On the Timeliness of Arithmetic Coding

Shirin Saeedi Bidokhti
Electrical and System Engineering
University of Pennsylvania, USA
saeedi@seas.upenn.edu

Aylin Yener
Electrical and Computer Engineering
The Ohio State University, USA
yener@ece.osu.edu

***Abstract*—Timeliness of information transfer is critical in real-time applications. Prioritizing timeliness, however, often comes at the cost of rate inefficiency, especially in block coding. In this work, motivated by the sequential nature of encoding and decoding in arithmetic source coding, the timeliness of arithmetic coding is investigated. For a generate-at-will source model, an upper bound is provided on the average peak age of information (PAoI). This upper bound builds on the arithmetic coding scheme of Shayevitz et al. (2007) which has a finite look-ahead parameter d. It captures interesting trade-offs between PAoI, compression rate, and the look-ahead parameter d. For periodic sources, rate efficiency is argued to be less critical than the look-ahead parameter d in minimizing the peak age, especially when the traffic load is moderate and small. Through simulations, two observations are made: (i) the optimal look-ahead parameter d is an increasing function of the traffic load, and (ii) asymptotically, as the traffic load gets close to its limit 1, the classical arithmetic coding (without a finite bound on d) performs better than the state-of-the-art age-optimal block codes.**

I. Introduction

Age of information (AoI) is an important performance metric for real-time sensing, estimation, and control applications where freshness of information is critical. AoI is a metric that is not directly captured by either rate or delay of a system, although it is affected by both. Originally targeting status updating systems, AoI captures how timely the information is from the receiver's perspective by tracking and accounting for the time since the generation of information at the source. Reference [1] and follow-up work, e.g., [2]–[4] use queuing theory to analyze average (or peak) age of information, see also [5], [6] and the references therein. Additionally, AoI has been extensively studied in information-freshness focused scheduling for wireless networks [7]–[17], in remote estimation and control [18]–[21], and in many other settings, e.g, [22]–[28]. Recent works have also considered new source and channel code designs to minimize the average age of information [29]–[35]. This work focuses on timely source coding. Below, we review literature related to our work.

A. Related Work

Reference [36] studies age minimization in lossless block coding for a streaming source with symbols that arrive periodically. The authors show that maximizing the error exponent is not equivalent to age minimization. They further use results from queuing theory to relate AoI to the first and second moments of codeword lengths. With this approach, they propose age-optimal block codes. The case of random source symbol arrivals is considered in [37]. Reference [38] considers a setting in which the encoder is informed of the busy/idle state at the channel interface. The authors propose a strategy in which the encoder switches among codebooks with different source blocklengths based on the backlog of symbols at the encoder, balancing delay and compression efficiency. In [39], variable length coding is investigated for random source arrivals and it is shown that it outperforms block coding when the traffic load is large. Distinct from the above-mentioned settings, [40] considers a generate-at-will model aiming only at age minimization. Compared to previous settings, this means that any source symbol that is generated while the system is busy will be discarded for newly generated packets make the old ones obsolete with respect to the age of information. In this model and through an optimization framework, [40] shows that asymptotic minimum average age can be attained up to a constant gap by Shannon codes that are designed for a tilted version of the original pmf.

B. Contributions

It is known that there are fundamental tradeoffs between rate efficiency and age efficiency, especially in block coding. This is because coding takes advantage of the typical behaviour of sequences while timeliness requires immediate encoding and decoding. In this sense, it is natural to look for sequential encoding/decoding strategies that encode/decode depending on the realization of the source as soon as they can. But this is just what arithmetic coding does with asymptotic optimality. As a matter of fact, it is known that arithmetic coding has the best known redundancy-delay tradeoff [41], where the term "delay" is simply the number of future source symbols the encoder needs to encode for successful decoding. To avoid confusion, we refer to this "delay" term as the look-ahead parameter of the scheme. Motivated by the sequential nature of encoding and decoding in arithmetic coding, we study its age performance in two source models: generate-at-will and periodic generation. Our findings are as follows:

- Under the generate-at-will model, we provide an upper bound on the average peak AoI, capturing tradeoffs between age, compression rate and the look-ahead parameter of the arithmetic coding proposed in [41].
- In periodic source models, we argue that minimizing the compression rate becomes less critical for (peak) age minimization than minimizing the look-ahead parameter of the scheme, especially when the traffic load is not large. We

then propose a scheme based on the mismatched arithmetic coding of [41] to trade rate optimality with the look-ahead parameter of the scheme.

- We numerically investigate the average AoI performance of the proposed scheme and show that the desired look-ahead d is an increasing function of the traffic load. Moreover, as the traffic load increases, asymptotically approaching 1, the classical arithmetic coding (matched to the original source) outperforms state-of-the-art age-efficient block coding schemes in terms of average AoI.

II. System Model and Preliminaries

A. Two Settings

Consider a discrete memoryless source X with alphabet $\mathcal{X} = \{0, 1, \ldots, K-1\}$, pmf $p(x)$, $x \in \mathcal{X}$, and entropy $H(X)$ that is to be compressed and communicated by a noiseless channel. We consider two information generation models. First, we consider a *generate-at-will* model in which the source generates an iid source symbol as soon as the channel becomes available. If the channel is already available, we account for a processing time of δ for the generation and/or processing of the next symbol. The second model we consider is a *periodic generation* processes in which source symbols arrive every time unit.

The source feeds information to an encoder which outputs L_i bits for the source symbol X_i and sends them to a decoder over a noiseless channel. L_i is a non-negative integer random variable. The delay of transmission per bit is assumed to be ν time units and the encoded bits go through the channel on a first come first serve basis.

Suppose that symbol X_i is generated at time $\tau(i)$. Under the periodic generation model, we simply have

$$\tau(i) = i \qquad \text{(periodic generation).}$$

By contrast, under the generate-at-will model, generation time $\tau(i)$ is a function of all past L_j, $j \le i$. More specifically, if $L_i = 0$, the channel remains available and the source immediately generates another symbol with δ processing time, i.e., X_{i+1} is generated at time $\tau(i+1) = \tau(i) + \delta$. Otherwise, $\tau(i+1) = \tau(i) + \nu L_i$. Under this model, the generation time of the i^{th} symbol is thus

$$\tau(i) = \sum_{j=1}^{i-1} \max(\nu L_j, \delta) \qquad \text{(generate-at-will).}$$

The decoder is interested in the timely recovery of the source symbols in a lossless manner. Denote the time at which the source symbol X_i is decoded by $t(i)$. Right after decoding X_i, the age of information drops at the decoder. It can happen that the decoder decodes multiple source symbols at a time. We assume a small processing time δ between decoded symbols. That is, the age is at peak right before X_i is decoded. This is a reasonable assumption since source symbols are often recovered sequentially in practice. This assumption simplifies our analysis. Figure 1 shows a sample age evolution

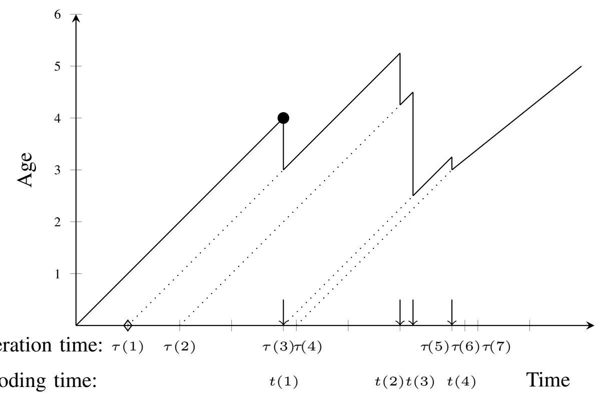

Fig. 1: A sample path for age of information under the generate-at-will source model.

in our generate-at-will model. Generation times $\{\tau(i)\}_i$ and delivered times $\{t(i)\}$ are marked on the figure.

To provide understanding on the timeliness of arithmetic coding and its tradeoff with rate optimality, we consider the metric of average (peak) age of information.

Consider time horizon $T \to \infty$ and suppose that the first $m(T)$ source symbols are decoded, where $m(T)$ is a random variable. Right before decoding each symbol X_i, $i = 1, \ldots, m(T)$, we have a peak age of information[1] denoted by $A(i)$. After delivery, age drops to $D(i)$ which encodes the delay in decoding X_i. The average peak age of information (PAoI) is

$$\text{PAoI} = \lim_{T\to\infty} \mathbb{E}\left[\frac{\sum_{i=1}^{m(T)} A(i)}{m(T)}\right]. \tag{1}$$

This metric is utilized in Sections III-IV. The average age of information (AoI) is

$$\text{AoI} = \lim_{T\to\infty} \mathbb{E}\left[\frac{\sum_{i=1}^{m(T)} \Delta(i)}{m(T)}\right]. \tag{2}$$

where $\Delta(i) = \frac{1}{2}\left(A(i)^2 - D(i-1)^2\right)$. This metric is used in Section V.

B. Arithmetic Coding Preliminaries

Arithmetic coding is known to be an asymptotically rate-optimal lossless data compression technique. The idea is to sequentially map a source sequence to a sequence of shrinking intervals in $[0, 1]$ that correspond to and contain the probability of the source sequence and can be represented by a binary sequence. It is known that there are infinitely many sequences for which arithmetic coding has unbounded "delay" where the term delay traditionally describes the lag between the encoded and decoded source symbols. The lag is associated to what [41] calls the "forbidden points". If the encoder interval contains a forbidden point, the lag can become unbounded.

[1]We assume a minimum processing time for decoding and, therefore, there is a peak right before decoding each source symbol

In [42], it was shown that the lag between the encoded and decoded source symbols is bounded in expectation. [41] made the lag finite by introducing a mismatched encoder with two additional (fictitious) source symbols. When the lag reaches the threshold d, then one of the fictitious symbols is sent by the mismatched encoder to force immediate decoding. These fictitious symbols are such that they change the encoding interval into a smaller one with no forbidden points and thus make decoding possible and reset the lag. We refer to d as the look-ahead parameter of the scheme.

Naturally, a bounded d comes at the cost of some redundancy in compression rate. This redundancy is due to (i) the fact that the arithmetic encoder with fictitious symbols is slightly mismatched with the original source and (ii) describing the fictitious symbols requires additional representative bits. By finding the right balance between these two sources of suboptimality, [41] finds tradeoffs between the look-ahead parameter d and redundancy.

We use the results of [41] and further build on the idea of mismatched arithmetic coding for age minimization. We refer to the mismatched source (with the two fictitious symbols) as $\tilde{X}$ with pmf $p_{\tilde{X}}(\tilde{x})$ and alphabet set $\tilde{X} \in \mathcal{X} \cup \{s_1, s_2\}$. For example, [41] designed $\tilde{X}$ to be given by

$$p_{\tilde{X}}(j) = \frac{p_X(j)}{1-2\epsilon}, \quad j \in \mathcal{X} \tag{3}$$
$$p_{\tilde{X}}(s_1) = p_{\tilde{X}}(s_2) = \epsilon$$

where $\epsilon = \alpha^d(1 + d\log(1/\alpha))\log e$.

Lemma 1 (cf. [41])**.** *Let L_i be the number of new bits that are output by the encoder after reading X_i. Under the scheme of [41], the compression rate is*

$$\lim_{m\to\infty} \frac{1}{m}\sum_{i=1}^{m} L_i = H(X) + \mathcal{R}(d)$$

where

$$\mathcal{R}(d) \le 4\alpha^{d-1-c(\alpha)}\left(1 + (d-1-c(\alpha))\log\left(\frac{1}{\alpha}\right)\right)^2 \tag{4}$$

$$\alpha = \max_{j\in\mathcal{X}} p(j) \tag{5}$$

$$c(\alpha) = \begin{cases} 0 & \alpha \le \frac{1}{16} \\ 2\lfloor\frac{4}{\log(1/\alpha)}\rfloor - 1 & \textit{otherwise.}\end{cases} \tag{6}$$

The following proposition is straightforward for the scheme of [41] based on the analysis in [42].

Lemma 2. *Let $n(i)$ be the number of bits received by the decoder right before $\tilde{X}_i$, $\tilde{X}_i \in \tilde{\mathcal{X}}$, is decoded following the scheme of [41]. The following holds.*

$$\mathbb{E}[n(m)] \le m\left(H(X) + \mathcal{R}(d)\right) + \log(4e). \tag{7}$$

Remark 1. *Note, as also remarked in [42], that $\mathbb{E}[n(m)]$ is different than the average number of bits produced, because $n(m)$ is the number of bits produced until a symbol is decoded. So the latter is often larger than the former.*

III. The Generate-at-Will Model

We now study the average peak age of information for the arithmetic coding scheme of [41] (with finite look-ahead d) under the generate-at-will model. After decoding a source symbol, the destination's information is updated with a more recent source symbol. Right before decoding, the age is at a peak. Suppose that the source symbol X_i is decoded. Recall that $t(i)$ is the time at which X_i is decoded. The peak age of information before decoding X_i is $t(i) - \tau(i-1)$ and the age of information drops to $t(i) - \tau(i)$ after decoding X_i. The peak age before decoding X_i can be upper bounded as follows:

$$\begin{aligned} A(i) &= t(i) - \tau(i-1) && (8)\\ &= t(i) - \sum_{j=1}^{i-2}\max(\nu L_j, \delta) && (9)\\ &\le \nu\left(n(i) - \sum_{j=1}^{i-2} L_j\right) + 2d\delta && (10)\end{aligned}$$

where $n(i)$ is the number of bits delivered before X_i is decoded and $\sum_{j=1}^{i-2} L_j$ is the number of bits communicated before X_{i-1} was generated. Besides the term $n(i) - \sum_{j=1}^{i-2} L_j$ that partly constitutes $t(i) - \tau(i-1)$, we have a term of order $O(d\delta)$. The precise description of this term which we have upper bounded by $2d\delta$ is $(d_i + u_i)\delta$ where d_i is the look-ahead needed for the decoding of X_i and u_i is the number of symbols decoded simultaneously at time $t(i)$. Since the look-ahead parameter of the of arithmetic coding in [41] is bounded by the constant d, both terms are bounded by d.

For arithmetic coding with finite look-ahead parameter d, we prove the following upper bound on PAoI.

Theorem 1. *Given a generate-at-will source model with transmission delay ν per bit, the average peak age of information (PAoI) is upper bounded as follows:*

$$\textit{PAoI} \le 2\nu\left(H(X) + \mathcal{R}(d)\right) + \nu\log(4e) + 2d\delta. \tag{11}$$

Proof. Recall that $m(T)$ denotes the number of decoded source symbols up to time T. Let

$$R^*(d) := H(X) + \mathcal{R}(d)$$

and define event $\mathcal{E}_T^d$ as follows:

$$\mathcal{E}_T^d = \left\{ \begin{array}{c} m(T) \ge \frac{1}{\nu R^*(d)}T(1 - O(\delta)) - o(T) \\ m(T) \le \mathbb{E}[m(T)] + o(T) \end{array}\right\} \tag{12}$$

where the term $O(\delta)$ is essentially the long term expected portion of the symbols X_i that are encoded with zero bits

($L_i = 0$). This event occurs with high probability as T gets large. The average peak AoI can now be written as follows:

$$\begin{aligned}
&\lim_{T\to\infty} \mathbb{E}\left[\frac{\sum_{i=1}^{m(T)} A(i)}{m(T)}\right] \\
&= \lim_{T\to\infty} \mathbb{E}\left[\frac{\sum_{i=1}^{m(T)} A(i)}{m(T)}\Big|\mathcal{E}_T^d\right] \Pr(\mathcal{E}_T^d) \\
&\quad + \lim_{T\to\infty} \mathbb{E}\left[\frac{\sum_{i=1}^{m(T)} A(i)}{m(T)}\Big|\mathcal{E}_T^{d\,c}\right] \Pr(\mathcal{E}_T^{d\,c}) \\
&\leq \lim_{T\to\infty} \mathbb{E}\left[\frac{\sum_{i=1}^{m(T)} A(i)}{m(T)}\Big|\mathcal{E}_T^d\right] \Pr(\mathcal{E}_T^d) + O(1)\Pr(\mathcal{E}_T^{d\,c})
\end{aligned}$$

where the inequality holds because the look-ahead parameter (and hence the delay and age $A(i)$) are bounded per design of [41]. Since the second term vanishes as T gets large, we focus on the first term.

$$\begin{aligned}
&\mathbb{E}\left[\frac{\sum_{i=1}^{m(T)} A(i)}{m(T)}\Big|\mathcal{E}_T^d\right] \Pr(\mathcal{E}_T^d) \\
&\overset{(a)}{\leq} \mathbb{E}\left[\frac{\nu R^*(d)\sum_{i=1}^{m(T)} A(i)}{T(1-O(\delta))-o(T)}\Big|\mathcal{E}_T^d\right] \Pr(\mathcal{E}_T^d) \\
&\overset{(b)}{\leq} \nu R^*(d)\mathbb{E}\left[\frac{\nu\sum_{i=1}^{m(T)}\left(n(i)-\sum_{j=1}^{i-2}L_j\right)+2m(T)d\delta}{T(1-O(\delta))-o(T)}\Big|\mathcal{E}_T^d\right]\Pr(\mathcal{E}_T^d) \\
&\overset{(c)}{\leq} \frac{\sum_{i=1}^{\mathbb{E}[m(T)]}\nu^2 R^*(d)\mathbb{E}\left[n(i)-\sum_{j=1}^{i-2}L_j\Big|\mathcal{E}_T^d\right]\Pr(\mathcal{E}_T^d)+o(T)}{T(1-O(\delta))-o(T)} \\
&\quad + \frac{2\nu R^*(d)\mathbb{E}[m(T)]d\delta}{T(1-O(\delta))-o(T)} \\
&\overset{(d)}{\leq} \frac{\sum_{i=1}^{\mathbb{E}[m(T)]}\nu^2 R^*(d)\mathbb{E}\left[n(i)-\sum_{j=1}^{i-2}L_j\right]+o(T)}{T(1-O(\delta))-o(T)} \\
&\quad + \frac{2\nu R^*(d)\mathbb{E}[m(T)]d\delta}{T(1-O(\delta))-o(T)}. \qquad (13)
\end{aligned}$$

In the above chain of inequalities, (a) follows by the definition of $\mathcal{E}_T^d$ in (12), (b) follows by (10), (c) holds by the definition of $\mathcal{E}_T^d$, in particular because $m(T) \leq \mathbb{E}[m(T)] + o(T)$ and $n(i) - \sum_{j=1}^{i-2} L_j$ is bounded by:

$$n(i) - \sum_{j=1}^{i-2} L_j \leq (d+2)\max L_i \leq (d+1)^2 \log(\frac{1}{p_{\min}}).$$

Finally (d) holds because $n(i) - \sum_{j=1}^{i-2} L_j$ is positive.

Taking the limit $T \to \infty$ in (13) for finite d, we obtain

$$\begin{aligned}
\text{PAoI} &\leq \lim_{T\to\infty} \frac{\sum_{i=1}^{\mathbb{E}[m(T)]}\nu^2 R^*(d)\mathbb{E}\left[n(i)-\sum_{j=1}^{i-2}L_j\right]}{T(1-O(\delta))} + 2d\delta \\
&\overset{(e)}{\leq} \lim_{T\to\infty} \frac{\mathbb{E}[m(T)]\nu^2 R^*(d)(2R^*(d)+\log(4e))}{T(1-O(\delta))} + 2d\delta \\
&= 2\nu R^*(d) + \nu\log(4e) + 2d\delta.
\end{aligned}$$

Step (e) follows from Lemma 2 and the fact that $\frac{1}{i}\sum_{j=1}^{i}\mathbb{E}[L_j] \approx R^*(d)$ for large i. This concludes the proof. □

IV. Periodic Generation

The second source model assumes periodic generation of the source symbols. There is a major difference between this model and the one we studied in Section III. Below, we will argue that even though efficiency in compression is still desired to minimize network congestion and timeliness, its role is not as dramatic in this setting.

The peaks of the age of information function $A(i)$ satisfy

$$\begin{aligned}
\mathbb{E}[A(i)] &= \mathbb{E}[t(i) - (i-1)] \\
&\leq 1 + \mathbb{E}[W_i] + \mathbb{E}[n(i) - n(i-1)] + \mathbb{E}[d_i\beta_i] + O(d\delta)
\end{aligned} \qquad (14)$$

where W_i is the waiting time for symbol X_i to be encoded, d_i is the look-ahead required for X_i to be decoded and β_i is the portion of time in which there is no bit transmission in time interval $[i, i+d_i)$. For example, if $W_i > d_i$, then $\beta_i = 0$. Analyzing $\mathbb{E}[W_i]$ is non-trivial because the number of bits produced for symbols $\{X_i\}_i$ (and hence the corresponding service times) are dependent and so typical results from queuing theory don't apply. Nonetheless, (14) still provides us with useful intuitions, and guides us to new arithmetic-type coding schemes.

There are multiple coupled factors playing role in (14): (i) W_i (which depends on the traffic load of the channel and the rate-efficiency of the encoder), (ii) $n(i) - n(i-1)$ which is the number of additional bits that the decoder needs after recovery of X_{i-1} to decode X_i (and is governed by the rate-efficiency of the encoder), and (iii) $d_i\beta_i$ (which is governed by the look-ahead parameter of the scheme, the traffic load of the channel and the rate-efficiency of the encoder). As traffic load increases, W_i increases and β_i decreases. In its extreme case, β_i vanishes and (14) would be minimized by minimizing the compression rate. When traffic load is small, however, W_i is small and the dominant factor governing (14) is d_i. So minimizing d as a finite upper bound on d_i, even at the cost of a larger compression rate, may be desired. This suggests devising schemes with a smaller look-ahead parameter d. We build on the mismatched arithmetic code design of [41]. Given a source X with pmf $p_X(.)$, add two fictitious source symbols and perturb the source pmf into $p_{\tilde{X}}(.)$ with the alphabet set $\tilde{\mathcal{X}} = \mathcal{X} \cup \{s_1, s_2\}$, such that the following assumption holds.

Assumption 1. *There is an ordering of the source alphabet through a map $g : \mathcal{X} \cup \{s_1, s_2\} \to \{0, 1, \ldots, K+1\}$ such that the cumulative distribution function $f_{\tilde{X}}$ satisfies*

$$\frac{3}{8} \leq f_{\tilde{X}}(g(s_1)-1), \quad f_{\tilde{X}}(g(s_1)) < \frac{1}{2} \qquad (15)$$

$$\frac{1}{2} < f_{\tilde{X}}(g(s_2)), \quad f_{\tilde{X}}(g(s_2)+1) \leq \frac{5}{8}. \qquad (16)$$

Using an arithmetic encoder matched to the source $\tilde{X}$ with properties stated in Assumption 1, we achieve a smaller d at the cost of a larger rate. More specifically, every time the look-ahead parameter reaches the target value d, the encoder inserts one of the two fictitious symbols. Assumption 1 makes immediate decoding possible (see also [41]).

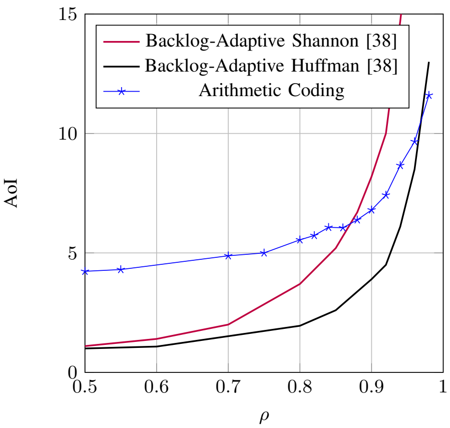

Fig. 2: AoI of a source with uniform generation process.

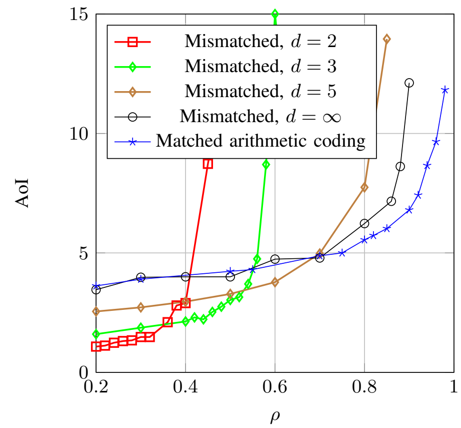

Fig. 3: Impact of the look-ahead parameter d.

Remark 2. *While Assumption 1 also underlies the designs in [41], the challenge in [41] is to choose the perturbation carefully so that rate efficiency is not compromised too much while keeping the look-ahead bounded. As discussed above, rate optimality is not always essential in the periodic source model. Therefore, we have the freedom to perturb the source however we want as long as Assumption 1 is satisfied.*

V. Simulation Results

In this section, we provide a summary of our simulation results, shedding light on the tradeoff between rate, age, and the look-ahead parameter d in arithmetic coding. While we worked with the metric of peak AoI in previous sections, we now consider average AoI for our simulations in order to have a fair comparison with the state-of-art.

In Figure 2, we consider a periodic source with pmf $p_X(0) = 0.6, p_X(1) = 0.3, p_X(2) = 0.1$ as was considered in [38]. The source symbols are generated every unit of time. Let $\rho = \nu H(X)$ be an indicator of the "traffic load". First, we compare the performance of arithmetic coding (in terms of AoI) with the backlog-adaptive schemes of [38]. We observe that in its classical form, arithmetic coding outperforms the schemes of [38] when ρ is very large. This is consistent with the asymptotic superiority of arithmetic coding, compared to block coding, in terms of the decoding lag [41]. When ρ is small, however, arithmetic coding does not perform well.

The reason behind the poor performance of arithmetic coding in the regime of small ρ is that while arithmetic coding seeks rate-optimality, it does so by encoding with a delay (waiting for future source symbols). The time elapsed in waiting for the generation of new source symbols contributes to the bad age performance of arithmetic coding. That time could have been used for transmitting bits without impacting AoI adversely. In other words, a smaller look-ahead parameter is desired even at the cost of a larger compression rate.

In Figure 3, we demonstrate the impact of d on AoI for different regimes of ρ. We encode the source using a mismatched arithmetic encoder with two fictitious symbols. For this purpose, we deviate from the design of [41] which targets finite (but not necessarily small) values of d and follows the approach we outlined in Section IV. For the plot in Figure 3, we have perturbed the original source pmf to $p_{\tilde{X}}(0) = 0.49$, $p_{\tilde{X}}(1) = 0.3$, $p_{\tilde{X}}(2) = 0.1$ in order to allow for $p_{\tilde{X}}(s_1) = 0.01$ and $p_{\tilde{X}}(s_2) = 0.1$. With this choice Assumption 1 holds and therefore s_1 and s_2 can be mapped to intervals without forbidden points. This perturbed source is not dependent on d and does not converge to the original source as d gets large. This explain why the black curve (for mismatched arithmetic coding with $d \to \infty$) is sub-optimal compared to the (classical) matched arithmetic encoder for large ρ. Optimal source perturbation methods remains open for further exploration. We observe in the plot of Figure 3 that imposing a judicious tradeoff between the look-ahead and rate of arithmetic coding, we can improve the AoI. In particular, as ρ increases, a larger look-ahead d is desired.

VI. Conclusion

In this paper, we have studied the interplay between two important performance metrics in source coding: efficiency in compression and timeliness of decoding. We have done so through the lens of arithmetic coding, which is sequential by nature and is known to have the (order-wise) best asymptotic delay performance. Looking at generate-at-will and periodic source models, we have studied the tradeoff between age, rate, and the look-ahead parameter of arithmetic coding. We have demonstrated that, for periodic source models, optimal compression is not always desired for age minimization especially in regimes where the traffic load is not very large. Consequently, we have outlined an approach building on the mismatched arithmetic encoding of [41] to improve the timeliness of classical arithmetic coding by trading rate efficiency for a smaller look-ahead parameter.

An interesting future research direction is to devise optimal mismatched encoding schemes that can adapt to the backlog of the system.

Acknowledgement

This work was supported by NSF-CRII award 1910594, NSF-CNS 1850356 and NSF CAREER award 2047482.

References

[1] S. Kaul, R. Yates and M. Gruteset, "Real-time status: How often should one update?" in *2012 Proceedings IEEE INFOCOM*, 2012, pp. 2731 – 2735.

[2] ——, "Status updates through queues," in *2012 46th Annual Conference on Information Sciences and Systems (CISS)*, 2012, pp. 1 – 6.

[3] M. Costa, M. Codreanu and A. Ephremides, "On the Age of Information in Status Update Systems With Packet Management," *IEEE Transactions on Information Theory*, vol. 62, no. 4, pp. 1897 – 1910, 2016.

[4] R. Yates and S. Kaul, "The Age of Information: Real-Time Status Updating by Multiple Sources," *IEEE Transactions on Information Theory*, vol. 65, no. 3, pp. 1807 – 1827, 2019.

[5] A. Kosta, N. Pappas and V. Angelakis, "Age of information: A new concept, metric, and tool," *Foundations and Trends in Networking*, vol. 12, no. 3, 2017.

[6] R. D. Yates, Y. Sun, D. R. Brown III, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *arXiv preprint arXiv:2007.08564*, 2020.

[7] I. Kadota, A. Sinha, E. Uysal-Biyikoglu, R. Singh, and E. Modiano, "Scheduling Policies for Minimizing Age of Information in Broadcast Wireless Networks," *IEEE/ACM Transactions on Networking*, vol. 26, pp. 2637 – 2650, 2018.

[8] I. Kadota, A. Sinha, E. Modiano, "Optimizing Age of Information in Wireless Networks with Throughput Constraints," in *IEEE INFOCOM 2018 - IEEE Conference on Computer Communications*, 2018, pp. 1844 – 1852.

[9] R. Talak, S. Karaman and E. Modiano, "Distributed Scheduling Algorithms for Optimizing Information Freshness in Wireless Networks," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2018, pp. 1 – 5.

[10] I. Kadota, A. Sinha, E. Modiano, "Scheduling Algorithms for Optimizing Age of Information in Wireless Networks with Throughput Constraints," *IEEE/ACM Transactions on Networking*, vol. 27, no. 4, pp. 1359 – 1372, 2019.

[11] I. Kadota and E. Modiano, "Minimizing the age of information in wireless networks with stochastic arrivals," *IEEE Transactions on Mobile Computing*, vol. Early Access, 2019.

[12] C. Joo and A. Eryilmaz, "Wireless Scheduling for Information Freshness and Synchrony: Drift-Based Design and Heavy-Traffic Analysis," *IEEE/ACM Transactions on Networking*, vol. 26, no. 6, pp. 2556 – 2568, 2018.

[13] Q. He, D. Yuan and A. Ephremides, "Optimal Link Scheduling for Age Minimization in Wireless Systems," *IEEE Transactions on Information Theory*, vol. 64, no. 7, pp. 5381 – 5394, 2018.

[14] X. Chen and S. Saeedi Bidokhti, "Benefits of Coding on Age of Information in Broadcast Networks," in *2019 IEEE Information Theory Workshop (ITW)*, 2019.

[15] S. Leng and A. Yener, "Age of information minimization for wireless ad hoc networks: A deep reinforcement learning approach," in *IEEE Global Communications Conference (GLOBECOM)*, 2019.

[16] ——, "An actor-critic reinforcement learning approach to minimum age of information scheduling in energy harvesting networks," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2021.

[17] X. Chen, R. Liu, S. Wang, and S. Saeedi Bidokhti, "Timely Broadcasting in Erasure Networks: Age-Rate Tradeoffs," in *2021 IEEE International Symposium on Information Theory (ISIT)*, 2021.

[18] T. Z. Ornee and Y. Sun, "Sampling for remote estimation through queues: Age of information and beyond," in *2019 International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOPT)*. IEEE, 2019, pp. 1–8.

[19] X. Chen, X. Liao, and S. S. Bidokhti, "Remote estimation in decentralized random access channels," *arXiv preprint arXiv:2007.03652*, 2020.

[20] A. Arafa, K. Banawan, K. G. Seddik and V. Poor, "Timely Estimation Using Coded Quantized Samples," arXiv: 2004.12982, 2020.

[21] ——, "Sample, Quantize and Encode: Timely Estimation over Noisy Channels," arXiv: 2007.10200, 2020.

[22] X. Chen, K. Gatsis, H. Hassani, and S. Saeedi Bidokhti, "Age of Information in Random Access Channels," arXiv:1912.01473, 2019.

[23] A. M. Bedewy, Y. Sun, and N. B. Shroff, "The age of information in multihop networks," *IEEE/ACM Transactions on Networking*, vol. 27, no. 3, pp. 1248–1257, 2019.

[24] S. Leng and A. Yener, "Age of information minimization for an energy harvesting cognitive radio," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 2, pp. 427–439, 2019.

[25] A. Baknina, O. Ozel, J. Yang, S. Ulukus, and A. Yener, "Sending information through status updates," in *Proceedings of IEEE International Symposium on Information Theory (ISIT)*, 2018.

[26] M. Bastopcu and S. Ulukus, "Minimizing age of information with soft updates," *Journal of Communications and Networks*, vol. 21, no. 3, pp. 233–243, 2019.

[27] R. D. Yates, P. Ciblat, A. Yener, and M. Wigger, "Age-optimal constrained cache updating," in *IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 141–145.

[28] A. Arafa, J. Yang, S. Ulukus, and H. V. Poor, "Age-minimal transmission for energy harvesting sensors with finite batteries: Online policies," *IEEE Transactions on Information Theory*, vol. 66, no. 1, pp. 534–556, 2020.

[29] K. Chen and L. Huang, "Age-of-Information in the Presence of Error," in *2016 IEEE International Symposium on Information Theory (ISIT)*, 2016.

[30] R. Yates, E. Najm, E. Soljanin and J. Zhong, "Timely updates over an erasure channel," in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 316 – 320.

[31] P. Parag, A. Taghavi and J. Chamberland, "On Real-Time Status Updates over Symbol Erasure Channels," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*, 2017, pp. 1 – 6.

[32] H. Sac, T. Bacinoglu, E. Uysal-Biyikoglu and G. Durisi, "Age-Optimal Channel Coding Blocklength for an M/G/1 Queue with HARQ," in *2018 IEEE 19th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, 2018, pp. 1 – 5.

[33] E. Najm, E. Telater and R. Nasser, "Optimal Age over Erasure Channels," arXiv: 1901.01573, 2019.

[34] S. Feng and J. Yang, "Adaptive Coding for Information Freshness in a Two-user Broadcast Erasure Channel," arXiv: 1905.00521, 2019.

[35] M. Costa and Y. Sagduyu, "Age of information with network coding," *Ad Hoc Networks*, vol. 86, pp. 15 – 22, 2019.

[36] J. Zhong and R. D. Yates, "Timeliness in lossless block coding," in *2016 Data Compression Conference (DCC)*, 2016, pp. 339–348.

[37] J. Zhong, R. D. Yates, and E. Soljanin, "Timely lossless source coding for randomly arriving symbols," 2018.

[38] J. Zhong, R. D. Yates, and E. Soljanin, "Backlog-adaptive compression: Age of information," in *2017 IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 566–570.

[39] U. Abend and A. Khina, "Real-time variable-to-fixed lossless source coding of randomly arriving symbols," 2020.

[40] P. Mayekar, P. Parag, and H. Tyagi, "Optimal source codes for timely updates," *IEEE Transactions on Information Theory*, vol. 66, no. 6, pp. 3714–3731, 2020.

[41] O. Shayevitz, E. Meron, M. Feder, and R. Zamir, "Delay and redundancy in lossless source coding," *IEEE Transactions on Information theory*, vol. 60, no. 9, pp. 5470–5485, 2014.

[42] R. Gallager, "Lecture Notes (Arithmetic Coding Supplementary Notes)," available on: http://web.mit.edu/gallager/www/notes/notes3.pdf, 1994.