Streaming with Autoregressive-Hamming Distortion for Ultra Short-Delay Communications

Roy Timo[†], Alex Grant[‡] and Badri N. Vellambi[‡]
Telecom ParisTech[†] and the University of South Australia[‡]
timo.roy-clinton@telecom-paristech.fr, {alex.grant, badri.vellambi}@unisa.edu.au

***Abstract*—A streaming communications model with an autoregressive distortion function is proposed. We show that the model is useful for delay-sensitive systems, and we present asymptotic and non-asymptotic achievability results that exhibit some fundamental tradeoffs between rate, reliability and delay.**

I. Setup & Motivation

The need for ultra short-delay high-rate communications systems in algorithmic trading, distributed control, cloud computing and vehicle-to-vehicle systems is driving significant new research and infrastructure investments [1]–[3]. In this paper, we investigate basic rate-reliability-delay tradeoffs for a streaming model of communications. The model is ideal for delay-sensitive systems and, for concreteness, we motivate it by a recent paradigm shift in finance — the algorithmic trading of financial instruments in the foreign exchange (FX) market [4]. Let us first give the problem setup.

A. Setup

A discrete memoryless source (DMS) emits a sequence of symbols $U_1, U_2, \ldots$, where each is uniformly distributed on a finite set $\mathcal{U} := \{0, 1, \ldots, M-1\}$. A transmitter streams the source symbols over a discrete memoryless channel (DMC) to a receiver operating with integer decoding delay (or, lookahead) $\Delta \geq 0$. Let $\mathcal{X}$ denote the channel's input alphabet, $\mathcal{Y}$ its output alphabet and $T_{Y|X}(y|x) := \mathbb{P}[Y = y|X = x]$ its transition probabilities. Encoding and decoding is performed in stages $n = 1, 2, \ldots$

Encoder (stage n): The DMS has output the first n source symbols, $\mathbb{U}_n := (U_1, U_2, \ldots, U_n)$. The transmitter sends

$$X_n := f^{(n)}(\mathbb{U}_n)$$

over the channel, where $f^{(n)} : \mathcal{U}^n \to \mathcal{X}$ is a deterministic map and $\mathcal{U}^n$ is the n-fold Cartesian product of $\mathcal{U}$. The receiver observes Y_n at the channel output.

Decoder (stage n): The channel has output the first $(n+\Delta)$ channel symbols, $\mathbb{Y}_{n+\Delta} := (Y_1, Y_2, \ldots, Y_{n+\Delta})$. The receiver attempts to reconstruct the first n source symbols $\mathbb{U}_n$ via

$$\hat{\mathbb{U}}_n^{(n)} := g^{(n)}(\mathbb{Y}_{n+\Delta}),$$

where $g^{(n)} : \mathcal{Y}^{n+\Delta} \to \mathcal{U}^n$ is a deterministic map and $\mathcal{Y}^{n+\Delta}$ is the $(n+\Delta)$-fold Cartesian product of $\mathcal{Y}$. The sequence $(f^{(1)}, g^{(1)}), (f^{(2)}, g^{(2)}), \ldots$ is called an (M, Δ)-*code*.

In this paper, we quantify the performance of an (M, Δ)-code using a novel autoregressive distortion function. Fix $\lambda \geq 0$, and define the distortion between $\mathbb{u}_n := (u_1, u_2, \ldots, u_n)$ and $\hat{\mathbb{u}}_n := (\hat{u}_1, \hat{u}_2, \ldots, \hat{u}_n)$ recursively by

$$\mathsf{d}_n(\hat{\mathbb{u}}_n, \mathbb{u}_n) := \lambda \mathsf{d}_{n-1}(\hat{\mathbb{u}}_{n-1}, \mathbb{u}_{n-1}) + \mathbb{1}\{\hat{\mathbb{u}}_n \neq \mathbb{u}_n\},$$

where $\mathsf{d}_1(\hat{u}_1, u_1) := \mathbb{1}\{\hat{u}_1 \neq u_1\}$ and $\mathbb{1}$ is the indicator function. The recursion provides a form of unequal error protection, where early errors are weighted more than late errors. We will see, in the next section, that this unequal protection is useful for tracking autoregressive processes. Let

$$\langle \mathsf{D}_N \rangle := \mathbb{E}\, \frac{1}{N} \sum_{n=1}^{N} \mathsf{d}_n(\hat{\mathbb{U}}_n^{(n)}, \mathbb{U}_n) \quad \text{and} \quad \langle \mathsf{R}_N \rangle := \frac{N \log M}{N + \Delta}$$

denote the average distortion and rate of the first N stages respectively. The main problem of interest is to determine the set of achievable rate-distortion-delay tuples in the following sense. Our results are summarised in Section II.

Definition 1: A rate-distortion-delay tuple (R, D, Δ) is said to be (N, λ)-*achievable* on a DMC if there exists an (M, Δ)-code such that $R \leq \langle \mathsf{R}_N \rangle$ and $D \geq \langle \mathsf{D}_N \rangle$.

Definition 1 captures tradeoffs between rate, reliability and delay. For example, the decoder initially estimates $\mathbb{U}_n$ from the first $(n + \Delta)$ channel outputs. The decoder then revisits and improves this estimate in each successive decoding stage, as more channel outputs become available. Intuitively, the rate at which the estimate improves is controlled by the distortion coefficient λ — a larger λ will demand a faster rate of improvement. Of course, one expects that increasing λ will lead to compromised achievable rates and delays. This paper seeks a better understanding of these tradeoffs. Definition 1 is related to Sahai's [5] notion of *anytime reliability*, see Appendix A. An application of Definition 1 is given next.

B. Motivation: Algorithmic FX Trading

The FX market is decentralised, massive and operates around the clock. Electronic "matchmaker" platforms connect thousands of buyers and sellers each second, and sophisticated computer algorithms track market movements and implement trading strategies. Some algorithms profit by arbitrage (exploiting small, short-lived, discrepancies in the exchange rates between currencies); some minimise the cost of executing large currency orders (e.g., a large company paying a foreign debt); and others predict market changes arising from external forces, such as government and corporate announcements, news and social network feeds. Profitability often requires short delay (e.g., microsecond) access to market data [1].

Consider a tick-by-tick sequence $Q_0, Q_1, \ldots$ of spot (ask or bid) prices of a currency pair. Let us assume the following:

978-1-4799-5186-4/14/$31.00

the spot prices arrive one-by-one at the matchmaker; the matchmaker (the *transmitter*) needs to stream N consecutive prices (the *source*) to a trader (the *receiver*) over a DMC with decoding delay Δ, and the initial price Q_0 is known to the transmitter and receiver. The DMC serves as a basic model for a noisy point-to-point channel or packet erasures in a congested network. Encoding and decoding is performed in stages, analogous to the setup of Definition 1: In stage n, the transmitter sends $X_n := f^{(n)}(Q_0, \ldots, Q_n)$ over the DMC and the receiver reconstructs $\hat{Q}_n := g^{(n)}(Y_1, \ldots, Y_{n+\Delta})$.

We now use an (M, Δ)-code and Definition 1 to give an achievable upper bound for the average L_1 log distortion, $(1/N)\sum_{n=1}^{N} \mathbb{E}|\log \hat{Q}_n - \log Q_n|$. The n-th spot price can be written as

$$Q_n = Q_{n-1}^{\lambda_n}(1 + \epsilon_n),$$

where ϵ_n is typically small, say $|\epsilon_n| \le 0.01$, and λ_n is close to one. To apply the autoregressive-distortion function d_n, let us consider the logarithmic prices $W_n := \log Q_n$ and

$$W_n = \log\left(Q_{n-1}^{\lambda_n}(1 + \epsilon_n)\right) \approx \lambda_n W_{n-1} + \epsilon_n, \qquad (1)$$

where $\log(1 + \epsilon_n) \approx \epsilon_n$ for small ϵ_n.

If $\epsilon_1, \epsilon_2, \ldots$ are iid random variables and λ_n is constant, then the right-most side of (1) is a first-order autoregressive process. Such processes are widely used in FX modelling because they are simply stated, difficult to outperform in forecasting and difficult to reject [6]. Let us assume that $\epsilon_1, \epsilon_2, \ldots$ are iid (uniform) on a finite set of M real numbers[1].

Log prices exhibit complicated cycles and non-stationary properties, which can be modelled by adapting λ_n. For example, λ_n might be determined by whether or not W_{n-1} sits above or below a threshold, or by the distance of W_{n-1} from some key economic indicator [6], [8]. Assume that $\lambda_n \le \lambda_{\max}$ for all n and that λ_n is known at the transmitter and receiver[2].

Suppose that the transmitter streams $\epsilon_1, \epsilon_2, \ldots$ with delay Δ using an (M, Δ)-code. At stage n, the receiver attempts to reconstruct the first n price changes $(\hat{\epsilon}_1^{(n)}, \hat{\epsilon}_2^{(n)}, \ldots, \hat{\epsilon}_n^{(n)})$. It then estimates W_n by setting $\hat{W}_0 := \log Q_0$ and

$$\hat{W}_n := \left(\prod_{i=1}^{n} \lambda_i\right) W_0 + \sum_{k=1}^{n-1}\left(\prod_{i=k+1}^{n} \lambda_i\right) \hat{\epsilon}_k^{(n)} + \hat{\epsilon}_n^{(n)}, \quad n \ge 1.$$

We can now apply Definition 1 to bound the average L_1 log error: If (R, D, Δ) is $(N, \lambda_{\max})$-achievable, then there exists a code for the log prices such that

$$\begin{aligned}\mathbb{E}\frac{1}{N}\sum_{n=1}^{N}|\hat{W}_n - W_n| &\le c\, \mathbb{E}\frac{1}{N}\sum_{n=1}^{N}\mathsf{d}_n\left((\hat{\epsilon}_1^{(n)} \ldots, \hat{\epsilon}_n^{(n)}), (\epsilon_1^{(n)} \ldots, \epsilon_n^{(n)})\right) \\ &\le c\, \langle \mathsf{D}_N \rangle \\ &\le c\, D,\end{aligned}$$

where $c := \max_{a,b' \in \mathcal{E}} |a - b|$. The reconstructions $\hat{Q}_1, \hat{Q}_2, \ldots$ at the receiver can be made arbitrarily accurate on average by applying Definition 1 with sufficiently small D.

[1]It is often assumed that $\epsilon_1, \epsilon_2, \ldots$ are log normally distributed for long time scales (e.g., day-to-day price changes), by way of the central limit theorem. However, high frequency prices are not log-normally distributed [7].

[2]For example, λ_n changes slowly relative to the tick frequency.

II. Main Results: Achievable Bounds

A. General DMCs

We now given three achievability results for Definition 1. Each result presents an achievable bound for $\langle \mathsf{D}_N \rangle$ for any given N, distortion coefficient λ, source cardinality M and delay Δ (and hence rate $\langle \mathsf{R}_N \rangle$). The following definitions are needed.

Given a pair of discrete random variables (A, B) on $\mathcal{A} \times \mathcal{B}$ with joint pmf $P_{A,B}$ and marginal pmfs P_A and P_B, the *information density* $\imath_{A;B} : \mathcal{A} \times \mathcal{B} \longrightarrow [-\infty, \infty]$ is

$$\imath_{A;B}(a; b) := \log \frac{P_{A,B}(a, b)}{P_A(a)\ P_B(b)}.$$

Fix M and a pmf P_X on $\mathcal{X}$, and consider the tuple

$$(X, Y, \tilde{X}) \sim P_X(x)\ T_{Y|X}(y|x)\ P_X(\tilde{x})$$

on $\mathcal{X} \times \mathcal{Y} \times \mathcal{X}$. For each $k = 1, 2, \ldots$ define

$$\tau(k) := \mathbb{E}\Big[\min\left\{1, M^{k-1}(M - 1)\ \zeta_k(\mathbb{X}_k, \mathbb{Y}_k, \tilde{\mathbb{X}}_k)\right\}\Big], \qquad (2)$$

where $(\mathbb{X}_k, \mathbb{Y}_k, \tilde{\mathbb{X}}_k)$ is a string of k i.i.d. $(X, Y, \tilde{X})$,

$$\zeta_k(\mathbb{X}_k, \mathbb{Y}_k, \tilde{\mathbb{X}}_k) := \mathbb{P}\left[\imath_{\mathbb{X}_k;\mathbb{Y}_k}(\tilde{\mathbb{X}}_k; \mathbb{Y}_k) \ge \imath_{\mathbb{X}_k;\mathbb{Y}_k}(\mathbb{X}_k; \mathbb{Y}_k)\Big|(\mathbb{X}_k, \mathbb{Y}_k)\right]$$

and the expectation in (2) is taken with respect to $(\mathbb{X}_k, \mathbb{Y}_k)$.

Theorem 1: For any N, λ, M, Δ and pmf P_X on $\mathcal{X}$, there exists an (M, Δ)-code with

$$\langle \mathsf{D}_N \rangle \le \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{k=1}^{n}\lambda^{n-k}\min\left\{1, \sum_{i=0}^{k-1}\tau(n + \Delta - i)\right\}\right). \qquad (3)$$

Proof: See Appendix B. ■

The proof of Theorem 1 is a streaming analogue of the *random coding union* bound for block codes [9, Thm. 17]. The bound is computable for reasonably large N on many channels, and we numerically evaluate it for a binary symmetric channel in Section II-B. Theorem 1 is not easy to visualise, so to help understand the various tradeoffs we now present two bounds based on the DMC's random-coding exponent.

Denote the channel capacity of the DMC $T_{Y|X}$ by

$$\mathsf{C} := \max_{P_X \in \mathcal{P}_\mathcal{X}} \mathbb{E}\, \imath_{X;Y}(X; Y),$$

where $\mathcal{P}_\mathcal{X}$ denotes the set of all pmfs on $\mathcal{X}$. The *random-coding exponent* of a DMC is defined by [10, p. 139]

$$\mathsf{E}_\mathrm{r}(R) := \max_{\rho \in [0,1]} \max_{P_X \in \mathcal{P}_\mathcal{X}} \Big[\mathsf{E}_\mathrm{o}(\rho, P_X) - \rho R\Big], \quad 0 \le R \le \mathsf{C},$$

where

$$\mathsf{E}_\mathrm{o}(\rho, P_X) := -\log\left(\sum_{y \in \mathcal{Y}}\left(\sum_{x \in \mathcal{X}} P_X(x)\Big(T_{Y|X}(y|x)\Big)^{\frac{1}{1+\rho}}\right)^{1+\rho}\right).$$

The next two corollaries are proved in Appendix C.

Corollary 1.1: For any N, Δ, $M = \lfloor \exp(R^*) \rfloor$ with $R^* < \mathsf{C}$, and $\lambda < \exp(\mathsf{E}_\mathrm{r}(R^*))$, there exists an (M, Δ)-code with

$$\langle \mathsf{D}_N \rangle \le \left(\frac{(e^{R^*} - 1)\ e^{\mathsf{E}_\mathrm{r}(R^*) - \log(\lambda)}}{e^{R^*}\ (e^{\mathsf{E}_\mathrm{r}(R^*)} - 1)\ (e^{\mathsf{E}_\mathrm{r}(R^*) - \log(\lambda)} - 1)}\right) e^{-\Delta \mathsf{E}_\mathrm{r}(R^*)}.$$

Corollary 1.1 demonstrates that bounded average distortion is achievable for $N \to \infty$ when $\lambda < \exp(\mathsf{E}_\mathrm{r}(R^*))$. However, the bound in Corollary 1.1 is rather weak; for example, it approaches a vertical asymptote as $\lambda \to \exp(\mathsf{E}_r(R^*))$ from below. Indeed, when $\lambda > \exp(\mathsf{E}_r(R^*))$ the bounds leading to Corollary 1.1 (viz. (16)) explode in N for every constant delay Δ. It is also worth noting that the condition $\lambda < \exp(\mathsf{E}_r(R^*))$ effectively constrains $\langle \mathsf{R}_N \rangle$ as a function of λ, since $\mathsf{E}_r(R^*)$ is monotonically decreasing in R^* [10].

The above explosion for $\lambda > \exp(\mathsf{E}_r(R^*))$ can be prevented by allowing the delay to be linear[3] in n; that is, $\Delta(n) := sn + \delta$, for some integers $s, \delta > 0$. Notice that linear delay does not imply vanishing average rate, since $\langle \mathsf{R}_N \rangle \longrightarrow \log M/(1+s)$ as $N \to \infty$. It can be shown that Theorem 1 holds with Δ replaced by $sn + \delta$. The next corollary shows that bounded average distortion is achievable with arbitrarily large N.

Corollary 1.2: For any N, $M = \lfloor \exp(R^*) \rfloor$ with $R^* < \mathsf{C}$, $\lambda > \exp(\mathsf{E}_r(R^*))$, and $\Delta(n) = sn + \delta$ with $\delta > 0$ and

$$s = \left\lceil \frac{\log \lambda - \mathsf{E}_r(R^*)}{\mathsf{E}_r(R^*)} \right\rceil$$

there exists an $(M, \Delta(n))$-code with

$$\langle \mathsf{D}_N \rangle \le \left(\frac{(e^{R^*} - 1)\; e^{\mathsf{E}_r(R^*) - \log \lambda}}{e^{R^*}\; (e^{\mathsf{E}_r(R^*)} - 1)\; (1 - e^{\mathsf{E}_r(R^*) - \log \lambda})} \right) e^{-\delta \mathsf{E}_r(R^*)}.$$

B. Example: Binary Symmetric Channel

Let the DMC $T_{Y|X}$ consists of κ independent parallel binary symmetric channels, each with crossover probability $0 < \varepsilon < 1/2$. We call κ the *bandwidth expansion factor*. Here $\mathcal{X} = \mathcal{Y} = \{0,1\}^\kappa$ and $T_{Y|X}(y|x) = \varepsilon^t\,(1-\varepsilon)^{\kappa - t}$, where $t \le \kappa$ is the Hamming distance between x and y.

Theorem 2: Fix N, λ, M, Δ and choose P_X uniform on $\mathcal{X}$. There exists an (M, Δ)-code such that the average distortion bound (3) holds for $\tau(k)$ given by

$$\tau(k) = \sum_{t=0}^{\kappa k} \tbinom{\kappa k}{t} \varepsilon^t (1-\varepsilon)^{\kappa k - t} \min\Big\{1, \; M^{k-1}(M-1)2^{-\kappa k} \sum_{s=0}^{t} \tbinom{\kappa k}{s} \Big\}.$$

The proof of Theorem 2 is omitted for brevity. Figure 1 depicts the bound (3) as per Theorem 2. Two values of the distortion coefficient, $\lambda = 2^{0.49}$ and $2^{0.52}$, are plotted for three delays $\Delta = 0$, 5 and 10. We fix $\varepsilon = 0.11$, $\kappa = 5$ and $M = 2$, and we notice that $0.49 \ln 2 < \mathsf{E}_r(\ln 2)$. As Corollary 1.1 suggests, the bounds converge for $\lambda = 2^{0.49}$ and explode for $\lambda = 2^{0.52}$.

Remark 1: When the DMC $T_{Y|X}$ consists of κ independent channels, the bounds in Corollaries 1.1 and 1.2 can be tightened via channel splitting and combining [11].

Appendix A
Anytime Capacity and Definition 1

Let us first restate some definitions in [5] using the nomenclature of this paper. Fix $a > 0$. A rate R is said to be *a-anytime reliable* [5] for a DMC if there exists a constant $b > 0$ and an $(M, 0)$-code[4] such that $R \le \log M$ and

$$\mathbb{P}[\hat{\mathbb{U}}_k^{(n)} \neq \mathbb{U}_k] \le b\; 2^{-a(n-k+1)} \quad \forall\, k \ge 1 \text{ and } n \ge k. \tag{4}$$

[3]In this case, Definition 1 will apply to $(M, \Delta(n))$-codes, where the stage-n decoder waits for $(n+\Delta(n))$ channel symbols, and the rate $\langle \mathsf{R}_N \rangle$ is defined with $\Delta(n)$ in place of Δ.

[4]The setup of [5] corresponds to zero-delay codes in this paper.

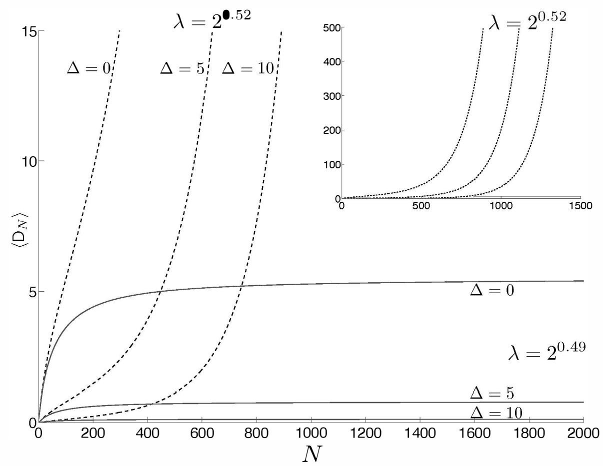

Fig. 1. Achievable bounds on the average distortion $\langle \mathsf{D}_N \rangle$ for the parallel BSC with $M = 2$, $\kappa = 5$ and $\varepsilon = 0.11$. The bounds are plotted as a function of the number of source symbols N (or, equivalently, the number of binary channel symbols κN).

The *anytime exponent*, for a fixed rate R, is

$$\mathsf{E}_{\text{Any}}(R) := \sup\big\{a \ge 0 : R \text{ is } a\text{-anytime reliable}\big\}.$$

The *anytime capacity*, for a fixed exponent a, is

$$\mathsf{C}_{\text{Any}}(a) := \sup\big\{R \ge 0 : \mathsf{E}_{\text{Any}}(R) \ge a\big\}.$$

To relate the anytime setup with Definition 1, it will be useful to tweak Sahai's definition of anytime reliability; specifically, let us require that the error probabilities in (4) decay with an exponent strictly greater than a.

Definition 2: Fix $\alpha \ge 0$. We say that a rate R is *α-anytime achievable* if there exists a constant $b > 0$, an exponent $a > \alpha$, and an $(M, 0)$-code such that (4) holds and $R \le \log M$.

It can be shown that the anytime exponent and capacity both remain unchanged under Definition 2. The next lemma provides a bridge to the setup of Definition 1 in this paper.

Lemma 3: Fix $\alpha \ge 0$. A rate R is *α-anytime achievable* if and only if there exists a constant $b > 0$, an exponent $a > \alpha$ and an $(M, 0)$-code such that $R \le \log M$ and

$$\sum_{n=k}^{\infty} 2^{a(n-k+1)}\; \mathbb{P}[\hat{\mathbb{U}}_k^{(n)} \neq \mathbb{U}_k] \le b, \quad \forall k \ge 1. \tag{5}$$

Proof: The forward *"if"* assertions follows directly from the non-negativity of each term in (5). Consider the reverse *"only if"* assertion. Fix R, and suppose we have an $(M, 0)$-code such that $R \le \log M$ and (4) holds for some $a > \alpha$ and $b > 0$. Pick $a' \in (\alpha, a)$. For each and every k we have

$$\sum_{n=k}^{\infty} 2^{a'(n-k+1)}\; \mathbb{P}[\hat{\mathbb{U}}_k^{(n)} \neq \mathbb{U}_k] \le b \sum_{n=k}^{\infty} 2^{-(a-a')(n-k+1)}, \tag{6}$$

where the RHS of (6) is finite and independent of k. ■

Lemma 3 asserts that the anytime achievability requirement of Definition 2 is equivalent to the boundedness of each and every positive series (indexed by $k = 1, 2, \ldots$) in (5); or, equivalently, there exists $b' > 0$ such that for all N

$$\sum_{n=k}^{N} 2^{a(n-k)}\; \mathbb{P}[\hat{\mathbb{U}}_k^{(n)} \neq \mathbb{U}_k] \le b', \quad \forall\, k \in \{1, 2, \ldots, N\}. \tag{7}$$

On the other hand, $\langle \mathsf{D}_N \rangle$ with $\lambda = 2^a$ can be written as an average version of (7); i.e.,

$$\langle \mathsf{D}_N \rangle = \frac{1}{N} \sum_{k=1}^{N} \left(\sum_{n=k}^{N} 2^{a(n-k)} \ \mathbb{P}[\hat{\mathbb{U}}_k^{(n)} \neq \mathbb{U}_k] \right).$$

The next proposition attempts to quantify the difference between Definitions 1 and 2 by showing that if $\langle \mathsf{D}_N \rangle$ is bounded for all N, then (7) holds for any fraction of the indices $k \in \{1, 2, \ldots, N\}$. The proof is similar to [13, Thm. 18] and is omitted.

Proposition 4: Fix $\gamma > 0$, $0 < \eta < 1$ and $b' = \gamma/(1-\eta)$. If there exists an (M, Δ)-code such that $\langle \mathsf{D}_N \rangle < \gamma \ \forall \ N$, then there exists a sequence of sets $\mathcal{S}_1, \mathcal{S}_2, \ldots$ with $\mathcal{S}_N \subseteq \{1, 2, \ldots, N\}$ and $|\mathcal{S}_N| = \lfloor \eta N \rfloor + 1$ such that (7) holds $\forall \ k \in \mathcal{S}_N$.

Finally, we note that (to the best of our knowledge) no achievability results for *deterministic* codes have been found for the anytime setup.

Appendix B
Proof of Theorem 1

A. M-ary Tree Codes

An (M, Δ)-code can be efficiently represented by an infinite M-ary rooted tree. Label the M edges (to child nodes) of each node with a unique $u \in \mathcal{U}$, and label each node by the unique edge path from the root to that node. For example, $\mathbb{u}_n = (u_1, u_2, \ldots, u_n)$ denotes the depth-n node identified by the edge path $(u_1, u_2, \ldots, u_n)$. An (M, Δ)-code is then an assignment of a channel code symbol to each node in the tree. Let $c(\mathbb{u}_n) \in \mathcal{X}$ denote the code symbol assigned to node $\mathbb{u}_n$. For example, if u_1, u_2 and u_3 are the first three symbols output by the source, then the first three code symbols sent over the channel are $c(u_1), c(u_1, u_2)$ and $c(u_1, u_2, u_3)$. Finally, denote the string of n code symbols on the path to node $\mathbb{u}_n$ by

$$\mathbb{c}(\mathbb{u}_n) = \big(c(\mathbb{u}_1), c(\mathbb{u}_2), \ldots, c(\mathbb{u}_n)\big),$$

where $\mathbb{u}_i$ represents the first i symbols of $\mathbb{u}_n$.

B. A Maximum-Likelihood (ML) Decoder

A key problem in the proof will be to estimate the first k source symbols $\mathbb{U}_k$ from the first n channel output symbols $\mathbb{Y}_n$, where $k \leq n$. We use an ML decoder that works as follows.

Fix an (M, Δ)-code and a pmf P_X on $\mathcal{X}$. Upon observing $\mathbb{y}_n = (y_1, y_2, \ldots, y_n)$ from the channel, the decoder computes the information density of every code string; i.e., it computes

$$\imath_{\mathbb{X}_n;\mathbb{Y}_n}\big(\mathbb{c}(\mathbb{u}_n); \mathbb{y}_n\big) = \sum_{i=1}^{n} \log \frac{T_{Y|X}\big(c(\mathbb{u}_i)|y_i\big)}{\sum_{x' \in \mathcal{X}} P_X(x') \ T_{Y|X}(y_i|x')}$$

for all $\mathbb{u}_n \in \mathcal{U}^n$. The decoder uniformly at random selects a string $\tilde{\mathbb{u}}_n$ from the set

$$\Big\{ \underset{\mathbb{u}_n \in \mathcal{U}^n}{\arg\max} \ \ \imath_{\mathbb{X}_n;\mathbb{Y}_n}\big(\mathbb{c}(\mathbb{u}_n); \mathbb{y}_n\big) \Big\},$$

and it declares the first k symbols of $\tilde{\mathbb{u}}_n$ to be its reconstruction, $\hat{\mathbb{u}}_k := \tilde{\mathbb{u}}_k$.

C. Random Coding

For a given (M, Δ)-code: Suppose that the first $n_\Delta := n + \Delta$ symbols output by the source are $\mathbb{U}_{n_\Delta} = \mathbb{u}_{n_\Delta}$. Let $\xi_k(\mathbb{u}_{n_\Delta})$ denote the conditional probability that the ML decoder *incorrectly* reconstructs the first k symbols of $\mathbb{U}_{n_\Delta}$.

Consider a random experiment in which all symbols of the (M, Δ)-code are selected i.i.d. $\sim P_X$. The code is random and $\xi_k(\mathbb{u}_{n_\Delta})$ is a real-valued random variable. We now upper bound the average conditional error probability for the all-zero source string, $\mathbb{E} \ \xi_k(\mathbb{0}_{n_\Delta}) = \mathbb{E} \ \xi_k(0, 0, \ldots, 0)$, where the expectation is taken over the code and channel. To do so, it is useful to define the *fan* of $\mathbb{u}_k \in \mathcal{U}^k$,

$$\text{fan}\,(\mathbb{u}_k) := \big\{\tilde{\mathbb{u}}_{n_\Delta} \in \mathcal{U}^{n_\Delta} : \mathbb{u}_k = \tilde{\mathbb{u}}_k\big\}.$$

If $\mathbb{c}(\mathbb{0}_{n_\Delta})$ is transmitted and $\mathbb{y}_{n_\Delta}$ observed, then a necessary condition for the ML decoder to make an error is that there exists a 'bad' source string $\mathbb{u}_{n_\Delta} \notin \text{fan}\,(\mathbb{0}_k)$ such that

$$\imath_{\mathbb{X}_{n_\Delta};\mathbb{Y}_{n_\Delta}}\big(\mathbb{c}(\mathbb{u}_{n_\Delta}); \mathbb{y}_{n_\Delta}\big) \geq \imath_{\mathbb{X}_{n_\Delta};\mathbb{Y}_{n_\Delta}}\big(\mathbb{c}(\tilde{\mathbb{u}}_{n_\Delta}); \mathbb{y}_{n_\Delta}\big)$$

holds for every string $\tilde{\mathbb{u}}_{n_\Delta} \in \text{fan}\,(\mathbb{0}_k)$. Taking in account randomness in the channel and code, we have

$$\begin{aligned}\mathbb{E}\xi_k(\mathbb{0}_{n_\Delta}) &\overset{(a)}{\leq} \mathbb{P}\Bigg[\bigcup_{\mathbb{u}_{n_\Delta} \notin \text{fan}(\mathbb{0}_k)} \Bigg(\bigcap_{\tilde{\mathbb{u}}_{n_\Delta} \in \text{fan}(\mathbb{0}_k)} \Big\{\imath\big(\mathbb{C}(\mathbb{u}_{n_\Delta}); \mathbb{Y}_{n_\Delta}\big) \\ &\qquad\qquad \geq \imath\big(\mathbb{C}(\tilde{\mathbb{u}}_{n_\Delta}); \mathbb{Y}_{n_\Delta}\big)\Big\}\Bigg)\Bigg] \\ &\overset{(b)}{\leq} \mathbb{P}\Bigg[\bigcup_{\mathbb{u}_{n_\Delta} \notin \text{fan}(\mathbb{0}_k)} \Big\{\imath\big(\mathbb{C}(\mathbb{u}_{n_\Delta}); \mathbb{Y}_{n_\Delta}\big) \geq \imath\big(\mathbb{C}(\mathbb{0}_{n_\Delta}); \mathbb{Y}_{n_\Delta}\big)\Big\}\Bigg] \quad (8)\end{aligned}$$

where (b) considers only the all-zero source string in the intersection. In both (a) and (b), we use uppercase $\mathbb{C}(\mathbb{u}_{n_\Delta})$ to emphasise that the code is random, and that we have omitted the information density subscripts $\mathbb{X}_{n_\Delta}, \mathbb{Y}_{n_\Delta}$.

If we define the set difference

$$\mathcal{K}_i := \begin{cases} \mathcal{U}^{n_\Delta} \backslash \text{fan}\,(\mathbb{0}_1)\,, & \text{for } i = 0 \\ \text{fan}\,(\mathbb{0}_i) \setminus \text{fan}\,(\mathbb{0}_{i+1})\,, & \text{for } i \geq 1, \end{cases}$$

then (8) can be written as

$$\begin{aligned}\mathbb{E}\xi(\mathbb{0}_{n_\Delta}) \leq \mathbb{P}\Bigg[\bigcup_{i=0}^{k-1} \Bigg(\bigcup_{\mathbb{u}_{n_\Delta} \in \mathcal{K}_i} \Big\{\imath\big(\mathbb{C}(\mathbb{u}_{n_\Delta}); \mathbb{Y}_{n_\Delta}\big) \\ \geq \imath\big(\mathbb{C}(\mathbb{0}_{n_\Delta}); \mathbb{Y}_{n_\Delta}\big)\Big\}\Bigg)\Bigg].\end{aligned}$$

Now apply the union bound

$$\begin{aligned}\mathbb{E}\xi(\mathbb{0}_{n_\Delta}) \leq \min\Bigg\{1, \sum_{i=0}^{k-1} \mathbb{P}\Bigg[\bigcup_{\mathbb{u}_{n_\Delta} \in \mathcal{K}_i} \Big\{\imath\big(\mathbb{C}(\mathbb{u}_{n_\Delta}); \mathbb{Y}_{n_\Delta}\big) \\ \geq \imath\big(\mathbb{C}(\mathbb{0}_{n_\Delta}); \mathbb{Y}_{n_\Delta}\big)\Big\}\Bigg]\Bigg\}. \quad (9)\end{aligned}$$

We now notice that for every source string $\mathbb{u}_{n_\Delta} \in \mathcal{K}_i$, the first i symbols of the corresponding code string $\mathbb{C}(\mathbb{u}_{n_\Delta})$ are identical to the first i symbols of $\mathbb{C}(\mathbb{0}_{n_\Delta})$. That is,

$$\mathbb{C}_{1,i}(\mathbb{u}_{n_\Delta}) = \mathbb{C}_{1,i}(\mathbb{0}_{n_\Delta}), \quad \forall \ \mathbb{u}_{n_\Delta} \in \mathcal{K}_i,$$

where we denote

$$\mathbb{C}_{j,j'}(\mathbb{u}_{n_\Delta}) := \big(\mathbb{C}(\mathbb{u}_j), \mathbb{C}(\mathbb{u}_{j+1}), \ldots, \mathbb{C}(\mathbb{u}_{j'})\big).$$

Thus, the error event due to $\mathcal{K}_i$ in (9) depends only on the last $(n_\Delta - i)$ symbols of $\mathbb{C}(\mathbb{u}_{n_\Delta})$ and $\mathbb{C}(\mathbb{0}_{n_\Delta})$,

$$\mathbb{E}\xi(\mathbb{0}_{n_\Delta}) \le \min\left\{1, \sum_{i=0}^{k-1} \mathbb{P}\left[\bigcup_{\mathbb{u}_{n_\Delta}\in\mathcal{K}_i} \Big\{\imath\big(\mathbb{C}_{i+1,n_\Delta}(\mathbb{u}_{n_\Delta}); \mathbb{Y}_{i+1,n_\Delta}\big) \ge \imath\big(\mathbb{C}_{i+1,n_\Delta}(\mathbb{0}_{n_\Delta}); \mathbb{Y}_{i+1,n_\Delta}\big)\Big\}\right]\right\}. \quad (10)$$

We now invoke a technique used by Polyanskiy in [9, Thm. 17]: Rewrite the probability in (10) as an expectation of a conditional probability, where the expectation is taken with respect to $(\mathbb{C}_{i+1,n_\Delta}(\mathbb{0}_{n_\Delta}), \mathbb{Y}_{i+1,n_\Delta})$. Doing so gives the bound in (11), and (12) follows by another application of the union bound. The probability term in (12) is the same for all $\mathbb{u}_{n_\Delta} \in \mathcal{K}_i$ and $|\mathcal{K}_i| = (M-1)M^{n_\Delta - i - 1}$, which leads to

$$\mathbb{E}\ \xi(\mathbb{0}_{n_\Delta}) \le \min\Big\{1, \sum_{i=0}^{k-1} \tau(n_\Delta - i)\Big\}, \quad (13)$$

where $\tau(\cdot)$ is defined in (2).

The bound (13) applies to the all-zero string $\mathbb{0}_{n_\Delta}$. By symmetry, the same bound holds for all $\mathbb{u}_{n_\Delta} \in \mathcal{U}^{n_\Delta}$. Moreover the source is uniform, so the average error probability satisfies

$$\mathbb{P}[\hat{\mathbb{U}}_k^{(n)} \ne \mathbb{U}_k] \le \min\Big\{1, \sum_{i=0}^{k-1} \tau(n_\Delta - i)\Big\}. \quad (14)$$

D. Completing the proof

The average distortion (averaged over the code, source and channel) satisfies

$$\langle \mathsf{D}_N \rangle = \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{k=1}^{n} \lambda^{n-k}\ \mathbb{P}[\hat{\mathbb{U}}_k^{(n)} \ne \mathbb{U}_k]\right)$$
$$\overset{(*)}{\le} \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{k=1}^{n} \lambda^{n-k}\ \min\left\{1, \sum_{i=0}^{k-1}\tau(n_\Delta - i)\right\}\right), \quad (15)$$

where (*) follows from (14). To complete the proof, we note that there must exist at least one deterministic (M,Δ)-code in the ensemble for which the bound (15) holds. ■

Appendix C
Proof of Corollaries 1.1 and 1.2

The next lemma can be distilled from Polyanskiy [9, p. 25] and Gallager [12, p. 5], we omit the details.

Lemma 5: Fix $\rho \in [0,1]$, $R^* < \mathsf{C}$, $M = \lfloor \exp(R^*) \rfloor$, and a pmf P_X on $\mathcal{X}$. We have

$$\tau(k) \le \left(\frac{e^{R^*}-1}{e^{R^*}}\right)\ e^{-k(\mathsf{E}_\text{o}(\rho,P_X) - \rho R^*)}, \quad \forall\ k \ge 1.$$

A. Proof Outline for Corollary 1.1

Fix N, P_X, Δ, $\rho \in [0,1]$, $\lambda = e^\alpha$, $M = \lfloor e^{R^*} \rfloor$. We have

$$\langle \mathsf{D}_N \rangle \le \frac{e^{R^*}-1}{e^{R^*}}\ \frac{1}{N}\sum_{n=1}^{N}\left(\sum_{k=1}^{n} e^{\alpha(n-k)} \left(\sum_{i=0}^{k-1} e^{-(n+\Delta-i)(\mathsf{E}_\text{o}(\rho,P_X)-\rho R^*)}\right)\right), \quad (16)$$

from Theorem 1 and Lemma 5. Selecting P_X and ρ in (16) to maximise $\mathsf{E}_\text{o}(\rho,P_X) - \rho R^*$ gives

$$\langle \mathsf{D}_N \rangle \le \frac{e^{R^*}-1}{e^{R^*}}\frac{1}{N}\sum_{n=1}^{N}\left(\sum_{k=1}^{n} e^{\alpha(n-k)}\Big(\sum_{i=0}^{k-1} e^{-(n+\Delta-i)\mathsf{E}_\text{r}(R^*)}\Big)\right).$$

The corollary follows from the above geometric sums. ■

B. Proof Outline for Corollary 1.2

Substitute $\Delta(n) = sn + \delta$ in (16). Pick ρ and P_X to maximise $\mathsf{E}_\text{o}(\rho,P_X) - \rho R^*$ and evaluate the geometric sums. ■

Acknowledgements

This work was supported by the city of Paris programme "Emergences" and the ARC Discovery Project DP120102123. Roy Timo would like to thank Thuy Ngoc Tran for optimising the algorithm used to plot Theorem 2 for large κN.

References

[1] D. Schneider, "The microsecond market," *IEEE Spectr.*, vol. 49, no. 6, 2012.

[2] P. Alexander, D. Haley, and A. Grant, "Cooperative intelligent transport systems: 5.9-GHz field trials," *Proc. IEEE*, vol. 99, no. 7, 2011.

[3] F. Poletti, N. V. Wheeler, M. N. Petrovich, N. Baddela, E. Numkam Fokoua, J. R. Hayes, D. R. Gray, Z. Li, R. Slavk, and D. J. Richardson, "Towards high-capacity fibre-optic communications at the speed of light in vacuum," *Nature Photon.*, vol. 7, 2013.

[4] A. Chaboud, E. Hjalmarsson, C. Vega, and B. Chiquoine, "Rise of the machines: algorithmic trading in the foreign exchange market," *Board of Governors of the Federal Reserve System, International Finance Discussion Paper*, no. 980, 2009.

[5] 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, no. 8, 2006.

[6] L. Kiliana and M. Taylor, "Why is it so difficult to beat the random walk forecast of exchange rates?" *Journal Intl. Economics*, vol. 60, 2003.

[7] R. Cont, "Empirical properties of asset returns: stylized facts and statistical issues," *Quantitative Finance*, vol. 1, 2001.

[8] R. S. Tsay, *Analysis of financial time series*, Wiley series in probability and statistics ed. Wiley, 2005, vol. 543.

[9] Y. Polyanskiy, "Channel coding: non-asymptotic fundamental limits," Thesis, Princeton University, 2010.

[10] R. Gallager, *Information Theory and Reliable Communication.* John Wiley and Sons, Inc. New York, NY, USA, 1968.

[11] E. Arikan, "Channel combining and splitting for cutoff rate improvement," *IEEE Trans. Inform. Theory*, vol. 52, no. 2, 2006.

[12] R. Gallager, "A simple derivation of the coding theorem and some applications," *IEEE Trans. Inform. Theory*, vol. 11, no. 1, 1965.

[13] S. Verdú, "Lecture notes on information theory 3, fundamental limits: channel coding," Princeton University, Fall, 2011.

$$\mathbb{E}\xi(\mathbb{0}_{n_\Delta}) \le \min\left\{1, \sum_{i=0}^{k-1} \mathbb{E}\left[\mathbb{P}\left[\bigcup_{\mathbb{u}_{n_\Delta}\in\mathcal{K}_i}\Big\{\imath\big(\mathbb{C}_{i+1,n_\Delta}(\mathbb{u}_{n_\Delta}); \mathbb{Y}_{i+1,n_\Delta}\big) \ge \imath\big(\mathbb{C}_{i+1,n_\Delta}(\mathbb{0}_{n_\Delta}); \mathbb{Y}_{i+1,n_\Delta}\big)\Big\}\,\Big|\,\big(\mathbb{C}_{i+1,n_\Delta}(\mathbb{0}_{n_\Delta}), \mathbb{Y}_{i+1,n_\Delta}\big)\right]\right]\right\} \quad (11)$$

$$\le \min\left\{1, \sum_{i=0}^{k-1} \mathbb{E}\left[\min\Big\{1, \sum_{\mathbb{u}_{n_\Delta}\in\mathcal{K}_i} \mathbb{P}\Big[\imath\big(\mathbb{C}_{i+1,n_\Delta}(\mathbb{u}_{n_\Delta}); \mathbb{Y}_{i+1,n_\Delta}\big) \ge \imath\big(\mathbb{C}_{i+1,n_\Delta}(\mathbb{0}_{n_\Delta}); \mathbb{Y}_{i+1,n_\Delta}\big)\,\Big|\,\big(\mathbb{C}_{i+1,n_\Delta}(\mathbb{0}_{n_\Delta}), \mathbb{Y}_{i+1,n_\Delta}\big)\Big]\Big\}\right]\right\} \quad (12)$$