

Timely Estimation Using Coded Quantized Samples

Ahmed Arafa¹, Karim Banawan², Karim G. Seddik³, and H. Vincent Poor⁴

¹Department of Electrical and Computer Engineering, University of North Carolina at Charlotte, USA

²Department of Electrical Engineering, Alexandria University, Egypt

³Electronics and Communications Engineering Department, American University in Cairo, Egypt

⁴Electrical Engineering Department, Princeton University, USA

Abstract—The effects of *quantization and coding* on the estimation quality of a Gauss-Markov, namely Ornstein-Uhlenbeck, process are considered. Samples are acquired from the process, quantized, and then encoded for transmission using either *infinite incremental redundancy* or *fixed redundancy coding* schemes. A fixed processing time is consumed at the receiver for decoding and sending feedback to the transmitter. Decoded messages are used to construct a minimum mean square error (MMSE) estimate of the process as a function of time. This is shown to be an increasing functional of the *age-of-information*, defined as the time elapsed since the sampling time pertaining to the latest successfully decoded message. Such (age-penalty) functional depends on the quantization bits, codeword lengths and receiver processing time. The goal, for each coding scheme, is to optimize sampling times such that the long term average MMSE is minimized. This is then characterized in the setting of *general increasing age-penalty functionals*, not necessarily corresponding to MMSE, which may be of independent interest in other contexts.

I. INTRODUCTION

Recent works have drawn connections between remote estimation of a time-varying process and the *age-of-information* (AoI) metric, which assesses the timeliness and freshness of the estimated data. While most works focus on transmitting *analog* samples for the purpose of estimation, this work focuses on using *quantized* and *coded* samples in that regard. We present optimal sampling methods that minimize the long term average minimum mean square error (MMSE) of a Gauss-Markov, namely Ornstein-Uhlenbeck (OU), process under specific coding schemes, taking into consideration receiver *processing* times consumed in decoding and sending feedback.

AoI, or merely age, is defined as the time elapsed since the latest useful data has reached its destination. An increasing number of works have used AoI as a latency performance metric in various contexts, including queuing analysis, scheduling in networks and optimization under different constraints, see, e.g., the general body of works in [1]–[26].

Of particular relevance to this paper are the works related to coding for AoI improvement, e.g., [27]–[40], and those related to the (inherent) role of AoI in remote estimation, e.g., [41]–[53]. Two main takeaway points from these works are: (1) optimal codes should strike a balance between using long codewords to combat channel errors and using short ones to minimize age; and (2) optimal sampling strategies should balance the need for frequent transmissions to minimize age as it relates to the correlation of the updates being received.

This work was supported in part by the U.S. National Science Foundation under Grants CCF-0939370 and CCF-1513915.

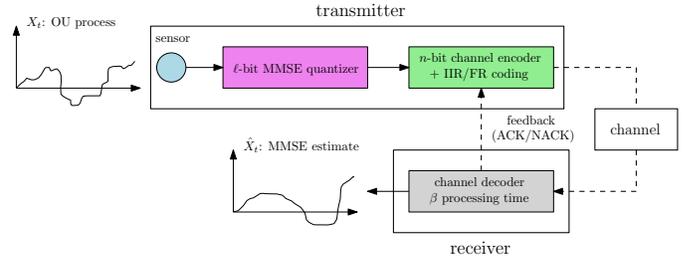


Fig. 1. System model considered for sampling, quantizing and encoding an OU process at the transmitter, and reconstructing it at the receiver.

The most closely-related works to ours are [47], [48], which derive optimal sampling methods to minimize the long term average MMSE for Wiener [47] and OU [48] processes. In both works, the communication channel introduces random delays, before perfect (distortion-free) samples are received. It is shown that if sampling times are independent of the instantaneous values of the process (signal-independent sampling) the MMSE reduces to AoI in case of Wiener [47], and to an increasing functional of AoI (age-penalty) in case of OU [48]. It is then shown that the optimal sampling policy has a threshold structure, in which a new sample is acquired only if the expected AoI in case of Wiener (or age-penalty in case of OU) surpasses a certain value. In addition, signal-dependent optimal sampling policies are also derived [47], [48].

In this work, we consider the transmission of quantized and coded samples of an OU process through a noisy channel.¹ Different from [48], not every sample has guaranteed reception, and received samples suffer from quantization errors. The receiver uses the received samples to construct an MMSE estimate for the OU process. Quantization and coding introduce a tradeoff: *few quantization levels and codeword bits would transmit samples faster, yet with high distortion and probability of error*. An optimal choice, therefore, needs to be made, which depends mainly on how fast the OU process varies as well as the channel errors. Different from related works, effects of having (fixed) *non-zero receiver processing times*, mainly due to decoding and sending feedback, are also considered in this work.

¹We note that we consider an OU process in our study since, unlike the conventional Wiener process, it has a bounded variance, leading to bounded quantization error as well. The OU process, in addition, is used to model various physical phenomena, and has relevant applications in control and finance (see, e.g., the discussion in [48]).

We focus on signal-independent sampling, together with an MMSE quantizer, combined with either *infinite incremental redundancy* (IIR) or *fixed redundancy* (FR) coding schemes; see Fig. 1. The MMSE of the OU process is first shown to be an increasing functional of AoI, as in [48], parameterized directly by the number of quantization bits ℓ , and indirectly by the number of codeword bits n and the receiver processing time β . We formulate two problems, one for IIR and another for FR, to choose sampling times so that the long term average MMSE is minimized. Focusing on stationary deterministic policies, we present optimal solutions for both problems in the case of *general increasing age-penalties*, not necessarily corresponding to MMSE, which may be useful in other contexts in which IIR and FR coding schemes are employed. The solution for IIR has a *threshold* structure, as in [16], [48], while that for FR is a *just-in-time* sampling policy that does not require receiver feedback. We finally discuss how to select ℓ and n , and show that the relatively simpler FR scheme can outperform IIR for relatively large values of β .

II. SYSTEM MODEL AND MMSE ESTIMATION

We consider a sensor that acquires time-stamped samples from an OU process. Given a value of X_s at time s , the OU process evolves as follows [54], [55]:

$$X_t = X_s e^{-\theta(t-s)} + \frac{\sigma}{\sqrt{2\theta}} e^{-\theta(t-s)} W_{e^{2\theta(t-s)} - 1}, \quad t \geq s, \quad (1)$$

where W_t denotes a Weiner process, while $\theta > 0$ and $\sigma > 0$ are fixed parameters. The sensor acquires the i th sample at time S_i and feeds it to an MMSE quantizer that produces an ℓ -bit message ready for encoding. We will use the term *message* to refer to a quantized sample of the OU process. Let \tilde{X}_{S_i} represent the quantized version of the sample X_{S_i} , and let Q_{S_i} denote the corresponding quantization error. Thus,

$$X_{S_i} = \tilde{X}_{S_i} + Q_{S_i}. \quad (2)$$

Each message is encoded and sent over a noisy channel to the receiver. The receiver updates an MMSE estimate of the OU process if decoding is successful. ACKs and NACKs are fed back following each decoding attempt, which is assumed to consume fixed $\beta \geq 0$ time units. Channel errors are independent and identically distributed (i.i.d.) across time/messages.

Two channel coding schemes are investigated. The first is IIR, in which a message transmission starts with an n -bit codeword, $n \geq \ell$, and then incremental redundancy (IR) bits are added one-by-one if a NACK is received until the message is eventually decoded and an ACK is fed back. The second scheme is FR, in which a message is encoded into fixed n -bit codewords, yet following a NACK the message in transmission is discarded and a *new* sample is acquired and used instead. Following ACKs, the transmitter may idly wait before acquiring a new sample and sending a new message.

Let D_i denote the reception time of the i th *successfully decoded* message. For the IIR scheme, each message is eventually decoded, and therefore

$$D_i = S_i + Y_i \quad (3)$$

for some random variable Y_i that represents the channel delay incurred due to the IR bits added. Let T_b denote the time units consumed per bit transmission. Hence,

$$Y_i = nT_b + \beta + r_i(T_b + \beta), \quad (4)$$

where $r_i \in \{0, 1, 2, \dots\}$ denotes the number of IR bits used until the i th message is decoded. Note that in the IIR scheme β is added for the original n -bit codeword transmission, and then for each IR transmission until successful decoding. Let

$$\bar{n} \triangleq nT_b + \beta \quad (5)$$

for conciseness. Channel delays Y_i 's are i.i.d. $\sim Y$, where

$$\mathbb{P}(Y = \bar{n}) = p_0, \quad (6)$$

$$\mathbb{P}(Y = \bar{n} + k(T_b + \beta)) = \prod_{j=0}^{k-1} (1 - p_j) p_k, \quad k \geq 1, \quad (7)$$

with p_j denoting the probability that an ACK is received when $r_i = j$. This depends on the channel code being used, and the model of the channel errors, yet it holds that $p_j \leq p_{j+1}$.

For the FR scheme, there can possibly be a number of transmission *attempts* before a message is eventually decoded. Let M_i denote the number of these attempts in between the $(i-1)$ th and i th successfully decoded messages, and let $S_{i,j}$ denote the sampling time pertaining to the j th attempt of which, $1 \leq j \leq M_i$. Therefore, only the M_i th message is successfully decoded, and the rest are all discarded. Since each message is encoded using fixed n -bit codewords, we have

$$D_i = S_{i,M_i} + \bar{n}, \quad \forall i. \quad (8)$$

Observe that in the FR scheme each successfully-decoded message incurs only *one* β , since each decoding attempt occurs on a message pertaining to a *fresh* sample. According to the notation developed for the IIR channel delays above, M_i 's are i.i.d. geometric random variables with parameter p_0 .

Upon successfully decoding a message at time D_i , the receiver constructs an MMSE estimate for the OU process. We restrict our attention to MMSE estimators that only use the latest-received information. For the IIR scheme this is

$$\hat{X}_t = \mathbb{E} \left[X_t \middle| S_i, \tilde{X}_{S_i} \right], \quad D_i \leq t < D_{i+1}. \quad (9)$$

Using (1) and (2), we have

$$\hat{X}_t = \mathbb{E} \left[\tilde{X}_{S_i} e^{-\theta(t-S_i)} + Q_{S_i} e^{-\theta(t-S_i)} + \frac{\sigma}{\sqrt{2\theta}} e^{-\theta(t-S_i)} W_{e^{2\theta(t-S_i)} - 1} \middle| S_i, \tilde{X}_{S_i} \right] \quad (10)$$

$$= \tilde{X}_{S_i} e^{-\theta(t-S_i)}, \quad D_i \leq t < D_{i+1}, \quad (11)$$

where the last equality follows by independence of the Weiner noise in $[D_i, t]$ from (S_i, \tilde{X}_{S_i}) , and that for the MMSE quantizer, the quantization error is zero-mean [56]. The MMSE is now given as follows for $D_i \leq t < D_{i+1}$: (see also [48])

$$\text{mse}(t, S_i) = \mathbb{E} \left[\left(X_t - \hat{X}_t \right)^2 \right] \quad (12)$$

$$= \mathbb{E} [Q_{S_i}^2] e^{-2\theta(t-S_i)} + \frac{\sigma^2}{2\theta} \left(1 - e^{-2\theta(t-S_i)}\right). \quad (13)$$

We see from the above that even if $D_i - S_i = 0$, i.e., if the i th sample is transmitted and received instantaneously, the MMSE estimate at $t = D_i$ would still suffer from quantization errors.

In the sequel, we consider $X_0 = 0$ without loss of generality, and hence, using (1), the variance of X_t is given by $\mathbb{E} [X_t^2] = \frac{\sigma^2}{2\theta} (1 - e^{-2\theta t})$, $t > 0$. Following a rate-distortion approach (note that X_t is Gaussian), the following relates the number of bits ℓ and the mean square quantization error [56]:

$$\mathbb{E} [Q_t^2] = \frac{\sigma^2}{2\theta} (1 - e^{-2\theta t}) 2^{-2\ell}, \quad t > 0. \quad (14)$$

Using the above in (13) and rearranging, we get that

$$\text{mse}(t, S_i) = \frac{\sigma^2}{2\theta} \left(1 - (1 - 2^{-2\ell} (1 - e^{-2\theta S_i})) e^{-2\theta(t-S_i)}\right), \quad (15)$$

We note that as $\ell \rightarrow \infty$, the above expression becomes the same as that derived for the signal-independent sampling scheme analyzed in [48]. However, since we consider practical coding aspects in this work, as $\ell \rightarrow \infty$, it holds that $n \rightarrow \infty$ as well and no sample will be received.

We focus on dealing with the system in *steady state*, in which both t and S_i are relatively large. In this case, the mean square quantization error in (14) becomes independent of time, and only dependent upon the steady state variance of the OU process $\sigma^2/2\theta$. Hence, in steady state, the MMSE becomes

$$\text{mse}(t, S_i) = \frac{\sigma^2}{2\theta} \left(1 - (1 - 2^{-2\ell}) e^{-2\theta(t-S_i)}\right) \quad (16)$$

$$\triangleq h_\ell(t - S_i), \quad D_i \leq t < D_{i+1}, \quad (17)$$

which is an increasing functional of the AoI $t - S_i$.

For the FR scheme, the analysis follows similarly, after adding one more random variable denoting the number of transmissions, M_i . Specifically, it holds that

$$\hat{X}_t = \tilde{X}_{S_i, M_i} e^{-\theta(t-S_i, M_i)}, \quad (18)$$

$$\text{mse}(t, S_i, M_i) = h_\ell(t - S_i, M_i), \quad D_i \leq t < D_{i+1}. \quad (19)$$

III. PROBLEM FORMULATION: GENERAL AGE-PENALTY

The main goal is to choose the sampling times, for given ℓ , n and β , such that the long term average MMSE is minimized. In this section, we formulate two problems to achieve such goal: one for IIR and another for FR, and present their solutions in the two upcoming sections. Then, in Section VI, we discuss how to choose the best ℓ and n , as well as compare the performances of IIR and FR in general. For both schemes, let us denote by an *epoch* the time elapsed in between two successfully received messages. Thus, the i th epoch starts at D_{i-1} and ends at D_i , with $D_0 \equiv 0$.

Remark 1 *Our analysis does not depend on the specific structure of the MMSE functional $h_\ell(\cdot)$; it extends to any differentiable increasing age-penalty functional $g(\cdot)$. Therefore, in what follows, we formulate our problems and present their solutions for the case of minimizing a long term average age-penalty, making the results applicable in other contexts.*

For the IIR scheme, the problem is formulated as

$$\min_{\{S_i\}} \limsup_{l \rightarrow \infty} \frac{\sum_{i=0}^l \mathbb{E} \left[\int_{D_i}^{D_{i+1}} g(t - S_i) dt \right]}{\sum_{i=0}^l \mathbb{E} [D_{i+1} - D_i]}, \quad (20)$$

where the numerator represents the total age-penalty (the MMSE in case of the OU process estimation) across all epochs, and the denominator represents the total time.

Let us define W_i as the waiting time at the beginning of the i th epoch before acquiring the i th sample. That is, $S_i = D_{i-1} + W_i$. Therefore, one can equivalently solve for the waiting times W_i 's instead of sampling times S_i 's. We focus on a class of *stationary deterministic* policies in which

$$W_i = f(g(D_{i-1} - S_{i-1})), \quad \forall i. \quad (21)$$

That is, *the waiting time in the i th epoch is a deterministic function of its starting age-penalty value*. Such focus is motivated by the fact that channel errors are i.i.d. and by its optimality in similar frameworks, e.g., [7], [8], [22]. Defining $w \triangleq f \circ g$ and noting that $D_{i-1} - S_{i-1} = Y_{i-1}$ we have

$$W_i = w(Y_{i-1}), \quad (22)$$

which induces a stationary distribution of D_i 's and the age-penalty across all epochs. Let us now (re)define notations used in a typical epoch. It starts at time \bar{D} with AoI \bar{Y} , and with the latest sample acquired at time \bar{S} , such that $\bar{D} = \bar{S} + \bar{Y}$. Then, a waiting time of $w(\bar{Y})$ follows, after which a new sample is acquired, quantized, and transmitted, taking Y time units to reach the receiver at time $D = \bar{D} + w(\bar{Y}) + Y$, which is the epoch's end time. Therefore, problem (20) now reduces to a minimization over a single epoch as follows:

$$\min_{w(\cdot) \geq 0} \frac{\mathbb{E} \left[\int_{\bar{D}}^{\bar{D} + w(\bar{Y}) + Y} g(t - \bar{S}) dt \right]}{\mathbb{E} [w(\bar{Y}) + Y]}. \quad (23)$$

Given the realization of \bar{Y} at time \bar{D} , the transmitter decides on the waiting time $w(\bar{Y})$ that minimizes the long term average age-penalty demonstrated in the fractional program above. We solve problem (23) in Section IV.

For the FR scheme, the formulated problem can be derived similarly, *with the addition of possible waiting times in between retransmissions*.² Specifically, let $W_{i,j}$ represent the waiting time before the j th transmission attempt in the i th epoch. A stationary deterministic policy here is such that

$$W_{i,1} = f(g(D_{i-1} - S_{i-1, M_{i-1}})) = w(\bar{n}) \equiv w_1, \quad (24)$$

$$W_{i,2} = w(w_1 + \bar{n}) \equiv w_2, \quad (25)$$

⋮

$$W_{i,j} = w(w_1 + \dots + w_{j-1} + \bar{n}) \equiv w_j, \quad (26)$$

and so on. Therefore, under the FR scheme, a stationary deterministic policy reduces to a countable sequence $\{w_j\}$.

²This is only amenable for FR since waiting leads to acquiring a fresher sample, and possibly reduced age-penalties. For IIR, waiting after a NACK is clearly suboptimal since it elongates the channel delay for the *same* sample.

Proceeding with the same notations for a given epoch as in the IIR scheme, let us define M as the number of transmission attempts in the epoch, \bar{M} as those in the previous epoch, and $\bar{S}_{\bar{M}}$ as the sampling time of the successful (last) transmission attempt in the previous epoch. The problem now becomes

$$\min_{\{w_j \geq 0\}} \frac{\mathbb{E} \left[\int_{\bar{D}}^{\bar{D} + \sum_{j=1}^M w_j + M\bar{n}} g(t - \bar{S}_{\bar{M}}) dt \right]}{\mathbb{E} \left[\sum_{j=1}^M w_j + M\bar{n} \right]}. \quad (27)$$

We solve problem (27) in Section V.

IV. THE IIR SCHEME: SOLUTION OF PROBLEM (23)

We follow Dinkelbach's approach to transform (23) into the following auxiliary problem for fixed $\lambda \geq 0$ [57]:

$$p^{IIR}(\lambda) \triangleq \min_{w(\cdot) \geq 0} \mathbb{E} \left[\int_{\bar{D}}^{\bar{D} + w(\bar{Y}) + Y} g(t - \bar{S}) dt \right] - \lambda \mathbb{E} [w(\bar{Y}) + Y]. \quad (28)$$

The optimal solution of (23) is then given by λ_{IIR}^* that solves $p^{IIR}(\lambda_{IIR}^*) = 0$, which can be found via bisection, since $p^{IIR}(\lambda)$ is decreasing [57]. We now have the following result:

Theorem 1 *The optimal solution of problem (28) is given by*

$$w^*(\bar{y}) = [G_{\bar{y}}^{-1}(\lambda)]^+, \quad (29)$$

where $[\cdot]^+ \triangleq \max(\cdot, 0)$, \bar{y} is the realization of the starting AoI \bar{Y} , and $G_{\bar{y}}(x) \triangleq \mathbb{E}[g(\bar{y} + x + Y)]$.

Theorem 1 can be proved using [16, Theorem 1]. We note, however, that its proof can be approached differently from that of [16, Theorem 1], in a way similar to the proof of Theorem 2 below. We omit such details due to space limits.

The optimal waiting policy for IIR has a *threshold* structure: a new sample is acquired only when the expected age-penalty by the end of the epoch is at least λ . Note that the optimal λ_{IIR}^* corresponds to the optimal long term average age-penalty.

V. THE FR SCHEME: SOLUTION OF PROBLEM (27)

We follow a similar approach here as in the IIR scheme and consider the following auxiliary problem:

$$p^{FR}(\lambda) \triangleq \min_{\{w_j \geq 0\}} \mathbb{E} \left[\int_{\bar{D}}^{\bar{D} + \sum_{j=1}^M w_j + M\bar{n}} g(t - \bar{S}_{\bar{M}}) dt \right] - \lambda \mathbb{E} \left[\sum_{j=1}^M w_j + M\bar{n} \right]. \quad (30)$$

The optimal solution of problem (27) is now given by λ_{FR}^* that solves $p^{FR}(\lambda_{FR}^*) = 0$, which we will provide in *closed-form*. The optimal waiting policy structure is provided next.

Theorem 2 *The optimal solution of problem (30) is given by*

$$w_1^* = [G^{-1}(\lambda)]^+, \quad (31)$$

$$w_j^* = 0, \quad j \geq 2, \quad (32)$$

where $G(x) \triangleq \mathbb{E}[g(\bar{n} + x + M\bar{n})]$. In addition, the optimal solution of problem (27), λ_{FR}^* , is such that $w_1^* = 0$.

Proof: We first simplify the terms of the objective function of (30). Using iterated expectations, it can be shown that

$$\mathbb{E} \left[\sum_{j=1}^M w_j + M\bar{n} \right] = \sum_{j=1}^{\infty} w_j (1 - p_0)^{j-1} + \frac{\bar{n}}{p_0}. \quad (33)$$

Now let us define

$$\zeta_m(\mathbf{w}_1^m) \triangleq \int_{\bar{D}}^{\bar{D} + \sum_{j=1}^m w_j + m\bar{n}} g(t - \bar{S}_{\bar{M}}) dt \quad (34)$$

and, leveraging iterated expectations on the first term of (30), introduce the following Lagrangian:³

$$\begin{aligned} \mathcal{L} = & \sum_{m=1}^{\infty} \zeta_m(\mathbf{w}_1^m) (1 - p_0)^{m-1} p_0 - \lambda \sum_{j=1}^{\infty} w_j (1 - p_0)^{j-1} \\ & - \lambda \frac{\bar{n}}{p_0} - \sum_{j=1}^{\infty} w_j \eta_j, \end{aligned} \quad (35)$$

where η_j 's are Lagrange multipliers. Now observe that, using Leibniz rule, it holds for $j \leq m$ that

$$\frac{\partial \zeta_m(\mathbf{w}_1^m)}{\partial w_j} = g \left(\bar{n} + \sum_{j=1}^m w_j + m\bar{n} \right). \quad (36)$$

Taking derivative of the Lagrangian with respect to w_j and equating to 0, we use the above to get

$$\sum_{m=j}^{\infty} g \left(\bar{n} + \sum_{j=1}^m w_j + m\bar{n} \right) (1 - p_0)^{m-j} p_0 = \lambda + \frac{\eta_j}{(1 - p_0)^{j-1}}. \quad (37)$$

Next, let us substitute $j = k$ and $j = k + 1$ above, $k \geq 1$, subtract them from each other, and rearrange to get

$$g \left(\bar{n} + \sum_{j=1}^k w_j + k\bar{n} \right) = \lambda + \frac{\eta_k - \eta_{k+1}}{(1 - p_0)^{k-1} p_0}. \quad (38)$$

Since $g(\cdot)$ is increasing, and λ is fixed, $\left\{ \frac{\eta_k - \eta_{k+1}}{(1 - p_0)^{k-1} p_0} \right\}$ is increasing. From there, one can conclude that $\eta_j > 0$, $j \geq 2$ must hold. Hence, by complementary slackness, $w_j^* = 0$, $j \geq 2$ [58]. Using (37) for $j = 1$, the optimal w_1^* now solves

$$G(w_1^*) = \lambda + \eta_1, \quad (39)$$

where $G(\cdot)$ is as defined in the theorem. Observe that $G(\cdot)$ is increasing and therefore the above has a unique solution. Now, if $\lambda \leq G(0)$, then we must have $\eta_1 > 0$, and hence $w_1^* = 0$ by complementary slackness; conversely, if $\lambda > G(0)$, then we must have $w_1^* > 0$, and hence $\eta_1 = 0$ by complementary slackness as well [58]. In the latter case, $w_1^* = G^{-1}(\lambda)$. Finally, observe that $\lambda \leq G(0) \iff G^{-1}(\lambda) \leq 0$. This concludes the proof of the first part of the theorem.

³Using monotonicity of $g(\cdot)$, it can be shown that problem (30) is convex.

To show the second part, all we need to prove now is that $G^{-1}(\lambda_{FR}^*) \leq 0$, or equivalently that $\lambda_{FR}^* \leq G(0)$. Toward that end, observe that $p_{FR}(\lambda)$ is decreasing, and therefore if $p_{FR}(G(0)) \leq 0$ then the premise follows. Now for $\lambda = G(0)$ we know from the first part of the proof that $w_1^* = 0$. Thus,

$$p_{FR}(G(0)) = \sum_{m=1}^{\infty} \zeta_m(0) (1-p_0)^{m-1} p_0 - G(0) \frac{\bar{n}}{p_0} \quad (40)$$

$$= \mathbb{E} \left[\int_{\bar{D}}^{\bar{D}+M\bar{n}} g(t - \bar{S}_M) dt \right] - G(0) \mathbb{E}[M] \bar{n} \quad (41)$$

$$= \mathbb{E} \left[\int_0^{M\bar{n}} g(\bar{n} + t) dt \right] - \mathbb{E} \left[\int_0^{M\bar{n}} G(0) dt \right] \quad (42)$$

$$= \mathbb{E} \left[\int_0^{M\bar{n}} \mathbb{E}[g(\bar{n} + t) - g(\bar{n} + M\bar{n})] dt \right], \quad (43)$$

where (42) follows by change of variables and (43) follows by definition of $G(\cdot)$. Finally, observe that by monotonicity of $g(\cdot)$, (43) is non-positive. This concludes the proof. \blacksquare

A closed-form expression for λ_{FR}^* can now be found via substituting $w_j = 0, \forall j$ in (27).

Theorem 2 shows that *zero-wait* policies are optimal for FR, which is quite intuitive. First, waiting is not optimal in between retransmissions, even though it would lead to acquiring fresher samples, since the AoI is already relatively high following failures. Second, since the epoch always starts with the same AoI, \bar{n} , one can optimize the long term average age-penalty to make waiting not optimal at the beginning of the epoch as well. We note, however, that such results do *not* follow from [7, Theorem 5], since there can be multiple transmissions in the same epoch. We also note that while zero-wait policies have been invoked in other works involving FR coding schemes, e.g., [29], [39], Theorem 2 provides a proof of their *optimality* for general increasing age-penalties.

Now that zero-waiting is optimal, we investigate whether it is necessary to wait for the receiver to decode and send feedback before sending the next message. It could be better, age-wise, to send a new message right away after the previous one is *delivered*, i.e., after nT_b time units instead of \bar{n} . However, this may not be optimal if β is relatively large, since it would lead to accumulating *stale* messages at the receiver's end as they wait for decoding to finish. The next lemma revolves this. The proof is similar, in essence, to that of Theorem 2, and is omitted due to space limits.

Lemma 1 *In the FR scheme, a new message is sent after the previous one's delivery by $[\beta - nT_b]^+$ time units.*

Lemma 1 shows that *just-in-time* updating is optimal. For $\beta \leq nT_b$, a new sample is acquired and transmitted just-in-time as the previous message is delivered. While for $\beta > nT_b$, a new sample is acquired and transmitted such that it is delivered just-in-time as the receiver finishes decoding the previous message. This way, delivered samples are always fresh, the receiver is never idle, and feedback is unnecessary.

Setting	$\epsilon = 0.1$:	IIR	FR	$\epsilon = 0.4$:	IIR	FR
$\theta = 0.01$		(5, 7)	(5, 7)		(4, 10)	(4, 6)
$\theta = 0.5$		(2, 4)	(2, 4)		(1, 3)	(2, 4)

TABLE I

OPTIMAL (ℓ^*, n^*) FOR IIR AND FR UNDER DIFFERENT SETTINGS.

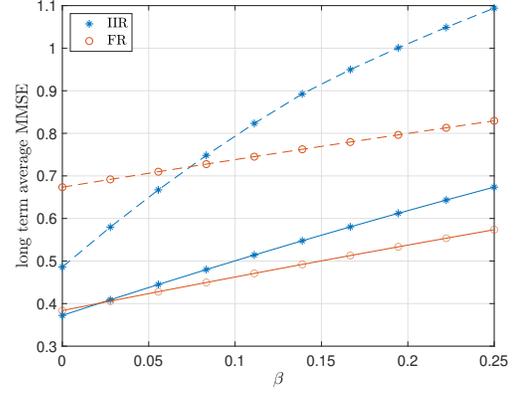


Fig. 2. $\ell = 3$; $\theta = 0.25$; dashed lines: $\epsilon = 0.4$, and solid lines: $\epsilon = 0.1$.

VI. DISCUSSION: WHICH SCHEME PERFORMS BETTER?

We now discuss the IIR and FR performances in the original context of OU estimation, i.e., when $g(\cdot) \equiv h_\ell(\cdot)$. Applying Theorem 1's result, the optimal waiting policy for IIR is

$$w^*(\bar{y}) = \left[\frac{1}{2\theta} \log \left(\frac{\frac{\sigma^2}{2\theta} (1 - 2^{-2\ell}) \mathbb{E}[e^{-2\theta Y}]}{\frac{\sigma^2}{2\theta} - \lambda_{IIR}^*} \right) - \bar{y} \right]^+ \quad (44)$$

One can also show that $\lambda_{IIR}^* \in \left[2^{-2\ell} \frac{\sigma^2}{2\theta}, \frac{\sigma^2}{2\theta} \right]$, facilitating the bisection search. Applying Theorem 2 and Lemma 1's results, the optimal long term average MMSE for FR is given by

$$\frac{\sigma^2}{2\theta} \left(1 - \frac{(1 - 2^{-2\ell}) e^{-2\theta \bar{n}} p_0}{2\theta K_{n,\beta}} \frac{1 - e^{-2\theta K_{n,\beta}}}{1 - (1 - p_0) e^{-2\theta K_{n,\beta}}} \right), \quad (45)$$

where $K_{n,\beta} \triangleq \max\{\beta, nT_b\}$. Derivation details are omitted.

We consider a binary symmetric channel with crossover probability $\epsilon \in (0, \frac{1}{2})$, and use MDS codes for transmission. This allows us to write $p_j = \sum_{l=0}^{\lfloor \frac{n+j-\ell}{2} \rfloor} \binom{n+j}{l} \epsilon^l (1-\epsilon)^{n+j-l}$, where $\lfloor x \rfloor$ denotes the highest integer not larger than x . We set $\sigma^2 = 1$, and $T_b = 0.05$ time units. For fixed $\beta = 0.15$, we vary ℓ and choose the best n for IIR and FR via grid search, and then choose the best all over ℓ . We repeat for $\theta = 0.01$ (slow process variation) and $\theta = 0.5$ (fast process variation). The results are shown in Table I. We see that the optimal ℓ^* that minimizes the MMSE is inversely proportional with θ . This is intuitive, since for slowly varying processes, one can tolerate larger waiting times to get high quality estimates, and vice versa. Similar behavior is noticed for relatively good ($\epsilon = 0.1$) and bad ($\epsilon = 0.4$) channel conditions.

In Fig. 2, we fix $\theta = 0.25$, $\ell = 3$ bits and vary β . We see that FR outperforms IIR for relatively large β . This is clearer in worse channel conditions ($\epsilon = 0.4$), which is expected since the β processing penalty is incurred, following NACKs, at the bit level for IIR and at the message level for FR.

REFERENCES

- [1] S. K. Kaul, R. D. Yates, and M. Gruteser. Real-time status: How often should one update? In *Proc. IEEE Infocom*, March 2012.
- [2] C. Kam, S. Kompella, and A. Ephremides. Age of information under random updates. In *Proc. IEEE ISIT*, July 2013.
- [3] R. D. Yates. Lazy is timely: Status updates by an energy harvesting source. In *Proc. IEEE ISIT*, June 2015.
- [4] M. Costa, M. Codreanu, and A. Ephremides. On the age of information in status update systems with packet management. *IEEE Trans. Inf. Theory*, 62(4):1897–1910, April 2016.
- [5] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis. Age and value of information: Non-linear age case. In *Proc. IEEE ISIT*, June 2017.
- [6] Y. Hsu, E. Modiano, and L. Duan. Age of information: Design and analysis of optimal scheduling algorithms. In *Proc. IEEE ISIT*, June 2017.
- [7] Y. Sun, E. Uysal-Biyikoglu, R. D. Yates, C. E. Koksal, and N. B. Shroff. Update or wait: How to keep your data fresh. *IEEE Trans. Inf. Theory*, 63(11):7492–7508, November 2017.
- [8] X. Wu, J. Yang, and J. Wu. Optimal status update for age of information minimization with an energy harvesting source. *IEEE Trans. Green Commun. Netw.*, 2(1):193–204, March 2018.
- [9] P. Mayekar, P. Parag, and H. Tyagi. Optimal lossless source codes for timely updates. In *Proc. IEEE ISIT*, June 2018.
- [10] A. Baknina, O. Ozel, J. Yang, S. Ulukus, and A. Yener. Sending information through status updates. In *Proc. IEEE ISIT*, June 2018.
- [11] B. Zhou and W. Saad. Optimal sampling and updating for minimizing age of information in the internet of things. In *Proc. IEEE Globecom*, December 2018.
- [12] R. D. Yates and S. K. Kaul. The age of information: Real-time status updating by multiple sources. *IEEE Trans. Inf. Theory*, 65(3):1807–1827, March 2019.
- [13] M. Zhang, A. Arafa, J. Huang, and H. V. Poor. How to price fresh data. In *Proc. WiOpt*, June 2019.
- [14] B. Buyukates, A. Soysal, and S. Ulukus. Age of information in multihop multicast networks. *J. Commun. Netw.*, 21(3):256–267, June 2019.
- [15] B. T. Bacinoglu, Y. Sun, E. Uysal-Biyikoglu, and V. Mutlu. Optimal status updating with a finite-battery energy harvesting source. *J. Commun. Netw.*, 21(3):280–294, June 2019.
- [16] Y. Sun and B. Cyr. Sampling for data freshness optimization: Non-linear age functions. *J. Commun. Netw.*, 21(3):204–219, June 2019.
- [17] S. Leng and A. Yener. Age of information minimization for an energy harvesting cognitive radio. *IEEE Trans. Cogn. Commun. Netw.*, 5(2):427–439, June 2019.
- [18] A. M. Bedewy, Y. Sun, and N. B. Shroff. The age of information in multihop networks. *IEEE/ACM Trans. Netw.*, 27(3):1248–1257, June 2019.
- [19] R. Talak and E. Modiano. Age-delay tradeoffs in single server systems. In *Proc. IEEE ISIT*, July 2019.
- [20] A. Arafa, R. D. Yates, and H. V. Poor. Timely cloud computing: Preemption and waiting. In *Proc. Allerton*, October 2019.
- [21] Y. Inoue, H. Masuyama, T. Takine, and T. Tanaka. A general formula for the stationary distribution of the age of information and its application to single-server queues. *IEEE Trans. Inf. Theory*, 65(12):8305–8324, December 2019.
- [22] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor. Age-minimal transmission for energy harvesting sensors with finite batteries: Online policies. *IEEE Trans. Inf. Theory*, 66(1):534–556, January 2020.
- [23] H. H. Yang, A. Arafa, T. Q. S. Quek, and H. V. Poor. Age-based scheduling policy for federated learning in mobile edge networks. In *Proc. IEEE ICASSP*, May 2020.
- [24] P. Zou, O. Ozel, and S. Subramaniam. Waiting before serving: A companion to packet management in status update systems. *IEEE Trans. Inf. Theory*. To appear.
- [25] A. Soysal and S. Ulukus. Age of information in G/G/1/1 systems: Age expressions, bounds, special cases, and optimization. Available Online: arXiv:1905.13743.
- [26] H. Tang, J. Wang, L. Song, and J. Song. Minimizing age of information with power constraints: Opportunistic scheduling in multi-state time-varying networks. Available Online: arXiv: 1912.05947.
- [27] P. Parag, A. Taghavi, and J.-F. Chamberland. On real-time status updates over symbol erasure channels. In *Proc. IEEE WCNC*, March 2017.
- [28] E. Najm, R. Yates, and E. Soljanin. Status updates through M/G/1/1 queues with HARQ. In *Proc. IEEE ISIT*, June 2017.
- [29] R. D. Yates, E. Najm, E. Soljanin, and J. Zhong. Timely updates over an erasure channel. In *Proc. IEEE ISIT*, June 2017.
- [30] A. Baknina and S. Ulukus. Coded status updates in an energy harvesting erasure channel. In *Proc. CISS*, March 2018.
- [31] E. T. Ceran, D. Gunduz, and A. Gyorgy. Average age of information with hybrid ARQ under a resource constraint. In *Proc. IEEE WCNC*, April 2018.
- [32] H. Sac, B. T. Bacinoglu, E. Uysal-Biyikoglu, and G. Durisi. Age-optimal channel coding blocklength for an M/G/1 queue with HARQ. In *Proc. IEEE SPAWC*, June 2018.
- [33] R. Devassy, G. Durisi, G. C. Ferrante, O. Simeone, and E. Uysal-Biyikoglu. Delay and peak-age violation probability in short-packet transmissions. In *Proc. IEEE ISIT*, June 2018.
- [34] S. Feng and J. Yang. Age-optimal transmission of rateless codes in an erasure channel. In *Proc. IEEE ICC*, May, 2019.
- [35] X. Chen and S. S. Bidokhti. Benefits of coding on age of information in broadcast networks. In *Proc. ITW*, August 2019.
- [36] A. Arafa, K. Banawan, K. G. Seddik, and H. V. Poor. On timely channel coding with hybrid ARQ. In *Proc. IEEE Globecom*, December 2019.
- [37] R. Wang, Y. Gu, H. Chen, Y. Li, and B. Vucetic. On the age of information of short-packet communications with packet management. In *Proc. IEEE Globecom*, December 2019.
- [38] S. Feng and J. Yang. Adaptive coding for information freshness in a two-user broadcast erasure channel. In *Proc. IEEE Globecom*, December 2019.
- [39] E. Najm, E. Telatar, and R. Nasser. Optimal age over erasure channels. Available Online: arXiv:1901.01573.
- [40] A. Javani, M. Zorghi, and Z. Wang. On the age of information in erasure channels with feedback. Available Online: arXiv:1911.05840.
- [41] J. Chakravorty and A. Mahajan. Distortion-transmission trade-off in real-time transmission of Gauss-Markov sources. In *Proc. IEEE ISIT*, June 2015.
- [42] X. Gao, E. Aykol, and T. Basar. Optimal estimation with limited measurements and noisy communication. In *Proc. IEEE CDC*, December 2015.
- [43] J. Yun, C. Joo, and A. Eryilmaz. Optimal real-time monitoring of an information source under communication costs. In *Proc. IEEE CDC*, December 2018.
- [44] O. Ayan, M. Vilgelm, M. Klugel, S. Hirche, and W. Kellerer. Age-of-information vs. value-of-information scheduling for cellular networked control systems. In *Proc. IEEE ICCPS*, April 2019.
- [45] A. Mitra, J. A. Richards, S. Bagchi, and S. Sundaram. Finite-time distributed state estimation over time-varying graphs: Exploiting the age-of-information. In *Proc. ACC*, July 2019.
- [46] J. Chakravorty and A. Mahajan. Remote estimation over a packet-drop channel with Markovian state. *IEEE Trans. Autom. Control*. To appear.
- [47] Y. Sun, Y. Polyanskiy, and E. Uysal-Biyikoglu. Remote estimation of the Wiener process over a channel with random delay. *IEEE Trans. Inf. Theory*. To appear.
- [48] T. Z. Ornee and Y. Sun. Sampling for remote estimation through queues: Age of information and beyond. Available Online: arXiv:1902.03552.
- [49] K. Huang, W. Liu, M. Shirvanimoghaddam, Y. Li, and B. Vucetic. Real-time remote estimation with hybrid ARQ in wireless networked control. Available Online: arXiv:1903.12472.
- [50] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides. The age of incorrect information: A new performance metric for status updates. Available Online: arXiv:1907.06604.
- [51] D. Ramirez, E. Erkip, and H. V. Poor. Age of information with finite horizon and partial updates. Available Online: arXiv:1910.00963.
- [52] M. Bastopcu and S. Ulukus. Age of information for updates with distortion: Constant and age-dependent distortion constraints. Available Online: arXiv:1912.13493.
- [53] M. Bastopcu and S. Ulukus. Partial updates: Losing information for freshness. Available Online: arXiv:2001.11014.
- [54] G. E. Uhlenback and L. S. Ornstein. On the theory of the Brownian motion. *Phys. Rev.*, 36:823–841, September 1930.
- [55] J. L. Doob. The Brownian movement and stochastic equations. *Ann. Math.*, 43(2):351–369, 1942.
- [56] T. Cover and J. A. Thomas. *Elements of Information Theory*. John Wiley & Sons, 2006.
- [57] W. Dinkelbach. On nonlinear fractional programming. *Management Science*, 13(7):492–498, 1967.
- [58] S. P. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.