Capacity of Time-Varying Channels with Channel Side Information at the Sender and Receiver.

Muriel Médard
University of Illinois at Urbana-Champaign
medard@comm.csl.uiuc.edu

Andrea J. Goldsmith
Stanford University
andrea@zinfandel.stanford.edu

Abstract. The capacity of time-varying asymptotically block memoryless channels with causal channel side information at the sender and receiver is considered. We briefly present a formal weak coding theorem and its converse. We then show that our coding theorem can be used to obtain the capacity of time-varying ISI channels, which cannot generally be considered using traditional decomposition methods. For perfect instantaneous side information, the form of the input distribution that maximizes capacity is found to be conditional Gaussian. We also consider the capacity of flat fading channels with imperfect channel estimates.

1. INTRODUCTION.

We study stationary and ergodic time-varying channels with memory where casual channel side information is known at each time sample by the sender and receiver. We consider asymptotically block memoryless channels. The variations of the channel suggest that the sender should adapt its transmission to the channel. Many practical transmission schemes adapt to channel conditions at the transmitter and/or receiver. However, determining the optimal adaptation is difficult because the channel memory may introduce self-interference, which the adaptation must consider ([1], [?]).

Most previous results on capacity of time-varying channels with side information either assume memoryless channels or side information at the receiver only. Memoryless channels with perfect channel state information at the transmitter and receiver ([3], [4], [5]), and at the receiver alone ([6], [7]) have been considered. The capacity of a memoryless arbitrarily time-varying channel was obtained in [14]. Other papers have considered the capacity of time-varying channels without side information ([8], [9], [10], [11], [12], [13]). In [15] a general channel capacity was derived. This result can be extended to the case of receiver ([16]), but not transmitter, side information. Lack of side information at the transmitter prevents adaptation to the changing channel. Our results also differ from those for a finite-state Markov channel (FSMC) with side information which is perfect at the receiver and delayed at the transmitter ([17]).

In Section 2, we define capacity and obtain a weak coding theorem and its converse, under some assumptions about channel decorrelation. Section 3 gives examples of channels to which our coding theorem applies. The first example is that of perfectly known ISI channels with bounded variance of the taps. Note that this result is not trivial, since traditional decomposition methods to btain capacity using decomposition techniques do not apply when we have time variations. The second example is that of a channel with imperfect chanel estimates.

2. MODEL AND CODING THEOREM.

We consider a discrete time-varying channels, with memory. The channel variations are stationary. At time n both transmitter and receiver have side information $k_n \in \mathcal{K}$, possibly imperfect and delayed, about the channel state, where the set $\mathcal{K}$ is finite. The channel input at time n is x_n and the output is y_n. We denote vectors thus: $\underline{v}^n = (v_1, \ldots, v_n)$, $\underline{v}_m^n = (v_m, \ldots, v_n)$. Lower case denotes sample values and upper case denotes r.v.s. We now define our channel properties.

Definition 1: The triplet $(\underline{y}^m, \underline{k}^m, \underline{x}^m)$ is ϵ-block-memoryless in n samples iff

$$\prod_{i=0}^{N-1} p\left(\underline{y}_{in+1}^{(i+1)n} | \underline{k}_{in+1}^{(i+1)n}, \underline{x}_{in+1}^{(i+1)n}\right) p\left(\underline{y}_{nN+1}^{m} | \underline{k}_{nN+1}^{m}, \underline{x}_{nN+1}^{m}\right)(1-\epsilon)^n \leq p\left(\underline{y}^m | \underline{k}^m, \underline{x}^m\right) \quad (1)$$

and the corresponding upper bound holds by replacing $-\epsilon$ by $+\epsilon$. Iff a triplet satisfies these bounds, we say that $(\underline{y}^m, \underline{k}^m, \underline{x}^m) \in \mathcal{D}_{n,\epsilon}(\underline{y}, \underline{k}, \underline{x})$.

Definition 2: A channel is asymptotically block memoryless iff for every input distribution, $\forall \epsilon_0, \epsilon_1$, an integer n_0, $\exists n \geq n_0$ and j s. t. $\forall m \geq jn$

$$\int_{(\underline{y}^m, \underline{k}^m, \underline{x}^m) \in \mathcal{D}_{n,\epsilon_0}(\underline{y}, \underline{k}, \underline{x})} p\left(\underline{y}^m, \underline{k}^m, \underline{x}^m\right) \geq 1 - \epsilon_1.$$

These definitions extend the notion of typical (message, received sequence), so errors may occur because, for a typical channel, we had an atypical transmission or because the channel was atypical. We assume the following input constraints. First, $\frac{1}{m}\mathrm{E}\left(\sum_{j=1}^{m} X_j^2\right) \leq S$, $\forall m$ (Input Energy Constraint). Second, $p\left(\underline{x}^m | \underline{k}^m\right) = \prod_{j=1}^{m} p\left(x_j | \underline{x}^{j-1}, \underline{k}^j\right)$ (Causal Constraint). Finally, for the input and output alphabets we use, $\exists A < \infty$, s. t.

$$p\left(\underline{y}^j | \underline{k}^j, \underline{x}^j\right) < A < \infty \ \forall \left(\underline{y}^j, \underline{k}^j, \underline{x}^j\right) \quad (2)$$

$$p\left(\underline{x}^j | \underline{k}^j, \underline{y}^j\right) < A < \infty \ \forall \left(\underline{y}^j, \underline{k}^j, \underline{x}^j\right) \quad (3)$$

(Bounded Probability Constraint). We define $C = lim_{n\to\infty} max_{p(\underline{x}^n|\underline{K}^n)} \left[\frac{1}{n} I\left(\underline{X}^n;\underline{Y}^n|\underline{K}^n\right)\right]$. If C exists, it is the channel capacity.

The gist of our coding theorem to achieve this capacity is to take n large enough so that we approach capacity over an interval of length n and m large enough that we can consider many IID components, for which the weal law of large numbers (WLLN) applies. The details of the coding theorem can be found in [2]. We first find n sufficiently large that we have $C - max_{p(\underline{x}^n|\underline{k}^n)} \left[\frac{1}{n} I\left(\underline{X}^n;\underline{Y}^n|\underline{K}^n\right)\right] \leq \frac{\epsilon}{3}$ where $p_n^*\left(\underline{x}^n|\underline{k}^n\right)$ is an input distribution which achieves maximum mutual information for a given n. The codeword is decoded using typical set decoding. We define the typical set A_ϵ^m as the set of all triplets $\left(\underline{y}^m,\underline{k}^m,\underline{x}^m\right)$ that satisfy $\left(\underline{y}^m,\underline{k}^m,\underline{x}^m\right) \in \mathcal{D}_{n,\epsilon_0}(\underline{y}^m,\underline{k}^m,\underline{x}^m)$ and $\left|\frac{i(\underline{x}^m;\underline{y}^m|\underline{k}^m)}{m} - C\right| < \epsilon$. We decode as follows. If there is a unique $\underline{x}^m$ s.t. $\left(\underline{y}^m,\underline{k}^m,\underline{x}^m\right) \in A_\epsilon^m$, we decide $\underline{x}^m$ was transmitted. Otherwise, we declare an error.

For our coding theorem, we design the code in blocks. Within each block the code achieves the maximum mutual information for the given side information sequence. Our coding theorem states that

Theorem 1: C is achievable.

We outline the proof in [2] below. We generate codebooks gradually, as side information becomes available. For each $\underline{k}^m$, we generate 2^{mR} codewords of length m in blocks of size n as follows. For each i we independently generate the first symbol $x_1(i)$ of the i^{th} codeword according to $p_1^*(x|k_1)$, yielding a partial codebook. We generate the next symbol by conditioning the distribution on $x_1(1), k_1, k_2$, and so on. The family of codebooks for each k^m is finally made known to the receiver. The state information alphabet is finite, so this transmission requires only a finite amount of time.

We now show that the average error probability is upper bounded by ϵ for any side information sequence. The average probability of error is obtained by averaging over the codewords of each codebook. WLOG, assume the first codeword is chosen. Let $E_i = \left\{\left(\underline{x}^m(i),\underline{y}^m,\underline{k}^m\right) \in A_\epsilon^m\right\}$ $\forall i \in \left\{1,\ldots,2^{mR}\right\}$. The error probability is upper bounded by $p\left(E_1^C\right) + \sum_{i\neq 1} p\left(E_i\right)$. We have

$$p\left(E_1^C\right) \leq p\left(\left(\underline{y}^m,\underline{k}^m,\underline{x}^m\right) \notin \mathcal{D}_{n,\epsilon_0}(\underline{k},\underline{x})\right) + p\left(\left|i\left(\underline{x}^m;\underline{y}^m|\underline{k}^m\right) - nC\ \right| \geq \epsilon\right) \tag{4}$$

Consider now the RHS of (4). The first term is bounded by $\epsilon/2$ by our decorrelation asumptions. Let us bound the second term. Let $N = \lfloor\frac{m}{n}\rfloor$. Then, from Definitions 1 and 2 and our block memoryless codebooks, $p(\underline{y}^m|\underline{k}^m) \geq \prod_{i=0}^{N-1} p(\underline{y}_{in+1}^{(i+1)n}|\underline{k}_{in+1}^{(i+1)n})p(\underline{y}_{Nn+1}^m|\underline{k}_{Nn+1}^m)(1-\epsilon_0)^m$ and upper bounded similarly by replacing $-\epsilon_0$ with $+\epsilon_0$. Hence,

$$\begin{aligned}
\frac{1}{m} i\left(\underline{x}^m;\underline{y}^m|\underline{k}^m\right) &= \frac{1}{m}\left[-ln\left(p\left(\underline{y}^m|\underline{k}^m\right)\right)\right. \\
&\quad \left. + ln\left(p\left(\underline{y}^m|\underline{x}^m,\underline{k}^m\right)\right)\right] \\
&\leq \frac{1}{m}\left[-ln\left(\prod_{i=0}^{N-1} p^*\left(\underline{y}_{in+1}^{(i+1)n}|\underline{k}_{in+1}^{(i+1)n}\right)\right.\right. \\
&\quad \left. p^*\left(\underline{y}_{Nn+1}^m|\underline{k}_{Nn+1}^m\right)(1-\epsilon_0)^m\right) \\
&\quad + ln\left(\prod_{i=0}^{N-1} p^*\left(\underline{y}_{in+1}^{(i+1)n}|\underline{k}_{in+1}^{(i+1)n},\underline{x}_{in+1}^{(i+1)n}\right)\right. \\
&\quad \left.\left. p^*\left(\underline{y}_{Nn+1}^m|\underline{k}_{Nn+1}^m,\underline{x}_{Nn+1}^m\right)(1+\epsilon_0)^m\right)\right] \\
&\leq \frac{1}{m}\left[-\sum_{i=0}^{N-1}\left[ln\left(p^*\left(\underline{y}_{in+1}^{(i+1)n}|\underline{k}_{in+1}^{(i+1)n}\right)\right)\right.\right. \\
&\quad \left. -ln\left(p^*\left(\underline{y}_{in+1}^{(i+1)n}|\underline{k}_{in+1}^{(i+1)n},\underline{x}_{in+1}^{(i+1)n}\right)\right)\right] \\
&\quad + -ln\left(p^*\left(\underline{y}_{Nn+1}^m|\underline{k}_{Nn+1}^m\right)\right) \\
&\quad \left. + ln\left(p*\left(\underline{y}_{Nn+1}^m|\underline{k}_{Nn+1}^m,\underline{x}_{Nn+1}^m\right)\right)\right] + \frac{\epsilon}{4},
\end{aligned} \tag{5}$$

A similar lower bound holds by replacing $\frac{\epsilon}{4}$ with $-\frac{\epsilon}{4}$. The following lemma, whose proof can be found in [2], allows us to handle the end terms in (5).

Lemma 1: Assume $n \in \mathcal{N}$ fixed. Then $\forall \mu \in \mathcal{N}$ s.t. $1 \leq \mu \leq n$, and $\forall \epsilon' > 0, \epsilon'' > 0$, $\exists \mathcal{M}$ s.t. $\forall m > \mathcal{M}$,

$$p^*\left(\left(\underline{y}^\mu,\underline{k}^\mu,\underline{x}^\mu\right) : \frac{1}{m}\left|ln\left(p^*\left(\underline{y}^\mu|\underline{k}^\mu\right)\right) - ln\left(p^*\left(\underline{y}^\mu|\underline{k}^\mu,\underline{x}^\mu\right)\right)\right| > \epsilon'\right) < \epsilon''. \tag{6}$$

Using our stationarity assumptions, let us now create two sets of IID r.v.s. The first set of r.v.s takes, with p.d.f. $p^*\left(\underline{y}_{in+1}^{(i+1)n},\underline{k}_{in+1}^{(i+1)n}\right)$, the value equal to the ln of the p.d.f. The second set of r.v.s takes, with p.d.f. $p^*\left(\underline{y}_{in+1}^{(i+1)n},\underline{k}_{in+1}^{(i+1)n},\underline{x}_{in+1}^{(i+1)n}\right)$, the value equal to the ln of the p.d.f. We then apply the WLLN to these sets of r.v.s. For n large enough, $p^*\left(\left|I\left(\underline{X}^n;\underline{Y}^n|\underline{K}^n\right) - nC\right| \geq \epsilon\right) < \frac{\epsilon}{2}$. Thus, each term on the RHS of (4) is bounded above by $\epsilon/2$, so we have $P\left(E_i^C\right) \leq \epsilon$.

We must still upper bound $\sum_{i\neq 1} P\left(E_i\right)$.

$$P\left(E_i\right) = \int_{\left(\underline{x}^m(j),\underline{y}^m,\underline{k}^m\right)\in A_\epsilon^m} p\left(\underline{x}^m|\underline{k}^m\right) p\left(\underline{y}^m|\underline{k}^m\right) p\left(\underline{k}^m\right) \tag{7}$$

since $\underline{x}^m(i)$ and $\underline{y}^m$ are independent for $i \neq 1$ conditioned on $\underline{k}^m$ and on $\underline{x}^m(1)$ having been transmitted. For $\left(\underline{y}^m,\underline{k}^m,\underline{x}^m\right) \in A_\epsilon^m$,

$$p\left(\underline{y}^m|\underline{k}^m\right) p\left(\underline{x}^m(j)|\underline{k}^m\right) \leq p\left(\underline{y}^m,\underline{x}^m(j)|\underline{k}^m\right) 2^{-m(C-\epsilon)}. \tag{8}$$

Substituting into (7), we have $\sum_{i\neq 1} P\left(E_i\right) \leq 2^{mR-m(C-\epsilon)}$. Since $C > R$ and ϵ is arbitrarily small, the above RHS $\to 0$ as $m \to \infty$.

We now state the converse theorem and sketch a proof for it. A detailed proof can be found in [2].

Theorem 2: Any sequence of $(2^{mR}, m, \epsilon_m)$ codes with average power S, causal side information k^m, and $\epsilon_m \to 0$ as $m \to \infty$ has rate $R < C$.

Consider a sequence of (2^{mR}) codes $\{x_W[i]\}_{i=1}^m$ with $W \in \{1,\ldots,2^{mR}\}$ and error probability $\epsilon_m \to 0$ as $m \to \infty$. Assume these codes satisfy the input constraints. Let W be uniformly distributed on $\{1,\ldots,2^{mR}\}$. For any side information sequence $\underline{K}^m$ independent of W, $mR \leq 1 + \epsilon_m mR + I(\underline{X}^m;\underline{Y}^m|\underline{K}^m)$, from Fano's inequality. Dividing both sides by m and letting $m \to \infty$ yields the desired result.

3. EXAMPLES OF CHANNELS.

3.1 ISI Channels.

The capacity of channels with time-invariant ISI has been thoroughly studied. The channel capacity is obtained by decomposing the ISI channel into a set of parallel independent memoryless channels for which the capacity is known. Gallager's [18] derivation uses a Karhunen-Loève (KL) decomposition and then applies the KMS theorem to obtain the maximum mutual information in the limit of infinite blocklength. An alternate derivation using a Fourier decomposition appears in [19]. Other approaches rely on limiting results of Toeplitz matrices [20].

One might hope to obtain the capacity of the time-varying ISI channel using similar decomposition techniques. Unfortunately, this approach will not work, even in the case where the channel realization is perfectly known a priori. In this case one can decompose the time-varying ISI channel over any *finite* time interval using the KL decomposition [18, Theorem 8.4.1]. Since the channels are time invariant, appropriate care must be taken to take into account the fact that more output degrees of freedom than input degrees of freedom are present [21], [22]. However, the decomposition cannot be used to obtain the channel capacity since the eigenvalues of the KL expansion may not converge as the time window increases to infinity. The only time when these eigenvalues converge to a limit is for the special case where the time-varying channel is just a time-invariant channel with a finite time window. In this case the limit of the eigenvalues is given by the KMS theorem [18], and application of the KMS theorem to this case yields the capacity of the time-invariant ISI channel in [18]. Note that the Toeplitz matrix approach is also not suited to time-varying channels, because limiting results may not hold.

Thus, we cannot use a decomposition to obtain the capacity of a time-varying ISI channel with a priori information about the channel variation. Moreover, we are interested in the case where we have causal side information about the channel, for which the KL decomposition is not applicable anyway. We now show that the capacity of the time-varying ISI channel with causal side information can be obtained from our capacity formula. This formula does not apply directly, since the ISI channel is not asymptotically block memoryless. However, our capacity formula can be used to obtain upper and lower bounds for the capacity of a time-varying ISI channel, which we then show to converge.

Consider a time-varying ISI channel with input $x(t)$, time-varying impulse response $h(t;\tau)$, and output $y(t)$ given by

$$y(t) = \int h(t;\tau)x(t-\tau)d\tau + n(t). \tag{9}$$

We assume that the sampling rate at the receiver is sufficiently fast [23] so that our channel can be represented by the discrete time model

$$y_k = \sum_{i=1}^{\mu} h_{k,i}x_{k-i} + n_k, \tag{10}$$

where $\mu < \infty$ is the memory of the ISI channel. We assume that the taps $h_{k,i}$ are stationary, ρth order Markov. The output has zero mean and power constraint $[\frac{1}{n}\sum_{i=1}^n y_n^2] \leq \sigma_Y^2$. We have perfect channel side information so that at time k, $K_k = (h_{k,1},\ldots,h_{k,\mu})$ is known at both the transmitter and receiver. The ρth order Markov assumption implies that K_k is independent of K_j if $|k-j| > \rho$. Note that this channel is not asymptotically block memoryless, i.e. it does not satisfy Definition 2.

We first lower bound the capacity of this channel. The lower bound is obtained by considering a similar ISI channel with the same outputs $\hat{y}_k = y_k$ as the original channel except that for any block length $m \geq \max(\mu,\rho)$, the output $\hat{y}_k = 0$ over the last μ outputs in the block. Specifically, the input/output relationship of our modified channel is

$$\hat{y}_k = \begin{cases} \sum_{i=1}^{\mu} h_{k,i}x_{k-i} + n_k & k \neq lm-\mu,\ldots,lm, \\ & \forall l \in \mathcal{Z} \\ 0 & \text{else} \end{cases} \tag{11}$$

Since we can discard the last μ outputs in any block of length m for our original channel, the capacity of the modified channel is clearly no larger than that of the original channel. This modified channel is easily shown to be block memoryless for a block length of m, so we can use our capacity theorem to obtain its capacity $C_{L,m} \leq C$. Specifically, this capacity is given by

$$C_{L,m} = \lim_{n\to\infty} max_{p(\underline{x}^n|\underline{K}^n)} \left[\frac{1}{n} I\left(\underline{X}^n;\hat{\underline{Y}}^n|\underline{K}^n\right)\right], \tag{12}$$

where the maximum is taken over distributions which satisfy our average energy, causal side information and bounded energy constraints. In [2] we show that $C_{L,m}$ converges to a finite value for every m, for input alphabets that are discrete or satisfy mild constraints. Since $C_{L,m} \leq C \; \forall m$ we get the lower bound

$$\sup_m C_{L,m} \leq C. \tag{13}$$

The upper bound is obtained by considering the capacity of the following set of two parallel independent channels. Fix any integer $m \geq \max(\rho,\mu)$. The first channel is the same as the channel used to obtain the lower bound as defined in (11). The second channel is the same as our original channel except that over each block of size m, the output $y_k = 0$ except over the last μ outputs in the block. Specifically,

$$\hat{\hat{y}}_k = \begin{cases} \sum_{i=1}^{\mu} h_{k,i}x_{k-i} + n_k & k = lm-\mu,\ldots,ln, \\ & \forall l \in \mathcal{Z} \\ 0 & \text{else} \end{cases} \quad (14)$$

Thus, the set of two parallel channels is described by (11) for the first channel and (14) for the second channel. We assume that each channel in the parallel set has the same average input power constraint as the original channel. The capacity of this parallel set is clearly at least as much as the capacity of our original channel since, if we use the same input x_k on both channels then $y_k = \hat{y}_k + \hat{\hat{y}}_k$, where y_k is the same output as on the original channel. Thus we can obtain the same data rate as on the original channel by restricting the input and output on our parallel set, so the parallel set has capacity greater than or equal to that of the original channel. Moreover, as long as $m \geq \mu$ it is easily shown that both channels in our parallel set are block memoryless and therefore we can apply our capacity theorem to both channels.

It is shown in [2] that the capacity of the parallel set is bounded above by the sum of capacities on each of the parallel channels. We will denote this sum capacity as $C_{U,m} \geq C$, which is given by

$$\begin{aligned} C_{U,m} &= C_{L,m} + \lim_{n\to\infty} \max_{p(\underline{x}^n|\underline{K}^n)} \frac{1}{n} I\left(X^n;\hat{\hat{Y}}^n|K^n\right) \\ &\leq C_L + \lim_{n\to\infty} \max_{p(\underline{x}^n|\underline{K}^n)} \frac{1}{n} I\left(X^n;\hat{\hat{Y}}^n|K^n\right). \quad (15) \end{aligned}$$

Thus, if we can show that the second term on the RHS of (15) goes to zero for any m, we have shown that the upper and lower bounds of C converge. By definition of $\hat{\hat{Y}}$ we have

$$\frac{1}{n} I(X^n;\hat{\hat{Y}}^n|K^n) = \frac{1}{n}\sum_{i=n-\mu}^{n} I(Y_i, X_{i-\mu}^i|Y^{i-1},K^n). \quad (16)$$

Define the quantity

$$y_{i,j} = h_{i,j}x_{i-j} + \frac{n_j}{\mu}. \quad (17)$$

Then $y_j = \sum_{i=1}^{\mu} y_{ij}$, so

$$\begin{aligned} \frac{1}{n} & I(Y_i;X_{i-\mu}^i|Y^{i-1},K^n) \\ \leq\ & I((Y_{1,j},\ldots,Y_{\mu,j});X_{i-\mu}^i|Y^{i-1},K^n) \\ \leq\ & \frac{1}{n}\sum_{j=1}^{\mu} I(Y_{i,j};X_{i-\mu}^i|Y^{i-1},Y_{1,j},\ldots,Y_{i-1,j},K^n) \\ =\ & \frac{1}{n}\sum_{j=1}^{\mu} H(Y_{i,j}|Y^{i-1},Y_{1,j},\ldots,Y_{i-1,j},K^n) \\ -\ & H(Y_{i,j}|Y^{i-1},Y_{1,j},\ldots,Y_{i-1,j},X_{i-\mu}^i,K^n) \\ =\ & \frac{1}{n}\sum_{j=1}^{\mu} H(Y_{i,j}|Y^{i-1},Y_{1,j},\ldots,Y_{i-1,j},K^n) \\ -\ & H(Y_{i,j}|Y^{i-1},Y_{1,j},\ldots,Y_{i-1,j},X_{i-j},K^n) \\ =\ & \frac{1}{n}\sum_{j=1}^{\mu} H(Y_{i,j}|Y^{i-1},Y_{1,j},\ldots,Y_{i-1,j},K^n) \\ -\ & H(Y_{i,j}|K_j,X_{i-j}) \\ \leq\ & \frac{1}{n}\sum_{j=1}^{\mu} H(Y_{i,j}|K_j) - H(n_j/\mu) \\ \leq\ & \frac{\mu}{2n}\log\left(1+\frac{\sigma_h^2 m\sigma_X^2}{\mu^{-2}\sigma_N^2}\right), \end{aligned} \quad (18)$$

where (18) follows from the fact that we can compress all of our average signal energy for the block of size m into these μ inputs. Combining (15), (18), and the output power constraint, we have

$$C \leq C_L + \lim_{n\to\infty}\frac{\mu}{2n}\log\left(1+\frac{m\sigma_Y^2}{\mu^{-2}\sigma_N^2}\right) \quad (19)$$

Using the fact that μ, σ_Y^2, and σ_N^2 are all finite we get that the second term in the RHS of (19) is zero for every m. Therefore we have the desired result $C \leq C_L$. Thus our upper and lower bounds converge to C_L, which equals the capacity of the original time-varying ISI channel.

We achieve the capacity $C = C_L$ of the time-varying ISI channel as follows. Fix any $\epsilon > 0$. We then find m such that $|C_L - C_{L,m}| < \epsilon/2$. Then, if we assume that the decoder discards the last μ outputs in every block of size m, by Theorem 1 there exists a coding strategy for this channel with rate $C_{L,m} - \epsilon/2$ and arbitrarily small error probability. We cannot achieve a higher rate on the time-varying ISI channel since its rate is bounded above by the rate of the parallel channel described above, and we have shown that we cannot achieve a rate higher than C_L on this parallel channel.

We now obtain a formula for C. By definition of our channel

$$\begin{aligned} I(\underline{X}^n,\underline{Y}^n|\underline{K}^n) &= h(\underline{Y}^n|\underline{K}^n) - h(\underline{Y}^n|\underline{K}^n,\underline{X}^n) \\ &= \sum_{i=1}^{n} h\left(Y_i|\underline{Y}^{i-1},\underline{K}^n\right) - \sum_{i=1}^{n} h(N_i) \end{aligned} \quad (20)$$

Thus, maximizing (20) is equivalent to maximizing $\sum_{i=1}^{n} h\left(Y_i|\underline{Y}^{i-1},\underline{K}^n\right)$, which in turn is equivalent to maximizing $\sum_{i=1}^{n} h\left(Y_i|\underline{Y}^{i-1},\underline{K}^i\right)$. Moreover, using the fact that conditioning reduces entropy and [24, pp. 232–234], we have that

$$\begin{aligned} & \textstyle\sum_{i=1}^{n} h\left(Y_i|\underline{Y}^{i-1},\underline{K}^i\right) \\ \leq\ & \textstyle\sum_{i=1}^{n} h\left(Y_i - \underline{\alpha}\underline{Y}^{i-1}|\underline{K}^i\right) \quad (21) \\ \leq\ & \textstyle\sum_{i=1}^{n} \frac{1}{2}ln\left(2\pi e \left| \Lambda_{Y_i-\underline{\alpha}\underline{Y}^{i-1}|\underline{K}^i} \right|\right) \quad (22) \end{aligned}$$

where $\underline{\alpha}$ is an arbitrary $n \times n$ vector dependent on $\underline{K}^i$ and Λ denotes an autocorrelation matrix, and $|\,|$ denotes absolute value of determinant. If $\underline{\alpha}$ is chosen to be the LLSE for Y_i, then (21) holds with equality. If $Y_i - \underline{\alpha}\underline{Y}^{i-1}$ conditioned on $\underline{K}^i$ is Gaussian , then (22) will hold with equality. Thus, $\sum_{i=1}^{n} h\left(Y_i|\underline{Y}^{i-1}, \underline{K}^i\right)$ will equal the RHS of (22) if we select, for all i, $\underline{X}^i$ to be a Gaussian random vector conditioned on $\underline{K}^i$. Thus, the distribution of the sender's signal is Gaussian *conditioned* on the channel up to the present time. The input distributions must comply with a total energy constraint

$$\frac{1}{n}\left[\sum_{j=1}^{i} x_j^2 + \sum_{j=i+1}^{n} \int p\left(x_j|\underline{x}^{i-1}, \underline{k}^i\right) x_j^2 dx d\left(\underline{k}\right)\right] \leq \mathbf{S} \tag{23}$$

which arises because we must allocate our energy judiciously so as to reserve energy for future transmissions.

3.2 Flat-Fading Channels with Imperfect Estimates

Consider a single-tap (flat-fading) channel with additive white Gaussian noise n_i, so

$$y_i = k_i x_i + n_i. \tag{24}$$

Using a discretization argument similar to that in the coding theorem proof of [3] we can show that our capacity results are applicable to this channel even if k_i and the corresponding side information are not restricted to finite sets. However, we still require that the channel be asymptotically block memoryless.

Suppose for this channel that at time i both the transmitter and receiver have a noisy or delayed estimate of k_i and assume that the K_is are i.i.d.. Such an assumption holds for frequency-hopped channels, such as the one considered in [25]. The channel is block memoryless for a block size of one and our theorem applies. Thus, the capacity is given by

$$\max_{p(x|K)} I(X;Y|K), \tag{25}$$

where the maximum is taken over all distributions that satisfy $E[X|K] \leq S$. The optimal conditional input distribution will depend on the relationship between k_i and Y_i, and will not in general be Gaussian. When k_i is known perfectly at both the transmitter and receiver the capacity has a simple form [3]:

$$C = \max_{S(\gamma):\int S(\gamma)p(\gamma)d\gamma=S} \int \frac{1}{2}\log\left(1 + \frac{S(\gamma)\gamma}{S}\right) p(\gamma)d\gamma, \tag{26}$$

where $p(\gamma) = p(K_i^2)$ and is independent of i by stationarity. In this case the optimal input distribution is Gaussian with power optimized using a water-filling in time.

4. CONCLUSIONS.

We have analyzed the capacity of a discrete time-varying channel with channel side information at the sender and receiver. The side information is updated at each time instant relative to the changing channel, which is asymptotic block memoryless. We proved a weak coding theorem and its converse. This coding theorem allows us to find the capacity of a large class of ISI channels with channel estimates. We also find the input distribution that maximizes capacity for these channels. Our theorem also allows us to consider the capacity of certain channels with imperfect channel estimates. The ability to treat ISI channels without the use of decomposition along a certain basis is crucial in allowing us to find capacity of channels with memory when the channel is time-varying.

REFERENCES

[1] M. Médard, A.J. Goldsmith, "Capacity of time-varying channels with channel side information", *ISIT 1997*.

[2] A.J. Goldsmith, M. Médard, "Capacity of Time-Varying Channels with Causal Channel Side Information", submitted to *IEEE Transactions on Information Theory*.

[3] A.J. Goldsmith, P. Varaiya, "Capacity of fading channels with channel side information," *IEEE Trans. Info. Theory*, Nov. 1997.

[4] J. Wolfowitz, *Coding Theorems of Information Theory*, 3^{rd} ed., Springer-Verlag, 1978.

[5] R.K. Mueller, G.J. Foschini, "The capacity of linear channels with additive Gaussian noise," *Bell Sys.Tech. J.*, Jan. 1970.

[6] L.H. Ozarow, S. Shamai, A.D. Wyner, "Information theoretic considerations for cellular mobile radio," *IEEE Trans. Vehic. Technol.*, Vol. VT-43, No. 2, May 1994.

[7] G. Kaplan, S. Shamai, "Error probabilities for the block-fading Gaussian channel," *AEU Arch. Elect.*, Vol. 49, No. 4, July 1995.

[8] R. Price, "Statistical theory applied to communication through multipath disturbances", Ph.D. Thesis, MIT, 1953.

[9] A.J. Goldsmith, P. Varaiya, "Capacity, mutual information, and coding for finite-state Markov channels," *IEEE Trans. Info. Theory*, May 1996.

[10] R.J. McEliece, W. E. Stark, "Channels with block interference," *IEEE Trans. Info. Theory*, Vol IT-30, No. 1, Jan. 1984.

[11] T.T. Kadota, A.D. Wyner, "Coding theorem for stationary, asymptotically memoryless, continuous-time channels", *The Annals of Math. Stats.* , vol. 43, no. 5, 1972.

[12] E. Pfaffelhuber, "Channels with asymptotically decreasing memory and anticipation", *IEEE Trans. Info. Theory*, vol. IT-17, no. 4, July 1971.

[13] M. Médard, "A coding theorem for multiple-access decorrelating channels", *ISIT 1998*.

[14] I. Csiszár, P. Narayan, The capacity of the Arbitrarily Varying Channel," *IEEE Trans. Info. Theory*, Vol. 37, No. 1, Jan. 1991.

[15] S. Verdú, T. S. Han. " A general formula for channel capacity", *IEEE Trans. Info.Theory*, IT-40(4), July 1994.

[16] M. Effros, A. Goldsmith, "The capacity of general channels with channel side information," *ISIT'98*.

[17] H. Viswanathan, "Capacity of time-varying channels with delayed feedback," preprint.

[18] R.G. Gallager, *Information Theory and Reliable Communications*, John Wiley & Sons, 1968.

[19] W. Hirt and J.L. Massey. Capacity of the discrete-time Gaussian channel with intersymbol interference. *IEEE Transactions on Information Theory*, IT-34(3), pp. 380–388, May. 1988.

[20] R.M. Gray, "On the Asymptotic Eigenvalue Distribution of Toeplitz Matrices", *IEEE Transactions on Infromation Theory*, vol. IT-18, no. 6, November 1972, pp. 725-729.

[21] M. Médard, R.G. Gallager, "The Issue of Spreading in Multipath Time-varying Channels", *IEEE Vehicular Technology Conference*, 1995.

[22] T. Kailath, *Sampling Models for Linear Time-variant Filters*, Technical Report 352, MIT RLE, Cambridge, MA, May 25, 1959.

[23] M. Médard, *The Capacity of Time Varying Multiple User Channels in Wireless Communications*, Sc. D. Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, September 1995.

[24] T.M. Cover, J.A. Thomas, *Elements of Information Theory*, John Wiley & Sons, 1991.

[25] I.C. Abou-Faycal, M.D. Trott, S. Shamai, "The Capacity of Discrete-Time Rayleigh Fading Channels", *International Symposium on Information Theory*, 1997. pg. 473.