GNN-Parametrized Diffusion Policies for Wireless Resource Allocation

Yiğit Berkay Uslu, **Samar Hadou**, Shirin Saeedi Bidokhti and Alejandro Ribeiro

NEURAL INFORMATION PROCESSING SYSTEMS

Penn, University of Pennsylvania

- **Goal:** Allocate network resources *optimally* for any given network state $\mathbf{H}$:

$$\mathrm{P}^*(\mathbf{H}) \;=\; \underset{\mathcal{D}_{\mathbf{x}}(\mathbf{H})}{\text{maximum}} \;\; \mathbb{E}_{\mathcal{D}_{\mathbf{x}}(\mathbf{H})}\left[f_0\big(\mathbf{x}(\mathbf{H}),\mathbf{H}\big)\right], \qquad \text{subject to } \;\mathbb{E}_{\mathcal{D}_{\mathbf{x}}(\mathbf{H})}\left[\mathbf{f}\big(\mathbf{x}(\mathbf{H}),\mathbf{H}\big)\right]. \tag{1}$$

$\Rightarrow$ $\mathcal{D}^*_{\mathbf{x}}(\mathbf{H})$ maximizes an *expected utility* while satisfying *expected requirements*.

$\Rightarrow$ QoS-optimality via *time-sharing* $\frac{1}{T}\sum_{\tau=1}^{T} f_0\big(\mathbf{x}_\tau(\mathbf{H}),\mathbf{H}\big) \approx \mathbb{E}_{\mathcal{D}_{\mathbf{x}}}\left[f_0\big(\mathbf{x}(\mathbf{H}),\mathbf{H}\big)\right]$.

- **Challenge:** We *cannot solve* directly for the *optimal distributions* $\mathcal{D}^*_{\mathbf{x}}(\mathbf{H})\mathcal{D}_{\mathbf{H}}$.
- **Solution:** Learn a **generative model of resource allocations** $\mathcal{D}_{\mathbf{x}}(\mathbf{H};\theta^*) \approx \mathcal{D}^*_{\mathbf{x}}(\mathbf{H})$.

$\Rightarrow$ Train a **conditional diffusion model policy** $\mathcal{D}^*_{\mathbf{x}}(\mathbf{H};\theta)$ to imitate the experts $\mathcal{D}^*_{\mathbf{x}}(\mathbf{H})$.

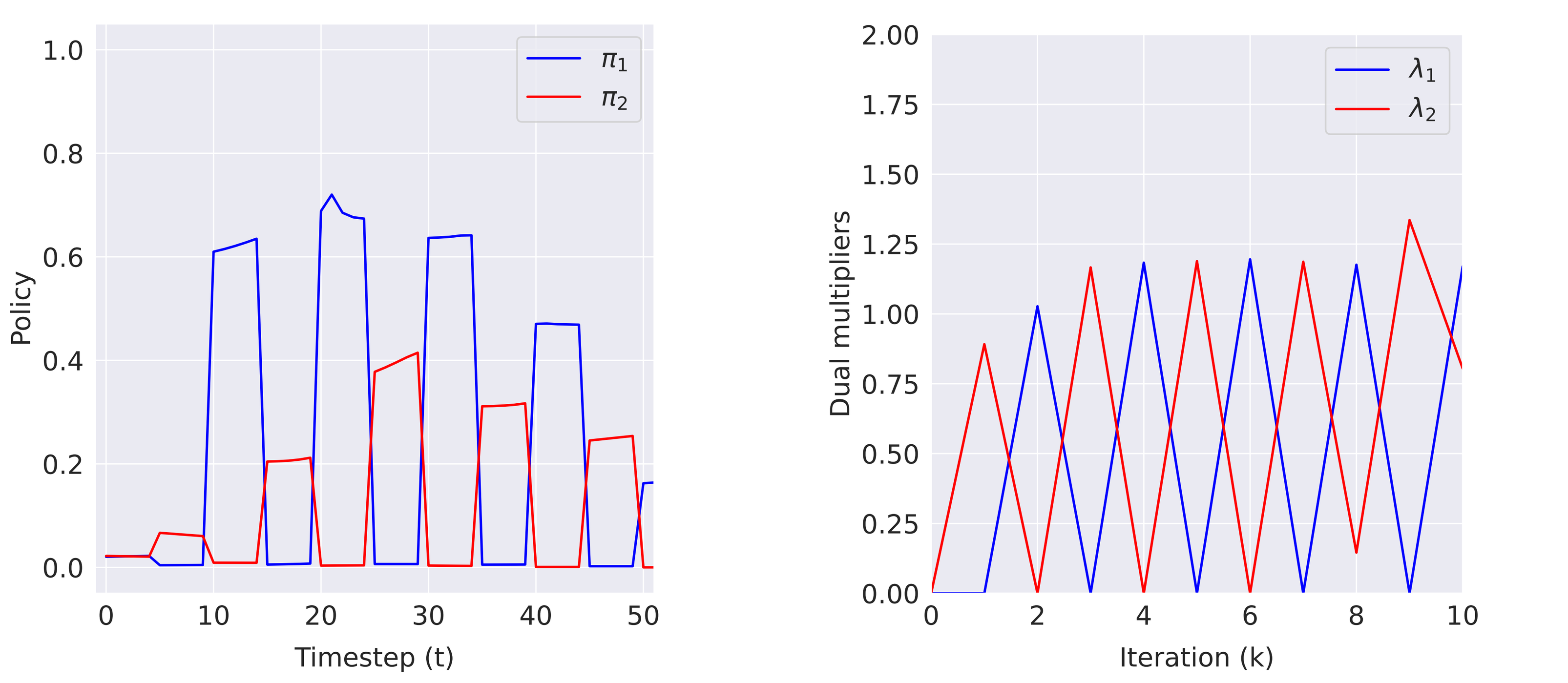
$\Rightarrow$ Utilize a **graph neural network (GNN) backbone for the diffusion model** to operate directly on graphs $\mathbf{H}$ and enable **learning families of policies** across $\mathcal{D}_{\mathbf{H}}$.

Imitation Learning of Stochastic Resource Allocation Policies

- A generative model learns to imitate an expert policy over a family of networks.

$$\mathcal{D}^*_{\mathbf{x}}(\mathbf{H};\theta) \;=\; \underset{\mathcal{D}_{\mathbf{x}}(\mathbf{H};\theta)}{\operatorname{argmin}} \;\mathbb{E}_{\mathcal{D}_{\mathbf{H}}}\Big[\mathrm{D}_{\mathrm{KL}}\big(\mathcal{D}^*_{\mathbf{x}}(\mathbf{H}) \,\big\|\, \mathcal{D}_{\mathbf{x}}(\mathbf{H};\theta)\big)\Big]. \tag{2}$$

- We leverage a (state-augmented) primal-dual algorithm as an expert policy that
  - $\Rightarrow$ generates a trajectory of optimal primal and dual variables $\big(\mathbf{x}_\tau(\mathbf{H}),\boldsymbol{\lambda}_\tau(\mathbf{H})\big)_{\tau\geq 1}^{\infty}$.
  - $\Rightarrow$ maintains a.s.-feasibility and near-optimality by policy randomization.
  - $\Rightarrow$ trades off objective optimality for fast transient dynamics.



- We collect an expert dataset $\left\{\mathbf{x}_1(\mathbf{H}^{(1)}),\ldots,\mathbf{x}_T(\mathbf{H}^{(1)}),\mathbf{x}_1(\mathbf{H}^{(2)}),\ldots,\mathbf{x}_T(\mathbf{H}^{(M)})\right\}$ of optimal solutions to (1) via (stationary) state-augmented dual descent roll-outs.
- We train a GDM policy to minimize (3) on the expert dataset.
- The trained GDM policy, parametrized by a GNN, generalizes to $\mathcal{D}_{\mathbf{H}}$.

GNN-Parametrized Generative Diffusion Model Policies

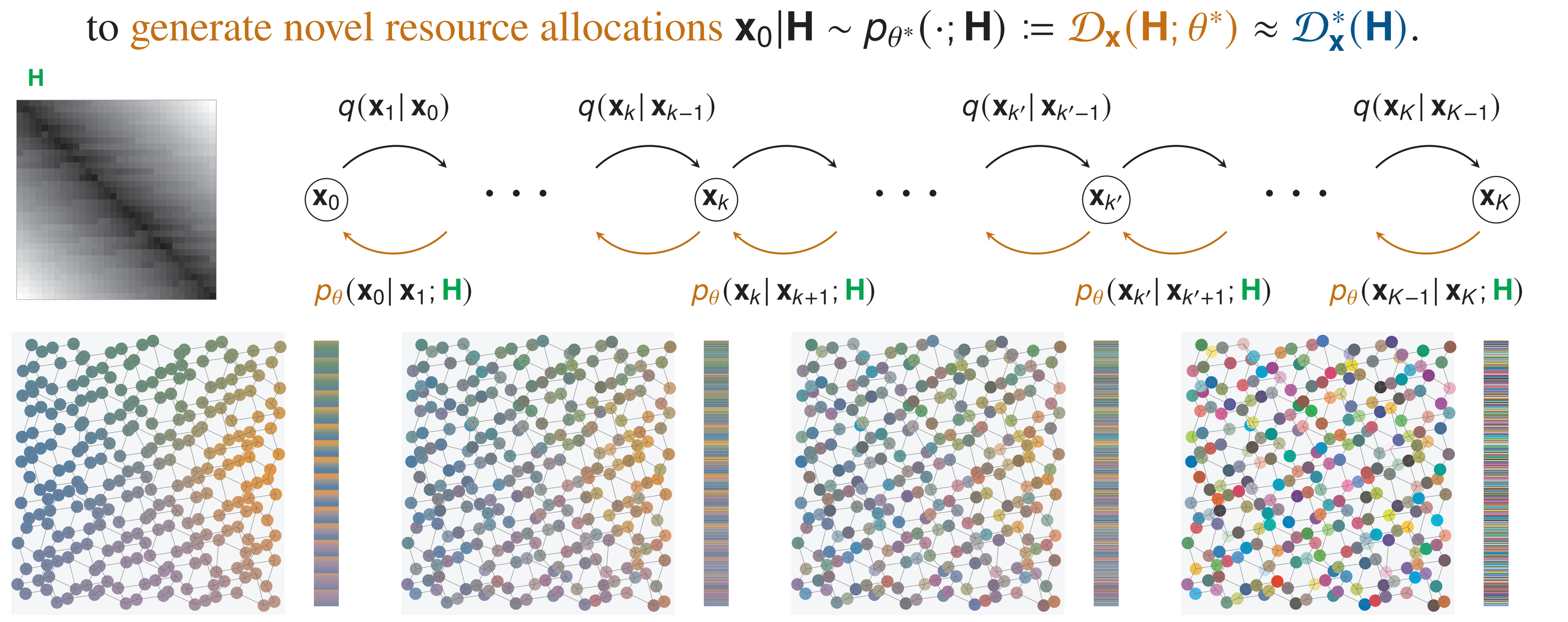
- Diffusion models learn a denoising chain that reverses a forward noising chain.
- We parametrize the reverse chain p_θ and learn a parametric denoiser $\boldsymbol{\epsilon}_{\theta^*}$,

$$\theta^* \in \underset{\theta}{\operatorname{argmin}}\; \mathcal{L}(\theta) := \mathbb{E}_{\mathbf{x}_0,\mathbf{H},k,\boldsymbol{\epsilon}}\left\|\boldsymbol{\epsilon}_\theta\big(\mathbf{x}_k(\mathbf{x}_0,\boldsymbol{\epsilon}),k;\mathbf{H}\big)-\boldsymbol{\epsilon}\right\|^2. \tag{3}$$

- We iterate the learned reverse chain $p_{\theta^*}(\mathbf{x}_{k-1}|\mathbf{x}_k;\mathbf{H})$ for $k=K,\ldots,1$, by updating

$$\mathbf{x}_{k-1} = \frac{1}{\sqrt{\alpha_k}}\left(\mathbf{x}_k - \frac{\beta_k}{\sqrt{1-\bar{\alpha}_k}}\boldsymbol{\epsilon}_{\theta^*}(\mathbf{x}_k,k;\mathbf{H})\right) + \sigma_k\mathbf{w}, \quad \mathbf{x}_K,\mathbf{w}\sim\mathcal{N}(\mathbf{0},\mathbf{I}), \tag{4}$$

to generate novel resource allocations $\mathbf{x}_0|\mathbf{H}\sim p_{\theta^*}(\cdot;\mathbf{H}) := \mathcal{D}_{\mathbf{x}}(\mathbf{H};\theta^*) \approx \mathcal{D}^*_{\mathbf{x}}(\mathbf{H})$.



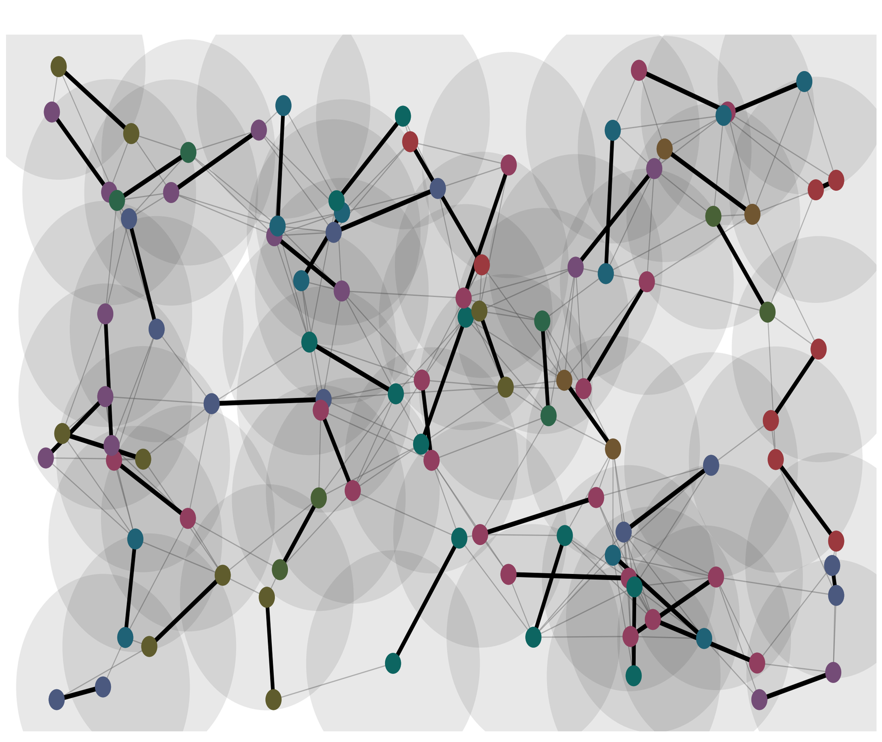
- We parametrize the denoiser $\boldsymbol{\epsilon}_\theta$ by a GNN architecture that cascades L graph convolutional layers with read-in $(\mathbf{x}_k,k)\mapsto\mathbf{Z}_0$ and read-out $\mathbf{Z}_L\mapsto\boldsymbol{\epsilon}_{\theta^*}$ layers:

$$\mathbf{Z}^{(\ell)} = \boldsymbol{\Psi}^{(\ell)}\left(\mathbf{Z}^{(\ell-1)};\mathbf{H},\Theta^{(\ell)}\right) = \varphi\left[\sum_{k=0}^{K}\mathbf{H}^k\mathbf{Z}^{(\ell-1)}\boldsymbol{\Theta}_k^{(\ell)}\right], \quad \ell=1,\ldots,L. \tag{5}$$

$\Rightarrow$ A graph signal generative model conditioned on input graphs $\mathbf{H}$ (via GSOs).

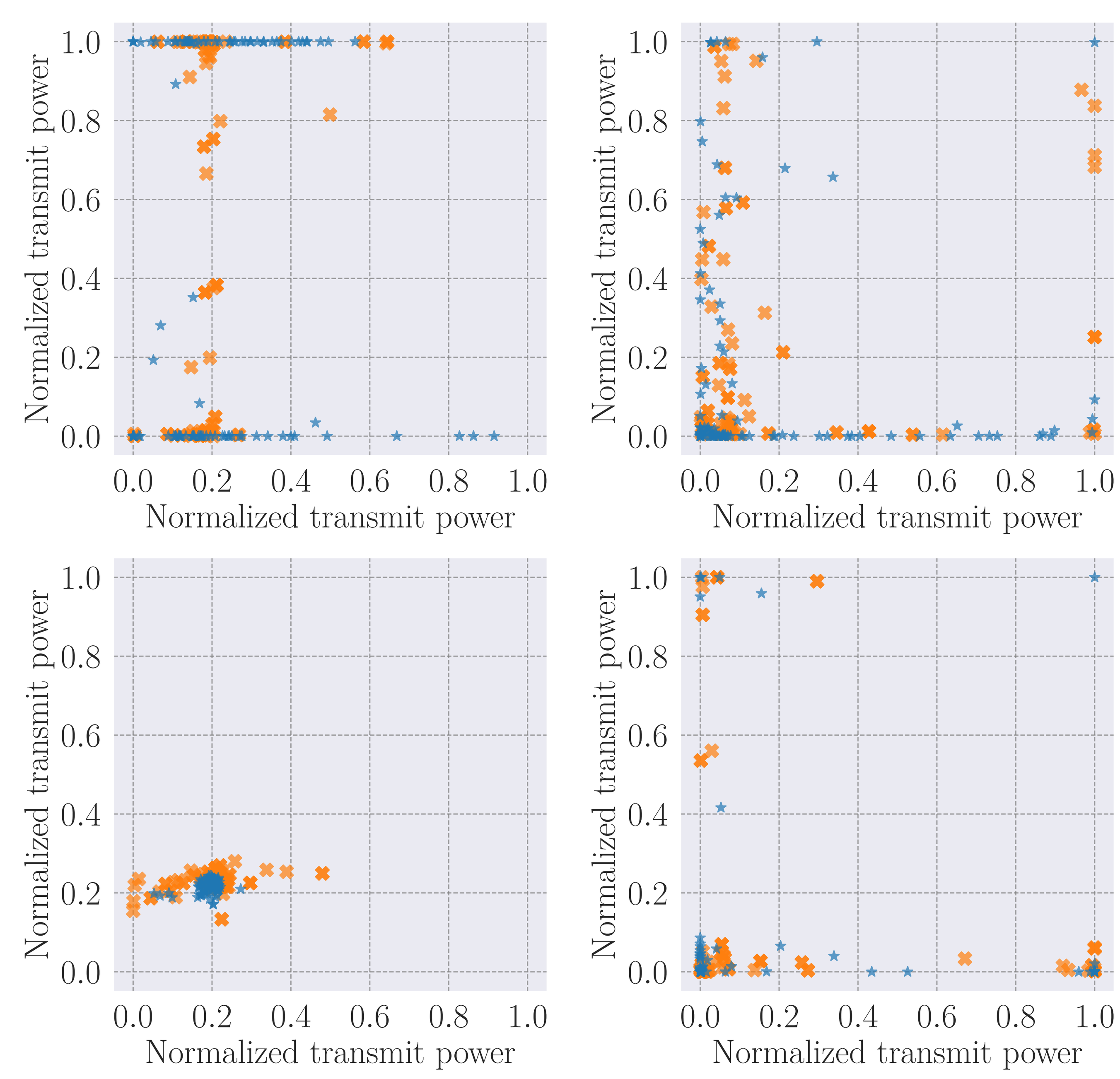
Numerical Results: Power Control

- A wireless (Rayleigh) fading channel with transmitter-receiver (tx-rx) pairs as users (nodes).
- Network state (GSO) $\mathbf{H}$ represented by the set of constant long-term channel gains.
- Actual channel gain from tx i to rx j at time τ fluctuates as $h_{ij,\tau}\sim\mathcal{D}_{\tilde{\mathbf{H}}|\mathbf{H}}(\mathbf{H})$ due to small-scale fading.
- Tx i allocates power $x_{i,\tau}\geq 0$ at time τ and causes interference to neighboring tx-rx pairs $j\in\mathcal{N}(i)$.
- Communication rate r determined by SINR at each rx j. $\Rightarrow$ $\mathrm{SINR}_{j,\tau} = \dfrac{h_{jj,\tau}\cdot x_{j,\tau}}{1+\sum_{i\in\mathcal{N}(j)}h_{ij,\tau}\cdot x_{i,\tau}}$.



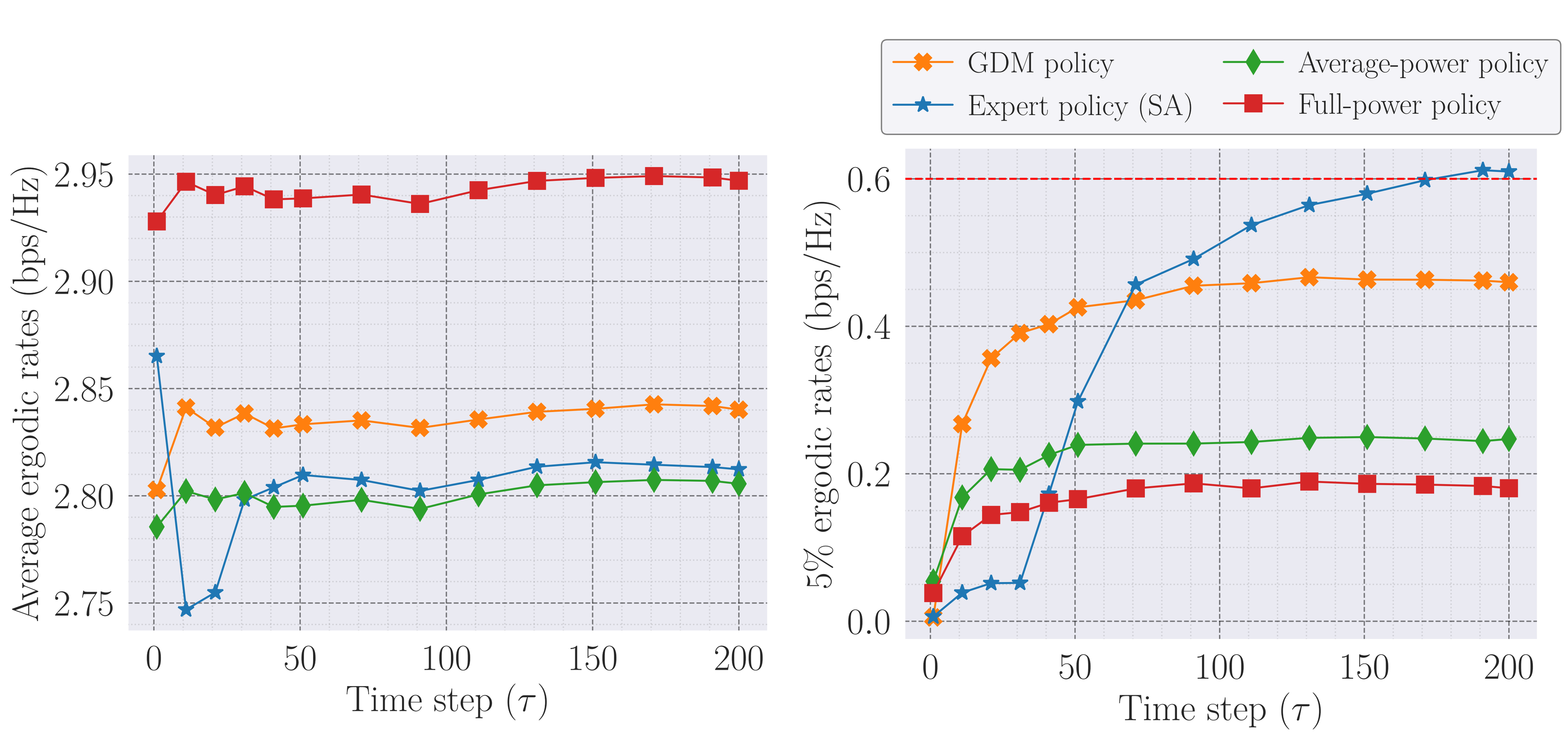
- Given $\mathbf{H}\sim\mathcal{D}_{\mathbf{H}}$, we want to allocate transmit powers $\mathbf{x}_\tau\sim\mathcal{D}_{\mathbf{x}}(\mathbf{H})$ over T time steps to maximize ergodic network-wide sum-rate, subject to minimum-rate requirements and a max. transmit power budget $x_{\max}$:

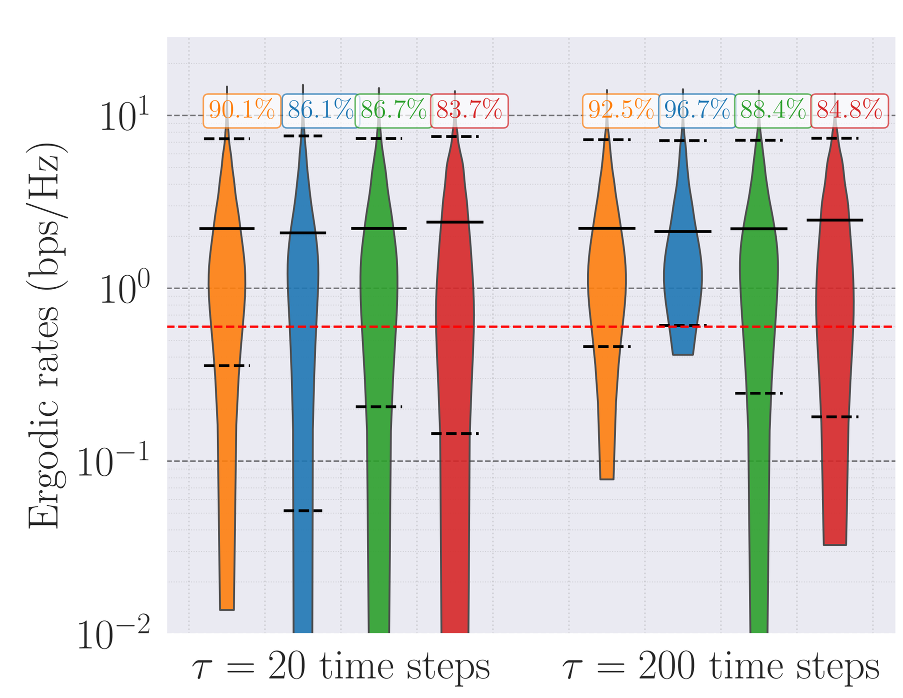
$$\begin{aligned}\mathrm{P}^*(\mathbf{H}) \;=\; &\max_{\mathcal{D}_{\mathbf{x}}(\mathbf{H})}\;\frac{1}{T}\sum_{\tau=1}^{T}\sum_j r(\mathrm{SINR}_{j,\tau})\\ &\text{s.t.}\quad \frac{1}{T}\sum_{\tau=1}^{T} r(\mathrm{SINR}_{j,\tau}) \;\geq\; r_{\min},\;\forall j,\\ &\text{s.t.}\quad 0\leq \mathbf{x}_{j,\tau}\leq x_{\max},\;\forall j,\tau=1,\ldots,T.\end{aligned}$$



- GDM policy samples optimal power control policies that tend to be probabilistic and involve multiple transmission modes.

- GDM policy $\mathbf{x}_\tau\sim\mathcal{D}_{\mathbf{x}}(\mathbf{H};\theta^*)$ achieve ergodic utility and requirement QoS close to the expert policy.
  - $\Rightarrow$ Deterministic baselines fail under challenging channel conditions.





- GDM policy bypasses suboptimal transients and samples from stationary dual descent dynamics.

Email: ybuslu@seas.upenn.edu