Delays and the Capacity of Continuous-time Channels

Sanjeev Khanna*
Dept. of Computer and Information Science
University of Pennsylvania
Philadelphia, PA
Email: sanjeev@cis.upenn.edu

Madhu Sudan
Microsoft Research New England
Cambridge, MA
Email: madhu@mit.edu

Abstract— Any physical channel of communication offers two potential reasons why its capacity (the number of bits it can transmit in a unit of time) might be unbounded: (1) (Uncountably) infinitely many choices of signal strength at any given instant of time, and (2) (Uncountably) infinitely many instances of time at which signals may be sent. However channel noise cancels out the potential unboundedness of the first aspect, leaving typical channels with only a finite capacity per instant of time. The latter source of infinity seems less extensively studied. A potential source of unreliability that might restrict the capacity also from the second aspect is "delay": Signals transmitted by the sender at a given point of time may not be received with a predictable delay at the receiving end. In this work we examine this source of uncertainty by considering a simple discrete model of delay errors. In our model the communicating parties get to subdivide time as microscopically finely as they wish, but still have to cope with communication delays that are macroscopic and variable. The continuous process becomes the limit of our process as the time subdivision becomes infinitesimal. We taxonomize this class of communication channels based on whether the delays and noise are stochastic or adversarial; and based on how much information each aspect has about the other when introducing its errors. We analyze the limits of such channels and reach somewhat surprising conclusions: The capacity of a physical channel is finitely bounded only if at least one of the two sources of error (signal noise or delay noise) is adversarial. In particular the capacity is finitely bounded only if the delay is adversarial, or the noise is adversarial and acts with knowledge of the stochastic delay. If both error sources are stochastic, or if the noise is adversarial and independent of the stochastic delay, then the capacity of the associated physical channel is infinite!

Keywords-Communication, Delays, Physical channels, Adversarial errors, Stochastic errors.

1. Introduction

It seems to be a folklore assumption that any physical medium of communication is constrained to communicating a finite number of bits per unit of time. This assumption forms the foundations of both the theory of communication [7] as well as the theory of computing [9]. The assumption also seems well-founded given the theory of signal processing. In particular the work of Shannon [8] explains reasons why such a statement may be true.

Any physical channel (a copper wire, an optical fiber, vaccuum etc.) in principle can be used by a *sender* to transmit a signal, i.e., a function $f : [0, T] \to [0, 1]$ for some time duration T. The *receiver* receives some function $\tilde{f} : [0, T] \to [0, 1]$, which tends to be a noisy, distorted version of the signal f. The goal of a communication system is to design encoders and decoders that communicate reliably over this channel. Specifically, one would like to find the largest integer k_T such that there exist functions $E : \{0, 1\}^{k_T} \to \{f : [0, T] \to [0, 1]\}$ and $D : \{\tilde{f} : [0, T] \to [0, 1]\} \to \{0, 1\}^{k_T}$ such that $\mathbf{Pr}_{\tilde{f}|E(m)}[D(\tilde{f}) \neq m] \to 0$ where m is chosen uniformly from $\{0, 1\}^{k_T}$ and $\tilde{f}$ is chosen by the channel given the input signal $E(m)$. The *capacity of the channel*, normalized per unit of time, is the $\limsup_{T\to\infty} k_T/T$.

In a typical such channel there are two possible sources of infinite capacity. The number of possible signal values $f(t)$, for any $t \in [0, T]$ is uncountably large and if the channel were not "noisy" this would lead to infinite capacity, even if time were discrete. But Shannon, in his works [7], [8], points out that usually $f(t)$ is not transmitted as is. Typical channels tend to add noise, typically a random function $\eta(t)$, which is modeled as a normally distributed random variable with mean zero and variance σ^2, and independent across different instances of time t. He points out that after this noise's effect is taken into account, the channel capacity is reduced to a finite number (proportional to $1/\sigma^2$) per instant of time.

Still this leaves a second possible way the channel capacity could be infinite, namely due to the availability of infinitely many time slots. This aspect has been considered before in the signal processing literature, and the works of Nyquist [6] and Hartley [4] (see the summary in [8]) once again point out that there is a finite limit. However the reason for this finite limit seems more axiomatic than physical. Specifically, these results come from the assumption that the signal f is a linear combination of a finite number of basis functions, where the basis functions are sinusoids with frequency that is an integral multiple of some minimal frequency, and upper bounded by some maximum frequency. This restriction is then translated into a "discretization" result showing it suffices to sample the signal at certain discrete time intervals, reducing the problem thus to a finite

* Supported in part by NSF Awards CCF-0635084 and IIS-0904314.

one.

In this work we attempt to explore the effects of "continuous time" more in the spirit of the obstacle raised in the context of the signal strength, namely that there is an obstacle also to assuming that time is preserved strictly accross the communication channel. We do so by introducing and studying a "delay channel" where signals transmitted by the sender arrive somewhat asynchronously at the receiver's end. We model and study this process as the limit of a discrete process.

In our discrete model the sender/receiver get to discretize time as finely as they wish, but there is uncertainty/unreliability associated with the delay between when a signal is sent and when it is received. Thus in this sense, there is timing noise, that is similar in spirit to the signal noise. A signal that is sent at time t is received at time $t+\eta(t)$ where $\eta(t)$ could be a random, or adversarial, amount of delay, but whose typical amount is a fixed constant (independent of the granularity of the discretization of time chosen by sender/receiver). Note that this could permute the bits in the sequence sent by the sender (or do more complex changes). We consider the effect of this delay on the channel capacity. For the sake of simplicity (and since this is anyway without loss of generality) we assume sender only sends a sequence of 0s and 1s. In addition to delays we also allow the channel to inject the usual noise.

We discuss our model and results more carefully in Section 2, but let us give a preview of the results here. It turns out that the question of when is the channel capacity finite is a function of several aspects of the model. Note there are two sources of error - the *signal error*, which we simply refer to as *noise*, and the *timing error*, which we refer to as *delay*. As either of these error sources could be probabilistic or adversarial, we get four possible channel models. Complicating things further is the dependence between the two – does either of the sources of error know about the error introduced by the other? Each setting ends up requiring a separate analysis. We taxonomize the many classes of channels that arise this way, and characterize the capacity of all the channels. The final conclusion is the following: If the delays are adversarial, or if the delay is stochastic and the noise is adversarial and acts with knowledge of the delay, then the channel capacity is finite (Theorem 2.1), else it is infinite (Theorem 2.2). In particular if both sources are adversarial then the channel capacity is finite; and perhaps most surprisingly and possibly the most realistic setting, if both sources are probabilistic, then the channel capacity is infinite: finer discretization always leads to increased capacity.

Organization: Section 2 formally describes our model and results. In Sections 3 and 4, we prove our results for finite and infinite channel capacity regimes respectively. Finally, we give some concluding thoughts in Section 5.

2. Preliminaries, Model, and Results

2.1. Continuous channels

We start by describing the basic entities in a communication system and how performance is measured. Most of the definitions are "standard"; the only novelty here is that we allow sender/receiver to choose the "granularity" of time. We first start with the standard definitions.

Channel(Generic): Given a fixed period of time T, a *signal* is a function $f : [0,T] \to \mathbb{R}$. We say the signal is bounded if its range is $[0,1]$. A time T (bounded-input) channel is given by a (possibly non-deterministic, possibly probabilistic, or a combination) function $\mathrm{channel}_T : f \mapsto \tilde{f}$ whose inputs is a bounded signal $f : [0,T] \to [0,1]$ and output is a signal $\tilde{f} : [0,T] \to \mathbb{R}$.

A *probabilistic channel* is formally given by a transition probability distribution which gives the probability of outputting $\tilde{f}$ given input f. An *adversarial channel* is given by a set of possible functions $\tilde{f}$ for each input f. We use $\tilde{f} = \mathrm{channel}_T(f)$ as shorthand for $\tilde{f}$ drawn randomly from the distribution specified by $\mathrm{channel}_T(f)$ in the case of probabilistic channels. For adversarial channels, we use the same notation $\tilde{f} = \mathrm{channel}_T(f)$ as shorthand for $\tilde{f}$ chosen adversarially (so as to minimize successful communication) from $\mathrm{channel}_T(f)$.

Channels can be composed naturally, leading to interesting mixes of adversarial and stochastic channels, which will lead to interesting scenarios in this work.

Encoder/Decoder: Given T and message space $\{0,1\}^{k_T}$, a time T encoder is a function $E : m \mapsto f$ where $m \in \{0,1\}^{k_T}$ and $f : [0,T] \to [0,1]$. Given T and message space $\{0,1\}^{k_T}$, a time T decoder is a function $D : f \mapsto m$ where $f : [0,T] \to \mathbb{R}$ and $m \in \{0,1\}^{k_T}$. (More generally, encoders, channels, and decoders should form composable functions.)

Success Criteria, Rate and Capacity: The decoding error probability of the system $(E_T, D_T, \mathrm{channel}_T)$ is the quantity $\mathbf{Pr}_{\mathrm{dec,T}}$ given by

$$\mathbf{Pr}_{m \leftarrow \{0,1\}^{k_T}, \mathrm{channel}}[m \neq D_T(\mathrm{channel}_T(E_T(m)))].$$

We say that the communication system is *reliable* if $\lim_{T\to\infty}\{\mathbf{Pr}_{\mathrm{dec,T}}\} = 0$.

The (asymptotic) *rate* of a communication system is the limit $\limsup_{T\to\infty}\{k_T/T\}$. The *capacity of a channel*, denoted by Cap, is defined to be the supremum of the rate of the communication system over all encoding/decoding schemes.

2.2. Channel models

We now move to definitions specific to our paper. We study continuous channels as a limit of discrete channels. To make the study simple, we restrict our attention to channels whose signal strength is already discretized, and indeed we will even restrict to the case where the channel only

transmits bits. The channel will be allowed to err, possibly probabilistically or adversarially, and ε will denote the error parameter.

We now move to the more interesting aspect, namely the treatment of time. Our model allows the sender and receiver to divide every unit of time into tiny subintervals, which we call *micro-intervals*, of length $\mu = 1/M$ (for some large integer M), and send arbitrary sequences of M bits per unit of time. This granularity is compensated for by the fact that the channel is allowed to introduce relatively large, random/adversarial, delays. However the channel is allowed to introduce uncertain delays into the system, where the delays average to some fixed constant Δ which is independent of μ. Given that all aspects are scalable, we scale time so that $\Delta = 1$. Again we distinguish between the adversarial case and the probabilistic case. In the adversarial case every transmitted symbol may be delayed by up to 1 unit of time (or by up to M microintervals). In the probabilistic case every transmitted symbol may be delayed by an amount which is a random variable distributed exponentially with mean 1. Finally, if multiple symbols end up arriving at the receiver at the same instant of time, we assume the receiver receives the sum of the value of the arriving symbols.

We now formally describe the model.

Encoding: For every T, the sender encodes k_T bits as MT bits by applying an encoding function $E_T : \{0,1\}^{k_T} \to \{0,1\}^{MT}$. The encoded sequence is denoted $X_1, \ldots, X_{MT}$.

Noise: The noise is given by a function $\xi : [MT] \to \{0,1\}$. The effect of the noise is denoted by the sequence $Z_1, \ldots, Z_{MT}$, where $Z_j = X_j \oplus \xi(j)$. (We stress that Z_j's are not necessarily "seen" by any physical entity — we just mention them since the notation is useful. Also, the $\oplus$ is merely a convenient notation and is not meant to suggest that the bits are elements of some finite field. We will be thinking of the bits as integers.)

Throughout this paper, when considering random noise at some rate $\varepsilon \in (0,1)$, we assume that for every packet released by the sender, the bit encoded in the packet is flipped with probability ε, independently of other packets. Similarly, when considering adversarial noise at some rate $\varepsilon \in (0,1)$, we assume that the noise adversary can flip encoded bits in up to ε-fraction of packets released by the sender.

Delay: The delay is modeled by a delay function $\Delta : [MT] \to \mathbb{Z}^{\geq 0}$ where $\mathbb{Z}^{\geq 0}$ denotes the non-negative integers.

Throughout this paper, when considering random delays, we assume that the packets delays are geometrically distributed with mean 1 – a packet released by the sender is transmitted over the channel in any micro-interval with probability $1/M$. Similarly, when the delay is adversarial, we assume that any packet released by the sender is transmitted by the delay adversary at an arbitrarily chosen micro-interval among the M micro-intervals that follow the release of the packet.

Received Sequence: The final sequence received by the receiver, on noise ξ and delay Δ, is the sequence $Y_1, \ldots, Y_{MT}$, where $Y_i = \sum_{j \leq i \text{ s.t. } j+\Delta(j)=i} Z_j$ and $Z_j = X_j \oplus \xi(j)$.

Decoding: The decoder is thus a function $D_T : (\mathbb{Z}^{\geq 0})^{MT} \to \{0,1\}^{k_T}$.

Note that while the notation suggests that the noise operates on the input first, and then the delay acts on it, we do not view this as an operational suggestion. Indeed the order in which these functions (ξ and Δ) are chosen will be crucial to our results.

Our channels are thus described as a composition of two channels, the noise-channel with parameter ε, denoted $N(\varepsilon)$ and the delay-channel D. Since each of these can be probabilistic or adversarial, this gives us four options. Furthermore a subtle issue emerges which is: Which channel goes first? Specifically if exactly one of the channels is adversarial, then does it get to choose its noise/delay before or after knowing the randomness of the other channel. We allow both possibilities which leads syntactically to eight possible channels (though only six of these are distinct).

Notation: We use D to denote the delay channel and $N(\varepsilon)$ to denote the noise channel with parameter ε. We use superscripts of A or P to denote adversarial or probabilistic errors respectively. We use the notation $X|Y$ to denote that channel X goes first and then Y acts (with knowledge of the effects of X). Thus the eight possible channels we consider are $N^P|D^P$, $D^P|N^P$, $D^A|N^P$, $N^A|D^P$, $D^P|N^A$, $N^P|D^A$, $N^A|D^A$, and $D^A|N^A$.

2.3. Our results

Given that the adversarial channels are more powerful than the corresponding random channels, and an adversary acting with more information is more powerful than one acting with less, some obvious bounds on the capacity of these channels follow:

$$\begin{aligned} \mathsf{Cap}(D^A|N^A) &= \mathsf{Cap}(N^A|D^A) \leq \mathsf{Cap}(N^P|D^A) \\ &\leq \mathsf{Cap}(D^A|N^P) \leq \mathsf{Cap}(D^P|N^P) \ , \quad (1) \end{aligned}$$

and

$$\begin{aligned} \mathsf{Cap}(D^A|N^A) &\leq \mathsf{Cap}(D^P|N^A) \leq \mathsf{Cap}(N^A|D^P) \\ &\leq \mathsf{Cap}(N^P|D^P) = \mathsf{Cap}(D^P|N^P) \ . \quad (2) \end{aligned}$$

The equalities above occur because if both channels are adversarial, or both are probabilistic, then ordering is unimportant.

Our main results are summarized by the following two theorems.

Theorem 2.1 (Finite Capacity Case): For every positive ε, the capacity of the channels $D^A|N(\varepsilon)^A$,$D^P|N(\varepsilon)^A$, $D^A|N(\varepsilon)^P$,and $N(\varepsilon)^P|D^A$ are finite. That is, for every one of these channel types, and $\varepsilon > 0$ there exists a capacity

$C < \infty$ such that for every μ, a μ-discretized encoder and decoder with rate $R > C$, there exists a $\gamma > 0$ such that the probability of decoding error $\mathbf{Pr}_{\text{dec}} \geq \gamma$.

Theorem 2.2 (Infinite Capacity Case): There exists a positive ε such that the capacity of the channels $D^P|N(\varepsilon)^P$ and $N(\varepsilon)^A|D^P$ are infinite. That is, for every one of these channel types, there exists an $\varepsilon > 0$ such that for every finite R, there exists a μ and a μ-discretized encoder and decoder achieving rate R, with decoding error probability $\mathbf{Pr}_{\text{dec}} \to 0$.

The theorems above completely characterize the case where the capacity is infinite. The theorems show that the capacity is infinite if either both channels are probabilistic (the most benign case) or if the noise is adversarial but acts without knowledge of the randomness of the probabilistic delay. On the other hand, the channel capacity is finite if the delay is adversarial, or if the noise is adversarial and acts with knowledge of the probabilistic delay.

Relying on the "obvious" inequalities given earlier, it suffices to give two finiteness bounds and one "infiniteness" bound to get the theorems above, and we do so in the next two sections. Theorem 2.1 follows immediately from Lemmas 3.2 and 3.3 (when combined with Equations (1) and (2)). Theorem 2.2 follows immediately from Lemma 4.1 (again using Equations (1) and (2)).

3. Finite Capacity Regime

In this section we prove that the capacity of our channels are finite, when the delay channel is adversarial (and acts without knowledge of the noise) or when the noise is adversarial and acts with knowledge of the delay. We consider the case of the adversarial noise first, and then analyze the case of the random errors adversarial delay first, and then consider the case of the adversarial noise. In both cases we use a simple scheme to show the capacity is limited. We show that with high probability, the channel can force the receiver to receive one of a limited number of signals.

The following simple lemma is then used to lower bound the probability of error.

Lemma 3.1: Consider a transmission scheme with the sender sending message from a set S with encoding scheme E, where the channel channel can select a set R of receiver signals such that

$$\mathbf{Pr}_{m \in S, \text{channel}}[\text{channel}(E(m)) \notin R] \leq \tau,$$

then the probability of decoding error is at least $1 - (\tau + |R|/|S|)$.

Proof: Let $\tilde{R}$ denote the space of all received signals and fix the decoding function $D : \tilde{R} \to S$. Now consider the event that a transmitted message m is decoded correctly. We claim this event occurs only if only of the two events listed below occur:

1) $\text{channel}(E(m)) \notin R$, which happens with probability at most τ.
2) $m = D(r)$ for some $r \in R$, which happens with probability at most $|R|/|S|$.

If neither of the events listed above occur then the received signal $r \in R$ and $D(r) \neq m$ implying the decoding is incorrect. The lemma follows immediately. ■

3.1. Random Delay followed by Adversarial Noise ($D^P|N^A$)

Here we consider the case where the delays are random, with expectation 1, and the noise is adversarial. In this section, it is useful to view the delay channel as a queueing system, under the noise channel's active control. To explain the queueing system, notice that geometric delays lead to a memoryless queue. At each microinterval of time, a packet enters the queue (the new bit sent by the sender). And then each packet in the queue chooses to depart, independent of other packets, with probability μ. (Note that the geometric delay/memorylessness renders the packets in the queue indistinguishable in terms of their arrival times.)

For the noise channel also, we will adopt a slightly different view. In principle, it is capable of looking at the entire sequence of bits in the order in which they depart the queue, and then decide which ones to flip. However our adversary will be much milder. It will divide time into small intervals with the total number of intervals being $N = O(T/\varepsilon)$. In each interval it will "hold" most arriving packets, releasing only those that are supposed to leave the queue. If packets are released during the interval, the noise adversary sets their value to 0. With the remaining packets it inserts them into the queue, with an integer multiple of $\varepsilon M/c$ of them being set to 1 (and flipping a few bits to 0 in the process as needed) for some constant $c = c(\varepsilon)$. The remaining departures from the queue will then be transmitted untampered to the receiver. We will show that this departure process can be simulated by just the knowledge of the number of 1s injected into the queue at the end of each interval, and the number of possibilities is just $(c+1)^N = c(\varepsilon)^{O(T/\varepsilon)})$ which is independent of M. The adversary will be able to carry out its plan with probability $1-\exp(-T)$, giving us the final result. The following lemma and proof formalize this argument.

Lemma 3.2: For every positive ε, there exists a capacity $C = C(\varepsilon)$ such that the capacity of the channel $D^P|N(\varepsilon)^A$ is bounded by C. Specifically, for every rate $R > C$, for every M (and $\mu = 1/M$), every T, every $k_T > R \cdot T$, and every pair of encoding/decoding functions $E_T : \{0,1\}^{k_T} \to \{0,1\}^{MT}$ and $D_T : (\mathbb{Z}^{\geq 0})^{MT} \to \{0,1\}^{k_T}$, the decoding error probability $\mathbf{Pr}_{\text{dec}} = 1 - \exp(-T)$.

Proof: We start with a formal description of the channel action, and then proceed to analyze the probability of decoding error and channel capacity.

Channel action: Let $X_1, \ldots, X_{MT}$ denote the MT bit string being sent be the sender. We will use $Z_j = X_j \oplus \xi(j)$ to denote the value of the jth bit after noise (even though the noise acts after the delay and so Z_j may not be the jth bit received by the receiver). We let $\Delta : [MT] \to \mathbb{Z}^{\geq 0}$ denote the delay function.

Let $\varepsilon' = \varepsilon/5$ and $L = \varepsilon' M$. The noise adversary partitions the MT microintervals into T/ε' intervals of length L each, where the ith interval $\Gamma_i = \{(i-1)L+1, \ldots, iL\}$. For every index $i \in \{1, \ldots, T/\varepsilon'\}$, the adversary acts as follows to set the noise function for packets from Γ_i:

1) Let n_i denote the Hamming weight of the string $X_{(i-1)L+1} \ldots X_{iL}$, i.e., the weight of the arrivals in the queue in interval i.
2) Let $\tilde{n}_i$ denote the rounding down of n_i to an integer multiple of $\varepsilon' \cdot L = (\varepsilon')^2 \cdot M$ (we assume all these are integers).
3) Let R_i denote the set of packets that arrive and leave in the ith interval, i.e., $R_i = \{j \in \Gamma_i | j + \Delta(j) \in \Gamma_i\}$.
4) For every $j \in R_i$, the adversary sets $Z_j = 0$ (or $\xi(j) = X_j$). Let $y_i = |R_i|$ and let $\hat{n}_i = \min\{\tilde{n}_i, L - y_i\}$.
5) The adversary flips the minimum number of packets from $\Gamma_i \setminus R_i$ so that exactly $\hat{n}_i$ of these are ones.

If at any stage the adversary exceeds its quota of εMT errors it stops flipping any further bits.

Error Analysis: We claim first that the probability that the adversary stops due to injecting too many errors is exponentially low. This is straightforward to bound. Notice that the number of bits of flipped in the ith interval due to early departures, is at most y_i, and $E[y_i] \leq \frac{1}{2}\varepsilon' L$. The number of bits flipped for packets that wait in the queue (i.e., from $\Gamma_i \setminus R_i$) is at most $\max\{y_i, n_i - \tilde{n}_i\}$. Again the expectation of this is bounded by the expectation of $y_i + (n_i - \tilde{n}_i)$ which is at most $\frac{3}{2}(\varepsilon')L$. Thus adding up the two kinds of errors, we find the expected number of bits flipped in the ith interval is at most $\frac{5}{2}\varepsilon' L$. Summing over all intervals and applying Chernoff bounds, we find the probability that we flip more than $(5\varepsilon' = \varepsilon)$-fraction of the bits is exponentially small in T.

Capacity Analysis: For the capacity analysis, we first note that the departure process from the delay queue (after the ξ function has been set) is completely independent of the encoding $X_1, \ldots, X_{MT}$, conditioned on $\tilde{n}_1, \ldots, \tilde{n}_{T/\varepsilon'}$ and on the event that the adversary does not exceed its noise bounds. Indeed for any fixing of the Δ function where the adversary does not exceed the noise bound, the output of $D^P|N^A$ channel on $X_1, \ldots, X_{MT}$ is the same as on the string $\tilde{X}_1 \ldots \tilde{X}_{MT}$, where for each i, the string $\tilde{X}_{(i-1)L+1} \ldots \tilde{X}_{iL}$ is set to $1^{\tilde{n}_i} 0^{L-\tilde{n}_i}$. Furthermore, note that the number of possible values of $\tilde{n}_i$ is at most $1/\varepsilon'$. We thus conclude that with all but exponentially small probability, the number of distinct distributions received by the receiver (which overcounts the amount of information received by the receiver) is at most $(1/\varepsilon')^{T/\varepsilon'} = (1/\varepsilon)^{O(T/\varepsilon)}$. An application of Lemma 3.1 now completes the proof.

■

3.2. Adversarial Delay followed by Random Noise ($D^A|N^P$)

Lemma 3.3: For every positive $\varepsilon \leq \frac{1}{2}$, there exists a capacity $C = C(\varepsilon)$ such that the capacity of the channel $D^A|N(\varepsilon)^P$ is bounded by C. Specifically, for every rate $R > C$, there exists a $\gamma > 0$ and $T_0 < \infty$ such that for every M (and $\mu = 1/M$), every $T \geq T_0$, and every pair of encoding/decoding functions $E_T : \{0,1\}^{k_T} \to \{0,1\}^{MT}$ and $D_T : (\mathbb{Z}^{\geq 0})^{MT} \to \{0,1\}^{k_T}$, the decoding error probability $\mathbf{Pr}_{\text{dec}} > \gamma$ if $k_T > R \cdot T$.

Proof Idea: We upper bound the channel capacity in two steps. In the first step we create an adversarial delay function that attempts to get rid of most of the "detailed" information being sent over the channel. The effect of this delay function is that most of the information being carried by the channel in M microintervals can be reduced to one of a constant (depending on ε) number of possibilities – assuming the errors act as they are expected to do. The resulting process reduces the information carrying capacity of the channel to that of a classical-style (discrete, memoryless) channel, and we analyze the capacity of this channel in the second step. We give a few more details below to motivate the definition of this classical channel.

We think of the delay function as a "queue", and the bits being communicated as "packets" arriving/departing from this queue. We call a packet a 0-*packet* if it was a zero under the encoding and as a 1-*packet* if it was a one under the encoding. Note that both types of packets, on release, get flipped with probability ε and the receiver receives one integer per time step representing the total number of ones received. The delay adversary clusters time into many large intervals and holds on to all packets received during an interval, and releases most of them at the end of the interval. In particular if it releases $\tilde{n_0}$ 0-packets and $\tilde{n_1}$ 1-packets at the end of an interval, it makes sure that $\varepsilon\tilde{n_0} + (1-\varepsilon)\tilde{n_1}$ takes on one of a "constant" number of values independent of M. (For some integer c, the actual value will be within $\pm\frac{1}{2}$ of an integer multiple of M/c due to integrality issues, but in this discussion we pretend we get an exact multiple of M/c.) Note that the quantity $\varepsilon\tilde{n_0} + (1-\varepsilon)\tilde{n_1}$ denotes the expected value of the signal received by the receiver when $\tilde{n_0}$ 0-packets and $\tilde{n_1}$ 1-packets are released, and so we refer to this quantity as the *signature* of the interval. If the errors were "deterministic" and flipped exactly the expected number of bits, then the channel would convey no information beyond the signature, and the total number of possible signatures over the course of all intervals would dictate the number of possible messages that could be distinguished from each other.

However the errors are not "deterministic" (indeed — it is not even clear what that would mean!). They are simply Bernoulli flips of the bits being transmitted, and it turns out that different pairs $(\tilde{n_0}, \tilde{n_1})$ with the same signature can be distinguished by the receiver due to the fact that they have different variance. This forces us to quantify the information carrying capacity of this "signal-via-noise" channel.

In the sequel, we first introduce this "signal-via-noise" channel (Definition 3.4) and bound its capacity (Lemmas 3.5 and 3.9). We then use this bound to give a proof of Lemma 3.3.

3.2.1. The "signal-via-noise" channel: We introduce the "signal-via-noise" channel which is a discrete memoryless channel, whose novelty is in the fact that it attempts to convey information using the variance of the signal. We assume that the reader is familiar with basic definitions from information theory, such as entropy, conditional entropy and mutual information (denoted $H(\cdot)$, $H(\cdot|\cdot)$ and $I(\cdot;\cdot)$ respectively) which we will use to bound the capacity of this channel. (These can be found in [2, Chapter 2] or in the full version of this paper.)

A discrete channel $\mathcal{C}$ is given by a triple $(\mathcal{X}, \mathcal{Y}, \mathcal{P})$, where $\mathcal{X}$ denotes the finite set of input symbols, $\mathcal{Y}$ denotes the finite set of output symbols, and $\mathcal{P}$ is a stochastic matrix with $\mathcal{P}_{ij}$ denoting the probability that the channel outputs $j \in \mathcal{Y}$ given $i \in \mathcal{X}$ as input. We use $\mathcal{C}(X)$ to denote the output of this channel on input X. The *information capacity* of such a channel is defined to be the maximum, over all distributions $\mathcal{D}$ on $\mathcal{X}$, of the mutual information $I(X;Y)$ between X drawn according to $\mathcal{D}$ and $Y = \mathcal{C}(X)$.

The information capacity turns out to capture the operational capacity (or just capacity as introduced in Section 2) of a channel when it is used many times (see Lemma 3.9 below). Our first lemma analyzes the information capacity of the "signal-via-noise" channel, which we define formally below.

In what follows, we fix positive integers M and c and a rational ε.

Definition 3.4: For integers M, c and $\varepsilon > 0$, the collection of (M, ε, c)-channels is given by $\{\mathcal{C}_\mu | \mu \in \{M/c, \ldots, M\}\}$, where the channel $\mathcal{C}_\mu = (\mathcal{X}_\mu, \mathcal{Y}_\mu, \mathcal{P}_\mu)$ is defined as follows: (a) $\mathcal{X}_\mu = \{(a,b) \in \mathbb{Z}^{\geq 0} \times \mathbb{Z}^{\geq 0}\}$ such that $a + b \leq M$ and

$$\mu - \frac{1}{2} < \varepsilon a + (1-\varepsilon) b \leq \mu + \frac{1}{2},$$

(b) $\mathcal{Y}_\mu = \{0, \ldots, M\}$, and (c) $\mathcal{P}_\mu$ is the distribution that, on input (a,b), outputs the random variable $Y = \sum_{i=1}^{a} U_i + \sum_{j=1}^{b} V_j$, where the U_i's and V_j's are independent Bernoulli random variable with $\mathbf{E}[U_i] = \varepsilon$ and $\mathbf{E}[V_j] = 1 - \varepsilon$.

Note that the expectation of the output of the channel $\mathcal{C}_\mu$ is roughly μ, and the only "information carrying capacity" is derived from the fact that the distribution over $\{0, \ldots, M\}$ is different (and in particular has different variance) depending on the choice of $(a,b) \in \mathcal{X}_\mu$. The following lemma shows that this information carrying capacity is nevertheless bounded as a function of ε and c (independent of M). Later we follow this lemma with a standard one from information theory showing that the information capacity does bound the functional capacity of this channel.

Lemma 3.5: For every $0 < \varepsilon \leq \frac{1}{2}$ and $c < \infty$, there exists $C_0 = C_0(\varepsilon, c)$ such that for all M the information capacity of every (M, ε, c)-channel is at most C_0.

Proof: The lemma follows from the basic inequality for any pair of random variables X and Y that $I(X;Y) = H(Y) - H(Y|X)$, where $H(\cdot)$ denotes the entropy function and $H(\cdot|\cdot)$ denotes the conditional entropy function. Thus to upper bound the capacity it suffices to give a lower bound on $H(Y|X)$ and an upper bound on $H(Y)$.

We prove below some rough bounds that suffice for us. Claim 3.6 proves $H(Y|X) \geq \frac{1}{2}\log_2 M - C_1(\varepsilon, c)$ and Claim 3.8 proves $H(Y) \leq \frac{1}{2}\log_2 M + C_2(\varepsilon, c)$. It immediately follows that the capacity of the channel $\mathcal{C}_\mu$ is at most $C_1(\varepsilon, c) + C_2(\varepsilon, c)$. We now proceed to prove Claims 3.6 and 3.8.

Claim 3.6: There exists $C_1(\varepsilon, c)$ such that for every $(a.b) \in \mathcal{C}_\mu$, $H(Y|X = (a,b)) \geq \frac{1}{2}\log_2 M - C_1(\varepsilon, c)$.

Proof: This part follows immediately from the claim below which asserts that for every $j \in \mathcal{Y}_\mu$, $\mathbf{Pr}[Y = j | X = (a,b)] \leq 8(c/\varepsilon)^{3/2} M^{-\frac{1}{2}}$. We thus conclude $H(Y|X = (a,b)] \geq \frac{1}{2}\log M - \frac{3}{2}\log\left(\frac{c}{\varepsilon}\right) - 3$.

Claim 3.7: For every $j \in \mathcal{Y}_\mu$, we have that

$$\mathbf{Pr}[Y = j | X = (a,b)] \leq 8(c/\varepsilon)^{3/2} M^{-\frac{1}{2}}.$$

The proof follows easily from an application of the Berry-Esséen theorem [3, Chapter 16]; we omit details from this version.

■

Claim 3.8: There exists C_2 such that for every random variable X supported on $\mathcal{X}_\mu$ and $Y = \mathcal{C}_\mu(X)$, we have $H(Y) \leq \frac{1}{2}\log_2 M + C_2$.

The claim roughly follows from the fact that the variance of Y is $O(M)$. Details omitted from this version.

■

Our analysis of the $D^A|N^P$ channel immediately yields a lower bound on the "operational capacity" of any sequence of channels $\{\mathcal{C}_{\mu_i}\}_{i=1}^{N}$. Standard bounds in information theory (see, for instance, [2, Chapter 8, Theorem 8.7.1]) imply immediately that a bound on the capacity also implies that any attempt to communicate at rate greater than capacity leads to error with positive probability. We summarize the resulting consequence below. (We note that while the theorem in [2] only considers a single channel and not a collection of channels, the proof goes through with only notational changes to cover a sequence of channels.)

Lemma 3.9: Transmission at rate R greater than C_0, the information capacity, leads to error with positive probability.

More precisely, for any $0 < \varepsilon \leq \frac{1}{2}$, let $C_0 = C_0(\varepsilon, c)$ be an upper bound on the information capacity of a collection of channels $\{\mathcal{C}_\mu|\mu\}$. Then for every $R > C_0$, there exists a $\gamma_0 > 0$ and $N_0 < \infty$ such that for every $N \geq N_0$ the following holds: For every sequence $\{\mathcal{C}_{\mu_i}\}_{i=1}^N$ of (M, ε, c) channels, and every encoding and decoding pairs $E : \{0,1\}^{RN} \to \prod_{i=1}^N \mathcal{X}_{\mu_i}$ and $D : \{0, \ldots, M\}^N \to \{0,1\}^{RN}$, the probability of decoding error $\mathbf{Pr}_{\text{dec}} \geq \gamma_0$.

3.2.2. Proof of Lemma 3.3:

Proof: We now formally describe the delay adversary and analyze the channel capacity. Let $c = 4/\varepsilon$ and let $C_0 = C_0(\varepsilon, c)$ be the bound on the capacity of (M, ε, c)-channels $\mathcal{C}_\mu$ from Lemma 3.5. We prove the lemma for $C(\varepsilon, c) = 2(C_0 + \log c)$.

Delay: Let $X_1, \ldots, X_{MT}$ denote the encoded signal the sender sends. The noise channel picks $\xi(j)$ independently for each j with $\xi(j)$ being 1 w.p. ε. We now describe the action of the delay channel (which acts without knowledge of ξ).

We divide time into $2T$ intervals, with the ith interval denoted $\Gamma_i = \{(i-1)(M/2)+1, \ldots, i(M/2)\}$. Let $n_1(i) = \sum_{j \in \Gamma_i} X_j$ and $n_0(i) = M/2 - n_1(i)$ denote the number of 1-packets and 0-packets that arrive in the queue in the ith interval. The delay adversary acts as follows:

1) Initialize $n_1'(1) = n_1(1)$ and $n_0'(1) = n_0(1)$.
2) For $i = 1$ to $2T$ do the following:
   a) If $n_1(i) \geq n_0(i)$ then set $\tilde{n}_0(i) = n_0'(i)$ and round $n_1'(i)$ down to $\tilde{n}_1(i)$ so that $(1-\varepsilon)\tilde{n}_1(i) + \varepsilon\tilde{n}_0(i)$ is within $\frac{1}{2}$ of the nearest integer multiple of M/c.
   b) Else let $\tilde{n}_1(i) = n_1'(i)$ and round $n_0'(i)$ down to $\tilde{n}_0(i)$ so that $(1-\varepsilon)\tilde{n}_1(i) + \varepsilon\tilde{n}_0(i)$ is within $\frac{1}{2}$ of the nearest integer multiple of M/c.
   c) Finally set $n_0'(i+1) = n_0(i+1) + n_0(i) - \tilde{n}_0(i)$. and $n_1'(i+1) = n_1(i+1) + n_1(i) - \tilde{n}_1(i)$.
   d) At the end of interval i, output $\tilde{n}_0(i)$ 0-packets and $\tilde{n}_1(i)$ 1-packets from the queue to the noise adversary. Formally, the delay channel outputs a set Λ_i of packets that are to be released at the end of interval Γ_i, where Λ_i includes all packets that arrived in Γ_{i-1} but were not included in Λ_{i-1}.
   e) The noise adversary simply flips the bits according to the noise function and outputs the sum of these bits. Specifically it sets $Z_j = X_j + \xi(j)$ and outputs $Y_i = \sum_{j \in \Lambda_i} Z_j$.

Analysis: We start by establishing that the delay adversary never delays any packet by more than M microintervals. Note that the number of packets that arrive in interval i, but are not released at the end of the interval is given by $(n_1'(i) - \tilde{n}_1(i)) + (n_0'(i) - \tilde{n}_0(i))$. One of the two summands is zero by construction, and the other *is at most* $M/(\varepsilon c) \leq M/4$ *by our construction*. Since the total number of packets arriving in an interval is $M/2$, this ensures that the total number released in an interval is never more than $3M/4 \leq M$ (as required for an (M, ε, μ)-channel). Next we note that packets delayed beyond their release interval do get released in the next interval. Again, suppose $n_0(i) > n_1(i)$. Then all 1-packets are released in interval i. And the number of 0-packets held back is at most $\frac{M}{\varepsilon c} \leq M/4$ which is less than $n_0(i)$ the total number of 0-packets arriving in interval Γ_i. Thus the adversary never delays any packet more than M microintervals, and the number of packets released in all intervals (except the final one) satisfy $\varepsilon\tilde{n}_0(i) + (1-\varepsilon)\tilde{n}_1(i)$ in an integer multiple of M/c.

For an encoded message $X_1, \ldots, X_{MT}$, let $\mu_i = [\epsilon\tilde{n}_0(i) + (1-\varepsilon)\tilde{n}_1(i)]$, where the notation $[x]$ indicates the nearest integer to x, denote the *signature* of the ith interval; and let $\vec{\mu} = (\mu_1, \ldots, \mu_{2T})$ denote its signature. Note that μ_i takes one of at most c distinct values (since it is between M/c and M and always an integer multiple of M/c). Thus the number of signatures is at most c^{2T}.

Now since the total number of distinct messages is 2^{k_T}, the average number of messages with a given signature sequence is at least $2^{k_T}/c^{2T}$. Furthermore, with probability at least $1 - \delta$, a random message is mapped to a signature sequence with at least $\delta 2^{k_T}/c^{2T}$ preimages. Suppose that such an event happens. Then, using the fact that $R > 2(C_0 + \log c) - \frac{1}{T}\log\delta$, we argue below that conditioned on this event the probability of correct decoding is at most $1 - \gamma_0$ (where $\gamma_0 > 0$ is the constant from Lemma 3.9). This yields the lemma for $\gamma = (1-\delta)\gamma_0$.

To see this, note that the signal $Y_1, \ldots, Y_{2T}$ received by the receiver is exactly the output of the channel sequence $\{\mathcal{C}_{\mu_i}\}_{i=1}^{2T}$ on input $\tilde{X}_1, \ldots, \tilde{X}_{2T}$ where $\tilde{X}_i = (1-\varepsilon)\tilde{n}_1(i) + \varepsilon\tilde{n}_0(i)$. If the receiver decodes the message (more precisely, its encoding) $X_1, \ldots, X_{MT}$ correctly from $Y_1, \ldots, Y_{2T}$, then we can also compute the sequence $\tilde{X}_1, \ldots, \tilde{X}_{2T}$ correctly (since the delay adversary is just a deterministic function of its input $X_1, \ldots, X_{MT}$). Thus correct decoding of the $D^A|N^P(\varepsilon)$ channel also leads to a correct decoding of the channel sequence $\{\mathcal{C}_{\mu_i}\}$. But the number of distinct messages being transmitted to this channel is $\delta 2^{k_T}/c^{2T}$. Denoting this by $2^{\tilde{R}\cdot 2\cdot T}$ and using the fact that $\tilde{R} > C_0$, we get that the channel must err with probability at least γ. ■

4. Infinite Capacity Regime

In this section we show that the capacity of the channel with adversarial noise followed by random delay $(N^A|D^P)$ is infinite. Specifically, we establish the following result:

Lemma 4.1: There exists a positive ε, such that the capacity of the channel $N(\varepsilon)^A|D^P$ is unbounded. Specifically, for every rate R, there exists a constant M (and $\mu = 1/M$), such that for sufficiently large T, there exist encoding and decoding functions $E_T : \{0,1\}^{k_T} \to \{0,1\}^{MT}$ and

$D_T : (\mathbb{Z}^{\geq 0})^{MT} \to \{0,1\}^{k_T}$, the decoding error probability $\mathbf{Pr}_{\text{dec}} \leq \exp(-T)$, with $k_T = R \cdot T$.

Proof Idea: The main idea here is that the encoder encodes a 0 by a series of 0s followed by a series of 1s and a 1 by a series of 1s followed by a series of 0s. Call such a pair of series a "block". If the noisy adversary doesn't corrupt too many symbols within such a block (and it can't afford to do so for most blocks), then the receiver can distinguish the two settings by seeing if the fraction of 1s being received went up in the middle of the block and then went down, or the other way around. This works with good enough probability (provided the delay queue has not accumulated too many packets) to allow a standard error-correcting code to now be used by sender and receiver to enhance the reliability.

Proof: We prove below the lemma for $\varepsilon < 1/64$ [1]. Let $k = k_T = RT$. We will set $M = O(R^5)$. Let $L = M^{4/5}$, and $L' = M^{3/4}$, $\Gamma_i = \{(i-1)L+1, \ldots, iL\}$, and $\Gamma'_i = \{iL - L' + 1, \ldots, iL\}$. As a building block for our sender-receiver protocol, we will use a pair of classical encoding and decoding algorithms, E' and D', that can handle up to $5/24$-fraction of adversarial errors. (Note that $5/24$ could be replaced with any constant less than $1/4$.) In particular, for each message $m \in \{0,1\}^k$, the algorithm E' outputs an encoding $E'(m)$ of length $N = \Theta(k)$ such that for any binary string s of length N that differs from $E'(m)$ in at most $(\frac{5}{24})N$ locations, $D'(s) = m$. We now describe our encoding and decoding protocols.

Sender Protocol: The encoding $E = E_T$ works as follows. Let $m \in \{0,1\}^k$ be the message that the sender wishes to transmit. The encoding $E(m)$ simply replaces every 0 in $E'(m)$ with the string $0^L 1^L$, and each 1 in $E'(m)$ with the string $1^L 0^L$. Thus $E(m)$ is a string of length $2LN = MT$. The sender transmits the string $E(m)$ over the channel.

Receiver Protocol: Recall that the receiver receives, at every microinterval of time $t \in [MT]$ the quantity $Y_t = \sum_{j \leq t | j + \Delta(j) = t} X_j \oplus \xi(j)$. For an interval $I \subseteq [MT]$, let $Y(I) = \sum_{j \in I} Y_j$. The decoding algorithm $D = D_T$, on input $Y_1, \ldots, Y_{MT}$ works as follows:

1) For $i = 1$ to N do:
   a) Let $\alpha_i = Y(\Gamma'_{2i-1})/L'$.
   b) If $Y(\Gamma'_{2i}) - Y(\Gamma'_{2i-1}) \leq (-\alpha_i + \frac{1}{2}) \cdot M^{11/20}$ then set $w_i = 1$, else set $w_i = 0$.
2) Output $D'(w)$.

Analysis: By the error-correction properties of the pair E', D', it suffices to show that for $(19/24)$-fraction of the indices $i \in [N]$, we have $w_i = E'(m)_i$.

Fix an $i \in [N]$ and let Q_i denote the number of 1's in the queue at the beginning of interval Γ'_{2i-1}. We enumerate a series of "bad events" for interval i and show that if none of them happen, then $w_i = E'(m)_i$. Later we show that with probability $(1 - \exp(-T))$ the number of bad i's is less than $(5/24)N$, yielding the lemma.

We start with the bad events:

$\mathcal{E}_1(i)$: $Q_i > cM$ (for appropriately chosen constant c). We refer i as *heavy* (or more specifically c-heavy) if this happens.

$\mathcal{E}_2(i)$: The number of errors introduced by the adversary in the interval Γ_{2i} is more than $16\varepsilon L$. We refer to i as *corrupted* if this happens.

$\mathcal{E}_3(i)$: i is not c-heavy but one of $Y(\Gamma'_{2i-1})$ or $Y(\Gamma'_{2i})$ deviates from its expectation by more than $\omega(M^{1/2})$. We refer to i as *deviant* if this happens.

In the absence of events $\mathcal{E}_1$, $\mathcal{E}_2$, $\mathcal{E}_3$, we first show that $w_i = E'(m)_i$. Denote i to be a 1-block if $E'(m)_i = 1$ and a 0-block otherwise. To see this, we first compute the expected values of $Y(\Gamma'_{2i-1})$, and $Y(\Gamma'_{2i})$ conditioned on i being a 0 block and i being a 1 block. (We will show that these expectations differ by roughly $M^{11/20}$, and this will overwhelm the deviations allowed for non-deviant i's.)

We start with the following simple claim (whose proof is omitted in this version).

Claim 4.2: Let ℓ_1, ℓ_2 be a pair of non-negative integers, and let $\mathcal{E}$ denote the event that a packet p that is in the delay queue at some time t leaves the queue during the interval $\{t + \ell_1 + 1, \ldots, t + \ell_1 + \ell_2\}$. Then $\left(1 - \frac{\ell_1}{M}\right)\left(\frac{\ell_2}{M} - \frac{\ell_2^2}{M^2}\right) \leq \mathbf{Pr}[\mathcal{E}] \leq \left(1 - \frac{\ell_1}{M} + \frac{\ell_1^2}{M^2}\right)\left(\frac{\ell_2}{M}\right)$. Thus if $\ell_1 = 0$ and $\ell_2 \ll M$, then $\frac{\ell_2}{M} - O\left(\left(\frac{\ell_2}{M}\right)^2\right) \leq \mathbf{Pr}[\mathcal{E}] \leq \frac{\ell_2}{M}$.

Let $Q_i = \tilde{\alpha} \cdot M$. We now analyze the expectations of the relevant $Y(\cdot)$'s. We analyze them under the conditions that $\tilde{\alpha}$ is bounded by the constant c (i.e. i is not heavy) and that i is not corrupt.

$\mathbf{E}[Y(\Gamma'_{2i-1})]$: The probability that a single packet leaves the queue in this interval is roughly $L'/M + O((L'/M)^2)$ (by Claim 4.2 above). The expected number of packets that were in the queue at the beginning of Γ'_{2i-1} that leave the queue in this interval is thus $(\tilde{\alpha} \cdot M \cdot L'/M) \pm O((L')^2/M)$. Any potential new packets that arrive during this phase contribute another $O((L')^2/M)$ potential packets, thus yielding $\mathbf{E}[Y(\Gamma'_{2i-1})] = \tilde{\alpha} L' \pm O(\sqrt{M}) = \tilde{\alpha} M^{3/4} \pm O(\sqrt{M})$.

$\mathbf{E}[Y(\Gamma'_{2i})]$ when i is a 1-block: Recall that a 1-block involves transmission of 1s in Γ_{2i-1} and 0s in Γ_{2i}. With the adversary corrupting up to $16\varepsilon L$ packets in Γ_{2i} and the addition of L' new 1s in the interval Γ'_{2i-1}, at most $L' + 16\varepsilon L$ new ones may be added to the queue at the beginning of the interval Γ'_{2i}. Using Claim 4.2 with $\ell_1 = L$ and $\ell_2 = L'$ to the Q_i packets from the beginning of interval Γ'_{2i-1}, and with $\ell_1 = 0$ and $\ell_2 = L'$ to the new packets that may have been added, we get that $\mathbf{E}[Y(\Gamma'_{2i})] \leq \tilde{\alpha} M^{3/4} - (\tilde{\alpha} - 16\varepsilon) M^{11/20} + O(\sqrt{M})$.

[1] For clarity of exposition, we do not make any attempt to optimize the bound on the value of ε.

$\mathbf{E}[Y(\Gamma'_{2i})]$ **when** i **is a** 0**-block:** In this case the Γ_{2i-1} is all 0s and Γ_{2i} is all 1s. So the number of 1s seen in the Γ'_{2i} should be more than the number of 1s seen in the 0-block case. In this case, the number of new 1s added to the queue in the intervals Γ'_{2i-1} and $\Gamma_{2i} - \Gamma'_{2i}$ is lower bounded by $L - L' - 16\varepsilon L$. Using Claim 4.2 again to account for the departures from Q_i as well as the new arrivals in the interval Γ'_{2i}, we get $\mathbf{E}[Y(\Gamma'_{2i})] \geq \tilde{\alpha} M^{3/4} - (\tilde{\alpha} - (1 - 16\varepsilon)) \cdot M^{11/20} + O(\sqrt{M})$.

Putting the above together we see that $\mathbf{E}[Y(\Gamma'_{2i-1}) - Y(\Gamma'_{2i})]$ has a leading term of $\tilde{\alpha} M^{3/4}$ in both cases (i being a 0-block or i being a 1-block), but the second order terms are different, and these are noticeably different. Now, if we take into account the fact that the event $\mathcal{E}_3(i)$ does not occur (i is not deviant), then we conclude that the deviations do not alter even the second order terms. We thereby conclude that if none of the events $\mathcal{E}_1(i)$ or $\mathcal{E}_2(i)$ or $\mathcal{E}_3(i)$ occur, then $w_i = E'(m)_i$.

We now reason about the probabilities of the three events. The simplest to count is $\mathcal{E}_2(i)$. By a simple averaging argument, at most $(1/8)$th of all indices i can be corrupt, since the total number of noise errors is bounded by $\varepsilon(2LN)$, and so the probability of $\mathcal{E}_2(i)$ is zero on at least $(7/8)$th fraction of indices. $\mathcal{E}_3(i)$ can be analyzed using standard tail inequalities. Conditioned on i being not c-heavy, each $Y(\cdot)$ is a sum of at most $(cM + L + L')$ independent random variables (each indicating whether a given packet departs queue in the specified interval). The probability that this sum deviates from its expectation by $\omega(\sqrt{M})$ is $o(1)$. Thus, the probability that $\mathcal{E}_3(i)$ happens for more than a $(1/24)$th fraction of indices i, can again be bounded by $\exp(-T)$ by Chernoff bounds.

The only remaining event is $\mathcal{E}_1(i)$. Lemma 4.3 below shows that we can pick c large enough to make sure the number of heavy i's is at most a $(1/24)$th fraction of all is, with probability at least $1 - \exp(-T)$. We conclude that with probability at least $1 - \exp(-T)$ the decoder decodes the message m correctly. ■

Lemma 4.3: For every $\delta > 0$, there exists a $c = c(\delta)$ such that the probability that more than δ-fraction of the indices i are c-heavy is at most $e^{-(MT)/4}$.

Proof: Recall that an interval i is *c-heavy* if $Q_i > cM$. We will show that the lemma holds for $c = 4/\delta$.

For each packet j, recall that $\Delta(j)$ indicates the number of microintervals for which the packet j stays in the queue. Let $W = \sum_j \Delta(j)$. We will bound the probability that W is "too large" and then use this to conclude that the probability that too many intervals are heavy is small.

Note W is the sum of MT identical and independent geometric variables (namely the $\Delta(j)$'s) with expectation of each being M. Thus the probability that $W > K$ (for any K) is exactly the probability that K independent Bernoulli random variables with mean $1/M$ sum to less than MT. We can bound the probability of this using standard Chernoff bounds. Setting $K = 2 \cdot M^2 \cdot T$, we thus get:

$$\mathbf{Pr}[W > 2M^2T] = \mathbf{Pr}[W > 2\mathbf{E}[W]] \leq \exp\left(-\frac{MT}{4}\right).$$

It then suffices to show that conditioned on $W \leq 2M^2T$, the fraction of c-heavy intervals (i.e., intervals where the queue contains more than $(4M)/\delta$ packets) is bounded by δ.

In order to bound the number of c-heavy intervals using the bound on W, we first note that $W = \sum_{t=1}^{MT} N_t$, where N_t denotes the number of packets in the queue at time t (counted in microintervals). Furthermore, since the number of packets in the queue can go up by at most one per microinterval, we see that heavy intervals contribute a lot to W. To make this argument precise, we partition time into chunks containing M/δ microintervals each (note that "chunks" are much larger than the "blocks"). We assume here that M/δ is an integer for notational simplicity. For $1 \leq \ell < \delta T$, the chunk C_ℓ spans the range $[\ell(M/\delta), (\ell+1)(M/\delta))$. We say a chunk C_ℓ is *bad* if the queue contains more than $(3M)/\delta$ packets at the beginning of the chunk, and say that it is *good* otherwise. On the one hand, if a chunk is good, then every interval contained inside the chunk has at most $(4M)/\delta$ packets in the queue, and is hence not c-heavy. On the other hand, if a chunk C_ℓ is bad, then its contribution to W (i.e., $\sum_{t \in C_\ell} N_t$) is at least $(M/\delta)(2M/\delta)$ (since this is the minimum of N_t for $t \in C_\ell$). This allows us to show that at most a δ-fraction of chunks can be bad. To see this, suppose δ_b is the fraction of bad chunks. Then we have

$$W = \sum_{\ell=1}^{MT} N_t \geq \delta_b(\delta T)(M/\delta)(2M/\delta) = 2(\delta_b/\delta)M^2T.$$

Now using $W \leq 2M^2T$, we get $\delta_b \leq \delta$. Finally note that if $1 - \delta$ fraction of the chunks are good, then $1 - \delta$ fraction of the blocks are not c-heavy, which completes the proof of the lemma. ■

5. Conclusions

Our findings, in particular the result that the channel capacity is unbounded in the setting of probabilistic error and delay, are surprising. They seem to run contrary to most traditional intuition about communication: all attempts at reliable communication, either in the formal theory of Shannon, or in the organic processes that led to the development of natural languages, are built on a discrete communication model (with finite alphabet and discrete time), even when implemented on physical (continuous time and alphabet) communication channels. In turn such assumptions also form the basis for our model of computing (the Turing model) and the discrete setting is crucial to its universality. In view of the central role played by the choice of finite alphabet in language and computation, it does make sense to ask how much of this is imposed by nature (and the

unreliability/uncertainty it introduces) and how much due to the convenience/utility of the model.

Of course, our results only talk about the capacity of a certain mathematical model of communication, and don't necessarily translate into the physical world. The standard assumption has been that a fixed communication channel, say a fixed copper wire, has an associated finite limit on its ability to transmit bits (reliably). We discuss below some of the potential reasons why this assumption may hold and how that contrasts with our results:

Finite Universe One standard working assumption in physics is that everything in the universe is finite and discrete and the continuous modeling is just a mathematical abstraction. While this may well be true, this points to much (enormously) larger communication capacities for the simple copper wire under consideration than the limits we have gotten to. Indeed in this case, infinity would be a pretty good abstraction also to the number of particles in the universe, and thus of the channel capacity. We note here that channel capacity has been studied from a purely physics perspective and known results give bounds on the communication rate achievable in terms of physical limits imposed by channel cross section, available power, Planck constant, and speed of light (see, for example, [1], [5]).

Expensive Measurements A second source of finiteness might be that precise measurements are expensive, and so increasing the capacity does come at increased cost. Again, this may well be so, but even if true suggests that we could stay with existing trans-oceanic cables and keep enhancing their capacity by just putting better signaling/receiving instruments at the two endpoints - a somewhat different assumption than standard ones that would suggest the wires have to be replaced to increase capacity.

Band-limited Communication A third possibility could be that signaling is inherently restricted to transmitting from the linear span of a discrete and bounded number of basis functions. As a physical assumption on nature, this seems somewhat more complex than the assumption of probabilistic noisiness, and, we believe, deserves further explanation/exploration.

Adversaries Everywhere Finally, there is always the possibility that the probabilistic modelling is too weak to model even nature and we should really consider the finite limits obtained in the adversarial setting as the correct limits. Despite our worst-case upbringing, this does seem a somewhat paranoid view of nature. Is there really an adversary sitting in every piece of copper wire?

ACKNOWLEDGMENTS

Thanks to Henry Cohn, Adam Kalai, Yael Kalai, Amos Lapidoth, David Tse, and Kuang Xu for helpful pointers and discussions.

REFERENCES

[1] J. D. Bekenstein. Black holes and information theory. Contemporary Physics, Vol. 45, Issue 1, pp. 31-43, Jan 2004.

[2] T. M. Cover and J. A. Thomas. Elements of Information Theory. Wiley Publishing, New York, 1991.

[3] W. Feller. An Introduction to Probability Theory and its Applications, Vol. 2. Wiley, New York, 1967.

[4] R. V. L. Hartley. Transmission of Information. Bell System Technical Journal, vol. 3, pp. 535–564, July 1928.

[5] S. Lloyd, V. Giovannetti, and L. Maccone. Physical Limits to Communication. Phys. Rev. Lett., 93(10), 100501, 2004.

[6] H. Nyquist. Certain topics in telegraph transmission theory. A.I.E.E. Transactions, p. 617; April, 1928.

[7] C. E. Shannon. A mathematical theory of communication. Bell System Technical Journal, vol. 27, pp. 379-423 and 623-656, July and October, 1948.

[8] C. E. Shannon. Communication in the presence of noise. In *Proc. Institute of Radio Engineers*, vol. 37 (1), pp. 10–21, Jan. 1949.

[9] A. M. Turing. On computable numbers. *Proceedings of the London Mathematical Society*, vol. 2 (42), pp. 230–265, 1936.