Anytime communication over the Gilbert-Eliot channel with noiseless feedback

Anant Sahai, Salman Avestimehr, Paolo Minero
Department of Electrical Engineering and Computer Sciences
University of California
Berkeley, CA 94720-1770, USA
Email: {sahai,avestime,minero}@eecs.berkeley.edu

***Abstract*— We study the reliability of sequential codes in a two-state Markov fading AWGN channel under the assumption of noiseless feedback and an average power constraint. We present a capacity achieving scheme with a doubly exponential anytime reliability function with respect to delay for every bit. The scheme is represented by a hybrid control system at the encoder in which the discrete system dynamics evolves based only on the channel state information while the continuous part of the state at the encoder reflects the evolution of the message uncertainty at the decoder. Whereas the classical Schalkwijk-Kailath scheme achieves double exponential reliability by exploiting the average nature of the power constraint to combat atypicality of the AWGN noise, our scheme also uses it to combat atypical fading realizations.**

I. Introduction

In this paper we propose a capacity-achieving sequential scheme with doubly exponential reliability in delay for a two-state Markov channel with channel state information available both at the receiver and, with unit delay, at the transmitter. In addition, we assume that noiseless feedback of the receiver's signal is available at the transmitter with a similar unit delay.

This channel models communication over a bandlimited wireless channel subject to fading, where the fades are not independent in time and the channel can be in a good or bad fade. The capacity of a discrete-time Markov channel with feedback has been studied in [7] and references therein. In [5], the authors presented a feedback communication scheme for finite-state Markov channel that can be viewed as a multiplexed Schalkwijk-Kailath scheme. However, atypical channel realizations were not considered in the analysis in [5], while in fact they are the dominant source of errors for that scheme. In this paper we show that the capacity for sequential communication (i.e. streaming) is equal to the capacity for nonstreaming showed in [5]. A sequential (anytime) generalization of the Schalkwijk-Kailath [1] scheme for the AWGN channel with feedback was introduced in [2] and was extended to the AWGN+erasure channel in [4]. These had message bits arriving at the encoder sequentially and required the encoder to be causal. The decoder produced updated estimates for all message bits so far and we required that the probability of error on each message bit goes to zero with increased delay. The rate of convergence to zero is considered the anytime reliability.

In this paper, we give a sequential scheme for anytime

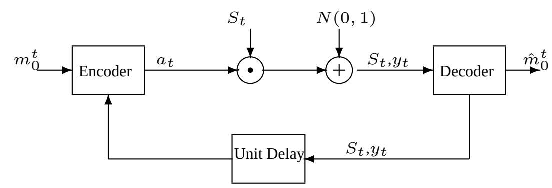


Fig. 1. Channel model.

communication over the two-state Markov fading AWGN channel with noiseless feedback. The code involves a hybrid control system consisting of a queue where the server can adjust its service rate based on the CSI and the number of bits waiting in the queue. The server pumps bits into a linear time varying system where the dynamics also change according to the channel state, and the noiseless feedback is used to apply a control signal so that the system stays stable. The decoder uses the CSI to run an unstable version of the encoder's system and reads off the message bit estimates from the state in the appropriate binary notation. This results in:

Theorem 1.1: We can communicate sequentially arriving bits reliably across a power constrained two-state Markov fading channel with noiseless feedback and perfect CSI at all rates $R < C$ while achieving at least doubly exponential anytime reliability in decoding delay:

$$\mathbb{P}\left\{\hat{m}_1^{t-d}(t) \neq m_1^{t-d}\right\} \leq c_1 e^{-c_2(e^{c_3 d})} \tag{1}$$

for some positive constants c_1, c_2 and c_3.

In Section II, we introduce the problem and then give the intuition behind the entire scheme in Section III. Section IV discusses the control model, while Section V gives the details of how bits are encoded and decoded from the control model. The proofs are omitted for lack of space, but will be found in the full version of this paper.[6]

II. Communication model

The feedback communication scheme we are considering is shown Figure 1. Messages m_i arrive one at a time and the decoder at time t produces estimates $\hat{m}_1^t(t)$ for all the past messages at each time. The Gilbert-Eliot fading process is modeled as the two-state Markov chain in Figure 4(a). The receiver has immediate access to the channel state, while the encoder has one-step-delayed information of both the fading state and the channel output. The channel state process S_t is stationary and ergodic with state space $\{G, B\}$. The one-step transition probabilities are given in Figure 4(a). The unique steady state distribution is denoted as $\pi = [\pi_B, \pi_G]$. We also consider the Markov chain corresponding to consecutive pairs of elements from the Markov chain in Figure 4(b). The space is $\{G, B\}^2$, the transition probabilities are shown in Figure 4(b), and we denote the steady state distribution as π_{ij} for $i, j \in \{G, B\}$. We assume that the noise process N_t is independent Gaussian with mean zero and unit variance. As shown in [7], the capacity of this channel, subject to an average input power constraint P, is given by

$$C = \max_{P} \frac{1}{2} E_{S_k S_{k-1}} \log\left(1 + S_k P(S_{k-1})\right) \qquad (2)$$

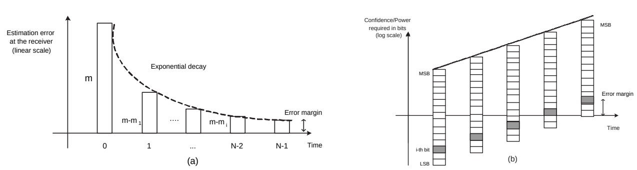
such that $E_S P(S) \leq P$. The optimal capacity-achieving power allocation consists of waterfilling as a function of the CSI. We denote this allocation as P_G and P_B. Hence, $C = \frac{1}{2} \sum_{i,j \in \{G,B\}} \pi_{ij} \log\left(1 + S_j P_i\right))$ where $\pi_B P_B + \pi_G P_G = P$.

III. Intuition

Because space is too short for proofs, it is important to present the intuition behind the result in a clear way.

A. Schalkwijk-Kailath for AWGN

We first review the Schalkwijk-Kailath scheme of [1]. Here there is no fading, and this is a block-coding scheme with the entire message known at the beginning. This message is encoded into binary and considered a real number $m \in [0, 1]$. In the first channel use, m is transmitted by scaling it to meet the power constraint. Since the channel is noisy, the receiver cannot decode perfectly the message m. Instead, it computes the ML estimate $\hat{m}$ of the original message at the price of some Gaussian estimation error that the transmitter tries to reduce in the remaining $N-1$ transmissions. After each transmission the message uncertainty at the receiver is reduced by a fixed



(a) Message uncertainty decreases exponentially.

(b) Confidence in each bit grows exponentially (log scale).

Fig. 2. Two pictorial views of the Schalkwijk-Kailath scheme for the AWGN channel

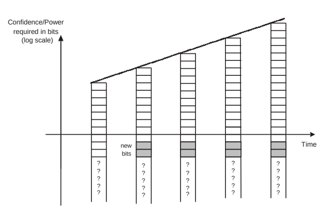
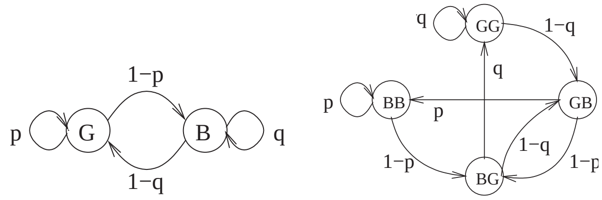


Fig. 3. Sequential Schalkwijk-Kailath scheme for the AWGN channel.



(a) Channel fading model

(b) Two consecutive channels

Fig. 4. Two-state fading process.

factor, proportional to the signal-to-noise ratio (SNR) at the receiver. Since the tail of a Gaussian is exponentially small, we get a doubly exponential reliability function.

We think of the Schalkwijk-Kailath scheme in another equivalent way. Order the bits in message m from the most significant bit (MSB) to the least significant bit (LSB). After the first channel use some uncertainty about the message remains at the decoder. We can imagine that the MSB in m have been decoded, while the LSB still need to resolved. As the scheme continues, each bit get more and more protection. The final error probability is dominated by getting an error on the LSB. This is illustrated in Figure 2(b) as an escalator that takes the reliability of each bit higher as time goes on.

B. Sequential Generalization for AWGN

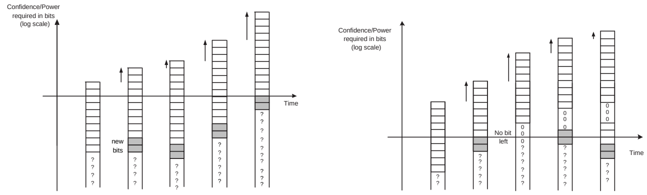
Although the derivation of the sequential scheme in [2] is done in terms of control systems, it can be understood directly by looking at Figure 3. Suppose that the message m had an infinite number of bits in it and we ran the Schalkwijk-Kailath scheme forever. From the point of view of the first bit, the story is effectively unchanged. It still gets a lot of protection very soon. In fact, any given bit of the message eventually gets arbitrarily well protected. Furthermore, it is clear that the very late bits do not really impact the initial transmission.

Thus, we can construct a “lazy” Schalkwijk-Kailath encoder that just sends the first few bits of the message initially and then tweaks its subsequent transmissions to make the receiver correctly estimate all the bits eventually. At any given time step, the analog level of this new perturbation can be made quite small and so this does not significantly impact the power consumption. This can be used to get around the limitation of only seeing the message unfold in real time. When message bits come in, they can be used to refine the message uncertainty with only a small perturbation.

Pictorially, the Schalkwijk-Kailath scheme is an escalator that steadily takes message bits to higher and higher reliabilities. It does not matter if the escalator is fully populated at the beginning or whether new bits get on at a suitably low floor.

C. Generalization to Gilbert-Eliot

For the Gilbert-Eliot channel with CSI at the receiver and unit-delayed CSI at the transmitter, the Schalkwijk-Kailath escalator moves at a variable speed. When the channel is good and we used a lot of power to transmit the current error signal, the message uncertainty at the decoder decreases by a larger factor than when the channel is bad or we used



(a) Bits are scaled by four different factors depending on the channel state.

(b) Dummy bits are inserted when the increments get too large.

Fig. 5. Sequential Schalkwijk-Kailath scheme for the two-state Markov channel.

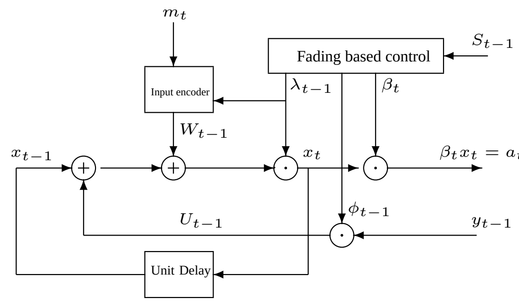


Fig. 6. Encoder structure. Encoder state is in x_t and past history of S_t.

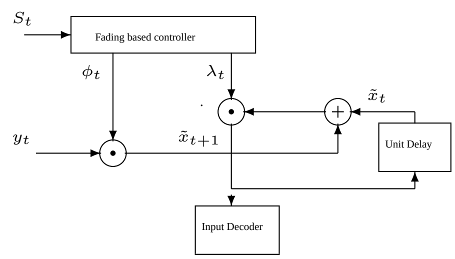


Fig. 7. Decoder structure. Decoder state is in x_t and past history of S_t.

less transmit power.[1] This variability is illustrated in Figure 5(a). We can still work sequentially by loading bits onto the escalator sequentially at a low floor so that they do not impact the overall power.

There are two challenges: first, we could have a run of good luck when the channel is in a good state for a long time. In that case, the escalator will have advanced by a lot while we have only had a small number of new message bits arriving at the encoder. If we try to load these message bits onto the escalator at the first open space, it would cause a large perturbation. This is resolved by just letting some spaces on the escalator go empty — fill them with 0's but insist on having the new bits get on board at a low floor so as not to perturb the power consumption by too much (Figure 5(b)). Because the escalator motion depends only on the fading sequence realization, it is known perfectly to both the transmitter and the receiver. Similarly, the message bit arrivals are deterministic and hence also known perfectly to both sides. Thus each one knows which positions in the escalator correspond to which message bits.

The second challenge is more serious. We could encounter a bad sequence of fading states which makes the escalator move very slowly. This results in a backlog of bits to transmit that still have not gotten on the escalator, or alternatively, are getting on the escalator at very low floors in the sub-basement. In such cases, we apply a special "turbo mode." When the fade has been bad for a while, we increase the transmit power even in the bad case so that the escalator moves up faster than the bits are arriving. This ensures that a bit has to spend only a bounded amount of time waiting before it gets a chance to ride the escalator. Because a long string of bad fades is very unlikely, this extra power does not make us violate our average power constraint.

The combination of these tricks allows every bit to get on the reliability escalator soon and thus experience the doubly exponential drop in error probability with delay shown in Theorem 1.

IV. Control model

In this Section we introduce the discrete-time control problem for stabilizing an unstable scalar linear hybrid system that is driven by a control sequence $\{U_t\}$ and a bounded disturbance $\{W_t\}$. The noisy channel in the feedback loop is modeled as the same two-state Markov AWGN channel as in the communication problem, and the state information is known casually at the decoder/controller and, with unit-delay, at the encoder. The closed loop state dynamic is described by the following time-varying linear equation

$$X_{t+1} = \lambda_t(X_t + W_t + U_t) \tag{3}$$

where $\{X_t\}$ is a $\mathbb{R}$-valued state process and the disturbance $|W_t| \leq \Omega$. We assume that the initial condition is known to be $X_0 = 0$. At each step the state of the plant is scaled with one of four different scalars $\{\lambda_{GG}, \lambda_{GB}, \lambda_{BG}, \lambda_{BB}\}$, depending on the channel state in the Markov chain in Figure 4(b). For instance, $\lambda_t = \lambda_{GB}$ if $S_{t-1} = G$ and $S_t = B$. We consider $\lambda_{i,j} > 1$ for $i, j \in \{B, G\}$ so the open-loop system is unstable. Define $\lambda_{max} = \max_{i,j} \lambda_{ij}$ for $i, j \in \{G, B\}$. The corresponding system can be thought as a hybrid control system with a common state whose dynamics depends on the channel state.

The control problem consists of designing an observer/encoder and a controller/decoder such that the closed loop system is stable and the average power constraint (II) is satisfied. We will use a linear observer and a linear controller, so the problem consists of choosing two time-varying coefficients, β_t at the observer and ϕ_t at the controller. The encoder and decoder structure are shown in Figure 6 and Figure 7. The control and the anytime communication problems are intimately connected, as shown in [3]. The goal is to reliably communicate at rate $R < C$ encoding the message bits causally into the disturbance W_t, so we can interpret the closed-loop state of the plant as the receiver's uncertainty

[1]There are 4 such cases depending on the transition we experience.

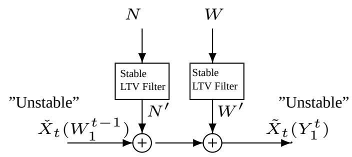


Fig. 8. N and W in stable filters.N' is Gaussian with variance K, while W' is bounded by Λ.

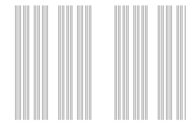


Fig. 9. The data bits are used to sequentially refine a point on a Cantor set. The Cantor set has finite gaps between all points corresponding to bit sequences that first differ in a particular bit position.

regarding previous bits.

We think of the state of the plant as the sum of the states of two different unstable linear time-varying plants. By linearity, we decompose the plant state as

$$X_t = (\tilde{X}_t + \check{X}_t). \tag{4}$$

It is assumed that $\tilde{X}_t$ is driven entirely by the control sequence U_t, while the plant with state $\check{X}_t$ is driven by the bounded disturbance W_t. Thus, $\tilde{X}_{t+1} = \lambda_t(\tilde{X}_t + U_t)$ and $\check{X}_{t+1} = \lambda_t(\check{X}_t + W_t)$. Similarly. we consider the following decomposition into two closed-loop systems:

$$X_t = (\breve{X}_t + \bar{X}_t), \tag{5}$$

where $\breve{X}_t$ is the closed-loop system driven entirely by the the Gaussian noise N_t, while $\bar{X}_t$ is driven by the bounded disturbance W_t.

As shown in Figure 8, $\tilde{X}_t$ can be thought as a noisy version of $\check{X}_{t+1}$.

V. Encoding and extracting data sequentially

Assume that the linear observer/controller pair stabilizes[2] the system with the closed loop equation (3). We now show how to use the noiseless feedback to communicate bits sequentially at rate $R < C$.

Following [3] the encoder sets

$$W_t = \gamma(\prod_{i=0}^{t} \lambda_i)(\sum_{k=\lfloor Rt \rfloor+1}^{\lfloor R(t+1) \rfloor} (2+\epsilon_1)^{-k} B_k)(2+\epsilon_1)^{-c_t} \tag{6}$$

where B_ks are ± 1 valued bits that we want to transmit, the constant γ and the integer-valued c_t are chosen to make the disturbance bounded. The sequence c_t must be known at the decoder as it is necessary for decoding the bits. We update c_t as follows $c_t = c_{t-1} + \text{inc}_t$ where inc_t is the smallest nonnegative integer [3] that satisfies

$$\gamma(\prod_{i=0}^{t} \lambda_i)(\sum_{k=\lfloor Rt \rfloor+1}^{\lfloor R(t+1) \rfloor} (2+\epsilon_1)^{-k})(2+\epsilon_1)^{-(c_{t-1}+\text{inc}_t)} < \Omega \tag{7}$$

From (7) it is clear that c_t only depends on $\{\lambda_i\}_{i=0}^{i=t}$ which are causally known at the decoder.

[2]Stabilization, involving the choice of the sequences $\{\phi_t\}$ and $\{\beta_t\}$, proceeds by using an MMSE control in short queue mode and a zero-forcing control in turbo mode. Standard Kalman filtering arguments show that the system is stable.

[3]Increasing inc_t makes the LHS of (7) smaller, so a nonnegative inc_t can always be found.

Consider the system decomposition in (4). $\check{X}_t$ is known to the decoder since it only depends on the control sequence U_t. Since the original observer/controller pair stabilizes the original system, we know that $|X_t| = |\check{X}_t - (-\tilde{X}_t)|$ remains small and hence $-\tilde{X}_t$ stays close to $\check{X}_t$.

This is why the sequential data encoded in the disturbance W_t can be decoded at the decoder based on $-\tilde{X}_t$. If we look at the system driven by the disturbance only, the following proposition characterizes the state of the plant at time t.

Proposition 5.1: By applying the disturbance in (6), to the system driven only by the disturbance, we have

$$\check{X}_t = \gamma(\prod_{i=0}^{t-1} \lambda_i) \sum_{k=0}^{\lfloor Rt \rfloor} (2+\epsilon_1)^{-(k+c_{\lfloor \frac{k}{R} \rfloor})} B_k$$

or

$$\check{X}_t = \gamma(\prod_{i=0}^{t-1} \lambda_i) \sum_{j=0}^{\lfloor Rt \rfloor + c_t} (2+\epsilon_1)^{-j} \bar{B}_j$$

where $B_i = \bar{B}_{i+c_{\lfloor \frac{i}{R} \rfloor}}$ for $i = 0, \ldots, \lfloor Rt \rfloor$ and $\bar{B}_j = 0$ in other positions.

This shows that $\check{X}_t$ takes value in a Cantor set[4] (Figure 9). Bits are decoded by figuring out the nearest neighbor in the Cantor set to the current $-\tilde{X}_t$. All the points in the Cantor set are separated by finite gaps, and the error performance is determined by the minimum gap between all points (Figure 11).

Proposition 5.2: The minimum gap between two $\check{X}_t$ having the same B_1^{i-1} prefix but differing in the i-th position is bounded below by $\gamma(\prod j = \frac{i}{R}^{t-1} \lambda_j) 2(1 - \frac{1}{1+\epsilon_1})$. This gap is called the error gap of position i at time t or $\text{gap}_i(t)$.

Since the gap grows, we can define:

Definition 5.1: The level of the $k-th$ bit at time $t > kR$ in the state of the system denoted by $L_{k,t}$ is:

$$L_{k,t} = \gamma(\prod_{i=0}^{t-1} \lambda_i)(2+\epsilon_1)^{-(k+c_{\lfloor \frac{k}{R} \rfloor})} \tag{8}$$

Lemma 5.3: The level $L_{k,t}$ evolution depends only on the fading process and not on the Gaussian noise.

Proposition 5.4: The level of a bit in the state of the system is an increasing function of time t.

Proof: $L_{k,t}$ has two parts: $(2+\epsilon_1)^{-(k+c_{\lfloor \frac{k}{R} \rfloor})}$ which is independent of t and $\gamma(\prod_{i=0}^{t-1} \lambda_i)$ which is an increasing function of t (as $\lambda_i > 1$). ■

[4]The Cantor set is used to avoid technical difficulties resulting from the "carry rippling" possible when real numbers are represented in binary. It bounds the ability of future bits to cause an error in decoding earlier ones.

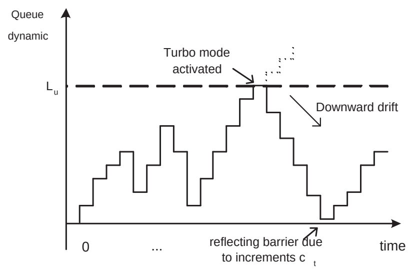


Fig. 10. Queue dynamic. Turbo mode limits the queue length, while c_t acts as a reflecting barrier.

So when a bit enters the system, it has an initial level and as the time goes on its level also increases. This suggests a FIFO queue model for the bits in the system: a bit enters the queue when it arrives and leaves the queue when its level in the state of the plane rises above a certain threshold, T_L. The queue size can be viewed as a random walk given by

$$Q_{t+1} = Q_t + R + \text{inc}_t - \log \lambda_t \tag{9}$$

where the inc_t provide a reflecting barrier at zero that does not let the random walk go under the zero level as is justified in the following proposition:

Proposition 5.5: Let the threshold $T_L < \frac{\Omega}{\lambda_{max}}$. If $\text{inc}_t > 0$, then no bit is waiting in the queue, i.e. $Q'_t = 0$.

Proof: We argue by contradiction. If there is at least one bit left in the queue, then the level of the last bit is less then the threshold T_L, so

$$L_{\lfloor Rt \rfloor, t} = \gamma(\prod_{i=0}^{t-1} \lambda_i)(2+\epsilon_1)^{-(\lfloor Rt \rfloor + c_t)} \leq T_L \tag{10}$$

From (6) we have the following bound on the disturbance

$$\begin{aligned} W_t &\leq \gamma(\prod_{i=0}^{t} \lambda_i)(\sum_{k=\lfloor Rt \rfloor+1}^{\lfloor R(t+1) \rfloor} (2+\epsilon_1)^{-k} B_k)(2+\epsilon_1)^{-c_t} \\ &\leq \gamma(\prod_{i=0}^{t} \lambda_i)((2+\epsilon_1)^{-(\lfloor Rt \rfloor + c_t)} \leq \lambda_{max} T_L < \Omega. \end{aligned}$$

which implies $\text{inc}_t = 0$ and establishes the contradiction. ■

Turbo Mode: To deal with atypical fading realizations:

Definition 5.2: When the size of the queue is greater then a threshold T_U, then the system enters *turbo mode*, such that it serves at rate larger than incoming rate, R. This is accomplished by choosing system parameters so that $\log \lambda_t > R$ no matter what the channel state does.

Turbo mode forces the number of the bits in the queue to remain bounded and thus allows us to bound the waiting time of a bit in the transmission queue.

Proposition 5.6: By using turbo mode, the time between the arrival and departure of a bit from the queue is bounded above by T_{max}.

Proposition 5.7: By using the turbo mode we can guarantee that $\prod_{i=t-j}^{t} \frac{\lambda_i}{\lambda} > k$ for all $j > j_{min}$ where k is a fixed number depending on the threshold set for the turbo mode.

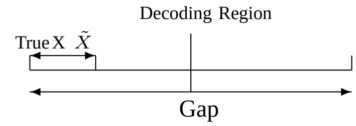


Fig. 11. A decoding error occurs when the noise brings $\tilde{X}$ beyond the decoding region.

Proof: By Proposition 5.6, for any $t > T_{max}$, the $\lfloor \frac{t-T_{max}}{R} \rfloor - th$ bit should leave the queue at time t. So by looking at the level of this bit at time t, we should have

$$\gamma(\prod_{i=0}^{t-1} \lambda_i)(2+\epsilon_1)^{-(\lfloor \frac{t-T_{max}}{R} \rfloor + c_{\lfloor \frac{t-T_{max}}{R} \rfloor})} > T_L \tag{11}$$

which implies that $\prod_{i=0}^{t-1} \lambda_i > k.(\lambda_{n_0})^{t-T_{max}}$ for some constants k, $\lambda_{n_0} > 1$. This completes the proof. ■

The dynamic of the queue with turbo mode are:

$$Q'_{t+1} = Q'_t + R + \text{inc}'_t - \log \lambda'_t \tag{12}$$

From the definition of the turbo mode it is obvious that $\lambda'_t \geq \lambda_t, \forall t$, thus by induction it follows that $Q'_t \leq Q_t, \forall t$ and also inc'_t is defined the same as inc_t but depends on the corresponding λ's. As Q_t is stable, we can conclude that Q'_t is also stable. Thus, $P\{Q'_t > k\}$ is a decreasing function of k. Hence, for every $\epsilon > 0$ there exists a threshold in the queue length T_U such that $P\{Q'_t > T_U\} < \epsilon$. Figure 10 shows the effect of the turbo mode and of inc in the queue dynamic.

So, by picking T_U large enough, we can make turbo mode as rare as would like. Thus, the extra power used by turbo mode will not impact the average power by that much. Meanwhile, proposition 5.7 shows that by using turbo mode we are sure that the error gaps are increasing exponentially with time regardless of the fading realizations. Thus, in order to have an error at delay d, the Gaussian noise has to be able push the closed-loop state away by at least an exponentially growing function of d. The closed-loop Gaussian noise can be made to have bounded variance regardless of the fading realizations and hence the probability of error goes to zero doubly exponentially in delay.

References

[1] J. P. M. Schalkwijk, T. Kailath, "A coding scheme for additive noise channels with feedback–I: No bandwidth constraint," *IEEE Trans. Info. Theory*, vol. IT-12, pp. 172 - 182, April 1966.

[2] A. Sahai, "Anytime Information Theory," Ph.D. dissertation, MIT 2001.

[3] A. Sahai, "The necessity and sufficiency of anytime capacity for scalar control over a noisy communication link," *IEEE Conference on Decision and Control*, December 2004.

[4] A. Sahai and Q. Xu, "The anytime reliability of constrained packet erasure channels with feedback," *Allerton Conference on Communication, Control, and Computing* October 2004.

[5] J. Liu, N. Elia, and S. Tatikonda, "Capacity-achieving Feedback Scheme for Markov Channels with Channel State Information", *2004 IEEE International Symposium on Information Theory* , July 2004.

[6] A. Sahai, Q. Xu, S. Avestimehr and P. Minero, "Anytime communication over fading channels with perfect CSI and noiseless feedback," *In preparation*.

[7] H. Viswanathan, "Capacity of Markov Channels with Receiver CSI and Delayed Feedback," *IEEE Trans. Info. Theory,* vol. IT-45, pp. 761 - 771, March 1999.