ENGINEERING Information Technology & Electronic Engineering
https://www.academax.com/EITEE; https://www.engineering.org.cn
https://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=11344065
E-mail: eng.itee@zju.edu.cn

Engineering Journals
Available on IEEE Xplore®

Correspondence

https://doi.org/10.1631/ENG.ITEE.2026.0238

Hardware–algorithm co-design for 3D arrays: unlocking spatial degrees of freedom with mutual coupling

Yuqi HU[§1], Ran JI[§2], Chongwen HUANG[1✉], Li WEI[1], Xiaoming CHEN[3], Wei E. I. SHA[1], Jiguang HE[4], Zhaoyang ZHANG[1], Mérouane DEBBAH[5]

[1]*College of Information Science and Electronic Engineering, Zhejiang University, Hangzhou 310027, China*
[2]*Electrical and Computer Engineering Department, Princeton University, Princeton 08544, USA*
[3]*School of Information and Communications Engineering, Xi'an Jiaotong University, Xi'an 710049, China*
[4]*School of Computing and Information Technology, Great Bay University, Dongguan 523000, China*
[5]*6G Research Center, Khalifa University, PO Box 127788, Abu Dhabi, United Arab Emirates*

Three-dimensional (3D) antenna arrays are crucial for enhancing the capacity of 6G systems, with sub-array-based hybrid beamforming widely adopted to control hardware costs. However, current precoding designs typically assume ideal conditions, neglecting the impact of the physical layout on electromagnetic (EM) mutual coupling. This correspondence highlights that compact layouts inherently induce near-field coupling, while the connection architecture dictates inter-layer leakage and interference. Consequently, sub-optimal hardware configurations severely constrain the performance of subsequent base band algorithms. To bridge the gap between theory and deployment, we advocate moving beyond traditional fixed-hardware assumptions toward a co-design of physical architectures and signal processing. Effectively mitigating adverse EM states requires geometry-aware coupling models, the introduction of dynamic sub-array reconfiguration, and the joint optimization of structural parameters and precoding weights. This hardware–algorithm co-design represents a critical pathway to unlocking the true potential of 3D arrays.

1 Evolution and bottlenecks of 3D arrays

Constrained by the physical dimensions of base stations, traditional two-dimensional (2D) planar arrays are approaching the theoretical Hannan limit (Björnson et al., 2024). Consequently, volumetric 3D arrays have emerged as a promising paradigm to further boost array gain by exploiting the extra spatial dimension (Ji et al., 2026). However, the massive number of antennas in 3D arrays incurs prohibitive hardware costs. To circumvent this bottleneck, hybrid beamforming architectures have been widely adopted in practical deployments.

In conventional signal processing, antenna arrays are often idealized, neglecting the mutual coupling and its resultant radiation pattern distortion. To bridge this gap, recent studies (Ji et al., 2024) have proposed EM-aware beamforming, incorporating mutual coupling into precoding optimization. While this represents significant progress, passively compensating for mutual coupling at the base band (BB) is fundamentally insufficient for 3D sub-connected architectures. This correspondence argues that effective mutual coupling is dictated by physical array geometry, whereas the radio frequency (RF) connection governs the distribution of coupling across RF chains. Therefore, the next critical frontier in unleashing the full potential of 3D arrays lies in breaking the traditional assumption of fixed hardware configurations, evaluating the heterogeneous intra- and inter-sub-array coupling, and leveraging these insights to drive a hardware–algorithm co-design.

2 Hardware limits on base band algorithms

While cost and power constraints necessitate sub-connected hybrid beamforming, current implementations fail to consider the dependence of effective coupling on complex 3D topologies. To compare different RF connection topologies, the average cross-chain coupling magnitude is defined as $\bar{c} = |\mathcal{P}|^{-1} \sum_{(m,n)\in\mathcal{P}} |C_{m,n}|$, where $\mathcal{P}$ denotes antenna pairs on different RF chains, and the coupling matrix $\boldsymbol{C}$ is obtained from the coupled and ideal electric fields exported from CST Studio Suite (Ji et al., 2024). For the $6 \times 4 \times 2$ interleaved

§ Equal contribution
✉ Chongwen HUANG, chongwenhuang@zju.edu.cn
Chongwen HUANG, https://orcid.org/0000-0001-8398-8437
CLC number: TN82; TN929.5

Received: July 28, 2026; Revision accepted: Sept. 14, 2026;
Crosschecked: Oct. 12, 2026

© The Authors 2026. Published by Zhejiang University Press Co., Ltd. This is an open access article distributed under the terms of the CC BY-NC-ND license (https://creativecommons.org/licenses/by-nc-nd/4.0/)

array with a $3\lambda \times 2\lambda$ aperture, $\bar{c}$ increases from 0.049 to 0.064 when vertically adjacent antennas are assigned to different RF chains, corresponding to an approximately 30.6% increase in the coupling metric. This indicates that different RF connection topologies change coupling conditions and thus may affect system performance.

Conventional multi-input multi-output (MIMO) models may capture spatial correlation statistically but generally omit EM effects such as pattern distortion and radiation-efficiency variation. EM-aware precoding incorporates mutual coupling but typically assumes a fixed RF topology, leaving potential hardware-reconfiguration gains unexploited.

In practical deployments, both traditional algorithms and EM-aware schemes with fixed topologies face identical hardware bottlenecks (El Ayach et al., 2014). Specifically, dense 3D geometries induce strong near-field coupling, which sub-connected architectures cannot mitigate due to their block-diagonal sparsity. Since isolated BB optimization cannot bypass these fundamental hardware limitations, fully unlocking 3D massive arrays necessitates co-design.

3 Proposed co-design framework

To overcome the performance bottlenecks caused by fixed hardware configurations, a hardware–algorithm co-design framework is proposed which essentially improves the EM environment by making static physical parameters tunable. Specifically, for a typical multi-user sum-rate maximization problem, the optimization can be formulated as

$$\begin{aligned}
&\max_{\boldsymbol{W}_{\mathrm{BB}},\boldsymbol{i},\boldsymbol{T},\boldsymbol{\Omega}} \quad \sum_{k=1}^{K} \log_2\left(1+\gamma_k\right) \\
&\text{s.t. } \gamma_k = \frac{e_{\mathrm{rad}}(\boldsymbol{\Omega})\left|\boldsymbol{h}_{s,k}^{\mathrm{T}}\boldsymbol{C}(\boldsymbol{\Omega})\boldsymbol{W}_{\mathrm{RF}}(\boldsymbol{i},\boldsymbol{T})\boldsymbol{w}_{\mathrm{BB},k}\right|^2}{e_{\mathrm{rad}}(\boldsymbol{\Omega})\sum_{j\neq k}\left|\boldsymbol{h}_{s,k}^{\mathrm{T}}\boldsymbol{C}(\boldsymbol{\Omega})\boldsymbol{W}_{\mathrm{RF}}(\boldsymbol{i},\boldsymbol{T})\boldsymbol{w}_{\mathrm{BB},j}\right|^2+\sigma^2}, \quad (1) \\
&\boldsymbol{W}_{\mathrm{RF}}(\boldsymbol{i},\boldsymbol{T}) \in \mathcal{B}(\boldsymbol{T}),\ \left\|\boldsymbol{W}_{\mathrm{RF}}(\boldsymbol{i},\boldsymbol{T})\boldsymbol{W}_{\mathrm{BB}}\right\|_{\mathrm{F}}^2 \le P_{\mathrm{total}}, \\
&\boldsymbol{T} \in \mathcal{T}_{\mathrm{feasible}}, \boldsymbol{\Omega} \in \mathcal{V}_{\mathrm{geometry}},
\end{aligned}$$

where $\boldsymbol{\Omega}$ specifies the array geometry, including the element layout and layer spacing, and determines the ideal spatial channel response $\boldsymbol{h}_{s,k}$, the radiation efficiency $e_{\mathrm{rad}}(\boldsymbol{\Omega})$, and the coupling matrix $\boldsymbol{C}(\boldsymbol{\Omega})$. The matrix $\boldsymbol{T}$ represents the RF topology, while $\boldsymbol{i}$ contains the corresponding complex analog excitation coefficients, which together determine the analog precoder $\boldsymbol{W}_{\mathrm{RF}}(\boldsymbol{i},\boldsymbol{T})$. Moreover, $\boldsymbol{W}_{\mathrm{BB}}$ denotes the digital precoder, γ_k is the signal-to-interference-plus-noise ratio (SINR) of user k, $\mathcal{B}(\boldsymbol{T})$, $\mathcal{T}_{\mathrm{feasible}}$, and $\mathcal{V}_{\mathrm{geometry}}$ define the feasible analog precoders, RF topologies, and array geometries, respectively, while σ^2 and P_{total} are the noise power and total power budget, respectively.

As illustrated in Fig. 1, the proposed framework coordinates the EM, RF, and BB domains over different timescales. Z_n and V_n denote the impedance and voltage source at the n^{th} RF chain, respectively. θ_l and ϕ_l are the elevation and azimuth angles associated with the array geometry $\boldsymbol{\Omega}$, respectively. The array geometry is generally determined offline and governs the coupling, radiation efficiency, and spatial response. Compact spacing can provide higher directivity under appropriately designed excitations, while stronger mutual coupling and reduced radiation efficiency may offset part of this benefit. At the RF level, the connection topology is reconfigured on a longer timescale in response to long-term EM and channel conditions, after which the analog excitation coefficients are optimized to shape the coupled array response. At the BB level, the digital precoder $\boldsymbol{W}_{\mathrm{BB}}$ is updated more frequently using the available channel state information to suppress residual multi-user interference. This multi-timescale organization captures cross-domain interactions without high overhead of simultaneous updates.

As a representative example, Fig. 2 evaluates a $3 \times 4 \times 2$ array with four RF chains. While coupling-unaware precoding suffers severe performance degradation due to unmitigated coupling, EM-aware design with a fixed topology yields substantial gains. Importantly, the proposed co-design further unlocks spatial degrees of freedom by optimizing the RF topology. It consistently outperforms fixed-hardware schemes, approaching the fully digital bound and providing a gain of 0.345 bit/(s · Hz) over the fixed-topology EM-aware baseline.

Although the formulation is illustrated through sum-rate maximization, the same design principle applies to other system objectives, including energy efficiency and integrated sensing and communication.

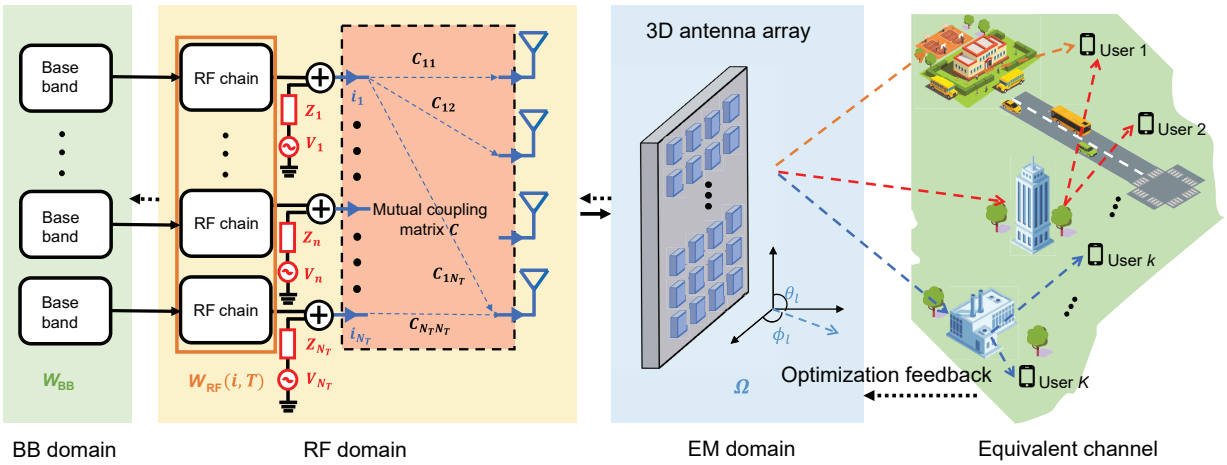


Fig. 1 Proposed joint optimization framework

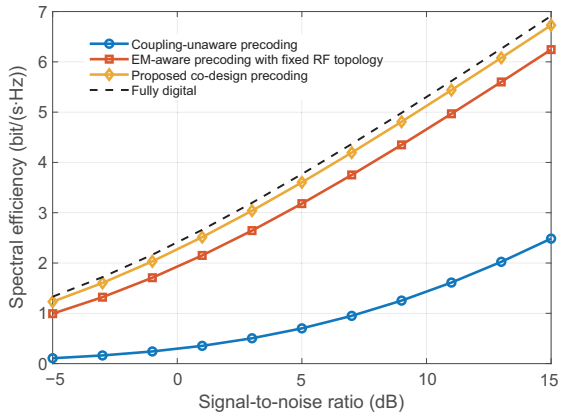


Fig. 2 Spectral efficiency of different precoding methods

4 Future pathways for co-design

While the proposed co-design paradigm provides a blueprint for overcoming hardware bottlenecks, translating this EM–RF–BB framework into practical deployments necessitates further breakthroughs. To fully realize this deep integration, future research should advance along three critical pathways:

1. Tractable analytical and surrogate EM models

Current co-design relies on accurate coupling matrices and radiation patterns. However, extracting these EM parameters traditionally requires computationally intensive full-wave simulations, which is unacceptable for real-time signal processing. Future research must bridge this computation latency gap by developing tractable analytical models or artificial intelligence-driven surrogate models. These models should instantly map specific 3D geometric layouts and discrete RF topologies to their equivalent channel responses, providing a mathematical foundation for BB optimization.

2. Low-overhead RF reconfiguration

While dynamic connections between antennas and RF chains unlock spatial degrees of freedom, realizing this in physical hardware introduces new challenges. Dense RF switch networks inevitably induce insertion losses, parasitic effects, and additional power consumption, especially at millimeter-wave and terahertz bands. Therefore, a critical future pathway is the design of low-overhead reconfigurable architectures. To keep the performance gains of this co-design from being negated by hardware overhead, it will be essential to pursue low-loss topological switching through novel materials, highly integrated packaging, and active metamaterials. Beyond the switch-based RF reconfiguration, stacked intelligent metasurfaces can perform wave-domain beamforming, providing an alternative co-design route (Liu et al., 2026). Their integration with 3D arrays requires addressing inter-layer coupling, propagation loss, calibration, and control overhead.

3. Real-time solvers for co-design

Mathematically, the hardware–algorithm co-design is inherently a highly complex mixed-integer nonlinear programming problem. Traditional alternating optimization is computationally prohibitive and struggles to meet the strict real-time constraints of 6G networks. Future algorithmic research must pivot toward data-driven and heuristic solvers. Specifically, leveraging graph neural networks to represent discrete topological connections, combined with deep reinforcement learning, can bypass exhaustive discrete searches. Developing such efficient, end-to-end solvers is the final crucial step toward deploying this co-design paradigm in practical dynamic environments.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Nos. 62331023 and 62394292), the Fundamental Research Funds for the Central Universities, China, and Zhejiang University Global Partnership Fund.

Author contributions

Yuqi HU and Ran JI developed the main ideas, performed the analysis, and drafted the paper. Chongwen HUANG and Li WEI supervised the research and revised the paper. Xiaoming CHEN, Wei E. I. SHA, Jiguang HE, Zhaoyang ZHANG, and Mérouane DEBBAH contributed to technical discussion and paper revision. All the authors read and approved the final paper.

Conflict of interest

Zhaoyang ZHANG is an editorial board member of *ENGINEERING Information Technology & Electronic Engineering* (*EITEE*); Chongwen HUANG and Xiaoming CHEN are junior editorial board members of *EITEE*. They were not involved with the peer review process of this paper. All the authors declare that they have no conflict of interest.

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Declaration on the use of generative AI tools

During the preparation of this work, the authors used DeepSeek and ChatGPT to improve language. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

References

Björnson E, Chae CB, Heath RWJr, et al., 2024. Towards 6G MIMO: massive spatial multiplexing, dense arrays, and interplay between electromagnetics and processing. https://doi.org/10.48550/arXiv.2401.02844

El Ayach O, Rajagopal S, Abu-Surra S, et al., 2014. Spatially sparse precoding in millimeter wave MIMO systems. *IEEE Trans Wirel Commun*, 13(3):1499-1513. https://doi.org/10.1109/TWC.2014.011714.130846

Ji R, Huang CW, Chen XM, et al., 2024. Electromagnetic hybrid beamforming for holographic MIMO communications. *IEEE Trans Wirel Commun*, 23(11):15973-15986. https://doi.org/10.1109/TWC.2024.3435844

Ji R, Huang CW, Chen XM, et al., 2026. Exploring Hannan limitation for 3D antenna array. *Sci China Inform Sci*, 69(5):152301. https://doi.org/10.1007/s11432-025-4705-7

Liu MB, Yuen C, Niyato D, et al., 2026. In-wave computation aided stacked intelligent metasurfaces in next-generation networks: challenges and opportunities. *IEEE Wirel Commun*, 33(4):56-63. https://doi.org/10.1109/MWC.2026.3672362