OFDM Phase Noise Cancellation via Approximate Probabilistic Inference

Darryl Dexu Lin, Yi Zhao and Teng Joon Lim

University of Toronto

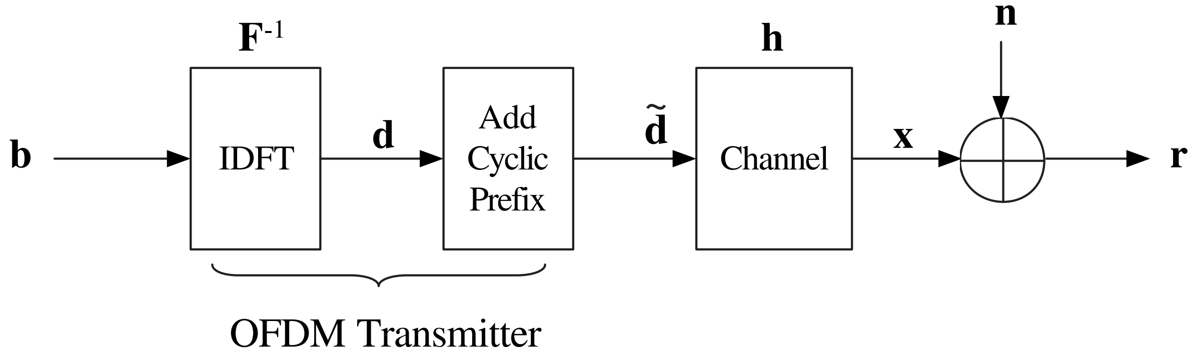
March 17, 2005

Outline

- System model and problem description.
- Conventional schemes and exact inference.
- Variational inference for phase noise (PHN) cancellation.
- Summary.

I. System Model: OFDM without PHN

- Discrete OFDM channel model with input $\mathbf{b} \in \mathbb{C}^{N\times 1}$ and output $\mathbf{r} \in \mathbb{C}^{N\times 1}$, sampled at $f_s = N/T$:



where N is the number of subcarriers and T is the OFDM symbol period.

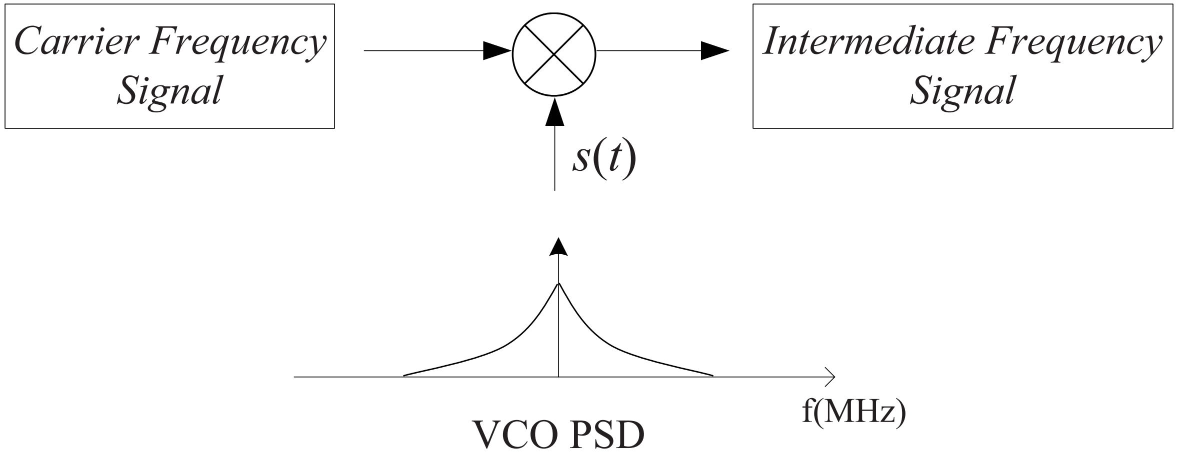
- Defining $\mathbf{H}$ to be circular convolution matrix, the mathematical model is:

$$\begin{aligned} \mathbf{r} &= \mathbf{h} * \tilde{\mathbf{d}} + \mathbf{n} = \mathbf{h} \otimes \mathbf{d} + \mathbf{n} \\ &= \mathbf{H}\mathbf{d} + \mathbf{n} = \mathbf{x} + \mathbf{n}. \end{aligned} \tag{1}$$

I. System Model: Source of PHN

- Voltage-controlled oscillator (VCO) converts the received signal from carrier frequency (CF) to intermediate frequency (IF).

- Small random phase variation is introduced by VCO at the receiver during frequency down-conversion.

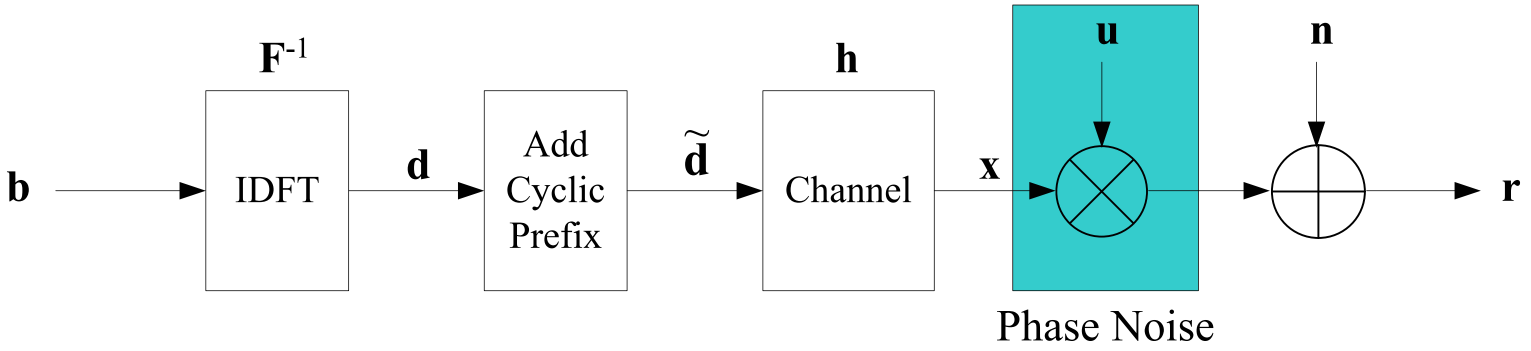


- VCO output with PHN: $s(t) = e^{j(2\pi f_o t + \theta(t))}$.

$$
\begin{aligned}
\Longrightarrow R_s(\tau) &\approx e^{-j2\pi f_o \tau} + R_\theta(\tau) e^{-j2\pi f_o \tau} \\
\Longrightarrow S_s(f) &\approx \delta(f - f_o) + S_\theta(f - f_o).
\end{aligned} \tag{1}
$$

I. System Model: OFDM with PHN

- Discrete OFDM channel model with phase noise (PHN):



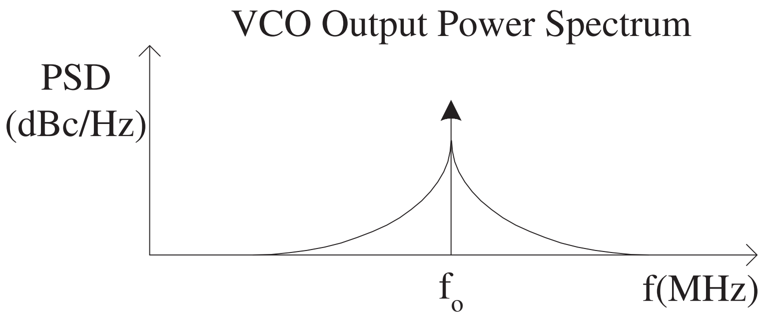
- $\mathbf{u} = [\exp(j\theta_1), \cdots, \exp(j\theta_N)]^T$ is the time domain PHN pattern.

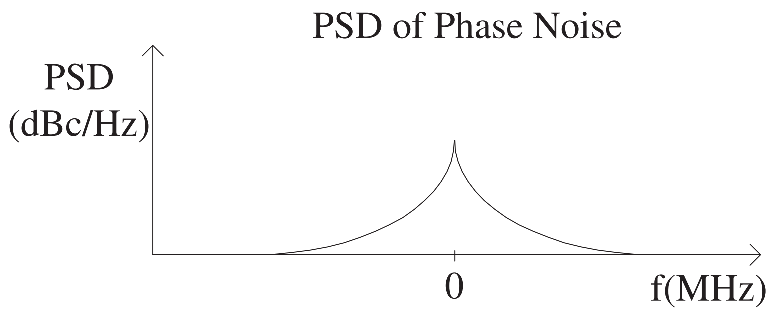
- Assuming small $\{\theta_n\}_{n=1}^N$, then $\mathbf{u} \approx \mathbf{1} + j\boldsymbol{\theta}$.Therefore,

$$\mathbf{r} = \mathbf{x} \circ \mathbf{u} + \mathbf{n} \approx \mathbf{x} \circ (\mathbf{1} + j\boldsymbol{\theta}) + \mathbf{n}. \tag{1}$$

I. System Model: Statistics of PHN

- What do we know about PHN?





- Since $S_s(f) = \delta(f - f_o) + S_\theta(f - f_o)$, we may extract $S_\theta(f)$ from the measured PSD.

- The autocorrelation function of $\theta(t)$, $R_\theta(\tau) = \mathcal{F}^{-1}(S_\theta(f))$.

- Assuming the sampled $\theta(t)$, $\boldsymbol{\theta}$, is a zero mean Gaussian process with distribution $\mathcal{N}(\mathbf{0}, \mathbf{\Phi})$, we may calculate $\mathbf{\Phi}$ as:

$$\mathbf{\Phi}_{i,j} = R_\theta\left(\mid i - j \mid \frac{T}{N} \right). \tag{1}$$

- System model and problem description.
- **Conventional schemes and exact inference.**
- Variational inference for phase noise (PHN) estimation.
- Summary.

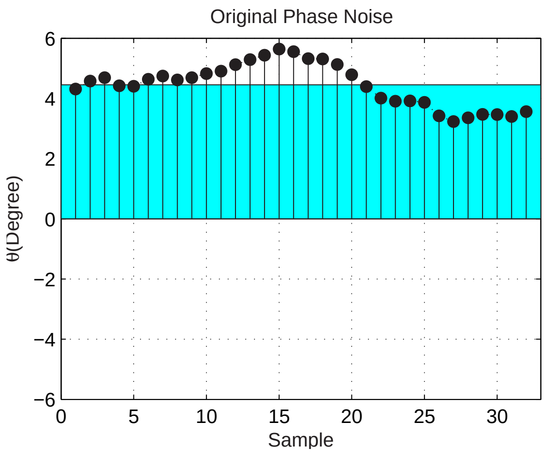
II. Conventional Scheme: Common PHN Removal

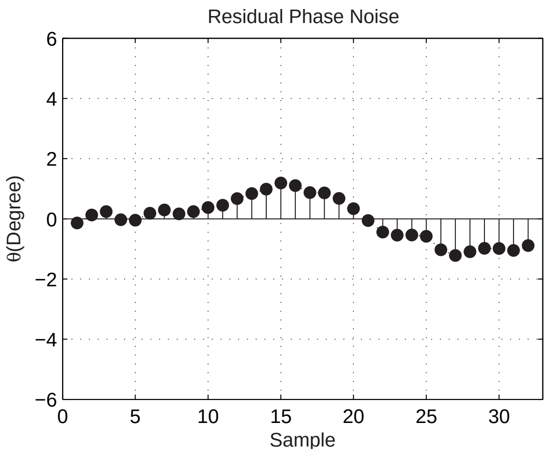
- Writing $\theta(t) = \theta_o + \theta_{random}(t)$ and assuming $\theta_{random}(t)$ is small, the conventional scheme focuses on removing θ_o, the *common PHN*.

- Ignoring $\theta_{random}(t)$, the system model becomes:

$$\mathbf{r} = \mathbf{x} \circ \mathbf{u} + \mathbf{n} \approx e^{j\theta_o}\mathbf{x} + \mathbf{n}. \tag{1}$$

Given perfect knowledge of $\mathbf{x} = \mathbf{Hd}$ (using pilot symbols or tentative data decisions), θ_o can be accurately estimated.

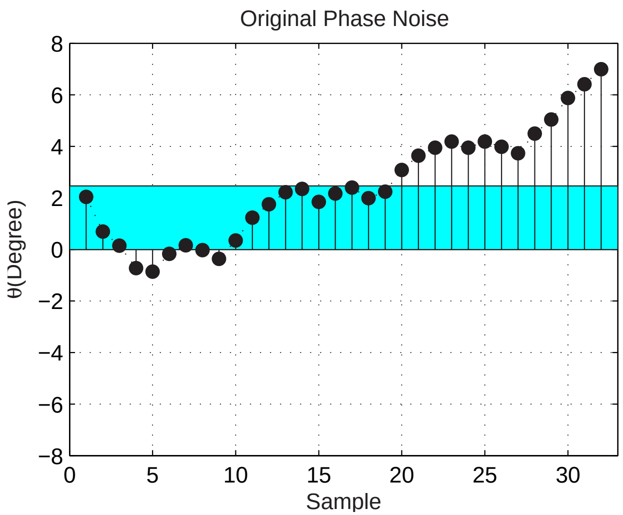


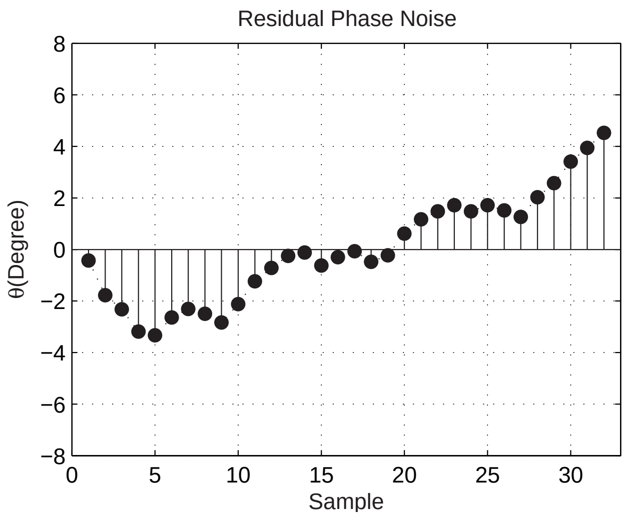


II. Conventional Scheme: Common PHN Removal

- Small $\theta_{random}(t)$ assumption is only realistic for good quality, low-bandwidth VCO, where $\theta(t)$ is a highly correlated process.

- In realistic scenarios, knowing θ_o is not sufficient. More sophisticated algorithm needs to be developed to estimate $\theta(t)$, or $\boldsymbol{\theta}$.





II. Exact Inference: Brute Force

- Consider the probabilistic model of $\mathbf{r} \approx \mathbf{x} \circ (\mathbf{1} + j\boldsymbol{\theta}) + \mathbf{n}$, where $\mathbf{x} = \mathbf{Hd}$.

- Assuming $p(\mathbf{d}) = \mathcal{CN}(\mathbf{0}, 2\rho^2\mathbf{I})$, we have $p(\mathbf{x}) = \mathcal{CN}(\mathbf{0}, 2\rho^2\mathbf{HH}^H)$.

- The joint distribution of the input and output is:

$$p(\mathbf{x}, \boldsymbol{\theta}, \mathbf{r}) = p(\mathbf{x})p(\boldsymbol{\theta})p(\mathbf{r}|\mathbf{x}, \boldsymbol{\theta}). \tag{1}$$

- The posterior estimate of $\mathbf{x}$, $p(\mathbf{x}|\mathbf{r})$ can be computed by eliminating $\boldsymbol{\theta}$:

$$\begin{aligned} p(\mathbf{x}|\mathbf{r}) &\propto p(\mathbf{x}, \mathbf{r}) = \int_{\boldsymbol{\theta}} p(\mathbf{x}, \boldsymbol{\theta}, \mathbf{r}) d\boldsymbol{\theta} \\ &= \mathcal{CN}(\mathbf{r}; \mathbf{x}, \mathbf{diag}(\mathbf{x})\boldsymbol{\Phi}\, \mathbf{diag}(\mathbf{x})^H + 2\sigma^2\mathbf{I}) \cdot \mathcal{CN}(\mathbf{x}; \mathbf{0}, 2\rho^2\mathbf{HH}^H). \end{aligned} \tag{2}$$

- Yet the optimal estimate $\hat{\mathbf{x}} = \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathbf{r})$ cannot be easily extracted due to the complicated expression.

- System model and problem description.
- Conventional schemes and exact inference.
- **Variational inference for phase noise (PHN) estimation.**
- Summary.

IV. Variational Inference: Motivation

- We have the joint posterior $p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{r})$, but it is too complex for us to extract optimal estimates $\hat{\mathbf{x}}$ and $\hat{\boldsymbol{\theta}}$!

- **Solution**: We can try to find a more manageable parameterized approximation, say $Q(\mathbf{x}, \boldsymbol{\theta}|\mathbf{r})$ (or written as $Q(\mathbf{x}, \boldsymbol{\theta})$ for simplicity).

- The *Kullback-Leibler* divergence $D[p(a)\|p(b)]$ offers a measure of similarity between two distributions. Thus we are looking for

$$\hat{Q}(\mathbf{x}, \boldsymbol{\theta}) = \arg\min_{Q} D[Q(\mathbf{x}, \boldsymbol{\theta})\|p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{r})]. \tag{1}$$

Or equivalently, $\hat{Q}(\mathbf{x}, \boldsymbol{\theta}) = \arg\min_Q D[Q(\mathbf{x}, \boldsymbol{\theta})\|p(\mathbf{x}, \boldsymbol{\theta}, \mathbf{r})]$, where $p(\mathbf{x}, \boldsymbol{\theta}, \mathbf{r})$ is the complete likelihood function we already have.

IV. Variational Inference: Free Energy

- We define $D[Q(\mathbf{x},\boldsymbol{\theta})\|p(\mathbf{x},\boldsymbol{\theta},\mathbf{r})]$ to be the *Variational Free Energy*, i.e.

$$\mathcal{F}(Q,p) = \int_{\mathbf{x},\boldsymbol{\theta}} Q(\mathbf{x},\boldsymbol{\theta}) \ \log \frac{Q(\mathbf{x},\boldsymbol{\theta})}{p(\mathbf{x},\boldsymbol{\theta},\mathbf{r})} d\mathbf{x} d\boldsymbol{\theta}. \tag{1}$$

- If no assumptions are made, we simply get $Q(\mathbf{x},\boldsymbol{\theta}) = p(\mathbf{x},\boldsymbol{\theta}|\mathbf{r})$.

- Introduce two convenient approximations to parameterize $Q(\mathbf{x},\boldsymbol{\theta})$:
  - **Mean-Field Approximation**

$$Q(\mathbf{x},\boldsymbol{\theta}) = Q(\mathbf{x})Q(\boldsymbol{\theta}) \tag{2}$$

  - **Gaussian Approximation**

$$Q(\mathbf{x}) = \mathcal{CN}(\mathbf{m}_x, \mathbf{S}_x); \quad Q(\boldsymbol{\theta}) = \mathcal{N}(\mathbf{m}_\theta, \mathbf{S}_\theta) \tag{3}$$

IV. Variational Inference: Parameter Update

- $\mathcal{F}(Q,p)$ can be obtained as

$$\begin{aligned}\mathcal{F} = \quad & \mathrm{tr}(\mathbf{\Sigma}^{-1}\mathbf{S}_x) + \mathbf{m}_x^H\mathbf{\Sigma}^{-1}\mathbf{m}_x + \tfrac{1}{2}\mathrm{tr}(\mathbf{\Phi}^{-1}\mathbf{S}_\theta) + \tfrac{1}{2}\mathbf{m}_\theta^T\mathbf{\Phi}^{-1}\mathbf{m}_\theta - \log|\mathbf{S}_\theta|^{\frac{1}{2}} - \log|\mathbf{S}_x| \\ & + \tfrac{1}{2\sigma^2}\{\mathrm{tr}[\mathbf{diag}(\mathbf{S}_\theta)\mathbf{S}_x] + \mathbf{m}_x^H\,\mathbf{diag}(\mathbf{S}_\theta)\mathbf{m}_x + \mathrm{tr}[(j\mathbf{M}_\theta + \mathbf{I})\mathbf{S}_x(j\mathbf{M}_\theta + \mathbf{I})^H] \\ & + [(j\mathbf{M}_\theta + \mathbf{I})\mathbf{m}_x - \mathbf{r}]^H[(j\mathbf{M}_\theta + \mathbf{I})\mathbf{m}_x - \mathbf{r}])\}, \end{aligned} \tag{1}$$

where $\mathbf{M}_\theta = \mathbf{diag}(\mathbf{m}_\theta)$, $\mathbf{\Sigma} = 2\rho^2\mathbf{H}\mathbf{H}^H$.

- The approximate posterior $Q(\mathbf{x})$ and $Q(\boldsymbol{\theta})$ may be found by minimizing $\mathcal{F}(Q,p)$ w.r.t. the parameters of $Q(\mathbf{x})$ and $Q(\boldsymbol{\theta})$.

- Since the optimal values of $\mathbf{m}_x, \mathbf{S}_x, \mathbf{m}_\theta$, and $\mathbf{S}_\theta$ are coupled, we take the coordinate descent approach by solving for the optimal parameters iteratively. For example, in the t-th iteration,

$$\hat{\mathbf{m}}_x^{(t)} = \arg\min_{\mathbf{m}_x} \mathcal{F}(\mathbf{m}_x, \hat{\mathbf{S}}_x^{(t-1)}, \hat{\mathbf{m}}_\theta^{(t-1)}, \hat{\mathbf{S}}_\theta^{(t-1)}) \tag{2}$$

IV. Variational Algorithm for PHN Cancellation

- In the t-th iteration, the parameter update equations are:

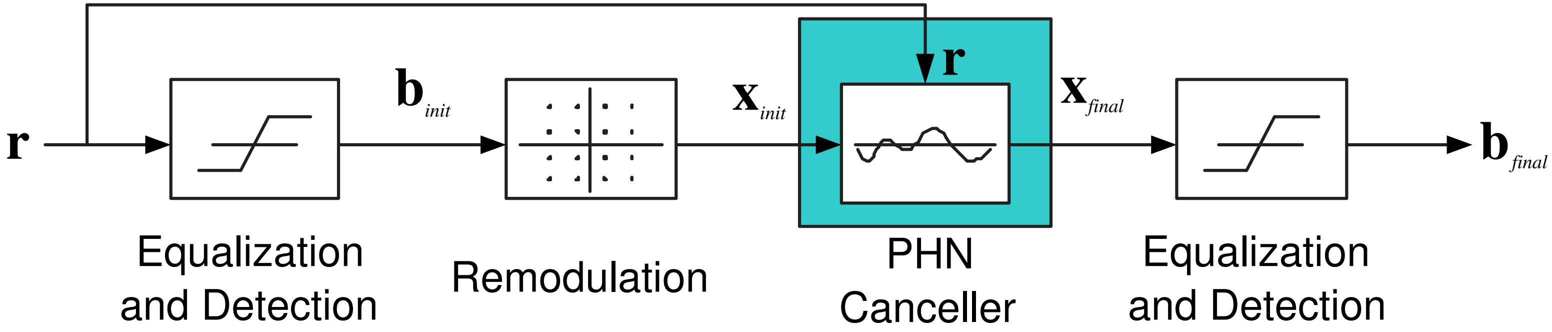
$$
\begin{aligned}
\mathbf{S}_\theta^{(t)} &= \sigma^2[\sigma^2\mathbf{\Phi}^{-1} + \mathbf{diag}(\mathbf{S}_x^{(t-1)}) + \mathbf{M}_x^{H(t-1)}\mathbf{M}_x^{(t-1)}]^{-1}; \\
\mathbf{m}_\theta^{(t)} &= \sigma^{-2}\mathbf{S}_\theta^{(t-1)} \cdot \mathrm{Re}[j\mathbf{M}_x^{H(t-1)}(\mathbf{m}_x^{(t-1)} - \mathbf{r}^*)]; \\
\mathbf{S}_x^{(t)} &= 2\sigma^2[\tfrac{\sigma^2}{\rho^2}(\mathbf{H}\mathbf{H}^H)^{-1} + \mathbf{diag}(\mathbf{S}_\theta^{(t-1)}) \\
&\quad +(j\mathbf{M}_\theta^{(t-1)} + \mathbf{I})^H(j\mathbf{M}_\theta^{(t-1)} + \mathbf{I})]^{-1}; \\
\mathbf{m}_x^{(t)} &= \tfrac{1}{2}\sigma^{-2}\mathbf{S}_x^{(t-1)}(j\mathbf{M}_\theta^{(t-1)} + \mathbf{I})^H\mathbf{r}.
\end{aligned}
\tag{1}
$$

- But to initialize the iteration, we need $\mathbf{m}_x^{(0)}$ and $\mathbf{S}_x^{(0)}$. Random initialization traps the algorithm into a poor local minimum.

- We may set $\mathbf{S}_x^{(0)} = \mathbf{0}$ and let $\mathbf{m}_x^{(0)} = \mathbf{x}_{init} = \mathbf{H}\mathbf{d}_{init}$, where $\mathbf{d}_{init}$ is a tentative symbol decision, obtained e.g. by ignoring PHN.

IV. Variational PHN Cancellation: Receiver Structure

- The iterative update of the parameters monotonically decreases $\mathcal{F}(Q, p)$.



- The PHN canceller takes the initial estimate $\mathbf{x}_{init}$ and produces a better estimate $\mathbf{x}_{final}$.

- $\mathbf{x}_{final}$ is now ideally clean of PHN distortion and can be used for final equalization and detection.

IV. ICM PHN Cancellation: A Simplified Alternative

- *Iterative Conditional Mode* (ICM) is a special case of variational inference with the Q-functions parameterized as δ functions instead of Gaussian. i.e.

$$Q(\mathbf{x}) = \delta(\mathbf{x}, \hat{\mathbf{x}}); \quad Q(\boldsymbol{\theta}) = \delta(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}) \tag{1}$$

- It can be shown that the minimization of $\mathcal{F}(Q, p)$ is equivalent to the maximization of $p(\mathbf{x}, \boldsymbol{\theta}, \mathbf{r})$ over $\mathbf{x}$ and $\boldsymbol{\theta}$.

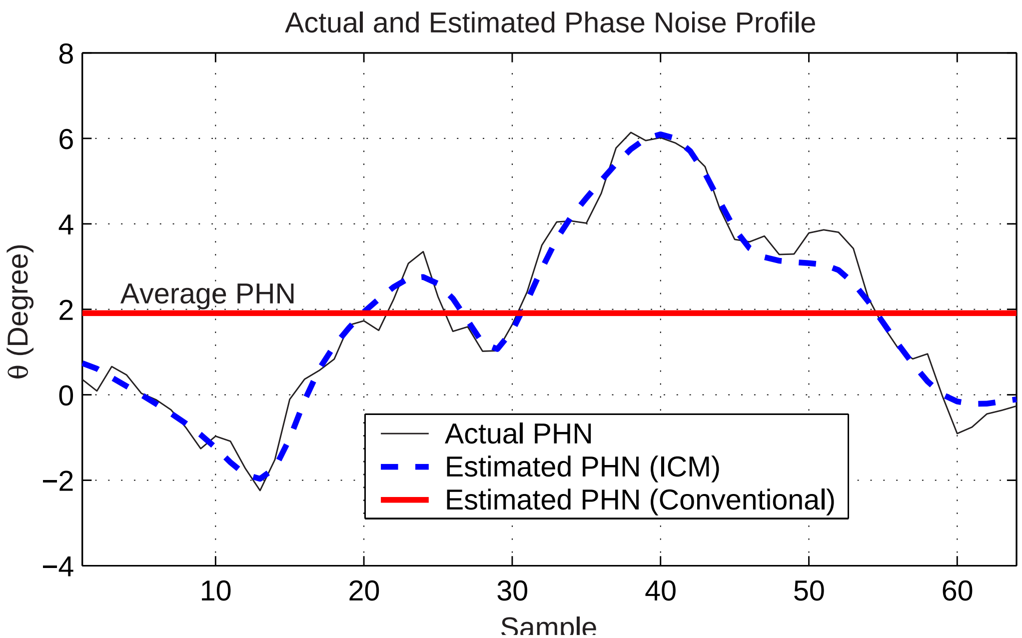
- In the t-th iteration, the parameter update equations are:

$$\begin{aligned} \boldsymbol{\theta}^{(t)} &= [\sigma^2 \boldsymbol{\Phi}^{-1} + \mathbf{X}^{H(t-1)} \mathbf{X}^{(t-1)}]^{-1} \cdot \mathrm{Re}[j \mathbf{X}^{H(t-1)} (\mathbf{x}^{(t-1)} - \mathbf{r})] \\ \mathbf{x}^{(t)} &= [\tfrac{\sigma^2}{\rho^2} (\mathbf{H}\mathbf{H}^H)^{-1} + (\mathbf{I} + j\boldsymbol{\Theta}^{(t-1)})^H (\mathbf{I} + j\boldsymbol{\Theta}^{(t-1)})]^{-1} \\ &\quad \cdot (\mathbf{I} + j\boldsymbol{\Theta}^{(t-1)})^H \mathbf{r}, \end{aligned} \tag{2}$$

where $\mathbf{X} = \mathbf{diag}(\mathbf{x})$ and $\boldsymbol{\Theta} = \mathbf{diag}(\boldsymbol{\theta})$.

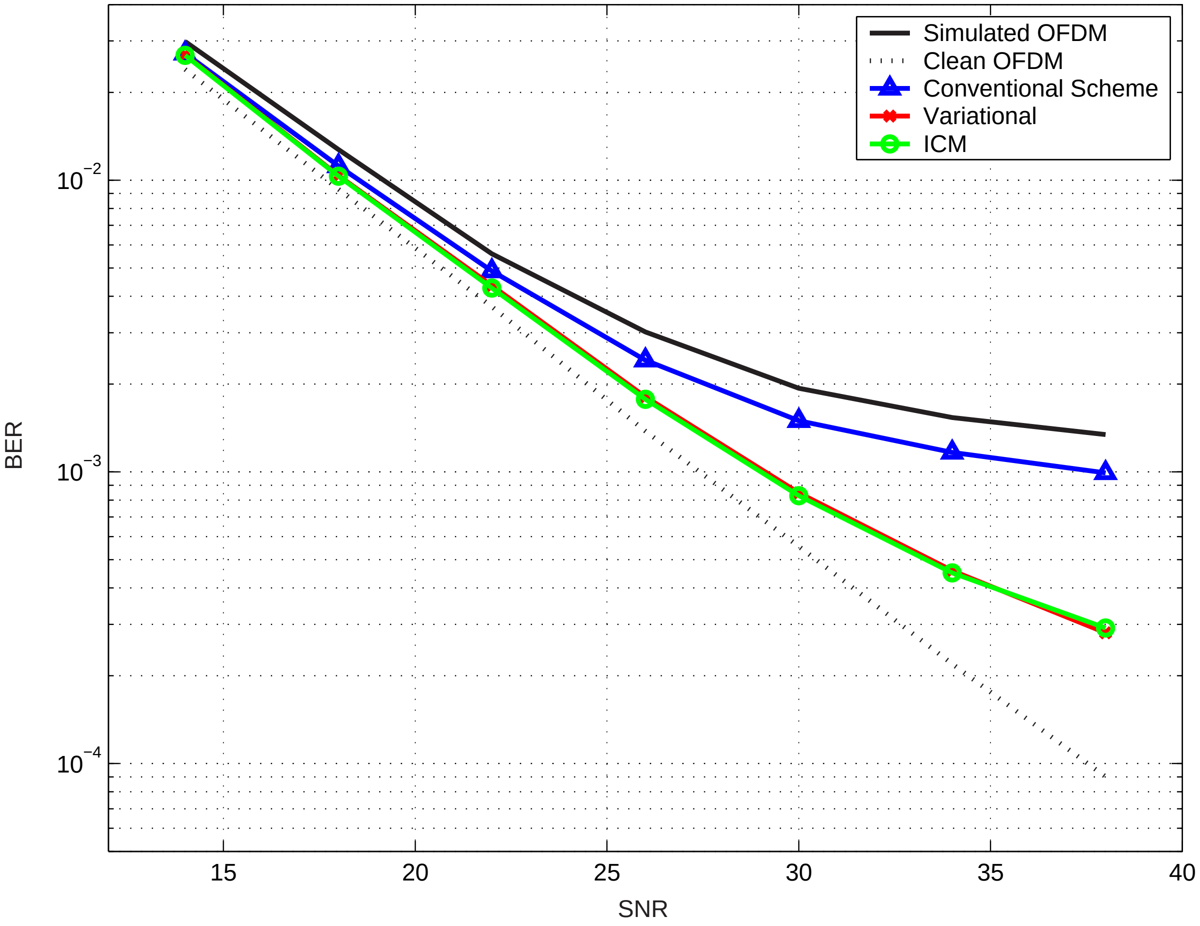
IV. Variational PHN Cancellation: Simulations

- Settings: 1) Rayleigh fading with 3 taps; 2) $N = 64$ subcarriers and each subcarrier 64-QAM; 3) Baseband $f_s = 20$MHz; 4) A phase-locked VCO with a PHN of 3 degrees RMS; 5) Random PHN generated as i.i.d. Gaussian samples through a 100KHz 3dB B/W Butterworth filter as recommended for IEEE 802.11g standard.



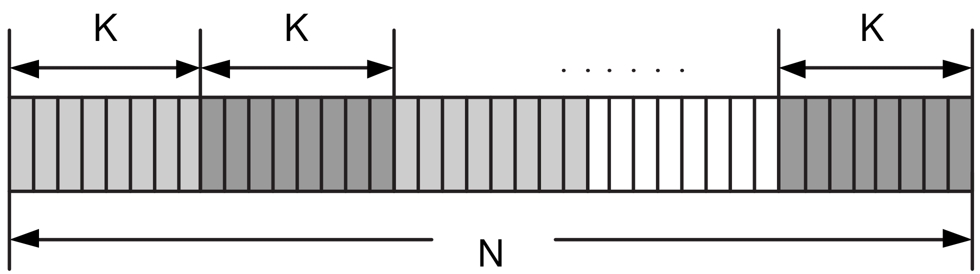
IV. Variational PHN Cancellation: Simulations

- Bit error rate (BER) comparison.



IV. Variational PHN Cancellation: Complexity

- The updates of parameters involve the inverse of $N \times N$ matrices, which has complexity $\mathcal{O}(N^3)$.

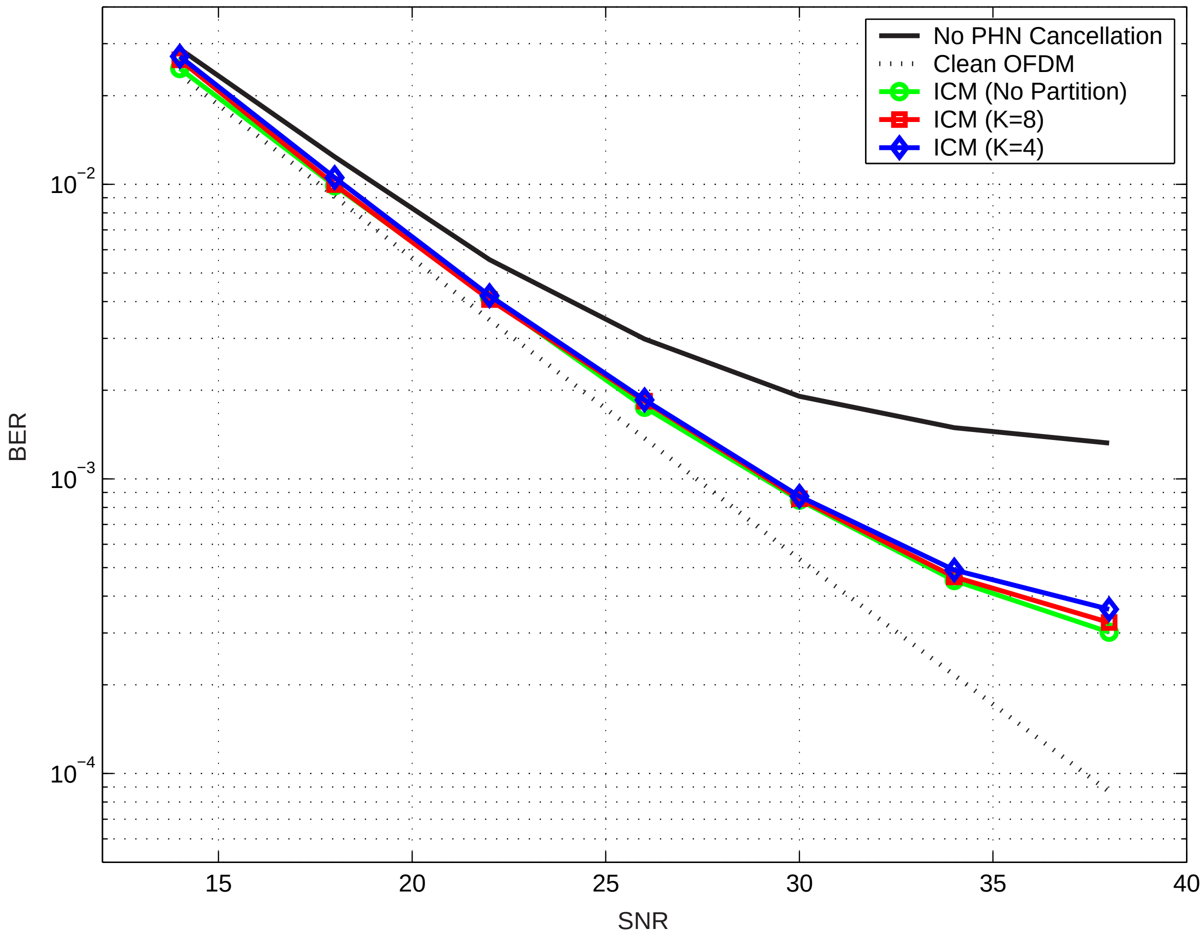


Received OFDM Symbol

- By partitioning the OFDM symbol into smaller blocks of size K, the complexity of PHN cancellation is reduced to $\mathcal{O}(N/K \times K^3) = \mathcal{O}(NK^2)$. A big difference for N in the thousands.

IV. Simplified PHN Cancellation: Simulations

- Bit error rate (BER) comparison.



V. Summary

- What is phase noise?
- What are the conventional and optimal solutions?
- How do we find practical solutions using variational inference?
- Future directions:
  - Further reduce the computational complexity.
  - Extend to carrier frequency offset and Doppler shift cancellation.
  - Use variational EM to remove residual errors in channel estimation.

Thank You!