# Status Prediction and Data Aggregation for AoI-Oriented Short-Packet Transmission in Industrial IoT

Qinqin Xiong, Xu Zhu, *Senior Member, IEEE*, Yufei Jiang, *Member, IEEE*, Jie Cao, *Member, IEEE*, Xiaogang Xiong, *Member, IEEE*, and Heng Wang, *Senior Member, IEEE*

***Abstract*— Age of information (AoI) is an effective performance metric for time-critical industrial Internet of things (IIoT) applications. We investigate status prediction and data aggregation with prediction error awareness, to enhance the AoI performance for short-packet transmission (SPT) in time-critical IIoT. A predict-compare (PredComp) transmission scheme is proposed, where proactive transmission termination is employed in case of prediction error, by comparing the predicted and real updates at source. It is proved to achieve a significant average AoI performance gain over the case without prediction, even under high prediction error probability. In addition, a predict-aggregate-compare (PredAggComp) transmission scheme is proposed, where two status updates are predicted with different prediction horizons and aggregated by utilizing their time correlation. That allows a good tradeoff between the prediction accuracy and the transmission error probability. A closed-form threshold that the PredAggComp scheme outperforms the PredComp scheme is derived. Moreover, prediction horizon adaptation is conducted to minimize the average AoI of the proposed transmission schemes. Simulation results verify the analytical results and show the superiority of the proposed PredComp and PredAggComp schemes, with an average AoI reduction of up to 64% over the case without prediction.**

Manuscript received 7 May 2022; revised 4 October 2022 and 7 November 2022; accepted 21 November 2022. Date of publication 1 December 2022; date of current version 16 January 2023. This work was supported in part by the Natural Science Foundation of China under Grant 62171161; in part by the International Science and Technology Cooperation Program of Guangdong Province under Grant 2022A0505050022; in part by the Natural Science Foundation of Guangdong Province under Grant 2021A1515011832; in part by the Shenzhen Key Laboratory under Grant ZDSYS20210623091808025; and in part by the Shenzhen Science and Technology Program under Grants GXWD20220817133854003, KQTD20190929172545139, JCYJ20210324133009027, JCYJ20210324131010028, JCYJ20190806145001754 and GXWD20200803230628015. An earlier version of this paper was presented at the IEEE WCNC 2022 [DOI: 10.1109/WCNC51071.2022.9771808]. The associate editor coordinating the review of this article and approving it for publication was H. Zhang. *(Corresponding author: Xu Zhu.)*

Qinqin Xiong is with the School of Electronic and Information Engineering, Harbin Institute of Technology, Shenzhen 518055, China (e-mail: 19b952021@stu.hit.edu.cn).

Xu Zhu and Yufei Jiang are with the School of Electronic and Information Engineering, Harbin Institute of Technology, Shenzhen 518055, China, and also with the Guangdong Provincial Key Laboratory, Aerospace Communication and Networking Technology, Shenzhen 518055, China (e-mail: xuzhu@ieee.org).

Jie Cao is with the Institute of Infocomm Research, Agency for Science, Technology and Research, Singapore 138632.

Xiaogang Xiong is with the School of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen 518055, China.

Heng Wang is with the Key Laboratory of Industrial Internet of Things and Networked Control, Chongqing University of Posts and Telecommunications, Chongqing 400065, China.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2022.3226188.

Digital Object Identifier 10.1109/TCOMM.2022.3226188

***Index Terms*— Industrial Internet of Things, short-packet transmission, age of information, status prediction, data aggregation.**

## I. Introduction

TIME-CRITICAL industrial Internet of things (IIoT) applications rely on acquisition of the latest status updates of the underlying physical processes to ensure real-time monitoring and accurate decision makings [2]. For example, in a real-time process tracking system for factory automation, a sensor captures the location and velocity status updates of a robot arm in the repetitive pick-and-place process. This information must be sensed and transmitted to the monitor in real time. The freshness of information is critical to these applications, since the outdated information is not of interest and may cause incorrect operations. Age of information (AoI) [3] is the measure of information freshness, which is defined as the time elapsed since the generation of the latest successfully received status update at the destination. Unlike the well-known delay metric, AoI incorporates the update rate, delay and reliability.

Most time-critical applications in IIoT require limited data size (typically 20-250 bytes [4]) and hence short-packet transmission (SPT) to reduce AoI as well as end-to-end delay [5], [6]. Compared to the conventional transmission assuming infinite blocklength, SPT suffers from an inevitable packet error [7] due to limited coding capability. This impairs the AoI performance. The data in IIoT, *e.g.*, the location and velocity sensor data, generally exhibit time correlation [8]. Hence, status prediction and data aggregation by exploiting the time correlation between data can help to improve the AoI performance.

### A. Related Work

The AoI performance of SPT has been investigated in [9], [10], [11], [12], [13], and [14]. The impact of blocklength on average AoI was analyzed in [9], and the optimal blocklength that minimizes the average AoI was derived. In [10], packet management strategies including preemption and retransmission were designed for a last-come-first-served (LCFS) system with SPT. In [11] and [12], the peak AoI (PAoI) violation probability was studied for short-packet systems operating with automatic repetition request (ARQ) protocol. Furthermore, the AoI tail distribution was characterized and optimized in [13]. In our previous work [14], the correlation between AoI and delay as well as their joint optimization were studied for an LCFS system. However, all the aforementioned work on

AoI-oriented SPT did not consider status prediction as well as data aggregation.

Status prediction is a promising approach to reducing the AoI, where source predicts a status update and transmits it to destination in advance, reducing the delay between information generation and reception. In [15], prediction of control commands at the controller was shown to improve the robustness of SPT systems against unreliable wireless links, where the predicted commands are sent to the actuator and executed if a real command is lost. In [8], by implementing a dual prediction algorithm at source and destination, the energy and bandwidth consumptions were reduced. The work of [15] and [8] respectively assumed no prediction error and no transmission error. These were considered in [16], where a low-latency communication scheme using status prediction was proposed for sources with multiple independent features.

Most existing work on status prediction [8], [15], [16] has focused on enhancing reliability, spectral/energy efficiency and delay performance by prediction. Very little work has been conducted to exploit the potential of status prediction in improving information freshness, especially in SPT. It was shown in [17] that with the assumption of no transmission and prediction errors, the prediction of user arrival patterns enables an average AoI reduction in mobile crowd-learning. The work in [18] further showed that status prediction reduces the average AoI of SPT based unmanned aerial vehicle relay systems in the presence of prediction and transmission errors. From an opposite perspective, the authors of [19] in a machine learning-based predictive control system analyzed the impact of AoI on the signal prediction credibility at controller/actuator as well as the control stability. In [18], a preliminary study of status prediction for AoI-oriented SPT was presented, while how to deal with the prediction error, and the tolerance of AoI performance against the prediction error were not analyzed. Also, the analysis of prediction horizon on AoI performance, as well as the prediction horizon adaptation were not provided in [18]. It is noteworthy that a longer prediction horizon leads to a lower delay between information generation and reception at the cost of a higher prediction error probability [16] and hence may impair the AoI performance.

Data aggregation, where several data are combined into a single packet to transmit to the destination, is envisioned to reduce the prediction error probability and hence improve the AoI performance of status-predicted SPT. This technique has been extensively investigated in IoT networks [8], [20], [21], [22], [23], [24], [25], [26], [27], [28]. It was shown that the aggregation of multiple data flows from different sources over the same link increases the average number of successful devices [20] and improves throughput [21], [22] and user fairness [23], by creating less interference within the shared spectrum. Such a multi-source data aggregation approach also saves energy and spectrum resources by eliminating recurring lower layer headers [24], [25]. In [8], data aggregation at cluster heads reduces traffic between cluster heads and sink nodes. In addition, the work in [26] showed that the accuracy of target tracking can be improved by implementing multi-source data aggregation. Unlike [8], [20], [21], [22], [23], [24], [25], and [26], data aggregation from a single source perspective was investigated in [27], in which the source temporarily stores the sampled data and aggregates them for transmission only if the data size reaches a threshold, leading to a significant energy saving. All the aforementioned work on data aggregation considered neither AoI nor SPT. In [28], multi-source data aggregation was studied for AoI-oriented SPT. Data aggregation of several predicted status updates with different prediction horizons can help to reduce the prediction error probability and improve the AoI performance. On the other hand, with a given blocklength, data aggregation introduces more information bits and less room for error correction code, and hence may result in a higher transmission error probability than the case of transmitting a single status update, which may impair the AoI performance. Therefore, it is of interest to investigate the ability of data aggregation to assist status prediction and to enhance the AoI performance, which has not been exploited for data aggregation in the previous work [8], [20], [21], [22], [23], [24], [25], [26], [27], [28].

In summary, the following open issues on AoI-oriented SPT remain: a)

1) How to design a prediction error-aware transmission scheme to enhance the AoI perormance, and what is the tolerance of the AoI performance against prediction error probability?
2) How to enhance the AoI performance of SPT via data aggregation alongside status prediction, and to make a tradeoff between prediction error probability and transmission error probability?
3) How to minimize the average AoI by prediction horizon adaptation?

### B. Contributions

Motivated by the above open issues, in this paper, we investigate status prediction and data aggregation with prediction error awareness for AoI-oriented SPT systems, to enable time-critical IIoT applications. Our contributions are as follows.

- A predict-compare (PredComp) transmission scheme with prediction error awareness is proposed to enhance the AoI performance, and a comprehensive performance analysis is presented correspondingly. A closed-form expression for the average AoI with respect to the prediction error probability is derived in the multi-feature source scenario with feature correlation. It is proved that the proposed PredComp scheme leads to a significant average AoI performance gain over the no prediction case, even under high prediction error probability. Simulation results verify the correctness of the analytical results, and show that the proposed PredComp scheme outperforms the status-prediction related work with no transmission termination [16], [18], and that there is also an optimal prediction horizon with regard to the average AoI.
- To further enhance the AoI performance, a predict-aggregate-compare (PredAggComp) transmission scheme is proposed by utilizing the time correlation of status updates. It allows a good tradeoff between the prediction error probability and the transmission error probability with the knowledge of the differential level of the predicted status updates. It is proved via a closed-form expression that the average AoI is monotonically

increasing with a higher differential level (related to the transmission error probability caused by the updated prediction and aggregation), with a given blocklength. A tight approximation of the differential level threshold is derived in closed form, below which PredAggComp is preferable over PredComp. The PredAggComp scheme demonstrates a 64% reduction in the average AoI over the no prediction case [9], [10], [11], [12], [13], [14] as well as a nearly doubled equivalent update rate at destination.

- To minimize the closed-form average AoI of the proposed PredComp and PredAggComp transmission schemes with respect to the prediction horizon, the prediction horizon is adapted. In the special case of a single-feature source, a tight approximation of the optimal prediction horizon of the PredComp scheme is derived in closed-form, and a joint prediction horizon and ratio optimization (JPHRO) algorithm is proposed for the PredAggComp scheme. In the multi-feature source scenario, the golden section search based to adapt the prediction horizon of the two proposed transmission schemes. The optimization results are shown to be near optimal by simulation, with a much lower complexity than exhaustive search.

The rest of the paper is organized as follows. Section II presents the system model. Section III presents the prediction error-aware PredComp transmission scheme and its average AoI performance analysis against the prediction error probability. Section IV presents the PredAggComp transmission scheme and the analysis of the preference region of prediction-aggregation in terms of the differential level. Section V is dedicated to adaptation of the prediction horizon for the two proposed transmission schemes based on their AoI performance analysis in Sections III and IV. Simulation results are given in Section VI, followed by conclusion in Section VII.

*Notations*: Throughout the paper, matrices and vectors are represented by boldface uppercase letters and boldface lowercase letters, respectively. $\mathbf{A} = [a_{i,j}]_{F\times F}$ builds an $F \times F$ metric, with $a_{i,j}$ representing the ($i$-th, $j$-th) element of matrix $\mathbf{A}$. $a_{i,j,(p)}$ denote the ($i$-th, $j$-th) element of metric $\mathbf{A}^p$. $|\cdot|$, $(\cdot)^{-1}$ and $[\cdot]^T$ denote magnitude, inverse and transpose operators, respectively. $\mathbb{E}[\cdot]$ and $\det(\cdot)$ refer to expectation and determinant, respectively. $(\cdot)^{\mathrm{PC}}$ and $(\cdot)^{\mathrm{PAC}}$ indicate the proposed PredComp scheme and PredAggComp scheme, respectively. $x \sim \mathcal{N}\left(0, \sigma^2\right)$ denotes that $x$ is a Gaussian random variable with zero mean and variance of $\sigma^2$. $\mathbf{x} \sim \mathcal{N}_F\left(\mathbf{0}, \mathbf{\Sigma}\right)$ denotes that $\mathbf{x}$ is an $F$-variate Gaussian random vector with mean vector $\mathbf{0}$ and covariance metric $\mathbf{\Sigma}$. $\Phi(u) = \int_{-\infty}^{u} (1/\sqrt{2\pi}) e^{-z^2/2} dz$ denotes the cumulative distribution function (CDF) of random variable $x \sim \mathcal{N}(0,1)$ evaluated at $u$. $\Phi_2(u_1, u_2; \rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \int_{-\infty}^{u_1}\int_{-\infty}^{u_2} e^{-\psi(\rho)} dz_2 dz_1$ denotes the joint CDF of random variables $x_1 \sim \mathcal{N}\left(0, \sigma_1^2\right)$ and $x_2 \sim \mathcal{N}\left(0, \sigma_2^2\right)$ evaluated at $\mathbf{u} = [u_1, u_2]^T$, with $\rho$ denoting the correlation coefficient of $x_1$ and $x_2$, and $\psi(\rho) = \frac{1}{2(1-\rho^2)}\left(\frac{z_1^2}{\sigma_1^2} - 2\rho\frac{z_1 z_2}{\sigma_1\sigma_2} + \frac{z_2^2}{\sigma_2^2}\right)$.

## II. System Model

We consider a point-to-point time-critical communication system of process tracking, where a single-antenna source predicts[1] the status updates of an underlying physical process and transmits them to a single-antenna destination in terms of short packets. The continues status updates of the physical process that includes a number of $F$ features are time-correlated. For example, the status updates of a sensor that captures the location and velocity of a robot arm in the repetitive pick-and-place process [6]. The destination which can be a monitor collects the latest packets to real-time monitor the status of the considered device. Assume that there are many time slots, each corresponding to a symbol duration of $T_{\mathrm{s}}$. The generated predicted update packet with $L$ bits of information is transmitted in terms of time slot. We consider a generate-at-will policy [28], [29], [30], [31] with deterministic service time, where a predicted update packet is generated at source and transmitted immediately with no queueing, once the last transmission is complete. This policy has been extensively adopted in the AoI related work [28], [29], [30], [31] and shown to fit continuous observation and monitoring of physical processes of IIoT [2], [28]. The blocklength is assumed to be fixed [9], [10], [14] to $N$ symbols, *i.e.*, channel uses (c.u.), for fair comparison between the proposed PredComp and PredAggComp schemes. Hence, the packet transmission time (*i.e.*, service time) without proactive transmission termination is fixed as $T = NT_{\mathrm{s}}$. Assume that there are a total of $K$ data transmissions between source and destination in the time interval of $(0, \tau]$. Each data transmission without proactive transmission termination is divided into $N$ time slots. Let $\mathrm{s}_k$ denote the first time slot of the $k$-th ($k = 1, 2, \ldots, K$) data transmission. Let $\mathrm{t}_k$ denote the time instant of the $k$-th data transmission start. Their relationship is given as $\mathrm{t}_k = (\mathrm{s}_k - 1)T_{\mathrm{s}}$. Let $x_f(\mathrm{s}_k)$ denote the $f$-th ($f = 1, 2, \ldots, F$) feature of status update of the physical process at time slot $\mathrm{s}_k$. Define $\mathbf{x}(\mathrm{s}_k) = [x_1(\mathrm{s}_k), x_2(\mathrm{s}_k), \ldots, x_F(\mathrm{s}_k)]^T$ as the update vector.

### A. State Transition Model

We consider a general linear state transition model, which is widely adopted in kinematics systems and control systems [16], [18], [19], [32]. The status update of the physical process at time slot $(\mathrm{s}_k + 1)$ is written as

$$\mathbf{x}(\mathrm{s}_k + 1) = \mathbf{A}\mathbf{x}(\mathrm{s}_k) + \mathbf{w}(\mathrm{s}_k), \tag{1}$$

where $\mathbf{A} = [a_{f,m}]_{F\times F}$ ($m = 1, 2, \ldots, F$) and $\mathbf{w}(\mathrm{s}_k) = [w_f(\mathrm{s}_k)]_{F\times 1}$ are the state transition matrix and transition noise vector, respectively. Assume that the elements of $\mathbf{A}$ are contant and the elements of $\mathbf{w}(\mathrm{s}_k)$ are independent identically distributed (i.i.d.) zero-mean Gaussian random variables [16], [18], [19] with variances $\sigma_{\mathrm{w},1}^2, \sigma_{\mathrm{w},2}^2, \ldots, \sigma_{\mathrm{w},F}^2$, respectively.

### B. Short Packet Transmission

The channel between source and destination is modeled to be quasi-static and block-fading Rayleigh, where the channel gain keeps constant during the transmission of a packet [33]. Let $g = \sqrt{\beta}h$ denote the channel gain, where $\beta$ represents

[1]Note that predicting the updates at source is suitable for time-critical IIoT applications of process tracking, as it requires no communication cost and enables proactive termination of a status update transmission in case of any prediction error. While prediction at destination is not aware of the real data and the prediction error in real time.

the large-scale fading channel gain that includes pathloss and shadowing, and $h$ represents the small-scale Rayleigh fading with zero mean and unit variance. Let $P_{\mathrm{t}}$ and $\sigma_{\mathrm{n}}^2$ denote the transmission power and the average AWGN power at destination, respectively. Then, the instantaneous signal to noise ratio (SNR) at destination is given by $\gamma = \beta|h|^2 P_{\mathrm{t}}/\sigma_{\mathrm{n}}^2$, and the average SNR is $\bar{\gamma} = \beta P_{\mathrm{t}}/\sigma_{\mathrm{n}}^2$. The achievable rate in SPT can be accurately approximated by [7]

$$R = \frac{L}{N} \approx \ln(1+\gamma) - \sqrt{\frac{V(\gamma)}{N}} Q^{-1}(\varepsilon_{\mathrm{te}}), \tag{2}$$

where $V(\gamma) = 1 - (1+\gamma)^{-2}$ is the channel dispersion and $Q(\varepsilon_{\mathrm{te}}) = \int_{\varepsilon_{\mathrm{te}}}^{\infty} (1/\sqrt{2\pi}) e^{-z^2/2} dz$ is the $Q$-function, with $\varepsilon_{\mathrm{te}}$ denoting the instantaneous transmission error probability. The average transmission error probability for Rayleigh fading channel can be expressed as [9]

$$\bar{\varepsilon}_{\mathrm{te}} = \int_0^{\infty} \frac{1}{\bar{\gamma}} e^{-\frac{z}{\bar{\gamma}}} Q\left( \frac{\sqrt{N}\left(\ln(1+z) - L/N\right)}{\sqrt{1-(1+z)^{-2}}} \right) dz. \tag{3}$$

Approximating $Q$ function in (3) yields a linear function [34]

$$\begin{cases} 1, & \text{if } z < \eta + \dfrac{1}{2\lambda}, \\ \lambda(\gamma - \eta) + 1/2, & \text{if } \eta + \dfrac{1}{2\lambda} \le z \le \eta - \dfrac{1}{2\lambda}, \\ 0, & \text{if } z > \eta - \dfrac{1}{2\lambda}, \end{cases}$$

with $\lambda = -\sqrt{N(2\pi(e^{2L/N}-1))^{-1}}$ and $\eta = e^{L/N} - 1$. The average transmission error probability can be tightly approximated by [9]

$$\bar{\varepsilon}_{\mathrm{te}} \approx 1 + \bar{\gamma}\lambda e^{-\frac{1}{\bar{\gamma}}(\eta + \frac{1}{2\lambda})} - \bar{\gamma}\lambda e^{-\frac{1}{\bar{\gamma}}(\eta - \frac{1}{2\lambda})}. \tag{4}$$

The transmission and prediction errors result in the lost of the predicted update packet. With no retransmission, the total error probability is given by

$$\varepsilon = 1 - (1 - \varepsilon_{\mathrm{pe}})(1 - \bar{\varepsilon}_{\mathrm{te}}), \tag{5}$$

where $\varepsilon_{\mathrm{pe}}$ is the prediction error probability.

### C. Performance Metric

The instantaneous AoI at time instant $t$ ($t \in (0, \tau]$) is defined as $\Delta(t) = t - u(t)$, with $u(t)$ denoting the generation time of the last successfully received update at destination [3]. With the ergodicity of the process $\Delta(t)$, the average AoI during time interval of (0, $\tau$] is defined as

$$\bar{\Delta} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_0^{\tau} \Delta(t) dt. \tag{6}$$

The average AoI reflects the effect of status errors occurring in the continuous process of transmission, as shown by Fig. 2.

## III. Analysis of Prediction Error Probability and Average AoI of the Proposed PredComp Transmission Scheme

In this section, we first describe the proposed PredComp transmission scheme, and derive the closed-form expression of prediction error probability with respect to prediction horizon in multi-feature source scenario. Based on the results obtained, we further derive the closed-form expression of average AoI, and analyze the tolerance of average AoI performance against prediction error probability.

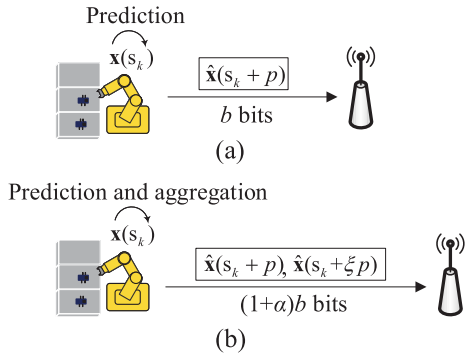



Fig. 1. Illustration of (a) the proposed PredComp transmission scheme and (b) the proposed PredAggComp transmission scheme in the $k$-th transmission at time slot $\mathrm{s}_k$.

### A. Proposed PredComp Transmission Scheme

We propose a PredComp transmission scheme to reduce AoI for SPT in IIoT. The PredComp scheme allows the source to not only collect an update, but also predict a late update in future. The predicted update is transmitted immediately at an early time, enabling the reduction of AoI between information generation and reception, severer than the state-of-the-art methods [9], [10], [11], [12], [13], [14] without prediction. In order to reduce the impact of prediction error on AoI, we also propose a proactive transmission termination strategy, if a prediction error is detected at source before the packet is received at destination.

The communication process of the proposed PredComp scheme in the $k$-th transmission is as follows. Assume that the prediction horizon is shorter than or equal to the transmission time of a predicted update packet. Let $p$ denote the prediction horizon, with $0 < p \le N$. Define $\hat{\mathbf{x}}(\mathrm{s}_k + p) = [\hat{x}_1(\mathrm{s}_k + p), \hat{x}_2(\mathrm{s}_k + p), \ldots, \hat{x}_F(\mathrm{s}_k + p)]^T$ as the predicted update vector. As shown in Fig. 1 (a), at time slot $\mathrm{s}_k$, immediately after sampling the update of $\mathbf{x}(\mathrm{s}_k)$, the source predicts the update of $\hat{\mathbf{x}}(\mathrm{s}_k + p)$ and transmits it to the destination. At time slot $(\mathrm{s}_k + p)$, the source samples the real update of $\mathbf{x}(\mathrm{s}_k + p)$ and determines whether or not to continue transmission of $\hat{\mathbf{x}}(\mathrm{s}_k + p)$. We introduce a prediction error vector

$$\begin{aligned} \mathbf{e}(\mathrm{s}_k + p) &= \mathbf{x}(\mathrm{s}_k + p) - \hat{\mathbf{x}}(\mathrm{s}_k + p) \\ &= [e_1(\mathrm{s}_k + p), e_2(\mathrm{s}_k + p), \ldots, e_F(\mathrm{s}_k + p)]^T, \end{aligned} \tag{7}$$

with $e_f(\mathrm{s}_k + p)$ denoting the element of the prediction error on the $f$-th feature. If $|e_f(\mathrm{s}_k + p)| \le \delta_f$ holds, where $\delta_f$ is the acceptable threshold of prediction error of the $f$-th feature, prediction is successful. The source continues to transmit $\hat{\mathbf{x}}(\mathrm{s}_k + p)$ until it is received by destination at time slot of $(\mathrm{s}_k + N)$. However, if $|e_f(\mathrm{s}_k + p)| > \delta_f$, a prediction error occurs. The source terminates the transmission of $\hat{\mathbf{x}}(\mathrm{s}_k + p)$ and immediately starts to predict $\hat{\mathbf{x}}(\mathrm{s}_k + 2p)$ based on the sampled real update of $\mathbf{x}(\mathrm{s}_k + p)$ and transmit $\hat{\mathbf{x}}(\mathrm{s}_k + 2p)$ to destination. Compared to the intuitive approach of transmitting the current $\mathbf{x}(\mathrm{s}_k + p)$, the proposed scheme transmitting the predicted $\hat{\mathbf{x}}(\mathrm{s}_k + 2p)$ reduces the AoI between information generation and reception. The real update of $\mathbf{x}(\mathrm{s}_k + p)$ is discarded by source at the end of time slot $(\mathrm{s}_k + p)$. The prediction error probability of the proposed PredComp scheme with prediction horizon of $p$ is given by

$$\varepsilon_{\mathrm{pe}}^{\mathrm{PC}} = 1 - \varepsilon_{\mathrm{ps}}(p), \tag{8}$$

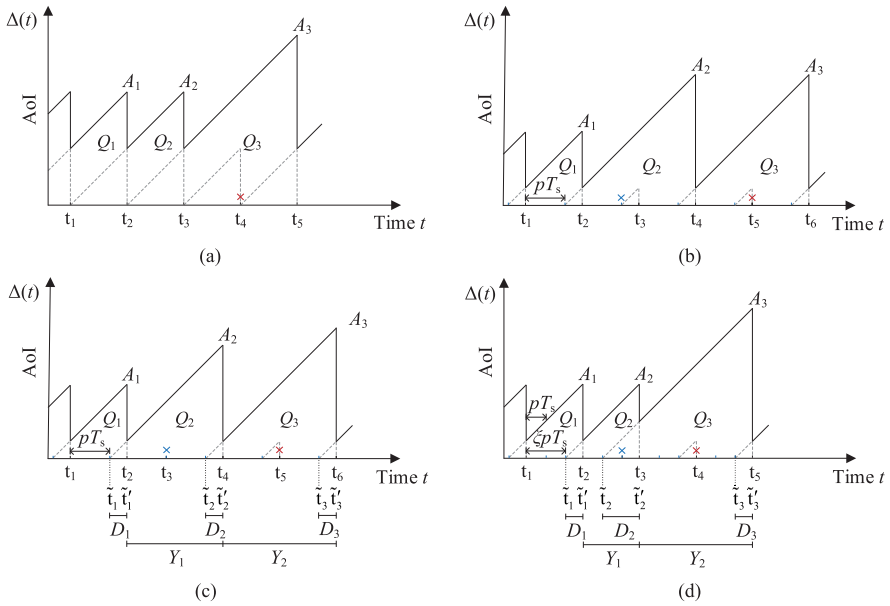



Fig. 2. Examples of $\Delta(t)$ evolution for (a) no prediction, (b) prediction without proactive transmission termination, (c) proposed PredComp scheme and (d) proposed PredAggComp scheme. At time instant $\mathrm{t}_k$, the source in (a) transmits the current update to destination, while the source in (b) and (c) transmits a predicted update with prediction horizon of $pT_\mathrm{s}$. The source in (d) transmits two predicted updates using data aggregation with prediction horizons of $pT_\mathrm{s}$ and $\xi pT_\mathrm{s}$. The transmission of the second update packet in (c) is proactively terminated at time instant $\mathrm{t}_3$ due to a prediction error is detected at source. The update packets that error in prediction and error in transmission lead to unsuccessful receptions at destination, as illustrated for the second and the fourth packets in (b-d), respectively. With a given blocklength, the packet transmission time of (b-d) is the same.

where $\varepsilon_{\mathrm{ps}}(p) = \Pr\left\{\bigcap_{f=1}^{F} |e_f(\mathrm{s}_k+p)| \le \delta_f\right\}$ denotes the prediction success probability for an $F$-feature status update with prediction horizon of $p$. Note that the prediction error affects reliability, which may impair the AoI performance.

### B. Prediction Error Probability

According to (1), the status update of the physical process at time slot of $(\mathrm{s}_k+p)$ is

$$\mathbf{x}(\mathrm{s}_k+p) = \mathbf{A}^p\mathbf{x}(\mathrm{s}_k) + \sum_{i=1}^{p}\mathbf{A}^{p-i}\mathbf{w}(\mathrm{s}_k+i-1). \quad (9)$$

As in [16] and [18], we consider a general linear prediction method. We predict the update at time slot of $(\mathrm{s}_k+p)$ from the update at time slot of $\mathrm{s}_k$ as

$$\hat{\mathbf{x}}(\mathrm{s}_k+p) = \mathbf{A}^p\mathbf{x}(\mathrm{s}_k). \quad (10)$$

Substituting (9) and (10) into (7) yields

$$\mathbf{e}(\mathrm{s}_k+p) = \sum_{i=1}^{p}\mathbf{A}^{p-i}\mathbf{w}(\mathrm{s}_k+i-1), \quad (11)$$

with the element of the prediction error on the $f$-th feature expressing as

$$e_f(\mathrm{s}_k+p) = \sum_{i=1}^{p}\sum_{j=1}^{F} a_{f,j,(p-i)}w_j(\mathrm{s}_k+i-1), \quad (12)$$

Since $e_f(\mathrm{s}_k+p)$ is a linear combination of state transition noises, we have $e_f(\mathrm{s}_k+p) \sim \mathcal{N}\left(0, \sigma^2_{\mathrm{e},f}(p)\right)$, with variance being

$$\sigma^2_{\mathrm{e},f}(p) = \sum_{i=1}^{p}\sum_{j=1}^{F}\left(a_{f,j,(p-i)}\right)^2\sigma^2_{\mathrm{w},j}. \quad (13)$$

As a consequence, we have $\mathbf{e}(\mathrm{s}_k+p) \sim \mathcal{N}_F(\mathbf{0}, \mathbf{\Sigma}(p))$, with covariance matrix being

$$\mathbf{\Sigma}(p) = \begin{bmatrix} \sigma_{\mathrm{e},11}(p) & \cdots & \sigma_{\mathrm{e},1F}(p) \\ \vdots & \ddots & \vdots \\ \sigma_{\mathrm{e},F1}(p) & \cdots & \sigma_{\mathrm{e},FF}(p) \end{bmatrix}, \quad (14)$$

where $\sigma_{\mathrm{e},fm}(p)$ is the covariance between $e_f(\mathrm{s}_k+p)$ and $e_m(\mathrm{s}_k+p)$, expressed as

$$\begin{aligned}\sigma_{\mathrm{e},fm}(p) &= \mathbb{E}\left[e_f(\mathrm{s}_k+p)e_m(\mathrm{s}_k+p)\right] \\ &= \sum_{i=1}^{p}\sum_{j=1}^{F} a_{f,j,(p-i)}a_{m,j,(p-i)}\sigma^2_{\mathrm{w},j}, \end{aligned} \quad (15)$$

with $\sigma_{\mathrm{e},ff}(p) \equiv \sigma^2_{\mathrm{e},f}(p)$. Therefore, the prediction success probability for an $F$-feature status update with prediction horizon of $p$ can be expressed as

$$\begin{aligned}&\varepsilon_{\mathrm{ps}}(p) \\ &= \frac{1}{\sqrt{(2\pi)^F \det\left(\mathbf{\Sigma}(p)\right)}}\int_{-\delta_1}^{\delta_1}\cdots\int_{-\delta_F}^{\delta_F} e^{-\frac{1}{2}\mathbf{z}_F^T\mathbf{\Sigma}^{-1}(p)\mathbf{z}_F}d\mathbf{z}_F, \end{aligned} \quad (16)$$

with $\mathbf{z}_F = [z_1, \cdots, z_F]^T$ and $d\mathbf{z}_F := dz_F\cdots dz_1$. There is feature correlation for general multi-feature source in IIoT, which however does not been considered in the state-of-the-art methods in [16] and [18]. When the number of features $F = 1$ and 2, the prediction success probability in (16) can be accurately expressed as (17a) and (17b), shown at the bottom of the next page, respectively, where $\Phi_2(\cdot,\cdot;\rho_{12}(p))$ is the joint CDF of $e_1(\mathrm{s}_k+p)$ and $e_2(\mathrm{s}_k+p)$, with $\rho_{12}(p) = \sigma_{\mathrm{e},12}(p)/(\sigma_{\mathrm{e},1}(p)\sigma_{\mathrm{e},2}(p))$ denoting their correlation coefficient. When $F > 2$, in order to solve the multivariate Gaussian integral in (16), the randomized quasi-Monte Carlo (RQMC) method [35, Ch. 4] is adopted. The prediction success proba-

bility in (16) is first decorrelated by applying a decorrelation transformation with respect to $\mathbf{z}_F$, and then converted into an integration over a unit hyper-cube by applying a series of variable transformations. We introduce a vector $\mathbf{o} = \mathbf{C}^{-1}\mathbf{z}_F$, where $\mathbf{C} = [c_{f,m}]_{F\times F}$ is the lower triangular Cholesky metric of $\mathbf{\Sigma}(p)$, with $\mathbf{\Sigma}(p) = \mathbf{C}\mathbf{C}^T$, and $\mathbf{o} = [o_1, \cdots, o_F]^T$. Using RQMC method, the prediction success probability $\varepsilon_{\text{ps}}(p)$ can be approximately expressed as (17c), shown at the bottom of the next page, where $f(\cdot) = \prod_{f=1}^{F}\Phi((\delta_f - \sum_{j=1}^{f-1} c_{f,j}o_f)/c_{f,f}) - \Phi((-\delta_f - \sum_{j=1}^{f-1} c_{f,j}o_f)/c_{f,f})$, $\Lambda$ is the number of random shifts, $\mathbf{u}_i$ consists of $F$ i.i.d. random variables that are uniformly distributed between 0 and 1, $\{\mathbf{x}\}$ denotes the remainder of each elements of the vector $\mathbf{x}$ modulo 1, and $\mathbf{l}_q = \{q\sqrt{\mathbf{p}}\}$, with $\mathbf{p} = (2, 3, 5, \ldots)$ consisting of the first $F$ prime numbers. It is shown in [35, Ch. 4] that the RQMC method has a convergence rate of $O(1/U)$, which is better than the convergence rate of $O(1/\sqrt{U})$ by the conventional Monte Carlo method. Substituting (17) into (8), the prediction error probability $\varepsilon_{\text{pe}}^{\text{PC}}$ of the proposed PredComp scheme can be obtained.

### C. Closed-Form Average AoI

We derive the closed-form expression for average AoI of the proposed PredComp scheme. Assume that there are a total of $V$ updates successfully received by destination in the time interval of $(0, \tau]$. We introduce a prediction horizon $p$, as shown in Fig. 2 (c). The delay of the $v$-th ($v = 1, 2, \ldots, V$) successfully received update is the time duration between generation time instant $\tilde{\text{t}}_v$ at source and reception time instant $\tilde{\text{t}}'_v$ at destination as $D_v = \tilde{\text{t}}'_v - \tilde{\text{t}}_v = T - pT_{\text{s}}$. The time interval between two consecutive successful receptions can be expressed as $Y_v = \tilde{\text{t}}'_{v+1} - \tilde{\text{t}}'_v = M_1 pT_{\text{s}} + M_2 T + T$, where $M_1$ and $M_2$ ($M_1, M_2 = 0, 1, 2, \cdots$) are the numbers of failed receptions at destination caused by prediction error and transmission error, respectively, in the time interval of $Y_v$. The integral $\int_0^\tau \Delta(t)dt$ in (6) can be interpreted as the area covered by $\Delta(t)$. From Figs. 2 (c) and 2 (d), that area can be calculated as the sum of the trapezoid part $Q_v$ ($v = 1, 2, \ldots, V$), with $Q_v = \frac{1}{2}Y_{v-1}(2D_{v-1} + Y_{v-1})$ for $v > 1$. With $Q_1$ being negligible as $\tau \to \infty$ and $Y_v$ and $D_v$ being independent of each other, the average AoI is calculated as

$$\bar{\Delta} = \lim_{\tau\to\infty} \frac{\sum_{v=1}^{V} Q_v}{\tau} = \lim_{V\to\infty} \frac{\sum_{v=2}^{V} Q_v}{\sum_{v=1}^{V-1} Y_v} = \lim_{V\to\infty} \frac{\sum_{v=2}^{V} \frac{1}{2}Y_{v-1}(2D_{v-1} + Y_{v-1})}{\sum_{v=1}^{V-1} Y_v} = \mathrm{E}[D_v] + \frac{\mathrm{E}[Y_v^2]}{2\mathrm{E}[Y_v]}. \tag{18}$$

With $\Pr\{Y_v = M_1 pT_{\text{s}} + M_2 T + T\} = \binom{M_1+M_2}{M_1}(\varepsilon_{\text{pe}}^{\text{PC}})^{M_1}((1-\varepsilon_{\text{pe}}^{\text{PC}})\bar{\varepsilon}_{\text{te}}^{\text{PC}})^{M_2}(1-\varepsilon^{\text{PC}})$, the expectations of $Y_v$ and $Y_v^2$ are respectively calculated in a recursive manner as

$$\begin{aligned}\mathbb{E}[Y_v] &= \sum_{M_1=0}^{\infty}\sum_{M_2=0}^{\infty}(M_1 pT_{\text{s}} + M_2 T + T)\tbinom{M_1+M_2}{M_1}(\varepsilon_{\text{pe}}^{\text{PC}})^{M_1} \\ &\quad\times((1-\varepsilon_{\text{pe}}^{\text{PC}})\bar{\varepsilon}_{\text{te}}^{\text{PC}})^{M_2}(1-\varepsilon^{\text{PC}}) \\ &= T + \frac{(1-\varepsilon_{\text{pe}}^{\text{PC}})\bar{\varepsilon}_{\text{te}}^{\text{PC}}T + \varepsilon_{\text{pe}}^{\text{PC}}pT_{\text{s}}}{1-\varepsilon^{\text{PC}}},\end{aligned} \tag{19}$$

and

$$\begin{aligned}\mathbb{E}[Y_v^2] &= \sum_{M_1=0}^{\infty}\sum_{M_2=0}^{\infty}(M_1 pT_{\text{s}} + M_2 T + T)^2\tbinom{M_1+M_2}{M_1}(\varepsilon_{\text{pe}}^{\text{PC}})^{M_1} \\ &\quad\times((1-\varepsilon_{\text{pe}}^{\text{PC}})\bar{\varepsilon}_{\text{te}}^{\text{PC}})^{M_2}(1-\varepsilon^{\text{PC}}) \\ &= \left(T + \frac{(1-\varepsilon_{\text{pe}}^{\text{PC}})\bar{\varepsilon}_{\text{te}}^{\text{PC}}T + \varepsilon_{\text{pe}}^{\text{PC}}pT_{\text{s}}}{1-\varepsilon^{\text{PC}}}\right)^2 \\ &\quad + \frac{1}{\varepsilon^{\text{PC}}}\left(\frac{(1-\varepsilon_{\text{pe}}^{\text{PC}})\bar{\varepsilon}_{\text{te}}^{\text{PC}}T + \varepsilon_{\text{pe}}^{\text{PC}}pT_{\text{s}}}{1-\varepsilon^{\text{PC}}}\right)^2.\end{aligned} \tag{20}$$

Substituting (19), (20) and $\mathbb{E}[D_v = T - pT_{\text{s}}]$ into (18), the average AoI of the proposed PredComp scheme with prediction horizon of $p$ is given by

$$\begin{aligned}\bar{\Delta}^{\text{PC}}(p) &= \frac{3}{2}T - pT_{\text{s}} + \frac{(1-\varepsilon_{\text{pe}}^{\text{PC}})\bar{\varepsilon}_{\text{te}}^{\text{PC}}T + \varepsilon_{\text{pe}}^{\text{PC}}pT_{\text{s}}}{2(1-\varepsilon^{\text{PC}})} \\ &\quad\times\left(1 + \frac{(1-\varepsilon_{\text{pe}}^{\text{PC}})\bar{\varepsilon}_{\text{te}}^{\text{PC}}T + \varepsilon_{\text{pe}}^{\text{PC}}pT_{\text{s}}}{\varepsilon^{\text{PC}}((1-\varepsilon_{\text{pe}}^{\text{PC}})T + \varepsilon_{\text{pe}}^{\text{PC}}pT_{\text{s}})}\right).\end{aligned} \tag{21}$$

With $pT_{\text{s}} \le T$, the term $\frac{(1-\varepsilon_{\text{pe}}^{\text{PC}})\bar{\varepsilon}_{\text{te}}^{\text{PC}}T + \varepsilon_{\text{pe}}^{\text{PC}}pT_{\text{s}}}{\varepsilon^{\text{PC}}((1-\varepsilon_{\text{pe}}^{\text{PC}})T + \varepsilon_{\text{pe}}^{\text{PC}}pT_{\text{s}})}$ is smaller or equal to 1, yielding a tight upper bound of $\bar{\Delta}^{\text{PC}}(p)$ as

$$\bar{\Delta}_{\text{ub}}^{\text{PC}}(p) = \frac{3}{2}T - pT_{\text{s}} + \frac{(1-\varepsilon_{\text{pe}}^{\text{PC}})\bar{\varepsilon}_{\text{te}}^{\text{PC}}T + \varepsilon_{\text{pe}}^{\text{PC}}pT_{\text{s}}}{1-\varepsilon^{\text{PC}}}. \tag{22}$$

The tightness is verified by Fig. 3 in the simulation results. For comparison, by setting $p = 0$ in (21), the average AoI with no prediction is given by

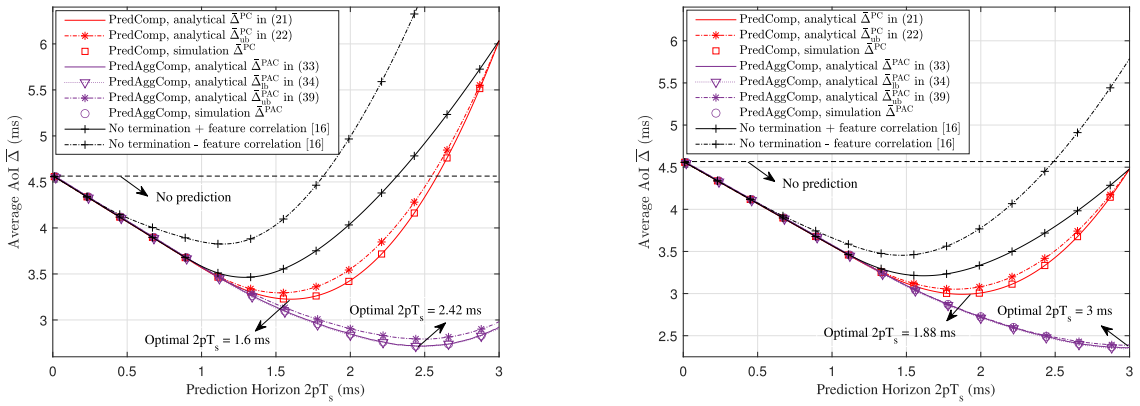
$$\bar{\Delta}^{\text{NP}} = \frac{3}{2}T + \frac{\bar{\varepsilon}_{\text{te}}^{\text{PC}}T}{1-\bar{\varepsilon}_{\text{te}}^{\text{PC}}}. \tag{23}$$

From (22) and (23), compared to without prediction, the proposed PredComp scheme achieves the reduction of average AoI as $\bar{\Delta}^{\text{PC}}(p) - \bar{\Delta}^{\text{NP}} \le \bar{\Delta}_{\text{ub}}^{\text{PC}}(p) - \bar{\Delta}^{\text{NP}} = \frac{(2-\bar{\varepsilon}_{\text{te}}^{\text{PC}})pT_{\text{s}}}{1-\varepsilon^{\text{PC}}}(\varepsilon_{\text{pe}}^{\text{PC}} - \frac{1-\bar{\varepsilon}_{\text{te}}^{\text{PC}}}{2-\bar{\varepsilon}_{\text{te}}^{\text{PC}}})$, which is smaller than zero with $\varepsilon_{\text{pe}}^{\text{PC}} < \frac{1-\bar{\varepsilon}_{\text{te}}^{\text{PC}}}{2-\bar{\varepsilon}_{\text{te}}^{\text{PC}}}$. Based on the result, we derive **Proposition 1** as follows.

*Proposition 1:* Given transmission error probability $\bar{\varepsilon}_{\text{te}}^{\text{PC}}$, the average AoI performance can benefit from status prediction as long as the prediction error probability $\varepsilon_{\text{pe}}^{\text{PC}} < \frac{1-\bar{\varepsilon}_{\text{te}}^{\text{PC}}}{2-\bar{\varepsilon}_{\text{te}}^{\text{PC}}}$.

*Proposition 1* implies that it is easy to obtain the average AoI performance gain from status prediction. This is also

$$\varepsilon_{\text{ps}}(p) = \begin{cases} 1 - 2Q\left(\delta/\sigma_{\text{e}}(p)\right), & \text{if F=1,} \quad (17a) \\ \Phi_2\left(\delta_1, \delta_2; \rho_{12}(p)\right) + \Phi_2\left(-\delta_1, -\delta_2; \rho_{12}(p)\right) - \Phi_2\left(-\delta_1, \delta_2; \rho_{12}(p)\right) - \Phi_2\left(\delta_1, -\delta_2; \rho_{12}(p)\right), & \text{if F=2,} \quad (17b) \\ \dfrac{1}{2\Lambda U}\displaystyle\sum_{i=1}^{\Lambda}\sum_{q=1}^{U}(f(|2\{\mathbf{l}_q + \mathbf{u}_i\} - 1|) + f(1 - |2\{\mathbf{l}_q + \mathbf{u}_i\} - 1|)), & \text{elsewhere} \quad (17c) \end{cases}$$




(a) Prediction error threshold $\delta = 2$ (b) Prediction error threshold $\delta = 2.5$

Fig. 3. Impact of prediction horizon on average AoI of the proposed PredComp and PredAggComp transmission schemes with a two-feature source. The longest prediction horizon 3 ms is equal to the packet transmission time $T$ without proactive transmission termination.

applicable to the case of high prediction error probability. For example, when $\bar{\varepsilon}^{\text{PC}}_{\text{te}} = 10^{-5}$, the maximum allowable prediction error probability is approximately 0.5. This conclusion can be intuitively explained by comparing the time-average AoI in the cases of no prediction, a failed prediction without proactive transmission termination, and a failed prediction with proactive transmission termination. The time-average AoI of the three cases can be calculated as the ratio of the area $Q_2$ to the time interval $Y_1$ shown in Figs. 2 (a)-(c), respectively. With $\bar{\varepsilon}^{\text{PC}}_{\text{te}} = 10^{-5}$, we have the ratio for different cases is:

- No prediction: $\frac{Q_2}{Y_1} = \frac{T(T+2T)/2}{T} = 1.5T$;
- A failed prediction without proactive transmission termination: $\frac{Q_2}{Y_1} = \frac{2T(2T+2(T-pT_{\text{s}}))/2}{2T} = 2T - pT_{\text{s}}$;
- A failed prediction with proactive transmission termination: $\frac{Q_2}{Y_1} = \frac{(T+pT_{\text{s}})(T-pT_{\text{s}}+2T)/2}{T+pT_{\text{s}}} = 1.5T - 0.5pT_{\text{s}}$.

The prediction error causes high error probability, which is supposed to increase the average AoI. However, we derive a surprising result that the high prediction error probability still allows the average AoI lower than the no prediction case, as long as the prediction horizon of $pT_{\text{s}}$ is longer than $0.5T$. The prediction significantly reduces the delay and the AoI between information generation and reception, showing a significant tolerance of prediction error probability. Furthermore, the action of proactive transmission termination reduces the impact of prediction error on AoI, which can reduce the average AoI.

## IV. Analysis of Average AoI and Preference Region of the Proposed PredAggComp Transmission Scheme

To further enhance the AoI performance, we propose a PredAggComp transmission scheme by utilizing the time correlation of status updates, where data aggregation is used to assist status prediction. We first derive the closed-form expression of prediction error probability of the proposed scheme, and then derive the closed-form expression of average AoI. Also, the preference region of the PredAggComp scheme is analyzed in terms of the differential level.

### A. Proposed PredAggComp Transmission Scheme

The proposed PredAggComp transmission scheme allows the source to predict two updates with different prediction horizons. The predicted two updates are aggregated into a single packet and transmitted by source, enabling a lower prediction error probability than the existing prediction methods without data aggregation [16], [18]. Prediction and aggregation of more updates require more frequent sensing and hence more energy consumption at source. We notice by simulation that the AoI performance gain increases marginally with the increase of the number of aggregated updates, while the sensing energy consumption increases linearly, which is exceptional for battery powered sensors. Therefore, we focus on two-update aggregation in this paper.

The communication process of the proposed PredAggComp scheme in the $k$-th transmission is given as follows. Let $\xi$ denote the ratio of the prediction horizon of the two predicted updates. For consistency, we denote the prediction horizons as $p$ and $\xi p$ with $\xi p \leq N$. At time slot $\text{s}_k$, the source predicts the updates of $\hat{\mathbf{x}}(\text{s}_k + p)$ and $\hat{\mathbf{x}}(\text{s}_k + \xi p)$, and aggregates their information into a single packet to transmit to the destination, using differential coding [36]. Assume that each predicted update can be represented by $b$ bits. Due to the temporal correlation, it is further assumed that the difference between any two predicted updates can be represented by $\alpha b$ bits, with $0 < \alpha \leq 1$ denoting the differential level, to indicate the degree of difference. Hence, the number of transmitted information bits in the proposed PredAggComp scheme is $L^{\text{PAC}} = b + \alpha b$, more than that in the proposed PredComp scheme with $L^{\text{PC}} = b$. At time slots $(\text{s}_k + p)$ and $(\text{s}_k + \xi p)$, the source respectively samples the real updates of $\mathbf{x}(\text{s}_k + p)$ and $\mathbf{x}(\text{s}_k + \xi p)$. Also, the source determines whether or not to continue transmission of the current packet at time slot of $(\text{s}_k + \xi p)$. If any one of the two predicted updates successes in prediction, the source continues to transmit the packet until it is received by destination at time slot $(\text{s}_k + N)$. However, if both predicted updates fail in prediction, the source terminates the transmission of the packet and immediately starts to predict $\hat{\mathbf{x}}(\text{s}_k + \xi p + p)$ and $\hat{\mathbf{x}}(\text{s}_k + 2\xi p)$ based on the sampled real update of $\mathbf{x}(\text{s}_k + \xi p)$ and transmit them to destination. The prediction error probability of the proposed PredAggComp scheme can be expressed as

$$\varepsilon^{\text{PAC}}_{\text{pe}} = 1 - \varepsilon_{\text{ps}}(p) - \varepsilon_{\text{ps}}(\xi p) + \varepsilon_{\text{ps,sim}}(p, \xi p), \tag{24}$$

where $\varepsilon_{\text{ps,sim}}(p, \xi p) = \Pr\left\{\bigcap_{f=1}^{F} \left(|e_f(\text{s}_k + p)| \leq \delta_f, |e_f(\text{s}_k + \xi p)| \leq \delta_f\right)\right\}$ denotes the probability that $\hat{\mathbf{x}}(\text{s}_k + p)$ and $\hat{\mathbf{x}}(\text{s}_k + \xi p)$ simultaneously success in prediction. Note that

if both $\hat{\mathbf{x}}(\mathrm{s}_k+p)$ and $\hat{\mathbf{x}}(\mathrm{s}_k+\xi p)$ success in prediction, the destination reserves $\hat{\mathbf{x}}(\mathrm{s}_k+\xi p)$ as the latest reception, to reduce AoI.

### *B. Prediction Error Probability*

With the derived $\varepsilon_{\mathrm{ps}}(p)$ in (17), it is easy to obtain $\varepsilon_{\mathrm{ps}}(\xi p)$ replacing $p$ by $\xi p$. In the following, we focus on the derivation of $\varepsilon_{\mathrm{ps,sim}}(p,\xi p)$. Since $\mathbf{e}\left(\mathrm{s}_k+p\right) \sim \mathcal{N}_F\left(\mathbf{0},\mathbf{\Sigma}(p)\right)$ and $\mathbf{e}\left(\mathrm{s}_k+\xi p\right) \sim \mathcal{N}_F\left(\mathbf{0},\mathbf{\Sigma}(\xi p)\right)$, then we can obtain $\left[\mathbf{e}\left(\mathrm{s}_k+p\right),\mathbf{e}\left(\mathrm{s}_k+\xi p\right)\right]^T \sim \mathcal{N}_{2F}\left(\mathbf{0},\mathbf{\Sigma}_{\mathrm{agg}}(p,\xi p)\right)$, with covariance metric being

$$\mathbf{\Sigma}_{\mathrm{agg}}(p,\xi p)=\begin{bmatrix}\mathbf{\Sigma}(p) & \mathbf{\Sigma}_{\mathrm{p},\xi\mathrm{p}}(p,\xi p)\\ \mathbf{\Sigma}_{\xi\mathrm{p},\mathrm{p}}(p,\xi p) & \mathbf{\Sigma}(\xi p)\end{bmatrix}, \tag{25}$$

with $\mathbf{\Sigma}_{\mathrm{p},\xi\mathrm{p}}(p,\xi p)=\mathbf{\Sigma}^T_{\xi\mathrm{p},\mathrm{p}}(p,\xi p)$, expressed as

$$\mathbf{\Sigma}_{\mathrm{p},\xi\mathrm{p}}(p,\xi p)=\begin{bmatrix}\sigma_{\mathrm{e},\widetilde{1}1}(p,\xi p) & \cdots & \sigma_{\mathrm{e},\widetilde{1}F}(p,\xi p)\\ \vdots & \ddots & \vdots\\ \sigma_{\mathrm{e},\widetilde{F}1}(p,\xi p) & \cdots & \sigma_{\mathrm{e},\widetilde{F}F}(p,\xi p)\end{bmatrix}, \tag{26}$$

where $\sigma_{\mathrm{e},\widetilde{\ell u}}(p,\xi p)$ $(\ell,u=1,2,\ldots,F)$ is the covariance between $e_\ell\left(\mathrm{s}_k+p\right)$ and $e_u\left(\mathrm{s}_k+\xi p\right)$, expressed as

$$\begin{aligned}\sigma_{\mathrm{e},\widetilde{\ell u}}(p,\xi p)&=\mathbb{E}\left[e_\ell(\mathrm{s}_k+p)e_u(\mathrm{s}_k+\xi p)\right]\\&=\sum_{i=1}^{p}\sum_{j=1}^{F}a_{\ell,j,(p-i)}a_{u,j,(\xi p-i)}\sigma^2_{\mathrm{w},j}.\end{aligned} \tag{27}$$

Therefore, we have

$$\begin{aligned}&\varepsilon_{\mathrm{ps,sim}}(p,\xi p)\\&=\frac{1}{\sqrt{(2\pi)^{2F}\det\left(\mathbf{\Sigma}_{\mathrm{agg}}(p,\xi p)\right)}}\\&\times\int_{-\delta_1}^{\delta_1}\cdots\int_{-\delta_F}^{\delta_F}\int_{-\delta_1}^{\delta_1}\cdots\int_{-\delta_F}^{\delta_F}e^{-\frac{1}{2}\mathbf{z}^T_{2F}\mathbf{\Sigma}^{-1}_{\mathrm{agg}}(p,\xi p)\mathbf{z}_{2F}}d\mathbf{z}_{2F},\end{aligned} \tag{28}$$

with $\mathbf{z}_{2F}=[z_1,\cdots,z_{2F}]^T$ and $d\mathbf{z}_{2F}:=dz_{2F}\cdots dz_1$. When the number of features $F\geq 2$, the multivariate Gaussian integral in (28) can be solved by RQMC method [35, Ch.4]. For the special case of a single-feature source, $\varepsilon_{\mathrm{ps,sim}}(p,\xi p)$ can be accurately expressed as

$$\begin{aligned}&\varepsilon_{\mathrm{ps,sim}}(p,\xi p)\,|_{F=1}\\&=\Phi_2\left(\delta,\delta;\rho_{\widetilde{1}1}(p,\xi p)\right)-\Phi_2\left(\delta,-\delta;\rho_{\widetilde{1}1}(p,\xi p)\right)\\&+\Phi_2\left(-\delta,-\delta;\rho_{\widetilde{1}1}(p,\xi p)\right)-\Phi_2\left(-\delta,\delta;\rho_{\widetilde{1}1}(p,\xi p)\right),\end{aligned} \tag{29}$$

where $\Phi_2\left(\cdot,\cdot;\rho_{\widetilde{1}1}(p,\xi p)\right)$ is the joint CDF of $e(\mathrm{s}_k+p)$ and $e(\mathrm{s}_k+\xi p)$, with $\rho_{\widetilde{1}1}(p,\xi p)=\sigma_{\mathrm{e},\widetilde{1}1}(p,\xi p)/(\sigma_{\mathrm{e}}(p)\sigma_{\mathrm{e}}(\xi p))$ denoting their correlation coefficient. Substituting $\varepsilon_{\mathrm{ps}}(p)$, $\varepsilon_{\mathrm{ps}}(\xi p)$ and $\varepsilon_{\mathrm{ps,sim}}(p,\xi p)$ into (24), the prediction error probability $\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}}$ of the proposed PredAggComp scheme can be obtained.

### *C. Closed-Form Average AoI*

We derive the closed-form expression of average AoI of the proposed PredAggComp scheme, by calculating the terms $\mathbb{E}[D_v]$, $\mathbb{E}[Y_v]$ and $\mathbb{E}[Y_v^2]$. We introduce two prediction horizons of $p$ and $\xi p$, as shown in Fig. 2 (d). The delay of the $v$-th successfully received update $D_v\in\{T-pT_{\mathrm{s}},T-\xi pT_{\mathrm{s}}\}$. Since $\Pr\{D_v=T-\xi pT_{\mathrm{s}}\}=\frac{\varepsilon_{\mathrm{ps}}(\xi p)}{1-\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}}}$ and $\Pr\{D_v=T-pT_{\mathrm{s}}\}=1-\frac{\varepsilon_{\mathrm{ps}}(\xi p)}{1-\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}}}$, the expectation of $D_v$ is given by

$$\begin{aligned}\mathbb{E}[D_v]&=(T-\xi pT_{\mathrm{s}})\frac{\varepsilon_{\mathrm{ps}}(\xi p)}{1-\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}}}+(T-pT_{\mathrm{s}})\left(1-\frac{\varepsilon_{\mathrm{ps}}(\xi p)}{1-\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}}}\right)\\&=T-\left(1+\frac{(\xi-1)\varepsilon_{\mathrm{ps}}(\xi p)}{1-\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}}}\right)pT_{\mathrm{s}}.\end{aligned} \tag{30}$$

The time interval between two consecutive successful receptions can be expressed as $Y_v=M_1\xi pT_{\mathrm{s}}+M_2T+T$. The probability of $Y_v$ is given by $\Pr\{Y_v=M_1\xi pT_{\mathrm{s}}+M_2T+T\}=\binom{M_1+M_2}{M_1}(\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}})^{M_1}((1-\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}})\bar{\varepsilon}^{\mathrm{PAC}}_{\mathrm{te}})^{M_2}(1-\varepsilon^{\mathrm{PAC}})$. The expectations of $Y_v$ and $Y_v^2$ are respectively calculated in a recursive manner as

$$\begin{aligned}\mathbb{E}[Y_v]&=\sum_{M_1=0}^{\infty}\sum_{M_2=0}^{\infty}(M_1\xi pT_{\mathrm{s}}+M_2T+T)\tbinom{M_1+M_2}{M_1}(\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}})^{M_1}\\&\times((1-\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}})\bar{\varepsilon}^{\mathrm{PAC}}_{\mathrm{te}})^{M_2}(1-\varepsilon^{\mathrm{PAC}})\\&=T+\frac{(1-\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}})\bar{\varepsilon}^{\mathrm{PAC}}_{\mathrm{te}}T+\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}}\xi pT_{\mathrm{s}}}{1-\varepsilon^{\mathrm{PAC}}},\end{aligned} \tag{31}$$

and

$$\begin{aligned}\mathbb{E}[Y_v^2]&=\sum_{M_1=0}^{\infty}\sum_{M_2=0}^{\infty}(M_1\xi pT_{\mathrm{s}}+M_2T+T)^2\tbinom{M_1+M_2}{M_1}(\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}})^{M_1}\\&\times((1-\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}})\bar{\varepsilon}^{\mathrm{PAC}}_{\mathrm{te}})^{M_2}(1-\varepsilon^{\mathrm{PAC}})\\&=\left(T+\frac{(1-\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}})\bar{\varepsilon}^{\mathrm{PAC}}_{\mathrm{te}}T+\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}}\xi pT_{\mathrm{s}}}{1-\varepsilon^{\mathrm{PAC}}}\right)^2\\&+\frac{1}{\varepsilon^{\mathrm{PAC}}}\left(\frac{(1-\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}})\bar{\varepsilon}^{\mathrm{PAC}}_{\mathrm{te}}T+\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}}\xi pT_{\mathrm{s}}}{1-\varepsilon^{\mathrm{PAC}}}\right)^2.\end{aligned} \tag{32}$$

Substituting (30)-(32) into (18) yields the average AoI of the proposed PredAggComp scheme with prediction horizons of $p$ and $\xi p$ and differential level of $\alpha$ as

$$\begin{aligned}\bar{\Delta}^{\mathrm{PAC}}(p,\xi p,\alpha)&=\frac{3}{2}T-\left(1+\frac{(\xi-1)\varepsilon_{\mathrm{ps}}(\xi p)}{1-\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}}}\right)pT_{\mathrm{s}}\\&+\frac{(1-\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}})\bar{\varepsilon}^{\mathrm{PAC}}_{\mathrm{te}}T+\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}}\xi pT_{\mathrm{s}}}{2(1-\varepsilon^{\mathrm{PAC}})}\\&\times\left(1+\frac{(1-\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}})\bar{\varepsilon}^{\mathrm{PAC}}_{\mathrm{te}}T+\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}}\xi pT_{\mathrm{s}}}{\varepsilon^{\mathrm{PAC}}((1-\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}})T+\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}}\xi pT_{\mathrm{s}})}\right).\end{aligned} \tag{33}$$

For convenience, we hereafter refer to the prediction horizon of the PredAggComp scheme as $\xi p$. From (33), we provide **Proposition 2** as follows.

*Proposition 2:* Given a blocklength $N$, $\bar{\Delta}^{\mathrm{PAC}}(p,\xi p,\alpha)$ monotonically increases with a higher differential level $\alpha$.

*Proof:* The first derivative of $\bar{\Delta}^{\mathrm{PAC}}(p,\xi p,\alpha)$ with respect to $\alpha$ is given by $\frac{\partial\bar{\Delta}^{\mathrm{PAC}}(p,\xi p,\alpha)}{\partial\alpha}=\frac{\partial\bar{\varepsilon}^{\mathrm{PAC}}_{\mathrm{te}}}{\partial\alpha}\frac{1-\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}}}{2(1-\varepsilon^{\mathrm{PAC}})^2}(\varsigma+\frac{\kappa}{\varsigma(\varepsilon^{\mathrm{PAC}})^2}[2\varepsilon^{\mathrm{PAC}}(1-\varepsilon^{\mathrm{PAC}})T-\kappa(1-2\varepsilon^{\mathrm{PAC}})]$, with $\varsigma=\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}}\xi pT_{\mathrm{s}}+(1-\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}})T$ and $\kappa=\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}}\xi pT_{\mathrm{s}}+(1-\varepsilon^{\mathrm{PAC}}_{\mathrm{pe}})\bar{\varepsilon}^{\mathrm{PAC}}_{\mathrm{te}}T$. Using the inequality $\kappa\leq\varepsilon^{\mathrm{PAC}}T$ and the fact that $\bar{\varepsilon}^{\mathrm{PAC}}_{\mathrm{te}}$ increases with a larger number of information bits [9] (*i.e.*, $\frac{\partial\bar{\varepsilon}^{\mathrm{PAC}}_{\mathrm{te}}}{\partial\alpha}>0$), we can obtain $\frac{\partial\bar{\Delta}^{\mathrm{PAC}}(p,\xi p,\alpha)}{\partial\alpha}>0$. ■

*Proposition 2* implies that a larger differential level leads to a higher value of average AoI of the proposed PredAggComp scheme. More information bits are transmitted with the increase of the differential level. For a given blocklength, the room for error correction code reduces and the transmission error probability increases, leading to an increase of AoI.

The proposed prediction-aggregation provides a tradeoff between prediction error probability and transmission error probability. Data aggregation of the predicted status updates with different prediction horizon leads to a lower prediction error probability. On the other hand, with a given blocklength, data aggregation introduces more information bits than the case of transmitting a single status update and therefore results in a higher transmission error probability. The proposed prediction-aggregation allows a good tradeoff between prediction error probability and transmission error probability with the knowledge of the differential level $\alpha$ of the predicted status updates, according to **Proposition 2**. Thus, this motivates us to analyze the preference region of prediction-aggregation in terms of differential level.

### D. Preference Region Analysis

We analyze the preference region of prediction-aggregation in terms of differential level $\alpha$ by comparing the average AoI of PredComp and PredAggComp schemes under the same prediction horizon of $\xi p$. Depending on the values of $\bar{\Delta}^{\mathrm{PAC}}(p,\xi p,\alpha)$ and $\bar{\Delta}^{\mathrm{PC}}(\xi p)$, there are two cases:

- If $\bar{\Delta}^{\mathrm{PAC}}(p,\xi p,1) \le \bar{\Delta}^{\mathrm{PC}}(\xi p)$, prediction-aggregation is preferable regardless of the differential level.
- If $\bar{\Delta}^{\mathrm{PAC}}(p,\xi p,1) > \bar{\Delta}^{\mathrm{PC}}(\xi p)$, according to **Proposition 2**, there exists a differential level threshold $\alpha_{\mathrm{thr}}$, only below which prediction-aggregation is preferable. $\alpha_{\mathrm{thr}}$ is the solution to $\bar{\Delta}^{\mathrm{PAC}}(p,\xi p,\alpha) = \bar{\Delta}^{\mathrm{PC}}(\xi p)$.

Due to the complex structure of $\bar{\varepsilon}_{\mathrm{te}}^{\mathrm{PAC}}$ as shown in (4), it is hard to derive a closed form $\alpha_{\mathrm{thr}}$ from $\bar{\Delta}^{\mathrm{PAC}}(p,\xi p,\alpha) = \bar{\Delta}^{\mathrm{PC}}(\xi p)$. To this end, we first rewrite (4) as $\bar{\varepsilon}_{\mathrm{te}}^{\mathrm{PAC}} \approx \varphi + \phi$, with $\varphi = 1 - e^{-\frac{1}{\bar{\gamma}}(\eta^{\mathrm{PAC}} + \frac{1}{2\lambda^{\mathrm{PAC}}})}$ and $\phi = (1+\bar{\gamma}\lambda^{\mathrm{PAC}})e^{-\frac{1}{\bar{\gamma}}(\eta^{\mathrm{PAC}} + \frac{1}{2\lambda^{\mathrm{PAC}}})} - \bar{\gamma}\lambda e^{-\frac{1}{\bar{\gamma}}(\eta^{\mathrm{PAC}} - \frac{1}{2\lambda^{\mathrm{PAC}}})}$. Equation (4) can be further simplified as $\bar{\varepsilon}_{\mathrm{te}}^{\mathrm{PAC}} \approx \varphi \triangleq \bar{\varepsilon}_{\mathrm{te,lb}}^{\mathrm{PAC}}$ using the fact that $\phi > 0$ is small compared with $\varphi$ [9]. Then, substituting $\bar{\varepsilon}_{\mathrm{te,lb}}^{\mathrm{PAC}}$ into (33) yields a lower bound of average AoI of the proposed PredAggComp scheme as

$$\bar{\Delta}_{\mathrm{lb}}^{\mathrm{PAC}}(p,\xi p,\alpha) = \bar{\Delta}^{\mathrm{PAC}}(p,\xi p,\alpha)\Big|_{\bar{\varepsilon}_{\mathrm{te}}^{\mathrm{PAC}} = \bar{\varepsilon}_{\mathrm{te,lb}}^{\mathrm{PAC}}}. \tag{34}$$

The tightness is verified by Fig. 3. By solving $\bar{\Delta}_{\mathrm{lb}}^{\mathrm{PAC}}(p,\xi p,\alpha) = \bar{\Delta}^{\mathrm{PC}}(\xi p)$, an approximate differential level threshold is given by

$$\alpha_{\mathrm{thr}} \approx \frac{1}{b} N \ln\left(\frac{\theta\vartheta - \sqrt{\theta\vartheta^2 - \theta + 1}}{\theta - 1}\right) - 1, \tag{35}$$

with $\theta = 2N/\pi$, $\vartheta = 1 - \bar{\gamma}\ln(1 - \frac{-\chi_{\mathrm{b}} + \sqrt{{\chi_{\mathrm{b}}}^2 - 4\chi_{\mathrm{a}}\chi_{\mathrm{c}}}}{2\chi_{\mathrm{a}}})$, $\chi_{\mathrm{a}} = T(1-\varepsilon_{\mathrm{pe}}^{\mathrm{PAC}})^2(T - \zeta - \varpi - \frac{\zeta\varepsilon_{\mathrm{pe}}^{\mathrm{PAC}}\xi p T_{\mathrm{s}}}{T(1-\varepsilon_{\mathrm{pe}}^{\mathrm{PAC}})})$, $\chi_{\mathrm{b}} = 2\varepsilon_{\mathrm{pe}}^{\mathrm{PAC}}\xi p T_{\mathrm{s}} T(1 - \varepsilon_{\mathrm{pe}}^{\mathrm{PAC}}) - \varpi(1-\varepsilon_{\mathrm{pe}}^{\mathrm{PAC}})(2\zeta - \varpi + 2T(1-\varepsilon_{\mathrm{pe}}^{\mathrm{PAC}}))$, $\chi_{\mathrm{c}} = (\varepsilon_{\mathrm{pe}}^{\mathrm{PAC}}\xi p T_{\mathrm{s}})^2 - \varpi(1-\varepsilon_{\mathrm{pe}}^{\mathrm{PAC}})(\zeta - \varepsilon_{\mathrm{pe}}^{\mathrm{PAC}}\xi p T_{\mathrm{s}})$, $\zeta = (1-\varepsilon_{\mathrm{pe}}^{\mathrm{PAC}})(2\bar{\Delta}^{\mathrm{PC}}(\xi p) - 3T) + \xi p T_{\mathrm{s}}(1 + \varepsilon_{\mathrm{ps}}(\xi p) - \varepsilon_{\mathrm{pe}}^{\mathrm{PAC}})$ and $\varpi = T(1-\varepsilon_{\mathrm{pe}}^{\mathrm{PAC}}) - \varepsilon_{\mathrm{pe}}^{\mathrm{PAC}}\xi p T_{\mathrm{s}}$.

## V. Adaptation of Prediction Horizon

For the proposed PredComp and PredAggComp transmission schemes, a long prediction horizon can reduce AoI. However, a too long prediction horizon increases prediction error probability [16], which limits the reduction of AoI. Therefore, the size of prediction horizon affects AoI. Based on the closed-form average AoI expressions respectively derived in Sections III and IV, we investigate prediction horizon adaptation for the proposed PredComp and PredAggComp transmission schemes in terms of average AoI.

Unfortunately, the prediction horizon affects the average AoI of the two proposed schemes in a complex manner, as shown in (21) and (33), especially with a multi-feature source. The prediction error probability is a complex multivariate Gaussian probability subject to high correlation between features to be significantly coupled together. As stated in [37] and [38], multivariate Gaussian probabilities are open problems, analytically intractable. Thus, it is extremely difficult to derive the closed-form solution to the problem of multivariate Gaussian probabilities with multiple coupling features. However, a single-feature source can relax such tough problem. We investigate closed-form expressions of the minimum average AoI of the two proposed schemes with a single-feature source, respectively, by adapting the prediction horizon. Also, we optimize the prediction horizon to minimize the average AoI of the schemes with a multi-feature source by golden section search based methods.

### A. Prediction Horizon Adaptation for the Proposed PredComp Transmission Scheme

Due to the complex form of the average AoI expression $\bar{\Delta}^{\mathrm{PC}}(p)$, we use its tight upper bound $\bar{\Delta}_{\mathrm{ub}}^{\mathrm{PC}}(p)$ as the optimization objective for the PredComp scheme with a single-feature source. The optimization problem is formulated as (36), shown at the bottom of the page, where (C2) represents that prediction horizon $p$ is a positive integer, shorter than or equal to the transmission time of a predicted update packet. Relaxing integer variable $p$ to a continuous variable, we provide **Proposition 3** and **Lemma 1** as follows.

*Proposition 3:* Optimization problem **P1** is convex with respect to $p$.

*Proof:* Please refer to Appendix A. ■

*Lemma 1:* The optimal prediction horizon $p_*^{\mathrm{PC}}$ for Problem **P1** is given by

$$p_*^{\mathrm{PC}} = \begin{cases} 1, & \text{if } G(1) > 0, \\ N, & \text{if } G(N) < 0, \\ \underset{p\in\{\lfloor\hat{p}\rfloor,\lceil\hat{p}\rceil\}}{\arg\min}\ \bar{\Delta}_{\mathrm{ub}}^{\mathrm{PC}}(p), & \text{otherwise}, \end{cases} \tag{37}$$

where $\hat{p}$ is the root of equation $G(p) = 0$, with $G(p)$ being the first derivative of $\bar{\Delta}_{\mathrm{ub}}^{\mathrm{PC}}(p)$ with respect to $p$ under $F = 1$, expressed as $G(p) = -T_{\mathrm{s}} + \frac{-\frac{\partial\varepsilon_{\mathrm{ps}}(p)}{\partial p} p T_{\mathrm{s}} + \varepsilon_{\mathrm{ps}}(p)(1-\varepsilon_{\mathrm{ps}}(p))T_{\mathrm{s}}}{(1-\bar{\varepsilon}_{\mathrm{te}}^{\mathrm{PC}})(\varepsilon_{\mathrm{ps}}(p))^2}$, and

$$\frac{\partial\varepsilon_{\mathrm{ps}}(p)}{\partial p} = \begin{cases} \frac{-\delta a^{2p}\ln a^2}{\sqrt{2\pi}\sigma_{\mathrm{w}}(a^{2p}-1)}\sqrt{\frac{1-a^2}{1-a^{2p}}}e^{-\frac{\delta^2(1-a^2)}{2\sigma_{\mathrm{w}}^2(1-a^{2p})}}, & \text{if } a^2 \neq 1, \\ \frac{-\delta}{\sqrt{2\pi}\sigma_{\mathrm{w}}}p^{-\frac{3}{2}}e^{-\frac{\delta^2}{2p\sigma_{\mathrm{w}}^2}}, & \text{if } a^2 = 1. \end{cases} \tag{38}$$

*Proof:* Please refer to Appendix B. ■

With a multi-feature source, the convexity of $\bar{\Delta}^{\mathrm{PC}}(p)$ with respect to $p$ is verified in Fig. 3. We use the golden section search method [39] to obtain the optimal prediction horizon of the proposed PredComp scheme based on the closed-form average AoI. Letting $\kappa = (\sqrt{5}-1)/2$ denote the golden ratio and $\widetilde{\theta}_{\mathrm{p}}$ the error tolerance of $p$ with $p \in [1, N]$, the complexity of the golden section search is $O\left(\left\lceil \frac{\ln(\widetilde{\theta}_{\mathrm{p}}/(N-1))}{\ln(\kappa)} \right\rceil\right)$, which is lower than the complexity of exhaustive search, $O(N)$.

### *B. Prediction Horizon Adaptation for the Proposed PredAggComp Transmission Scheme*

The closed-form average AoI of the PredAggComp in (33) can be upper bounded by

$$\begin{aligned}\bar{\Delta}^{\mathrm{PAC}}(p,\xi p,\alpha) &\overset{(a)}{\leq} \frac{3}{2}T - \left(1+\frac{(\xi-1)\varepsilon_{\mathrm{ps}}(\xi p)}{1-\varepsilon_{\mathrm{pe}}^{\mathrm{PAC}}}\right)pT_{\mathrm{s}} \\ &\quad + \frac{(1-\varepsilon_{\mathrm{pe}}^{\mathrm{PAC}})\bar{\varepsilon}_{\mathrm{te}}^{\mathrm{PAC}}T + \varepsilon_{\mathrm{pe}}^{\mathrm{PAC}}\xi pT_{\mathrm{s}}}{1-\varepsilon^{\mathrm{PAC}}} \\ &\overset{(b)}{\leq} \frac{3}{2}T - \left(1+\frac{(\xi-1)\varepsilon_{\mathrm{ps}}(\xi p)}{\varepsilon_{\mathrm{ps}}(p)}\right)pT_{\mathrm{s}} \\ &\quad + \frac{\varepsilon_{\mathrm{ps}}(p)\bar{\varepsilon}_{\mathrm{te}}^{\mathrm{PAC}}T + (1-\varepsilon_{\mathrm{ps}}(p))\xi pT_{\mathrm{s}}}{\varepsilon_{\mathrm{ps}}(p)(1-\bar{\varepsilon}_{\mathrm{te}}^{\mathrm{PAC}})} \\ &\triangleq \bar{\Delta}_{\mathrm{ub}}^{\mathrm{PAC}}(p,\xi p,\alpha), \end{aligned} \tag{39}$$

where (a) uses the inequality $\xi pT_{\mathrm{s}} \leq T$ and (b) follows $\varepsilon_{\mathrm{pe}}^{\mathrm{PAC}} \leq 1-\varepsilon_{\mathrm{ps}}(p)$. Similar to the optimization objective function in (36), for the PredAggComp scheme with a single-feature source, we choose to minimize the upper bound on its average AoI in (39), whose tightness is verified by Fig. 3. The optimization problem is given by

$$\begin{aligned}\mathbf{P2}: \min_{p,\xi}\ & \bar{\Delta}_{\mathrm{ub}}^{\mathrm{PAC}}(p,\xi p,\alpha) \\ \text{s.t.}\ & (\mathrm{C3}): 0 < \xi pT_{\mathrm{s}} \leq NT_{\mathrm{s}}, p \in \mathbb{N}_+, \xi p \in \mathbb{N}_+, \\ & \xi > 1, \text{and } (\mathrm{C1}). \end{aligned} \tag{40}$$

Problem **P2** is not convex with respect to both $p$ and $\xi$. To address this issue, we decompose Problem **P2** as

$$\mathbf{P2.1}: \min_{p}\ \bar{\Delta}_{\mathrm{ub}}^{\mathrm{PAC}}(p,\xi p,\alpha)\ \text{ s.t. } (\mathrm{C1}) \text{ and } (\mathrm{C3}), \tag{41}$$

$$\mathbf{P2.2}: \min_{\xi}\ \bar{\Delta}_{\mathrm{ub}}^{\mathrm{PAC}}(p,\xi p,\alpha)\ \text{ s.t. } (\mathrm{C1}) \text{ and } (\mathrm{C3}). \tag{42}$$

Relaxing the integer constraint of $p$ into continuous space, we provide **Proposition 4** as follows.

*Proposition 4:* Given a ratio $\xi$ of the prediction horizons of the updates, Problem **P2.1** is convex with respect to $p$. Given a prediction horizon $p$, Problem **P2.2** is convex with respect to $\xi$.

**Algorithm 1** The JPHRO Algorithm

1 **Initialize:** Iteration index $i = 1$, $F$, $\xi^{(i)}$, $p^{(i)}$, maximum number of iterations $I_{\max}$, and error tolerances $\widetilde{\theta}_{\mathrm{p}}$ and $\widetilde{\theta}_{\xi}$;
2 **Repeat**
3 If $F = 1$, calculate $p^{(i)}$ using (43) with a given $\xi^{(i-1)}$, and then calculate $\xi^{(i)}$ using (44) with given $p^{(i)}$;
4 If $F > 1$, obtain $p^{(i)}$ by golden section search with a given $\xi^{(i-1)}$, and then obtain $\xi^{(i)}$ by golden section search with given $p^{(i)}$;
5 $i = i + 1$;
6 **Until** $i \geq I_{\max}$ or $|\xi^{(i)} - \xi^{(i-1)}| < \widetilde{\theta}_{\xi}$ & $|p^{(i)} - p^{(i-1)}| < \widetilde{\theta}_{\mathrm{p}}$

*Proof:* Please refer to Appendix C. ■

Based on **Proposition 4**, similar to the proof of **Lemma 1**, we can provide **Lemma 2** and **Lemma 3**.

*Lemma 2:* Given a ratio $\xi$ of the prediction horizons of the updates, the optimal prediction horizon $p_*^{\mathrm{PAC}}$ for Problem **P2.1** is given by

$$p_*^{\mathrm{PAC}} = \begin{cases} 1, & \text{if } H(1) > 0, \\ N/\xi, & \text{if } H(N/\xi) < 0, \\ \underset{p\in\{\lfloor\tilde{p}\rfloor,\lceil\tilde{p}\rceil\}}{\arg\min}\ \bar{\Delta}_{\mathrm{ub}}^{\mathrm{PAC}}(p,\xi p,\alpha), & \text{otherwise}, \end{cases} \tag{43}$$

where $\tilde{p}$ is the root of $H(p) = 0$, with $H(p)$ being the first derivative of $\bar{\Delta}_{\mathrm{ub}}^{\mathrm{PAC}}(p,\xi p,\alpha)$ with respect to $p$ under $F = 1$, expressed as $H(p) = -T_{\mathrm{s}} - (\xi-1)\frac{\varepsilon_{\mathrm{ps}}(\xi p)T_{\mathrm{s}} + \frac{\partial\varepsilon_{\mathrm{ps}}(\xi p)}{\partial p}\xi pT_{\mathrm{s}}}{\varepsilon_{\mathrm{ps}}(p)} + (\xi-1)pT_{\mathrm{s}}\frac{\partial\varepsilon_{\mathrm{ps}}(p)}{\partial p}\frac{\varepsilon_{\mathrm{ps}}(\xi p)}{(\varepsilon_{\mathrm{ps}}(p))^2} + \xi\frac{\varepsilon_{\mathrm{ps}}(p)(1-\varepsilon_{\mathrm{ps}}(p))T_{\mathrm{s}} - \frac{\partial\varepsilon_{\mathrm{ps}}(p)}{\partial p}pT_{\mathrm{s}}}{(1-\bar{\varepsilon}_{\mathrm{te}}^{\mathrm{PAC}})(\varepsilon_{\mathrm{ps}}(p))^2}$.

*Lemma 3:* Given a prediction horizon $p$, the optimal ratio $\xi_*^{\mathrm{PAC}}$ of the prediction horizons of the updates for Problem **P2.2** is given by

$$\xi_*^{\mathrm{PAC}} = \begin{cases} 1, & \text{if } J(1) > 0, \\ N/p, & \text{if } J(N/p) < 0, \\ \tilde{\xi}, & \text{otherwise}, \end{cases} \tag{44}$$

where $\tilde{\xi}$ is the root of $J(\xi) = 0$, with $J(\xi)$ being the first derivative of $\bar{\Delta}_{\mathrm{ub}}^{\mathrm{PAC}}(p,\xi p,\alpha)$ with respect to $\xi$ under $F = 1$, expressed as $J(\xi) = \frac{(1-\varepsilon_{\mathrm{ps}}(p))pT_{\mathrm{s}}}{\varepsilon_{\mathrm{ps}}(p)(1-\bar{\varepsilon}_{\mathrm{te}}^{PAC})} - \frac{pT_{\mathrm{s}}}{\varepsilon_{\mathrm{ps}}(p)}(\varepsilon_{\mathrm{ps}}(\xi p) + p^2(\xi-1)\frac{\partial\varepsilon_{\mathrm{ps}}(\xi p)}{\partial p})$.

With **Lemma 2** and **3**, Problem **P2** can be solved in an iterative manner, as described in the proposed JPHRO algorithm in Algorithm 1.

In the multi-feature source scenario, the convexity of $\bar{\Delta}^{\mathrm{PAC}}(p,\xi p,\alpha)$ with respect to $p$ and $\xi$ are verified in Fig. 3 and Fig. 7, respectively. We can use an iterative method

$$\begin{aligned}\mathbf{P1}: \min_{p}\ & \bar{\Delta}_{\mathrm{ub}}^{\mathrm{PC}}(p) \\ \text{s.t.}\ & (\mathrm{C1}): \varepsilon_{\mathrm{ps}}(p)\,|_{F=1} = \begin{cases} 1-2Q\left(\frac{\delta}{\sigma_{\mathrm{w}}}\sqrt{\frac{1-a^2}{1-a^{2p}}}\right), & \text{if } a^2 \neq 1, \\ 1-2Q\left(\frac{\delta}{\sigma_{\mathrm{w}}\sqrt{p}}\right), & \text{if } a^2 = 1, \end{cases} \\ & (\mathrm{C2}): 0 < pT_{\mathrm{s}} \leq NT_{\mathrm{s}}, p \in \mathbb{N}_+, \end{aligned} \tag{36}$$

based on the golden section search to obtain the optimal prediction horizon and prediction horizon ratio of the proposed PredAggComp scheme, as shown in Algorithm 1.

The complexity of the proposed JPHRO algorithm with $F = 1$ source feature depends on the maximum number of iterations $I_{\max}$ and the complexity of solving the non-linear equations of $H(p) = 0$ and $J(\xi) = 0$. Using Newton's method [40] to solve the two equations, the total complexity is $O\left(I_{\max}(||\mu_{\rm p}|| + ||\mu_{\xi}||)\right)$, where $\mu_{\rm p}$ and $\mu_{\xi}$ are the gaps between the initialized solution and the corresponding exact solution of the two equations, respectively. The complexity of the JPHRO algorithm with $F > 1$ source features depends on the value of $I_{\max}$ and the golden section search in steps 3 and 4. With $p \in [1, \lfloor N/\xi \rfloor]$ given a $\xi$, the computation number in step 3 is $I_{\rm p} = 2 + \left\lceil \frac{\ln\left(\tilde{\theta}_{\rm p}/(\lfloor N/\xi \rfloor - 1)\right)}{\ln(\kappa)} \right\rceil$. With $\xi \in [\frac{p+1}{p}, \frac{p+2}{p}, \ldots, \frac{N}{p}]$ given a $p$, the computation number in step 4 is $I_{\xi} = 2 + \left\lceil \frac{\ln\left(\tilde{\theta}_{\xi}/((N-p-1)/p)\right)}{\ln(\kappa)} \right\rceil$. Hence, the total complexity is $O\left(I_{\max}(I_{\rm p} + I_{\xi})\right)$. In contract, exhaustive search for the proposed PredAggComp scheme requires a much higher complexity of $O\left((N-1)N/2\right)$.

## VI. Simulation Results

In this section, simulation results are presented to verify the performance analysis of average AoI of the proposed PredComp and PredAggComp transmission schemes, and evaluate the performance of the proposed prediction horizon adaptation methods.

Similar to [16], we take a steady one-dimensional movement process as an example, where the source senses velocity and location of a target device. In this two-feature source scenario, the state transition model can be expressed as

$$\begin{bmatrix} x_{\rm l}(\mathrm{s}_k+1) \\ x_{\rm v}(\mathrm{s}_k+1) \end{bmatrix} = \begin{bmatrix} 1 & T_{\rm s} \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_{\rm l}(\mathrm{s}_k) \\ x_{\rm v}(\mathrm{s}_k) \end{bmatrix} + \begin{bmatrix} 0 \\ w_{\rm v}(\mathrm{s}_k) \end{bmatrix},$$

where $x_{\rm l}(\mathrm{s}_k)$, $x_{\rm v}(\mathrm{s}_k)$ and $w_{\rm v}(\mathrm{s}_k)$ are the location, velocity and Gaussian noise of velocity at time slot $\mathrm{s}_k$, respectively. During predictions, the standard deviation of velocity noise is $\sigma_{\rm w,v} = 0.12$ mm/ms, and the prediction error thresholds of location and velocity are $\delta_{\rm l}$ mm and $\delta_{\rm v}$ mm/ms, respectively. Without loss of generality, we set $\delta_{\rm l} = \delta_{\rm v} = \delta$. The ratio of the prediction horizons of the updates in the proposed PredAggComp scheme is $\xi = 2$, except for Figs. 7, 8, 9 and 10. Hence, the prediction horizon is set to $2pT_{\rm s}$, for fair comparison between the proposed PredComp and PredAggComp schemes. The differential level of two predicted updates is set to $\alpha = 0.1$, except for Figs. 4 and 5. We consider the number of information bits per predicted update $b = 160$, and the blocklength $N = 300$ c.u. except for Fig. 5. The bandwidth is 100 KHz [41]. Hence, the symbol duration is $T_{\rm s} = 0.01$ ms and the packet transmission time without proactive transmission termination is $T = NT_{\rm s} = 3$ ms. To compute the multivariate Gaussian integral in (28), $\Lambda$ and $U$ are chosen as 8 and 1000, respectively [42]. The simulation results are obtained by Monte Carlo simulation with $10^5$ realizations of packet transmission. The pathloss is modeled as $35.3 + 37.6\log_{10}(d) + S$, where $d = 200$ m is the distance between source and destination and $S$ is the shadowing which follows log normal distribution with zero mean and standard deviation of 8 [16]. The noise power spectrum density is $-174$ dBm/Hz [16]. Transmission power is $P_{\rm t} = 13$ dBm, yielding the average SNR to be $\bar{\gamma} = \beta P_{\rm t}/\sigma_{\rm n}^2 = 15$ dB.

### A. Average AoI Analysis

Fig. 3 shows the impact of prediction horizon on average AoI performance of the proposed PredComp and PredAggComp transmission schemes, as compared to the previous work [9], [10], [11], [12], [13], [14] without prediction and the existing status-prediction related work [16], [18] without proactive transmission termination. To better illustrate the impact of feature correlation on the performance analysis, the work of [16] and [18] is shown under two cases: i) with and ii) without feature correlation, which are labeled as 'No termination + feature correlation' and 'No termination - feature correlation', respectively. It is observed that there exists a big gap between the above two cases, implying that the results of existing methods considering no feature correlation [16], [18] cannot be directly applied to the case with feature correlations. Also, it is observed that the analytical results of average AoI in (21) and (33) match with simulation results, and the analytical upper bounds in (22) and (39) and lower bound in (34) are tight to the simulation results. The proposed PredComp scheme achieves a significantly lower average AoI than that without prediction [9], [10], [11], [12], [13], [14]. Also, it outperforms the existing status-prediction schemes [16], [18] due to proactive transmission termination. The proposed PredAggComp scheme enables a reduction of average AoI with the increase of prediction horizon. There are optimal prediction horizons to minimize the values of the average AoI of the PredComp and PredAggComp schemes, respectively, consistent with our analysis in Section V.

Fig. 4 shows the preference regions of the proposed PredComp and PredAggComp transmission schemes in terms of the differential level under different prediction horizons and prediction error thresholds, and demonstrates the tightness of the approximate differential level threshold in (35). The approximate differential level threshold is close to the precise threshold by exhaustive search, with a much lower complexity. A longer prediction horizon and a lower prediction error threshold lead to a larger differential level threshold, as the benefit of prediction error probability reduction by the proposed prediction-aggregation increases.

Fig. 5 shows the preference regions of the proposed PredComp and PredAggComp transmission schemes in terms of the differential level under different blocklengths and average SNRs. The approximate differential level threshold is tight at medium to high SNR. At blocklength $N = 150$ c.u., the gaps between the precise and approximate thresholds at average SNR $\bar{\gamma} = 15$, 10 and 5 dB are 0.4%, 3.7% and 31%, respectively. The differential level threshold increases with a larger blocklength and a higher average SNR. The transmission error probability caused by the update aggregation reduces. For a given prediction horizon, the tolerance of the differential level increases.

Fig. 6 shows the impact of prediction horizon on the equivalent update rate at destination. The proposed PredAggComp scheme achieves a nearly double equivalent update rate at

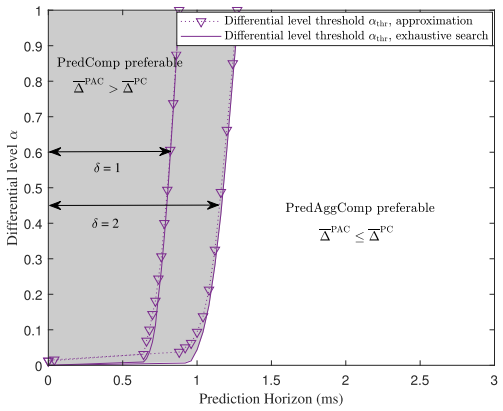



Fig. 4. Preference regions of the proposed PredComp and PredAggComp transmission schemes in terms of the differential level with a two-feature source at average SNR $\bar{\gamma} = 25$ dB and prediction error threshold $\delta = 1$ and 2.

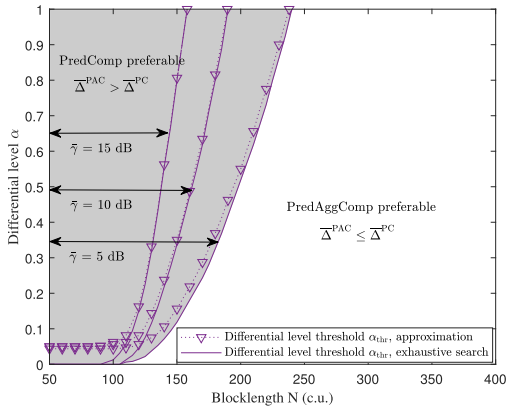



Fig. 5. Preference regions of the proposed PredComp and PredAggComp transmission schemes in terms of the differential level with a two-feature source at prediction horizon equal to packet transmission time, prediction error threshold $\delta = 2$ and average SNR $\bar{\gamma} = 15$, 10 and 5 dB.

destination than the existing work [9], [10], [11], [12], [13], [14] without prediction. The proposed PredAggComp scheme allows transmissions of two predicted updates, while there is a single update to be transmitted for the existing work [9], [10], [11], [12], [13], [14]. This is of significance for time-critical IIoT applications such as real-time monitoring. The increasing number of historical updates is helpful to exploit and predict the device state at destination.

Fig. 7 shows the impact of ratio of the prediction horizons of the updates on average AoI of the proposed PredAggComp scheme. Given a $p$, there is an optimal ratio of the prediction horizons to achieve the purpose of the minimized value of the average AoI of the PredAggComp scheme, consistent with **Proposition 4**.

### *B. Prediction Horizon Adaptation*

In Figs. 8, 9, and 10, we show the performance of the proposed prediction horizon adaptation methods with $F = 1$ and 2 source features. The state transition model in the single-feature source scenario simplifies to $x_{\text{v}}(\text{s}_k + 1) = x_{\text{v}}(\text{s}_k) + w_{\text{v}}(\text{s}_k)$. The maximum number of iterations in the proposed JPHRO algorithm is set to 3.

Figs. 8 and 9 show the impacts of average SNR and prediction error threshold of the velocity on the minimum average AoI of the proposed PredComp and PredAggComp

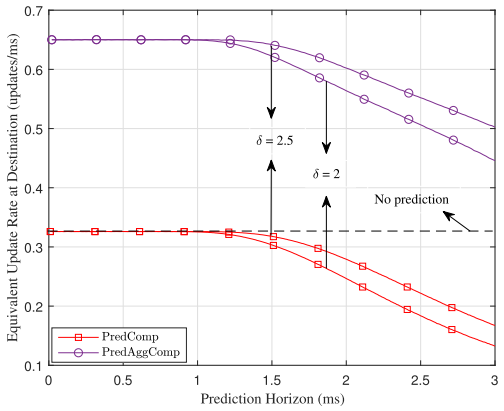



Fig. 6. Impact of prediction horizon on the equivalent update rate at destination with a two-feature source at prediction error threshold $\delta = 2$ and 2.5.

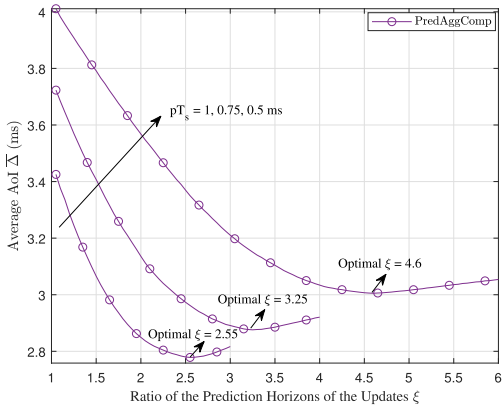



Fig. 7. Impact of ratio of the prediction horizons of the updates on average AoI of the proposed PredAggComp scheme with a two-feature source at prediction error threshold $\delta = 2$. With constraint $\xi p T_{\text{s}} \leq T = 3$ ms, the maximum ratio $\xi$ for prediction horizon $pT_{\text{s}} = 1$, 0.75 and 0.5 ms are 3, 4 and 6, respectively.

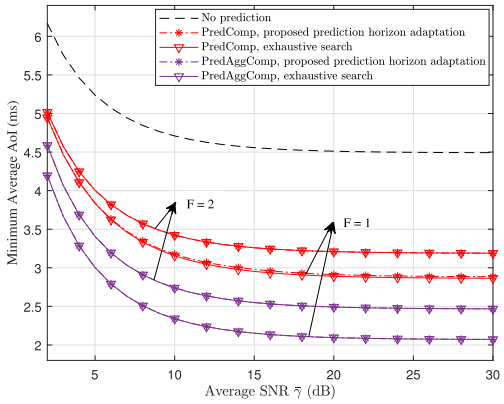



Fig. 8. Impact of average SNR on the minimum average AoI of the proposed PredComp and PredAggComp transmission schemes by the proposed prediction horizon adaptation methods with $F = 1$ and 2 source features and prediction error threshold $\delta = 2$.

transmission schemes by the proposed prediction horizon adaptation methods, respectively. With a single-feature source, the upper bounds on the average AoI are minimized by the proposed adaptation methods, while the exact average AoI expressions are minimized by exhaustive search as a benchmark. It can be observed in Figs. 8 and 9 that the two sets of the prediction horizon results are very close, demonstrating

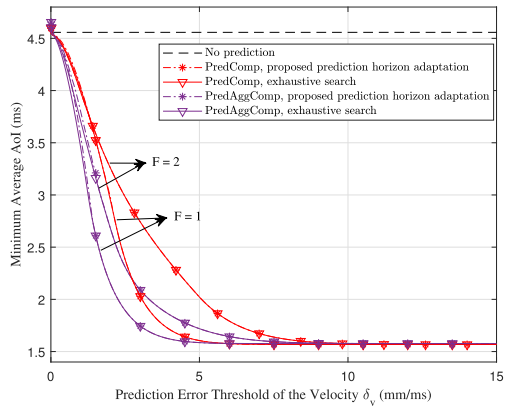



Fig. 9. Impact of prediction error threshold of the velocity on the minimum average AoI of the proposed PredComp and PredAggComp transmission schemes by the proposed prediction horizon adaptation methods with $F = 1$ and 2 source features.

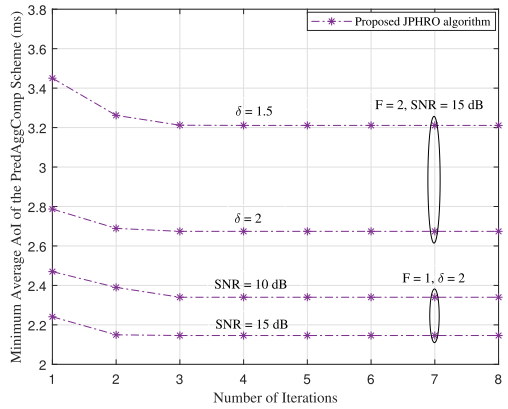



Fig. 10. Convergence behavior of the proposed JPHRO algorithm with $F = 1$ and 2 source features.

the effectiveness of the suboptimal problem formulation and solution for the two schemes. With a two-feature source, the proposed golden section search based adaptation methods also achieve prediction horizon results close to exhaustive search. In Fig. 8, the minimum average AoI decreases with a higher average SNR, as transmission error probability decreases. In Fig. 9, the proposed PredComp and PredAggComp schemes achieve the average AoI reduction by up to 64% over the no prediction case [9], [10], [11], [12], [13], [14] at single-feature source with $\delta_v > 5$ mm/ms and at two-feature source with $\delta_v > 9$ mm/ms, respectively. In addition, when the prediction error threshold is very low or very high, the minimum average AoI of the PredAggComp scheme is close to that of the PredComp scheme, as the capability of prediction-aggregation to reduce prediction error probability is limited in this situation.

Fig. 10 shows the convergence behavior of the proposed JPHRO algorithm with $F = 1$ and 2 source features. The JPHRO algorithm converges fast with 2-3 iterations.

## VII. Conclusion

In this paper, the PredComp transmission scheme has been proposed to enhance the AoI performance of SPT, where proactive transmission termination with prediction error awareness is employed at source. Furthermore, the PredAggComp scheme has been proposed, by aggregation of two predicted time-correlated status data. Closed-form expressions for their average AoI have been presented. The closed-form threshold that the PredAggComp scheme outperforms the PredComp scheme has been derived with regard to the differential level of the predicted updates. Also, prediction adaption has been conducted to minimize the average AoI and the optimal prediction horizons and their optimal ratio have been derived for the PredAggComp scheme. Simulation results show that the PredComp scheme significantly outperforms the previous status-prediction related schemes [16], [18], in terms of the average AoI. The PredComp and PredAggComp schemes enable an average AoI redution of up to 64% over the no prediction case [9], [10], [11], [12], [13], [14] and the performance gain is also robust, even under a high prediction error probability of up to 0.5. In the future, this work will be extended to the multi-source scenario.

## Appendix A
## Proof of Proposition 3

The second derivative of $\bar{\Delta}_{\text{ub}}^{\text{PC}}(p)$ with respect to $p$ under $F = 1$ is given by $\bar{\Delta}_{\text{ub}}^{\text{PC}\prime\prime}(p) = \frac{(1-\varepsilon_{\text{pe}}^{\text{PC}})Z(p)+2(\frac{\partial\varepsilon_{\text{ps}}(p)}{\partial p})^2 pT_{\text{s}}}{(1-\bar{\varepsilon}_{\text{te}}^{\text{PC}})(\varepsilon_{\text{ps}}(p))^3}$, where $Z(p) = -\frac{\partial^2\varepsilon_{\text{ps}}(p)}{\partial p^2}pT_{\text{s}} - 2T_{\text{s}}\frac{\partial\varepsilon_{\text{ps}}(p)}{\partial p}$, $\frac{\partial\varepsilon_{\text{ps}}(p)}{\partial p}$ and $\frac{\partial^2\varepsilon_{\text{ps}}(p)}{\partial p^2}$ are the first derivative and second derivative of $\varepsilon_{\text{ps}}(p)$, respectively. $\frac{\partial^2\varepsilon_{\text{ps}}(p)}{\partial p^2}$ is expressed as

$$\frac{\partial^2\varepsilon_{\text{ps}}(p)}{\partial p^2} = \begin{cases} \frac{1}{2}\frac{\partial\varepsilon_{\text{ps}}(p)}{\partial p}\left(\frac{(2+a^{2p})\ln a^2}{1-a^{2p}} - \frac{(1-a^{2p})\Gamma^2(p)}{a^{2p}\ln a^2}\right), & \text{if } a^2 \neq 1, \\ \frac{1}{2}\frac{\partial\varepsilon_{\text{ps}}(p)}{\partial p}\left(\frac{\delta^2}{p^2\sigma_{\text{w}}^2} - \frac{3}{p}\right), & \text{if } a^2 = 1, \end{cases} \tag{45}$$

with $\Gamma(p) = -\frac{\delta a^{2p}\ln a^2}{\sigma_{\text{w}}(1-a^{2p})}\sqrt{\frac{1-a^2}{1-a^{2p}}}$. With $\frac{\partial\varepsilon_{\text{ps}}(p)}{\partial p} < 0$ as shown in (38), we can obtain $Z(p) = -\frac{1}{2}T_{\text{s}}\frac{\partial\varepsilon_{\text{ps}}(p)}{\partial p}(1+\frac{\delta^2}{p\sigma_{\text{w}}^2}) > 0$ if $a^2 = 1$, and $Z(p) = -\frac{1}{2}T_{\text{s}}\frac{\partial\varepsilon_{\text{ps}}(p)}{\partial p}(\frac{(2+a^{2p})\ln a^{2p}}{1-a^{2p}} - \frac{p(1-a^{2p})(\Gamma(p))^2}{a^{2p}\ln a^2} + 4) > -\frac{1}{2}T_{\text{s}}\frac{\partial\varepsilon_{\text{ps}}(p)}{\partial p}(5 - \frac{p(1-a^{2p})(\Gamma(p))^2}{a^{2p}\ln a^2}) > 0$ if $a^2 \neq 1$. As a result, $\bar{\Delta}_{\text{ub}}^{\text{PC}\prime\prime}(p) > 0$. Moreover, the constraint (C2) is an affine set, and hence **Proposition 3** is proved.

## Appendix B
## Proof of Lemma 1

As Problem **P1** is convex, the optimal prediction horizon $p_*^{\text{PC}}$ can be obtained by applying the Karush-Kuhn-Tucker (KKT) condition. The dual Lagrangian function of Problem **P1** is given by $\Theta = \frac{3}{2}T - pT_{\text{s}} + \frac{(1-\varepsilon_{\text{pe}}^{\text{PC}})\bar{\varepsilon}_{\text{te}}^{\text{PC}}T+\varepsilon_{\text{pe}}^{\text{PC}}pT_{\text{s}}}{1-\varepsilon^{\text{PC}}} + \nu(1-p) + \iota(p-N)$, with $\nu \geq 0$ and $\iota \geq 0$. According to the KKT condition, the optimal prediction horizon $p_*^{\text{PC}}$ should satisfy

$$\left.\frac{d\Theta}{dp}\right|_{p=p_*^{\text{PC}}} = G(p) - \nu^* + \iota^* = 0, \tag{46a}$$

$$\nu^*(1-p_*^{\text{PC}}) = 0, \tag{46b}$$

$$\iota^*(p_*^{\text{PC}} - N) = 0. \tag{46c}$$

Since the objective function is convex, $G(p)$ increases with $p$, *i.e.*, $G(N) \geq G(p) \geq G(1)$. If $G(N) < 0$, then $G(p) < 0$, $\forall p$. It follows that $\iota^*$ must greater than zero, in order to satisfy (46a). According to (46c), we can obtain $p_*^{\text{PC}} = N$.

Similarly, if $G(1) > 0$, then $G(p) > 0, \forall p$. To guarantee (46a), $\nu^*$ needs to be greater than zero, and $p_*^{\mathrm{PC}} = 0$ can be obtained from (46b). If $G(1) \leq 0$ and $G(N) \geq 0$, there exists a unique $p_*^{\mathrm{PC}}$ such that $\bar{\Delta}_{\mathrm{ub}}^{\mathrm{PC}}(p)$ is minimized, which is the root of $G(p) = 0$. With integer constraint of $p$, **Lemma 1** is proved.

## APPENDIX C
## PROOF OF PROPOSITION 4

The second derivative of $\bar{\Delta}_{\mathrm{ub}}^{\mathrm{PAC}}(p, \xi p, \alpha)$ with respect to $p$ is given by $\frac{\partial^2 \bar{\Delta}_{\mathrm{ub}}^{\mathrm{PAC}}(p,\xi p,\alpha)}{\partial p^2} = \xi(\xi-1)\frac{Z(\xi p)}{\varepsilon_{\mathrm{ps}}(p)} - (\xi-1)\frac{\varepsilon_{\mathrm{ps}}(\xi p)Z(p) - 2\frac{\partial \varepsilon_{\mathrm{ps}}(p)}{\partial p}\frac{\partial \varepsilon_{\mathrm{ps}}(\xi p)}{\partial p}\xi p T_{\mathrm{s}}}{(\varepsilon_{\mathrm{ps}}(p))^2} - 2(\xi-1)pT_{\mathrm{s}}(\frac{\partial \varepsilon_{\mathrm{ps}}(p)}{\partial p})^2\frac{\varepsilon_{\mathrm{ps}}(\xi p)}{(\varepsilon_{\mathrm{ps}}(p))^3} + \xi\frac{\varepsilon_{\mathrm{ps}}(p)Z(p) + 2(\frac{\partial \varepsilon_{\mathrm{ps}}(p)}{\partial p})^2 pT_{\mathrm{s}}}{(1-\bar{\varepsilon}_{\mathrm{te}}^{\mathrm{PAC}})(\varepsilon_{\mathrm{ps}}(p))^3}$. Using the fact $1 - \bar{\varepsilon}_{\mathrm{te}}^{\mathrm{PAC}} < 1$ yields $\frac{\partial^2 \bar{\Delta}_{\mathrm{ub}}^{\mathrm{PAC}}(p,\xi p,\alpha)}{\partial p^2} > \xi(\xi-1)\frac{Z(\xi p)}{\varepsilon_{\mathrm{ps}}(p)} + 2pT_{\mathrm{s}}(\frac{\partial \varepsilon_{\mathrm{ps}}(p)}{\partial p})^2\frac{1+(\xi-1)(1-\varepsilon_{\mathrm{ps}}(\xi p))}{(\varepsilon_{\mathrm{ps}}(p))^3} + \frac{(1+(\xi-1)(1-\varepsilon_{\mathrm{ps}}(\xi p)))Z(p) + 2(\xi-1)\frac{\partial \varepsilon_{\mathrm{ps}}(p)}{\partial p}\frac{\partial \varepsilon_{\mathrm{ps}}(\xi p)}{\partial p}\xi p T_{\mathrm{s}}}{(\varepsilon_{\mathrm{ps}}(p))^2}$. Since we have proved $Z(p) > 0$ holds for all $p \leq N$ in the proof of Proposition 3. With $\frac{\partial \varepsilon_{\mathrm{ps}}(p)}{\partial p} < 0$ as shown in (38), we can obtain $\frac{\partial^2 \bar{\Delta}_{\mathrm{ub}}^{\mathrm{PAC}}(p,\xi p,\alpha)}{\partial p^2} > 0$.

The second derivative of $\bar{\Delta}_{\mathrm{ub}}^{\mathrm{PAC}}(p, \xi p, \alpha)$ with respect to $\xi$ is given by $\frac{\partial^2 \bar{\Delta}_{\mathrm{ub}}^{\mathrm{PAC}}(p,\xi p,\alpha)}{\partial \xi^2} = \frac{p^3 T_{\mathrm{s}} Y(\xi p)}{\varepsilon_{\mathrm{ps}}(p)}$, with $Y(\xi p) = -p(\xi-1)\frac{\partial^2 \varepsilon_{\mathrm{ps}}(\xi p)}{\partial p^2} - 2\frac{\partial \varepsilon_{\mathrm{ps}}(\xi p)}{\partial p}$. With derived $\frac{\partial \varepsilon_{\mathrm{ps}}(p)}{\partial p}$ and $\frac{\partial^2 \varepsilon_{\mathrm{ps}}(p)}{\partial p^2}$ in (38) and (45), we can obtain $\frac{\partial \varepsilon_{\mathrm{ps}}(\xi p)}{\partial p}$ and $\frac{\partial^2 \varepsilon_{\mathrm{ps}}(\xi p)}{\partial p^2}$ replacing $p$ with $\xi p$. Since $\frac{\partial \varepsilon_{\mathrm{ps}}(\xi p)}{\partial p} < 0$, we have $Y(\xi p) = -\frac{1}{2}\frac{\partial \varepsilon_{\mathrm{ps}}(\xi p)}{\partial p}\left(p(\xi-1)(\frac{\delta}{\sigma_{\mathrm{w}}\xi p})^2 + \frac{\xi+3}{\xi}\right) > 0$ if $a^2 = 1$, and $Y(\xi p) = -\frac{1}{2}\frac{\partial \varepsilon_{\mathrm{ps}}(\xi p)}{\partial p}(\frac{(\xi-1)(2+a^{2\xi p})\ln a^{2p}}{1-a^{2\xi p}} - \frac{(\xi-1)(1-a^{2\xi p})p(\Gamma(\xi p))^2}{a^{2\xi p}\ln a^2} + 4) > -\frac{1}{2}\frac{\partial \varepsilon_{\mathrm{ps}}(\xi p)}{\partial p}(\xi + 3 - \frac{p(\xi-1)(1-a^{2\xi p})(\Gamma(\xi p))^2}{a^{2\xi p}\ln a^2}) > 0$ if $a^2 \neq 1$. As a consequence, $\frac{\partial^2 \bar{\Delta}_{\mathrm{ub}}^{\mathrm{PAC}}(p,\xi p,\alpha)}{\partial \xi^2} > 0$. Moreover, the constraint (C3) is an affine set. Thus, **Proposition 4** is proved.

## REFERENCES


[1] Q. Xiong, X. Zhu, Y. Jiang, J. Cao, and Y. Wang, "Status prediction for age of information oriented short-packet transmission in industrial IoT," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Austin, TX, USA, Apr. 2022, pp. 794–799.

[2] M. A. Abd-Elmagid, N. Pappas, and H. S. Dhillon, "On the role of age of information in the Internet of Things," *IEEE Commun. Mag.*, vol. 57, no. 12, pp. 72–77, Dec. 2019.

[3] S. Kaul, R. Yates, and M. Gruteser, "Real-time status: How often should one update?" in *Proc. IEEE INFOCOM*, Orlando, FL, USA, Mar. 2012, pp. 2731–2735.

[4] *Service Requirements for Cyber—Physical Control Applications in Vertical Domains; Stage-1 (Release 17)*, 3GPP, document TS 22.104, Jun. 2019.

[5] G. Durisi, T. Koch, and P. Popovski, "Toward massive, ultrareliable, and low-latency wireless communication with short packets," *Proc. IEEE*, vol. 104, no. 9, pp. 1711–1726, Aug. 2016.

[6] B. Holfeld et al., "Wireless communication for factory automation: An opportunity for LTE and 5G systems," *IEEE Commun. Mag.*, vol. 54, no. 6, pp. 36–43, Jun. 2016.

[7] Y. Polyanskiy, H. V. Poor, and S. Verdú, "Channel coding rate in the finite blocklength regime," *IEEE Trans. Inf. Theory*, vol. 56, no. 5, pp. 2307–2359, May 2010.

[8] A. Jarwan, A. Sabbah, and M. Ibnkahla, "Data transmission reduction schemes in WSNs for efficient IoT systems," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1307–1324, Mar. 2019.

[9] B. Yu, Y. Cai, D. Wu, and Z. Xiang, "Average age of information in short packet based machine type communication," *IEEE Trans. Veh. Technol.*, vol. 69, no. 9, pp. 10306–10319, Sep. 2020.

[10] 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 Global Commun. Conf. (GLOBECOM)*, Waikoloa, HI, USA, Dec. 2019, pp. 1–6.

[11] J. Ostman, R. Devassy, G. Durisi, and E. Uysal, "Peak-age violation guarantees for the transmission of short packets over fading channels," in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Paris, France, Apr. 2019, pp. 109–114.

[12] R. Devassy, G. Durisi, G. C. Ferrante, O. Simeone, and E. Uysal, "Reliable transmission of short packets through queues and noisy channels under latency and peak-age violation guarantees," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 4, pp. 721–734, Apr. 2019.

[13] M. K. Abdel-Aziz, S. Samarakoon, C.-F. Liu, M. Bennis, and W. Saad, "Optimized age of information tail for ultra-reliable low-latency communications in vehicular networks," *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1911–1924, Mar. 2020.

[14] J. Cao, X. Zhu, Y. Jiang, Z. Wei, and S. Sun, "Information age-delay correlation and optimization with finite block length," *IEEE Trans. Commun.*, vol. 69, no. 11, pp. 7236–7250, Nov. 2021.

[15] X. Tong, G. Zhao, M. A. Imran, Z. Pang, and Z. Chen, "Minimizing wireless resource consumption for packetized predictive control in real-time cyber physical systems," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, Kansas City, MO, USA, May 2018, pp. 1–6.

[16] Z. Hou, C. She, Y. Li, L. Zhuo, and B. Vucetic, "Prediction and communication co-design for ultra-reliable and low-latency communications," *IEEE Trans. Wireless Commun.*, vol. 19, no. 2, pp. 1196–1209, Feb. 2020.

[17] Z. Yuan, B. Li, and J. Liu, "Can we improve information freshness with predictions in mobile crowd-learning?" in *Proc. IEEE Conf. Comput. Commun. Workshops (INFOCOM WKSHPS)*, Toronto, ON, Canada, Jul. 2020, pp. 702–709.

[18] Y. Zhang, Y. Chen, B. Yu, X. Diao, and Y. Cai, "Minimizing age of information based on predictions and short packet communications in UAV relay systems," in *Proc. 13th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Changsha, China, Oct. 2021, pp. 1–5.

[19] A. M. Girgis, J. Park, M. Bennis, and M. Debbah, "Predictive control and communication co-design via two-way Gaussian process regression and AoI-aware scheduling," *IEEE Trans. Commun.*, vol. 69, no. 10, pp. 7077–7093, Oct. 2021.

[20] J. Guo, S. Durrani, X. Zhou, and H. Yanikomeroglu, "Massive machine type communication with data aggregation and resource scheduling," *IEEE Trans. Commun.*, vol. 65, no. 9, pp. 4012–4026, Sep. 2017.

[21] Z. Sun, Z. Wei, N. Yang, and X. Zhou, "Two-tier communication for UAV-enabled massive IoT systems: Performance analysis and joint design of trajectory and resource allocation," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 4, pp. 1132–1146, Apr. 2021.

[22] T. Wang, Y. Wang, C. Wang, Z. Yang, and J. Cheng, "Group-based random access and data transmission scheme for massive MTC networks," *IEEE Trans. Commun.*, vol. 69, no. 12, pp. 8287–8303, Dec. 2021.

[23] D. Han, H. Minn, U. Tefek, and T. J. Lim, "Network dimensioning, QoE maximization, and power control for multi-tier machine-type communications," *IEEE Trans. Commun.*, vol. 67, no. 1, pp. 859–872, Jan. 2019.

[24] H. Wang, Q. Xiong, and M. Li, "Selection cooperation using packet aggregation with equal rate feedback in industrial wireless sensor networks," *IEEE Commun. Lett.*, vol. 22, no. 12, pp. 2531–2534, Dec. 2018.

[25] T. Zhu, J. Li, H. Gao, and Y. Li, "Data aggregation scheduling in battery-free wireless sensor networks," *IEEE Trans. Mobile Comput.*, vol. 17, no. 12, pp. 2835–2852, Dec. 2018.

[26] W. Wei et al., "Impact of in-network aggregation on target tracking quality under network delays," *IEEE J. Sel. Areas Commun.*, vol. 31, no. 4, pp. 808–2421, Apr. 2013.

[27] J. L. Zechinelli-Martini, P. Bucciol, and G. Vargas-Solar, "Energy aware data aggregation in wireless sensor networks," in *Proc. IEEE Wireless VITAE*, Chennai, India, Feb. 2011, pp. 1–5.

[28] H. Pan, S. C. Liew, J. Liang, V. C. M. Leung, and J. Li, "Coding of multi-source information streams with age of information requirements," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1427–1440, May 2021.

[29] 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*, vol. 63, no. 11, pp. 7492–7508, Nov. 2017.

[30] J. P. Champati, R. R. Avula, T. J. Oechtering, and J. Gross, "Minimum achievable peak age of information under service preemptions and request delay," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1365–1379, May 2021.

[31] R. V. Bhat, R. Vaze, and M. Motani, "Throughput maximization with an average age of information constraint in fading channels," *IEEE Trans. Wireless Commun.*, vol. 20, no. 1, pp. 481–494, Jan. 2021.
[32] S. M. Kay, *Fundamentals of Statistical Signal Processing*, vol. 1. Upper Saddle River, NJ, USA: Prentice-Hall, 1993.
[33] M. Cheffena, "Propagation channel characteristics of industrial wireless sensor networks," *IEEE Antennas Propag. Mag.*, vol. 58, no. 1, pp. 66–73, Feb. 2016.
[34] B. Makki, T. Svensson, and M. Zorzi, "Finite block-length analysis of spectrum sharing networks using rate adaptation," *IEEE Trans. Commun.*, vol. 63, no. 8, pp. 2823–2835, Aug. 2015.
[35] A. Genz and F. Bretz, *Computation of Multivariate Normal and T Probabilities*. 1st ed. Berlin, Germany: Springer, 2009.
[36] S. Poojary, S. Bhambay, and P. Parag, "Real-time status updates for Markov source," *IEEE Trans. Inf. Theory*, vol. 65, no. 9, pp. 5737–5749, Sep. 2019.
[37] J. Cao, M. G. Genton, D. E. Keyes, and G. M. Turkiyyah, "Hierarchical-block conditioning approximations for high-dimensional multivariate normal probabilities," *Statist. Comput.*, vol. 29, no. 3, pp. 585–598, May 2019.
[38] G. Gallego, C. Cuevas, R. Mohedano, and N. García, "On the Mahalanobis distance classification criterion for multidimensional normal distributions," *IEEE Trans. Signal Process.*, vol. 61, no. 17, pp. 4387–4396, Sep. 2013.
[39] D. Zhai, H. Li, X. Tang, R. Zhang, Z. Ding, and F. R. Yu, "Height optimization and resource allocation for NOMA enhanced UAV-aided relay networks," *IEEE Trans. Commun.*, vol. 69, no. 2, pp. 962–975, Feb. 2021.
[40] B. Yu, Y. Cai, and D. Wu, "Joint access control and resource allocation for short-packet-based mMTC in status update systems," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 3, pp. 851–865, Mar. 2021.
[41] L. Liu and W. Yu, "A D2D-based protocol for ultra-reliable wireless communications for industrial automation," *IEEE Trans. Wireless Commun.*, vol. 17, no. 8, pp. 5045–5058, Aug. 2018.
[42] R. S. Isaac and N. B. Mehta, "Efficient computation of multivariate Rayleigh and exponential distributions," *IEEE Wireless Commun. Lett.*, vol. 8, no. 2, pp. 456–459, Apr. 2019.

**Qinqin Xiong** received the B.S. degree in electrical engineering and the M.S. degree in control theory and engineering from the Chongqing University of Posts and Telecommunications, Chongqing, China, in 2016 and 2019, respectively. She is currently pursuing the Ph.D. degree with the School of Electrical and information Engineering, Harbin Institute of Technology, Shenzhen, China. Her current research interests include age of information, short packet communication, and ultra reliable low latency communication.

**Xu Zhu** (Senior Member, IEEE) received the B.Eng. degree (Hons.) in electronics and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 1999, and the Ph.D. degree in electrical and electronic engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2003. She was a Reader at the Department of Electrical Engineering and Electronics, University of Liverpool, Liverpool, U.K. She is currently a Professor with the School of Electronic and Information Engineering, Harbin Institute of Technology, Shenzhen, China. She has more than 210 peer-reviewed publications on communications and signal processing. Her research interests include MIMO, channel estimation and equalization, ultra reliable low latency communication, resource allocation, and cooperative communication. She was a recipient of the Best Paper Award of the IEEE GLOBECOM 2019. She is currently a 2022 Distinguished Lecturer of the IEEE Vehicular Technology Society. She has acted as the chair for various international conferences, such as the Vice-Chair of the 2006 and 2008 ICARN International Workshops, the Program Chair of ICSAI 2012, the Symposium Co-Chair of IEEE ICC 2016, IEEE ICC 2019, and IEEE GLOBECOM 2021, and the Track Co-Chair of IEEE WCNC 2022. She has served as an Editor for the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and a Guest Editor for several international journals such as *Electronics*.

**Yufei Jiang** (Member, IEEE) received the Ph.D. degree in electrical engineering and electronics from the University of Liverpool, Liverpool, U.K., in 2014. From 2014 to 2015, he was the Post-Doctoral Researcher at the Department of Electrical Engineering and Electronics, University of Liverpool. From 2015 to 2017, he was a Research Associate at the Institute for Digital Communications, University of Edinburgh, Edinburgh, U.K. He is currently an Associate Professor with the Harbin Institute of Technology, Shenzhen, China. His research interests include Li-Fi, synchronization, full-duplex, age of information, and blind source separation.

**Jie Cao** (Member, IEEE) received the B.S. degree from the Harbin Institute of Technology, Weihai, China, in 2016, the M.S. degree from the Harbin Institute of Technology, Harbin, China, in 2018, and the Ph.D. degree from the Harbin Institute of Technology, Shenzhen, China, in 2022. He is currently a Research Scientist with the Institute for Infocomm Research (I2R), Singapore. His current research interests include ultrareliable and low-latency communication, short packet communication, age of information, and task-oriented communication.

**Xiaogang Xiong** (Member, IEEE) received the Ph.D. (Eng.) degree in mechanical and science engineering from Kyushu University, Fukuoka, Japan, in 2014. From 2014 to 2015, he was a Researcher at the Kyushu Institute of Technology, Iizuka, Japan. From 2015 to 2016, he was a Research Fellow with the Singapore Institute of Manufacturing Technology, Singapore. He is currently with the Department of Mechanical Engineering and Automation, Harbin Institute of Technology, Shenzhen, China. His research interests include human–robot coordination, non-smooth systems, and robot real-time control.

**Heng Wang** (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in communication engineering from Chongqing University, Chongqing, China, in 2003, 2006, and 2010, respectively. He is currently a Full Professor with the Chongqing University of Posts and Telecommunications, Chongqing. His current research interests include Industrial Internet of Things, wireless sensor networks, clock synchronization, and real-time scheduling.