Multifaceted UAV Fleet Automation via Asynchronous Deep RL

Bharath Keshavamurthy   Dr. Nicolò Michelusi

Ira A. Fulton Schools of Engineering | Arizona State University | Tempe, AZ

System Deployment Model | Policy Computation Algorithmic Control Flow

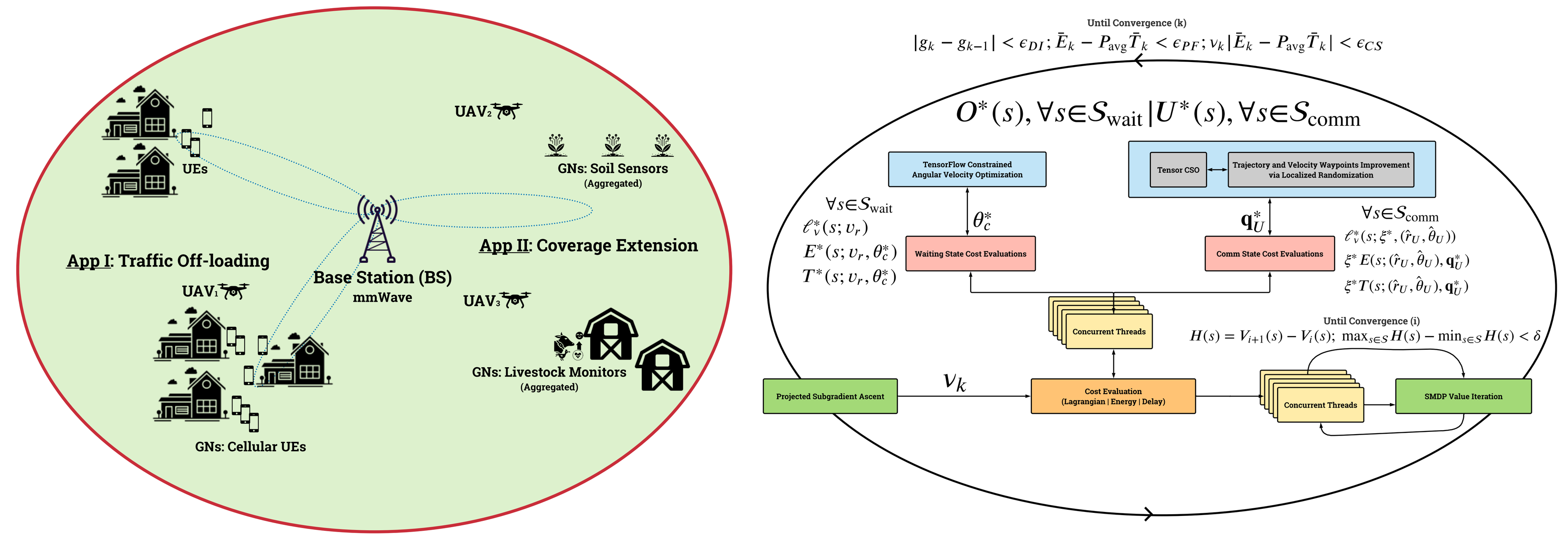

Figure 1. The deployment architecture under analysis: A cellular Base Station (BS) at the center of a circular cell of radius a meters, with multiple UAVs serving as BS relays for traffic off-loading and coverage extension (a); The algorithmic flowchart of our Semi-Markov Decision Process and Hierarchical Competitive Swarm Optimization framework to solve the **single-agent two-stage scheduling and trajectory optimization problem** (b).

Single-Agent Optimal Policy Computation: Mathematical Construction

- UAV Mobility Power Model: $P_{\text{mob}}(V) = P_1\left(1+\frac{3V^2}{U_{\text{tip}}^2}\right) + P_2\left(\sqrt{1+\frac{V^4}{4v_0^4}} - \frac{V^2}{2v_0^2}\right)^{1/2} + P_3V^3$
- A2G Channel Model (with Rate Adaptation):

$$P_{\text{LoS}}(\varphi) = \frac{1}{1+z_1\cdot\exp\left(-z_2\left[\varphi - z_1\right]\right)},\quad P_{\text{NLoS}}(\varphi) = 1 - P_{\text{LoS}}(\varphi); C(h) = B\cdot\log_2\left(1+\frac{|h|^2P}{\sigma^2\Gamma}\right)$$

$$P_{\text{out}}(\Upsilon,\beta,K) = 1 - Q_1\left(\sqrt{2K}, \sqrt{2(K+1)u(\Upsilon,\beta)}\right); u(\Upsilon,\beta) \triangleq \sigma^2\Gamma(2^{\Upsilon/B}-1)/(\beta P)$$

$$R(\Upsilon,\beta,K) = \Upsilon\cdot(1-P_{\text{out}}(\Upsilon,\beta,K)) = \Upsilon\cdot Q_1\left(\sqrt{2K}, \sqrt{2(K+1)u(\Upsilon,\beta)}\right)$$

With $Z \triangleq \sqrt{\frac{2\beta P}{\sigma^2\Gamma}u(\Upsilon,\beta)},\ \Upsilon = B\log_2\left(1+\frac{1}{2}Z^2\right) \triangleq f(Z)$ and $\Upsilon^*(\beta,K) = f(Z^*(\beta,K))$

$Z^*(\beta,K) \triangleq \arg\min_{Z\geq 0} -\ln f(Z) - \ln Q_1\left(\sqrt{2K}, \sqrt{\frac{(K+1)\sigma^2\Gamma}{\beta P}Z}\right)$ [Bisection Method]

$$\bar{R}(d,\varphi) \triangleq P_{\text{LoS}}(\varphi)\cdot R^*(\beta_{\text{LoS}}(d), K(\varphi)) + P_{\text{NLoS}}(\varphi)\cdot R^*(\beta_{\text{NLoS}}(d), 0)$$

Two-Stage Decision Framework | A Semi-Markov Decision Process (SMDP) formulation

- Waiting states ($s\in\mathcal{S}_{\text{wait}}$):
  - Outer decision | Optimal UAV radial velocity (v_r^*) | **SMDP Value Iteration**
  - Inner decision | Optimal UAV angular velocity (θ_c^*) | **Solve**

$$\ell_\nu^*(s;v_r) = \min_{\theta_c} \nu\left(P_{\text{mob}}\left(\sqrt{v_r^2 + r_U^2\cdot\theta_c^2}\right) - P_{\text{avg}}\right)\Delta_0 \text{ s.t. } \sqrt{v_r^2+r_U^2\cdot\theta_c^2} \leq V_{\max}.$$

- Communication states ($s\in\mathcal{S}_{\text{comm}}$):
  - Outer decision | Optimal UAV end position $[(\hat{r}_U^*, \hat{\theta}_U^*)]$ | **SMDP Value Iteration**
  - Inner decision I | Optimal scheduling strategy (ξ^*) | **Pick the minimizing action between**

$$\ell_\nu^*(s;r_U,0) = L/\bar{R}_{GB}(r) \text{ and } \ell_\nu^*(s;\hat{r}_U,1) = \min_{\Delta,\mathbf{q}_U,t_p} (1-\nu P_{\text{avg}})\Delta + \nu\int_0^\Delta P_{\text{mob}}\left(\sqrt{r_U'(\eta)^2 + r_U^2(\eta)\cdot\theta_U'(\eta)^2}\right)\mathrm{d}\eta$$

s.t. data payload, max velocity, and start & end position constraints

  - Inner decision II | Optimal trajectory | **Hierarchical Competitive Swarm Optimization (HCSO)**

$$\mathbf{p}^*,\mathbf{v}^* = \underset{\mathbf{p},\mathbf{v}\in[V_{\text{low}},V_{\max}]^M}{\operatorname{argmin}} (1-\nu P_{\text{avg}})\sum_{m=0}^{M-1}\frac{\|\mathbf{x}_{m+1}-\mathbf{x}_m\|_2}{v_m} + \nu\sum_{m=0}^{M-1}\frac{\|\mathbf{x}_{m+1}-\mathbf{x}_m\|_2}{v_m}P_{\text{mob}}(v_m)$$

s.t. data payload, max velocity, and start & end position constraints

Scenario Evaluation: Optimal SMDP-HCSO Policy + Multi-Agent Heuristics

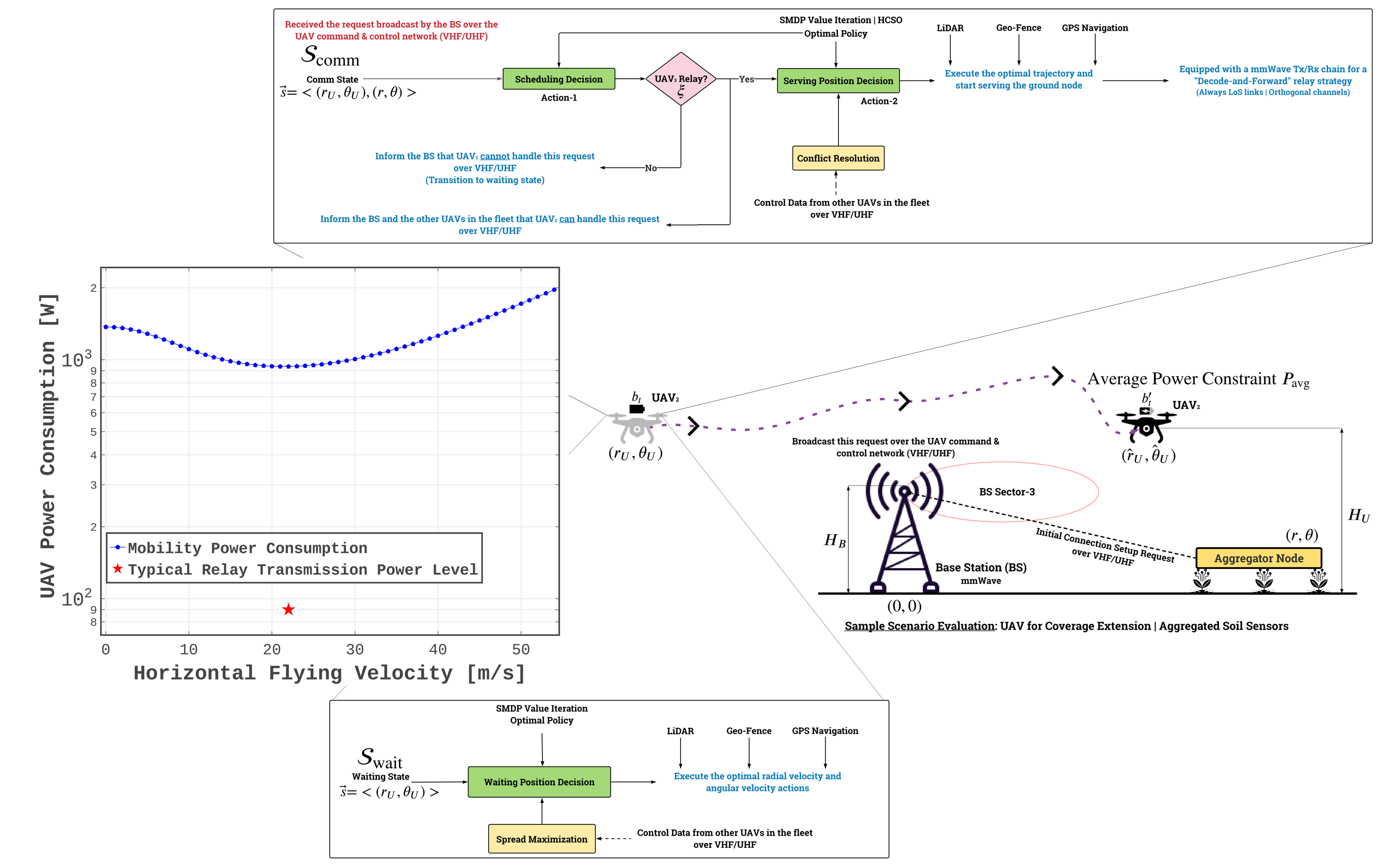

Figure 2. The multi-agent evaluation of optimal policy execution in smart agriculture from the perspective of UAV_2: coverage extension for soil sensors as Ground Nodes (via an aggregator node).

Visualizations: Delay-Power Trade-offs | Optimal Trajectories

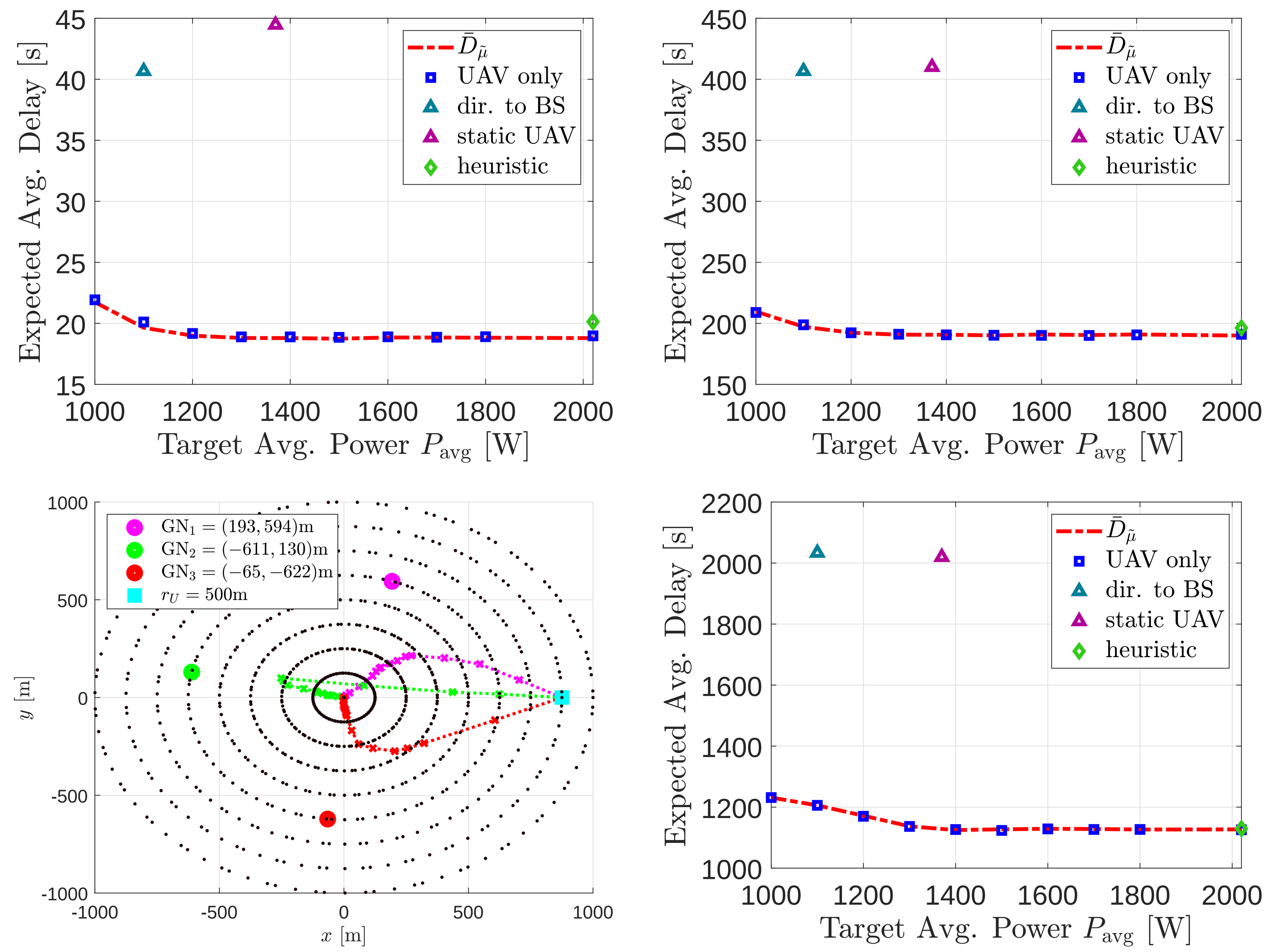

Figure 3. (Clockwise from top-left) The expected service delay versus the target average power constraint for varying packet lengths: 0.1Mb; 1.0Mb; 5.0Mb; The optimal UAV trajectories under 0.1Mb | P_{avg}=1.1kW.

[M. Bliss and N. Michelusi, "Power-Constrained Trajectory Optimization for Wireless UAV Relays with Random Requests," IEEE ICC 2020]

Asynchronous Actor-Critic in a Double Deep-Q Network Framework

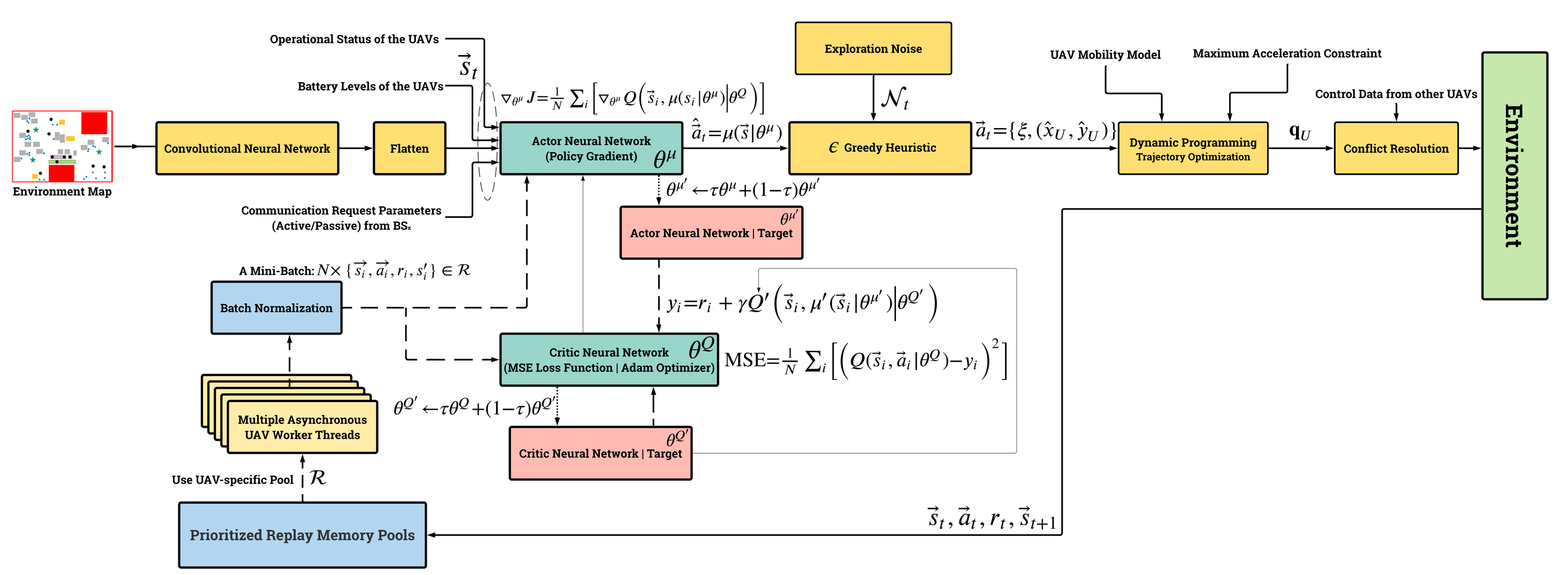

Figure 4. A different approach to UAV fleet automation: an asynchronous Actor-Critic structure with a Deep Deterministic Policy Gradient (DDPG) in the Actor and a Mean Square Error (MSE) driven Adam optimizer in the Critic—coupled with target networks, thread-specific hyper-parameter training, thread-specific prioritized experiential replay memory pools, and exploration noise (via an ϵ−greedy heuristic). Equipped with advanced UAV power & mobility models, additional constraints such as maximum UAV acceleration are imposed in the trajectory optimization paradigm—solved via Dynamic Programming. *[T. P. Lillicrap, et al., "Continuous Control with Deep RL," ICLR 2016]*

Companion Research: NSF POWDER Millimeter-Wave Propagation Modeling

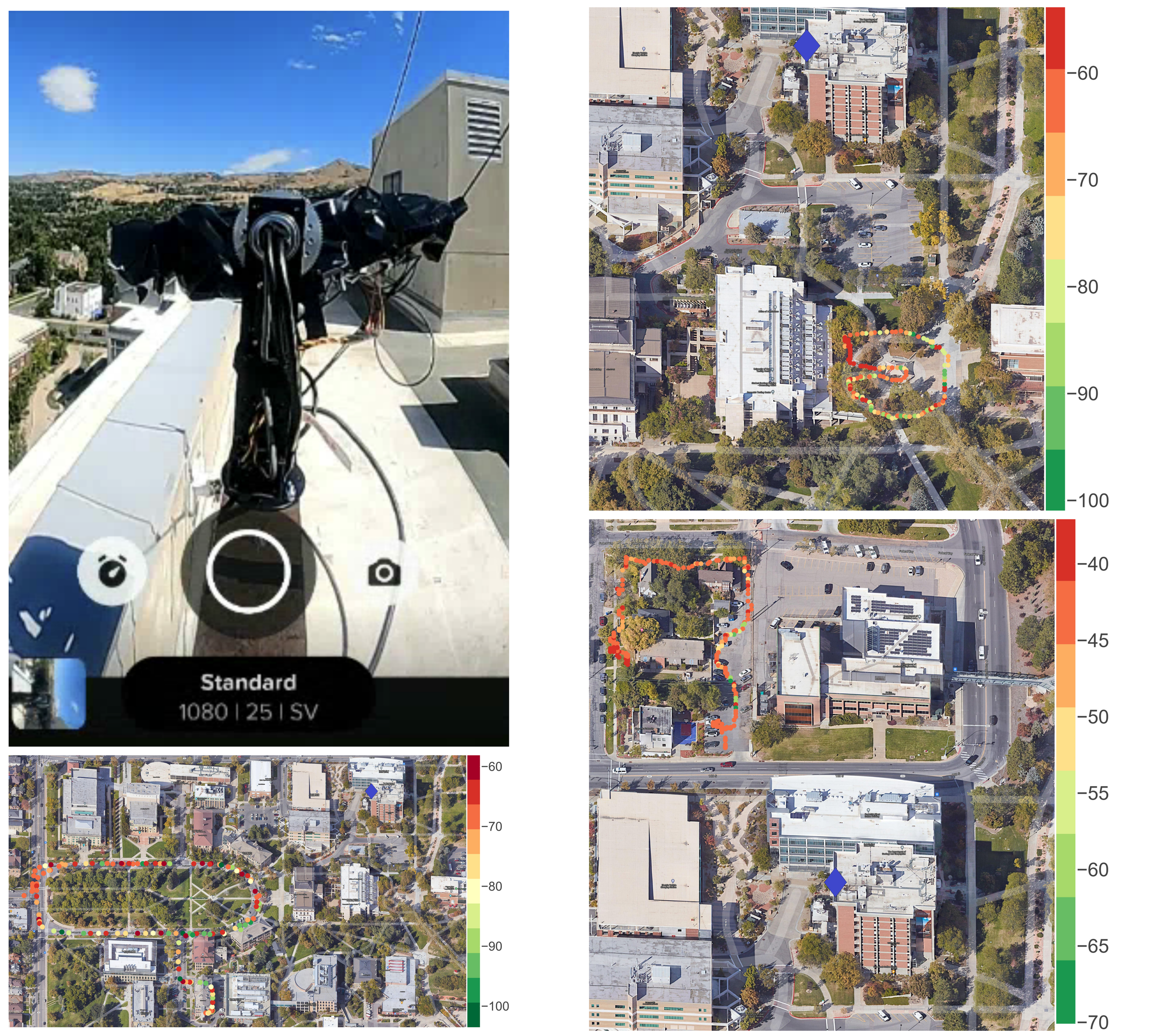

Figure 5. (Clockwise from top-left) Our remote monitoring & troubleshooting interface—via an Android Debug Bridge—exhibiting the deployment of our Tx on the roof-top of the William Browning Building; The illustrations of the received signal power values superimposed on a Google Hybrid map of the sites under analysis.

[B. Keshavamurthy, et. al., "A Robotic Antenna Alignment and Tracking System for Millimeter Wave Propagation Modeling," USNC-URSI NRSM 2022]

Funded by NSF under grants CNS-1642982 and CNS-2129015

NSF AERPAW Poster Presentation 2021, Raleigh — NC 27606

github.com/bharathkeshavamurthy/Odin