Received XX Month, XXXX; revised XX Month, XXXX; accepted XX Month, XXXX; Date of publication XX Month, XXXX; date of current version 4 July, 2026.

Digital Object Identifier 10.1109/OJCOMS.2024.011100

Mobility Aware Dual Waveguide ISAC Pinching Antenna NOMA for Resilient Mission Critical Networks

Derek Chui *(Student Member, IEEE)*, Ramon Anaya, Pedro Aguirre-Salas, Monica Reyes, Collin Fiske, Duy P. Ngo *(Student Member, IEEE)*, AND Krishna Murthy Kattiyan Ramamoorthy *(Member, IEEE)*

[1]Department of Computer Science and Engineering, Santa Clara University, Santa Clara, CA 95053 USA

CORRESPONDING AUTHOR: Krishna Ramamoorthy (e-mail: kkattiyanramamoorthy@scu.edu).

This work was supported by Santa Clara University.

ABSTRACT Sixth generation wireless networks must support mission critical services such as disaster response and public safety communications, where mobile users, heterogeneous data priorities, and rapidly varying wireless channels make conventional static and throughput oriented resource allocation insufficient. This paper proposes a mobility aware resource allocation framework for pinching antenna systems using a dual waveguide integrated sensing and communication architecture, in which one waveguide performs radar like tracking of mobile users while a second waveguide serves them through non orthogonal multiple access transmission, with pinching antennas dynamically repositioned along both waveguides. We derive a physically grounded sensing channel model from first principles, establish a tracking accuracy constraint, and formulate a joint optimization problem spanning user grouping, antenna assignment, power allocation, and antenna repositioning under a shared sensing and communication power budget. We then show that this constraint is met with roughly 30 dB of margin throughout, so the resource that limits the deployment is not the sensing power share but the age of the position estimates the allocator acts on, since user displacement between updates exceeds the sensing ranging error by an order of magnitude. To solve this problem under continuous mobility, we develop a deep reinforcement learning framework that extends double deep Q networks with a mobility and sensing aware state, action, and reward design. We characterize the learned policy against an exhaustive search benchmark, the exactly solvable decoupled allocation, a cost matched sequential greedy heuristic, a published sum rate optimal power allocation, and a random floor, on identical held out scenes and under both nominal and contended load, all while holding users above the sensing SNR threshold required for reliable tracking. We evaluate the framework in mission critical public safety scenarios modeled after disaster response and stadium emergencies, with results also connecting to a secondary sports analytics testbed for live tactical decision making.

INDEX TERMS deep reinforcement learning, integrated sensing and communication (ISAC), mission critical networks, mobility aware resource allocation, non orthogonal multiple access (NOMA), pinching antenna systems (PASS), public safety communications

I. INTRODUCTION

SIXTH GENERATION (6G) wireless networks are envisioned to support a new class of mission critical services. This ranges from first responder coordination during disaster response, triage during mass casualty events, and command to control links for search and rescue teams. Such cases demand both timeliness and priority in the delivered information, where the difference can be a matter of life and safety. These environments share three key characteristics. First a dense population of mobile users (responders, victims, sensors, vehicles) whose positions change continuously and often unpredictably. Second, heterogeneous data streams with various levels of urgency (ex. medical alert versus routine status). Third, severe and time varying interference

driven by that same mobility we have just mentioned. Meeting these demands calls for resource allocation that is simultaneously both mobility and priority aware, as well as being spectrally efficient. All this is a combination that conventional static and throughput only scheduling simply cannot provide.

A first consequence of that mobility is that channel state information ages. Beamforming, scheduling, user pairing, and power allocation are all optimized against estimates that are already stale by the time they are applied, and in dense deployments where movement is both continuous and correlated across users the degradation compounds rather than averaging out. Resolving this by external localization is a dependency mission critical deployments cannot assume, since the infrastructure that would provide it is precisely what fails in the scenarios of interest. Integrated sensing and communication (ISAC) removes that dependency by having the communication infrastructure estimate user positions and velocities itself while it transmits, so that allocation decisions are made against a current picture of the environment rather than a remembered one.

Sensing on its own only closes half the loop. Knowing where users have moved is of limited use if the radiating aperture cannot follow them, and a fixed array can only reweight over a geometry it cannot change. At millimeter wave over a service area of this size the dominant term is large scale path loss, which is set by distance rather than by steering, and that is the term a fixed geometry cannot address. Pinching antenna systems (PASS) have recently emerged as a low cost and physically reconfigurable alternative to conventional massive MIMO and RIS architectures, in which pinching antennas (PAs) are repositioned along a dielectric waveguide to reshape the channel directly. This offers large scale path loss control that fixed antenna arrays cannot match [1], and it is what converts a fresh position estimate into a shorter link rather than only into a better aimed one.

What remains is contention. Mission critical operations concentrate many users with sharply different priorities into the same area at the same time, and orthogonal access spends a dedicated resource slice on each of them regardless of what they carry. Non orthogonal multiple access (NOMA) instead superposes user signals in the power domain and separates them by successive interference cancellation at the receiver, which both raises spectral efficiency and makes the intra group power split an explicit lever over which user is served better. Paired with PASS the repositioned aperture serves both members of a group at once, so the spatial and power domain decisions act on the same allocation. The three are complementary in a way none of them is alone. ISAC supplies the position and velocity information, PASS acts on it by physically reconfiguring the aperture, and NOMA determines how the resulting link budget is divided among users of differing priority. Prior work on PASS-NOMA, including our own [2], has shown the benefits of jointly optimizing user grouping, PA placement, and power allocation, but only under static user positions and a single purpose communication only waveguide, which is exactly the pair of assumptions a mission critical deployment cannot make.

A. CONTRIBUTIONS

We propose a dual waveguide ISAC PASS architecture that separates sensing and communication onto independently reconfigurable waveguides. A physically grounded sensing channel model for the sensing waveguide is derived from first principles through its round trip, giving a tracking fidelity constraint the deployment must meet before its position estimates can be used. On that basis we formulate the joint mobility aware allocation problem over user grouping, PA assignment, NOMA power allocation and PA repositioning on both waveguides under a shared sensing and communication power budget, and develop a deep reinforcement learning framework extending the double deep Q network that learns the grouping, the power split and the antenna placement as a single discrete policy conditioned on user positions and velocities.

We further establish the structure of the per slot problem itself, showing that a decoupled form of it reduces to maximum weight perfect matching and is therefore exactly solvable in polynomial time, and identifying contention for the finite pinching antenna pool and the cross group spacing constraint as the two mechanisms that take the deployed system out of that class. This both bounds what any allocator ignoring contention can achieve and locates the part of the problem a learned sequential policy exists to capture.

Results are obtained on the analytical PASS channel of Section III against exhaustive, greedy, throughput maximizing, published sum rate optimal and random baselines, and confirmed on a site specific ray traced reconstruction of Levi's Stadium driven by real player tracking, which reproduces the analytical trends to within roughly 0.7 dB. The scenarios modeled are first responder coordination in disaster environments and large scale mass gathering emergencies such as stadium evacuation, with a secondary sports analytics testbed for live tactical decision making.

II. RELATED WORK

Pinching antenna systems have been formalized as a physically reconfigurable alternative to fixed arrays and RIS, with [1] providing a tutorial treatment and confirming that PAs can also act as receivers, a property our sensing waveguide relies on directly. Surveys of the paradigm [3], [4] and the related flexible antenna framing [5] situate PASS within the wider reconfigurable antenna literature. Multi-waveguide extensions include waveguide division multiple access [6] and multi-mode PASS [7]. On power allocation, [8] derives a closed form KKT solution for PASS NOMA sum rate maximization and [9] formulates power minimization with explicit inter-waveguide interference, both of which inform the channel model of Section III. PA assignment has been studied directly in [10], which optimizes activation under

a sum rate objective, and [11], which adds per user QoS guarantees, both complementary to the joint grouping, assignment and power problem we build on from [2]. Beyond throughput, [12] treats finite blocklength short packet transmission, relevant to the small payload traffic characteristic of mission critical alerts, and [13] addresses physical layer security through a PPO based continuous action formulation that we revisit in Section V.

NOMA user grouping and power allocation have been studied jointly with learning based methods, including RIS assisted NOMA with K-means++ grouping under imperfect SIC [14], grant free uplink NOMA with multi-RIS support in dense IIoT deployments [15], and sum rate improvement through partial correlation detection under asynchronous transmission [16]. Classical grouping baselines remain useful comparison points. Dynamic user clustering with power allocation is proposed in [17] and an improved pairing scheme building on it in [18], both of which we have used as baseline pairing strategies in companion ns3 work, while [19] shows that joint grouping and power optimization extends naturally to objectives beyond throughput. None of these consider PASS style spatial reconfigurability or explicit user mobility.

Semantic aware NOMA weights delivered information by its relevance to the receiving application rather than by raw rate, and is the framework we develop in Section III and previously applied in a static setting [2], [20]. Deep learning based semantic communication departs from bit accurate transmission [21], and [22], [23] carry those principles into NOMA by exploiting semantic redundancy and heterogeneous semantic and bit user coexistence, with [24] optimizing power efficiency for the coexisting case. Auction based [25] and Stackelberg based [26] formulations offer game theoretic alternatives to centralized power allocation. Collectively these establish that application level quality is a saturating and nonlinear function of link quality rather than a proportional one, which is the empirical basis for the fidelity model of Section III, though none address mobility or sensing driven allocation.

Integrated sensing and communication is a natural complement to NOMA. A broad survey of ISAC channel modeling [27] covers the two way sensing channel and radar cross section characterization our derivation builds on, and [28] demonstrates joint sensing and communication through STBC assisted NOMA. At the system level ISAC has been positioned as the physical layer enabler of digital twins, in which continuous two way sensing keeps a virtual model synchronized with its environment [31], which motivates the extension outlined in Section VII. Mobility aware allocation has meanwhile been explored outside the PASS and ISAC context, notably in vehicular networks, where [29] applies reinforcement learning to joint scheduling and power allocation for low latency NOMA V2X and [30] applies deep Q learning under a random waypoint mobility model, confirming that reinforcement learning suits the temporal coupling mobility introduces. No prior work, however, combines PASS spatial reconfigurability, ISAC based mobility tracking and deep reinforcement learning for mission critical, mobility dominated deployments, which is the gap this paper addresses.

III. SYSTEM MODEL

A. NETWORK ARCHITECTURE

We consider a downlink PASS system where a base station (BS) feeds two co located dielectric waveguides carried on an elevated overhead cable structure, forming a dual waveguide integrated sensing and communication (ISAC) architecture, as illustrated in Fig. 1. The two waveguides serve distinct functions.

The sensing waveguide $\mathcal{W}_s$ is equipped with N_s pinching antennas (PAs) for radar like probing of mobile users, which provides real time position estimates to guide resource allocation. This goes hand in hand with the communication waveguide $\mathcal{W}_c$, which is given N_c PAs for downlink non orthogonal multiple access (NOMA) transmission to user groups.

Let $\mathcal{K} = \{1, 2, \ldots, K\}$ denote the set of K single antenna mobile users distributed in a rectangular service area of dimensions $L \times W$ in the ground plane. Instead of having the waveguides as low fixed rails along the area perimeter, which illuminate ground users at shallow grazing angles and via severe mutual occlusion, we adopt an elevated converging cable geometry. Four dielectric cables are suspended from towers of height H_t at the four corners of the service area and converge at a common feed hub $\tilde{\boldsymbol{\psi}}_0 = [L/2,\, W/2,\, H_c]^T$ above the area center, at convergence height $H_c < H_t$. Viewed from above the four cables cross at the hub to form an X. This overhead geometry gives near line of sight visibility to ground users and because the anchors are widely separated horizontally and differ in height, well conditioned geometry diversity for position estimation. In a stadium this is realized by a spidercam style suspended cable rig. In mission critical deployments the same geometry is used with rapidly deployable masts.

Each cable carries both a sensing waveguide $\mathcal{W}_s$ and a communication waveguide $\mathcal{W}_c$ co located along its span and fed from the hub, so the BS distributes the sensing power P_s and communication power P_c across the four cables from the common feed point $\tilde{\boldsymbol{\psi}}_0$. A PA on waveguide $\ell \in \{s, c\}$ is placed at arc length $s_{\ell,n}(t) \in [0, L_\ell]$ measured from the hub along its host cable, where L_ℓ is the cable length. Neglecting cable sag, its position is

$$\boldsymbol{\psi}_{\ell,n}(t) = \tilde{\boldsymbol{\psi}}_0 + \frac{s_{\ell,n}(t)}{L_\ell}\left(\boldsymbol{c}_{\ell,n} - \tilde{\boldsymbol{\psi}}_0\right) \tag{1}$$

where $\boldsymbol{c}_{\ell,n} = [c^x_{\ell,n},\, c^y_{\ell,n},\, H_t]^T$ is the corner tower top of the cable hosting PA n, with $(c^x_{\ell,n}, c^y_{\ell,n}) \in \{0, L\} \times \{0, W\}$. The N_s sensing and N_c communication PAs are distributed evenly over the four cables so that each cable hosts $N_s/4$ sensing and $N_c/4$ communication PAs. The time varying position of user k is $\boldsymbol{\psi}_k(t) = [x_k(t),\, y_k(t),\, 0]^T$, where t

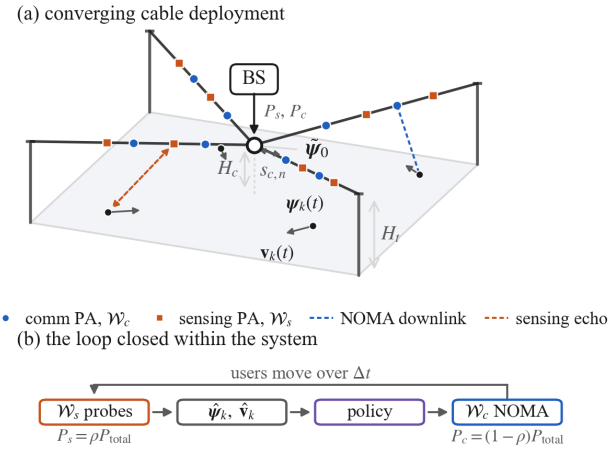

FIGURE 1. Dual waveguide ISAC PASS architecture. (a) Four dielectric cables are suspended from corner towers of height H_t and converge at the common feed hub $\tilde{\psi}_0$ at height H_c above the center of the $L \times W$ service area, crossing in an X when viewed from above. Each cable carries both the sensing waveguide $\mathcal{W}_s$ and the communication waveguide $\mathcal{W}_c$ along its span, hosting $N_s/4$ sensing and $N_c/4$ communication pinching antennas located by arc length $s_{s,n}$ and $s_{c,n}$ measured from the hub, with the base station feeding both from that hub. User positions $\psi_k(t)$ evolve with velocity $\mathbf{v}_k(t)$. The vertical scale is exaggerated relative to the ground plane for legibility. (b) The allocation loop closes inside the system. $\mathcal{W}_s$ supplies the position and velocity estimates on which the policy conditions, the policy fixes the grouping, the intra-group power split and the PA arc positions, $\mathcal{W}_c$ transmits, and the users have moved on by the next slot. The two waveguides draw on one budget through the split ρ.

indexes discrete time slots of duration Δt. The straight single waveguide geometry of prior PASS work [2] is recovered as the specific case of one cable.

The two waveguides operate as a closed loop within each slot. At the start of slot t the sensing waveguide probes the service area and the echo model below is integrated coherently over T_{int} to produce a position estimate for every user whose sensing SNR meets the tracking threshold $\Gamma_{s,k}(t) \geq \Gamma_{\min}$. Those estimates, and the velocities they induce across consecutive slots, form the state on which the allocation policy of Section V conditions. The policy then fixes the user grouping and the intra group power split, the PAs on both waveguides are repositioned along their host cables, and the communication waveguide transmits for the remainder of the slot. Because the position information originates from the system's own sensing waveguide rather than from an external localization service, the architecture is self contained, and the sensing power P_s is the price paid for the channel knowledge the allocator consumes. Raising P_s sharpens those estimates but reduces the P_c available to serve users, and that balance is swept directly in Section VI. We assume throughout that a user meeting the tracking threshold is localized accurately enough for its estimate to stand in for its true position, and we do not propagate the residual error through the allocation decision.

B. MOBILITY MODEL

Unlike prior static PASS NOMA formulations [2], we instead explicitly model user mobility. The position of user k evolves according to

$$\boldsymbol{\psi}_k(t+1) = \boldsymbol{\psi}_k(t) + \mathbf{v}_k(t)\,\Delta t \tag{2}$$

where $\mathbf{v}_k(t) = [v_k^x(t),\, v_k^y(t)]^T$ is the velocity vector of user k at time t. We adopt the Gauss Markov mobility model, in which the velocity evolves as

$$\mathbf{v}_k(t+1) = \alpha_v\,\mathbf{v}_k(t) + (1-\alpha_v)\,\bar{\mathbf{v}}_k + \sigma_v\sqrt{1-\alpha_v^2}\,\mathbf{w}_k(t) \tag{3}$$

where $\alpha_v \in [0,1]$ is the memory parameter controlling temporal correlation, $\bar{\mathbf{v}}_k$ is the asymptotic mean velocity, σ_v is the velocity standard deviation, and $\mathbf{w}_k(t) \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_2)$ is the Gaussian noise. When $\alpha_v = 1$, the model produces constant velocity trajectories. When $\alpha_v = 0$, the velocity is in turn memoryless. Intermediate values produce smooth and correlated trajectories that are well suited for modeling realistic user movement, including structured mobility patterns in sporting environments.

C. CHANNEL MODEL

1) FREE SPACE CHANNEL

The distance between user k and PA n on waveguide $\ell \in \{s, c\}$ at time t is given by

$$\begin{aligned} d_{\ell,k,n}(t) &= \left\|\boldsymbol{\psi}_k(t) - \boldsymbol{\psi}_{\ell,n}\right\| \\ &= \sqrt{\left(x_k(t) - x_{\ell,n}\right)^2 + \left(y_k(t) - y_\ell\right)^2 + h^2}\,. \end{aligned} \tag{4}$$

Assuming line of sight (LoS) propagation based on spherical wave assumptions [1], [6], the free space channel coefficient from PA n on waveguide ℓ to user k at time t is modeled as

$$h_{\ell,k,n}(t) = \frac{\sqrt{\eta}}{d_{\ell,k,n}(t)} \exp\!\left(-j\frac{2\pi}{\lambda}\, d_{\ell,k,n}(t)\right) \tag{5}$$

where $\eta = (\lambda/4\pi)^2$ is the free space path loss constant at the reference distance of 1 m and λ is the carrier wavelength.

2) WAVEGUIDE PROPAGATION

As the signal propagates from the BS feed point along waveguide ℓ to PA n, it accumulates a guided wave phase shift. The propagation coefficient for PA n on waveguide ℓ is given by

$$g_{\ell,n} = \exp\!\left(-j\frac{2\pi}{\lambda_g}\, s_{\ell,n}\right) \tag{6}$$

where $\lambda_g = \lambda/n_{\text{eff}}$ is the guided wavelength in the dielectric waveguide and n_{eff} is the effective refractive index. We assume negligible waveguide propagation loss, which is valid for dielectric waveguides at millimeter wave frequencies [1].

3) EFFECTIVE COMMUNICATION CHANNEL

We then consider downlink NOMA transmission with two user groups. Let $P = K/2$ denote the number of groups

(assuming K is even), indexed by $\mathcal{P} = \{1, \ldots, P\}$. A pairing is represented by a mapping $\pi(t) : \mathcal{K} \to \mathcal{P}$ such that each group contains exactly two users at time t.

Each communication PA n is assigned to at most one group or left inactive. Let $a_n(t) \in \{0\} \cup \mathcal{P}$ denote the assignment label of PA n at time t, where $a_n = 0$ indicates inactivity and $a_n = p$ indicates service to group p. The set of PAs assigned to group p is

$$\mathcal{M}_p(t) \triangleq \{n \in \mathcal{N}_c : a_n(t) = p\} \tag{7}$$

The total BS transmit power P_{tx} allocated to the communication waveguide is P_c. For multi waveguide deployment, the power is first divided across waveguides, then evenly among active PAs. The per PA power budget for group p is $P_m = P_c/(|\mathcal{M}_p(t)| \cdot P)$ [2]. Given $\mathcal{M}_p(t)$, the effective channel of user k in group p is

$$\tilde{h}_{k,p}(t) \triangleq \sum_{n \in \mathcal{M}_p(t)} g_{c,n}\, h_{c,k,n}(t)\, \sqrt{P_m} \tag{8}$$

which aggregates the contributions of all PAs assigned to group p, including both free space propagation and in waveguide phase accumulation.

D. PASS BASED SENSING MODEL

A dedicated sensing waveguide supplies the real time position estimates on which mobility aware optimization of the communication waveguide depends. We state the round trip model that produces those estimates and the fidelity constraint they must satisfy.

The BS transmits a normalized probing signal $s_s(t)$ with $\mathbb{E}[|s_s(t)|^2] = 1$ into the sensing waveguide with total power P_s, distributed equally among N_s sensing PAs. The signal reaching user k from PA n is the cascade of the guided wave phase accumulated to that PA (6) and free space propagation to that user (5). A portion of the incident energy scatters back with radar cross section $\sigma_{\text{RCS},k}$ [27], contributing an amplitude factor $\sqrt{\sigma_{\text{RCS},k}/4\pi\eta}$ in which the $1/4\pi$ accounts for isotropic re-radiation and the division by η removes the receive aperture factor $\lambda^2/4\pi$ embedded in the Friis coefficient (5), since RCS is referred to the power density incident on the target rather than to the power intercepted by an isotropic antenna at it. That aperture factor applies only on the return leg, where the PA physically collects the echo. The return path is otherwise identical to the forward path [1], so every medium is traversed twice and the round trip contribution of PA n carries $g_{s,n}^2 h_{s,k,n}^2(t)$, complex squares that preserve the doubled phase rather than magnitude squares. Coherently combining the N_s returns, received at round trip delay $\tau_k = 2d_{s,k,n}(t)/c$ in additive white Gaussian noise of power σ_s^2, gives the aggregate sensing channel

$$\beta_k(t) \triangleq \sqrt{\frac{\sigma_{\text{RCS},k}}{4\pi\eta}} \sum_{n=1}^{N_s} g_{s,n}^2\, h_{s,k,n}^2(t) \tag{9}$$

All K users are illuminated by the same probing signal, so the receiver sees a superposition of K simultaneous reflections rather than an isolated single target return. The BS carries a predicted position $\hat{\psi}_k(t \mid t-1)$ for every user, propagated from the previous slot estimate through (2) and (3), and the delay this implies match filters the aggregate into a per user echo. The quantities $\beta_k(t)$ and $\Gamma_{s,k}(t)$ are therefore outputs of that per target filtering stage applied to one physical aggregate, not K physically separate sensing channels. Because the PAs lie at different distances from user k, their echoes combine with different phases, so the sensing PA positions enter $|\beta_k(t)|$ through both the per leg path loss $d_{s,k,n}^{-2}(t)$ and the relative phase alignment across PAs.

Correlating the echo against the transmitted waveform over an integration window T_{int} at bandwidth B yields a coherent processing gain

$$G_p = T_{\text{int}} B \tag{10}$$

which is essential rather than incidental, since the two way link incurs d^{-4} spreading loss and the single pulse echo SNR is consequently well below unity at practical ranges. The post integration sensing SNR for user k is then

$$\Gamma_{s,k}(t) = \frac{P_s}{N_s}\, G_p\, \frac{|\beta_k(t)|^2}{\sigma_s^2} \tag{11}$$

where P_s/N_s is the per PA transmit power and $|\beta_k(t)|^2$ the round trip channel power gain. Expanding $|\beta_k(t)|^2$ through (9) exposes its dependence on the sensing PA positions $\{s_{s,n}\}$ through $g_{s,n}$ and on the user positions through $d_{s,k,n}(t)$, so $\Gamma_{s,k}(t)$ varies over time as users move. Reliable tracking of every user requires the sensing fidelity constraint

$$\Gamma_{s,k}(t) \geq \Gamma_{\min}, \quad \forall k \in \mathcal{K},\ \forall t, \tag{12}$$

below which the estimated position $\hat{\psi}_k(t)$ becomes unreliable and degrades the quality of user pairing and PA assignment on the communication waveguide. We set $\Gamma_{\min} = 15$ dB, slightly above the $\sim$13 dB detection threshold commonly adopted in the ISAC literature [**?**], as a conservative margin. Because $\{s_{s,n}\}$ enters through $d_{s,k,n}(t)$, the sensing PAs can be repositioned as users move in order to maintain the constraint, which exploits the spatial reconfigurability of PASS for adaptive sensing.

The sensing and communication subsystems share a common power budget $P_s + P_c \leq P_{\text{total}}$. We adopt a fixed split ratio $P_s = \rho P_{\text{total}}$ and $P_c = (1-\rho)P_{\text{total}}$, treating ρ as a fixed system parameter instead of an additional decision variable, which keeps the core optimization problem and resulting DRL state and action design tractable. Sensitivity to that choice is characterized directly by sweeping ρ and reporting the utility versus tracking accuracy tradeoff in Section VI.

E. NOMA DOWNLINK TRANSMISSION WITH SIC

We adopt a clustered NOMA structure in which users are partitioned into disjoint groups of two, power domain superposition with successive interference cancellation (SIC) is applied within each group, and distinct groups are separated orthogonally. Two mechanisms enforce that separation. Each group is allocated its own disjoint subset of communication PAs drawn from the shared pool, so no PA radiates for more than one group in a given slot, and groups are assigned orthogonal frequency resources. Consequently the interference seen by a user originates only from its own group partner, and the SINR expressions below carry an intra-group interference term alone. This intra-cluster NOMA with inter cluster orthogonality structure is the standard formulation in the PASS NOMA literature [8], [10], [11] and keeps the grouping decision separable across groups, which is what makes the per slot allocation tractable. Residual inter cluster leakage through the shared waveguide, which would couple the groups and destroy that separability, is outside the present model and identified as future work in Section VII.

Within a group p at time t containing users $\{u, v\}$, let $w(p,t)$ and $s(p,t)$ index the members with the smaller and the larger effective channel power $|\tilde{h}_{k,p}(t)|^2$, and let $\alpha_{p,w}(t)$ and $\alpha_{p,s}(t)$ be the intra-group power fractions assigned to them, satisfying $\alpha_{p,w}(t) + \alpha_{p,s}(t) = 1$ and $\alpha_{p,w}(t) \geq \alpha_{p,s}(t) > 0$. Successful SIC requires the weak user stream to be decodable at both receivers. Its SINR at either receiver has the form $|\tilde{h}|^2\alpha_{p,w}(t)/(|\tilde{h}|^2\alpha_{p,s}(t)+\sigma^2)$, which is strictly increasing in $|\tilde{h}|^2$, so the binding case is the weak user itself and decodability at the strong user follows by construction. The weak user therefore achieves

$$\gamma_{p,w}(t) = \frac{|\tilde{h}_{w(p),p}(t)|^2\,\alpha_{p,w}(t)}{|\tilde{h}_{w(p),p}(t)|^2\,\alpha_{p,s}(t) + \sigma^2} \tag{13}$$

while the strong user, having canceled the weak user signal, decodes its own message free of intra-group interference with

$$\gamma_{p,s}(t) = \frac{|\tilde{h}_{s(p),p}(t)|^2\,\alpha_{p,s}(t)}{\sigma^2} \tag{14}$$

giving achievable rates in bits/s/Hz of

$$R_{p,w}(t) = \log_2\big(1+\gamma_{p,w}(t)\big), \quad R_{p,s}(t) = \log_2\big(1+\gamma_{p,s}(t)\big) \tag{15}$$

F. SEMANTIC AWARE UTILITY MODEL

Building on achievable rates in (15), we incorporate application level semantics into the resource allocation objective. Each user k is assigned a semantic importance weight $w_k \geq 0$ which reflects the priority of its data stream. We model the semantic fidelity $\xi(\gamma) \in [a_1, a_2]$ as a logistic function of the SINR γ

$$\xi(\gamma) = a_1 + \frac{a_2 - a_1}{1 + \exp\big(-(c_1\gamma + c_2)\big)} \tag{16}$$

where (a_1, a_2) are the min and max fidelity bounds, and (c_1, c_2) are shaping parameters that control the sensitivity of fidelity to SINR variations.

The logistic form is not chosen for analytical convenience. Learned semantic transceivers exhibit task accuracy that is empirically sigmoidal in channel quality, collapsing below a task dependent threshold and saturating well before the bit error rate does [21], [22], and it is that saturation that makes rate a poor proxy for delivered value. Each parameter carries a direct operational reading. The lower bound $a_1 = 0.30$ is the residual fidelity of a degraded but still actionable transmission, such as a coarse position fix recovered from a partially decoded stream, and the upper bound $a_2 = 0.98$ is capped below unity because a semantic decoder saturates at its own task accuracy ceiling rather than at perfect reconstruction. The ratio $-c_2/c_1$ places the inflection at $\gamma = 3.2$, or 5.1 dB, a deliberate margin above the per user rate floor $R_{\min} = 1$ bit/s/Hz ($\gamma = 0$ dB) of (19i), so that fidelity transitions across the region where the QoS constraint is marginal rather than saturating on either side of it. The slope $c_1 = 0.25$ sits between a step response, which would make the objective discontinuous in the allocation, and a near linear response, which would collapse the model back to weighted rate.

The semantic weights w_k are likewise not assigned arbitrarily. They enumerate the traffic classes of the deployment, ordered by the operational cost of delaying them, with $w_k = 1.0$ for life critical medical telemetry, 0.8 for command and control, 0.7 for responder status and biometrics, 0.6 for vehicle and sensor telemetry, and 0.3 for bulk or archival streams. The five discrete levels used in Section VI are these five classes, so the weight of a user is a property of the service it carries rather than a free parameter of the optimization. In the secondary sports analytics testbed the same five levels are grounded in observable match state, with the highest reserved for injury and collision alerts and the rest ordered by proximity to the active phase of play.

Because the fidelity model is a modeling choice rather than a measured curve, we verify that the conclusions do not depend on it. Sweeping the inflection point over $\gamma^\star \in \{1.6, 3.2, 6.4\}$, the slope over $c_1 \in \{0.125, 0.25, 0.5\}$ and the bounds (a_1, a_2) over $\{(0.10, 1.00), (0.50, 0.90)\}$, the ordering of the compared policies is preserved in every configuration. Absolute utilities shift with the parameterization as expected, but no qualitative conclusion of this paper rests on the adopted values.

For group p at time t, the semantic fidelities of both weak and strong users are $\xi_{p,w}(t) = \xi\big(\gamma_{p,w}(t)\big)$ and $\xi_{p,s}(t) = \xi\big(\gamma_{p,s}(t)\big)$ respectively. The semantic aware contribution of each user is then

$$\begin{aligned} G_{p,w}(t) &= w_{w(p)}\,\xi_{p,w}(t)\,R_{p,w}(t), \\ G_{p,s}(t) &= w_{s(p)}\,\xi_{p,s}(t)\,R_{p,s}(t). \end{aligned} \tag{17}$$

The total semantic utility of the system at time t is $\sum_{p=1}^{P}\big[G_{p,w}(t) + G_{p,s}(t)\big]$.

G. PROBLEM FORMULATION

Taking all this into consideration, we now formulate the mobility aware joint optimization problem that captures user grouping, PA assignment, NOMA power allocation, and the sensing communication resource tradeoff. Specifically, our goal is to jointly optimize time varying user pairing $\pi(t)$, communication PA assignment $\mathbf{a}(t) = [a_1(t), \ldots, a_{N_c}(t)]$, the intra group NOMA power split $\boldsymbol{\alpha}(t)$, and communication PA arc length positions $\mathbf{s}_c(t) = [s_{c,1}(t), \ldots, s_{c,N_c}(t)]$ to maximize the time averaged total semantic utility. The sensing PA arc length positions $\mathbf{s}_s(t) = [s_{s,1}(t), \ldots, s_{s,N_s}(t)]$ are placed by a fixed geometric rule on the tracked user set and are not decision variables.

A convenient closed form power split, used in the static formulation of our prior work [2], sets the fractions directly from the intra group channel disparity,

$$\begin{aligned} \bar{\alpha}_{p,w}(t) &= \frac{|\tilde{h}_{s(p),p}(t)|^2}{|\tilde{h}_{w(p),p}(t)|^2 + |\tilde{h}_{s(p),p}(t)|^2}, \\ \alpha_{p,w}(t) &= \min\{\max\{\bar{\alpha}_{p,w}(t),\, \alpha_{\min}\},\, \alpha_{\max}\}, \\ \alpha_{p,s}(t) &= 1 - \alpha_{p,w}(t), \end{aligned} \tag{18}$$

where $\bar{\alpha}_{p,w}(t)$ is the unclipped channel disparity ratio and $[\alpha_{\min}, \alpha_{\max}] = [0.55, 0.95]$. This allocates power to the weak user in proportion to channel disparity and satisfies (19c) to (19d) by construction. This is convenient but not utility optimal as it is blind to the semantic weights w_k and to the logistic fidelity $\xi(\cdot)$ that shape the objective. We thus retain (18) only as a baseline and as the split used by the exhaustive greedy benchmark in Section VI, and instead treat the power split as a learned action. The policy selects $\alpha_{p,w}(t)$ from the discrete grid $\mathcal{A}_\alpha = \{0.55, 0.60, \ldots, 0.95\}$ jointly with the user pairing. Discretizing the split keeps the action space finite while letting the policy learn a semantic aware allocation the closed form cannot express.

$$\max_{\substack{\pi(t),\, \mathbf{a}(t),\\ \boldsymbol{\alpha}(t),\, \mathbf{s}_c(t)}} \frac{1}{T} \sum_{t=1}^{T} \sum_{p=1}^{P} \Big[G_{p,w}(t) + G_{p,s}(t) \Big] \tag{19a}$$

$$\text{s.t. } a_n(t) \in \{0\} \cup \mathcal{P}, \quad \forall n \in \mathcal{N}_c,\ \forall t, \tag{19b}$$

$$\alpha_{p,w}(t) + \alpha_{p,s}(t) = 1, \quad \forall p \in \mathcal{P},\ \forall t, \tag{19c}$$

$$\alpha_{p,w}(t) \geq \alpha_{p,s}(t) > 0, \quad \forall p \in \mathcal{P},\ \forall t, \tag{19d}$$

$$s_{\ell,n+1}(t) - s_{\ell,n}(t) \geq \Delta_{\min},\ \forall n, t,\ \ell \in \{s, c\}, \tag{19e}$$

$$s_{\ell,n}(t) \in [0,\, L_\ell],\ \forall n, t,\ \ell \in \{s, c\}, \tag{19f}$$

$$P_s + P_c \leq P_{\text{total}}, \tag{19g}$$

$$\Gamma_{s,k}(t) \geq \Gamma_{\min}, \quad \forall k \in \mathcal{K},\ \forall t, \tag{19h}$$

$$R_k(t) \geq R_{\min}, \quad \forall k \in \mathcal{K},\ \forall t. \tag{19i}$$

Constraint (19b) lets each communication PA serve one group or remain inactive, and (19c) with (19d) enforces the standard NOMA structure in which the weak user receives the larger share for reliable SIC decoding. Constraints (19e) and (19f) impose a minimum inter PA spacing $\Delta_{\min} = \lambda/2$ to mitigate mutual coupling [6] and confine PAs to the physical cable length. Constraint (19g) carries the sensing communication tradeoff, since raising P_s sharpens tracking at the cost of the P_c available to serve users, with the split $\rho = P_s/P_{\text{total}}$ held as a fixed system parameter and swept in Section VI rather than optimized online. Constraint (19h) keeps the position estimates feeding the communication decisions reliable under mobility. Because $\mathbf{s}_s(t)$ is set geometrically rather than chosen, it enters as a per slot feasibility condition on the deployment rather than as a quality the allocator trades against, and Section VI reports the margin with which it is met. Constraint (19i) enforces a minimum per user rate for quality of service.

Relative to the static formulation, (19) is a sequential decision problem. The pairing, the PA assignment and the power allocation must be re-optimized every slot as the positions $\boldsymbol{\psi}_k(t)$ evolve, and the communication PAs can be repositioned along their cables to follow the users while the sensing PAs track the same motion under the geometric rule, which exploits the spatial reconfigurability of PASS. The problem is non convex through the coupled SINR expressions, the combinatorial pairing and assignment decisions, and the temporal dependence introduced by mobility, and each slot's decisions must be made from position estimates that are themselves imperfect. This motivates the deep reinforcement learning approach developed in Section V.

IV. PROBLEM STRUCTURE AND THE DECOUPLED OPTIMUM

Before developing a learned solution it is worth establishing what problem (19) actually is. Firstly, what the simplest version of the allocation problem the deployment could have posed, and what the standard solution to that version is. Then which specific features of the deployed system remove it from that class, such that the standard solution stops being exact. Lastly why the resulting gap is not a fixed overhead but rather one that grows as the system is scaled, which is what makes a learned sequential policy vital.

The answer to the first is that the restricted problem is the static single waveguide allocation of our prior work [2], which is solved exactly in polynomial time by a maximum weight matching. This is explained below, then we show the problem it raises.

Consider the case in which the block of communication PAs serving group p is fixed by a rule that depends only on the members of p, for instance placed at the arc length whose ground projection is nearest that group's own centroid, independently of what any other group is assigned. Because distinct groups are separated orthogonally, the SINRs (13) and (14) of the users in group p then depend only on p's own members, its own PA block, and its own split $\alpha_{p,w}$. The objective (19a) consequently decomposes into a sum of per group terms, and the following structural result holds. It is the default solution to the simple version of the problem,

in the sense that no better allocation exists and none needs to be searched for.

Under the group orthogonality and a per group PA placement rule depending only on that group, the optimal per slot allocation is a maximum weight perfect matching on the complete graph over $\mathcal{K}$, with the weight of edge (i, j) given by

$$u(i,j) = \max_{\alpha \in \mathcal{A}_\alpha} \left[G_{p,w} + G_{p,s} \right]_{p=\{i,j\}} \tag{20}$$

and is therefore obtainable exactly in $O(K^3)$ time by Edmonds' blossom algorithm.

The consequence is sharp. In this decoupled case the combinatorial enumeration of Section D is unnecessary. The exact optimum is reachable in polynomial time at every load, so a learned policy would have nothing to contribute beyond a constant factor, and any claim that the pairing decision is intractable would not survive scrutiny.

The deployed architecture is not in that class and two mechanisms take it out. Firstly the communication PAs are a finite shared resource. Each group draws its block from the pool left by the groups formed before it, such that a later group's effective channel (8) is a function of which PAs earlier groups consumed, and the value of a pair is no longer defined independently of the rest of the allocation. Second, the minimum spacing constraint (19e) binds across groups rather than within them. Arc length claimed on a cable by one group is removed from the feasible placements available to every later group sharing that cable, so the action set itself contracts as the slot proceeds. Under either mechanism the objective fails to decompose, the edge weights (20) are not well defined, and the matching formulation degrades from an exact solution to a relaxation.

What the deployment pays for assuming the decomposition is not a fixed overhead, which is what determines whether a learned policy is valid. Both mechanisms above are driven by contention for a finite aperture, so their severity is set by how much of the communication PA pool and of the usable arc length a slot's groups collectively need. Under light load the groups formed in a slot draw on a pool with slack remaining. Few placements are foreclosed for the groups after, and a pair valued in isolation is valued very nearly correctly. The matching solution of (20) is then a good approximation and there is little for a sequential policy to recover. As load rises, the arc length available to a group is increasingly determined by what earlier groups claimed rather than by that group's own geometry. At that point the edge weights (20) are not merely inexact but can be ordered incorrectly, since the pair with the highest isolated value may be precisely the pair whose placement most restricts every group formed after it. The error compounds across the commitments made within a slot instead of averaging out over them.

Mobility acts on the same mechanism along a second axis. A block position chosen without regard to later groups

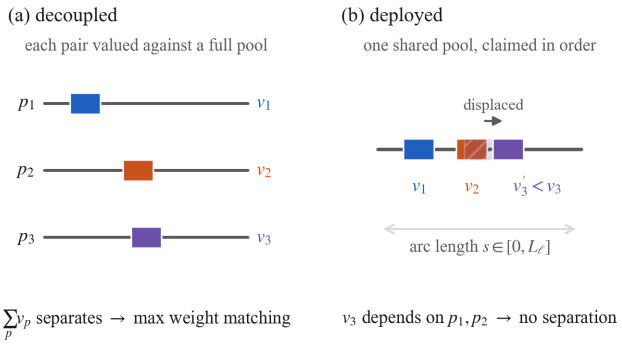

FIGURE 2. The decoupled case against the deployed case. Left shows each group's PA block is placed by a rule depending only on its own members, the objective separates over groups, and the allocation is an exact maximum weight matching. On the right, groups draw sequentially from a shared pool, so the arc length available to the third group is what the first two left, and the value of a pair is no longer defined independently of the rest of the allocation.

persists for the remainder of the slot. Under motion the configuration it forecloses is the one a later group would have wanted as its users move into it. The cost of a myopic per pair valuation is thus paid once per slot at a lighter load, and repeatedly once load and motion act together, which is what this deployment is built for. The load sweep of Section VI is designed to expose exactly this, and the margin between the decoupled optimum and the proposed policy does widen with K rather than hold constant, rising from 100.2% of the decoupled optimum at $K = 4$, where a pair valued in isolation is valued almost correctly, to 123.5% at $K = 12$.

We therefore retain the decoupled optimum as an explicit baseline in Section VI rather than discarding it. It is the exact solution of the allocation problem as posed in the static single waveguide setting of our prior work [2], and the margin by which the coupled system exceeds it isolates precisely what contention for a finite reconfigurable aperture is worth. That margin is also the part of the problem a learned sequential policy exists to capture, since it is the part that a per pair valuation, however computed, structurally cannot see. Fig. 2 illustrates the two regimes, and Fig. 5 reports how the margin between them behaves as load increases.

V. PROPOSED DEEP REINFORCEMENT LEARNING FRAMEWORK

The per slot allocation is non convex and combinatorial. We solve it with a policy applied independently at each slot and conditioned on the current user positions and velocities, such that as users move the allocation tracks them. Unlike the static PASS NOMA problem in [2], which was solved with tabular Q learning, the high dimensional state (positions, velocities, per user channel quality, and sensing feedback) motivates a deep reinforcement learning (DRL) approach with function approximation.

To keep the action space finite we do not learn every variable of problem (19) jointly. The sensing PAs are repositioned each slot by a geometric rule that biases them outward

along their host cable for a wider tracking baseline, and are not part of the learned decision. The learned policy controls the three decisions that are combinatorial, coupled across groups, and semantic aware. That being which users to group, the NOMA power split $\alpha_{p,w}(t)$ of each group, and where along the cables that group's block of communication PAs is placed. The last of these is what makes the communication PA positions $\mathbf{s}_c(t)$ a genuine decision variable of (19) rather than a side effect of the pairing, and it is the decision through which the cross group coupling of Section IV acts, since arc length claimed by one group constrains every group formed after it. All three are discrete, so we adopt a double deep Q network (DDQN) rather than a continuous action method such as the PPO design of [13]. DDQN is also off policy and reuses past transitions via replay for sample efficiency and its decoupled action selection and evaluation corrects the value overestimation that a large combinatorial action space would otherwise induce. This extends the tabular Q learning of our static formulation [2] to a setting too large for a lookup table.

A. STATE, ACTION, AND REWARD DESIGN

We assign this per slot allocation as a sequential pairing process solved by a single Q network. Starting from all K users unpaired and all N_c communication PAs free, the agent forms one group per step until fewer than two users remain. The semantic utility accumulated along this sequence is the return it learns to maximize.

1) STATE

At each step a state is formed for every candidate pair (u, v) by concatenating the two users' feature rows with a summary of what remains. Each user k contributes a seven dimensional row

$$\boldsymbol{\phi}_k = \big[\, x_k,\ y_k,\ v_k^x,\ v_k^y,\ \rho_k,\ w_k,\ \bar{g}_k \,\big] \tag{21}$$

Its grounded position, shadow fading ρ_k, semantic weight w_k, and communication channel gain $\bar{g}_k = 10\log_{10}|\tilde{h}_k(t)|^2$ (in dB) to the currently positioned communication PAs, offset by a constant for input conditioning. An eight dimensional context vector $\mathbf{c}(t)$ summarizes what remains. The fraction of users still unpaired, the fraction of communication PAs still free, the mean and spread of remaining users' channel gains, weights, and positions, and the arc length already claimed on each cable by groups formed earlier in the slot. That last component lets the policy observe the contention identified in Section IV rather than merely suffer it. The full state for pair (u, v) at candidate offset δ is $\mathbf{s}_{u,v,\delta} = [\boldsymbol{\phi}_u, \boldsymbol{\phi}_v, \mathbf{c}(t), \boldsymbol{\psi}_{u,v,\delta}] \in \mathbb{R}^{27}$. A five dimensional block $\boldsymbol{\psi}_{u,v,\delta}$ carries what the candidate offset would do to this pair's channel. The coherent gain each user would see from the block placed there, and its distance to the nearest resulting PA. The gain must be the coherent sum rather than a path loss proxy, since over a ± 20 m slide the $1/d$ envelope moves only a few dB while the in waveguide phase $e^{-j2\pi s/\lambda_g}$ scrambles completely.

2) ACTION

For a given state, the candidate PA block offset $\delta \in \mathcal{A}_\delta$ (9 levels spanning ± 20 m of arc) enters the state rather than the output, and the network emits one Q value per candidate power split $\alpha_{p,w} \in \mathcal{A}_\alpha$ (9 levels). The $|\mathcal{A}_\alpha||\mathcal{A}_\delta| = 81$ actions available to a candidate pair are therefore scored as $|\mathcal{A}_\delta|$ rows of $|\mathcal{A}_\alpha|$ outputs in one batched pass. The offset is measured relative to the arc length at which the group's own centroid projects onto each cable, which makes the decision translation invariant across scenes. Offsets that would place the block within $\Delta_{\min}$ of arc length already claimed by an earlier group are masked out before the maximization, which enforces (19e) across groups rather than only within them. Over all candidate pairs available at the current step, the agent takes the pair, split, and level of largest value,

$$(u^\star, v^\star, \alpha^\star, \delta^\star) = \arg \max_{\substack{(u,v),\ \alpha\in\mathcal{A}_\alpha,\\ \delta\in\mathcal{A}_\delta(\mathcal{O}_t)}} Q\big(\mathbf{s}_{u,v,\delta},\ \alpha\big) \tag{22}$$

where $\mathcal{O}_t$ is the set of arc length intervals already occupied at this point in the slot and $\mathcal{A}_\delta(\mathcal{O}_t) \subseteq \mathcal{A}_\delta$ the offsets that remain feasible against it.

3) REWARD

Forming group $p = \{u, v\}$ at split $\alpha^\star$ yields the immediate reward

$$r = \big(G_{p,w} + G_{p,s}\big) - \lambda_R \sum_{i\in\{u,v\}} \big[R_{\min} - R_i\big]^+ - \lambda_\Gamma \sum_{i\in\{u,v\}} \big[\Gamma_{\min} - \Gamma_{s,i}\big]^+ \tag{23}$$

the group's semantic utility (17) minus soft penalties for violating the per user rate floor (19i) and the sensing SNR constraint (12), where $[\cdot]^+ = \max(0, \cdot)$, the sensing term is taken in dB, and $(\lambda_R, \lambda_\Gamma) = (2, 1)$ weight the two penalties. Folding both constraints into the reward keeps them active during learning without an explicit constrained solver. The rate term depends on the chosen pair, split, and placement and therefore shapes the policy directly. The sensing term depends only on the user positions and the sensing PA placement, both of which are fixed within a slot under the geometric placement rule, so it is invariant to the pairing decision and enters as a per slot feasibility monitor rather than as a shaping signal. Only the pure semantic utility $G_{p,w} + G_{p,s}$ is reported as performance. Equation (23) is the immediate reward r_t, evaluated at the current slot geometry. Because the deployment is mobile and the objective (19a) is a time average over the trajectory, the reward is shaped to reflect how a decision fares once the users move. After the round is decided, the scene is advanced one mobility slot under the Gauss–Markov mobility model, and the same

action is re-scored at the moved positions to give r_{t+1}. The reward used for learning is the blend

$$r = (1 - \eta\, r_t + \eta\, r_{t+1}, \qquad \eta = 0.3, \tag{24}$$

over a one slot look ahead horizon. This is a one step surrogate for the time averaged objective (19a): a decision that is favorable at t but degrades once the users move is penalized, so the velocity components of the state become informative and the allocation is conditioned on user motion rather than treating velocity as an inert feature.

4) TRANSITION

Once a group is formed, its two users are removed, its assigned communication PAs are withdrawn from the free pool, and the arc length its block occupies is removed from the placements available to later groups on the same cable. Later groups in the same slot therefore inherit both a smaller user set and a depleted, partially obstructed aperture. The episode terminates when fewer than two users remain. With discount γ, the value of an early pairing accounts for the degraded pool it leaves behind, which is exactly the cross group coupling identified in Section IV and the reason a per pair valuation is insufficient.

B. NETWORK ARCHITECTURE AND TRAINING

1) NETWORK

A single network $Q(\mathbf{s};\theta)$, the JointQNet, maps the 27 dimensional state to nine Q values, one per candidate power split level, through two hidden layers of 256 and 128 ReLU units, for $41\,225$ parameters in total. The 81 joint actions available to a candidate pair are scored as nine such rows, one per placement offset, in a single batched pass. A target copy $Q(\cdot;\theta^-)$ supplies bootstrap targets and is synchronized to θ every 250 episodes.

2) TRAINING

We train from scratch with the double DQN update. Each episode draws a fresh scene of $N \in \{2,4,6,8,10,12\}$ users, with positions uniform over the $L \times W$ area, Gauss Markov velocities, semantic weights sampled from $\{0.3, 0.6, 0.7, 0.8, 1.0\}$, and log normal shadowing, then pairs the scene off end to end as above. Actions are ϵ greedy with ϵ annealed from 1.0 to 0.05 over the first $15\,000$ episodes. Each completed step also writes counterfactual rows for the remaining legal (offset, split) pairs into the slot buffer $\mathcal{B}$, off policy data the double DQN update consumes without extra gradient steps. The replay buffer $\mathcal{D}$ holds 4×10^5 transitions and is sampled in mini batches of 128. Once the round is complete, the scene is advanced one mobility slot and the same decisions are re-scored at the moved positions; each transition's reward is set to the look ahead blend of (**??**) before it is written to the replay buffer $\mathcal{D}$. For a sampled transition the learning target is

$$y = r + \gamma\, Q\Big(\mathbf{s}',\, \arg\max_{a'} Q(\mathbf{s}', a'; \theta);\ \theta^-\Big) \tag{25}$$

with $y = r$ at episode termination. The online net selects the next action and the target net evaluates it, the double DQN correction to value overestimation. The feasibility mask is applied to the online net's selection as well, so a bootstrap target is never formed from an action the successor state could not legally take. We minimize the mean squared error between $Q(\mathbf{s}, a; \theta)$ and y using Adam (learning rate 10^{-3}, gradient norm clipped at 1.0) with $\gamma = 0.99$. Drawing a new random scene each episode exposes the policy to the full mobility induced distribution of positions and velocities. At deployment the trained policy is applied greedily per slot as the scene evolves. Training runs for 10^5 episodes and every 500 episodes the policy is graded against exhaustive greedy on held out scenes with only the best ratio checkpoint retained, which guards against a late unlucky update.

C. ALGORITHM SUMMARY

Algorithm 1 summarizes training. One episode allocates one slot: the inner loop steps over the pairing decisions taken within the frozen geometry of that slot rather than over time, so the discount γ acts across the groups formed inside a slot and not across slots. No slot to slot transition appears anywhere in the update, and mobility reaches the policy only through the velocity components of (21). Deployment is one greedy pass ($\epsilon = 0$) of lines 5 to 10 per slot on the current users.

D. COMPLEXITY AND SCALABILITY

The motivation for a learned policy is that the per slot allocation problem is combinatorial in the number of users. The number of ways to partition K users into pairs is the double factorial $(K-1)!! = 105$ at $K = 8$ and $10\,395$ at $K = 12$, and searching the power split and the antenna placement jointly multiplies this by $(|\mathcal{A}_\alpha||\mathcal{A}_\delta|)^{K/2}$, giving $105 \cdot 81^4 \approx 4.52 \times 10^9$ objective evaluations at $K = 8$ and $10\,395 \cdot 81^6 \approx 2.94 \times 10^{15}$ at $K = 12$. Exhaustive search is therefore usable as an offline benchmark but not as an online allocator, and the growth is super exponential rather than merely steep. This count is only meaningful because the coupling established in Section IV rules out the polynomial time matching solution that the decoupled case admits. Absent that coupling the enumeration would be avoidable entirely, and it is the shared PA pool and the cross group spacing constraint, not the raw size of the search space, that make the per slot problem genuinely hard.

The proposed policy replaces that search with a fixed cost evaluation. Groups are formed sequentially from the candidate pool, and a single forward pass of Q_θ scores one candidate pair across all $|\mathcal{A}_\alpha| = 9$ power splits simultaneously, since the split indexes the output layer rather than the input. The number of forward passes per slot is thus $\sum_{m=0}^{K/2-1} \binom{K-2m}{2}$, which is 50 at $K = 8$ and 161 at $K = 12$, growing as $O(K^3)$. Against exhaustive joint search this is a reduction of roughly eight orders of magnitude at $K = 8$ and thirteen at $K = 12$, and the gap widens monotonically

Algorithm 1 Mobility Aware Semantic DDQN (training). One episode allocates one slot.

Initialize online $Q(\cdot;\theta)$, target $\theta^- \leftarrow \theta$, replay $\mathcal{D}$, $\epsilon \leftarrow 1$
for episode $= 1$ to E **do**
  Sample one slot: N users (positions, velocities, weights, fading), place sensing PAs by the geometric rule, reset free PA pool and cable occupancy $\mathcal{O}$, slot buffer $\mathcal{B} \leftarrow \varnothing$
  while ≥ 2 users remain in the slot **do**
    Build candidate pairs and states $\mathbf{s}_{u,v} = [\boldsymbol{\phi}_u, \boldsymbol{\phi}_v, \mathbf{c}]$, mask offsets infeasible against $\mathcal{O}$
    With prob. ϵ pick a random feasible $(\text{pair}, \alpha, \delta)$, else take $\arg\max Q$ (22)
    Place the group's PA block at offset δ, deduct from pool, claim its arc length in $\mathcal{O}$, then get immediate reward r_t
    Remove the two users and form $\mathbf{s}'$, append $(\mathbf{s}, a, r_t, \mathbf{s}', \text{done})$ to slot buffer $\mathcal{B}$
  end while
  Propagate positions one slot via the Gauss–Markov model: $\mathbf{p} \rightarrow \mathbf{p}'$
  for all $(\mathbf{s}, a, r_t, \mathbf{s}', \text{done}) \in \mathcal{B}$ **do**
    Re-score action at $\mathbf{p}'$ to get r_{t+1}; $r \leftarrow (1-\eta)\, r_t + \eta\, r_{t+1}$; store $(\mathbf{s}, a, r, \mathbf{s}', \text{done})$ in $\mathcal{D}$
  end for
  Sample minibatch; compute target y (25); SGD step on MSE
  Every C episodes set $\theta^- \leftarrow \theta$; decay ϵ
end for

with load, which is the regime that matters for the dense deployments this work targets.

Memory is likewise bounded and independent of K. The network of Section V holds $41\,225$ parameters across its three layers, occupying under 200 kB in single precision, and this footprint is fixed because the state is a 27 dimensional vector whose width does not depend on the number of users in the scene. This is the decisive advantage over the tabular formulation of our prior static design [2], [20], in which the table is indexed by a discretization of the state and therefore grows exponentially in the state dimension. Discretizing the present 27 dimensional continuous state at even ten bins per dimension would require on the order of 10^{27} table entries, a fraction of which no amount of training data could visit. Function approximation is what removes that barrier, and it is the reason the framework extends to user counts at which a tabular controller cannot be instantiated at all. Per slot inference time measured on a single CPU core is reported alongside utility results in Table 2.

VI. SIMULATION RESULTS

Unless stated otherwise all system parameters take the values in Table 1, which is referenced for both the analytical model and the ray traced environment.

Results are reported on the analytical dual waveguide channel of Section III, which admits the large number of independent scenes that training and held out evaluation require. The site specific ray traced environment described below confirms that the analytical trends survive realistic propagation, with the ray traced and analytical utilities agreeing to within 0.7 dB.

A. SIMULATION SETUP

1) DEPLOYMENT GEOMETRY

The service area is a regulation pitch of $L \times W = 105 \times 68$ m. Four cables are anchored at towers of height $H_t = 20$ m at the corners and converge at the hub $\tilde{\boldsymbol{\psi}}_0 = [52.5,\ 34,\ 12]^T$ m, which gives a cable length of $L_\ell = 63.1$ m. The $N_s = 8$ sensing and $N_c = 8$ communication PAs are distributed evenly, such that each cable hosts two of each, and every PA is constrained to its host cable's arc length interval $[0, L_\ell]$ with minimum spacing $\Delta_{\min} = \lambda/2$.

2) USERS AND MOBILITY

Each scene draws K users with positions uniform over the pitch, semantic weights drawn from $\{0.3, 0.6, 0.7, 0.8, 1.0\}$, and independent log normal shadowing with 2 dB standard deviation. Velocities follow the Gauss Markov model (3) with memory $\alpha_v = 0.9$ and $\sigma_v = 1.7$ m/s, set to match the per axis velocity spread measured from the real player tracking data, whose median speed is 1.4 m/s and 95th percentile 4.8 m/s, updated every slot of duration $\Delta t = 40$ ms. The nominal configuration is $K = 8$. To probe the effect of load we sweep $K \in \{2, 4, 6, 8, 10, 12\}$, which spans the regime from ample spectral headroom to contention in which the allocator must choose which streams to prioritize.

3) CHANNEL

The analytical model realizes the free space PASS response of Section III, combining the in waveguide phase accumulated to each PA's arc length position with the over the air propagation term to the user, summed coherently over the PAs assigned to a group. The sensing return uses the two way echo model with radar cross section $\sigma_{\mathrm{RCS},k} = 1.0$ m^2 and coherent processing gain $G_p = T_{\mathrm{int}} B = 4 \times 10^6$ over the $T_{\mathrm{int}} = 40$ ms integration window. The channel and sensing computations were independently reimplemented in MATLAB directly from the governing equations; the two implementations agree to within 1.8×10^{-14} dB in sensing SNR and 1.2×10^{-15} in relative effective channel power, that is, to double precision accuracy.

4) TRAINING PROTOCOL

Training follows the protocol of Section V-B without modification. Fig. 3 reports the resulting learning curves, evaluated both on the frozen slot and after one mobility step.

5) EVALUATION PROTOCOL

All reported results use held out scenes generated from seeds disjoint from training, with the policy applied greedily ($\epsilon =$

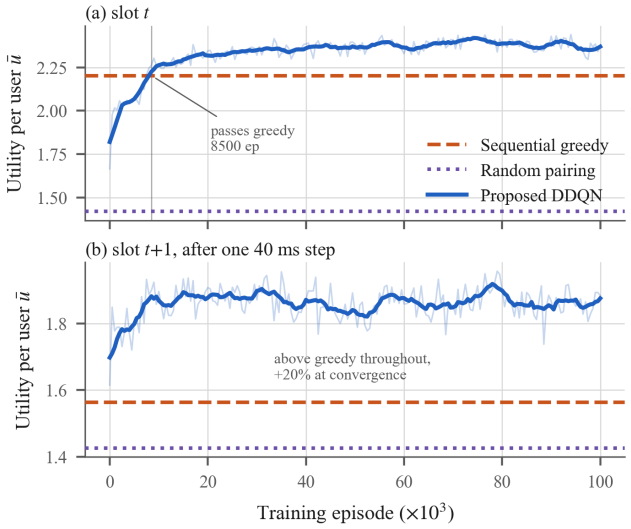

FIGURE 3. Training convergence of the DDQN policy, evaluated every 500 episodes on a held out scene set. (a) On the frozen slot the policy passes the sequential greedy heuristic at roughly $8{,}500$ episodes. (b) The same decisions re-scored after one 40 ms mobility step, with no allocator permitted to re-optimize. A held allocation decays in proportion to how tightly it was fitted to the frozen slot: the greedy heuristic loses 29.0% of its utility, the policy 20.4%, and random pairing 0.4%. Panel (b) therefore measures robustness to staleness rather than a learned anticipation of motion, and the cost of mobility itself is quantified by the periodic re-optimization rows of Table 2. Raw evaluations are shown faint, with a moving average overlaid. The policy's $t+1$ advantage in (b) follows from the blended look ahead reward of (??), which trains it to prefer allocations that remain effective once users move.

0). Every method under comparison is scored on the identical scene set, so differences are attributable to the allocation policy and not to scene sampling. We report the average semantic utility per served user,

$$\bar{u} = \frac{1}{2|\mathcal{P}|} \sum_{p \in \mathcal{P}} \left[G_{p,w} + G_{p,s} \right] \tag{26}$$

with $G_{p,w}$ and $G_{p,s}$ as defined in (17). Normalizing per served user makes the metric comparable across the load sweep. The rate and sensing penalties of (23) shape learning but are excluded from the reported score, and constraint satisfaction is instead reported separately as the fraction of users meeting the rate floor $R_{\min}$ (19i) and the sensing threshold $\Gamma_{\min}$ (19h). For consistency the search based benchmarks of Table 2 optimize this same reported metric rather than the penalized objective.

6) SENSING COMMUNICATION POWER SPLIT

The budget constraint (19g) is exercised by sweeping the split ratio $\rho = P_s/P_{\text{total}}$ over $[0.1, 0.5]$ at the nominal $\rho = 0.3$ of Table 1, reporting semantic utility and the sensing SNR margin $\Gamma_{s,k} - \Gamma_{\min}$ jointly, such that the operating point can be read against both objectives rather than either alone. This sweep is shown in Fig. 6.

For the stadium and football application scenarios, the target evaluation is a high fidelity and site specific ray traced reconstruction of Levi's Stadium, built in Blender and

TABLE 1. System Parameters

Parameter	Symbol	Value
Carrier frequency	f_c	28 GHz
Wavelength	λ	10.71 mm
Effective refractive index	n_{eff}	1.4
Guided wavelength	λ_g	7.65 mm
Path loss constant	$\eta = (\lambda/4\pi)^2$	7.27×10^{-7}
Signal bandwidth	B	100 MHz
Min. PA spacing	$\Delta_{\min} = \lambda/2$	5.36 mm
Noise figure	NF	8 dB
Noise power	$\sigma^2 = \sigma_s^2$	2.5×10^{-12} W (-86 dBm)
Total power budget	P_{total}	1 W (30 dBm)
Power split ratio	ρ	0.3
Sensing / comm. power	P_s / P_c	0.3 W / 0.7 W
Radar cross section	$\sigma_{\text{RCS},k}$	1.0 m^2
Integration window	T_{int}	40 ms
Processing gain	$G_p = T_{\text{int}} B$	4×10^6 (+66 dB)
Min. sensing SNR	$\Gamma_{\min}$	15 dB
Number of users	K	8
Number of groups	$P = K/2$	4
Sensing / comm. PAs	N_s / N_c	8 / 8
Pitch dimensions	$L \times W$	105 m $\times$ 68 m
Tower height	H_t	20 m
Convergence height	H_c	12 m
Cable length	L_ℓ	63.1 m
Fidelity bounds	(a_1, a_2)	$(0.30,\ 0.98)$
Fidelity shaping	(c_1, c_2)	$(0.25,\ -0.8)$
Semantic weights	w_k	$\{0.3, 0.6, 0.7, 0.8, 1.0\}$
Slot duration	Δt	40 ms (25 Hz)
Number of slots	T	1500 (60 s)
Velocity memory	α_v	0.9
Velocity std. dev.	σ_v	1.7 m/s
Min. per user rate	$R_{\min}$	1 bit/s/Hz

imported into Sionna [33] for channel modeling (Fig. 4), with user trajectories replayed from open real world player tracking data [34]. We emphasize that this environment is a physically faithful simulation driven by replayed player tracking data rather than a live digital twin, as it maintains no real time bidirectional link to an instrumented venue. Coupling the ISAC sensing waveguide to such a live feed to realize a full digital twin is a direction we outline in Section VII.

B. BASELINE COMPARISONS

Each baseline replaces exactly one component of the proposed framework with a reduced version, such that the resulting gap attributes performance to that component rather than to the system as a whole. All baselines are evaluated on the identical held out scenes.

Three search benchmarks bound what is achievable per slot without learning. Exhaustive search enumerates all $(K-1)!!$ feasible pairings and takes the best, with the power split fixed by the closed form (18). This is our primary point of comparison since it is the strongest allocation reachable without learning the power split, and is an offline bound rather than a deployable allocator. A sequential greedy

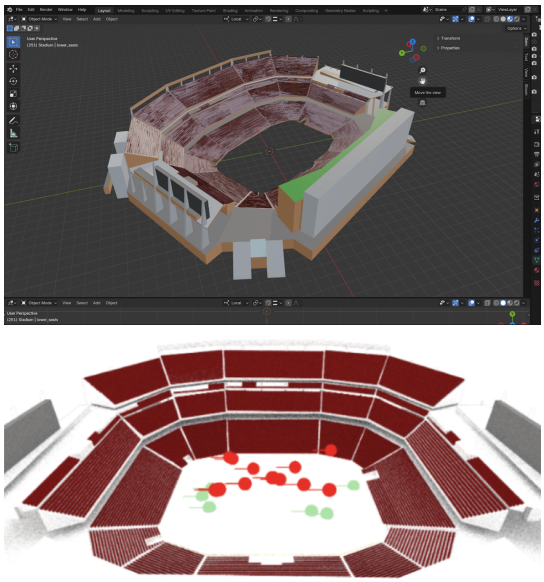

FIGURE 4. Target evaluation environment. Top shows a ray traced reconstruction of Levi's Stadium built in Blender and imported into Sionna. The bottom shows the same reconstruction with replayed player positions in green and pinching antenna positions in red.

heuristic instead commits the single best remaining pair at each step without revisiting earlier commitments, which is the decision layer used by the static formulation of [2]. It costs $O(K^3)$ rather than $O((K-1)!!)$ and is therefore the fair cost matched point of comparison for a learned policy. Finally, an oracle split benchmark searches the grid $\mathcal{A}_\alpha$ for every group in addition to the pairing, maximizing the same pure semantic utility that is reported as performance in (26) rather than the penalized objective of (23). It is thus a strong offline reference rather than a proven bound, since it fixes the block position by an inner search rather than optimizing placement jointly with the pairing, and a joint enumeration at $K = 6$ exceeds it by 1.55% on average and 13.90% at worst. It is not implementable online since it requires evaluating every pairing at every split.

Two further comparisons bracket the semantic claim itself. A throughput maximizing allocator runs the identical machinery with all semantic weights set to $w_k \equiv 1$ such that it maximizes rate alone and is blind to which stream is decision relevant, isolating what the semantic weighting contributes. A uniformly random feasible pairing at the closed form split establishes the performance floor.

Beyond these internal reductions we reimplement the closed form KKT power allocation derived for NOMA assisted PASS in [8] and evaluate it on the identical scenes. That solution maximizes the sum rate subject to the weak user's target rate, and admits the closed form $\alpha^*_{p,s} = \max\{0, \min\{\tilde{\alpha}_{p,s}, 1/2\}\}$ with $\tilde{\alpha}_{p,s} = (\rho_w + 1 - 2^{R_{\min}})/(\rho_w 2^{R_{\min}})$, where $\rho_w = |\tilde{h}_{w(p),p}|^2/\sigma^2$ is the weak user's receive SNR. We evaluate it in two configurations. The first retains the semantically aware pairing and substitutes only the split, which isolates the power allocation rule against our own closed form (18). The second additionally drives the pairing by sum rate, which gives a fully rate driven allocator and a published point of comparison against which the semantic objective could be read. Both of those obtain their pairing by exhaustive enumeration and are therefore offline references. A third configuration substitutes the same split into the sequential greedy pairing, which is the strongest allocator built from published components that remains affordable at the slot rate and is thus the cost matched comparison for the learned policy.

Three reduced architectures isolate the physical design. The single straight waveguide baseline collapses the deployment to the degenerate one cable geometry of prior PASS work [2], with the sensing and communication rails fixed at the convergence height and PAs sliding along a single axis, which isolates what the converging cable geometry buys. The static PA position baseline freezes all PAs at their slot zero placement while still re-optimizing pairing and split each slot, which isolates the contribution of dynamic repositioning under mobility. The communication waveguide only baseline disables sensing entirely ($P_s = 0$, with the full budget to communication), which quantifies the utility cost of reserving power for the sensing function. We note the allocator is supplied with exact user positions in both configurations, so this baseline bounds the cost of sensing but does not measure the value of tracking.

Finally, a periodic re-optimization baseline applies the full policy once every M slots and holds it constant in between, for $M \in \{5, 10, 25\}$. This quantifies how much of the gain genuinely depends on tracking mobility at the slot rate rather than on the quality of any single allocation, which matters directly for deployments where per slot re-optimization is not affordable.

Table 2 reports mean semantic utility per served user (26) at the nominal $K = 8$ and under contention at $K = 12$, alongside constraint satisfaction and per slot inference cost.

C. RESULTS AND DISCUSSION

The first question Table 2 is built to answer is what the semantic weighting itself contributes, which the throughput maximizing row isolates by running identical machinery at $w_k \equiv 1$ and scoring the result under the true weights. Because that row optimizes a different objective over the same feasible set as the semantically aware search and is then scored on the latter's metric, it is bounded above by it at every load by construction, and the quantity of interest is the size of the shortfall rather than its sign. The mechanism that governs that size is spectral headroom. When the deployment can serve every user near its fidelity ceiling the logistic $\xi(\gamma)$ of (16) saturates for all of them, the weighted and unweighted objectives select nearly the same allocation, and ignoring w_k costs little. Once demand exceeds what the waveguides can deliver, some users must be served below saturation, the allocator is forced to choose whom to starve, and only the weighted objective encodes that preference. Measured against the semantically aware exhaustive row of Table 2, the shortfall is 1.8% at $K = 8$ and 1.4% at $K = 12$.

TABLE 2. Baseline comparison at seeds disjoint from training, on the analytical PASS channel. Semantic utility per served user (26). Paired entries are reported as K=8 / K=12. Rows above the rule use 200 held out scenes at five seeds, except the four search based rows at $K = 12$, which use 30 scenes at three seeds because each scene enumerates $(K-1)!! = 10{,}395$ matchings with an inner placement search. Sensing $\Gamma_{\min}$ satisfaction exceeds 99.8% for every method listed, since sensing PA placement is geometric and invariant to the allocation decision, and is omitted as a column for that reason. Rows below the rule isolate one component each and are measured under their own protocols, so they are read against the proposed policy under the matching protocol rather than against each other. The temporal rows roll mobility tracks at the slot rate, and the geometry row uses the search based allocator because a policy trained on one rig does not transfer to another. The sequential greedy pairing, KKT split row performs the identical pairing and placement work as the sequential greedy row and differs from it only in a closed form scalar, so it is reported at that row's measured inference cost.

Method	$\bar{u}$ (K=8)	$\bar{u}$ (K=12)	$R_{\min}$ sat. (%)	Inference (ms/slot)
Proposed DDQN	2.487	2.062	91.9 / 89.0	0.87 / 1.60
Exhaustive search, closed form α	2.443	1.971	100.0 / 99.5	—
Exhaustive search, oracle α	2.650	2.248	95.5 / 92.7	—
Sequential greedy heuristic	2.257	1.700	99.4 / 99.5	18.77 / 53.08
Throughput maximizing	2.400	1.944	100.0 / 99.7	—
Semantic pairing, KKT split [8]	2.649	2.250	100.0 / 100.0	—
Sum rate pairing, KKT split [8]	2.598	2.229	100.0 / 100.0	—
Sequential greedy pairing, KKT split [8]	2.344	1.811	99.3 / 100.0	18.77 / 53.08
Random pairing	1.966	1.597	98.0 / 99.7	—
Single straight waveguide	2.457	2.220	100.0 / 100.0	—
Static PA positions	2.035	1.837	85.0 / 92.1	0.978 / 1.600
Communication waveguide only	2.697	2.294	94.4 / 95.6	—
Periodic re optimization (M=10)	2.150	1.985	87.5 / 87.2	0.099 / 0.189
Decoupled optimum, max weight matching	2.111	1.695	99.3 / 98.9	—

A second and more actionable observation concerns which decision the semantic weights belong in. Substituting the KKT split of [8] while retaining the semantically aware pairing raises $\bar{u}$ to 2.649 at $K = 8$ and 2.250 at $K = 12$, exceeding our closed form (18) by 8.4% and 14.2% respectively and matching the oracle split reference to within 0.1% at both loads while satisfying the rate floor of every user. Driving the pairing by sum rate as well returns it to 2.598 and 2.229. Both configurations obtain their pairing by exhaustive enumeration, which is why their inference column is empty, so neither answers whether the KKT split is worth adopting in a deployable allocator. Substituting it into the sequential greedy pairing does answer that, and it raises $\bar{u}$ to 2.344 at $K = 8$ and 1.811 at $K = 12$ at that heuristic's own cost. The proposed policy remains 6.1% and 13.9% above it. The KKT advantage visible in Table 2 therefore rests on the exhaustive pairing it is combined with rather than on the split rule alone, and the learned policy keeps its margin against the strongest allocator available at comparable cost. Part of that margin is bought against the rate floor, which the greedy and KKT combination meets for 99.3% and 100.0% of users against the policy's 91.9% and 89.0%.

Read together these two configurations say that the grouping and the split want different objectives. The grouping decision should be semantic since w_k enters the utility multiplicatively and determines which users are worth co-scheduling at all. The intra-group split should be rate driven, since $\xi(\gamma)$ saturates and the split's remaining leverage sits almost entirely on the $\log_2(1+\gamma)$ factor. The closed form (18) allocates by channel inversion and therefore favors the weak user, sacrificing more rate than the semantic weighting recovers, whereas the KKT form solves for the weak user's rate to meet $R_{\min}$ with equality, which drives $\alpha_{p,w}$ to 0.520 on average at $K = 8$ and 0.535 at $K = 12$, mostly below the floor of our discrete grid $\mathcal{A}_\alpha$ and spends nothing on margin the objective does not reward. This decomposition is the clearest design guidance the evaluation yields. The temporal rows carry a second one. Applying the policy once every ten slots rather than every slot cuts the amortized inference cost from 0.908 to 0.099 ms per slot, a factor of nine, for a 17.2% loss in utility, and the loss saturates by that point rather than growing with the holding interval.

Against this the learned policy is positioned by both quality and cost, and Section IV determines which comparison is the meaningful one. A policy that merely approached exhaustive search would not justify itself, since the decoupled optimum is already reachable exactly and cheaply. What it has to do is exceed the allocation that a per pair valuation can reach, by exploiting the contention structure that valuation cannot see. Set against the 4.52×10^9 and 2.94×10^{15} objective evaluations that exhaustive joint search requires at two loads (Section D), the policy is the only allocator in the table that remains a candidate for online use at the slot rate. The degradation with load is diagnostic rather than incidental. The policy commits groups sequentially from a shrinking candidate pool, so the number of irrevocable decisions per slot grows with K while each is made from a fixed width context summary, and the preceding paragraph

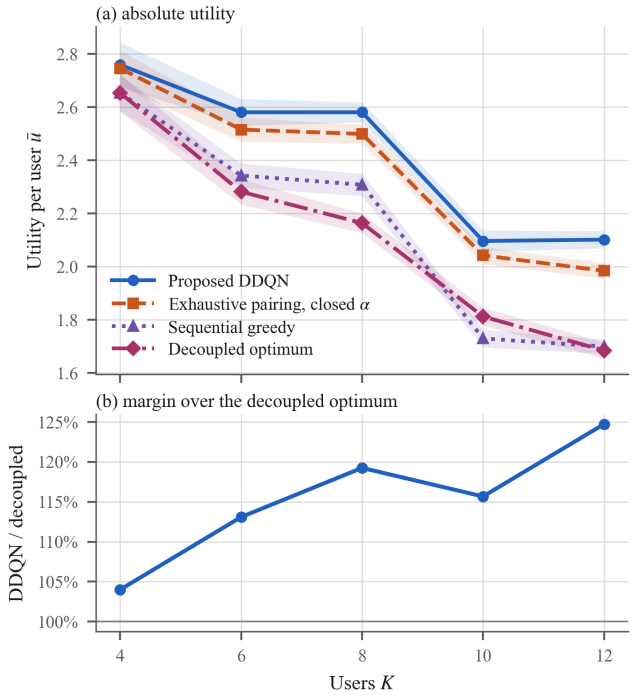

FIGURE 5. **Allocation quality against load, all four allocators on identical scenes. (a) Absolute utility per served user, shaded by standard error. The exhaustive row enumerates pairings at the fixed closed form split (18), so the proposed policy exceeds it by optimizing the split as well. The oracle split benchmark of Table 2, which searches the split too, is not attained. (b) The margin over the decoupled optimum, the quantity Section IV predicts, rising from 104% at $K = 4$ to 125% at $K = 12$. The policy is flat between $K = 6$ and $K = 8$ because the antenna pool grants two communication PAs per group at both ends, so the per group structure is unchanged. Panel (a) plots the policy against the three allocators that frame the structural claim of Section IV, namely exhaustive enumeration at the fixed split, the cost matched sequential greedy heuristic, and the decoupled optimum. The alternative split rules of Table 2 vary the power allocation rather than the allocation structure and are reported there.**

localizes the residual gap further. A fixed analytic split already recovers most of the available split gain, so the learned split head is expending capacity on a function that admits a near optimal closed form. Retraining against the KKT split as the behavioral reference, rather than against channel inversion, is the natural next step and is expected to recover a substantial share of the remaining gap.

The sensing constraint is likewise inactive as a discriminator. $\Gamma_{\min}$ satisfaction is 100.0% at $K = 8$ and 99.8% at $K = 12$ for every method, since sensing PA placement follows the geometric rule of Section III and is invariant to the allocation decision. It becomes a discriminating constraint only once $\mathbf{s}_s(t)$ enters the action space, which we identify as future work.

Fig. 6 reports the sweep. At the nominal $\rho = 0.3$ the sensing SNR margin is 30.8 dB above the $\Gamma_{\min} = 15$ dB threshold, so tracking clears with wide headroom while 70% of the budget serves communication. As ρ decreases the average utility rises monotonically, reaching $\bar{u} = 2.66$ at $\rho = 0.01$, while the margin narrows but never crosses the threshold: even when sensing receives only 1% of the budget

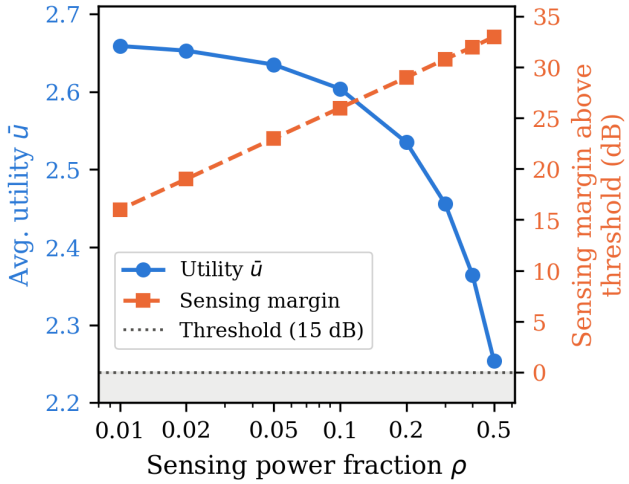

FIGURE 6. **Power split resilience at fixed total power. As the sensing power fraction ρ is reduced from 0.5 to 0.01, average utility $\bar{u}$ rises while the sensing margin above the $\Gamma_{\min} = 15$ dB detection threshold falls but stays positive throughout. Even when sensing is allocated only 1% of the budget ($\rho = 0.01$), detection clears the threshold by 16 dB.**

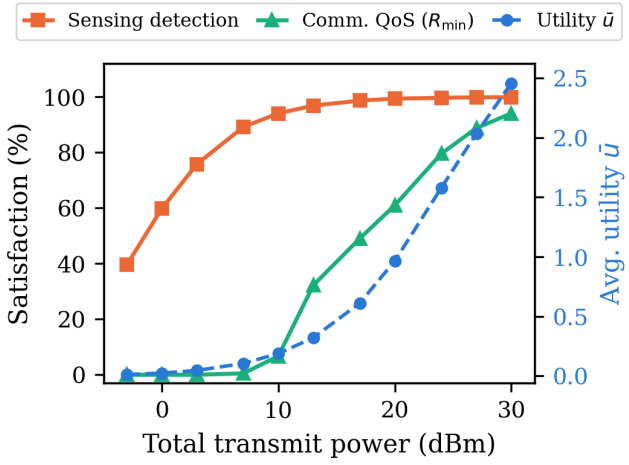

FIGURE 7. **Total power resilience at fixed split ($\rho = 0.3$). As the total transmit power is reduced, sensing detection stays near perfect down to about 0 dBm (a thousand fold reduction) while communication QoS ($R_{\min}$ satisfaction) collapses far earlier, showing that communication, not sensing, is the binding constraint under power scarcity.**

it still clears detection by 16 dB. Conversely, raising ρ to 0.5 trades utility down to 2.25 for a 33.0 dB margin. The tracking constraint is therefore slack across the entire swept range, which shows the fixed 0.3 split is conservative on the sensing side and could be pushed toward communication without compromising tracking.

A complementary stress test shrinks the entire power budget with the split held at $\rho = 0.3$ (Fig. 7). Sensing detection remains essentially perfect as the total transmit power falls from 30 dBm to about 0 dBm (1 mW), a thousand fold reduction, and degrades only below that. Communication QoS collapses far earlier: the fraction of users meeting $R_{\min}$ drops from 94% at 30 dBm to 32% by 13 dBm and to near zero below 10 dBm. This asymmetry–graceful sens-

TABLE 3. Real-data validation across two matches and three windows each.

Match (squad)	Window	$\bar{u}$	Sensing
Game 1 (home)	0–1 min	2.65	100.0%
Game 1 (home)	20–21 min	2.77	100.0%
Game 1 (home)	40–41 min	2.73	100.0%
Game 2 (away)	0–1 min	2.61	100.0%
Game 2 (away)	20–21 min	2.46	99.9%
Game 2 (away)	40–41 min	2.53	99.9%

ing degradation against an abrupt communication collapse–confirms that communication rather than sensing is the binding constraint under power scarcity, consistent with the large coherent processing gain that makes sensing inexpensive in this deployment.

One limitation bounds the scope of these results. The learned policy is trained on frozen per slot scenes with velocity entering only as a state feature, so the reported figures characterize a per slot policy conditioned on motion rather than a policy trained through mobility. The periodic re-optimization rows of Table 2 quantify what that costs directly, by holding an allocation for M slots and re-scoring it against the users as they move. Those rows, rather than the $t+1$ panel of Fig. 3, are the measurement of what mobility costs, since that panel compares held decisions whose decay tracks how tightly each allocator fitted the slot it was given. Fig. 8 shows the behavior those rows summarize, with the grouping and the communication PA positions following real player motion across three snapshots of a single possession.

D. REAL DATA VALIDATION AND GENERALIZATION

To confirm the policy operates on real motion rather than the synthetic Gauss-Markov model it trained against, we run the trained DDQN unmodified on two independent Metrica matches (different teams, three one-minute windows each). Table 3 reports the result.

Per served user utility stays consistent within each match, averaging 2.72 for Game 1 and 2.53 for Game 2, and transfers across matches with no retraining, while sensing clears between 99.9% and 100.0% of users in every window.

To check the analytical channel model against full-wave propagation, we re-traced the deployment in Sionna against a Blender reconstruction of Levi's Stadium comprising 34 meshes at $98{,}665$ faces, with custom radio materials for the field, seating, goal frames and scoreboard, and compared the resulting effective channel to the analytical model on real player trajectories. Over 300 tracking slots drawn from five windows spanning the match, the ray traced effective channel power exceeds the analytical model by $+0.72$ dB on average, with a per slot standard deviation of 0.94 dB arising from multipath, and the sensing SNR by $+1.95$ dB. Pooling across windows is necessary rather than cosmetic. The per window means span only $+0.58$ to $+0.80$ dB for the channel, so that figure is stable, whereas the sensing means span $+1.24$ to $+3.27$ dB and a single window would not characterize it. The agreement is close in the mean and shows no systematic attenuation, but it is not a line of sight channel. Individual links depart from the analytical prediction by up to ± 13 dB as players move through the stadium, and the close agreement is a property of the average over links rather than any of them.

E. SENSING COVERAGE AND THE VALUE OF THE RIG

Section III motivates the converging cable geometry on the grounds that widely separated anchors at differing heights give better visibility of the whole service area than a low perimeter rail. Fig. 10 tests that directly, by ray tracing the sensing field over the pitch for both rigs in the site specific Levi's Stadium reconstruction. Both are given the same number of sensing PAs, the same power, and crucially the same placement rule. The antennas are positioned by the geometric rule of Section V against the tracked players at one real slot of the Metrica data, so the comparison is between deployments rather than between a deployment and a frozen baseline.

The straight rail concentrates its coverage where its antennas happen to be gathered and leaves the far third of the pitch 15 to 20 dB weaker, with several tracked players standing in it. The converging rig spreads its field across the whole area, because arc length on that geometry couples horizontal reach with height. Sliding a PA outward to cover a distance corner also raises it. Pooled over 10 slots spanning the match on a 3 m grid, the converging rig is ahead by 2.74 dB at the median in every slot, and by 4.18 dB at the fifth percentile in 9 of the 10. The advantage grows on a finer grid, to 4.4 and 8.1 dB in the slot shown, because a coarse grid steps over the nulls and the rail's nulls are the deeper ones. We claim no advantage in the single worst cell, which is positive in 7 of the 10 slots but with a standard deviation of 5.5 dB across them, and is therefore not a reliable effect.

F. RECOVERING PLAYER KINEMATICS FROM THE ECHOES

The allocator consumes position and velocity estimates that the sensing waveguide produces, so it is worth asking how good those estimates are at the SNR the deployment actually achieves. Section III assumes a user meeting the tracking threshold is localized accurately enough to substitute its estimate for its true position and does not model the residual error. Here we characterize that error without feeding it back, so no result of Section VI is affected. Ranging accuracy follows the Cramér-Rao bound $\sigma_R = c/(2B\sqrt{2\Gamma_{s,k}})$ at the measured two way SNR, and the sensing PA constellation converts it to a position error through the geometric dilution $\sqrt{\mathrm{tr}((\mathbf{H}^T\mathbf{H})^{-1})}$, with $\mathbf{H}$ the ground projected unit vectors from the user to each sensing PA.

Fig. 11 reports both. Over 1500 slots of real tracking the position error has a median of 6.6 mm and a 95th percentile of 22.8 mm, which is a fraction of the 10.71 mm wavelength and over an order of magnitude below the 80 mm a player covers in one slot. Position is therefore not the limitation. Velocity is, because it is obtained by differencing

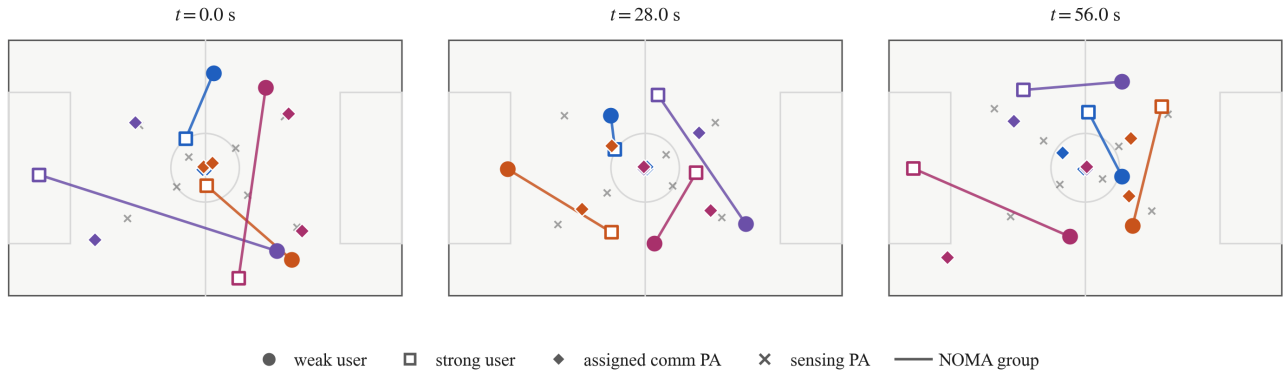

FIGURE 8. **Allocation tracking player motion over a single possession. Three snapshots from the Metrica tracking data show user positions, the current grouping, and the communication PA block positions along each cable, illustrating how the placement follows the users between slots.**

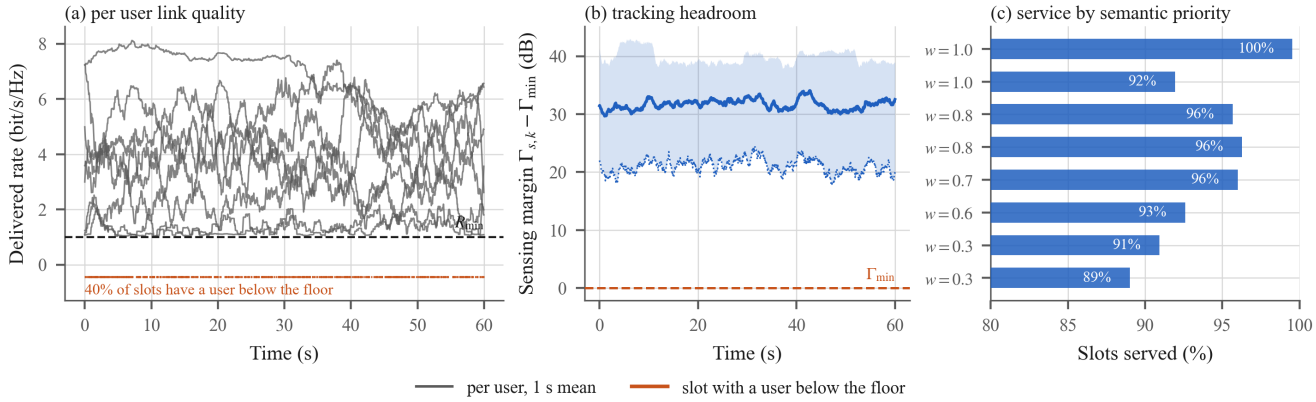

FIGURE 9. **Mission critical application layer built on the allocator output, over one minute of real tracking. (a) Delivered rate per user as a one second mean, with the rate floor marked and a rug at full slot resolution marking every slot in which some user falls below it. (b) Tracking headroom, median and range across users, sitting some 30 dB above $\Gamma_{\min}$ throughout. (c) Fraction of slots served against semantic priority, rising from 90% at $w = 0.3$ to 96% at $w \geq 0.7$.**

two position estimates one slot apart, which amplifies the error by $\sqrt{2}/\Delta t = 35$ s^{-1} and yields a per slot velocity uncertainty of 0.23 m/s, roughly 15% of the median player speed. The aggregate speed distribution nonetheless matches the recorded one closely, since that uncertainty is small against the 0 to 5 m/s spread of the data. The practical consequence is that a single slot's velocity is not trustworthy while its statistics are, which is consistent with the age of the estimates rather than their precision being the binding quantity for this deployment.

G. RESILIENCE UNDER PARTIAL FAILURE

The evaluation so far assumes the deployment is intact. Because the target application is mission critical, the more revealing question is what the allocator does when capacity is suddenly lost. We remove half of the communication pinching antennas from the pool at the start of each scene, leaving everything else unchanged, and compare the proposed policy against a semantics blind allocator that optimizes the same objective with every w_k set to unity. Both allocators observe the reduced pool, so this measures adaptation rather than surprise. The outcome is reported as delivered rate rather than semantic utility. The utility of (17) already contains w_k, so measuring the protection in utility would be true by construction, whereas delivered rate contains no semantic term and can therefore falsify the claim.

Fig. 12 reports the result over 300 scenes at three seeds. With the deployment intact the semantics blind allocator delivers essentially flat 3.7 bit/s/Hz irrespective of priority, a separation of $1.0\times$ between the lowest and highest weighted users, while the proposed policy separates them by $3.2\times$. Removing half the antennas costs both allocators roughly half of the delivered rate in aggregate, which is expected, since the capacity is simply gone. What differs is who absorbs the loss. The blind allocator degrades every user by the same proportion and its separation stays flat at $1.1\times$. The proposed policy instead widens its separation to $3.8\times$, holding 3.17 bit/s/Hz for the highest weighted users while the lowest fall to 0.84.

The reality is that semantic awareness does not create capacity under failure but rather decides who keeps it. The blind allocator in fact delivers more to the lowest weighted users, 1.73 against 0.57 bit/s/Hz, precisely because it does not distinguish them from the rest. For a deployment whose

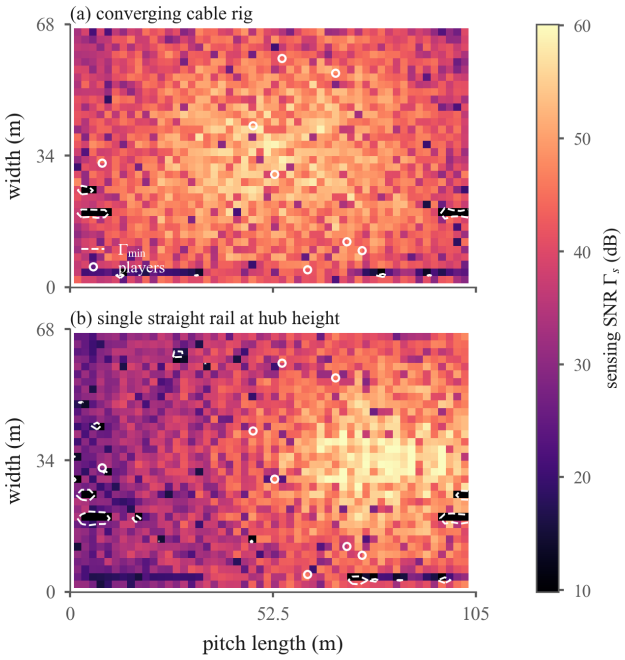

FIGURE 10. Ray traced sensing SNR over the pitch at one slot of real player tracking, on a 2 m grid, for (a) the converging cable rig and (b) the single straight rail at the convergence height. Both carry N_s sensing PAs placed by the same geometric rule against the same tracked players, marked as open circles. The dashed contour is the tracking threshold $\Gamma_{\min}$. The rail leaves the far third of the pitch weak. The converging rig covers the area more evenly. Pooled over 10 slots on a 3 m grid the converging rig leads by 2.74 dB at the median in every slot and by 4.18 dB at the fifth percentile in 9 of 10.

value is carrying the streams that matter when conditions are worst, that trade is the intended behavior, and it is the concrete sense in which the architecture is resilient rather than merely high throughput.

VII. CONCLUSION

This paper developed a mobility aware resource allocation framework for a dual waveguide ISAC pinching antenna system. We proposed an elevated converging cable deployment carrying co located sensing and communication waveguides, which recovers the straight single waveguide geometry of prior PASS work as a degenerate case, and derived the sensing channel through its round trip to a tracking constraint that is met with roughly 30 dB of margin, so that the binding resource is the age of the position estimates rather than the sensing power share. On this we formulated the joint allocation problem over user grouping, PA assignment, NOMA power split and PA repositioning under a shared power budget, and developed a double deep Q network that learns the grouping and the split as a single discrete policy conditioned on user positions and velocities.

The evaluation yielded three findings that we regard as more informative than an aggregate performance figure. First, the per slot allocation separates over groups only under a decoupled placement rule, where it reduces to maximum weight perfect matching and is exactly solvable in polynomial time. Contention for the finite pinching antenna pool and the spacing constraint acting across groups are what remove the deployed system from that class, and are therefore what a learned policy has to earn its place on. Second, the grouping and the power split call for different objectives. Retaining a semantically aware pairing while adopting the sum rate optimal KKT split of [8] outperformed our channel inversion closed form by 8.4% at nominal load and 14.2% under contention, and matched the per slot oracle to within 0.1% while satisfying the rate floor for every user. Third, the learned policy is evaluated against the decoupled optimum rather than against enumeration alone, since that optimum is both exactly attainable and inexpensive and is consequently the standard any online allocator must clear to be worth deploying.

These results also delimit the work. The reported figures are produced on the analytical free space channel, which the site specific ray traced environment corroborates to within 0.7 dB in the mean, and retraining the policy against the KKT split rather than channel inversion remains a natural next step. The ray traced reconstruction is moreover an empty stadium, and because spectators are the dominant blockers at this carrier, populating the stands is the change most likely to separate the two channel models.

Several modeling choices made here to keep the per slot problem tractable are candidates for relaxation. Promoting the split ρ from a fixed parameter to a learned action, or resolving it through a Stackelberg or bargaining formulation between the two objectives, would let the system trade sensing quality against served utility online, and promoting the sensing PA positions $\mathbf{s}_s(t)$ into the action space would turn the sensing constraint from a feasibility monitor into a learning signal. Blending the immediate reward with the same quantity re-evaluated after the users have advanced would train each placement against the geometry it will actually face, at the cost of a look ahead horizon that has to be tuned against the slot duration.

Finally, the ISAC sensing waveguide supplies exactly the real time position feedback that a live digital twin requires. Instrumenting a venue so that this sensing loop continuously updates the ray traced model would elevate the present site specific simulation into a full ISAC enabled digital twin [31], [32], which we leave to future work.

REFERENCES

[1] Y. Liu, H. Jiang, X. Xu, Z. Wang, J. Guo, and C. Ouyang, "Pinching-antenna systems (PASS): A tutorial," *IEEE Trans. Commun.*, vol. 74, pp. 4881–4918, Jan. 2026, doi: 10.1109/TCOMM.2026.3658289.

[2] D. Chui, C. Fiske, W. Wang, and K. Ramamoorthy, "Joint user grouping, power allocation and antenna assignment for NOMA-enabled pinching antenna systems for next-generation multiple access," unpublished manuscript, 2026.

[3] Z. Yang, N. Wang, Y. Sun, Z. Ding, R. Schober, et al., "Pinching antennas: Principles, applications and challenges," arXiv:2501.10753, Jan. 2025.

[4] Y. Liu, Z. Wang, X. Mu, C. Ouyang, X. Xu, and Z. Ding, "Pinching-antenna systems: Architecture designs, opportunities, and outlook," *IEEE Commun. Mag.*, 2025, doi: 10.1109/MCOM.001.2500037.

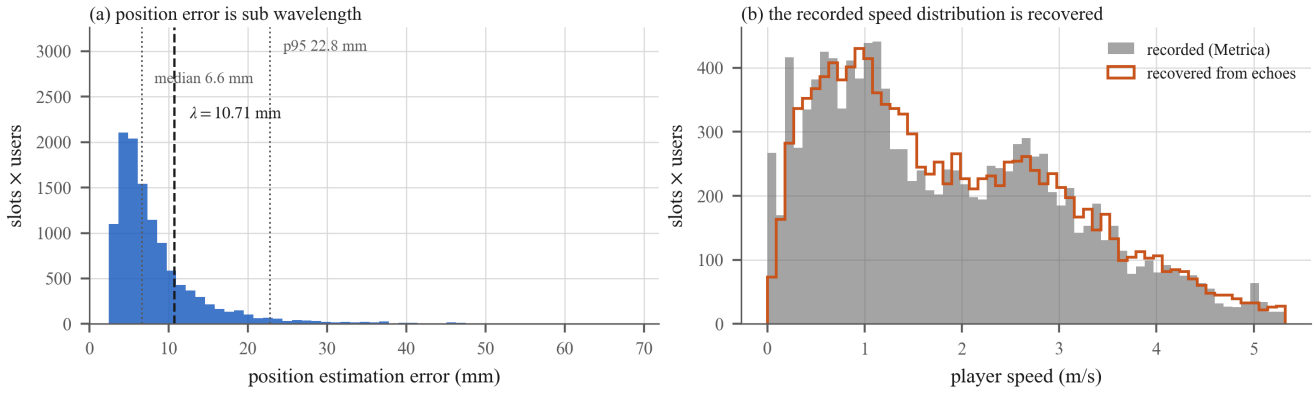

FIGURE 11. **Kinematics recovered from the sensing echoes over 1500 slots of Metrica tracking. (a) Position estimation error against the carrier wavelength. The median of 6.6 mm is well inside a wavelength. (b) The recovered speed distribution against the recorded one. The per slot velocity uncertainty is $\sqrt{2}\,\sigma_p/\Delta t = 0.23$ m/s, about 15% of the median speed, so individual slot velocities are noisy while the distribution is faithfully recovered. The trajectories are replayed from the same recording, so this measures the estimator rather than the realism of the mobility model.**

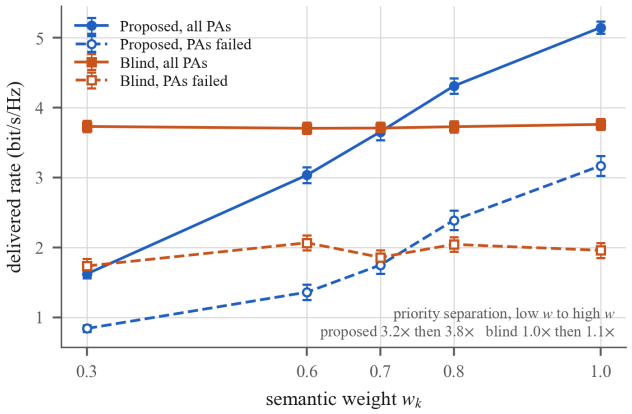

FIGURE 12. **Delivered rate against semantic weight, with the deployment intact (solid) and after half the communication PAs fail (dashed), over 300 scenes at three seeds. Error bars are the standard error of the mean. The semantics blind allocator is flat in both states, separating the lowest and highest priority users by $1.0\times$ and $1.1\times$. The proposed policy separates them by $3.2\times$ intact and widens to $3.8\times$ under failure. Delivered rate is reported rather than semantic utility because the utility already contains w_k. The metric shown is neutral with respect to priority, so the separation is a measured effect rather than a definitional one.**

[5] Z. Ding, R. Schober, and H. V. Poor, "Flexible-antenna systems: A pinching-antenna perspective," *IEEE Trans. Commun.*, vol. 73, no. 10, p. 9236, Oct. 2025.

[6] J. Zhao, X. Mu, K. Cai, Y. Zhu, and Y. Liu, "Waveguide division multiple access for pinching-antenna systems (PASS)," arXiv:2502.17781, Feb. 2025.

[7] X. Xu, X. Mu, Y. Liu, and A. Nallanathan, "Multi-mode pinching-antenna systems: Mode selection or mode combining?," arXiv:2603.08472, Mar. 2026.

[8] Z. Zhou, Z. Yang, G. Chen, and Z. Ding, "Sum-rate maximization for NOMA-assisted pinching-antenna systems," *IEEE Wireless Commun. Lett.*, vol. 14, no. 9, p. 2728, Sept. 2025, doi: 10.1109/LWC.2025.3578312.

[9] Y. Fu, F. He, Z. Shi, and H. Zhang, "Power minimization for NOMA-assisted pinching antenna systems with multiple waveguides," arXiv:2503.20336, Mar. 2025.

[10] K. Wang, Z. Ding, and R. Schober, "Antenna activation for NOMA assisted pinching-antenna systems," *IEEE Wireless Commun. Lett.*, vol. 14, no. 5, p. 1526, May 2025.

[11] Y. Xu, Z. Ding, and D. Cai, "QoS-aware NOMA design for downlink pinching-antenna systems," *IEEE Trans. Commun.*, 2025, doi: 10.1109/TCOMM.2025.3605466.

[12] S. Kumar, K. Singh, C.-P. Li, and Z. Ding, "Pinching-antenna-assisted short packet NOMA communication systems," *IEEE Wireless Commun. Lett.*, vol. 15, p. 2954, 2026, doi: 10.1109/LWC.2026.3689693.

[13] Y. Li, Z. Hu, H. Li, X. Qin, D. Li, and X. Mu, "Security-aware pinching-antenna systems (PASS) for physical-layer security," unpublished.

[14] K. Tang, S. He, X. Xiu, B. Zheng, W. Feng, W. Che, and Q. Xue, "Throughput maximization design for RIS-assisted mmWave WPCN-NOMA systems," *IEEE Trans. Green Commun. Netw.*, vol. 10, 2026, doi: 10.1109/TGCN.2025.3540888.

[15] Y. Kaura, J. Jose, B. Lall, R. K. Mallik, A. Singhal, and V. Bhatia, "Efficient grant-free uplink NOMA with multi-IRS and sidelink relays for IIoT-driven smart factories," *IEEE Wireless Commun. Lett.*, vol. 15, p. 21, 2026, doi: 10.1109/LWC.2025.3617114.

[16] T. Kurayama and T. Miyajima, "Sum-rate improvement of asynchronous NOMA downlink through partial correlation detection," *IEEE Wireless Commun. Lett.*, vol. 15, p. 3074, 2026, doi: 10.1109/LWC.2026.3692136.

[17] Z. Ali, et al., "Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems," *IEEE Access*, 2016, doi: 10.1109/ACCESS.2016.2604821.

[18] S. Mounchili and S. Hamouda, "New user grouping scheme for better user pairing in NOMA systems," in *Proc. Int. Wireless Commun. Mobile Comput. (IWCMC)*, Limassol, Cyprus, 2020, pp. 820–825, doi: 10.1109/IWCMC48107.2020.9148054.

[19] Y. Cao, S. Wang, M. Jin, N. Zhao, Y. Chen, and Z. Ding, "Joint user grouping and power optimization for secure mmWave-NOMA systems," *IEEE Trans. Wireless Commun.*, vol. 21, no. 5, p. 3307, May 2022.

[20] D. Chui, W. Wang, and K. M. Kattiyan Ramamoorthy, "Semantic-aware learning-based NOMA user grouping for programmable 6G RANs," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Kuala Lumpur, Malaysia, 2026, pp. 1–6, doi: 10.1109/WCNCW67598.2026.11555565.

[21] H. Xie, Z. Qin, G. Y. Li, and B.-H. Juang, "Deep learning enabled semantic communication systems," *IEEE Trans. Signal Process.*, vol. 69, p. 2663, 2021.

[22] X. Mu and Y. Liu, "Exploiting semantic communication for non-orthogonal multiple access," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 8, p. 2563, Aug. 2023.

[23] X. Mu, Y. Liu, L. Guo, and N. Al-Dhahir, "Heterogeneous semantic and bit communications: A semi-NOMA scheme," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 1, p. 155, Jan. 2023.

[24] X. Xie, F. Fang, and L. Zhang, et al., "Power-efficient optimization for coexisting semantic and bit-based users in NOMA networks," *IEEE Trans. Commun.*, vol. 73, no. 12, p. 14183, Dec. 2025.

[25] K. M. K. Ramamoorthy, W. Wang, and K. Sohraby, "A power auction approach for non-orthogonal multiple access wireless relay communications," in *Proc. IEEE ICC Workshops*, 2023.

[26] J. Duan, K. M. K. Ramamoorthy, W. Wang, and Y. Zhao, "When semantics meet strategy: Stackelberg game-theoretic power trading in NOMA wireless networks," submitted to *IEEE Trans. Cogn. Commun. Netw.*, 2026.

[27] Z. Wei, J. Jia, Y. Niu, L. Wang, H. Wu, H. Yang, and Z. Feng, "Integrated sensing and communication channel modeling: A survey," arXiv:2404.17462, Apr. 2024.

[28] A. Bal, H. Cai, H. Guo, Y. Zheng, and X. Zhang, "Enabling joint sensing and communication via STBC assisted NOMA in ISAC systems," in *Proc. IEEE 22nd Int. Conf. Mobile Ad-Hoc Smart Syst. (MASS)*, 2025, p. 407, doi: 10.1109/MASS66014.2025.0064.

[29] H. Ding and K.-C. Leung, "Resource allocation for low-latency NOMA-V2X networks using reinforcement learning," in *Proc. IEEE INFOCOM Workshops (GI)*, 2021, doi: 10.1109/INFOCOMWK-SHPS51825.2021.9484529.

[30] J. Deng, X. Tang, X. Wei, P. Li, J. Li, X. Li, and Z. Ghassemlooy, "Power allocation for NOMA-based visible light communication systems with DQN," in *Proc. 14th Int. Symp. Commun. Syst. Netw. Digit. Signal Process. (CSNDSP)*, 2024, p. 512, doi: 10.1109/CSNDSP60683.2024.10636475.

[31] Y. Cui, W. Yuan, Z. Zhang, J. Mu, and X. Li, "On the physical layer of digital twin: An integrated sensing and communications perspective," *IEEE J. Sel. Areas Commun.*, vol. 41, no. 11, p. 3474, Nov. 2023, doi: 10.1109/JSAC.2023.3314826.

[32] V. C. Andrei, P. Susarla, A. Djuhera, N. Vaara, J. Mustaniemi, C. Álvarez Casado, X. Li, U. J. Mönich, H. Boche, and M. Bordallo López, "Demonstration of a continuously updated, radio-compatible digital twin for robotic integrated sensing and communications," in *Proc. IEEE Int. Symp. Joint Commun. Sensing (JC&S)*, 2025, pp. 1–2, doi: 10.1109/JCS64661.2025.10880640.

[33] J. Hoydis, S. Cammerer, F. Ait Aoudia, et al., "Sionna: An open-source library for next-generation physical layer research," arXiv:2203.11854, 2022.

[34] Metrica Sports, "Sample tracking and event data," 2020. [Online]. Available: https://github.com/metrica-sports/sample-data