

Distributed Architecture for XL-Metasurface Beamforming

Kofi Anane Boateng*, Di Zhang[†], and Kyungchun Lee*

*Department of Electrical and Information Engineering and the Research Center for Electrical and Information Technology, Seoul National University of Science and Technology, Seoul 01811, Republic of Korea. Email: kaboateng, kcllee@seoultech.ac.kr

[†]School of Intelligent Systems Engineering, Sun Yat Sen University, Shenzhen, China. Email: zhangd263@mail.sysu.edu.cn

Abstract—This paper presents a distributed beamforming framework for extremely large (XL) dynamic metasurface antenna (DMA) systems, offering an energy-efficient and scalable solution for 6G deployments. The transmit array is partitioned into clusters, each managed by a distributed node (DN) that estimates local channel state information (CSI) and designs DMA weights using global CSI provided by a central node (CN). The CN also performs digital precoding based on the global CSI. A downlink capacity maximization problem is formulated, and an efficient joint optimization algorithm for digital precoding and DMA weights is developed. Simulations show that the proposed approach achieves near-optimal spectral efficiency with substantially reduced interconnection overhead, offering a practical solution for scalable 6G deployments.

Index Terms—Distributed beamforming, downlink communication, extremely large antenna array (ELAA), and massive multiple-input multiple-output (mMIMO).

I. INTRODUCTION

The limitations of fifth-generation (5G) networks in supporting emerging applications such as extended reality (XR), holographic communication, the internet of everything (IoE), and autonomous systems have driven the evolution toward sixth-generation (6G) wireless networks. These applications demand hyper-reliable, low-latency, and high-capacity communication beyond the capabilities of the current 5G systems. Massive multiple-input multiple-output (mMIMO) has been a key technology in 5G, offering gains in spectral and energy efficiency through large-scale antenna arrays [1]. Along this line of thought, to meet the stringent requirements of 6G, extremely large antenna array (ELAA) systems have been proposed. ELAA extends the mMIMO concept by deploying hundreds to thousands of antenna elements, either centrally or distributed across an environment in a coordinated manner, enabling enhanced spatial multiplexing and significantly improved spectral efficiency [2], [3].

This large-scale deployment introduces critical challenges, including increased power consumption, hardware complexity, and limited scalability in signal processing. Centralized

baseband precoding (CBP), where all signal processing is performed at a central unit, leads to excessive interconnection overhead. Studies have shown that the raw data throughput in such systems can exceed 1 Tbps, surpassing the capacity of current base station interconnect standards like enhanced common public radio interface (eCPRI) [4]–[6]. This limitation renders CBP impractical for ELAA, motivating the shift toward distributed baseband precoding (DBP), which partitions processing across multiple nodes to reduce overhead and enable scalable ELAA implementation in 6G networks [4], [7].

DBP partitions the antenna array into multiple clusters, each paired with a dedicated baseband unit (BBU), forming distributed nodes (DNs). These DNs are interconnected through topologies such as star or daisy-chain to enable coordination. This decentralized architecture reduces interconnection overhead and computational complexity by enabling local processing of lower-dimensional signals, thereby improving scalability and robustness. DBP has been investigated in the context of massive MIMO systems [4]–[6], [8], where it has shown considerable advantages in minimizing interconnection load and reducing processing complexity compared to centralized approaches. Most of these solutions, however, employ fully digital architectures that have a dedicated radio frequency (RF) chain to each antenna element. While this configuration enables high spatial resolution and flexible digital control, the resulting hardware cost and power consumption strongly escalate with the array size, limiting its feasibility for ELAA deployments.

Dynamic metasurface antennas (DMAs), which leverage sub-wavelength spacing and real-time reconfigurability of metamaterial elements for efficient signal transmission and reception [9]–[11], have emerged as a cost-effective and energy-efficient alternative to fully digital transceiver architectures. By eliminating the need for a dedicated RF chain per antenna element, DMAs significantly reduce hardware complexity and power consumption in large-scale systems. However, prior works on large-scale DMA deployments [12]–[14] rely on centralized precoding architectures, which result in substantial data storage demands, increased computational complexity, and significant interconnection overhead.

Contributions: This paper proposes a distributed beamform-

This work was supported in part by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education under Grant RS-2019-NR040071 and in part by the NRF grant funded by the Korean Government (MSIT)(RS-2025-21812978). The study of Di Zhang was supported by the Henan Natural Science Foundation for Excellent Young Scholar under Grant 242300421169.

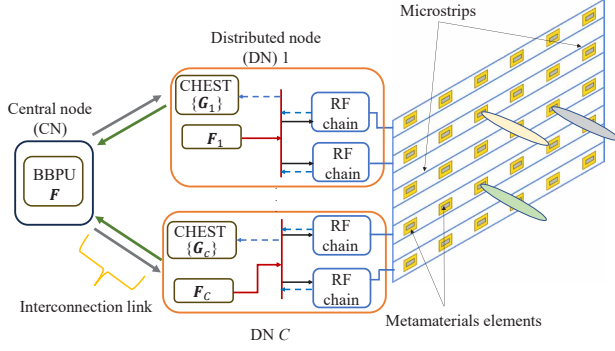


Fig. 1. Distributed star architecture for XL-DMA.

ing architecture for XL-DMA systems to reduce interconnection overhead while achieving near-optimal system capacity. A downlink point-to-point (P2P) capacity maximization problem is formulated, and an efficient distributed beamforming algorithm is developed. In the proposed framework, the central node (CN) collects global channel state information (CSI) from DNs and designs the digital precoder. Each DN then independently computes its DMA weights using the global CSI shared by the CN.

Notations: Scalars, vectors, and matrices are denoted by lower-case italics letters, boldface lower-case letters, and boldface upper-case letters, respectively. For a scalar a , $|a|$ and $\arg(a)$ represent its magnitude and argument, respectively. For a vector $\mathbf{a}$, the i -th element is denoted by $[\mathbf{a}]_i$. For a matrix $\mathbf{A}$, the element in the i -th row and j -th column is denoted by $[\mathbf{A}]_{i,j}$. Also, $|\mathbf{A}|$, $\|\mathbf{A}\|_F$, and $\mathbf{A}^{-1}$ denote its determinant, Frobenius norm, and inverse, respectively. The operators $(\cdot)^*$, $(\cdot)^T$, and $(\cdot)^H$ represent the complex conjugate, transpose, and conjugate transpose, respectively. The identity matrix of size $N \times N$ is denoted by $\mathbf{I}_N$, and the operator $\text{diag}(\cdot)$ forms a diagonal matrix using its argument as the main diagonal.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. Proposed architecture description

The proposed system architecture, depicted in Fig. 1, features a BS equipped with XL-DMA, partitioned into $C \geq 1$ independent clusters. Each cluster contains $N_c = N_{d,c} \times N_{e,c}$ radiating elements, where $N_{d,c}$ and $N_{e,c}$ are the number of microstrips and metamaterial elements within the c -th cluster, respectively. The system-wide totals for radiating elements, microstrips, and metamaterial elements are given by $N = \sum_{c=1}^C N_c$, $N_d = \sum_{c=1}^C N_{d,c}$, and $N_e = \sum_{c=1}^C N_{e,c}$.

Each cluster is paired with dedicated RF and baseband processing hardware, collectively referred to as a DN, which performs local channel estimation via the channel estimation unit (CHEST) and DMA weights design. The DNs are linked to a CN via bidirectional communication links, enabling coordinated transmission. The CN aggregates local CSI from all DNs to form global CSI, which it uses to design digital precoders. These precoders are then forwarded to the corresponding DNs. This architecture enables efficient coordination with reduced

computational load and communication overhead, while still leveraging the benefits of centralized digital beamforming and decentralized DMA weights design.

B. Problem formulation

For a single user with M antennas served by the BS equipped with the proposed architecture in the downlink, the received signal can be expressed as

$$\mathbf{y} \in \mathbb{C}^{M \times 1} = \sum_{c=1}^C \mathbf{G}_c \mathbf{H}_c \mathbf{Q}_c \mathbf{F}_c \mathbf{s} + \mathbf{n}, \quad (1)$$

where $\mathbf{G}_c \in \mathbb{C}^{M \times N_c}$ is the channel between cluster c at the BS and the user, $\mathbf{H}_c \in \mathbb{C}^{N_c \times N_c}$ is the diagonal matrix representing the signal propagation effect of the DMA in cluster c , $\mathbf{Q}_c \in \mathbb{C}^{N_c \times N_{d,c}}$ is the block diagonal matrix containing the DMA weights within cluster c , $\mathbf{F}_c \in \mathbb{C}^{N_{d,c} \times N_s}$ is the transmit precoder for cluster c for data stream $\mathbf{s} \in \mathbb{C}^{N_s \times 1}$, and $\mathbf{n} \sim \mathcal{CN}(0, \sigma^2 \mathbf{I})$ is the circularly symmetric complex additive white Gaussian noise with zero mean and variance $\sigma^2 \mathbf{I}$ at the receiver. The structures of $\mathbf{H}_c$ and $\mathbf{Q}_c$ defined according to [10]. To maximize the capacity of the system, we design the transmit precoders for all clusters $\mathbf{F} = [\mathbf{F}_1, \mathbf{F}_2, \dots, \mathbf{F}_C]^T$ and DMA weights for all clusters $\mathbf{Q} = [\mathbf{Q}_1, \mathbf{Q}_2, \dots, \mathbf{Q}_C]$. The capacity maximization problem is formulated as

$$\max_{\{\mathbf{F}, \mathbf{Q}\}} \log_2 \left| \mathbf{I} + \frac{1}{\sigma^2} \sum_{c=1}^C \mathbf{G}_c \mathbf{H}_c \mathbf{Q}_c \mathbf{F}_c \left(\sum_{c=1}^C \mathbf{G}_c \mathbf{H}_c \mathbf{Q}_c \mathbf{F}_c \right)^H \right|, \quad (2a)$$

$$\text{s.t.} \quad \sum_{c=1}^C \text{Tr}(\mathbf{H}_c \mathbf{Q}_c \mathbf{F}_c \mathbf{F}_c^H \mathbf{Q}_c^H \mathbf{H}_c^H) \leq P, \quad (2b)$$

$$\mathbf{Q}_c \in \Phi = \left\{ \frac{j + e^{j\phi_{i,l}}}{2} \mid \phi_{i,l} \in [0, 2\pi] \right\}, \quad (2c)$$

where P is the total transmit power budget.

III. PROPOSED SOLUTION

A. Precoder design at CN

As previously described, the CN aggregates local CSI from all clusters to construct the global CSI and designs the digital precoder accordingly. The resulting global CSI and corresponding precoders are then transmitted to the respective DNs. To facilitate tractable analysis, we focus on the digital precoder design by fixing $\mathbf{Q}_c, \forall c$ and relax the power constraint in (2b). The constraint is subsequently enforced by applying an appropriate scaling to the designed precoder. Under this setting, the optimization problem in (2a) can be reformulated as

$$\max_{\{\mathbf{F}\}} \log_2 \left| \mathbf{I} + \frac{1}{\sigma^2} \mathbf{G}_{global} \mathbf{F} \mathbf{F}^H \mathbf{G}_{global}^H \right|, \quad (4a)$$

$$\text{s.t.} \quad \sum_{c=1}^C \text{Tr}(\mathbf{F}_c \mathbf{F}_c^H) \leq P, \quad (4b)$$

where $\mathbf{G}_{global} = [\mathbf{G}_1 \mathbf{H}_1 \mathbf{Q}_1, \mathbf{G}_2 \mathbf{H}_2 \mathbf{Q}_2, \dots, \mathbf{G}_C \mathbf{H}_C \mathbf{Q}_C]$. The optimal digital precoder for the relaxed problem in

$$\max_{\{\mathbf{W}_c\}} \log_2 \left| \mathbf{I} + \frac{1}{4\sigma^2} (\tilde{\mathbf{Z}} + 2\mathbf{P}\mathbf{W}_c^H \mathbf{G}_c^H + 2(\mathbf{P}\mathbf{W}_c^H \mathbf{G}_c^H)^H + \mathbf{R}\mathbf{W}_c^H \mathbf{G}_c^H + (\mathbf{R}\mathbf{W}_c^H \mathbf{G}_c^H)^H + (\mathbf{G}_c \mathbf{W}_c \mathbf{T}_c)(\mathbf{G}_c \mathbf{W}_c \mathbf{T}_c)^H) \right| \quad (3a)$$

$$\text{s.t. } |\mathbf{W}_c|_{i,i} = 1. \quad (3b)$$

(4) can be obtained by applying truncated singular value decomposition (SVD) to $\mathbf{G}_{global}$, followed by water-filling power allocation [14]–[16]. After designing the digital precoders, the CN computes the effective global channel as $\tilde{\mathbf{G}} = \sum_{c=1}^C \mathbf{G}_c \mathbf{H}_c \mathbf{Q}_c \mathbf{F}_c$, where $\tilde{\mathbf{G}}$ represents the aggregated global CSI.

B. DMA weights design at DN

To design the DMA weights, the optimization problem in (2) is decomposed into parallel subproblems at the cluster level, since each DN independently configures its DMA weights. The per-cluster optimization problem is formulated as

$$\max_{\{\mathbf{Q}_c\}} \log_2 \left| \mathbf{I} + \frac{1}{\sigma^2} (\mathbf{B}\mathbf{B}^H + \mathbf{B}\mathbf{F}_c^H \mathbf{Q}_c^H \mathbf{H}_c^H \mathbf{G}_c^H + \mathbf{G}_c \mathbf{H}_c \mathbf{Q}_c \mathbf{F}_c \mathbf{B}^H + \mathbf{G}_c \mathbf{H}_c \mathbf{Q}_c \mathbf{F}_c \mathbf{F}_c^H \mathbf{Q}_c^H \mathbf{H}_c^H \mathbf{G}_c^H) \right| \quad (5a)$$

$$\text{s.t. (2c)}, \quad (5b)$$

where $\mathbf{B} = \sum_{k \neq c} \mathbf{G}_k \mathbf{H}_k \mathbf{Q}_k \mathbf{F}_k$. Each DN obtains $\mathbf{B}$ by subtracting its local CSI from the aggregated global CSI, $\tilde{\mathbf{G}}$. To facilitate analysis, the structures of $\mathbf{H}_c$ and $\mathbf{Q}_c$ are reformulated into diagonal matrices $\tilde{\mathbf{H}}_c \in \mathbb{C}^{N_c \times N_{d,c}}$ and $\tilde{\mathbf{Q}}_c \in \mathbb{C}^{N_c \times N_c}$, respectively, with elements defined according to [17]. For further analytical tractability, $\tilde{\mathbf{Q}}_c$ is expressed as $\tilde{\mathbf{Q}}_c = \frac{j \cdot \mathbf{I}_{N_c} + \mathbf{W}_c}{2}$, where $\mathbf{W}_c = \text{diag}(\mathbf{w}_c)$ is the optimizable term of $\tilde{\mathbf{Q}}_c$ with each element unit modulus constrained, and $\mathbf{w}_c = [w_1, w_2, \dots, w_{N_c}]^T$. The resulting objective function is then given by (3), where $\tilde{\mathbf{Z}} = 4\mathbf{B}\mathbf{B}^H + 2\tilde{\mathbf{B}} + 2\tilde{\mathbf{B}}^H + \mathbf{Z}$, $\tilde{\mathbf{B}} = \mathbf{B}\mathbf{F}_c^H \tilde{\mathbf{H}}_c^H \mathbf{\Lambda}^H \mathbf{G}_c^H$, $\mathbf{\Lambda} = j \cdot \mathbf{1}_{N_c}$, $\mathbf{Z} = \mathbf{G}_c \mathbf{\Lambda} \tilde{\mathbf{H}}_c \mathbf{F}_c \mathbf{F}_c^H \tilde{\mathbf{H}}_c^H \mathbf{\Lambda}^H \mathbf{G}_c^H$, $\mathbf{P} = \mathbf{B}\mathbf{F}_c^H \tilde{\mathbf{H}}_c^H$, $\mathbf{R} = \mathbf{G}_c \mathbf{\Lambda} \tilde{\mathbf{H}}_c \mathbf{F}_c \mathbf{F}_c^H \tilde{\mathbf{H}}_c^H$, and $\mathbf{T}_c = \tilde{\mathbf{H}}_c \mathbf{F}_c$. Given the diagonal structure of $\mathbf{W}_c$, the objective function can be reformulated in an element-wise manner as shown in (6). In this formulation, $\mathbf{p}_i$, $\mathbf{r}_i$, $\mathbf{g}_i$, and $\mathbf{t}_i$ represent the i -th column of $\mathbf{P}$, i -th column $\mathbf{R}$, i -th column of $\mathbf{G}_c$, and i -th column of $\mathbf{T}_c^H$, respectively. In (6), $\stackrel{(a)}{=}$ follows from reformulating the objective as a per-element optimization problem, where each element of $\mathbf{W}_c$ is optimized individually while keeping the remaining elements fixed, where $\Omega_i = 2\mathbf{g}_i \mathbf{p}_i^H + \mathbf{g}_i \mathbf{r}_i^H + \mathbf{g}_i \mathbf{t}_i^H \mathbf{E}_i^H$, $\mathbf{E}_i = \sum_{n \neq i}^{N_c} w_n \mathbf{g}_n \mathbf{t}_n^H$, $\Psi_i = \tilde{\mathbf{Z}} + 2 \sum_{n \neq i}^{N_c} w_n^* \mathbf{p}_n \mathbf{g}_n^H + 2 \sum_{n \neq i}^{N_c} (w_n^* \mathbf{p}_n \mathbf{g}_n^H)^H + \sum_{n \neq i}^{N_c} w_n^* \mathbf{r}_n \mathbf{g}_n^H + \sum_{n \neq i}^{N_c} (w_n^* \mathbf{r}_n \mathbf{g}_n^H)^H + \sum_{n \neq i}^{N_c} w_n \mathbf{g}_n \mathbf{t}_n^H \left(\sum_{n \neq i}^{N_c} w_n \mathbf{g}_n \mathbf{t}_n^H \right)^H + w_i \mathbf{g}_i \mathbf{t}_i^H (w_i \mathbf{g}_i \mathbf{t}_i^H)^H$, $\mathbf{A}_i = \mathbf{I} + \frac{1}{4\sigma^2} \Psi_i$, and $\tilde{\Omega}_i = \frac{1}{4\sigma^2} \Omega_i$.

From (6b), it can be observed that the second term is independent of w_i , allowing the problem to be further simplified as

Algorithm 1 Proposed distributed beamforming algorithm

```

1: Input:  $\mathbf{G}_c, \mathbf{H}_c, \mathbf{Q}_c, \forall c, N_c, P$ , and  $C$ .
2: Repeat
3:   At the CN:
4:   Perform SVD on  $\mathbf{G}_{global}$  and waterfilling to obtain  $\mathbf{F}$ .
5:   Compute  $\tilde{\mathbf{G}}$ .
6:   At the DN:
7:   for  $i = 1$  to  $N_c$  do
8:     Compute  $w_i$  according to (8).
9:   end for
10: Until Convergence.
11:  $\mathbf{F}_c = \sqrt{\frac{P}{\sum_{c=1}^C \text{Tr}(\mathbf{H}_c \mathbf{Q}_c \mathbf{F}_c \mathbf{F}_c^H \mathbf{Q}_c^H \mathbf{H}_c^H)}} \mathbf{F}_c$ .
12: Output:  $\mathbf{F}_c, \mathbf{Q}_c, \forall c$ .

```

$$\max_{\{w_i\}} \log_2 \left| \mathbf{I} + w_i^* \mathbf{A}_i^{-1} \tilde{\Omega}_i^H + w_i \mathbf{A}_i^{-1} \tilde{\Omega}_i \right| \quad (7a)$$

$$\text{s.t. } |w_i| = 1. \quad (7b)$$

The simplified problem in (7) resembles those in [14], [16]. By following similar procedures, the optimal value of w_i is given by

$$w_i = e^{-j \arg(\lambda_i)}, \quad (8)$$

where λ_i denotes the sole non-zero eigenvalue of $\mathbf{A}_i^{-1} \tilde{\Omega}_i$. A summary of the proposed algorithm is provided in Algorithm 1.

IV. NUMERICAL RESULTS

This section presents the numerical results to evaluate the proposed distributed XL-DMA architecture. We consider the narrowband near-field communication channel model as described in [18], the pathloss parameters adopted from [17]. The simulation parameters are set as follows: $M = 4$, $P = 40$ dBm, $\alpha = 0.6 \text{ m}^{-1}$, $\beta = 827.67 \text{ m}^{-1}$, $\sigma^2 = -80$ dBm, and $C = 5$. To assess the effectiveness of the proposed architecture (Prop. DSP), we compare it with the conventional centralized fully digital ELAA architecture (Fully digital) and the centralized conventional XL-DMA architecture (Conv. DMA) [14].

In Fig. 2 (a) and (b), we evaluate the impact of N_d on the SE and the interconnection overhead per cluster, respectively. We set $N_e = 25$, $N_{e,c} = 5$, and $N_{d,c} = \frac{N_d}{C}$. As illustrated in Fig. 2 (a), in terms of SE, all considered architectures exhibit increasing SE with increasing N_d , while the fully digital architecture achieves the highest performance. This gain is

$$\max_{\{w_i\}} \log_2 \left| \mathbf{I} + \frac{1}{4\sigma^2} \left(\tilde{\mathbf{Z}} + 2 \sum_{i=1}^{N_c} w_i \mathbf{p}_i \mathbf{g}_i^H + 2 \sum_{i=1}^{N_c} (w_i \mathbf{p}_i \mathbf{g}_i^H)^H + \sum_{i=1}^{N_c} w_i \mathbf{r}_i \mathbf{g}_i^H + \sum_{i=1}^{N_c} (w_i \mathbf{r}_i \mathbf{g}_i^H)^H + \sum_{i=1}^{N_c} w_i \mathbf{g}_i \mathbf{t}_i^H \left(\sum_{i=1}^{N_c} w_i \mathbf{g}_i \mathbf{t}_i^H \right)^H \right) \right| \quad (6a)$$

$$\stackrel{(a)}{=} \max_{\{w_i\}} \log_2 |\mathbf{A}_i + w_i^* \tilde{\mathbf{\Omega}}_i^H + w_i \tilde{\mathbf{\Omega}}_i| = \max_{\{w_i\}} \log_2 |\mathbf{I} + w_i^* \mathbf{A}_i^{-1} \tilde{\mathbf{\Omega}}_i^H + w_i \mathbf{A}_i^{-1} \tilde{\mathbf{\Omega}}_i| + \log_2 |\mathbf{A}_i^{-1}| \quad (6b)$$

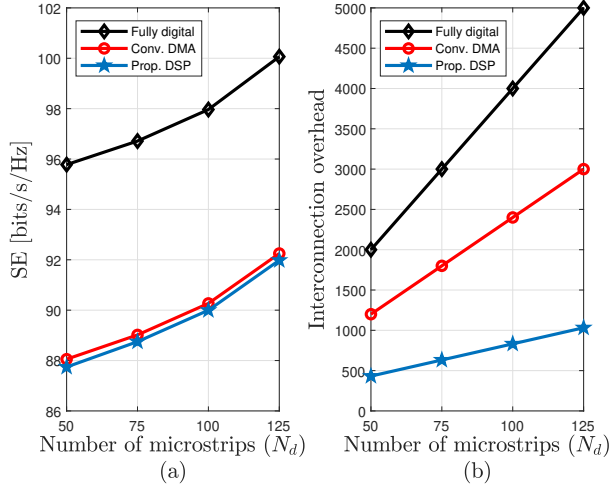


Fig. 2. (a) SE (b) Interconnection overhead versus number of microstrips with $N_e = 25$, $N_{e,c} = 5$, and $N_{d,c} = \frac{N_d}{C}$.

primarily attributed to its use of N RF chains, which enhances spatial resolution and beamforming capability.

Among the DMA-based architectures, specifically the conv. DMA and Prop. DSP architectures have comparable SE performance. This indicates that the proposed distributed architecture is capable of deploying XL-DMA systems in a decentralized fashion without incurring significant performance degradation.

In Fig. 2 (b), we observe the Prop.DSP has the lowest interconnection overhead per cluster. This behavior stems from the decentralized nature of the architecture, where communication overhead is evenly distributed among clusters.

V. CONCLUSION

This paper proposed a distributed architecture for XL-DMA systems. A novel and efficient distributed beamforming algorithm was developed to address the capacity maximization problem, with digital precoders designed at the CN and DMA weights optimized locally at the DNs. Simulation results show that the proposed scheme achieves comparable capacity to centralized XL-DMA systems while significantly reducing interconnection overhead.

REFERENCES

[1] H. Jin, K. Liu, M. Zhang, L. Zhang, G. Lee, E. N. Farag, D. Zhu, E. Onggosanusi, M. Shafi, and H. Tataria, "Massive MIMO evolution toward 3GPP release 18," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 6, pp. 1635–1654, 2023.

[2] Z. Wang, J. Zhang, H. Du, D. Niyato, S. Cui, B. Ai, M. Debbah, K. B. Letaief, and H. V. Poor, "A tutorial on extremely large-scale MIMO for 6G: Fundamentals, signal processing, and applications," *IEEE Commun. Surveys Tuts.*, vol. 26, no. 3, pp. 1560–1605, 2024.

[3] Q. Cui, X. You, N. Wei, G. Nan, X. Zhang, J. Zhang, X. Lyu, M. Ai, X. Tao, Z. Feng *et al.*, "Overview of AI and communication for 6G network: fundamentals, challenges, and future research opportunities," *Sci China Inf Sci*, vol. 68, no. 7, p. 171301, 2025.

[4] X. Zhao, M. Li, Y. Liu, T.-H. Chang, and Q. Shi, "Communication-efficient decentralized linear precoding for massive MU-MIMO systems," *IEEE Trans. Signal Process.*, vol. 71, pp. 4045–4059, 2023.

[5] J. R. Sánchez, F. Rusek, O. Edfors, M. Sarajlić, and L. Liu, "Decentralized massive MIMO processing exploring daisy-chain architecture and recursive algorithms," *IEEE Trans. Signal Process.*, vol. 68, pp. 687–700, 2020.

[6] M. Sarajlić, F. Rusek, J. R. Sánchez, L. Liu, and O. Edfors, "Fully decentralized approximate zero-forcing precoding for massive MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 8, no. 3, pp. 773–776, 2019.

[7] Y. Xu, E. G. Larsson, E. A. Jorswieck, X. Li, S. Jin, and T.-H. Chang, "Distributed signal processing for extremely large-scale antenna array systems: State-of-the-art and future directions," *IEEE J. Sel. Top. Signal Process.*, 2025.

[8] K. Li, C. Jeon, J. R. Cavallaro, and C. Studer, "Feedforward architectures for decentralized precoding in massive MU-MIMO systems," in *Proc. 52nd Conf. Signals, Syst., Comput.* IEEE, 2018, pp. 1659–1665.

[9] N. Shlezinger, G. C. Alexandropoulos, M. F. Imani, Y. C. Eldar, and D. R. Smith, "Dynamic metasurface antennas for 6G extreme massive MIMO communications," *IEEE Wireless Commun.*, vol. 28, no. 2, pp. 106–113, 2021.

[10] S. F. Kimaryo and K. Lee, "Downlink beamforming for dynamic metasurface antennas," *IEEE Trans. Wireless Commun.*, vol. 22, no. 7, pp. 4745–4755, 2022.

[11] A. Matemu and K. Lee, "Spatial modulation and generalized spatial modulation for dynamic metasurface antennas," *IEEE Trans. Wireless Commun.*, 2024.

[12] J. Xu, L. You, G. C. Alexandropoulos, X. Yi, W. Wang, and X. Gao, "Near-field wideband extremely large-scale MIMO transmissions with holographic metasurface-based antenna arrays," *IEEE Trans. Wireless Commun.*, 2024.

[13] Y. Li, S. Gong, H. Liu, C. Xing, N. Zhao, and X. Wang, "Near-field beamforming optimization for holographic XL-MIMO multiuser systems," *IEEE Trans. Commun.*, vol. 72, no. 4, pp. 2309–2323, 2023.

[14] S. F. Kimaryo and K. Lee, "Uplink and downlink capacity maximization of a P2P DMA-based communication system," *IEEE Trans. Wireless Commun.*, 2024.

[15] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[16] S. Zhang and R. Zhang, "Capacity characterization for intelligent reflecting surface aided MIMO communication," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 8, pp. 1823–1838, 2020.

[17] S. F. Kimaryo and K. Lee, "Weighted sum rate maximization for DMA-based MU-MIMO system," in *Proc. IEEE 15th Int. Conf. Inf. Commun. Technol. Convergence, Jeju Island, South Korea*. IEEE, 2024, pp. 1171–1176.

[18] Y. Liu, Z. Wang, J. Xu, C. Ouyang, X. Mu, and R. Schober, "Near-field communications: A tutorial review," *IEEE Open J. Commun. Soc.*, vol. 4, pp. 1999–2049, 2023.