \( \eta \) is a valid channel parameter for the given family.

The mutual information for any channel with memory, including Markov channels, is given by

\[
I(X;Y) = \lim_{n \to \infty} \frac{1}{n} I(X^n;Y^n)
\]

where \( I(X;Y) \) represents the mutual information between the length-\( n \) vector random variables \( X \) and \( Y \), so long as the limit exists. In this paper, we assume that the input distribution \( p(x) \) is i.i.d., and that the Markov chain is regular, which are sufficient to ensure that the limit in (6) exists. For a fixed i.i.d. input distribution \( p(x) \), the mutual information is dependent on the channel parameters \( \eta \). To make this dependence explicit, we will write \( I[c;X;Y] \) to represent the mutual information.

In the remainder of the paper, we will assume that the i.i.d. input distribution \( p(x) \), as well as the family corresponding to \( p(y_t | s_t, x_t) \), are given, and our mutual information comparisons will be made on that basis. In other words, we are not interested in comparing the mutual information of systems with different input distributions, nor are we interested in comparing the mutual information in, e.g., BSC families with the mutual information in Gaussian families.

B. Mixing Operator

In describing our main results, we use the \( (\leftrightarrow, \mu_{12}, \mu_{21}) \) operator, originally defined in [10] (note that this operator is parametrized by \( \mu_{12} \) and \( \mu_{21} \), where \( 0 \leq \mu_{12} \leq 1 \) and \( 0 \leq \mu_{21} \leq 1 \)). This operator is used to “mix” a pair of Markov channels to produce a new Markov channel with a larger state alphabet, and is described briefly as follows. Let \( \mathcal{C}_1, \mathcal{C}_2 \in \mathcal{M} \) represent two Markov channels from the same family, where \( \mathcal{C}_1 = (P_1, \eta_1) \) and \( \mathcal{C}_2 = (P_2, \eta_2) \). Let \( S_1 \) represent the state alphabet for \( \mathcal{C}_1 \), and let \( p_1 \) represent the steady-state probability distribution over \( S_1 \) (resp., \( S_2 \) and \( p_2 \) for \( \mathcal{C}_2 \)). Let the notation \( [p_1]_j \), \( p_1 \) represent a matrix with \( j \) rows, where each row is identical and equal to \( p_1 \) (with \( [p_2]_j \) defined similarly). Then the statement

\[
c = \mathcal{C}_1 \leftrightarrow_{\mu_{12}, \mu_{21}} \mathcal{C}_2
\]

(7)

signifies that \( c = (P, \eta) \), where

\[
P = \begin{bmatrix} (1 - \mu_{12})P_1 & \mu_{12}[p_2]_{S_1} \\ \mu_{21}[p_1]_{S_2} & (1 - \mu_{21})P_2 \end{bmatrix}
\]

(8)

and

\[
\eta = [\eta_1, \eta_2].
\]

(9)

Thus, \( c \in \mathcal{M} \), and the state alphabet for \( c \) contains \( |S_1| + |S_2| \) states.

The reader is directed to [10] for the details, but the operator \( \leftrightarrow \) has the following physical interpretation. Suppose the probabilistic state machines for \( \mathcal{C}_1 \) and \( \mathcal{C}_2 \) exist in parallel. Given a starting state in \( S_1 \), with probability \( (1 - \mu_{12}) \) the next state is chosen from \( S_1 \), according to the Markov transition probabilities \( P_1 \) (i.e., the usual evolution of the probabilistic state machine); otherwise, with probability \( \mu_{12} \), the next state is chosen from \( S_2 \), at random from the elements of \( S_2 \) according to the steady-state probabilities \( p_2 \), and independently of any previous state. For a starting state in \( S_2 \), the same situation applies, exchanging \( S_1 \) and \( S_2 \), and substituting \( \mu_{21} \) for \( \mu_{12} \). That is, from one state to the next, the Markov chain “jumps” from \( \mathcal{C}_1 \) to \( \mathcal{C}_2 \) with probability \( \mu_{12} \), or from \( \mathcal{C}_2 \) to \( \mathcal{C}_1 \) with probability \( \mu_{21} \); if no jump occurs, the state evolves within the state machine from \( \mathcal{C}_1 \) or \( \mathcal{C}_2 \) as usual.
III. MAIN RESULT

For each Markov channel $c = (P, \eta)$, given an i.i.d. input distribution $p(x)$ and a family corresponding to $p(y \mid s_t, x_t)$, let $D(c, p(x), p(y \mid s_t, x_t)) \subseteq M$ represent the set of channels with equal or smaller mutual information, that is,

$$D(c, p(x), p(y \mid s_t, x_t)) = \{c^* : c^* \in M, I[c^*(X,Y)] \leq I[c](X,Y)\}. \quad (10)$$

Clearly, membership or nonmembership in $D(c, p(x), p(y \mid s_t, x_t))$ represents an ordering, comparing each channel in $M$ with $c$. The challenge is then to find the members of $D(c, p(x), p(y_{s_t, x_t}))$. This challenge is partially addressed by the following theorem, which is the main result of this paper.

**Theorem 1:** For any $c^* \in D(c, p(x), p(y_{s_t, x_t}))$

$$c^* \leftrightarrow c^* \in D(c, p(x), p(y_{s_t, x_t})) \quad (11)$$

for all $\mu_{12}, \mu_{21}$ such that $0 \leq \mu_{12} \leq 1, 0 \leq \mu_{21} \leq 1$.

Prior to proving the theorem, we give a lemma that is useful to the proof of the main result. For a vector random variable $X$, let the notation $X_1^k$ represent the segment of that vector from index 1 to index $k$, inclusive. Then we have the following.

**Lemma 1:** Let $c$ represent a Markov channel. Then, if the input distribution is i.i.d.

$$I[c](X;Y) \leq I[c](X_1^k;Y_1^k) \quad (12)$$

for any $k < \infty$.

This lemma makes rigorous the intuition that observing the channel briefly never gives more mutual information than observing the channel over an asymptotically long period of time. The proof for the Lemma is contained in the Appendix.

The proof of Theorem 1 is then given as follows.

**Proof:** For convenience, let $c' = [c \leftrightarrow c^*]$. Let $S$ and $S^*$ represent the state alphabets corresponding to $c$ and $c^*$, respectively. At each time $t$, the mixed channel $c'$ is in a state belonging either to $S$, or to $S^*$. Let $U_t$ be a random variable which indicates the alphabet containing $s_t$, defined as

$$u_t = \begin{cases} 1, & s_t \in S \\ 2, & s_t \in S^* \end{cases} \quad (13)$$

Furthermore, let $U = [U_1, U_2, \ldots, U_n]$ represent the corresponding random vector.

From the definition of $U_t$ and the description of $\mu_{12}, \mu_{21}$, note that

$$p(s_{t+1} \mid s_t, u_t) = p(s_{t+1} \mid s_t, u_{t+1}, u_t) \quad (14)$$

$$= \begin{cases} p(s_{t+1} \mid s_t), & u_{t+1} = u_t \\ p(s_{t+1}), & u_{t+1} \neq u_t \end{cases} \quad (15)$$

where $p(s_{t+1} \mid s_t)$ represents the usual Markov relationship in the Markov channel corresponding to $u_{t+1}$, and $p(s_{t+1})$ represents the marginal steady-state probabilities in the Markov channel corresponding to $u_{t+1}$. From (15), each time $u_{t+1} \neq u_t$, the Markov chain is partitioned into independent segments. Let $J$ represent an index set corresponding to the independent segments, let the subscript $i, j$ represent the $i$th and $j$th symbol in the $j$th segment, and let $l(j)$ represent the length of the $j$th segment. Then we can write

$$p(s \mid u) = \prod_{j \in J} \prod_{i=1}^{l(j)-1} p(s_{i+1,j} \mid s_{i,j}) \quad (16)$$

For each $j \in J$, let $c_j$ represent the Markov channel, which is either $c$ or $c^*$. From (16), since the channel input distribution is i.i.d., and since the elements of $Y$ are conditionally independent given $X$ and $S$, it is apparent that $p(y \mid x, u)$ and $p(y \mid u)$ are also partitioned into independent segments. Thus, letting $X_1^{(j)}$ (resp., $Y_1^{(j)}$) represent the first through $l(j)$th channel inputs (resp., outputs) of segment $j$, we can write

$$I[c_j](X;Y \mid U) = E \sum_{j \in J} I[c_j](X_1^{(j)};Y_1^{(j)}) \quad (17)$$

where the expectation is taken over $J$ and $l(j)$, which are functions of the random variables $U$.

Since $U$ is independent of $X$, it is true that

$$I[c_j](X;Y) \leq I[c_j](X;Y, U) \quad (18)$$

$$= I[c_j](X;Y \mid U) + I[c_j](X;U) \quad (19)$$

$$= I[c_j](X;Y) \quad (20)$$

Because the distribution of $X$ is i.i.d., and because each Markov chain segment starts out with the steady-state distribution from (16), the distributions of $X_1^{(j)}$ and $Y_1^{(j)}$ are only dependent on $j$ through the channel $c_j$ and the segment length $l(j)$. That is, $X_1^{(j)}$ has the same distribution as any length-$l(j)$ vector of channel inputs (written $X_1^{(j)}$), and, similarly, $Y_1^{(j)}$ has the same distribution as any length-$l(j)$ vector of channel outputs from $c_j$ (written $Y_1^{(j)}$). Thus, we can rewrite (17) as

$$I[c_j](X;Y \mid U) = E \left[ \sum_{j \in J} I[c_j](X_1^{(j)};Y_1^{(j)}) \right] \quad (21)$$

Now, we can write

$$E \left[ \sum_{j \in J} I[c_j](X_1^{(j)};Y_1^{(j)}) \right] \leq E \left[ \sum_{j \in J} l(j) I[c_j](X;Y) \right] \quad (22)$$

$$\leq E \left[ \sum_{j \in J} l(j) I[c](X;Y) \right] \quad (23)$$

$$= n I[c](X;Y) \quad (24)$$

where inequality (22) follows from Lemma 1; inequality (23) follows from the fact that $c_j$ is either $c$ or $c^*$, which from (10) means that $I[c_j](X;Y) \leq I[c](X;Y)$ for all $j$; and equality
(24) follows from the fact that the sum of the lengths $l(j)$ of all the segments equal the length $n_2$ of the sequence, regardless of how the sequence is divided.

From (20)–(24), and the definition of $I[c’](X; Y)$, we have that

$$I[c’](X; Y) = \lim_{n \to \infty} \frac{1}{n} I[c’](X; Y)$$

$$\leq \lim_{n \to \infty} \frac{1}{n} I[c](X; Y | U)$$

$$\leq I[c](X; Y)$$

which proves the theorem.

It follows from Theorem 1 and (11) that

$$I \left[ c \longleftrightarrow c^* \right](X; Y) \leq I[c](X; Y),$$

(28)

In fact, we can sharpen the bound in (28) somewhat. Letting $\alpha$ represent the number of channel uses in $c$, and letting $\alpha^*$ represent the number of channel uses in $c^*$, the right-hand side of (22) can be rewritten as

$$E[\alpha] I[c](X; Y) + E[\alpha^*] I[c^*](X; Y),$$

(29)

The random process governing jumps between $c$ and $c^*$ is itself a two-state Markov chain; thus, the steady-state probabilities of $c$ and $c^*$ are $\mu_{21}(\mu_{12} + \mu_{21})$ and $\mu_{12}(\mu_{12} + \mu_{21})$, respectively [2]. As a result, $E[\alpha] = \mu_{21}/(\mu_{12} + \mu_{21})$, and $E[\alpha^*] = \mu_{12}/(\mu_{12} + \mu_{21})$. Then it is true that

$$I \left[ c \longleftrightarrow c^* \right](X; Y) \leq \frac{\mu_{21} I[c](X; Y) + \mu_{12} I[c^*](X; Y)}{\mu_{12} + \mu_{21}}$$

(30)

as a corollary to Theorem 1. Since the quantity on the right-hand side of (30) is an average between $I[c](X; Y)$ and $I[c^*](X; Y)$, and since $I[c^*](X; Y) \leq I[c](X; Y)$ from (10), the right-hand side of (30) is always less than or equal to $I[c](X; Y)$. 

IV. EXAMPLE AND DISCUSSION

In this section, we give an example to illustrate the use of the main result, and discuss the relationship of the result with existing papers in the literature, especially [10] and [2].

A. Example Illustrating the Main Result

Here, we give a straightforward example to illustrate the use of the main result. Suppose $c$ is a GE channel, with parameters

$$c = (P, \eta) = \left( \begin{array}{cc} 0.9 & 0.1 \\ 0.1 & 0.9 \end{array} \right), [0.1, 0.4].$$

(31)

Suppose $c^*$ is a channel formed by concatenating $c$ with a BSC, which has inversion probability $\nu = 0.1$. Then $c^*$ is also a GE channel with parameters [4]

$$c^* = (P^*, \eta^*) = \left( \begin{array}{cc} 0.9 & 0.1 \\ 0.1 & 0.9 \end{array} \right), [0.18, 0.42].$$

(32)

Then $c^*$ is degraded with respect to $c$, and so $c^* \in \mathcal{D}(c; p(x), p(y|x_1, x_4))$.

Let $\mu_{12} = 0.1$, and let $\mu_{21} = 0.2$. The marginal state probability vector is given by $\nu = [0.5, 0.5]$ for $c$, and this vector is the same in $c^*$. Forming $\mu_{12} \mu_{21}$, from (8) and (9) we have that

$$c \longleftrightarrow c^* (P^*, \eta^*) = (P', \eta'),$$

(33)

where

$$P' = \begin{bmatrix} 0.81 & 0.09 & 0.05 & 0.05 \\ 0.09 & 0.81 & 0.05 & 0.05 \\ 0.10 & 0.10 & 0.72 & 0.08 \\ 0.10 & 0.10 & 0.08 & 0.72 \end{bmatrix}$$

(34)

and

$$\eta' = [0.1, 0.4, 0.18, 0.42].$$

(35)

Finally, from Theorem 1, we can conclude that $c \rightarrow c^* \in \mathcal{D}(c; p(x), p(y|x_1, x_4))$, and so

$$I[c](X; Y) \geq I \left[ c \longleftrightarrow c^* \right](X; Y).$$

(36)

Under an equiprobable i.i.d. input distribution, we have that $I[c](X; Y) = 0.203$, $I[c^*](X; Y) = 0.124$, and $I[c \leftrightarrow c^*](X; Y) = 0.170$, which is in line with both Theorem 1 and (30).

We make two remarks on this example. First, we note that the BSC inversion probability $\nu$ is a parameter of $c^*$, and $\mu_{12}$ and $\mu_{21}$ are parameters of the mixing operator. Since any valid setting of these three parameters gives a new Markov channel in $\mathcal{D}(c; p(x), p(y|x_1, x_4))$, the operator $\mu_{12} \mu_{21}$ casts a three-dimensional “shadow” of degraded channels in the parameter space $\mathcal{M}$. Second, since the new channel returns to $\mathcal{D}(c; p(x), p(y|x_1, x_4))$, we can use the operator recursively to find channels with an arbitrary number of states that are also in $\mathcal{D}(c; p(x), p(y|x_1, x_4))$.

B. Relationship to [10] and [2]

As we mentioned in earlier sections, the operator $\mu_{12} \mu_{21}$ was introduced in [10], where it was used to prove an analogous ordering result for the decoding of codes on sparse graphs, such as LDPC codes. Furthermore, in [10], it was shown that the operator can be generalized to mix more than two Markov channels, and the ordering in this paper can be easily adapted to the generalized operator. However, the result in [10] can only be used to analyze channels mixed with $\mu_{12} \mu_{21}$ if both channels started out as members of a set of channels with particular properties, known as fuzzy-edge degraded channels. In this paper, meaningful results are obtained for any pair of channels, so long as it is known which one has the larger mutual information. As a result, the ordering given in this paper has a much wider scope than the one given in [10].

A related result, constrained to the GE channel, was presented in [2, Theorem 5], while the proof of that result contained a technique similar to our mixing operator (indeed, that technique formed the inspiration for our mixing operator). In particular,
suppose \( \mathcal{C} \) is a GE channel; then using our notation, the method from [2] operates by forming the mixed channel \( \mathcal{C} \rightarrow \mathcal{C} \) (that is, \( \mathcal{C} \) is mixed with itself, where \( \mu_{12} \) and \( \mu_{21} \) are both equal to \( \mu \)). From (10), note that \( \mathcal{C} \in \mathcal{D}(C; p(x), p(y | s_t, x_t)) \), so \( \mathcal{C} \rightarrow \mathcal{C} \) is \( \mathcal{C} \in \mathcal{D}(C; p(x), p(y | s_t, x_t)) \) from Theorem 1. Since \( \mathcal{C} \rightarrow \mathcal{C} \) is a GE channel, it initially appears that \( \mathcal{C} \rightarrow \mathcal{C} \) is a four-state binary Markov channel, which, from (8) and (9), is given by

\[
\mathcal{C} \rightarrow \mathcal{C} = \left( \begin{array}[]{c}
(1 - \mu) \mathbf{P}^\mathcal{C} \mu_{12} \mathbf{P}^\mathcal{C} \mu_{21} \mathbf{P}^\mathcal{C}
\end{array} \right).
\]  

(37)

However, for any pair of channel inputs \( x \in \{0, 1\}^n \) and channel outputs \( y \in \{0, 1\}^n \), it can be shown that \( p(y | x) \) is always the same in both the channel from (37) and in a GE channel \( \mathcal{C}' \), where \( \mathcal{C}' \) is given by

\[
\mathcal{C}' = ((1 - \mu) \mathbf{P}^\mathcal{C} + \mu_{12} \mathbf{P}^\mathcal{C} \mu_{21} \mathbf{P}^\mathcal{C}) \eta.
\]  

(38)

Thus, we can conclude that \( I[c] \rightarrow \mathcal{C} | X; Y \rangle = I[c'] | X; Y \rangle \), so \( \mathcal{C}' \in \mathcal{D}(C; p(x), p(y | s_t, x_t)) \). The property that

\[
I[c'] | X; Y \rangle \leq I[c] | X; Y \rangle
\]

was the main idea expressed in [2, Theorem 5], which was stated as an ordering on GE channels.

In [10], the term degenerate Markov channel was used to describe any Markov channel which could be represented with fewer states. For instance, here the four-state Markov channel in (37) is degenerate, since it can be represented as the two-state Markov channel \( \mathcal{C}' \). Note that if \( \mu_{12} \neq \mu_{21} \), \( \mathcal{C} \rightarrow \mathcal{C} \) is no longer degenerate in general, and any nondegenerate (or non-binary) case is beyond what was considered in [2]. The result from [2] may therefore be thought of as a special case of this result.

V. CONCLUSION

In this paper, we presented a method for ordering finite-state Markov channels with respect to mutual information, which generalizes previous work from [10] and [2]. Mutual information for a Markov channel is generally not available in closed form, so the result in this paper is useful to quickly characterize the large parameter spaces of these channels.

APPENDIX

PROOF OF LEMMA 1

Suppose a length-\( n \) vector of channel inputs \( X \) (where \( n = dk \) for integer \( d \)), and the corresponding vector of channel observations \( Y \) are broken up into blocks of \( k \), so that for each \( i \in \{1, 2, \ldots, d\} \), there is a block of channel inputs \( X^{ik}_{t+k} \), and there is a corresponding block of channel outputs \( Y^{ik}_{t+k} \). Considering block \( i \) in isolation, the mutual information for all its \( k \) channel uses is

\[
I[c] \left( X^{ik}_{t+k} ; Y^{ik}_{t+k} \right) = E \left[ \log \frac{p(Y^{ik}_{t+k+1} | X^{ik}_{t+k+1})}{p(Y^{ik}_{t+k+1})} \right] \tag{39}
\]

where \( E[] \) represents expectation. Thus, the total throughput simultaneously achievable by examining all \( d \) blocks in isolation is

\[
\sum_{i=1}^{d} I[c] \left( X^{ik}_{t+k} ; Y^{ik}_{t+k} \right) = E \left[ \log \frac{p(Y^{ik}_{t+1} | X^{ik}_{t})}{p(Y^{ik}_{t+1})} \right]. \tag{40}
\]

Since the Markov chain starts out in steady state, and since the input distribution is i.i.d., the distributions of \( X^{ik}_{t+k} \) and \( Y^{ik}_{t+k+1} \) are independent of \( i \), so

\[
I[c] \left( X^{ik}_{t+k} ; Y^{ik}_{t+k+1} \right) = I[c] \left( X^{ik}_{t+k} ; Y^{ik}_{t+1} \right). \tag{41}
\]

Thus, (40) becomes

\[
\sum_{i=1}^{d} I[c] \left( X^{ik}_{t+k} ; Y^{ik}_{t+k+1} \right) = dI[c] \left( X^{k} ; Y^{k} \right). \tag{42}
\]

However, the right-hand side of (40) can be rewritten as

\[
E \left[ \log \frac{p(Y^{ik}_{t+k+1} | X^{ik}_{t+k+1})}{p(Y^{ik}_{t+k+1})} \right] = E \left[ \log \frac{p(Y^{ik}_{t+1} | X^{ik}_{t})}{p(Y^{ik}_{t+1})} \right]. \tag{43}
\]

From the auxiliary channel lower bound (see [11]), it is true that

\[
E \left[ \log \frac{p(Y^{ik}_{t+k+1} | X^{ik}_{t+k+1})}{p(Y^{ik}_{t+k+1})} \right] \leq E \left[ \log \frac{p(Y^{ik} | X^{ik})}{p(Y^{ik})} \right]. \tag{44}
\]

Finally

\[
I[c] \left( X^{k} ; Y^{k} \right) = \lim_{d \to \infty} \frac{1}{d} E \left[ \log \frac{p(Y^{ik} | X^{ik})}{p(Y^{ik})} \right] \tag{45}
\]

\[
I[c] \left( X^{k} ; Y^{k} \right) = \lim_{d \to \infty} \frac{1}{d} E \left[ \log \frac{p(Y^{ik} | X^{ik})}{p(Y^{ik})} \right] \tag{46}
\]

\[
I[c] \left( X^{k} ; Y^{k} \right) \geq \lim_{d \to \infty} \frac{1}{d} dI[c] \left( X^{k} ; Y^{k} \right) \tag{47}
\]

\[
I[c] \left( X^{k} ; Y^{k} \right) \geq \frac{1}{k} I[c] \left( X^{k} ; Y^{k} \right) \tag{48}
\]

where (46) follows from the definition of mutual information for Markov channels, and (48) follows from (40)–(45).

REFERENCES

[1] E. O. Elliott, “Estimates of error rates for codes on burst-noise channels,” Bell Syst. Tech. J., vol. 42, pp. 1977–1997, Sep. 1963.

[2] M. Mushkin and I. Bar-David, “Capacity and coding for the Gilbert-Elliott channels,” IEEE Trans. Inf. Theory, vol. 35, no. 6, pp. 1277–1290, Nov. 1989.
[3] J. Garcia-Frias and J. D. Villasenor, “Turbo decoding of Gilbert-Elliot channels,” *IEEE Trans. Commun.*, vol. 50, no. 3, pp. 357–363, Mar. 2002.

[4] A. W. Eckford, F. R. Kschischang, and S. Pasupathy, “Analysis of low-density parity-check codes for the Gilbert-Elliott channel,” *IEEE Trans. Inf. Theory*, vol. 51, no. 11, pp. 3872–3889, Nov. 2005.

[5] A. J. Goldsmith and P. P. Varaiya, “Capacity, mutual information, and coding for finite-state Markov channels,” *IEEE Trans. Inf. Theory*, vol. 42, no. 3, pp. 868–886, May 1996.

[6] P. Bergamo, D. Maniezzo, A. Giovanardi, G. Mazini, and M. Zorzi, “Improved Markov model for Rayleigh fading,” *Electron. Lett.*, vol. 38, pp. 477–478, May 2002.

[7] W. Turin and M. Zorzi, “Performance analysis of delay-constrained communications over slow Rayleigh fading channels,” *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 801–807, Oct. 2002.

[8] R. K. Guha and S. Sarkar, “Characterizing temporal SNR variation in 802.11 networks,” in *Proc. IEEE Wireless Communications and Networking Conf.*, Las Vegas, NV, Apr. 2006, pp. 1408–1413.

[9] H. Lou and J. Garcia-Frias, “Low-density generator matrix codes for indoor and Markov channels,” *IEEE Trans. Wireless Commun.*, vol. 6, no. 7, pp. 1436–1445, Apr. 2007.

[10] A. W. Eckford, F. R. Kschischang, and S. Pasupathy, “A partial ordering of general finite-state Markov channels under LDPC decoding,” *IEEE Trans. Inf. Theory*, vol. 53, no. 6, pp. 2072–2087, Jun. 2007.

[11] D. M. Arnold, H.-A. Loeliger, P. O. Vontobel, A. Kavčič, and W. Zeng, “Simulation-based computation of information rates for channels with memory,” *IEEE Trans. Inf. Theory*, vol. 52, no. 8, pp. 3498–3508, Aug. 2006.

Andrew W. Eckford (M’96–S’97–M’04) is originally from Edmonton, AB, Canada. He received the B.Eng. degree in electrical engineering from the Royal Military College of Canada, Kingston, ON, Canada, in 1996, and the M.A.Sc. as well as Ph.D. degrees in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 1999 and 2004, respectively.

He is currently an Assistant Professor of Computer Science and Engineering at York University, Toronto, ON, Canada, a post he has held since July 2006.