

Streaming data over fading wireless channels: The Diversity-Multiplexing Tradeoff

Ashish Khisti

ECE Dept.

University of Toronto

Toronto, ON, M1B 5P1

akhisti@comm.utoronto.ca

Stark C. Draper

ECE Dept.

University of Wisconsin, Madison

Madison, WI, 53076

sdraper@ece.wisc.edu

Abstract—We study delay constrained sequential streaming over block fading channels. The transmitter observes a stream of messages, one message in each coherence block, and the receiver needs to output a sequence of messages, each with a fixed delay of T coherence blocks. We characterize the associated diversity-multiplexing tradeoff (DMT) for this model. The proposed coding scheme involves a semi-infinite random Gaussian tree-code and a sequential decision directed decoder. The converse applies an *outage amplification argument* that exploits the delay constraint to amplify the error event associated with a single message to an entire sequence of messages.

I. INTRODUCTION

Multimedia streaming over wireless links is a challenging problem, both because of the stringent requirements of streaming data and the rapid fluctuations of wireless links. Understanding fundamental limits and characterizing good engineering architectures of such systems is highly relevant to several emerging applications. 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 and identically distributed source messages, one in each coherence block. The transmitted signal produced by our *streaming encoder* can only be a causal function of the observed source packets. The *delay constrained decoder* is required to output each source packet with a maximum delay of T coherence blocks. We characterize the diversity-multiplexing tradeoff [1] of this delay-constrained streaming model.

In related literature, a framework for delay-universal streaming has been introduced by Sahai and his co-authors [2]–[5]. These works investigate error exponents in the limit of large decoding delays. There are however some important differences with our work. First, these papers consider a maximum likelihood decoder, whereas for the outage analysis we propose a decision directed decoder based on joint-typicality. Secondly, except in some special cases [6], the characterization of error exponents remains an open problem. In contrast, the present paper develops a converse based on an *outage amplification argument* and thus provides a complete characterization of

the DMT in the streaming setup. Finally [2]–[4] also develop connections to control with noisy-feedback. We do not pursue such connections in the present paper.

II. MODEL

In the proposed model the transmitter has N_t transmit antennas, the receiver has N_r receive antennas, and the underlying channel model is an independent identically distributed (i.i.d.) block fading channel model with a coherence period of M . The channel output is

$$\mathbf{Y}_k = \mathbf{H}_k \cdot \mathbf{X}_k + \mathbf{Z}_k, \quad (1)$$

where $k = 0, 1, \dots$, denotes the index of the coherence block of the fading channel. Each coherence block consists of M time-slots. The matrix $\mathbf{H}_k \in \mathbb{C}^{N_r \times N_t}$ denotes the channel transfer matrix in coherence period k , $\mathbf{X}_k = [\mathbf{X}_k(1) \mid \dots \mid \mathbf{X}_k(M)] \in \mathbb{C}^{N_t \times M}$ is a matrix whose j -th column, $\mathbf{X}_k(j)$, denotes the vector transmitted in time-slot j in block k and similarly $\mathbf{Y}_k \in \mathbb{C}^{N_r \times M}$ is a matrix whose j -th column, $\mathbf{Y}_k(j)$ denotes the vectors received in time-slot j in block k . The additive noise matrix is $\mathbf{Z}_k \in \mathbb{C}^{N_r \times M}$. Thus (1) can also be expressed as,

$$\mathbf{Y}_k(j) = \mathbf{H}_k \cdot \mathbf{X}_k(j) + \mathbf{Z}_k(j), \quad j = 1, \dots, M. \quad (2)$$

We assume that all entries of \mathbf{H}_k are sampled independently from the complex Gaussian distribution¹ 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{Z}_k are also sampled i.i.d. $\mathcal{CN}(0, 1)$. Finally all the channel matrices \mathbf{H}_k are revealed to the decoder, but not to the encoder.

We assume an average power constraint $E[\|\mathbf{X}_k\|_F^2] \leq M\rho$, where $\|\cdot\|_F$ denotes the Frobenius norm of the matrix \mathbf{X}_k . Note that ρ denotes the *average* signal-to-noise-ratio (SNR) at each receiving antenna. 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:

The work of A. Khisti was supported by a National Science and Engineering Research Council (NSERC) Discovery Grant. The work of S. C. Draper was supported by the National Science Foundation under grant CAREER 0844539.

¹While we only focus on this particular flat-fading Rayleigh channel model, our result in Theorem 2 easily extends to other channel models. The assumption of Rayleigh fading is primarily for convenience.

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

1. A sequence of messages $\{w_k\}_{k \geq 0}$ each distributed uniformly over the set $\mathcal{I} = \{1, 2, \dots, 2^{MR}\}$.
2. A sequence of encoding functions $\mathcal{F}_k : \mathcal{I}^{k+1} \rightarrow \mathbb{C}^{N_t \times M}$,

$$\mathbf{X}_k = \mathcal{F}_k(w_0, \dots, w_k), \quad k = 0, 1, \dots \quad (3)$$

that maps the input message sequence to a codeword of length M .

3. A sequence of decoding functions $\mathcal{G}_k : \mathbb{C}^{M(k+T)} \rightarrow \mathcal{I}$ that outputs message \hat{w}_k based on the first $k + T$ observations, i.e.,

$$\hat{w}_k = \mathcal{G}_k(\mathbf{Y}_0, \dots, \mathbf{Y}_{k+T-1}), \quad k = 0, 1, \dots \quad (4)$$

We now define the diversity-multiplexing tradeoff (DMT) [1] associated with the streaming code $\mathcal{C}(R, T)$. Let the error probability for the k -th message be $p_k = \Pr(w_k \neq \hat{w}_k)$ where \hat{w}_k is the delay- T decoder output (4). Let $\Pr(e) = \max_k p_k$ denote the maximum error probability in the streaming setup. The DMT tradeoff [1] of (r, d) is achievable with delay T if there exists a sequence of codebooks $\mathcal{C}(R = r \log \rho, T)$ such that

$$d = \lim_{\rho \rightarrow \infty} \frac{-\log \Pr(e)}{\log \rho}, \quad r = \lim_{\rho \rightarrow \infty} \frac{R(\rho)}{\log \rho}.$$

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 [1]. We reproduce the result below for the convenience of the reader.

Theorem 1: ([1]) For the quasi-static fading channel, the optimal DMT tradeoff curve $d_1(r)$ is given by the piecewise linear function connecting the points $(k, d_1(k))$ for $k = 0, 1, \dots, \min(N_r, N_t)$, where $d_1(k) = (N_r - k)(N_t - k)$.

As our main result, we establish that 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.

Remark 1: 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 normalized by the number of channels. This equivalence is surprising, as neither the achievability nor the converse follows from the result involving parallel channels. In a parallel-channel code, the information is coded jointly over available channels whereas in the streaming setup, the output only depends causally on the message stream. The decoder for parallel channels outputs all messages simultaneously, whereas the decoder in the streaming setup is only required to output each message by its individual deadline.

In the remainder of this paper we provide a sketch of the coding theorem and the converse. Complete details will be provided in the journal version of this paper.

IV. CODING THEOREM

We first provide details of the code construction and then provide the analysis of the DMT for our proposed scheme.

A. Coding Scheme

Our proposed streaming code, $\mathcal{C}(R, T)$, consists of a semi-infinite sequence of codebooks $\{\mathcal{C}_0, \mathcal{C}_1, \dots, \mathcal{C}_k, \dots\}$, where \mathcal{C}_k is the codebook to be used in coherence block k when messages (w_0, \dots, w_k) are revealed. Codebook \mathcal{C}_k consists of a total of $2^{MR(k+1)}$ codewords and each codeword is assigned to one element in the set

$$\mathcal{I}^{k+1} = \{ (w_0, \dots, w_k) : w_0 \in \mathcal{I}, \dots, w_k \in \mathcal{I} \}. \quad (5)$$

All codewords are length M sequences whose symbols are sampled i.i.d. from $\mathcal{CN}(0, \frac{\rho}{N_t})$. In coherence block k , the encoder observes w_0, \dots, w_k , maps it to the codeword $\mathbf{X}_k(w_0^k) \in \mathbb{C}^{N_t \times M}$ in \mathcal{C}_k , and transmits each of the M columns of \mathbf{X}_k over M channel uses. The entire transmitted sequence up to and including block k is denoted by

$$\mathbf{X}_0^k(w_0^k) \triangleq \{ \mathbf{X}_0(w_0), \mathbf{X}_1(w_0^1), \dots, \mathbf{X}_k(w_0^k) \}, \\ \mathbf{X}_0^k(w_0^k) \in \mathbb{C}^{N_t \times (k+1)M} \quad (6)$$

For decoding message w_k , our proposed decoder does not rely on previously decoded messages, but instead computes a new estimate of all the messages \hat{w}_0^k at time $T_k = k + T - 1$ using the entire received sequence $\mathbf{Y}_0^{T_k} = (\mathbf{Y}_0, \dots, \mathbf{Y}_{T_k})$. First it searches for a message \hat{w}_0 by searching over all message sequences $\hat{w}_0^{T_k}$ such that $\{ \mathbf{X}_0^{T_k}(\hat{w}_0^{T_k}), \mathbf{Y}_0^{T_k} \}$ are jointly typical. If each such sequence has a unique prefix \bar{w}_0 then \bar{w}_0 is selected as the message in block 0. Otherwise an error is declared.

The decoder then proceeds sequentially, producing estimates $\bar{w}_1, \dots, \bar{w}_k$. In determining w_l , $l \leq k$, the decoder uses the already-determined vector of estimates \bar{w}_0^{l-1} . The decoder 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

$$(\mathbf{X}_l^{T_k}(\bar{w}_0^{l-1}, \hat{w}_l^{T_k}), \mathbf{Y}_l^{T_k}) \in \mathcal{T}_{l, T_k}, \quad (7)$$

where the set $\mathcal{T}_{l, l'}$ is the set of all jointly typical sequences [7],

$$\mathcal{T}_{l, l'} = \left\{ (\mathbf{X}_l^{l'}, \mathbf{Y}_l^{l'}) : \mathbf{X}_l^{l'} \in \mathcal{T}(p_{\mathbf{X}_l^{l'}}), \mathbf{Y}_l^{l'} \in \mathcal{T}(p_{\mathbf{Y}_l^{l'}}), \right. \\ \left. \left| \frac{-\sum_{k=l}^{l'} [\log p_{\mathbf{X}_k, \mathbf{Y}_k}(\mathbf{X}_k, \mathbf{Y}_k) - h(p_{\mathbf{X}_k, \mathbf{Y}_k})]}{M(l' - l + 1)} \right| \leq \varepsilon \right\}, \quad (8)$$

where $\mathcal{T}(p_{\mathbf{X}_l^{l'}})$ 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 where $h(p_{\mathbf{X}_k, \mathbf{Y}_k})$ denotes the differential entropy of jointly Gaussian random variables. These sets are determined by the input distribution, the noise statistics, and the known channel matrices $\{\mathbf{H}_k\}$.

If the list of all message sequences $\hat{w}_l^{T_k}$ that satisfy (7) 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, message \bar{w}_k is declared to be the output message.

Remark 2: We note that our decoder produces a fresh estimate of all the messages in each step. By using a longer decoding window we improve the reliability of the previous messages to an extent that the error event is dominated by the failure of the most recent message.

Remark 3: 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 $\mathbf{Y}_0^{T_k}$. Such an approach generally fails, since if $T_k \gg T$ then if $\hat{w}_k \neq w_k$, but $\hat{w}_0, \dots, \hat{w}_{k-1}$ are all correct, only the subsequence $\mathbf{X}_k^{T_k}(\hat{w}_0^{T_k})$ is different from the transmitted sequence. This is a negligible portion of the overall sequence when $T_k \gg T$ and the resulting $(\mathbf{X}_0^{T_k}(\hat{w}_0^{T_k}), \mathbf{Y}_0^{T_k})$ will likely appear typical, leading to an error.

B. Analysis of error probability

We show that for any $\delta > 0$ and $0 \leq r \leq N_r \cdot N_t$, the error probability averaged over the ensemble of codebooks $\mathcal{C}(R = (r-\delta) \log_2 \rho, T)$, satisfies $\Pr(\mathcal{E}) \leq \rho^{-d(r)}$. This shows that there exists at least one codebook in the ensemble that meets the diversity order. By symmetry we can assume without loss of generality that a particular fixed message sequence is transmitted i.e., $w_0^k = w_0^k$.

For the analysis of error probability we define the events

$$\mathcal{E}_l = \left\{ w_0^l : (\bar{w}_0, \dots, \bar{w}_{l-1}) = (w_0, \dots, w_{l-1}), \bar{w}_l \neq w_l \right\},$$

$$0 \leq l \leq k \quad (9)$$

and note that

$$\Pr\{\bar{w}_k \neq w_k\} \leq \sum_{l=0}^k \Pr(\mathcal{E}_l), \quad (10)$$

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 (10).

We further express $\mathcal{E}_l = \mathcal{A}_l \cup \mathcal{B}_l$, where

$$\mathcal{A}_l = \left\{ (\mathbf{X}_l^{T_k}(w_0^{T_k}), \mathbf{Y}_l^{T_k}) \notin \mathcal{T}_{l, T_k} \right\} \quad (11)$$

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

$$\mathcal{B}_l = \left\{ \exists \bar{w}_0^{T_k} : \bar{w}_0^{l-1} = w_0^{l-1}, \bar{w}_l \neq w_l, \right. \\ \left. (\mathbf{X}_l^{T_k}(\bar{w}_0^{T_k}), \mathbf{Y}_l^{T_k}) \in \mathcal{T}_{l, T_k} \right\} \quad (12)$$

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 A, using an appropriate Chernoff bound we can express,

$$\Pr(\mathcal{A}_l) \leq 2^{-M(T_k-l+1)f(\varepsilon)} \quad (13)$$

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 transmit subsequence $\mathbf{X}_l^{T_k}(\bar{w}_0^{T_k})$ is independent of $\mathbf{Y}_l^{T_k}$. Hence from the joint typicality theorem [7], we have that for any sequence $\bar{w}_0^{T_k}$ with $\bar{w}_l \neq w_l$

$$\Pr\left((\mathbf{X}_l^{T_k}(\bar{w}_0^{T_k}), \mathbf{Y}_l^{T_k}) \in \mathcal{T}_{l, T_k} \mid \mathbf{H}_l^{T_k} \right) \\ \leq 2^{-M\left(\sum_{j=l}^{T_k} I(\mathbf{X}_j(1); \mathbf{Y}_j(1)) - 3\varepsilon\right)} \\ = 2^{-M\left(\sum_{j=l}^{T_k} C_j(\rho) - 3\varepsilon\right)}$$

where we have introduced,

$$C_j(\rho) = \log \det \left(\mathbf{I} + \frac{\rho}{N_t} \mathbf{H}_j \mathbf{H}_j^\dagger \right) \quad (14)$$

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

$$\Pr(\mathcal{B}_l \mid \mathbf{H}_l^{T_k}) \leq 2^{-M\left(\sum_{j=l}^{T_k} C_j(\rho) - (T_k-l+1)R - 3\varepsilon\right)}. \quad (15)$$

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

$$\mathcal{O}_l = \left\{ (\mathbf{H}_l, \dots, \mathbf{H}_{T_k}) : \sum_{i=l}^{T_k} C_i(\rho) \leq \right. \\ \left. (k+T-l)r \log \rho + (k-l)\Delta(r) \log \rho + 4\varepsilon \log \rho \right\} \quad (16)$$

where

$$\Delta(r) = \frac{(N_t - r)(N_r - r)}{2(N_t + N_r - 2r)}, \quad 0 \leq r < \min(N_t, N_r). \quad (17)$$

The following lemma, which can also be interpreted as the diversity-multiplexing tradeoff of symmetric parallel MIMO channels is useful in our analysis.

Lemma 1: For any $\beta = 1, 2, \dots$ and $s \geq 0$ we have that

$$\Pr\left(\sum_{i=l}^{l+\beta-1} C_i(\rho) \leq s \cdot \log \rho \right) \doteq \rho^{-\beta d_1(s/\beta)} \quad (18)$$

where $d_1(\cdot)$ denotes the DMT of the single MIMO link in Theorem 1. \square

Taking expectation of the conditional probability in (15), and upper bounding the associated events,

$$\Pr(\mathcal{B}_l) \leq \Pr(\mathcal{B}_l \mid \mathbf{H}_l^{T_k} \in \mathcal{O}_l^c) + \Pr(\mathbf{H}_l^{T_k} \in \mathcal{O}_l) \quad (19)$$

From (16) and (15) we have

$$\Pr(\mathcal{B}_l \mid \mathbf{H}_l^{T_k} \in \mathcal{O}_l^c) \leq 2^{-M(k-l)\Delta(r) \log \rho - M\varepsilon \log \rho} \quad (20)$$

$$\doteq \rho^{-M\varepsilon - M(k-l)\Delta(r)}. \quad (21)$$

Thus it remains to bound $\Pr(\mathcal{O}_l)$ in (19). Substituting $\beta = T_k - l + 1$ and $s = \beta r + (k-l)\Delta(r) + 4\varepsilon$, in (18),

$$\beta \cdot d_1\left(\frac{s}{\beta}\right) = \beta \left(N_t - \frac{s}{\beta}\right) \left(N_r - \frac{s}{\beta}\right) \quad (22)$$

$$\geq \beta(N_t - r)(N_r - r) - (k-l)(N_t + N_r - 2r)\Delta(r) - o_\varepsilon(1) \quad (23)$$

$$= T(N_t - r)(N_r - r) + \frac{(k-l)}{2}(N_t - r)(N_r - r) - o_\varepsilon(1) \quad (24)$$

$$= Td_1(r) + \frac{k-l}{2}d_1(r) - o_\varepsilon(1) \quad (25)$$

where we substituted (17) for $\Delta(r)$ in (24) and let $o_\varepsilon(1)$ be a function of ε that vanishes as $\varepsilon \rightarrow 0$.

Thus we have

$$\Pr(\mathcal{O}_l) \leq \rho^{-(Td_1(r) + \frac{(k-l)}{2}d_1(r) + o_\varepsilon(1))}. \quad (26)$$

From (19) and substituting (21) and (26) and using $\mathcal{E}_l = \mathcal{A}_l \cup \mathcal{B}_l$ we have

$$\Pr(\mathcal{E}_l) \leq \Pr(\mathcal{A}_l) + \Pr(\mathcal{B}_l) \quad (27)$$

$$\leq 2^{-M(T_k - l + 1)f(\varepsilon)} + \rho^{-M\varepsilon - M(k-l)\Delta(r)} + \rho^{-Td_1(r) - \frac{(k-l)}{2}d_1(r) + o_\varepsilon(1)}. \quad (28)$$

From the union bound,

$$\Pr(\mathcal{E}) \leq \sum_{l=0}^k \Pr(\mathcal{E}_l) \quad (29)$$

$$\leq \sum_{l=0}^k 2^{-M(T_k - l + 1)f(\varepsilon)} + \sum_{l=0}^k \rho^{-M\varepsilon - M(k-l)\Delta(r)} + \sum_{l=0}^k \rho^{-Td_1(r) - \frac{(k-l)}{2}d_1(r) + o_\varepsilon(1)} \quad (30)$$

Substituting for $T_k = k + T - 1$, we can express the first term in (30) as

$$\sum_{l=0}^k 2^{-M(k+T-l)f(\varepsilon)} = \sum_{l=0}^k 2^{-M(l+T)f(\varepsilon)} \quad (31)$$

$$\leq \sum_{l=0}^{\infty} 2^{-M(l+T)f(\varepsilon)} = 2^{-MTf(\varepsilon)+1}, \quad (32)$$

which vanishes as $M \rightarrow \infty$. By a similar argument we can simplify the second and third terms in (30) to get

$$\Pr(\mathcal{E}) \leq o_M(1) + \rho^{-Td_1(r) + o_\varepsilon(1)}. \quad (33)$$

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

V. CONVERSE

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

$$\Pr(e) \geq \rho^{-Td_1(r)}$$

where $d_1(r)$ is the DMT tradeoff associated with a single-link MIMO channel. It is convenient to define $\mathcal{E}_l = \{w_l \neq \hat{w}_l\}$, the error event associated with block l and note that $\Pr(e) = \max_{k \geq 0} \Pr(\mathcal{E}_k)$.

We begin by lower bounded the probability of the error event \mathcal{E}_k associated with message w_k . Recall that this message needs to be decoded after T coherence blocks indexed as $t \in \{k, \dots, k+T-1\}$. Let $\mathbf{H}_k^{k+T-1} = \mathbf{H}_k^{k+T-1}$ be the realization of the transfer channel matrices in this interval. Applying Fano's Inequality for message w_k and using the fact that w_0^{k-1} is independent of w_k and of the channel matrices we have

$$\Pr(\mathcal{E}_k; \mathbf{H}_k^{k+T-1} = \mathbf{H}_k^{k+T-1}) \geq 1 - \frac{1}{Mr \log \rho} - \frac{I(w_k; \mathbf{Y}_k^{k+T-1} | w_0^{k-1}, \mathbf{H}_k^{k+T-1} = \mathbf{H}_k^{k+T-1})}{Mr \log \rho}.$$

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

$$\mathcal{H}_\delta = \left\{ \mathbf{H} : \log \det \left(\mathbf{I} + \frac{\rho}{M} \mathbf{H} \mathbf{H}^\dagger \right) \leq (r - \delta) \log \rho \right\} \quad (34)$$

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

$$\Pr(\mathbf{H}_k^{k+T-1} \in \mathcal{H}_\delta^T) = P_\delta^T \quad (35)$$

where $P_\delta = \Pr(\mathbf{H} \in \mathcal{H}_\delta)$. And, from single link DMT in Theorem 1,

$$P_\delta = \Pr(\mathbf{H} \in \mathcal{H}_\delta) \doteq \rho^{-d_1(r-\delta)}. \quad (36)$$

Further as we establish in the full paper, we can show that

$$\Pr(\mathcal{E}_k) \geq P_\delta^T \cdot \left(1 - \frac{I(w_k; \mathbf{Y}_0^{N-1} | w_0^{k-1}, \mathbf{H}_0^{N-1} \in \mathcal{H}_\delta^N)}{Mr \log \rho} \right) \quad (37)$$

Now we combine the error events. Since the maximum is always larger than the average,

$$\begin{aligned} \max_{0 \leq k \leq N-T-1} \Pr(\mathcal{E}_k) &\geq \frac{1}{N-T} \sum_{k=0}^{N-T-1} \Pr(\mathcal{E}_k) \\ &\geq P_\delta^T \left(1 - \frac{\sum_{k=0}^{N-T-1} I(w_k; \mathbf{Y}_0^{N-1} | w_0^{k-1}, \mathbf{H}_0^{N-1} \in \mathcal{H}_\delta^N)}{(N-T)Mr \log \rho} \right) \\ &= P_\delta^T \left(1 - \frac{I(w_0^{N-T-1}; \mathbf{Y}_0^{N-1} | \mathbf{H}_0^{N-1} \in \mathcal{H}_\delta^N)}{(N-T)Mr \log \rho} \right) \\ &\geq P_\delta^T \left(1 - \frac{I(\mathbf{X}_0^{N-1}; \mathbf{Y}_0^{N-1} | \mathbf{H}_0^{N-1} \in \mathcal{H}_\delta^N)}{(N-T)Mr \log \rho} \right) \end{aligned} \quad (38)$$

where (38) follows from the data processing inequality since $w_0^{N-T-1} \rightarrow \mathbf{X}_0^{N-1} \rightarrow \mathbf{Y}_0^{N-1}$ holds. And, since the channel is memoryless,

$$\begin{aligned} & \max_{0 \leq k \leq N-T-1} \Pr(\mathcal{E}_k) \\ & \geq P_\delta^T \left(1 - \frac{\sum_{i=0}^{N-1} I(\mathbf{X}_i; \mathbf{Y}_i | \mathbf{H}_i \in \mathcal{H}_\delta)}{(N-T)Mr \log \rho} \right) \end{aligned} \quad (39)$$

$$\geq P_\delta^T \left(1 - \frac{NM(r-\delta) \log \rho}{(N-T)Mr \log \rho} \right) \quad (40)$$

$$\doteq \left(1 - \frac{N(r-\delta)}{(N-T)r} \right) \rho^{-Td_1(r-\delta)} \quad (41)$$

where (40) follows by substituting in (34).

Now we arrive at the bound. For any $\delta > 0$, by selecting $N > T \frac{r}{\delta}$ the term inside the brackets is strictly positive. Since $\delta > 0$ is arbitrary, it follows that the diversity order greater than $Td_1(r)$ cannot be achieved.

VI. CONCLUSIONS

We studied the problem of delay constrained streaming over a block fading channel and established that the associated diversity multiplexing tradeoff. The proposed code construction involves a semi-infinite tree construction. The associated decoder involves a decision-directed joint-typicality based decision rule. It exploits the exponentially increasing reliability in decoding of the past messages to prevent error propagation. We also provide a new converse based on an outage-amplification argument that amplifies the error associated with a single message to an entire sequence. 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 (13)

Recall that $\mathcal{A}_{l,l'}$ the event that the true codeword is not jointly typical with the received sequence. If we ignore the marginal typicality constraints we can upper bound this event as

$$\begin{aligned} & \Pr(\mathcal{A}_{l,l'}) \leq \\ & \Pr \left(\left| \frac{\sum_{k=l}^{l'} [\log p_{\mathbf{X}_k, \mathbf{Y}_k}(\mathbf{X}_k, \mathbf{Y}_k) - h(p_{\mathbf{X}_k, \mathbf{Y}_k})]}{M(l'-l+1)} \right| > \varepsilon \right), \end{aligned} \quad (42)$$

which follows since the codewords are sampled i.i.d. Note that as $\mathbf{Y}_k = \mathbf{H}_k \mathbf{X}_k + \mathbf{Z}_k$, the \mathbf{H}_k are known to the decoder, and the noise sequence $\{\mathbf{Z}_k\}$ is independent,

$$p_{\mathbf{X}_k, \mathbf{Y}_k}(\mathbf{X}_k, \mathbf{Y}_k) = p_{\mathbf{X}}(\mathbf{X}_k) p_{\mathbf{Y}_k | \mathbf{X}_k}(\mathbf{Y}_k | \mathbf{X}_k) \quad (43)$$

$$= p_{\mathbf{X}}(\mathbf{X}_k) p_{\mathbf{Z}}(\mathbf{Y}_k - \mathbf{H}_k \cdot \mathbf{X}_k) \quad (44)$$

$$= p_{\mathbf{X}}(\mathbf{X}_k) p_{\mathbf{Z}}(\mathbf{Z}_k), \quad (45)$$

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

$$\left| \sum_{k=l}^{l'} [\log p_{\mathbf{X}_k, \mathbf{Y}_k}(\mathbf{X}_k, \mathbf{Y}_k) - h(p_{\mathbf{X}_k, \mathbf{Y}_k})] \right| \quad (46)$$

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

$$\leq \left| \sum_{k=l}^{l'} [\log p_{\mathbf{X}}(\mathbf{X}_k) - h(p_{\mathbf{X}})] \right| + \left| \sum_{k=l}^{l'} [\log p_{\mathbf{Z}}(\mathbf{Z}_k) - h(p_{\mathbf{Z}})] \right| \quad (47)$$

where the last step follows from the triangular inequality. Substituting (47) into (42) and using the union bound we have

$$\Pr(\mathcal{A}_{l,l'}) \leq \Pr(\mathcal{A}_{l,l'}^X) + \Pr(\mathcal{A}_{l,l'}^Z) \quad (48)$$

where we define

$$\mathcal{A}_{l,l'}^X = \left\{ \mathbf{X}_{l'}^l : \left| \frac{\sum_{k=l}^{l'} [\log p_{\mathbf{X}}(\mathbf{X}_k) - h(p_{\mathbf{X}})]}{M(l'-l+1)} \right| \geq \varepsilon \right\}, \quad (49)$$

$$\mathcal{A}_{l,l'}^Z = \left\{ \mathbf{Z}_{l'}^l : \left| \frac{\sum_{k=l}^{l'} [\log p_{\mathbf{Z}}(\mathbf{Z}_k) - h(p_{\mathbf{Z}})]}{M(l'-l+1)} \right| \geq \varepsilon \right\}. \quad (50)$$

Since \mathbf{X}_k is a sequence of M i.i.d. random vectors each sampled from $\mathcal{CN}(0, \frac{\rho}{N} \mathbf{I})$ and $E[\log p_{\mathbf{X}}(\mathbf{X}_k)] = h(p_{\mathbf{X}})$. Similarly, $E[\log p_{\mathbf{Z}}(\mathbf{Z}_k)] = h(p_{\mathbf{Z}})$. Then, from the Chernoff theorem, there exist *rate functions* $f_X(\varepsilon) > 0$ and $f_Z(\varepsilon) > 0$ such that for sufficiently large value of $M(l'-l+1)$

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

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

Setting $f(\varepsilon) = \max(f_X(\varepsilon), f_Z(\varepsilon))$ establishes (13).

REFERENCES

- [1] 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.
- [2] 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.
- [3] 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.
- [4] —, "Anytime information theory," Ph.D. dissertation, Mass. Instit. of Tech., 2001.
- [5] S. Draper, C. Chang, and A. Sahai, "Sequential random binning for streaming distributed source coding," in *Proc. Int. Symp. Inform. Theory*, 2005.
- [6] 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.
- [7] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. John Wiley and Sons, 1991.