MmWave Channel Estimation via Atomic Norm Minimization for Multi-User Hybrid Precoding

Junquan Deng*, Olav Tirkkonen* and Christoph Studer†

*Department of Communications and Networking, Aalto University, Finland

†School of Electrical and Computer Engineering, Cornell University, NY, USA

April 17, 2018

5G Requirements and NR

- IMT-2020 enhanced Mobile Broadband (eMBB)[1]

KPI	**Value**	
Peak data rate	DL: 20 Gbps,	UL: 10 Gbps
Peak spectral efficiency	DL: 30 bps/Hz,	UL: 15 bps/Hz
5% user spectral efficiency	DL: 0.225 bps/Hz,	UL: 0.15 bps/Hz
Average spectral efficiency	DL: 7.80 bps/Hz,	UL: 5.40 bps/Hz

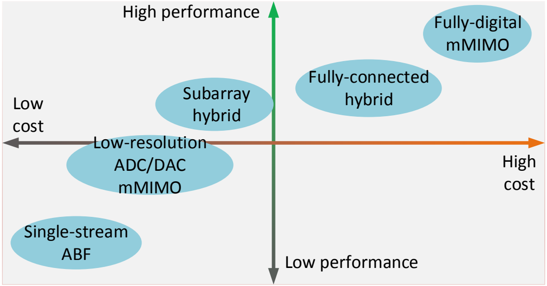
- Three Key Technologies for 5G eMBB
  - Millimeter-Wave (mmWave) with large continuous bandwidths
  - Beamforming and MU-MIMO with large antenna arrays
  - Ultra-Dense Network for seamless coverage
- 3GPP 5G New Radio (NR)
  - FR1 (450-6000 MHz) and FR2 (24250-52600 MHz)
  - OFDM with flexible subcarrier spacing to support extreme wideband transmissions in high frequencies
  - Support large number of antenna ports (i.e massive MIMO)

---

[1] ITU-R, "Minimum Requirements Related to Technical Performance for IMT-2020 Radio Interface(s)," Report M.2410-0, Nov. 2017.

Cost-efficient Solutions Are Necessary

- Single-stream Analog Beamforming (ABF) with one RF chain
- Massive MIMO systems with low-precision ADCs/DACs
- Hybrid beamforming architectures with phase shifters and a small number of RF chains



- Performance of low-complexity architectures depends heavily on mmWave channel characteristics

Urban Outdoor MmWave Channel Model

Channel matrix on an OFDM subcarrier for a UE k is given by

$$\mathbf{H}_k = \sum_{l=1}^{L} \alpha_l \mathbf{a}_{\mathrm{UE}}(\theta_l) \mathbf{a}_{\mathrm{BS}}^{\mathrm{H}}(\phi_l).$$

$$\mathbf{a}_{\mathrm{BS}}(\phi) = [1, \mathrm{e}^{j\frac{2\pi}{\lambda} d \sin(\phi)}, \ldots, \mathrm{e}^{j(N-1)\frac{2\pi}{\lambda} d \sin(\phi)}]^{\mathrm{T}},$$

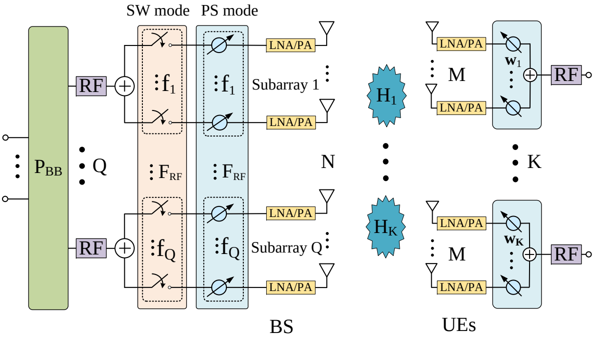
$$\mathbf{a}_{\mathrm{UE}}(\theta) = [1, \mathrm{e}^{j\frac{2\pi}{\lambda} d \sin(\theta)}, \ldots, \mathrm{e}^{j(M-1)\frac{2\pi}{\lambda} d \sin(\theta)}]^{\mathrm{T}},$$

- MmWave channel models highly depend on the environments
- Channels are dominated by LoS and/or low-order reflection multi-path components (MPCs) in urban outdoor
- Angular spread in elevation is much smaller than in azimuth, and users need to be separated in azimuth domain

Channel Estimation Challenges in mmWave MU-MIMO

- Full Channel State Information (CSI) for each user is generally required at the BS for inter-user interference mitigation in MU-MIMO
- Channel estimation in hybrid architectures is challenging as received reference signals are beamformed
- Fortunately, mmWave MIMO channels are approximately low-rank (dominated by LoS and low-order reflections), and Comprehensive Sensing (CS) methods can be utilized
- Design of measurement matrix and recovery algorithm is important

Proposed BS and UE Architectures



System architectures for the BS and UEs. Each UE has a single RF chain with M phase shifters. The BS uses the switch-phase-shifter-subarray architecture, which has two operation modes:

- phase-shifter-based mode (PS mode) for beamforming transmission
- switch-based mode (SW mode) for channel estimation

Signaling for Channel Estimation

- Assume that the kth UE transmits a pilot sequence $\mathbf{s}_k \in \mathbb{C}^{1\times T_\mathrm{s}}$ using its best beam $\mathbf{w}_k$. The received training signal at BS is

$$\mathbf{Y} = \mathbf{F}_\mathrm{RF}^\mathrm{H}\left(\sum\nolimits_{k=1}^{K} \mathbf{H}_k^\mathrm{H}\mathbf{w}_k\mathbf{s}_k + \mathbf{N}_\mathrm{b}\right) = \mathbf{F}_\mathrm{RF}^\mathrm{H}\left(\sum\nolimits_{k=1}^{K} \mathbf{h}_k\mathbf{s}_k + \mathbf{N}_\mathrm{b}\right)$$

- Assume orthogonality for pilots with $\mathbf{s}_k\mathbf{s}_{k'}^\mathrm{H} = T_\mathrm{s}\rho_\mathrm{UE}\delta_{k,k'}$, we have

$$\mathbf{y}_k = \mathbf{Y}\mathbf{s}_k^\mathrm{H} = T_\mathrm{s}\rho_\mathrm{UE}\mathbf{F}_\mathrm{RF}^\mathrm{H}\mathbf{h}_k + \mathbf{F}_\mathrm{RF}^\mathrm{H}\mathbf{N}_\mathrm{b}\mathbf{s}_k^\mathrm{H}$$

- T snapshots of measurements

$$\mathbf{z} = \begin{bmatrix} \mathbf{y}_{1,k} \\ \mathbf{y}_{2,k} \\ \vdots \\ \mathbf{y}_{T,k} \end{bmatrix} = \underbrace{T_\mathrm{s}\rho_\mathrm{UE}\begin{bmatrix} \mathbf{F}_{1,\mathrm{RF}}^\mathrm{H} \\ \mathbf{F}_{2,\mathrm{RF}}^\mathrm{H} \\ \vdots \\ \mathbf{F}_{T,\mathrm{RF}}^\mathrm{H} \end{bmatrix}}_{\mathbf{\Phi}} \mathbf{h}_k + \underbrace{\begin{bmatrix} \mathbf{F}_{1,\mathrm{RF}}^\mathrm{H}\mathbf{N}_{1,\mathrm{b}} \\ \mathbf{F}_{2,\mathrm{RF}}^\mathrm{H}\mathbf{N}_{2,\mathrm{b}} \\ \vdots \\ \mathbf{F}_{T,\mathrm{RF}}^\mathrm{H}\mathbf{N}_{T,\mathrm{b}} \end{bmatrix}\mathbf{s}_k^*}_{\mathbf{n}},$$

On-grid Channel Estimation

- $\mathbf{h}_k = \mathbf{H}_k^{\mathrm{H}} \mathbf{w}_k = \sum_{l=1}^{L} \beta_l \mathbf{a}_{\mathrm{BS}}(\phi_l)$ is effective MISO channel, it contains fewer significant paths compared to the full MIMO channel $\mathbf{H}_k$
- In grid-based CS methods, a discrete dictionary $\mathbf{\Psi}_{\mathrm{BS}} = [\mathbf{a}_{\mathrm{BS}}(\phi_1), \ldots, \mathbf{a}_{\mathrm{BS}}(\phi_{G_{\mathrm{b}}})]$ with G_{b} bases is used to represent the channel $\mathbf{h}_k$ as

$$\mathbf{h}_k = \mathbf{\Psi}_{\mathrm{BS}} \mathbf{h}_{\mathrm{v}},$$

- Denoting $\mathbf{A} = \mathbf{\Phi} \mathbf{\Psi}_{\mathrm{BS}}$, to estimate the sparse virtual channel $\mathbf{h}_{\mathrm{v}}$, one can formulate the following optimization problem:

$$\underset{\mathbf{h}_{\mathrm{v}}}{\text{minimize}} \, \|\mathbf{h}_{\mathrm{v}}\|_1 \quad \text{s.t.} \quad \|\mathbf{z} - \mathbf{A}\mathbf{h}_{\mathrm{v}}\|_2^2 \leq \eta \qquad \text{(P1)}$$

- P1 can be efficiently solved via Orthogonal Matching Pursuit (OMP) algorithm

Gridless Channel Estimation via ANM

- Using $\mathbf{\Psi}_{\mathrm{BS}}$ introduces a basis mismatch problem
- Let us consider a continuous dictionary as

$$\mathcal{A} = \left\{ \underbrace{\mathbf{a}_{\mathrm{BS}}(\phi)\,\alpha}_{Atom} : \ \phi \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right], \alpha \in \mathbb{C}, |\alpha| = 1 \right\}$$

- The atomic norm of a channel $\mathbf{h}$ is defined as

$$\begin{aligned} \|\mathbf{h}\|_{\mathcal{A}} &= \inf \{ g > 0 : \mathbf{h} \in g \cdot \mathsf{conv}(\mathcal{A}) \} \\ &= \inf \left\{ \sum\nolimits_i b_i : \mathbf{h} = b_i \sum\nolimits_i \mathbf{a}_i, b_i > 0, \mathbf{a}_i \in \mathcal{A} \right\}. \end{aligned}$$

- Based on measurements, one can formulate the following optimization problem without introducing a discrete dictionary

$$\underset{\mathbf{h}}{\text{minimize}} \ \|\mathbf{h}\|_{\mathcal{A}} \quad \text{s.t.} \ \|\mathbf{z} - \mathbf{\Phi}\mathbf{h}\|_2^2 \leq \eta. \qquad \text{(P2)}$$

Gridless Channel Estimation via ANM

- $\|\mathbf{h}\|_{\mathcal{A}}$ defined in (P2) equals the optimal value of the following matrix trace minimization problem:

$$\underset{\boldsymbol{u},t}{\text{minimize}}\ \frac{1}{2}(t+u_1) \quad \text{s.t.} \quad \begin{bmatrix} \mathcal{T}(\boldsymbol{u}) & \mathbf{h} \\ \mathbf{h}^{\mathrm{H}} & t \end{bmatrix} \succeq \mathbf{0}, \qquad \text{(P3)}$$

  where $\mathcal{T}(\boldsymbol{u})$ is a Hermitian Toeplitz matrix with the first row as $\boldsymbol{u} = [u_1, \ldots, u_N]^{\mathrm{T}}$.
- In the noisy case, using the atomic norm, a regularized optimization can be formulated as

$$\begin{aligned} &\underset{\mathbf{D}\succeq\mathbf{0}}{\text{minimize}}\ \frac{\xi}{2}(t+u_1)+\frac{1}{2}\|\mathbf{z}-\mathbf{\Phi}\mathbf{h}\|_2^2 \\ &\text{s.t.}\ \ \mathbf{D} = \begin{bmatrix} \mathcal{T}(\boldsymbol{u}) & \mathbf{h} \\ \mathbf{h}^{\mathrm{H}} & t \end{bmatrix}. \end{aligned} \qquad \text{(P4)}$$

- P4 is a SDP, and can be solved by off-the-shelf convex optimization tools in polynomial time

Antenna Domain Sub-Sampling (ADSS)

- $\mathbf{A} = \mathbf{\Phi}\mathbf{\Psi}_{\mathrm{BS}}$ with measurement matrix $\mathbf{\Phi}$ and dictionary matrix $\mathbf{\Psi}_{\mathrm{BS}}$. OMP can recover $\mathbf{h}_{\mathrm{v}}$ in the noiseless case if

$$\mu(\mathbf{A}) = \max_{i \neq j} \frac{|\mathbf{a}_i^{\mathrm{H}} \mathbf{a}_j|}{\|\mathbf{a}_i\|_2 \cdot \|\mathbf{a}_j\|_2} < \frac{1}{2L-1},$$

  where $\mathbf{a}_i$ and $\mathbf{a}_j$ are two different columns of $\mathbf{A}$, and $\mu(\mathbf{A})$ is the coherence of $\mathbf{A}$

- Defining $\mathcal{A}_{\mathbf{\Phi}} = \left\{\mathbf{\Phi}\mathbf{a}_{\mathrm{BS}}(\phi) : \phi \in \left(-\frac{\pi}{2}, \frac{\pi}{2}\right]\right\}$, the problem (P2) with $\eta = 0$ in the noiseless setting has a unique solution if the number of paths satisfies

$$L < \frac{\mathrm{spark}(\mathcal{A}_{\mathbf{\Phi}})}{2},$$

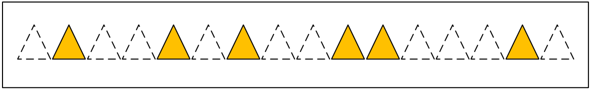
  where $\mathrm{spark}(\mathcal{A}_{\mathbf{\Phi}})$ is defined as the smallest number of atoms which are linearly dependent in $\mathcal{A}_{\mathbf{\Phi}}$

Antenna Domain Sub-Sampling (ADSS)

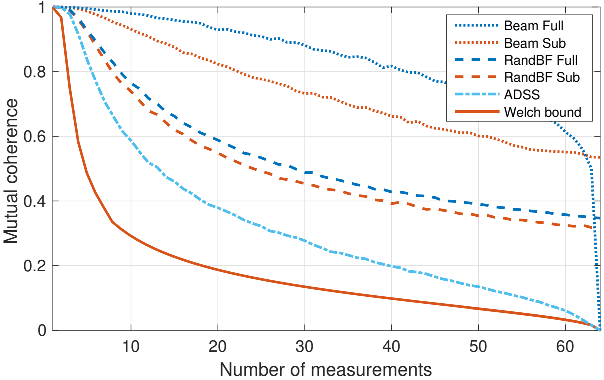
- To construct $\mathbf{A}$ with low coherence and $\mathcal{A}_{\mathbf{\Phi}}$ with large spark, $\mathbf{\Phi}$ and $\mathbf{\Psi}_{\mathrm{BS}}$ (or $\mathcal{A}$) should be constructed in different "domains"
- As the dictionary is in angular domain (beam domain), beam-domain measurement is unable to achieve low-coherence
- We consider antenna domain for construction of $\mathbf{\Phi}$

$$\text{ADSS}: \quad \mathbf{\Phi} = T_{\mathrm{s}}\rho_{\mathrm{UE}}[\mathbf{e}_{i_{1,1}}, \mathbf{e}_{i_{1,2}}, \ldots, \mathbf{e}_{i_{1,Q}}, \ldots, \mathbf{e}_{i_{T,Q}}]^{\mathrm{H}}$$

- ADSS measurement matrix and $\mathbf{\Psi}_{\mathrm{BS}}$ (or $\mathcal{A}$) are mutually incoherent



Mutual Coherence of ADSS



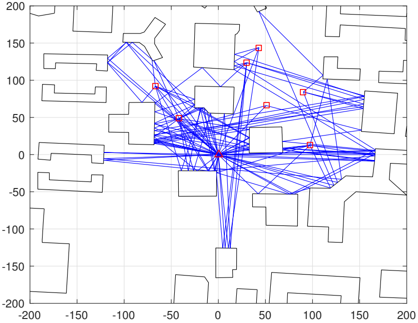
- Beam-steering (Beam) with a fully-connected array
- Beam-steering with contiguous subarrays
- Random phase-shifting (RandBF) with a fully-connected array
- Random phase-shifting with contiguous subarrays
- ADSS via switches

mmWave MU-MIMO precoding/combining

RF Beam-tracking + baseband ZF

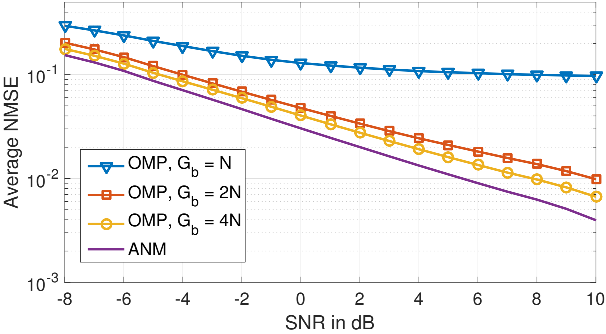
- BS transmits Q DL pilots using Q sub-arrays
- Each UE receives DL pilots using codewords from UE codebook $\mathcal{U} = \{\mathbf{w}^{(1)}, \ldots, \mathbf{w}^{(P)}\}$, and find a best UE beam $\mathbf{w}_k$ which maximize the aggregate received power
- UE k transmits UL pilot to BS using $\mathbf{w}_k$, the BS estimates the effective user channel $\mathbf{h}_k$ using OMP or ANM methods; the BS has $\hat{\mathbf{H}} = [\hat{\mathbf{h}}_1, \hat{\mathbf{h}}_2, \ldots, \hat{\mathbf{h}}_K]^{\mathrm{H}}$
- For qth RF chain, select UE $k = q$, $\mathbf{f}_q = \mathrm{argmax}_{\mathbf{f}\in\mathcal{F}_q} |\hat{\mathbf{h}}_k^{\mathrm{H}}\mathbf{f}|^2$, where $\mathcal{F}_q$ is the RF beam-steering codebook subject to low phase-shifter resolution
- Design baseband precoder as $\mathbf{P}_{\mathrm{BB}} = \frac{(\hat{\mathbf{H}}\mathbf{F}_{\mathrm{RF}})^{-1}}{\|\mathbf{F}_{\mathrm{RF}}(\hat{\mathbf{H}}\mathbf{F}_{\mathrm{RF}})^{-1}\|_{\mathrm{F}}}$ using Zero-Forcing

Performance Evaluation



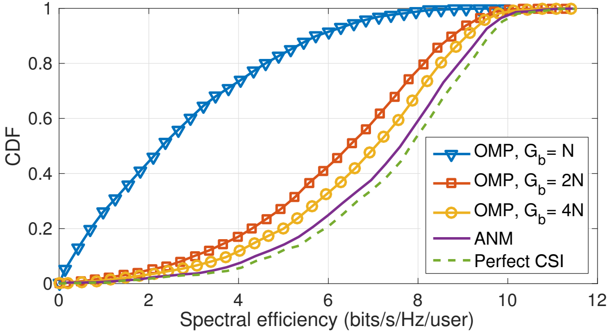
- $M = 8$ for UEs, and $N = 64$, $Q = 8$ for BS
- Ray-tracing is used to generate MPCs; up to 3rd order reflections are taken into account
- Blocked UEs which cannot find a LoS or a reflected path are not considered; 8000 UEs are randomly divided into 1000 groups and 8 UEs in each group are served simultaneously.

NMSE for Channel Estimation



- Normalized Mean Square Error (NMSE) $\mathbb{E}\{\|\mathbf{h}_k - \hat{\mathbf{h}}_k\|_2^2 / \|\mathbf{h}_k\|_2^2\}$
- Only 16 antennas are sampled via the switch network randomly
- When a dictionary with $G_b = N$ is used, OMP suffers severely from basis mismatch
- As G_b increases, the gap between OMP and ANM decreases

Multi-User Spectral Efficiency Performance



- Performance of multi-user hybrid precoding with sub-arrays, with different channel estimates
- Spectral efficiency is estimated as $\log_2(1+\gamma)$ with γ the signal-to-interference-plus-noise ratio

Concluding Remarks

- We proposed a low-complexity hybrid architecture in which an inexpensive switch network is added to the subarrays to facilitate channel estimation.
- We formulated the mmWave channel estimation as an ANM problem which can be solved via SDP with polynomial complexity.
- A low-coherence measurement matrix is constructed via sub-sampling in antenna domain by the cheap switch network.
- ANM achieves better channel estimation accuracy compared to grid-based CS methods
- Better CSI accuracy can help mmWave MU-MIMO hybrid precoding to achieve better user spectral efficiency performance