Learning-Aided Joint Channel Estimation and Detection in OTFS with Graph Neural Networks

Xinwei Qu

***Abstract*—**

***Index Terms*—OTFS joint channel estimation and symbol detection, graph neural network, 6G**

I. Introduction

A. Related Works

B. Contributions

The main contributions of this paper are summarized as follows.

- we propose an OTFS frame design that increase the transmission efficiency and overcomes the overspreading channel issue
- We propose a novel neural network-based framework for joint channel estimation and symbol detection for OTFS systems, referred as Joint Parallel Interference Cancellation Network (JPICNet) framework. The proposed framework integrates the joint PIC scheme and graph neural network.
- we incorporate attention mechanism into neural network-based framework to improve the performance.

C. Notations

a, $\mathbf{a}$ and $\mathbf{A}$ demote scalar, vector, and matrix respectively. $\mathbf{I}$ denotes an identity matrix. $\mathbb{C}^{M\times N}$ denotes the set of $M\times N$ dimensional complex matrix. $(\cdot)^{\mathrm{T}}$, $(\cdot)^{\mathrm{H}}$, $(\cdot)^{*}$, and $[\cdot]_M$ represent the transpose, Hermitian transpose, conjugate, and mod-M operations. $\odot$ denotes Hadamard multiplication. $\mathsf{diag}(\boldsymbol{a})$ denotes the operation to diagonize a vector $\boldsymbol{a}$, $\mathsf{off}(\mathbf{A})$ forces all diagonal elements to zero. We define $\mathbf{a}=\mathsf{vec}(\mathbf{A})$ as the column-wise vectorization of matrix $\mathbf{A}$. For any real number, $\lfloor\cdot\rfloor$ denotes the greatest integer less than or equal.

II. System Model

We first introduce the OTFS mod/demod and its frame structure. Subsequently, we derive the input-output relation in the delay-Doppler (DD) domain for the two most widely adopted pulse-shaping waveforms.

A. OTFS System

We consider a single input single output (SISO) OTFS system as illustrated in Fig. 1. The transmitter operates an OTFS frame (detailed in II-B), $\mathbf{X}[k,l]\in\mathbb{C}^{K\times L}$, with $k=0,\cdots,K-1$ and $l=0,\cdots,L-1$ indexing discretized Doppler and delay shifts, respectively. After transposition. the frame is converted to the time-frequency (TF) domain via the inverse symplectic finite Fourier transform (ISFFT), mapping the data on $L\times K$ grids with uniform intervals Δf (Hz) and $T=1/\Delta f$ (seconds). The time-domain signal is synthesized using (discrete) Heisenberg transform with a pulse-shaping waveform employing a single initial cyclic prefix spanning the full OTFS frame duration. The time-domain signal is transmitted over a time-varying wireless channel characterized by the delay-Doppler impulse response $h(\tau,v)$ as [**?**],

$$h(\tau,v)=\sum_{i=1}^{P}h_i\delta(\tau-\tau_i)\delta(v-v_i), \tag{1}$$

where $\delta(\cdot)$ denotes the Dirac delta function, $h_i\sim\mathcal{N}(0,\frac{1}{P})$ is the gain of the i-th propagation path, and P represents the total number of paths. Each path is characterized by distinct delay and/or Doppler shifts, modeling the channel response between the receiver and either moving reflectors or the transmitting source. The delay and Doppler shifts are given as,

$$\tau_i=l_i\frac{T}{L}, v_i=k_i\frac{\Delta f}{K}, \tag{2}$$

respectively. Let the integers $l_i\in[0,l_{\max}]$ and $k_i\in[-k_{\max},k_{\max}]$ represent the delay and Doppler shift indices, respectively, where $l_{\max}$ and $k_{\max}$ denote the maximum delay index and maximum Doppler shift index across all propagation paths. Note that we restrict our consideration to integer-valued indices, as fractional delay and Doppler shifts can be equivalently represented through virtual integer taps in the delay-Doppler domain using the techniques described in [**? ? ?**].

At the receiver, the time-domain signal is first transformed to the time-frequency (TF) domain using a matched filter followed by Wigner transform. The resulting signal is then transposed and converted to the delay-Doppler (DD) domain through the symplectic finite Fourier transform (SFFT), producing the received frame $\mathbf{Y}[k,l]\in\mathbb{C}^{K\times L}$. The DD domain input-output relationship can be formulated in a vector form as [**?**],

$$\mathbf{y}=\mathbf{H}\mathbf{x}+\mathbf{z}, \tag{3}$$

where $\mathbf{x}=\mathsf{vec}(\mathbf{X}^{\mathrm{T}})$, $\mathbf{y}=\mathsf{vec}(\mathbf{Y}^{\mathrm{T}})$, and $\mathbf{z}\sim\mathcal{CN}(0,\sigma^2)$ is an independent and identically distributed (i.i.d.) Gaussian noise [**? ? ?**].

B. OTFS Frame Structure

As illustrated in Fig. 2, a superimposed OTFS frame structure is considered, where pilot and data symbols are jointly embedded over delay-Doppler grids, i.e.,

$$\mathbf{X}=\mathbf{X}_d+\mathbf{X}_p, \tag{4}$$

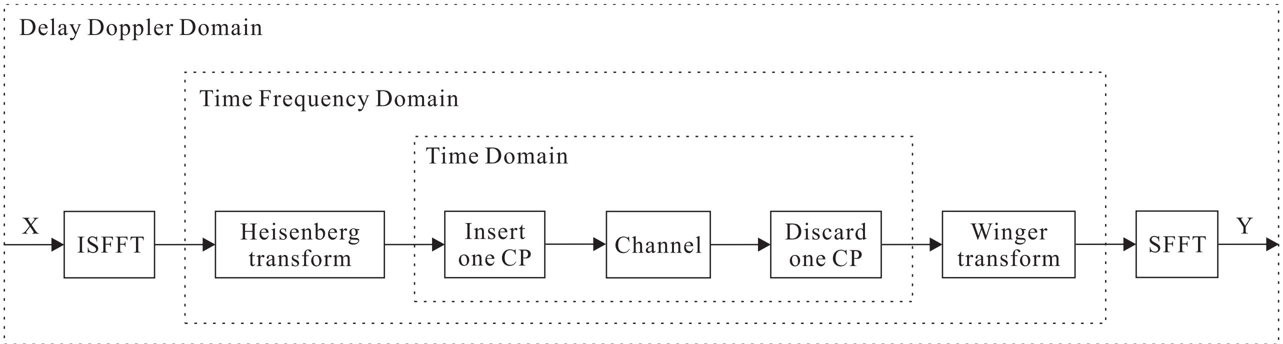

Fig. 1: OTFS mode/demod

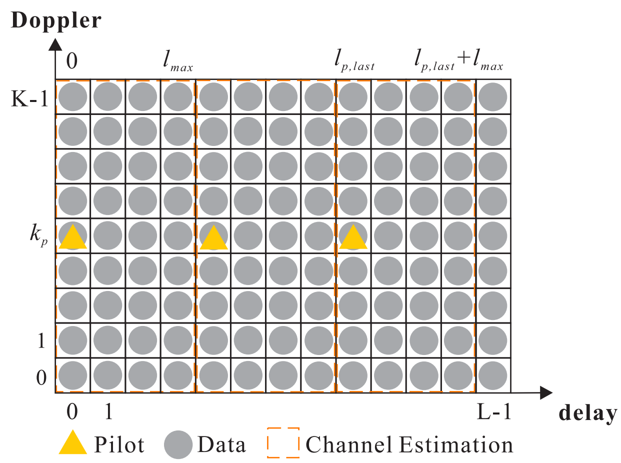

Fig. 2: OTFS frame structure with the last pilot delay index at $l_{p,\text{last}} = (N_p - 1)(l_{\max} + 1)$

where $\mathbf{X}_d[k,l] \in \mathbb{C}^{K \times L}$ denotes the data frame composed of quadrature amplitude modulation (QAM) symbols drawn from a constellation $\mathcal{A}$ with average energy E_d. The pilot frame $\mathbf{X}_p[k,l]$ contains nonzero elements only at designated positions, i.e.,

$$\mathbf{X}_p[k,l] = \begin{cases} x_p, & k = k_p,\ l = l_p, \\ 0, & \text{otherwise}, \end{cases} \tag{5}$$

where x_p is the pilot symbol with energy E_p, $k_p = \lfloor (K-1)/2 \rfloor$ is the Doppler index of all pilots, and $l_p = i(l_{\max}+1)$ for $i = 0, \ldots, N_p - 1$ are their delay indices. Here, $N_p = \lfloor L/(l_{\max}+1) \rfloor$ denotes the total number of pilots. Each pilot facilitates channel estimation over a region of size $K \times (l_{\max}+1)$ in the DD domain.

III. OTFS Channel Estimation using VB

A. OTFS Channel Estimation using VB

We assume the channel follows the Gaussian distribution, i.e.,

$$p(h|\gamma) = \prod_{i=0}^{P_{\max}-1} p(h_i|\gamma_i) = \prod_{i=0}^{P-1} \mathcal{CN}(h_i; 0, \gamma_i^{-1}), \tag{6}$$

where $P_{\max} = (l_{\max}+1)(2k_{\max}+1)$, $\gamma = [\gamma_0, \gamma_1, \cdots, \gamma_{P_{\max}-1}]$ is the precision vector of h. γ follows the Gamma distribution, i.e.,

$$p(\gamma) = \text{Gamma}(\gamma; a, b) \tag{7}$$

where a is the shape parameter and b is the inverse scale parameter. Also, we assume the noise obeys the Gaussian distribution, i.e.,

$$p(z) = \mathbf{z} \sim \mathcal{CN}(0, \alpha^{-1}\boldsymbol{I}) \tag{8}$$

where α obeys the Gamma distribution, i.e.,

$$p(\alpha) = \text{Gamma}(\alpha; c, d) \tag{9}$$

where c is the shape parameter and d is the inverse scale parameter. For all parameters, we define $\Theta = \{h, \gamma, \alpha\}$.

Here, we use the mean field assumption to estimate the channel, i.e.,

$$\begin{aligned} p(y,\Theta) &= p(y|h,\alpha)p(h|\gamma)p(\gamma)p(\alpha) \\ lnp(y,\Theta) &= lnp(y|h,\alpha) + lnp(h|\gamma) + lnp(\gamma) + lnp(\alpha) \end{aligned} \tag{10}$$

Here, we use a distribution family Q over Θ to approximate $p(\Theta|y)$, i.e.,

$$\begin{aligned} p(\Theta|y) &= \underset{q(\Theta)\in Q}{\arg\min}\, KL(q(\Theta)||p(\Theta|y)) \\ &= \underset{q(\Theta)\in Q}{\arg\min} - \int_{\Theta} q(\Theta)\ln\frac{p(\Theta|y)}{q(\Theta)} d\Theta \end{aligned} \tag{11}$$

where $q(\Theta)$ follows the mean field assumption, i.e.,

$$q(\Theta) = q(h)q(\gamma)q(\alpha) \tag{12}$$

Therefore, we can update the probability functions as follows:

$$q^{(t+1)}(\alpha) \propto exp(\mathbb{E}_{q_{-\alpha}^{(t)}}[\ln p(y,\Theta)]\}) \tag{13}$$

$$q^{(t+1)}(h) \propto exp(\mathbb{E}_{q_{-h}^{(t)}}[\ln p(y,\Theta)]\}) \tag{14}$$

$$q^{(t+1)}(\gamma) \propto exp(\mathbb{E}_{q_{-\gamma}^{(t)}}[\ln p(y,\Theta)]\}) \tag{15}$$

The update is computed as follows

1) Update $q(\alpha)$

$$q^{(t+1)}(\alpha) \propto exp(\mathbb{E}_{q_{-\alpha}^{(t)}}[\ln p(y,\Theta)]\}) \tag{16}$$

$$lnq^{(t+1)}(\alpha) \propto \mathbb{E}_{q_{-\alpha}^{(t)}}[\ln p(y,\Theta)] \tag{17}$$

$$lnq^{(t+1)}(\alpha) \propto \mathbb{E}_{q_{h}^{(t)}}[\ln p(y|h,\alpha)] + lnp(\alpha) \tag{18}$$

Here,

$$p(y|h,\alpha) = \mathcal{CN}(\boldsymbol{y}_p; \boldsymbol{\Phi}_p \boldsymbol{h}, \alpha^{-1}\boldsymbol{I}) \tag{19}$$

2) 2
3) 3

B. Variational Bayes Interference

IV. Joint Parallel Interference Cancellation Network

In this section, we propose a novel JPICNet framework that enhance the joint PIC scheme with graph neural network. Fig.**??** illustrates the process of the proposed framework. The proposed framework consists of channel parameters estimation (CPE), the GNN-aided symbol detection (GSD) and the GNN-aided channel estimation (GCHE) modules. CPE runs only once to generate initial input for GSD, while GSD and GCHE iteratively exchange outputs in a feedback loop.

A. Channel Parameters Estimation

Estimation area $k \in [0, K-1], l \in [0, lmax]$

Based on the transmission arrangement, we can express the received signal in the channel estimation area as,

$$\begin{aligned} y[k,l] =& h[k-k_p, l-l_{p0}]x_p \\ &+ \sum_{i}^{P} h_i \beta[k,l,k_i,l_i] x_d[[k-k_i]_K, [l-l_i]_L] \\ &+ \tilde{z}[k,l]. \end{aligned} \tag{20}$$

The parameter $\beta[k,l,k_i,l_i]$ has distinct forms for different OTFS waveforms:

i) *Ideal waveform*:

$$\beta_{\text{ideal}}[k,l,k_i,l_i] = e^{-j2\pi\frac{k_i l_i}{KL}}, \tag{21}$$

ii) *Rectangular waveform*:

$$\beta_{\text{rect}}[k,l,k_i,l_i] = \begin{cases} e^{j2\pi\frac{k_i(l-l_i)}{KL}}, & l_i \le l < L, \\ e^{-j2\pi\left(\frac{[k-k_i]_K}{K}\right)} e^{j2\pi\frac{k_i(l-l_i)}{KL}}, & 0 \le l < l_i. \end{cases} \tag{22}$$

From (20), the second term arises from the interference between pilot and data symbols. For simplicity, this cross-interference can be modeled as an additive noise term in subsequent analysis, i.e.,

$$\mathcal{I}[k,l] = \sum_{i}^{P} h_i \beta[k,l,k_i,l_i] x_d[[k+k_i]_K, [l+l_i]_L]. \tag{23}$$

The interference term $\mathcal{I}[k,l]$ is zero-mean ($\mathbb{E}[\mathcal{I}\{[k,l]\} = 0$) with its variance given by,

$$\text{Var}(\mathcal{I}[k,l]) = E_d. \tag{24}$$

Proof: The proof is given in Appendix A. □

Based on the received signal model in (20), we propose a threshold-based channel estimation algorithm as follows. For each delay-Doppler bin (k,l), if the received energy $|y[k,l]|^2$ falls below a predefined threshold $\mathcal{T}$, we declare an empty bin (no path present). Otherwise, we obtain an initial channel estimate and its variance, i.e.,

$$\hat{h}[k-k_p, l-l_{p0}] = y[k,l]/x_p, \tag{25a}$$

$$\hat{hv}[k-k_p, l-l_{p0}] = \frac{E_d+\sigma^2}{E_p}. \tag{25b}$$

Therefore, the threshold provides a mask vector $\mathbf{h}_m$ to indicate whether there is a path along the delay vector $\mathbf{k}$ and Doppler vector $\mathbf{l}$, i.e,

$$\begin{aligned} \mathbf{h}_m =& [h_{m0}, \cdots, h_{mi}, \cdots, h_{mP_{max}}], & (26a)\\ \mathbf{k} =& [-k_{max}, \cdots, 0, \cdots, k_{max}, \cdots, k_i, \cdots, \\ & -k_{max}, \cdots, 0, \cdots, k_{max}], & (26b)\\ \mathbf{l} =& [0, 1, \cdots, l_{max}, \cdots, l_i, \cdots, 0, 1, \cdots, l_{max}]. & (26c) \end{aligned}$$

where $h_{mi}=1$ denotes a path at (k_i,l_i) or $h_{mi}=0$ otherwise, and $P_{max} = (l_{max}+1)(2k_{max}+1)$.

B. GNN-aided Symbol Detection

We need to

At each iteration t, the symbol vector $\boldsymbol{x}$ defined in (3) is treated as a random vector, with its statistical parameters estimated from the observation $\boldsymbol{y}$. based on the same equation. The conditional Gaussian probability distribution function (PDF) is subsequently derived as,

$$p^{(t)}(\mathbf{x}|\mathbf{y}) = \mathcal{N}(\mathbf{x}, \boldsymbol{\mu}_{\mathbf{x}}^{(t)}; \boldsymbol{\Sigma}_{\mathbf{x}}^{(t)}), \tag{27}$$

where

$$\begin{aligned} \boldsymbol{\mu}_{\mathbf{x}}^{(t)} =& (\mathbf{I} \odot \hat{\mathbf{H}}^{(t)^{\mathrm{H}}}\hat{\mathbf{H}}^{(t)})^{-1}(\hat{\mathbf{H}}^{(t)^{\mathrm{H}}}\mathbf{y} - \\ & \mathsf{off}(\hat{\mathbf{H}}^{(t)^{\mathrm{H}}}\hat{\mathbf{H}}^{(t)})\hat{\mathbf{X}}^{(t-1)} - \hat{\mathbf{H}}^{(t)^{\mathrm{H}}}\hat{\mathbf{H}}^{(t)}\mathbf{X}_p), & (28a)\\ \boldsymbol{\Sigma}_{\mathbf{x}}^{(t)} =& (\sigma^2 + \sigma_x^2 \mathbf{D}_{\Delta\hat{\boldsymbol{H}}})(\mathbf{I} \odot \hat{\mathbf{H}}^{(t)^{\mathrm{H}}}\hat{\mathbf{H}}^{(t)})^{-1}. & (28b) \end{aligned}$$

where

$$\mathbf{D}_{\Delta\hat{\mathbf{H}}} = \mathsf{diag}(\sum_{i=0}^{KL-1} \boldsymbol{\Sigma}_{\hat{\mathbf{H}}}[:,i]), \tag{29}$$

$\boldsymbol{\Sigma}_{\hat{\mathbf{H}}}$ denotes the variance matrix of $\hat{\mathbf{H}}$, and $\sum_{i=0}^{KL-1}\boldsymbol{\Sigma}_{\hat{\mathbf{H}}}[:,i]$ means we sum all columns of $\boldsymbol{\Sigma}_{\hat{\mathbf{H}}}$ together to build a column vector.

Proof: The proof is given in Appendix B. □

C. GNN-aided Channel Estimation

Rewriting (3) for channel estimation, we have

$$\mathbf{y} = \boldsymbol{\Phi}\mathbf{h} + \mathbf{z}, \tag{30}$$

where $\mathbf{h}$ is the path gain in the time domain,

$$\boldsymbol{\Phi} = \begin{bmatrix} \Phi_{k_{m0},l_{m0}}(1,1) & \cdots & \Phi_{k_{mP_{max}},l_{k_{mP_{max}}}}(KL,P_{max}) \\ \vdots & \ddots & \vdots \\ \Phi_{k_{m0},l_{m0}}(KL,1) & \cdots & \Phi_{k_{mP_{max}},l_{k_{mP_{max}}}}(KL,P_{max}) \end{bmatrix},$$

where $(kL+l,i)$th entry of $\mathbf{\Phi} \in \mathcal{C}^{(KL-1)\times P_{max}}$ is [?],

$$\Phi_{k_{mi},l_{mi}}(kL+l,i) = X([k-k_{mi}]_K,[l - l_{mi}]_L)\beta[k,l,k_{mi},l_{mi}]. \tag{31}$$

As $\mathbf{\Phi}$ is unavailable, its estimate at the t-th iteration is employed for channel estimation, i.e.,

$$\begin{aligned}\mathbf{y} &= \mathbf{\Phi}^{(t)}\mathbf{h} + \mathbf{\Delta}\Phi^{(t)}\mathbf{h} + \mathbf{z}, \\ &= \mathbf{\Phi}^{(t)}\mathbf{h} + \mathbf{z}_2,\end{aligned} \tag{32}$$

where $\mathbf{\Phi} = \mathbf{\Phi}^{(t)} + \mathbf{\Delta}\Phi^{(t)}$ and $\mathbf{z}_2 = \mathbf{\Delta}\Phi^{(t)}\mathbf{h} + \mathbf{z}$ is assumed to be the effective noise.

Inspired by [? ?], we implement a linear minimum-mean-squared error (MMSE) channel estimation , i.e.,

$$\hat{\mathbf{h}}^{(t)} = \mathbf{R}_{\mathbf{h}}\mathbf{\Phi}^{(t)^{\mathrm{H}}}(\mathbf{\Phi}^{(t)}\mathbf{R}_{\mathbf{h}}\mathbf{\Phi}^{(t)^{\mathrm{H}}} + \frac{1}{P}\mathsf{diag}(\mathbf{h}_m)\mathbf{D}_{\Delta\mathbf{\Phi}^{(t)}} + \sigma^2\mathbf{I})^{-1}\mathbf{y}, \tag{33a}$$

$$\mathsf{Var}(\hat{\mathbf{h}}^{(t)}) = \mathbf{R}_{\mathbf{h}} - \mathbf{R}_{\mathbf{h}}\mathbf{\Phi}^{(t)^{\mathrm{H}}}(\mathbf{\Phi}^{(t)}\mathbf{R}_{\mathbf{h}}\mathbf{\Phi}^{(t)^{\mathrm{H}}} + \frac{1}{P}\mathsf{diag}(\mathbf{h}_m)\mathbf{D}_{\Delta\mathbf{\Phi}^{(t)}} + \sigma^2\mathbf{I}))^{-1}\mathbf{\Phi}^{(t)}\mathbf{R}_{\mathbf{h}}, \tag{33b}$$

where $\mathbf{\Phi}^{(t)}$ is the estimation of $\mathbf{\Phi}$ in t-th iteration, and,

$$\mathbf{D}_{\Delta\mathbf{\Phi}^{(t)}} = \mathsf{diag}(\sum_{i=0}^{KL-1}\mathbf{\Sigma}_{\Delta\mathbf{\Phi}^{(t)}}[:,i]). \tag{34}$$

$\sum_{i=0}^{KL-1}\mathbf{\Sigma}_{\Delta\mathbf{\Phi}^{(t)}}[:,i]$ denotes the row-wise sum of $\mathbf{\Sigma}_{\Delta\mathbf{\Phi}^{(t)}}$ as a column vector, and $\mathbf{\Sigma}_{\Delta\mathbf{\Phi}^{(t)}}$ denotes the variance matrix of $\mathbf{\Phi}^{(t)}$.

Proof: The proof is given in Appendix C. □

V. Simulation Results

VI. Conclusion

VII. Acknowledgment

This research was supported by the research training program stipend from The University of Sydney. The work of Branka Vucetic was supported in part by the Australian Research Council Laureate Fellowship grant number FL160100032.

Appendix A
CPE Cross-Interference

$$\mathbb{E}\{\mathcal{I}[k,l]\} = \sum_{i=1}^{P}\underbrace{\mathbb{E}\{h_i\}}_{0}\beta[k,l,k_i,l_i] \underbrace{\mathbb{E}\{x_d[[k+k_i]_K,[l+l_i]_L]\}}_{0} = 0 \tag{35}$$

$$\mathsf{Var}(\mathcal{I}[k,l]) = \mathbb{E}\left\{|\mathcal{I}[k,l]|^2\right\}, \tag{36}$$

where

$$\begin{aligned}|\mathcal{I}[k,l]|^2 = &\sum_{i=1}^{P}|h_i|^2|\beta[k,l,k_i,l_i]|^2|x_d[[k+k_i]_K,[l+l_i]_L]|^2 \\ &+ \sum_{i\neq j}h_ih_j^*\beta[k,l,k_i,l_i]\beta[k,l,k_j,l_j]^* \cdot \\ &x_d[[k+k_i]_K,[l+l_i]_L]x_d[[k+k_j]_K,[l+l_j]_L]^*\end{aligned} \tag{37}$$

The expectation of the second term in (37) vanishes because $\mathbb{E}\{h_ih_j^*\}=0$, i.e.,

$$\begin{aligned}\mathsf{Var}(\mathcal{I}[k,l]) &= \sum_{i=1}^{P}\underbrace{|h_i|^2}_{1/P}\cdot\underbrace{|\beta[k,l,k_i,l_i]|^2}_{1}\cdot \\ &\quad \underbrace{|x_d[[k+k_i]_K,[l+l_i]_L]|^2}_{E_d} \\ &= \sum_{i=1}^{P}1/P\cdot E_d = E_d\end{aligned} \tag{38}$$

Appendix B
GSD Variance

For simplification, we denote $\hat{\mathbf{H}}^{(t)}$ as $\hat{\mathbf{H}}$ and the power of $\mathbf{x}$ as σ_x^2. Therefore, the estimated symbol is

$$\begin{aligned}\boldsymbol{\mu}_{\mathbf{x}} &= (\mathbf{I}\odot\hat{\mathbf{H}}^{\mathrm{H}}\hat{\mathbf{H}})^{-1}\hat{\mathbf{H}}^{\mathrm{H}}\mathbf{y}, \\ &= (\mathbf{I}\odot\hat{\mathbf{H}}^{\mathrm{H}}\hat{\mathbf{H}})^{-1}\hat{\mathbf{H}}^{\mathrm{H}}(\mathbf{H}\mathbf{x}+\mathbf{z}), \\ &= \underbrace{(\mathbf{I}\odot\hat{\mathbf{H}}^{\mathrm{H}}\hat{\mathbf{H}})^{-1}\hat{\mathbf{H}}^{\mathrm{H}}}_{\mathbf{W}_{GSD}}\times \\ &\quad (\hat{\mathbf{H}}\mathbf{x}+\mathbf{H}\mathbf{x}-\hat{\mathbf{H}}\mathbf{x}+\mathbf{z}), \\ &= \mathbf{W}_{GSD}(\hat{\mathbf{H}}\mathbf{x}+\Delta\hat{\mathbf{H}}\mathbf{x}+\mathbf{z}), \\ &= \mathbf{W}_{GSD}\hat{\mathbf{H}}\mathbf{x}+\mathbf{W}_{GSD}\Delta\mathbf{H}\mathbf{x}+\mathbf{W}_{GSD}\mathbf{z}.\end{aligned} \tag{39}$$

The estimation error is

$$\begin{aligned}\mathbf{e} &= \boldsymbol{\mu}_{\mathbf{x}} - \mathbf{x}, \\ &= (\mathbf{W}_{GSD}\hat{\mathbf{H}}-\mathbf{I})\mathbf{x}+\mathbf{W}_{GSD}\Delta\mathbf{H}\mathbf{x}+\mathbf{W}_{GSD}\mathbf{z}.\end{aligned} \tag{40}$$

Please note $\hat{\mathbf{H}}$ is the estimation of $\mathbf{H}$, i.e., $\mathbb{E}\{\Delta\hat{\mathbf{H}}\}=0$ and $\mathbb{E}\{\mathbf{z}\}=0$. Therefore, the corvariance of $\boldsymbol{\mu}_{\mathbf{x}}$ is,

$$\begin{aligned}\mathsf{Cov}(\boldsymbol{\mu}_{\mathbf{x}}) &= \mathbb{E}\{\mathbf{e}\mathbf{e}^{\mathrm{H}}\}, \\ &= \mathbb{E}\{(\mathbf{W}_{GSD}\hat{\mathbf{H}}-\mathbf{I})\mathbf{x}\mathbf{x}^{\mathrm{H}}(\mathbf{W}_{GSD}\hat{\mathbf{H}}-\mathbf{I})^{\mathrm{H}}\} \\ &\quad + \mathbb{E}\{\mathbf{W}_{GSD}\Delta\hat{\mathbf{H}}\mathbf{x}\mathbf{x}^{\mathrm{H}}\Delta\hat{\mathbf{H}}^{\mathrm{H}}\mathbf{W}_{GSD}^{H}\} \\ &\quad + \mathbb{E}\{\mathbf{W}_{GSD}\mathbf{z}\mathbf{z}^{\mathrm{H}}\mathbf{W}_{GSD}^{\mathrm{H}}\}, \\ &= \sigma_x^2\mathbb{E}\{(\mathbf{W}_{GSD}\hat{\mathbf{H}}-\mathbf{I})(\mathbf{W}_{GSD}\hat{\mathbf{H}}-\mathbf{I})^{\mathrm{H}}\} \\ &\quad + \sigma_x^2\mathbb{E}\{\mathbf{W}_{GSD}\Delta\hat{\mathbf{H}}\Delta\hat{\mathbf{H}}^{\mathrm{H}}\mathbf{W}_{GSD}^{\mathrm{H}}\} \\ &\quad + \sigma^2\mathbb{E}\{\mathbf{W}_{GSD}\mathbf{W}_{GSD}^{\mathrm{H}}\}.\end{aligned} \tag{41}$$

Since the estimated channel coefficients are assumed to be mutually independent, the first term in (41) can thus be simplified as,

$$
\begin{aligned}
&\sigma_x^2\mathbb{E}\{(\mathbf{W}_{GSD}\hat{\mathbf{H}}-\mathbf{I})(\mathbf{W}_{GSD}\hat{\mathbf{H}}-\mathbf{I})^{\mathrm{H}}\} \\
&=\sigma_x^2\mathbb{E}\{\mathbf{W}_{GSD}\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathrm{H}}\mathbf{W}_{GSD}^{\mathrm{H}}-\mathbf{I}\}, \\
&=\sigma_x^2\mathbb{E}\{(\mathbf{I}\odot\hat{\mathbf{H}}^{\mathrm{H}}\hat{\mathbf{H}})^{-1}\hat{\mathbf{H}}^{\mathrm{H}}\hat{\mathbf{H}}\hat{\mathbf{H}}^{\mathrm{H}}\hat{\mathbf{H}}(\mathbf{I}\odot\hat{\mathbf{H}}^{\mathrm{H}}\hat{\mathbf{H}})^{-1} \\
&\quad -\mathbf{I}\}, \\
&=\sigma_x^2\mathbb{E}\{\mathbf{I}-\mathbf{I}\}=\mathbf{0}.
\end{aligned} \tag{42}
$$

Similarly, the last term in (41) can be simplified as,

$$
\begin{aligned}
&\sigma^2\mathbb{E}\{\mathbf{W}_{GSD}\mathbf{W}_{GSD}^{\mathrm{H}}\} \\
&=\sigma^2\mathbb{E}\{(\mathbf{I}\odot\hat{\mathbf{H}}^{\mathrm{H}}\hat{\mathbf{H}})^{-1}\hat{\mathbf{H}}^{\mathrm{H}}\hat{\mathbf{H}}(\mathbf{I}\odot\hat{\mathbf{H}}^{\mathrm{H}}\hat{\mathbf{H}})^{-1}\}, \\
&=\sigma^2(\mathbf{I}\odot\hat{\mathbf{H}}^{\mathrm{H}}\hat{\mathbf{H}})^{-1}.
\end{aligned} \tag{43}
$$

Therefore, (41) can be written as,

$$
\begin{aligned}
\mathrm{Cov}(\boldsymbol{\mu}_{\mathbf{x}})&=\sigma_x^2\mathbb{E}\{\mathbf{W}_{GSD}\mathbf{D}_{\Delta\hat{\mathbf{H}}}\mathbf{W}_{GSD}^{\mathrm{H}}\} \\
&\quad+\sigma^2(\mathbf{I}\odot\hat{\mathbf{H}}^{\mathrm{H}}\hat{\mathbf{H}})^{-1},
\end{aligned} \tag{44}
$$

The variance is the diagonal of $\mathrm{Cov}(\boldsymbol{\mu}_{\mathbf{x}})$, i.e.,

$$
\begin{aligned}
\mathrm{Var}(\boldsymbol{\mu}_{\mathbf{x}i})=\sigma^2&\sum_{j=0}^{KL-1}|\hat{H}[j,i]|^{-2}+ \\
&\sigma_x^2\sum_{j=0}^{KL-1}|\hat{H}[j,i]|^{-2}\sum_{j'=0}^{KL-1}\sigma_{\hat{H}}^2[i,j'],
\end{aligned} \tag{45}
$$

where $\sigma_{\hat{H}}^2[j,i']$ is the (j,i')th entry of $\boldsymbol{\Sigma}_{\hat{\mathbf{H}}}$.

Appendix C
GCHE

In t-th iteration, (30) can be written as,

$$
\begin{aligned}
\mathbf{y}&=(\boldsymbol{\Phi}^{(t)}+\boldsymbol{\Phi}-\boldsymbol{\Phi}^{(t)})\mathbf{h}+\mathbf{z}, \\
&=\boldsymbol{\Phi}^{(t)}\mathbf{h}+\Delta\boldsymbol{\Phi}^{(t)}\mathbf{h}+\mathbf{z}, \\
&=\boldsymbol{\Phi}^{(t)}\mathbf{h}+\mathbf{z}^{(t)},
\end{aligned} \tag{46}
$$

where $\mathbb{E}\{\Delta\boldsymbol{\Phi}^{(t)}\}=0$. The mean of $\mathbf{z}^{(t)}$ is,

$$
\begin{aligned}
\mathbb{E}\{\mathbf{z}^{(t)}\}&=\mathbb{E}\{\Delta\boldsymbol{\Phi}^{(t)}\mathbf{h}\}+\mathbb{E}\{\mathbf{z}\}, \\
&=\Delta\boldsymbol{\Phi}^{(t)}\mathbb{E}\{\mathbf{h}\}+\mathbb{E}\{\mathbf{z}\}, \\
&=\mathbf{0}.
\end{aligned} \tag{47}
$$

The covariance of $\mathbf{z}^{(t)}$ is,

$$
\begin{aligned}
\mathrm{Cov}(\mathbf{z}^{(t)})&=\mathbf{R}_{\mathbf{z}^{(t)}}, \\
&=\mathbf{R}_{\Delta\boldsymbol{\Phi}^{(t)}\mathbf{h}}+\mathbf{R}_{\mathbf{z}}, \\
&=\mathbb{E}\{\Delta\boldsymbol{\Phi}^{(t)}\mathbf{h}\mathbf{h}^{\mathrm{H}}\Delta\boldsymbol{\Phi}^{(t)^{\mathrm{H}}}\}+\sigma^2\mathbf{I}, \\
&=\frac{1}{P}\mathbb{E}\{\Delta\boldsymbol{\Phi}^{(t)}\mathsf{diag}(\mathbf{h}_m)\Delta\boldsymbol{\Phi}^{(t)^{\mathrm{H}}}\}, \\
&=\frac{1}{P}\mathsf{diag}(\mathbf{h}_m)\mathbf{D}_{\Delta\Phi^{(t)}}.
\end{aligned} \tag{48}
$$

Therefore, the linear MMSE estimation is,

$$
\begin{aligned}
\hat{\mathbf{h}}^{(t)}=\mathbf{R}_{\mathbf{h}}\boldsymbol{\Phi}^{(t)^H}(\boldsymbol{\Phi}^{(t)}\mathbf{R}_{\mathbf{h}}\boldsymbol{\Phi}^{(t)^H}+\frac{1}{P}\mathsf{diag}(\mathbf{h}_m)\mathbf{D}_{\Delta\Phi^{(t)}} \\
+\sigma^2\mathbf{I})^{-1}\mathbf{y}.
\end{aligned} \tag{49}
$$

The covariance of the estimation error is,

$$
\begin{aligned}
\mathrm{Var}(\mathbf{e})&=\mathbb{E}\{(\mathbf{h}-\hat{\mathbf{h}}^{(t)})(\mathbf{h}-\hat{\mathbf{h}}^{(t)})^H\}, \\
&=\mathbb{E}\{\mathbf{h}\mathbf{h}^H\}-\mathbb{E}\{\hat{\mathbf{h}}\hat{\mathbf{h}}^H\}, \\
&=\mathbf{R}_{\mathbf{h}}-\mathbf{R}_{\mathbf{h}}\boldsymbol{\Phi}^{(t)^H}\mathbf{R}_{\mathbf{y}}^{-1}\mathbf{R}_{\mathbf{y}}\mathbf{R}_{\mathbf{y}}^{-1}\boldsymbol{\Phi}^{(t)}\mathbf{R}_{\mathbf{h}}, \\
&=\mathbf{R}_{\mathbf{h}}-\mathbf{R}_{\mathbf{h}}\boldsymbol{\Phi}^{(t)^H}\mathbf{R}_{\mathbf{y}}^{-1}\boldsymbol{\Phi}^{(t)}\mathbf{R}_{\mathbf{h}}, \\
&=\mathbf{R}_{\mathbf{h}}-\mathbf{R}_{\mathbf{h}}\boldsymbol{\Phi}^{(t)^H}(\boldsymbol{\Phi}^{(t)}\mathbf{R}_{\mathbf{h}}\boldsymbol{\Phi}^{(t)^H}+ \\
&\quad\frac{1}{P}\mathsf{diag}(\mathbf{h}_m)\mathbf{D}_{\Delta\boldsymbol{\Phi}^{(t)}}+\sigma^2\mathbf{I}))^{-1}\boldsymbol{\Phi}^{(t)}\mathbf{R}_{\mathbf{h}},
\end{aligned} \tag{50}
$$