Sparse Bayesian Learning for Site-Specific Hybrid MIMO Channel Estimation

ICARUS: Özlem Tuğfe Demir[*], Cenk M. Yetis[**], Emil Björnson[*,***], Pontus Giselsson[**]

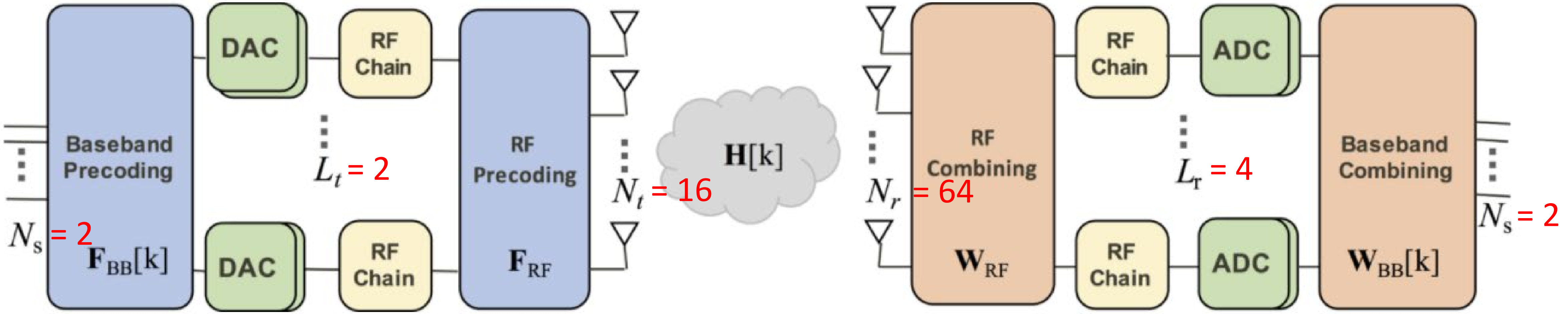
[*]Department of Electrical Engineering (ISY), Linköping University, Sweden

[**] Department of Automatic Control, Lund University, Sweden

[***]Department of Computer Science, KTH Royal Institute of Technology, Sweden

WASP | WALLENBERG AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

Challenge: Site-specific channel estimation with hybrid MIMO [AI5G]



- Frequency selective hybrid millimeter wave MIMO-OFDM, K=256 subcarriers
- **Raymobtime** [RayMob] for collecting realistic datasets by ray-tracing with mobility
- **Off-line training** → 10,000 channels and 100 corresponding received pilots in frequency domain for SNR = -15 dB, -10 dB, and -5 dB
- **Testing** → 9 collections of received pilots at different SNRs and pilot lengths

[AI5G] https://research.ece.ncsu.edu/ai5gchallenge/
[RayMob] https://www.lasse.ufpa.br/raymobtime/

System model [FPVH18]

ℓth delay tap of the channel → $\mathbf{H}_\ell = \mathbf{A}_\mathrm{R}\mathbf{\Delta}_\ell\mathbf{A}_\mathrm{T}^\mathrm{H}, \quad \ell = 0, \ldots, L-1$

$\mathbf{\Delta}_\ell \in \mathbb{C}^{N_\mathrm{ray} \times N_\mathrm{ray}}$ → Diagonal matrix with complex path gains

The millimeter wave channel is **sparse**:

- The number of non-zeros taps is much smaller than *L*!
- Each non-zero tap only contains beams from a small number of angles!

Extended virtual channel model: $\mathbf{H}_\ell \approx \tilde{\mathbf{A}}_\mathrm{R}\mathbf{\Delta}_\ell^v\tilde{\mathbf{A}}_\mathrm{T}^\mathrm{H}$

$\mathbf{\Delta}_\ell^v \in \mathbb{C}^{G_r \times G_t}$ → Sparse matrix with path gains of the quantized spatial frequencies

Frequency-domain channel at subcarrier k

$$\mathbf{H}[k] \approx \tilde{\mathbf{A}}_\mathrm{R}\left(\sum_{\ell=0}^{L-1}\mathbf{\Delta}_\ell^v e^{-\mathrm{j}\frac{2\pi k}{K}\ell}\right)\tilde{\mathbf{A}}_\mathrm{T}^\mathrm{H} = \tilde{\mathbf{A}}_\mathrm{R}\mathbf{\Delta}^v[k]\tilde{\mathbf{A}}_\mathrm{T}^\mathrm{H}$$

Same non-zero indices for the sparse matrices $\mathbf{\Delta}^v[k]$ for $k = 1, \ldots, K$

We will exploit this in our proposed **sparse Bayesian learning** method to improve the channel estimation at **low SNRs** for the **sparse channels in the considered site**.

[FPVH18] J. Rodríguez-Fernandez, N. González-Prelcic, K. Venugopal, and R. W. Heath, "Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems," IEEE Transactions on Wireless Communications, vol. 17, no. 5, pp. 2946–2960, 2018.

System model [FPVH18]

Concatenated received signals during M training intervals

$$\underbrace{\begin{bmatrix} \mathbf{y}^{(1)}[k] \\ \vdots \\ \mathbf{y}^{(M)}[k] \end{bmatrix}}_{\mathbf{y}[k]} = \underbrace{\begin{bmatrix} \mathbf{\Phi}^{(1)} \\ \vdots \\ \mathbf{\Phi}^{(M)} \end{bmatrix}}_{\mathbf{\Phi}} \mathbf{\Psi}\mathbf{h}^{v}[k] + \underbrace{\begin{bmatrix} \mathbf{n}_c^{(1)}[k] \\ \vdots \\ \mathbf{n}_c^{(M)}[k] \end{bmatrix}}_{\mathbf{n}_c[k]}$$

Pilot signal | Transmit precoder | Receive combiner

$$\mathbf{\Phi}^{(m)} = \left(\mathbf{q}^{(m)^{\mathrm{T}}} \mathbf{F}_{\mathrm{tr}}^{(m)^{\mathrm{T}}} \otimes \mathbf{W}_{\mathrm{tr}}^{(m)^{\mathrm{H}}} \right)$$

$$\mathbf{\Psi} = (\tilde{\mathbf{A}}_{\mathrm{T}}^{*} \otimes \tilde{\mathbf{A}}_{\mathrm{R}})$$

Dictionary matrix:

- known,
- fixed grid, BUT
- different spatial frequencies during off-line training and testing in the proposed method

$\mathbf{h}^{v}[k] = \mathrm{vec}\{\mathbf{\Delta}^{v}[k]\} \in \mathbb{C}^{G_t G_r} \longrightarrow$ Sparse vector

Colored noise, the correlation matrix of $\mathbf{n}_c[k]/\sigma$:

$$\mathbf{C}_{\mathbf{w}} = \mathrm{blkdiag}\left\{ \mathbf{W}_{\mathrm{tr}}^{(1)^{\mathrm{H}}} \mathbf{W}_{\mathrm{tr}}^{(1)}, \; \ldots, \; \mathbf{W}_{\mathrm{tr}}^{(M)^{\mathrm{H}}} \mathbf{W}_{\mathrm{tr}}^{(M)} \right\}$$

[FPVH18] J. Rodríguez-Fernandez, N. González-Prelcic, K. Venugopal, and R. W. Heath, “Frequency-domain compressive channel estimation for frequency-selective hybrid millimeter wave MIMO systems,” IEEE Transactions on Wireless Communications, vol. 17, no. 5, pp. 2946–2960, 2018.

Proposed Framework:
Conversion of the received signal to time-domain and whitening

First Step: Apply inverse DFT to the frequency-domain received signals and scale them:

$$\widetilde{\mathbf{y}}[\ell] = \frac{1}{\sqrt{K}} \left(\sum_{k=0}^{K-1} \mathbf{y}[k] e^{\mathrm{j}\frac{2\pi k}{K}\ell} \right)$$
$$= \boldsymbol{\Phi}\boldsymbol{\Psi}\widetilde{\mathbf{h}}^{v}[\ell] + \widetilde{\mathbf{n}}_c[\ell], \quad \ell \in \mathcal{L}$$

Due to sparse channel, most of the time-domain received signals contain only noise!

Second Step: Keep only a subset of delay taps that dominate the power of the received signal:

$$\mathcal{L} \subset \{0, \ldots, K-1\}$$

By applying a simple thresholding on the total energy of the signals $\widetilde{\mathbf{y}}[\ell]$, for $\ell = 0, \ldots, K-1$

Third Step: Whiten the time-domain signals:

$$\mathbf{y}_{\mathrm{w}}[\ell] = \mathbf{C}_{\mathrm{w}}{}^{1/2}\widetilde{\mathbf{y}}[\ell] = \mathbf{C}_{\mathrm{w}}{}^{1/2}\boldsymbol{\Phi}\boldsymbol{\Psi}\widetilde{\mathbf{h}}^{v}[\ell] + \underbrace{\mathbf{n}_{\mathrm{w}}[\ell]}_{\sim \mathcal{N}_{\mathbb{C}}(\mathbf{0}_{ML_r}, \sigma^2 \mathbf{I}_{ML_r})}$$

Proposed pattern-coupled hierarchical model

- We extend the pattern-coupled sparse Bayesian learning in [FZL16] for our problem by
  - Introducing sparsity connections between consecutive AoAs and AoDs
  - Modeling the hyper-parameters to exploit the common sparsity
- Noisy measurements: $\mathbf{y}^{\ell} = \mathbf{A}\mathbf{x}^{\ell} + \mathbf{w}^{\ell}, \quad \ell \in \mathcal{L}$

Sparse unknown vectors ($\mathbf{x}^{\ell}$)

i.i.d. Gaussian noise with zero-mean and unknown noise variance ($\mathbf{w}^{\ell}$)

$$\mathbf{y}^{\ell} = \mathbf{y}_{\mathrm{W}}[\ell] \qquad \mathbf{A} = \mathbf{C}_{\mathrm{W}}{}^{1/2}\mathbf{\Phi}\mathbf{\Psi}$$

$$\mathbf{x}^{\ell} = \widetilde{\mathbf{h}}^{v}[\ell] \qquad \mathbf{w}^{\ell} = \mathbf{n}_{\mathrm{W}}[\ell]$$

[FZL16] J. Fang, L. Zhang, and H. Li, "Two-dimensional pattern-coupled sparse Bayesian learning via generalized approximate message passing," IEEE Transactions on Image Processing, vol. 25, no. 6, pp. 2920–2930, 2016.

Proposed pattern-coupled hierarchical model

$$\mathbf{x}^{\ell} = \begin{bmatrix} x^{\ell}_{1,1} \\ \vdots \\ x^{\ell}_{G_r,1} \\ x^{\ell}_{1,2} \\ \vdots \\ x^{\ell}_{G_r,2} \\ \vdots \\ x^{\ell}_{1,G_t} \\ \vdots \\ x^{\ell}_{G_r,G_t} \end{bmatrix}, \quad \ell \in \mathcal{L}$$

- We exploit the block-sparse structure along AoAs, AoDs, and the common sparsity for all the delay taps, we define the prior

$$p(\mathbf{x}|\boldsymbol{\alpha}) = \prod_{g_r=1}^{G_r} \prod_{g_t=1}^{G_t} \prod_{\ell \in \mathcal{L}} \mathcal{N}_{\mathbb{C}}(x^{\ell}_{g_r,g_t} | 0, \eta^{-1}_{g_r,g_t})$$

$$\mathbf{x} \triangleq \{\mathbf{x}^{\ell} : \ell \in \mathcal{L}\}$$

- Pattern-coupling:

$$\eta_{g_r,g_t} = \alpha_{g_r,g_t} + \beta_r \alpha_{g_r-1,g_t} + \beta_r \alpha_{g_r+1,g_t} + \beta_t \alpha_{g_r,g_t-1} + \beta_t \alpha_{g_r,g_t+1}$$

$\{\alpha_{g_r,g_t}\}$ → Hyperparameters controlling the sparsity

$\beta_r \in [0,1]$ and $\beta_t \in [0,1]$ → Pattern relevance parameters

[FZL16] J. Fang, L. Zhang, and H. Li, “Two-dimensional pattern-coupled sparse Bayesian learning via generalized approximate message passing,” IEEE Transactions on Image Processing, vol. 25, no. 6, pp. 2920–2930, 2016.

Expectation maximization (EM) algorithm for the proposed sparse Bayesian learning method

- Introduce the inverse of noise variance as another hyperparameter:

$$\gamma = 1/\sigma^2 \qquad \gamma \sim \mathcal{U}[\gamma_{\text{low}}, \gamma_{\text{upp}}]$$

- Utilize EM algorithm for learning the sparse signal $\mathbf{x}$ and the hyperparameters $\Theta \triangleq \{\alpha, \gamma\}$

- Treat the sparse signal as hidden variable and maximize a lower bound on the posterior probability $p(\Theta|\mathbf{y})$ (called Q-function)

- Alternate between E-step and M-step.

E-Step: Compute the posterior distribution of the sparse signal conditioned on the observed data and the hyperparameters

Multivariate Gaussian distribution → compute the mean and covariance matrix in closed-form!

M-Step: Estimate the hyperparameters by maximizing the Q-function

$$\Theta^{(t+1)} = \arg\max_{\Theta} Q(\Theta|\Theta^{(t)})$$
$$= \arg\max_{\Theta} \mathbb{E}_{\mathbf{x}|\mathbf{y},\Theta^{(t)}}\{\ln p(\Theta|\mathbf{x}, \mathbf{y})\}$$

Only one suboptimal update, all closed-form simple updates!

Learning the joint relations between AoAs and AoDs

Apply the same EM maximization algorithm by arranging the system model such that the observed variables are true channels that are available from the off-line training data

- Use the uniform grid on $[0, \pi]$ with $G_r = 96$ and $G_t = 24$
- We add a small-variance noise ($\gamma = 10^4$) to obtain the original sparse model

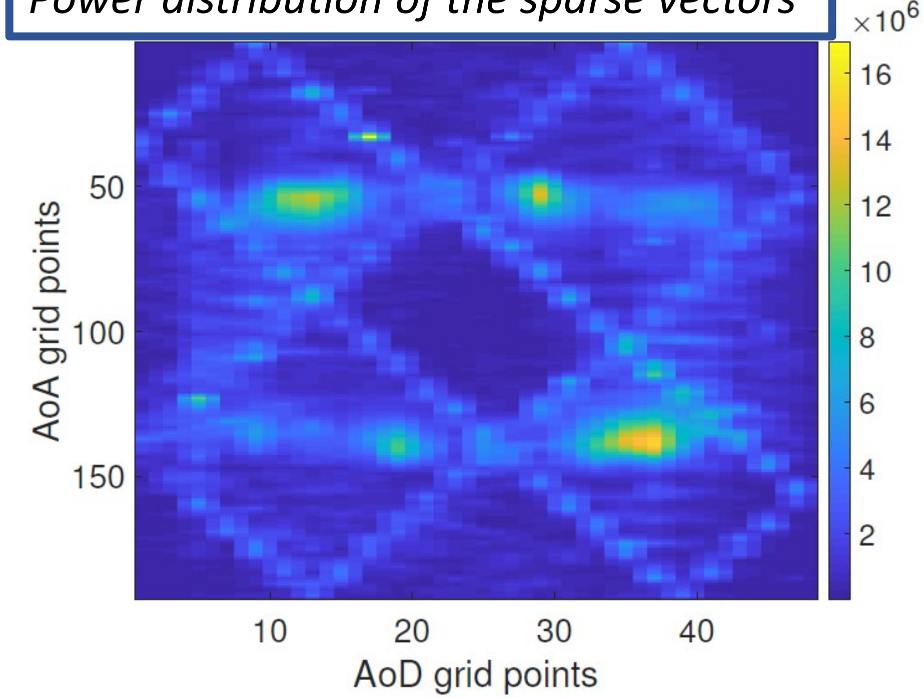
$$\mathbf{y}_{\text{training}}^{\ell} = \mathbf{\Psi}\mathbf{x}_{\text{training}}^{\ell} + \mathbf{w}_{\text{training}}^{\ell}$$

- Obtain 10,000 sparse estimates $\widehat{\mathbf{x}}_{\text{training}}^{\ell}$ and estimate the power distribution along $2G_r = 192$ AoA and $2G_t = 48$ AoD points (interpolation by 2 along both AoA and AoDs)

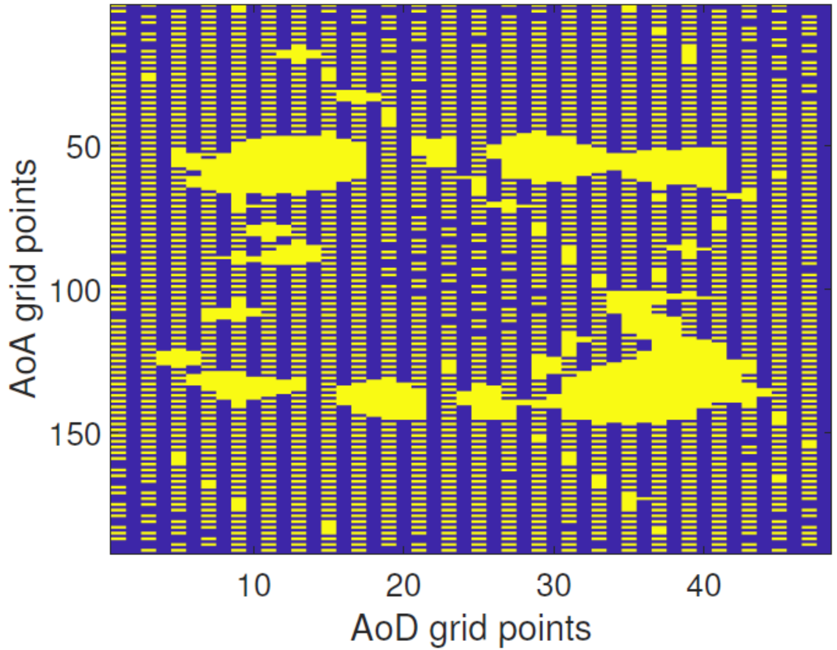
Learning the joint relations between AoAs and AoDs

- We propose a grid construction algorithm to adjust the grid points according to the power distribution.
- Start with a uniform grid with 96x24 points.
- Assign additional 96x8 grid points to the most yellowish regions by sorting the power values accordingly.
- Move the points to the places where the power of the sparse vectors obtained from the training data is greater.
- At the same time, try to prevent the neighboring grid points from being far away by some tuning.

Power distribution of the sparse vectors



Selected grid points for testing stage (yellow pixels)



Evaluation: Test scores and final ranking

- 9 unknown data sets of received pilots obtained at SNRs ranging from -20 to 0 dB
- 1000 channels in each data set
- Normalized mean square error (NMSE) scores:

	[-20 dB, -11 dB]	[-11 dB, -6 dB]	[-6 dB, 0 dB]
Pilot length: 20	-8.94 dB	-9.99 dB	-10.31 dB
Pilot length: 40	-10.82 dB	-11.33 dB	-11.89 dB
Pilot length: 80	-11.74 dB	-12.47 dB	-12.98 dB

Using weights 0.5, 0.3, and 0.2, giving more weight to the more challenging settings (lower SNR and less training):

Final performance score (PS) is -9.48 dB

Rank	Team-Name
1	ML-DOJO
2	ICARUS
3	Learned Chester

Thank you!

ICARUS: Özlem Tuğfe Demir, Cenk M. Yetis, Emil Björnson, Pontus Giselsson

E-mail: ozlem.tugfe.demir@liu.se

WASP | WALLENBERG AUTONOMOUS SYSTEMS AND SOFTWARE PROGRAM

LiU LINKÖPING UNIVERSITY

LUND UNIVERSITY

KTH VETENSKAP OCH KONST