

Scaling mmWave MU-MIMO: Information-Theoretic Guidance using Real-World Data

Canan Cebeci*, Oveys Delafrooz Noroozi* and Upamanyu Madhow
ECE Dept., University of California, Santa Barbara, CA, U.S.A.
{ccebci, oveys, madhow}@ucsb.edu

Abstract—Advances in millimeter wave (mmWave) radio frequency integrated circuits (RFICs) potentially enable the realization of one RF chain per antenna for arrays with hundreds of antenna elements, opening up the possibility of fully digital beamforming for truly massive multiuser (MU-)MIMO. In order to scale adaptive multiuser detection to such regimes, an attractive option, explored in several recent studies, is to reduce complexity via “beam-space” techniques which exploit the sparsity of mmWave channels. In this paper, we derive information-theoretic benchmarks using measured mmWave channels, comparing the capacity with ideal channel knowledge without complexity constraints, and several related benchmarks, with that attainable with low-complexity linear adaptive multiuser detection strategies in beam-space. We hope that the resulting insights regarding the price paid for various levels of simplification in processing will guide performance/complexity tradeoffs in the design of scalable signal processing algorithms.

Index Terms—mmWave, MU-MIMO, beam-space processing

I. INTRODUCTION

mmWave MU-MIMO holds great potential for the next-generation wireless systems, combining large available bandwidths with aggressive spatial reuse. The small wavelengths enable realization of compact antenna arrays with a large number of elements, while recent advances in mmWave RFICs open the path to fully digital beamforming with one RF chain for each antenna element. However, as we scale the number of antennas and users, classical multiuser detection algorithms incur drastically increased computational complexity. In addition to the computational burden, the length of the training overhead required to achieve reasonable performance even with “simple” linear adaptive multiuser detection scales up with the signal space dimension [1]. Several recent works [2]–[6] seek to address this issue by exploiting the sparsity of mmWave channels which typically consist of a small number of dominant paths. The sparsity results from occurs larger propagation losses incurred by reflections (surfaces “look rougher” at smaller wavelengths, which results in more scattering) and path blockage (obstacles “look larger” at smaller wavelengths).

For arrays with regular geometries (such as linear or planar arrays with uniformly spaced antennas), this sparsity is conveniently brought out in “beam-space” via a spatial FFT. Specifically, since the spatial response of each incoming path is

a complex exponential in space, “beam-space transformation” of the observations by taking a spatial FFT across antennas implements approximate spatial matched filtering for each path. For a sparse channel, the energy of dominant paths is concentrated into a small number of DFT bins, which makes it possible to reduce signal dimension. This reduces the computational complexity of MU-MIMO processing as we scale the system. Dimension reduction also reduces training overhead for channel estimation/adaptive implementations.

In this paper, we seek to quantify the potential performance loss due to such complexity reduction techniques by establishing fundamental information-theoretic benchmarks for sparse mmWave channels using real-world measured data. Our effort leverages recent efforts of industry, academia and government institutions to collaborate on consolidating mmWave channel measurements from a number of measurement campaigns. By analyzing the capacity of mmWave MU-MIMO using real-world measurements, and channel models abstracted from such measurements, we hope to establish a principled basis for guiding MU-MIMO transceiver and system design. We report on capacity analysis for an urban environment, comparing information-theoretic benchmarks with unconstrained complexity and ideal channel knowledge against the performance attainable using low-complexity adaptive beam-space techniques. Our analysis motivates further research into exploring performance/complexity tradeoffs in beam-space transceiver architectures and algorithms.

For the MU-MIMO systems we consider here, the term “capacity” denotes achievable spectral efficiency, averaged across the band of interest, as well as across users.

II. MMWAVE CHANNEL MODEL

The mmWave channel measurements at 28 GHz used in our capacity calculations are from Charbonnier *et al* [7]. They were collected in an urban environment in downtown Boulder, Colorado. The transmitter (Tx) and receiver (Rx) comprise the NIST 28-GHz switched-array channel sounder. While the Tx is kept stationary, the Rx array with multiple horn antennas is moved along a linear path that has a line-of-sight (LOS) to the Tx. For each ray arriving at the Rx, the path gain, delay, azimuth and elevation angle of arrival (AoA) are extracted via SAGE algorithm [8].

*Equal contribution.

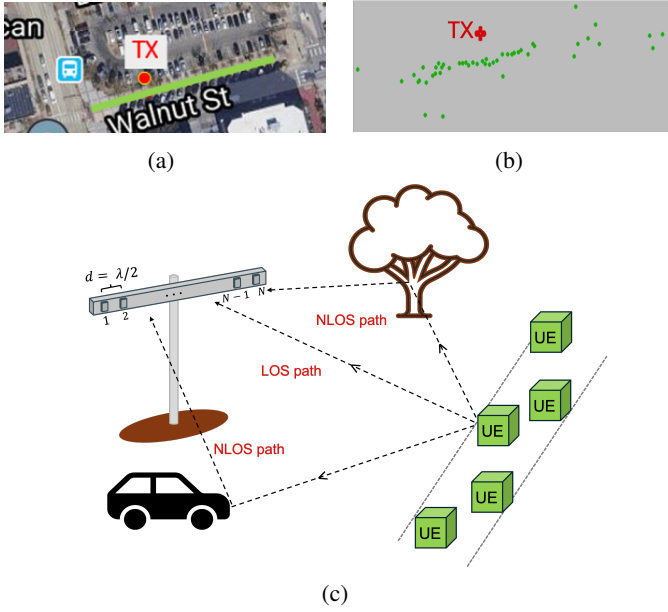


Fig. 1: (a) Environment where the channel measurements are collected. The green line indicates the path on which the Rx was moved [7]. (b) 50 Rx locations with respect to the TX derived from the path losses and azimuth angles of arrival. (c) mmWave MIMO uplink channel with multiple UEs, each having multipaths.

From these channel measurements, we abstract a channel model for a mmWave massive MU-MIMO uplink in an urban picocell. Among the 50 Rx locations used in the measurements (Fig. 1b), we randomly select a K -element subset for channel simulations and use them as our user equipment (UE) locations. We flip the roles of Tx and Rx, using the original Rx locations at ground level as our UE locations, and similarly the original elevated Tx location becomes the location of our base station (BS) receiver. The uplink channel we construct in this fashion consists of K simultaneous single-antenna UEs communicating with a BS equipped with a N -antenna uniform linear array (ULA) with inter-element spacing of d as depicted in Fig. 1c.

The channel impulse response for the k^{th} user with L_k number of paths can be expressed as

$$\mathbf{h}_k(t) = \sum_{l=1}^{L_k} \alpha_{k,l} \mathbf{a}(\Omega_{k,l}) \delta(t - \tau_{k,l}), \quad (1)$$

$$\mathbf{a}(\Omega_{k,l}) = [1 e^{j\Omega_{k,l}} e^{j2\Omega_{k,l}} \dots e^{j(N-1)\Omega_{k,l}}]^T,$$

where $\alpha_{k,l}$, $\Omega_{k,l} = \frac{2\pi d \sin \theta_{k,l}}{\lambda}$, $\theta_{k,l}$ and $\tau_{k,l}$ are the complex path gain, spatial frequency, AoA, and delay for the l^{th} path of the k^{th} user, respectively. We use the standard inter-element spacing $d = \lambda/2$ for our numerical results.

In order to simplify discrete time modeling in this initial exploration, $\tau_{k,l}$ are quantized to the nearest integer sample, and $\tau_{k,1} = 0$ for $1 \leq k \leq K$. The channel matrix in the frequency domain for K users is then given by:

$$\mathbf{H}(f)_{(N \times K)} = [\mathbf{H}_1(f) \quad \dots \quad \mathbf{H}_K(f)], \quad (2)$$

$$\mathbf{H}_k(f) = \sum_{l=1}^{L_k} \alpha_{k,l} \mathbf{a}(\Omega_{k,l}) e^{-j2\pi f \tau_{k,l}} \quad (3)$$

The antenna domain received signal can be expressed as:

$$\mathbf{y}[n] = \sum_{k=1}^K \sum_{l=1}^{L_k} \mathbf{h}_{k,l} \mathbf{b}_k[n - \tau_{k,l}] + \mathbf{n}[n], \quad (4)$$

$$\mathbf{h}_{k,l} = \alpha_{k,l} \mathbf{a}(\Omega_{k,l}),$$

where $\{\mathbf{b}_k[n]\}$ denotes the symbol stream for user k , and $\mathbf{n}[n] \sim \mathcal{CN}(0, 2\sigma^2 \mathbf{I})$ is the complex AWGN noise vector. Applying the DFT across antenna elements (i.e. multiplying the antenna space received signal by the N -point DFT matrix \mathbf{D}_N), we obtain the beamspace received vector as:

$$\tilde{\mathbf{y}}[n] = \mathbf{D}_N \mathbf{y}[n]. \quad (5)$$

The channel matrix constructed from the measurements exhibits sparsity in beamspace as expected (Fig. 2). We can discern a single dominant path for each user in Fig. 2b even in the presence of more than a hundred paths per user.

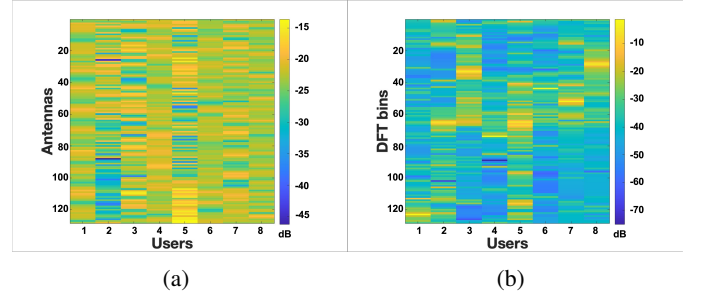


Fig. 2: mmWave channel matrix ($\mathbf{H}(f)$) in the (a) antenna space (b) beamspace. $f = 28$ GHz, $N = 128$, $K = 8$, the number of paths for each user ranges from 110 to 153.

III. MIMO CAPACITY UPPER BOUND

The spectral efficiency (η_B) of a K -user MU-MIMO system over a bandwidth B is defined by the integral expression:

$$\eta_B = \frac{1}{KB} \int_B \log_2 \det \left(\mathbf{I} + \frac{\mathbf{H}(f) \mathbf{H}^H(f)}{2\sigma^2} \right) df. \quad (6)$$

where we have averaged across users as well.

The sparsity of the channel matrix in beamspace, as depicted in Fig. 2, motivates a dimension reduction strategy in which a small window of size W in beamspace is chosen to capture most of the energy for each user. In our setting, we choose this window around the strongest DFT bin (corresponding to the dominant path) for each user. The window selection is depicted in Figure 3 in the context of the linear beamspace receivers that we consider in the next section. However, for the purpose of deriving a MIMO capacity upper bound (unconstrained by detector/decoder complexity), we define a reduced size beamspace channel matrix taking the union of the windows across users. The number of rows in this reduced matrix is therefore bounded by KW . We can now derive

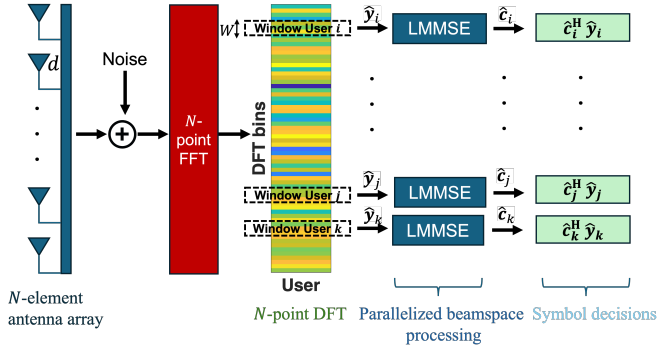


Fig. 3: Windowed beamspace receiver architecture. Different from the conventional antenna space detection, a N -point FFT is applied to the received signal. W number of DFT bin indices containing the most of the energy of each user are extracted from the beamspace received signal.

an information-theoretic benchmark associated with dimension reduction by recalculating spectral efficiency using Eq. 6 for this reduced size beamspace channel matrix.

IV. LMMSE RECEPTION

The capacity benchmarks in the preceding section may be interpreted as MIMO-OFDM with optimal multiuser detection and decoding without any complexity constraints. We now consider a far simpler class of schemes: linear MMSE (LMMSE) multiuser detection, with separate decoding for each user (Fig. 3). For each user, we employ a spatial LMMSE correlator aimed at estimating the symbol modulating its dominant path, so that other multipath components for that user (which are modulated by other, uncorrelated, symbols) act as interference, as do the signals from all other users. This approach is consistent with power-efficient single-carrier modulation, since we make no attempt to “gather” energy from the weaker multipath components of a given user. We compute the signal-to-interference-plus-noise ratio (SINR) for the LMMSE correlator, which yields a lower bound on capacity under a worst-case model of residual interference plus noise as Gaussian.

In the presence of channel side information, it is well known that the LMMSE receiver is given by the expression:

$$\mathbf{c}_k = \mathbf{R}^{-1} \mathbf{p}_k \quad (7)$$

$$\mathbf{R} = E[\mathbf{y}[n]\mathbf{y}^H[n]] = E_b \sum_{k=1}^K \sum_{l=1}^{L_k} \mathbf{h}_{k,l} \mathbf{h}_{k,l}^H + 2\sigma^2 \mathbf{I}$$

$$\mathbf{p}_k = E[\mathbf{y}[n]b_k^*[n]] = E_b \mathbf{h}_{k,1}$$

where E_b is the average symbol energy. The estimated symbol for user k is then $\hat{b}_k[n] = \mathbf{c}_k^H \mathbf{y}[n]$.

For windowed beamspace LMMSE correlation [2], the preceding formulation applies after a suitable linear projection of the original received vector. Specifically, for each user k ,

Eq. 7 can be recalculated using the windowed beamspace received vector:

$$\hat{\mathbf{y}}_k[n] = \mathbf{W}_k \tilde{\mathbf{y}}[n]. \quad (8)$$

Here, \mathbf{W}_k is a $W \times N$ matrix that picks the desired DFT bins related to the k^{th} user. This transformation simplifies the calculation of the LMMSE receiver for the user k , resulting in significantly reduced computational complexity.

The SINR for any linear receiver after dimension reduction can be computed using the original “antenna space” model for the received vector by “lifting” the reduced dimension correlator back to that space. Using Eq. 8, the equivalent LMMSE correlator in the antenna space is given by:

$$\tilde{\mathbf{c}}_k = \mathbf{D}_N^H \mathbf{W}_k^H \hat{\mathbf{c}}_k \quad (9)$$

where the $\hat{\mathbf{c}}_k$ denotes the LMMSE receiver in windowed beamspace for user k .

A particularly attractive feature of LMMSE reception is that it can be adaptively implemented based on a training sequence for each user, without requiring explicit channel information, which is difficult to obtain. Specifically, we consider least squares adaptation, in which the LMMSE correlator for user k is given by the same expression as Eq. 7, replacing \mathbf{R} and \mathbf{p}_k by empirical averages over the training period. Rewriting Eq. 4 to emphasize the dominant path of the desired user k , we obtain:

$$\mathbf{y}[n] = \mathbf{h}_{k,1} b_k[n] + \mathbf{I}_k[n] + \mathbf{n}[n] \quad (10)$$

where interference $\mathbf{I}_k[n]$ is given by:

$$\mathbf{I}_k[n] = \sum_{l=2}^{L_k} \mathbf{h}_{k,l} b_k[n - \tau_{k,l}] + \sum_{j \neq k} \sum_{l=1}^{L_j} \mathbf{h}_{j,l} b_j[n - \tau_{j,l}] \quad (11)$$

Since the symbol streams are uncorrelated, the SINR for user k is:

$$SINR_k = \frac{E_b |\mathbf{c}_k^H \mathbf{h}_{k,1}|^2}{\mathbf{c}_k^H \mathbf{R}_{\mathbf{I}_k} \mathbf{c}_k + 2\sigma^2 \mathbf{c}_k^H \mathbf{c}_k}. \quad (12)$$

Here, $\mathbf{R}_{\mathbf{I}_k} = E[\mathbf{I}_k[n]\mathbf{I}_k^H[n]]$ and E_b is the average symbol energy. We can now obtain a lower bound on capacity by assuming that the residual interference and noise at the correlator output follows a Gaussian distribution. The (lower bound on) capacity of the system over the number of users is therefore given by:

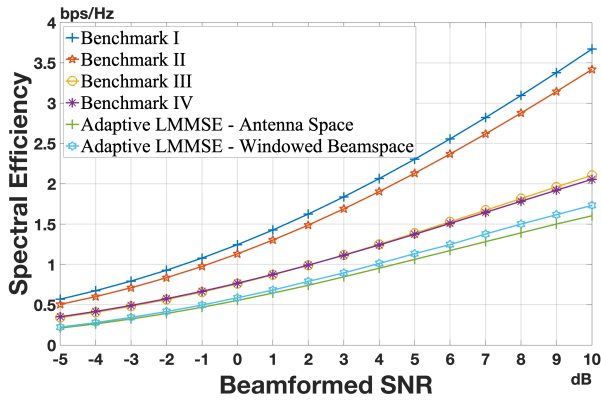
$$C = \frac{1}{K} \sum_{k=1}^K \log_2(1 + SINR_k). \quad (13)$$

V. CAPACITY COMPARISON AND DISCUSSION

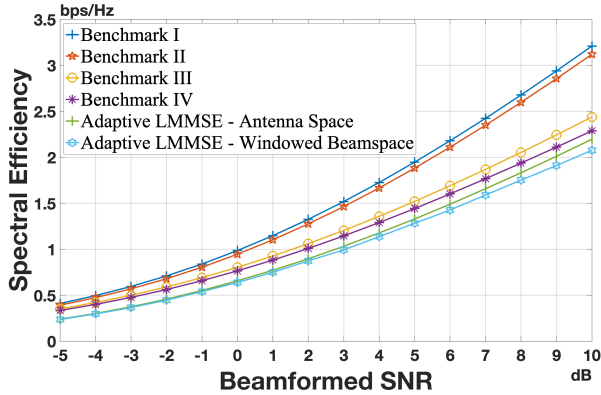
We compare a number of capacity benchmarks for the mmWave uplink channel described in Section II, where a 64-antenna BS communicates with 16 single-antenna UEs simultaneously over a bandwidth of 1 GHz. For each benchmark, we plot the capacity for two scenarios:

Multipath model: each user has multiple paths as described in Section II (Fig. 4a);

Single path model: only the dominant path for each user is retained in the model (Fig. 4b).



(a) Multipath scenario



(b) Single path scenario

Fig. 4: Capacity benchmarks, averaged across bandwidth and users. For each curve, $N = 64$ and $K = 16$, and the dominant path of each user is scaled to have the same beamformed SNR. Note that a per-user spectral efficiency of 1.5 bps/Hz on these curves, for example, translates to a spectral efficiency of $1.5K = 24$ bps/Hz aggregated over users.

Benchmark I (Capacity with unconstrained complexity):

For the MIMO channel expressed in the form of Eq. 2, the unconstrained-complexity capacity is calculated using Eq. 6. While this benchmark provides an upper bound on performance, it is impractical for several reasons. Ideal channel estimation for sparse multipath channels in which most of the paths are weak is practically infeasible, as is optimal space-time processing for multiuser detection, equalization and decoding. A direct application of MIMO-OFDM in parallel across subcarriers does not scale due to the difficulties of channel estimation and computational complexity. While channel estimation is less challenging for the single path model, the computational load of optimal multiuser detection and decoding still render this benchmark infeasible. It is worth noting that this benchmark is higher for the multipath model than for the single path model, since ideal unconstrained processing is able to optimally utilize all paths for all users. This is in contrast to the simple linear receivers considered later, which seek to utilize the dominant path for each user, treating all other paths as interference to be suppressed.

Benchmark II (Capacity with unconstrained complexity for a dimension-reduced windowed beamspace receiver):

This capacity benchmark is based on beamspace observations, taking the union of beamspace windows (with $W = 4$) across users. In principle, the choice of beamspace window for a user can vary across frequency. However, as in the linear receivers considered later, we use a frequency-independent window for each user, setting it based on the spatial frequency at the carrier frequency corresponding to the AoA for its dominant path. The achieved capacity is less than 1 dB below the unconstrained capacity benchmark for the multipath model and exhibits an even smaller loss for the single path model. This implies that beamspace dimension reduction in itself does not lead to significant loss in performance for mmWave channels, as long as we employ sophisticated enough space-time-frequency strategies for multiuser detection and equalization.

Benchmark III (LMMSE with ideal channel information):

Using the full-size received signal in the LMMSE calculations, a lower bound on the capacity is obtained by Eq. 13. When compared with the preceding capacity benchmarks, in the multipath scenario, we note a significant gap in performance due to restriction to linear processing that is focused on the dominant path for each user, treating the other paths for that user, as well as all paths for all other users, as interference.

Benchmark IV (Windowed beamspace LMMSE with ideal channel information):

We now calculate Eq. 13 employing the windowed beamspace received signal in LMMSE detection for each user in parallel. This corresponds to windowed beamspace “local” LMMSE reception [2] with ideal channel information. We use a beamspace window of size $W = 4$ around the dominant path for each user. For both the multipath and single path models, we observe no significant difference compared to Benchmark III. This suggests that, while restricting to linear processing of the dominant path causes a performance drop in Benchmark III relative to Benchmarks I and II, further dimension reduction via windowed beamspace does not detrimentally affect performance.

Adaptive windowed beamspace LMMSE: We now consider *adaptive* windowed beamspace local LMMSE reception [2], in which channel information is not available. Rather, the LMMSE correlators are trained in parallel for each user, using, as in Benchmark IV, a beamspace window of size $W = 4$ around the dominant path for each user. We obtain a lower bound on capacity per user (i.e. Eq. 13) based on the SINR obtained by its receiver. Because of drastic dimension reduction and parallelism, we are able to obtain an effective adaptive implementation using a training sequence of $N/2 = 32$ symbols. We note that the capacity is quite close to the Benchmark III in both scenarios, indicating that using an adaptive windowed beamspace approach does not lead to a significant performance loss, even with a small number of training symbols.

Adaptive antenna space LMMSE: Finally, we plot SINR-based lower bound on capacity obtained by Eq. 13 for an adaptive LMMSE implementation in antenna space using $10N = 640$ training symbols. For the multipath model, even

with a far greater number of training symbols, the capacity is *worse* than parallel beamspace adaptive LMMSE. This highlights that dimension reduction via beamspace, in addition to reducing complexity, also has the advantage of requiring far less training overhead. For the single path model, however, $10N = 640$ training symbols suffice for learning the LMMSE correlator effectively in antenna space, and the capacity is slightly better than that of the beamspace adaptive LMMSE receiver with $N/2 = 32$ training symbols.

VI. CONCLUSIONS

Our information-theoretic computations for measured channels at 28 GHz show that dimension reduction via beamspace is indeed an effective strategy for leveraging the sparsity of mmWave channels. There is only about a dB difference between the unconstrained capacity before and after dimensionality reduction. Similarly, the capacity of LMMSE detection focusing on the dominant component for ideal channel information is close to that of adaptive implementations of low-complexity windowed beamspace LMMSE, which enjoys the benefits of both reduced computational complexity and training overhead. However, there is a significant gap between the unconstrained capacity (with and without dimension reduction) and various flavors of LMMSE detection focused on the dominant component.

Part of the gap between unconstrained capacity and linear processing can be reduced by higher layer protocols that ensure a minimum angular separation between simultaneous users. For example, in our setting, enforcing a 3 degree minimum angular separation reduces the gap between the unconstrained capacity Benchmark I and the linear processing Benchmark III by about 2 dB at a spectral efficiency of 2 bps/Hz for the multipath model. However, there is still a substantial remaining gap, so that architecture and algorithm design for closing it are an important area for future research.

ACKNOWLEDGMENT

This work was supported in part by the Center for Ubiquitous Connectivity (CUBIC), sponsored by Semiconductor Research Corporation (SRC) and Defense Advanced Research Projects Agency (DARPA) under the JUMP 2.0 program, and in part by the National Science Foundation under grant CNS-2148303 under the RINGS program.

REFERENCES

- [1] M. L. Honig *et al.*, *Advances in multiuser detection*. Wiley Online Library, 2009.
- [2] M. Abdelghany, U. Madhow, and A. Tölli, "Beamspace local LMMSE: An efficient digital backend for mmwave massive MIMO," in *2019 IEEE 20th International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2019, pp. 1–5.
- [3] S. H. Mirfarshbafan and C. Studer, "Sparse beamspace equalization for massive MU-MIMO mmwave systems," in *ICASSP 2020-2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2020, pp. 1773–1777.
- [4] P. V. Amadori and C. Masouros, "Low RF-complexity millimeter-wave beamspace-MIMO systems by beam selection," *IEEE Transactions on Communications*, vol. 63, no. 6, pp. 2212–2223, 2015.
- [5] R. Pal, A. K. Chaitanya, and K. V. Srinivas, "Low-complexity beam selection algorithms for millimeter wave beamspace MIMO systems," *IEEE Communications Letters*, vol. 23, no. 4, pp. 768–771, 2019.
- [6] J. Han, C. Cebeci, W. Tang, Z. Zhang, and U. Madhow, "Tiled beamspace processing for scaling mmwave massive MU-MIMO," in *2024 IEEE 100th Vehicular Technology Conference (VTC2024-Fall)*, 2024, pp. 1–6.
- [7] R. Charbonnier, C. Lai, T. Tenoux, D. Caudill, G. Gougeon, J. Senic, C. Gentile, Y. Corre, J. Chuang, and N. Golmie, "Calibration of ray-tracing with diffuse scattering against 28-GHz directional urban channel measurements," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 12, pp. 14 264–14 276, 2020.
- [8] T. Someya and T. Ohtsuki, "SAGE algorithm for channel estimation and data detection with tracking the channel variation in MIMO system," in *IEEE Global Telecommunications Conference, 2004. GLOBECOM '04.*, vol. 6, 2004, pp. 3651–3655 Vol.6.