SPADE: Sparsity-Adaptive Equalization for MMwave Massive MU-MIMO

Conference Paper

Author(s):
Mirfarshbafan, Seyedhadi (iD); Studer, Christoph (iD)

SPADE: SPARSITY-ADAPTIVE EQUALIZATION FOR MMWAVE MASSIVE MU-MIMO

Seyed Hadi Mirfarshbafan and Christoph Studer

Department of Information Technology and Electrical Engineering
ETH Zürich, Switzerland; e-mail: mirfarshbafan@iis.ee.ethz.ch and studer@ethz.ch

ABSTRACT

We propose SParsity-ADaptive Equalization (SPADE), a novel approach to reduce the effective number of multiplications in sparse inner products by adaptively skipping multiplications that have little to no effect on the result. We apply SPADE to beamspace linear minimum mean square error (LMMSE) spatial equalization in all-digital millimeter-wave (mmWave) massive multiuser multiple-input multiple-output (MU-MIMO) systems. We propose a SPADE-based architecture that mutes insignificant multiplications to offer power savings. We use simulation results with line-of-sight (LoS) and non-LoS mmWave channel models to demonstrate that SPADE-LMMSE performs on par with state-of-the-art beamspace equalizers in terms of bit error-rate, while requiring significantly lower preprocessing complexity.

1. INTRODUCTION

Millimeter-wave (mmWave) communication [1,2] and massive multiuser (MU) multiple-input multiple-output (MIMO) [3] are two core technologies of fifth-generation (5G) and beyond 5G wireless systems. While mmWave communication provides access to large portions of unused bandwidth, it suffers from high propagation losses [4]. MU-MIMO [3] is able to (i) compensate for the high propagation losses via fine-grained beamforming and (ii) support simultaneous communication with multiple user equipments (UEs) in the same frequency band. However, the large number of basestation (BS) antennas combined with the high baseband sampling rates create new challenges for mmWave massive MU-MIMO hardware design.

While hybrid digital-analog architectures [5, 6] can reduce the hardware complexity, all-digital architectures with low-resolution data converters [7–9] and low-resolution baseband processing [10] achieve higher spectral efficiency, provide more flexibility, and simplify radio-frequency (RF) circuitry and baseband processing [11, 12]. However, to reduce the complexity and power consumption of baseband processing in all-digital BS designs, novel algorithms and architectures are necessary. Fortunately, mmWave channels typically consist of only a few dominant propagation paths [1, 4], which can be exploited to simplify BS design [13–18]. Concretely, by taking

a spatial discrete Fourier transform (DFT) over the antenna array, which converts the antenna domain into beamspace domain, one can drastically simplify some of the most complex baseband processing tasks, including channel estimation [18] and spatial equalization [19–21]. Existing beamspace equalizers either suffer from a notable performance degradation compared to antenna domain spatial equalizers [20] or result in high preprocessing complexity [21].

1.1. Contributions

We propose SParsity-ADaptive Equalization (SPADE), a novel scheme to reduce the effective number of multiplications in sparse inner products. SPADE compares the entries of the two vectors and adaptively skips scalar multiplications which have negligible impact on the final result. When applied to beamspace linear minimum mean square error (LMMSE) equalization, SPADE offers significant reductions in the number of multiplications, without increasing the preprocessing complexity of LMMSE. We propose an architecture for SPADE-LMMSE that adaptively mutes multiplications to reduce power consumption. Furthermore, we use simulations with line-of-sight (LoS) and non-LoS mmWave channels to demonstrate that SPADE performs on par with the state-of-the-art beamspace equalization algorithms.

1.2. Notation

Boldface lowercase and uppercase letters represent column vectors and matrices, respectively. For a matrix $\mathbf{A}$, the transpose and Hermitian transpose are $\mathbf{A}^\mathsf{T}$ and $\mathbf{A}^\mathsf{H}$, respectively, the kth column is denoted by $\mathbf{a}_k$, and the Frobenius norm is $\|\mathbf{A}\|_F$. For a vector $\mathbf{a}$, the kth entry is denoted by a_k, the real and imaginary parts are denoted by $\mathbf{a}^\mathcal{R}$ and $\mathbf{a}^\mathcal{I}$, respectively, and the ℓ_∞-norm and $\ell_{\widetilde{\infty}}$-norm are defined as $\|\mathbf{a}\|_\infty \triangleq \max_k |a_k|$ and $\|\mathbf{a}\|_{\widetilde{\infty}} \triangleq \max\{\|\mathbf{a}^\mathcal{R}\|_\infty, \|\mathbf{a}^\mathcal{I}\|_\infty\}$, respectively [22]. The $N \times N$ identity and the unitary $N \times N$ discrete Fourier transform (DFT) matrices are denoted by $\mathbf{I}_N$ and $\mathbf{F}_N$, respectively.

2. BACKGROUND

2.1. Antenna-Domain and Beamspace System Models

We focus on an all-digital mmWave massive MU-MIMO uplink system as depicted in Figure 1. We assume frequency-flat
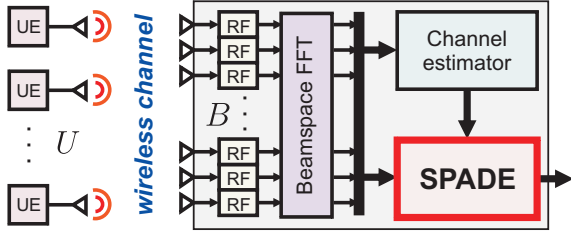
Fig. 1. Beamspace massive MU-MIMO uplink system with a sparsity-adaptive equalization (SPADE) receiver.

channels in which U single-antenna user equipments (UEs) transmit information simultaneously to a B-antenna basestation (BS) in the same frequency band. We model the *antenna domain* received signal vector $\bar{\mathbf{y}} \in \mathbb{C}^B$ at the BS as $\bar{\mathbf{y}} = \bar{\mathbf{H}}\mathbf{s} + \bar{\mathbf{n}}$, where $\bar{\mathbf{H}} \in \mathbb{C}^{B \times U}$ is the uplink channel matrix, $\mathbf{s} \in \mathcal{S}^U$ is the vector of the UEs' data symbols taken from a discrete constellation set $\mathcal{S}$ (e.g., 16-QAM) with $E_s \triangleq \mathbb{E}\left[|s_u|^2\right]$, $u = 1, \ldots, U$. The noise vector $\bar{\mathbf{n}} \in \mathbb{C}^B$ has i.i.d. circularly-symmetric complex Gaussian entries with variance N_0. The average receive signal-to-noise ratio (SNR) at the BS is defined as $SNR \triangleq E_s \|\bar{\mathbf{H}}\|_F^2 / (BN_0)$.

In what follows, we consider mmWave propagation conditions with a B-antenna uniform linear array (ULA). To obtain the *beamspace* input-output relation, one applies a spatial DFT [23] to the received antenna-domain vector $\bar{\mathbf{y}}$ as follows:

$$\mathbf{y} = \mathbf{F}_B \bar{\mathbf{y}} = \mathbf{H}\mathbf{s} + \mathbf{n}. \tag{1}$$

Here, $\mathbf{y} = \mathbf{F}_B \bar{\mathbf{y}}$ is the beamspace receive vector, $\mathbf{H} = \mathbf{F}_B \bar{\mathbf{H}}$ is the (typically sparse) beamspace channel matrix, and $\mathbf{n}$ is the beamspace noise vector with the same statistics as the antenna-domain noise vector $\bar{\mathbf{n}}$. Since the beamspace DFT must be computed for every receive vector $\bar{\mathbf{y}}$, a practical system would rely on streaming fast Fourier transforms (FFTs) [23].

To see why the beamspace transform sparsifies mmWave channel matrices, consider the following plane-wave model for the antenna-domain channel vector $\bar{\mathbf{h}}_u$ of the uth UE [24]:

$$\bar{\mathbf{h}}_u = \sum_{\ell=0}^{L-1} \alpha_\ell \bar{\mathbf{a}}(\phi_\ell). \tag{2}$$

Here, L stands for the number of propagation paths , $\alpha_l \in \mathbb{C}$ is the channel gain of the ℓth propagation path, and

$$\bar{\mathbf{a}}(\phi_\ell) = \left[1, e^{j\phi_\ell}, e^{j2\phi_\ell}, \ldots, e^{j(B-1)\phi_\ell}\right], \tag{3}$$

where the spatial frequency ϕ_ℓ is determined by the ℓth path's incident angle to the ULA. Since L is small for line-of-sight (LoS) mmWave channels, taking a DFT $\mathbf{h}_u = \mathbf{F}_B \bar{\mathbf{h}}_u$ reveals the sparse structure of such channel vectors, i.e., most of the vector's energy is concentrated on the entries of $\mathbf{h}_u$ associated with the spatial frequencies ϕ_ℓ.

2.2. Beamspace LMMSE Equalization

Linear data detection consists of two phases: (i) *preprocessing*, which is performed only once per channel coherence interval

and produces a $U \times B$ equalization matrix $\mathbf{W}^{\mathrm{H}}$ and (ii) *spatial equalization*, which is performed at baseband sampling rate (for each received signal vector $\mathbf{y}$) in order to obtain estimates of the transmitted symbol vectors according to $\hat{\mathbf{s}} = \mathbf{W}^{\mathrm{H}}\mathbf{y}$. In what follows, we focus on beamspace LMMSE equalization for which the equalization matrix is given by

$$\mathbf{W}^{\mathrm{H}} = \left(\mathbf{H}^{\mathrm{H}}\mathbf{H} + \frac{N_0}{E_s}\mathbf{I}_U\right)^{-1}\mathbf{H}^{\mathrm{H}}. \tag{4}$$

In order to support high-bandwidth communication at mmWave frequencies, the spatial equalization stage must be carried out at extremely high baseband sampling rates. Hence, to keep power consumption within reasonable bounds, efficient methods to calculate $\hat{\mathbf{s}} = \mathbf{W}^{\mathrm{H}}\mathbf{y}$ must be deployed in practice.

2.3. Existing Sparsity-Exploiting Spatial Equalizers

In recent years, a number of sparsity-exploiting beamspace equalization methods have been proposed [19–21]. All of these methods exploit the fact that for sparse beamspace channel matrices $\mathbf{H}$, the associated LMMSE equalization matrices $\mathbf{W}^{\mathrm{H}}$ tend to be sparse as well. This observation enables design of beamspace equalization algorithms that produce *sparse* equalization matrices $\tilde{\mathbf{W}}^{\mathrm{H}}$ with a given *density coefficient* $\delta \triangleq \|\tilde{\mathbf{W}}^{\mathrm{H}}\|_0 / (BU)$, where $\|\tilde{\mathbf{W}}\|_0$ is the number of nonzero entries of $\tilde{\mathbf{W}}$. Such sparsity-exploiting spatial equalizers reduce the number of multiplications required when calculating $\hat{\mathbf{s}} = \tilde{\mathbf{W}}^{\mathrm{H}}\mathbf{y}$, which reduces power consumption and/or implementation complexity. Among such methods, the entry-wise orthogonal matching pursuit (EOMP) proposed in [21] achieves the highest sparsity (lowest δ) and hence highest complexity reduction during spatial equalization. EOMP and related algorithms, however, considerably increase the complexity of preprocessing, resulting in inefficient circuitry. We next propose SPADE-LMMSE, a novel sparsity-adaptive spatial equalization method that competes with EOMP in terms of complexity reduction during spatial equalization while directly using the conventional LMMSE equalization matrix (4) which means that the preprocessing complexity does not increase.

3. SPADE: SPARSITY-ADAPTIVE EQUALIZATION

Consider the inner product of two B-dimensional real-valued vectors $\langle \mathbf{w}, \mathbf{y} \rangle = \sum_{b=1}^{B} w_b y_b$. Intuitively, if B is large, then we can skip partial products $w_b y_b$ of small magnitude, without incurring a large relative error in the result, assuming that the exact result is bounded away from zero. However, we cannot eliminate partial products based on their magnitude, as this requires actually performing the multiplication. Therefore, we propose to set thresholds for the absolute values of the two operands w_b and y_b, and skip (or mute) multiplications if the absolute values of both operands are below their respective thresholds. The same approach can be extended to complex-valued case, noting that each complex-valued inner product can be decomposed into four real-valued inner products. Since for mmWave channels, the rows $\mathbf{w}_u^{\mathrm{H}}, u = 1, \ldots, U$ of beamspace

LMMSE matrices $\mathbf{W}^H$ in (4) and the receive vectors $\mathbf{y}$ exhibit sparsity, a large number of partial products when computing $\hat{s}_u = \mathbf{w}_u^H \mathbf{y}$, $u = 1, \ldots, U$, will be small; hence, a large number of multiplications can be skipped—this is the general idea behind SPADE. In order to simplify hardware implementation of SPADE, we propose to use two fixed thresholds $\tau_y \in \mathbb{R}$ and $\tau_w \in \mathbb{R}$, for the real and imaginary parts of all entries of the receive vector $\mathbf{y}$ and the equalization matrix $\mathbf{W}^H$, respectively.

3.1. Setting the SPADE Thresholds

The thresholds τ_y and τ_w are used to trade arithmetic precision for reduction in the effective number of multiplications. Setting these thresholds close to zero will result in high precision, but will increase the number of active multiplications. In contrast, setting these thresholds to large values will result in precision loss, but will lower the number of active multiplications. To improve the effectiveness of SPADE, we propose the following techniques:

1) Since we use a single threshold τ_y for all the entries of $\mathbf{y}$, and the statistics of $\mathbf{y}$ change dynamically, we propose to determine a fixed threshold τ_y' and then set $\tau_y = \|\mathbf{y}\|_{\widetilde{\infty}} \tau_y'$ in order to incorporate fluctuations of $\mathbf{y}$ into τ_y.

2) Since we use a single threshold τ_w for all the entries of $\mathbf{W}^H$, we scale the rows $\mathbf{w}_u^H$, $u = 1, \ldots, U$, so that their $\ell_{\widetilde{\infty}}$-norms are equal. Such a row-scaling scheme was proposed in [10] to reduce the dynamic range of the entries of equalization matrices. For SPADE, we scale the rows of the LMMSE equalization matrix according to $\mathbf{V}^H = \mathrm{diag}(\boldsymbol{\alpha}) \mathbf{W}^H$ such that the rows $\mathbf{v}_u^H$, $u = 1, \ldots, U$, of the scaled matrix $\mathbf{V}^H$ satisfy $\|\mathbf{v}_u\|_{\widetilde{\infty}} < 1$. This can be accomplished by $\alpha_u = 1/(\|\mathbf{w}_u^H\|_{\widetilde{\infty}} + \varepsilon)$, where $\varepsilon > 0$ is a small constant that ensures that $\|\mathbf{v}_u\|_{\widetilde{\infty}}$ is just below one. With this approach, estimates of the transmit vector are computed as $\hat{\mathbf{s}} = \mathrm{diag}(\boldsymbol{\alpha})^{-1} \mathbf{V}^H \mathbf{y}$, which corresponds to post-multiplying the uth entry of $\mathbf{V}^H \mathbf{y}$ by $1/\alpha_u$ for $u = 1, \ldots, U$. The threshold τ_w is applied to the entries of the scaled matrix $\mathbf{V}^H$.

3.2. SPADE-Based Architecture

Figure 2 shows a high-level architecture of a fully-unrolled beamspace LMMSE equalizer employing SPADE to adaptively mute multipliers. We emphasize that in high-bandwidth mmWave systems with multi-GHz baseband sampling rates, fully-unrolled architectures become a natural choice to deliver the desired throughput while minimizing data buffering and control overhead [23]. In Figure 2, the LMMSE preprocessing block receives beamspace channel estimates and computes $\mathbf{V}^H$ once per channel coherence interval. This block also performs the comparison of real and imaginary entries of $\mathbf{V}^H$ with τ_w and provides the comparison bits to the SPADE-based equalizer block. The FFT block takes the antenna domain received vectors $\bar{\mathbf{y}}$ and produces the beamspace domain vectors $\mathbf{y}$, along with the comparison bits for each real and imaginary part of entries of $\mathbf{y}$ with τ_y. The fully-unrolled matrix-vector multiplication engine comprises U inner product engines, each consisting
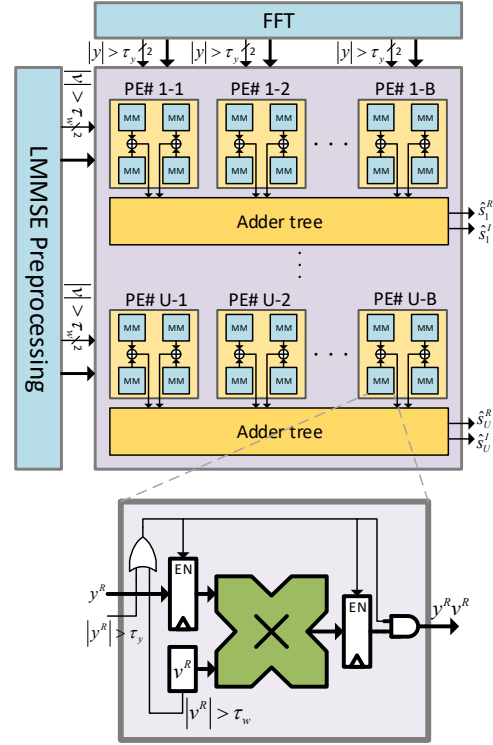

Fig. 2. High-level architecture of a SPADE-LMMSE equalizer (top) with mutable multiplier (MM) details (bottom).

of B processing elements (PEs). Each PE is a complex-valued multiplier with four real-valued mutable multipliers (labeled as "MM"), whose internal architecture is depicted in Figure 2. At the beginning of each channel coherence interval, the weight registers of the PEs are loaded with the entries of $\mathbf{V}^H$ and the comparison bits provided by the LMMSE preprocessing block. Then, for each beamspace receive vector coming from the FFT block, each MM conditionally disables the registers before and after the multiplier if the comparison bits indicate that both operands have absolute values smaller than their thresholds. Consequently, such a matrix-vector multiplication engine adaptively saves power by nulling the switching activity of unused multipliers—this will reduce the dynamic power consumption. For correct functionality, the output of each muted multiplier must be set to zero, which is implemented by the AND gate at the output of each MM as shown in Figure 2. The number of muted multipliers depends on the sparsity of $\mathbf{y}$ and $\mathbf{v}_u^H$, $u = 1, \ldots, U$, which is determined by the channel conditions. For example, as shown in Section 4, in a system with $B = 128$ BS antennas and $U = 16$ UEs with LoS channel conditions, it is possible to mute 80% of the multiplications while incurring no more than $0.1\,\mathrm{dB}$ SNR loss at an uncoded bit-error-rate (BER) of 1%.

We reiterate that LMMSE preprocessing is carried out only once per channel coherence interval and the spatial FFT can be implemented efficiently using the fully-unrolled multiplierless architecture proposed in [23]. However, the matrix-vector multiplication engine must operate constantly and at baseband
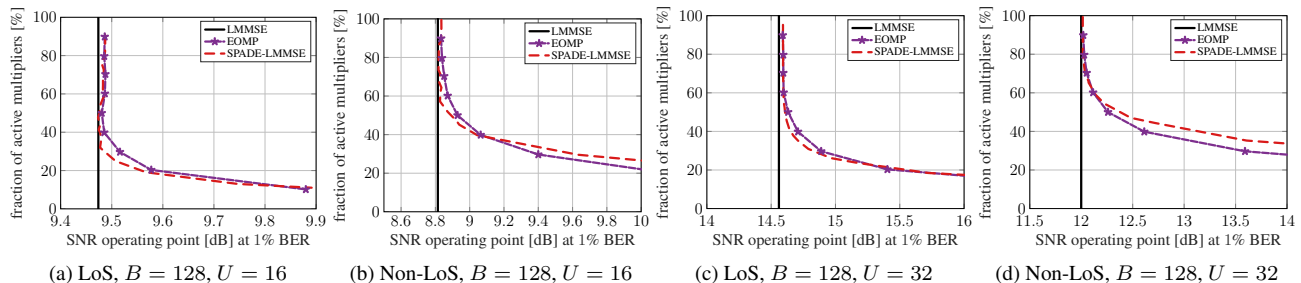
Fig. 3. Trade-off between SNR operating point at 1% uncoded BER and multiplier activity rate.

sampling rate, which emphasizes the importance of saving power consumption in the spatial equalization block.

4. SIMULATION RESULTS

To evaluate the performance-complexity trade-off offered by SPADE-LMMSE, we provide simulation results in Figure 3. We consider a mmWave massive MU-MIMO basestation with a $B = 128$ antenna $\lambda/2$-spaced ULA and 16-QAM transmission from $U \in \{16, 32\}$ single-antenna UEs. We carried out Monte-Carlo simulations with LoS and non-LoS channels generated by the QuaDRiGa mmMAGIC UMi model [25] at a carrier frequency of 60 GHz. The UEs are placed randomly in a $120°$ sector within 10 m to 110 m from the BS array, with a minimum of $1°$ angular separation. We use power control so that the variation in receive power of UEs is limited to ± 3 dB. For channel estimation, we used pilot-based least squares (LS) followed by BEACHES [18]. We simulated the uncoded BER for (i) conventional antenna domain LMMSE equalization, (ii) EOMP-based beamspace equalization with density coefficient δ ranging from 10% to 90%, and (iii) beamspace SPADE-LMMSE with different values for the threshold pair (τ'_y, τ_w). The values of the threshold pair (τ'_y, τ_w) corresponding to each simulated point of the SPADE-LMMSE curves in Figure 3 are obtained via a 2-dimensional grid search over a range of plausible candidates and picking the Pareto-optimal parameters.

4.1. Comparison with EOMP

Figure 3 shows the SNR operating point required to achieve an uncoded BER of 1% versus the fraction of active multipliers (which coincides with the density coefficient δ of EOMP). We see that the performance of SPADE-LMMSE is very close to EOMP, meaning that for a given fraction of active multipliers, they both achieve an uncoded BER of 1% at similar SNRs. We also see that SPADE-LMMSE and EOMP require fewer active multipliers for LoS channels than for non-LoS channels. For a negligible SNR loss (compared to antenna domain LMMSE), up to 80% of multipliers can be muted in a 128×16 LoS scenario, whereas about 50% can be muted for a non-LoS scenario. Note that the SNR operating point required to achieve an even lower uncoded BER of 0.1% behaves very similarly, but is omitted due to space constraints.

A key advantage of SPADE-LMMSE over EOMP is the fact that EOMP requires external control of the density coefficient δ dependent on the channel conditions to avoid a large SNR loss. In contrast, SPADE-LMMSE can simply use the same pair of threshold values (τ'_y, τ_w) while automatically adapting to the required number of multipliers based on the instantaneous channel conditions. For example, in the 128×16 LoS setting, choosing $\tau'_y = 1/2$ and $\tau_w = 1/20$ results in an SNR gap of only 0.25 dB with respect to antenna domain LMMSE equalization while requiring only 17% active multipliers. The same threshold parameters in the non-LoS setting results in an SNR gap of less than 0.15 dB while requiring only 45% active multipliers. Another key advantage is that SPADE simply uses conventional LMMSE preprocessing whereas the preprocessing complexity of EOMP is considerably higher. Using the complexity expressions provided in [21], the preprocessing complexity of EOMP—measured in terms of the number of real-valued multiplications—is about $15\times$ higher than that of conventional LMMSE for a $B = 128$ BS antenna and $U = 16$ UE system with density factor $\delta = 20\%$.

5. CONCLUSIONS

We have proposed SParsity-ADaptive Equalization (SPADE), a novel spatial equalization approach that adaptively reduces the number of multiplications based upon the instantaneous channel conditions. SPADE offers the following advantages: (i) the preprocessing complexity is the same as the conventional LMMSE equalization, (ii) the performance degradation with respect to antenna-domain equalization is negligible for suitably chosen thresholds, and (iii) the method adaptively reduces complexity and lowers power based on the instantaneous channel conditions. For LoS and non-LoS mmWave massive MU-MIMO channel models, we have demonstrated that SPADE performs on par with EOMP [21], but requires significantly lower preprocessing complexity. We emphasize that SPADE is not only suitable for spatial equalization in mmWave massive MU-MIMO systems, but finds potential use in many other applications that carry out approximate sparse matrix-vector products. A hardware implementation of the proposed SPADE-LMMSE is ongoing work.

6. REFERENCES

[1] T. S. Rappaport, S Sun, R. Mayzus, H. Zhao, Y. Azar, K. Wang, G. N. Wong, J. K. Schulz, M. Samimi, and F. Gutierrez, "Millimeter wave mobile communications for 5G cellular: It will work!," *IEEE Access*, vol. 1, pp. 335–349, May 2013.

[2] T. S. Rappaport, Y. Xing, G. R. MacCartney, A. F. Molisch, E. Mellios, and J. Zhang, "Overview of millimeter wave communications for fifth-generation (5G) wireless networks—with a focus on propagation models," *IEEE Trans. Antennas Propag.*, vol. 65, no. 12, pp. 6213–6230, Dec. 2017.

[3] E. G. Larsson, F. Tufvesson, O. Edfors, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[4] T. S. Rappaport, Robert W. Heath Jr., R. C. Daniels, and J. N. Murdock, *Millimeter Wave Wireless Communications*, Prentice Hall, 2015.

[5] F. Sohrabi and W. Yu, "Hybrid digital and analog beamforming design for large-scale antenna arrays," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 501–513, Apr. 2016.

[6] R. W. Heath Jr., N. González Prelcic, S. Rangan, W. Roh, and A. Sayeed, "An overview of signal processing techniques for millimeter wave MIMO systems," *IEEE J. Sel. Topics Signal Process.*, vol. 10, no. 3, pp. 436–453, Feb. 2016.

[7] C. Risi, D. Persson, and E. G. Larsson, "Massive MIMO with 1-bit ADC," *arXiv preprint: 1404.7736*, Apr. 2014.

[8] C. Studer and G. Durisi, "Quantized massive MU-MIMO-OFDM uplink," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2387–2399, June 2016.

[9] S. H. Mirfarshbafan, S. A. Nezamalhosseini, M. Shabany, and C. Studer, "Algorithm and VLSI design for 1-bit data detection in massive MIMO-OFDM," *IEEE Open J. Circuits Syst. I*, vol. 1, pp. 170–184, Sept. 2020.

[10] O. Castañeda, S. Jacobsson, G. Durisi, T. Goldstein, and C. Studer, "Finite-alphabet MMSE equalization for all-digital massive MU-MIMO mmWave communication," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 9, pp. 2128–2141, June 2020.

[11] H. Yan, S. Ramesh, T. Gallagher, C. Ling, and D. Cabric, "Performance, power, and area design trade-offs in millimeter-wave transmitter beamforming architectures," *IEEE Circuits Syst. Mag.*, vol. 19, no. 2, pp. 33–58, May 2019.

[12] P. Skrimponis, S. Dutta, M. Mezzavilla, S. Rangan, S. H. Mirfarshbafan, C. Studer, J. Buckwalter, and M. Rodwell, "Power consumption analysis for mobile mmwave and sub-THz receivers," in *2020 IEEE 6G Wireless Summit*, Mar. 2020, pp. 1–5.

[13] A. Alkhateeb, O. El Ayach, G. Leus, and R. W. Heath Jr., "Channel estimation and hybrid precoding for millimeter wave cellular systems," *IEEE J. Sel. Topics Signal Process.*, vol. 8, no. 5, pp. 831–846, Oct. 2014.

[14] P. Schniter and A. Sayeed, "Channel estimation and precoder design for millimeter-wave communications: The sparse way," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Nov. 2014, pp. 273–277.

[15] J. Deng, O. Tirkkonen, and C. Studer, "mmWave channel estimation via atomic norm minimization for multi-user hybrid precoding," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2018, pp. 1–6.

[16] A. Alkhateeb, G. Leus, and Robert W. Heath Jr., "Limited feedback hybrid precoding for multi-user millimeter wave systems," *IEEE Trans. Wireless Commun.*, vol. 14, no. 11, pp. 6481–6494, Nov. 2015.

[17] J. Lee, G. Gil, and Y. H. Lee, "Channel estimation via orthogonal matching pursuit for hybrid MIMO systems in millimeter wave communications," *IEEE Trans. Commun.*, vol. 64, no. 6, pp. 2370–2386, June 2016.

[18] S. H. Mirfarshbafan, A. Gallyas-Sanhueza, R. Ghods, and C. Studer, "Beamspace channel estimation for massive MIMO mmWave systems: Algorithm and VLSI design," *IEEE Trans. Circuits Syst. I*, pp. 1–14, Sept. 2020.

[19] M. Abdelghany, U. Madhow, and A. Tölli, "Beamspace local LMMSE: An efficient digital backend for mmWave massive MIMO," in *Proc. IEEE Int. Workshop Signal Process. Advances Wireless Commun. (SPAWC)*, Aug. 2019, pp. 1–5.

[20] M. Mahdavi, O. Edfors, V. Öwall, and L. Liu, "Angular-domain massive MIMO detection: Algorithm, implementation, and design tradeoffs," *IEEE Trans. Circuits Syst. I*, vol. 67, no. 6, pp. 1948–1961, Jan. 2020.

[21] S. H. Mirfarshbafan and C. Studer, "Sparse beamspace equalization for massive MU-MIMO mmWave systems," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. (ICASSP)*, May 2020, pp. 1773–1777.

[22] D. Seethaler and H. Bölcskei, "Performance and complexity analysis of infinity-norm sphere-decoding," *IEEE Trans. Inf. Theory*, vol. 56, no. 3, pp. 1085–1105, Mar. 2010.

[23] S. H. Mirfarshbafan, S. Taner, and C. Studer, "SMUL-FFT: a streaming multiplierless fast Fourier transform," *IEEE Trans. Circuits Syst. II*, vol. 68, no. 5, pp. 1715–1719, Mar. 2021.

[24] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*, Cambridge Univ. Press, 2005.

[25] S. Jaeckel, L. Raschkowski, K. Börner, L. Thiele, F. Burkhardt, and E. Eberlein, "QuaDRiGa - quasi deterministic radio channel generator user manual and documentation," Tech. Rep. v2.0.0, Fraunhofer Heinrich Hertz Institute, Aug. 2017.