Massive MIMO-OFDM Systems with Low Resolution ADCs: Cramér-Rao Bound, Sparse Channel Estimation, and Soft Symbol Decoding Supplementary Material

Sai Subramanyam Thoota and Chandra R. Murthy

I. Proof of Theorem 1

The conditional probability distribution of $\widetilde{\mathbf{Y}}$ given $\widetilde{\mathbf{X}}$, $\widetilde{\mathbf{\Phi}}$ and $\widetilde{\sigma}_w$ is

$$\begin{aligned} p(\widetilde{\mathbf{Y}}|\widetilde{\mathbf{X}};\widetilde{\mathbf{\Phi}},\widetilde{\sigma}_w^2) &= \prod_{\ell=1}^{T} p(\widetilde{\mathbf{y}}_\ell|\widetilde{\mathbf{x}}_\ell;\widetilde{\mathbf{\Phi}},\widetilde{\sigma}_w^2) \\ &= \prod_{\ell=1}^{T}\int_{\widetilde{\mathbf{z}}_\ell\in\mathbb{R}^{\widetilde{M}}} p(\widetilde{\mathbf{y}}_\ell|\widetilde{\mathbf{z}}_\ell)p(\widetilde{\mathbf{z}}_\ell|\widetilde{\mathbf{x}}_\ell;\widetilde{\mathbf{\Phi}},\widetilde{\sigma}_w^2)\,\mathrm{d}\widetilde{\mathbf{z}}_\ell \\ &= \prod_{\ell=1}^{T}\prod_{m=1}^{\widetilde{M}} B_{m\ell}, \end{aligned} \tag{52}$$

where

$$B_{m\ell} \triangleq \int_{\frac{\widetilde{z}_{m\ell}^{(\mathrm{lo})}-\sum_{n=1}^{\widetilde{N}}\widetilde{\Phi}_{mn}\widetilde{x}_{n\ell}}{\widetilde{\sigma}_w}}^{\frac{\widetilde{z}_{m\ell}^{(\mathrm{hi})}-\sum_{n=1}^{\widetilde{N}}\widetilde{\Phi}_{mn}\widetilde{x}_{n\ell}}{\widetilde{\sigma}_w}} \frac{1}{\sqrt{2\pi}}\exp\left[-\frac{\widetilde{z}_{m\ell}^2}{2}\right]\mathrm{d}\widetilde{z}_{m\ell} \tag{53}$$

and $\widetilde{z}_{m\ell}^{(\mathrm{lo})}$ and $\widetilde{z}_{m\ell}^{(\mathrm{hi})}$ are the lower and upper quantization thresholds for the $(m,\ell)^{\text{th}}$ entry of $\widetilde{\mathbf{Y}}$, respectively. Also, $\widetilde{\Phi}_{mn}$ and $\widetilde{x}_{n\ell}$ denote the $(m,n)^{\text{th}}$ and $(n,\ell)^{\text{th}}$ entries of $\widetilde{\mathbf{\Phi}}$ and $\widetilde{\mathbf{X}}$, respectively.

Note that, since we estimate $\widetilde{N}T$ parameters in total, the BIM is block diagonal matrix of size $\widetilde{N}T\times\widetilde{N}T$, with T blocks each of size $\widetilde{N}\times\widetilde{N}$. Computing it requires the gradient and Hessian of the joint probability distribution w.r.t. $\mathbf{x}_\ell\ \forall\ell$. Since the columns of $\widetilde{\mathbf{X}}$ are independent of each other, we express the logarithm of the joint distribution using the chain rule as shown in (54). In (54), we omit the terms that do not depend on $\widetilde{\mathbf{X}}$ for brevity. We can verify that the joint probability distribution in (54) fall into the exponential family of distributions which satisfies the regularity conditions in Sec. 5.2.3 of [1]. We apply Leibniz integral rule to compute the first and second derivatives of $\log B_{ml}$ with respect to (w.r.t.) $\widetilde{x}_{k\ell}$ and $\widetilde{x}_{j\ell}$ shown in (55) and (56), respectively, where

$$\widetilde{\eta}_{m\ell}^{(\mathrm{hi})} \triangleq \frac{\widetilde{z}_{m\ell}^{(\mathrm{hi})}-\sum_{n=1}^{\widetilde{N}}\widetilde{\Phi}_{mn}\widetilde{x}_{n\ell}}{\widetilde{\sigma}_w}, \tag{57}$$

$$\widetilde{\eta}_{m\ell}^{(\mathrm{lo})} \triangleq \frac{\widetilde{z}_{m\ell}^{(\mathrm{lo})}-\sum_{n=1}^{\widetilde{N}}\widetilde{\Phi}_{mn}\widetilde{x}_{n\ell}}{\widetilde{\sigma}_w}. \tag{58}$$

In (56), $f(\cdot)$ and $F(\cdot)$ denote the PDF and CDF of a standard normal random variable, respectively. Writing in matrix form, the ℓ^{th} diagonal block of the BIM, denoted by $\widetilde{\mathbf{M}}_\ell(\widetilde{\mathbf{\Phi}},a,r,\widetilde{\sigma}_w^2)$, is shown in (59).

II. Proof of Lemma 1

As $\delta\to 0$, both (12) and (13) become indeterminate forms. Further, both the numerators and denominators in the left hand sides of (12) and (13) are differentiable at 0. Applying L'Hôpital's and Leibniz integral rules (for differentiating the denominators), we get the right hand sides in (12) and (13).

III. Proof of Lemma 2

The proof follows by using the lower and upper thresholds of the 1-bit quantizer as follows:

$$\widetilde{z}_{m\ell}^{(\mathrm{lo})} = \begin{cases} 0 & \text{if } \widetilde{y}_{m\ell}=+1 \\ -\infty & \text{if } \widetilde{y}_{m\ell}=-1 \end{cases} \tag{60}$$

and

$$\widetilde{z}_{m\ell}^{(\mathrm{hi})} = \begin{cases} \infty & \text{if } \widetilde{y}_{m\ell}=+1 \\ 0 & \text{if } \widetilde{y}_{m\ell}=-1. \end{cases} \tag{61}$$

Substituting (60) and (61) in (6), after straightforward algebraic manipulation and using the facts that $F(\infty)=1$, $F(-\infty)=0$ and $F(\widetilde{\eta}_{m\ell})=1-F(-\widetilde{\eta}_{m\ell})\ \forall\, m,\ell$, we get (15).

IV. Channel Estimation And Data Detection As Statistical Inference

We formulate the received system as probabilistic graphical models, and infer the posterior distributions of the channel and data symbols given the quantized pilot and data observations. We represent these Bayesian network graphical models in Figures 2, 3 and 4. As our goal is to obtain the posterior beliefs or LLRs of the data symbols that will be input to a channel decoder, a statistical inference framework is a suitable approach to solve our problems. We use shaded circles, transparent circles, and squares to represent the observations, latent variables, and deterministic variables. In our channel estimation and data detection problem, the quantized received pilot and data signals are the observations, channel and data symbols are the latent variables, pilot symbols and noise variance are the deterministic variables.

As mentioned earlier, the computational intractability of joint channel estimation and data detection problem necessitates us to adopt an iterative algorithm. So, we use separate

$$\log p(\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}}, \widetilde{\mathbf{P}}; \widetilde{\mathbf{\Phi}}, \widetilde{\sigma}_w^2, a, r) = \log p(\widetilde{\mathbf{Y}} | \widetilde{\mathbf{X}}; \widetilde{\mathbf{\Phi}}, \widetilde{\sigma}_w^2) + \log p(\widetilde{\mathbf{X}}; \widetilde{\mathbf{P}}) + \log p(\boldsymbol{\alpha}; a, r),$$
$$\propto \sum_{\ell=1}^{T} \sum_{m=1}^{\widetilde{M}} \log \int_{\frac{\widetilde{z}_{m\ell}^{(\mathrm{lo})} - \sum_{n=1}^{\widetilde{N}} \widetilde{\Phi}_{mn} \widetilde{x}_{n\ell}}{\widetilde{\sigma}_w}}^{\frac{\widetilde{z}_{m\ell}^{(\mathrm{hi})} - \sum_{n=1}^{\widetilde{N}} \widetilde{\Phi}_{mn} \widetilde{x}_{n\ell}}{\widetilde{\sigma}_w}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{\widetilde{z}_{m\ell}^2}{2}\right] \mathrm{d}\widetilde{z}_{m\ell} - \sum_{\ell=1}^{T} \frac{\widetilde{\mathbf{x}}_\ell^T \widetilde{\mathbf{P}} \widetilde{\mathbf{x}}_\ell}{2}. \tag{54}$$

$$\frac{\partial}{\partial \widetilde{x}_{k\ell}} \log B_{m\ell} = -\frac{\widetilde{\Phi}_{mk}}{\sqrt{2\pi} B_{m\ell} \widetilde{\sigma}_w} \left[\exp\left(-\frac{(\widetilde{z}_{m\ell}^{(\mathrm{hi})} - \sum_{n=1}^{\widetilde{N}} \widetilde{\Phi}_{mn} \widetilde{x}_{n\ell})^2}{2\widetilde{\sigma}_w^2}\right) - \exp\left(-\frac{(\widetilde{z}_{m\ell}^{(\mathrm{lo})} - \sum_{n=1}^{\widetilde{N}} \widetilde{\Phi}_{mn} \widetilde{x}_{n\ell})^2}{2\widetilde{\sigma}_w^2}\right)\right]. \tag{55}$$

$$-\frac{\partial^2}{\partial \widetilde{x}_{j\ell} \widetilde{x}_{k\ell}} \log B_{m\ell} = \frac{\widetilde{\Phi}_{mk} \widetilde{\Phi}_{mj}}{\widetilde{\sigma}_w^2} \left[\frac{\widetilde{\eta}_{m\ell}^{(\mathrm{hi})} f(\widetilde{\eta}_{m\ell}^{(\mathrm{hi})}) - \widetilde{\eta}_{m\ell}^{(\mathrm{lo})} f(\widetilde{\eta}_{m\ell}^{(\mathrm{lo})})}{F(\widetilde{\eta}_{m\ell}^{(\mathrm{hi})}) - F(\widetilde{\eta}_{m\ell}^{(\mathrm{lo})})} + \left(\frac{f(\widetilde{\eta}_{m\ell}^{(\mathrm{hi})}) - f(\widetilde{\eta}_{m\ell}^{(\mathrm{lo})})}{F(\widetilde{\eta}_{m\ell}^{(\mathrm{hi})}) - F(\widetilde{\eta}_{m\ell}^{(\mathrm{lo})})}\right)^2\right], \tag{56}$$

$$\widetilde{\mathbf{M}}_\ell(\widetilde{\mathbf{\Phi}}, a, r, \widetilde{\sigma}_w^2) = \mathbb{E}\left[-\frac{\partial^2}{\partial \widetilde{\mathbf{x}}_\ell \widetilde{\mathbf{x}}_\ell^T} \log p(\widetilde{\mathbf{Y}}, \widetilde{\mathbf{X}}; \widetilde{\mathbf{P}}, \widetilde{\mathbf{\Phi}}, \widetilde{\sigma}_w^2, a, r)\right]$$
$$= \widetilde{\mathbf{\Phi}}^T \mathrm{diag}\left(\frac{1}{\widetilde{\sigma}_w^2} \mathbb{E}\left[\frac{\widetilde{\eta}_{m\ell}^{(\mathrm{hi})} f(\widetilde{\eta}_{m\ell}^{(\mathrm{hi})}) - \widetilde{\eta}_{m\ell}^{(\mathrm{lo})} f(\widetilde{\eta}_{m\ell}^{(\mathrm{lo})})}{F(\widetilde{\eta}_{m\ell}^{(\mathrm{hi})}) - F(\widetilde{\eta}_{m\ell}^{(\mathrm{lo})})} + \left(\frac{f(\widetilde{\eta}_{m\ell}^{(\mathrm{hi})}) - f(\widetilde{\eta}_{m\ell}^{(\mathrm{lo})})}{F(\widetilde{\eta}_{m\ell}^{(\mathrm{hi})}) - F(\widetilde{\eta}_{m\ell}^{(\mathrm{lo})})}\right)^2\right]\right)_{m=1}^{\widetilde{M}} \widetilde{\mathbf{\Phi}} + \mathbb{E}[\widetilde{\mathbf{P}}]. \tag{59}$$

Bayesian network models for the channel estimation and data decoding problems. We explain the intractability issue mathematically here. The posterior distribution of channel $\mathbf{H}$ and data $\{\mathbf{x}_1^{(d)}[\tau_p+1], \ldots, \mathbf{x}_K^{(d)}[\tau_p+\tau_d]\}$ given the observations $\mathbf{Y}^{(p)}$, $\mathbf{Y}^{(d)}$ and pilots $\mathbf{X}^{(p)}[1], \ldots, \mathbf{X}^{(p)}[\tau_p]$ is given by (62), where $\mathbf{Y}^{(p)}$ and $\mathbf{Y}^{(d)}$ are the marginal likelihoods as shown in (63) and (64), respectively.[1]

Exact computation of the posterior distributions using the above is computationally intractable, as it requires solving high dimensional integrals over $\mathbf{H}$, $\mathbf{x}_1^{(d)}[\tau_p+1], \ldots, \mathbf{x}_K^{(d)}[\tau_p+\tau_d]$ to obtain the partition functions $P(\mathbf{Y}^{(p)})$ and $P(\mathbf{Y}^{(d)})$. Moreover, we estimate the UEs' channels in their lag domain, and use their frequency domain representation for data detection, which complicates the joint channel estimation and data detection problems further. These difficulties motivate the need to employ approximate inference techniques to solve the channel estimation and data detection problems.

V. Proof of Lemma 3

To obtain the posterior distribution $q_{\mathbf{H}}(\mathbf{H})$, we first compute the posterior distribution of each of the factors $q_{\mathbf{h}_n}(\mathbf{h}_n)$, $n = \{1, \ldots, N_r\}$. We calculate the expectation of the joint distribution in (31) with respect to the posterior distributions $\{q_{\mathbf{Z}}(\mathbf{Z}^{(p)}), q_{\mathbf{h}_1}(\mathbf{h}_1), \ldots, q_{\mathbf{h}_{n-1}}(\mathbf{h}_{n-1}), q_{\mathbf{h}_{n+1}}(\mathbf{h}_{n+1}), \ldots,$ $q_{\mathbf{h}_{N_r}}(\mathbf{h}_{N_r}), q_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})\}$ as follows:

$$\ln q_{\mathbf{h}_n}(\mathbf{h}_n) = \Big\langle \ln\, p(\mathbf{Y}^{(p)} \,|\, \mathbf{Z}^{(p)}) + \ln\, p(\mathbf{Z}^{(p)} \,|\, \mathbf{H}; \mathbf{\Phi}^{(p)}, \sigma_w^2) + \ln\, p(\mathbf{H}|\mathbf{P}) + \ln\, p(\boldsymbol{\alpha}; a, r)\rangle \tag{65}$$
$$\propto \langle \ln\, p(\mathbf{Z}^{(p)} \,|\, \mathbf{H}; \mathbf{\Phi}^{(p)}, \sigma_w^2) + \ln\, p(\mathbf{H}|\mathbf{P})\rangle, \tag{66}$$

where $\langle\cdot\rangle$ denotes the expectation operation w.r.t. the posterior distributions of all the latent variables except $q_{\mathbf{h}_n}(\mathbf{h}_n)$. We obtain (66) from (65) by including only the terms that do not depend on $\mathbf{h}_n$ as proportionality constants such that $q_{\mathbf{h}_n}(\mathbf{h}_n)$ becomes a probability distribution. Simplifying (66) by separating the terms that depend only on $\mathbf{h}_n$, we get

$$\ln q_{\mathbf{h}_n}(\mathbf{h}_n) \propto \left\langle -\frac{1}{\sigma_w^2}\left(\mathbf{h}_n^H \mathbf{\Phi}^{(p)H} \mathbf{\Phi}^{(p)} \mathbf{h}_n - \mathbf{z}_n^{(p)H} \mathbf{\Phi}^{(p)} \mathbf{h}_n - \mathbf{h}_n^H \mathbf{\Phi}^{(p)H} \mathbf{z}_n^{(p)}\right) - \mathbf{h}_n^H \mathbf{P} \mathbf{h}_n \right\rangle \tag{67}$$
$$\propto -\left(\mathbf{h}_n^H \left(\frac{1}{\sigma_w^2} \mathbf{\Phi}^{(p)H} \mathbf{\Phi}^{(p)} + \langle\mathbf{P}\rangle\right) \mathbf{h}_n - \frac{1}{\sigma_w^2} \left\langle \mathbf{z}_n^{(p)} \right\rangle^H \mathbf{\Phi}^{(p)} \mathbf{h}_n - \frac{1}{\sigma_w^2} \mathbf{h}_n^H \mathbf{\Phi}^{(p)H} \left\langle \mathbf{z}_n^{(p)} \right\rangle\right), \tag{68}$$

where $\left\langle \mathbf{z}_n^{(p)} \right\rangle$ is the mean of $q_{\mathbf{z}_n}\left(\mathbf{z}_n^{(p)}\right)$, $\langle\mathbf{P}\rangle = \mathrm{diag}(\langle\boldsymbol{\alpha}\rangle)$, and $\langle\boldsymbol{\alpha}\rangle$ is the mean of $q_{\boldsymbol{\alpha}}(\boldsymbol{\alpha})$. Taking exponentials on both sides of (68), and by completing the squares, we can deduce from the structure of the resulting expression that $q_{\mathbf{h}_n}(\mathbf{h}_n)$ is complex normal distributed with covariance matrix and mean given by

$$\mathbf{\Sigma}_{\mathbf{H}} = \left(\frac{1}{\sigma_w^2} \mathbf{\Phi}^{(p)H} \mathbf{\Phi}^{(p)} + \langle\mathbf{P}\rangle\right)^{-1}, \tag{69}$$
$$\langle\mathbf{h}_n\rangle = \frac{1}{\sigma_w^2} \mathbf{\Sigma}_{\mathbf{H}} \mathbf{\Phi}^{(p)H} \left\langle \mathbf{z}_n^{(p)} \right\rangle, \tag{70}$$

respectively. Note that, the covariance matrix $\mathbf{\Sigma}_{\mathbf{H}}$ is independent of n. So, we can write the posterior mean of $q_{\mathbf{h}_n}(\mathbf{h}_n)$, $n = \{1, \ldots, N_r\}$ in a matrix form to get (37) and (38). ■

VI. Proof of Lemma 4

To obtain the posterior distribution $q_{\mathbf{Z}}(\mathbf{Z}^{(p)})$, we first compute the posteriors of each of its factors $q_{\mathbf{z}_n}(\mathbf{z}_n^{(p)})$, $n = \{1, \ldots, N_r\}$. We calculate the expectation of the log

[1] $\mathbf{X}^{(d)}$ comes from a discrete M-QAM constellation, but we use integrals here for convenience. In the actual derivation, the integrals are replaced by summations.

$$p\left(\mathbf{H},\mathbf{x}_1^{(d)}[\tau_p+1],\ldots,\mathbf{x}_K^{(d)}[\tau_p+\tau_d]\,|\,\mathbf{Y}^{(p)},\mathbf{Y}^{(d)};\mathbf{X}^{(p)}[1],\ldots,\mathbf{X}^{(p)}[\tau_p]\right)$$
$$=\frac{p(\mathbf{Y}^{(p)}\,|\,\mathbf{H};\mathbf{X}^{(p)}[1],\ldots,\mathbf{X}^{(p)}[\tau_p])\,p(\mathbf{Y}^{(d)}\,|\,\mathbf{H},\mathbf{x}_1^{(d)}[\tau_p+1],\ldots,\mathbf{x}_K^{(d)}[\tau_p+\tau_d])p(\mathbf{H})\,\prod_{k=1}^{K}\prod_{t=\tau_p+1}^{\tau_p+\tau_d}p(\mathbf{x}_k^{(d)}[t])}{p(\mathbf{Y}^{(p)};\mathbf{X}^{(p)}[1],\ldots,\mathbf{X}^{(p)}[\tau_p])\,p(\mathbf{Y}^{(d)})}. \tag{62}$$

$$p(\mathbf{Y}^{(p)})=\int p(\mathbf{Y}^{(p)}\,|\,\mathbf{H};\mathbf{X}^{(p)}[1],\ldots,\mathbf{X}^{(p)}[\tau_p])\,p(\mathbf{H})\,\mathrm{d}\mathbf{H}, \tag{63}$$
$$p(\mathbf{Y}^{(d)})=\int p(\mathbf{Y}^{(d)}\,|\,\mathbf{H},\mathbf{x}_1^{(d)}[\tau_p+1],\ldots,\mathbf{x}_K^{(d)}[\tau_p+\tau_d])\,p(\mathbf{H})\prod_{k=1}^{K}\prod_{t=\tau_p+1}^{\tau_p+\tau_d}p(\mathbf{x}_k^{(d)}[t])\,\mathrm{d}\mathbf{x}_k^{(d)}[t]\,\mathrm{d}\mathbf{H}. \tag{64}$$

of the joint probability distribution in (31) with respect to the posterior distributions of all the latent variables except $q_{\mathbf{z}_n}(\mathbf{z}_n^{(p)})$ as follows:

$$\begin{aligned}&\ln q_{\mathbf{z}_n}(\mathbf{z}_n^{(p)})\\&=\Big\langle \ln\, p(\mathbf{Y}^{(p)}\,|\,\mathbf{Z}^{(p)})+\ln\, p(\mathbf{Z}^{(p)}\,|\,\mathbf{H};\mathbf{\Phi}^{(p)},\sigma_w^2)\\&\qquad+\ln\, p(\mathbf{H}|\mathbf{P})+\ln\, p(\boldsymbol{\alpha};a,r)\rangle &(71)\\&\propto\Big\langle \ln\, p(\mathbf{y}_n^{(p)}\,|\,\mathbf{z}_n^{(p)})+\ln\, p(\mathbf{z}_n^{(p)}\,|\,\mathbf{h}_n;\mathbf{\Phi}^{(p)},\sigma_w^2)\Big\rangle &(72)\\&\propto \ln\mathbb{1}(\mathbf{z}_n^{(p)}\in(\mathbf{z}_n^{(\mathrm{lo})},\mathbf{z}_n^{(\mathrm{hi})}))-\frac{1}{\sigma_w^2}\left\langle\left\|\mathbf{z}_n^{(p)}-\mathbf{\Phi}^{(p)}\mathbf{h}_n\right\|^2\right\rangle. &(73)\end{aligned}$$

By expanding the second term in (73), completing the squares, and taking exponential on both sides, $q_{\mathbf{z}_n}(\mathbf{z}_n^{(p)})$ is truncated complex normal distributed with mean given by

$$\left\langle\mathbf{z}_n^{(p)}\right\rangle=\mathbf{\Phi}^{(p)}\langle\mathbf{h}_n\rangle+\frac{\sigma_w}{\sqrt{2}}\frac{f\left(\frac{\mathbf{z}_n^{(\mathrm{lo})}-\mathbf{\Phi}^{(p)}\langle\mathbf{h}_n\rangle}{\sigma_w/\sqrt{2}}\right)-f\left(\frac{\mathbf{z}_n^{(\mathrm{hi})}-\mathbf{\Phi}^{(p)}\langle\mathbf{h}_n\rangle}{\sigma_w/\sqrt{2}}\right)}{F\left(\frac{\mathbf{z}_n^{(\mathrm{hi})}-\mathbf{\Phi}^{(p)}\langle\mathbf{h}_n\rangle}{\sigma_w/\sqrt{2}}\right)-F\left(\frac{\mathbf{z}_n^{(\mathrm{lo})}-\mathbf{\Phi}^{(p)}\langle\mathbf{h}_n\rangle}{\sigma_w/\sqrt{2}}\right)}. \tag{74}$$

We have included the second order terms of $q_{\mathbf{h}_n}(\mathbf{h}_n)$ as part of the proportionality constant to arrive at (74). By writing the posterior means of $q_{\mathbf{z}_n}(\mathbf{z}_n^{(p)})$ in matrix form, we get (39). ■

VII. Proof of Lemma 5

We follow similar steps that are used to compute $q_{\mathbf{Z}}(\mathbf{Z}^{(p)})$ and $q_{\mathbf{H}}(\mathbf{H})$ to obtain $q_{\alpha_k}(\alpha_k)$, $1\le k\le KL$ as follows:

$$\begin{aligned}\ln q_{\alpha_k}(\alpha_k)&=\Big\langle \ln\, p(\mathbf{Y}^{(p)}\,|\,\mathbf{Z}^{(p)})+\ln\, p(\mathbf{Z}^{(p)}\,|\,\mathbf{H};\mathbf{\Phi}^{(p)},\sigma_w^2)\\&\qquad+\ln\, p(\mathbf{H}|\mathbf{P})+\ln\, p(\boldsymbol{\alpha};a,r)\rangle &(75)\\&\propto\langle\ln\, p(\mathbf{H}|\mathbf{P})+\ln\, p(\boldsymbol{\alpha};a,r)\rangle &(76)\\&\propto(a+N_r-1)\ln\alpha_k-\Big(r+\sum_{n=1}^{N_r}\langle|h_{kn}|^2\rangle\Big)\alpha_k. &(77)\end{aligned}$$

From the structure of (77), we see that $q_{\boldsymbol{\alpha_k}}(\alpha_k)\;\forall k$ is Gamma distributed with shape and rate parameters given by (40). The mean $\langle\alpha_k\rangle$ is computed as $\widetilde{a}_k/\widetilde{r}_k$, where $\widetilde{a}_k$ and $\widetilde{r}_k$ are as defined in (40). ■

VIII. Proof of Lemma 6

We obtain the posterior distribution $q_x\left(x_{kt}^{(d)}\right)$ by computing the expectation of the log of the joint distribution with respect to the posterior distributions of all the latent variables except $q_{x_{kt}}$.

$$\begin{aligned}&\ln q_{x_{kt}}\left(x_{kt}^{(d)}\right)\\&\quad=\Big\langle\ln p\left(\mathbf{Y}^{(d)}|\mathbf{Z}^{(d)}\right)+\ln p\left(\mathbf{Z}^{(d)}|\mathbf{X}^{(d)},\mathbf{D},\sigma_w^2\right)\\&\qquad\qquad+\ln p\left(\mathbf{X}^{(d)}\right)\Big\rangle\\&\quad\propto-\frac{1}{\sigma_w^2}\Big(\|\mathbf{D}_{:,k}\|^2\left|x_{kt}^{(d)}\right|^2-2\Re\Big[\mathbf{D}_{:,k}^H\Big(\left\langle\mathbf{z}_t^{(d)}\right\rangle\\&\qquad-\sum_{\substack{k'=1\\k'\neq k}}^{KN_c}\mathbf{D}_{:,k'}\left\langle x_{k't}^{(d)}\right\rangle\Big)x_{kt}^{(d)*}\Big]\Big)+\ln p(x_{kt}^{(d)}). &(78)\end{aligned}$$

We include all the terms that do not depend on $x_{kt}^{(d)}$ as part of the proportionality constant. Now, we substitute the values of $x_{kt}^{(d)}$ from a discrete constellation set in (78) and take exponential on both the sides to get an expression for the probability mass at a constellation point, and normalize it to obtain the posterior probability mass function of $x_{kt}^{(d)}$ given in (46). ■

References

[1] J. H. Dauwels, *On graphical models for communications and machine learning: Algorithms, bounds, and analog implementation.* ETH Zurich, 2006, vol. 17.