The Streaming-DMT of Fading Channels

Ashish Khisti Member, IEEE, and Stark Draper Member, IEEE

Abstract

We study sequential transmission of a stream of messages over a block-fading multi-input-multi-output (MIMO) channel. A new message arrives in each coherence block and the decoder is required to output each message after a delay of $T$ coherence blocks. We establish the optimal diversity-multiplexing tradeoff (DMT) and show that it can be achieved using a simple interleaving of messages. The converse is based on an outage amplification technique which appears to be new. We also discuss another coding scheme based on a sequential tree-code. This coding scheme only requires the knowledge of delay at the decoder and yet realizes the optimal DMT. We finally discuss some extensions when multiple messages at uniform intervals arrive within each coherence period.

I. INTRODUCTION

Many multimedia applications require real-time encoding of the source stream and a sequential reconstruction of each source frame by its playback deadline. Both the fundamental limits and optimal communication techniques for such streaming systems can be very different from classical communication systems. In recent years there has been a growing interest in characterizing information theoretic limits for delay constrained communication over wireless channels. When the transmitter has channel state information (CSI), a notion of delay-limited capacity can be defined [2]. For slow fading channels, the delay-limited capacity is achieved using channel inversion at the transmitter [3]. In absence of transmitter CSI, an outage capacity can be defined [4], [5]. An alternative notion of expected capacity using a broadcast strategy has also been proposed [6] in such scenarios. Each fading state maps to a virtual receiver and a broadcast coding technique is used at the encoder. This approach has been further treated in e.g., [7]–[10]. For related work on joint source-channel coding over fading channels, see e.g., [11]–[21] and the references therein.

The present paper studies delay constrained streaming over multi-antenna wireless channels. We assume a block fading channel model and assume that the transmitter observes a sequence of independent messages, one in each coherence block. The encoded signal is a causal function of the messages. The decoder is required to output each message with a maximum delay of $T$ coherence blocks. As our main contribution we characterize the diversity-multiplexing tradeoff of this delay-constrained streaming model and refer to it as streaming-DMT.

Diversity-multiplexing tradeoff (DMT) of the quasi-static (slow-fading) channel model was first introduced in [22]. The authors propose diversity order and multiplexing gain as two fundamental metrics for communication over a wireless channel, and establish a tradeoff between these. A significant body of literature on DMT for quasi-static

Part of this work was presented at the International Symposium on Information Theory, St. Petersburg, Russia, 2011 [1].

A. Khisti’s work was supported by an NSERC (Natural Sciences Engineering Research Council) Discovery Grant and an Ontario Early Researcher Award. The corresponding author is A. Khisti (ashish.khisti@gmail.com)
delay-universal streaming has been studied in e.g., [39]–[45]. A tree-based code is proposed to encode a sequence of messages. The decoder is required to produce an estimate of each past message (at each time) under the constraint that the error probability decreases exponentially with the delay. Such a delay-universal (anytime) constraint is motivated by an application of stabilizing a control plant over a noisy channel. A maximum likelihood decoder is studied and the associated error exponent is characterized for a variety of discrete memoryless channels. In contrast the present work focuses on the streaming-DMT of block fading channels. We show that for a fixed decoding delay, the optimal DMT is achieved using a surprisingly simple interleaving technique, and does not require a tree code. We also study the performance achieved by a tree based encoder and a decision feedback based decoder and show that such a scheme can achieve the streaming-DMT in a delay universal manner. For related work on erasure channels we point the reader to e.g., [46]–[52] and references therein.

II. MODEL

Our setup is illustrated in Fig. 1. We consider an independent identically distributed (i.i.d.) block fading channel model with a coherence period of $M$:

$$Y_k = H_k \cdot X_k + Z_k,$$

where $k = 0, 1, \ldots, M$ denotes the index of the coherence block of the fading channel. The matrix $H_k \in \mathbb{C}^{N_r \times N_t}$ denotes the channel transfer matrix in coherence period $k$. We assume that the transmitter has $N_t$ transmit antennas and the receiver has $N_r$ receive antennas.

$$X_k = [X_k(1) | \ldots | X_k(M)] \in \mathbb{C}^{N_t \times M}$$

is a matrix whose $j$-th column, $X_k(j)$, denotes the vector transmitted in time-slot $j$ in the coherence block $k$ and similarly $Y_k \in \mathbb{C}^{N_r \times M}$ is a matrix whose $j$-th column, $Y_k(j)$ denotes the vectors received in time-slot $j$ in block $k$. The additive noise matrix is $Z_k \in \mathbb{C}^{N_r \times M}$. Thus (1) can also be expressed as,

$$Y_k(j) = H_k \cdot X_k(j) + Z_k(j), \quad j = 1, \ldots, M.$$
We assume that all entries of $H_k$ are sampled independently from the complex Gaussian distribution with zero-mean and unit-variance i.e., $CN(0,1)$. The channel remains constant during each coherence block and is sampled independently across blocks. All entries of the additive noise matrix $Z_k$ are also sampled i.i.d. $CN(0,1)$. Finally the realization of the channel matrices $H_k$ is revealed to the decoder, but not to the encoder.

We assume an average (short-term) power constraint $E[\sum_{i=1}^{M} ||X_k(i)||^2] \leq M\rho$. Note that $\rho$ denotes the transmit signal-to-noise-ratio (SNR). We will limit our analysis to the case where $M$ is sufficiently large so that random coding arguments can be invoked within each coherence block. A delay-constrained streaming code is defined as follows:

**Definition 1 (Streaming Code):** A rate $R$ streaming code with delay $T$, $C(R,T)$, consists of
1. A sequence of messages $\{w_k\}_{k \geq 0}$ each distributed uniformly over the set $\mathcal{I}_M = \{1,2,\ldots,2^{MR}\}$.
2. A sequence of encoding functions $F_k : \mathcal{I}_1^{k+1} \rightarrow \mathbb{C}^{N_t \times M}$, $X_k = F_k(w_0, \ldots, w_k)$, $k = 0,1,\ldots$ (3)
   that maps the input message sequence to a codeword $X_k \in \mathbb{C}^{N_t \times M}$.
3. A sequence of decoding functions $G_k : \mathbb{C}^{M(k+T)} \rightarrow \mathcal{I}_M$ that outputs message $\hat{w}_k$ based on the first $k + T$ observations, i.e., $\hat{w}_k = G_k(Y_0, \ldots, Y_{k+T-1})$, $k = 0,1,\ldots$ (4)

We now define the diversity-multiplexing tradeoff (DMT) [22] associated with the streaming code $C(R,T)$. Let the error probability for the $k$-th message be $p_k = \text{Pr}(w_k \neq \hat{w}_k)$ where $\hat{w}_k$ is the decoder output (4) and the error probability be averaged over the random channel gains. Let $\text{Pr}(e) = \sup_{k \geq 0} p_k$ denote the worst-case error probability. The DMT tradeoff [22] of $(r,d)$ is achievable with delay $T$ if there exists a sequence of codebooks $C(R = r \log \rho, T)$ such that
$$d = \lim_{\rho \rightarrow \infty} -\log \text{Pr}(e) \log \rho, \quad r = \lim_{\rho \rightarrow \infty} \frac{R(\rho)}{\log \rho}$$

Of interest, is the optimal diversity-multiplexing tradeoff, denoted by $d_T(r)$.

### III. Main Result

The optimal tradeoff between diversity and multiplexing (DMT) for the quasi-static fading channel was characterized in [22]. We reproduce the result below for the convenience of the reader.

**Theorem 1:** ([22]) For the quasi-static fading channel
$$Y(t) = H \cdot X(t) + Z(t)$$

\footnote{While we only focus on the Rayleigh channel model, our result easily extends to other channel models.}

\footnote{We caution the reader that this is not the maximum error probability with respect to a single realization of the fading state sequence. This later quantity is clearly 1 as in any sufficiently long realization, we will eventually find at-least one block that leads to an outage. In our definition we fix an index, $k$ and find the error probability $p_k$ averaged over the channel gains. We subsequently search for an index $k$ with maximum error probability. For time-invariant coding schemes, due to symmetry, $p_k$ will be independent of $k$.}
where the entries of \( H \in \mathbb{C}^{N_r \times N_t} \) are sampled i.i.d. \( \mathcal{CN}(0, 1) \), the optimal DMT tradeoff \( d_1(r) \) is a piecewise linear function connecting the points \((k, d_1(k))\) for \( k = 0, 1, \ldots, \min(N_r, N_t) \), where \( d_1(k) = (N_t - k)(N_t - k) \).

In our analysis the following simple generalization of the quasi-static DMT to \( L \) parallel channels is useful.

**Corollary 1:** Consider a collection of \( L \) parallel quasi-static fading channel
\[
Y_l(t) = H_l \cdot X_l(t) + Z_l(t), \quad l = 1, \ldots, L
\]
where the entries of \( H_l \in \mathbb{C}^{N_r \times N_t} \) are all sampled i.i.d. \( \mathcal{CN}(0, 1) \). The DMT tradeoff is given by \( d_L^1(r) = L \cdot d_1 \left( \frac{r}{L} \right) \) for any \( r \in (0, L \min(N_r, N_t)) \).

We provide a proof of Corollary 1 in Appendix A.

Our main result establishes the optimal DMT for a block fading channel model with a delay constraint of \( T \) coherence blocks.

**Theorem 2:** The optimal DMT tradeoff for a streaming code in Definition 1 with a delay of \( T \) coherence blocks is given by \( d_T(r) = T \cdot d_1(r) \), where \( d_1(r) \) is the optimal DMT of the underlying quasi-static fading channel.

The result in Theorem 2 illustrates that the DMT of a streaming source under a delay constraint of \( T \) coherence blocks is identical to the DMT of a system with \( T \) independent and parallel MIMO channels if the rate of the latter system is suitably normalized. Indeed our achievability scheme exploits this connection. We show that the DMT can be achieved by interleaving messages in a suitable manner to reduce the system to a parallel channel setup. We also present another scheme based on a tree code that achieves the DMT. The converse however does not appear to follow from earlier results. We present, what appears to be a new idea called “outage-amplification” which is specific to our streaming setup.

In the remainder of this paper we present the converse in Section IV and two coding schemes for achieving the optimal DMT in Sections V and VI respectively. We discuss an extension to the case when two messages arrive within each coherence block in Section VII.

### IV. Converse

**A. Discussion**

We first present simple bounds that suffice to bound the maximum diversity and maximum multiplexing gain, but are not tight in between. We then provide a heuristic explanation of the new dimension of our proof in tying together the bounding results for individual messages to take into account their overlapping transmission times. A formal proof of the converse is presented in section IV-B.

For our discussion we consider a single-antenna channel model when \( T = 2 \). Consider the decoding of \( w_0 \) from blocks 0 and 1. One upper bound is obtained by revealing \( w_1 \) to the decoder resulting in \( d_1^1(r) = 2 - r \). Another upper bound is obtained by revealing \( w_1 \) to the encoder at time \( t = 0 \), revealing \( w_2 \) to the decoder and relaxing the
Fig. 2. Simple bounds on the single-antenna model for $T = 2$. The actual DMT is $d_2(r) = 2(1 - r)$. Among the sequence of upper bounds $d^+_k(r)$, note that $d^+_0(r)$ is tight at the maximum diversity point while $d^+_\infty(r)$ is tight at the maximum multiplexing point. However, the intersection of these bounds is not tight enough to yield $d_2(r)$.

We illustrate the main idea of our proposed technique in Fig. 3 and Fig. 4 and provide some intuition below. Assume that a DMT larger than that claimed in Theorem 2 is achievable; assume that $d(r) = 2(1 - r + 2\delta)$ for some $\delta > 0$. This implies that $\Pr(\hat{w}_k \neq w_k) \leq \rho^{-2(1-r+2\delta)}$ holds for each $k \geq 0$. Now suppose that gains $h_{k+1}$ belongs to the set $\mathcal{H}_k$ defined as:

$$\mathcal{H}_k = \left\{ (h_0, \ldots, h_{k+1}) : |h_0|^2 \geq 1, \ldots, |h_{k-1}|^2 \geq 1, |h_k|^2 \leq \rho^{-(1-r+\delta)}, |h_{k+1}|^2 \leq \rho^{-(1-r+\delta)} \right\}$$ (7)
Fig. 3. Streaming setup when \( T = 2 \). The source data stream is observed sequentially and mapped into the channel stream. If we assume that \( d(r) > 2(1 - r) \) holds then message \( w_k \) can be recovered even if the channel gains in blocks \( k \) and \( k + 1 \) (shaded) are in outage. This is indicated in the upper figure with blocks \( k \) and \( k + 1 \) shaded. Our outage amplification argument amplifies this effect and shows that we can decode every data packet by its deadline even if all the channel gains are in outage, as illustrated in the lower figure.

Fig. 4. Illustration of the outage amplification argument. Assume that the channel gains \( h_0, \ldots, h_4 \) are all in outage. For decoding \( w_0 \) only the values of \( h_0 \) and \( h_1 \) matter. The message \( w_0 \) thus can be decoded from blocks 0 and 1. At this point we can reconstruct \( X_0 \) as it only depends on \( w_0 \) and treat block 0 as if it were not in outage. Then using blocks 1 and 2 we can decode \( w_1 \), recover \( X_1 \). Repeating this procedure we can proceed to decode all the messages sequentially.

From standard analysis \[22\] we have

\[
\Pr(h_{k+1}^k \in \mathcal{H}_k) \leq \rho^{2(1-r+\delta)} > \Pr(w_k \neq w_k).
\]

Thus the receiver cannot declare an outage event for the set \( \mathcal{H}_k \). It turns out that given this fact, one can exploit the streaming nature of the problem to show successful decoding over a much weaker set of channels. For each \( N \geq 1 \) let:

\[
\mathcal{H}_N^\otimes = \left\{ h_0, \ldots, h_N : |h_k|^2 \leq \rho^{-(1-r+\delta)}, \ 0 \leq k \leq N \right\}
\]

(10)
denotes the set that all the \( N + 1 \) links are simultaneously in outage. We reason that the above decoder must in fact recover all the messages \( w_0, w_1, \ldots, w_{N-1} \) by their respective deadlines.

\( ^3 \)Throughout we use the notation \( \doteq \) to denote equality in the exponential sense. The function \( f_1(\rho) \doteq f_2(\rho) \) if \( \lim_{\rho \to \infty} \frac{\log f_1(\rho)}{\log f_2(\rho)} = 1 \). We define \( \preceq \) and \( \succeq \) in a similar manner.
Consider any \( h_0^N \in \mathcal{H}_{N}^\ominus \). Since \( h_0^0 \in \mathcal{H}_0 \) the decoder can recover \( w_0 \) at the end of block 1. Upon recovering \( w_0 \), the decoder also recovers the transmitted codeword \( X_0(w_0) \) and can replace the original channel output \( Y_0 \) with \( Y_0 = 1 \cdot X_0 + Z_0 \). The effective sequence of channel gains \( \hat{h}_0^{N} \) with \( \hat{h}_0 = 1 \) and \( \hat{h}_1^2 = h_1^2 \) now satisfies \( \hat{h}_0^{N} \in \mathcal{H}_1 \). Hence the decoder can decode message \( w_1 \) at the end of block 2. Continuing this argument, for the decoding of message \( w_k \), we assume that \( w_0, \ldots, w_{k-1} \) have already been decoded and therefore the associated sequence \( \hat{h}_0^{k+1} \) with \( \hat{h}_j = 1 \) for \( j = 1, 2, \ldots, k-1 \) is contained in \( \mathcal{H}_k \). Therefore \( w_k \) is also decoded. Thus we see that the entire sequence of messages \( w_0^{N-1} \) can be decoded.

However for sufficiently large \( N \), this leads to a contradiction. Notice that when \( h_0^N \in \mathcal{H}_{N}^\ominus \),

\[
\frac{1}{M} I(X_0^N : Y_0^N) = \sum_{k=0}^{N} \log(1 + |h_k|^2 \rho) \\
= (N + 1)(r - \delta) \log \rho
\]

whereas the total message rate is \( Nr \log \rho \). Thus if \( N > \frac{r}{\delta} - 1 \), we have that the information rate decoded over the channel exceeds the instantaneous capacity. This contradicts Fano’s inequality. Hence our assumption that \( d(r) \geq 2(1 - r + 2\delta) \) cannot be true. Our formal proof that applies for any \( N_r, N_t \) and \( T \) is presented next.

**B. Proof**

We establish that a lower bound on the error probability for any \( \mathcal{C}(R = r \log \rho, T) \) code in Definition [1] is

\[
\Pr(e) \geq \rho^{-T d_1(r)}
\]

where \( d_1(r) \) is the DMT tradeoff associated with a single-link MIMO channel. Define \( E_k = \{ w_k \neq w_k \} \), the error event associated with message \( w_k \), and note that \( \Pr(e) = \sup_{k \geq 0} \Pr(E_k) \).

We begin by lower bounding \( E_k \) associated with message \( w_k \). Recall that this message needs to be decoded after \( T \) coherence blocks indexed as \( t \in \{ k, \ldots, k + T - 1 \} \). Let \( H_k^{k+T-1} = H_k^{k+T-1} \) be the realization of the channel matrices in this interval. Applying Fano’s Inequality [53] for message \( w_k \) and using the fact that \( w_0^{k-1} \) is independent of \( w_k \) and \( H_k^{k+T-1} \) we have

\[
\Pr(E_k; H_k^{k+T-1} = H_k^{k+T-1}) \geq 1 - \frac{1}{M r \log \rho} - \frac{I(Y_k^{k+T-1}; H_k^{k+T-1} | w_k^{k-1}, H_k^{k+T-1} = H_k^{k+T-1})}{M r \log \rho}
\]

Since the second term vanishes as the coherence period \( M \to \infty \), we ignore it in our analysis. To bound the remaining terms we let

\[
\mathcal{H}_\delta = \left\{ H : \log \det \left( I + \frac{\rho}{M} H H^\dagger \right) \leq (r - \delta) \log \rho \right\}
\]

and use \( \mathcal{H}_\delta^T \) to denote the \( T \)-fold Cartesian product of the set \( \mathcal{H}_\delta \). Furthermore since the channel gains are sampled i.i.d.

\[
\Pr(H_k^{k+T-1} \in \mathcal{H}_\delta^T) = (P_\delta)^T
\]

where \( P_\delta = \Pr(H \in \mathcal{H}_\delta) \). From the single link DMT in Theorem [1]

\[
P_\delta = \Pr(H \in \mathcal{H}_\delta) \approx \rho^{-d_1(r-\delta)}.
\]
where $d_1(\cdot)$ is the associated DMT function in Theorem 1. The average error probability \( \Pr(\mathcal{E}_k) = E_{H_k} [\Pr(\mathcal{E}_k; H_k^{k+T-1})] \), can be lower bounded as follows.

\[
\Pr(\mathcal{E}_k) \geq \Pr(H_k^{k+T-1} \in \mathcal{H}_\delta^T) \cdot \left( 1 - \frac{I(w_k; Y_k^{k+T-1}|w_0^{k-1}, H_k^{k+T-1} \in \mathcal{H}_\delta^T)}{Mr \log \rho} \right) \tag{16}
\]

\[
= (P_3)^T \left( 1 - \frac{I(w_k; Y_k^{k+T-1}|w_0^{k-1}, H_k^{k+T-1} \in \mathcal{H}_\delta^T)}{Mr \log \rho} \right)
\]

\[
= (P_3)^T \left( 1 - \frac{I(w_k; Y_0^{N-1}|w_0^{k-1}, H_0^{N-1} \in \mathcal{H}_\delta^N)}{Mr \log \rho} \right) \tag{17}
\]

\[
\geq (P_3)^T \left( 1 - \frac{I(w_k; Y_0^{N-1}|w_0^{k-1}, H_0^{N-1} \in \mathcal{H}_\delta^N)}{Mr \log \rho} \right) \tag{18}
\]

where \(17\) follows from the fact that the channel gains \((H_0^{k-1}, H_0^{N-1})\) are independent of \((w_0^{k}, Y_0^{k+T-1}, H_0^{k+T-1})\).

Now we combine the error events.

\[
\max_{0 \leq k \leq N-T-1} \Pr(\mathcal{E}_k) \geq \sum_{k=0}^{N-T-1} \Pr(\mathcal{E}_k) \geq (P_3)^T \left( 1 - \frac{\sum_{k=0}^{N-T-1} I(w_k; Y_0^{N-1}|w_0^{k-1}, H_0^{N-1} \in \mathcal{H}_\delta^N)}{(N-T)Mr \log \rho} \right)
\]

\[
= (P_3)^T \left( 1 - \frac{I(w_0^{N-T-1}; Y_0^{N-1}|H_0^{N-1} \in \mathcal{H}_\delta^N)}{(N-T)Mr \log \rho} \right)
\]

\[
\geq (P_3)^T \left( 1 - \frac{I(X_0^{N-1}; Y_0^{N-1}|H_0^{N-1} \in \mathcal{H}_\delta^N)}{(N-T)Mr \log \rho} \right) \tag{19}
\]

where \(19\) follows from the data processing inequality since \(w_0^{N-T-1} \rightarrow X_0^{N-1} \rightarrow Y_0^{N-1}\) holds. Finally since the fading across different blocks is independent,

\[
\max_{0 \leq k \leq N-T-1} \Pr(\mathcal{E}_k) \geq (P_3)^T \left( 1 - \frac{\sum_{i=0}^{N-1} I(X_i; Y_i|H_i \in \mathcal{H}_\delta)}{(N-T)Mr \log \rho} \right) \tag{20}
\]

\[
\geq (P_3)^T \left( 1 - \frac{NM(r-\delta) \log \rho}{(N-T)Mr \log \rho} \right) \tag{21}
\]

\[
\geq \left( 1 - \frac{N(r-\delta)}{(N-T)r} \right) \rho^{-Td_1(r-\delta)} \tag{22}
\]

where \(21\) follows by substituting in \(13\).

For any \(\delta > 0\), by selecting \(N > T_D r\) the term inside the brackets is strictly positive. Since \(\delta > 0\) is arbitrary, it follows that the diversity order greater than \(T d_1(r)\) cannot be achieved.

V. CODING THEOREM: INTERLEAVING SCHEME

We now present an interleaving based scheme that achieves the DMT stated in Theorem 2. Our codebook $C$ maps each message $w_k \in \mathcal{I}_M$ to $T$ codewords $X_0(w_k), X_1(w_k), \ldots, X_{T-1}(w_k)$ of length $\frac{M}{T}$, i.e., $C = C_0 \times C_1 \ldots \times C_{T-1}$ and $X_k \in \mathcal{C}_k$. 
Coherence Block: K

\[ X_0(w_k) \quad X_1(w_{k-1}) \quad X_2(w_{k-2}) \quad X_3(w_{k-3}) \quad X_4(w_{k-4}) \quad X_5(w_{k-5}) \]

Coherence Block: K+1

\[ X_0(w_{k+1}) \quad X_1(w_k) \quad X_2(w_{k-1}) \quad X_3(w_{k-2}) \quad X_4(w_{k-3}) \quad X_5(w_{k-4}) \]

Coherence Block: K+2

\[ X_0(w_{k+2}) \quad X_1(w_{k+1}) \quad X_2(w_k) \quad X_3(w_{k-1}) \quad X_4(w_{k-2}) \quad X_5(w_{k-3}) \]

Coherence Block: K+3

\[ X_0(w_{k+3}) \quad X_1(w_{k+2}) \quad X_2(w_{k+1}) \quad X_3(w_k) \quad X_4(w_{k-1}) \quad X_5(w_{k-2}) \]

Coherence Block: K+4

\[ X_0(w_{k+4}) \quad X_1(w_{k+3}) \quad X_2(w_{k+2}) \quad X_3(w_{k+1}) \quad X_4(w_k) \quad X_5(w_{k-1}) \]

Fig. 5. Interleaving based coding scheme for \( T = 6 \). Each coherence block is divided into \( T \) sub-intervals and each sub-interval is dedicated to transmission of one message. The transmission of message \( w_k \) spans coherence blocks \( k, k+1, \ldots, k+T-1 \) using codewords of \( X_0(w_k), \ldots, X_{T-1}(w_k) \) as shown by the shaded blocks.

For transmission of these messages, we assume that each coherence block of length \( M \) is further divided into \( T \) sub-blocks of length \( \frac{M}{T} \), as indicated in Fig. 5. Let \( I_{k,0}, \ldots, I_{k,T-1} \) denote these intervals. The codeword \( X_0(w_k) \) is transmitted in the first sub-block \( I_{k,0} \) of coherence block \( k \). The codeword \( X_1(w_k) \) is transmitted in the sub-block \( I_{k+1,1} \) of coherence block \( k+1 \) and likewise \( X_j(w_k) \) is transmitted in the \( j \)-th sub-block of coherence block \( \{k+j\} \).

The corresponding output sequences associated with message \( w_k \) are denoted as

\[ Y_{k,j} = H_{k,j}X_j(w_k) + Z_{k,j}, \quad j = 0, \ldots, T-1. \] (23)

The decoder finds a message \( w_k \) such that for each \( j \in \{0, \ldots, T-1\} \), \( (X_j(w_k), Y_{k,j}) \) are jointly typical. The outage event at the decoder is given by:

\[ \left\{ \frac{1}{T} \sum_{j=k}^{k+T-1} C_j(\rho) \leq r \log \rho \right\} \] (24)

where \( C_j(\rho) = \log \det \left( I + \frac{\rho}{N_t} H_j H_j^\dagger \right) \). Since (24) precisely corresponds to the outage event of a quasi-static parallel MIMO fading channel, with \( T \) channels and a multiplexing gain of \( T \cdot r \), the DMT follows from Corollary I.

VI. CODING SCHEME: SEQUENTIAL TREE CODES

We propose a second construction which is inspired by the sequential tree codes proposed in [39]–[45]. This approach has the advantage that the encoder does not need to be revealed the delay. The delay constraint only needs to be revealed to the decoder and yet the optimal DMT is attained.

Our proposed streaming code, \( C(R, T) \), consists of a semi-infinite sequence of codebooks \( \{C_0, C_1, \ldots, C_k, \ldots\} \),
Fig. 6. The left hand figure illustrates construction of our proposed codebook. The message $w_0$ is mapped to one of $2^{nR}$ codewords in the first level, the message pair $(w_0, w_1)$ is mapped to one of $2^{2nR}$ codewords in the second level of the tree etc., While decoding $w_k$ the decoder starts at the root of the tree. It first finds all possible transmit paths of depth $k + T$ in the tree typical with the received sequence. If a unique prefix codeword in level 1 is determined then the corresponding message $w_0$ is decoded. At this point the decoder moves along the path of $w_0$ and finds all possible codewords from level 2 to $k + T$ that are typical with the received codeword. A unique message $w_1$ is determined if there is a unique prefix codeword in level 2. This process continues till level $k$ is reached and $w_k$ is determined.

where $C_k$ is the codebook to be used in coherence block $k$ when messages $(w_0, \ldots, w_k)$ are revealed. Codebook $C_k$ consists of a total of $2^{MR(k+1)}$ codewords and each codeword is assigned to one element in the set

$$\mathcal{I}_M^{k+1} = \{ (w_0, \ldots, w_k) : w_0 \in \mathcal{I}_M, \ldots, w_k \in \mathcal{I}_M \}.$$  \hspace{1cm} (25)$$

where $\mathcal{I}_M \triangleq \{1, 2, \ldots, 2^{MR}\}$. All codewords are length $M$ sequences whose symbols are sampled i.i.d. from $\mathcal{CN}(0, \frac{1}{Nt})$. In coherence block $k$, the encoder observes $w_0, \ldots, w_k$, maps it to the codeword $X_k(w_0^k) \in \mathbb{C}^{N_k \times M}$ in $C_k$, and transmits each of the $M$ columns of $X_k$ over $M$ channel uses. The entire transmitted sequence up to and including block $k$ is denoted by

$$X_0^k(w_0^k) \triangleq \{X_0(w_0), X_1(w_0^1), \ldots, X_k(w_0^k)\}, \quad X_0^k(w_0^k) \in \mathbb{C}^{N_k \times (k+1)M}$$  \hspace{1cm} (26)$$

For decoding message $w_k$, our proposed decoder does not rely on previously decoded messages, but instead

4We will make the practically relevant assumption that the communication terminates after a sufficiently large but fixed number of coherence blocks.
computes a new estimate of the all the messages $\hat{w}_0^k$ at time

$$T_k = k + T - 1$$

(27)

using the entire received sequence $Y_0^{T_k'} = (Y_0, \ldots, Y_{T_k'})$. First it searches for a message $\hat{w}_0$ by searching over all message sequences $\hat{w}_0^{T_k}$ such that $\{X_0^{T_k'}(\hat{w}_0^{T_k}), Y_0^{T_k'}\}$ are jointly typical. If each such sequence has a unique prefix $\hat{w}_0$ then $\hat{w}_0$ is selected as the message in block 0. Otherwise an error is declared. The decoder then proceeds sequentially, producing estimates of $w_1, \ldots, w_k$. In determining $w_l, l \leq k$, the decoder uses the already-determined vector of estimates $\hat{w}_0^{l-1}$. The decoder searches for a sequence of messages $\hat{w}_l^{T_k}$ such that the corresponding transmit sequence $X_l^{T_k}((\hat{w}_0^{l-1}, \hat{w}_l^{T_k}))$ has the property that the sub-sequence between $l$ to $T_k$ (the suffix) satisfies

$$(X_l^{T_k}(\hat{w}_0^{l-1}, \hat{w}_l^{T_k}), Y_l^{T_k}) \in T_{l, T_k},$$

(28)

where the set $T_{l, T_k}$ is the set of all jointly typical sequences $[53]$

$$T_{l, T_k} = \left\{ (X_l', Y_l') : X_l' \in T(p_{X_k, Y_k}), Y_l' \in T(p_{Y_k'}), \left| -\sum_{k=1}^l \frac{\log p_{X_k, Y_k}(X_k, Y_k) - h(p_{X_k, Y_k})}{M(l' - l + 1)} \right| \leq \varepsilon \right\},$$

(29)

where $T(p_{X_k, Y_k})$ and $T(p_{Y_k'})$ denotes the set of typical $\{X_l'\}$ and $\{Y_l'\}$ sequences respectively and where $h(p_{X_k, Y_k})$ denotes the differential entropy of jointly Gaussian random variables.

Remark 1: Our decoder is a decision directed decoder. In estimating $\hat{w}_0^k$, it first estimates $\hat{w}_0$ based on $Y_0^{T_k'}$. It next makes a conditional estimate of $\hat{w}_l$ based on $Y_l^{T_k}$ with $\hat{w}_0^l$ fixed, and continues along in $k + 1$ steps. One may be tempted to try a simpler decoding scheme that avoids the $k + 1$ steps and directly search for a unique prefix $\hat{w}_0^k$ such that the resulting transmit sequence $X_0^{T_k}$ is jointly typical with the received sequence $Y_0^{T_k}$ i.e.,

$$\left\{ -\sum_{k=0}^{T_k} \frac{\log p_{X_k, Y_k}(X_k, Y_k) - h(p_{X_k, Y_k})}{M(k + T)} \leq \varepsilon \right\}$$

(30)

Such an approach will not guarantee the recovery of true $\hat{w}_k$. This is because for $k \gg 1$ the contribution of the terms before $\hat{w}_k$ will dominate. Even when $\hat{w}_k \neq w_k$ but $\hat{w}_0^k = w_0^k$, the pair $(\hat{X}_k, Y_k)$ will in general satisfy (30) as the contribution of the suffix associated with $\hat{w}_k$ will be negligible.

Our proposed decision directed decoder guarantees that when decoding $w_k$ we do not include the bias introduced by $w_0^k$ in (29).

A. Analysis of error probability

We show that for any $\delta > 0$ and $0 < r < \min(N_1, N_1)$, the error probability averaged over the ensemble of codebooks $C(R = (r - \delta) \log_2 \rho, T)$, satisfies $\Pr(\mathcal{E}) \leq \rho^{-(T+d(r))}$. By symmetry we will assume, without loss of generality, that a particular message sequence $w_0^k = w_0^k$ is transmitted.
For the analysis of error probability we define the events
\[ E_l = \left\{ w_l : (\tilde{w}_0, \ldots, \tilde{w}_{l-1}) = (w_0, \ldots, w_{l-1}), \tilde{w}_l \neq w_l \right\}, \quad 0 \leq l \leq k \] (31)
and note that
\[ \Pr \{ \tilde{w}_k \neq w_k \} \leq \sum_{l=0}^{k} \Pr(\mathcal{E}_l), \] (32)
where \( \mathcal{E}_l \) corresponds to the event that our proposed decoder fails in step \( l \) of the decoding process. We develop an upper bound on \( \mathcal{E}_l \) for each \( 0 \leq l \leq k \) and substitute these bounds in (32).

We further express \( \mathcal{E}_l = \mathcal{A}_l \cup \mathcal{B}_l \), where
\[ \mathcal{A}_l = \left\{ (X_i^{T_k} (w_0^{T_k}), Y_i^{T_k}) \notin \mathcal{T}_l \right\} \] (33)
denotes the event that a decoding failure happens because the transmitted sub-sequence starting from position \( l \) fails to be typical with the received sequence whereas
\[ \mathcal{B}_l = \left\{ \exists w_0^{T_k} : \tilde{w}_0^{l-1} = w_0^{l-1}, \tilde{w}_l \neq w_l, (X_i^{T_k} (\tilde{w}_0^{T_k}), Y_i^{T_k}) \in \mathcal{T}_l \right\} \] (34)
denotes the event that the decoding failure happens because a transmit sequence corresponding to a message sequence with \( \tilde{w}_l \neq w_l \) appears typical with the received sequence.

As shown in the Appendix B using an appropriate Chernoff bound we can express,
\[ \Pr(\mathcal{A}_l) \leq 2^{-M(T_k-l+1)f(\epsilon)} \] (35)
where \( f(\epsilon) \) is a function that satisfies \( f(\epsilon) > 0 \) for each \( \epsilon > 0 \).

To bound \( \Pr(\mathcal{B}_l) \) we begin by noting that by our code construction, we are guaranteed that whenever \( \tilde{w}_l \neq w_l \), the associated transmit subsequence \( X_i^{T_k} (\tilde{w}_0^{T_k}) \) is sampled independently from \( Y_i^{T_k} \). Hence from the joint typicality analysis [53], we have that for any sequence \( \tilde{w}_0^{T_k} \) with \( \tilde{w}_l \neq w_l \)
\[ \Pr \left( (X_i^{T_k} (\tilde{w}_0^{T_k}), Y_i^{T_k}) \in \mathcal{T}_l \right| H_i^{T_k}) \leq 2^{-M \left( \sum_{j=1}^{T_k} I(H_j | Y_j (1)) - 3\epsilon \right)} \]
\[ = 2^{-M \left( \sum_{j=1}^{T_k} C_j(\rho) - 3\epsilon \right)} \]
where
\[ C_j(\rho) = \log \det \left( I + \frac{\rho}{N_k} H_j H_j^\dagger \right) \] (36)
is the associated mutual information between the input and output in the \( j \)-th coherence block when the channel matrix equals \( H_j = H_j \). Applying the union bound we have that
\[ \Pr(\mathcal{B}_l \mid H_i^{T_k}) \leq 2^{-M \left( \sum_{j=1}^{T_k} C_j(\rho) - (T_k-l+l+1)R - 3\epsilon \right)}. \] (37)

To bound \( \Pr(\mathcal{B}_l) \) we define
\[ \mathcal{O}_l = \left\{ (H_l, \ldots, H_{T_k}) : \sum_{i=l}^{T_k} C_i(\rho) \leq (k + T - l)R \log \rho + (k - l)\Delta(r) \log \rho + 4\epsilon \log \rho \right\} \] (38)
Thus it remains to bound Corollary 1 with order is given by:

\[
\Delta(r) = \frac{(N_t - r)(N_r - r)}{2(N_t + N_r - 2r)}, \quad 0 \leq r < \min(N_t, N_r).
\] (39)

Note that

\[
\Pr(\mathcal{B}_l) \leq \Pr(\mathcal{B}_l | \mathbf{H}^{T_k}_l \in \mathcal{O}_l^c) + \Pr(\mathbf{H}^{T_k}_l \in \mathcal{O}_l)
\] (40)

From (38) and (37) we have

\[
\Pr(\mathcal{B}_l | \mathbf{H}^{T_k}_l \in \mathcal{O}_l^c) \leq 2^{-M(k-l)\Delta(r)} \log \rho - M \varepsilon \log \rho
\] (41)

\[
\Pr(\mathbf{H}^{T_k}_l \in \mathcal{O}_l) \leq \rho^{-M\varepsilon-M(k-l)\Delta(r)}.
\] (42)

Thus it remains to bound \(\Pr(\mathcal{O}_l)\) in (40). Note that \(\mathcal{O}_l\) is precisely corresponds to the parallel MIMO channel in Corollary 1 with \(L = T_k - l + 1\) and the multiplexing gain of \(s = Lr + (k-l)\Delta(r) + 4\varepsilon\). The associated diversity order is given by:

\[
L \cdot d_1 \left(\frac{8}{L}\right) = L \left(N_t - \frac{8}{L}\right) \left(N_r - \frac{8}{L}\right)
\] (43)

\[
> L(N_t - r)(N_r - r) - (k-l)(N_t + N_r - 2r)\Delta(r) - o_\varepsilon(1)
\] (44)

\[
= T(N_t - r)(N_r - r) + \frac{(k-l)}{2}(N_t - r)(N_r - r) - o_\varepsilon(1)
\] (45)

\[
= T d_1(r) + \frac{k-l}{2} d_1(r) - o_\varepsilon(1)
\] (46)

where we substituted (39) for \(\Delta(r)\) in (45) and let \(o_\varepsilon(1)\) be a function of \(\varepsilon\) that vanishes as \(\varepsilon \to 0\).

Thus we have

\[
\Pr(\mathcal{O}_l) \leq \rho^{-(T d_1(r) + \frac{(k-l)}{2} d_1(r)) + o_\varepsilon(1)}.
\] (47)

From (40) and substituting (43) and (47) and using \(\mathcal{E}_l = \mathcal{A}_l \cup \mathcal{B}_l\) we have

\[
\Pr(\mathcal{E}_l) \leq \Pr(\mathcal{A}_l) + \Pr(\mathcal{B}_l)
\] (48)

\[
\leq 2^{-M(T_k - l + 1)f(\varepsilon)} + \rho^{-M\varepsilon-M(k-l)\Delta(r)} + \rho^{-T d_1(r) - \frac{(k-l)}{2} d_1(r) + o_\varepsilon(1)}.
\] (49)

From the union bound,

\[
\Pr(\mathcal{E}) \leq \sum_{l=0}^{k} \Pr(\mathcal{E}_l)
\] (50)

\[
\leq \sum_{l=0}^{k} 2^{-M(T_k - l + 1)f(\varepsilon)} + \sum_{l=0}^{k} \rho^{-M\varepsilon-M(k-l)\Delta(r)} + \sum_{l=0}^{k} \rho^{-T d_1(r) - \frac{(k-l)}{2} d_1(r) + o_\varepsilon(1)}
\] (51)

Substituting for \(T_k = k + T - 1\), we can express the first term in (51) as

\[
\sum_{l=0}^{k} 2^{-M(k+T-l)f(\varepsilon)} = \sum_{l=0}^{k} 2^{-M(l+T)f(\varepsilon)}
\] (52)

\[
\leq \sum_{l=0}^{\infty} 2^{-M(l+T)f(\varepsilon)} = 2^{-MTf(\varepsilon) + 1},
\] (53)
The first message, \( w_{k,1} \), arrives \( \Delta M \) symbols from the start of the coherence block while the second message \( w_{k,2} \) arrives \( M/2 \) symbols later. We assume a decoding delay of \( M \) symbols for each message. The left figure shows the case when \( \Delta = 0 \), while the right figure shows the case when \( \Delta = \frac{M}{2} \).

which vanishes as \( M \to \infty \). By a similar argument we can simplify the second and third terms in (51) to get

\[
\Pr(\mathcal{E}) \leq o_M(1) + \rho^{-Td_1(r) + o(1)}. \tag{54}
\]

where \( o_M(1) \to 0 \) as \( M \to \infty \). Since \( \epsilon > 0 \) is arbitrary we have established that the DMT of \( Td_1(r) \) is achievable.

**VII. Multiple Messages per Coherence Block**

Our primary focus in this paper has been the case when there is one message per coherence block i.e., we assume that the message \( w_k \) arrives at the beginning of coherence block \( k \), and needs to be reconstructed after a delay of \( T \cdot M \) symbols. In this section we will consider the case when two messages, say \( w_{k,1} \) and \( w_{k,2} \) arrive in each coherence block. We assume that \( w_{k,1} \) arrives at time \( t_{k,1} = kM + \Delta M \) and \( w_{k,2} \) arrives at time \( t_{k,2} = kM + (\Delta + \frac{1}{2}) M \) where \( \Delta \in [0, 1/2] \). We will assume that each message \( w_{k,i} \) is uniformly distributed in the set \( \mathcal{I}_M = \{1, 2, \ldots, 2^{M/2}\} \) and each message has a decoding delay of \( T \cdot M \) symbols. We will restrict our discussion to the case when \( T = 1 \).

In general when multiple messages arrive in each coherence block there exists an asymmetry in channel conditions experienced by these messages. For example in Fig. 7 when \( \Delta = 0 \) the message \( w_{k,1} \) only span one coherence block whereas the message \( w_{k,2} \) span across two coherence blocks and thus see two independent fading gains. Therefore the simple interleaving technique which was optimal in the case of a single message may not be optimal when there are multiple messages in each block. The following result shows that this is indeed the case.

**Proposition 1:** The optimal DMT of the SISO streaming setup with two messages per coherence block, and with \( \Delta = 0 \) and \( T = 1 \), is

\[
d(r) = \min \left(1 - \frac{r}{2}, 2 - 2r\right), \quad r \in [0, 1]. \tag{55}
\]

**Proof:** The upper bound is based on the following observation. The bound \( d(r) = 1 - r/2 \) follows by revealing every message \( w_{k,2} \) to the destination. The bound \( d(r) = 2 - 2r \) follows by revealing message \( w_{k,2} \) at the start of coherence block \( k \) and relaxing the deadline of \( w_k \) and \( w_{k+1} \) such that both only need to be recovered at the end of the coherence block \( k + 1 \). From Theorem 2 the associated DMT of this setup is \( d(r) = 2 - 2r \). The upper bound follows.
The achievability is as follows. We split each message $w_{k,1}$ into two equal sized messages $(w_{k,1}^1, w_{k,1}^2)$ of rate $R_0 = R/4$. We do not split the messages $w_{k,2}$. We sample three Gaussian codebooks as follows.

- The codebook $C_A$ consisting of $2^{3MR_0}$ codewords $x_A^{M/2}$ sampled i.i.d. from $\mathcal{CN}(0, \rho)$. Each pair $(w_{k,1}^1, w_{k,1}^2)$ is mapped to a unique codeword $x_A(w_{k,1}^1, w_{k,1}^2)$.

- The codebook $C_B$ consisting of $2^{MR_0}$ codewords $x_B^{M/2}$ sampled i.i.d. from $\mathcal{CN}(0, \rho)$. Each message $w_{k,1}^2$ is mapped to a unique codeword $x_B(w_{k,1}^2)$.

- The codebook $C_C$ consisting of $2^{2MR_0}$ codewords $x_C^{M/2}$ sampled i.i.d. from $\mathcal{CN}(0, \rho^{1-\beta})$. Each message $w_{k,2}$ is mapped to a unique codeword $x_C(w_{k,2})$. We will select $\beta = r/2$.

In coherence block $k$, the transmitter transmits $x_{k,i} = x_A(w_{k,1}^1, w_{k,1}^2)$ in the first half of the coherence block and $x_{k,II} = x_B(w_{k,1}^2) + x_C(w_{k,2})$ in the second half of the coherence block. The receiver observes $y_{k,i} = h_k x_{k,i} + z_{k,i}$ for $i \in \{I, II\}$. The decoding of the messages is as follows.

- The receiver decodes $w_{k,1}^2$ from $y_{k,II}$ treating $x_C$ as additional noise. An outage happens if

$$\left\{ \frac{1}{2} \log \left( 1 + \frac{|h_k|^2 \rho}{1 + |h_k|^2 \rho^{1-\beta}} \right) \leq \frac{r}{4} \log \rho \right\} \quad (56)$$

setting $|h_k|^2 = \rho^{-(1-\alpha)}$ and $\beta = r/2$, we can show that $\frac{56}{(56)}$ is equivalent to

$$\left\{ \log \left( 1 + \frac{\rho^\alpha}{1 + \rho^{\alpha}} \right) \leq \frac{r}{2} \log \rho \right\} \quad (57)$$

from which it can be shown that $d(r) = 1 - \frac{r}{2}$.

- After decoding $w_{k,1}^2$, the decoder subtracts $x_B(w_{k,1}^2)$ from $y_{k,II}$ i.e., $\tilde{y}_{k,II} = y_{k,II} - h_k x_B$. The decoder searches for a pair $(w_{k,2}, w_{k+1,1}^1)$ such that $(x_C(w_{k,2}), \tilde{y}_{k,II})$ are jointly typical and $(x_A(w_{k+1,1}^1, w_{k,2}))$ are jointly typical. An outage happens if

$$\left\{ \frac{1}{2} \log \left( 1 + \rho^{1-\beta}|h_k|^2 \right) + \frac{1}{2} \log(1 + \rho |h_{k+1}|^2) \leq \frac{3}{4} r \log \rho \right\} \quad (58)$$

Setting $|h_k|^2 = \rho^{-(1-\alpha_1)}$ and $|h_{k+1}|^2 = \rho^{-(1-\alpha_2)}$ and $\beta = r/2$ in $\frac{58}{(58)}$.

$$\log(1 + \rho^{\alpha_1-r/2}) + \log(1 + \rho^{\alpha_2}) \leq \frac{3r}{2} \log \rho \quad (59)$$

The associated DMT is given by

$$d_2(r) = \min_{(\alpha_1, \alpha_2) \in A} (1 - \alpha_1)^+ + (1 - \alpha_2)^+ \quad (60)$$

where $A = \{(\alpha_1, \alpha_2) \geq 0 : (\alpha_1 - r/2)^+ + \alpha_2 \leq 3r/2\}$.

It can be deduced that $d_2(r) = 2 - 2r$. Thus the DMT associated with the decoding of $(w_{k,2}, w_{k+1,1}^1)$ equals $\min(1 - r/2, 2 - 2r)$. Since this analysis can be applied for each $k$ $\frac{55}{(55)}$ follows.

We note that if $\Delta = 1/2$, and the decoding delay equals $T = M$ symbols, then message $w_1$ spans across two coherence blocks whereas $w_2$ only spans one coherence block. By reversing the role of $w_{k,1}$ and $w_{k,2}$ in the coding scheme in Prop. $\Box$ we can still achieve the DMT in $\frac{55}{(55)}$. However the following result shows that we cannot have a universal coding scheme oblivious of $\Delta$ that achieves the same DMT.
Proposition 2: Consider the SISO channel model with two messages in each coherence block as in Prop. 1. Assume that either \( \Delta = 0 \) or \( \Delta = 1/2 \), but the actual value of \( \Delta \) is known only to the receiver. The DMT for this setup equals \( d(r) = 1 - r \).

Proof: The achievability is straightforward. Each message \( w_{k,j} \) is mapped to a codeword of length \( M/2 \) of a Gaussian codebook and transmitted immediately. Since each message is of rate \( r_2 \log \rho \) the DMT of \( d(r) = 1 - r \) is achievable.

For the converse, we consider a multicast setup with two receivers. In coherence block \( k \) the transmitter transmits \( x_{k,I} \) in the first half of the coherence block and transmits \( x_{k,II} \) in the second half i.e., \( x_k = [x_{k,I} \ x_{k,II}] \), where both \( x_{k,I}, x_{k,II} \in \mathbb{C}^{M/2} \). Receiver 1 observes \( y_k = [y_{k,I} \ y_{k,II}] \) in coherence block \( k \) as follows:

\[
y_{k,I} = h_k x_{k,I} + n_{k,I,1}, \tag{61}
y_{k,II} = h_k x_{k,II} + n_{k,II,1} \tag{62}
\]

where receiver 2 observes \( v_k = [v_{k,I} \ v_{k,II}] \) in coherence block \( k \) as follows:

\[
v_{k,I} = h_k x_{k,I} + n_{k,I,2}, \tag{63}
v_{k,II} = h_{k+1} x_{k,II} + n_{k,II,2} \tag{64}
\]

where the noise variables \( n_{k,j,l} \) have i.i.d. \( \mathcal{CN}(0,1) \) entries. For both receivers, message \( w_{k,1} \) must be decoded at the end of the coherence block \( k \) and message \( w_{k,2} \) must be decoded in the middle of coherence block \( k + 1 \). Note that the duration of \( w_{k,1} \) spans only one fading state \( h_k \), for receiver 1 while \( w_{k,2} \) spans only one fading state \( h_{k+1} \), for receiver 2. By construction any feasible coding scheme for the original channel where the transmitter is oblivious of \( \Delta \) must be simultaneously feasible for the two receivers on the multicast channel. We show that under this constraint \( d(r) = 1 - r \) is the maximum possible DMT.

We begin by considering Fano’s inequality for receiver 1 for message \( w_{0,1} \) and rate \( \frac{M r_1}{2} \log \rho \):

\[
\Pr(\mathcal{E}_{0,1}; h_0 = h_0) \geq 1 - \frac{2}{\frac{M r_1}{2} \log \rho} - \frac{2I(w_{0,1}; y_1|h_0 = h_0)}{\frac{M r_1}{2} \log \rho} \tag{65}
\]

Ignoring the second term, which goes to zero as \( M \to \infty \) and using the same sequence of steps leading to (16) we have with \( P_\delta = \rho^{-(1-r+\delta)} \)

\[
\Pr(\mathcal{E}_{0,1}) \geq P_\delta \left( 1 - \frac{2I(w_{0,1}; y_0|h_0 \in \mathcal{H}_\delta)}{\frac{M r_1}{2} \log \rho} \right) \tag{66}
\]

\[
= P_\delta \left( 1 - \frac{2I(w_{0,1}; y_0|h_0^N \in \mathcal{H}_\delta^{N+1})}{\frac{M r_1}{2} \log \rho} \right) \tag{67}
\]

\[
\geq P_\delta \left( 1 - \frac{2I(w_{0,1}; y_0^N, v_0^N|h_0^N \in \mathcal{H}_\delta^{N+2})}{\frac{M r_1}{2} \log \rho} \right) \tag{68}
\]

where (67) follows from the fact that \( h_1^N \) is independent of \( (w_0, y_0, h_0) \).
Similarly applying Fano’s inequality for receiver 2 for message \( w_{0,2} \) we have

\[
\Pr(\mathcal{E}_{0,2}) \geq P_\delta \left( 1 - \frac{2I(w_{0,2}; \mathbf{x}_1, w_{0,1}, h_1 \in \mathcal{H}_\delta)}{Mr \log \rho} \right) \tag{69}
\]

\[
= P_\delta \left( 1 - \frac{2I(w_{0,2}; \mathbf{x}_1, w_{0,1}, h_0^{N+1} \in \mathcal{H}_\delta^{N+2})}{Mr \log \rho} \right) \tag{70}
\]

\[
\geq P_\delta \left( 1 - \frac{2I(w_{0,2}; y_0^N, w_0^N | w_{0,1}, h_0^{N+1} \in \mathcal{H}_\delta^{N+2})}{Mr \log \rho} \right) \tag{71}
\]

Likewise we can show that for each \( k \leq N - 1 \)

\[
\Pr(\mathcal{E}_{k,1}) \geq P_\delta \left( 1 - \frac{2I(w_{k,1}; y_0^N, w_0^N | h_0^{N+1} \in \mathcal{H}_\delta^{N+2}, w_{k-1}^k)}{Mr \log \rho} \right) \tag{72}
\]

\[
\Pr(\mathcal{E}_{k,2}) \geq P_\delta \left( 1 - \frac{2I(w_{k,2}; y_0^N, w_0^N | h_0^{N+1} \in \mathcal{H}_\delta^{N+2}, w_{k-1}^k, w_{k,1})}{Mr \log \rho} \right) \tag{73}
\]

Thus we have that

\[
\max_{0 \leq k \leq N-1} \max \{ \Pr(\mathcal{E}_{k,1}, \Pr(\mathcal{E}_{k,2})) \} \geq \frac{1}{2N} \sum_{k=0}^{N-1} \{ \Pr(\mathcal{E}_{k,1}) + \Pr(\mathcal{E}_{k,2}) \} \tag{74}
\]

\[
\geq P_\delta \left( 1 - \frac{I(w_0^{N-1}; y_0^N, w_0^N | h_0^{N+1} \in \mathcal{H}_\delta^{N+1})}{Nm \log \rho} \right) \tag{75}
\]

\[
\geq P_\delta \left( 1 - \frac{I(x_0^N; y_0^N, w_0^N | h_0^{N+1} \in \mathcal{H}_\delta^{N+1})}{Nm \log \rho} \right) \tag{76}
\]

\[
\geq P_\delta \left( 1 - \frac{\sum_{k=0}^{N} I(x_k; y_k, w_k | h_k) + I(x_{k,1}; y_{k,1}, w_{k,1} | h_k, h_{k+1})}{Nm \log \rho} \right) \tag{77}
\]

\[
\geq \rho^{-(1-r+\delta)} \left( 1 - \frac{N + 1 + (N + 1)(r - \delta) \log \rho}{N \rho} \right) \tag{78}
\]

where the steps leading to (78) are similar to (21) and hence not elaborated. For \( N \) sufficiently large the expression in the brackets in (78) is positive. This establishes that \( d(r) \geq 1 - r + \delta \) must hold. Since \( \delta > 0 \) is arbitrary this concludes the converse in Prop. 2.

We conclude this section with the following remark. When there are multiple messages, at equal intervals arriving in each coherence block different messages observe different channel conditions. Prop. 1 shows that coding schemes that exploits this asymmetry across the messages indeed improve the DMT. On the other hand such schemes crucially rely on where the messages arrive in each block. If such information is not available the DMT is in general smaller, as established in Prop. 2.

VIII. CONCLUSIONS

We studied the problem of delay constrained streaming over a block fading channel and established that the associated diversity multiplexing tradeoff when there is one message arriving in each coherence block. The converse
is based on an outage-amplification argument and does not follow from simple reductions to known upper bound. The DMT can be achieved using an interleaving scheme that reduces the system to a set of parallel independent channels. We also show that another coding scheme, that uses a sequential tree code, can achieve the DMT in a delay universal fashion. We also discuss some extensions when multiple messages arrive in each coherence block.

We believe that the fundamental limits of delay-constrained streaming over fading channels are not well understood and the techniques developed in this work can be a useful starting point for many other scenarios in wireless communications.

**Appendix A**

**Proof of Corollary**

Let us define:

\[ C_l(\rho) = \log \det \left( \mathbb{I} + \rho \frac{H_l H_l^\dagger}{N_t} \right) , \quad l = 1, \ldots, L \]  

(79)

which is the maximum mutual information over channel \( l \). Using Theorem 1 we have that

\[ \Pr(C_l(\rho) \leq r \log \rho) = \rho^{-d_1(r)} , \]

(80)

Let \( \Sigma(\rho) = \sum_{l=1}^L C_l(\rho) \).

\[ \Pr(\Sigma(\rho) \leq s \log \rho) \]

\[ = \Pr \left( \bigcup_{\sum_{l=1}^L r_l \leq s, r_l \geq 0} \bigcap_{l=1}^L \{ C_l(\rho) \leq r_l \log \rho \} \right) \]

(81)

\[ \leq \sum_{\sum_{l=1}^L r_l \leq s, r_l \geq 0} \Pr \left( \bigcup_{l=1}^L \{ C_l(\rho) \leq r_l \log \rho \} \right) \]

(82)

\[ \leq \sum_{\sum_{l=1}^L r_l \leq s, r_l \geq 0} \prod_{l=1}^L \Pr(\{C_l(\rho) \leq r_l \log \rho\}) \]

(83)

\[ = \sum_{\sum_{l=1}^L r_l \leq s, r_l \geq 0} \rho^{-\sum_{l=1}^L d_1(r_l)} \]

(84)

where (82) follows from the union bound, (83) follows because the random variables in (79) are i.i.d. random variables and (84) follows by substituting in (80).

Applying Vardhan’s Lemma [22] we have that

\[ \Pr(\Sigma(\rho) \leq s \log \rho) = \rho^{-d(s)} , \]

(85)

where

\[ d(s) = \min_{\sum_{l=1}^L r_l \leq s, r_l \geq 0} \sum_{l=1}^L d_1(r_l) . \]

(86)
Since the objective function on the right hand side of (86) is symmetric and convex in \( r_1, \ldots, r_L \), the minimum happens when \( r_1 = r_2 = \ldots = r_L = s/L \), thus yielding
\[
d(s) = Ld_1 \left( \frac{s}{L} \right),
\]
as required.

**APPENDIX B**

**PROOF OF (35).**

Our proof is based on the Chernoff-Cramer theorem of large deviations stated below.

**Theorem 3:** Suppose that \( x_1, \ldots, x_N \) are i.i.d. random variables with a rate function \( f_x(\cdot) \) defined as
\[
f_x(t) = \sup_\theta \left\{ \theta \cdot t - \log E_x [\exp(\theta \cdot x)] \right\},
\]
and let \( M_n = \frac{1}{n} \sum_{i=1}^{n} x_i \). Then there exists a constant \( N > 0 \) such that for all \( n \geq N \) we have,
\[
\Pr(M_n \geq t) \leq e^{-nf_x(t)}
\]
\(\square\)

Note that as \( Y_k = H_kX_k + Z_k \), the \( H_k \) are known to the decoder, and the noise sequence \( \{Z_k\} \) is independent,
\[
p_{X_k, Y_k}(X_k, Y_k) = p_{X_k}(X_k)p_{Y_k|X_k}(Y_k|X_k)
\]
\[
= p_{X_k}(X_k)p_{Z_k}(Y_k - H_k \cdot X_k)
\]
\[
= p_{X_k}(X_k)p_{Z_k}(Z_k)
\]
where the last equality holds since the codewords are sampled i.i.d. and the noise is also i.i.d. Thus \( h(p_{X_k, Y_k}) = h(p_X) + h(p_Z) \). And so
\[
\left| \sum_{k=l}^{l'} [\log p_{X_k, Y_k}(X_k, Y_k) - h(p_{X_k, Y_k})] \right| = \left| \sum_{k=l}^{l'} [\log p_{X_k}(X_k) + \log p_{Z_k}(Z_k) - h(p_X) - h(p_Z)] \right|
\]
\[
\leq \left| \sum_{k=l}^{l'} [\log p_{X_k}(X_k) - h(p_X)] \right| + \left| \sum_{k=l}^{l'} [\log p_{Z_k}(Z_k) - h(p_Z)] \right|
\]

Recall that \( A_{l,l'} \) is the event that the true codeword is not jointly typical with the received sequence. To upper bound the probability we can ignore the marginal typicality constraints and use
\[
\Pr(A_{l,l'}) \leq \Pr \left( \left| \frac{\sum_{k=l}^{l'} [\log p_{X_k, Y_k}(X_k, Y_k) - h(p_{X_k, Y_k})]}{M(l' - l + 1)} \right| > \varepsilon \right).
\]
where the last step follows from the triangular inequality. Substituting (95) into (90) and using the union bound we have

$$\Pr(A_{l,l'}) \leq \Pr(A_{l,l'}^X) + \Pr(A_{l,l'}^Z)$$  \hspace{1cm} (96)

where we define

$$A_{l,l'}^X = \left\{ X_{l'} : \left| \sum_{k=l'} \log p_X(X_k) - h(p_X) \right| \geq \varepsilon \right\},$$  \hspace{1cm} (97)

$$A_{l,l'}^Z = \left\{ Z_{l'} : \left| \sum_{k=l'} \log p_Z(Z_k) - h(p_Z) \right| \geq \varepsilon \right\},$$  \hspace{1cm} (98)

Note that $X_k$ is a sequence of $M$ i.i.d. random vectors each sampled from $\mathcal{CN}(0, \rho^N \mathbf{I})$ and $E[\log p_X(X_k)] = h(p_X)$. Similarly, $E[\log p_Z(Z_k)] = h(p_Z)$. Then using Theorem 3 there exist functions $f_X(\varepsilon)$ and $f_Z(\varepsilon)$ such that for sufficiently large $N = M(l' - l + 1)$

$$\Pr(A_{l,l'}^X) \leq \exp\{-M(l' - l + 1)f_X(\varepsilon)\},$$

$$\Pr(A_{l,l'}^Z) \leq \exp\{-M(l' - l + 1)f_Z(\varepsilon)\}.$$  

Furthermore by directly using (88) we can show that $f_X(\varepsilon) > 0$ and $f_Y(\varepsilon) > 0$. Setting $f(\varepsilon) = \max(f_X(\varepsilon), f_Z(\varepsilon))$ establishes (35).

**References**

[1] A. Khisti and S. C. Draper, “Streaming data over fading wireless channels: The diversity-multiplexing tradeoff,” in ISIT, 2011, pp. 1111–1115.

[2] S. V. Hanly and D. N. C. Tse, “Multiaccess Fading Channels-Part II: Delay-Limited Capacities,” IEEE Transactions on Information Theory, vol. 44, no. 7, pp. 2816–2831, 1998.

[3] G. Caire, G. Taricco, and E. Biglieri, “Optimum power control over fading channels,” IEEE Transactions on Information Theory, vol. 45, no. 5, pp. 1468–1489, 1999.

[4] E. Biglieri, J. G. Proakis, and S. Shamai, “Fading channels: Information-theoretic and communication aspects,” IEEE Transactions on Information Theory, vol. 44, no. 6, pp. 2619–2692, 1998.

[5] D. Tse and P. Viswanath, Fundamentals of Wireless Communication. Cambridge University Press, 2005.

[6] S. Shamai and A. Steiner, “A broadcast approach for a single-user slowly fading MIMO channel,” IEEE Transactions on Information Theory, vol. 49, no. 10, pp. 2619–2635, 2003.

[7] P. A. Whiting and E. M. Yeh, “Broadcasting over uncertain channels with decoding delay constraints,” IEEE Transactions on Information Theory, vol. 52, no. 3, pp. 904–921, 2006.

[8] J. W. Yoo, T. Liu, and S. S. (Shitz), “Worst-case expected-rate loss of slow-fading channels,” in ISIT, 2012, pp. 2186–2190.

[9] K. M. Cohen, A. Steiner, and S. S. (Shitz), “The broadcast approach under mixed delay constraints,” in ISIT, 2012, pp. 209–213.

[10] G. Cocco, D. Gündüz, and C. Ibars, “Streaming transmitter over block-fading channels with delay constraint,” CoRR, vol. abs/1203.2870, 2012.

[11] C. T. Ng, D. Gunduz, A. Goldsmith, and E. Erkip, “Distortion minimization in Gaussian layered broadcast coding with successive refinement,” IEEE Trans. Inform. Theory, vol. 55, pp. 5074–5086, Nov. 2009.

[12] G. Caire and K. Narayan, “On the distortion SNR exponent of hybrid digital analog space time coding,” IEEE Trans. Inform. Theory, vol. 53, pp. 2867–2878, Aug. 2007.
[13] K. Bhattad, G. Caire, and K. Narayan, “On the distortion SNR exponent of some layered transmission schemes,” *IEEE Trans. Inform. Theory*, vol. 54, pp. 2943–2958, Jul. 2008.

[14] T. Kim, M. Skoglund, and G. Caire, “On source transmission over MIMO channels with limited feedback,” *IEEE Trans. Signal Processing*, vol. 57, pp. 324–341, Jan. 2009.

[15] M. Effros, A. Goldsmith, and Y. Liang, “Generalizing capacity: New definitions and capacity theorems for composite channels,” *IEEE Trans. Inform. Theory*, vol. 56, pp. 3069–3087, Jul. 2010.

[16] F. Etemadi and H. Jafarkhani, “A unified framework for layered transmission over fading and packet erasure channels,” vol. 55, pp. 565–573, Apr. 2008.

[17] C. Tian, A. Steiner, S. Shamai, and S. Diggavi, “Successive refinement via broadcast: Optimizing expected distortion of a Gaussian source over a Gaussian fading channel,” *IEEE Trans. Inform. Theory*, vol. 54, pp. 2903–2918, Jul. 2008.

[18] D. Gunduz and E. Erkip, “Joint source-channel codes for MIMO block fading channels,” *IEEE Trans. Inform. Theory*, vol. 54, pp. 116–134, Jan. 2008.

[19] A. Kwasinski, Z. Han, and K. J. R. Liu, “Cooperative multimedia communications: joint source coding and collaboration,” in *GLOBECOM*, Nov. 2005, pp. 374–378.

[20] C. Bilen, E. Erkip, and Y. Wang, “Layered video multicast using diversity embedded space time codes,” in *Sarnoff Symposium*, Apr. 2009, pp. 1–5.

[21] X. Yu, H. Wang, E. Yang, and M. Uysal, “Tradeoff between diversity and decodable-rate: Diversity multiplexing tradeoff for fixed encoding schemes,” in *CWIT*, May 2009, pp. 79–82.

[22] L. Zheng and D. Tse, “Diversity and multiplexing: A fundamental tradeoff in multiple antenna channels,” *IEEE Trans. Inform. Theory*, vol. 49, pp. 1073–1096, May 2003.

[23] K. Azarian, H. E. Gamal, and P. Schniter, “On the achievable diversity-multiplexing tradeoff in half-duplex cooperative channels,” *IEEE Trans. Inform. Theory*, vol. 51, pp. 4152–4172, Dec. 2005.

[24] A. Bletsas, A. Khisti, D. Reed, and A. Lippman, “A simple cooperative diversity method based on network path selection,” *IEEE J. Select. Areas Commun.*, vol. 24, pp. 659–672, Mar. 2005.

[25] H. E. Gamal, G. Caire, and M. O. Damen, “The MIMO ARQ channel: Diversity-multiplexing-delay tradeoff,” *IEEE Trans. Inform. Theory*, vol. 52, pp. 3601–3621, Aug. 2006.

[26] J. Laneman, D. Tse, and G. Wornell, “Cooperative diversity in wireless networks: Efficient protocols and outage behavior,” *IEEE Trans. Inform. Theory*, vol. 50, pp. 3062–3080, Dec. 2004.

[27] A. Lozano and N. Jindal, “Transmit diversity vs. spatial multiplexing in modern MIMO systems,” *IEEE Transactions on Wireless Communications*, vol. 9, pp. 186–197, Jan. 2010.

[28] K. Azarian and H. E. Gamal, “The throughput-reliability tradeoff in block-fading mimo channels,” *IEEE Trans. Inform. Theory*, vol. 54, pp. 488–501, Feb. 2007.

[29] E. Stauffer, O. Oyman, R. Narasimhan, and A. Paulraj, “Finite-SNR diversity-multiplexing tradeoffs in fading relay channels,” *IEEE J. Select. Areas Commun.*, vol. 25, pp. 245–257, Feb. 2007.

[30] T. Kim and M. Skoglund, “Diversity-multiplexing tradeoff in MIMO channels with partial CSIT,” *IEEE Trans. Inform. Theory*, vol. 53, pp. 2743–2759, Aug. 2007.

[31] H. E. Gamal, G. Caire, and M. O. Damen, “Lattice coding and decoding achieve the optimal diversity-multiplexing tradeoff of MIMO channels,” *IEEE Trans. Inform. Theory*, vol. 50, pp. 968–985, Jun. 2004.

[32] P. Elia, K. R. Kumar, S. A. Pawar, P. V. Kumar, and H.-F. Lu, “Explicit space-time codes achieving the diversity-multiplexing gain tradeoff,” *IEEE Trans. Inform. Theory*, vol. 52, pp. 3869–3884, Sep. 2006.

[33] P. Elia, B. A. Sethuraman, and P. V. Kumar, “Perfect space-time codes for any number of antennas,” *IEEE Trans. Inform. Theory*, vol. 55, pp. 3853–3868, Nov. 2007.

[34] P. Elia, S. Kittipiyakul, and T. Javidi, “On the responsiveness-diversity-multiplexing tradeoff,” in *WoOpt*, Apr. 2007.

[35] S. Kittipiyakul, P. Elia, and T. Javidi, “High-SNR analysis of outage-limited communications with bursty and delay-limited information,” *IEEE Trans. Inform. Theory*, vol. 55, pp. 746–763, Feb. 2009.
[37] S. Dusad and S. N. Diggavi, “On successive refinement of diversity for fading ISI channels,” IEEE Transactions on Communications, vol. 59, no. 7, pp. 1820–1829, 2011.

[38] Y. Yona and M. Feder, “The fundamental limits of infinite constellations in MIMO fading channels,” in ISIT, 2011, pp. 1836–1840.

[39] A. Sahai and S. Mitter, “The necessity and sufficiency of anytime capacity for stabilization of a linear system over a noisy communication link Part I: Scalar systems,” IEEE Trans. Inform. Theory, vol. 52, pp. 3369–3395, Aug. 2006.

[40] A. Sahai, “Why do block length and delay behave differently if feedback is present?” IEEE Trans. Inform. Theory, vol. 54, pp. 1860–1886, May 2008.

[41] ——, “Anytime information theory,” Ph.D. dissertation, Mass. Instit. of Tech., 2001.

[42] A. Sahai, S. Avestimehr, , and P. Minero, “Anytime communication over the Gilbert-Eliot channel with noiseless feedback,” in Proc. Int. Symp. Inform. Theory, 2005.

[43] C. Chang and A. Sahai, “Sequential random coding error exponents for multiple access channels,” in WirelessCom, 2005.

[44] S. Draper, C. Chang, and A. Sahai, “Sequential random binning for streaming distributed source coding,” in Proc. Int. Symp. Inform. Theory, 2005.

[45] A. Sahai, “Anytime coding on the infinite bandwidth AWGN channel: A sequential semi-orthogonal optimal code,” unpublished, http://arxiv.org/abs/cs/0610151, Oct. 2006.

[46] E. Martinian and G. Wornell, “Universal codes for minimizing per-user delay on streaming broadcast channels,” in Proc. Allerton Conf. Commun., Cont., Computing, 2003.

[47] E. Martinian and M. Trott, “Delay-optimal burst erasure code construction,” in Proc. Int. Symp. Inform. Theory, 2007.

[48] A. Badr, A. Khisti, and E. Martinian, “Diversity embedded streaming erasure codes (DE-SCo): Constructions and optimality,” IEEE J. Select. Areas Commun., May 2011.

[49] A. Badr, A. Khisti, W.-T. Tan, and J. Apostolopoulos, “Streaming codes for channels with burst and isolated erasures,” CoRR, vol. abs/1208.0072, 2012.

[50] A. Khisti and J. Singh, “On multicasting with streaming burst-erasure codes,” in ISIT, 2009, pp. 2887–2891.

[51] A. Badr, D. Lui, and A. Khisti, “Multicast streaming codes (Mu-SCo) for burst erasure channels,” in 48th Annual Allerton Conference on Communication, Control and Computing, Oct 2010, pp. 383 –390.

[52] D. Leong and T. Ho, “Erasure coding for real-time streaming,” in ISIT, 2012, pp. 289–293.

[53] T. M. Cover and J. A. Thomas, Elements of Information Theory. John Wiley and Sons, 1991.