Managing Conflicts Among Black-Box RAN Apps via Multi-Fidelity Game-Theoretic Optimization

Abstract

Modern open and softwarized networks—such as O-RAN telecom systems—host independently developed applications, known as xApps, with distinct and potentially conflicting objectives. Coordinating their behavior to ensure stable operation is challenging, especially when each application's utility is only accessible via costly black-box evaluations. In this work, we consider a centralized controller that suggests joint configurations to multiple apps, modeled as strategic players, with the goal of aligning their incentives toward a stable outcome. This setting is modeled as a Stackelberg game in which the central controller (leader) lacks analytical forms of the players' utility functions, and must learn them through sequential, multi-fidelity queries. We propose MF-UCB-PNE, a novel multi-fidelity Bayesian optimization strategy that efficiently approximates a pure Nash equilibrium (PNE) under a limited query budget. MF-UCB-PNE balances exploration of cheap, lower-fidelity utility observations with exploitation of high-fidelity evaluations, enabling convergence to incentive-compatible configurations. We provide theoretical guarantees in terms of no-regret learning of equilibrium, as well as empirical results on wireless networking problems, demonstrating that MF-UCB-PNE identifies high-quality equilibrium solutions under strict cost budgets.

1 Introduction

1.1 Context and Contribution

Open and software-defined network architectures allow multiple independent applications to control shared resources. These applications, or apps for short, often have *conflicting* objectives, which can lead to unstable behavior. For instance, in an O-RAN telecom network, specialized applications, known as xApps, run on a centralized controller, each generally optimizing a different goal [1]. For example, as seen in Fig. 1, one xApp might increase transmit power to maximize throughput, while another may try to shut down radios to save energy, causing a conflict over the use of shared resources.

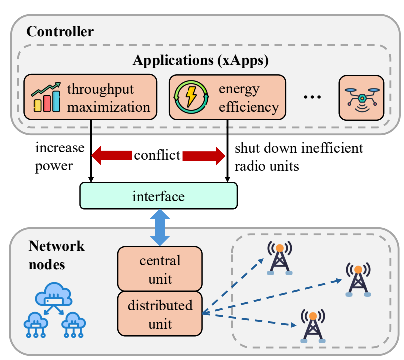

Figure 1: **O-RAN scenario with two conflicting xApps.** Using the methods introduced in this paper, a joint configuration can be chosen by the controller so that neither app has incentive to deviate unilaterally.

As illustrated in Fig. 1, the central controller can mitigate such conflicts by recommending joint configurations for all apps, aiming for a working point where no app has incentive to deviate from it on its own. We model this interaction as a hierarchical game in which the central controller plays the role of a *leader* that proposes a configuration profile, and the N apps are *followers* that respond according to their individual utility functions. A stable operating point corresponds to a *pure Nash equilibrium* (PNE) of this game, i.e., a joint configuration of actions from which no single player can unilaterally improve its utility. The controller's goal is to find a PNE, so that its recommendation will be self-enforcing.

However, a key challenge is that each player's utility function is a black-box from the perspective of the controller: it is not given in closed form and can only be observed by running experiments or simulations. Accordingly, each query can be *expensive* in time or resources. To address these challenges, in this work we consider the *multi-fidelity* learning setting illustrated in Fig. 2. In it, the controller can observe utility values at various fidelity levels. Low-fidelity evaluations, e.g., coarse simulations or partial measurements, are cheaper but less accurate, while high-fidelity evaluations, e.g., detailed simulations or field experiments, provide accurate feedback at higher cost. By judiciously using low- and high-fidelity queries, the controller can save budget while still homing in on a stable configuration.

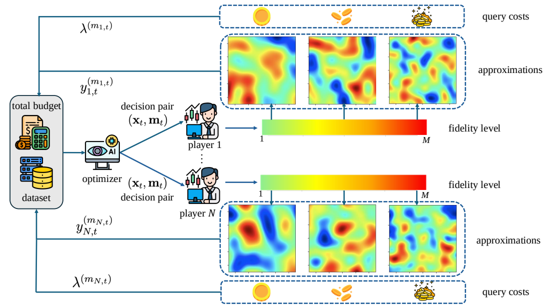

Figure 2: **Multi-fidelity Stackelberg game setup.** A central optimizer aims at identifying a joint configuration for the apps that constitute an approximate Nash equilibrium for the utilities of the apps, while having access only to multi-fidelity estimators of expensive-to-evaluate black-box utility functions.

In this context, we introduce **MF-UCB-PNE**, a multi-fidelity Bayesian optimization algorithm to efficiently learn players' utilities and approximate a PNE under a total query cost constraint. We show that this approach leads to provably no-regret convergence to an approximate equilibrium as the budget grows. We also demonstrate the effectiveness of MF-UCB-PNE on two wireless network case studies (power control and random access), where it outperforms single-fidelity baselines in finding high-quality equilibria within limited budgets.

1.2 Related Work

Single-objective black-box optimization. *Bayesian optimization* (BO) [2] provides a sample-efficient sequential decision-making framework that is widely adopted in expensive-to-evaluate black-box optimization problems. Uncertainty-aware acquisition functions include upper confidence bound (UCB) [3], expected improvement (EI) [4], and entropy search (ES) [5].

Multi-fidelity single-objective black-box optimization. In order to reduce the query cost incurred in the acquisition process, multi-fidelity BO (MFBO) constructs multi-fidelity surrogate models and define the cost-aware acquisition functions to adaptively prioritize queries across fidelity levels [6, 7].

Nash equilibrium (NE) evaluation with unknown utility functions. Reference [8] characterized the query complexity associated with evaluating a NE for unknown utilities across various game classes.Subsequent works in [9, 10] proposed a BO-based optimizer that can approximate NE solutions with formal regret guarantees. However, these methods assume access to a single fidelity of utility feedback and do not explore the potential benefits of multi-fidelity observations, which can offer more cost-effective and informative sampling strategies. A longer version of this work can be found in the unpublished manuscript [11].

2 Problem Formulation

Strategic game model. We consider N players, representing apps in a softwarized wireless system, with app n characterized by a configuration setting x_n, also referred to as action, from its strategy set $\mathcal{X}_n$. For example, for a random access app, the configuration x_n may represent the transmission probability or a back-off parameter. Let $\mathbf{x} = (x_1, \ldots, x_N) \in \mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$ denote the joint *action profile*.

The utility for player n is an unknown function $u_n(\mathbf{x})$, which depends on all players' actions. We assume the utilities $u_n(\mathbf{x})$ to be *black-box* functions: the controller cannot directly observe $u_n(\mathbf{x})$ except by querying the outcome of action profile $\mathbf{x}$. This is a typical situation in wireless systems in which *key performance indicators* (KPIs) can be only measured based on simulations or real-world developments.

A recommended action profile $\mathbf{x}$ is considered stable if no player can significantly increase its utility by deviating. Accordingly, the *dissatisfaction* of player n at $\mathbf{x}$ is defined as

$$f_n(\mathbf{x}) = \max_{x'_n \in \mathcal{X}_n} u_n(x'_n, \mathbf{x}_{-n}) - u_n(\mathbf{x}) \geq 0, \tag{1}$$

where $\mathbf{x}_{-n}$ denotes the actions of all players except n. The function $f_n(\mathbf{x})$ can be interpreted as quantifying the incentive for player n to unilaterally deviate from $\mathbf{x}$. A joint action $\mathbf{x}^*$ is a *pure Nash equilibrium (PNE)* if $f_n(\mathbf{x}^*) = 0$ for all players n—i.e., no player can improve its utility by deviating.

In many games a PNE may not exist or may be hard to find, so we consider approximate equilibria. For $\epsilon \geq 0$, an ϵ-*PNE* is an action profile $\mathbf{x}^{(\epsilon)}$ such that the dissatisfaction levels satisfy the inequalities $f_n(\mathbf{x}^{(\epsilon)}) \leq \epsilon$ for all n. In an ϵ-PNE, no player can gain more than ϵ by unilaterally deviating, so ϵ quantifies the stability of the configuration. Let ϵ^* be the smallest achievable dissatisfaction level

$$\epsilon^* = \inf\{\epsilon \in \mathbb{R} | \mathcal{X}^{(\epsilon)} \neq \emptyset\}, \tag{2}$$

where

$$\mathcal{X}^{(\epsilon)} := \{\mathbf{x}^{(\epsilon)} \in \mathcal{X} | f_n(\mathbf{x}^{(\epsilon)}) \leq \epsilon \text{ for all } n\} \tag{3}$$

is the corresponding ϵ-PNE set. Our goal is to identify an action profile $\mathbf{x}^*$ that lies in the ϵ^*-PNE set, i.e., an approximately stable joint configuration with minimum possible ϵ^*.

Multi-fidelity utility observations. We assume the controller can sequentially query the utilities by proposing action profiles and observing the resulting payoffs with chosen fidelity. Specifically, at each time step $t = 1, 2, \ldots$, the controller selects a joint action $\mathbf{x}_t$ and a fidelity level $m_{n,t} \in \mathcal{M} = \{1, \ldots, M\}$ for each player n. We write $\mathbf{m}_t = (m_{1,t}, \ldots, m_{N,t})$. As a result of this query, the controller observes noisy utility feedback $y_{n,t}^{(m_{n,t})}$ for each player n as

$$y_{n,t}^{(m_{n,t})} = u_n^{(m_{n,t})}(\mathbf{x}_t) + \zeta_{n,t}, \tag{4}$$

where $u_n^{(m)}(\mathbf{x})$ denotes a *fidelity-m approximation* of the true utility $u_n(\mathbf{x})$, and $\zeta_{n,t} \sim \mathcal{N}(0, \sigma^2)$ is observational noise. We assume the highest fidelity $m = M$ corresponds to the actual utility, i.e., $u_n^{(M)}(\mathbf{x}) = u_n(\mathbf{x})$, whereas lower fidelities $m < M$ provide cheaper but biased estimates.

Each fidelity level m has an associated cost $\lambda^{(m)} > 0$ representing the resource expenditure (e.g., computation time or experimental cost) for one evaluation at fidelity m. We assume $0 < \lambda^{(1)} \leq \lambda^{(2)} \leq \cdots \leq \lambda^{(M)}$, so that higher fidelity is costlier. The overall querying process is subject to the total cost budget Λ as

$$\sum_{t=1}^{T} \sum_{n=1}^{N} \lambda^{(m_{n,t})} \leq \Lambda, \tag{5}$$

where T is the (random) total number of queries performed. This budget constraint enforces that the controller balance the benefit of information against the cost of its queries.

In summary, the controller faces a sequential decision problem: it must adaptively select which joint configuration $\mathbf{x}_t$ and fidelity levels $\mathbf{m}_t$ to query at each step, based on past observations, in order to efficiently learn the utility functions $u_n(\mathbf{x})$ and approach an ϵ^*-PNE solution within the budget Λ.

3 The MF-UCB-PNE Algorithm

MF-UCB-PNE is a BO scheme designed for the hierarchical game setting described in the previous section. The algorithm maintains a probabilistic surrogate model for the unknown utilities and uses it to decide on informative queries. We leverage a multi-output Gaussian process (MOGP) model to jointly capture all players' utility functions. The GP prior provides an uncertainty estimate for each utility at any candidate action, which we exploit to guide exploration.

MF-UCB-PNE proceeds in a sequence of episodes, indexed by $j = 1, 2, \ldots$. In each *episode* j, the algorithm first executes an *exploration phase* by collecting a batch of low-cost observations, and then it conducts an *evaluation phase* by querying the highest-fidelity utilities to evaluate a candidate action. At the start of episode j, let $\mathcal{D}_{j,0}$ be the set of observations collected in all previous episodes. The exploration phase of episode j consists of a series of queries $(\mathbf{x}_{j,\tau}, \mathbf{m}_{j,\tau})$ for $\tau = 1, 2, \ldots, t_j$, where at least one player n is sampled at lower than max fidelity, i.e., $m_{n,j,\tau} < M$. These points are chosen by an acquisition function that trades off information gain and cost, i.e.,

$$(\tilde{\mathbf{x}}_{j,\tau}, \tilde{\mathbf{m}}_{j,\tau}) = \arg \max_{\mathbf{x}\in\mathcal{X},\mathbf{m}\in\mathcal{M}} \frac{\sum_{n=1}^{N} \mathrm{I}(y_n^{(m_n)}; u_n(\mathbf{x})|\mathbf{x}, m_n, \mathcal{D}_{n,j,\tau-1})}{\sum_{n=1}^{N} \lambda^{(m_n)}} \tag{6a}$$

$$\text{s.t.} \quad \sum_{n=1}^{N} \lambda^{(m_n)} \leq \Lambda - \sum_{j'=1}^{j-1} \Lambda_{j'} - \sum_{n=1}^{N} \sum_{\tau'=1}^{\tau-1} \lambda^{(m_{n,j,\tau'})} - N\lambda^{(M)}, \tag{6b}$$

where details on computing information gain $\mathrm{I}(y_n^{(m_n)}; u_n(\mathbf{x})|\mathbf{x}, m_n, \mathcal{D}_{n,j,\tau-1})$ via MOGP can be found in the longer version [11].

By (6), MF-UCB-PNE adopts an information-theoretic criterion adapted to the multi-fidelity setting: at each exploration step, it selects the joint action and fidelity that maximize the information-to-cost ratio (6a), until a certain stopping condition is met [11]. Intuitively, the exploration phase continues as long as lower-fidelity queries can significantly reduce uncertainty about the optimal ϵ^*-PNE, without exceeding a cost threshold.

After exploration, the algorithm performs an *evaluation* step by selecting the most promising action $\tilde{\mathbf{x}}_j$ based on the surrogate MOGP model. This is done by selecting the action that appears to minimize the maximum dissatisfaction $\max_n f_n(\mathbf{x})$ using an upper confidence bound (UCB) criterion. This yields an evaluation of the actual joint utility outcome and the corresponding dissatisfaction levels $f_n(\tilde{\mathbf{x}}_j)$. If the maximum dissatisfaction $\max_n f_n(\tilde{\mathbf{x}}_j)$ is below a pre-specified threshold η, the algorithm terminates. Otherwise, the knowledge gained from this evaluation is incorporated into the dataset, and a new episode begins if budget remains.

Theoretical guarantees. To analyze the theoretical performance guarantee of MF-UCB-PNE, we first define the *loss* accrued by the selection $(\mathbf{x}, \mathbf{m})$ as

$$\ell(\mathbf{x}, \mathbf{m}) = \begin{cases} \max_n f_n(\mathbf{x}) & \text{if } \mathbf{m} = M \cdot \mathbf{1} \quad \text{(evaluation)} \\ C & \text{otherwise} \quad \text{(exploration)} \end{cases} \tag{7}$$

with constant C representing the cost of exploration at lower fidelity levels, satisfying $C \geq \max_{n,\mathbf{x}} f_n(\mathbf{x})$. Defining the *regret* after using budget Λ as $R(\Lambda) = \sum_{t=1}^{T} \ell(\mathbf{x}_t, \mathbf{m}_t) - \frac{\Lambda\epsilon^*}{N\lambda^{(M)}}$, we have the following upper bound on the regret.

Theorem 1 (Regret Bound of MF-UCB-PNE) *Assume the utility function $u_n(\mathbf{x})$ for each player $n = 1, ..., N$ lies in the reproducing kernel Hilbert space (RKHS) associated with the same kernel function assumed by the MOGP prior. The regret incurred by MF-UCB-PNE is upper bounded as $\tilde{O}(1/\sqrt{\Lambda})$ (up to log factors) with probability at least $1 - N\delta$ for $\delta \in (0, 1/N)$.*

Theorem 1 implies that as the budget Λ grows, MF-UCB-PNE converges to an ϵ^*-PNE with high probability, obtaining, asymptotically, a vanishing regret. Due to space constraints, we refer to the full paper [11] for a detailed statement and a proof of Theorem 1.

4 Experiments

We demonstrate the performance of MF-UCB-PNE on two wireless network problems: a power control game in an interference network, and a random access game for medium access control. In each case, we compare our approach against baseline schemes, including a single-fidelity BO algorithm (UCB-PNE) that always queries at the highest fidelity [9], and the Probability-of-Equilibrium (PE) strategy [12] that greedily uses full-fidelity evaluations without exploration.

Power allocation game. We consider $N = 20$ interfering wireless links (transmitter-receiver pairs). Each link n chooses a transmit power level x_n in the range $[-13, , 23]$ dBm. The utility $u_n(\mathbf{x})$ is defined as the long-term average spectral efficiency (rate) of link n minus a penalty term proportional to its power usage. Specifically, the utility is $u_n(\mathbf{x}) = \mathbb{E}\big[\log_2(1 + \mathsf{SINR}_n)\big] - \xi_n x_n$, where SINR_n is the signal-to-interference-plus-noise ratio for link n given the power profile x, and $\xi_n = 0.1$ is a weight reflecting the importance of power consumption. The expectation is with respect to random channel gains: each link has a Rayleigh fading channel, and interference between any two links is governed by an interference gain factor ψ. In our setup, the direct channel gains are $\mathcal{CN}(0,1)$ variables and interference gains are $\mathcal{CN}(0,\psi)$ variables, with ψ varying from -30 dB (negligible interference) up to 0 dB (strong interference). Higher values of parameter ψ means more interference coupling between links, making the game more challenging.

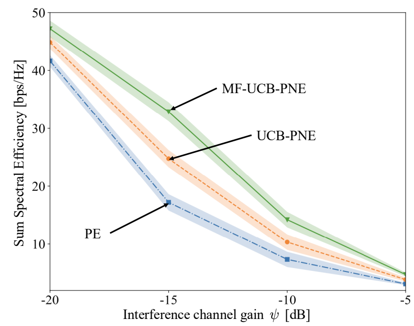

Figure 3: Sum spectral efficiency attained at the approximate equilibrium vs. interference channel gain ψ, in a power allocation setting for an interference channel.

To emulate multi-fidelity feedback, we let fidelity level $m \in 1, 2, 3, 4, 5$ correspond to averaging the SINR over $\lambda^{(m)}$ i.i.d. channel realizations, where $\lambda^{(1)} = 1$ sample (cheapest, no averaging) and $\lambda^{(5)} = 100$ samples (highest cost). Thus, the utility $u_n^{(m)}(\mathbf{x})$ is the sample average of the rate $\log_2(1 + \mathsf{SINR}_n)$ over $\lambda^{(m)}$ draws, providing an unbiased estimate of the true expected rate. We assign query costs equal to the number of samples, i.e. cost $\lambda^{(m)}$ for fidelity m. A high-fidelity query ($m = 5$) is $100\times$ more costly than a low-fidelity one ($m = 1$).

In Figure 3, we plot the *sum spectral efficiency* (sum of all links' rates) achieved by the equilibrium solutions found, as a function of the interference gain ψ. All methods were given the same total budget $\Lambda = 36000$. At low interference, the links' utilities are nearly independent, and all algorithms find similar near-optimal power allocations. In moderate interference regimes, MF-UCB-PNE (green solid line) significantly outperforms UCB-PNE (orange dashed) and the greedy PE baseline (blue dash-dotted), achieving higher sum-rate by finding a better balance between interference and power use. At very high interference, all schemes see degraded performance, but MF-UCB-PNE still maintains a slight advantage. These results show that multi-fidelity exploration allows MF-UCB-PNE to discover improved equilibria especially when interference coupling is non-trivial, by efficiently using the budget on informative low-cost queries.

Random access game. Our second experiment involves optimizing a slotted ALOHA random access protocol. We consider $N = 5$ mobile terminals sharing a wireless channel. Each terminal n has two action parameters: $x_{n,1} \in [0, 1]$ is the probability the terminal is active in a slot (duty cycle), and $x_{n,2} \in [0, 1]$ is the conditional probability of attempting a transmission when active. If terminal n transmits while another is also transmitting, a collision occurs. The success (packet reception) probability for terminal n is given by $T_n(\mathbf{x}) = \frac{x_{n,1}x_{n,2}}{1-x_{n,1}x_{n,2}} \prod_{n' \neq n}(1 - x_{n',1}x_{n',2})$, which accounts for the event that n transmits and no other active terminal transmits in the same slot [13]. The expected energy consumed by terminal n in a slot is $E_n(\mathbf{x}) = x_{n,1}(c_1 + c_2 x_{n,2})$, where $c_1 = 50$ and $c_2 = 70$ (in normalized units) are the energy costs for being active and for transmitting, respectively. We set per-terminal maximum energy budgets such that $E_n(\mathbf{x}) \leq 60, 55, 50, 45, 40$ for $n = 1, \ldots, 5$. Each terminal's utility is $u_n(\mathbf{x}) = T_n(\mathbf{x}) - \xi_n E_n(\mathbf{x})$, where $\xi_n = 6.5 \times 10^{-4}$ is a weight tuning the importance of energy consumption relative to throughput.

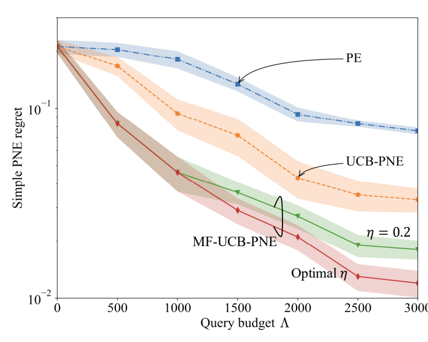

Figure 4: Simple PNE regret (maximum dissatisfaction) in a random access setting as a function of total query budget Λ.

We define three fidelity levels for simulation: at fidelities 1, 2, and 3, the throughput function $T_n(\mathbf{x})$ is replaced by a coarser approximation using factors $\omega^{(m)}$ instead of the true parameters, while fidelity 4 uses the exact $T_n(x)$. The costs are $\lambda^{(1)} = 1$, $\lambda^{(2)} = 5$, $\lambda^{(3)} = 10$, and $\lambda^{(4)} = 20$.

Figure 4 shows the *simple PNE regret* of each method as a function of the total query budget Λ. Regret here is defined as the gap between the achieved performance and the ideal equilibrium. We compare MF-UCB-PNE with a fixed dissatisfaction threshold $\eta = 0.2$ (green line) and with an optimized threshold η (red line), against UCB-PNE (orange dashed) and PE (blue dash-dotted). MF-UCB-PNE attains the lowest regret across all budget values, substantially outperforming both baselines. Notably, UCB-PNE and PE, which always use high-fidelity queries, exhaust their budgets quickly and then stagnate at higher regret. In contrast, MF-UCB-PNE continues to improve as it efficiently allocates budget to informative low-cost queries. For example, at $\Lambda = 3000$, MF-UCB-PNE achieves a regret of about 5×10^{-3}, whereas UCB-PNE stays above 1.5×10^{-2}. This demonstrates that multi-fidelity exploration enables finding better approximate equilibria under tight budget constraints, compared to strategies that naively focus on high-fidelity evaluations.

5 Conclusions

We have proposed MF-UCB-PNE, a novel MFBO strategy that efficiently approximates PNE solutions by leveraging cheap low-fidelity evaluations for broad exploration, transitioning to high-fidelity queries for refined exploitation. Theoretical analysis proves that MF-UCB-PNE obtains a vanishing regret under suitable conditions, while empirical results on wireless networking problems validate the high-quality equilibrium identified by MF-UCB-PNE under limited cost budgets. Looking forward, it would be intriguing to extend the strategy to the evaluation of mixed Nash equilibria [9].

References

[1] Pietro Brach del Prever, Salvatore D'Oro, Leonardo Bonati, Michele Polese, Maria Tsampazi, Heiko Lehmann, and Tommaso Melodia. Pacifista: Conflict evaluation and management in open RAN. *IEEE Trans. Mobile Comput.*, pages 1–15, 2025.

[2] Peter I Frazier. A tutorial on Bayesian optimization. arXiv preprint arXiv:1807.02811, 2018.

[3] Niranjan Srinivas, Andreas Krause, Sham M Kakade, and Matthias W Seeger. Information-theoretic regret bounds for Gaussian process optimization in the bandit setting. *IEEE Trans. Inf. Theory*, 58(5):3250–3265, 2012.

[4] Donald R Jones, Matthias Schonlau, and William J Welch. Efficient global optimization of expensive black-box functions. *J. Glob. optim.*, 13:455–492, 1998.

[5] Philipp Hennig and Christian J Schuler. Entropy search for information-efficient global optimization. *J. Mach. Learn. Res.*, 13(1):1809–1837, 2012.

[6] Jialin Song, Yuxin Chen, and Yisong Yue. A general framework for multi-fidelity Bayesian optimization with Gaussian processes. In *Proc. Int. Conf. Artif. Intell. Stat.*, pages 3158–3167, Naha, Japan, 2019.

[7] Yunchuan Zhang, Sangwoo Park, and Osvaldo Simeone. Multi-fidelity Bayesian optimization with across-task transferable max-value entropy search. *IEEE Trans. Signal Process.*, 73:418–432, 2025.

[8] John Fearnley, Martin Gairing, Paul W Goldberg, and Rahul Savani. Learning equilibria of games via payoff queries. *J. Mach. Learn. Res.*, 16:1305–1344, 2015.

[9] Sebastian Shenghong Tay, Quoc Phong Nguyen, Chuan Sheng Foo, and Bryan Kian Hsiang Low. No-regret sample-efficient Bayesian optimization for finding Nash equilibria with unknown utilities. In *Proc. Int. Conf. Artif. Intell. Stat.*, pages 3591–3619, Valencia, Spain, 2023.

[10] Minbiao Han, Fengxue Zhang, and Yuxin Chen. No-regret learning of Nash equilibrium for black-box games via Gaussian processes. In *Proc. 40th Conf. Uncertainty Artif. Intell.*, pages 1541–1557, Barcelona, Spain, 2024.

[11] Yunchuan Zhang, Osvaldo Simeone, and H Vincent Poor. Multi-fidelity Bayesian optimization for Nash equilibria with black-box utilities. *arXiv preprint arXiv:2505.11265*, 2025.

[12] Victor Picheny, Mickael Binois, and Abderrahmane Habbal. A Bayesian optimization approach to find Nash equilibria. *J. Glob. Optim.*, 73:171–192, 2019.

[13] Lazaros Gkatzikis, Georgios S Paschos, and Iordanis Koutsopoulos. Medium access games: The impact of energy constraints. In *Proc. Int. Conf. NETw. Games, Control. Optim. (NetGCooP 2011)*, pages 1–8, Paris, France, 2011.