The Streaming-DMT of Fading Channels

Ashish Khisti *Member, IEEE,* and Stark C. Draper *Member, IEEE*

***Abstract*—We consider the sequential transmission of a stream of messages over a block-fading multi-input-multi-output (MIMO) channel. A new message arrives at the beginning of each coherence block, and the decoder is required to output each message sequentially, after a delay of T coherence blocks. In the special case when $T = 1$, the setup reduces to the quasi-static fading channel. We establish the optimal diversity-multiplexing tradeoff (DMT) in the high signal-to-noise-ratio (SNR) regime, and show that it equals T times the DMT of the quasi-static channel. The converse is based on utilizing the delay constraint to amplify a local outage event associated with a message, globally across all the coherence blocks. This approach appears to be new. We propose two coding schemes that achieve the optimal DMT. The first scheme involves interleaving of messages, such that each message is transmitted across T consecutive coherence blocks. This scheme requires the knowledge of the delay constraint at both the encoder and decoder. Our second coding scheme involves a sequential tree code and is delay-universal i.e., the knowledge of the decoding delay is not required by the encoder. However, in this scheme we require the coherence block-length to increase as $\log(\mathrm{SNR})$, in order to attain the optimal DMT. Finally, we discuss the case when multiple messages arrive at uniform intervals *within* each coherence period. Through a simple example we exhibit the sub-optimality of interleaving, and propose another scheme that achieves the optimal DMT.**

***Index Terms*—Real-Time Streaming Communication, Diversity-Multiplexing Tradeoff, Block-Fading, Tree Codes, Interleaving.**

I. Introduction

Multimedia applications require real-time encoding of a source stream and a sequential reconstruction of each source-packet 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]. Unfortunately the characterization of the outage capacity is in general a challenging problem, even in point-to-point settings [6]. A somewhat coarse metric for studying the outage capacity is the diversity-multiplexing tradeoff (DMT), first introduced in [7]. The authors propose *diversity order* and *multiplexing gain* as two fundamental metrics for communication over a wireless channel, and establish a tradeoff between these for quasi-static, multi-input-multi-output (MIMO) fading channels, in the high signal-to-noise-ratio (SNR) regime. A significant body of literature on DMT now exists, see e.g., [5, Chapter 9]. Of particular interest in this work is the case of T independent parallel MIMO fading channels where the optimal DMT equals T times the DMT of the quasi-static MIMO fading channel, with a suitably normalized multiplexing gain [7]. Practical code constructions for parallel fading channels have been proposed in [8]–[10]. Interestingly, when the parallel channels are correlated, the DMT analysis is far more intricate and only special cases are known [11], [12].

In the present paper we consider the problem of real-time streaming over a MIMO block-fading wireless channel. We assume that the transmitter observes a sequence of independent messages. One message arrives per coherence block, right at the start of the block. The input into the channel can depend on all the past messages, but not on any future messages. The decoder is required to output each message with a maximum delay of T coherence blocks. When $T = 1$, each message sees only one fading realization and the setup reduces to the quasi-static fading channel model. In general, each message experiences T independent fading blocks; however it must be multiplexed together with messages arriving in other coherence blocks. We declare a message to be in outage if it cannot be decoded by its deadline. We establish the optimal DMT in this streaming setting and show that it equals T times the DMT of the quasi-static MIMO fading channel. The optimal DMT can be achieved by a simple interleaving of messages across coherence blocks and transmitting each message over T parallel, MIMO fading channels. We also propose an alternative tree-code that attains the optimal DMT. In this scheme the delay constraint only needs to be revealed to the decoder, and not to the encoder, and thus it is suitable for applications where a common source stream must be transmitted to multiple receivers with different decoding delays. However in order to achieve the optimal DMT, the coherence block length for tree-codes must increase with $\log(\mathrm{SNR})$ and thus this scheme appears to require long coherence periods.

Tree codes for streaming communication over a discrete memoryless channels (DMC) have been studied previously in [13]–[17]. These works, however, consider maximum likelihood and universal decoders. In contrast, our analysis of the tree code is based on a very different outage analysis paired with a decision directed decoder. We express our error probability as a sum of two terms — one term decreases exponentially in $\log(\mathrm{SNR})$ while the other decreases exponentially in the coherence block-length. By suitably balancing the two exponents we establish that our proposed scheme attains the optimal DMT. Another recent work, reference [18], studies a related setup when the transmitter sequentially observes a

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. S. Draper's work was supported by the National Science Foundation under CAREER grant CCF 0844539 and by an NSERC Discovery Grant.

stream of messages, but assumes that all the messages have a common deadline. A variety of coding techniques such as adaptive joint encoding, memoryless transmission, time sharing and superposition transmission are compared in different delay and SNR regimes. Another followup work [19] considers the case when all the messages are available to the encoder, but have different playback deadline at the receiver. In contrast our proposed setup requires that each incoming message must be reconstructed after a fixed decoding deadline, which is relevant in applications such as real-time voice and video streaming. Finally in yet another related work [20], [21], the authors study the transmission of bursty and delay-sensitive data source over a constant-rate MIMO fading channel and establish an optimal operating point on the DMT that balances the channel outage and *delay-violation* probabilities. However the results are valid only for asymptotically large decoding delays. Furthermore it appears that the coding techniques considered in these works do not retransmit the same information bits across multiple coherence blocks, a key idea exploited in the present paper.

In the rest of the paper we describe the system model in Section II, and the main result, that characterizes the streaming-DMT in Section III. We provide the proof of the converse in Section IV. The coding schemes based on interleaving and tree-codes are presented in Section V. We discuss extension to the case of multiple messages in Section VI and provide conclusions in Section VII.

Throughout the paper we will use the following notation. Upper case bold-font will be reserved for matrices (e.g., $\mathbf{H}$) whereas lower case bold-font (e.g., $\mathbf{x}$) will be used for vectors. Scalar symbols will be denoted using lower case non-bold fonts. We will use the sans serif font for random variables e.g., x. A sequence of symbols $x_i, x_{i+1}, \ldots, x_j$ will be denoted using the notation x_i^j. Throughout the paper the symbol $\doteq$ will be reserved to denote equality in the exponential sense i.e., we express, $f(\rho) \doteq \rho^b$, if $\lim_{\rho\to\infty} \frac{\log f(\rho)}{\log \rho} = b$ holds. The symbols $\dot{\leq}$ and $\dot{\geq}$ will be defined in a similar fashion.

II. Model

We consider an independent identically distributed (i.i.d.) block fading channel model with a coherence period of M:

$$\mathbf{\mathsf{Y}}_k = \mathbf{\mathsf{H}}_k \cdot \mathbf{X}_k + \mathbf{\mathsf{Z}}_k, \tag{1}$$

where $k = 0, 1, \ldots$, denotes the index of the coherence block of the fading channel. The matrix $\mathbf{\mathsf{H}}_k \in \mathbb{C}^{N_\mathrm{r}\times N_\mathrm{t}}$ 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.

$$\mathbf{\mathsf{X}}_k = [\mathbf{\mathsf{x}}_k(1) \mid \ \ldots \ \mid \mathbf{\mathsf{x}}_k(M)] \in \mathbb{C}^{N_t \times M}$$

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

$$\mathbf{\mathsf{y}}_k(j) = \mathbf{\mathsf{H}}_k \cdot \mathbf{\mathsf{x}}_k(j) + \mathbf{\mathsf{z}}_k(j), \qquad j = 1, \ldots, M. \tag{2}$$

We assume that all entries of $\mathbf{\mathsf{H}}_k$ are sampled independently from the complex Gaussian distribution[1] with zero-mean and unit-variance i.e., $\mathcal{CN}(0, 1)$. The channel remains constant during each coherence block and is sampled independently across blocks. All entries of the additive noise matrix $\mathbf{\mathsf{Z}}_k$ are also sampled i.i.d. $\mathcal{CN}(0, 1)$. Finally the realization of the channel matrices $\mathbf{\mathsf{H}}_k$ is revealed to the decoder, but not to the encoder.

We assume an average (short-term) power constraint $E[\sum_{i=1}^M ||\mathbf{\mathsf{x}}_k(i)||^2] \leq M\rho$. Note that ρ denotes the transmit power which will serve as our SNR parameter. A delay-constrained streaming code is defined as follows:

Definition 1 (Streaming Code): A rate R streaming code with delay T, $\mathcal{C}(R, T)$, consists of

1. A sequence of messages $\{\mathsf{w}_k\}_{k\geq 0}$ each distributed uniformly over the set $\mathcal{I}_M = \{1, 2, \ldots, 2^{MR}\}$.
2. A sequence of encoding functions $\mathcal{F}_k : \mathcal{I}_M^{k+1} \to \mathbb{C}^{N_\mathrm{t}\times M}$,
$$\mathbf{\mathsf{X}}_k = \mathcal{F}_k(\mathsf{w}_0, \ldots, \mathsf{w}_k), \quad k = 0, 1, \ldots \tag{3}$$
that maps the input message sequence to the channel input matrix $\mathbf{\mathsf{X}}_k \in \mathbb{C}^{N_\mathrm{t}\times M}$.
3. A sequence of decoding functions $\mathcal{G}_k : \mathbb{C}^{M(k+T)} \to \mathcal{I}_M$ that outputs message estimate $\hat{\mathsf{w}}_k$ based on the first $k+T$ observations, i.e.,
$$\hat{\mathsf{w}}_k = \mathcal{G}_k(\mathbf{\mathsf{Y}}_0, \ldots, \mathbf{\mathsf{Y}}_{k+T-1}), \quad k = 0, 1, \ldots \tag{4}$$

Fig. 1 illustrates such a setup for the case when $T = 2$. One message w_k arrives at the start of each coherence block. The codeword transmitted in block k, $\mathbf{\mathsf{X}}_k(\mathsf{w}_0^k)$ can depend on all the past messages, but not on any future messages. Since $T = 2$, the receiver must decode message w_k at the end of coherence block $k + 1$ i.e., $\hat{\mathsf{w}}_k = \mathcal{G}_k(\mathbf{\mathsf{Y}}_0, \ldots, \mathbf{\mathsf{Y}}_{k+1})$.

We now define the diversity-multiplexing tradeoff (DMT) associated with the streaming code $\mathcal{C}(R, T)$. The error probability for the k-th message is $\Pr[\mathsf{w}_k \neq \hat{\mathsf{w}}_k]$ where $\hat{\mathsf{w}}_k$ is the decoder output (4) and the error probability is averaged over the random channel gains. In the DMT both error probability and rate are studied as a function of the SNR parameter ρ. Let $e_{\max}(\rho) = \sup_{k\geq 0} \Pr[\hat{\mathsf{w}}_k \neq \mathsf{w}_k]$ denote the worst-case error probability of a rate-$R(\rho)$ code. A DMT tradeoff [7] of (r, d) is said to be achievable with delay T if there exists a sequence of codebooks $\mathcal{C}(R(\rho), T)$ achieving $e_{\max}(\rho)$ such that

$$r = \lim_{\rho\to\infty} \frac{R(\rho)}{\log \rho}, \quad d = \lim_{\rho\to\infty} \frac{-\log e_{\max}(\rho)}{\log \rho}. \tag{5}$$

Of interest, is the optimal diversity-multiplexing tradeoff, denoted by $d_T(r)$. [2]

One class of wireless systems that motivates this model is frequency-hopping orthogonal frequency-division multiple

[1] While we only focus on the Rayleigh channel model, our results easily extend to other channel models.

[2] We caution the reader that in the above discussion $e_{\max}(\rho)$ 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 there will eventually be a block fade that induces an outage. In our definition we fix an index, k, and find the error probability $\Pr[\hat{\mathsf{w}}_k \neq \mathsf{w}_k]$ averaged over the channel gains. We subsequently search for the index k with the maximum error probability. E.g., for time-invariant coding schemes operating in the steady-state regime, due to symmetry, $\Pr[\hat{\mathsf{w}}_k \neq \mathsf{w}_k]$ will not depend on k.

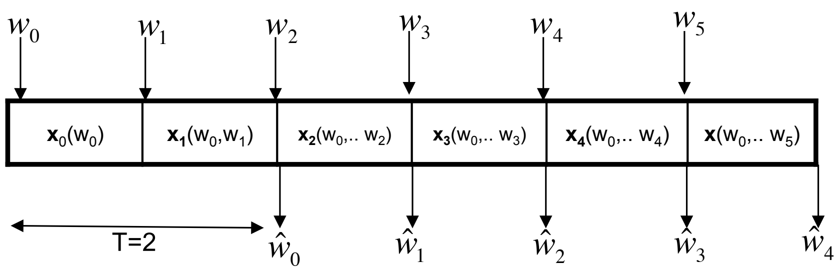

Fig. 1. Proposed Streaming Model. One new message arrives at the start of each coherence block. The message stream is encoded sequentially and each message needs to be output at the receiver after T coherence blocks. In the above figure $T = 2$.

access (OFDMA) systems. Here the frequency bands (or "sub-channel") allocated to a user are changed at regular intervals and in a randomized fashion. In frequency-selective channels the randomized sub-channel allocation means that the channels the the user experiences pre- and post-hop are approximately independent, as long as the expected recurrence time of a particular sub-channel exceeds the coherence time of the channel. In this setting we can characterize the playback deadline in terms of the number of hops T until the message w_k must be estimated by the receiver. This can be translated into a number of channel uses, TM, where M denotes the number of symbols transmitted in each hop, and hence into time. In practical systems the value of M may be fixed and thus not under the control of the application.

III. Main Result

The optimal tradeoff between diversity and multiplexing (DMT) for the quasi-static fading channel was characterized in [7]. We reproduce the result below for the convenience of the reader.

Theorem 1: (Zheng and Tse, [7]) For the quasi-static fading channel

$$\mathbf{y}(t) = \mathbf{H} \cdot \mathbf{x}(t) + \mathbf{z}(t) \tag{6}$$

where the entries of $\mathbf{H} \in \mathbb{C}^{N_\mathrm{r} \times N_\mathrm{t}}$ 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 $k = 0, 1, \ldots, \min(N_\mathrm{r}, N_\mathrm{t})$ where

$$d_1(k) = (N_\mathrm{r} - k)(N_\mathrm{t} - k). \tag{7}$$

□

In our analysis the following generalization of the quasi-static DMT to L parallel channels, see [7, Corollary 8] [8], [9] is useful.

Corollary 1: Consider a collection of L parallel quasi-static fading channel

$$\mathbf{y}_l(t) = \mathbf{H}_l \cdot \mathbf{x}_l(t) + \mathbf{z}_l(t), \quad l = 1, \ldots, L \tag{8}$$

where the entries of $\mathbf{H}_l \in \mathbb{C}^{N_\mathrm{r} \times N_\mathrm{t}}$ are all sampled i.i.d. $\mathcal{CN}(0,1)$. The DMT tradeoff is given by $d_L^{||}(r) = L \cdot d_1\left(\frac{r}{L}\right)$ for any $r \in (0, L\min(N_\mathrm{r}, N_\mathrm{t}))$.

□

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 (cf. 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 (7).

□

Comparing the results of Theorem 2 with that of Corollary 1, we observe 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, one of our achievability schemes 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. However, the converse does not follow from earlier results since the length-T playback deadlines of successive messages are only partially overlapping. We present a new approach that addresses the overlapping character of the playback deadlines. The technique is specific to the streaming setup and appears novel.

Remark 1: In our system model, we assumed that the coherence block-length M can be arbitrarily large. It is well known [7] that for quasi-static channel, as well as its extension to L parallel channel [9], the DMT in Theorem 1 holds for any coherence block-length $M \geq N_\mathrm{r} + N_\mathrm{t} - 1$. In a similar fashion our result in Theorem 2 holds for any $M \geq N_\mathrm{r} + N_\mathrm{t} - 1$. In particular the converse in Section IV holds for any M. The interleaved coding scheme in Section V-A reduces the setup to parallel channels and applies to any $M \geq N_\mathrm{r} + N_\mathrm{t} - 1$. However this scheme requires the knowledge of T at both the encoder and decoder. Our second coding scheme, which is based on a tree code and only requires the knowledge of T at the decoder, does require M to be sufficiently large. In particular our analysis for this scheme requires that M must increase as $\log \mathrm{SNR}$ to achieve the optimal DMT.

IV. Converse

In this section we establish a lower bound on the error probability for any streaming code in Definition 1. We thereby upper bound the achievable DMT. In particular we show that

$$\Pr[\text{error}] \dot{\geq} \rho^{-Td_1(r)}$$

where $d_1(r)$ is the DMT tradeoff associated with a single-link MIMO channel. For the purpose of establishing a contradiction, we will assume that a DMT *better* than $d_T(r)$ is achievable, say $T \cdot d_1(r-\delta)$. We show that for any $\delta > 0$ a contradiction will build up if we operate the system over a sufficiently large number of blocks N. The smaller δ is the longer it takes the contradiction to build up.

The steps in our proof are the following, illustrated in Fig. 2.

1) FANO: Apply Fano's inequality to each w_k individually. A decision on w_k must be made at time $T_k = k+T-1$.
2) GENIE: Condition the decoding of w_k on all previous messages w_0^{k-1}. This can only help the decoder (thereby increasing the DMT) because the decoder knows exactly the value of all earlier messages. This step can be thought of as a genie-helper.
3) SUFFIX OUTAGE: Next we condition on the event that the *suffix* of the codeword is in outage. By suffix we mean the symbols transmitted in blocks $k, k+1, \ldots, T_k = T+k-1$. We bound this event using the standard DMT analysis.
4) COMBINE EVENTS: Finally, using standard information manipulations, we combine events up to message w_{N-T+1} for some large N (to be determined).
5) CONTRADICTION: Finally, using the statistical description of the channel law, we find that for any $\delta > 0$ we can identify a finite N sufficiently large such that a contradiction arises and this demonstrates that a DMT of $T \cdot d_1(r-\delta)$ is not achievable.

Following the approach outlined above, in our first step we apply Fano's inequality [22, Chapter 2] to lower bound the error probability associated with message w_k. To do this, we define $\mathcal{E}_k$ to be the event that $\hat{\mathsf{w}}_k \neq \mathsf{w}_k$ and note that $\Pr[\text{error}] = \sup_{k \geq 0} \Pr[\mathcal{E}_k]$. We start by indexing the error events pointwise in the possible channel realizations. Strictly speaking, the summation over the channel gains must be an integral, since the channel gains are continuous valued. We however use summations, so that the expressions are easier to follow. All the steps in this section easily follow when the summations are replaced by corresponding multi-integrals.

$$\Pr[\mathcal{E}_k] = \sum_{\mathbf{H}_0^{T_k}} \Pr[\mathsf{H}_0^{T_k} = \mathbf{H}_0^{T_k}] \Pr[\mathcal{E}_k | \mathsf{H}_0^{T_k} = \mathbf{H}_0^{T_k}]. \quad (9)$$

Message w_k needs to be decoded at time $T_k = k+T-1$. The observations accumulated to that time are the channel outputs $\mathsf{Y}_0^{T_k}$. Recognizing that the log-cardinality of the message set from which w_k is chosen is $Mr\log\rho$, we apply Fano's inequality to each channel realization to get

$$\begin{aligned}
\Pr[\mathcal{E}_k | \mathsf{H}_0^{T_k} = \mathbf{H}_0^{T_k}] &\geq \frac{-1 + H(\mathsf{w}_k | \mathsf{Y}_0^{T_k}, \mathsf{H}_0^{T_k} = \mathbf{H}_0^{T_k})}{Mr\log\rho} \\
&= \frac{-1 + H(\mathsf{w}_k) - H(\mathsf{w}_k) + H(\mathsf{w}_k | \mathsf{Y}_0^{T_k}, \mathsf{H}_0^{T_k} = \mathbf{H}_0^{T_k})}{Mr\log\rho} \\
&\geq 1 - \frac{1}{Mr\log\rho} - \frac{H(\mathsf{w}_k | \mathsf{w}_0^{k-1}, \mathsf{H}_0^{T_k} = \mathbf{H}_0^{T_k})}{Mr\log\rho} + \\
&\qquad \frac{H(\mathsf{w}_k | \mathsf{Y}_0^{T_k}, \mathsf{H}_0^{T_k} = \mathbf{H}_0^{T_k})}{Mr\log\rho},
\end{aligned}$$

where the latter inequality follows since $H(\mathsf{w}_k) = H(\mathsf{w}_k | \mathsf{w}_0^{k-1}, \mathsf{H}_0^{T_k} = \mathbf{H}_0^{T_k})$.

The second step in our proof is the genie-aided step. We condition the last term in the above on all previous messages yielding the further lower bound:

$$\begin{aligned}
&\Pr[\mathcal{E}_k | \mathsf{H}_0^{T_k} = \mathbf{H}_0^{T_k}] & (10)\\
&\geq 1 - \frac{1}{Mr\log\rho} - \frac{H(\mathsf{w}_k | \mathsf{w}_0^{k-1}, \mathsf{H}_0^{T_k} = \mathbf{H}_0^{T_k})}{Mr\log\rho} \\
&\qquad + \frac{H(\mathsf{w}_k | \mathsf{w}_0^{k-1}, \mathsf{Y}_0^{T_k}, \mathsf{H}_0^{T_k} = \mathbf{H}_0^{T_k})}{Mr\log\rho} \\
&= 1 - \frac{1}{Mr\log\rho} - \frac{I(\mathsf{w}_k; \mathsf{Y}_0^{T_k} | \mathsf{w}_0^{k-1}, \mathsf{H}_0^{T_k} = \mathbf{H}_0^{T_k})}{Mr\log\rho} & (11)\\
&= 1 - \frac{1}{Mr\log\rho} - \frac{I(\mathsf{w}_k; \mathsf{Y}_k^{T_k} | \mathsf{w}_0^{k-1}, \mathsf{H}_k^{T_k} = \mathbf{H}_k^{T_k})}{Mr\log\rho} & (12)
\end{aligned}$$

To get the last equality we note the following conditions. First, since message and channel realizations are independent $H(\mathsf{w}_k | \mathsf{w}_0^{k-1}, \mathsf{H}_0^{T_k} = \mathbf{H}_0^{T_k}) = H(\mathsf{w}_k | \mathsf{w}_0^{k-1}, \mathsf{H}_k^{T_k} = \mathbf{H}_k^{T_k})$. Second, we note the following Markov relationship: $\mathsf{w}_k \leftrightarrow \mathsf{w}_0^{k-1}, \mathsf{Y}_k^{T_k}, \mathsf{H}_k^{T_k} \leftrightarrow \mathsf{Y}_0^{k-1}, \mathsf{H}_0^{k-1}$. This relation holds due to the causal nature of the encoder and the i.i.d. nature of the channel. In particular, note that causal encoding means that the channel inputs $\mathsf{X}_k^{T_k}$ are a function of w_k and w_0^{k-1} while past channel inputs are a function only of w_0^{k-1}. Thus, since the channel is memoryless the past channel output and state information $(\mathsf{Y}_0^{k-1}, \mathsf{H}_0^{k-1})$ provides no information about w_k that $(\mathsf{w}_0^{k-1}, \mathsf{Y}_k^{T_k}, \mathsf{H}_k^{T_k})$ does not provide.

In the third step we condition on the suffix being in outage. In particular, define the single-block outage set

$$\mathcal{H}_\delta = \left\{ \mathbf{H} : I(\mathsf{x}; \mathsf{y} | \mathsf{H} = \mathbf{H}) \leq (r-\delta)\log\rho \right\} \quad (13)$$

By the classic outage analysis [23], [24], which underlies the DMT of Theorem 1, we know that

$$P_\delta = \Pr[\mathsf{H} \in \mathcal{H}_\delta] \doteq \rho^{-d_1(r-\delta)} \quad (14)$$

where $d_1(\cdot)$ is the DMT specified in Theorem 1 and the exponential equality is at high SNR. By "suffix outage" we mean that H_j is in outage for *every* block $j = k, \ldots, T_k$, in other words $\cap_{j=k}^{T_k}(\mathsf{H}_j \in \mathcal{H}_\delta)$. Using $\mathcal{H}_\delta^T$ to denote the $T-$fold Cartesian product of the set $\mathcal{H}_\delta$, and recalling that the channel gains are sampled in an i.i.d. fashion across blocks, we have

$$\Pr\left[\cap_{j=k}^{T_k}(\mathsf{H}_j \in \mathcal{H}_\delta)\right] = \Pr\left[\mathsf{H}_k^{T_k} \in \mathcal{H}_\delta^T\right] = (P_\delta)^T \doteq \rho^{-Td_1(r-\delta)}. \quad (15)$$

We next incorporate the effect of outage into our lower bound. In Appendix A we show that

$$\begin{aligned}
&\Pr[\mathcal{E}_k] \geq \Pr[\mathsf{H}_k^{T_k} \in \mathcal{H}_\delta^T] \times \\
&\left(1 - \frac{1}{Mr\log\rho} - \frac{I(\mathsf{w}_k; \mathsf{Y}_k^{T_k} | \mathsf{w}_0^{k-1}, \mathsf{H}_k^{T_k}, \mathsf{H}_k^{T_k} \in \mathcal{H}_\delta^T)}{Mr\log\rho}\right). \quad (16)
\end{aligned}$$

where the expression $\mathsf{H}_k^{T_k} \in \mathcal{H}_\delta^T$ in the conditioning indicates that the sequence $\mathsf{H}_k^{T_k}$ belongs to the outage set $\mathcal{H}_\delta^T$.

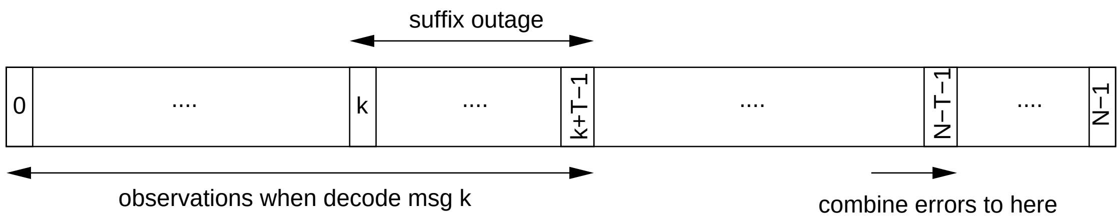

Fig. 2. In the converse, we consider a total of N coherence blocks and $N-T$ messages. For each message k, we consider the event that the coherence blocks $k,\ldots,k+T-1$ are in outage and lower bound the error probability in (16). We then combine the error probabilities associated with all the messages to obtain a lower bound on the maximum error.

Since all the terms in the mutual information expression in (16) are independent of the channel gains: $\{\mathbf{H}_0^{k-1}, \mathbf{H}_{T_k+1}^{N-1}\}$ we can express

$$\begin{aligned} &I(w_k;\mathbf{Y}_k^{T_k}|w_0^{k-1},\mathbf{H}_k^{T_k},\mathbf{H}_k^{T_k}\in\mathcal{H}_\delta^T) \\ &= I(w_k;\mathbf{Y}_k^{T_k}|w_0^{k-1},\mathbf{H}_0^{N-1},\mathbf{H}_0^{N-1}\in\mathcal{H}_\delta^N) \qquad (17) \\ &\leq I(w_k;\mathbf{Y}_0^{N-1}|w_0^{k-1},\mathbf{H}_0^{N-1},\mathbf{H}_0^{N-1}\in\mathcal{H}_\delta^N). \qquad (18) \end{aligned}$$

where the last step follows from the fact that the mutual information is non-negative. We will see that the final loosening in (18) doesn't weaken our bound for two reasons. The first is because we study the max error probability. The second is because the information about the messages embedded in later channel uses, i.e., $\mathbf{Y}_{T_k+1}^{N-1}$, must be used to decrease the entropy of later messages. It cannot be focused exclusively on reducing the uncertainty of w_k without detrimental effects on the ability to estimate later messages. This coupling of errors across time is what we call the *outage amplification* effect.

In the fourth step we combine events. Substituting (15) and (18) into (16) we find

$$\Pr[\mathcal{E}_k] \geq P_\delta^T \times \left[1-\frac{1}{Mr\log\rho}-\frac{I(w_k;\mathbf{Y}_0^{N-1}|w_0^{k-1},\mathbf{H}_0^{N-1},\mathbf{H}_0^{N-1}\in\mathcal{H}_\delta^N)}{Mr\log\rho}\right]. \qquad (19)$$

And, since the max error is at least as large as the average error, we can arrive at

$$\max_{0\leq k\leq N-T-1}\Pr[\mathcal{E}_k] \geq P_\delta^T \times \left[1-\frac{1}{Mr\log\rho}-\frac{I(\mathbf{X}_0^{N-1};\mathbf{Y}_0^{N-1}|\mathbf{H}_0^{N-1},\mathbf{H}_0^{N-1}\in\mathcal{H}_\delta^N)}{(N-T)Mr\log\rho}\right] \qquad (20)$$

as shown in the steps between (21)–(22). where (20) follows from the data processing inequality since $w_0^{N-T-1}\to\mathbf{X}_0^{N-1}\to\mathbf{Y}_0^{N-1}$ holds regardless of the channel realization.

In the final step we apply the channel statistics to get a contradiction. In particular, since fading across different blocks is independent we can break the mutual information term in (20) into a simple sum

$$\begin{aligned} &\max_{0\leq k\leq N-T-1}\Pr[\mathcal{E}_k] \\ &\geq P_\delta^T\left[1-\frac{1}{Mr\log\rho}-\frac{\sum_{j=0}^{N-1}I(\mathbf{X}_j;\mathbf{Y}_j|\mathbf{H}_j,\mathbf{H}_j\in\mathcal{H}_\delta)}{(N-T)Mr\log\rho}\right] \qquad (23) \\ &\geq P_\delta^T\left[1-\frac{1}{Mr\log\rho}-\frac{NM(r-\delta)\log\rho}{(N-T)Mr\log\rho}\right] \qquad (24) \\ &\doteq \left[1-\frac{1}{Mr\log\rho}-\frac{N(r-\delta)}{(N-T)r}\right]\rho^{-Td_1(r-\delta)} \qquad (25) \end{aligned}$$

where (24) follows from the definition (13) of $\mathcal{H}_\delta$ and we recall that there are M channel uses in each coherence interval.

To see the contradiction we assume high SNR, so that the second term vanishes. Then, for any $\delta>0$, by selecting $N>T\frac{r}{\delta}$ the term $\left[1-\frac{N(r-\delta)}{(N-T)r}\right]$ is strictly positive. Since $\delta>0$ is arbitrary, it follows that a diversity order greater than $Td_1(r)$ cannot be achieved.

We observe that the N required to realize a contradiction is inversely proportional to δ. This means that if you operate your system to exceed the DMT by a very small amount it will take some time for a contradiction to build up. A coding scheme can be designed so that early message can borrow channel resources from later message to ensure their reliability. But, eventually, the borrowing builds up and later generations cannot meet their obligations. The parameter N indexes the generation that runs into difficulty.

V. Coding Theorem

We present two approaches for achieving the DMT stated in Theorem 2. As mentioned in the introduction, the first approach is based on interleaving the last T messages across coherence block while the second approach is based on a delay-universal tree code construction.

A. Interleaving Scheme

We show that a simple interleaving based scheme suffices to achieve the DMT stated in Theorem 2. Our codebook $\mathcal{C}$ maps each message $w_k\in\mathcal{I}_M\triangleq\{1,2,\ldots,2^{Mr\log\rho}\}$ to T codewords $\{\mathbf{X}_0(w_k),\mathbf{X}_1(w_k),\ldots,\mathbf{X}_{T-1}(w_k)\}$ where each $\mathbf{X}_j\in\mathbb{C}^{N_t\times\frac{M}{T}}$. Thus the overall code is a Cartesian product: $\mathcal{C}=\mathcal{C}_0\times\mathcal{C}_1\ldots\times\mathcal{C}_{T-1}$, where $\mathbf{X}_k\in\mathcal{C}_k$. We will assume that

$$\max_{0\le k\le N-T-1}\Pr[\mathcal{E}_k] \ge \frac{1}{N-T}\sum_{k=0}^{N-T-1}\Pr[\mathcal{E}_k] \tag{21}$$

$$\ge P_\delta^T\left[1-\frac{1}{Mr\log\rho}-\frac{\sum_{k=0}^{N-T-1} I(w_k;\mathbf{Y}_0^{N-1}|w_0^{k-1},\mathbf{H}_0^{N-1},\mathbf{H}_0^{N-1}\in\mathcal{H}_\delta^N)}{(N-T)Mr\log\rho}\right]$$

$$= P_\delta^T\left[1-\frac{1}{Mr\log\rho}-\frac{I(w_0^{N-T-1};\mathbf{Y}_0^{N-1}|\mathbf{H}_0^{N-1},\mathbf{H}_0^{N-1}\in\mathcal{H}_\delta^N)}{(N-T)Mr\log\rho}\right]$$

$$\ge P_\delta^T\left[1-\frac{1}{Mr\log\rho}-\frac{I(\mathbf{X}_0^{N-1};\mathbf{Y}_0^{N-1}|\mathbf{H}_0^{N-1},\mathbf{H}_0^{N-1}\in\mathcal{H}_\delta^N)}{(N-T)Mr\log\rho}\right] \tag{22}$$

each codebook $\mathcal{C}_j$ is sampled i.i.d. according to a complex normal $\mathcal{CN}(0,\frac{\rho}{N_t})$ distribution[3]

For transmission of each message, 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. 3. Suppose that $\mathcal{I}_{k,0},\ldots,\mathcal{I}_{k,T-1}$ denote these intervals. The codeword $\mathbf{X}_0(w_k)$ is transmitted in the first sub-block $\mathcal{I}_{k,0}$ of coherence block k. The codeword $\mathbf{X}_1(w_k)$ is transmitted in the sub-block $\mathcal{I}_{k+1,1}$ of coherence block $k+1$ and likewise $\mathbf{X}_j(w_k)$ is transmitted in the j-th sub-block, $\mathcal{I}_{k,j}$, of coherence block $k+j$. The corresponding output sequences associated with message w_k are denoted by:

$$\mathbf{Y}_{k,j} = \mathbf{H}_{k+j}\mathbf{X}_j(w_k)+\mathbf{Z}_{k,j},\quad j=0,\ldots,T-1. \tag{26}$$

The decoder finds the message $\hat{w}_k$ such that for each $j\in\{0,\ldots,T-1\}$ the sequence pair $(\mathbf{X}_j(\hat{w}_k),\mathbf{Y}_{k,j})$ is jointly weakly typical [22]. The shaded boxes in Fig. 3 denote the intervals used for the decoding of w_k. The outage event at the decoder, associated with $\hat{w}_k$, is given by:

$$\left\{\frac{1}{T}\sum_{j=k}^{k+T-1} C_j(\rho)\le r\log\rho\right\} \tag{27}$$

where $C_j(\rho)=\log\det\left(I+\frac{\rho}{N_t}\mathbf{H}_j\mathbf{H}_j^\dagger\right)$. Since (27) 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 achievability of the DMT in Theorem 2 follows from Corollary 1.

B. Sequential Tree Codes

Our second scheme is based on a sequential tree code construction. This approach has the advantage that the encoder does not require the knowledge of T. The delay constraint only needs to be revealed to the decoder, and yet the optimal DMT is attained.

Our proposed construction consists of a sequence of codebooks $\{\mathcal{C}_0,\mathcal{C}_1,\ldots,\mathcal{C}_k,\ldots\}$, where $\mathcal{C}_k$ is the codebook to be used in coherence block k when messages $(w_0,\ldots,w_k)$ are revealed to the encoder. Codebook $\mathcal{C}_k$ consists of a total of $2^{MR(k+1)}$ codeword sequences, with one codeword for each element in the set:

$$\mathcal{I}_M^{k+1}=\{\ (w_0,\ldots,w_k): w_0\in\mathcal{I}_M,\ldots,w_k\in\mathcal{I}_M\}. \tag{28}$$

where $\mathcal{I}_M\triangleq\{1,2,\ldots,2^{MR}\}$. All codewords are sampled i.i.d. from $\mathcal{CN}\left(0,\frac{\rho}{N_t}\right)$ and are revealed to both the encoder and the decoder in advance[4]. In coherence block k, the encoder maps $w_0,\ldots,w_k$ to the codeword $\mathbf{X}_k(w_0^k)\in\mathbb{C}^{N_t\times M}$ in $\mathcal{C}_k$, and transmits it over M channel uses. The entire transmitted sequence up to and including block k is denoted by

$$\mathbf{X}_0^k(w_0^k)\triangleq\left\{\mathbf{X}_0(w_0),\mathbf{X}_1(w_0^1),\ldots,\mathbf{X}_k(w_0^k)\right\},\qquad \mathbf{X}_0^k(w_0^k)\in\mathbb{C}^{N_t\times(k+1)M} \tag{29}$$

The decoder uses a sequential, decision-directed decoding rule. We focus on the decoding of message w_k at the end of coherence block $T_k=k+T-1$, which corresponds to the deadline of message w_k. The decoder considers the entire received sequence $\mathbf{Y}_0^{T_k}=(\mathbf{Y}_0,\ldots,\mathbf{Y}_{T_k})$ and computes a fresh estimate of all the messages up to time k in $k+1$ steps as follows. In the first step, the decoder searches over all message sequences $\hat{w}_0^{T_k}$ such that the pair $(\mathbf{X}_0^{T_k}(\hat{w}_0^{T_k}),\mathbf{Y}_0^{T_k})$ is jointly typical. If each such message sequence has a unique prefix, say $\bar{w}_0$, then $\bar{w}_0$ is selected as the message in block 0. Otherwise an error is declared. Once the message $\bar{w}_0$ is fixed in the first step, the decoder then proceeds to the second step. It searches for the message sequences $\hat{w}_1^{T_k}$ such that the pair $(\mathbf{X}_1^{T_k}(\bar{w}_0,\hat{w}_1^{T_k}),\mathbf{Y}_1^{T_k})$ is jointly typical. If each such message sequence has a unique prefix, say $\bar{w}_1$ then it is selected as the message in block 1. Otherwise an error is declared. Once the message $\bar{w}_1$ is fixed the decoder proceeds sequentially, producing $\bar{w}_2,\ldots,\bar{w}_k$. In determining $\bar{w}_l$, with $l\le k$, the decoder fixes $\bar{w}_0^{l-1}$ and searches for a sequence of messages $\hat{w}_l^{T_k}$ such that the corresponding transmit sequence $\mathbf{X}_0^{T_k}(\bar{w}_0^{l-1},\hat{w}_l^{T_k})$ has the property that the sub-sequence between l to T_k (the suffix) satisfies

$$(\mathbf{X}_l^{T_k}(\bar{w}_0^{l-1},\hat{w}_l^{T_k}),\mathbf{Y}_l^{T_k})\in\mathcal{T}_{l,T_k}, \tag{30}$$

where the set $\mathcal{T}_{l,l'}$ is the set of all jointly typical se-

[3]We note however that any space-time code that achieves the DMT for independent parallel MIMO fading channels can be used for the sub-codebooks $\mathcal{C}_0,\ldots,\mathcal{C}_{T-1}$. In particular the non-vanishing determinant (NVD) code in [8] can be used for these sub-codebooks instead of the random Gaussian codebook.

[4]We will make the assumption that the communication terminates after a sufficiently large but fixed number of coherence blocks.

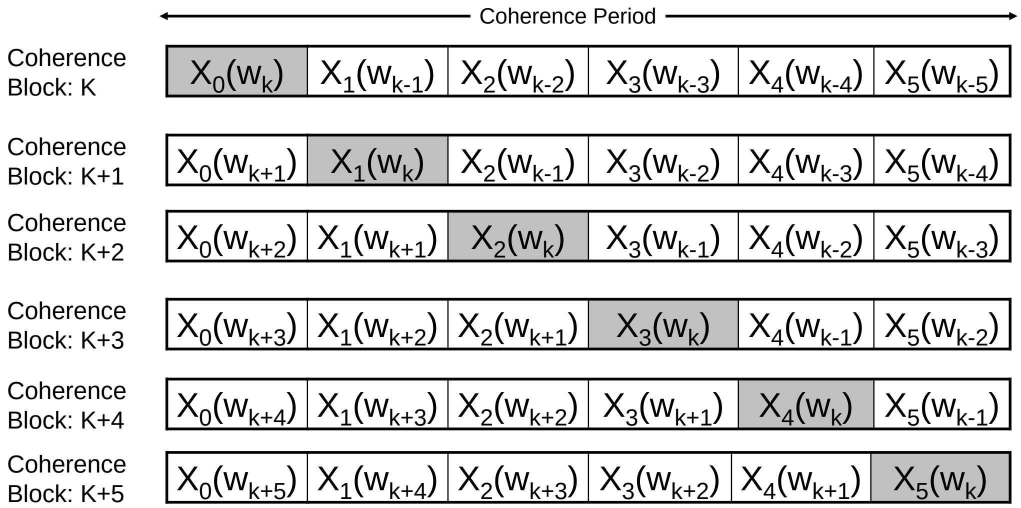

Fig. 3. Interleaving based coding scheme for $T = 6$. Each coherence block is divided into T sub-intervals and each sub-interval is dedicated to transmission of one message. The transmission of message w_k spans coherence blocks $k, k+1, \ldots, k+T-1$ using codewords of $\mathbf{X}_0(w_k), \ldots, \mathbf{X}_{T-1}(w_k)$ as shown by the shaded blocks.

quences [22],

$$\mathcal{T}_{l,l'} = \Bigg\{ (\mathbf{X}_l^{l'}, \mathbf{Y}_l^{l'}) : \mathbf{X}_l^{l'} \in \mathcal{T}(p_{\mathbf{X}_l^{l'}}), \mathbf{Y}_l^{l'} \in \mathcal{T}(p_{\mathbf{Y}_l^{l'}}), \Bigg| \frac{\sum_{k=l}^{l'} [-\log p_{\mathbf{X}_k,\mathbf{Y}_k}(\mathbf{X}_k, \mathbf{Y}_k) - h(p_{\mathbf{X}_k,\mathbf{Y}_k})]}{M(l'-l+1)} \Bigg| \le \varepsilon \Bigg\}. \quad (31)$$

In (31) $\mathcal{T}(p_{\mathbf{X}_l^{l'}})$ and $\mathcal{T}(p_{\mathbf{Y}_l^{l'}})$ denotes the set of typical $\{\mathbf{X}_l^{l'}\}$ and $\{\mathbf{Y}_l^{l'}\}$ sequences respectively and $h(p_{\mathbf{X}_k,\mathbf{Y}_k})$ denotes the differential entropy of jointly Gaussian random variables.

If the list of all message sequences $\hat{w}_l^{T_k}$ that satisfy (30) have a unique prefix $\bar{w}_l$ then we concatenate $\bar{w}_l$ with $\bar{w}_0^{l-1}$ to get $\bar{w}_0^l$, otherwise an error is declared. When the process continues to step $k+1$ without declaring an error, $\bar{w}_k$ is declared to be the output message estimate, i.e., $\hat{w}_k = \bar{w}_k$.

Fig. 4 illustrates the codebook construction and the proposed sequential decoding rule. The figure on the left hand side illustrates the sequential tree code. The right figures illustrate the sequential decoding of w_0, w_1 and w_2. When decoding w_0, we consider all possible paths in the tree typical with the received sequence. If all such paths lead to a unique prefix $\bar{w}_0$, we declare this to be the message. Otherwise an error is declared. Once $\bar{w}_0$ is fixed, we move along the path of $\bar{w}_0$ in the tree. Thereafter we search for all paths in the tree from level 1 to $k+T-1$ that are typical with the received sequence. This process continues until level k is reached and $\bar{w}_k$ is determined.

Remark 2: Our decoder is a decision directed decoder. In estimating $\bar{w}_0^k$, it first estimates $\bar{w}_0$ based on $\mathbf{Y}_0^{T_k}$. It next makes a conditional estimate of $\bar{w}_1$ based on $\mathbf{Y}_1^{T_k}$ with $\bar{w}_0$ 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 $\mathbf{X}_0^{T_k}$ is jointly typical with the received sequence $\mathbf{Y}_0^{T_k}$ i.e.,

$$\left\{ \left| \frac{\sum_{k=0}^{T_k} [-\log p_{\mathbf{X}_k,\mathbf{Y}_k}(\mathbf{X}_k, \mathbf{Y}_k) - h(p_{\mathbf{X}_k,\mathbf{Y}_k})]}{M(k+T)} \right| \le \varepsilon \right\}. \quad (32)$$

Such an approach will not guarantee the recovery of the true w_k with high probability. 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 $\hat{w}_0^{k-1} = w_0^{k-1}$, the pair $(\hat{\mathbf{X}}^{T_k}, \mathbf{Y}^{T_k})$ will in general satisfy (32) as for $k \gg 1$ the contribution of the suffix associated with $\hat{w}_k$ will be negligible. In other words, our proposed decision directed decoder in (30) fixes the older messages and guarantees that when decoding w_k we do not include the bias introduced by w_0^{k-1} in (32).

1) Analysis of error probability: We show that for any $\delta > 0$ and $0 < r < \min(N_\mathrm{r}, N_\mathrm{t})$, the error probability averaged over the ensemble of codebooks satisfies $\Pr(\hat{w}_k \neq w_k) \dot{\le} \rho^{-T \cdot d(r)}$. In our analysis, we exploit symmetry in the code construction, as well as the encoding and decoding functions. To lay out the analysis assume, without loss of generality, that a particular message sequence $w_0^k = w_0^k$ has been transmitted. Define the events[5]

$$\mathcal{E}_l = \Big\{ w_0^l : (w_0, \ldots, w_{l-1}) = (w_0, \ldots, w_{l-1}),\ \bar{w}_l \neq w_l \Big\}, \qquad 0 \le l \le k \quad (33)$$

and note that

$$\Pr\{\bar{w}_k \neq w_k\} \le \sum_{l=0}^{k} \Pr(\mathcal{E}_l), \quad (34)$$

where $\mathcal{E}_l$ corresponds to the event that our proposed decoder fails in step l of the decoding process. We develop an upper

[5]All the error events $\mathcal{E}_l$ are defined for the decoder at time $T+k-1$. However we suppress this dependence to keep the notation compact.

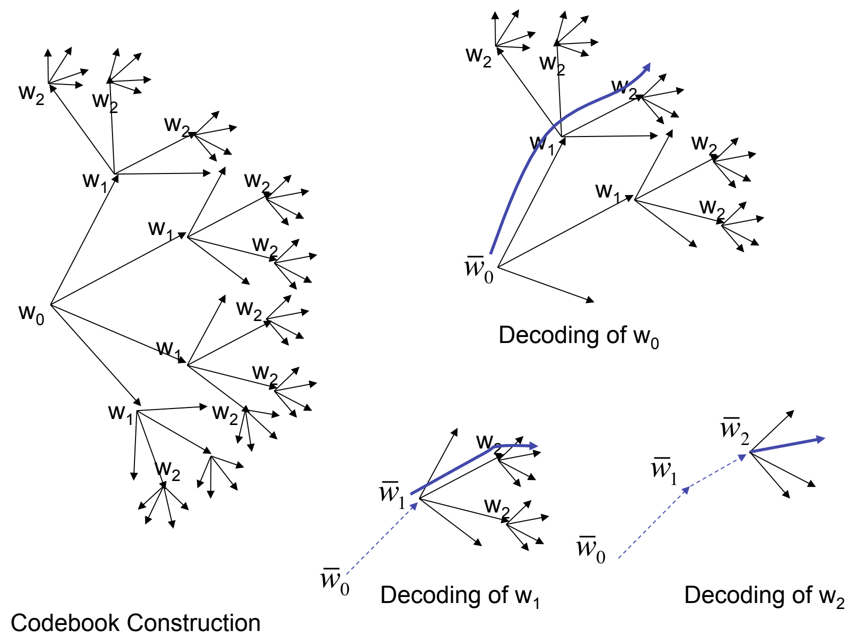

Fig. 4. The left figure illustrates the tree-codebook. The message w_0 is mapped to one of 2^{MR} codewords in the first level, the message pair (w_0, w_1) is mapped to one of 2^{2MR} codewords in the second level etc., While decoding w_k the decoder starts at the root of the tree. It first finds all possible transmit paths of depth $k+T-1$ in the tree, typical with the received sequence. If a unique prefix codeword $\bar{w}_0$ is determined then the decoder moves along the path of $\bar{w}_0$ and finds all possible codewords from level 1 to $k+T-1$ that are typical with the received codeword. A unique message $\bar{w}_1$ is determined if there is a unique prefix codeword in this level. This process continues till level k is reached and $\bar{w}_k$ is determined.

bound on $\mathcal{E}_l$ for each $0 \le l \le k$ and substitute these bounds in (34). We further express $\mathcal{E}_l = \mathcal{A}_l \cup \mathcal{B}_l$, where

$$\mathcal{A}_l = \left\{ (\mathbf{X}_l^{T_k}(w_0^{T_k}), \mathbf{Y}_l^{T_k}) : (\mathbf{X}_l^{T_k}(w_0^{T_k}), \mathbf{Y}_l^{T_k}) \notin \mathcal{T}_{l,T_k} \right\} \tag{35}$$

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

$$\mathcal{B}_l = \left\{ \bar{w}_0^{T_k} : \bar{w}_0^{l-1} = w_0^{l-1}, \bar{w}_l \neq w_l, (\mathbf{X}_l^{T_k}(\bar{w}_0^{T_k}), \mathbf{Y}_l^{T_k}) \in \mathcal{T}_{l,T_k} \right\} \tag{36}$$

denotes the event that the decoding failure happens because a transmit sequence corresponding to a message sequence with $\bar{w}_l \neq w_l$ appears typical with the received sequence.

As shown in the Appendix B, using an appropriate Chernoff bound we can express,

$$\Pr(\mathcal{A}_l) \le 2^{-M(T_k-l+1)f(\varepsilon)} = 2^{-M(T+k-l)f(\varepsilon)} \tag{37}$$

where $f(\varepsilon)$ is a function that satisfies $f(\varepsilon) > 0$ for each $\varepsilon > 0$.

To bound $\Pr(\mathcal{B}_l)$ we begin by noting that by our code construction, we are guaranteed that whenever $\bar{w}_l \neq w_l$, the associated transmitted subsequence $\mathbf{X}_l^{T_k}(\bar{w}_0^{T_k})$ is sampled independently of $\mathbf{Y}_l^{T_k}$. Hence from the joint typicality analysis [22], we have that for any sequence $\bar{w}_0^{T_k}$ with $\bar{w}_l \neq w_l$

$$\begin{aligned} &\Pr\left((\mathbf{X}_l^{T_k}(\bar{w}_0^{T_k}), \mathbf{Y}_l^{T_k}) \in \mathcal{T}_{l,T_k} \mid \mathbf{H}_l^{T_k} = \mathbf{H}_l^{T_k} \right) \\ &\le 2^{-M\left(\sum_{j=l}^{T_k} I(\mathbf{x}_j;\mathbf{y}_j|\mathbf{H}_j=\mathbf{H}_j) - 3\varepsilon\right)} \\ &= 2^{-M\left(\sum_{j=l}^{T_k} C_j(\rho;\mathbf{H}_j) - 3\varepsilon\right)} \end{aligned}$$

where

$$C_j(\rho; \mathbf{H}_j) \triangleq \log\det\left(\mathbf{I} + \frac{\rho}{N_\mathrm{t}} \mathbf{H}_j \mathbf{H}_j^\dagger\right) \tag{38}$$

is the associated mutual information between the input and output in the j-th coherence block when the channel matrix $\mathbf{H}_j = \mathbf{H}_j$. Applying the union bound we have that

$$\Pr(\mathcal{B}_l \mid \mathbf{H}_l^{T_k} = \mathbf{H}_l^{T_k}) \tag{39}$$

$$\le \sum_{\bar{w}_l^{T_k} \in \mathcal{I}_M^{T_k-l+1}} \Pr\left((\mathbf{X}_l^{T_k}(\bar{w}_0^{T_k}), \mathbf{Y}_l^{T_k}) \in \mathcal{T}_{l,T_k} \mid \mathbf{H}_l^{T_k} = \mathbf{H}_l^{T_k} \right) \tag{40}$$

$$\le \sum_{\bar{w}_l^{T_k} \in \mathcal{I}_M^{T_k-l+1}} 2^{-M\left(\sum_{j=l}^{T_k} C_j(\rho;\mathbf{H}_j) - 3\varepsilon\right)} \tag{41}$$

$$\le \left(2^{M(T_k-l+1)R}\right) 2^{-M\left(\sum_{j=l}^{T_k} C_j(\rho;\mathbf{H}_j) - 3\varepsilon\right)} \tag{42}$$

$$\le 2^{-M\left(\sum_{j=l}^{T_k} C_j(\rho;\mathbf{H}_j) - (T_k-l+1)R - 3\varepsilon\right)}. \tag{43}$$

To bound $\Pr(\mathcal{B}_l)$ we define

$$\mathcal{O}_l = \left\{ (\mathbf{H}_l, \ldots, \mathbf{H}_{T_k}) : \sum_{i=l}^{T_k} C_i(\rho; \mathbf{H}_i) \le (k+T-l) r \log\rho + (k-l)\Delta(r)\log\rho + 4\epsilon\log\rho \right\} \tag{44}$$

where

$$\Delta(r) = -\frac{d_1(r)}{2d_1'(r)} \tag{45}$$

where we recall that $d_1(r)$ denotes the quasi-static DMT (7) of the MIMO fading channel and we use $d_1'(r)$ to denote its right derivative. Note that $d_1'(r) < 0$ for all $r \in [0, \min(N_\mathrm{t}, N_\mathrm{r})]$ as the DMT is a decreasing function of r. Thus it follows that $\Delta(r) > 0$.

Note that

$$\Pr(\mathcal{B}_l) \leq \Pr(\mathcal{B}_l \mid \mathbf{H}_l^{T_k} \in \mathcal{O}_l^c) + \Pr(\mathbf{H}_l^{T_k} \in \mathcal{O}_l). \tag{46}$$

From (44) and (43) we have

$$\Pr(\mathcal{B}_l \mid \mathbf{H}_l^{T_k} \in \mathcal{O}_l^c) \leq 2^{-M(k-l)\Delta(r)\log\rho - M\varepsilon\log\rho} \tag{47}$$

$$= \rho^{-M\varepsilon - M(k-l)\Delta(r)}. \tag{48}$$

We next upper bound the second term in (46). Note that $\mathcal{O}_l$ precisely corresponds to the parallel MIMO channel in Corollary 1 with $L \triangleq T_k - l + 1 = k + T - l$ channels, and multiplexing gain $s = Lr + (k-l)\Delta(r) + 4\varepsilon$. The associated DMT satisfies:

$$L \cdot d_1\left(\frac{s}{L}\right) = L \cdot d_1\left(r + \frac{k-l}{L}\Delta(r) + \frac{4\varepsilon}{L}\right) \tag{49}$$

$$= L \cdot d_1\left(r + \frac{k-l}{L}\Delta(r)\right) + o_\varepsilon(1) \tag{50}$$

$$\geq L\left(d_1(r) + d_1'(r)\frac{k-l}{L}\Delta(r)\right) + o_\varepsilon(1) \tag{51}$$

$$= Ld_1(r) - \frac{(k-l)}{2}d_1(r) + o_\varepsilon(1) \tag{52}$$

$$= Td_1(r) + (k-l)d_1(r) - \frac{(k-l)}{2}d_1(r) + o_\varepsilon(1) \tag{53}$$

$$= Td_1(r) + \frac{(k-l)}{2}d_1(r) + o_\varepsilon(1) \tag{54}$$

where we use the continuity of $d_1(r)$ in (50) and let $o_\varepsilon(1)$ be a function of ε that vanishes as $\varepsilon \to 0$. We use the convexity of $d_1(r)$ in (51). We substitute (45) for $\Delta(r)$ in (52) and substitute $L = T + k - l$ in (53). Thus we have

$$\Pr(\mathcal{O}_l) \dot{\leq} \rho^{-(Td_1(r) + \frac{(k-l)}{2}d_1(r)) + o_\varepsilon(1)}. \tag{55}$$

From (46) and substituting (48) and (55) and using $\mathcal{E}_l = \mathcal{A}_l \cup \mathcal{B}_l$ we have

$$\Pr(\mathcal{E}_l) \leq \Pr(\mathcal{A}_l) + \Pr(\mathcal{B}_l) \tag{56}$$

$$\dot{\leq} 2^{-M(T+k-l)f(\varepsilon)} + \rho^{-M\varepsilon - M(k-l)\Delta(r)} + \rho^{-Td_1(r) - \frac{(k-l)}{2}d_1(r) + o_\varepsilon(1)} \tag{57}$$

From the union bound,

$$\Pr(\mathcal{E}) \leq \sum_{l=0}^{k} \Pr(\mathcal{E}_l) \tag{58}$$

$$\leq \sum_{l=0}^{k} 2^{-M(T+k-l)f(\varepsilon)} + \sum_{l=0}^{k} \rho^{-M\varepsilon - M(k-l)\Delta(r)} + \sum_{l=0}^{k} \rho^{-Td_1(r) - \frac{(k-l)}{2}d_1(r) + o_\varepsilon(1)}. \tag{59}$$

We upper bound the first term in (59) as

$$\sum_{l=0}^{k} 2^{-M(k+T-l)f(\varepsilon)} = \sum_{l=0}^{k} 2^{-M(l+T)f(\varepsilon)} \tag{60}$$

$$\leq \sum_{l=0}^{\infty} 2^{-M(l+T)f(\varepsilon)} \leq 2^{-MTf(\varepsilon)+1}, \tag{61}$$

which vanishes as $M \to \infty$. By a similar argument we can upper bound the second term as

$$\sum_{l=0}^{k} \rho^{-M\varepsilon - M(k-l)\Delta(r)} = \rho^{-M\varepsilon} \sum_{l=0}^{k} \rho^{-Ml\Delta(r)} \tag{62}$$

$$\leq \rho^{-M\varepsilon} \sum_{l=0}^{\infty} \rho^{-Ml\Delta(r)} \leq 2\rho^{-M\varepsilon} \tag{63}$$

for sufficiently large ρ and M such that $\rho^{-M\Delta(r)} \leq \frac{1}{2}$. In a similar fashion we can upper bound the third term in (59) as

$$\sum_{l=0}^{k} \rho^{-Td_1(r) - \frac{(k-l)}{2}d_1(r) + o_\varepsilon(1)} \leq \rho^{-Td_1(r) + o_\varepsilon(1)} \sum_{l=0}^{k} \rho^{-\frac{l}{2}d_1(r)} \tag{64}$$

$$\leq 2\rho^{-Td_1(r) + o_\varepsilon(1)}. \tag{65}$$

From (59) we have that

$$\Pr(\mathcal{E}) \leq 2(2^{-MTf(\varepsilon)} + \rho^{-Td_1(r) + o_\varepsilon(1)} + \rho^{-M\varepsilon}) \tag{66}$$

By selecting $M \geq \frac{d_1(r)\log\rho}{f(\varepsilon)}$, we have that $2^{-MTf(\varepsilon)} \leq \rho^{-Td_1(r) + o_\varepsilon(1)}$. Finally since $\varepsilon > 0$ can be selected as small as required and $o_\varepsilon(1) \to 0$ as $\varepsilon \to 0$ we have that $\Pr(\mathcal{E}) \dot{\leq} \rho^{-Td_1(r)}$. This completes the error analysis for the sequential tree code.

VI. Multiple Messages per Coherence Block

Our focus so far has been on the case where only one message arrives in each coherence block. In this section, we consider some special cases when two messages, say $\mathsf{w}_{k,1}$ and $\mathsf{w}_{k,2}$ arrive in each coherence block. Each message must be decoded after $M \cdot T$ channel uses, where T denotes the delay in coherence blocks. In such a setup, the number of coherence blocks seen by $\mathsf{w}_{k,1}$ before its deadline, will be different from $\mathsf{w}_{k,2}$. Thus a simple interleaving techniques such as is presented in Section V-A is no longer optimal; more sophisticated coding techniques that exploit the asymmetry between $\mathsf{w}_{k,1}$ and $\mathsf{w}_{k,2}$ will be required.

Assume that $\mathsf{w}_{k,1}$ arrives at time $t_{k,1} = (k-1)M + M \cdot \Delta$, while $\mathsf{w}_{k,2}$ arrives at time $t_{k,2} = (k-1)M + M\left(\Delta + \frac{1}{2}\right)$ where $\Delta \in [0, 1/2]$ denotes the offset relative to the start of the coherence block when $\mathsf{w}_{k,1}$ arrives. Fig. 5 denotes the streaming setup with two messages per block corresponding to $\Delta = 0$ and $\Delta = 1/2$ respectively. We obtain the optimal DMT for the SISO channel with $T = 1$ and $\Delta = 0$, which corresponds to the first case in Fig. 5, in Section VI-A. By the symmetry of the problem, the same result also applies when $\Delta = 1/2$, illustrated in the second case in Fig 5. In subsection VI-B we show that if either $\Delta = 0$ or $\Delta = 1/2$, but its actual value is not known to the encoder, the DMT is strictly smaller.

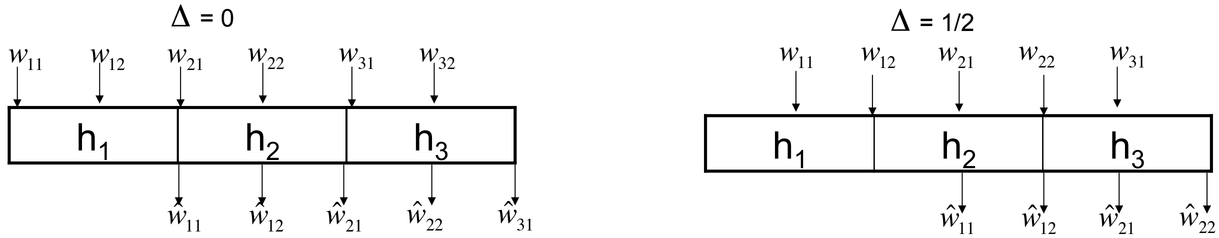

Fig. 5. Streaming setup with two messages arriving in each coherence block. In coherence block k two messages, $w_{k,1}$ and $w_{k,2}$, arrive as shown in the figure. We assume that $w_{k,1}$ arrives $\Delta \cdot M$ symbols after the start of the coherence block k, while $w_{k,2}$ arrives $\left(\Delta + \frac{1}{2}\right) \cdot M$ symbols from the start of coherence block k. We assume a decoding delay of one coherence block for each message as shown. The left figure illustrates the case when $\Delta = 0$, while the right figure shows the case when $\Delta = \frac{1}{2}$.

A. DMT when $\Delta = 0$.

We assume that each message $w_{k,i}$ is uniformly distributed in the set $\mathcal{I}_M = \{1, 2, \ldots, 2^{MR/2}\}$, so that a total of MR information bits arrive in each coherence block of length M. We consider a SISO channel model. Let h_k denote the channel gain in coherence block k, and denote the corresponding input sequence as x_k^M. We split $x_k^M = \left(x_{k,1}^{M/2}, x_{k,2}^{M/2}\right) \triangleq (\mathbf{x}_{k,1}, \mathbf{x}_{k,2})$ into two subsequences, each of length $M/2$. The input sequence $\mathbf{x}_{k,1}$ can depend on $(w_{k,1}, w_{k-1,1}, w_{k-1,2}, \ldots)$, while $\mathbf{x}_{k,2}$ can depend on $(w_{k,1}, w_{k,2}, w_{k-1,1}, w_{k-1,2}, \ldots)$. The received sequence y_k^M is also partitioned into $\left(y_{k,1}^{M/2}, y_{k,2}^{M/2}\right) \triangleq (\mathbf{y}_{k,1}, \mathbf{y}_{k,2})$. Recall that for the SISO channel, we have $\mathbf{y}_{k,i} = h_k \cdot \mathbf{x}_{k,i} + \mathbf{z}_{k,i}$ where the additive noise sequence $\mathbf{z}_{k,i}$ is sampled i.i.d. from $\mathcal{CN}(0,1)$. We will assume that each message $w_{k,i}$ must be decoded with a delay of $T = 1$ coherence period. Thus, $w_{k,1}$ must be decoded at the end of coherence block k whereas $w_{k,2}$ must be decoded in the middle of coherence block $k+1$, as is illustrated in Fig. 5. The following result shows how to exploit the asymmetry in channel conditions experienced by $w_{k,1}$ and $w_{k,2}$ to attain a higher DMT than that which can be obtained through simple interleaving.[6]

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

$$d(r) = \min\left(1 - \frac{r}{2}, 2 - 2r\right), \quad r \in [0, 1]. \tag{67}$$

Converse: The upper bound is based on two genie aided arguments. The bound $d(r) = 1 - r/2$ follows by revealing every message $w_{k,2}$ to the destination. Thus message $w_{k,1}$ needs to be decoded at the end of coherence block k. Since $w_{k,1}$ is uniformly distributed in $\{1, 2, \ldots, 2^{MR/2}\}$ and has a rate of $R/2$, it follows that the DMT for this genie aided channel equals $d(r) = 1 - r/2$.

To establish the other upper bound of $d(r) = 2 - 2r$ we consider another genie aided channel. We reveal message $w_{k,2}$ at the start of coherence block k and relax the deadline of $w_{k,1}$ and $w_{k,2}$ such that both only need to be decoded at the end of the coherence block $k+1$. Such an assumption can clearly only improve the DMT. However the setup now is identical to that considered in Theorem 2 where the message $w_k = (w_{k,1}, w_{k,2})$ arrive at the start of coherence block k and must be decoded with a delay of $T = 2$ coherence blocks. The associated DMT, $d(r) = 2 - 2r$ for this channel, is thus an upper bound for the original setup. This completes the justification of the converse.

Achievability: We next present a coding scheme that attains the DMT in (67). We first split each message $w_{k,1}$ into two equal sized messages $w_{k,1} = (w_{k,1}^1, w_{k,1}^2)$, where each sub-message is of rate $R_0 = R/4$. Thus we can assume that both $w_{k,1}^1$ and $w_{k,1}^2$ are independent and sampled uniformly from $\mathcal{J}_M = \{1, 2, \ldots, 2^{MR/4}\}$. We do not split the messages $w_{k,2}$ and assume that it is sampled uniformly from $\mathcal{I}_M = \{1, 2, \ldots, 2^{MR/2}\}$. We sample three Gaussian codebooks as follows:

- The codebook $\mathcal{C}_A$ 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 $\mathbf{x}_A(w_{k,1}^1, w_{k-1,2})$ i.e.,

$$\mathcal{C}_A = \left\{\mathbf{x}_A(w_{k,1}^1, w_{k-1,2})\right\}_{w_{k,1}^1 \in \mathcal{J}_M, w_{k-1,2} \in \mathcal{I}_M} \tag{68}$$

- The codebook $\mathcal{C}_B$ 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 $\mathbf{x}_B(w_{k,1}^2)$ i.e.,

$$\mathcal{C}_B = \left\{\mathbf{x}_B(w_{k,1}^2)\right\}_{w_{k,1}^2 \in \mathcal{J}_M} \tag{69}$$

- The codebook $\mathcal{C}_C$ 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 $\mathbf{x}_C(w_{k,2})$.

$$\mathcal{C}_C = \{\mathbf{x}_C(w_{k,2})\}_{w_{k,2} \in \mathcal{I}_M} \tag{70}$$

  We will select $\beta = r/2$. Note that the total power in the second block is $\rho + \rho^{1-r/2} \doteq \rho$, since $\rho^{1-r/2} \ll \rho$.

In coherence block k, the transmitter transmits $\mathbf{x}_{k,1} = \mathbf{x}_A(w_{k,1}^1, w_{k-1,2})$ in the first half of the coherence block and $\mathbf{x}_{k,2} = \mathbf{x}_B(w_{k,1}^2) + \mathbf{x}_C(w_{k,2})$ in the second half of the coherence block. The receiver observes $\mathbf{y}_{k,i} = h_k \cdot \mathbf{x}_{k,i} + \mathbf{z}_{k,i}$ for $i \in \{1, 2\}$. The decoding of the messages is as follows.

- In the second half of coherence block k, the receiver decodes $w_{k,1}^2$ using $\mathbf{y}_{k,2}$, by treating $\mathbf{x}_C(w_{k,2})$ as additional noise. It searches for a unique message $\hat{w}_{k,1}^2 \in \mathcal{J}_M$ such that $\left(\mathbf{x}_B(\hat{w}_{k,1}^2), \mathbf{y}_{k,2}\right) \in \mathcal{T}_{\varepsilon,2}^{M/2}$. The error event $\mathcal{E}_{k,1}^2$

[6]We will drop the subscript $d_T(\cdot)$ in this section for the DMT since we fix $T = 1$.

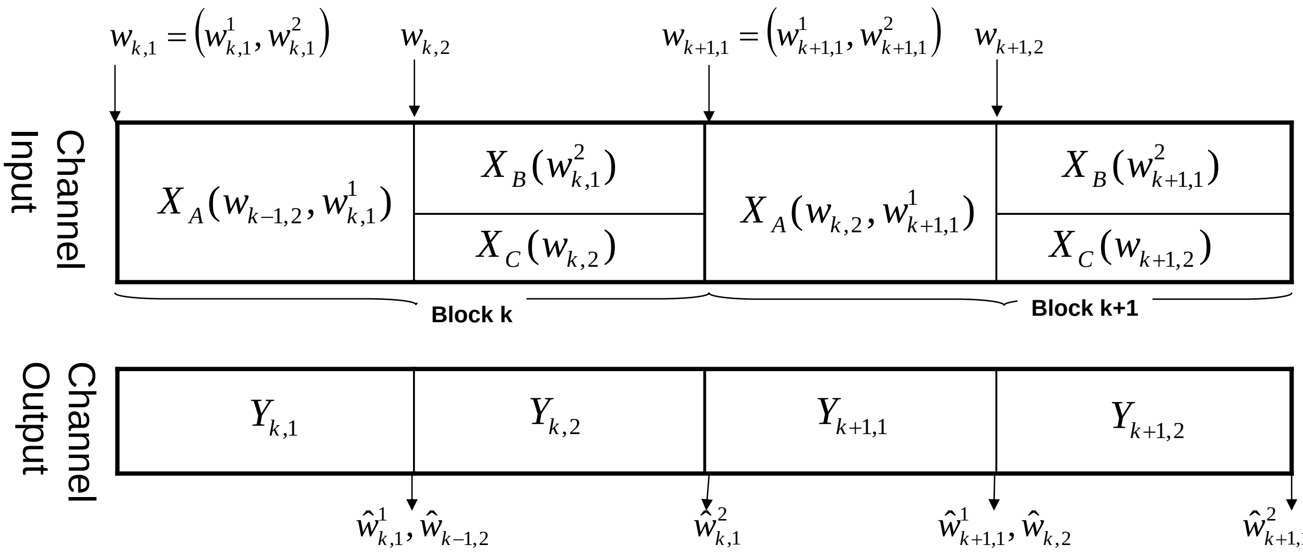

Fig. 6. Coding scheme for two messages per coherence block with $\Delta = 0$. The first message $w_{k,1}$ is split into two sub-messages $(w^1_{k,1}, w^2_{k,1})$ of equal size, while the second message $w_{k,2}$ is not split. In the first half of the coherence block, we transmit the codeword $\mathbf{x}_A(w_{k-1,2}, w^1_{k,1})$ while in the second half of the coherence block we transmit the sum $\mathbf{x}_B(w^2_{k,1}) + \mathbf{x}_C(w_{k,2})$.

denotes the event that $\hat{w}^2_{k,1} \neq w^2_{k,1}$ and let $\mathcal{O}^2_{k,1}$ denote the outage event:

$$\left\{\frac{1}{2}\log\left(1+\frac{|h_k|^2\rho}{1+|h_k|^2\rho^{1-\beta}}\right) \leq \frac{r}{4}\log\rho\right\}. \quad (71)$$

- After decoding $\hat{w}^2_{k,1}$ the decoder subtracts $\mathbf{x}_B(\hat{w}^2_{k,1})$ from $\mathbf{y}_{k,2}$ i.e., $\tilde{\mathbf{y}}_{k,2} = \mathbf{y}_{k,2} - h_k\mathbf{x}_B(\hat{w}^2_{k,1})$. The decoder uses the second half of coherence block k and the first half of coherence block $k+1$ to decode the message pair $(w_{k,2}, w^1_{k+1,1})$. In particular, it searches for a pair $(\hat{w}_{k,2}, \hat{w}^1_{k+1,1})$ such that $(\mathbf{x}_C(\hat{w}_{k,2}), \tilde{\mathbf{y}}_{k,2}) \in \mathcal{T}^{M/2}_{\varepsilon,3}$ and $(\mathbf{x}_A(\hat{w}^1_{k+1,1}, \hat{w}_{k,2}), \mathbf{y}_{k+1,1}) \in \mathcal{T}^{M/2}_{\varepsilon,4}$ are jointly typical. The error analysis involves two events, $\mathcal{E}_{k,2}$ and $\mathcal{E}^1_{k+1,1}$, associated with the error in decoding $w_{k,2}$ and $w^1_{k+1,1}$ respectively. In particular, let $\mathcal{E}_{k,2} = \{\hat{w}_{k,2} \neq w_{k,2}\}$ and the outage event $\mathcal{O}_{k,2}$ be given by:

$$\left\{\frac{1}{2}\log\left(1+\rho^{1-\beta}|h_k|^2\right)+\frac{1}{2}\log(1+\rho|h_{k+1}|^2) \leq \frac{3}{4}r\log\rho\right\}. \quad (72)$$

Similarly let $\mathcal{E}^1_{k+1,1} = \{\{\hat{w}_{k,2} = w_{k,2}\} \cap \{\hat{w}^1_{k+1,1} \neq w^1_{k+1,1}\}\}$ be the event that the message $w_{k,2}$ is decoded correctly, but an error occurs in the decoding of $w^1_{k+1,1}$ and let $\mathcal{O}^1_{k,1}$ be the corresponding outage event:

$$\left\{\frac{1}{2}\log(1+\rho|h_k|^2) \leq \frac{r}{4}\log\rho\right\}. \quad (73)$$

It suffices to show that the error probability satisfies $\Pr\left(\mathcal{E}^2_{k,1} \cup \mathcal{E}_{k,2} \cup \mathcal{E}^1_{k+1,1}\right) \dot{\leq} \rho^{-d(r)} + o_M(1)$ where $d(r)$ is defined in (67) and $o_M(1)$ approaches zero as $M \to \infty$. By selecting M sufficiently large (for each fixed ρ), the proposed DMT is then achievable.

We first consider the event $\mathcal{E}^2_{k,1} = \{\hat{w}^2_{k,1} \neq w^2_{k,1}\}$ and use the following upper bound:

$$\Pr\left(\mathcal{E}^2_{k,1}\right) \leq \Pr\left(\mathcal{E}^2_{k,1} \mid \mathcal{O}^{2,c}_{k,1}\right) + \Pr(\mathcal{O}^2_{k,1}) \quad (74)$$

where $\mathcal{O}^2_{k,1}$ is defined in (71), with $\beta = r/2$ as:

$$\left\{\frac{1}{2}\log\left(1+\frac{|h_k|^2\rho}{1+|h_k|^2\rho^{1-r/2}}\right) \leq \frac{r}{4}\log\rho\right\}. \quad (75)$$

From standard arguments, the first term in (74) decreases to zero as $M \to \infty$, and thus we only need to upper bound the second term. Letting $|h_k|^2 = \rho^{-(1-\alpha)}$ we have that (75) is equivalent to

$$\log\left(1+\frac{\rho^\alpha}{1+\rho^{\alpha-r/2}}\right) \leq \frac{r}{2}\log\rho \quad (76)$$

which in turn implies that $\alpha < \frac{r}{2}$ as $\rho \to \infty$. Thus we have that

$$\Pr(\mathcal{O}^2_{k,1}) \dot{\leq} \rho^{-(1-r/2)} \quad (77)$$

and in turn

$$\Pr(\mathcal{E}^2_{k,1}) \dot{\leq} \rho^{-(1-r/2)} + o_M(1) \quad (78)$$

holds.

Next we upper bound the probability of the event $\mathcal{E}_{k,2} = \{w_{k,2} \neq \hat{w}_{k,2}\}$. We can express

$$\Pr(\mathcal{E}_{k,2}) = \Pr(\mathcal{E}_{k,2}|\mathcal{O}^c_{k,2}) + \Pr(\mathcal{O}_{k,2}) \quad (79)$$

where recall that the event $\mathcal{O}_{k,2}$ in (72) is defined, with $\beta = r/2$ as:

$$\left\{\frac{1}{2}\log\left(1+\rho^{1-r/2}|h_k|^2\right)+\frac{1}{2}\log(1+\rho|h_{k+1}|^2) \leq \frac{3}{4}r\log\rho\right\}. \quad (80)$$

Note that whenever $\{w_{k,2} \neq \hat{w}_{k,2}\}$, we have that $(\mathbf{x}_C(\hat{w}_{k,2}), \tilde{\mathbf{y}}_{k,2})$ are mutually independent and furthermore $(\mathbf{x}_A(\hat{w}^1_{k+1,1}, \hat{w}_{k,2}), \mathbf{y}_{k+1,1})$ are mutually independent. It can be shown through a standard union bound argument that $\Pr(\mathcal{E}_{k,2}|\mathcal{O}^c_{k,2})$ vanishes to zero as $M \to \infty$. To upper bound

$\mathcal{O}_{k,2}$ we let $|h_k|^2 = \rho^{-(1-\alpha_1)}$ and $|h_{k+1}|^2 = \rho^{-(1-\alpha_2)}$ and note that (80) reduces to:

$$\log(1+\rho^{\alpha_1 - r/2}) + \log(1+\rho^{\alpha_2}) \leq \frac{3r}{2}\log\rho. \quad (81)$$

The associated DMT is given by

$$d_2(r) = \min_{(\alpha_1,\alpha_2)\in\mathcal{A}} (1-\alpha_1)^+ + (1-\alpha_2)^+ \quad (82)$$

where $\mathcal{A} = \{(\alpha_1,\alpha_2) \geq 0 : (\alpha_1 - r/2)^+ + \alpha_2 \leq 3r/2\}$ and $(v)^+$ equals 0 is $v < 0$. It can be deduced that $d_2(r) = 2-2r$ and thus

$$\Pr(\mathcal{O}_{k,2}) \dot{\leq} \rho^{-2(1-r)} \quad (83)$$

holds. Thus we have that

$$\Pr(\mathcal{E}_{k,2}) \dot{\leq} o_M(1) + \rho^{-2(1-r)} \quad (84)$$

holds.

Finally we consider the event $\mathcal{E}^1_{k+1,1} = \{\{\hat{w}_{k,2} = w_{k,2}\} \cap \{\hat{w}^1_{k+1,1} \neq w^1_{k+1,1}\}\}$ which corresponds to an error in message estimate $\hat{w}^1_{k+1,1}$ in the first half of the coherence block. Under this event the codeword $\mathbf{x}_C(\hat{w}_{k,2})$ is decoded correctly however the pair $(\mathbf{x}_A(\hat{w}^1_{k+1,1}, \hat{w}_{k,2}), \mathbf{y}_{k+1,1})$ is mutually independent. Using $\mathcal{O}^1_{k,1}$ be defined in (73), we can express

$$\Pr\left(\mathcal{E}^1_{k,1}\right) \leq \Pr(\mathcal{O}^1_{k,1}) + \Pr\left(\mathcal{E}^1_{k,1} \;\middle|\; \mathcal{O}^{1,c}_{k,1}\right). \quad (85)$$

It follows from standard arguments that $\Pr(\mathcal{O}^1_{k,1}) \doteq \rho^{-(1-r/2)}$ and furthermore $\Pr\left(\mathcal{E}^1_{k,1} \;\middle|\; \mathcal{O}^{1,c}_{k,1}\right)$ vanishes to zero as $M \to \infty$. Thus we have that

$$\Pr\left(\mathcal{E}^1_{k,1}\right) \dot{\leq} \rho^{-(1-r/2)} + o_M(1). \quad (86)$$

This completes our achievability.

B. Unknown Offset

We consider the case when either $\Delta = 0$ or $\Delta = 1/2$, but when the actual value of Δ is not known to the transmitter. Such a setup applies when simultaneously transmitting to two users whose coherence blocks are offset by $M/2$ symbols. The following result shows that we cannot have a universal coding scheme oblivious of Δ 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 where the actual value of Δ 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 $\frac{r}{2}\log\rho$ a 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 $\mathbf{x}_{k,1}$ in the first half of the coherence block and transmits $\mathbf{x}_{k,2}$ in the second half, i.e., $\mathbf{x}_k = [\mathbf{x}_{k,1}\ \mathbf{x}_{k,2}]$, where both $\mathbf{x}_{k,1}, \mathbf{x}_{k,2} \in \mathbb{C}^{M/2}$. Receiver 1 observes $\mathbf{y}_k = [\mathbf{y}_{k,1}\ \mathbf{y}_{k,2}]$ in coherence block k as follows:

$$\mathbf{y}_{k,1} = h_k \mathbf{x}_{k,1} + \mathbf{n}_{k,1}, \quad (87)$$

$$\mathbf{y}_{k,2} = h_k \mathbf{x}_{k,2} + \mathbf{n}_{k,2}. \quad (88)$$

Receiver 2 observes $\mathbf{v}_k = [\mathbf{v}_{k,1}\ \mathbf{v}_{k,2}]$ in coherence block k as follows:

$$\mathbf{v}_{k,1} = h_k \mathbf{x}_{k,1} + \mathbf{z}_{k,1}, \quad (89)$$

$$\mathbf{v}_{k,2} = h_{k+1} \mathbf{x}_{k,2} + \mathbf{z}_{k,2}. \quad (90)$$

The noise variables $\mathbf{n}_{k,j}$ and $\mathbf{z}_{k,j}$ have i.i.d. $\mathcal{CN}(0,1)$ entries. For both receivers we have $T = 1$ and message deadlines are shown in Fig. 5. Note that the duration of $w_{k,1}$ spans only one coherence block for receiver 1, but $w_{k,2}$ spans two coherence blocks. Likewise $w_{k,2}$ spans only one coherence block for receiver 2, but $w_{k,1}$ spans two blocks. We show that under this constraint the maximum possible DMT is $d(r) = 1 - r$

We begin by considering Fano's inequality for receiver 1 for message $w_{0,1}$ and rate $\frac{Mr}{2}\log\rho$:

$$\Pr(\mathcal{E}_{0,1}|h_0 = h_0) \geq 1 - \frac{2}{Mr\log\rho} - \frac{2I(w_{0,1};\mathbf{y}_1|h_0 = h_0)}{Mr\log\rho}. \quad (91)$$

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 $P_\delta \doteq \rho^{-(1-r+\delta)}$

$$\Pr(\mathcal{E}_{0,1}) \geq P_\delta\left(1 - \frac{2I(w_{0,1};\mathbf{y}_0|h_0, h_0 \in \mathcal{H}_\delta)}{Mr\log\rho}\right) \quad (92)$$

$$= P_\delta\left(1 - \frac{2I(w_{0,1};\mathbf{y}_0|h_0^{N+1}, h_0^{N+1} \in \mathcal{H}_\delta^{N+2})}{Mr\log\rho}\right) \quad (93)$$

$$\geq P_\delta\left(1 - \frac{2I(w_{0,1};\mathbf{y}_0^N, \mathbf{v}_0^N|h_0^{N+1}, h_0^{N+1} \in \mathcal{H}_\delta^{N+2})}{Mr\log\rho}\right) \quad (94)$$

where (93) follows from the fact that h_1^{N+1} is independent of $(w_0, \mathbf{y}_0, h_0)$.

Similarly, applying Fano's inequality to receiver 2 for message $w_{0,2}$ we have

$$\Pr(\mathcal{E}_{0,2}) \geq P_\delta\left(1 - \frac{2I(w_{0,2};\mathbf{v}_{0,2},\mathbf{v}_{1,1}|w_{0,1}, h_1, h_1 \in \mathcal{H}_\delta)}{Mr\log\rho}\right) \quad (95)$$

$$= P_\delta\left(1 - \frac{2I(w_{0,2};\mathbf{v}_{0,2},\mathbf{v}_{1,1}|w_{0,1}, h_0^{N+1}, h_0^{N+1} \in \mathcal{H}_\delta^{N+2})}{Mr\log\rho}\right) \quad (96)$$

$$\geq P_\delta\left(1 - \frac{2I(w_{0,2};\mathbf{y}_0^N, \mathbf{v}_0^N|w_{0,1}, h_0^{N+1}, h_0^{N+1} \in \mathcal{H}_\delta^{N+2})}{Mr\log\rho}\right). \quad (97)$$

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

$$\Pr(\mathcal{E}_{k,1}) \geq P_\delta\left(1 - \frac{2I(w_{k,1}; \mathbf{y}_0^N, \mathbf{v}_0^N | h_0^{N+1}, h_0^{N+1} \in \mathcal{H}_\delta^{N+2}, w_0^{k-1})}{Mr\log\rho}\right) \tag{98}$$

$$\Pr(\mathcal{E}_{k,2}) \geq P_\delta\left(1 - \frac{2I(w_{k,2}; \mathbf{y}_0^N, \mathbf{v}_0^N | h_0^{N+1}, h_0^{N+1} \in \mathcal{H}_\delta^{N+2}, w_0^{k-1}, w_{k,1})}{Mr\log\rho}\right). \tag{99}$$

Thus we have that

$$\max_{0\leq k\leq N-1} \max\{\Pr(\mathcal{E}_{k,1}), \Pr(\mathcal{E}_{k,2}))\}$$
$$\geq \frac{1}{2N}\sum_{k=0}^{N-1}\{\Pr(\mathcal{E}_{k,1}) + \Pr(\mathcal{E}_{k,2})\} \tag{100}$$
$$\geq P_\delta\left(1 - \frac{I(w_0^{N-1}; \mathbf{y}_0^N, \mathbf{v}_0^N | h_0^{N+1} \in \mathcal{H}_\delta^{N+1})}{NMr\log\rho}\right) \tag{101}$$
$$\geq P_\delta\left(1 - \frac{I(\mathbf{x}_0^N; \mathbf{y}_0^N, \mathbf{v}_0^N | h_0^{N+1} \in \mathcal{H}_\delta^{N+1})}{NMr\log\rho}\right) \tag{102}$$
$$\geq P_\delta \times \left(1 - \frac{\sum_{k=0}^N I(\mathbf{x}_{k,1}; \mathbf{y}_{k,1}, \mathbf{v}_{k,1}|h_k) + I(\mathbf{x}_{k,2}; \mathbf{y}_{k,2}, \mathbf{v}_{k,2}|h_k, h_{k+1})}{NMr\log\rho}\right) \tag{103}$$
$$\geq \rho^{-(1-r+\delta)}\left(1 - \frac{N+1+(N+1)(r-\delta)\log\rho}{Nr\log\rho}\right). \tag{104}$$

The steps leading to (104) are similar to (25) and hence are not elaborated. For N sufficiently large the expression in the brackets in (104) 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 that arrive at equal intervals in each coherence block, different messages observe different channel conditions. Prop. 1 shows that coding schemes that exploit this asymmetry between messages can improve the DMT. On the other hand such schemes depend crucially on where the messages arrive in each block. If such information is not available the DMT is, in general, smaller, as is established in Prop. 2.

VII. Conclusions

In this paper we study the problem of delay-constrained streaming over a block fading channel. We establish the diversity multiplexing tradeoff when there is one message arriving in each coherence block. The converse is based on a novel "outage-amplification" argument that builds up a contradiction, over a sufficiently large duration, if we assume a larger DMT. We propose two coding schemes for achieving the optimal DMT. The first uses an interleaving scheme that reduces the system to a set of parallel independent channels. The advantage of this scheme is its simplicity. The disadvantage is that the playback deadline T must be known in advance. The second scheme pairs a sequential decoder with a tree code. This scheme also attains that DMT, and in a delay-universal fashion, but is more computationally complex and appears to require sufficiently long coherence blocks. Finally, we discuss some extensions when multiple messages arrive in each coherence block.

The fundamental limits of delay-constrained streaming over fading channels are not well understood in general. We hope that the techniques developed in this work can serve as a useful starting point for other investigations.

Appendix A
Proof of the Lower Bound in (16)

In this Appendix we incorporate the effect of outage into our lower bound. We continue from (9), dividing the channel realizations into sets in which the suffix is in outage, and when it is not, and dropping the latter. Thus,

$$\Pr[\mathcal{E}_k] \geq \sum_{\mathbf{H}_0^{T_k}:\mathbf{H}_k^{T_k}\in\mathcal{H}_\delta^T} \Pr[\mathbf{H}_0^{T_k} = \mathbf{H}_0^{T_k}]\Pr[\mathcal{E}_k|\mathbf{H}_0^{T_k} = \mathbf{H}_0^{T_k}].$$

Applying (12) and marginalizing out over the prefixes $\{\mathbf{H}_0^{k-1}\}$ we get

$$\Pr[\mathcal{E}_k] \geq \sum_{\mathbf{H}_k^{T_k}:\mathbf{H}_k^{T_k}\in\mathcal{H}_\delta^T} \Pr[\mathbf{H}_k^{T_k} = \mathbf{H}_k^{T_k}]\times \left(1 - \frac{1}{Mr\log\rho} - \frac{I(w_k; \mathbf{Y}_k^{T_k}|w_0^{k-1}, \mathbf{H}_k^{T_k} = \mathbf{H}_k^{T_k})}{Mr\log\rho}\right). \tag{105}$$

We can further express the right hand side of (105) as follows. Note that

$$\sum_{\mathbf{H}_k^{T_k}:\mathbf{H}_k^{T_k}\in\mathcal{H}_\delta^T} \Pr[\mathbf{H}_k^{T_k} = \mathbf{H}_k^{T_k}]\left(1 - \frac{1}{Mr\log\rho}\right) \tag{106}$$
$$= \left(1 - \frac{1}{Mr\log\rho}\right)\sum_{\mathbf{H}_k^{T_k}:\mathbf{H}_k^{T_k}\in\mathcal{H}_\delta^T} \Pr[\mathbf{H}_k^{T_k}] \tag{107}$$
$$= \left(1 - \frac{1}{Mr\log\rho}\right)\Pr[\mathbf{H}_k^{T_k} \in \mathcal{H}_\delta^T] \tag{108}$$

For the second term in (105), note that

$$\sum_{\mathbf{H}_k^{T_k}:\mathbf{H}_k^{T_k}\in\mathcal{H}_\delta^T} \Pr[\mathbf{H}_k^{T_k} = \mathbf{H}_k^{T_k}]I(w_k; \mathbf{Y}_k^{T_k}|w_0^{k-1}, \mathbf{H}_k^{T_k} = \mathbf{H}_k^{T_k}) \tag{109}$$
$$= \sum_{\mathbf{H}_k^{T_k}:\mathbf{H}_k^{T_k}\in\mathcal{H}_\delta^T} \Pr[\mathbf{H}_k^{T_k} = \mathbf{H}_k^{T_k}]\times I(w_k; \mathbf{Y}_k^{T_k}|w_0^{k-1}, \mathbf{H}_k^{T_k} = \mathbf{H}_k^{T_k}, \mathbf{H}_k^{T_k} \in \mathcal{H}_\delta^T) \tag{110}$$
$$= \Pr[\mathbf{H}_k^{T_k} \in \mathcal{H}_\delta^T]\sum_{\mathbf{H}_k^{T_k}:\mathbf{H}_k^{T_k}\in\mathcal{H}_\delta^T} \Pr[\mathbf{H}_k^{T_k} = \mathbf{H}_k^{T_k}|\mathbf{H}_k^{T_k} \in \mathcal{H}_\delta^T]\times I(w_k; \mathbf{Y}_k^{T_k}|w_0^{k-1}, \mathbf{H}_k^{T_k} = \mathbf{H}_k^{T_k}, \mathbf{H}_k^{T_k} \in \mathcal{H}_\delta^T) \tag{111}$$
$$= \Pr[\mathbf{H}_k^{T_k} \in \mathcal{H}_\delta^T]I(w_k; \mathbf{Y}_k^{T_k}|w_0^{k-1}, \mathbf{H}_k^{T_k}, \mathbf{H}_k^{T_k} \in \mathcal{H}_\delta^T), \tag{112}$$

where in (110), we use the fact that the indicator random variable $\mathbf{H}_k^{T_k} \in \mathcal{H}_\delta^T$ is a deterministic function of $\mathbf{H}_k^{T_k}$, and hence can be added as in (110). In (111), we note that for each $\mathbf{H}_k^{T_k} \in \mathcal{H}_\delta^T$ we have:

$$\Pr[\mathsf{H}_k^{T_k} = \mathbf{H}_k^{T_k} | \mathsf{H}_k^{T_k} \in \mathcal{H}_\delta^T] \tag{113}$$

$$= \frac{\Pr[\mathsf{H}_k^{T_k} = \mathbf{H}_k^{T_k}] \Pr[\mathsf{H}_k^{T_k} \in \mathcal{H}_\delta^T | \mathsf{H}_k^{T_k} = \mathbf{H}_k^{T_k}]}{\Pr[\mathsf{H}_k^{T_k} \in \mathcal{H}_\delta^T]} \tag{114}$$

$$= \frac{\Pr[\mathsf{H}_k^{T_k} = \mathbf{H}_k^{T_k}]}{\Pr[\mathsf{H}_k^{T_k} \in \mathcal{H}_\delta^T]}. \tag{115}$$

Thus substituting (115) in (110) we have that (111) follows.

Substituting (108) and (112) into (105), we have

$$\Pr[\mathcal{E}_k] \geq \Pr[\mathsf{H}_k^{T_k} \in \mathcal{H}_\delta^T] \times \left(1 - \frac{1}{Mr \log \rho} - \frac{I(w_k; \mathsf{Y}_k^{T_k} | w_0^{k-1}, \mathsf{H}_k^{T_k}, \mathsf{H}_k^{T_k} \in \mathcal{H}_\delta^T)}{Mr \log \rho}\right). \tag{116}$$

This completes the justification of (16).

Appendix B
Proof of (37).

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

Theorem 3: Suppose that $\mathsf{x}_1, \dots, \mathsf{x}_N$ are i.i.d. random variables with a rate function $f_\mathsf{x}(\cdot)$ defined as

$$f_\mathsf{x}(t) = \sup_\theta \left\{ \theta \cdot t - \log E_\mathsf{x} \left[\exp(\theta \cdot \mathsf{x})\right] \right\}, \tag{117}$$

and let $M_n = \frac{1}{n} \sum_{i=1}^n \mathsf{x}_i$. Then there exists a constant $N > 0$ such that for all $n \geq N$

$$\Pr(M_n \geq t) \leq e^{-n f_\mathsf{x}(t)}. \tag{118}$$

□

Recall that $\mathcal{A}_{l,l'}$ 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

$$\Pr(\mathcal{A}_{l,l'}) \leq \Pr\left(\left| \frac{\sum_{k=l}^{l'} [-\log p_{\mathsf{X}_k, \mathsf{Y}_k}(\mathbf{X}_k, \mathbf{Y}_k) - h(p_{\mathsf{X}_k, \mathsf{Y}_k})]}{M(l'-l+1)} \right| > \varepsilon \right). \tag{119}$$

Note that as $\mathsf{Y}_k = \mathsf{H}_k \mathsf{X}_k + \mathsf{Z}_k$, the H_k are known to the decoder, and the noise sequence $\{\mathsf{Z}_k\}$ is independent,

$$p_{\mathsf{X}_k, \mathsf{Y}_k}(\mathbf{X}_k, \mathbf{Y}_k) = p_{\mathsf{X}_k}(\mathbf{X}_k) p_{\mathsf{Y}_k | \mathsf{X}_k}(\mathbf{Y}_k | \mathbf{X}_k) \tag{120}$$

$$= p_{\mathsf{X}_k}(\mathbf{X}_k) p_{\mathsf{Z}_k}(\mathbf{Y}_k - \mathbf{H}_k \cdot \mathbf{X}_k) \tag{121}$$

$$= p_{\mathsf{X}}(\mathbf{X}_k) p_{\mathsf{Z}_k}(\mathbf{Z}), \tag{122}$$

where the last equality holds since the codewords are sampled i.i.d. and the noise is also i.i.d. Thus $h(p_{\mathsf{X}_k, \mathsf{Y}_k}) = h(p_\mathsf{X}) + h(p_\mathsf{Z})$. And so

$$\left| \sum_{k=l}^{l'} [-\log p_{\mathsf{X}_k, \mathsf{Y}_k}(\mathbf{X}_k, \mathbf{Y}_k) - h(p_{\mathsf{X}_k, \mathsf{Y}_k})] \right| \tag{123}$$

$$= \left| \sum_{k=l}^{l'} [-\log p_\mathsf{X}(\mathbf{X}_k) - \log p_\mathsf{Z}(\mathbf{Z}_k) - h(p_\mathsf{X}) - h(p_\mathsf{Z})] \right|$$

$$\leq \left| \sum_{k=l}^{l'} [-\log p_\mathsf{X}(\mathbf{X}_k) - h(p_\mathsf{X}) \right| + \left| \sum_{k=l}^{l'} [-\log p_\mathsf{Z}(\mathbf{Z}_k) - h(p_\mathsf{Z})] \right| \tag{124}$$

where the last step follows from the triangle inequality. Substituting (124) into (119) and using using the union bound we have

$$\Pr(\mathcal{A}_{l,l'}) \leq \Pr(\mathcal{A}_{l,l'}^X) + \Pr(\mathcal{A}_{l,l'}^Z) \tag{125}$$

where we define

$$\mathcal{A}_{l,l'}^X = \left\{ \mathbf{X}_l^{l'} : \left| \frac{\sum_{k=l}^{l'} [-\log p_\mathsf{X}(\mathbf{X}_k) - h(p_\mathsf{X})]}{M(l'-l+1)} \right| \geq \varepsilon \right\}, \tag{126}$$

$$\mathcal{A}_{l,l'}^Z = \left\{ \mathbf{Z}_{l'}^{l} : \left| \frac{\sum_{k=l}^{l'} [-\log p_\mathsf{Z}(\mathbf{Z}_k) - h(p_\mathsf{Z})}{M(l'-l+1)} \right| \geq \varepsilon \right\}. \tag{127}$$

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

$$\Pr(\mathcal{A}_{l,l'}^X) \leq \exp\{-M(l'-l+1) f_X(\varepsilon)\},$$
$$\Pr(\mathcal{A}_{l,l'}^Z) \leq \exp\{-M(l'-l+1) f_Z(\varepsilon)\}.$$

Furthermore by directly using (117) we can show that $f_X(\varepsilon) > 0$ and $f_Y(\varepsilon) > 0$. Setting $f(\varepsilon) = \max(f_X(\varepsilon), f_Z(\varepsilon))$ establishes (37).

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] A. Lozano, A. Tulino, and S. Verdu, "Multiantenna capacity: Myths and realities," in *Space-Time Wireless Systems: From Array Processing to MIMO Communications*, C. P. H. Bolcskei, D. Gesbert and A. J. van der Veen, Eds. Cambridge University Press, 2006, pp. 87–107.

[7] 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.

[8] S. Yang and J. Belfiore, "Optimal space-time codes for the mimo amplify-and-forward cooperative channel," *IEEE Trans. Inform. Theory*, vol. 53, no. 2, pp. 647–663, 2007.

[9] H. Lu, "Constructions of multiblock space-time coding schemes that achieve the diversity-multiplexing tradeoff," *IEEE Trans. Inform. Theory*, vol. 54, no. 8, pp. 3790–3796, 2008.

[10] L. Mroueh, "Space time coding in multiple input multiple output systems: challenges and applications," in *Communication systems: New research*. Nova Publisher, 2012.

[11] P. Coronel and H. Bolcskei, "Diversity-Multiplexing tradeoff in selective-fading MIMO channels," in *Proc. Int. Symp. Inform. Theory*, 2007, pp. 2841–2845.

[12] H. Bolsckei, “Personal communication.” June 2013.
[13] 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.
[14] ——, “Anytime information theory,” Ph.D. dissertation, Mass. Instit. of Tech., 2001.
[15] S. C. Draper, C. Chang, and A. Sahai, “Sequential random binning for streaming distributed source coding,” in *Proc. Int. Symp. Inform. Theory*, 2005.
[16] ——, “Lossless coding for distributed streaming sources,” *IEEE Trans. Inform. Theory*, vol. 60, pp. 1447–1474, Mar. 2014.
[17] R. Sukhavasi and B. Hassibi, “Linear error correcting codes with anytime reliability,” in *Proc. Int. Symp. Inform. Theory*, 2011, pp. 1748–1752.
[18] G. Cocco, D. Gunduz, and C. Ibars, “Streaming transmitter over block fading channels with delay constraint,” *IEEE Trans. Wireless Communications*, vol. 12, p. 4315, 2013.
[19] ——, “Throughput and delay analysis in video streaming over block-fading channels,” in *International Conference on Communications*, 2013, pp. 5578–5582.
[20] S. Kittipiyakul and T. Javidi, “Optimal operating point in MIMO channel for delay-sensitive and bursty traffic,” in *Proc. Int. Symp. Inform. Theory*, 2006.
[21] 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, no. 2, pp. 746–763, 2009.
[22] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley and Sons, 1991.
[23] L. Ozarow, S. Shamai, and A. Wyner, “Information-theoretic considerations for cellular mobile radio,” *IEEE Trans. Vehic. Technol.*, vol. 43, pp. 359–378, May 1994.
[24] E. Telatar, “Capacity of multi-antenna Gaussian channels,” *European Transactions on Telecommunications*, vol. 10, pp. 585–595, Nov 1999.

Ashish Khisti Ashish Khisti is an assistant professor in the Electrical and Computer Engineering (ECE) department at the University of Toronto, Toronto, Ontario Canada. He received his BASc degree in Engineering Sciences from University of Toronto and his S.M and Ph.D. Degrees from the Massachusetts Institute of Technology (MIT), Cambridge, MA, USA. His research interests span the areas of information theory, wireless physical layer security and streaming in multimedia communication systems. At the University of Toronto, he heads the signals, multimedia and security laboratory. For his graduate studies he was a recipient of the NSERC postgraduate fellowship, HP/MIT alliance fellowship, Harold H. Hazen Teaching award and the Morris Joseph Levin Masterworks award.

Stark C. Draper (S'99-M'03) received the M.S. and Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), and the B.S. and B.A. degrees in electrical engineering and history, respectively, from Stanford University.

He is an Associate Professor of Electrical and Computer Engineering at the University of Toronto, Canada. From 2007-2014 he was an Associate Professor at the University of Wisconsin, Madison. Before moving to the University of Wisconsin he was with the Mitsubishi Electric Research Laboratories (MERL), Cambridge, MA. He has held postdoctoral positions in the Wireless Foundations, University of California, Berkeley, and in the Information Processing Laboratory, University of Toronto. He has worked at Disney Research, Cambridge, MA, Arraycomm, San Jose, CA, the C. S. Draper Laboratory, Cambridge, MA, and Ktaadn, Newton, MA. His research interests include communication and information theory, error-correction coding, statistical signal processing and optimization, security, and application of these disciplines to computer architecture.

Dr. Draper has received the NSERC Discovery Award, the NSF CAREER Award, the 2010 MERL President's Award, departmental teaching awards from the University of Toronto and the University of Wisconsin, the MIT Carlton E. Tucker Teaching Award, an Intel Graduate Fellowship, Stanford's Frederick E. Terman Engineering Scholastic Award, and a U.S. State Department Fulbright Fellowship.