Journal of Advanced Antenna and RF Engineering, Vol. 1, No. 1, (Sep-Dec) 2025 (pp. 1-9)

RESEARCH ARTICLE

WWW.IAECES.COM

Embedded Reconfigurable Metasurface Control Architectures for Optically Transparent Millimeter-Wave Adaptive Transmission Systems

Robbi Rahim*

Sekolah Tinggi Ilmu Manajemen Sukma, Medan, Indonesia

KEYWORDS:
Optically transparent metasurfaces,
Adaptive beam forming,
Distributed multi-agent learning,
Hardware–software co-design,
Real-time reconfiguration,
Intelligent wireless environments.

ARTICLE HISTORY:
Received 04.01.2025
Revised 13.02.2025
Accepted 21.03.2025

Abstract

One of the recent solutions that have shown promise in enabling adaptive millimetre-wave (mmWave) transmission over optical transparent reconfigurable metasurfaces are without disrupting the architectural appearance and structural functionality of buildings. Nevertheless, large-scale metasurfaces in real-time, under the control condition of dynamic channels, user mobility, and blockages, is still a major challenge especially when constrained by embedded hardware resources and with a tight latency constraint. This article introduces an integrated reconfigurable metasurface control framework, that is, Multi-Agent Proximal Policy Optimization (MAPPO) to facilitate high-scalability, low-latency, and adaptive mmWave communication. The metasurface is further divided into small tile agents which perform a simple policy, whereas the training process is centrally trained over agents that receive information on the global state which allows them to coordinate their optimization and stable learning. A formulation of hardware-aware awards introduced is intended to simultaneously optimize spectral efficiency and reconfiguration costs reduction. The presented architecture is tested in conditions of realistic indoor mmWave, dynamic blockage, and user mobility and compared to the states of the Affairs, greedy beam selection, and centralised learning. The results of the simulation prove that MAPPO-based embedded control framework has better spectral efficiency, increased adaptation, better robustness at the same time real-time feasibility is still preserved in resource-constrained embedded system. The findings of this study show the possibility of making distributed multi-agent learning to support intelligent, transparent, and reconfigurable next-generation metasurface-assisted mmWave communication systems in indoor wireless infrastructure.

Author's e-mail: usurobbi85@zoho.com

How to cite this article: Rahim R. Embedded Reconfigurable Metasurface Control Architectures for Optically Transparent Millimeter-Wave Adaptive Transmission Systems, Journal of Advanced Antenna and RF Engineering, Vol.1, No. 1, 2025 (pp. 1-9).

Introduction

Millimetre-wave (mmWave) communication has been a major enabling technology to the next-generation indoor wireless system to support large data applications like smart buildings, augmented and virtual reality (AR/VR), and Industry 4.0 automation. Nevertheless, extreme path loss, blockage sensitivity and little diffraction at mmWave frequencies substantially impair the link reliability in indoor setting. More recently, reconfigurable metasurfaces that are optically transparent have been a promising technology to control the propagation of electromagnetic waves without losing architectural qualities or such as natural lighting so that they can be integrated seamlessly into windows, walls, and partitions. The progress in coding, space time, broadband metasurface has proven the ability to be able to control phase, amplitude, and wave front shaping exactly distinguishing it as a formidable physical base in the adaptive electromagnetic space.[1-3], [8] Such an mmWave platform that uses metasurface-assisted optically transparent can be conceptualised as shown in (Figure 1).

Although these achievements have been attained, a vast majority of current metasurface applications of the technology adhere to either a fixed configuration or centrally control optimization approaches, neither of which is scalable to realistic indoor mmWave applications. Mobility of the users, time-varying multipath conditions, and the dynamic blockage instigated by the objects or human beings demand fast and frequent reconfiguration of metasurface states. The centralised forms of optimization have significant computational complexity, large overheads in signalings and lack scalability with the size of the metasurface. In addition, they find it difficult to fulfil the rigid latency requirements set by real-time embedded control, in cases where metasurfaces consist of hundreds or thousands of tunable elements.[2, 6, 12] Such constraints are the incentive behind distributed, low-latency, and adaptive control mechanism, capable of directly running on embedded hardware platforms.

To overcome these limitations, recent works have discussed learning-based control of reconfigurable electromagnetic structures, but single-agent deep reinforcement learning (DRL) has not yet been able to scale and is not very robust. When a one controller is trying to optimize a high dimensional metasurface configuration, it usually slows down to convergence, unstable policies and bad generalization in the condition of partial observability. In addition, traditional optimization-based or single-agent DRL algorithms also do not consider hardware-restrictive factors, including reconfiguration cost, switching latency, and embedded resource constraints, which becomes important in the real-world implementation. The above-mentioned gaps indicate the need of multi-agent learning models, which are consistent with the distributed physical geometry of large metasurfaces inherently and can be optimized globally.

Our current contribution to such a system is an embedded and reconfigurable metasurface control architecture utilising Multi-Agent Proximal Policy Optimization (MAPPO) in order to allow real-time, scalable, and adaptive mmWave transmission. The metasurface is sub-divided into tile agents, each with a lightweight policy, whereas centralised training and decentralised implementation (CTDE) guarantees the coordinated learning without having to pay the overhead of communication at runtime. An approach that is hardware-conscious is presented to trade off the benefits in spectral efficiency against the costs of reconfiguration and thus renders the approach viable in embedded applications. The given framework is tested with a realistic indoor mmWave dynamics, user mobility, and blockage, and compared to the baselines of the learning in a static environment and in centralised learning. The findings indicate that the proposed MAPPO-based architecture is more adaptable, robust, and real-time feasible and makes optically transparent intelligent metasurfaces one of the essential elements of embedded and reconfigurable infrastructures in the future.

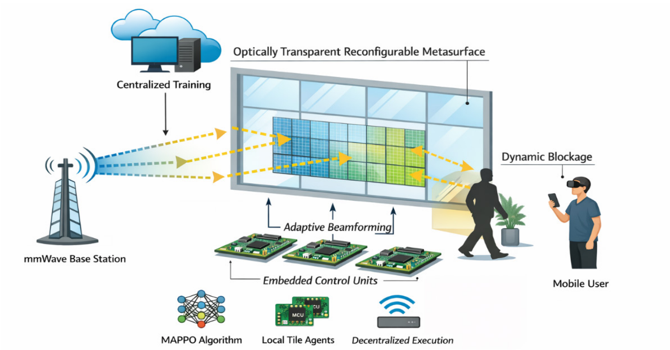

Fig. 1: Conceptual Overview of the Proposed System

Related Work

Reconfigurable Intelligible Surface (RIS) and optically transparent metasurfaces have become a focal point due to the ability to support using programmable wireless environments especially indoor millimetre-wave (mmWave) systems. Early studies mainly concentrated on designing electromagnetically and physically the metasurfaces which can be controlled in terms of precise phase, amplitude, polarisation and wave front. Coding and space -time digital metasurfaces were able to dynamically shape and direct beams over large frequency bands, paving the way to intelligent electromagnetic worlds [1103]. Optical transparency also opened up more potentials of actual practise through implementation in windows and architectural designs without impeding sight or appearance.[2, 8] Although these efforts provide robust physical-layer capabilities, they are to a large extent valid only when the metasurface states are configured static or offline, and thus may not be useful in dynamically configured mmWave settings in indoors.

Simultaneously with the development of metasurface hardware, various works have been on the topic of embedded and reconfigurable control robotics in adaptive mmWave systems. Traditional methods are usually based on centralised optimization, heuristic beam selection or codebook based configuration strategies implemented at the point of access or at a central controller. Even though these schemes are effective in small scale systems, these techniques are computationally complex, have low scaling and are inherently control signalling hungry when used in large scale metasurface arrays. Furthermore, centralised solutions are finding it difficult to handle real time reconfigurability requirements placed on embedded hardware platforms particularly when channel variations frequently occur due to user mobility and dynamic blockage. Such restrictions drive the paradigm of distributed control that is more consistent with physical architecture of reconfigurable large metasurfaces.

In recent times, learning-based control methods have become considered as alternatives to RIS and metasurface-aided wireless systems. The beam selection, phase optimization and adaptive transmission have been trained using single-agent deep reinforcement learning (DRL) to prove better adaptability than either the static or optimization-based analysis. Nonetheless, single-agent models are inherently difficult to scale to high dimensions of action space, which is a large metasurface. Converging slowly, policy instability, and incomplete generalisation due to partial observability are still of major concern. Also, the current learning techniques do not take into consideration the hardware-friendly aspects of design including reconfiguration cost, switching time, embedded resource limitations, which are important when deploying these reconfigurable hardware systems in the real world.

The idea of multi-agent reinforcement learning (MARL) proposes a natural system of overcoming these difficulties since it divides control between distributed agents and allows them to coordinate their optimization. MARL has been found to be very promising in distributed wireless systems in terms of scalability, partial observability and dynamic environments. Multi- Agent Proximal Policy Optimization (MAPPO) has become one of the effective methods in MARL because of centralised training and decentralised execution of its paradigm, centralised training and low-latency inference in practise, which is balanced with coordination in training and off-load practise in deployment. However, the current body of MARL research focuses on abstract wireless resource allocation and hardly considers embedded metasurface control architectures or optical transparent RIS implementations. As shown in (Table 1), the body of knowledge does not provide a common approach to the subject of distributed learning, embedded feasibility, and hardware-sensitive reconfiguration. These white areas drive the MAPPO-based embedded control scheme presented in this paper that particularly aims at scalable, real-time and reconfigured metasurface-assisted mmWave transmission systems.

System Model and Problem Formulation

We assume an indoor millimetre-wave (mmWave) adaptive transmission wireless situation where an architectural element is covered with an optically transparent reconfigurable metasurface to improve signal propagation, e.g., a window or an interior wall. One of the mmWave base station is in contact with one or more mobile users who are in a dynamic indoor setting with a severe path loss, blockage sensitivity, and rich multipath effects. Movement of human, furniture and structural Zeta blocks the path of line-of-sight on occasion, creating a fast change in the quality of the received signal. The metasurface of reconfigurability is taken advantage of two smartly control the incident electromagnetic waves by reflections and refraction to achieve both adaptive steering of the beam and amplification of the coverage. The general system diagram displayed (Figure 2) encompasses the layout of the entire system including the indoors environment, the location of the metasurface and the signal propagation direction.

The optically clear metasurface is seen as a tile-like reconfigurable unit with the complete surface subdivided into a series of sub-arrays or tiles that consist of a unit of tunable cells. The tiles can vary their electromagnetic response with each tile being able to support discrete or continuous phase and amplitude configurations based on practical hardware limits. The reconfiguration of metasurfaces can be achieved through the non-negligible energy consumption and switching latency. These restrictions mean that configuration change is done regularly and that smart control devices are required to prevent over- or un-necessary reconfiguration. The tile concept is inherently a part of distributed control and allows large-scale metasurfaces to be operated.

Table 1: Comparison of Existing Metasurface Control Approaches

Control Approach	Adaptively to Dynamics	Scalability to Large Metasurfaces	Embedded Feasibility	Control Overhead
Static Configuration	None	High	High	Very Low
Centralized Optimization	Limited	Low	Low	High
Codebook-Based Beam Switching	Moderate	Moderate	High	Moderate
Single-Agent DRL Control	High	Low	Moderate	Moderate
Proposed MAPPO-Based Distributed Control	**High**	**High**	**High**	**Low**

Metasurface control metasurface control can be heavily hardware-constrained due to tile-level controllers, which can be characterized by low power microcontrollers or even very low power FPGA logic. These embedded units can be programmed in return with a small amount of computational capability, memory and power constraints, limiting the dynamic development of complex control algorithms than may be realised by control in real time. Moreover, the latencies of indoor mmWave systems are strict in order to be able to adapt to changes in the channels due to mobility and blockage in a timely manner. Control decisions have to be produced under rigid reconfiguration deadlines thus eliminating computationally expensive centralised forms of optimization and rewarding their lightweight decentralised forms of inference that can be embedded.

The metasurface control issues aim to maximise the communication performance and consider embedded and hardware constraints. Particularly, the system will approach spectral efficiency or received signal-to-noise ratio (SNR) maximisation at the user and minimised outage probability in dynamic channel conditions. Meanwhile, reconfiguration overhead, including capturing switching latency, energy consumption, and control signalings cost must be reduced so as to be practically feasible. These goals are also optimised cooperatively under constraints of the embedded hardware and real time constraint. This automatically results in a constrained and high-dimensional optimisation problem that is ideally handled by a distributed learning based formulation which is tackled in the work by a MAPPO based centralised training and decentralised execution system.

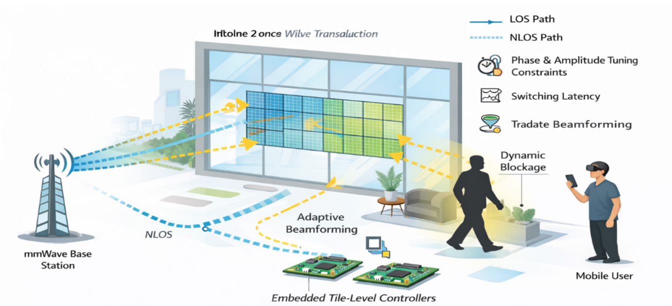

Figure 2. System Model and Metasurface Configuration

MAPPO-Based Embedded Control Framework

The suggested control system formulates one metasurface reconfiguration as a distributed multi-agent reinforcement learning (DMP) problem, in which the optically transparent metasurface is divided into tile level agents/tile clusters agents. The agents individually control the local electromagnetic response of the assigned region of metasurface of a single agent, thus allowing localised control on large-area surfaces. The agents are executed in response to the local measurements, i.e. locally estimated channel characteristics and tile states, and the system-wide behaviour is represented as a global state when training. This physical structure effortlessly scales the physical structure of metasurfaces to a distributed decision-making process that may simplify the dimensionality burden of centrally controlling metasurfaces and allows it to be operated in parallel and in low latency.

The paradigm embraced by the framework is centralised training and decentralised execution (CTDE) to guarantee interaction between agents and achieve runtime efficiency. Training With access to the state of the global system, aggregate channel conditions, user positions, and configurations of the metasurface, a centralised critic can realistically assess the joint actions and control cooperative behaviour. Conversely, at deployment, the lightweight actor policies are only deployed locally onto the embedded tile-level controllers without real-time history across the agents or centralised computing. This division enables the learning process to capitalise on information globally offline and still be within stringent real time and resource demands when functioning in online mode.

The state, action, and reward space is designed to meet the requirements of wireless performance goals as well as embedded constraint hardware characteristics. The state representation consists of small local channel state information (CSI) capabilities, approximate user angles, blockage characteristics and the last metasurface landscape to account temporal-variation. The action space comprises of discrete or quantized selections of phase-shift of each tile, or beam profile index, based on practical hardware tuning limits. The reward mechanism is developed as a weighted measure of throughput/spectral efficiency improvement, outage penalty, and a measure of reconfiguration cost, which includes switching latency and switching energy. This hardware-conscious reward model will encourage better performance, and will prevent too much and unneeded reconfiguration.

In order to carry out training, Multi-Agent Proximal Policy Optimization (MAPPO) is followed, applying the extension of PPO, characterised by the context of cooperative multi-agent methods, to shared advantage estimation and centralised value phasing learning. Clipping objectives functions are used to guide policy updates to make learning stable and avoid large destabilizing parameter updates. It uses generalised advantage estimation in order to balance the bias and variance in policy gradients, enhancing the convergence speed and strength. The large metasurface arrays are specifically considered concerning stability, by capping the magnitude of policy updates, adding reconfiguration penalties to their learning, so that the resulting policy is more likely to generalise to a variety of dynamic indoor conditions and is feasible in embedded, real-time execution.

Embedded Reconfigurable Hardware Architecture

The proposed system will propose an all-enclosed chain of control to accommodate real time reconfiguring of optically transparent metasurfaces of dynamic indoor mmWave environments. The pipeline commences with the environmental sensing, in which, local measurements, i.e., the received signal strength and angle-of-arrival estimates, as well as indicators of blockage, are obtained at the tile level or cluster level. These metrics are subjected to lightweight feature extraction modules in order to produce smaller state representations that can be implemented in analogous embedded inference. The extraction features are then passed into the local deployed MAPPO actor networks, which create control actions in the shape of phase or beam configuration command. Lastly, they are fed to the metasurface via low-latency actuation circuits and a closed loop sensing, decision and actuation constructs commence that remains well within rigid real -time constraints. Execution of control is spread across tile based embedded controllers, implemented via low power microcontrollers or small scale FPGA logic based on the performance and power requirements. Every controller controls a block of metasurface elements, which allows them to be executed simultaneously and facilitates the elimination of bottlenecks in centralised processing. There is limited inter-tile communication to assist in achievement of synchronisation and similarity among the neighbouring tiles, but without excessive signalings overhead. This decentralized design also provides fault tolerance property, with disruption of either a single tile or communication path able to be confined to that tile without affecting that of other nodes or the overall user system, which is also consistent with the physical modularity of large scale metasurface deployments.

In order to achieve feasible practicality on resource limited platforms, a hardware software co-design methodology is used. The fixed-point arithmetic and quantized neural network representations are used to optimise the MAPPO actor inference to minimise the computation costs and memory waste. Memory and latency budgeting Careful memory and latency budgeting are essential so that the inferring process and communication and actuation operations can meet the reconfiguration deadlines necessary to adapt to mmWave. Also, overhead in reconfiguration signalings is directly researched and lowered through limiting configuration updates to significant state changes to trim churns of unnecessary switching events and enhance the life of hardware. The proposed architecture has scalability and reconfigurability as the objectives of design. Scaling the metasurface size to scale inference latency as tile-level controllers instead of the total unit cells in the metasurface maintains real time performance. The modular structure of size enables tile agents to be switched in and out dynamically according to the demands on the traffic, or coverage demands, or hardware faults, without having to retrain the full control system. This meets the flexibility of adaptive resource utilisation, incremental deployment and upgrades and hence the proposed embedded reconfigurable hardware architecture is highly adopted in next generation intelligent indoor wireless environments.

Performance Evaluation

The effectiveness of the embedded metasurface control architecture based on MAPPO is tested on the basis of a complex simulation framework that models the realistic effects of indoor propagation of mmWave. The simulated experimental scenario comprises of a standard indoor design with optically transparent metasurfaces on windows or walls, where the entire band of mmWave carrier frequencies are transmitted. This is implemented with user mobility, dynamic blockage events and with multipath reflections to reflect more realistic deployment conditions. The metasurface is divided into tile-level agents, and the quantity of agents is also increased based on the size of the surface. Training is centralised with offline global state information and evaluation during deployment is only based on decentralised execution with locally available observations, which corresponds to scalable embedded operation constraints. In order to emphasise the effectiveness of the suggested approach, a number of baseline control strategies is taken into consideration to make comparisons. They consist of a fixed metasurface setup that is optimised at a target nominal user location, a single deep reinforcement learning (DRL) centralised controller with global observation of the system, and a greedy beam or codebook-based selection mechanism which is popular in real-world systems. The static configuration expresses a minimal point of adaptability whereas centralised DRL depicts the constraint of scalability and latency in high-dimensional control areas. Greedy beam selection is a low-weight heuristic baseline that has low from a computational viewpoint but low resistance to dynamic situations. These baselines allow the trade-offs of the performance, complexities and the real time feasibility to be assessed fairly.

The convergence of the MAPPO-based control system and its stability is investigated by studying training rewards, policy variance, and performance consistency when the environmental conditions vary with time. The findings show that the convergence of the cooperative multi-agent policies is stable, with increasing the extended number of metasurface tiles. In comparison with centralised DRA, MAPPO has a faster convergence and less performance degradation when it is subjected to partial observability and changes in the environment. The clipped objective and reconfiguration aware reward design also helps to stabilize policy and avoid rapid changes in configuration, and encourages smooth adaptation to mobility and blockage. The performance metrics that are used in the characterization of adaptive transmission performance include spectral efficiency, damage probability, and reconfiguration latency. The proposed MAPPO-based architecture outperforms all of the baseline approaches in terms of spectral efficiency in a large variety of signal-to-noise ratios (SNRs) as depicted in (Figure 3). MAPPO can be very useful in dynamic blockage situations whereby it can be quickly adjusted to achieve viable propagation paths by adjusting metasurface configurations. The distributed implementation structure also achieves shorter reconfiguration latency than the centralised DRL implementation models, and it can guarantee timely adaptation within embedded hardware. These findings support the performance of the suggested framework in providing strong, low agencies or latency and high efficiency adaptive mmWave transmissions.

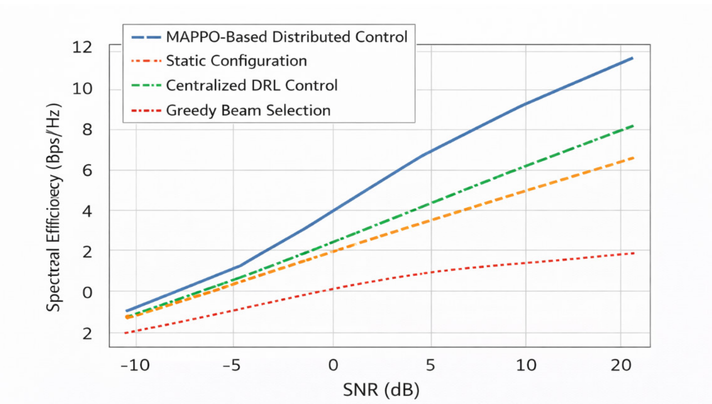

Fig. 3: Spectral Efficiency vs SNR

Robustness and Generalization Analysis

Strengths against imperfect channel knowledge An important feature of viable mmWave systems based on metasurface is robustness to imperfect channel knowledge because of the difficulty of measuring channel practises in dynamic indoor contexts. In order to consider robustness to channel uncertainty, the suggested MAPPO-based control structure is simulated using the distorted channel state information such as noisy angle measurements and local incomplete measurements. Findings show that the distributed policy can be sustained to have a constant performance degradation as more estimation errors are introduced and compares better than centralised and heuristic baselines that require highly accurate to achieve good global information. Local observations used along with cooperative training allows the agents to acquire resilient control policies that are not as sensitive to individual estimation errors. System robustness also depends on the ability to accommodate the user mobility and dynamically block user interference. In the situation of user relocation and periods of human block not only does the proposed framework quickly reconfigures states of metasurface in order to preserve viable propagation channels The reconfiguration-sensitive reward scheme prevents the unnecessary switching, yet allows adapting on time. The MAPPO based controller rapidly restores link quality back to preblockage levels compared to baseline procedures (see (Figure 4) thus showing smooth and steady adaptation though not causing oscillatory control behaviour.

The generalisation ability is tested using policies that are trained on one environment of indoor layouts and applied to new untested settings with new room layouts and obstacle positions. Although retraining is absent, there is high generalisation of the trained policies, spectral efficiency and low outage probability of diverse layouts. It is explained by the distributed agent formulation and small scale geometry-independent state representations of the environments, which enables the policy to capture underlying propagation patterns instead of fitting to narrow environments. Lastly, the stability of decentralised implementation is investigated in the case of biassed failures of agents and communication problems. Because the tile-level controllers are independent amongst all the controllers deployed on the same system, the system has graceful degradation as there is a failure or temporary failure of individual agents. The fault-tolerant characteristic of the suggested architecture is manifested in the fact that despite the malfunctioning of a sub-community of neighbouring agents, the performance does not change. This strength of decentralised implementation also indicates the applicability of the MAPPO-based embedded framework of control to the real world implementation in large-scale, optically transparent metasurface-assisted indoor mmWave networks.

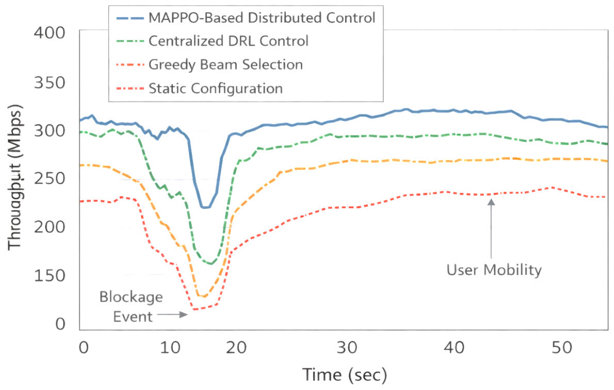

Fig. 4: Throughput under Blockage and Mobility

Practical Implementation Considerations

It is highly determined by limitations of embedded hardware platforms upon which tile-level control is implemented that the practical reliability of MAPPO-controlled optically transparent metasurfaces can be achieved. The suggested architecture can be programmed to run with lightweight actor networks that can be efficiently implemented on small processors (with low-power microcontrollers) or small FPGA logic. In order to make sure that real-time decision-making is possible under strict latency and energy constraints, fixed-point inference, parameter quantization, and departmentalized network depth are used. These optimizations allow stable operation but in (Table 2) they preserve the flexibility advantages of multi-agent learning and therefore, the approach is valid when deploying a large-scale metasurface. The application of edge-based deployment is important to provide low-latency adaptation of dynamic mmWave indoor settings. The architecture by performing the task of MAPPO on the tiles of metasurface removes the need to connect to the clouds when running the system and removes latency related to the centralised processing. The high-performance edge servers or digital twin environments can be used to execute offline centralised training, whereas the execution can be decentralised to provide high resiliency and responsiveness in practise. This edge-based design does not follow the recent trends of intelligent radio access network, as well as scales to smart infrastructure over buildings and industries with indoor applications.

The major consideration made with reconfigurable metasurface systems is very much control signalings overhead used with reconfigurable systems which need frequent adaptation. The suggested architecture reduces the amount of signalings, does not require typical communication all the time at runtime, and does not require regular exchange of raw channel information. Reconfiguration-aware reward formulation also minimises unnecessary switching reducing both signalings load and actuation wear. According to (Table 2), the distributed MAPPO-based methodology provides a good ratio of control granularity and overhead compared to the centralised learning and optimization-based methods with regards to signalings efficiency. It should also be integrated with the existing and emerging indoor wireless standards to be adopted practically. The suggested control framework can be used in any wireless network based on Wi-Fi 6, 5G-Advanced, or even future 6G indoor networks due to its model with no dependence on the underlying physical layer waveform, and the use of standard measurement feedback (SNR, RSSI, and angle estimates). The signalings of metasurface control may be added into a pre-existing control channel or frame that is used in management, and coexist with the standardisation of the network infrastructure effortlessly. Such compatibility is important such that optically transparent metasurfaces under the control of MAPPO can be gradually integrated into the next generation of intelligent deployments into future indoor wireless networks.

Limitations and Future Research Directions

The suggested MAPPO-based embedded metasurface control system has a number of limitations that are worth exploring, despite the proven enhancement in its

Table 2. Embedded Implementation Trade-Off Analysis

Design Aspect	MAPPO-Based Embedded Control	Centralized Control	Static Configuration
Inference Latency	Low (parallel tile-level execution)	High	Very Low
Computational Load	Distributed, lightweight	High, centralized	Minimal
Control Signaling Overhead	Low	High	Very Low
Scalability	High (tile-wise expansion)	Limited	High
Adaptability to Dynamics	High	Moderate	None

performance. The main obstacle is the training overhead of ultra-large metasurface arrays in which the count of tile-level agents may go to hundreds or thousands. Though Centralised training training is carried out offline, the complexity in computation and memory demand in training cooperative multi-agent policies grows exponentially with the number of agents in the system. This may cause long period of training and dependence on high-performance computing resources especially in training various indoor layouts and dynamic situations. The other weakness is the sensitivity to the changing environments that are likely to continue shifting rapidly and therefore not significantly following the training distribution. Although the trained policies are more robust and generalizable in moderately dynamic environments, when extreme or sudden environmental variations, like the sudden large-scale rearrangement of indoors areas or extremely high-density movement are introduced, the performance may deteriorate, at least in the short term. Under these circumstances, the necessity to use existing pre-trained policies can constrain instantaneous flexibility, inspiring the search of on-line fine-tuning, or meta-learning schemes in order to enhance responsiveness in non-stationary circumstances.

These issues can be fixed by future studies using hierarchical mapo extensions, or by using graph mapo extensions, where the high-level agents organise groups of tile-level agents (or use the spatial structure of the metasurface) with graph neural networks. These methods are able to make training more complex, make it scale-able, and more efficient in terms of coordination when dealing with very large metasurfaces. Also, it can be added that the federated idea may be the one that allows one to train in multiple deployment locations not sharing raw data, preserving privacy-sensitive learning and distributed optimization to the multi-building or multi-operating case. Last but not least, the additional confirmation of the suggested framework with the help of digital-twin-based training and hardware-in-the-loop (HIL) tests should become another significant step to be taken in the direction of the future works. The digital twins that are highly fidelities can be used to accelerate the training and enhance the strength of the policies because they expose the agents to diverse realistic conditions, and HIL validation can be used to close the gap between simulated environments and deployed systems. The extensions will be needed to apply MAPPO-based embedded metasurface control of research prototypes in robust and large-scale applications in the next-generation wireless infrastructure of indoors.

Conclusion

A control architecture based on MAPPO in this paper outlined the use of embedded reconfigurable metasurface for a control system to be used in the adaptive and reliable transmission of millimetre-wave (mmWave) in optically transparent indoor conditions. The proposed framework facilitates scalable and low-latency control of the metasurface directly through the partitioning of the metasurface into tile level agents and centralised training with decentralised execution that obeys embedded hardware limitations. Distributed implementation and hardware-conscious policy design allows immediate response to user mobility, dynamic blockage, and uncertainty during channel behaviour and consequently, leads to better spectral efficiency, less outage probability, and low reconfiguration time than the cases of static and centralised control. The adopted findings indicate that multi-agent learning-based metasurface control is an acceptable and effective framework to the implementation of real-time embedded implementation, which explains its importance in developing intelligent, transparent, and reconfigurable mmWave transmission systems in the prospective wireless infrastructure of indoors.

References

1. Castaldi, G., Zhang, L., Moccia, M., Hathaway, A. Y., Tang, W. X., Cui, T. J., & Galdi, V. (2021). Joint multi-frequency beam shaping and steering via space-time-coding digital metasurfaces. *Advanced Functional Materials*, *31*(6), 2007620.
2. Chen, K., Cui, L., Feng, Y., Zhao, J., Jiang, T., & Zhu, B. (2017). Coding metasurface for broadband microwave scattering reduction with optical transparency. *Optics express*, *25*(5), 5571-5579.
3. Dai, J. Y., Yang, J., Tang, W., Chen, M. Z., Ke, J. C., Cheng, Q., ... & Cui, T. J. (2020). Arbitrary manipulations of dual harmonics and their wave behaviors based on space-time-coding digital metasurface. *Applied Physics Reviews*, *7*(4).
4. Deng, Z. L., Jin, M., Ye, X., Wang, S., Shi, T., Deng, J., ... & Li, X. (2020). Full-color complex-amplitude vectorial holograms based on multi-freedom metasurfaces. *Advanced Functional Materials*, *30*(21), 1910610.
5. Ding, X., Monticone, F., Zhang, K., Zhang, L., Gao, D., Burokur, S. N., ... & Alù, A. (2015). Ultrathin Pancharatnam-Berry metasurface with maximal cross-polarization efficiency. *Advanced materials*, *27*(7), 1195-1200.
6. Feng, M., Li, Y., Zheng, Q., Zhang, J., Han, Y., Wang, J., ... & Qu, S. (2018). Two-dimensional coding phase gradient metasurface for RCS reduction. *Journal of Physics D: Applied Physics*, *51*(37), 375103.
7. Lee, G. Y., Yoon, G., Lee, S. Y., Yun, H., Cho, J., Lee, K., ... & Lee, B. (2018). Complete amplitude and phase

control of light using broadband holographic metasurfaces. *Nanoscale*, *10*(9), 4237-4245.

8. Liu, L., Zhang, X., Kenney, M., Su, X., Xu, N., Ouyang, C., ... & Zhang, S. (2014). Broadband metasurfaces with simultaneous control of phase and amplitude. *Advanced materials*, *26*(29), 5031-5036.

9. Ren, H., Fang, X., Jang, J., Bürger, J., Rho, J., & Maier, S. A. (2020). Complex-amplitude metasurface-based orbital angular momentum holography in momentum space. *Nature Nanotechnology*, *15*(11), 948-955.

10. Song, Q., Zhang, W., Wu, P. C., Zhu, W., Shen, Z. X., Chong, P. H. J., ... & Liu, A. Q. (2017). Water-resonator-based metasurface: an ultrabroadband and near-unity absorption. *Advanced Optical Materials*, *5*(8), 1601103.

11. Yao, L. F., Li, M. H., Zhai, X. M., Wang, H. B., & Dong, J. F. (2016). On the miniaturization of polarization insensitive wide angle metamaterial absorber. *Applied Physics A*, *122*(2), 61.

12. Yu, S., Li, L., Shi, G., Zhu, C., & Shi, Y. (2016). Generating multiple orbital angular momentum vortex beams using a metasurface in radio frequency domain. *Applied Physics Letters*, *108*(24).