2D-AoI: Age-of-Information of Distributed Sensors for Spatio-Temporal Processes

Markus Fidler, Flavio Gallistl, Jaya Prakash Champati, Joerg Widmer

***Abstract*—The freshness of sensor data is critical for all types of cyber-physical systems. An established measure for quantifying data freshness is the Age-of-Information (AoI), which has been the subject of extensive research. Recently, there has been increased interest in multi-sensor systems: redundant sensors producing samples of the same physical process, sensors such as cameras producing overlapping views, or distributed sensors producing correlated samples. When the information from a particular sensor is outdated, fresh samples from other correlated sensors can be helpful. To quantify the utility of distant but correlated samples, we put forth a two-dimensional (2D) model of AoI that takes into account the sensor distance in an age-equivalent representation. Since we define 2D-AoI as equivalent to AoI, it can be readily linked to existing AoI research, especially on parallel systems. We consider physical phenomena modeled as spatio-temporal processes and derive the 2D-AoI for different Gaussian covariance kernels. For a basic exponential product kernel, we find that spatial distance causes an additive offset of the AoI, while for other kernels the effects of spatial distance are more complex and vary with time. Using our methodology, we evaluate the 2D-AoI of different spatial topologies and sensor densities.**

***Index Terms*—Age-of-information, information freshness, spatio-temporal processes, correlated sources, Gaussian process regression**

I. Introduction

The Internet of Things (IoT) aims to achieve seamless integration between the physical and digital worlds. Physical devices, equipped with sensors and network connectivity, exchange information that is collected and processed in cyberspace. This information is used to gather data about physical phenomena, make decisions, augment reality, or control devices in the physical world through actuators. Examples of such cyber-physical systems include environmental monitoring, intelligent transportation, augmented reality, robotics, and networked feedback control. Common to these is the need to maintain fresh information about a physical phenomenon. An accepted measure for quantifying information freshness is the Age-of-Information (AoI).

The concept of AoI was initially defined in the field of vehicular communications [1] and motivated the investigation of AoI in queueing systems [2]. Today, the AoI is known for a variety of system models, see the recent surveys [3], [4], including wireless channels [5], [6], wireless networks [7], random access channels [8]–[10], queueing systems [11], [12],

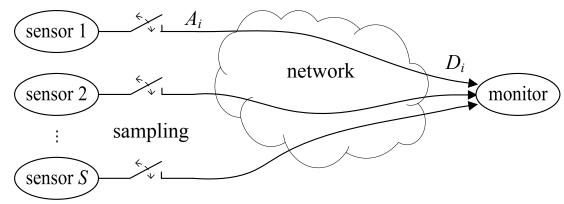


Fig. 1. System with S networked sensors that observe a physical phenomenon. The sensors generate samples, indexed by i, at times A_i. The samples are transmitted via a network and received by the monitor at times D_i.

and queueing networks [13]. While the focus has been on mean AoI and mean peak AoI, there are a number of papers that derive the CCDF [12], [14]–[16] or tail bounds of the AoI [17], [18].

The classic configuration studied in AoI involves a single sensor that observes a physical phenomenon. The sensor takes samples at random times and sends them as packets over a network to a monitor. During transmission, packets can be delayed or even lost. As a result, the information available to the monitor has a certain age, which is quantified by the AoI. This system is a special case, $S = 1$, of the general S sensor system shown in Fig. 1.

In many instances of practical significance, however, there are multiple sensors, $S > 1$, as illustrated in Fig. 1. Sensors can be spatially distributed, redundant, or of different types, and can be connected in a variety of ways, including independent network paths or a shared wireless channel. Environmental monitoring involves many spatially distributed sensors that generate correlated information and can be connected via LoRaWAN or 5G mMTC, for example. Modern vehicles are equipped with automotive Ethernet and an array of sensors, including cameras, lidar, radar, and ultrasound, that observe the environment and create overlapping or correlated views. Vehicles share sensor information over Wifi to create a collective perception, to name a few examples. Despite this, the effects of the AoI when multiple statistically correlated sensors are available have only recently begun to be explored [19]–[23], and so far there is no mathematical derivation that incorporates spatial correlation into the AoI.

In this study, we examine physical phenomena represented as correlated spatio-temporal processes sampled by a number of distributed sensors S. The spatial dimension can be Euclidean space, although this is not a necessary condition. The same methodology can be applied to include sensors that observe different but correlated physical phenomena, such as road traffic flow and travel time or air quality. In general, the spatial and temporal sampling densities are constrained by the

M. Fidler and F. Gallistl are with the Department of Electrical Engineering and Computer Science, Leibniz University Hannover. J. P. Champati is with the University of Victoria. J. Widmer is with the IMDEA Networks Institute, Madrid. This work was supported in part by the German Research Foundation (DFG) under grant FI 1236/9-1.

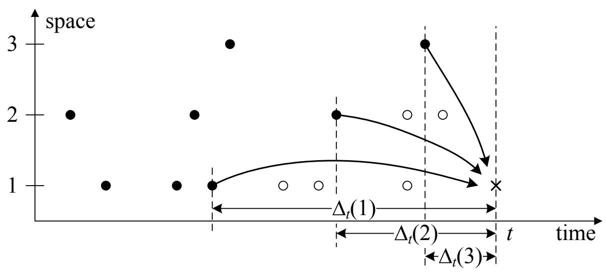


Fig. 2. Example timeline. The times A_i at which sensors take samples are indicated by circles. The circle is filled, if the sample is available at the monitor at time t, i.e., if the departure time from the network D_i satisfies $D_i \leq t$. The most recent sample of sensor s has AoI $\Delta_t(s)$. Assume the physical process of sensor 1 and time t is of interest. A stale sample of sensor 1 with AoI $\Delta_t(1)$ can be used, or a prediction can be made using fresher samples from remote sensors 2 or 3.

limited capacity of the network. This raises the question of which sensor should be sampled and when. A related question is whether it is preferable to use outdated information obtained at a specific location or to utilize the most recent data from a different location. Fig. 2 illustrates the example. To make progress on these issues, it is necessary to formalize the trade-off between AoI and the spatial distance of sensors.

To include the spatial dimension into the temporal analysis of AoI, we define a two-dimensional (2D) model of AoI and derive an age-equivalent representation of spatial distance. The distance measure can be directly incorporated into the AoI, providing a strong link to the existing AoI literature and a foundation for further investigation. In particular, our work achieves a mapping of the AoI of spatio-temporal sampling to a suitably adapted model of AoI in parallel systems. The spatio-temporal process model we start from is generic and can be defined in different ways depending on the specific analysis being performed. In the course of this paper we use Gaussian processes, with distinct covariance kernels, but other models are also possible. An important finding of our work is that spatial distance can have very different effects depending on the type of covariance kernel. In the basic case of exponential product kernels, the spatial distance between the sensors leads to an additive shift of the AoI, that is independent of time. In contrast, in the case of squared exponential and rational quadratic kernels, the impact of spatial distance on the AoI is observed to diminish over time.

By employing our model of 2D-AoI, we demonstrate the potential benefits and limitations of distributed sensors. In our analysis, we consider different network models, including both lossy ALOHA multi-access links and lossless M/M/1 queues. We evaluate the impact of spatial topology and sensor density on the AoI performance. We use our method to show the trade-off between refining the temporal or spatial resolution and which is most advantageous under a capacity constraint. We believe that our methodology is also suitable for studying other relevant systems.

The main contributions of our work are summarized below:

1) We extend the temporal concept of AoI by a spatial dimension to include multiple, distributed sensors. Compared to related works, the key difference is that we derive an age-equivalent representation of distance, which we integrate into the standard model of AoI. This allows us to effectively extend AoI results that exist for a variety of different channel and queueing models to distributed sensors.
2) We use Gaussian processes to model temporal and spatial correlation and derive how sensor distance affects the AoI. We discover that the effect of distance shows fundamental differences, depending on the type of covariance kernel. This finding goes beyond state-of-the-art studies that use an a priori assumption of a linear attenuation factor of the AoI to model distance.
3) We validate our results using a real freeway traffic scenario by training a kernel function with traffic counts from distributed loop detectors.
4) We evaluate the effect of the spatial density of sensors on the AoI. Given a communication capacity constraint, we weigh the spatial and temporal sampling intensities against each other to minimize the AoI.

The remainder of this work is structured as follows. We start with a discussion of related works in Sec. II. In Sec. III we define the notion of 2D-AoI. We show how the 2D-AoI can be evaluated for a network model with independent channels in Sec. IV. In Sec. V, we give a brief background on Gaussian processes to characterize spatio-temporal physical phenomena and we use these to instantiate the 2D-AoI model. Sec. VI considers relevant covariance kernels and their impact on the AoI. We show an application of Gaussian processes to real-world road traffic data in Sec. VII. Evaluation results are presented in Sec. VIII, where the spatial density of the sensors is one of the key factors taken into account. Brief conclusions are provided in Sec. IX. Further results and a non-Gaussian model of 2D-AoI can be found in the technical report [24].

II. Related Work

The freshness of information is of general interest in various areas, including vehicular networks [1], [2], remote state estimation [23], [25], [26], networked feedback control [27]–[29], and cyber-physical systems in general, databases, caches, to micro-blogging [30]. The work [1] on the AoI of status information broadcast by vehicles in a vehicular network marks the beginning of comprehensive research on the topic of AoI. Subsequently, the AoI of various queueing models was investigated [2], and today a catalog of results for various systems is available, see e.g., [12] and the surveys [3], [4].

1) Definition of AoI: For a basic definition of AoI, consider a sensor that is connected to a monitor via a network, as shown in Fig. 1 for the special case $S = 1$. The sensor generates samples $i \geq 1$ at times A_i and sends them to the network, which delivers them to the monitor at times D_i. The AoI at time $t \geq D_1$ is defined as

$$\Delta_t = t - \max_{i \geq 1}\{A_i : D_i \leq t\}. \tag{1}$$

The AoI exhibits a distinctive piecewise linear sawtooth shape. Upon the arrival of a new sample at the monitor, the AoI drops to the delay of that sample, that is $\Delta_t = D_i - A_i$ if $t = D_i$. Subsequently, the AoI increases at unit rate with time, until a fresher sample is received.

2) Value-of-Information: Different from the linear aging assumed by the definition of AoI (1), there are also systems that exhibit nonlinear behavior as the information ages. Important examples arise in remote state estimation of a system or a process [25]–[29]. The difference to (1) is often formalized in terms of nonlinear functions of the AoI [3]. An age-penalty is defined in [31] as a non-decreasing function of the AoI to model the dissatisfaction caused by stale information. The concept of the age of incorrect information [32] uses a product of two functions, an age-penalty and an error-penalty, where the latter evaluates the deviation of the monitor's estimate from the correct state of the physical process. Mirroring age-penalty functions, the non-increasing value of aging information is modeled as value-of-information in [33].

Nonlinear functions of the AoI have been used in networked feedback control systems to characterize the mean squared error that arises in remote state estimation of a linear system with Gaussian disturbances [27]–[29]. Related to this are the works [25], [26] on remote estimation of the state of Gaussian processes. Gaussian processes are also used to model spatial phenomena. In this case, the process at one location of interest can be predicted from another sensor location. The prediction variance is used in [34] to define the value-of-information of that sensor.

3) Overlapping Sensors with Common Observations: In recent research [19]–[23], [35]–[51], there is a growing interest in the AoI of multi-sensor systems. The works [37]–[39], [41]–[45], [50], [51] are based on a model of overlapping sensors that can generate common observations. They differ in their assumptions regarding the overlap of the sensors. In [35], [36] multiple sensors that generate independent information flows are considered. The flows share a single queue, and thus influence each other's AoI. Multiple sensor sources that share a single queue are also examined in [50], [51], with the difference that a fraction of the updates of each source are also valid updates for the other sources.

The AoI of systems in which multiple sensors monitor the same physical phenomenon but transmit their samples via independent channels is analyzed in [38], [39]. In [38], an energy harvesting system is investigated that can choose between multiple sensors, each of which offers an individual trade-off between AoI and energy. A dual-updating system, where two sensors generate and transmit samples independently, is considered in [39], and the AoI is derived for Markovian and deterministic queues with feedback. Directly related to multi-sensor systems that observe a single physical phenomenon are parallel systems in which the samples from a single sensor are split between two (or more) independent channels [6], [52]. As only the most recent sample is decisive for the AoI, the main characteristic of parallel systems is that their AoI is the minimum of the AoI of the individual subsystems. In contrast, [40] studies a system with S sensors with the constraint that a sample from each of the sensors is required to reconstruct the physical process. In this system, the AoI is the maximum of the AoI of the individual sensors.

The AoI of systems with overlapping sensors that generate common observations is studied in [37], [41]–[45]. In [41] a multi-view camera system is considered, where S cameras observe K different scenes. A single camera can observe multiple scenes and the cameras' fields of view overlap so that scenes are observed by multiple cameras simultaneously. A similar model of sensors with overlapping monitoring areas is used in [45] and in [37], where K information sources are sensed by S servers. A related model with non-deterministic sensors, where sensor $s \in \{1, 2, \ldots, S\}$ detects physical phenomenon $k \in \{1, 2, \ldots, K\}$ with probability p_{sk}, is studied in [42], [43] and also in [44] which models the coupling of the AoI processes of the sources.

4) Correlated Sensors: More closely related to our work are the recent papers [19]–[22] that use Gaussian processes to model the spatial and temporal correlations of multiple distributed sensors. A common feature of these papers is that the monitor uses only the most recent sample from each sensor. This is consistent with the definition of AoI. The AoI is then used as input to a nonlinear function that expresses an age-penalty or value-of-information: [19] employs the estimation error resulting from using aged information from a possibly remote sensor, similarly [21], [22] apply the mean squared estimation error, and [20] uses the mutual information that applies over a spatial span and temporal lag. A related work is also [23], which uses Kalman filters with inputs from multiple sensors to estimate the state of several random walks with spatially correlated Gaussian increments. Unlike [19]–[22], our consideration of spatially distributed sensors takes place in the AoI domain, i.e., we transform spatial distance into a representation that is equivalent to AoI. This provides a direct link to the AoI literature, where our result generalizes the AoI of parallel systems [6], [52] by considering correlated spatio-temporal processes instead of a single temporal process.

In more detail [19] considers two sensor nodes, one at the point of interest and one distant. The sensors generate samples periodically with update period T and a defined offset τ. The AoI is deterministic in $[0, T)$ meaning there are no network delays or losses. Regarding the estimation error, a key insight is that there is a break-even point at which a fresh sample from the far sensor is better than an outdated sample from the near sensor. This imposes a deterministic condition on the update period and distance for the far sensor to take effect. In our work, we show that similar conditions can be established in the case of random network delays as they occur, e.g., in multi-access channels or queueing networks.

The papers [20], [21] differ from our work in that they consider the optimal scheduling of S sensors connected via a satellite link with deterministic delay or a wireless sensor network with zero delay, respectively. For a random topology where sensors are deployed as a Poisson point process, [22] investigates how the density of exponential temporal and spatial sampling affects the estimation error. Analytical solutions are provided for a one-dimensional sensor network and exponential service times. Among other things, we also evaluate the density of a sensor network, but our results are expressed in terms of AoI.

5) Modeling Spatial Distance in the AoI Domain: The works [45]–[47], like our work, differ from [19]–[22], by modeling the spatial distance of the sensors in the AoI domain: [45] multiplies the AoI of a sensor by an attenuation factor

$a \in [0, 1]$ when a more recent sample from another correlated sensor becomes available. Similarly, [47] defines a hyper-parameter a as a weight of the AoI of a neighbor having fresher information, and [46] uses a weighting factor a of the AoI, which is determined from a spatial model of a vehicular network consisting of spatial distance and angle of vehicles. Different from these works, we do not make the a priori assumption that spatial distance can be expressed by a constant attenuation or weighting factor of the AoI. Instead, we start with a defined correlation model and derive a function that characterizes the influence of spatial distance on the AoI. To parameterize the function, we evaluate several spatio-temporal covariance kernels. Stationary exponential product kernels, which are also used in [19], [20], [22], have the basic effect of causing a constant additive offset of the AoI of distant sensors. For squared exponential and rational quadratic kernels, on the other hand, we find that the influence of spatial distance on the AoI diminishes as the AoI increases. These behaviors, demonstrated through the use of different kernel functions, are fundamentally different from those assumed in related studies. In practice, once we have estimated the covariance kernel for given real-world data, we can develop strategies to optimize the AoI for this scenario. We demonstrate this for the case of road traffic monitoring.

III. Definition of 2D-AoI

We consider a set of sensors $\mathbb{S} = \{1, 2, \dots, S\}$. In a basic scenario, sensors are homogeneous and differ only by their position x_s where $s \in \mathbb{S}$ is the index of the sensor. The position x_s can be scalar, i.e. lie on a one-dimensional line, but it can also lie in two- or three-dimensional Euclidean space. To simplify notation, we generally refer to sensors by their index. In important cases, which we will analyze below, only the sensor position x_s referred to by index s is relevant. However, other parameters such as different types of sensors can also be included, resulting in a higher dimensional model.

The sensors sample the physical process and transmit the samples to a monitor. The samples may be subject to noise, which we will consider in Sec. V-D. During transmission, samples may be delayed or may even get lost. We denote samples that are available at the monitor at time $t \geq 0$ by indicator variables $\Gamma_t(s, \tau) \in \{0, 1\}$, where $s \in \mathbb{S}$ and $\tau \in [0, t]$ is the generation time of the sample. Time can be continuous or discrete. The monitor has a time horizon T that can be unlimited $T = \infty$ or limited $T < \infty$, in which case samples that are older than T are discarded. The AoI of sensor $s \in \mathbb{S}$ at time $t \geq 0$ is

$$\Delta_t(s) = \inf\{\tau \in [0, \min\{t, T\}] : \Gamma_t(s, t - \tau) = 1\}. \quad (2)$$

If no sample of sensor s is available at the monitor at time t, we define $\Delta_t(s) = \infty$, that is also obtained from (2) as the result of taking the infimum of the empty set.

The AoI $\Delta_t(s)$ can be thought of as a penalty when predicting the physical process at position x_s and time t from the latest sample that is available for that position at the monitor. This is the sample generated at time $\tau = t - \Delta_t(s)$. The actual quality of the prediction depends on the temporal correlation.

If the point of interest is x_s, but we have a sensor at position $x_\varsigma \neq x_s$ (instead of or in addition to the sensor at position x_s), the physical process at position x_s and time t can also be predicted from a sample of the sensor at position x_ς. The latest sample of that sensor was generated at time $\tau = t - \Delta_t(\varsigma)$, see the example in Fig. 2. Now, the quality of the prediction depends on the temporal correlation and on the spatial correlation of the process at positions x_ς and x_s. This motivates a notion of two-dimensional (2D) AoI, expressed by the function $\Delta_t^{\text{2D}}(\varsigma, s)$. The 2D-AoI can be thought of as a spatio-temporal distance of a sample. In the special case $\varsigma = s$ we have $\Delta_t^{\text{2D}}(\varsigma, \varsigma) = \Delta_t(\varsigma)$, while for $\varsigma \neq s$ in the general case $\Delta_t^{\text{2D}}(\varsigma, s)$ is different from $\Delta_t(\varsigma)$.

The difference $\Lambda_t(\varsigma, s) = \Delta_t^{\text{2D}}(\varsigma, s) - \Delta_t(\varsigma)$ represents the effect of the distance of the sensor at position x_ς on the AoI at position x_s. We call $\Lambda_t(\varsigma, s)$ the Age-equivalent (of the) Distance (AeD). In the simplest case, a distant sensor perceives a physical phenomenon delayed by a deterministic time $\Lambda_t(\varsigma, s)$. This is the case with seismic waves, for example, which travel at a few kilometres per second and are detected by distributed sensors. How to select $\Lambda_t(\varsigma, s)$ in non-trivial cases based on statistical correlation is shown in the following Secs. V and VI. In some relevant cases, we find that the AeD $\Lambda_t(\varsigma, s) = \Lambda(\varsigma, s)$ is only a function of space and independent of time, while in others it is a more complex function that changes with time.

Using the definition of AeD, the 2D-AoI is expressed as a sum of AeD and AoI as

$$\Delta_t^{\text{2D}}(\varsigma, s) = \Lambda_t(\varsigma, s) + \Delta_t(\varsigma). \quad (3)$$

Given the monitor has samples from several sensors $S > 1$, an evident choice is to select the sample that has the minimal 2D-AoI

$$\Delta_t^{\text{2D}}(s) = \min_{\varsigma \in \mathbb{S}} \{\Delta_t^{\text{2D}}(\varsigma, s)\}. \quad (4)$$

To keep the notation simple, we use Δ_t^{2D} again, but with a single argument. In the case that the AeD is just a function of space, it simply contributes as a time-independent offset, and we have

$$\Delta_t^{\text{2D}}(s) = \min_{\varsigma \in \mathbb{S}} \{\Lambda(\varsigma, s) + \Delta_t(\varsigma)\}. \quad (5)$$

Equation (5) permits a very intuitive visualization, which is shown in Fig. 3. In the example, there are $S = 2$ sensors and the minimal 2D-AoI $\Delta_t^{\text{2D}}(1)$ for sensor 1 is derived. A continuous time model is used. As before, A_i denotes the time of generation of sample i, that is the arrival time of the sample to the network, and D_i the time of its departure from the network to the monitor. Samples 2 and 4 are taken by sensor 1 and samples 1, 3, and 5 by sensor 2. The thin green saw-tooth function represent $\Delta_t^{\text{2D}}(1, 1) = \Delta_t(1)$, that is the AoI of sensor 1. The thin blue saw-tooth represents $\Delta_t^{\text{2D}}(2, 1) = \Lambda(2, 1) + \Delta_t(2)$, that is the AoI of sensor 2 shifted upwards by an offset that considers the AeD of sensor 2 with respect to sensor 1. At the monitor the sample that has the minimal 2D-AoI is selected, marked by the thick saw-tooth function that is either green, if the most recent sample of sensor 1 is used, or blue, if the most recent sample of sensor 2 is used.

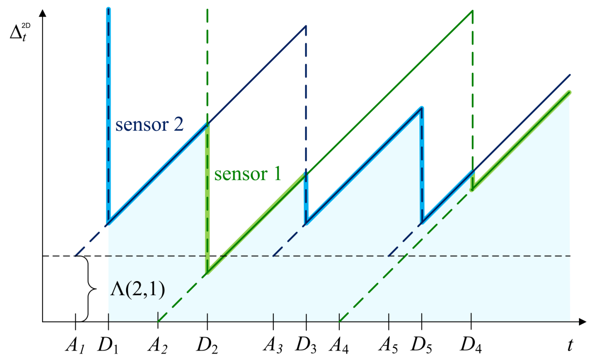


Fig. 3. Example of the evolution of the 2D-AoI $\Delta_t^{2D}(1)$ for $S = 2$ sensors.

It is evident that the utility of sensor 2 depends on the condition that the AoI of sensor 1 exceeds the AeD of sensor 2. In a deterministic network with constant delay, it is straightforward to devise a sampling schedule that adheres to the specified condition. This result is consistent with the identification of a break-even point in Gaussian process estimation from two sensors and deterministic sampling as detailed in [19], see the discussion in Sec. II. In the following, we derive statistics on the 2D-AoI under random sampling and random transmission delays.

IV. 2D-AoI for Independent Channels

We consider a network that connects each sensor $\varsigma \in \mathbb{S}$ to the monitor via an independent channel. In this case, the distribution of the minimal 2D-AoI from (4) is

$$\begin{aligned}\mathsf{P}[\Delta_t^{2D}(s) > y] &= \mathsf{P}\Big[\min_{\varsigma\in\mathbb{S}}\{\Delta_t^{2D}(\varsigma, s)\} > y\Big] \\ &= \prod_{\varsigma\in\mathbb{S}} \mathsf{P}[\Delta_t^{2D}(\varsigma, s) > y] \qquad (6)\\ &= \prod_{\varsigma\in\mathbb{S}} \mathsf{P}[\Delta_t(\varsigma) > y - \Lambda(\varsigma, s)]. \qquad (7)\end{aligned}$$

In (7) we have substituted (3) where we consider the AeD $\Lambda(\varsigma, s)$ as a deterministic, time-independent function of space and $\Delta_t(\varsigma)$ is the random AoI process. Note that the AeD can also be modeled as random, in which case (7) evaluates the sum of two random variables $\mathsf{P}[\Delta_t(\varsigma) + \Lambda(\varsigma, s) > y]$. For the special case $\Lambda(\varsigma, s) = 0$ for all $\varsigma, s \in \mathbb{S}$, (7) recovers the AoI of a parallel system, in which independent samples of the same physical process are transmitted to a monitor via parallel transmission channels. Parallel systems have previously been studied in [6], [52]. To evaluate (7), we need the CCDF of the AoI, as given in [12] for the M/M/1 queue, for example.

In the following, we investigate spatio-temporal processes where $\Lambda(\varsigma, s) > 0$ for $\varsigma \neq s$. We consider the stationary distribution of $\Delta_t^{2D}(s)$ and omit the subscript t when it is nonambiguous. The expected value of the 2D-AoI is obtained by integration of the tail distribution as

$$\mathsf{E}[\Delta^{2D}(s)] = \int_0^\infty \mathsf{P}[\Delta^{2D}(s) > y]dy. \qquad (8)$$

For a first evaluation, we consider a set of spatially distributed sensors that are each connected to the monitor via an independent M/M/1 queue with unlimited buffer space and FCFS policy. The CCDF of the AoI of the M/M/1 queue is [12], [14]

$$\mathsf{P}[\Delta > y] = e^{-(1-\rho)\mu y} - \left(\frac{1}{1-\rho} + \rho\mu y\right)e^{-\mu y} + \frac{1}{1-\rho}e^{-\lambda y}, \qquad (9)$$

where the time parameter $y \geq 0$ is continuous, $\lambda > 0$ is the arrival rate, $\mu > 0$ is the service rate, with $\mu > \lambda$ for stability, and $\rho = \lambda/\mu$ is the utilization. The expected value of the AoI of the M/M/1 queue is $\mathsf{E}[\Delta] = (1 + 1/\rho + \rho^2/(1-\rho))/\mu$. It is minimized for $\rho \approx 0.53$ [2]. Regarding the CCDF, there is no single value of ρ that minimizes (9) for all $y \geq 0$. However, $\rho = 0.53$ is a good choice, as it is generally close to optimal.

By insertion of (9) into (7), we obtain the CCDF of the 2D-AoI $\Delta^{2D}(s)$. By integration of the tail distribution (8) the expected value $\mathsf{E}[\Delta^{2D}(s)]$ follows.

We note that the CCDF of the AoI or tail bounds of the AoI are also known for a number of other queueing systems and wireless channel models [12], [15]–[18], which can be used in the same way. We show the two cases of M/M/1/1 and M/M/1* queues below.

As the AoI is determined by the last sample available at the monitor, it is often better to discard packets that cannot be transmitted immediately than to buffer them. This makes network systems without buffers particularly interesting. The M/M/1/1 system has no buffer and if the system is busy, a new arrival is discarded. The CCDF of the AoI is [16]

$$\mathsf{P}[\Delta > y] = \frac{\lambda\mu}{\mu+\lambda}\left(\frac{\mu^2(e^{-\lambda y} - e^{-\mu y})}{\lambda(\mu-\lambda)^2} + e^{-\mu y}\left(\frac{1}{\lambda} + \frac{1}{\mu} - \frac{\lambda y}{\mu - \lambda}\right)\right),$$

for $\lambda \neq \mu$ and

$$\mathsf{P}[\Delta > y] = \frac{\lambda\mu^2}{2(\mu+\lambda)}e^{-\mu y}\left(y + \frac{2}{\mu}\right)^2,$$

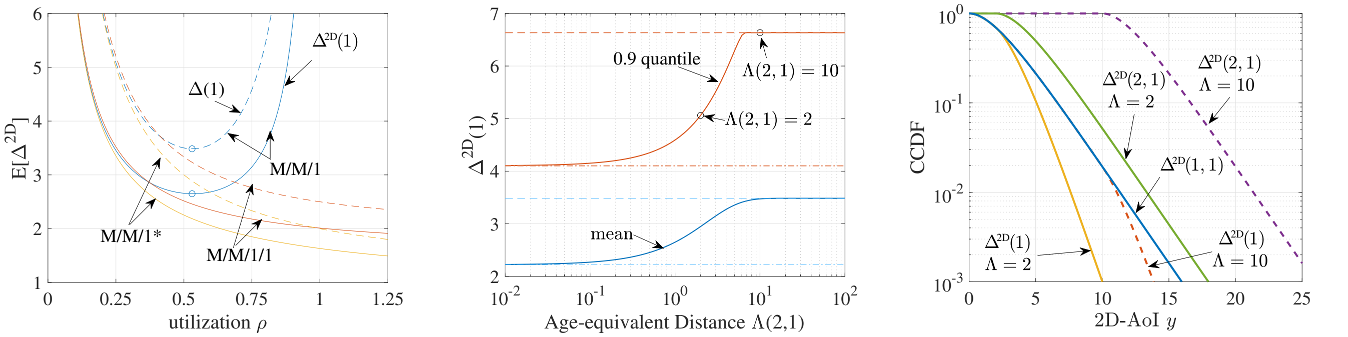
for $\lambda = \mu$.

The M/M/1* system is a variant without buffer and with preemption, in which a newly arriving packet preempts a packet that may already be in service. The CCDF of the AoI is known to be [12]

$$\mathsf{P}[\Delta > y] = \frac{\mu}{\mu-\lambda}e^{-\lambda y} - \frac{\lambda}{\mu-\lambda}e^{-\mu y},$$

for $\lambda \neq \mu$ and $\mathsf{P}[\Delta > y] = (1 + \lambda y)e^{-\lambda y}$ for $\lambda = \mu$.

In Fig. 4, we consider $S = 2$ sensors and evaluate the effect of the AeD $\Lambda(2, 1)$ on the 2D-AoI $\Delta^{2D}(1)$. In Fig. 4(a) we compare the expected value of the 2D-AoI $\Delta^{2D}(1)$ obtained in case of independent M/M/1, M/M/1/1, and M/M/1* systems. The service rate is set to $\mu = 1$ and the arrival rate λ is adjusted to evaluate the impact of the utilization. The AeD of sensor 2 is set to $\Lambda(2, 1) = 1$. The dashed lines show corresponding results that are obtained if only sensor 1 is used. It can be seen that including the more distant sensor 2 improves the 2D-AoI $\Delta^{2D}(1)$. For systems without buffer, i.e., M/M/1/1 and M/M/1*, a higher update rate λ and thus an increase in ρ generally improves the AoI. The situation is different with M/M/1 queues, where waiting times in the buffer affect the



(a) Mean

(b) Mean and 0.9 quantile for M/M/1 queues

(c) CCDF for M/M/1 queues

Fig. 4. 2D-AoI $\Delta^{2D}(\varsigma, s)$ and minimal 2D-AoI $\Delta^{2D}(s) = \min_\varsigma\{\Delta^{2D}(\varsigma, s)\}$ for two sensors $\varsigma, s \in \{1, 2\}$ that are connected to a monitor via independent M/M/1 queues. Fig. 4(a) compares the expected 2D-AoI for different types of systems. The AeD is $\Lambda(2, 1) = 1$. The dashed lines mark the case where only sensor 1 is used. Fig. 4(b) evaluates the impact of the AeD $\Lambda(2, 1)$ on $\Delta^{2D}(1)$ for the M/M/1 case. Fig. 4(c) shows the effect of two values $\Lambda(2, 1) \in \{2, 10\}$ on the CCDFs.

AoI. As the arrival rate increases, the AoI first decreases, but when the utilization surpasses 0.53, the AoI increases again.

In Fig. 4(b), we show how the mean and the 0.9 quantile of $\Delta^{2D}(1)$ increase with $\Lambda(2, 1)$ for the case of two homogeneous M/M/1 queues with parameters $\lambda = 0.53$ and $\mu = 1$. We do not show corresponding results for the case of M/M/1/1 and M/M/1* queues, since they exhibit the same trend when $\Lambda(2, 1)$ is increased. In case of $\Lambda(2, 1) = 0$, both sensors generate samples of the same process and $\Delta^{2D}(1)$ equals the AoI of a basic parallel system, in which samples of a single sensor are transmitted in parallel via two independent M/M/1 queues [52]. For $\Lambda(2, 1) \to \infty$ the samples of sensor 2 become irrelevant and $\Delta^{2D}(1)$ approaches the AoI of a single M/M/1 queue. The cases $\Lambda(2, 1) = 0$ and $\Lambda(2, 1) \to \infty$ mark lower and upper bounds, which are indicated in the figure by dashed dotted lines and dashed lines, respectively. In between, $\Lambda(2, 1)$ has a noticeable effect. The extent is determined by $\Lambda(2, 1)$ and by the speed of the tail decay of $\Delta_t(\varsigma)$ as given by (9).

We explain the effect in Fig. 4(c) where CCDFs for the cases $\Lambda(2, 1) = 2$ and $\Lambda(2, 1) = 10$, marked by circles in Fig. 4(b), are compared. For $\Lambda(2, 1) = 2$ (solid lines), the CCDF of $\Delta^{2D}(2, 1)$ decays from 1.0 starting at $y = 2$, where the CCDF of $\Delta^{2D}(1, 1)$ has already dropped to approximately 0.6. At this point, sensor 2 begins to contribute and the CCDF of the minimal 2D-AoI $\Delta^{2D}(1)$ decays faster than $\Delta^{2D}(1, 1)$ from that point onward. For $\Lambda(2, 1) = 10$ (dashed lines) the same effect occurs. However, the CCDF of $\Delta^{2D}(1, 1)$ has already dropped to about 0.02 for $y = 10$ before sensor 2 contributes, so that the benefit of sensor 2 is almost negligible.

Essentially, the utility of sensor 2 depends on how far the CCDF of the AoI of sensor 1 has dropped before the AeD of sensor 2 is compensated and sensor 2 becomes useful. We note that the speed of the tail decay of the CCDFs depends on the arrival and service rates λ and μ of the M/M/1 queues, where we used $\lambda = 0.53\,\mu$, see the exponents in (9).

V. Gaussian Spatio-Temporal Process Model

We defined 2D-AoI $\Delta^{2D}(\varsigma, s)$ in Sec. III as an abstract and generic concept that can be tailored to different applications. In this section, we show a way in which Gaussian processes can be used to instantiate the 2D-AoI. The use of Gaussian processes also makes a connection of AoI with regression and machine learning. Depending on the application, alternative process models and metrics can also be useful for specifying the 2D-AoI.

The physical phenomenon that we consider is a spatio-temporal process that is modeled as a two-dimensional Gaussian process. Additional dimensions, e.g., the correlation of different types of sensors can be included as well. We use the posterior variance of the Gaussian process model to establish a link between space and time and to derive the Age-equivalent of the Distance (AeD) of a sensor. The posterior variance is associated with the AoI when the most recent sample of a sensor that is available at the monitor is used for prediction. In the case of spatio-temporal processes (or general higher-dimensional Gaussian processes), we use the posterior variance resulting from the most recent sample of a distant sensor to derive the 2D-AoI as the equivalent age (in terms of posterior variance) of a sample at the position of interest.

Before deriving the 2D-AoI, we provide a brief overview of Gaussian process regression. A comprehensive treatment is [53]. For an introductory lecture see, e.g., [54]. We present an application of Gaussian processes to a real-world dataset in Sec. VII.

A. Gaussian Process Regression

The following Gaussian regression model is commonly used for prediction in machine learning [53].

1) Process Definition: We consider a Gaussian process $f(\mathbf{x})$ that takes an input vector $\mathbf{x}$ to a function value f. The Gaussian process is completely specified by its mean function $\mu(\mathbf{x}) = \mathsf{E}[f(\mathbf{x})]$ and covariance function $k(\mathbf{x}, \mathbf{x}') = \mathsf{E}[(f(\mathbf{x}) - \mu(\mathbf{x}))(f(\mathbf{x}') - \mu(\mathbf{x}'))]$.

For a set of values $\{(\mathbf{x}_i, f(\mathbf{x}_i), i = 1, \ldots, m\}$ the Gaussian process implies a multivariate Gaussian distribution $\mathbf{f} = \{f(\mathbf{x}_1), f(\mathbf{x}_2), \ldots, f(\mathbf{x}_m)\} \sim \mathcal{N}(\boldsymbol{\mu}_\mathbf{f}, \Sigma_\mathbf{ff})$ where $\boldsymbol{\mu}_\mathbf{f}$ is the mean vector and $\Sigma_\mathbf{ff}$ the covariance matrix of $\mathbf{f}$ obtained by evaluating the mean function and covariance function at all points $\{\mathbf{x}_1, \mathbf{x}_2, \ldots \mathbf{x}_m\}$.

2) Regression/Prediction: Given a set of m observed values $\mathbf{f}$, the goal is to predict a set of m_* values $\mathbf{f}_*$ that are not observed. The joint prior distribution of $\mathbf{f}$ and $\mathbf{f}_*$ is

$$\begin{bmatrix}\mathbf{f}\\ \mathbf{f}_*\end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix}\boldsymbol{\mu}_{\mathbf{f}}\\ \boldsymbol{\mu}_{\mathbf{f}_*}\end{bmatrix}, \begin{bmatrix}\Sigma_{\mathbf{ff}} & \Sigma_{\mathbf{ff}_*}\\ \Sigma_{\mathbf{f}_*\mathbf{f}} & \Sigma_{\mathbf{f}_*\mathbf{f}_*}\end{bmatrix}\right),$$

where $\Sigma_{\mathbf{ff}_*}$ is the $m \times m_*$ covariance matrix of all pairs of values in $\mathbf{f}$ and $\mathbf{f}_*$ and so on. Conditioning on the observations $\mathbf{f}$, the posterior distribution of $\mathbf{f}_*$ is [53, Eq. (2.19)]

$$\begin{aligned}[\mathbf{f}_* \mid \mathbf{f}] \sim \mathcal{N}(&\boldsymbol{\mu}_{\mathbf{f}_*} + \Sigma_{\mathbf{f}_*\mathbf{f}}\Sigma_{\mathbf{ff}}^{-1}(\mathbf{f} - \boldsymbol{\mu}_{\mathbf{f}}),\\ &\Sigma_{\mathbf{f}_*\mathbf{f}_*} - \Sigma_{\mathbf{f}_*\mathbf{f}}\Sigma_{\mathbf{ff}}^{-1}\Sigma_{\mathbf{ff}_*}).\end{aligned} \tag{10}$$

In particular, the posterior covariance does not depend on the actual values that are observed. The term $\Sigma_{\mathbf{f}_*\mathbf{f}}\Sigma_{\mathbf{ff}}^{-1}\Sigma_{\mathbf{ff}_*}$ quantifies the reduction in uncertainty that is achieved by the observations $\mathbf{f}$.

3) Standard Kernels and Combined Kernels: The covariance function is also called the kernel and there are a number of well-known kernel functions in the literature [55], such as exponential, squared exponential, and rational quadratic, that are characteristic of different processes. As a first example, we consider an exponential covariance function with a scalar input x, which is known as the Ornstein-Uhlenbeck process [53]. It is defined as

$$k(x, x') = \sigma^2 e^{-\frac{1}{l}|x-x'|}, \tag{11}$$

where $\sigma^2 > 0$ is the variance and parameter $l > 0$ determines the length scale of the correlation. The kernel depends only on the difference $x - x'$, i.e., it is stationary. The mean function is frequently normalized as $\mu(\mathbf{x}) = 0$.

If the process depends on a vector of inputs, several individual kernels can be combined to build a new multi-dimensional kernel. Multiplication of kernels gives a kernel that has large values only, if all of the individual kernels have a large value. This can be thought of as ‘and’ operation [55].

A relevant case arises when we consider a system of spatially distributed sensors whose values are correlated in space and time. If the spatial and temporal correlations are each modeled by an exponential covariance function with indices s and t, respectively, then the 2D kernel that is obtained by multiplication is

$$k(\mathbf{x}, \mathbf{x}') = \sigma^2 e^{-\frac{1}{l_s}|x_s - x'_s| - \frac{1}{l_t}|x_t - x'_t|}.$$

Other higher-dimensional models are also possible.

In practice, the observation of the function values $f(\mathbf{x})$ may be subject to noise. With additive iid Gaussian observation noise ϵ, with zero mean and variance σ_N^2, the covariance matrix of the observations is [53, Eq. (2.20)]

$$\Sigma_{\mathbf{ff}} = k(\mathbf{x}, \mathbf{x}') + I\sigma_N^2. \tag{12}$$

It is common to learn the type of kernel and its hyper-parameters, such as the length scales l_s, l_t, variance σ^2, and noise variance σ_N^2, from training data. A standard method is to apply Bayesian inference and estimate the hyper-parameters by maximizing the marginal likelihood [53, Sec. 5.41].

B. Prediction Variance

We characterize the samples that are available at the monitor at time $t \geq 0$ by indicator variables $\Gamma_t(\varsigma, \tau)$, where $\varsigma \in \mathbb{S}$ denotes the sensor and $\tau \in [0, t]$ is the generation time of the sample. Assuming the monitor does not have knowledge of the physical process of sensor s at time t, it uses the samples $\Gamma_t(\varsigma, \tau)$ to make a prediction and we denote $\Phi_t(\Gamma, s)$ as the prediction variance of the estimate. For a Gaussian process we use (10) to derive the prediction variance

$$\Phi_t(\Gamma, s) = \Sigma_{\mathbf{f}_*\mathbf{f}_*} - \Sigma_{\mathbf{f}_*\mathbf{f}}\Sigma_{\mathbf{ff}}^{-1}\Sigma_{\mathbf{ff}_*}, \tag{13}$$

where $\mathbf{f}$ comprises the samples that are used to make the prediction, these are the samples marked in Γ, and $\mathbf{f}_*$ is the prediction, here of the process at sensor s and time t.

Similarly, we use $\Phi_t(\varsigma, s)$ to denote the prediction variance of the process at sensor s and time t, given only the most recent sample from sensor ς with AoI $\Delta_t(\varsigma)$. The variance $\Phi_t(\varsigma, s)$ can be viewed as a penalty that is due to to the use of a sample from a different sensor ς and the AoI of that sample. Conversely, the term $\Sigma_{\mathbf{f}_*\mathbf{f}}\Sigma_{\mathbf{ff}}^{-1}\Sigma_{\mathbf{ff}_*}$ is a measure of the value-of-information of the sample, as it quantifies the reduction in uncertainty that is achieved by the sample. Since the definition of AoI is based only on the generation time of the most recent sample available at the monitor, we use the same approach to derive a relation between the posterior variance and the 2D-AoI. We note that the monitor can still use all available samples and (13) to reduce the prediction variance further.

C. 2D-AoI of Product Kernels

Equation (13) provides a means of calculating the prediction variance for any given covariance kernel $k(\mathbf{x}, \mathbf{x}')$. Evaluating the kernel at defined points in time and space yields the covariance matrices $\Sigma_{\mathbf{ff}_*}$ etc., which are used in (13). In the following, we employ a number of well-established kernel functions in order to derive closed-form results. In particular, we formalize the relationship between space and time, thereby obtaining the AeD of sensors. We use stationary product kernels [55] to model spatio-temporal processes. To include a range of common kernel functions, we consider product kernels with the following structure

$$k(\Delta_t, x_\varsigma, x_s) = \sigma^2 k_t(\Delta_t(\varsigma)) k_s(x_\varsigma, x_s), \tag{14}$$

where functions k_t and k_s satisfy $k_t(0) = 0$, $k_s(x_s, x_s) = 1$ for all $s \in \mathbb{S}$, and σ^2 is the process variance. Function k_t denotes the temporal correlation and k_s the spatial correlation. Additional dimensions, such as non-homogeneous sensors, can be included in k_s by extending the index set, or represented by multiplication of additional kernel functions.

In line with the general assumption of AoI, that a new sample replaces previous samples, we assume that function k_t is decreasing. This means that a more recent sample is better suited for prediction than any previous one. We define k_t^{-1} to be the inverse of k_t. Since k_t is decreasing, it follows that k_t^{-1} is decreasing, too. To see this, define $y_i = k_t(x_i)$ and $x_i = k_t^{-1}(y_i)$. Then for k_t decreasing and $y_1 < y_2$

$$\begin{aligned} y_1 < y_2 &\Leftrightarrow k_t(x_1) < k_t(x_2) \\ \Leftrightarrow x_1 > x_2 &\Leftrightarrow k_t^{-1}(y_1) > k_t^{-1}(y_2). \end{aligned}$$

The prediction variance given the most recent sample of sensor ς follows by insertion of (14) into (13) as

$$\Phi_t(\varsigma, s) = \sigma^2\Big(1 - \big(k_t(\Delta_t(\varsigma))k_s(x_\varsigma, x_s)\big)^2\Big), \tag{15}$$

where we used that $\Sigma_{\mathbf{ff}} = \Sigma_{\mathbf{f}_*\mathbf{f}_*} = \sigma^2$, $\Sigma_{\mathbf{ff}}^{-1} = 1/\sigma^2$, and $\Sigma_{\mathbf{ff}_*} = \Sigma_{\mathbf{f}_*\mathbf{f}} = \sigma^2 k_t(\Delta_t(\varsigma))k_s(x_\varsigma, x_s)$.

The next step is to leverage the structure of the prediction variance in (15) to develop a definition of 2D-AoI that transforms space into an equivalent of time. This is achieved by setting

$$\Delta_t^{\text{2D}}(\varsigma, s) := k_t^{-1}(k_t(\Delta_t(\varsigma))k_s(x_\varsigma, x_s)). \tag{16}$$

The definition implies that

$$k_t(\Delta_t^{\text{2D}}(\varsigma, s)) = k_t(\Delta_t(\varsigma))k_s(x_\varsigma, x_s),$$

i.e., the 2D-AoI transfers the spatial correlation $k_s(x_\varsigma, x_s)$ to the temporal domain, expressed by function $k_t(\Delta_t^{\text{2D}}(\varsigma, s))$. Thus, (15) can be written as

$$\Phi_t(\varsigma, s) = \sigma^2\Big(1 - \big(k_t(\Delta_t^{\text{2D}}(\varsigma, s))\big)^2\Big), \tag{17}$$

using only the temporal covariance function k_t. This means that for predicting the physical phenomenon at position x_s and time t, a sample from position x_ς with AoI $\Delta_t(\varsigma)$ is as good as using a sample from position x_s with AoI $\Delta_t(s) = \Delta_t^{\text{2D}}(\varsigma, s)$. This achieves the intended transformation of the spatial distance into an age-equivalent representation. Given the prediction variance, (17) can be solved for the 2D-AoI

$$\Delta_t^{\text{2D}}(\varsigma, s) = k_t^{-1}\left(\sqrt{1 - \frac{\Phi_t(\varsigma, s)}{\sigma^2}}\right).$$

As defined in (3) we have $\Delta_t^{\text{2D}}(\varsigma, s) = \Lambda_t(\varsigma, s) + \Delta_t(\varsigma)$ and by insertion of (16) into (3) we obtain the AeD

$$\Lambda_t(\varsigma, s) = k_t^{-1}(k_t(\Delta_t(\varsigma))k_s(x_\varsigma, x_s)) - \Delta_t(\varsigma). \tag{18}$$

We will derive $\Lambda_t(\varsigma, s)$ and hence $\Delta_t^{\text{2D}}(\varsigma, s)$ for different commonly used kernel functions in Sec. VI, where we find that in some but not all cases $\Lambda_t(\varsigma, s)$ is a function of space only, i.e., $\Lambda_t(\varsigma, s) = \Lambda(\varsigma, s)$ as depicted in Fig. 3.

Before we move on to specific kernel functions, we conclude this section with some general results. With (3) the CCDF of $\Delta_t^{\text{2D}}(\varsigma, s)$ is

$$\mathsf{P}[\Delta_t^{\text{2D}}(\varsigma, s) > y] = \mathsf{P}[\Delta_t(\varsigma) > y - \Lambda_t(\varsigma, s)], \tag{19}$$

for $y \geq \Lambda_t(\varsigma, s)$. The CCDF of the prediction variance follows from (17) as

$$\begin{aligned} \mathsf{P}[\Phi_t(\varsigma, s) > z\sigma^2] &= \mathsf{P}[1 - k_t(\Delta_t^{\text{2D}}(\varsigma, s))^2 > z] \\ &= \mathsf{P}\big[k_t(\Delta_t^{\text{2D}}(\varsigma, s)) < \sqrt{1-z}\big] \\ &= \mathsf{P}\big[\Delta_t^{\text{2D}}(\varsigma, s) > k_t^{-1}\big(\sqrt{1-z}\big)\big], \end{aligned}$$

where we normalized the CCDF with the variance, so that $z \in [0, 1]$. Above, since k_t^{-1} is a decreasing function the order of the inequality is inverted. Solving for $\Delta_t(\varsigma)$ the CCDF of the prediction variance can also be written as

$$\mathsf{P}[\Phi_t(\varsigma, s) > z\sigma^2] = \mathsf{P}\left[\Delta_t(\varsigma) > k_t^{-1}\left(\frac{\sqrt{1-z}}{k_s(x_\varsigma, x_s)}\right)\right]. \tag{20}$$

The sample that has the minimal 2D-AoI is the most recent sample of the sensor ς that minimizes the prediction variance $\arg\min_\varsigma\{\Phi_t(\varsigma, s)\}$. In the case that the sensors are connected via independent channels, (6) applies accordingly to the prediction variance and the CCDF of the minimal prediction variance $\Phi_t(s) = \min_\varsigma\{\Phi_t(\varsigma, s)\}$ is

$$\begin{aligned} \mathsf{P}[\Phi_t(s) > z\sigma^2] &= \mathsf{P}\Big[\min_{\varsigma\in\mathbb{S}}\{\Phi_t(\varsigma, s)\} > z\sigma^2\Big] \\ &= \prod_{\varsigma\in\mathbb{S}} \mathsf{P}[\Phi_t(\varsigma, s) > z\sigma^2]. \end{aligned} \tag{21}$$

The mean prediction variance follows by integration of the tail distribution as $\mathsf{E}[\Phi_t(s)] = \int_0^1 \mathsf{P}[\Phi_t(s) > z\sigma^2]dz$.

D. Observation Noise

Assuming noisy observations $\mathbf{f}$ with additive iid Gaussian noise ϵ with zero mean and variance σ_N^2, the covariance of the observations, given by (12), is $\Sigma_{\mathbf{ff}} = k(\mathbf{x}, \mathbf{x}') + I\sigma_N^2$. The prediction variance follows by insertion of $\Sigma_{\mathbf{ff}}$ into (13). Here we assume that the noise of all sensors is homogeneous with variance σ_N^2. Sensors with different noise variance can be taken into account at the expense of additional notation. When using the most recent sample of sensor ς for prediction, we have $\Sigma_{\mathbf{ff}} = \sigma^2 + \sigma_N^2$ and by insertion of (14) into (13) the prediction variance turns into

$$\Phi_t(\varsigma, s) = \sigma^2\big(1 - \eta(k_t(\Delta_t(\varsigma))k_s(x_\varsigma, x_s))^2\big), \tag{22}$$

with $\eta = \frac{\sigma^2}{\sigma^2 + \sigma_N^2}$. Here we used that $\Sigma_{\mathbf{f}_*\mathbf{f}_*} = \sigma^2$, $\Sigma_{\mathbf{ff}}^{-1} = 1/(\sigma^2 + \sigma_N^2)$, and $\Sigma_{\mathbf{ff}_*} = \Sigma_{\mathbf{f}_*\mathbf{f}} = \sigma^2 k_t(\Delta_t(\varsigma))k_s(x_\varsigma, x_s)$.

Substituting (22) into the key steps of the definition of 2D-AoI in (15)-(17), we see that the inclusion of observation noise has no effect on the definition of 2D-AoI, but only on the prediction variance, and (20) is shifted so that

$$\mathsf{P}[\Phi_t(\varsigma, s) > z\sigma^2] = \mathsf{P}\left[\Delta_t(\varsigma) > k_t^{-1}\left(\frac{\sqrt{1-z}}{\sqrt{\eta}k_s(x_\varsigma, x_s)}\right)\right].$$

This allows our model to easily incorporate independent Gaussian noise. For convenience, we use $\eta = 1$ in the following, i.e., we assume noise-free observations or negligible noise.

VI. 2D-AoI of Common Kernels

Using the general results from Sec. V we derive specific solutions for products of some widely used kernel functions including exponential, squared exponential, and rational quadratic. Empirical data have often been used to parameterize these kernels, e.g., exponential product kernels have been used to model spatio-temporal temperature and humidity data in [19], squared exponential kernels for road speed correlation [34], and product kernels including squared exponential and rational quadratic for atmospheric CO_2 concentration [53], to name a few.

A. Exponential Kernels

For the case of exponential product kernels we instantiate (14) with $k_t(\Delta_t(\varsigma)) = e^{-\frac{1}{l_t}\Delta_t(\varsigma)}$ and $k_s(x_\varsigma, x_s) = e^{-\frac{1}{l_s}|x_\varsigma - x_s|}$, where $|x_\varsigma - x_s|$ is the Euclidean distance of sensors ς and s, and $l_t, l_s > 0$ are the length scales of the spatial and temporal correlation, respectively. The inverse function of the temporal covariance function follows as

$$k_t^{-1}(y) = -l_t \ln y,$$

for $y \in (0, 1]$, and by insertion into (18) the AeD is

$$\begin{aligned}\Lambda(\varsigma, s) &= -l_t \ln\left(e^{-\frac{1}{l_s}|x_\varsigma - x_s|} e^{-\frac{1}{l_t}\Delta_t(\varsigma)}\right) - \Delta_t(\varsigma)\\ &= \frac{l_t}{l_s}|x_\varsigma - x_s|. \qquad (23)\end{aligned}$$

The CCDF of the 2D-AoI follows directly by insertion of (23) into (19) as

$$\mathsf{P}[\Delta_t^{\text{2D}}(\varsigma, s) > y] = \mathsf{P}\left[\Delta_t(\varsigma) > y - \frac{l_t}{l_s}|x_\varsigma - x_s|\right], \qquad (24)$$

for $y \geq \frac{l_t}{l_s}|x_\varsigma - x_s|$. The CCDF of the prediction variance from (20) is

$$\begin{aligned}&\mathsf{P}[\Phi_t(\varsigma, s) > z\sigma^2]\\ &= \mathsf{P}\left[\Delta_t(\varsigma) > -l_t \ln\left(\frac{\sqrt{1-z}}{e^{-\frac{1}{l_s}|x_\varsigma - x_s|}}\right)\right]\\ &= \mathsf{P}\left[\Delta_t(\varsigma) > -\frac{l_t}{2}\ln(1-z) - \frac{l_t}{l_s}|x_\varsigma - x_s|\right], \qquad (25)\end{aligned}$$

for $z \geq 1 - e^{-\frac{2}{l_s}|x_\varsigma - x_s|}$ and $z < 1$. For exponential product kernels, the AeD $\Lambda(\varsigma, s)$ in (23) is a function of space only. This is the basic case that we depicted already in Fig. 3. Numerical results for this case have been presented in Sec. IV.

B. Squared Exponential Kernels

For squared exponential kernels we substitute $k_t(\Delta_t(\varsigma)) = e^{-\frac{1}{2l_t^2}(\Delta_t(\varsigma))^2}$ and $k_s(x_\varsigma, x_s) = e^{-\frac{1}{2l_s^2}(x_\varsigma - x_s)^2}$, with $l_t, l_s > 0$, into (14). It follows that

$$k_t^{-1}(y) = l_t\sqrt{-2\ln y},$$

for $y \in (0, 1]$. By insertion into (18) we obtain

$$\begin{aligned}\Lambda_t(\varsigma, s) &= \left(-2l_t^2 \ln\left(e^{-\frac{1}{2l_s^2}(x_\varsigma - x_s)^2} e^{-\frac{1}{2l_t^2}\Delta_t(\varsigma)^2}\right)\right)^{\frac{1}{2}} - \Delta_t(\varsigma)\\ &= \sqrt{\frac{l_t^2}{l_s^2}(x_\varsigma - x_s)^2 + \Delta_t(\varsigma)^2} - \Delta_t(\varsigma). \qquad (26)\end{aligned}$$

In comparison to the behavior observed in (23), the behavior exhibited by (26) is quite distinct. In this case, the AeD $\Lambda_t(\varsigma, s)$ is a function not only of space but also of time, specifically the AoI at the given time point t. When $\Delta_t(\varsigma) = 0$, (26) is identical to (23). However, the magnitude of $\Lambda_t(\varsigma, s)$ in (26) is reduced as $\Delta_t(\varsigma)$ increases, with $\Lambda_t(\varsigma, s) \to 0$ for $\Delta_t(\varsigma) \to \infty$. The effect is illustrated in the schematic in Fig. 5, which can be directly compared with Fig. 3. In contrast to Fig. 3, the spatial offset of the blue curve representing sensor 2 gradually diminishes over time in Fig. 5. This indicates that when sensors communicate less frequently, samples from distant sensors are of greater value in the case of squared exponential kernels than in the case of exponential kernels.

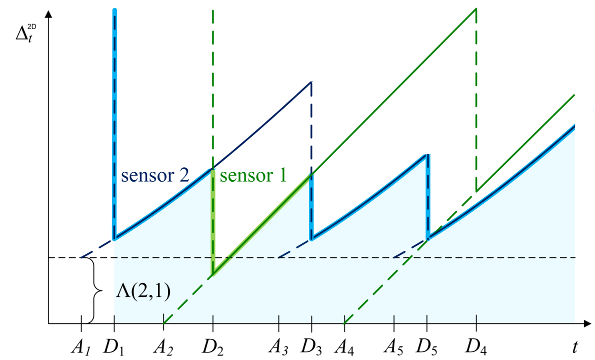


Fig. 5. Minimal 2D-AoI $\Delta_t^{\text{2D}}(1)$ for the case of $S = 2$ sensors as in Fig. 3 but for squared exponential instead of exponential kernels.

The CCDF of the 2D-AoI follows by insertion of (26) into (19) as

$$\mathsf{P}[\Delta_t^{\text{2D}}(\varsigma, s) > y] = \mathsf{P}\left[\Delta_t(\varsigma) > \left(y^2 - \frac{l_t^2}{l_s^2}(x_\varsigma - x_s)^2\right)^{\frac{1}{2}}\right], \quad (27)$$

for $y \geq \frac{l_t}{l_s}|x_\varsigma - x_s|$, and the CCDF of the minimal 2D-AoI is obtained by insertion into (6). The CCDF of the prediction variance in (20) of the squared exponential kernel becomes

$$\begin{aligned}&\mathsf{P}[\Phi_t(\varsigma, s) > z\sigma^2]\\ &= \mathsf{P}\left[\Delta_t(\varsigma) > \left(-2l_t^2 \ln\left(\frac{\sqrt{1-z}}{e^{-\frac{1}{2l_s^2}(x_\varsigma - x_s)^2}}\right)\right)^{\frac{1}{2}}\right]\\ &= \mathsf{P}\left[\Delta_t(\varsigma) > \left(-l_t^2 \ln(1-z) - \frac{l_t^2}{l_s^2}(x_\varsigma - x_s)^2\right)^{\frac{1}{2}}\right], \quad (28)\end{aligned}$$

for $z \geq 1 - e^{-\frac{1}{l_s^2}(x_\varsigma - x_s)^2}$ and $z < 1$.

Results for the scenario evaluated in Sec. IV, where two sensors are spaced apart and connected to the monitor via independent M/M/1 queues, are shown in Fig. 6 for squared exponential kernels. The parameters of the kernels are selected so that $(x_2 - x_1)^2 l_t^2 / l_s^2 = 100$, i.e., for $\Delta_t(2) = 0$ the AeD is $\Lambda_t(2, 1) = 10$ from (26). For comparison, the dashed lines show the previous results from Fig. 4(c) that apply for exponential kernels with the same parameters x_1, x_2, l_t, l_s.

Fig. 6 clearly demonstrates the impact of the squared exponential kernel compared to the exponential kernel. The offset of $\Delta^{\text{2D}}(2, 1)$ of the CCDF at 1.0 is identical to $y = 10$ in both cases. However, for $y > 10$, that is when $\Delta_t(2)$ increases, the CCDF of $\Delta^{\text{2D}}(2, 1)$ of the squared exponential kernel decays more rapidly. This is a consequence of (26). As the value of y increases, the CCDFs of $\Delta^{\text{2D}}(2, 1)$ and $\Delta^{\text{2D}}(1, 1)$ converge for the squared exponential kernel, showing the decreasing effect of the distance of sensor 2. The CCDF of $\Delta^{\text{2D}}(2, 1)$ of the exponential kernel (dashed lines) has the same speed of tail decay but will remain separated from $\Delta^{\text{2D}}(1, 1)$ by $y = 10$. While $\Delta^{\text{2D}}(1, 1)$ is identical for both types of kernels, since it depends on the AoI $\Delta(1)$ only, the minimal 2D-AoI $\Delta^{\text{2D}}(1)$ benefits from the more rapid decay of $\Delta^{\text{2D}}(2, 1)$ for the

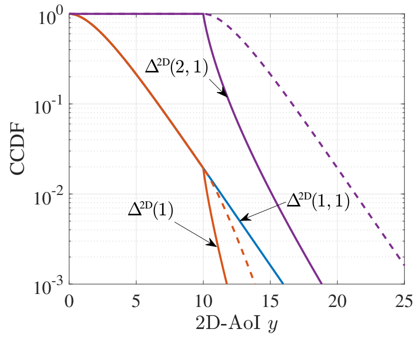


Fig. 6. CCDF of the 2D-AoI like Fig. 4(c) but with squared exponential kernels. The dashed lines show the corresponding result when using an exponential kernel.

squared exponential kernel, beginning at $y = 10$ and a tail probability of about 0.02.

C. Rational Quadratic and Mixed Product Kernels

There are other commonly used kernel functions that can be arbitrarily combined to form different types of product kernels (14). We conclude this section by considering a rational quadratic temporal kernel $k_t(\Delta_t(\varsigma)) = (1 + \Delta_t(\varsigma)^2/(2\alpha l_t^2))^{-\alpha}$, where $\alpha, l_t > 0$, combined with a not explicitly specified spatial kernel $k_s(x_\varsigma, x_s)$ that can be any of, e.g., exponential, squared exponential, or rational quadratic. Following the established procedure, we obtain

$$k_t^{-1}(y) = \sqrt{2\alpha l_t^2\big(y^{-\frac{1}{\alpha}} - 1\big)},$$

for $y \in (0, 1]$. By insertion into (18) the AeD is

$$\Lambda_t(\varsigma, s) = \sqrt{2\alpha l_t^2\big(k_s(x_\varsigma, x_s)^{-\frac{1}{\alpha}} - 1\big) + k_s(x_\varsigma, x_s)^{-\frac{1}{\alpha}}\Delta_t(\varsigma)^2} - \Delta_t(\varsigma). \tag{29}$$

The CCDF of the 2D-AoI follows by insertion of (29) into (19)

$$\mathsf{P}[\Delta_t^{\text{2D}}(\varsigma, s) > y] = \mathsf{P}\left[\Delta_t(\varsigma) > \Big(k_s(x_\varsigma, x_s)^{\frac{1}{\alpha}}\big(y^2 + 2\alpha l_t^2\big) - 2\alpha l_t^2\Big)^{\frac{1}{2}}\right], \tag{30}$$

for $y \geq (2\alpha l_t^2(k_s(x_\varsigma, x_s)^{-\frac{1}{\alpha}} - 1))^{\frac{1}{2}}$ and the CCDF of the prediction variance follows from (20).

The AeD of the rational quadratic kernel in (29) shows a dependence on the AoI $\Delta_t(\varsigma)$ as also observed for the AeD of the squared exponential kernel in (26). For high α the rational quadratic kernel converges against the squared exponential kernel. Consequently letting $\alpha \to \infty$, (29) is verified to be the same as (26) for the squared exponential kernel.

VII. Model Selection for Freeway Traffic Flow

We illustrate how Gaussian process regression can be used to model road traffic data collected by the traffic control centers. A common technology for recording traffic counts is loop detectors. This data is then transmitted to the control center via wired or wireless communication.

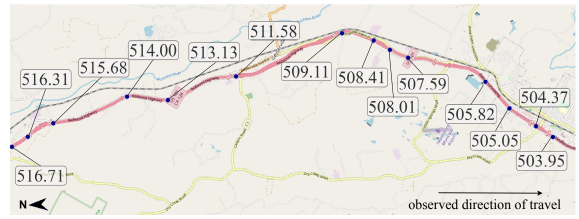


Fig. 7. Interstate US101-S, south direction. Segment between Cloverdale and Healdsburg in direction Santa Rosa, California. Numbers indicate the absolute postmiles. Sensors are indexed $s = 1, 2, \ldots, 14$ in increasing order of their postmile position. Map data © OpenStreetMap contributors, CC BY-SA [58].

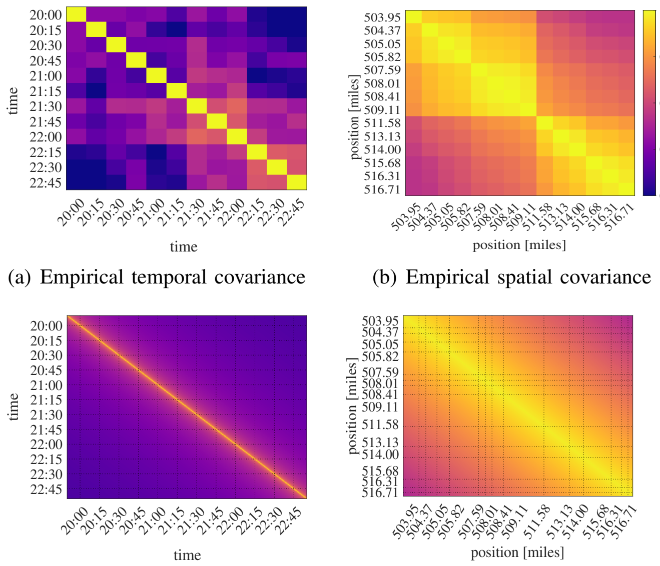
For cost efficiency, wireless IoT technologies, such as NB-IoT and LoRaWAN, are of particular interest for data collection. However, these solutions only support small data volumes, have low data rates, and experience high latency. For example, see the study [56] on using LoRaWAN to connect loop detectors. Given the limited update rates and the delays that occur, it is important to understand the AoI and its implications.

The dataset that we use was collected by the Performance Measurement System (PeMS) of the California Department of Transportation (Caltrans) on freeways [57]. We train spatio-temporal product kernels on the dataset and estimate their hyper-parameters. Road traffic data, specifically speed, was also used in [34]. Other publicly available datasets that have been modeled by Gaussian processes include atmospheric carbon dioxide [53], temperature, and humidity [19].

1) Traffic Dataset: PeMS data is collected from around 40,000 detectors on freeways in major metropolitan areas in California. The data includes estimates of detector reliability, and we have omitted data from detectors categorized as unreliable. The dataset we use covers $S = 14$ sequential traffic detectors located along a 13-mile section of the Interstate US101-S heading south from Cloverdale in direction Santa Rosa. The data relates to the specified direction of travel, with all lanes taken into account by the detectors. Fig. 7 shows the section of the freeway, the sensors $s \in \{1, 2, \ldots, 14\}$ and their positions x_s.

The data we use for training was recorded from 1 April to 14 May 2025, i.e., it comprises 44 consecutive days. The days from 15 May onward are available for testing. We use a three-hour time slot each day, between 20:00 and 23:00. During this time, traffic flow is usually moderate. If the daily course of the traffic flow is of interest, an additional sinusoidal kernel can be added [53], [55]. However, these time scales fall outside the scope of our investigation of AoI.

2) Data Preprocessing: In short, inductive loop detectors recognize and count vehicles that pass over the detector, i.e. the output data is the number of vehicles counted in a given time interval. PeMS data from traffic detectors is available at various levels of temporal aggregation. On short timescales, traffic detector data fluctuates heavily, but starts to show more structured behavior when aggregated over five to fifteen minutes. These aggregation levels are typical for traffic detector data [34], [59]. In our experiments, we use 15-minute time intervals as we have found that these provide a clearer



(a) Empirical temporal covariance
(b) Empirical spatial covariance
(c) Temporal kernel function
(d) Spatial kernel function

Fig. 8. Empirical covariance matrices and estimated kernel functions.

structure, and obtain $T = 12$ intervals per three-hour time slot. Using $S = 14$ traffic detectors, this results in $ST = 168$ samples for each of the 44 days. We denote the samples $\tilde{f}^i_{\text{raw}}(s,t)$, where $s \in \{1, 2, \dots, 14\}$ is the sensor index, $t \in \{1, 2, \dots, 12\}$ the time index, and $i \in \{1, 2, \dots, 44\}$ refers to the day. We use superscript tilde to indicate empirical data and statistics derived from it, and subscript 'raw' refers to data before preprocessing.

To bring out the structure of the data better and make it more accessible for learning, we standardize the data by subtracting the average value and dividing by its standard deviation. We calculate $\tilde{f}^i_{\text{std}}(s,t) = (\tilde{f}^i_{\text{raw}}(s,t) - \tilde{\mu}(s,t))/\tilde{\sigma}(s,t)$, where $\tilde{\mu}(s,t)$ and $\tilde{\sigma}(s,t)$ are the empirical average and standard deviation, respectively, obtained by averaging over all days, i.e., $\tilde{\mu}(s,t) = \sum_{i=1}^{44} \tilde{f}^i_{\text{raw}}(s,t)/44$. The standardized data have a mean of zero and a variance of one. A positive (negative) value $\tilde{f}^i_{\text{std}}(s,t)$ means that the number of vehicles counted by sensor s in time-interval t is larger (smaller) than the average observed by that sensor at that time of the day. The magnitude indicates how large the deviation from the average is in relation to the standard deviation. Standardization is reversible, meaning that standardized data can be transformed back into its original form.

To illustrate the degree of similarity between the data points $\tilde{f}^i_{\text{std}}(s,t)$, we present the empirical temporal and spatial covariance matrices in Figs. 8(a) and 8(b). Since the data has been standardized, the magnitude of the covariance is limited to one, as seen on the diagonal. Generally, the covariance is larger near the diagonal, and decreases with increasing temporal or spatial distance, respectively.

Regarding the spatial covariance matrix in Fig. 8(b), a few large, rectangular-shaped structures are apparent. These are due to the fact that the traffic detectors are not evenly spaced along the freeway section, see Fig. 7, and notice the influence of the comparably large distance of the traffic detectors at mileage 509.11 and 511.58 in Fig. 8(b).

As can be seen in Fig. 8(a), the temporal correlation diminishes more rapidly, with a significant decrease occurring even after one time step. This emphasizes the importance of AoI and shows how useful fresh samples from nearby sensors are to compensate for outdated data.

3) Fitting the Kernel: Given the empirical data, the aim is to find a kernel model that can reflect the essential characteristics of the data. Fitting the kernel by maximization of the marginal likelihood is implemented in well-known software frameworks, such as the Python packages *scikit-learn* [60] for machine learning as well as *GPy* [61] and *GPyTorch* [62] for Gaussian Processes, that offer a catalogue of kernel models.

We assume a stationary two-dimensional product kernel to characterize the spatio-temporal road traffic dataset and estimate the hyper-parameters from the dataset using the *GPyTorch* implementation. We use multi-task learning, where several datasets that share the same hyper-parameters are used for training. To do this, we divided the data into 44 individual datasets, one dataset for each day. We trained several standard kernels and found that the spatial and temporal characteristics of the data are best matched by a product of an exponential and a rational quadratic kernel

$$k(\Delta_t, x_\varsigma, x_s) = \sigma^2 e^{-\frac{1}{l_s}|x_\varsigma - x_s|}\left(1 + \frac{\Delta_t^2}{2\alpha l_t^2}\right)^{-\alpha}.$$

We specify distance in miles and time in seconds, i.e., the scale parameters l_s and l_t are defined in terms of these units.

We show the individual kernel functions in Fig. 8(c) and 8(d). Unlike the empirical covariance, that is a matrix, the kernels are continuous functions. The estimated hyper-parameters are $\sigma^2 = 0.988$ with noise variance $\sigma_N^2 = 0.012$ and for the rational quadratic temporal kernel $l_t = 26.45$ and $\alpha = 0.028$. The temporal kernel in Fig. 8(c) shows quite good agreement with the empirical covariance in Fig. 8(a). The irregularities that can be observed in the empirical covariance matrix cannot be reproduced by a stationary kernel.

The hyper-parameter of the exponential spatial kernel is $l_s = 52.24$. The match of Fig. 8(d) with Fig. 8(b) is less obvious. This is because the spatial kernel is a function of position, whereas the covariance matrix is accessed via sensor indices, and the traffic detectors are not spaced equidistantly. For reference, we mark the positions of the traffic detectors in the kernel function in Fig. 8(d). In the upper left corner, two groups of four sensors in close proximity can be seen, whereas the six sensors in the lower right corner are spaced further apart. This causes the rectangular structures in the empirical covariance in Fig. 8(b). As the kernels are continuous functions, they enable the prediction of function values for input vectors that were not previously observed and that may even lie outside the observed range.

VIII. Evaluation

We present two sets of evaluation results in which the sensors are arranged in a line topology, corresponding to the freeway scenario described in Sec. VII. Sec. VIII-A considers a basic wireless channel model for sensor connectivity and Sec. VIII-B evaluates the impact of spatial sensor density using

M/M/1 queues with a sum rate constraint. Further results for a grid topology are included in the technical report [24].

A. Traffic Detectors connected via ALOHA Channels

We assume that the loop detectors positioned along the section of the freeway are connected via a low-power wide-area wireless network. Relevant technologies include NB-IoT and LoRaWAN. The connection of loop detectors via LoRaWAN has been investigated, for example, in [56]. LoRaWAN uses a range of spreading factors to adjust transmission robustness according to distance. This results in payload data rates ranging from several kbps down to 0.25 kbps over a distance of up to 10 miles. Including time overhead for transmission of preamble, physical layer header, and cyclic redundancy check, even small messages of a couple of bytes can have transmission times of up to one second or more. Channel access is performed using a basic ALOHA access scheme. To ensure fair use, the LoRaWAN alliance recommends to limit the senders' duty cycle to at most 1%. Furthermore, reducing the duty cycle decreases the sensors' energy consumption.

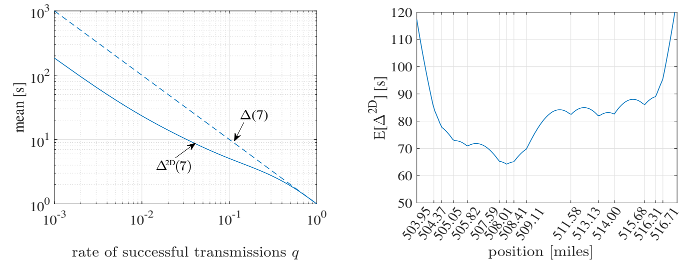
1) 2D-AoI Analysis: To facilitate the assessment, we adopt a basic slotted ALOHA channel access model and a generate-at-will sampling policy. Each time a sensor accesses the channel, it takes a fresh sample and transmits it as a packet. The transmission time of a packet is set to one second, and this is also the unit of one time slot. Sensors access the channel at random at a mean rate of λ, where $\lambda = 1/300$ seconds^{-1}, for example, corresponds to an average of one update message every five minutes and a duty cycle of 0.33%. The transmission is subject to errors and possibly packet collisions, resulting in an error probability of p_{err}. Consequently, $(1 - p_{\text{err}})$ is the probability of a successful transmission, and we denote the mean rate of successful transmissions of a sensor as $q = \lambda(1 - p_{\text{err}})$. The probability that the AoI of the sensor is larger than a threshold y equals the probability that there was no successful transmission of that sensor during the past y time slots (otherwise the AoI would be less than or at most equal to y), resulting in the CCDF

$$\mathsf{P}[\Delta > y] = (1 - q)^{\lfloor y \rfloor}, \tag{31}$$

for $y \geq 0$. By rounding y down to the nearest integer, time is discretized and the transmission time of one time slot is taken into account.

The 2D-AoI is obtained by substitution of (31) into (30) for a product kernel, which is composed of a rational quadratic temporal kernel and an exponential spatial kernel. We use the hyper-parameters that we obtained when training the kernels with the Caltrans PeMS data, see Figs. 8(c) and 8(d). We assume that sensors are connected via statistically independent channels and obtain the minimal 2D-AoI by insertion of (30) into (6) and then (8) for the expected value.

We present numerical results in Fig. 9. Fig. 9(a) compares the AoI of sensor 7 at postmile 508.41 with the 2D-AoI for that position. Compared to the AoI, the 2D-AoI uses samples from all sensors. The impact of the rate of successful transmissions q is evaluated. As q approaches one, there is always a fresh sample available from sensor 7, and data from more distant



(a) Position x_7 = postmile 508.41 (b) All positions, $q = 0.003$

Fig. 9. AoI and 2D-AoI for sensors connected via slotted ALOHA channels, where q is the rate of successful transmissions. The sensors are placed along the freeway as displayed in Fig. 7. The learned spatio-temporal kernel functions shown in Figs. 8(c) and 8(d) are used.

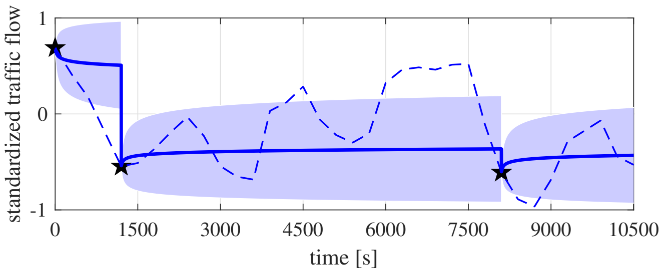
sensors cannot offer any additional benefit. Consequently, AoI and 2D-AoI coincide in this case. Since (31) is geometric with parameter q, the AoI increases with $1/q$ as q decreases. As fresh samples from sensor 7 become scarcer, those from the other sensors become more useful and the 2D-AoI shows a significant improvement over the AoI for medium to small q.

In Fig. 9(b), we consider an update rate $\lambda = 1/300$ seconds^{-1} and an error probability $p_{\text{err}} = 0.1$ so that $q = 0.003$ seconds^{-1}. Fig. 9(b) shows the corresponding mean 2D-AoI along the freeway segment displayed in Fig. 7. Note that 2D-AoI is also defined for positions between sensors and for positions outside the sensor arrangement. The 2D-AoI shows small dips at the sensor positions and generally lower values at positions where sensor density is higher. Outside the sensor arrangement, the 2D-AoI increases quickly. For comparison, the mean AoI of any single sensor is $1/q \approx 333$ seconds, and the 2D-AoI shows the benefits of including additional sensors, even if they are remote.

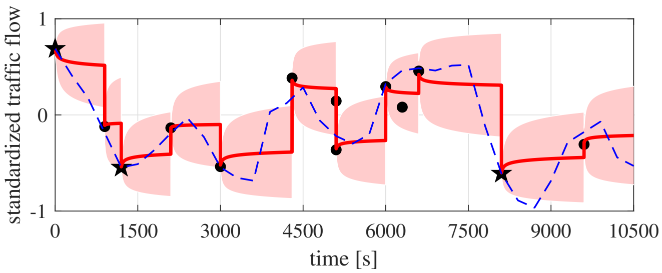
2) Prediction Evaluation: The 2D-AoI of a sensor sample determines how useful it is for predicting the process at a specific location. The prediction variance Φ is a function of the 2D-AoI (17), and the sample that has the minimal 2D-AoI across all sensors is the sample that minimizes the prediction variance (21). We first illustrate an example of the prediction before we calculate numerical averages.

This example uses PeMS data recorded on 15 May 2025, which was not included in training the kernel. The monitor receives a few randomly selected samples of the data and uses these to predict the traffic flow at sensor 14's position at postmile 516.71. Fig. 10 shows the traffic flow (dashed line), the samples that are used (dot markers), the prediction mean (solid line), and the prediction variance (colored area). The prediction is made either using the most recent sample from sensor 7 only (Fig. 10(a)), or the sample with the minimal 2D-AoI across all sensors (Fig. 10(b)).

The traffic data shows considerable variability over time. This was also reflected in the covariance analysis. The prediction in Fig. 10(a), which uses only samples taken at the position of interest, achieves moderate accuracy. This is due to the limited number of samples of sensor 14 and the characteristic decay of the temporal covariance, see Fig. 8. However, the spatial covariance between the sensors is relatively high (although it is also affected by the decay of the temporal



(a) 3 samples of sensor $s = 14$ at the point of interest



(b) 13 samples of sensors $s = 14, 9, 14, 13, 12, 11, 6, 12, 13, 2, 10, 14, 10$

Fig. 10. Prediction of the standardized traffic flow at sensor 14's position at postmile 516.71. Fig. 10(a) uses only the most recent samples from sensor 14, marked by stars. Fig. 10(b) uses the sample with the minimal 2D-AoI of all sensors. Samples from sensors other than sensor 14 are marked by dots. Some of these samples do not contribute, as they never reach the minimal 2D-AoI, i.e., the samples of sensor 6 at time 5100 and sensor 2 at time 6300.

correlation), and the prediction that uses the samples with the minimal 2D-AoI of all sensors performs significantly better. Samples from remote sensors differ more or less from the traffic flow at sensor 14's position and some of the samples do not contribute, as they do not reach the minimal 2D-AoI.

For Fig. 10 we used a very low sampling rate and took only a few samples so as not to overload the visualization. We evaluate the mean prediction variance for a (almost 40 times) higher rate of $q = 0.003$. If only the sample with the smallest AoI from sensor 14 is used, the mean prediction variance at position $x_{14} = 516.71$ is $\mathsf{E}[\Phi_t(14, 14)] = 0.31$. Using the best sample, i.e., the sample that has the smallest 2D-AoI of all sensors, reduces the variance to $\mathsf{E}[\Phi_t(14)] = 0.22$. Optionally, (13) can make a prediction using all samples $\Gamma_t(\varsigma, \tau)$ that are available at the monitor at time t. There is no analytical solution for this case and we have to use simulation. We set the time horizon of the monitor to $T = 3600$ seconds, i.e., the monitor does not use any samples that are older than one hour. On average, the monitor has $qTS = 151.2$ samples available. We ran the simulation for 10^6 time slots and obtained an average prediction variance $\mathsf{E}[\Phi_t(\Gamma, 14)] = 0.12$. We find comparable results for other positions, e.g., at $x_7 = 508.41$ in the middle of the freeway segment, we obtain $\mathsf{E}[\Phi_t(7, 7)] = 0.31$, $\mathsf{E}[\Phi_t(7)] = 0.19$, and $\mathsf{E}[\Phi_t(\Gamma, 7)] = 0.11$.

The improvement achieved by the sample with the smallest 2D-AoI of all sensors compared to the sample with the smallest AoI of the sensor at the position of interest leads to a slight increase in complexity. Selecting the best sample from (4) has a complexity in $\mathcal{O}(S)$ for S sensors, where the 2D-AoI and AeD are determined by insertion of the respective

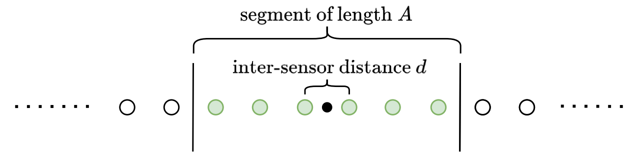


Fig. 11. Regular line topology. Sensors (green and white dots) are placed at an inter-sensor distance d. A monitor predicts the process at the point of interest (black dot) using the S sensors (green dots) that are located within a segment of length A. The communication capacity per segment is fixed.

kernel function into (16) and (18). However, if all available samples are used, the matrix inversion $\Sigma_{\mathbf{ff}}^{-1}$ in (13) dominates the complexity. The number of samples that can be used for prediction is upper bounded by ST and the complexity is in $\mathcal{O}((ST)^3)$ using, e.g., Gauss-Jordan elimination.

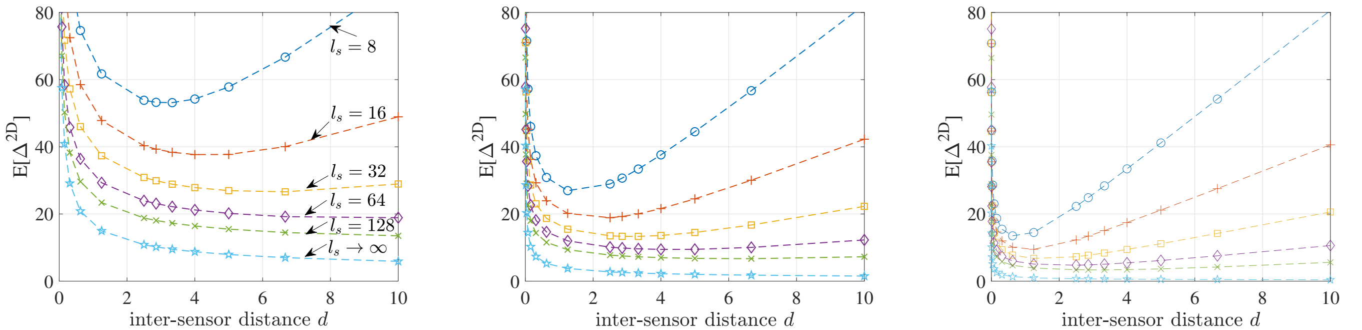
B. Spatial Density of Sensors under a Capacity Constraint

Next, we evaluate the impact of the spatial density of sensors on the 2D-AoI. To this end, we employ a line topology with equidistantly spaced sensors. The communication capacity is limited so that the service rate per sensor decreases as the spatial density of sensors increases. This limits the temporal intensity of sampling for each sensor. We evaluate this trade-off between spatial and temporal sampling intensity and present results that show the optimal density of sensors. Random sensor topologies, such as the one-dimensional Poisson point process studied in [22], can be taken into account by treating the AeD $\Lambda_t(\varsigma, s)$ as a random variable (due to the random sensor positions x_ς).

The line topology is illustrated in Fig. 11. The sensors are positioned at fixed intervals, separated by a distance of d. The line is divided into segments, each of length A. Each segment comprises S sensors, marked by green dots in Fig. 11. We only evaluate values of $S \in \{2, 4, 6, \dots\}$ where the sensors are arranged regularly along the segment with sensors at the positions $(i + 0.5)\, d$ for $i \in \{0, 1, \dots, S-1\}$ and $d = A/S$. For example, for $A = 40$ and $S = 4$ we have sensors at positions 5, 15, 25, and 35.

The sensor nodes of a segment are connected to the monitor via independent M/M/1 queues. Each segment has a fixed total communication capacity μ, so that the S sensors within the segment each have a service rate of $\mu_i = \mu/S$. The samples of the sensors are used to estimate a physical process at certain points of interest. We select the center of the segment at position $A/2$, marked by a black dot in Fig. 11, as the point of interest to evaluate its 2D-AoI. The distance of the nearest sensor to the point of interest is $d/2$.

In the evaluation, we set the segment length $A = 40$. We use a Gaussian product kernel composed of an exponential temporal and an exponential spatial kernel. The temporal correlation parameter is fixed to a value of $l_t = 128$ while we consider different values for the spatial correlation parameter $l_s \in \{8, 16, 32, 64, 128, \infty\}$. The range of l_s under consideration extends from one-fifth to approximately three times the segment length A. The variance of the process is set to $\sigma^2 = 1$. The sampling rate of each sensor equals the arrival rate to the



(a) $\mu = 1$ (b) $\mu = 4$ (c) $\mu = 16$

Fig. 12. Monitoring of a spatio-temporal Gaussian process with exponential product kernel, temporal correlation parameter $l_t = 128$ and different spatial correlation parameters l_s. The sensors are placed at an inter-sensor distance of d along a segment of length $A = 40$. The sensors are connected to the monitor via independent M/M/1 queues with a total service rate of μ.

queue λ_i, where $\lambda_i = \rho\mu_i$ and we set the utilization $\rho = 0.53$. We consider distance values in the range of $d \in (0, 10]$ and per segment capacity values $\mu \in \{1, 4, 16\}$. Numerical results are provided by substituting (9) for the M/M/1 queue into (24) for the exponential kernel. The expected value is obtained by integration of the tail (8). The results are shown in Fig. 12.

In all cases in Fig. 12, a significant influence of the sensor distance d on the expected 2D-AoI $\mathsf{E}[\Delta^{2D}]$ can be seen, which has a characteristic u-shape. The 2D-AoI is composed of AoI and AeD, and the u-shape of the 2D-AoI is a consequence of the fact that the AoI is a decreasing and the AeD an increasing function of d. If d is small, the spatial density of sensors is high and there are sensors in the vicinity of the point of interest that have a small AeD. However, since the sensors continue to cover the entire segment length A, not all of them are close to the point of interest. Due to the high density of sensors, the service rate per sensor is low, as is the sampling rate. As can be seen from (9), a small μ_i as well as a small λ_i result in a considerable increase of the AoI. In comparison, the AeD of nearby sensors is insignificant so that the AoI dominates the 2D-AoI for small d.

In contrast, when d is large, the spatial density of sensors is low, yet the sampling intensity is high, as more capacity is available to each sensor. As a result, the impact of the AoI decreases significantly, while the AeD becomes considerable and dominates the 2D-AoI. For the exponential kernel, the AeD is obtained from (23) as $\Lambda = |x_\varsigma - x_s|(l_t/l_s)$. For the sensors in closest proximity to the point of interest $\Lambda = (d/2)(l_t/l_s)$. At an inter-sensor distance $d = 10$, the AoI is small and the 2D-AoI curves are dominated by the AeD of the closest sensors that is $\Lambda = 5(l_t/l_s)$, respectively, 40 for $l_s = 16$. Due to their larger AeD (by a factor of 3 or more), the more distant sensors cannot make a noticeable contribution in this case. The convergence towards the AeD is faster when μ is larger, see Fig. 12(c), but it is also clearly visible in Figs. 12(b) and 12(a), where μ is smaller. For small μ the influence of the AoI remains, however, more noticeable.

Between the two extremes of very small and very large sensor distances, the ratio of AoI and AeD of sensors that are close to the point of interest is more balanced and we can determine a sensor distance d that minimizes the mean 2D-AoI for the given parameters (except for the special case $l_s \to \infty$, in which the spatial distance is irrelevant). In a concrete example for $\mu = 4$ and $l_s = 16$ in Fig. 12(b), the minimal expected 2D-AoI of 18.80 is found at a sensor distance of $d = 2.22$ corresponding to $S = 18$ sensors. Here, the AeD is 8.88 for the closest sensors.

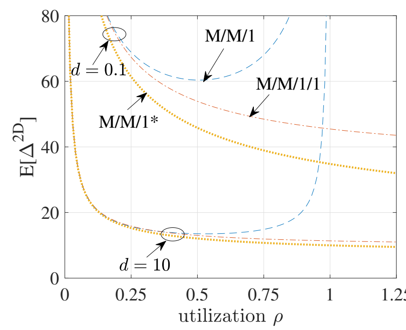


Fig. 13. Comparison of the expected 2D-AoI of M/M/1, M/M/1/1, and M/M/1* queues. The service rate is $\mu = 1$ as in Fig. 12(a), the spatial correlation parameter $l_s = 128$ and the inter-sensor distance $d \in \{0.1, 10\}$. The impact of the utilization ρ is evaluated.

A range of other queueing models are also possible, only the CCDF of the AoI of the chosen model is required. In Fig. 13 we compare the 2D-AoI of the line topology for M/M/1, M/M/1/1, and M/M/1* queues as introduced in Sec. IV. The parameters are $\mu = 1$ as in Fig. 12(a), $l_s = 128$, and $d \in \{0.1, 10\}$. Different from Fig. 12, we vary the sampling rate to evaluate the impact of the utilization ρ. In Fig. 12, we generally used $\rho = 0.53$ that is close to optimal for the case of M/M/1 queues. For M/M/1/1 and M/M/1* queues with loss and preemption, respectively, increasing the sampling rate improves the AoI even for $\rho > 1$, see Fig. 13. While Fig. 13 shows results for a small and a large sensor distance $d \in \{0.1, 10\}$, we also evaluated the effect of d for different ρ in detail. We refrain from presenting the results because, although the M/M/1/1 and M/M/1* queues show an improvement in the 2D-AoI, the shape of the curves is largely the same as that for M/M/1 queues in Fig. 12.

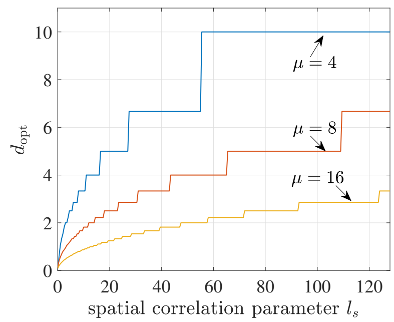


Fig. 14. Sensor distance d_{opt} that minimizes the mean 2D-AoI for $\mu \in \{1, 4, 16\}$, $l_t = 128$ and different l_s.

We show the sensor distance d_{opt} that minimizes the 2D-AoI of the M/M/1 queue in Fig. 14. It is evident that an increase of the overall capacity μ allows for greater spatial and temporal sampling intensity, which improves the 2D-AoI. As μ increases, the sensor distance that achieves the minimal 2D-AoI decreases. In addition, higher spatial correlation helps to improve the 2D-AoI, as samples from distant sensors still provide useful information. In this case it is beneficial to increase the intensity of temporal sampling at the expense of the spatial density. Consequently, the optimal sensor distance increases with l_s.

In the limit $l_s \to \infty$ the sensor distance is completely insignificant as the AeD is always zero regardless of d. In this case the S sensors can be viewed as one sensor that transmits samples via S parallel queues. The AoI of parallel systems was analyzed in [6], [52]. In brief, a main finding of [52] is that adding a second parallel M/M/1 queue with the same service rate reduces the mean AoI by a factor of approximately $5/8$. On the other hand, doubling the service rate of an M/M/1 queue reduces the mean AoI by half. In all cases, the utilization ρ that minimizes the AoI is chosen and the sum of the individual arrival rates $\lambda_1 + \lambda_2$ of the parallel system matches the arrival rate λ of the single system. While the mean update rate of both systems is the same, the single system with pooled capacity offers shorter service times. This becomes most obvious when considering an arrival at an empty system. In the case of a parallel system, this arrival has to be assigned to one queue, leaving the other queue idle. As a consequence, it is more effective to pool the capacity in a single server than to split it among parallel servers. This explains why the 2D-AoI for $l_s \to \infty$ decreases as the number of sensors S decreases, respectively, as the inter-sensor distance d increases, see Fig. 12.

IX. Conclusions

This work presented a model of AoI for two-dimensional processes. The physical phenomenon can be a spatio-temporal process observed by distributed sensors. By employing a Gaussian process model we have shown how the spatial distance between sensors can be transformed into an Age-equivalent Distance (AeD). The AeD can be a constant offset that does not change over time, depending on the covariance kernel, or it can diminish as time progresses. This means that a sample from a distant sensor becomes increasingly valuable when the AoI of a nearby sensor is high. By transforming distance into the domain of AoI, one can leverage a variety of established results for AoI and AoI of parallel systems. We presented evaluation results for different models of communications links, different network topologies, and different kernel functions. These make it possible to optimize sensor density, to decide on the necessity for additional sensors, and to configure the network capacity to ensure that desired performance criteria are met. The 2D-AoI model can be instantiated in different ways extending the utility of the established theory to a range of potential applications.

References

[1] S. Kaul, M. Gruteser, V. Rai, and J. Kenney, "Minimizing age of information in vehicular networks," in *Proc. of IEEE SECON*, Jun. 2011, pp. 350–358.

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

[3] R. D. Yates, Y. Sun, D. R. Brown, S. K. Kaul, E. Modiano, and S. Ulukus, "Age of information: An introduction and survey," *IEEE J. Sel. Areas Commun.*, vol. 39, no. 5, pp. 1183–1210, May 2021.

[4] A. Kosta, N. Pappas, and V. Angelakis, *Age of Information: A New Concept, Metric, and Tool.* now publishers, 2017.

[5] J. Pan, A. M. Bedewy, Y. Sun, and N. B. Shroff, "Age-optimal scheduling over hybrid channels," *IEEE Trans. Mob. Comput.*, pp. 1–17, 2022.

[6] M. Fidler, J. P. Champati, J. Widmer, and M. Noroozi, "Statistical age-of-information bounds for parallel systems: When do independent channels make a difference?" *IEEE J. Sel. Topics Inf. Theory*, vol. 4, pp. 591–606, Nov. 2023.

[7] R. Talak, S. Karaman, and E. Modiano, "Optimizing information freshness in wireless networks under general interference constraints," *IEEE/ACM Trans. Netw.*, vol. 28, no. 1, pp. 15–28, Feb. 2020.

[8] S. K. Kaul and R. D. Yates, "Status updates over unrelaible multiaccess channels," in *Proc. of IEEE ISIT*, 2017, pp. 331–335.

[9] N. Pappas and M. Kountouris, "Delay violation probabiliy and age of information interplay in the two-user multiple access channel," in *Proc. of IEEE SPAWC Workshop*, Jul. 2019, pp. 1–5.

[10] X. Chen, K. Gatsis, H. Hassani, and S. S. Bidokhti, "Age of information in random access channels," *IEEE Trans. Inf. Theory*, vol. 68, no. 10, pp. 6548–6568, Oct. 2022.

[11] A. Soysal and S. Ulukus, "Age of information in G/G/1/1 systems: Age expressions, bounds, special cases, and optimization," *IEEE Trans. Inf. Theory*, vol. 67, no. 11, pp. 7477–7489, 2021.

[12] Y. Inoue, H. Masuyama, T. Takine, and T. Tanaka, "A general formula for the stationary distribution of the age of information and its application to single-server queues," *IEEE Trans. Inf. Theory*, vol. 65, no. 12, pp. 8305–8324, Dec. 2019.

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

[14] L. Hu, Z. Chen, Y. Dong, L. Liang, and M. Wang, "Status update in IoT networks: Age-of-information violation probability and optimal update rate," *IEEE Internet Things J.*, vol. 8, no. 14, pp. 11 329–11 344, Jul. 2021.

[15] A. Rizk and J.-Y. L. Boudec, "A Palm calculus approach to the distribution of the age of information," *IEEE Trans. Inf. Theory*, vol. 69, no. 12, pp. 8097–8110, Dec. 2023.

[16] J. P. Champati, H. Al-Zubaidy, and J. Gross, "On the distribution of AoI for the GI/GI/1/1 and GI/GI/1/2* systems: Exact expressions and bounds," in *Proc. of IEEE INFOCOM*, Apr. 2019, pp. 37–45.

[17] ——, "Statistical guarantee optimization for AoI in single-hop and two-hop FCFS systems with periodic arrivals," *IEEE Trans. Commun.*, vol. 69, no. 1, pp. 365–381, Jan. 2021.

[18] M. Noroozi and M. Fidler, "A min-plus model of age-of-information with worst-case and statistical bounds," in *Proc. of IEEE ICC*, 2022.

[19] J. Hribar, M. Costa, N. Kaminski, and L. A. DaSilva, "Using correlated information to extend device lifetime," *IEEE Internet of Things Journal*, vol. 6, no. 2, pp. 2439–2448, 2018.

[20] Y. Li, Y. Xu, Q. Zhang, and Z. Yang, “Age-driven spatially temporally correlative updating in the satellite-integrated Internet of Things via Markov decision process,” *IEEE Internet of Things Journal*, vol. 9, no. 15, pp. 13 612–13 625, 2022.

[21] V. W. Håkansson, N. K. D. Venkategowda, S. Werner, and P. K. Varshney, “Optimal scheduling of multiple spatiotemporally dependent observations for remote estimation using age of information,” *IEEE Internet of Things Journal*, vol. 9, no. 20, pp. 20 308–20 321, Oct. 2022.

[22] Z. Jiang and S. Zhou, “Status from a random field: How densely should one update?” in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, 2019, pp. 1037–1041.

[23] R. V. Ramakanth, V. Tripathi, and E. Modiano, “Monitoring correlated sources: AoI-based scheduling is nearly optimal,” in *Proc. of IEEE INFOCOM*, May 2024.

[24] M. Fidler, F. Gallistl, J. P. Champati, and J. Widmer, “2D-AoI: Age-of-information of distributed sensors for spatio-temporal processes,” Tech. Rep. arXiv:2412.12789, Dec. 2024.

[25] Y. Sun, Y. Polyanskiy, and E. Uysal, “Sampling of the Wiener process for remote estimation over a channel with random delay,” *IEEE Trans. Inf. Theory*, vol. 66, no. 2, pp. 1118–1135, Feb. 2020.

[26] T. Z. Ornee and Y. Sun, “Sampling and remote estimation for the Ornstein-Uhlenbeck process through queues: Age of information and beyond,” *IEEE/ACM Trans. Netw.*, vol. 29, no. 5, pp. 1962–1975, 2021.

[27] J. P. Champati, M. Mamduhi, K. Johansson, and J. Gross, “Performance characterization using AoI in a single-loop networked control system,” in *Proc. of IEEE INFOCOM AoI Workshop*, Apr. 2019, pp. 197–203.

[28] M. Klügel, M. H. Mamduhi, S. Hirche, and W. Kellerer, “AoI-penalty minimization for networked control systems with packet loss,” in *Proc. of IEEE INFOCOM AoI Workshop*, Apr. 2019, pp. 189–196.

[29] O. Ayan, M. Vilgelm, M. Klügel, S. Hirche, and W. Kellerer, “Age-of-information vs. value-of-information scheduling for cellular networked control systems,” in *Proc. of ACM/IEEE ICCPS*, Apr. 2019.

[30] E. Altman, R. El-Azouzi, D. S. Menasche, and Y. Xu, “Forever young: Aging control for hybrid networks,” in *Proc. of ACM Mobihoc*, Jul. 2019, pp. 91–100.

[31] 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.

[32] A. Maatouk, S. Kriouile, M. Assaad, and A. Ephremides, “The age of incorrect information: A new performance metric for status updates,” *IEEE/ACM Trans. Netw.*, vol. 28, no. 5, pp. 2215–2228, Oct. 2020.

[33] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, “Age and value of information: Non-linear age case,” in *Proc. of IEEE International Symposium on Information Theory (ISIT)*, 2017, pp. 326–330.

[34] A. Krause, E. Horvitz, A. Kansal, and F. Zhao, “Toward community sensing,” in *Proc. of ACM/IEEE IPSN*, 2008.

[35] R. D. Yates and S. Kaul, “Real-time status updating: Multiple sources,” in *Proc. of IEEE ISIT*, 2012, pp. 2666–2670.

[36] Y. Sun, E. Uysal-Biyikoglu, and S. Kompella, “Age-optimal updates of multiple information flows,” in *Proc. of IEEE INFOCOM AoI Workshop*, 2018.

[37] A. Javani, M. Zorgui, and Z. Wang, “Age of information in multiple sensing,” in *Proc. of IEEE Globecom*, 2019.

[38] E. Gindullina, L. Badia, and D. Gündüz, “Age-of-information with information source diversity in an energy harvesting system,” *IEEE Trans. Green Com. and Netw.*, vol. 5, no. 3, pp. 1529–1540, Sep. 2021.

[39] Z. Chen, D. Deng, H. H. Yang, N. Pappas, L. Hu, Y. Jia, M. Wang, and T. Q. S. Quek, “Analysis of age of information in dual updating systems,” *IEEE Trans. Wireless Commun.*, vol. 22, no. 11, pp. 8003–8019, Nov. 2023.

[40] B. Zhou and W. Saad, “On the age of information in internet of things systems with correlated devices,” in *Proc. of IEEE Globecom*, 2020.

[41] Q. He, G. Dán, and V. Fodor, “Joint assignment and scheduling for minimizing age of correlated information,” *IEEE/ACM Trans. Netw.*, vol. 27, no. 5, pp. 1887–1900, 2019.

[42] A. E. Kalør and P. Popovski, “Minimizing the age of information from sensors with common observations,” *IEEE Wireless Communications Letters*, vol. 8, no. 5, pp. 1390–1393, 2019.

[43] ——, “Timely monitoring of dynamic sources with observations from multiple wireless sensors,” *IEEE/ACM Trans. Netw.*, vol. 31, no. 3, pp. 1263–1276, Jun. 2023.

[44] V. Tripathi and E. Modiano, “Optimizing age of information with correlated sources,” *IEEE/ACM Trans. Netw.*, vol. 32, no. 6, pp. 4660–4675, Dec. 2024.

[45] J. Tong, L. Fu, and Z. Han, “Age-of-information oriented scheduling for multichannel IoT systems with correlated sources,” *IEEE Trans. Wireless Commun.*, vol. 21, no. 11, pp. 9775–9790, Nov. 2022.

[46] J. Heinovski, J. T. Gómez, and F. Dressler, “Focusing on information context for ITS using a spatial age of information model,” *Computer Communications*, vol. 209, pp. 203–216, 2023.

[47] A. Zancanaro, G. Cisotto, and L. Badia, “Modeling value of information in remote sensing from correlated sources,” *Computer Communications*, vol. 203, pp. 289–297, Apr. 2023.

[48] J. Holm, F. Chiariotti, A. E. Kalør, B. Soret, T. B. Pedersen, and P. Popovski, “Goal-oriented scheduling in sensor networks with application timing awareness,” *IEEE Trans. Commun.*, vol. 71, no. 8, pp. 4513–4527, Aug. 2023.

[49] J. Luo and N. Pappas, “Semantic-aware remote estimation of multiple Markov sources under constraints,” Tech. Rep. arXiv:2403.16855v1, Mar. 2024.

[50] L. Crosara and L. Badia, “Cost and correlation in strategic wireless sensing driven by age of information,” in *Proc. of European Wireless Conference*, Sep. 2022.

[51] L. Crosara, A. Zancanaro, G. Cisotto, N. Laurenti, and L. Badia, “Analytical evaluation of age of information in networks of correlated sources,” in *IEEE Workshop on Metrology for Agriculture and Forestry*, Nov. 2022, pp. 323–328.

[52] A. Bhati, S. R. B. Pillai, and R. Vaze, “On the age of information of a queuing system with heterogeneous servers,” in *National Conference on Communications, Kanpur, India*, 2021.

[53] C. E. Rasmussen and K. I. Williams, *Gaussian Processes for Machine Learning*. MIT Press, 2006.

[54] R. Turner, “Gaussian processes,” https://www.youtube.com/watch?v=92-98SYOdlY, Nov. 2016. [Online]. Available: https://www.youtube.com/watch?v=92-98SYOdlY

[55] D. K. Duvenaud, “Automatic model construction with gaussian processes,” Ph.D. dissertation, Univ. of Cambridge, Jun. 2014.

[56] A. Aneiba, J. Hayes, B. Nangle, and M. Albaarini, “Real-time IoT urban road traffic data monitoring using LoRaWAN,” in *Proc. of International Conference on the Internet of Things*, Oct. 2019.

[57] C. D. of Transportation, “Caltrans performance measurement system (PeMS),” https://pems.dot.ca.gov/, 2025.

[58] OpenStreetMap, https://www.openstreetmap.org/, 2025.

[59] A. M. El-Geneidy and R. L. Bertini, “Toward validation of freeway loop detector speed measurements using transit probe data,” in *Proc. of IEEE Intelligent Transportation Systems Conference*, Oct. 2004, pp. 779–784.

[60] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, “Scikit-learn: Machine learning in Python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[61] GPy, “GPy: A Gaussian process framework in python,” http://github.com/SheffieldML/GPy, since 2012.

[62] J. R. Gardner, G. Pleiss, D. Bindel, K. Q. Weinberger, and A. G. Wilson, “GPyTorch: Blackbox matrix-matrix Gaussian process inference with GPU acceleration,” in *Advances in Neural Information Processing Systems*, 2018.

Markus Fidler received the doctoral degree in computer engineering from RWTH Aachen University, Germany, in 2004. He was a Post-Doctoral Fellow of the Mittag-Leffler Institute Stockholm, Sweden, in 2004, NTNU Trondheim, Norway, in 2005 and the University of Toronto, Canada, in 2006. From 2007, he was a DFG Emmy Noether Group Leader at the Technical University of Darmstadt, Germany. Since 2009, he has been a Professor of Communication Networks at Leibniz University Hannover, Germany. In 2012, he received an ERC Starting Grant.

Flavio Gallistl has obtained his master's degree in computer engineering from Leibniz University Hannover, Germany, in 2023. He is currently pursuing his Ph.D. with the Institute of Communications Technology at Leibniz University Hannover, Germany since 2023. His research interests include Age-of-Information and deterministic communication systems.

Jaya Prakash Champati is an assistant professor at the Department of Computer Science, University of Victoria (UVic), Canada. Before his current position, he was a research assistant professor at IMDEA Networks Institute in Spain. He did his postdoctoral research at EECS, KTH Royal Institute of Technology in Sweden, where he significantly contributed to the field of Age of Information Analysis and Optimization. He earned his PhD in ECE from the University of Toronto, Canada, in 2017, where his research on scheduling parallel processors earned him the Doctoral Completion Award and the Paul Biringer Scholarship. Dr. Champati holds a master's degree in technology from the Indian Institute of Technology (IIT) Bombay, India. He also has industry experience at Broadcom Communications, where he played a key role in developing the 4G LTE MAC layer. He was a Marie Skłodowska-Curie Actions (MSCA) postdoctoral fellow and received the best paper award at the IEEE National Conference on Communications in India in 2011. His current research focus is on efficient device-centric AI.

Joerg Widmer is Research Professor and Research Director of IMDEA Networks in Madrid, Spain. Before, he held positions at DOCOMO Euro-Labs in Munich, Germany and EPFL, Switzerland. He was a visiting researcher at the International Computer Science Institute in Berkeley, USA, University College London, UK, and TU Darmstadt, Germany. His research focuses on wireless networks, ranging from extremely high frequency millimeter-wave communication and MAC layer design to mobile network architectures. Joerg Widmer authored more than 250 conference and journal papers and three IETF RFCs, and holds 14 patents. He was awarded an ERC consolidator grant, the Friedrich Wilhelm Bessel Research Award of the Alexander von Humboldt Foundation, a Mercator Fellowship of the German Research Foundation, a Spanish Ramon y Cajal grant, as well as 16 best paper awards. He is an IEEE Fellow and Distinguished Member of the ACM.