On the Value of Information and Mean Squared Error for Noisy Gaussian Models

Zijing Wang, *Student Member, IEEE,* Mihai-Alin Badiu, *Member, IEEE,* and Justin P. Coon, *Senior Member, IEEE*

***Abstract*—The relationship between the age of information (AoI) and the mean squared error (MSE) in optimisation problems has been widely investigated for various Gaussian Markov models. Recently, we put the AoI in an information-theoretic context and proposed a mutual information-based value of information (VoI) framework to characterise how valuable the status updates are. In this paper, we consider general noisy Gaussian models and investigate the connection between the VoI and MSE. We show that the VoI can be presented by a logarithmic function of the MSE. We also study the rate of change of the two with respect to the signal-to-noise ratio (SNR) and the AoI. The results hold for both discrete-time and continuous-time stationary Gaussian processes. This paper illuminates a useful relationship between information theory and signal processing theory in the presence of the AoI and additive noise, and provides a possible way to infer the information value by estimating the MSE for practical applications.**

***Index Terms*—Value of information, mean squared error, age of information, noisy Gaussian models.**

I. Introduction

THE timeliness of data is of great importance for next-generation wireless communication systems. Innovative real-time applications largely rely on time-critical information for timely monitoring and control. A metric named age of information (AoI) has been introduced to characterise the data freshness at the receiver [1]. It is defined as the time difference between the current time and the generation time of the latest received data sample. The AoI and its applications have been widely investigated in various queueing models and wireless networks [2]–[5]. However, the AoI may not be useful in all applications since it is a process-independent metric that grows linearly with unit slope. This means that the AoI is unable to capture non-linearities caused by correlation properties of the underlying physical process.

For the reasons mentioned above, other measures have been proposed in the literature. One method is to use monotonic non-linear AoI functions to measure the freshness or staleness of data. Exponential, logarithmic and step functions have been commonly used as the AoI-dependent metric [6], [7]. The other method is to employ information theory or signal processing theory to characterise the importance of fresh information. The estimation error [8]–[13], mutual information [14] and conditional entropy [15] have been utilised as non-linear functions to characterise the information value. The relationship between the AoI and the mean squared error (MSE) for remote estimation has been investigated. The authors in [9]–[11] showed that the MSE could be written as a function of the AoI for a Wiener process, an Ornstein-Uhlenbeck (OU) process and a linear time-invariant (LTI) system, respectively.

Despite their novelty and contributions, existing works mostly focus on Markov models in which the system performance only relates to the most recent received data sample and samples are assumed to be observable at the receiver. Since the status updates can be negatively affected by noise or other features in the transmission system, it may not be reasonable to assume that they are directly visible. In [16], the authors explored the performance of the MSE in the presence of the noise but only considered the latest single observation. However, in this case, all the past observations hold valuable information about the underlying random process. Therefore, we require a more appropriate metric to evaluate the information value for latent variable models.

In our previous work [17], [18], a mutual information-based value of information (VoI) framework was proposed to characterise how valuable the status updates are for latent variable models. We explored this framework in the context of a specific noisy Gaussian process (Ornstein–Uhlenbeck process) and studied its relationship to the AoI.

In this work, we focus on more general noisy Gaussian models to investigate the VoI and its connection to MSE. In [19], an important relationship between MSE and the input-output mutual information was established in the absence of transmission delay. In that work, data samples are treated individually and the data value of each sample does not change with time. In this work, we extend this result to status update systems in the presence of transmission delay. We show that the VoI equals a logarithmic function of MSE, a result that holds for both discrete-time and continuous-time Gaussian processes. We also explore the connection between the rate of change of the VoI and MSE under high and low correlation conditions. This paper illuminates interesting relationships between the timeliness of information and remote estimation in an information-theoretic context, which provides a possible way to infer the information value by utilising the estimation error.

II. VoI and MSE Formulation

We consider the single transmitter–receiver case in which the sensor device is deployed to observe a random process

This work was supported by EPSRC grant number EP/T02612X/1. For the purpose of open access, the authors have applied a Creative Commons Attribution (CC BY) license to any Accepted Manuscript version arising. The authors also gratefully acknowledge the support of the Clarendon Fund Scholarships of the University of Oxford.

The authors are with the Department of Engineering Science, University of Oxford, Oxford OX1 3PJ, U. K. (e-mail: {zijing.wang, mihai.badiu, and justin.coon}@eng.ox.ac.uk).

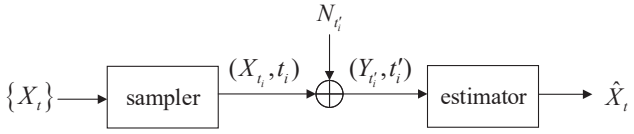


Fig. 1. System model.

$\{X_t\}$. The sensor needs to continuously sample data to get the latest status update of a physical phenomenon and transmit data samples in a timely manner to the receiver. The system model is shown in Fig. 1. We denote X_{t_i} as the status update of the underlying random process and t_i as its generation time at the transmitter. Samples are observed through an additive noise channel. We denote $Y_{t_i'}$ as the corresponding observation and t_i' as its receiving time at the receiver[1]. The estimator uses the observations to provide the optimal estimation $\hat{X}_t$ of the current status of the underlying process. Due to the transmission delay and the noise in the transmission channel, we have $t_i' > t_i$ and $Y_{t_i'} \neq X_{t_i}$. For a given time instant t, we denote n as the index of the most recent received data sample (i.e., $t > t_n'$), and n is also the total number of data samples observed at the receiver. In this case, the AoI is given as $t - t_n$.

For given time instants t, t_i and t_i' $(1 \leq i \leq n)$, the value of information at time t is defined as [17]

$$v(t) = \mathrm{I}(X_t; Y_{t_n'}, \cdots, Y_{t_{n-m+1}'}), \quad t > t_n'. \tag{1}$$

Here, X_t is the current status of the random process at the transmitter, m represents the memory of the system, and $\{Y_{t_n'}, \cdots, Y_{t_{n-m+1}'}\}$ is the collection of the most recent m of n observations that are captured at the receiver. The concept of the value of information is defined as the mutual information between the current status of the underlying hidden process and a sequence of past noisy measurements. This definition is interpreted as the reduction in the uncertainty of the current hidden status based on some noisy observations. The VoI characterises how valuable the status updates are from the receiver's perspective.

For remote estimation of the underlying hidden process, the mean squared error at time t is defined as

$$\mathrm{mse}(t) = \mathrm{E}[(X_t - \hat{X}_t)^2] \tag{2}$$

where $\hat{X}_t$ is the estimator of the current status X_t. The minimum mean squared error (MMSE) is achieved when

$$\hat{X}_t = \mathrm{E}[X_t | Y_{t_n'}, \cdots, Y_{t_{n-m+1}'}]. \tag{3}$$

The MSE measures the error of an estimator of the current unobserved status of the underlying process given that we have some past noisy observations.

III. VoI and MSE for Noisy Gaussian Models

A. Noisy Gaussian Model

The underlying hidden process $\{X_t\}$ is considered to be a stationary, Gaussian process. We assume that samples $\{X_{t_i}\}$ are Gaussian variables with zero mean and constant variance σ_x^2. Due to the property of stationarity, we denote ρ_τ $(0 < \rho_\tau < 1)$ as the auto-correlation coefficient of two variables X_{t_i} and X_{t_j} where $\tau = |t_j - t_i|$ (Theorem 6.8.2, [20]), i.e.,

$$\rho_\tau = \frac{\mathrm{E}[X_{t_i} X_{t_j}]}{\sigma_x^2}. \tag{4}$$

If X_{t_i} and X_{t_j} are less correlated, ρ_τ approaches 0; otherwise, it approaches 1. Let vector $\boldsymbol{X} = [X_{t_n}, \cdots, X_{t_{n-m+1}}]^{\mathrm{T}}$ be the sequence of status updates of the underlying random process and let $\mathbf{K}_{\boldsymbol{X}}$ be its auto-correlation matrix, i.e.,

$$\mathbf{K}_{\boldsymbol{X}} = \begin{bmatrix} 1 & \rho_{t_n - t_{n-1}} & \cdots & \rho_{t_n - t_{n-m+1}} \\ \rho_{t_n - t_{n-1}} & 1 & \cdots & \rho_{t_{n-1} - t_{n-m+1}} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{t_n - t_{n-m+1}} & \rho_{t_{n-1} - t_{n-m+1}} & \cdots & 1 \end{bmatrix}. \tag{5}$$

We assume that this random process is observed through an additive Gaussian noise channel. We denote $N_{t_i'}$ as the noise sample induced by the receiver at t_i'. Then, the corresponding noisy measurement is defined as

$$Y_{t_i'} = X_{t_i} + N_{t_i'}. \tag{6}$$

The samples $\{N_{t_i'}\}$ are assumed to be independent and identically distributed (i.i.d.) Gaussian variables with zero mean and constant variance σ_n^2. Let γ be

$$\gamma = \frac{\sigma_x^2}{\sigma_n^2} \tag{7}$$

which is used to represent the signal-to-noise ratio (SNR) over the Gaussian channel. Similarly, let vectors $\boldsymbol{Y} = [Y_{t_n'}, \cdots, Y_{t_{n-m+1}'}]^{\mathrm{T}}$ and associated $\boldsymbol{N} = [N_{t_n'}, \cdots, N_{t_{n-m+1}'}]^{\mathrm{T}}$ be the noisy observations and noise samples, respectively. Let $\boldsymbol{\Sigma}_{\boldsymbol{Y}}$ be the auto-covariance matrix of $\boldsymbol{Y}$. Thus, observations and covariance matrix can be collectively represented by

$$\boldsymbol{Y} = \boldsymbol{X} + \boldsymbol{N}, \quad \boldsymbol{\Sigma}_{\boldsymbol{Y}} = \sigma_x^2 \left(\mathbf{K}_{\boldsymbol{X}} + \frac{1}{\gamma} \mathbf{I} \right) \tag{8}$$

where $\mathbf{I}$ is the identity matrix.

B. Value of Information

Let the m-dimensional vector $\boldsymbol{u}$ be given by

$$\boldsymbol{u} = \left[\rho_{t-t_n}, \rho_{t-t_{n-1}}, \cdots, \rho_{t-t_{n-m+1}} \right]^{\mathrm{T}}. \tag{9}$$

Lemma 1. *The VoI for the noisy Gaussian model is given as*

$$v(t) = -\frac{1}{2} \log \left[1 - \boldsymbol{u}^{\mathrm{T}} \left(\mathbf{K}_{\boldsymbol{X}} + \frac{1}{\gamma} \mathbf{I} \right)^{-1} \boldsymbol{u} \right]. \tag{10}$$

Proof: See the appendix. ∎

Here, γ is the SNR in the transmission environment. The vector $\boldsymbol{u}$ represents the correlation between the current status and all the m most recent observations. The first element ρ_{t-t_n} relates to the time difference $t - t_n$, which is the AoI. When the AoI is very small or the underlying random process is highly correlated, ρ_{t-t_n} is close to 1.

[1]Here, $\{X_t\}$ can be either a discrete-time or a continuous-time random process. For the discrete-time random process, $\{t_i\}$ and $\{t_i'\}$ are positive integers; for the continuous-time random process, $\{t_i\}$ and $\{t_i'\}$ are positive real values.

Lemma 1 gives the VoI expression and links it to the AoI for general noisy Gaussian models. The VoI relates the correlation of the underlying random process, the noise over the transmission channel, the number of observations, and the AoI. Similar results hold for specific Gaussian hidden Markov models; details can be found in our previous work [17], [18]. From the operational perspective, the VoI can be interpreted as how many bits of information are contained in a latent state by observing a sequence of visible noisy measurements.

C. Mean Squared Error

For remote estimation, the estimator is the conditional expectation given some noisy measurements, i.e., $\hat{X}_t = \mathrm{E}[X_t|\boldsymbol{Y}=\boldsymbol{y}]$, which is given by (Theorem 11.1, [21])

$$\hat{X}_t = \sigma_x^2 \boldsymbol{u}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} \boldsymbol{y}. \tag{11}$$

Based on (11), we can state the following result.

Lemma 2. *The MSE for the noisy Gaussian model is given by*

$$\mathrm{mse}(t) = \sigma_x^2 \left[1 - \boldsymbol{u}^{\mathrm{T}} \left(\mathbf{K}_{\boldsymbol{X}} + \frac{1}{\gamma}\mathbf{I} \right)^{-1} \boldsymbol{u} \right]. \tag{12}$$

Proof:

$$\begin{aligned} \mathrm{mse}(t) =& \mathrm{E}[(X_t - \hat{X}_t)^2] \\ =& \mathrm{E}[(X_t - \sigma_x^2 \boldsymbol{u}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} \boldsymbol{y})^2] \\ =& \mathrm{E}[X_t^2] + \sigma_x^4 \boldsymbol{u}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} \mathrm{E}[\boldsymbol{y}\boldsymbol{y}^{\mathrm{T}}] \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} \boldsymbol{u} \\ & - 2\sigma_x^2 \boldsymbol{u}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} \mathrm{E}[X_t \boldsymbol{y}] \\ =& \sigma_x^2 + \sigma_x^4 \boldsymbol{u}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} \boldsymbol{\Sigma}_{\boldsymbol{Y}} \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} \boldsymbol{u} - 2\sigma_x^4 \boldsymbol{u}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} \boldsymbol{u} \\ =& \sigma_x^2 (1 - \sigma_x^2 \boldsymbol{u}^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{Y}}^{-1} \boldsymbol{u}). \end{aligned} \tag{13}$$

■

The MSE in Lemma 2 shows a similar behaviour to the VoI in Lemma 1. It also relates the SNR γ, the auto-correlation matrix $\mathbf{K}_{\boldsymbol{X}}$, and the vector $\boldsymbol{u}$. For the single observation case, it is easy to find that large SNR or highly correlated data source can lead to small MSE. Moreover, we have the following result regarding the connection between the MSE and the AoI in this case.

Remark 1. *If the correlation coefficient ρ_τ is non-increasing when τ increases, the MSE is a non-decreasing function of the AoI.*

Proof: The derivative of the MSE with respect to vector $\boldsymbol{u}$ is given as

$$\frac{\partial \,\mathrm{mse}}{\partial \boldsymbol{u}} = -2\sigma_x^2 \boldsymbol{u}^{\mathrm{T}} \left(\mathbf{K}_{\boldsymbol{X}} + \frac{1}{\gamma}\mathbf{I} \right)^{-1}. \tag{14}$$

For the single observation case, vector $\boldsymbol{u}$ is ρ_{t-t_n}, which is non-increasing with the AoI $\Delta = t - t_n$ and we have

$$\frac{\partial \rho_{t-t_n}}{\partial \Delta} \leq 0. \tag{15}$$

Thus, the derivative of the MSE with respect to the AoI is given as

$$\frac{\partial \,\mathrm{mse}}{\partial \Delta} = \frac{\partial \,\mathrm{mse}}{\partial \rho_{t-t_n}} \frac{\partial \rho_{t-t_n}}{\partial \Delta} = -2\sigma_x^2 \rho_{t-t_n} \frac{\gamma}{1+\gamma} \frac{\partial \rho_{t-t_n}}{\partial \Delta} \geq 0. \tag{16}$$

■

Remark 1 shows that the MSE can be presented by a non-decreasing function of the AoI for noisy Gaussian models. This remark holds for specific Gaussian processes (e.g., Wiener process [9] and OU process [10]). It also holds for exponentially correlated data source in LTI systems [11].

IV. Relationship between VoI and MSE

Based on Lemmas 1 and 2, we can state the following main result.

Proposition 1. *For a fixed time instant t, the relationship between the VoI and the MSE for noisy Gaussian models satisfies*

$$v(t) = \frac{1}{2} \log \frac{\sigma_x^2}{\mathrm{mse}(t)}. \tag{17}$$

This proposition links the VoI to a logarithmic function of the MSE. In practice, if we think about a communication system and try to determine how valuable the status update are, it may not be easy to estimate the mutual information but it is easy to estimate the MSE. Therefore, by estimating the MSE and using this link, it is possible to infer what the value of information is.

Proposition 1 illustrates the general relationship between the VoI and the MSE with multiple observations. In the following part, we consider the single observation case to explore this result further. The VoI and MSE can be calculated by replacing the m-dimensional vector $\boldsymbol{u}$ and matrix $\mathbf{K}_{\boldsymbol{X}}$ with the single variable ρ_{t-t_n} and 1, respectively. Then, we have the following corollaries.

Corollary 1. *If the latest sample X_{t_n} and the current status X_t of the underlying random process are highly correlated (i.e., $\rho_{t-t_n} \to 1$), the relationship between the VoI and MSE can be written as*

$$\frac{\mathrm{d}v}{\mathrm{d}\gamma} = \frac{\mathrm{mse}}{2\sigma_x^2}. \tag{18}$$

If the latest sample X_{t_n} and the current status X_t of the underlying random process are uncorrelated (i.e., $\rho_{t-t_n} \to 0$), we have

$$\mathrm{mse} \to \sigma_x^2, \quad \frac{\mathrm{d}v}{\mathrm{d}\gamma} \to 0. \tag{19}$$

In this corollary, (18) links the derivative of the VoI with respect to the SNR to the MSE for highly correlated data samples. A high correlation condition means that either the AoI is very small (i.e., $t - t_n \to 0$) or the underlying data source does not change quickly with time (e.g., the random process $\{X_t\}$ is a constant). This relationship covers the result in [19], which links the derivative of the input-output mutual information to the MSE in the absence of transmission delay. For less correlated data samples, (19) shows that the MSE is independent of the SNR. Moreover, a low correlation condition yields zero rate of change of the VoI with increasing SNR.

Corollary 2. *The derivative of the VoI with respect to the AoI can be written as*

$$\frac{\partial v}{\partial \Delta} = -\frac{1}{2\,\mathrm{mse}} \frac{\partial \,\mathrm{mse}}{\partial \Delta}. \tag{20}$$

This corollary shows that the rate of change of the VoI is not equivalent to the MSE when increasing the AoI. Based on Remark 1, it is easy to see that the VoI is a non-increasing function of the AoI. When the estimation error is large, the rate of change of the VoI is 0, and it does not relate to the AoI.

V. Conclusions

In this paper, we studied the connection between the VoI and the MSE for general noisy Gaussian models. The proposed mutual information-based VoI framework represents the reduction in the uncertainty of the current status given some past noisy observations. This paper showed that the VoI can be presented as a logarithmic function of the MSE, which gives a possible method to evaluate the information value by estimating the MSE for practical applications. Moreover, the rate of change of the VoI and MSE with respect to the SNR and the AoI were investigated in high and low correlation regimes, which further illustrated connections among the AoI, VoI and MSE. Future work can be focused on exploring the effect of different autocorrelation functions on the VoI and its relationship to the MSE.

Appendix

The VoI definition given in (1) follows from the relation $\mathrm{I}(X_t;\boldsymbol{Y}) = \mathrm{h}(X_t) + \mathrm{h}(\boldsymbol{Y}) - \mathrm{h}(X_t,\boldsymbol{Y})$. Since $(X_t,\boldsymbol{Y}^{\mathrm{T}})$ is multivariate Gaussian and stationary distributed, the VoI can be further written as (Example 8.5.1, [22])

$$v(t) = \frac{1}{2}\log\frac{\sigma_x^2\det(\boldsymbol{\Sigma}_{\mathbf{Y}})}{\det(\boldsymbol{\Sigma}_{X_t,\mathbf{Y}})} \tag{21}$$

where $\boldsymbol{\Sigma}_{X_t,\mathbf{Y}}$ is the covariance matrix of $(X_t,\boldsymbol{Y}^{\mathrm{T}})^{\mathrm{T}}$. The covariance between X_t and $Y_{t_i'}$ is given as

$$\mathrm{E}[X_tY_{t_i'}] = \mathrm{E}[X_t(X_{t_i}+N_{t_i'})] = \mathrm{E}[X_tX_{t_i}] = \sigma_x^2\rho_{t-i}. \tag{22}$$

Therefore, $\boldsymbol{\Sigma}_{X_t,\mathbf{Y}}$ can be presented by $\boldsymbol{\Sigma}_{\mathbf{Y}}$, i.e.,

$$\boldsymbol{\Sigma}_{X_t,\mathbf{Y}} = \begin{bmatrix} \sigma_x^2 & \sigma_x^2\boldsymbol{u}^{\mathrm{T}} \\ \sigma_x^2\boldsymbol{u} & \boldsymbol{\Sigma}_{\mathbf{Y}} \end{bmatrix}. \tag{23}$$

Substituting (23) into (21), the VoI is given as

$$\begin{aligned} v(t) &\overset{(a)}{=} \frac{1}{2}\log\frac{\sigma_x^2\det(\boldsymbol{\Sigma}_{\mathbf{Y}})}{\sigma_x^2\det(\boldsymbol{\Sigma}_{\mathbf{Y}}-\sigma_x^2\boldsymbol{u}\boldsymbol{u}^{\mathrm{T}})} \\ &\overset{(b)}{=} \frac{1}{2}\log\frac{1}{1-\sigma_x^2\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}\boldsymbol{u}} \\ &= -\frac{1}{2}\log\left(1-\sigma_x^2\boldsymbol{u}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathbf{Y}}^{-1}\boldsymbol{u}\right) \end{aligned} \tag{24}$$

where (a) is obtained by the block matrix determinant and (b) is obtained by the determinant lemma.

References

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

[2] R. D. Yates and S. K. Kaul, "The age of information: Real-time status updating by multiple sources," *IEEE Trans. Inf. Theory*, vol. 65, no. 3, pp. 1807–1827, 2019.

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

[4] Y. Sun, I. Kadota, R. Talak, and E. Modiano, *Age of Information: A New Metric for Information Freshness*. San Rafael, CA, USA: Morgan & Claypool, 2019.

[5] 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, 2021.

[6] Y. Sun and B. Cyr, "Sampling for data freshness optimization: Non-linear age functions," *J. Commun. Netw.*, vol. 21, no. 3, pp. 204–219, 2019.

[7] A. Kosta, N. Pappas, A. Ephremides, and V. Angelakis, "The cost of delay in status updates and their value: Non-linear ageing," *IEEE Trans. Commun.*, vol. 68, no. 8, pp. 4905–4918, 2020.

[8] C. Kam, S. Kompella, G. D. Nguyen, J. E. Wieselthier, and A. Ephremides, "Towards an effective age of information: Remote estimation of a Markov source," in *Proc. IEEE INFOCOM Workshops*, 2018, pp. 367–372.

[9] 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, 2020.

[10] 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.

[11] J. P. Champati, M. H. Mamduhi, K. H. Johansson, and J. Gross, "Performance characterization using AoI in a single-loop networked control system," in *Proc. IEEE INFOCOM Workshops*, 2019, pp. 197–203.

[12] 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. ACM/IEEE ICCPS*, 2019, p. 109–117.

[13] M. Wang, W. Chen, and A. Ephremides, "Real-time reconstruction of a counting process through first-come-first-serve queue systems," *IEEE Trans. Inf. Theory*, vol. 66, no. 7, pp. 4547–4562, 2020.

[14] Y. Sun and B. Cyr, "Information aging through queues: A mutual information perspective," in *Proc. IEEE SPAWC*, 2018, pp. 1–5.

[15] T. Soleymani, S. Hirche, and J. S. Baras, "Optimal self-driven sampling for estimation based on value of information," in *International Workshop on Discrete Event Systems (WODES)*, 2016, pp. 183–188.

[16] A. Arafa, K. Banawan, K. G. Seddik, and H. V. Poor, "Sample, quantize, and encode: Timely estimation over noisy channels," *IEEE Trans. Commun.*, vol. 69, no. 10, pp. 6485–6499, 2021.

[17] Z. Wang, M.-A. Badiu, and J. P. Coon, "A framework for characterising the value of information in hidden Markov models," *IEEE Trans. Inf. Theory*, 2022.

[18] ——, "Relationship between age and value of information for a noisy Ornstein–Uhlenbeck process," *Entropy*, vol. 23, no. 8, 2021.

[19] D. Guo, S. Shamai, and S. Verdu, "Mutual information and minimum mean-square error in Gaussian channels," *IEEE Trans. Inf. Theory*, vol. 51, no. 4, pp. 1261–1282, 2005.

[20] K. I. Park, *Fundamentals of Probability and Stochastic Processes with Applications to Communications*. Cham, Switzerland: Springer, 2018.

[21] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[22] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY, USA: John Wiley & Sons, 2006.