The Anytime Reliability of the AWGN+erasure channel with Feedback

Anant Sahai
Department of Electrical Engineering and Computer Science
University of California
Berkeley, CA 94720-1770
sahai@eecs.berkeley.edu

Qing Xu
Department of Mechanical Engineering
University of California
Berkeley, CA 94720-1770
qingxu@me.berkeley.edu

Abstract

We study the feedback anytime reliability of a discrete-time channel with additive white Gaussian noise where the channel output is also subject to iid erasure. The encoder has noiseless access to the past channel outputs, which includes perfect information regarding which transmissions were erased. There is an average power constraint on the channel input. We show that the anytime reliability is constant at all rates up to the Shannon capacity of the channel and that this constant is essentially the logarithm of the probability of erasure. In order to show achievability of this reliability, we give a construction involving a hybrid control system consisting of a FIFO queue where the server can adjust its service rate based on the number of bits awaiting transmission. The server takes the data bits and uses them to drive a scalar linear control system with continuous state where the dynamics can switch between fast and slow based on the service rate.

I. Introduction

The discrete-time power constrained additive white Gaussian noise channel is one of the most useful idealizations in communications theory. It models communication over a bandlimited wireless channel. Real wireless channels are not so simple. They can be subject to fading, wherein the transmitted signal is attenuated more substantially at some times. One of the simplest models of fading has independent fades from time to time, where the fade is either 1 or 0. To further simplify the model, we will assume knowledge of the fade at the receiver. This model can be thought of as the AWGN+erasure channel in that we can consider the output to be "erased" whenever the fade is 0.

For this simplified model, rather than considering the fade as side information known at receiver, we can further simplify the model by viewing the fade as occurring after the additive noise. Since the channel noise in the model is continuous, the fade is immediately apparent whenever the output is 0.

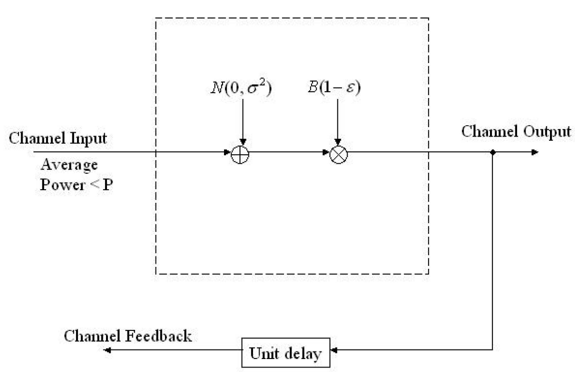

Fig. 1. The AWGN+erasure channel with feedback

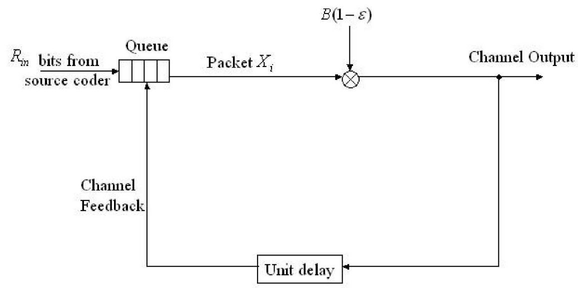

Fig. 2. Packet erasure channel with feedback, fed by a queue

Figure 1 shows the model we are studying in this submission. We assume noiseless feedback[1] so that we can study the feedback anytime reliability.

In [6], we showed that the feedback anytime capacity is the relevant sense of capacity when we are using a noisy channel in the feedback loop for controlling an unstable plant. The anytime capacity of a channel relates the bit error with the delay in a communication system where we require every bit to get through eventually. We review the definition of anytime capacity:[5]

Definition 1: $C_{anytime}(\alpha)$, the α-anytime capacity, is the max rate at which the channel can be used to communicate with a bit error probability that drops with delay exponentially at a rate of α.

$$C_{anytime}(\alpha) = \sup\{R|\exists \mathscr{E}^{\mathscr{R}}, K > 0, \forall N, \exists \mathscr{D}^{\mathscr{R}}_{\mathscr{N}}, P_{error}(\mathscr{E}^{\mathscr{R}}, \mathscr{D}^{\mathscr{R}}_{\mathscr{N}}) < K2^{-\alpha N}\} \quad (1)$$

In above definition $\mathscr{E}^{\mathscr{R}}$ is the anytime encoder, $\mathscr{D}^{\mathscr{R}}_{\mathscr{N}}$ is the decoder, and N is the delay that a bit experiences in units of channel uses. The parameter α is called the *anytime reliability.*[2] Since the probability of error on every single bit goes to zero with increasing delay, it is clear that the anytime capacity is always less than or equal to the classical Shannon capacity.

Fundamentally, what we have is a region of achievable (α, R) pairs — the region between the α axis and the anytime capacity curve. Whether we choose to look at maximizing R as a function of α or as maximizing α as a function of R is a matter of convenience for the problem at hand. Once we know the anytime capacity we know the anytime achievable region, and vice versa.

Essentially, to hold the η-moment of the state of an unstable plant finite, it is necessary and sufficient for the feedback channel's anytime capacity corresponding to anytime-reliability $\alpha = \eta \log_2 \lambda$ to be greater than $\log_2 \lambda$ where λ is the unstable eigenvalue of the plant.[6] Recall Kailath and Schalkwijk's famous paper [9] showing that the double-exponential convergence of probability of error to zero with delay was possible for the AWGN channel with feedback. This was extended in [5] to the streaming context (similar in spirit, though not details to Horstein's sequential transmission [3]) with no explicit block-length — so that every bit will eventually be decoded correctly.

In the companion submission [8] to this one, we study the feedback anytime reliability of the variable sized packet erasure channel as depicted in Figure 2. There, we show that despite having a moment constraint on the packet size, we could substantially improve anytime reliability by transmitting larger packets in the rare events when the queue of bits awaiting transmission was long. In this submission, we extend that basic strategy to the AWGN+erasure channel. Intuitively, the double-exponential convergence of the probability of bit error to zero possible with feedback over the AWGN channel allows us to treat the channel as a virtual packet erasure channel. Furthermore, the average nature of the power constraint allows us to treat the outgoing data rate as variable with an appropriate

[1] The feedback is delayed by one unit of discrete time to avoid any causality issues.

[2] The anytime reliability with noiseless feedback is fundamentally different from both the classical error exponents of Gallager and the exponents for variable delay decoding given by Burnashev[1]. See [7] for more discussion on this.

constraint. It is this strategy that is sketched out in this submission, with the detailed proofs being available in [10]. The results here are related in spirit to a specialized case of [2], though by going to the anytime framework, we are able to cover many more senses of stability than just second moment.

II. The main result and how to achieve it

Theorem 1: The feedback anytime capacity of the AWGN + erasure channel with average power constraint P, erasure probability ε, and noise variance σ^2 is

$$C_{\text{anytime}}(\alpha) = \begin{cases} \frac{(1-\varepsilon)}{2}\log_2\left(1+\frac{P}{\sigma^2}\right) & \text{if } 0 \leq \alpha < -\log_2 \varepsilon \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

$$\alpha^*(R) = \begin{cases} -\log_2 \varepsilon & \text{if } R < \frac{(1-\varepsilon)}{2}\log_2\left(1+\frac{P}{\sigma^2}\right) \\ 0 & \text{otherwise} \end{cases} \quad (3)$$

This result is illustrated in Figure 3. Since there is an $\varepsilon^d = 2^{d\log_2 \varepsilon}$ probability that the last d time steps were completely faded, we can do no better than guessing on any bit that was first received by the encoder d time steps ago. So $\alpha^* \leq -\log_2 \varepsilon$. The anytime capacity is always less than the Shannon capacity and so by the independence of the additive noise and multiplicative fade, we know that $C_{\text{anytime}}(\alpha) \leq \frac{(1-\varepsilon)}{2}\log_2\left(1+\frac{P}{\sigma^2}\right)$. The intersection of these two gives us the entire shaded region.

To achieve arbitrary points within this region, we use the approach illustrated in Figure 4. Bits arrive into a FIFO queue. If the queue is short (i.e. no bits have been waiting for too long), then a certain number of bits corresponding to a rate R_1 are tentatively taken out of the queue. If the queue is very long (i.e. many old bits are still awaiting a shot at transmission), then a larger number of bits $R_2 = nR_1$ are tentatively taken out of the queue. The bits that are tentatively taken out of the queue are used to drive a special simulated source connected to a joint source-channel encoder. If the feedback comes back and shows that the transmission was erased, the bits tentatively taken out are put back into the head of the queue and the evolution of the simulated source is backed out as though discrete time had not advanced at all. The reason that $R_2 = nR_1$ is so we can think of the long-queue behavior as attempting to make the simulated time run n times faster.

The state of the encoding system consists of two parts: the discrete valued state of the queue and the continuous valued state within the simulated source. The goal of the continuous valued state is to provide appropriate continuous valued inputs into the channel and taken together, to try and realize the packet-level abstraction shown in Figure 5. The analysis of that abstraction alone is given in [8].

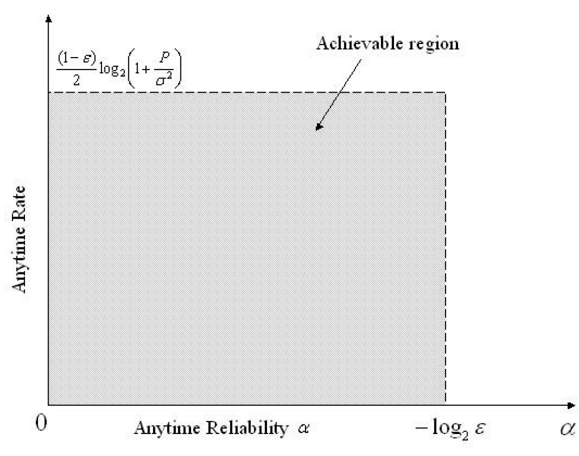

Fig. 3. The Anytime Capacity of The AWGN+Erasure Channel

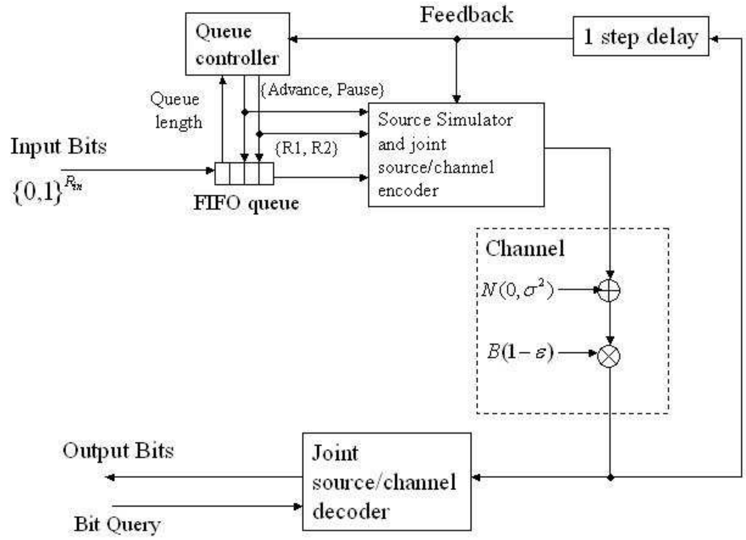

Fig. 4. Encoding and Decoding System Overview

Figure 6 shows the inside of the part that deals with the continuous valued state and figures 7 and 8 show the open-loop dynamics and actions of the controllers respectively. The open-loop dynamics are unstable with $a_2 = a_1^n$ to effectively let it take n time-steps all at once. The controllers apply the control designed to stabilize the system over the noisy feedback link when time should advance. When there is an erasure, the state is left unchanged.

The decoder (illustrated in Figure 9) does not have access to the exact input bits nor the exact continuous state of the encoder. The discrete state (queue length) depends only on the sequence of erasures so far and so the decoder does have access to that. In response to a query about the value of any particular bit, the decoder checks to see if that bit is still waiting in the queue. If not, it gives its best estimate of that bit's value.

To extract an estimate of the bit from the received channel outputs so far, the decoder maintains an internal state corresponding to how the encoder's state would have evolved if there had been no

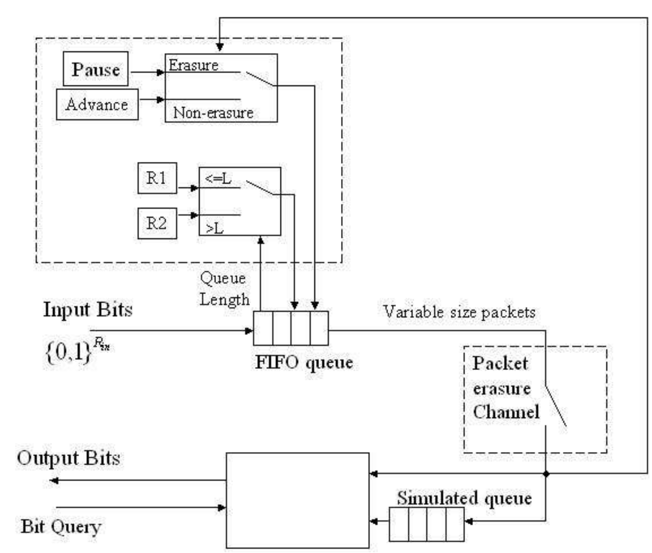

Fig. 5. Queue Level Abstraction: Bits are removed from the queue when a non-erasure happens and n-times larger packets are used when the queue is long.

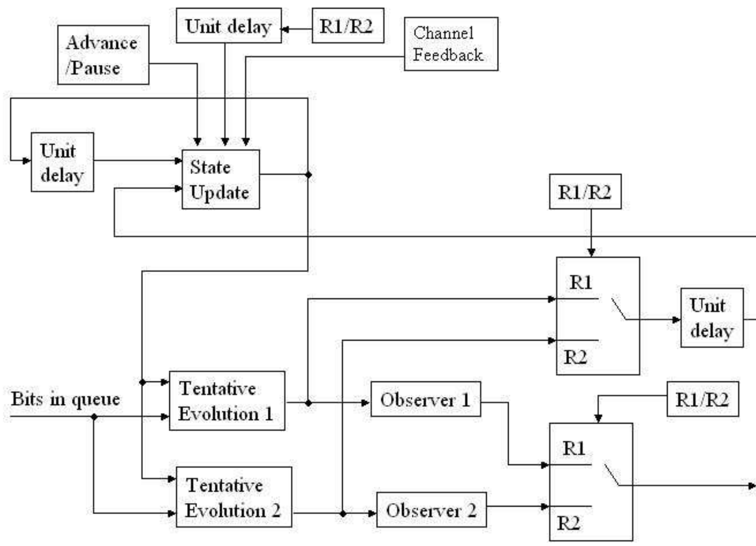

Fig. 6. Source Simulator and Joint Source/Channel Encoder

inputs. By linearity, the sum of the response of the system to only the controls and the response of the system to only the inputs is the actual state of the encoder. Although both of those terms individually represent the outputs of unstable processes, their sum is stable. Thus, the response to the controls alone has to track closely the response to the inputs alone. As described more fully in [5] and [6], the unstable state can be thought to consist of an integer whose binary expansion is the desired data bits.[3]

This system works because once a bit is out of the queue, the decoder's estimate of it will converge doubly exponentially in the number of subsequent non-erased channel outputs. As such, the dominant source of errors is the erasures. The average power constraint is met by making the "long queue" behavior rare enough by choosing a suitably large threshold L between the two behaviors.

III. Simulation illustrating behavior

In our system, there are three different notions of internal time:

- Real time: t_r which advances by one tick with every channel use. Bits are arriving in real time and so is the discrete encoder state.
- Simulation time: t which only advances by one tick with every non-erased channel output. The continuous state evolves only in simulation time.

[3]To see this, consider $a_1 = 2$ and $a_2 = 8 = 2^3$. Then the unstable state can hold a single bit per unit time when a_1 is used and three bits per unit time when a_2 is used.

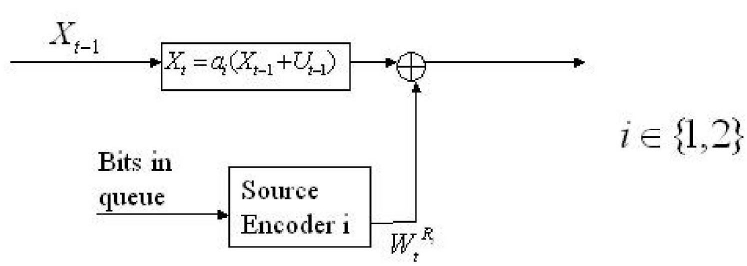

Fig. 7. Tentative Evolution: Unstable Open-Loop Dynamics

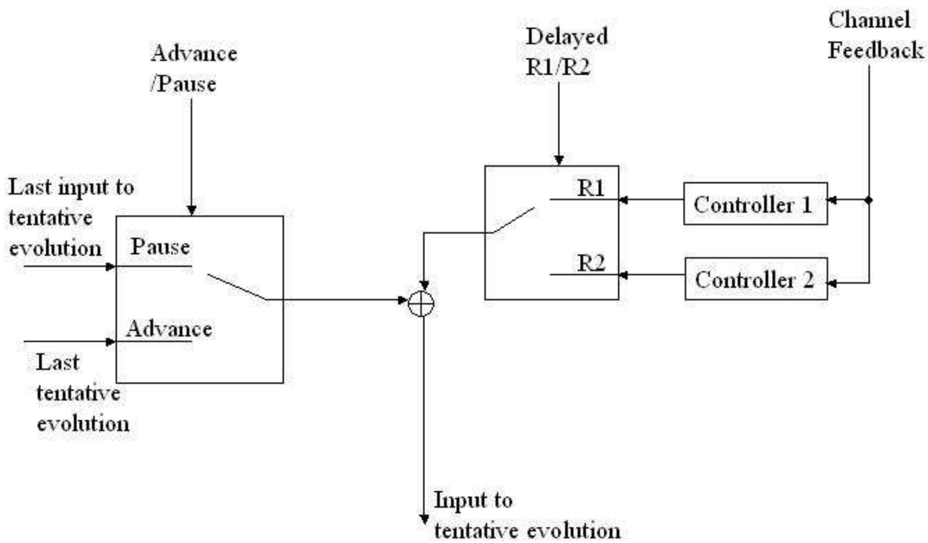

Fig. 8. State Update: Advancing Time and Applying Controls

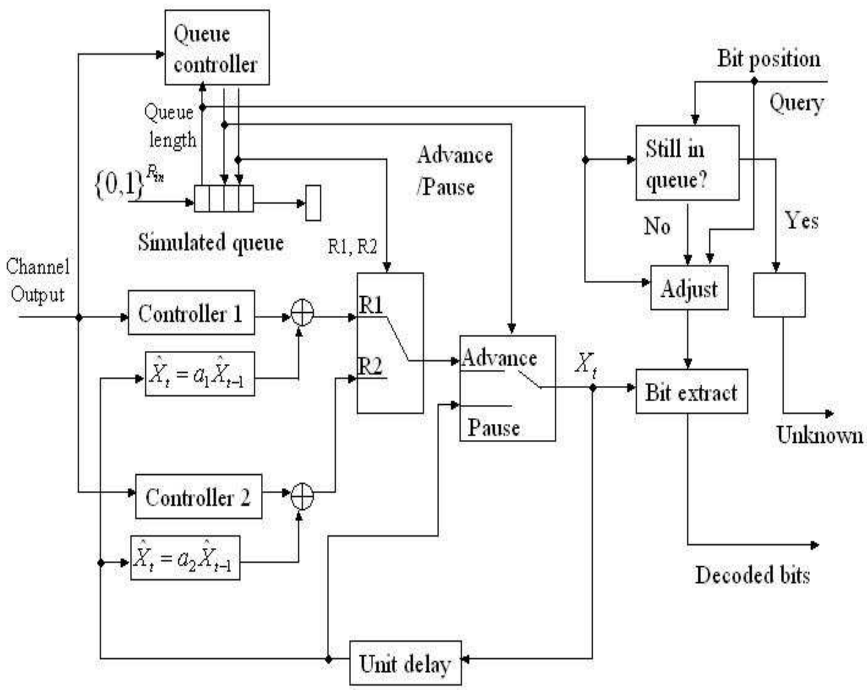

Fig. 9. Decoder

- Virtual time: t_v which indicates the last bit to have become incorporated into the continuous state. Virtual time can only advance with the simulation time, but it sometimes moves faster than simulation time if the queue is long.

The difference between $R_{in}t_r$ and t_v represents the number of bits still waiting in the queue. The reason we have the speedup in the encoder is to help keep this difference small since it is the queuing delay that causes the dominant term in the probability of bit error.

Figure 10 shows the lag between virtual time and real time for a simulated run and compares the case with speedup to the case without it. We can see that before the first non-erasure after the system enters the long queue mode for the first time, the lag in the two cases are the same. But when a

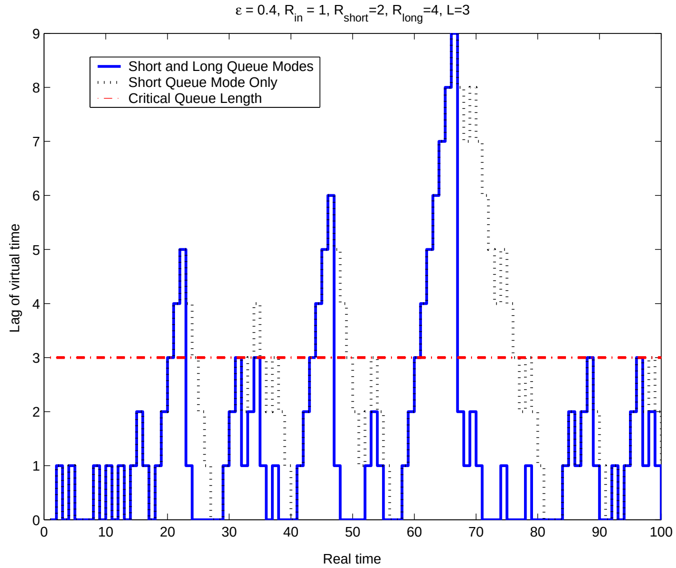

Fig. 10. Lag between Real Time and Virtual Time with one mode and two modes: $\varepsilon = 0.4$, $R_{in} = 1$ bit, $R_{short} = 2$ bits, $R_{long} = 4$ bits, $L_c = 3$

successful transmission occurs, the two-mode system transmits more bits than the one-mode system, and reduces the lag faster.

Figure 11 shows the average power corresponding to the channel realization in Figure 10. We can see that when the system stays in the short queue mode, the average[4] transmission power increases gradually and converges to P_1, and when the system stays in the long queue mode, the average transmission power is $P_2 > P_1$ due to the higher data rate. At the instant the system transits from the long queue to short queue mode the average power is $P_1^{trans} < P_1$ by design. As the system stays in the short queue mode, the power increases from P_1^{trans} and converges back up to P_1. Similarly the

[4]The average in this figure is taken over the realizations of the Gaussian channel noise. The erasure noise corresponds to one particular realization.

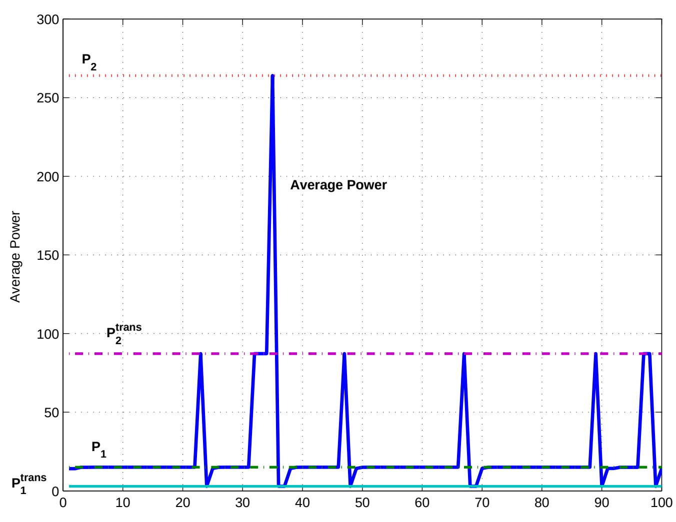

Fig. 11. Average Power: $\varepsilon = 0.4$, $R_{in} = 1$ bit, $R_{short} = 2$ bits, $R_{long} = 4$ bits, $L_c = 3$

average power changes to P_2^{trans} at the instant the system transits to the long queue mode. Then the average power goes to P_2 in the long queue mode until the system transits back to the short queue mode. As such, it is P_1 which dominates the overall average transmission power of the system as the P_2^{trans} or P_2 spikes can be made as rare as we would like by increasing L_c.

IV. Proof Sketch

Lemma 1: Let the input rate be R_{in} and the system have this queuing rule:

- *Short queue mode*: When the queue length is smaller than L_c, in units of R_{in}, the system removes $R_1 > \frac{R_{in}}{(1-\varepsilon)}$ bits per non-erasure from the queue.
- *Long queue mode*: When the queue length is larger than L_c, in units of R_{in}, the system removes $R_2 > R_1$ bits per non-erasure from the queue.

Then the tail probability the queue length is bounded:

$$P(L > dR_{in}) \leq T_1 2^{-\alpha_1^* d}$$

where α_1^* is the feedback anytime reliability of the R_1-sized packet erasure channel corresponding to a rate of R_{in} from [8] and T_1 is some positive constant that does not depend on L_c.

Furthermore, when the queue length $L > L_c R_{in}$ bits, the tail probability has a tighter bound:

$$P(L > dR_{in}) \leq T_2' 2^{-\alpha_2^* d}$$

where α_2^* is the feedback anytime reliability corresponding to a rate of R_{in} using a R_2-size packet erasure channel and T_2' is some positive constant that depends on L_c.

Proof: While the full proof can be found in [10] and [8], the key insight behind the first part is that this queue must work at least as well as one that only uses the slower service rate. Since the queue is stable even at the first service rate, the tail probability is bounded by an exponential. For the second part, the probability of very large queue lengths dies at a rate determined by the higher service rate used in that regime. As a result, the anytime reliability will be dominated by the higher service rate, while the average power used will be dominated by the lower service rate. At the cost of a constant delay, we get the best of both worlds! ■

The next key insight has to do with the role of virtual time. By suitably choosing the two unstable modes to have $a_2 = (a_1)^n$ and using $R_2 = nR_1$ bits, we can make each step of the faster mode feel like n steps of the slower mode, except taken all at once and thereby using substantially more power. Since we are already burning substantial power in the fast mode, we feel free to burn even more and use a control strategy designed to drive the controlled state to zero in one step rather than trying to minimize the channel input power.

By doing this, we see that:

1) Without noise n_t, the un-controlled state of the system depends only on the virtual time.
2) The controlled state depends on the virtual time, the data to be transmitted, and the channel noise since the last queue mode transition.

The last point is true because in the fast mode, the only channel noise that matters is the one at the immediate past time. With these, we get:

Lemma 2: The probability of bit error at the i-th bit with delay d and queue length l, denoted as $P_{error}(d,l,i)$, is bounded by

$$P_{error}(d,l,i) \leq f\left(\frac{\varepsilon_1}{2+2\varepsilon_1} 2^{(\log_2 a_1)d-l}\right)$$

where $f(\cdot) \triangleq \max(f_1(\cdot), f_2(\cdot))$ where f_1 is the distribution that bounds the tail of the controlled system state when only the slow dynamics are used, and f_2 is the counterpart for the fast dynamics.

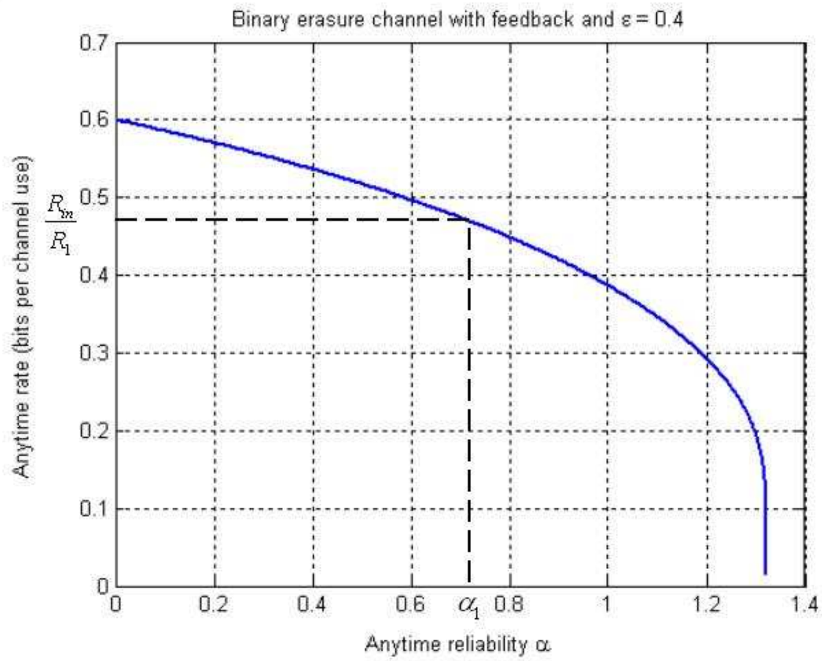

Fig. 12. Finding α_1 in the anytime capacity curve of binary erasure channel

Proof: This is a consequence of Theorem 3.3 in [6] or equivalently, Corollary 7.1.2 in [5], together with the realization that the lag between the virtual time and the real time is the queue length. See [10] for more details. ■

Now, f_1 and f_2 are not just anything. Both represent the probability of large deviations for a control system driven by a bounded disturbance with Gaussian noise in the observation equation. The tail probability is dominated by the Gaussian and is therefore at least exponential. The maximum of two exponentials is bounded by the larger exponential. So the above tells us that the probability of error is dying at least doubly exponentially in $(\log_2 a_1)d - L$ where d is the delay and L is the random variable denoting the queue length.

As such, the dominant term is going to be the probability of having a very long queue and Lemma 1 tells us that is dominated by α_2 coming from the rare high-power mode of operation. All that remains is to give the order of setting the parameters $R_1, a_1, \alpha_1, R_2, a_2, P_1, P_2$ and L_c to achieve a target (R_{in}, α) pair within the achievable region. Without loss of generality, change units so that $\sigma^2 = 1$ so P is the average SNR constraint.

1) P_1 Select an arbitrary P_1 such that $2^{\frac{2R_{in}}{1-\varepsilon}} - 1 < P_1 < P$. This is always possible since the given R_{in} is in the achievable region.
2) R_1 and a_1: $R_1 = \frac{R_{in}}{1-\varepsilon} + \delta_1$, where δ_1 is an arbitrarily small positive number. Select an a_1 which satisfies $2^{\frac{R_{in}}{1-\varepsilon}+\delta_1} < a_1 \leq \sqrt{P_1 + 1}$. This will enable our channel input power to always be less than P_1 on average whenever we are in the short queue mode.
   The resulting α_1 is obtained by looking up the anytime capacity curve of the R_1-size packet erasure channel with noiseless feedback at rate R_{in}. You can also do this by using Figure 12.
3) α_2: Choose α_2 such that $\alpha < \alpha_2 < -\log_2 \varepsilon$.
4) n and R_2: Choose n such that $R_2 = nR_1, n \in \mathbb{Z}^+, n > 1$ with the feedback anytime reliability of the R_2-size erasure channel evaluated at R_{in} being α_2 or larger. This can be done graphically by looking up $R = \frac{R_{in}}{nR_1}$ on the binary erasure anytime capacity given by Figure 12.
   $a_2 = a_1^n$ and set $P_2 = a_2^2(1 + C_2^2)$ where C_2 is the gain inside the second observer and is set high

enough so that the channel noise looks quite small in comparison. Details are in [10].

5) L_c: Let $P_{2,max} \triangleq \max(P_2, P_2^{trans})$ where P_2^{trans} is the transient power when making a short queue to long queue transition. $P_{2,max}$ does not depend on L_c. We require

$$\begin{aligned} E\{y_t^2\} &= Prob(l \leq L_c)P_1 + Prob(l > L_c)P_{2,max} \\ &\leq P_1 + Prob(l > L_c)P_{2,max} \\ &\leq P_1 + T_1 2^{-\alpha_1 L_c} P_{2,max} \\ &\leq P \end{aligned}$$

Since P_1 is selected to be smaller than P, there are always L_c's to hit the last inequality.

V. Conclusion and Future Work

By using more transmit power when we have many bits awaiting transmission, we can achieve any desired point in the anytime reliability region corresponding to having the feedback anytime reliability as close to $-\log_2 \varepsilon$ as we would like for all rates up to the Shannon capacity of the AWGN+erasure channel. This is because the doubly-exponential vanishing of the probabilities of error with delay lets us effectively conceptualize the channel as a noiseless packet erasure channel, though there are many technical steps along the way with the full details available in [10]. The corresponding result for variable size packet erasure channels is given in the companion submission [8], where we also imposed a peak-packet size constraint.

We believe that this style of analysis can be extended to give us the anytime reliabilities of many Gaussian channels with noiseless feedback and channel state side information available at both the transmitter and receiver. In particular, we believe it can be extended to cover the case of finite state Markov-fading channels[5] with side information and thereby give a variation of Kailath-Schalkwijk[9] style error-convergence for some such channels. In addition, for control problems over noisy channels, it should be possible to take the hybrid data-communication strategy we have developed here and turn it into a hybrid-control strategy that works with the system state directly. This may better allow the stabilization, though not necessarily in the second-moment sense, of unstable systems over fading Gaussian channels.

References

[1] M. V. Burnashev. Data transmission over a discrete channel with feedback, random transmission time. *Problemy Perdachi Informatsii*, 12(4):10–30, 1976.

[2] N. Elia. Feedback stabilization in the presence of fading channels. *Proceedings of the American Control Conference*, pages 4438–4443, June 2003.

[3] M. Horstein. Sequential transmission using noiseless feedback. *IEEE Trans. on Information Theory*, 12:448–455, October 1966.

[4] J. Liu, N. Elia, and S.C. Tatikonda. Capacity-achieving feedback scheme for markov channels with channel state information. *Proceedings of the International Symposium on Information Theory*, 2004.

[5] A. Sahai. *Anytime Information Theory*. PhD thesis, Massachusetts Institute of Technology, February 2001.

[6] A. Sahai. The necessity and sufficiency of anytime capacity for scalar control over a noisy communication link. *To appear at the CDC*, 2004.

[7] A. Sahai and T. Şimşek. On the reliability function of discrete memoryless channels with feedback and variable-delay decoding. *To Appear in the Information Theory Workshop*, 2004.

[8] A. Sahai and Q. Xu. The anytime reliability of constrained packet erasure channels with feedback. *Also appeared in the Allerton Conference*, 2004.

[9] J. P. M. Schalkwijk and T. Kailath. A coding scheme for additive noise channels with feedback – i: No bandwidth constraint. *IEEE Trans. on Information Theory*, 12:172–182, April 1966.

[10] Q. Xu. The anytime capacity of AWGN+erasure channel with feedback. Master's thesis, University of California at Berkeley, August 2004.

[11] S. Yang, A. Kavcic, and S.C. Tatikonda. Linear gaussian channels: Feedback capacity under power constraints. *Proceedings of the International Symposium on Information Theory*, 2004.

[5] Of the type studied by [4] and perhaps even those of [11]. The results given in those were capacity achieving, but because of their reliance on typicality, they did not achieve the kind of reliability exponents that we believe are truly possible in the limit.