Anytime Information Theory

by

Anant Sahai

B.S., University of California at Berkeley (1994)
S.M., Massachusetts Institute of Technology (1996)

Submitted to the Department of Electrical Engineering and Computer Science
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2001

Author . . . .
Department of Electrical Engineering and Computer Science
November 30, 2000

Certified by . . . .
Sanjoy K. Mitter
Professor
Thesis Supervisor

Accepted by . . . .
Arthur C. Smith
Chairman, Department Committee on Graduate Students

Anytime Information Theory
by
Anant Sahai

Submitted to the Department of Electrical Engineering and Computer Science
on November 30, 2000, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Abstract

We study the reliable communication of delay-sensitive bit streams through noisy channels. To bring the issues into sharp focus, we focus on the specific problem of communicating the values of an unstable real-valued discrete-time Markov random process through a finite-capacity noisy channel so as to have finite average squared error from end-to-end. On the source side, we give a coding theorem for such unstable processes that shows that we can achieve the rate-distortion bound even in the infinite horizon case if we are willing to tolerate bounded delays in encoding and decoding. On the channel side, we define a new parametric notion of capacity called anytime capacity. The corresponding sense of reliable transmission is stronger than the traditional Shannon capacity sense but is less demanding than the sense underlying zero-error capacity. We show that anytime capacity exists for memoryless channels without feedback and is connected to standard random coding error exponents. The main result of the thesis is a new source/channel separation theorem that encompasses unstable processes and establishes that the stronger notion of anytime capacity is required to be able to deal with delay-sensitive bit streams. This theorem is applied in the control systems context to show that anytime capacity is required to evaluate channels if we intend to use them as part of a feedback link from sensing to actuation. Finally, the theorem is used to shed light on the concept of "quality of service requirements" by examining a toy mathematical example for which we prove the fundamental necessity of differentiated service without appealing to human preferences.

Thesis Supervisor: Sanjoy K. Mitter
Title: Professor

Acknowledgments

I thank my advisor Professor Sanjoy Mitter for being such an inspiring role model as a researcher. Not only did his ideas and insights help shape this work, but watching his aesthetic sense and intuition in action taught me what good research is truly about. He was also very supportive and patient with the time and twists that this research program has taken. I would also like to thank Professors Robert Gallager, John Tsitsiklis, and Dave Forney for their helpful comments and guidance as members of my committee. I also appreciate the many interactions I had with Professor Vivek Borkar. It was some of our work with him that helped prompt me to revisit the whole idea of capacity and take a fresh look. More recently, the discussions I have had with Professors Venkat Anantharam, Shlomo Shamai, and Emre Telatar have helped me present the ideas in this thesis in a clearer way.

I thank all my fellow students at LIDS and MIT who helped make my time here very enlightening. In particular, I valued my close collaborations with my office-mate Sekhar Tatikonda who not only served as a useful sounding board for ideas, but also was ever helpful with references and pointers into the literature. S.R. Venkatesh has also been a friendly source of different perspectives and interesting discussions. The many people I met through the student organizations Sangam and Asha helped broaden my non-academic interests. The administrative staff of LIDS, the EECS department, and MIT have also been a great help in navigating the system.

Last but not least, I want to thank my brother, parents, and the Gods for their constant love and support.

This research was partially supported by a DOD/ONR fellowship from the U.S. Government and by grants ECS-9873451 from NSF and contracts DAAD19-00-1-0466 and DAAL03-92-G-0115 from the Army Research Office. The flexibility and academic freedom made possible by these sponsors are much appreciated.

Contents

List of Figures

Chapter 1

Introduction

"A bit is a bit is a bit" seems to have become the motto of this digital age. Though we often take it for granted, the existence of such a common currency for information is far from obvious. While in practice it has been justified by the rapid advancement of digital electronics and the rise of a communications infrastructure to match, its meaningfulness rests upon the fundamental theorems of information theory — the philosophical foundation that helps us gain insight into problems of communication.

Broadly speaking, there are two sides to information theory. The better known side deals with the transmission of bits over channels. The noisy channel coding theorems establish the fundamental limitations for reliable transmission of bits over noisy channels. But there is another side to information theory. It relates to the encoding of sources with respect to some criterion of fidelity. The rate distortion theorems establish fundamental tradeoffs between fidelity and the length of the encoding in bits. These two sides of information theory are joined together by the information transmission theorems relating to source/channel separation. Philosophically speaking, these theorems establish that the notion of "reliable transmission" used for discussing channels is compatible with the notion of "bits" used for encoding sources.

Of course, information theory has come a long way since Shannon's seminal papers.[60, 62] On the channel side, there has been much work over the years in extending this theory to cover broader classes of channels[67], by now covering very useful approximations to unreliable media that are encountered in practice, including many situations with memory. Though less dramatic, the source side has also advanced considerably and now covers many general classes of sources with memory.[37, 7] The limiting factor on source coding has traditionally been a matter of finding good theoretical approximations to real world situations. The information transmission theorems have also been extended along the way[6, 66], even partially into contexts of networked collaborative computing.[58, 52]

However, there are still many issues in modern communication systems for which information theoretic understanding eludes us. Networking in particular has a whole host of them, leading Ephremides and Hajek to entitle their recent survey article "Information Theory and Communication Networks: An Unconsummated Union!" [15]. They comment:

> The interaction of source coding with network-induced delay cuts across the classical network layers and has to be better understood. The interplay between the distortion of the source output and the delay distortion induced on the queue that this source output feeds into may hold the secret of a deeper connection between information theory. Again, feedback and delay considerations are important.

These issues center around "information streams" where the data source continues to generate information forever rather than being a single isolated message. In these contexts, people have noticed that streams emerging from different real-world sources seem to behave differently. What is appropriate for a classic file transfer may not be suitable for real-time multimedia signals.

Simultaneously, an important class of mathematical sources has eluded complete understanding in an information theoretic sense: unstable sources. By these, we are referring to certain non-stationary processes which often arise in control and other system-theoretic modeling contexts. Although they have proven themselves to be valuable parts of the system modeler's toolbox, the fundamental difficulty with unstable processes is that they tend to grow with time and they have asymptotically infinite variance. Because of this, their "streaming" character through time cannot be ignored. We believe that extending information theory to address unstable sources might be the conceptual key to understanding the deeper issue of "streaming" in general.

In control contexts, the objective is to keep the modeled system stable. To do this, communicating the values of the underlying unstable process from one place to another is an important issue[69, 70, 54] since the physical location for applying actions is generally somewhat removed from the location where observations are being made. In practice, people are already using quantizers and digital implementations to control unstable physical systems and there has been some isolated theoretical work on trying to extend source coding theory to unstable processes. Most recently, the work of Sekhar Tatikonda [63] has taken us significantly forward in understanding the information theoretic limitations of control systems. Even so, a major gap remains. Till now, there have been no information transmission theorems to tie the source coding of unstable processes to communication over noisy channels.

This question has remained open for quite some time. To make matters concrete, we start with a very simple example of a random walk on the infinite line with known initial condition. We then show how even with a noiseless channel, a block source code cannot be used to track this unstable process. All block source codes have asymptotically infinite distortion, regardless of the rate or how large a block length is chosen. For our particular example however, there is an obvious causal source encoder with memory which tracks the source perfectly over a noiseless link. To illustrate the issues that arise with noisy links without feedback, we show how traditional "linear" approaches to estimation fail because the effective signal-to-noise ratio goes to zero with time. The point of this section is to show that even in this simple case, the problem is far from trivial.

After these introductory comments, we introduce the general problem of tracking unstable scalar Markov processes over noisy channels, ask the relevant questions, and informally state the major thrusts and results of this thesis.

In the second chapter of this thesis, we review definitions of sources, channels, codes, etc. After reviewing channel coding in a streaming context, we illustrate how any attempt to cascade our simple source encoder with a conventional channel encoder is doomed to failure because the Shannon sense of reliable transmission is not sufficient for our task. This will show that in the context of unstable processes, a "bit" is not necessarily a "bit" and we will make this idea much more precise in the next chapter by introducing a delay-dependent notion of intrinsic meaning for streams of bits that shows how bit-streams can be fundamentally different from each other even if they have the same bit-rate.

Next, we introduce a new parametric notion of reliable transmission and its associated channel capacity that we call anytime capacity and suggest how anytime decoders can be used to avoid the problem for our simple random walk. Unlike classical decoders, an anytime decoder will eventually correctly recover the values for *all* the bits originally sent. We then use a random coding argument to show that the anytime capacity of a binary erasure channel and the power-constrained AWGN channel is non-zero even without any feedback but implicitly assuming access to common randomness shared by the encoder and decoder. After that motivation, we quickly establish the connection between anytime capacity and general block random coding error exponents. We also show the existence of deterministic codes with the required properties.

Finally, we show that the random walk can be tracked using an anytime channel code to transport the bit-stream and give the main result of this thesis: the relevant information transmission theorem for unstable scalar Markov processes. This information transmission theorem establishes that the unstable processes are fundamentally different from memoryless or stationary processes. The traditional source-channel separation theorems imply that the bit-streams used to encode all memoryless or stationary processes are qualitatively alike as long as we are willing to tolerate delays. All they need is a channel with sufficient capacity. But even if we are willing to live with delay, unstable processes require more from a channel than merely Shannon capacity.

This understanding is extended to simple control problems and then used to establish the anytime capacities of the erasure channel with feedback and the AWGN channel with feedback. In the erasure case, we are able to get a parametric closed-form expression for the true anytime capacity. For the AWGN case, we are able to recover a result reminiscent of Kailath and Schalkwijk[59] which gives a doubly-exponential in delay convergence of errors to zero. The difference is that our scheme requires no blocking of the bits and works for all delays at once! The same technique lets us get a lower bound on the zero-error capacity of power-constrained additive noise channels where the channel noise has bounded support.

In a sense, the parameter in our definition of anytime capacity can be interpreted as an additional Quality of Service requirement. We illustrate that in the penultimate chapter by giving a simple example where the need for differentiated service can be established mathematically. To our knowledge, this is the first time where a simple class of random processes have been exhibited which intrinsically require different Qualities of Service apart from just bit-rate. This requirement is not posited *a-priori* or by appeal to human preferences, but emerges from the nature of the source and distortion measure.

1.1 "Streaming" and Why We Study Unstable Processes

At one level, unstable processes are interesting in their own right since they emerge naturally in the course of system modeling. Addressing them from an information theoretic perspective is useful to other fields like control which might have communication problems embedded in their contexts. But at a deeper level, we are interested in them because they represent a conceptually simple and yet extreme case. For unstable processes, the issues of streaming, delays, and "real-time" are absolutely essential to the problem of end-to-end communication. For us, the non-stationarity of unstable processes is a way of forcing attention on their streaming aspects since we show that the standard tricks of dividing the source stream into a sequence of independent message blocks cannot work.

In the literature, there are other information theoretic approaches that shed some light on the general issue of "streaming." Here we will briefly mention two of them: Tse's approach of adjusting the rate of source coding in response to buffer states and Shulman's approach to interactive computation over noisy channels.

1.1.1 Adjustable Length Source Codes: Tse

Tse's dissertation [65] shows how adjusting the coding of streaming sources in response to the state of the communication system that they are feeding into can give much improved performance. There are fewer buffer overflows while we still maintain good end-to-end distortion. Since the delays in the model are due to buffering, this strategy also reduces the end-to-end delays, even when the buffer sizes are infinite.

From our perspective, Tse's approach represents an important step towards understanding "streaming" when the acceptable end-to-end delay is thought of as being smaller than the end-to-end delay that would be introduced by a fixed-rate source coding system connected to the same communication system. If we restrict ourselves to the narrow classical information-theoretic perspective that we do not care about end-to-end delay as long as it is finite, Tse's approach does not buy us anything in terms of either average rate or distortion. But by introducing a loose coupling between the source and channel coding, it does give us a much better tradeoff between delay and performance than the classical completely separated approach.

1.1.2 Interactive Computation: Schulman

Schulman's work [58] addresses the problem of how to interactively compute using distributed processors if the communication links connecting the processors are noisy channels. The issue of "streaming" is central here since the messages to be sent from one processor to the other depend on the messages already received from the other processor. Grouping the messages into blocks for traditional block channel-encoding is out of the question since there is no way to know what messages we will want to send in the future without already sending and receiving messages now. He shows that for any sufficiently large problem size, there exist good coding strategies that allow the computation to proceed with an arbitrarily low probability of error.

As such, the essential "streaming" aspect is understood to emerge from the interaction between the two streams of messages. The results are conceived in terms of a single finite-sized problem instance rather than infinite duration streams. This line of work has effectively given new information transmission theorems that cover the case of interactive computation and have been extended by others [52]. In the spirit of computer science, one of the goals has been to make sure that the complexity does not grow by more than a polynomial factor. In contrast, our work on unstable processes in this thesis covers a problem with a single information stream in a standard one-way communication problem. However, it still shares a lot with Schulman's approach in spirit and technique.

1.2 The Simplest Case: Tracking a Simple Random Walk

Random walks are among the simplest examples of unstable processes. To make matters concrete, let us consider the following extremely simple discrete time real valued source X_t which is generated as follows: $X_0 = 0$ and $X_{t+1} = X_t + W_t$ where the $\{W_t\}$ are i.i.d. fair coin tosses with $P(W_t = -1) = P(W_t = 1) = \frac{1}{2}$. It should be clear that $\{X_t\}$ is not stationary and can wander over all the integers (a countable set) in its random walk. In fact, it turns out that while $E[X_t] = 0$ for all t, $E[X_t^2] = t$ and thus tends to infinity by construction.

To talk about tracking this random walk, we need to have a per-letter distortion measure. We will use the usual squared-error distortion $\rho(X, \hat{X}) = (\hat{X} - X)^2$. The following definition specifies what we mean by "tracking."

Definition 1.2.1 *A random process $\{\hat{X}_t\}$ is said to* ρ-track *another process $\{X_t\}$ if $\exists D < \infty$ such that $\sup_{t>0} E[\rho(X_t, \hat{X}_t)] = D < \infty$.*

We are interested in the problem of tracking this random walk through a "finite capacity" link. We will first consider noiseless finite rate binary channels and then consider a noisy real valued channel to show some of the different issues that can arise.

1.2.1 Block-codes are not enough

A great deal of attention in coding and information theory has been dedicated to block-codes and analyzing their properties.[23] In order to show their shortcomings, we need to know what we mean by a "block-code" for sources.

Definition 1.2.2 *A* block source-code *with block-length n and rate R (where both n and nR are assumed to be non-negative finite integers) is a pair of functions (F_n, G_n) such that F_n maps a vector $(X_{in}, X_{in+1}, \ldots, X_{in+n-1})$ (representing n consecutive symbols from the source sequence) into a binary string s of length nR (thought of as a non-negative integer less than 2^{nR}) and G_n maps binary strings s of length nR into a reconstruction vector $(\hat{X}_{in}, \hat{X}_{in+1}, \ldots, \hat{X}_{in+n-1})$. We call F_n the* block source-encoder *and G_n the* block source-decoder.

This is slightly different from some standard definitions of a block source-code which do not explicitly include the decoder and instead use a single function going directly from a block of source samples to a block of reconstructions. We use our definition because it makes it easier to see what it means to interconnect a source-code with a channel-code since both explicitly use the common currency of binary bits.

Now, we are ready to state and prove the initial negative result.

Proposition 1.2.1 *No block source-code can track the random walk $\{X_t\}$, regardless of the block-length n or the rate R.*

Proof: Let (F_n^R, G_n^R) be a block-code of rate R and block-length n. Let the maximum size element of the reconstruction vector given input s be denoted by $M(s) = \max_{j<n} \|G_n^R(s)_j\|$. Since by definition the domain of G_n^R is the finite set $\{0, 1, \ldots, 2^{nR} - 1\}$, we know that there exists a finite integer M such that $M(s) < M$. This means that for all t, $\|\hat{X}_t\| < M$.

But the random walk X_t will eventually spend arbitrarily large amounts of time outside any bound. Since X_t is the sum of t i.i.d. random variables with zero mean and unit variance, by a simple application of the central limit theorem we know $\forall T > 0$ there exists $N_T > 0$ such that $\forall t > N_T$, we have $P(|X_t| > \sqrt{t} > T) > \frac{1}{4}$. So, $\limsup_{t\to\infty} E[\rho(X_t, \hat{X}_t)] > \frac{1}{4}(T - M)^2$. But T was arbitrary and so we have $\limsup_{t\to\infty} E[\rho(X_t, \hat{X}_t)] = \infty$ for all block-codes, thereby proving the theorem. □

Notice that the proof only relied upon the fact that the *source decoder* was a deterministic function on a finite set and hence had a finite range. Thus, we can immediately extend this result to the case of block joint-source-channel-codes in those cases where the channel has a finite output alphabet giving us the following Corollary, which reduces to the previous theorem in the case of a noiseless channel.

Corollary 1.2.1 *If the communication channel has a finite output alphabet, no block joint-source-channel-code can track the random walk $\{X_t\}$, regardless of the block-length n.*

This rejection of block codes when dealing with unstable processes tells us that the arguments of Krich [40, 41] on sources with delay-dependent fidelity criteria are not applicable directly since he restricts his attention to block-codes throughout. We discuss this more in Section 2.2.4.

1.2.2 A Source Code with Memory

The negative results for block codes are interesting because we intuitively know that this source should be able to be perfectly tracked with a rate 1 causal encoder that simply encodes $W_t = X_{t+1} - X_t$ with a single bit corresponding to its sign. Mathematically, we have:

$$S_t = F_t(X_1^t) = \frac{(X_t - X_{t-1} + 1)}{2} \tag{1.1}$$

As we can see, the encoder only needs access to the current and previous value of the source, not the entire history.

All the decoder has to do is to add up the W_t values up till now. So:

$$\hat{X}_t = G_t(S_1^t) = \sum_{i=1}^{t}(2S_i - 1) \tag{1.2}$$

which can also be done in a a simple recursive manner using an internal integer variable to track $\hat{X}_{t-1}$. In this example, the reconstruction clearly tracks the original perfectly, since $\hat{X}_t = G_t(F_1^t(X_1^t)) = \sum_{i=1}^{t}(2\frac{(X_i - X_{i-1}+1)}{2} - 1) = \sum_{i=1}^{t}(X_i - X_{i-1}) = X_t$. Of course, this is all *assuming that the input to the decoder is exactly the output of the encoder.*

1.2.3 Tracking across Noisy Channels

The simple random walk example is clearly a well behaved linear system on the real line. Our experience with filtering for linear systems[36] might suggest that it is easy to track given linear observations with additive noise. If we want to use a linear encoder with Gaussian noise, our observation equation will be of the form: $Y_t = c_t X_t + V_t$ where $c_t X_t$ is our linear encoding of the signal and $\{V_t\}$ is a white Gaussian process with zero mean and unit variance.

But things are not so simple since real communication channels have constraints on their inputs. The additive white Gaussian noise channel (defined formally in section A.4.1) is traditionally considered with a power constraint P on the average power in the channel input signal. It turns out that if our encoder has access to noiseless feedback then the problem is not that difficult. Our analysis in [55] can be used to show that a simple simulated control system can be used to track the random walk process while still meeting the power constraint. In that case, we do not transmit $c_t X_t$ but rather $c_t u_t$ where $\{u_t\}$ is the simulated control signal used to keep the simulated plant stable. This particular scheme is discussed further in Section 7.2.2.

The case without access to feedback is far more tricky. Without feedback, there is no straightforward way to have a simulated control system and so we have to transmit functions of X_t itself. Since $E[X_t^2] = t$, it is clear that no constant c_t will meet the finite power constraint. We have to use a time-varying encoder with $c_t \leq \frac{\sqrt{P}}{\sqrt{t}}$ in order to meet it.

We can then generate estimates using the corresponding time-varying Kalman filter.[36] Since we know that the estimates can only get better if c_t is larger, it suffices to look at the case where c_t is equal to its upper bound. If we explicitly write out the recursive expression for the Kalman filter error covariance, we get:

$$\begin{aligned}\sigma_{t+1}^2 &= \frac{\sigma_t^2+1}{\frac{P}{t+1}(\sigma_t^2+1)+1} \\ &= \frac{1}{\frac{P}{t+1}+\frac{1}{\sigma_t^2+1}}\end{aligned}$$

with an initial condition of $\sigma_0^2 = 0$ since we know exactly where the random walk starts. If we evaluate this recursion numerically on a computer, we see that the variance grows as $O(\sqrt{t})$ with time t.

We can also see this more rigorously. As a function of t, the estimation variance σ_t^2 can either get arbitrarily large or stay bounded. To show that this variance has no finite upper bound, we first show that σ_t^2 increases with time as long as $\sigma_t^2 < -\frac{1}{2}+\sqrt{\frac{t+1}{P}+\frac{1}{4}}$ by noticing that:

$$\begin{aligned}1 &< \frac{\sigma_{t+1}^2}{\sigma_t^2} \\ 1 &< \frac{1}{\frac{\sigma_t^2 P}{t+1}+\frac{\sigma_t^2}{\sigma_t^2+1}} \\ \frac{\sigma_t^2 P}{t+1} &< \frac{1}{\sigma_t^2+1} \\ (\sigma_t^2)^2 P + \sigma_t^2 P - t - 1 &< 0 \\ \sigma_t^2 &< -\frac{1}{2}+\sqrt{\frac{t+1}{P}+\frac{1}{4}}\end{aligned}$$

are equivalent statements if we know that σ_t^2 must be positive. If we assume that σ_t^2 stays bounded, we know that for large t it must be strictly increasing up to a some threshold M.

To now see that it must exceed all thresholds M, we just consider a time $t > 2(M+\frac{3}{4})^2P - 1$ and consider what happens if it is within $\frac{1}{4}$ of the threshold M at time t. Then:

$$\begin{aligned}\sigma_{t+1}^2 &= \frac{\sigma_t^2+1}{\frac{P}{t+1}(\sigma_t^2+1)+1} \\ &> (M-\frac{1}{4}+1)(1-\frac{\frac{(\sigma_t^2+1)P}{t+1}}{\frac{(\sigma_t^2+1)P}{t+1}+1} \\ &> (M+\frac{3}{4})(1-\frac{\frac{(M+\frac{3}{4})P}{t+1}}{\frac{(\sigma_t^2+1)P}{t+1}+1} \\ &> (M+\frac{3}{4})(1-\frac{(M+\frac{3}{4})P}{t+1}) \\ &= M+\frac{3}{4}-\frac{(M+\frac{3}{4})^2P}{t+1}\end{aligned}$$

$$
\begin{aligned}
&> \quad M + \frac{3}{4} - \frac{(M+\frac{3}{4})^2 P}{(2(M+\frac{3}{4})^2 P - 1) + 1} \\
&= \quad M + \frac{1}{4}
\end{aligned}
$$

showing that it crosses the threshold M. But M was arbitrary and so we can take an increasing sequence starting at 0 and going up by $\frac{1}{4}$ each to show that all thresholds are eventually crossed regardless of the power constraint P. This proves the following theorem:

Theorem 1.2.1 *For all power constraints $P > 0$, no time-varying linear memoryless encoder can be used to track our simple random walk across an additive white Gaussian noise channel with power constraint P without access to feedback.*

Other methods based on encoding the innovation signal directly suffer the problem of having to keep the encoder and decoder perfectly "in sync." We will show this more precisely in Section 2.4.2. This difficulty has long been recognized. In his definitive book on rate-distortion theory, Toby Berger writes:([6] pg 247)

> It is worth stressing that we have proved only a source coding theorem for the Wiener process, not an information transmission theorem. If uncorrected channel errors were to occur, even in extremely rare instances, the user would eventually lose track of the Wiener process completely. It appears (*although it has never been proved* [emphasis mine]) that, even if a *noisy* [emphasis in original] feedback link were provided, it still would not be possible to achieve a finite MSE per letter as $t \to \infty$.

One of the goals of this thesis is to disprove this longstanding conjecture.

1.3 General Problem Statement

The above discussion of the simple random walk was just intended to motivate the general issues and to show that they are far from trivial. In general, we are interested in the problem of communicating a Markov process across a noisy channel:

Given a scalar discrete-time Markov source $\{X_t\}$ with parameter a and driven by noise $\{W_t\}$:

$$X_{t+1} = AX_t + W_t$$

and a discrete-time $(0, \tau, \mathcal{A}, \mathcal{B})$ noisy channel with 0 delay and sampling time τ, that maps inputs from $\mathcal{A}$ randomly into outputs in $\mathcal{B}$, is it possible to design encoders and decoders within a specified finite end-to-end delay constraint so that the output of the decoder $\{\hat{X}_t\}$ achieves a desired mean-squared performance $\sup_{t>0} E[(\hat{X}_t - X_t)^2] = D$?

As illustrated in Figure 1-1, this problem has a few important features. The first is that the source sequence $\{X_t\}_{t=1}^{\infty}$ is not realized all at once, but is observed in "real time" as time progresses. In the same way, the channel is also only available to be used once every τ units of time. As a result, it only makes sense that the reconstructed estimates $\{\hat{X}_t\}_{t=1}^{\infty}$ are also generated one at a time in a somewhat "real time" fashion. By finite end-to-end delay constraint d, we mean that the estimate $\hat{X}_t$ for X_t must be ready at the decoder within d time units — that is, it must be ready before time $t + d$. In other words, all the blocks in the above diagram can and should be viewed as being causal in the appropriate sense. Their outputs are not allowed to depend on random variables that have not occurred yet. The following questions arise naturally and are answered in this Thesis:

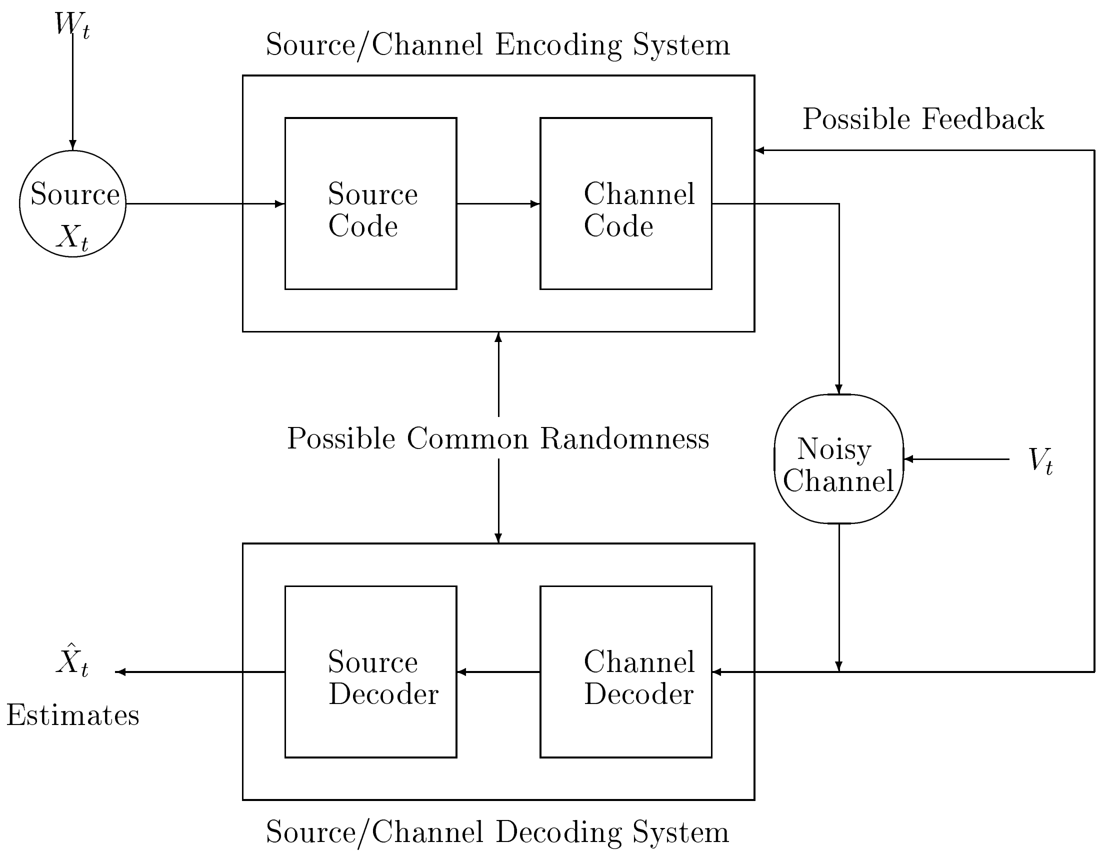

Figure 1-1: The estimation problem across a noisy channel

- For a finite end-to-end delay constraint, can we ever get D finite if the process is unstable (ie $A \geq 1$)?

- Is it possible to accomplish this task using encoders which have access to noiseless feedback of the channel outputs?

- Is it possible to accomplish this task using encoders which do not have access to any feedback?

- Is there a necessary and sufficient condition on the channel that must hold for finite D to exist even if we impose no restrictions on the joint source/channel encoding and decoding systems?

- Can we express this condition in terms of reliably communicating "bits" in some sense?

- Is it possible to somehow factor solutions through "bits" — giving the joint encoder and decoder systems in terms of separate source and channel codes?

- What do the encoder and decoder need? Do they have to be random and have access to common randomness?

- Can we bound the performance and the optimal D if we allow ourselves to consider the limit of increasing end-to-end delays as we do in existing rate-distortion theory?

- Is this bound achievable over a wide class of suitable noisy channels?

- Can we extend these results to performance measures other than the expected second moment of the error?

- What are the implications for Control, "Quality of Service," and more general problems?

1.4 Main Results

After introductory and review material, the thesis has four main theoretical thrusts. Here we informally state the main result of each. While the source coding and channel coding discussions are intelligible even in isolation, the main contribution of the thesis is to give a unified picture that lets us look at the whole problem and to see how the pieces fit together in the spirit of Shannon's original program for understanding communication. The idea is to look at extremely simple situations and use them to develop insights that shed light on certain core problems of communication.

In the final chapters of this thesis, we use the seemingly simple unstable scalar Markov sources and the results described below to illuminate issues of "real-time," "streaming sources," "control over communication links," and "differentiated quality of service." Without looking at the whole picture, it is impossible to understand the real issues and interconnections underlying such problems.

1.4.1 Source Coding

Lossy source coding is about translating the original source process into an approximate version of that process by first going through an intermediate process expressed in the language of bits. The intermediate process consists of a stream of bits generated based on the realization of the original source. It is subsequently mapped into the reconstruction alphabet to get the approximate version. We can naturally speak of the bit-rate for the intermediate stream, while distortion is measured between the original process and the final approximating version. In Chapter 4, a source coding theorem is proved for unstable Markov processes. The theorem works for the infinite-horizon average distortion case and thereby resolves a problem left open since [29].

Result 1.4.1 *For the unstable scalar Markov source, there exist variable rate source codes, all with finite end-to-end delay, which can approach the fundamental rate-distortion bounds in both rate and distortion.*

The important feature of this result is that we are dealing with a streaming context. Unlike the standard setup [30], we cannot assume that the source realization is known in advance at time 0. It is only observed as time goes on. Furthermore our approach works even though the source is severely non-stationary and exhibits strong dependencies that never fade away between any two blocks of time.

1.4.2 Delay Sensitivity

In Chapter 3, a formal definition is given for the delay and error dependent "complexity" of source codes and thus also of the bit-streams that emerge from them. This definition focuses on the sensitivity of a code to the propagation of errors. It results in a sense of intrinsic complexity for a pair consisting of a random source together with the distortion measure. In some ways, this parallels the way that the theory of computational complexity is able to use a complexity measure for algorithms to determine the intrinsic complexity for problems. This error-propagation complexity is explored for the Markov sources and is used to establish:

Result 1.4.2 *For the unstable scalar Markov source, all source codes which track the process in the mean-squared sense are sensitive to bit-errors. If any of the bits for the last d time units have been altered, the resulting additional distortion can grow exponentially as A^{2d}.*

This sort of fundamental sensitivity to errors tells us that standard approaches to cascading source and channel codes can never be successful because even a single uncorrected error, no matter how rare, can eventually cause unboundedly large distortions. We call bitstreams "weak" if they result from codes that have this cascading-error property. In contrast, bitstreams coming from traditional codes where errors cannot cascade catastrophically are called "strong."

1.4.3 Anytime Channel Coding

Channel coding is about reliable transporting a stream of bits across a noisy channel. Recognizing that traditional conceptions of reliable transmission are inadequate for "weak" bitstreams, a definition is given for a new stronger sense of reliable transmission and its associated capacity measure that we call anytime capacity. Rather than thinking of "reliable transmission" as being able to achieve a specified low probability of error, we think of the probability of error as being determined only when the user of the information has committed to a particular delay or wait. The notion of reliability can be parametrized by a function which tends to zero with increasing delay. Even if the probability never actually reaches 0 for a finite wait, if it goes down fast enough we can show that eventually every bit will be decoded correctly. This sense is in some ways the natural generalization of the sense in which TCP/IP achieves reliable transport of bit streams across a noisy Internet.

Result 1.4.3 *For discrete memoryless channels without feedback, encoders exist for which the decoder has the freedom to choose the delay of decoding and so that the probability of error for any bit position tends to zero exponentially with increasing delay.*

We study this anytime capacity for particular channels using random coding arguments and show it to be non-zero. In the general case, we relate it to standard block random coding error exponents and show how this sense of reliable transmission can be interpreted as a "universal" error exponent, where the universality of the encoder is over the delay that is acceptable to the decoder. This is related to work showing that the random coding error exponents govern the rate at which the probability of error for finite state convolutional codes goes down if we force decoding to occur with a fixed delay rather than letting Viterbi decoding take its course.[19]

1.4.4 Separation Theorem

In chapter 6, a new information transmission theorem is developed that uses the notions developed so far and ties them together to give the relevant source/channel separation for unstable Markov processes.

Result 1.4.4 *For the unstable scalar Markov source with parameter $A > 1$ to be tracked in the finite mean-squared error sense across a noisy channel by some encoding/decoding system, it is necessary for the channel to be able to carry at least $\log_2 A$ bits per unit time with a probability of error that tends to zero exponentially as $2^{-(2\log_2 A)d}$ with delay $d > 0$ chosen at the decoder. Furthermore, this condition is also sufficient and tracking can be accomplished with a system that is divided into a source code (which outputs bits) and a channel code which focuses on getting the bits across the channel.*

This establishes that the sense of reliable transmission that we have defined is fundamental for this problem. The key to this result is the proof of the converse theorem. While for traditional Shannon capacity and rate-distortion theory the converse is a simple consequence of the properties of mutual information, in our case the converse is much more difficult to prove since we do not have a mutual-information characterization of anytime capacity. We give a constructive proof of the converse that is able to relate the two operational notions to each other without having to go through another characterization.

By rigorously giving us something other than just the bit-rate to match between source code and channel code, this theorem is a significant step. It brings us several steps closer towards a rigorous understanding of prior ideas, motivated by intuition, of loosely coupled joint source/channel coding like those in [32].