

A Variational Bayes Approach to Adaptive Radio Tomography

Donghoon Lee, *Member, IEEE*, and Georgios B. Giannakis, *Fellow, IEEE*

Abstract—Radio tomographic imaging (RTI) is an emerging technology for localization of physical objects in a geographical area covered by wireless networks. With attenuation measurements collected at spatially distributed sensors, RTI capitalizes on spatial loss fields (SLFs) measuring the absorption of radio frequency waves at spatial locations along the propagation path. These SLFs can be utilized for interference management in wireless communication networks, environmental monitoring, and survivor localization after natural disasters such as earthquakes. Key to the success of RTI is to accurately model shadowing as the weighted line integral of the SLF. To learn the SLF exhibiting statistical heterogeneity induced by spatially diverse environments, the present work develops a Bayesian framework entailing a piecewise homogeneous SLF with an underlying hidden Markov random field model. Utilizing variational Bayes techniques, the novel approach yields efficient field estimators at affordable complexity. A data-adaptive sensor selection strategy is also introduced to collect informative measurements for effective reconstruction of the SLF. Numerical tests using synthetic and real datasets demonstrate the capabilities of the proposed approach to radio tomography and channel-gain estimation.

Index Terms—Radio tomography, channel-gain estimation, variational Bayes, active learning, Bayesian inference

I. INTRODUCTION

Tomography is imaging by sectioning through the use of a penetrating wave, and has been widely appreciated by natural sciences, notably in medical imaging [2]. The principles underpinning radio tomographic methods have been carried over to construct what are termed *spatial loss fields* (SLFs), which are maps quantifying the attenuation experienced by electromagnetic waves in radio frequency (RF) bands at every spatial position [3]. To this end, pairs of collaborating sensors are deployed over the area of interest to estimate the attenuation introduced by the channel between those pairs of sensors. Different from conventional methods, radio tomography relies on *incoherent* measurements containing no phase information, e.g., the received signal strength (RSS). Such simplification saves costs for synchronization needed to calibrate phase differences among waveforms received at different sensors.

SLFs are instrumental in several tasks including radio tomography [4] and channel-gain cartography [5]. Absorption captured by the SLF allows one to discern objects located in

the area of interest, thus enabling radio tomographic imaging (RTI). Benefiting from the ability of RF waves to penetrate physical structures such as trees and buildings, RTI provides a means of device-free passive localization [6], [7], and has found diverse applications in disaster response for e.g., detecting individuals trapped in buildings or smoke [8]. SLFs are also useful in channel-gain cartography to provide channel-state information (CSI) for a link between any two locations even where no sensors are present [5]. Such maps can be employed by cognitive radios to control the interference that a secondary network inflicts to primary users that do not transmit—a setup encountered with television broadcast systems [9], [10], [11]. The non-collaborative nature of primary users precludes training-based channel estimation between a secondary transmitter and a primary receiver, and vice versa. Note that channel-gain cartography is also instrumental for interference management in the Internet-of-things (IoT) [12].

The key premise behind RTI is that spatially close radio links exhibit similar shadowing due to the presence of common obstructions. This shadowing correlation is related to the geometry of objects present in the area that waves propagate through [3], [13]. As a result, shadowing is modeled as the weighted line integral of the underlying two-dimensional SLF. The weights in the integral are determined by a function depending on the transmitter-receiver locations [3], [14], [15], which models the SLF effect on shadowing over a link. Inspired by this SLF model, various tomographic imaging methods were proposed [4], [8], [16], [17]. To detect locations of changes in the propagation environment, one can use the difference between the SLF across consecutive time slots [4], [16]. To cope with multipath in a cluttered environment, multi-channel measurements can be utilized to enhance localization accuracy [17]. Although these are calibration-free approaches, they cannot reveal static objects in the area of interest. It is also possible to replace the SLF with a label field indicating presence (or absence) of objects in motion on each voxel [8], and leverage the influence that moving objects on the propagation path have, on the variance of a RSS measurement. On the other hand, the SLF itself was reconstructed in [14], [18], [19], [20] to depict static objects in the area of interest, but calibration was necessary by using extra measurements (e.g., collected in free space). One can avoid extra data for calibration by estimating the SLF together with pathloss components [15], [21].

Another body of work leveraging the SLF model is that of channel-gain cartography when employing tomography based approaches [5], [15], [20], [22]. Linear interpolation techniques such as kriging were further employed to estimate

Parts of this work were presented at the IEEE International Conference on Acoustics, Speech and Signal Processing, held in Brighton, UK, during May 12-17, 2019 [1].

D. Lee and G. B. Giannakis are with the Department of Electrical and Computer Engineering and the Digital Technology Center, University of Minnesota, Minneapolis, MN 55455, USA. Emails: {leex6962, georgios}@umn.edu.

The work in this paper was supported in part by NSF grants 1508993, 1711471, and 1901134.

shadowing based on spatially correlated measurements [22], while spatio-temporal dynamics were tracked via Kalman filtering [5]. SLFs with regular patterns of objects have also been modeled as a superposition of a low-rank matrix plus a sparse matrix capturing structure irregularities [20]. While the aforementioned methods rely on heuristic criteria to choose the weight function, [15] provides blind algorithms to learn the weight function using a non-parametric kernel regression.

Conventionally, the SLF is learned via least-squares (LS) estimation regularized by the propagation environment [14], [16], [20]. The resultant ridge-regularized LS solution can be interpreted as a maximum a posteriori (MAP) estimator when the SLF is statistically homogeneous and modeled as a zero-mean Gaussian random field. However, these estimators are less effective when the propagation environment is spatially diverse due to a combination of free space and objects in different sizes and materials (e.g., as in urban areas), which subsequently induces statistical heterogeneity in the SLF. To account for environmental heterogeneity, we proposed in [19] a Bayesian approach to learn a piecewise homogeneous SLF through a binary hidden Markov random field (MRF) model [23] via Markov chain Monte Carlo (MCMC) [24]. But this approach does not scale because MCMC is computationally demanding.

Aiming at efficient field estimators at affordable complexity, we propose a variational Bayes (VB) framework for radio tomography to approximate the analytically intractable minimum mean-square error (MMSE) or MAP estimators. Instead of considering the binary hidden MRF to model statistical heterogeneity in the SLF [19], we further generalize the SLF model by considering K -ary piecewise homogeneous regions for $K \geq 2$, to address a richer class of environmental heterogeneity. Besides developing efficient and affordable solutions for RTI, another contribution here is a data-adaptive sensor selection technique, with the goal of reducing uncertainty in the SLF, by cross-fertilizing ideas from the fields of experimental design [25] and active learning [26]. The conditional entropy of the SLF is considered as an uncertainty measure, giving rise to a novel sensor selection criterion. Although this criterion is intractable especially when the size of the SLF is large, its efficient proxy can be obtained thanks to the availability of an approximate posterior model from the proposed VB algorithm.

The rest of the paper is organized as follows. Sec. II reviews the radio tomography model and states the problem. The Bayesian model and the resultant field reconstruction are the subjects of Sec. III, together with the proposed sensor selection method. Numerical tests with synthetic as well as real measurements are provided in Sec. IV. Finally, Sec. V summarizes the main conclusions.

Notation. Bold uppercase (lowercase) letters denote matrices (column vectors). Calligraphic fonts are used for sets; $\mathbf{I}_n$ is the $n \times n$ identity matrix. Operator $(\cdot)^\top$ represents the transpose a matrix $\mathbf{X} \in \mathbb{R}^{N_x \times N_y}$; $|\cdot|$ is used for the cardinality of a set, the magnitude of a scalar, or the determinant of a matrix; and $\text{vec}(\mathbf{X})$ produces a column vector $\mathbf{x} \in \mathbb{R}^{N_x N_y}$ by stacking the columns of a matrix one below the other ($\text{unvec}(\mathbf{x})$ denotes the reverse process).

II. BACKGROUND AND PROBLEM STATEMENT

Consider a set of sensors deployed over a two-dimensional geographical area $\mathcal{A} \subset \mathbb{R}^2$. After averaging out small-scale fading effects, the channel-gain measurement over a link between a transmitter located at $\mathbf{x} \in \mathcal{A}$ and a receiver located at $\mathbf{x}' \in \mathcal{A}$ can be represented (in dB) as

$$g(\mathbf{x}, \mathbf{x}') = g_0 - \gamma 10 \log_{10} d(\mathbf{x}, \mathbf{x}') - s(\mathbf{x}, \mathbf{x}') \quad (1)$$

where g_0 is the path gain at unit distance; $d(\mathbf{x}, \mathbf{x}') := \|\mathbf{x} - \mathbf{x}'\|$ is the Euclidean distance between the transceivers at $\mathbf{x}$ and $\mathbf{x}'$; γ is the pathloss exponent; and $s(\mathbf{x}, \mathbf{x}')$ is the attenuation due to shadow fading.¹

A tomographic shadow fading model is [3], [14], [20]

$$s(\mathbf{x}, \mathbf{x}') = \int_{\mathcal{A}} w(\mathbf{x}, \mathbf{x}', \tilde{\mathbf{x}}) f(\tilde{\mathbf{x}}) d\tilde{\mathbf{x}} \quad (2)$$

where $f : \mathcal{A} \rightarrow \mathbb{R}$ denotes the *spatial loss field* (SLF) capturing the attenuation at location $\tilde{\mathbf{x}}$, and $w : \mathcal{A} \times \mathcal{A} \times \mathcal{A} \rightarrow \mathbb{R}$ is a weight function describing how the SLF at $\tilde{\mathbf{x}}$ contributes to the shadowing experienced over the link $\mathbf{x} - \mathbf{x}'$. Typically, w confers a greater weight $w(\mathbf{x}, \mathbf{x}', \tilde{\mathbf{x}})$ to those locations $\tilde{\mathbf{x}}$ lying closer to the link $\mathbf{x} - \mathbf{x}'$. Examples of the weight function include the *normalized ellipse model* [16]

$$w(\mathbf{x}, \mathbf{x}', \tilde{\mathbf{x}}) := \begin{cases} 1/\sqrt{d(\mathbf{x}, \tilde{\mathbf{x}}) + d(\mathbf{x}', \tilde{\mathbf{x}})}, & \text{if } d(\mathbf{x}, \tilde{\mathbf{x}}) + d(\mathbf{x}', \tilde{\mathbf{x}}) \\ < d(\mathbf{x}, \mathbf{x}') + \lambda/2 \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where $\lambda > 0$ is a tunable parameter. The value of λ is commonly set to the wavelength to assign non-zero weights only within the first Fresnel zone. In radio tomography practice, the integral in (2) is approximated by a finite sum as

$$s(\mathbf{x}, \mathbf{x}') \simeq c \sum_{i=1}^{N_g} w(\mathbf{x}, \mathbf{x}', \tilde{\mathbf{x}}_i) f(\tilde{\mathbf{x}}_i) \quad (4)$$

where $\{\tilde{\mathbf{x}}_i\}_{i=1}^{N_g}$ is a grid of points over $\mathcal{A}$ and c is a constant that can be set to unity without loss of generality by absorbing any scaling factor in f . Clearly, (4) shows that $s(\mathbf{x}, \mathbf{x}')$ depends on f only through its values at the grid points.

The model in (2) describes how the spatial distribution of obstructions in the propagation path influences the attenuation between a pair of locations. The usefulness of (2) is twofold: i) as f represents absorption across space, it can be used for imaging; and ii) once f and w are known, the gain between any two points $\mathbf{x}$ and $\mathbf{x}'$ can be recovered through (1) and (2), which is precisely the objective of channel-gain cartography.

The goal of radio tomography is to obtain a tomogram by estimating f . To this end, N sensors located at $\{\mathbf{x}_1, \dots, \mathbf{x}_N\} \in \mathcal{A}$ collaboratively obtain channel-gain measurements. At time slot τ , the radios indexed by $n(\tau)$ and $n'(\tau)$ measure the channel-gain $\tilde{g}_\tau := g(\mathbf{x}_{n(\tau)}, \mathbf{x}_{n'(\tau)}) + \nu_\tau$ by exchanging training sequences known to both transmitting and receiving radios, where $n(\tau), n'(\tau) \in \{1, \dots, N\}$ and ν_τ denotes measurement

¹The channel model in (1) applies to either single-path propagation, or, to the first arriving path of multi-path propagation, with the remaining paths lumped to the noise ν_τ in (5): see also [14]. Although not considered here, multi-path diversity can be leveraged as in e.g., [17], at the expense of collecting and processing additional data.

noise. It is supposed that g_0 and γ have been estimated during a calibration stage [14]. After subtracting known components (g_0 and pathloss) from $\tilde{g}_\tau$, the shadowing estimate is found as

$$\begin{aligned}\tilde{s}_\tau &:= g_0 - \gamma 10 \log_{10} d(\mathbf{x}_{n(\tau)}, \mathbf{x}_{n'(\tau)}) - \tilde{g}_\tau \\ &= s(\mathbf{x}_{n(\tau)}, \mathbf{x}_{n'(\tau)}) - \nu_\tau.\end{aligned}\quad (5)$$

Having available $\tilde{\mathbf{s}}_t := [\tilde{s}_1, \dots, \tilde{s}_t]^\top \in \mathbb{R}^t$ along with the known set of links $\{(\mathbf{x}_{n(\tau)}, \mathbf{x}_{n'(\tau)})\}_{\tau=1}^t$ and the weight function w at the fusion center, the problem is to estimate $\mathbf{f}$, and thus $\mathbf{f} := [f(\tilde{\mathbf{x}}_1), \dots, f(\tilde{\mathbf{x}}_{N_g})]^\top \in \mathbb{R}^{N_g}$ using (4).

Conventional regularized LS estimators of $\mathbf{f}$ solve [14], [16]

$$\min_{\mathbf{f}} \sum_{\tau=1}^t \left(\tilde{s}_\tau - \sum_{i=1}^{N_g} w(\mathbf{x}_{n(\tau)}, \mathbf{x}_{n'(\tau)}, \tilde{\mathbf{x}}_i) f(\tilde{\mathbf{x}}_i) \right)^2 + \rho_f \mathcal{R}(\mathbf{f}) \quad (6)$$

where $\mathcal{R} : \mathbb{R}^{N_g} \rightarrow \mathbb{R}$ is a generic regularizer to promote a known attribute of $\mathbf{f}$, and $\rho_f \geq 0$ is a regularization scalar to reflect compliance of $\mathbf{f}$ with this attribute. Although (6) has been successfully applied to radio tomography after customizing the regularizer to the propagation environment, how accurate approximation is provided by a regularized solution of (6) is unclear, especially when the propagation environment exhibits inhomogeneous characteristics. Recent RTI developments include deep learning based solutions [27], [28], that require training of a neural network for RTI with a dataset collected offline. However, sufficiently large training datasets may not be available, while the generalization capability of the trained network for arbitrary propagation environments has not been validated yet.

To overcome this and improve the SLF estimator performance, prior knowledge on the heterogeneous structure of $\mathbf{f}$ will be exploited next, using a Bayesian approach.

III. ADAPTIVE BAYESIAN RADIO TOMOGRAPHY

In this section, we view $\mathbf{f}$ as random, and introduce a two-layer Bayesian SLF model, along with a VB-based approach to inference. We further develop a data-adaptive sensor selection method to collect informative measurements.

A. Bayesian model and problem formulation

Let $\mathcal{A}$ consist of K disjoint homogeneous regions $\mathcal{A}_k := \{\mathbf{x} | \mathbb{E}[f(\mathbf{x})] = \mu_{f_k}, \text{Var}[f(\mathbf{x})] = \sigma_{f_k}^2\}$ for $k = 1, \dots, K$, giving rise to a latent random label field $\mathbf{z} := [z(\tilde{\mathbf{x}}_1), \dots, z(\tilde{\mathbf{x}}_{N_g})]^\top \in \{1, \dots, K\}^{N_g}$ with K -ary entries $z(\tilde{\mathbf{x}}_i) = k$ if $\tilde{\mathbf{x}}_i \in \mathcal{A}_k \forall i, k$. Each $\mathcal{A}_k$ includes objects with generally different materials and in various spatial density, leading to distinct values of μ_{f_k} and $\sigma_{f_k}^2 \forall k$. With $K = 2$ and $\mathcal{A}$ corresponding to an urban area, $\mathcal{A}_2$ may include densely populated regions with buildings, while $\mathcal{A}_1$ with $\mu_{f_1} < \mu_{f_2}$ may capture the less obstructive open spaces. For such a paradigm, we model the conditional distribution of $f(\tilde{\mathbf{x}}_i)$ as

$$p(f(\tilde{\mathbf{x}}_i) | z(\tilde{\mathbf{x}}_i) = k) = \mathcal{N}(\mu_{f_k}, \sigma_{f_k}^2) \forall k. \quad (7)$$

We further assign Potts prior to $\mathbf{z}$ in order to capture the dependency among spatially correlated labels. By the Hammersley-

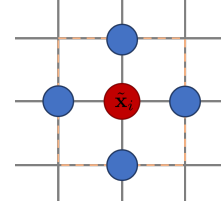


Fig. 1: Four-connected MRF with $z(\tilde{\mathbf{x}}_i)$ marked red and its neighbors in $\mathcal{N}(\tilde{\mathbf{x}}_i)$ marked blue.

Clifford theorem [29], Potts prior of $\mathbf{z}$ follows a Gibbs distribution

$$p(\mathbf{z}; \beta) = \frac{1}{C(\beta)} \exp \left[\sum_{i=1}^{N_g} \sum_{j \in \mathcal{N}(\tilde{\mathbf{x}}_i)} \beta \delta(z(\tilde{\mathbf{x}}_j) - z(\tilde{\mathbf{x}}_i)) \right] \quad (8)$$

where $\mathcal{N}(\tilde{\mathbf{x}}_i)$ is a set of indices comprising 1-hop neighbors of $\tilde{\mathbf{x}}_i$ on the rectangular grid in Fig. 1, β is a granularity coefficient controlling the degree of homogeneity in $\mathbf{z}$, $\delta(\cdot)$ is Kronecker's delta, and the normalization constant

$$C(\beta) := \sum_{\mathbf{z} \in \mathcal{Z}} \exp \left[\sum_{i=1}^{N_g} \sum_{j \in \mathcal{N}(\tilde{\mathbf{x}}_i)} \beta \delta(z(\tilde{\mathbf{x}}_j) - z(\tilde{\mathbf{x}}_i)) \right] \quad (9)$$

is the partition function with $\mathcal{Z} := \{1, \dots, K\}^{N_g}$. Such a prior locally promotes statistical homogeneity among proximal locations in $\mathbf{f}$ through $\mathbf{z}$, to efficiently handle global statistical heterogeneity in $\mathcal{A}$. To ease exposition, β is assumed known or fixed a priori; see e.g., [30], [31], [19] for a means of estimating β . If $\{f(\tilde{\mathbf{x}}_i)\}_{i=1}^{N_g}$ are conditionally independent given $\mathbf{z}$, the model reduces to the Gauss-Markov-Potts model [32]. Such model with $K = 3$ is depicted in Fig. 2 with the measurement model in (4).

Noise ν_τ in (5) is assumed independent and identically distributed (i.i.d.) Gaussian with zero mean and variance σ_ν^2 . Here, we correspondingly consider precisions of ν_τ and $\{f_k\}_{k=1}^K$ that are denoted as $\varphi_\nu := 1/\sigma_\nu^2$ and $\varphi_{f_k} := 1/\sigma_{f_k}^2 \forall k$, respectively. Let also $\boldsymbol{\theta}$ be a hyperparameter vector comprising φ_ν and $\boldsymbol{\theta}_f := [\boldsymbol{\mu}_f^\top, \boldsymbol{\varphi}_f^\top]^\top$ with $\boldsymbol{\mu}_f := [\mu_{f_1}, \dots, \mu_{f_K}]^\top \in \mathbb{R}^K$ and $\boldsymbol{\varphi}_f := [\varphi_{f_1}, \dots, \varphi_{f_K}]^\top \in \mathbb{R}^K$. Assuming the independence among entries of $\boldsymbol{\theta}$, we deduce that

$$p(\boldsymbol{\theta}) = p(\varphi_\nu) p(\boldsymbol{\mu}_f) p(\boldsymbol{\varphi}_f) = p(\varphi_\nu) \prod_{k=1}^K p(\mu_{f_k}) p(\varphi_{f_k}) \quad (10)$$

where the priors $p(\varphi_\nu)$, $p(\boldsymbol{\mu}_f)$, and $p(\boldsymbol{\varphi}_f)$ are as follows.

1) *Noise precision φ_ν* : With additive Gaussian noise having fixed mean, it is common to assign a conjugate prior to φ_ν that can reproduce a posterior in the same family of its prior. The gamma distribution for $\varphi_\nu \in \mathbb{R}^+$ serves this purpose, as

$$p(\varphi_\nu) = \mathcal{G}(a_\nu, b_\nu) := \frac{1}{\Gamma(a_\nu) b_\nu^{a_\nu}} (\varphi_\nu)^{a_\nu-1} e^{-\varphi_\nu/b_\nu} \quad (11)$$

where a_ν is referred to as the shape parameter, b_ν as the scale parameter, and $\Gamma(\cdot)$ denotes the gamma function.

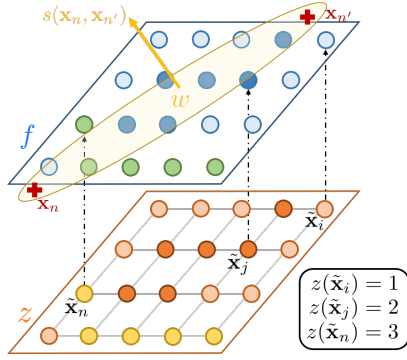


Fig. 2: Gauss-Markov-Potts radio tomography model with $K = 3$, together with measuring sensors located at $(\mathbf{x}_n, \mathbf{x}_{n'})$.

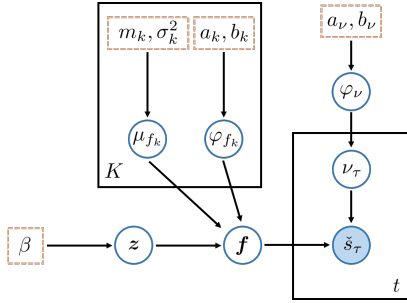


Fig. 3: Graphical model representation of the hierarchical Bayesian model for (hyper) parameters (those in dashed boxes are fixed).

2) *Hyperparameters θ_f of the SLF* : While the prior for μ_{f_k} is assumed to be Gaussian with mean m_k and variance σ_k^2 (see also [32]), similar to the noise, the prior for $\varphi_{f_k} \in \mathbb{R}^+$ is the Gamma distribution parameterized by $\{a_k, b_k\}$; that is,

$$p(\mu_{f_k}) = \mathcal{N}(m_k, \sigma_k^2), \quad k = 1, \dots, K \quad (12)$$

$$p(\varphi_{f_k}) = \mathcal{G}(a_k, b_k), \quad k = 1, \dots, K. \quad (13)$$

We stress that analytical tractability is the main motivation behind selecting the conjugate priors in (11)–(13).

Our goal of inferring $\mathbf{f}$, relies on the following posterior distribution that can be factored (within a constant) as

$$p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} | \tilde{\mathbf{s}}_t) \propto p(\tilde{\mathbf{s}}_t | \mathbf{f}, \varphi_\nu) p(\mathbf{f} | \mathbf{z}, \boldsymbol{\theta}_f) p(\mathbf{z}; \beta) p(\boldsymbol{\theta}) \quad (14)$$

where $p(\tilde{\mathbf{s}}_t | \mathbf{f}, \varphi_\nu) \sim \mathcal{N}(\mathbf{W}_t^\top \mathbf{f}, \sigma_\nu^2 \mathbf{I}_t)$ is the data likelihood with the weight matrix $\mathbf{W}_t \in \mathbb{R}^{N_g \times t}$ formed from columns $\mathbf{w}_\tau^{(n, n')} := [w(\mathbf{x}_{n(\tau)}, \mathbf{x}_{n'(\tau)}, \tilde{\mathbf{x}}_1), \dots, w(\mathbf{x}_{n(\tau)}, \mathbf{x}_{n'(\tau)}, \tilde{\mathbf{x}}_{N_g})]^\top := [w_{\tau,1}, \dots, w_{\tau, N_g}]^\top \in \mathbb{R}^{N_g}$ of the link $\mathbf{x}_{n(\tau)} - \mathbf{x}_{n'(\tau)}$ for $\tau = 1, \dots, t$. Fig. 3 depicts our hierarchical Bayesian model for $\{\tilde{\mathbf{s}}_t, \mathbf{f}, \mathbf{z}, \boldsymbol{\theta}\}$ as a directed acyclic graph, where the dependence between (hyper) parameters is indicated with an arrow.

Given the posterior in (14), the conditional minimum mean-square error (MMSE) estimator of the field is

$$\hat{\mathbf{f}}_{\text{MMSE}} := \mathbb{E}[\mathbf{f} | \mathbf{z} = \hat{\mathbf{z}}_{\text{MAP}}, \tilde{\mathbf{s}}_t] \quad (15)$$

where the maximum a posteriori (MAP) label estimator is

$$\hat{\mathbf{z}}_{\text{MAP}} := \arg \max_{\mathbf{z}} p(\mathbf{z} | \tilde{\mathbf{s}}_t) \quad (16)$$

and the MMSE estimators of $\boldsymbol{\theta}$ entries are

$$\widehat{\varphi_\nu}_{\text{MMSE}} := \mathbb{E}[\varphi_\nu | \tilde{\mathbf{s}}_t] \quad (17)$$

$$\widehat{\mu_{f_k}}_{\text{MMSE}} := \mathbb{E}[\mu_{f_k} | \tilde{\mathbf{s}}_t], \quad k = 1, \dots, K \quad (18)$$

$$\widehat{\varphi_{f_k}}_{\text{MMSE}} := \mathbb{E}[\varphi_{f_k} | \tilde{\mathbf{s}}_t], \quad k = 1, \dots, K. \quad (19)$$

B. Radio tomography using variational Bayes

Although the estimator forms in (15)–(19) have been considered also in [33], obtaining estimates in practice is not tractable because the complex posterior in (14) is not amenable to marginalization or maximization. To overcome this hurdle, one can resort to approximate Bayesian inference methods such as MCMC [24] that relies on samples of $\{\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}\}$ drawn from a complex distribution. Although MCMC can asymptotically approach an exact target distribution, such as the sought one in (14), it can be computationally demanding and thus does not scale well. Aiming at a scalable alternative, we will adopt the so-termed variational Bayes (VB) approach [34].

VB is a family of techniques to approximate a complex distribution by a tractable one termed variational distribution. This allows one to efficiently incorporate complex yet accurate statistical models in various signal processing and machine learning tasks, including variational autoencoder designs [35], and Bayesian neural network training modules [36]. A typical choice of an approximation criterion is to find the variational distribution q minimizing the Kullback-Leibler (KL) divergence ($D_{\text{KL}}(q \| p)$) to a target distribution q . The variational distribution q is further assumed to belong to a certain family $\mathcal{Q}$ of distributions possessing a simpler form of dependence between variables than the original one; see also [37], [38] for the so-termed mean-field approximation.

Tailored to the posterior in (14) the variational one, solves

$$\min_{q(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) \in \mathcal{Q}} D_{\text{KL}}(q(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) \| p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} | \tilde{\mathbf{s}}_t)) \quad (20)$$

Using that $D_{\text{KL}}(q \| p) := -\mathbb{E}_q[\ln(p/q)]$, the latter reduces to

$$(P1) \quad \max_{q(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) \in \mathcal{Q}} \underbrace{\mathbb{E}_{q(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta})} \left[\ln \left(\frac{p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}, \tilde{\mathbf{s}}_t)}{q(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta})} \right) \right]}_{=: \text{ELBO}(q(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}))} \quad (21)$$

where we drop the constant $p(\tilde{\mathbf{s}}_t)$ from the posterior that resulted in the so-termed *evidence lower bound* (ELBO) in (P1), which involves the joint $p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}, \tilde{\mathbf{s}}_t)$ factored as in the right-hand side (RHS) of (14). We choose the family $\mathcal{Q}$ as

$$\mathcal{Q} := \left\{ q : q(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}) := q(\mathbf{f} | \mathbf{z}) q(\mathbf{z}) q(\boldsymbol{\theta}) \right. \\ \left. = \prod_{i=1}^{N_g} q(f_i | z_i) \prod_{i=1}^{N_g} q(z_i) q(\boldsymbol{\theta}) \right\} \quad (22)$$

where $f_i := f(\tilde{\mathbf{x}}_i)$ and $z_i := z(\tilde{\mathbf{x}}_i) \forall i$ for simplicity, and

$$q(\boldsymbol{\theta}) := q(\varphi_\nu) q(\boldsymbol{\mu}_f) q(\boldsymbol{\varphi}_f) = q(\varphi_\nu) \prod_{k=1}^K q(\mu_{f_k}) \prod_{k=1}^K q(\varphi_{f_k}). \quad (23)$$

Following the general VB steps [39], we will solve our (P1) here via coordinate minimization among factors of $q(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta})$. Within a constant c , the optimal solutions have the form

$$\ln q^*(f_i | z_i) = \mathbb{E}_{-q(f_i | z_i)} [\ln p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}, \check{\mathbf{s}}_t)] + c \quad \forall i \quad (24)$$

$$\ln q^*(z_i) = \mathbb{E}_{-q(z_i)} [\ln p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}, \check{\mathbf{s}}_t)] + c \quad \forall i \quad (25)$$

$$\ln q^*(\boldsymbol{\theta}) = \mathbb{E}_{-q(\boldsymbol{\theta})} [\ln p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}, \check{\mathbf{s}}_t)] + c \quad (26)$$

where the expectation in (24) is over the variational pdf of $\mathbf{f}_{-i}$, $\mathbf{z}$, and $\boldsymbol{\theta}$, that is $\prod_{j \neq i} q(f_j | z_j) q(\mathbf{z}) q(\boldsymbol{\theta})$. Similar expressions are available for (25) and (26). The solutions in (24)–(26) are intertwined since the evaluation of one requires the others. We show in Appendices A–E that the optimal solutions can be obtained iteratively; that is, per iteration $\ell = 1, 2, \dots$, we have

$$q^{(\ell)}(f_i | z_i = k) = \mathcal{N}(\check{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i), \check{\sigma}_{f_k}^{2(\ell)}(\tilde{\mathbf{x}}_i)) \quad \forall k \quad (27)$$

$$q^{(\ell)}(z_i = k) =: \check{\zeta}_k^{(\ell)}(\tilde{\mathbf{x}}_i) = \frac{\check{\zeta}_k^{(\ell)}(\tilde{\mathbf{x}}_i)}{\sum_{k=1}^K \check{\zeta}_k^{(\ell)}(\tilde{\mathbf{x}}_i)} \quad \forall k \quad (28)$$

$$q^{(\ell)}(\varphi_\nu) = \mathcal{G}(\check{a}_\nu, \check{b}_\nu^{(\ell)}) \quad (29)$$

$$q^{(\ell)}(\mu_{f_k}) = \mathcal{N}(\check{m}_k^{(\ell)}, \check{\sigma}_k^{2(\ell)}) \quad \forall k \quad (30)$$

$$q^{(\ell)}(\varphi_{f_k}) = \mathcal{G}(\check{a}_k^{(\ell)}, \check{b}_k^{(\ell)}) \quad \forall k \quad (31)$$

with variational parameters

$$\check{\sigma}_{f_k}^{2(\ell)}(\tilde{\mathbf{x}}_i) = \left(\check{\varphi}_\nu^{(\ell-1)} \sum_{\tau=1}^t w_{\tau,i}^2 + \check{\varphi}_{f_k}^{(\ell-1)} \right)^{-1} \quad \forall k \quad (32)$$

$$\check{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i) = \check{f}_i^{(\ell-1)} + \check{\sigma}_{f_k}^{2(\ell)}(\tilde{\mathbf{x}}_i) \left[(\check{m}_k^{(\ell-1)} - \check{f}_i^{(\ell-1)}) \check{\varphi}_{f_k}^{(\ell-1)} + \check{\varphi}_\nu^{(\ell-1)} \sum_{\tau=1}^t w_{\tau,i} (\check{s}_\tau - \check{s}_\tau^{(\ell-1)}) \right] \quad \forall k \quad (33)$$

$$\check{\zeta}_k^{(\ell)}(\tilde{\mathbf{x}}_i) = \exp \left\{ -\frac{\check{\varphi}_{f_k}^{(\ell-1)}}{2} \left[\check{\sigma}_{f_k}^{2(\ell)}(\tilde{\mathbf{x}}_i) + \left(\check{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i) \right)^2 - 2\check{m}_k^{(\ell-1)} \check{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i) + \check{\sigma}_k^{2(\ell-1)} + \left(\check{m}_k^{(\ell-1)} \right)^2 \right] + \frac{1}{2} \left(\psi(\check{a}_k^{(\ell-1)}) + \ln \check{b}_k^{(\ell-1)} \right) + \sum_{j \in \mathcal{N}(\tilde{\mathbf{x}}_i)} \beta \check{\zeta}_k^{(\ell-1)}(\tilde{\mathbf{x}}_j) \right\} \quad \forall k \quad (34)$$

$$\check{a}_\nu = a_\nu + \frac{t}{2} \quad (35)$$

$$\check{b}_\nu^{(\ell)} = \left\{ \frac{1}{b_\nu} + \frac{1}{2} \sum_{\tau=1}^t \check{s}_\tau^2 - 2\check{s}_\tau \check{s}_\tau^{(\ell)} + \sum_{i=1}^{N_g} w_{\tau,i}^2 \right. \\ \times \left[\sum_{k=1}^K \check{\zeta}_k^{(\ell)}(\tilde{\mathbf{x}}_i) \left(\check{\sigma}_{f_k}^{2(\ell)}(\tilde{\mathbf{x}}_i) + \left(\check{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i) \right)^2 \right) - \left(\check{f}_i^{(\ell)} \right)^2 \right] + \left(\check{s}_\tau^{(\ell)} \right)^2 \left. \right\}^{-1} \quad (36)$$

$$\check{\sigma}_k^{2(\ell)} = \left(\frac{1}{\sigma_k^2} + \sum_{i=1}^{N_g} \check{\zeta}_k^{(\ell)}(\tilde{\mathbf{x}}_i) \check{\varphi}_{f_k}^{(\ell-1)} \right)^{-1} \quad \forall k \quad (37)$$

$$\check{m}_k^{(\ell)} = \check{\sigma}_k^{2(\ell)} \left(\frac{m_k}{\sigma_k^2} + \sum_{i=1}^{N_g} \check{\zeta}_k^{(\ell)}(\tilde{\mathbf{x}}_i) \check{\varphi}_{f_k}^{(\ell-1)} \check{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i) \right) \quad \forall k \quad (38)$$

$$\check{a}_k^{(\ell)} = a_k + \frac{1}{2} \sum_{i=1}^{N_g} \check{\zeta}_k^{(\ell)}(\tilde{\mathbf{x}}_i) \quad \forall k \quad (39)$$

$$\check{b}_k^{(\ell)} = \left[\frac{1}{b_k} + \frac{1}{2} \sum_{i=1}^{N_g} \check{\zeta}_k^{(\ell)}(\tilde{\mathbf{x}}_i) \left(\check{\sigma}_{f_k}^{2(\ell)}(\tilde{\mathbf{x}}_i) + \left(\check{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i) \right)^2 - 2\check{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i) \check{m}_k^{(\ell-1)} + \check{\sigma}_k^{2(\ell-1)} + \left(\check{m}_k^{(\ell-1)} \right)^2 \right) \right]^{-1} \quad \forall k \quad (40)$$

where $\psi(\cdot)$ is the digamma function, $\check{f}_i^{(\ell)} := \sum_{k=1}^K \check{\zeta}_k^{(\ell)}(\tilde{\mathbf{x}}_i) \check{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i) \quad \forall i$, $\check{s}_\tau^{(\ell)} := \sum_{i=1}^{N_g} w_{\tau,i} \check{f}_i^{(\ell)} \quad \forall \tau$, $\check{\varphi}_\nu^{(\ell)} := \mathbb{E}_{q^{(\ell)}(\varphi_\nu)}[\varphi_\nu] = \check{a}_\nu \check{b}_\nu^{(\ell)}$, and $\check{\varphi}_{f_k}^{(\ell)} := \mathbb{E}_{q^{(\ell)}(\varphi_{f_k})}[\varphi_{f_k}] = \check{a}_k^{(\ell)} \check{b}_k^{(\ell)} \quad \forall k$; see Appendices A–E for detailed derivation of the variational factors and parameters in (27)–(40).

Upon convergence of the iterative solvers, the (approximate) MAP estimator of $\mathbf{z}$ can be obtained as

$$\hat{\mathbf{z}}_{\text{MAP},i} = \arg \max_{z_i \in 1, \dots, K} q^*(z_i) \quad \forall i, \quad (41)$$

and then the (approximate) MMSE estimator of $\mathbf{f}$ as

$$\hat{\mathbf{f}}_{\text{MMSE},i} \simeq \mathbb{E}_{q^*(f_i | \hat{\mathbf{z}}_{\text{MAP},i})} [f_i] = \check{\mu}_{f_{\hat{\mathbf{z}}_{\text{MAP},i}}}^*(\tilde{\mathbf{x}}_i) \quad \forall i \quad (42)$$

while $\boldsymbol{\theta}$ is estimated using the marginal MMSE estimators

$$\widehat{\varphi}_\nu^{\text{MMSE}} \simeq \mathbb{E}_{q^*(\varphi_\nu)} [\varphi_\nu] = \check{a}_\nu^* \check{b}_\nu^* \quad (43)$$

$$\widehat{\mu}_{f_k}^{\text{MMSE}} \simeq \mathbb{E}_{q^*(\mu_{f_k})} [\mu_{f_k}] = \check{m}_k^* \quad \forall k \quad (44)$$

$$\widehat{\varphi}_{f_k}^{\text{MMSE}} \simeq \mathbb{E}_{q^*(\varphi_{f_k})} [\varphi_{f_k}] = \check{a}_k^* \check{b}_k^* \quad \forall k. \quad (45)$$

The VB algorithm to obtain $\{\hat{\mathbf{f}}_{\text{MMSE},i}\}_{i=1}^{N_g}$, $\{\hat{\mathbf{z}}_{\text{MAP},i}\}_{i=1}^{N_g}$, $\hat{\boldsymbol{\theta}}_{\text{MMSE}}$, and $q^*(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta})$ is tabulated in Alg. 1.

Remark 1. (Assessing convergence). The steps of Alg. 1 guarantee that the ELBO monotonically increases across iterations ℓ [40]. Hence, convergence of the solution can be assessed by monitoring the change in the ELBO of (P1) in (21), which for a preselected threshold $\xi > 0$ suggests stopping at iteration ℓ if $\text{ELBO}(q^{(\ell)}(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta})) - \text{ELBO}(q^{(\ell-1)}(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta})) \leq \xi$.

Remark 2. (Computational complexity). For Alg. 1, the complexity order to update $q(f_i | z_i = k) \quad \forall i, k$ per iteration ℓ is $\mathcal{O}(tKN_g)$ to compute $\check{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i)$ in (33), while updating $\check{\zeta}_k^{(\ell)}(\tilde{\mathbf{x}}_i) \quad \forall i, k$ via (28) incurs complexity $\mathcal{O}(KN_g)$. In addition, updating $q(\boldsymbol{\theta})$ has complexity $\mathcal{O}(tKN_g)$ that is dominated by the computation of $\check{b}_\nu$ in (36). Overall, the per-iteration complexity of Alg. 1 is $\mathcal{O}((2t+1)KN_g)$.

Note that a sample-based counterpart of Alg. 1 via MCMC in [19] incurs complexity in the order of $\mathcal{O}(N_g^3)$. For conventional methods to estimate $\mathbf{f}$, the ridge regularized LS [14] has a one-shot (non-iterative) complexity of $\mathcal{O}(N_g^3)$, while the total variation (TV) regularized LS via the alternating direction method of multipliers (ADMM) in [41] incurs complexity of $\mathcal{O}(N_g^3)$ per iteration ℓ ; see also [20], [15] for details. This means that Alg. 1 incurs the lowest per-iteration complexity, which becomes more critical as N_g increases.

Algorithm 1 Field estimation via variational Bayes

Input: $\check{\mathbf{s}}_t, \mathbf{W}_t, \left\{ a_\nu, b_\nu, \{m_k, \sigma_k^2, a_k, b_k\}_{k=1}^K \right\}$, and N_{Iter} .

- 1: Initialize $q^{(0)}(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta})$ and set $\ell = 0$
- 2: Obtain $\check{a}_\nu$ with (35)
- 3: **while** ELBO has not converged and $\ell \leq N_{\text{Iter}}$ **do**
- 4: Set $\ell \leftarrow \ell + 1$
- 5: Obtain $\check{\sigma}_{f_k}^{2(\ell)}(\tilde{\mathbf{x}}_i) \forall i, k$ via (32)
- 6: Obtain $\check{\mu}_{f_k}^{(\ell)}(\tilde{\mathbf{x}}_i) \forall i, k$ via (33)
- 7: Obtain $q_{f_k}^{(\ell)}(z_i = k) \forall i, k$ via (28)
- 8: Obtain $\check{b}_\nu^{(\ell)}$ via (36)
- 9: Obtain $\check{\sigma}_k^{2(\ell)} \forall k$ via (37)
- 10: Obtain $\check{m}_k^{(\ell)} \forall k$ via (38)
- 11: Obtain $\check{a}_k^{(\ell)} \forall k$ via (39)
- 12: Obtain $\check{b}_k^{(\ell)} \forall k$ via (40)
- 13: **end while**
- 14: Set $q^*(f_i|z_i) = q^{(\ell)}(f_i|z_i)$ and $q^*(z_i) = q^{(\ell)}(z_i) \forall i$
- 15: Set $q^*(\boldsymbol{\theta}) = q^{(\ell)}(\boldsymbol{\theta})$
- 16: Estimate $\hat{z}_{\text{MAP},i} = \arg \max_{z_i \in \{1, \dots, K\}} q^*(z_i) \forall i$
- 17: Estimate $\hat{f}_{i,\text{MMSE}} = \check{\mu}_{f_{\hat{z}_{\text{MAP},i}}}^*(\tilde{\mathbf{x}}_i) \forall i$
- 18: Estimate $\hat{\boldsymbol{\theta}}_{\text{MMSE}} = \mathbb{E}_{q^*(\boldsymbol{\theta})}[\boldsymbol{\theta}]$ via (43)–(45)
- 19: **return** $\hat{\mathbf{f}}_{\text{MMSE}}, \hat{\mathbf{z}}_{\text{MAP}}, \hat{\boldsymbol{\theta}}_{\text{MMSE}}, q^*(\mathbf{f}|\mathbf{z}), q^*(\mathbf{z}),$ and $q^*(\boldsymbol{\theta})$

C. Data-adaptive sensor selection via uncertainty sampling

Here we deal with cost-effective radio tomography as new data are collected by interactively querying the location of sensing radios to acquire a minimal but most informative measurements. To this end, a measurement (or a mini-batch of measurements) can be adaptively collected using a set of available sensing radio pairs, with the goal of reducing the uncertainty of $\mathbf{f}$. Since the proposed Bayesian framework accounts for the uncertainty through $\check{\sigma}_{f_k}^2(\tilde{\mathbf{x}}_i)$ in (38), we adopt the conditional entropy [42] to serve as an uncertainty measure of $\mathbf{f}$ at time slot τ , namely,

$$H(\mathbf{f}|\mathbf{z}, \check{\mathbf{s}}_\tau; \hat{\boldsymbol{\theta}}_\tau) = \sum_{\mathbf{z}' \in \mathcal{Z}} \int p(\mathbf{z}', \check{\mathbf{s}}'_\tau; \hat{\boldsymbol{\theta}}_\tau) \times H(\mathbf{f}|\mathbf{z} = \mathbf{z}', \check{\mathbf{s}}_\tau = \check{\mathbf{s}}'_\tau; \hat{\boldsymbol{\theta}}_\tau) d\check{\mathbf{s}}'_\tau, \quad (46)$$

where $\hat{\boldsymbol{\theta}}_\tau$ is the estimate obtained via (43)–(45) per slot τ , and

$$\begin{aligned} H(\mathbf{f}|\mathbf{z} = \mathbf{z}', \check{\mathbf{s}}_\tau = \check{\mathbf{s}}'_\tau; \hat{\boldsymbol{\theta}}_\tau) \\ := - \int p(\mathbf{f}|\mathbf{z} = \mathbf{z}', \check{\mathbf{s}}_\tau = \check{\mathbf{s}}'_\tau; \hat{\boldsymbol{\theta}}_\tau) \ln p(\mathbf{f}|\mathbf{z} = \mathbf{z}', \check{\mathbf{s}}_\tau = \check{\mathbf{s}}'_\tau; \hat{\boldsymbol{\theta}}_\tau) d\mathbf{f} \\ = \frac{1}{2} \ln \left| \check{\Sigma}_{f|\mathbf{z}', \check{\mathbf{s}}'_\tau; \hat{\boldsymbol{\theta}}_\tau} \right| + \frac{N_g}{2} \left(1 + \ln 2\pi \right) \end{aligned} \quad (47)$$

since $p(\mathbf{f}|\mathbf{z}, \check{\mathbf{s}}_\tau; \hat{\boldsymbol{\theta}}_\tau)$ is Gaussian with covariance matrix $\check{\Sigma}_{f|\mathbf{z}, \check{\mathbf{s}}_\tau; \hat{\boldsymbol{\theta}}_\tau} := \left(\widehat{\varphi}_\nu \mathbf{W}_\tau \mathbf{W}_\tau^\top + \widehat{\Phi}_{f|\mathbf{z}} \right)^{-1}$ with $\widehat{\Phi}_{f|\mathbf{z}} := \text{diag} \left(\left\{ \widehat{\varphi}_{f_{z_i}} \right\}_{i=1}^{N_g} \right)$ [19]. Then, using the matrix determinant

identity lemma [43, Chap. 18], it is not hard to show that

$$\begin{aligned} H(\mathbf{f}|\mathbf{z}, \check{\mathbf{s}}_{\tau+1}; \hat{\boldsymbol{\theta}}_\tau) &= H(\mathbf{f}|\mathbf{z}, \check{\mathbf{s}}_\tau; \hat{\boldsymbol{\theta}}_\tau) - \frac{1}{2} \sum_{\mathbf{z}' \in \mathcal{Z}} \int p(\mathbf{z}', \check{\mathbf{s}}'_\tau; \hat{\boldsymbol{\theta}}_\tau) \\ &\quad \times \ln \left(1 + \widehat{\varphi}_\nu \mathbf{w}_{\tau+1}^{(n, n')}^\top \check{\Sigma}_{f|\mathbf{z}', \check{\mathbf{s}}'_\tau; \hat{\boldsymbol{\theta}}_\tau} \mathbf{w}_{\tau+1}^{(n, n')} \right) d\check{\mathbf{s}}'_\tau. \end{aligned} \quad (48)$$

To obtain $\check{\mathbf{s}}_{\tau+1}$, we choose a pair of sensors (n^*, n'^*) , or equivalently find $\mathbf{w}_{\tau+1}^{(n^*, n'^*)}$ minimizing $H(\mathbf{f}|\mathbf{z}, \check{\mathbf{s}}_{\tau+1}; \hat{\boldsymbol{\theta}}_\tau)$.

Given $\check{\mathbf{s}}_\tau$, we then find $\mathbf{w}_{\tau+1}^{(n^*, n'^*)}$ by solving

$$(P2) \quad \max_{\substack{\mathbf{w}_{\tau+1}^{(n, n')}: \\ (n, n') \in \mathcal{M}_{\tau+1}}} \mathbb{E}_{p(\mathbf{z}|\check{\mathbf{s}}_\tau; \hat{\boldsymbol{\theta}}_\tau)} \left[h(\mathbf{z}, \check{\mathbf{s}}_\tau, \mathbf{w}_{\tau+1}^{(n, n')}; \hat{\boldsymbol{\theta}}_\tau) \right] \quad (49)$$

where $h(\mathbf{z}, \check{\mathbf{s}}_\tau, \mathbf{w}; \hat{\boldsymbol{\theta}}_\tau) := \ln \left(1 + \widehat{\varphi}_\nu \mathbf{w}^\top \check{\Sigma}_{f|\mathbf{z}, \check{\mathbf{s}}_\tau; \hat{\boldsymbol{\theta}}_\tau} \mathbf{w} \right)$ and $\mathcal{M}_\tau := \{(n, n') | \exists (\mathbf{x}_n - \mathbf{x}_{n'}) \text{ at } \tau, (n, n') \in \{1, \dots, N\}^2\}$ denotes the set of available sensing radio pairs at slot τ .

Clearly, (P2) in (49) cannot be directly solved because $p(\mathbf{z}|\check{\mathbf{s}}_\tau; \hat{\boldsymbol{\theta}}_\tau)$ is not tractable e.g., by marginalizing the posterior in (14). Hence, evaluating the cost of (P2) is intractable for large N_g as $|\mathcal{Z}| = 2^{N_g}$. Fortunately, we show next how (P2) can be approximately reformulated using the variational distribution $q(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta})$. Consider first that

$$p(\mathbf{f}|\mathbf{z}, \check{\mathbf{s}}_\tau, \boldsymbol{\theta}) = \frac{p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}|\check{\mathbf{s}}_\tau)}{p(\mathbf{z}, \boldsymbol{\theta}|\check{\mathbf{s}}_\tau)} \approx \frac{q(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta})}{q(\mathbf{z}, \boldsymbol{\theta})} = q(\mathbf{f}|\mathbf{z}), \quad (50)$$

which yields the approximation of H in (47), as

$$\begin{aligned} H(\mathbf{f}|\mathbf{z} = \mathbf{z}', \check{\mathbf{s}}_\tau = \check{\mathbf{s}}'_\tau; \hat{\boldsymbol{\theta}}_\tau) \\ \approx \frac{1}{2} \ln \left| \check{\Sigma}_{f|\mathbf{z}', \check{\mathbf{s}}'_\tau; \hat{\boldsymbol{\theta}}_\tau} \right| + \frac{N_g}{2} \left(1 + \ln 2\pi \right) \end{aligned} \quad (51)$$

with $\check{\Sigma}_{f|\mathbf{z}, \check{\mathbf{s}}_\tau; \hat{\boldsymbol{\theta}}_\tau} := \text{diag} \left(\left\{ \check{\sigma}_{f_{z_i}}^2(\tilde{\mathbf{x}}_i) \right\}_{i=1}^{N_g} \right)$; and subsequently, that of $H(\mathbf{f}|\mathbf{z}, \check{\mathbf{s}}_\tau; \hat{\boldsymbol{\theta}}_\tau)$ by substituting (51) into (46).

Similar to (48), we then show in Appendix F that

$$\begin{aligned} H(\mathbf{f}|\mathbf{z}, \check{\mathbf{s}}_{\tau+1}; \hat{\boldsymbol{\theta}}_\tau) &\approx H(\mathbf{f}|\mathbf{z}, \check{\mathbf{s}}_\tau; \hat{\boldsymbol{\theta}}_\tau) \\ &- \frac{1}{2} \sum_{\mathbf{z}' \in \mathcal{Z}} \int p(\mathbf{z}', \check{\mathbf{s}}'_\tau; \hat{\boldsymbol{\theta}}_\tau) \ln \left| \mathbf{I}_{N_g} + \tilde{\varphi}_\nu \Delta_{w_{\tau+1}} \check{\Sigma}_{f|\mathbf{z}', \check{\mathbf{s}}'_\tau; \hat{\boldsymbol{\theta}}_\tau} \right| d\check{\mathbf{s}}'_\tau \end{aligned} \quad (52)$$

where $\Delta_{w_{\tau+1}} := \text{diag} \left(\mathbf{w}_{\tau+1}^{(n, n')} \circ \mathbf{w}_{\tau+1}^{(n, n')} \right)$, with $\circ$ denoting the Hadamard product. Given $\check{\mathbf{s}}_\tau$, and using the approximation $p(\mathbf{z}|\check{\mathbf{s}}_\tau; \hat{\boldsymbol{\theta}}_\tau) \approx q(\mathbf{z})$, we can reformulate (P2) as (cf. (52))

$$(P2') \quad \max_{\substack{\mathbf{w}_{\tau+1}^{(n, n')}: \\ (n, n') \in \mathcal{M}_{\tau+1}}} \underbrace{\sum_{i=1}^{N_g} \mathbb{E}_{q(z_i)} \left[\ln \left(1 + \tilde{\varphi}_\nu \check{\sigma}_{f_{z_i}}^2(\tilde{\mathbf{x}}_i) w_{\tau+1, i}^2 \right) \right]}_{=: \bar{h}(\mathbf{w}_{\tau+1}^{(n, n')})}.$$

A greedy solver of (P2') yields the pair of sensors (n^*, n'^*) associated with $\mathbf{w}_{\tau+1}^{(n^*, n'^*)}$, to collect the datum $\check{\mathbf{s}}_{\tau+1}$ carrying the largest amount of information about uncertain parts of $\mathbf{f}$.

The overall algorithm for adaptive radio tomography via VB is tabulated in Alg. 2.

Remark 3. (Mini-batch setup). The proposed data-adaptive sensor selection scheme can be easily extended to a mini-batch setup of size N_{Batch} per time slot τ as follows: i) find weight vectors $\left\{ \mathbf{w}_{\tau+1}^{(n^{(m)}, n'^{(m)})} \right\}_{m=1}^{N_{\text{Batch}}}$ for $\left\{ (n^{(m)}, n'^{(m)}) \right\}_{m=1}^{N_{\text{Batch}}} \subset$

Algorithm 2 Adaptive radio tomography via variational Bayes

Input: $\tilde{\mathbf{s}}^{(0)}$, $\mathbf{W}^{(0)}$, $\left\{a_\nu, b_\nu, \{m_k, \sigma_k^2, a_k, b_k\}_{k=1}^K\right\}$, and N_{Iter} .

- 1: Set $\tilde{\mathbf{s}}_0 = \tilde{\mathbf{s}}^{(0)}$ and $\mathbf{W}_0 = \mathbf{W}^{(0)}$
- 2: **for** $\tau = 0, 1, \dots$ **do**
- 3: Obtain $\hat{\mathbf{f}}_{\text{MMSE}}$, $\hat{\boldsymbol{\theta}}_{\text{MMSE}}$, and $q^*(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta})$ via Alg. 1 $\left(\tilde{\mathbf{s}}_\tau, \mathbf{W}_\tau, \left\{a_\nu, b_\nu, \{m_k, \sigma_k^2, a_k, b_k\}_{k=1}^K\right\}, N_{\text{Iter}}\right)$
- 4: Evaluate $\bar{h}(\mathbf{w}_{\tau+1}^{(n,n')})$ in (P2') $\forall \{n, n'\} \in \mathcal{M}_{\tau+1}$
- 5: Collect $\tilde{\mathbf{s}}_{\tau+1}$ from (n^*, n'^*) with $\max \bar{h}(\mathbf{w}_{\tau+1}^{(n,n')})$
- 6: Set $\tilde{\mathbf{s}}_{\tau+1} = [\tilde{\mathbf{s}}_\tau^\top, \tilde{\mathbf{s}}_{\tau+1}^\top]^\top$ and $\mathbf{W}_{\tau+1} = [\mathbf{W}_\tau, \mathbf{w}_{\tau+1}^{(n^*, n'^*)}]$
- 7: **end for**
- 8: **return** $\hat{\mathbf{f}}_{\text{MMSE}}$

$\mathcal{M}_{\tau+1}$ associated with N_{Batch} largest values of $\bar{h}(\mathbf{w}_{\tau+1}^{(n,n')})$ in (P2'), and collect $\{\tilde{\mathbf{s}}_{\tau+1}^{(m)}\}_{m=1}^{N_{\text{Batch}}}$ from pairs of sensors revealed from those weight vectors (steps 4–5 in Alg. 2); and ii) aggregate those measurements below $\tilde{\mathbf{s}}_\tau$ to construct $\tilde{\mathbf{s}}_{\tau+1} := [\tilde{\mathbf{s}}_\tau^\top, \tilde{\mathbf{s}}_{\tau+1}^{(1)\top}, \dots, \tilde{\mathbf{s}}_{\tau+1}^{(N_{\text{Batch}})\top}]^\top$ (step 6 in Alg. 2). Numerical tests are presented next to assess the mini-batch operation of Alg. 2.

IV. NUMERICAL TESTS

Performance of the proposed algorithms was assessed through numerical tests using MATLAB on synthetic and real datasets. Comparisons were carried out with existing methods, including the ridge-regularized SLF estimate given by $\hat{\mathbf{f}}_{\text{LS}} = (\mathbf{W}_t \mathbf{W}_t^\top + \rho_f \mathbf{C}_f^{-1})^{-1} \mathbf{W}_t \tilde{\mathbf{s}}_t$ [14], where $\mathbf{C}_f$ is a spatial covariance matrix modeling the similarity between points $\tilde{\mathbf{x}}_i$ and $\tilde{\mathbf{x}}_j$ in area $\mathcal{A}$. We further tested the TV-regularized LS scheme in [41], which solves the problem in (6) with

$$\mathcal{R}(\mathbf{f}) = \sum_{i=1}^{N_x-1} \sum_{j=1}^{N_y} |F_{i+1,j} - F_{i,j}| + \sum_{i=1}^{N_x} \sum_{j=1}^{N_y-1} |F_{i,j+1} - F_{i,j}| \quad (53)$$

where $\mathbf{F} := \text{unvec}(\mathbf{f}) \in \mathbb{R}^{N_x \times N_y}$ and $F_{i,j} := [\mathbf{F}]_{i,j}$. We also tested an MCMC-based counterpart of Alg. 2 for estimating the posterior in (14), and solving (P2) in (49); see e.g., [19], [44] for details.

We further compared the proposed data-adaptive sensor selection with simple *random sampling* for both regularized LS estimators, by selecting $\{(n^{(m)}, n'^{(m)})\}_{m=1}^{N_{\text{Batch}}}$ uniformly at random to collect $\{\tilde{\mathbf{s}}_{\tau+1}^{(m)}\}_{m=1}^{N_{\text{Batch}}} \forall \tau$. Alg. 2 after replacing steps 4–5 with random sampling is termed *non-adaptive VB algorithm*, and will be compared with the proposed method throughout synthetic and real data tests.

A. Test with synthetic data

This section validates the proposed algorithm using synthetic datasets. Random tomographic measurements were collected from $N = 200$ sensors uniformly deployed on the boundary of $\mathcal{A} := [0.5, 60.5] \times [0.5, 60.5]$. Using these measurements, the SLF was reconstructed over the grid $\{\mathbf{x}_i\}_{i=1}^{3,600} := \{1, \dots, 60\}^2$. To generate the ground-truth SLF $\mathbf{f}_0$, the ground-truth label field $\mathbf{z}_0$ was generated via

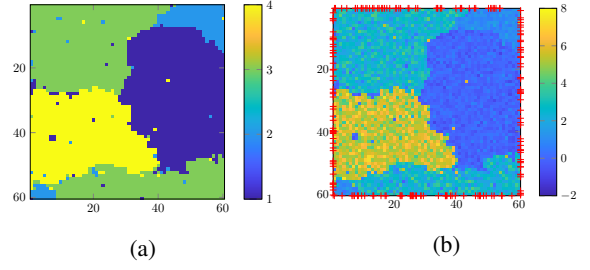


Fig. 4: True fields for synthetic tests: (a) hidden label field $\mathbf{Z}_0$ and (b) spatial loss field $\mathbf{F}_0$ with $N = 200$ sensor locations marked with crosses.

Gibbs sampling [45] by using Potts prior of $\mathbf{z}$ in (8) with $\beta = 1.5$ and $K = 4$. Given $\boldsymbol{\theta}_f := [\boldsymbol{\mu}_f^\top, \boldsymbol{\varphi}_f^\top]^\top$ with $\boldsymbol{\mu}_f = [0, 1, 2.5, 5.5]^\top$ and $\boldsymbol{\varphi}_f = [10, 10, 2, 2]^\top$, vector $\mathbf{f}_0$ was constructed to have $f(\mathbf{x}_i) \sim \mathcal{N}(\mu_{f_k}, \varphi_{f_k}^{-1}) \forall \mathbf{x}_i \in \mathcal{A}_k, \forall k$ conditioned on the labels in $\mathbf{z}_0$. The resulting hidden label field $\mathbf{Z}_0 := \text{unvec}(\mathbf{z}_0) \in \{1, 2, 3, 4\}^{60 \times 60}$, and the true SLF $\mathbf{F}_0 := \text{unvec}(\mathbf{f}_0) \in \mathbb{R}^{60 \times 60}$ are depicted in Fig. 4 with sensor locations marked by crosses. By deploying sensors on the boundaries of the area of interest, it becomes to reveal the object pattern through the reconstructed SLF based on non-invasively collected measurements. The effects of calibration are not accounted for, meaning that g_0 and γ are assumed to be known, and the fusion center directly uses shadowing measurements $\tilde{\mathbf{s}}_\tau$. Under the mini-batch operation, each measurement $\tilde{\mathbf{s}}_\tau^{(m)} \forall \tau, m$ was generated according to (5), where s_τ was obtained using (4) with w set to the normalized ellipse model in (3) with $\lambda = 0.39$, while ν_τ was set to follow a zero-mean Gaussian pdf with $\varphi_\nu = 20$. To construct $\mathcal{M}_{\tau+1}$ per time slot τ , $|\mathcal{M}_{\tau+1}| = 200$ pairs of sensors were uniformly selected at random with replacement. Then, $N_{\text{Batch}} = 100$ shadowing measurements were collected at $\{(n^{(m)}, n'^{(m)})\}_{m=1}^{N_{\text{Batch}}} \subset \mathcal{M}_{\tau+1}$ to run Alg. 2 for $\tau = 0, 1, \dots, 8$.

In all synthetic tests, the simulation parameters were set to $N_{\text{Iter}} = 3,000$ and $\xi = 10^{-6}$; hyper-hyper parameters of ν_t were set to $a_\nu = 1,300$ and $b_\nu = 2$; and those of $\boldsymbol{\theta}_f$ were set as listed in Table. I. To execute Alg. 1, variational parameters of $q^{(0)}(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta})$ were initialized as follows: $\{\check{\mu}_{f_k}^{(0)}(\tilde{\mathbf{x}}_i)\}_{i=1}^{N_g} \forall k$, $\check{b}_\nu^{(0)}$, $\{\check{\sigma}_k^{(0)}\}_{k=1}^4$, and $\{\check{a}_k^{(0)}, \check{b}_k^{(0)}\}_{k=1}^4$ were drawn from the uniform distribution $\mathcal{U}(0, 1)$, while $\check{m}_k^{(0)} = m_k \forall k$; and it was set to $\check{\zeta}_k^{(0)}(\tilde{\mathbf{x}}_i) = 1/4 \forall i, k$. Furthermore, $\tilde{\mathbf{s}}^{(0)}$ was collected from 800 pairs of sensors selected at random, which determined $\mathbf{W}^{(0)}$. To find ρ_f of the competing alternatives, the L-curve [46, Chapter 26] was used for the ridge regularization, while the generalized cross-validation [47] was adopted for the TV regularization. The hyper-hyper parameters of $\boldsymbol{\theta}$ used for the proposed algorithm were also adopted to run its MCMC-based counterpart.

The first experiment is performed to validate Alg. 2. Estimates of SLFs $\hat{\mathbf{F}} := \text{unvec}(\hat{\mathbf{f}})$ and the associated hidden label fields $\hat{\mathbf{Z}} := \text{unvec}(\hat{\mathbf{z}})$ at time slot $\tau = 8$ obtained via Alg. 2, and the competing alternatives, are depicted in Figs. 5a–5j. One-shot estimates of the SLF and associated

TABLE I: Hyper-parameters of θ_f for synthetic data tests.

m_1	m_2	m_3	m_4	σ_1^2	σ_2^2	σ_3^2	σ_4^2
0	0.9	2.7	5.3	10^{-4}	10^{-4}	10^{-4}	10^{-4}
a_1	a_2	a_3	a_4	b_1	b_2	b_3	b_4
0.8	0.8	0.8	0.8	1	1	0.5	0.5

hidden field, denoted as $\hat{\mathbf{F}}_{\text{full}}$ and $\hat{\mathbf{Z}}_{\text{full}}$, respectively, are also displayed in Figs. 5k and 5l, which were obtained via Alg. 2 by using the entire set of 2,400 measurements collected till $\tau = 8$. Clearly, satisfactory results were obtained only by the approximate Bayesian inference methods including MCMC and VB because every piecewise homogeneous region was accurately classified through the hidden label field. As discussed in Remark 2 however, the proposed algorithm is computationally much more efficient than the ones using MCMC. Per-iteration execution time was 0.04 (sec) for Alg. 2 on average, while that was 3.64 (sec) for the MCMC method. On the other hand, the regularized LS solutions were unable to accurately reconstruct the SLF, as depicted in Figs. 5a and 5b.

To test the proposed sensor selection method, $\hat{\mathbf{F}}$ and $\hat{\mathbf{Z}}$ found using the non-adaptive VB algorithm are depicted in Figs. 5e and 5f. Visual comparison of Figs. 5c and 5e reveals that the reconstruction performance for $\mathbf{F}$ can be improved with the same number of measurements by adaptively selecting pairs of sensors. Accuracy of $\hat{\mathbf{z}}$ was also quantitatively measured by the labeling-error, defined using the entrywise Kronecker delta $\delta(\cdot)$, as $\|\delta(\mathbf{z}_0 - \hat{\mathbf{z}})\|_1/N_g$. Progression of the labeling error averaged over 20 Monte Carlo (MC) runs is displayed in Fig. 6a, where the proposed method consistently outperforms the non-adaptive one. This shows that informative measurements adaptively collected to decrease uncertainty of $\mathbf{f}$ given a current estimate of θ improve accuracy of $\hat{\mathbf{f}}$ and $\hat{\mathbf{z}}$ in the next time slot. As a result, the SLF reconstruction accuracy of Alg. 2 improves accordingly with fewer measurements, as confirmed by comparing Figs. 5c and 5k.

The next experiment tests robustness of the proposed algorithms against measurement noise ν_τ . We adopted the labeling-error for $\mathbf{z}$ averaged over sensor locations and realizations of $\{\nu_\tau\}_\tau^t$ to quantify the reconstruction performance. Fig. 6b shows the progression of the labeling error at $\tau = 8$ as a function of the noise precision φ_ν averaged over 20 MC runs. Note that Figs. 5d and 5f correspond to the rightmost point of the x-axis of Fig. 6b. Clearly, the reconstruction performance does not severely decrease as φ_ν decreases, or equivalently σ_ν^2 increases. This confirms that the proposed algorithm is reasonably robust against measurement noise.

Averaged estimates of θ and associated standard deviation denoted with $\pm$ are listed in Table II. Together with Fig. 5, the high estimation accuracy of hyperparameters implies that the proposed method can effectively reveal patterns of objects in $\mathcal{A}$ by correctly inferring the underlying statistical properties of each piecewise homogeneous region in the SLF. Note that φ_f entries are overestimated in Table II. This can be intuitively understood in the sense that minimizing the KL divergence in (20) leads to $q(\theta_f)$ avoiding regions in which $p(\mathbf{f}|\mathbf{z}, \theta_f)p(\theta_f)$ is small by setting each φ_{f_k} to a large value $\forall k$, which corroborates the result in [48, p. 468].

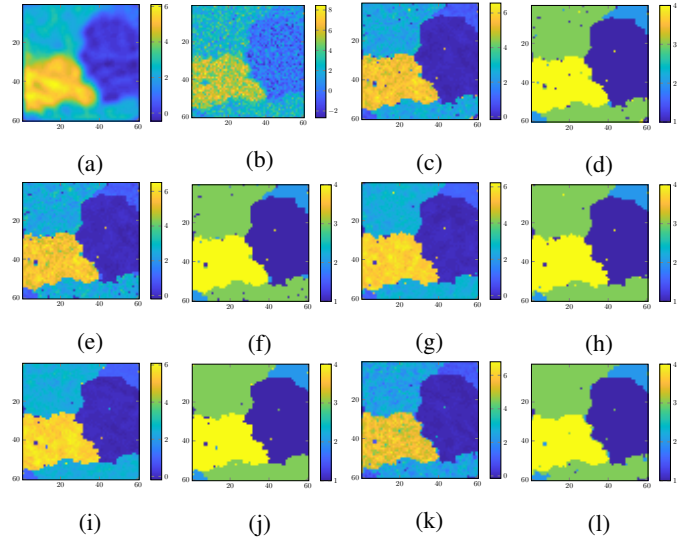


Fig. 5: SLF estimates $\hat{\mathbf{F}}$ at $\tau = 8$ (with 1,600 measurements) via; (a) ridge-regularized LS ($\rho_f = 0.015$ and $\mathbf{C}_f = \mathbf{I}_{3,600}$); (b) TV-regularized LS ($\rho_f = 10^{-11}$); (c) Alg. 2 through (d) estimated hidden field $\hat{\mathbf{Z}}$; (e) non-adaptive VB algorithm through (f) $\hat{\mathbf{Z}}$; (g) adaptive MCMC algorithm through (h) $\hat{\mathbf{Z}}$; (i) non-adaptive MCMC algorithm through (j) $\hat{\mathbf{Z}}$; and (k) $\hat{\mathbf{F}}_{\text{full}}$ and (l) $\hat{\mathbf{Z}}_{\text{full}}$ obtained by using the full data (with 2,400 measurements) via Alg. 2.

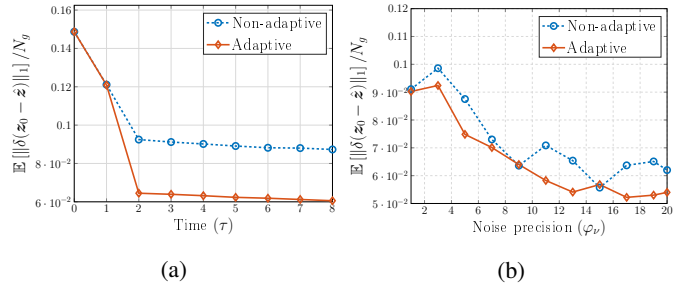


Fig. 6: Progression of estimation error of $\mathbf{z}$ versus (a) time τ ; and (b) noise precision φ_ν , averaged over 20 MC runs.

TABLE II: True θ and estimated $\hat{\theta}$ via Alg. 2 (setting of Fig. 5c); and non-adaptive VB algorithm (setting of Fig. 5e) averaged over 20 independent MC runs.

θ	True	Est. (Alg. 2)	Est. (non-adaptive)
φ_ν	20	$18.329 \pm 6 \times 10^{-3}$	$18.461 \pm 4.6 \times 10^{-3}$
μ_{f_1}	0	$0.022 \pm 1.2 \times 10^{-2}$	$0.018 \pm 1.9 \times 10^{-2}$
μ_{f_2}	1	$0.957 \pm 1.7 \times 10^{-2}$	$0.962 \pm 1.6 \times 10^{-2}$
μ_{f_3}	2.5	$2.573 \pm 1.7 \times 10^{-2}$	$2.578 \pm 2.6 \times 10^{-2}$
μ_{f_4}	5.5	$5.399 \pm 2.7 \times 10^{-2}$	$5.374 \pm 7 \times 10^{-3}$
φ_{f_1}	10	$40.178 \pm 3 \times 10^{-3}$	$42.352 \pm 2 \times 10^{-3}$
φ_{f_2}	10	$14.634 \pm 1.4 \times 10^{-2}$	$15.845 \pm 1.2 \times 10^{-2}$
φ_{f_3}	2	$7.712 \pm 2.7 \times 10^{-2}$	$7.493 \pm 2.2 \times 10^{-2}$
φ_{f_4}	2	$4.620 \pm 6.1 \times 10^{-2}$	$5.451 \pm 4 \times 10^{-2}$

Next, we will validate the efficacy of the proposed algorithms for channel-gain cartography using the setup of Fig. 5. From the estimate $\hat{\mathbf{f}}_{\text{MMSE}}$ obtained through Alg. 2, we found the shadowing attenuation $\hat{s}(\mathbf{x}, \mathbf{x}')$ between two arbitrary points $\mathbf{x}$ and $\mathbf{x}'$ in $\mathcal{A}$ using (4) after replacing $\mathbf{f}$ with

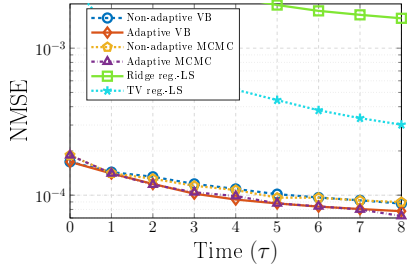


Fig. 7: Progression of channel-gain estimation error.

$\hat{f}_{\text{MMSE}}$. Subsequently, we obtained the estimated channel-gain $\hat{g}(\mathbf{x}, \mathbf{x}')$ after substituting $\hat{s}(\mathbf{x}, \mathbf{x}')$ into (1).

Since g_0 and γ are known, obtaining $s(\mathbf{x}, \mathbf{x}')$ is equivalent to finding $g(\mathbf{x}, \mathbf{x}')$; cf. (1). This suggests adopting a performance metric quantifying the mismatch between $s(\mathbf{x}, \mathbf{x}')$ and $\hat{s}(\mathbf{x}, \mathbf{x}')$, using the normalized mean-square error

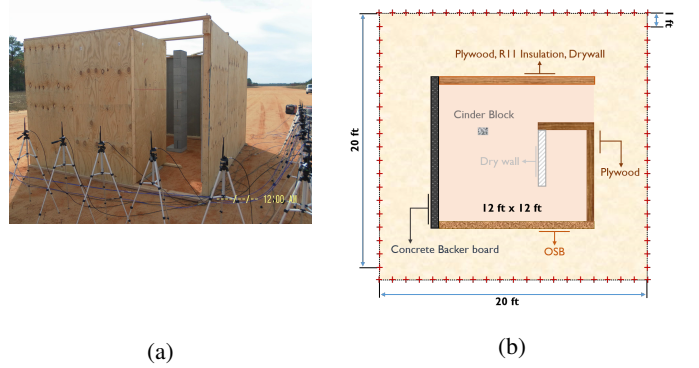
$$\text{NMSE} := \frac{\mathbb{E} \left[\int_{\mathcal{A}} (s(\mathbf{x}, \mathbf{x}') - \hat{s}(\mathbf{x}, \mathbf{x}'))^2 d\mathbf{x} d\mathbf{x}' \right]}{\mathbb{E} \left[\int_{\mathcal{A}} s^2(\mathbf{x}, \mathbf{x}') d\mathbf{x} d\mathbf{x}' \right]} \quad (54)$$

where the expectation is over the set $\{\mathbf{x}_n\}_{n=1}^N$ of sensor locations and realizations of $\{\nu_\tau\}_\tau$. The integrals are approximated by averaging the integrand over 500 pairs of $(\mathbf{x}, \mathbf{x}')$ chosen independently and uniformly at random on the boundary of $\mathcal{A}$. The expectations are estimated by averaging simulated deviates over 20 MC runs.

Fig. 7 depicts the NMSE of the proposed method and those of competing alternatives. Clearly, the approximate Bayesian inference methods outperform the regularized LS solutions. Furthermore, the performance of the VB methods is comparable to those of the MCMC methods. Noticeably, the adaptive VB method consistently exhibits lower NMSE than both non-adaptive ones, which highlights the efficacy in estimating channel-gain via the data-adaptive sensor selection. This suggests that the proposed VB framework is a viable solution for both radio tomography and channel-gain cartography, while enjoying low computational complexity.

B. Test with real data

This section validates the proposed method using the real dataset in [14]. The test setup is depicted in Fig. 8, where $\mathcal{A} = [0.5, 20.5] \times [0.5, 20.5]$ is a square with sides of 20 feet (ft), over which a grid $\{\tilde{\mathbf{x}}_i\}_{i=1}^{3,721} := \{1, \dots, 61\}^2$ of $N_g = 3,721$ points is defined. A collection of $N = 80$ sensors measure the RSS at 2.425 GHz between pairs of sensor positions, marked with the $N = 80$ crosses in Fig. 8b. To estimate g_0 and γ using the approach in [14], a first set of 2,400 measurements was obtained before placing objects. Estimates $\hat{g}_0 = 54.6$ (dB) and $\hat{\gamma} = 0.276$ were obtained during the calibration phase. Afterwards, the structure comprising one pillar and six walls of different materials was assembled as shown in Fig. 8b, and $T = 2,380$ measurements $\{\tilde{g}_{\tau'}\}_{\tau'=1}^T$ were collected. Calibrated measurements $\{\tilde{s}_{\tau'}\}_{\tau'=1}^T$ were then obtained from $\{\tilde{g}_{\tau'}\}_{\tau'=1}^T$ after substituting $\hat{g}_0$ and $\hat{\gamma}$ into (5). The weights $\{\mathbf{w}_{\tau'}^{(n,n')}\}_{\tau'=1}^T$ were constructed with w in (3) by using known locations of sensor pairs. Note that τ' is

Fig. 8: (a) Setup for the real data collection [14]; and (b) configuration of the testbed with $N = 80$ sensor locations marked with red crosses.

introduced to distinguish indices of the real data from τ used to index time slots in numerical tests.

We randomly selected 1,380 measurements from $\{\tilde{s}_{\tau'}\}_{\tau'=1}^T$ to initialize $\tilde{\mathbf{s}}^{(0)}$ and $\mathbf{W}^{(0)}$, and used the remaining 1,000 measurements to run the proposed algorithm under the mini-batch operation for $\tau = 0, 1, \dots, 5$. At each time slot τ , $\mathcal{M}_{\tau+1}$ was formed by sensors corresponding to $|\mathcal{M}_{\tau+1}| = 200$ weight vectors uniformly selected at random from $\{\mathbf{w}_{\tau'}^{(n,n')}\}_{\tau'}$ associated with the remaining 1,000 measurements without replacement. Then, $N_{\text{Batch}} = 100$ measurements were chosen from $\{\tilde{s}_{\tau'}\}_{\tau'}$ associated with $\mathcal{M}_{\tau+1}$.

Simulation parameters were set to $N_{\text{Iter}} = 3,000$, $\xi = 10^{-6}$, and $K = 3$; and hyper-hyper parameters of θ were set to $a_\nu = b_\nu = 10^{-3}$, $[m_1, m_2, m_3]^\top = [0, 0.035, 0.05]^\top$, $\sigma_k^2 = 10^{-4} \forall k$, and $a_k = b_k = 0.1 \forall k$, respectively. To execute Alg. 1, variational parameters of $q^{(0)}(\mathbf{f}, \mathbf{z}, \theta)$ were initialized as follows: $\{\tilde{\mu}_{f_k}^{(0)}(\tilde{\mathbf{x}}_i)\}_{i=1}^{N_g} \forall k$, $\tilde{b}_\nu^{(0)}$, $\{\tilde{\sigma}_k^{2(0)}\}_{k=1}^3$, and $\{\tilde{a}_k^{(0)}, \tilde{b}_k^{(0)}\}_{k=1}^3$ were drawn from the uniform distribution $\mathcal{U}(0, 1)$, while $\tilde{m}_k^{(0)} = m_k \forall k$ and $\tilde{\zeta}_k^{(0)}(\tilde{\mathbf{x}}_i) = 1/3 \forall i, k$.

Following [13], [14], a spatial covariance matrix was used for $\mathbf{C}_f$ of the ridge-regularized LS estimator, which models the similarity between points $\tilde{\mathbf{x}}_i$ and $\tilde{\mathbf{x}}_j$ as $[\mathbf{C}_f]_{ij} = \sigma_s^2 \exp[-\|\tilde{\mathbf{x}}_i - \tilde{\mathbf{x}}_j\|_2 / \kappa]$ [13] with $\sigma_s^2 = \kappa = 1$, and $\rho_f = 0.015$ found with the L-curve [46, Chapter 26]. For the TV-regularized LS estimator, it was set to $\rho_f = 6$ found through the generalized cross validation [47]. To assess the efficacy of our Bayesian model with the K -ary hidden label field, we tested the adaptive MCMC method in [19] with $K = 2$.

Figs. 9a–9h depict SLF estimates $\hat{\mathbf{F}}$ and associated hidden fields $\hat{\mathbf{Z}}$ at $\tau = 5$ obtained via the proposed algorithms and competing alternatives. As a benchmark, one-shot estimates of the SLF $\hat{\mathbf{F}}_{\text{full}}$ and associated hidden field $\hat{\mathbf{Z}}_{\text{full}}$ are also displayed in Figs. 9i and 9j obtained via Alg. 2 by using the entire set of 2,380 measurements. Comparing Figs. 9e and 9i (or Figs. 9f and 9j) shows that the proposed method accurately reveals the structural pattern of the testbed by using fewer number of measurements; e.g., the cinder block in the testbed was not captured by the SLF in Fig. 9g, but that in Fig. 9e. Such similarity of object patterns in Figs. 8b and 9e visually

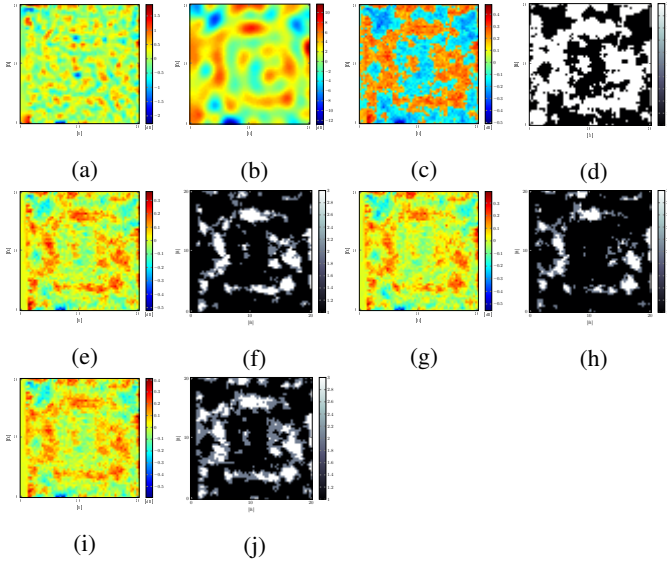


Fig. 9: SLF estimates $\hat{\mathbf{F}}$ at $\tau = 5$ (with 1,880 measurements) via; (a) ridge-regularized LS ($\rho_f = 0.015$ and $\mathbf{C}_f = \mathbf{I}_{3,600}$); (b) TV-regularized LS ($\rho_f = 6$); (c) adaptive MCMC algorithm in [19] with $K = 2$ through (d) estimated hidden field $\hat{\mathbf{Z}}$; (e) Alg. 2 through (f) $\hat{\mathbf{Z}}$; (g) non-adaptive VB algorithm through (h) $\hat{\mathbf{Z}}$; and (i) $\hat{\mathbf{F}}_{\text{full}}$ and (j) $\hat{\mathbf{Z}}_{\text{full}}$ obtained by using the full data (with 2,380 measurements).

validates the efficacy of the proposed algorithm for RTI as well. For competing alternatives, the testbed structure was not captured through the SLFs in Figs. 9a and 9b estimated via both regularized LS methods. On the other hand, the MCMC method reveals the structure through $\hat{\mathbf{F}}$ and $\hat{\mathbf{Z}}$ in Figs. 9c and 9d, although they are less accurately delineated than those from the proposed method. This illustrates the benefits of considering a general Bayesian model with $K \geq 2$ addressing a richer class of spatial heterogeneity.

Efficacy of the data-driven sensor selection scheme is further analyzed. Specifically, the accuracy of $\hat{\mathbf{z}}$ measured by the labeling error $\|\delta(\hat{\mathbf{z}}_{\text{full}} - \hat{\mathbf{z}})\|_1/N_g$ with $\hat{\mathbf{z}}_{\text{full}} := \text{vec}(\hat{\mathbf{Z}}_{\text{full}})$ was used as performance metric. Progression of the labeling error for Alg. 2 is depicted in Fig. 10 with that for the non-adaptive VB algorithm, where the proposed method consistently outperforms the non-adaptive one for every τ . This implies that the proposed sensor selection strategy helps to reveal object patterns more accurately while reducing data collection costs.

To corroborate the hyperparameter estimation capability of the proposed algorithm, estimates of θ averaged over 20 MC runs are listed in Table III. Estimated $\hat{\theta}$ obtained by using the full data was considered as a benchmark, to demonstrate that Alg. 2 yields estimates θ closer to the benchmark than its non-adaptive counterpart (except φ_ν). Note that the level of measurement noise is high since $\hat{\sigma}_\nu^2 = \hat{\varphi}_\nu^{-1} \approx 15$. This can be justified because the testbed structure was accurately revealed in $\hat{\mathbf{F}}$ and $\hat{\mathbf{Z}}$ from the proposed method by incorporating imperfect calibration effects in the measurement noise.

The last simulation assesses performance of the proposed algorithms for channel-gain cartography. The set of shadowing measurements and setup was the one used in the first simulated

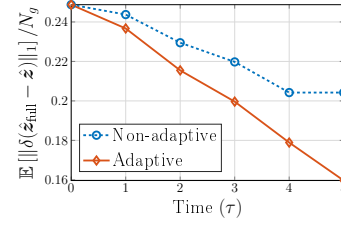


Fig. 10: Progression of a mismatch between $\hat{\mathbf{z}}$ and $\hat{\mathbf{z}}_{\text{full}}$.

TABLE III: Estimated $\hat{\theta}$ via benchmark algorithm (setting of Fig. 9i); Alg. 2 (setting of Fig. 9e); and non-adaptive VB algorithm (setting of Fig. 9g), averaged over 20 independent MC runs.

θ	Benchmark	Est. (Alg. 2)	Est. (non-adaptive)
φ_ν	0.075 ± 10^{-16}	0.068 ± 0.13	0.071 ± 0.24
μ_{f_1}	-0.001 ± 10^{-17}	-0.001 ± 10^{-7}	-0.001 ± 10^{-7}
μ_{f_2}	0.032 ± 10^{-17}	0.032 ± 10^{-8}	0.032 ± 10^{-8}
μ_{f_3}	0.045 ± 10^{-17}	0.046 ± 10^{-8}	0.046 ± 10^{-8}
φ_{f_1}	5.524 ± 10^{-18}	4.951 ± 10^{-3}	$4.789 \pm 1.9 \times 10^{-3}$
φ_{f_2}	5.524 ± 10^{-18}	4.942 ± 10^{-3}	$4.782 \pm 1.7 \times 10^{-3}$
φ_{f_3}	5.524 ± 10^{-18}	4.935 ± 10^{-3}	$4.775 \pm 1.7 \times 10^{-3}$

tests of this section. A channel-gain map is constructed to portray the channel-gain between every point in the map $\mathbf{x}$, and a fixed receiver location $\mathbf{x}_{\text{rx}}$. Specifically, Alg. 2 is executed and estimates $\{\hat{s}(\tilde{\mathbf{x}}_i, \mathbf{x}_{\text{rx}})\}_{i=1}^{N_g}$ are obtained by substituting $\hat{\mathbf{f}}$ and w into (4). Subsequently, $\{\hat{g}(\tilde{\mathbf{x}}_i, \mathbf{x}_{\text{rx}})\}_{i=1}^{N_g}$ are obtained by substituting $\{\hat{s}(\tilde{\mathbf{x}}_i, \mathbf{x}_{\text{rx}})\}_{i=1}^{N_g}$ into (1) with $\hat{g}_0$ and $\hat{\gamma}$. Upon defining $\hat{\mathbf{g}} := [\hat{g}(\tilde{\mathbf{x}}_1, \mathbf{x}_{\text{rx}}), \dots, \hat{g}(\tilde{\mathbf{x}}_{N_g}, \mathbf{x}_{\text{rx}})]^\top \in \mathbb{R}^{N_g}$, we construct the channel-gain map $\hat{\mathbf{G}} := \text{unvec}(\hat{\mathbf{g}})$ with the receiver located at $\mathbf{x}_{\text{rx}}$.

Let $\hat{\mathbf{S}} := \text{unvec}(\hat{\mathbf{s}})$ denote a shadowing map with $\hat{\mathbf{s}} := [\hat{s}(\tilde{\mathbf{x}}_1, \mathbf{x}_{\text{rx}}), \dots, \hat{s}(\tilde{\mathbf{x}}_{N_g}, \mathbf{x}_{\text{rx}})]^\top \in \mathbb{R}^{N_g}$. Fig. 11 displays estimated shadowing maps and corresponding channel-gain maps constructed via Alg. 2 and the competing alternatives, when the receiver is located at $\mathbf{x}_{\text{rx}} = (10.3, 10.7)$ (ft) marked by the cross. In every channel-gain map of Fig. 11, stronger attenuation is observed when signals propagate through either more building materials (bottom-right side of $\hat{\mathbf{G}}$), or the concrete wall (left side of $\hat{\mathbf{G}}$). On the other hand, only the channel-gain maps in Figs. 11f, 11h, 11j, and 11l constructed by the approximate Bayesian inference methods exhibit less attenuation along the entrance of the structure (top-right side of $\hat{\mathbf{G}}$); this cannot be seen through the channel-gain maps in Figs. 9a and 9b constructed by both regularized LS methods. The reason is that free space and objects are more distinctively delineated in $\hat{\mathbf{F}}$ by the proposed method. All in all, the simulation results confirm that our approach could provide more site-specific information of the propagation medium, and thus endows the operation of cognitive radio networks with more accurate interference management.

V. CONCLUSIONS

This paper developed a variational Bayes approach to adaptive radio tomography, which estimates the spatial loss field of the tomographic model at affordable complexity by using measurements collected from sensing radio pairs that

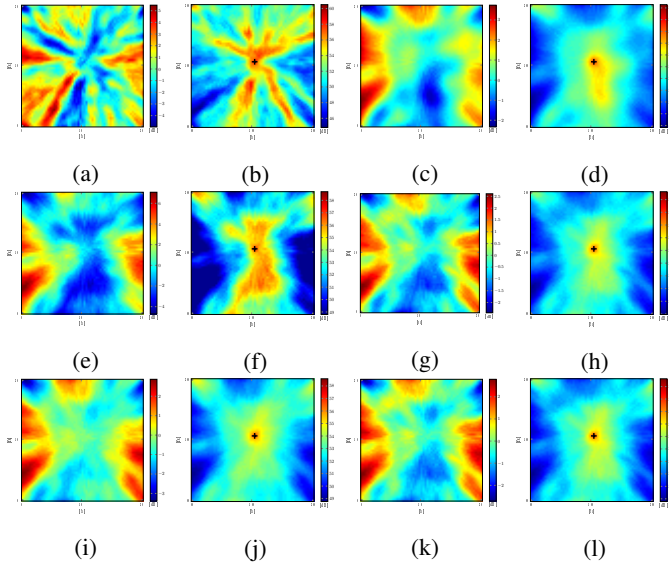


Fig. 11: Estimated shadowing maps $\hat{\mathbf{S}}$ and corresponding channel-gain maps $\hat{\mathbf{G}}$ at $\tau = 5$ via (a)-(b) ridge-regularized LS (setting of Fig. 9a); (c)-(d) TV-regularized LS (setting of Fig. 9b); (e)-(f) adaptive MCMC algorithm in [19] with $K = 2$ (setting of Fig. 9c); (g)-(h) Alg. 2 (setting of Fig. 9e); and (i)-(j) non-adaptive VB algorithm (setting of Fig. 9g); and (k)-(l) benchmark algorithm (setting of Fig. 9i), with the receiver location at $\mathbf{x}_{\text{rx}} = (10.3, 10.7)$ (ft) marked with the black cross.

are adaptively chosen with an uncertainty sampling criterion. Extensive synthetic and real data tests corroborated the efficacy of the proposed novel algorithm for imaging and channel-gain cartography applications. Future research will include an online approach to radio tomography for streaming data.

APPENDIX

Here we derive the variational distributions in (22). Terms not related to a target variable will be lumped in a generic constant c . The iteration index ℓ will be omitted for simplicity.

A. Variational distribution of the SLF $q(\mathbf{f}|\mathbf{z})$

Recall that the conditional posterior obeys $p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}|\check{\mathbf{s}}_t) \propto p(\check{\mathbf{s}}_t|\mathbf{f}, \varphi_\nu)p(\mathbf{f}|\mathbf{z}, \boldsymbol{\theta}_f)$. The first factor in (22), is expressed as

$$\ln q(\mathbf{f}|\mathbf{z}) = \sum_{i=1}^{N_g} \ln q(f_i|z_i) = \sum_{k=1}^K \sum_{i: \tilde{\mathbf{x}}_i \in \mathcal{A}_k} \ln q(f_i|z_i = k) \quad (55)$$

where $\ln q(f_i|z_i = k)$ can be written as

$$\begin{aligned} \ln q(f_i|z_i = k) &= \mathbb{E}_{-q(f_i|z_i=k)} [\ln p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}|\check{\mathbf{s}}_t)] + c \\ &= \mathbb{E}_{-q(f_i|z_i=k)} [\ln p(\check{\mathbf{s}}_t|\mathbf{f}, \varphi_\nu)] \\ &\quad + \mathbb{E}_{-q(f_i|z_i=k)} [\ln p(\mathbf{f}|\mathbf{z}, \boldsymbol{\theta}_f)] + c. \end{aligned} \quad (56)$$

Each term on the RHS in (56) is thus given by

$$\begin{aligned} &\mathbb{E}_{-q(f_i|z_i=k)} [\ln p(\check{\mathbf{s}}_t|\mathbf{f}, \varphi_\nu)] \\ &\leftrightarrow \frac{\tilde{\varphi}_\nu}{2} \sum_{\tau=1}^t \left[w_{\tau,i}^2 f_i^2 - 2 \left(\check{s}_\tau - \sum_{j \neq i} w_{\tau,j} \bar{f}_j \right) w_{\tau,i} f_i \right] \end{aligned} \quad (57)$$

where $\bar{f}_j := \sum_{k=1}^K \zeta_k(\tilde{\mathbf{x}}_j) \check{\mu}_{f_k}(\tilde{\mathbf{x}}_j)$, and

$$\begin{aligned} &\mathbb{E}_{-q(f_i|z_i=k)} [\ln p(\mathbf{f}|\mathbf{z}, \boldsymbol{\theta}_f)] \\ &\leftrightarrow \mathbb{E}_{-q(f_i|z_i=k)} \left[\frac{\varphi_{f_k}}{2} (f_i^2 - 2\mu_{f_k} f_i) \right] = \frac{\tilde{\varphi}_{f_k}}{2} (f_i^2 - 2\check{m}_k f_i). \end{aligned} \quad (58)$$

After substituting (57) and (58) into (56), the pdf $q(f_i|z_i = k)$ can be expressed as

$$\begin{aligned} q(f_i|z_i = k) &\propto \exp \left\{ -\frac{1}{2} \left(\tilde{\varphi}_\nu \sum_{\tau=1}^t w_{\tau,i}^2 + \tilde{\varphi}_{f_k} \right) f_i^2 \right. \\ &\quad \left. + \left[\tilde{\varphi}_\nu \sum_{\tau=1}^t \left(\check{s}_\tau - \sum_{j \neq i} w_{\tau,j} \bar{f}_j \right) w_{\tau,i} + \tilde{\varphi}_{f_k} \check{m}_k \right] f_i \right\}. \end{aligned} \quad (59)$$

By completing the square, it can be readily verified that $q(f_i|z_i = k) = \mathcal{N}(\check{\mu}_{f_k}(\tilde{\mathbf{x}}_i), \check{\sigma}_{f_k}^2(\tilde{\mathbf{x}}_i)) \forall k$, where

$$\check{\sigma}_{f_k}^2(\tilde{\mathbf{x}}_i) = \left(\tilde{\varphi}_\nu \sum_{\tau=1}^t w_{\tau,i}^2 + \tilde{\varphi}_{f_k} \right)^{-1} \quad (60)$$

$$\begin{aligned} \check{\mu}_{f_k}(\tilde{\mathbf{x}}_i) &= \check{\sigma}_{f_k}^2(\tilde{\mathbf{x}}_i) \\ &\times \left[\tilde{\varphi}_\nu \sum_{\tau=1}^t \left(\check{s}_\tau - \sum_{j \neq i} w_{\tau,j} \bar{f}_j \right) w_{\tau,i} + \tilde{\varphi}_{f_k} \check{m}_k \right]. \end{aligned} \quad (61)$$

Upon defining $\bar{s}_\tau := \sum_{i=1}^{N_g} w_{\tau,i} \bar{f}_i$, $\check{\mu}_{f_k}(\tilde{\mathbf{x}}_i)$ in (61), it follows that

$$\begin{aligned} \check{\mu}_{f_k}(\tilde{\mathbf{x}}_i) &= \bar{f}_i + \check{\sigma}_{f_k}^2(\tilde{\mathbf{x}}_i) \\ &\times \left[(\check{m}_k - \bar{f}_i) \tilde{\varphi}_{f_k} + \tilde{\varphi}_\nu \sum_{\tau=1}^t w_{\tau,i} (\check{s}_\tau - \bar{s}_\tau) \right]. \quad \blacksquare \end{aligned} \quad (62)$$

B. Variational distribution of the hidden label field $q(\mathbf{z})$

Since $q(\mathbf{z}) = \prod_{i=1}^{N_g} q(z_i)$ in (22) because z_i and $z_j \forall i \neq j$ are independent, we focus on the derivation of $q(z_i)$. By proportionality of the conditional posterior $p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}|\check{\mathbf{s}}_t) \propto p(\mathbf{f}|\mathbf{z}, \boldsymbol{\theta}_f)p(\mathbf{z}; \beta)$ wrt $\mathbf{z}$, after singling out the terms that involve $q(z_i)$, we arrive at

$$\begin{aligned} \ln q(z_i) &= \mathbb{E}_{-q(z_i)} [\ln p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta}|\check{\mathbf{s}}_t)] + c \\ &\leftrightarrow \mathbb{E}_{-q(z_i)} [\ln p(\mathbf{f}|\mathbf{z}, \boldsymbol{\theta}_f)] + \mathbb{E}_{-q(z_i)} [\ln p(z_i|\mathbf{z}_{-i}; \beta)]. \end{aligned} \quad (63)$$

For $z_i = k$, each term on the RHS in (63) becomes

$$\begin{aligned} &\mathbb{E}_{-q(z_i)} [\ln p(\mathbf{f}|\mathbf{z}, \boldsymbol{\theta}_f)] \leftrightarrow \frac{1}{2} \mathbb{E}_{-q(z_i)} [\ln \varphi_{f_k} - \varphi_{f_k} (f_i - \mu_{f_k})^2] \\ &= -\frac{\tilde{\varphi}_{f_k}}{2} \left[\underbrace{\mathbb{E}_{-q(z_i)} [f_i^2]}_{=\check{\sigma}_{f_k}^2(\tilde{\mathbf{x}}_i) + \check{\mu}_{f_k}^2(\tilde{\mathbf{x}}_i)} - 2\check{m}_k \check{\mu}_{f_k}(\tilde{\mathbf{x}}_i) + \underbrace{\mathbb{E}_{-q(z_i)} [\mu_{f_k}^2]}_{=\check{\sigma}_{f_k}^2(\tilde{\mathbf{x}}_i) + \check{m}_k^2} \right] \\ &\quad + \frac{1}{2} \underbrace{\mathbb{E}_{-q(z_i)} [\ln \varphi_{f_k}]}_{=\psi(\check{a}_k) + \ln \check{b}_k} \end{aligned} \quad (64)$$

and

$$\begin{aligned} &\mathbb{E}_{-q(z_i)} [\ln p(z_i = k|\mathbf{z}_{-i}, \beta)] \\ &\leftrightarrow \mathbb{E}_{-q(z_i)} \left[\beta \sum_{j \in \mathcal{N}(\tilde{\mathbf{x}}_i)} \delta(z_j - k) \right] = \beta \sum_{j \in \mathcal{N}(\tilde{\mathbf{x}}_i)} \zeta_k(\tilde{\mathbf{x}}_j). \end{aligned} \quad (65)$$

All in all, the variational pdf $q(z_i = k)$ becomes

$$q(z_i = k) \propto \exp \left\{ -\frac{\tilde{\varphi}_{f_k}}{2} \left[\check{\sigma}_{f_k}^2(\tilde{\mathbf{x}}_i) + \check{\mu}_{f_k}^2(\tilde{\mathbf{x}}_i) - 2\check{m}_k\check{\mu}_{f_k}(\tilde{\mathbf{x}}_i) + \check{\sigma}_k^2 + \check{m}_k^2 \right] + \frac{1}{2}(\psi(\check{a}_k) + \ln \check{b}_k) + \sum_{j \in \mathcal{N}(\tilde{\mathbf{x}}_i)} \beta \check{\zeta}_k(\tilde{\mathbf{x}}_j) \right\}. \quad (66)$$

which leads to the update rule of $q(z_i = k)$ in (28). ■

C. Variational distribution of the noise precision $q(\varphi_\nu)$

As the conditional posterior $p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} | \tilde{\mathbf{s}}_t)$ is proportional to $p(\tilde{\mathbf{s}}_t | \mathbf{f}, \varphi_\nu) p(\varphi_\nu)$ wrt φ_ν , we can write

$$\begin{aligned} \ln q(\varphi_\nu) &= \mathbb{E}_{-q(\varphi_\nu)} [\ln p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} | \tilde{\mathbf{s}}_t)] + c \\ &\leftrightarrow \mathbb{E}_{-q(\varphi_\nu)} [\ln p(\tilde{\mathbf{s}}_t | \mathbf{f}, \varphi_\nu)] + \mathbb{E}_{-q(\varphi_\nu)} [\ln p(\varphi_\nu)] \end{aligned} \quad (67)$$

where

$$\begin{aligned} \mathbb{E}_{-q(\varphi_\nu)} [\ln p(\tilde{\mathbf{s}}_t | \mathbf{f}, \varphi_\nu)] &\leftrightarrow \frac{t}{2} \ln \varphi_\nu - \frac{\varphi_\nu}{2} \|\tilde{\mathbf{s}}_t - \mathbf{W}_t^\top \mathbf{f}\|_2^2 \\ &= \frac{t}{2} \ln \varphi_\nu - \frac{\varphi_\nu}{2} \sum_{\tau=1}^t \check{s}_\tau^2 - 2\check{s}_\tau \bar{s}_\tau + \mathbb{E}_{-q(\varphi_\nu)} \left[(\mathbf{w}_\tau^{(n,n')\top} \mathbf{f})^2 \right] \end{aligned} \quad (68)$$

and

$$\mathbb{E}_{-q(\varphi_\nu)} [\ln p(\varphi_\nu)] \leftrightarrow (a_\nu - 1) \ln \varphi_\nu - \frac{\varphi_\nu}{b_\nu}. \quad (69)$$

After substituting (68) and (69) into (67), we can easily see that $q(\varphi_\nu) = \mathcal{G}(\check{a}_\nu, \check{b}_\nu)$ with $\check{a}_\nu := a_\nu + t/2$, and

$$\check{b}_\nu := \left(\frac{1}{b_\nu} + \frac{1}{2} \sum_{\tau=1}^t \check{s}_\tau^2 - 2\check{s}_\tau \bar{s}_\tau + \mathbb{E}_{-q(\varphi_\nu)} \left[(\mathbf{w}_\tau^{(n,n')\top} \mathbf{f})^2 \right] \right)^{-1} \quad (70)$$

where

$$\begin{aligned} &\mathbb{E}_{-q(\varphi_\nu)} \left[(\mathbf{w}_\tau^{(n,n')\top} \mathbf{f})^2 \right] \\ &= \text{Var} \left[\mathbf{w}_\tau^{(n,n')\top} \mathbf{f} \right] + \left(\mathbb{E}_{-q(\varphi_\nu)} \left[\mathbf{w}_\tau^{(n,n')\top} \mathbf{f} \right] \right)^2 \end{aligned} \quad (71)$$

$$= \sum_{i=1}^{N_g} w_{\tau,i}^2 \left[\sum_{k=1}^K \check{\zeta}_k(\tilde{\mathbf{x}}_i) (\check{\sigma}_{f_k}^2(\tilde{\mathbf{x}}_i) + \check{\mu}_{f_k}^2(\tilde{\mathbf{x}}_i) - \bar{f}_i^2) \right] + \bar{s}_\tau^2 \quad (72)$$

by the law of total variance on $\text{Var} \left[\mathbf{w}_\tau^{(n,n')\top} \mathbf{f} \right]$ [49, p. 401]. ■

D. Variational distribution of the field means $q(\boldsymbol{\mu}_f)$

Since the conditional posterior $p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} | \tilde{\mathbf{s}}_t)$ is proportional to $p(\mathbf{f} | \mathbf{z}, \boldsymbol{\theta}_f) p(\boldsymbol{\mu}_f)$ wrt $\boldsymbol{\mu}_f$, the entries of $\boldsymbol{\mu}_f$ are iid, we have

$$\begin{aligned} \ln q(\boldsymbol{\mu}_f) &= \mathbb{E}_{-q(\boldsymbol{\mu}_f)} [\ln p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} | \tilde{\mathbf{s}}_t)] + c \\ &\leftrightarrow \mathbb{E}_{-q(\boldsymbol{\mu}_f)} [\ln p(\mathbf{f} | \mathbf{z}, \boldsymbol{\theta}_f)] + \mathbb{E}_{-q(\boldsymbol{\mu}_f)} \left[\sum_{k=1}^K \ln p(\mu_{f_k}) \right] \end{aligned} \quad (73)$$

where

$$\begin{aligned} &\mathbb{E}_{-q(\boldsymbol{\mu}_f)} [\ln p(\mathbf{f} | \mathbf{z}, \boldsymbol{\theta}_f)] \\ &\leftrightarrow \sum_{k=1}^K \sum_{i=1}^{N_g} \check{\zeta}_k(\tilde{\mathbf{x}}_i) \tilde{\varphi}_{f_k} \left(\mu_{f_k}^2 - 2\check{\mu}_{f_k}(\tilde{\mathbf{x}}_i) \mu_{f_k} \right) \end{aligned} \quad (74)$$

and

$$\mathbb{E}_{-q(\boldsymbol{\mu}_f)} \left[\sum_{k=1}^K \ln p(\mu_{f_k}) \right] \leftrightarrow \sum_{k=1}^K \frac{1}{\sigma_k^2} (\mu_{f_k}^2 - 2\mu_{f_k} m_k). \quad (75)$$

Together with (74) and (75), $\ln q(\boldsymbol{\mu}_f)$ becomes

$$\begin{aligned} \ln q(\boldsymbol{\mu}_f) &\leftrightarrow \sum_{k=1}^K \left[\left(\frac{1}{\sigma_k^2} + \sum_{i=1}^{N_g} \check{\zeta}_k(\tilde{\mathbf{x}}_i) \tilde{\varphi}_{f_k} \right) \mu_{f_k}^2 \right. \\ &\quad \left. - 2 \left(\frac{m_k}{\sigma_k^2} + \sum_{i=1}^{N_g} \check{\zeta}_k(\tilde{\mathbf{x}}_i) \tilde{\varphi}_{f_k} \check{\mu}_{f_k}(\tilde{\mathbf{x}}_i) \right) \mu_{f_k} \right]. \end{aligned} \quad (76)$$

After completing the square of the summand in (76), we find $q(\mu_{f_k}) = \mathcal{N}(\check{m}_k, \check{\sigma}_k^2) \forall k$ with

$$\check{\sigma}_k^2 := \left(\frac{1}{\sigma_k^2} + \sum_{i=1}^{N_g} \check{\zeta}_k(\tilde{\mathbf{x}}_i) \tilde{\varphi}_{f_k} \right)^{-1} \quad (77)$$

$$\check{m}_k := \check{\sigma}_k^2 \left(\frac{m_k}{\sigma_k^2} + \sum_{i=1}^{N_g} \check{\zeta}_k(\tilde{\mathbf{x}}_i) \tilde{\varphi}_{f_k} \check{\mu}_{f_k}(\tilde{\mathbf{x}}_i) \right) \quad (78)$$

by inspection since $q(\boldsymbol{\mu}_f) = \prod_{k=1}^K q(\mu_{f_k})$, as in (22). ■

E. Variational distribution of the field precisions $q(\varphi_f)$

Similar to $q(\boldsymbol{\mu}_f)$, the pdf $q(\varphi_f)$ can be expressed as

$$\begin{aligned} \ln q(\varphi_f) &= \mathbb{E}_{-q(\varphi_f)} [\ln p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} | \tilde{\mathbf{s}}_t)] + c \\ &\leftrightarrow \mathbb{E}_{-q(\varphi_f)} [\ln p(\mathbf{f} | \mathbf{z}, \boldsymbol{\theta}_f)] + \mathbb{E}_{-q(\varphi_f)} \left[\sum_{k=1}^K \ln p(\varphi_{f_k}) \right] \end{aligned} \quad (79)$$

by appealing to the proportionality of the conditional posterior $p(\mathbf{f}, \mathbf{z}, \boldsymbol{\theta} | \tilde{\mathbf{s}}_t) \propto p(\mathbf{f} | \mathbf{z}, \boldsymbol{\theta}_f) p(\varphi_f)$ w.r.t. φ_f . Each term on the RHS in (79) can be thus expressed as

$$\begin{aligned} &\mathbb{E}_{-q(\varphi_f)} [\ln p(\mathbf{f} | \mathbf{z}, \boldsymbol{\theta}_f)] \\ &= \frac{1}{2} \sum_{k=1}^K \sum_{i=1}^{N_g} \check{\zeta}_k(\tilde{\mathbf{x}}_i) \left[\ln \varphi_{f_k} - \varphi_{f_k} \mathbb{E}_{-q(z_i)} [(f_i - \mu_{f_k})^2] \right] + c \end{aligned} \quad (80)$$

where

$$\begin{aligned} \mathbb{E}_{-q(z_i)} [(f_i - \mu_{f_k})^2] &= \check{\sigma}_{f_k}^2(\tilde{\mathbf{x}}_i) + \check{\mu}_{f_k}^2(\tilde{\mathbf{x}}_i) - 2\check{m}_k\check{\mu}_{f_k}(\tilde{\mathbf{x}}_i) \\ &\quad + \check{\sigma}_k^2 + \check{m}_k^2, \end{aligned} \quad (81)$$

and

$$\mathbb{E}_{-q(\varphi_f)} \left[\sum_{k=1}^K \ln p(\varphi_{f_k}) \right] = \sum_{k=1}^K \left[(a_k - 1) \ln \varphi_{f_k} - \frac{\varphi_{f_k}}{b_k} \right] + c. \quad (82)$$

After substituting (80) and (82) into (79), φ_f can be shown to follow $q(\varphi_{f_k}) = \mathcal{G}(\check{a}_k, \check{b}_k) \forall k$ with

$$\check{a}_k := a_k + \frac{1}{2} \sum_{i=1}^{N_g} \check{\zeta}_k(\tilde{\mathbf{x}}_i) \quad (83)$$

$$\check{b}_k := \left[\frac{1}{b_k} + \frac{1}{2} \sum_{i=1}^{N_g} \check{\zeta}_k(\tilde{\mathbf{x}}_i) \right. \quad (84)$$

$$\left. \times \left(\check{\sigma}_{f_k}^2(\tilde{\mathbf{x}}_i) + \check{\mu}_{f_k}^2(\tilde{\mathbf{x}}_i) - 2\check{\mu}_{f_k}(\tilde{\mathbf{x}}_i) \check{m}_k + \check{\sigma}_k^2 + \check{m}_k^2 \right) \right]^{-1}$$

where we used that $q(\varphi_f) = \prod_{k=1}^K q(\varphi_{f_k})$, as in (22). ■

F. Derivation of the cross-entropy $H(\mathbf{f}|\mathbf{z}, \check{\mathbf{s}}_{\tau+1}; \hat{\boldsymbol{\theta}}_{\tau})$

To establish the expression for $H(\mathbf{f}|\mathbf{z}, \check{\mathbf{s}}_{\tau+1}; \hat{\boldsymbol{\theta}}_{\tau})$ in (52), consider that at time slot $\tau + 1$. Similar to (51), we have

$$H(\mathbf{f}|\mathbf{z} = \mathbf{z}', \check{\mathbf{s}}_{\tau+1} = \check{\mathbf{s}}'_{\tau+1}; \hat{\boldsymbol{\theta}}_{\tau}) \approx \frac{1}{2} \ln \left| \check{\Sigma}_{\mathbf{f}|\mathbf{z}', \check{\mathbf{s}}'_{\tau+1}; \hat{\boldsymbol{\theta}}_{\tau}} \right| + \frac{N_g}{2} (1 + \ln 2\pi). \quad (85)$$

With $\Delta_{w_{\tau+1}} := \text{diag}(\mathbf{w}_{\tau+1}^{(n, n')} \circ \mathbf{w}_{\tau+1}^{(n, n')})$, and using the construction of $\check{\sigma}_{f_k}^2(\tilde{\mathbf{x}}_i)$ in (60), we can write

$$\check{\Sigma}_{\mathbf{f}|\mathbf{z}', \check{\mathbf{s}}'_{\tau+1}; \hat{\boldsymbol{\theta}}_{\tau}} = \left[\check{\Sigma}_{\mathbf{f}|\mathbf{z}', \check{\mathbf{s}}'_{\tau}; \hat{\boldsymbol{\theta}}_{\tau}}^{-1} + \tilde{\varphi}_{\nu} \Delta_{w_{\tau+1}} \right]^{-1} \quad (86)$$

from which we deduce that

$$\left| \check{\Sigma}_{\mathbf{f}|\mathbf{z}', \check{\mathbf{s}}'_{\tau+1}; \hat{\boldsymbol{\theta}}_{\tau}} \right|^{-1} = \left| \mathbf{I}_{N_g} + \tilde{\varphi}_{\nu} \Delta_{w_{\tau+1}} \check{\Sigma}_{\mathbf{f}|\mathbf{z}', \check{\mathbf{s}}'_{\tau}; \hat{\boldsymbol{\theta}}_{\tau}} \right| \left| \check{\Sigma}_{\mathbf{f}|\mathbf{z}', \check{\mathbf{s}}'_{\tau}; \hat{\boldsymbol{\theta}}_{\tau}}^{-1} \right| \quad (87)$$

by using the matrix determinant identity lemma [43, Chapter 18]. Further substituting (87) into (85), leads to

$$H(\mathbf{f}|\mathbf{z} = \mathbf{z}', \check{\mathbf{s}}_{\tau+1} = \check{\mathbf{s}}'_{\tau+1}; \hat{\boldsymbol{\theta}}_{\tau}) \approx H(\mathbf{f}|\mathbf{z} = \mathbf{z}', \check{\mathbf{s}}_{\tau} = \check{\mathbf{s}}'_{\tau}; \hat{\boldsymbol{\theta}}_{\tau}) - \frac{1}{2} \ln \left| \mathbf{I}_{N_g} + \tilde{\varphi}_{\nu} \Delta_{w_{\tau+1}} \check{\Sigma}_{\mathbf{f}|\mathbf{z}', \check{\mathbf{s}}'_{\tau}; \hat{\boldsymbol{\theta}}_{\tau}} \right|. \quad (88)$$

It follows from the conditional entropy definition in (46) that

$$\begin{aligned} H(\mathbf{f}|\mathbf{z}, \check{\mathbf{s}}_{\tau+1}; \hat{\boldsymbol{\theta}}_{\tau}) &\approx \sum_{\mathbf{z}' \in \mathcal{Z}} \int p(\mathbf{z}', \check{\mathbf{s}}'_{\tau+1}; \hat{\boldsymbol{\theta}}_{\tau}) \left(H(\mathbf{f}|\mathbf{z} = \mathbf{z}', \check{\mathbf{s}}_{\tau} = \check{\mathbf{s}}'_{\tau}; \hat{\boldsymbol{\theta}}_{\tau}) \right. \\ &\quad \left. - \frac{1}{2} \ln \left| \mathbf{I}_{N_g} + \tilde{\varphi}_{\nu} \Delta_{w_{\tau+1}} \check{\Sigma}_{\mathbf{f}|\mathbf{z}', \check{\mathbf{s}}'_{\tau}; \hat{\boldsymbol{\theta}}_{\tau}} \right| \right) d\check{\mathbf{s}}'_{\tau+1} \\ &\stackrel{(e1)}{=} H(\mathbf{f}|\mathbf{z}, \check{\mathbf{s}}_{\tau}; \hat{\boldsymbol{\theta}}_{\tau}) \\ &\quad - \sum_{\mathbf{z}' \in \mathcal{Z}} \int p(\mathbf{z}', \check{\mathbf{s}}'_{\tau}; \hat{\boldsymbol{\theta}}_{\tau}) \frac{1}{2} \ln \left| \mathbf{I}_{N_g} + \tilde{\varphi}_{\nu} \Delta_{w_{\tau+1}} \check{\Sigma}_{\mathbf{f}|\mathbf{z}', \check{\mathbf{s}}'_{\tau}; \hat{\boldsymbol{\theta}}_{\tau}} \right| d\check{\mathbf{s}}'_{\tau}, \end{aligned} \quad (89)$$

where (e1) is obtained after marginalizing out $\check{\mathbf{s}}_{\tau+1}$ from $p(\mathbf{z}', \check{\mathbf{s}}'_{\tau+1}; \hat{\boldsymbol{\theta}}_{\tau})$ as the RHS of (88) is not a function of $\check{\mathbf{s}}_{\tau+1}$. ■

REFERENCES

- [1] D. Lee and G. B. Giannakis, "A variational Bayes approach to adaptive channel-gain cartography," in *Proc. 2019 Intl. Conf. on Acoust., Speech, and Signal Process.*, Brighton, U.K., May 2019.
- [2] N. B. Smith and A. Webb, *Introduction to Medical Imaging: Physics, Engineering and Clinical Applications*. Cambridge University Press, 2010.
- [3] N. Patwari and P. Agrawal, "Effects of correlated shadowing: Connectivity, localization, and RF tomography," in *Int. Conf. Info. Process. Sensor Networks*, St. Louis, MO, Apr. 2008, pp. 82–93.
- [4] J. Wilson, N. Patwari, and O. G. Vasquez, "Regularization methods for radio tomographic imaging," in *Virginia Tech Symp. Wireless Personal Commun.*, Blacksburg, VA, Jun. 2009.
- [5] S.-J. Kim, E. Dall'Anese, and G. B. Giannakis, "Cooperative spectrum sensing for cognitive radios using kriged Kalman filtering," *IEEE J. Sel. Topics Sig. Process.*, vol. 5, no. 1, pp. 24–36, Feb. 2011.
- [6] K. Woyach, D. Puccinelli, and M. Haenggi, "Sensorless sensing in wireless networks: Implementation and measurements," in *Int. Symp. Modeling Optimization Mobile, Ad Hoc Wireless Netw.*, Apr. 2006, pp. 1–8.
- [7] M. Youssef, M. Mah, and A. Agrawala, "Challenges: Device-free passive localization for wireless environments," in *Proc. ACM MobiCom*, Sep. 2007, pp. 222–229.
- [8] J. Wilson and N. Patwari, "See-through walls: Motion tracking using variance-based radio tomography networks," *IEEE Trans. Mobile Comput.*, vol. 10, no. 5, pp. 612–621, 2011.
- [9] Q. Zhao and B. M. Sadler, "A survey of dynamic spectrum access," *IEEE Sig. Process. Mag.*, vol. 24, no. 3, pp. 79–89, 2007.
- [10] Federal Communications Commission, "FCC 11-131," *Unlicensed Operation in the TV Broadcast Bands*, 2011.
- [11] S.-J. Kim, E. Dall'Anese, J. A. Bazerque, K. Rajawat, and G. B. Giannakis, "Advances in spectrum sensing and cross-layer design for cognitive radio networks," in *Academic Press Library in Signal Processing: Communications and Radar Signal Processing: 2*. Academic Press, 2013, ch. 9, pp. 471–497.
- [12] A. A. Khan, M. H. Rehmani, and A. Rachedi, "Cognitive-radio-based Internet of Things: Applications, architectures, spectrum related functionalities, and future research directions," *IEEE Wireless Commun.*, vol. 24, no. 3, pp. 17–25, Jun. 2017.
- [13] P. Agrawal and N. Patwari, "Correlated link shadow fading in multi-hop wireless networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 9, pp. 4024–4036, Aug. 2009.
- [14] B. R. Hamilton, X. Ma, R. J. Baxley, and S. M. Matechik, "Propagation modeling for radio frequency tomography in wireless networks," *IEEE J. Sel. Topics Sig. Proc.*, vol. 8, no. 1, pp. 55–65, Feb. 2014.
- [15] D. Romero, D. Lee, and G. B. Giannakis, "Blind radio tomography," *IEEE Trans. Sig. Proc.*, vol. 66, no. 8, pp. 2055–2069, Apr. 2018.
- [16] J. Wilson and N. Patwari, "Radio tomographic imaging with wireless networks," *IEEE Trans. Mobile Comput.*, vol. 9, no. 5, pp. 621–632, 2010.
- [17] O. Kaltiokallio, M. Bocca, and N. Patwari, "Enhancing the accuracy of radio tomographic imaging using channel diversity," in *Proc. of the 9th IEEE Int. Conf. on Mobile Ad hoc and Sensor Syst.*, Las Vegas, NV, Oct. 2012, pp. 254–262.
- [18] A. Gonzalez-Ruiz and Y. Mostofi, "Cooperative robotic structure mapping using wireless measurements - a comparison of random and coordinated sampling patterns," *IEEE Sensors J.*, vol. 13, no. 7, pp. 2571–2580, Jul. 2013.
- [19] D. Lee, D. Berberidis, and G. B. Giannakis, "Adaptive Bayesian radio tomography," *IEEE Trans. Sig. Proc.*, vol. 67, no. 8, pp. 1964–1977, Apr. 2019.
- [20] D. Lee, S.-J. Kim, and G. B. Giannakis, "Channel gain cartography for cognitive radios leveraging low rank and sparsity," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5953–5966, Nov. 2017.
- [21] B. Beck, X. Ma, and R. Baxley, "Ultrawideband tomographic imaging in uncalibrated networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6474–6486, Sep. 2016.
- [22] E. Dall'Anese, S.-J. Kim, and G. B. Giannakis, "Channel gain map tracking via distributed kriging," *IEEE Trans. Veh. Technol.*, vol. 60, no. 3, pp. 1205–1211, 2011.
- [23] D. Higdon, "Spatial applications of Markov chain Monte Carlo for Bayesian inference," Ph.D. dissertation, Dept. Stat., Univ. Washington, Seattle, WA, 1994.
- [24] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall, 1996.
- [25] V. V. Fedorov, *Theory of Optimal Experiments*. New York, NY: Academic Press, 1972.
- [26] D. MacKay, "Information-based objective functions for active data selection," *Neural Comput.*, vol. 4, no. 4, pp. 590–604, 1992.
- [27] J. Li, R. L. Ewing, and X. Shen, "Radio frequency tomographic reconstruction based on convolutional neural networks," in *Proc. of IEEE Nat. Aerosp. and Electron. Conf.*, Dayton, OH, 2018.
- [28] L. Zhao, C. Su, H. Huang, Z. Han, S. Ding, and X. Li, "Intrusion detection based on device-free localization in the era of IoT," *Symmetry*, vol. 11, no. 630, pp. 1–15, May 2019.
- [29] J. Hammersley and P. Clifford, "Markov field on finite graphs and lattices," 1971. [Online]. Available: <http://www.statslab.cam.ac.uk/~grg/books/hammfest/hamm-cliff.pdf>
- [30] X. Descombes, R. D. Morris, J. Zerubia, and M. Berthod, "Estimation of Markov random field prior parameters using Markov Chain Monte Carlo maximum likelihood," *IEEE Trans. Sig. Proc.*, vol. 8, no. 7, pp. 954–963, Jul. 1999.
- [31] C. A. McGrory, D. M. Titterton, R. Reeves, and A. N. Pettitt, "Variational Bayes for estimating the parameters of a hidden Potts model," *Stat. Comput.*, vol. 19, no. 3, pp. 329–340, 2009.
- [32] H. Ayasso and A. Mohammad-Djafari, "Joint NDT image restoration and segmentation using Gauss-Markov-Potts models and variational Bayesian computation," *IEEE Trans. Image Process.*, vol. 19, no. 9, pp. 2265–2277, Sep. 2010.

- [33] G. Kail, J.-Y. Tournet, F. Hlawatsch, and N. Dobigeon, "Blind deconvolution of sparse pulse sequences under a minimum distance constraint: a partially collapsed Gibbs sampler method," *IEEE Trans. Sig. Proc.*, vol. 60, no. 6, pp. 2727 – 2743, 2012.
- [34] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Found. Trends Mach. Learn.*, vol. 1, no. 1–2, pp. 1 – 305, 2008.
- [35] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," in *Proc. Intl. Conf. of Learn. Representations*, Banff, Canada, Apr. 2014.
- [36] C. Blundell, J. Cornebise, K. Kavukcuoglu, and D. Wierstra, "Weight uncertainty in neural networks," in *Proc. Intl. Conf. of Mach. Learn.*, Lille, France, Jul. 2015, pp. 1613–1622.
- [37] M. Oppor and D. Saad, *Advanced mean field methods: Theory and practice*. Cambridge, MA: MIT Press, 2001.
- [38] E. Riegler, G. E. Korkelund, C. N. Manchon, M.-A. Badiu, and B. H. Fleury, "Merging belief propagation and the mean field approximation: A free energy approach," *IEEE Trans. Info. Theory*, vol. 59, no. 1, pp. 588 – 602, Jan. 2013.
- [39] K. P. Murphy, *Machine Learning: A Probabilistic Perspective*. Cambridge, MA: MIT Press, 2012.
- [40] M. J. Beal, *Variational Algorithms for Approximate Bayesian Inference*. London, U.K: Gatsby Comput. Neurosci. Unit, Univ. College London, 2003.
- [41] Z. Qin, D. Goldfarb, and S. Ma, "An alternating direction method for total variation denoising," *Optim. Method Softw.*, vol. 30, no. 3, pp. 594–615, 2015.
- [42] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. New York, NY: USA:Wiley, 1991.
- [43] D. A. Harville, *Matrix Algebra From a Statistician's Perspective*. New York, NY: Springer-Verlag, 1997.
- [44] M. Pereyra, N. Dobigeon, H. Batatia, and J.-Y. Tournet, "Estimating the granularity coefficient of a Potts-Markov random field within a Markov chain Monte Carlo algorithm," *IEEE Trans. Image Process.*, vol. 2, no. 6, pp. 2385 – 2397, Jun. 2013.
- [45] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, no. 6, pp. 721–741, Nov. 1984.
- [46] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*. Philadelphia, PA: SIAM, 1974.
- [47] G. H. Golub, M. Heath, and G. Wahba, "Generalized cross-validation as a method for choosing a good ridge parameter," *Technometrics*, vol. 21, no. 2, pp. 215 – 223, May 1979.
- [48] C. M. Bishop, *Pattern Recognition and Machine Learning*, ser. Information Science and Statistics. Springer, 2006. [Online]. Available: <https://books.google.com/books?id=kTNoQgAACAAJ>
- [49] J. K. Blitzstein and J. Hwang, *Introduction to Probability*. Boca Raton, FL: USA:CRC, 2014.



Georgios B. Giannakis (Fellow'97) received his Diploma in Electrical Engr. from the Ntl. Tech. Univ. of Athens, Greece, 1981. From 1982 to 1986 he was with the Univ. of Southern California (USC), where he received his MSc. in Electrical Engineering, 1983, MSc. in Mathematics, 1986, and Ph.D. in Electrical Engr., 1986. He was a faculty member with the University of Virginia from 1987 to 1998, and since 1999 he has been a professor with the Univ. of Minnesota, where he holds an ADC Endowed Chair, a University of Minnesota McKnight Presidential Chair in ECE, and serves as director of the Digital Technology Center.

His general interests span the areas of statistical learning, communications, and networking - subjects on which he has published more than 465 journal papers, 765 conference papers, 25 book chapters, two edited books and two research monographs. Current research focuses on Data Science, and Network Science with applications to the Internet of Things, and power networks with renewables. He is the (co-) inventor of 33 issued patents, and the (co-) recipient of 9 best journal paper awards from the IEEE Signal Processing (SP) and Communications Societies, including the G. Marconi Prize Paper Award in Wireless Communications. He also received the IEEE-SPS Nobert Wiener Society Award (2019); EURASIP's A. Papoulis Society Award (2020); Technical Achievement Awards from the IEEE-SPS (2000) and from EURASIP (2005); the IEEE ComSoc Education Award (2019); the G. W. Taylor Award for Distinguished Research from the University of Minnesota, and the IEEE Fourier Technical Field Award (2015). He is a Fellow of the National Academy of Inventors, the European Academy of Sciences, IEEE and EURASIP. He has served the IEEE in a number of posts, including that of a Distinguished Lecturer for the IEEE-SPS.



Donghoon Lee (Member'20) received his B.Eng. and M.Eng. degrees in electrical engineering from Korea University, Seoul, Korea in 2010 and 2013, respectively. From 2013 to 2019, he was with the University of Minnesota (UMN), Minneapolis, MN, USA, where he received the Ph.D. degree in electrical and computer engineering. From 2010 to 2011, he was an invited researcher at the Imaging Media Research Center in the Korea Institute of Science and Technology (KIST), Seoul, Korea.

He is currently a senior engineer with Samsung Electronics America. His research interests include statistical signal processing and machine learning, with applications to wireless communication and networking.