A Variational Bayes Approach to Adaptive Radio Tomography

Donghoon Lee, Student 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 [37]. 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 [33]. 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 [40] and channel-gain cartography [24]. 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 [41], [42], and has found diverse applications in disaster response for e.g., detecting individuals trapped in buildings or smoke [39]. 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 [24]. 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 [43], [10], [23]. 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) [22].

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 [33], [1]. 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 [33], [16], [36], which models the SLF effect on shadowing over a link. Inspired by this SLF model, various tomographic imaging methods were proposed [40], [39], [38], [21]. To detect locations of changes in the propagation environment, one can use the difference between the SLF across consecutive time slots [40], [38]. To cope with multipath in a cluttered environment, multi-channel measurements can be utilized to enhance localization accuracy [21]. 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 [39], 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 [15], [16], [26], [28] 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 [4], [36].

Another body of work leveraging the SLF model is that of channel-gain cartography when employing tomography based approaches [24], [36], [8], [28]. Linear interpolation techniques such as kriging were further employed to estimate...
shadowing based on spatially correlated measurements [8], while spatio-temporal dynamics were tracked via Kalman filtering [24]. 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 [28]. While the aforementioned methods rely on heuristic criteria to choose the weight function, [36] 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 [16], [28], [38]. 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 [26] a Bayesian approach to learn a piecewise homogeneous SLF through a binary hidden Markov random field (MRF) model [19] via Markov chain Monte Carlo (MCMC) [13]. 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, 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 [11] and active learning [29]. 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 solution can be set to unity without loss of generality by absorbing the pathloss exponent, $\gamma$, where $\lambda$ is a tunable parameter. The value of $\lambda$ is commonly set to the wavelength to assign non-zero weights only within the first Fresnel zone. In radio tomography, the integral in (2) is approximated by a finite sum as

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

where $\{\tilde{x}_i\}_{i=1}^{N_g}$ is a grid of points over $A$ 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 $\{\mathbf{x}_1, \ldots, \mathbf{x}_N\} \subset A$ collaborate to obtain channel-gain measurements. At time slot $\tau$, the radios indexed by $n(\tau)$ and $n'(\tau)$ measure the channel-gain $\hat{g}_{\tau} := g(\mathbf{x}_{n(\tau)}, \mathbf{x}_{n'(\tau)}) + \nu_{\tau}$ by exchanging training sequences known to both transmitting and receiving radios, where $n(\tau), n'(\tau) \in \{1, \ldots, N\}$ and $\nu_{\tau}$ denotes measurement noise. It is supposed that $g_0$ and $\gamma$ have been estimated during a calibration stage. After subtracting known components from $\hat{g}_{\tau}$, the shadowing estimate is found as

$$\hat{s}_{\tau} := g_0 - \gamma 10 \log_{10} d(\mathbf{x}_{n(\tau)}, \mathbf{x}_{n'(\tau)}) - \hat{g}_{\tau} = s(\mathbf{x}_{n(\tau)}, \mathbf{x}_{n'(\tau)}) - \nu_{\tau}.$$
Having available $\mathbf{s}_t := [\mathbf{s}_1, \ldots, \mathbf{s}_k]^T \in \mathbb{R}^k$ along with the known set of links $\{(\mathbf{x}_{n(\tau)}, \mathbf{x}_{n(\tau)'})\}_{\tau=1}^K$ and the weight function $w$ at the fusion center, the problem is to estimate $f$, and thus $f := [f(\mathbf{x}_1), \ldots, f(\mathbf{x}_N)]^T \in \mathbb{R}^N$ using (4).

Conventional regularized LS estimators of $f$ solve [16], [38]

$$\min_{f} \sum_{\tau=1}^T \left( s_{\tau} - \sum_{i=1}^N w(\mathbf{x}_{n(\tau)}, \mathbf{x}_{n(\tau)'}, \mathbf{x}_i) f(\mathbf{x}_i) \right)^2 + \rho_f \mathcal{R}(f)$$

(6)

where $\mathcal{R} : \mathbb{R}^N \rightarrow \mathbb{R}$ is a generic regularizer to promote a known attribute of $f$, and $\rho_f \geq 0$ is a regularization scalar to reflect compliance of $f$ 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.

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

III. ADAPTIVE BAYESIAN RADIO TOMOGRAPHY

In this section, we view $f$ 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 $\mathcal{A}_k := \{\mathbf{x} \in \mathbb{R}^N | \mu_{f_k} = \mu_{f_k}^{\text{true}}\}$ for $k = 1, \ldots, K$, giving rise to a latent random label field $z := [z(\mathbf{x}_1), \ldots, z(\mathbf{x}_N)]^T \in \{1, \ldots, K\}^N$ with $K$-ary entries $z(\mathbf{x}_i) = k$ if $\mathbf{x}_i \in \mathcal{A}_k \forall i, k$. The $K$ separate regions will model heterogeneous environments. 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 $f(\mathbf{x}_i)$ as

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

We further assign the Potts prior to $z$ in order to capture the dependency among spatially correlated labels. By the Hammersley-Clifford theorem [17], the Potts prior of $z$ follows a Gibbs distribution

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

(8)

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

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

(9)

is the partition function with $\mathcal{Z} := \{1, \ldots, K\}^N$. To ease exposition, $\beta$ is assumed known or fixed a priori; see e.g., [9], [30], [26] for a means of estimating $\beta$. If $\{f(\mathbf{x}_i)\}_{i=1}^N$ are conditionally independent given $z$, the model reduces to the Gauss-Markov-Potts model [2]. Such a model with $K = 3$ is depicted in Fig. 2 with the measurement model in (4).

Noise $\nu_t$ in (5) is assumed independent and identically distributed (i.i.d.) Gaussian with zero mean and variance $\sigma_{\nu_t}^2$. Here, we correspondingly consider precisions of $\nu_t$ 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 $\theta$ be a hyperparameter vector comprising $\varphi_{\nu}$ and $\theta_f := [\varphi_{f_1}, \varphi_{f_2}]$ with $\mu_f := [\mu_{f_1}, \ldots, \mu_{f_K}]^T \in \mathbb{R}^K$ and $\varphi_f := [\varphi_{f_1}, \ldots, \varphi_{f_K}]^T \in \mathbb{R}^K$. Assuming the independence among entries of $\theta$, we deduce that

$$p(\theta) = p(\varphi_{\nu}) p(\mu_f) p(\varphi_f) = p(\varphi_{\nu}) \prod_{k=1}^K p(\mu_{f_k}) p(\varphi_{f_k})$$

(10)

where the priors $p(\varphi_{\nu})$, $p(\mu_f)$, and $p(\varphi_f)$ are as follows.

1) Noise precision $\varphi_{\nu}$: With additive Gaussian noise having fixed mean, it is common to assign a conjugate prior to $\varphi_{\nu}$
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
\[
p(\varphi_{\nu}) = \mathcal{G}(b_{\nu}, a_{\nu}) := \frac{1}{\Gamma(a_{\nu})b_{\nu}^{a_{\nu}}}(\varphi_{\nu})^{a_{\nu}-1}e^{-\varphi_{\nu}/b_{\nu}}
\]
where \( a_{\nu} \) is referred to as the shape parameter, \( b_{\nu} \) as the scale parameter, and \( \Gamma(\cdot) \) denotes the gamma function.

2) Hyperparameters \( \theta_f \) of the SLF: While the prior for \( \mu_{\nu} \) is assumed to be Gaussian with mean \( \mu_{\nu} \) and variance \( \sigma_{\nu}^2 \) (see also [2]), similar to the noise, the prior for \( \varphi_{\nu} \in \mathbb{R}^+ \) is the Gamma distribution parameterized by \( \{a_{\nu}, b_{\nu}\} \); that is,
\[
p(\varphi_{\nu}) = \mathcal{G}(a_{\nu}, b_{\nu}), \quad k = 1, \ldots, K
\]
\[
p(\varphi_{\nu}) = \mathcal{G}(a_{\nu}, b_{\nu}), \quad k = 1, \ldots, K
\]
We stress that analytical tractability is the main motivation behind selecting the conjugate priors in (11)–(13).

Our goal of inferring \( f \), relies on the following posterior distribution (within a constant) as
\[
p(f, z, \theta|s_t) \propto p(s_t|f, \varphi_{\nu})p(f|z, \theta_f)p(z; \beta)p(\theta)
\]
where \( p(s_t|f, \varphi_{\nu}) \sim \mathcal{N}(w^T_f \varphi_{\nu}, \sigma^2_{\nu} \mathbf{I}_k) \) is the data likelihood with the weight matrix \( w^T_f \in \mathbb{R}^{N_t \times K} \) formed from columns \( w^T_{n,\tau} := [w(x_{n,(\tau)}^1, x_{n,(\tau)}^2, \ldots, x_{n,(\tau)}^m)]_T \): where we drop the constant \( p(s_t) \) from the posterior that resulted in the so-called evidence lower bound (ELBO) in (P1), which involves the joint \( p(f, z, \theta, s_t) \) factored as in the right-hand side (RHS) of (14). We choose the family \( Q \) as
\[
Q := \left\{ q : q(f, z, \theta) := q(f|z)q(z|\theta) \right\}
\]
where \( f := f(\mathbf{x}_t) \) and \( z := z(\mathbf{x}_t) \) \( \forall i \) for simplicity, and
\[
q(\theta) := q(\varphi_{\nu})q(\mu_f)q(\varphi_f) = q(\varphi_{\nu}) \prod_{k=1}^K q(\mu_{f_k}) \prod_{k=1}^K q(\varphi_{f_k}).
\]

B. Radio tomography using variational Bayes

Although the estimator forms in (15)–(19) have been considered also in [20], 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 [13] that relies on samples of \( \{f, z, \theta\} \) 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. Aimed at a scalable alternative, we will adopt the so-called variational Bayes (VB) approach.

VB is a family of techniques to approximate a complex distribution by a tractable one termed variational distribution. A typical choice of an approximation criterion is to find the variational distribution \( q \) minimizing the Kullback-Leibler (KL) divergence \( (D_{KL}(q||p)) \) to a target distribution \( p \). The variational distribution \( q \) is further assumed to belong to a certain family \( Q \) of distributions possessing a simpler form of dependence between variables than the original one; see also [32] for the so-called mean-field approximation.

Tailored to the posterior in (14) the variational one, solves
\[
\min_{q(f, z, \theta) \in Q} D_{KL}(q(f, z, \theta)||p(f, z, \theta|s_t))
\]
Using that \( D_{KL}(q||p) := -E_q[\ln(p/q)] \), the latter reduces to
\[
\max_{q(f, z, \theta) \in Q} E_q[f, z, \theta] \left[ \ln \left( \frac{p(f, z, \theta|s_t)}{q(f, z, \theta)} \right) \right]
\]
where \( \ell := \text{ELBO}(q(f, z, \theta)) \)

Following the general VB steps [31], we will solve our (P1) here via coordinate minimization among factors of \( q(f, z, \theta) \). Within a constant \( c \), the optimal solutions have the form
\[
\ln q^*(f_i|z_i) = E_{-q(f_i|z_i)}[\ln p(f, z, \theta, s_t)] + c \quad \forall i
\]
\[
\ln q^*(z_i) = E_{-q(z_i)}[\ln p(f, z, \theta, s_t)] + c \quad \forall i
\]
\[
\ln q^*(\theta) = E_{-q(\theta)}[\ln p(f, z, \theta, s_t)] + c
\]
where the expectation in (24) is over the variational pdf of \( f_i \), \( z_i \), and \( \theta \), that is, \( \prod_{i \neq i} q(f_j|z_j)q(z_j|\theta) \). 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 \( \ell = 1, 2, \ldots \), we have
\[
q^{(\ell)}(f_i|z_i = k) = \mathcal{N}(\mu_{f_{ik}}^{(\ell)}(\mathbf{x}_t), \sigma^2_{f_{ik}}^{(\ell)}(\mathbf{x}_t)) \quad \forall k
\]
\[
q^{(\ell)}(z_i = k) = \mathcal{N}(\mu_{f_{ik}}^{(\ell)}(\mathbf{x}_t), \sigma^2_{f_{ik}}^{(\ell)}(\mathbf{x}_t)) \quad \forall k
\]
\[
q^{(\ell)}(\varphi_{\nu}) = \mathcal{G}(\mu_{\nu}^{(\ell)}(\mathbf{x}_t), \sigma^2_{\nu}^{(\ell)}(\mathbf{x}_t)) \quad \forall k
\]
\[
q^{(\ell)}(\mu_{f_k}) = \mathcal{G}(\mu_{f_{ik}}^{(\ell)}(\mathbf{x}_t), \sigma^2_{f_{ik}}^{(\ell)}(\mathbf{x}_t)) \quad \forall k
\]
\[
q^{(\ell)}(\varphi_{f_k}) = \mathcal{G}(\mu_{f_{ik}}^{(\ell)}(\mathbf{x}_t), \sigma^2_{f_{ik}}^{(\ell)}(\mathbf{x}_t)) \quad \forall k
\]
with variational parameters

\[
\tilde{\sigma}_{f_k}^2(\tilde{x}_i) = \left( \tilde{\varphi}_\nu(\ell) + \varphi_{f_k}(\ell) \right)^{-1} \quad \forall k
\] (32)

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

\[
\tilde{\varphi}_{\nu}(\ell) = \exp \left\{ - \frac{\varphi_{f_k}(\ell)}{2} \left[ \frac{1}{\tilde{\sigma}_{f_k}^2(\tilde{x}_i)} + \left( \tilde{\mu}_{f_k}(\tilde{x}_i) \right)^2 \right] \right\}
\] (34)

\[
\tilde{\varphi}_{f_k}(\ell) = 2\tilde{m}_k(\ell) - \tilde{f}_i(\ell) + \tilde{\varphi}_{f_k}(\ell) \sum_{\tau=1}^{t} w_{\tau,i} \left( \tilde{s}_\tau - \tilde{\varphi}_{\nu}(\ell) \right) \quad \forall k
\] (35)

where the \( \tilde{\psi}(\cdot) \) is the digamma function, \( \tilde{f}_i(\ell) := \sum_{k=1}^{K} \tilde{z}_k(\tilde{x}_i) \tilde{\varphi}_{f_k}(\ell) \) \( \forall i, \) \( \tilde{s}_\tau := \sum_{i=1}^{N_g} w_{\tau,i} \tilde{f}_i(\ell) \) \( \forall \tau, \)

\[
\tilde{\varphi}_{\nu}(\ell) := \mathbb{E}_{q^{(\varphi_{\nu})}(r_{\nu})} [ \varphi_{\nu}] = \tilde{a}_\nu \tilde{b}_\nu^{-1}
\] (36)

\[
\tilde{\varphi}_{f_k}(\ell) := \mathbb{E}_{q^{(\varphi_{f_k})}(r_{f_k})} [ \varphi_{f_k}] = \tilde{a}_k^{(f_k)} \tilde{b}_k^{-1}
\] (37)

\[
\tilde{m}_k(\ell) = \tilde{\sigma}_{f_k}^2(\tilde{x}_i) \left( \tilde{\mu}_{f_k}(\tilde{x}_i) \right) \tilde{\varphi}_{f_k}(\ell) - \tilde{\varphi}_{f_k}(\ell) \sum_{\tau=1}^{t} w_{\tau,i} \left( \tilde{s}_\tau - \tilde{\varphi}_{\nu}(\ell) \right) \quad \forall k
\] (38)

\[
\tilde{a}_k(\ell) = a_k + \sum_{i=1}^{N_g} \tilde{z}_k(\tilde{x}_i) \tilde{\mu}_{f_k}(\tilde{x}_i) \] (39)

and the (approximate) MMSE estimator of \( f \) as

\[
\hat{f}_{\text{MMSE}, i} := \mathbb{E}_{q^{(f_i)}(r_{f_i})} [ f_i ] = \tilde{\mu}_{f_{\text{MAP}, i}}(\tilde{x}_i) \quad \forall i
\] (42)

while \( \theta \) is estimated using the marginal MMSE estimators

\[
\tilde{\varphi}_{\text{MMSE}} := \mathbb{E}_{q^{(\varphi_{\nu})}} [ \varphi_{\nu}] = \tilde{a}_\nu \tilde{b}_\nu^{-1}
\] (43)

\[
\tilde{\mu}_{f_k, \text{MMSE}} := \mathbb{E}_{q^{(\varphi_{f_k})}} [ \varphi_{f_k}] = \tilde{m}_k \forall k
\] (44)

\[
\tilde{\varphi}_{f_k, \text{MMSE}} := \mathbb{E}_{q^{(\varphi_{f_k})}} [ \varphi_{f_k}] = \tilde{a}_k^{(f_k)} \tilde{b}_k^{-1} \quad \forall k.
\] (45)

The VB algorithm to obtain \( \{ \hat{f}_{\text{MMSE}, i} \}_{i=1}^{N_g}, \{ \tilde{\varphi}_{\nu} \}_{i=1}^{N_g}, \tilde{\varphi}_{\text{MMSE}}, \tilde{\mu}_{f_{\text{MAP}, i}}, \theta_{\text{MMSE}}, \) and \( q^{(f, \theta, z)} \) is tabulated in Alg. 1.

**Remark 1.** (Assessing convergence.) The steps of Alg. 1 guarantee that the ELBO monotonically increases across iterations \( \ell \) [3]. 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 \( \ell \) if \( \text{ELBO}(q^{(\ell)}, f, z, \theta) < \xi \).

**Remark 2. (Computational complexity.)** For Alg. 1, the complexity order to update \( q(f_i | z_i = k) \) \( \forall i, k \) per iteration \( \ell \) is \( O((2K + 1)NK) \) to compute \( \tilde{\mu}_{f_k}(\tilde{x}_i) \) in (33), while updating \( \tilde{z}_k(\tilde{x}_i) \) \( \forall i, k \) via (28) incurs complexity \( O(N_g K) \). In addition, updating \( q(\theta) \) has complexity \( O((2K + 1)NK) \) that is dominated by the computation of \( \tilde{b}_\nu \) in (36). Overall, the per-iteration complexity of Alg. 1 is \( O((2K + 1)NK) \).

Note that a sample-based counterpart of Alg. 1 via MCMC in [26] incurs complexity in the order of \( O(N_g^3) \). For conventional methods to estimate \( f \), the ridge regularized LS [16] has a one-shot (non-iterative) complexity of \( O(N_g^3) \), while the total variation (TV) regularized LS via the alternating direction method of multipliers (ADMM) in [35] incurs complexity of \( O(N_g^3) \) per iteration \( \ell \); see also [28], [36] for details. This means that Alg. 1 incurs the lowest per-iteration complexity, which becomes more critical as \( N_g \) increases.

**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 \( f \). Since the proposed Bayesian framework accounts for the uncertainty through \( \tilde{\sigma}_{f_k}^2(\tilde{x}_i) \) in (38), we adopt the conditional entropy [7] to serve as an uncertainty measure of \( f \) at time slot \( \tau \), namely

\[
H(f | z, \tilde{s}_\tau; \hat{\theta}_\tau) := \sum_{z' \in Z} \int p(z', \tilde{s}_\tau'; \hat{\theta}_\tau) \times H(f | z = z', \tilde{s}_\tau = \tilde{s}_\tau'; \hat{\theta}_\tau) dz',
\] (46)

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

\[
H(f | z = z', \tilde{s}_\tau = \tilde{s}_\tau'; \hat{\theta}_\tau) := - \int p(f | z = z', \tilde{s}_\tau = \tilde{s}_\tau'; \hat{\theta}_\tau) \ln p(f | z = z', \tilde{s}_\tau = \tilde{s}_\tau'; \hat{\theta}_\tau) df
\] (47)
Algorithm 1 Field estimation via variational Bayes

Input: $\mathbf{s}_t$, $\mathbf{W}_t$, \left\{a_v, b_v, \{m_k, \sigma_k^2, a_k, b_k\}_{k=1}^K\right\}$, and $N_{\text{iter}}$

1. Initialize $q^{(0)}(f, z, \theta)$ and set $\ell = 0$
2. Obtain $\hat{\theta}_\ell$ with (35)
3. while ELBO has not converged and $\ell \leq N_{\text{iter}}$ do
   4. Set $\ell \leftarrow \ell + 1$
   5. Obtain $\hat{\theta}^{(\ell)}(\tilde{x}_i)$ i.e., $k$ via (32)
   6. Obtain $\hat{\theta}^{(\ell)}(\tilde{x}_i)$ i.e., $k$ via (33)
   7. Obtain $q^{(\ell)}(z_k = k)$ i.e., $k$ via (28)
   8. Obtain $\hat{\theta}^{(\ell)}(\tilde{x}_i)$ via (36)
   9. Obtain $\hat{\theta}^{(\ell)}(\tilde{x}_i)$ via (37)
  10. Obtain $\hat{\theta}^{(\ell)}(\tilde{x}_i)$ via (38)
  11. Obtain $\hat{\theta}^{(\ell)}(\tilde{x}_i)$ via (39)
  12. Obtain $\hat{\theta}^{(\ell)}(\tilde{x}_i)$ via (40)
4. end while
13. Set $q^{*}(f) = q^{(\ell)}(f)$ and $q^{*}(z) = q^{(\ell)}(z)$ i.e., $i$
14. Set $q^{*}(\theta) = q^{(\ell)}(\theta)$
15. Estimate $\hat{\theta}_{\text{MAP},i} = \arg\max_{z_i \in \{1, \ldots, K\}} q^{*}(z_i)$ i.e., $i$
16. Estimate $\hat{f}_{\text{MMSE}} = \hat{f}_{\text{MAP},(\tilde{x}_i)}$ i.e., $i$
17. Estimate $\hat{\theta}_{\text{MMSE}} = \mathbb{E}_{\hat{\phi}}[q^{*}(\theta)]$ via (43)–(45)
18. return $\hat{f}_{\text{MMSE}}, \hat{\theta}_{\text{MMSE}}, q^{*}(f|z), q^{*}(z),$ and $q^{*}(\theta)$

since $p(f|z, \hat{s}_t; \hat{\theta}_\ell)$ is Gaussian with covariance matrix

$$
\Sigma_{f|z, \hat{s}_t; \hat{\theta}_\ell} : = \left( \tilde{\Sigma}_f \mathbf{W}_t \mathbf{W}_t^\top + \Phi_{f|z} \right)^{-1}
$$

with $\Phi_{f|z} := \text{diag}\left(\left\{\tilde{\Sigma}_f\right\}_{i=1}^{N_y}\right)$ [26]. Then, using the matrix determinant identity lemma [18, Chap. 18], it is not hard to show that

$$
H(f|z, \hat{s}_{t+1}; \hat{\theta}_\ell) = H(f|z, \hat{s}_t; \hat{\theta}_\ell) - \frac{1}{2} \sum_{z' \in \tilde{Z}} \int p(z', \hat{s}_t; \hat{\theta}_\ell) \log \left(1 + \tilde{\Sigma}_f \mathbf{W}_t \mathbf{W}_t^\top \right)^{-1} p(z' | \hat{s}_t; \hat{\theta}_\ell) dz'.
$$

To obtain $\hat{s}_{t+1}$, we choose a pair of sensors $(n^*, n'^*)$, or equivalently find $\mathbf{w}_{t+1}^{(n^*, n'^*)}$ minimizing $H(f|z, \hat{s}_{t+1}; \hat{\theta}_\ell)$. Given $\hat{s}_t$, we then find $\mathbf{w}_{t+1}^{(n^*, n'^*)}$ by solving

$$
\max_{(n^*, n'^*)} \mathbb{E}_{p(z|\hat{s}_t; \hat{\theta}_\ell)} \left[ h(z, \hat{s}_t, \mathbf{w}_{t+1}^{(n^*, n'^*)}; \hat{\theta}_\ell) \right]
$$

where $h(z, \hat{s}_t, \mathbf{w}_{t+1}; \hat{\theta}_\ell) := \ln \left(1 + \tilde{\Sigma}_f \mathbf{W}_t \mathbf{W}_t^\top \right)^{-1} \mathbf{w}_{t+1}^{(n^*, n'^*)}$ and $\tilde{Z}$ denotes the set of available sensing radio pairs at slot $\tau$.

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

$$
p(f|z, \hat{s}_t, \theta) = \frac{p(f, z, \theta|\hat{s}_t)}{p(z, \theta)} \approx \frac{q(f, z, \theta)}{q(z, \theta)} = q(f|z),
$$

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

$$
H(f|z) \approx \frac{1}{2} \ln \left| \Sigma_{f|z, \hat{s}_t; \hat{\theta}_\ell} \right| + \frac{N_y}{2} \left(1 + \ln 2\pi\right)
$$

with $\Sigma_{f|z, \hat{s}_t; \hat{\theta}_\ell} := \text{diag}\left(\left\{\tilde{\Sigma}_f\right\}_{i=1}^{N_y}\right)$; and subsequently, that of $H(f|z, \hat{s}_t; \hat{\theta}_\ell)$ by substituting (51) into (46).

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

$$
H(f|z, \hat{s}_t; \hat{\theta}_\ell) \approx H(f|z, \hat{s}_t; \hat{\theta}_\ell) - \frac{1}{2} \sum_{z' \in \tilde{Z}} \int p(z', \hat{s}_t; \hat{\theta}_\ell) \ln \left| \Sigma_{f|z', \hat{s}_t; \hat{\theta}_\ell} \right| + \frac{N_y}{2} \left(1 + \ln 2\pi\right)
$$

where $\Delta_{w_{t+1}} := \text{diag}\left(\left\{\tilde{\Sigma}_f\right\}_{i=1}^{N_y}\right)$, and $\mathbf{w}_{t+1}$ is defined in (52) as (cf. (52))

$$
\mathbf{w}_{t+1}^{(n^*, n'^*)} = \max_{(n, n') \in \mathcal{M}_{t+1}} \frac{1}{N_y} \sum_{i=1}^{N_y} \mathbb{E}_{q(z_i)} \left[ \ln \left(1 + \tilde{\Sigma}_f \mathbf{W}_t \mathbf{W}_t^\top \right)^{-1} \mathbf{w}_{t+1}^{(n^*, n'^*)} \right].
$$

Solving (P2) using a greedy search, we obtain the pair of sensors $(n^*, n'^*)$ associated with $\mathbf{w}_{t+1}^{(n^*, n'^*)}$, based on which we collect the informative measurement $\hat{s}_{t+1}$.

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 $N_{\text{batch}}$ per time slot $\tau$ as follows: i) find weight vectors $\left\{\mathbf{w}_{t+1}^{(m,n,m')}\right\}_{m=1}^{N_{\text{batch}}}$ for $\left\{(n^m, n^{m'})\right\}_{m=1}^{N_{\text{batch}}} \subset \mathcal{M}_{t+1}$ associated with $N_{\text{batch}}$ largest values of $h(\mathbf{w}_{t+1}^{(n^m, n^{m'})})$ in (P2*), and collect $\left\{\mathbf{s}_{t+1}^{(m,n)}\right\}_{m=1}^{N_{\text{batch}}}$ from pairs of sensors revealed from those weight vectors (steps 4–5 in Alg. 2); and ii) aggregate those measurements below $\hat{s}_t$ to construct $\hat{s}_{t+1} := \left[\hat{s}_t^T, \hat{s}_{t+1}^T, \ldots, \hat{s}_{t+1}^{(N_{\text{batch}})}\right]^T$ (step 6 in Alg. 2). Numerical tests are presented next to assess the mini-batch operation of Alg. 2.
conditioned on the labels in $F$ through numerical tests using MATLAB datasets. Comparisons were carried out with existing methods, including the ridge-regularized SLF estimate given by $f_{LS} = (W_i W_i^\top + \rho_f C_f^{-1})^{-1} W_i \hat{s}_i$ [16], where $C_f$ is a spatial covariance matrix modeling the similarity between points $\hat{x}_i$ and $\tilde{x}_i$ in area $A$. We further tested the TV-regularized LS scheme in [35], which solves the problem in (6) with
\begin{equation}
\mathcal{R}(f) = \sum_{i=1}^{N_h} \sum_{j=1}^{N_v} |F_{i+1,j} - F_{i,j}| + \sum_{i=1}^{N_h} \sum_{j=1}^{N_v-1} |F_{i,j+1} - F_{i,j}|
\end{equation}
(53)
where $F := \text{unvec}(f) \in \mathbb{R}^{N_h \times N_v}$ and $F_{i,j} := [F]_{i,j}$. We also tested an MCMC-based counterpart of Alg. 2 for estimating the posterior in (14), and solving (P2) in (49); see e.g., [26], [34] for details.

We further compared the proposed data-adaptive selection with random sampling and the ridge-regularized SLF estimators, by selecting $\{\{n(m), n(t(m))\}_{m=1}^{N_{\text{data}}}\}$ uniformly at random to collect $\{\hat{s}(m)\}_{m=1}^{N_{\text{data}}} \forall \tau$. 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 $A := [0.5, 60.5] \times [0.5, 60.5]$. Using these measurements, the SLF was reconstructed over the grid $\{x_i\}_{i=1}^{3600} := \{1, \ldots, 60\}^2$. To generate the ground-truth SLF $f_0$, the ground-truth label field $z_0$ was generated via Gibbs sampling [12] by using the Potts prior of $z$ in (8) with $\beta = 1.5$ and $K = 4$. Given $\theta_f := [\mu_f, \varphi_f]^{\top}$ with $\mu_f = [0, 1, 2.5, 5.5]^\top$ and $\varphi_f = [10, 10, 2, 2]^\top$, vector $f_0$ was constructed to have $f(x_i) \sim \mathcal{N}(\mu_f, \varphi_f^{-1}) \forall x_i \in A_k$, $\forall k$ conditioned on the labels in $z_0$. The resulting hidden label field $Z_0 := \text{unvec}(z_0) \in \{1, 2, 3, 4\}^{60 \times 60}$, and the true SLF $F_0 := \text{unvec}(f_0) \in \mathbb{R}^{60 \times 60}$ are depicted in Fig. 4 with sensor locations marked by crosses. The effects of calibration are not accounted for, meaning that $g_0$ and $\gamma$ are assumed to be known, and the fusion center directly uses shadowing measurements $\hat{s}_\tau$. Under the mini-batch operation, each measurement $\hat{s}(m) \forall m$ was generated according to (5), where $s_\tau$ was obtained using (4) with $w$ set to the normalized ellipse model in (3) with $\Lambda = 0.39$, while $\nu$ was set to follow a zero-mean Gaussian pdf with $\varphi_\nu = 20$. To construct $M_{\tau+1}$ per time slot $\tau$, $|M_{\tau+1}| = 200$ pairs of sensors were uniformly selected at random with replacement. Then, $N_{\text{Batch}} = 100$ shadowing measurements were collected at $\{\{n(m), n(t(m))\}_{m=1}^{N_{\text{data}}}\} \subset M_{\tau+1}$ to run Alg. 2 for $\tau = 0, 1, \ldots, 8$.

In all synthetic tests, the simulation parameters were set to $N_{\text{data}} = 3,000$ and $\xi = 10^{-6}$; hyper-hyper parameters of $\nu_t$ were set to $a_\nu = 1, 300$ and $b_\nu = 2$; and those of $\theta_f$ were set as listed in Table I. To execute Alg. 1, variational parameters of $q_f(\cdot; z, \theta)$ were initialized as follows: $\{\hat{\mu}_{f_k}(\tilde{x}_i)\}_{i=1}^{N_q}$, $\hat{b}_q$, $\hat{\beta}_q$, $\hat{\gamma}_k^0$, $\tau^0$, $\pi_k^0$, and $\tilde{\gamma}_k$ were drawn from the uniform distribution $\mathcal{U}(0, 1)$, while $m_k^0 = m_k \forall k$; and it was set to $\theta_k^0(\tilde{x}_i) = 1/4 \forall i, k$. Furthermore, $\hat{\theta}(0)$ was collected from 800 pairs of sensors selected at random, which determined $W(0)$. To find $\rho_f$ of the competing alternatives, the L-curve [25, Chapter 26] was used for the ridge regularization, while the generalized cross-validation [14] was adopted for the TV regularization. The hyper-hyper parameters of $\theta$ 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 $F \hat{=} \text{unvec}(f)$ and the associated hidden label fields $\hat{Z} := \text{unvec}(\hat{z})$ 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 hidden field, denoted as $\hat{F}_\text{full}$ and $\hat{Z}_\text{full}$, 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, $\hat{F}$ and $\hat{Z}$ found using the non-adaptive VB algorithm are depicted in Figs. 5c and 5f. Visual comparison of Figs. 5c and 5e reveals that the reconstruction performance for $\hat{F}$ can be improved with the same number of measurements by adaptively selecting pairs of sensors. Accuracy of $\hat{z}$ was also quantitatively measured by the labeling-error, defined using the entrywise Kronecker delta $\delta(\cdot)$, as $\|\delta(z_0 - \hat{z})\|_{1/N_{\text{data}}}$. The proportion 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 $f$ given a current estimate of $\theta$ improve accuracy of $\hat{f}$ and $\hat{z}$ in the
TABLE I: Hyper-parameters of $\theta_f$ for synthetic data tests.

| $m_1$ | $m_2$ | $m_3$ | $m_4$ | $\alpha^2_0$ | $\alpha^2_1$ | $\alpha^2_2$ | $\alpha^2_3$ |
|-------|-------|-------|-------|-------------|-------------|-------------|-------------|
| 0.8   | 0.8   | 0.8   | 0.8   | 1           | 1           | 1           | 1           |

TABLE II: True $\theta$ 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.

| $\theta$  | True       | Est. (Alg. 2) | Est. (non-adaptive) |
|-----------|------------|---------------|---------------------|
| $\varphi_{\nu}$ | 20  | 18.329 ± 6 × 10^{-3} | 18.461 ± 4.6 × 10^{-3} |
| $\mu_{f_1}$  | 0           | 0.022 ± 1.2 × 10^{-2} | 0.018 ± 1.9 × 10^{-2} |
| $\mu_{f_2}$  | 1           | 0.957 ± 1.7 × 10^{-2} | 0.962 ± 1.6 × 10^{-2} |
| $\mu_{f_3}$  | 2.5         | 2.573 ± 1.7 × 10^{-2} | 2.578 ± 2.6 × 10^{-2} |
| $\mu_{f_4}$  | 5.5         | 5.399 ± 2.7 × 10^{-2} | 5.374 ± 7.9 × 10^{-3} |
| $\nu_{f_1}$  | 10          | 40.178 ± 3 × 10^{-1}  | 42.352 ± 2 × 10^{-1}  |
| $\nu_{f_2}$  | 10          | 11.634 ± 1.4 × 10^{-2} | 15.845 ± 1.2 × 10^{-1} |
| $\nu_{f_3}$  | 2           | 7.712 ± 2.7 × 10^{-2} | 7.493 ± 2.2 × 10^{-2} |
| $\nu_{f_4}$  | 2           | 4.620 ± 6.1 × 10^{-2} | 5.451 ± 4 × 10^{-2} |

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 $\nu_{\tau}$. We adopted the labeling-error for $z$ averaged over sensor locations and realizations of $\{\nu_{\tau}\}_{\tau}$ to quantify the reconstruction performance. Fig. 6b shows the progression of the labeling error at $\tau = 8$ as a function of the noise precision $\varphi_{\nu}$ 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 $\varphi_{\nu}$ decreases, or equivalently $\sigma_{\nu}^2$ increases. This confirms that the proposed algorithm is reasonably robust against measurement noise.

Averaged estimates of $\theta$ 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 $A$ by correctly inferring the underlying statistical properties of each piecewise homogeneous region in the SLF. Note that $\varphi_{\nu}$ 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 $p(f|z, \theta_f)p(\theta_f)$ is small by setting each $\varphi_{\nu}$ to a large value $\forall \nu$, which corroborates the result in [5, p. 468].

Next, we will validate the efficacy of the proposed algorithms for channel-gain cartography using the setup of Fig. 5. From the estimate $\hat{f}_{\text{MMSE}}$ obtained through Alg. 2, we found the shadowing attenuation $\hat{s}(x, x')$ between two arbitrary points $x$ and $x'$ in $A$ using (4) after replacing $f$ with $\hat{f}_{\text{MMSE}}$. Subsequently, we obtained the estimated channel-gain $\hat{g}(x, x')$ after substituting $\hat{s}(x, x')$ into (1).

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

$$\text{NMSE} := \frac{\mathbb{E} \left[ \int_A \left( s(x, x') - \hat{s}(x, x') \right)^2 dxdx' \right]}{\mathbb{E} \left[ \int_A s^2(x, x') dxdx' \right]}$$

where the expectation is over the set $\{x_n\}_{n=1}^N$ of sensor locations and realizations of $\{\nu_{\tau}\}_{\tau}$. The integrals are approximated by averaging the integrand over 500 pairs of $(x, x')$ chosen independently and uniformly at random on the boundary of $A$. 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 [16]. The test setup is depicted in Fig. 8, where \(A = [0.5, 20.5] \times [0.5, 20.5]\) is a square with sides of 20 feet (ft), over which a grid \(\{\hat{x}_i\}_{i=1}^{3721} = \{1, \ldots, 61\}^2\) 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. 8. To estimate \(g_0\) and \(\gamma\) using the approach in [16], 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. 8, and \(T = 2,380\) measurements \(\{\tilde{g}_\tau\}_{\tau=1}^{T}\) were collected. Calibrated measurements \(\{\hat{s}_\tau\}_{\tau=1}^{T}\) were then obtained from \(\{\tilde{g}_\tau\}_{\tau=1}^{T}\) after substituting \(\hat{g}_0\) and \(\hat{\gamma}\) into (5). The weights \(\{\hat{w}_{\tau,n}\}_{\tau=1}^{T}\) were constructed with \(w\) in (3) by using known locations of sensor pairs. Note that \(\tau\) is introduced to distinguish indices of the real data from \(\tau\) used to index time slots in numerical tests.

We randomly selected 1,380 measurements from \(\{\hat{s}_\tau\}_{\tau=1}^{T}\) to initialize \(\hat{s}^{(0)}\) and \(\hat{W}^{(0)}\), and used the remaining 1,000 measurements to run the proposed algorithm under the minibatch operation for \(\tau = 0, 1, \ldots, 5\). At each time slot \(\tau\), \(M_{\tau+1}\) was formed by sensors corresponding to \(|M_{\tau+1}| = 200\) weight vectors uniformly selected at random from \(\{\hat{w}_{\tau,n}\}_{\tau=1}^{T}\), associated with the remaining 1,000 measurements without replacement. Then, \(N_{\text{batch}} = 100\) measurements were chosen from \(\{\hat{s}_\tau\}_{\tau=1}^{T}\) associated with \(M_{\tau+1}\).

Simulation parameters were set to \(N_{\text{iter}} = 3,000\), \(\xi = 10^{-6}\), and \(K = 3\); and hyper-hyper parameters of \(\theta\) were set to \(a_\nu = b_\nu = 10^{-3}\), \([m_1, m_2, m_3]^{\top} = [0.035, 0.05, 0.5]^{\top}\), \(\sigma_k^2 = 10^{-4}\) \(\forall k\), and \(a_k = b_k = 0.1\) \(\forall k\), respectively. To execute Alg. 1, variational parameters of \(q^{(0)}(f, z, \theta)\) were initialized as follows: \(\{\bar{b}_k^{(0)}(\hat{x}_i)\}_{k=1}^{N_g}\), \(\{\bar{b}_k^{(0)}(\hat{x}_i)\}_{k=1}^{N_g}\), and \(\{\bar{b}_k^{(0)}(\hat{x}_i)\}_{k=1}^{N_g}\) were drawn from the uniform distribution \(\mathcal{U}(0,1)\), while \(\tilde{m}_k = m_k \forall k\) and \(\tilde{c}_k(\hat{x}_i) = 1/3 \forall i, k\).

Following [1], [16], a spatial covariance matrix was used for \(C_f\) of the ridge-regularized LS estimator, which models the similarity between points \(\hat{x}_i\) and \(\hat{x}_j\) as \(C_f = \rho_f^2 \exp[-||\hat{x}_i - \hat{x}_j||^2/\nu]\) [1] with \(\sigma^2 = \kappa = 1\), and \(\rho_f = 0.015\) found with the L-curve [25, Chapter 26]. For the TV-regularized LS estimator, it was set to \(\rho_f = 6\) found through the generalized cross validation [14]. To assess the efficacy of our Bayesian model with the \(K\)-ary hidden label field, we tested the adaptive MCMC method in [26] with \(K = 2\).

Figs. 9a–9h depict SLF estimates \(\hat{F}\) and associated hidden fields \(\hat{Z}\) at \(\tau = 5\) obtained via the proposed algorithms and competing alternatives. As a benchmark, one-shot estimates of the SLF \(\hat{F}_{\text{full}}\) and associated hidden field \(\hat{Z}_{\text{full}}\) 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. 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 \(\hat{F}\) and \(\hat{Z}\) 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 \(\hat{z}\) measured by the labeling error \(|\delta(\hat{z}_{\text{full}} - \hat{z})|/N_g\) with \(\hat{z}_{\text{full}} := \text{vec}(\hat{Z}_{\text{full}})\) 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 \(\tau\). 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 \(\theta\) averaged over 20 MC runs are listed in Table III. Estimated \(\theta\) obtained by using the full data was considered as a benchmark, to demonstrate that Alg. 2 yields estimates \(\theta\) closer to the benchmark than its non-adaptive counterpart (except \(\varphi_{\nu}\)). Note that the level of measurement noise is high since \(\sigma^2_{\nu} = \varphi_{\nu}^{-1} \approx 15\). This can be justified because the testbed structure was accurately revealed in \(\hat{F}\) and \(\hat{Z}\) 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 tests of this section. A channel-gain map is constructed to portray the channel-gain between every point in the map \(x\), and a fixed receiver location \(x_{\text{rx}}\). Specifically, Alg. 2 is executed.
and estimates \( \{\tilde{s}(\mathbf{x}_i, \mathbf{x}_r)\}^{N_y}_{i=1} \) are obtained by substituting \( \tilde{f} \) and \( \tilde{w} \) into (4). Subsequently, \( \{\tilde{g}(\mathbf{x}_i, \mathbf{x}_r)\}^{N_y}_{i=1} \) are obtained by substituting \( \{\tilde{s}(\mathbf{x}_i, \mathbf{x}_r)\}^{N_y}_{i=1} \) into (1) with \( \tilde{g}_0 \) and \( \tilde{c}_t \). Upon defining \( \tilde{g} := [\tilde{g}(\mathbf{x}_1, \mathbf{x}_r), \ldots, \tilde{g}(\mathbf{x}_{N_y}, \mathbf{x}_r)]^\top \in \mathbb{R}^{N_y} \), we construct the channel-gain map \( \mathbf{G} := \text{unvec}(\tilde{g}) \) with the receiver located at \( \mathbf{x}_r \).

Let \( \mathbf{S} := \text{unvec}(\tilde{s}) \) denote a shadowing map with \( \tilde{s} := [\tilde{s}(\mathbf{x}_1, \mathbf{x}_r), \ldots, \tilde{s}(\mathbf{x}_{N_y}, \mathbf{x}_r)]^\top \in \mathbb{R}^{N_y} \). 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 \( \mathbf{x}_r = (10.3, 10.7) \) (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 \( \mathbf{G} \)), or the concrete wall (left side of \( \mathbf{G} \)). 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 \( \mathbf{G} \)); 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 \( \mathbf{F} \) 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

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

| \( \theta \) | Benchmark | Est. (Alg. 2) | Est. (non-adaptive) |
|-----------|-----------|---------------|---------------------|
| \( \varphi_1 \) | 0.075 \( \pm 10^{-1} \) | 0.068 \pm 0.13 | 0.071 \pm 0.24 |
| \( \mu_1 \) | 0.001 \( \pm 10^{-1} \) | -0.001 \( \pm 10^{-3} \) | -0.001 \( \pm 10^{-3} \) |
| \( \mu_2 \) | 0.062 \( \pm 10^{-1} \) | 0.032 \( \pm 10^{-8} \) | 0.063 \( \pm 10^{-8} \) |
| \( \mu_3 \) | 0.045 \( \pm 10^{-1} \) | 0.046 \( \pm 10^{-8} \) | 0.046 \( \pm 10^{-8} \) |
| \( \varphi_2 \) | 5.524 \( \pm 10^{-18} \) | 4.951 \( \pm 10^{-3} \) | 4.789 \( \pm 1.9 \times 10^{-3} \) |
| \( \varphi_3 \) | 5.524 \( \pm 10^{-18} \) | 4.942 \( \pm 10^{-3} \) | 4.782 \( \pm 1.7 \times 10^{-3} \) |

Fig. 10: Progression of a mismatch between \( \tilde{z} \) and \( \tilde{z}_{\text{full}} \).

Fig. 11: Estimated shadowing maps \( \mathbf{S} \) and corresponding channel-gain maps \( \mathbf{G} \) 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 [26] 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 \( \mathbf{x}_r = (10.3, 10.7) \) (ft) marked with the black cross.

field of the tomographic model at affordable complexity by using measurements collected from sensing radio pairs that 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 \( \ell \) will be omitted for simplicity.

A. Variational distribution of the SLF \( q(f|z) \)

Recall that the conditional posterior obeys \( p(f, z, \theta|s_t) \propto p(s_t, f, \varphi_n)p(f|z, \theta_f) \). The first factor in (22), is expressed as

\[
\ln q(f|z) = \sum_{i=1}^{N_g} \ln q(f_i|z_i) = \sum_{k=1}^{K} \sum_{i : i \in A_k} \ln q(f_i|z_i = k)
\]

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

\[
\ln q(f_i|z_i = k) = E_{-q(f_i|z_i = k)}[\ln p(f_i|z, \theta_f)] \]

(55)

where \( \bar{f}_j := \sum_{k=1}^{K} z_k(\bar{x}_j)\tilde{m}_k f_k(\bar{x}_j) \)

and

\[
E_{-q(f_i|z_i = k)}[\ln p(f|z, \theta_f)] = \frac{-\nabla_{f_i}}{2} (f_i^2 - 2\mu_k f_i).
\]

(56)

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

\[
E_{-q(f_i|z_i = k)}[\ln p(\bar{s}_i|f, \varphi_n)]
\]

\[
\exp \left\{ -\frac{1}{2} \left( \frac{\bar{\varphi}_\nu}{2} \sum_{\tau=1}^{t} \left[ \bar{s}_\tau - \sum_{j \neq i} w_{\tau,j} \bar{f}_j \right] w_{\tau,i} f_i \right) \right\}
\]

(57)

By completing the square, it can be readily verified that

\[
q(f_i|z_i = k) \propto \mathcal{N}(\tilde{\mu}_f(k), \tilde{\sigma}_f^2(k))
\]

\[
\tilde{\mu}_f(k) = \frac{1}{\tilde{\varphi}_\nu} \sum_{\tau=1}^{t} \left[ \bar{s}_\tau - \sum_{j \neq i} w_{\tau,j} \bar{f}_j \right] w_{\tau,i} f_i
\]

(60)

\[
\tilde{\sigma}_f^2(k) = \left( \frac{1}{\tilde{\varphi}_\nu} \sum_{\tau=1}^{t} w_{\tau,i}^2 + \frac{\bar{\varphi}_\nu}{2} \right)\left( \frac{1}{\tilde{\varphi}_\nu} \sum_{\tau=1}^{t} \left[ \bar{s}_\tau - \sum_{j \neq i} w_{\tau,j} \bar{f}_j \right] w_{\tau,i} \right)^{-1}
\]

(61)

Upon defining \( \bar{s}_\tau := \sum_{i=1}^{N_g} w_{\tau,i} f_i \), it follows that

\[
\tilde{\mu}_f(k) = \bar{f}_i + \frac{\tilde{\varphi}_\nu}{2} \sum_{\tau=1}^{t} \left[ \bar{s}_\tau - \bar{x}_i \right] w_{\tau,i}
\]

(62)

B. Variational distribution of the hidden label field \( q(z) \)

Since \( q(z) = \prod_{i=1}^{N_g} q(z_i) \) 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 \( p(f, z, \theta|s_t) \propto p(f|z, \theta_f)p(z; \beta) \) wrt \( z \), after singling out the terms that involve \( z_i \), we arrive at

\[
\ln q(z_i) = E_{-q(z_i)}[\ln p(f, z, \theta|s_t)] + c
\]

(63)

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

\[
E_{-q(z_i)}[\ln p(f|z, \theta_f)] = \frac{1}{2} E_{-q(z_i)}[\ln \varphi_{f_k} - \varphi_{f_k}(f_k - \mu_{f_k})^2] \]

\[
= \frac{1}{2} \left( \frac{\tilde{\varphi}_f}{\varphi_n} \left[ E_{-q(z_i)}[f_i^2] - 2\tilde{\mu}_f(k) f_i \right] \right) = \frac{1}{2} \left( \frac{\tilde{\varphi}_f}{\varphi_n} \left[ E_{-q(z_i)}[f_i^2] - 2\tilde{\mu}_f(k) f_i \right] \right)
\]

(64)

(65)

and

\[
E_{-q(z_i)}[\ln p(z|k, \beta)] = \beta \sum_{j \in N(z_i)} \delta(z_j - k)
\]

(66)

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

C. Variational distribution of the noise precision \( q(\varphi_n) \)

As the conditional posterior \( p(f, z, \theta|s_t) \) is proportional to \( p(s_t|f, \varphi_n)p(\varphi_n)p(\varphi_n) \), we can write

\[
\ln q(\varphi_n) = E_{-q(\varphi_n)}[\ln p(f, z, \theta|s_t)] + c
\]

(67)

where

\[
E_{-q(\varphi_n)}[\ln p(s_i|f, \varphi_n)] = \frac{t}{2} \ln \varphi_n - \frac{\tilde{\varphi}_\nu}{2} \|s_i - W_f^\top f\|^2
\]

(68)

\[
= \frac{t}{2} \ln \varphi_n - \frac{\tilde{\varphi}_\nu}{2} \sum_{\tau=1}^{t} \bar{s}_\tau^2 - 2\bar{s}_\tau \bar{x}_\tau + E_{-q(\varphi_n)} \left[ \left( W_{\tau}^{(n,n')}^\top f \right)^2 \right]
\]

and

\[
E_{-q(\varphi_n)}[\ln p(\varphi_n)] \leftrightarrow (a_\nu - 1) \ln \varphi_n - \frac{\tilde{\varphi}_\nu}{b_\nu}
\]

(69)

After substituting (68) and (69) into (67), we can easily see that

\[
q(\varphi_n) = \mathcal{G}(\tilde{\alpha}_\nu, \tilde{\beta}_\nu) \]

\[
= \frac{1}{b_\nu} + \frac{1}{2} \frac{\tilde{\varphi}_\nu}{b_\nu} \sum_{\tau=1}^{t} \bar{s}_\tau^2 - 2\bar{s}_\tau \bar{x}_\tau + E_{-q(\varphi_n)} \left[ \left( W_{\tau}^{(n,n')}^\top f \right)^2 \right]
\]

(70)
where
\[
E_{q(\varphi)} \left[ (w_{\tau}^{(n,n')^T} f)^2 \right] \\
= \text{Var} \left[ w_{\tau}^{(n,n')^T} f \right] + \left( E_{q(\varphi)} \left[ w_{\tau}^{(n,n')^T} f \right] \right)^2
\]
(71)
\[
= \sum_{i=1}^{N_q} w_{\tau,i}^2 \left[ \sum_{k=1}^{K} \zeta_k(\tilde{x}_i) \left( \sigma_{f_k}^2(\tilde{x}_i) + \mu_{f_k}^2(\tilde{x}_i) \right) - f_i^2 \right] + \sigma_r^2
\]
(72)
by the law of total variance on \( \text{Var} \left[ w_{\tau}^{(n,n')^T} f \right] \) [6, p. 401].

D. Variational distribution of the field means \( q(\mu_f) \)

Since the conditional posterior \( p(f, z, \theta|\tilde{s}_i) \) is proportional to \( p(f|z, \theta_f)p(\mu_f) \) w.r.t. \( \mu_f \), the entries of \( \mu_f \) are iid, we have
\[
\ln q(\mu_f) = E_{q(\mu_f)} [\ln p(f, z, \theta|\tilde{s}_i)] + c
\]
\[
\leftrightarrow E_{q(\mu_f)} [\ln p(f|z, \theta_f)] + E_{q(\mu_f)} \left[ \sum_{k=1}^{K} \ln p(\mu_{f_k}) \right]
\]
(73)
where
\[
E_{q(\mu_f)} [\ln p(f|z, \theta_f)]
\]
\[
\leftrightarrow \sum_{k=1}^{K} \sum_{i=1}^{N_q} \zeta_k(\tilde{x}_i) \tilde{\varphi}_{f_k} \left( \sigma_{f_k}^2(\tilde{x}_i) - 2\tilde{\mu}_{f_k}(\tilde{x}_i) \mu_{f_k} \right)
\]
(74)
and
\[
E_{q(\mu_f)} \left[ \sum_{k=1}^{K} \ln p(\mu_{f_k}) \right] \leftrightarrow \sum_{k=1}^{K} \frac{1}{\sigma_k^2} \left( \sigma_{f_k}^2(\tilde{x}_i) - 2\tilde{\mu}_{f_k}(\tilde{x}_i) \mu_{f_k} \right).
\]
(75)
Together with (74) and (75), \( \ln q(\mu_f) \) becomes
\[
\ln q(\mu_f) \leftrightarrow \sum_{k=1}^{K} \left[ \left( \frac{1}{\sigma_k^2} + \sum_{i=1}^{N_q} \zeta_k(\tilde{x}_i) \tilde{\varphi}_{f_k} \right) \sigma_k^2(\tilde{x}_i) - 2 \left( \frac{m_k}{\sigma_k^2} + \sum_{i=1}^{N_q} \zeta_k(\tilde{x}_i) \tilde{\varphi}_{f_k} \mu_{f_k}^{(n,n')} \right) \right].
\]
(76)
After completing the square of the summand in (76), we find
\( q(\mu_{f_k}) = N(\tilde{\mu}_{f_k}, \tilde{\sigma}_{f_k}^2) \) \( \forall k \) with
\[
\tilde{\sigma}_{f_k}^2 := \left( \frac{1}{\sigma_k^2} + \sum_{i=1}^{N_q} \zeta_k(\tilde{x}_i) \tilde{\varphi}_{f_k} \right)^{-1}
\]
(77)
and
\[
\tilde{m}_k := \tilde{\sigma}_{f_k}^2 \left( \frac{m_k}{\sigma_k^2} + \sum_{i=1}^{N_q} \zeta_k(\tilde{x}_i) \tilde{\varphi}_{f_k} \mu_{f_k} \right).
\]
(78)
by inspection since \( q(\mu_f) = \prod_{k=1}^{K} q(\mu_{f_k}) \), as in (22). 

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

Similar to \( q(\mu_f) \), the pdf \( q(\varphi_f) \) can be expressed as
\[
\ln q(\varphi_f) = E_{q(\varphi_f)} [\ln p(f, z, \theta|\tilde{s}_i)] + c
\]
\[
\leftrightarrow E_{q(\varphi_f)} [\ln p(f|z, \theta_f)] + E_{q(\varphi_f)} \left[ \sum_{k=1}^{K} \ln p(\varphi_{f_k}) \right]
\]
(79)
by appealing to the proportionality of the conditional posterior \( p(f, z, \theta|\tilde{s}_i) \propto p(f|z, \theta_f)p(\varphi_f) \) w.r.t. \( \varphi_f \). Each term on the RHS in (79) can be thus expressed as
\[
E_{q(\varphi_f)} [\ln p(f|z, \theta_f)]
\]
\[
= \frac{1}{2} \sum_{k=1}^{K} \sum_{i=1}^{N_q} \zeta_k(\tilde{x}_i) \left[ \ln \varphi_{f_k} - \varphi_{f_k} E_{q(\varphi_{f_k})} [(f_i - \mu_{f_k})^2] \right] + c
\]
(80)
where
\[
E_{q(\varphi_{f_k})} [(f_i - \mu_{f_k})^2] = \sigma_{f_k}^2(\tilde{x}_i) + \mu_{f_k}^2(\tilde{x}_i) - 2\mu_{f_k}(\tilde{x}_i) \mu_{f_k} + \sigma_{f_k}^2 + m_k^2,
\]
(81)
and
\[
E_{q(\varphi_f)} \left[ \sum_{k=1}^{K} \ln p(\varphi_{f_k}) \right] = \sum_{k=1}^{K} \left[ \ln \varphi_{f_k} - \frac{\varphi_{f_k}}{\beta_k} \right] + c.
\]
(82)
After substituting (80) and (82) into (79), \( \varphi_f \) can be shown to follow \( q(\varphi_f) = \mathcal{G}(\tilde{\mu}_k, \tilde{\beta}_k) \) \( \forall k \) with
\[
\tilde{a}_k := a_k + \frac{1}{2} \sum_{i=1}^{N_q} \zeta_k(\tilde{x}_i)
\]
(83)
and
\[
\tilde{b}_k := \left[ \frac{1}{\beta_k} + \frac{1}{2} \sum_{i=1}^{N_q} \zeta_k(\tilde{x}_i) \right]^{-1}
\]
(84)
where we used that \( q(\varphi_f) = \prod_{k=1}^{K} q(\varphi_{f_k}) \), as in (22). 

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

To establish the expression for \( H(f|z, \tilde{s}_{\tau+1}; \tilde{\theta}_\tau) \) in (52), consider that at time slot \( \tau + 1 \). Similar to (51), we have
\[
H(f|z = z', \tilde{s}_{\tau+1}; \tilde{\theta}_\tau) \approx \frac{1}{2} \ln |\Sigma_{f|z', \tilde{s}_{\tau+1}, \tilde{\theta}_\tau}| + \frac{N_q}{2} \left( 1 + \ln 2\pi \right).
\]
(85)
With \( \Delta_{w_{\tau+1}} := \text{diag} \left( w_{\tau+1}^{(n,n')} \right) \), and using the construction of \( \sigma_{f_k}^2(\tilde{x}_i) \) in (60), we can write
\[
|\Sigma_{f|z', \tilde{s}_{\tau+1}; \tilde{\theta}_\tau}| = |\Sigma_{f|z', \tilde{s}_{\tau+1}; \tilde{\varphi}_f \Delta_{w_{\tau+1}}}|^{-1}
\]
(86)
from which we deduce that
\[
|\Sigma_{f|z', \tilde{s}_{\tau+1}; \tilde{\theta}_\tau}|^{-1} = |I_{N_q} + \tilde{\varphi}_f \Delta_{w_{\tau+1}} \Sigma_{f|z', \tilde{s}'_{\tau+1}; \tilde{\varphi}_f} |^{-1} \Sigma_{f|z', \tilde{s}'_{\tau+1}; \tilde{\varphi}_f}
\]
(87)
by using the matrix determinant identity lemma [18, Chapter 18]. Further substituting (87) into (85), leads to
\[
H(f|z = z', \tilde{s}_{\tau+1} = \tilde{s}'_{\tau+1}; \tilde{\theta}_\tau) \approx \frac{1}{2} \ln |I_{N_q} + \tilde{\varphi}_f \Delta_{w_{\tau+1}} \Sigma_{f|z', \tilde{s}'_{\tau+1}; \tilde{\theta}_\tau}|
\]
(88)
It follows from the conditional entropy definition in (46) that

\[
H(f|z, \tilde{s}_{r+1}; \theta_r) \\
\approx \sum_{z' \in Z} \int p(z', \tilde{s}_{r+1}; \theta_r) \left( H(f|z = z', \tilde{s}_r = \tilde{s}_{r+1}; \theta_r) \\
- \frac{1}{2} \ln |\mathcal{N}_y + \sum_{f} \Sigma f z', \tilde{s}_{r+1}; \theta_r| \right) d\tilde{s}_{r+1} \\
H(f|z, \tilde{s}_r; \theta_r) \\
(89)
\]

where (e1) is obtained after marginalizing out \( \tilde{s}_{r+1} \) from \( p(z', \tilde{s}_{r+1}; \theta_r) \) as the RHS of (88) is not a function of \( \tilde{s}_{r+1} \).

REFERENCES

[1] 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.

[2] 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.

[3] J. Besag, “Variance Approximations for Approximate Bayesian Inference,” London, U.K.: Gatsby Comput. Neurosci. Unit, Univ. College London, 2003.

[4] 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.

[5] C. M. Bishop, Pattern Recognition and Machine Learning, ser. Information Science and Statistics. Springer, 2006, [Online]. Available: https://books.google.com/books?id=kTNsQgAACAAJ

[6] J. K. Blitzstein and J. Hwang, Introduction to Probability. Boca Raton, FL: USA/CRC, 2014.

[7] T. M. Cover and J. A. Thomas, Elements of Information Theory. New York, NY: USA/Wiley, 1991.

[8] 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.

[9] 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. 951 – 963, Jul. 1999.

[10] Federal Communications Commission, “FCC 11-131,” Unlicensed Operation in the TV Broadcast Bands, 2011.

[11] V. V. Fedorov, Theory of Optimal Experiments. New York, NY: Academic Press, 1972.

[12] 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.

[13] W. R. Gilks, S. Richardson, and D. J. Spiegelhalter, Markov Chain Monte Carlo in Practice. London: Chapman and Hall, 1996.

[14] 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.

[15] 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.

[16] 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.

[17] 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

[18] D. A. Harville, Matrix Algebra From a Statistician’s Perspective. New York: Springer-Verlag, 1997.

[19] D. Higdon, “Spatial applications of Markov chain Monte Carlo for Bayesian inference,” Ph.D. dissertation, Dept. Stat., Univ. Washington, Seattle, WA, 1994.