RIS Phase Optimization via Generative Flow Networks

Charbel Bou Chaaya and Mehdi Bennis

***Abstract*—This letter introduces a new Machine Learning (ML) technique to learn phase shifting patterns for Reconfigurable Intelligent Surfaces (RISs). We leverage the Generative Flow Network (GFlowNet) paradigm and adapt it so as to compose a RIS phase control resulting in high communication rate. To generalize our approach for different physical layer scenarios, we use a channel chart as a latent representation of the wireless spatial environment to condition the GFlowNet. As such, the GFlowNet learns a scalable policy over RIS configurations that tailors the propagation environment in real-time. We evaluate our solution by means of simulations on a synthetic dataset, and the results corroborate its superiority compared to benchmarks, achieving more than 15% higher communication rates.**

***Index Terms*—Reconfigurable Intelligent Surface, Machine Learning, Generative Flow Networks, Channel Charting.**

I. Introduction

THE stringent demands of forthcoming wireless applications entail innovative communication schemes [1]. As such, researchers are seeking novel technologies that can cater to those demands, with minimal power footprint. A propitious candidate is Reconfigurable Intelligent Surfaces (RISs), also called meta-surfaces. Technically, a RIS is a two-dimensional array of sub-wavelength scattering elements, namely thin layers of meta-material, whose reflection properties can be tuned [2]. By smartly controlling these reflection coefficients, a RIS can be used to meticulously steer impinging signals. What makes this technology appealing is its independence of power amplifiers and radio frequency chains. Moreover, unlike the half-duplex operation of relays, a RIS affects the propagation environment in real-time. Hence, the deployment of these surfaces is expected to enhance the performance of wireless communication systems, as they allow the reconfigurability of its stochastic scattering environment, using low-cost and power efficient hardware.

Aside from their promising potential, integrating a RIS in wireless systems brings about challenging optimization problems. On the one hand, the phase shifts induced by its scattering elements on incident waves must be jointly optimized so that the aggregated signals add constructively. On the other hand, the number of reflecting elements must scale in the order of hundreds or thousands to achieve a noticeable performance improvement [3]. In addition, practical RIS implementations limit its actual phase shifts to a set of discrete values. Such constraints induce a very large and intractable feasibility set for RIS optimization problems, that classical optimization tools fail to manage efficiently.

This work was funded in part by the ERA-NET CHIST-ERA project (MUSE-COM^2: AI-enabled MUltimodal SEmantic COMmunications and COMputing), in part by the Research Council of Finland (former Academy of Finland) project (Vision-Guided Wireless Communication), and in part by the European Union through the project 6G-INTENSE (G.A. no. 101139266).

The authors are with the Centre for Wireless Communications, University of Oulu, Finland. (emails: {charbel.bouchaaya, mehdi.bennis}@oulu.fi).

To this extent, a considerable part of the literature focuses on exploiting recent Machine Learning (ML) paradigms to solve RIS tuning problems in a convenient data-driven manner. From the plethora of ML machinery, supervised learning has been extensively used to configure the RIS [4], [5]. This is mostly due to its relatively simple implementation and excellent results [6]. However, supervised learning techniques require a large training set labeled by the optimal configuration, on which a Neural Network (NN) model is fitted. Such training datasets are extremely expensive to collect, and the trained models are known to perform poorly for out-of-generalization [7].

As a remedy, Reinforcement Learning (RL) trains a RIS control policy by interacting with the wireless environment [8]. Although RL avoids labeled datasets, it suffers from an unstable performance, particularly when the RIS size is large resulting in a huge action space. In fact, the action space increases exponentially with the number of scattering elements. Practically, the number of RIS elements must be as large as possible to guarantee a reliable communication, a regime which RL approaches cannot handle efficiently. This is due to their greedy approach of seeking a policy that maximizes the expected reward, which could be trapped in local optima. Furthermore, RL procedures necessitate lengthy training epochs to converge, consuming a considerable amount of training samples [9].

To deal with the drawbacks of existing studies, we propose a novel ML approach to solve the RIS configuration problem, relying on the recently proposed Generative Flow Networks (GFlowNets) [10]. Instead of maximizing the downstream payoff, GFlowNets draw compositional action samples proportional to the reward they obtain. Hence, sampling readily from GFlowNets renders favorable actions, since they explore all the modes of the reward function.

In terms of physical layer literature, [11] is the only work that uses GFlowNets to select antenna activation schemes providing a desired beampattern. Although the proposed solution is proven to be computationally friendly and data efficient, a critical shortcoming is its focus on a single beampattern. Essentially, various steering directions must be learned by the controller to deal with the different environment observations it captures in real-time. In this regard, we condition the GFlowNet on a latent representation of the spatial wireless environment, namely its channel chart [12]. Adopting the chart as a conditional embedding elicits many benefits other than RIS control, as it can be used for sensing and proactive radio resource management [13]. While the works [14]–[16] exploited the chart for beamforming purposes, the proposed ML models are trained in a supervised manner, whereas our approach is label-free.

II. Preliminaries – Generative Flow Networks

A GFlowNet is a variational inference algorithm that beholds the problem of sampling compositional objects from an intractable target distribution as a sequence of constructive steps aggregating elements of the object. Accordingly, it follows a trajectory-based generative process, where selected actions iteratively modulate a state variable representing the object to compose. Formally, we consider a Directed Acyclic Graph (DAG) $\mathcal{G} = (\mathcal{S}, \mathcal{A})$, where $\mathcal{S}$ is a finite set of states and $\mathcal{A} \subseteq \mathcal{S} \times \mathcal{S}$ represents directed edges, also referred to as actions that transition from a state to another. The DAG is endowed with an initial state with no parents, denoted $\mathbf{s}_0$, whereas states having no outgoing edges are referred to as terminal states, and their collection is a set $\mathcal{X}$. A complete trajectory is a sequence $\tau = (\mathbf{s}_0 \to \mathbf{s}_1 \to \cdots \to \mathbf{s}_n) \in \mathcal{T}$, such that $\forall i, (\mathbf{s}_i, \mathbf{s}_{i+1}) \in \mathcal{A}$ and $\mathbf{s}_n \in \mathcal{X}$.

The GFlowNet forward policy is a choice of distribution $P_F(\mathbf{s}'|\mathbf{s})$ over the states $\mathbf{s}' \in \mathcal{S} \setminus \mathcal{X}$ children of $\mathbf{s}$. As such, an object $\mathbf{x} \in \mathcal{X}$ can be generated by starting from $\mathbf{s}_0$ and sequentially drawing actions from P_F. Note that the forward policy defines a distribution over complete trajectories given by $P_F(\tau) = \prod_{i=0}^{n-1} P_F(\mathbf{s}_{i+1}|\mathbf{s}_i)$. Akin to the forward policy, the backward policy $P_B(\mathbf{s}|\mathbf{s}')$ is a distribution over the parents $\mathbf{s}$ of a non-initial state $\mathbf{s}'$.

The forward policy further induces a marginal policy over terminal states via $\pi(\mathbf{x}) = \sum_{\tau \to \mathbf{x}} P_F(\tau)$, with the sum taken over all trajectories terminating at $\mathbf{x} \in \mathcal{X}$. A reward function R is a non-negative mapping over the set of generated objects, which can be seen as an unnormalized probability mass function over $\mathcal{X}$. The problem approximated by a GFlowNet is that of learning a policy P_F such that the marginal likelihood of sampling any object matches its reward, i.e. $\pi(\mathbf{x}) \propto R(\mathbf{x}), \forall \mathbf{x} \in \mathcal{X}$. In simpler terms, the GFlowNet fits a policy P_F so that the induced marginal $\pi(\mathbf{x}) \approx \frac{R(\mathbf{x})}{Z}$, where $Z = \sum_{\mathbf{x} \in \mathcal{X}} R(\mathbf{x})$ denotes the partition function. The problem is particularly challenging since the partition function Z is unknown and the marginal policy π is intractable to compute exactly, given the forward policy P_F.

In certain cases, the reward is determined by some conditioning information $\mathbf{c} \in \mathcal{C}$, wherein each realization prompts a reward function $R(\mathbf{x}|\mathbf{c})$. Analogous to GFlowNets, reward-conditional GFlowNets [17] learn a policy conditioned on the observation of $\mathbf{c}$, simultaneously modeling the family of conditional rewards. Accordingly, having $\mathbf{c}$ as an input, we denote $P_F(\mathbf{s}'|\mathbf{s}, \mathbf{c})$ and $P_B(\mathbf{s}|\mathbf{s}', \mathbf{c})$ as the conditional forward and backward policies, $\pi(\mathbf{x}|\mathbf{c})$ as the marginal likelihood of composing $\mathbf{x}$ given $\mathbf{c}$, and $Z(\mathbf{c}) = \sum_{\mathbf{x} \in \mathcal{X}} R(\mathbf{x}|\mathbf{c})$ as the conditional partition function. The objective of a conditional GFlowNet is to estimate the conditional forward policy, such that $\pi(\mathbf{x}|\mathbf{c}) \propto R(\mathbf{x}|\mathbf{c})$.

III. System Model and Problem Formulation

We consider the downlink of an Orthogonal Frequency Division Multiplexing (OFDM) communication system between a single-antenna transmitter and single-antenna receiver. Further, we assume that the direct link between them is blocked, and the communication is managed by a RIS comprising N reflecting elements, as shown in Fig. 1. While we consider that both are always in Line-of-Sight (LoS) with the RIS, the transmitter is assumed to be fixed, whereas the receiver is mobile and its position is denoted $\mathbf{p} \in \mathbb{R}^3$. We consider transmissions over W evenly spaced subcarriers spanning a bandwidth B with a central frequency f_c. Let $\mathbf{g} \in \mathbb{C}^{NW}$ represent the channel between the transmitter and the RIS, and $\mathbf{h} \in \mathbb{C}^{NW}$ designate the channel between the RIS and the receiver. The received signal at the destination is:

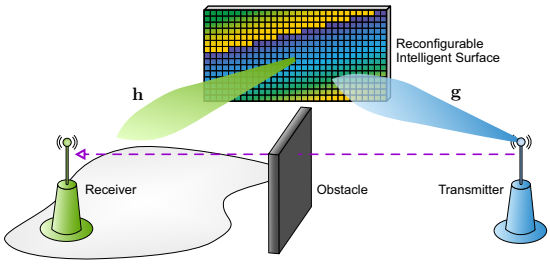


Figure 1. System Model.

$$y = \sqrt{\gamma}\left(\mathbf{g}^{\mathsf{H}}\mathbf{\Phi}\mathbf{h}\right)s + n, \tag{1}$$

where s is the unit-power information signal, γ is the transmit power, and $n \sim \mathcal{CN}(0, \sigma^2)$ is the receiver noise. The RIS reflection matrix $\mathbf{\Phi} = \mathrm{diag}(\boldsymbol{\phi})$, where $\boldsymbol{\phi} = [\phi_1, \ldots, \phi_N]^{\mathsf{T}}$ is a vector consisting of the reflection coefficients of the RIS elements. Essentially, each element applies a phase shift on its incident signal. We consider a realistic case where the phase shifting resolution is finite, i.e.

$$\phi_i \in \mathcal{P} = \left\{e^{j2^{1-\nu}\pi a}\right\}_{a=0}^{2^\nu - 1}, \quad i = 1, \ldots, N, \tag{2}$$

where ν is the phase quantization step in bits.

In this work, we focus on the downlink rate maximization problem, formulated as[1]:

$$\underset{\boldsymbol{\phi} \in \mathcal{P}^N}{\text{maximize}} \qquad \varrho(\mathbf{g}, \boldsymbol{\phi}, \mathbf{h}) = \log_2\left(1 + \frac{|\mathbf{g}_c^{\mathsf{H}}\mathbf{\Phi}\mathbf{h}_c|^2\gamma}{\sigma^2}\right), \tag{3}$$

where $\mathbf{g}_c$ and $\mathbf{h}_c$ are the central subcarrier channels. Deriving an analytical solution to (3) is intricate, since the phases take discrete values. Heuristic search mechanisms can find the optimal solution, however their complexity requires excessive computing resources, scaling as $O\left(2^{\nu N}\right)$. Recently, [18] proposed an algorithm to optimally solve (3) with linear complexity $O(N)$. Although its complexity is much lower compared to other heuristics, N is still very large, which might hinder its effectiveness. Added to that, we show that our proposed approach is more versatile, as it allows reusing the communication spectrum for other purposes, while solving (3).

Note that all these solutions assume that the RIS can estimate both channels $\mathbf{g}$ and $\mathbf{h}$, which can be done using various methods [19], and is outside the scope of this work.

[1]For sake of simplicity, we consider passive beamforming to maximize the rate at the central subcarrier only. However, one can consider the sum rate over all subcarriers and extend our solution in a straightforward manner.

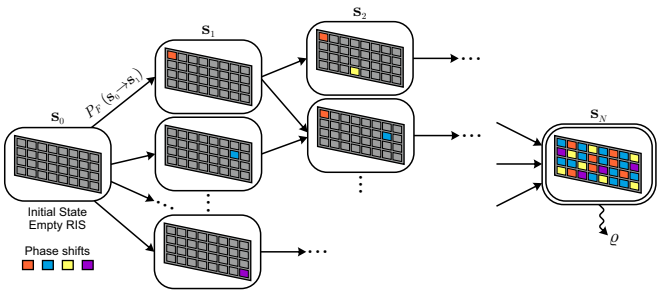


Figure 2. GFlowNet model to compose a RIS phase.

IV. Channel Chart Conditional GFlowNet

A. GFlowNets for RIS Phase Optimization

We propose to use a GFlowNet to sample candidate RIS phases $\boldsymbol{\phi} \in \mathcal{P}^N$. In particular, since the search space is large, a GFlowNet alleviates the problem of RL algorithms that seek to maximize (3). Instead, we identify favorable phase shifts solutions. Since GFlowNets compose objects in proportion to the reward, they explore all its modes. Thus, the sampled objects provide a good coverage of the reward modes, resulting in a diverse set of candidate solutions.

Designing a GFlowNet that constructs a RIS phase vector in a compositional manner is carried out as follows. We consider a DAG whose initial state $\mathbf{s}_0$ is an empty RIS phase vector. The states $\mathcal{S}$ of the DAG represent partially constructed RISs, and each action in $\mathcal{A}$ modifies the phase vector $\boldsymbol{\phi}$ by picking a discrete phase from $\mathcal{P}$. We restrict the GFlowNet to select one phase per RIS element (once a phase is selected, it cannot be changed), and constrain terminal states to fully composed RISs. Finally, we consider the reward function that maps every terminal object $\boldsymbol{\phi} \in \mathcal{P}^N$ to its provided downlink rate $\varrho(\mathbf{g}, \boldsymbol{\phi}, \mathbf{h})$. A model of the GFlowNet is shown in Fig. 2, where $\nu = 2$.

It is worth noting that for practical implementations, we consider an exponentiated rate as a reward function ϱ^β instead of simply ϱ, where $\beta > 0$ is a temperature parameter. As such, the GFlowNet samples RIS configurations proportionally to the modified reward, i.e. $\pi(\boldsymbol{\phi}) \propto \varrho^\beta$. As the reward's values are exponentiated, relatively higher values become largely higher, while relatively lower values become slightly higher. Thus, this parameter controls the diversity of sampled solutions. For instance, by increasing β the GFlowNet is incentivized to sample more likely from the modes of ϱ, which is important to obtain high rate phase shifting solutions. Whereas by decreasing β, the generated samples are more diverse. Controlling this parameter mediates the quality of the sampled solutions between diversity and higher reward.

Consequently, for given realizations of $\mathbf{g}$ and $\mathbf{h}$, one can use the proposed GFlowNet to draw multiple candidates of phase vectors $\boldsymbol{\phi}$, and select the one that maximizes (3). Nevertheless, this approach still assumes particular observations of the channels, whereas in our case the RIS-receiver channel $\mathbf{h}$ varies with the receiver's mobility. To mitigate this problem, one can consider a conditional GFlowNet, where the conditional variable is the channel $\mathbf{h}$. However, the channel's size is very large, and comprises redundant features, which hinders its efficiency as a conditional variable. Hence, we propose conditioning the GFlowNet on a low-dimensional embedding of the channels that conserves their latent structure.

Algorithm 1 Training channel chart conditional GFlowNets

Initialize: Conditional GFlowNet with parameters $\boldsymbol{\theta}$ $(P_F(\cdot|\cdot, \mathbf{z}), P_B(\cdot|\cdot, \mathbf{z}), \log Z(\mathbf{z}))$
repeat
  Estimate channel $\mathbf{h}$ and compute $\mathbf{z} = \zeta(\mathbf{h})$
  Sample trajectory τ following $P_F(\cdot|\cdot, \mathbf{z})$ leading to $\boldsymbol{\phi}$
  Compute reward $\varrho^\beta(\mathbf{g}, \boldsymbol{\phi}, \mathbf{h})$ for generated $\boldsymbol{\phi}$
  Update GFlowNet parameters $\boldsymbol{\theta}$ via gradient step on $\ell_{\boldsymbol{\theta}}(\tau, \mathbf{z})$
until A stopping criterion is met

B. Channel Charting

Channel charting is a dimensionality reduction procedure that aims at projecting high-dimensional wireless channels into a low dimensional embedding space, called a channel chart [12]. It is based on the fact that channel realizations are governed by the manifold hypothesis, where they actually lie in a low-dimensional latent manifold although their original space is high dimensional. Formally, we seek a forward charting function:

$$\begin{array}{rccc} \zeta: & \mathbb{C}^{NW} & \longrightarrow & \mathbb{R}^d \\ & \mathbf{h} & \longmapsto & \mathbf{z} \end{array} \tag{4}$$

such that spatial neighborhoods are preserved as much as possible. In other words, considering distinct receiver positions $\mathbf{p}, \mathbf{p}'$ for which the RIS respectively collects channel observations $\mathbf{h}, \mathbf{h}'$,

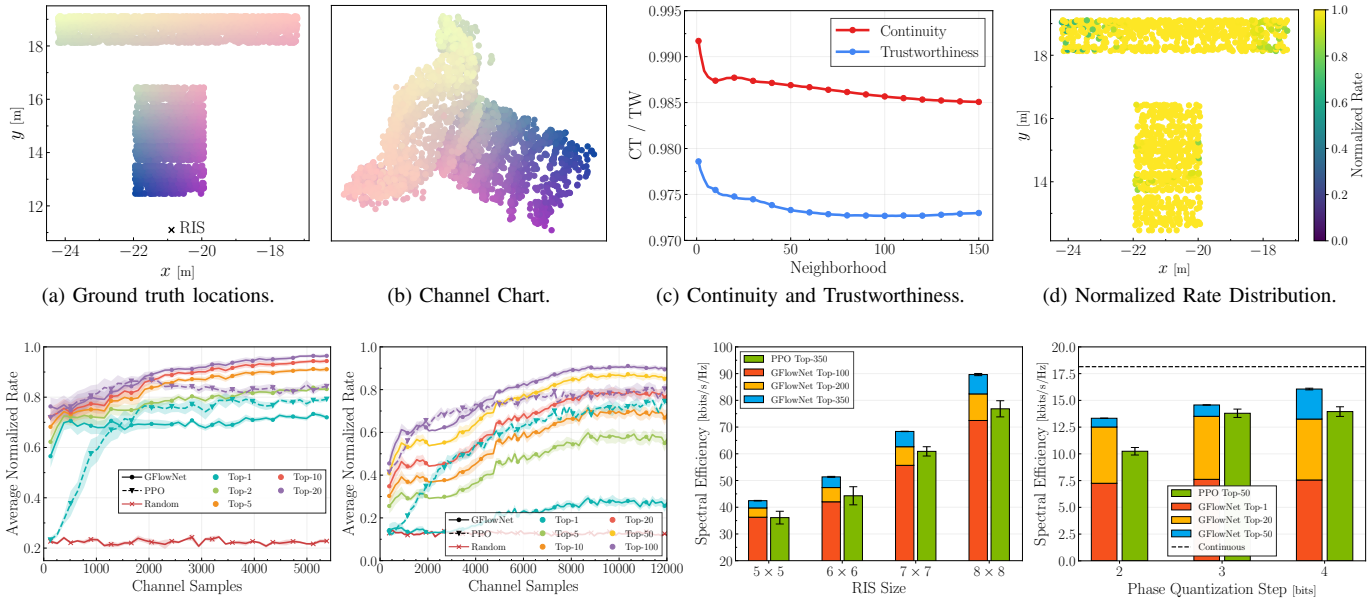
$$\mathbf{p} \approx \mathbf{p}' \iff \zeta(\mathbf{h}) \approx \zeta(\mathbf{h}'), \tag{5}$$

meaning that learned representations corresponding to spatially close channels must remain close in the chart; conversely, neighboring chart points must correspond to neighboring receiver positions. The chart's dimension d is a user-defined parameter, usually set to 2.

Finding the channel chart can be done using many dimensionality reduction techniques [13]. In this work, we rely on the approach presented in [16], as it is computationally efficient and can be easily extended for unseen channel samples. First, the RIS estimates an initial set of M channels $\{\mathbf{h}^k\}_{k=1}^M$, as representative as possible of the receiver's spatial environment. Then, we apply the dimensionality reduction algorithm Isomap over the phase-insensitive channel distance [20, Equation 11]. The estimated channels are organized as columns of a matrix $\mathbf{H} \in \mathbb{C}^{NW \times M}$, and their corresponding initial chart is denoted $\mathbf{Z} \in \mathbb{R}^{2 \times M}$. Subsequently, any new channel sample $\mathbf{h}$ is projected to the chart as the convex combination of the chart points of its closely correlated channel calibration samples, i.e. $\zeta(\mathbf{h}) = \mathbf{Z}\mathbf{d}$, where $\mathbf{d}$ is a normalized vector containing the l-largest values of $|\mathbf{H}^{\mathrm{H}}\mathbf{h}|$, and l is a hyperparameter.

C. Overall Algorithm

Henceforth, we project downstream channel observations $\mathbf{h}$ to the channel chart, and treat their embeddings $\mathbf{z}$ as



(a) Ground truth locations. (b) Channel Chart. (c) Continuity and Trustworthiness. (d) Normalized Rate Distribution.

(e) Convergence curves – 3×3 RIS. (f) Convergence curves – 4×4 RIS. (g) Spectral efficiency for different N. (h) Spectral efficiency for different ν.

Figure 3. Simulation Results

a GFlowNet conditional variable. To train the proposed GFlowNet, we consider the trajectory balance objective [21], that learns the forward and backward conditional policies $P_F(\cdot|\cdot,\mathbf{z})$, $P_B(\cdot|\cdot,\mathbf{z})$, and the log-partition function $\log Z(\mathbf{z})$, parametrized by $\boldsymbol{\theta}$. For a given channel realization $\mathbf{h}$, its representation $\mathbf{z} = \zeta(\mathbf{h})$, and complete trajectory $\tau = (\mathbf{s}_0 \to \mathbf{s}_1 \to \cdots \to \mathbf{s}_N = \boldsymbol{\phi})$, the trajectory balance loss is:

$$\ell_{\boldsymbol{\theta}}(\tau,\mathbf{z}) = \left(\log \frac{Z_{\boldsymbol{\theta}}(\mathbf{z}) \prod_{i=0}^{N-1} P_F(\mathbf{s}_{i+1}|\mathbf{s}_i,\mathbf{z})_{\boldsymbol{\theta}}}{\varrho^{\beta}(\mathbf{g},\boldsymbol{\phi},\mathbf{h}) \prod_{i=0}^{N-1} P_B(\mathbf{s}_i|\mathbf{s}_{i+1},\mathbf{z})_{\boldsymbol{\theta}}}\right)^2 \quad (6)$$

The trajectory balance theorem [21, Proposition 1] states that if $\ell_{\boldsymbol{\theta}}(\tau,\mathbf{z}) = 0$ for all complete trajectories, the induced marginal $\pi(\boldsymbol{\phi}|\mathbf{z}) \propto \varrho^{\beta}(\mathbf{g},\boldsymbol{\phi},\mathbf{h})$. Algorithm 1 recapitulates the training procedure of GFlowNets.

V. Simulation Results

We carry out simulations in scenario 'Indoor 1' of the DeepMIMO datatset [22]. The RIS is placed on the ceiling of an indoor room, while the receiver's position varies along two pre-defined grids, and the transmitter's location is fixed. We consider a communication over $W = 16$ subcarriers, spread over a bandwidth of $B = 20$ Mhz and a central frequency $f_c = 2.5$ GHz. The RIS calibrates the chart using $M = 3000$ estimated channels.

The GFlowNet policies are parameterized using the Multilayer Perceptron (MLP) architecture with 4 hidden layers of 64 neurons and ReLU activations. The model is trained with a batch size of 128 and a learning rate of 0.001.

We compare our approach with the Proximal Policy Optimization (PPO) method [23], a deep RL algorithm that has been previously used to control RISs [24]. PPO also learns a stochastic policy and is therefore a proper benchmark for GFlowNets. We parameterize the actor and critic using two separate MLPs, with 3 hidden layers of 64 neurons and ReLU activations. We consider a discount factor of 0.9, a clipping factor of 0.2, a value function coefficient of 0.5, and an entropy coefficient of 0.05. The model is trained with a batch size of 128 and a learning rate of 0.002.

Channel Charting: We start by providing channel charting results. Fig. 3a shows the ground truth positions of the receiver in the room, while Fig. 3b displays the learned chart. Gradient coloring is used to distinguish local neighborhoods. To test the faithfulness of the chart, we report the continuity (CT) and trustworthiness (TW) metrics as defined in [12]. They respectively characterize the validity of the forward and backward implications in (5), and should be closer to 1 for optimal structure preservation. One can remark from Fig. 3c that the channel chart is a reliable representation for the receiver's location, as the CT and TW values are very close to 1.

Impact of the chart: We start by considering a tractable case where the RIS comprises $N = 9$ elements and $\nu = 1$ bit. Thus, possible phase configurations sum up to 512. We use an exhaustive search approach to find the optimal configurations for a set of receiver test channels. Fig. 3d shows the distribution of the rate provided by the GFlowNet divided by the optimal rate. Clearly, by using the chart as a conditional variable, the GFlowNet samples good candidate solution for each channel embedding, and is capable of rendering almost optimal rates regardless of the receiver's position.

Sample Efficiency: Fig. 3e displays the evolution of the average normalized rate versus the number of observed training channel samples for the same 3×3 RIS with $\nu = 1$ bit. We report the Top-k performance of GFlowNet and PPO, where k is the number of candidate solutions generated for each test channel. We further add the performance of a random RIS control. We notice that for a small RIS, the GFlowNet does not require many training samples to provide a reliable

rate, whereas PPO consumes at least 1000 samples to match its Top-1 performance. More interestingly, even the Top-20 performance of PPO only matches the Top-2 performance of GFlowNet, after convergence. This is due to its greedy approach of fitting the policy that finds the maximal reward, and thus its converged performance only gains 5% when more candidate solutions are tested (Top-1 and Top-20). On the other hand, a GFlowNet draws diverse candidate solutions in proportion to the reward's modes, and thus its performance increases with the number of tested candidates by 15% for instance, comparing Top-2 and Top-10. Similar observations hold for a 4×4 RIS and are presented in Fig. 3f. Since the number of possible configurations increases to 65536, more sampling from the GFlowNet is needed to guarantee a high-rate communication. For example, the Top-20 performance of GFlowNet converges to an average normalized rate of 80%, which is the rate achieved by PPO Top-100. In contrast, sampling 100 candidates from the GFlowNet achieves 90% of the optimal rate on average.

Spectral Efficiency: We now increase the RIS size from 5×5 to 8×8, while keeping $\nu = 1$ bit. Since the number of feasible phases becomes intractable, we show the average spectral efficiency obtained by the two algorithms in Fig. 3g. Clearly, in this scaling regime, the GFlowNet provides much better solutions than PPO. In fact, for a 5×5 RIS, the Top-350 performance of PPO is only comparable to the Top-100 of GFlowNet. Likewise, for a 8×8 RIS, the Top-350 performance of PPO is only 8% higher than the Top-100 performance of GFlowNet. Insofar, by sampling enough from the GFlowNet, its Top-200 and Top-350 rates are 10% and 20% higher than that of PPO. Added to that, one can notice the high variance in the PPO performance underscoring its instability due to the large action space. Despite that, the GFlowNet only shows negligible variance. We finally consider a 3×3 RIS, and vary ν from 2 to 4 bits. We compare the obtained results in Fig. 3h, with an ideal case of continuous phases ($\nu \to \infty$). We notice, for instance, that the PPO algorithm requires 3 phase quantization bits to provide the performance of a GFlowNet with 2 quantization bits, in terms of Top-50 average rate. We again note the instability in the performance of PPO. It is also worth highlighting that when $\nu = 4$, the GFlowNet controlled RIS achieves 91% of the performance of a continuous-phase RIS.

VI. Conclusion

In this letter, we proposed a novel ML approach to configure discrete phase RISs. We designed a GFlowNet that composes RIS phase patterns, and used a latent representation of the wireless environment, particularly a channel chart, to condition its operation. We showed that the GFlowNet performance is very sample efficient, and can conveniently scale with large surfaces, avoiding the drawbacks of RL algorithms in large action spaces. Extensions of this work include adaptations and performance evaluations for GFlowNets under different wireless problems, such as MIMO.

References

[1] M. Chafii, L. Bariah, S. Muhaidat, and M. Debbah, "Twelve scientific challenges for 6g: Rethinking the foundations of communications theory," *IEEE Communications Surveys & Tutorials*, vol. 25, no. 2, pp. 868–904, 2023.

[2] M. D. Renzo, M. Debbah, D.-T. Phan-Huy, A. Zappone, M.-S. Alouini, C. Yuen, V. Sciancalepore, G. C. Alexandropoulos, J. Hoydis, H. Gacanin *et al.*, "Smart radio environments empowered by reconfigurable ai meta-surfaces: An idea whose time has come," *EURASIP Journal on Wireless Communications and Networking*, vol. 2019, no. 1, pp. 1–20, 2019.

[3] E. Björnson, Ö. Özdogan, and E. G. Larsson, "Intelligent reflecting surface versus decode-and-forward: How large surfaces are needed to beat relaying?" *IEEE Wireless Communications Letters*, vol. 9, no. 2, pp. 244–248, 2019.

[4] G. C. Alexandropoulos, S. Samarakoon, M. Bennis, and M. Debbah, "Phase configuration learning in wireless networks with multiple reconfigurable intelligent surfaces," in *2020 IEEE Globecom Workshops (GC Wkshps*. IEEE, 2020, pp. 1–6.

[5] Ö. Özdoğan and E. Björnson, "Deep learning-based phase reconfiguration for intelligent reflecting surfaces," in *2020 54th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2020, pp. 707–711.

[6] H. Zhou, M. Erol-Kantarci, Y. Liu, and H. V. Poor, "A survey on model-based, heuristic, and machine learning optimization approaches in ris-aided wireless networks," *IEEE Communications Surveys & Tutorials*, 2023.

[7] S. Samarakoon, J. Park, and M. Bennis, "Robust reconfigurable intelligent surfaces via invariant risk and causal representations," in *2021 IEEE 22nd International Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*. IEEE, 2021, pp. 301–305.

[8] G. C. Alexandropoulos, K. Stylianopoulos, C. Huang, C. Yuen, M. Bennis, and M. Debbah, "Pervasive machine learning for smart radio environments enabled by reconfigurable intelligent surfaces," *Proceedings of the IEEE*, vol. 110, no. 9, pp. 1494–1525, 2022.

[9] A. Taha, Y. Zhang, F. B. Mismar, and A. Alkhateeb, "Deep reinforcement learning for intelligent reflecting surfaces: Towards standalone operation," in *2020 IEEE 21st international workshop on signal processing advances in wireless communications (SPAWC)*. IEEE, 2020, pp. 1–5.

[10] E. Bengio, M. Jain, M. Korablyov, D. Precup, and Y. Bengio, "Flow network based generative models for non-iterative diverse candidate generation," *Advances in Neural Information Processing Systems*, vol. 34, pp. 27 381–27 394, 2021.

[11] S. Evmorfos, Z. Xu, and A. Petropulu, "Gflownets for sensor selection," in *2023 IEEE 33rd International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2023, pp. 1–6.

[12] C. Studer, S. Medjkouh, E. Gonultaş, T. Goldstein, and O. Tirkkonen, "Channel charting: Locating users within the radio environment using channel state information," *IEEE Access*, vol. 6, pp. 47 682–47 698, 2018.

[13] P. Ferrand, M. Guillaud, C. Studer, and O. Tirkkonen, "Wireless channel charting: Theory, practice, and applications," *IEEE Communications Magazine*, vol. 61, no. 6, pp. 124–130, 2023.

[14] T. Ponnada, P. Kazemi, H. Al-Tous, Y.-C. Liang, and O. Tirkkonen, "Best beam prediction in non-standalone mm wave systems," in *2021 Joint European Conference on Networks and Communications & 6G Summit (EuCNC/6G Summit)*. IEEE, 2021, pp. 532–537.

[15] L. Le Magoarou, T. Yassine, S. Paquelet, and M. Crussière, "Channel charting based beamforming," in *2022 56th Asilomar Conference on Signals, Systems, and Computers*. IEEE, 2022, pp. 1185–1189.

[16] T. Yassine, B. Chatelier, V. Corlay, M. Crussiere, S. Paquelet, O. Tirkkonen, and L. L. Magoarou, "Model-based deep learning for beam prediction based on a channel chart," *arXiv preprint arXiv:2312.02239*, 2023.

[17] Y. Bengio, S. Lahlou, T. Deleu, E. J. Hu, M. Tiwari, and E. Bengio, "Gflownet foundations," *Journal of Machine Learning Research*, vol. 24, no. 210, pp. 1–55, 2023.

[18] D. Zhao, G. Wang, J. Wang, and C. Wang, "Optimal passive beamforming for reconfigurable intelligent surface-assisted communications with discrete phase shifts," *IEEE Wireless Communications Letters*, 2023.

[19] B. Zheng, C. You, W. Mei, and R. Zhang, "A survey on channel estimation and practical passive beamforming design for intelligent reflecting surface aided wireless communications," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 2, pp. 1035–1071, 2022.

[20] L. Le Magoarou, "Efficient channel charting via phase-insensitive distance computation," *IEEE Wireless Communications Letters*, vol. 10, no. 12, pp. 2634–2638, 2021.

[21] N. Malkin, M. Jain, E. Bengio, C. Sun, and Y. Bengio, "Trajectory balance: Improved credit assignment in gflownets," *Advances in Neural Information Processing Systems*, vol. 35, pp. 5955–5967, 2022.

[22] A. Alkhateeb, "DeepMIMO: A generic deep learning dataset for millimeter wave and massive MIMO applications," in *Proc. of Information Theory and Applications Workshop (ITA)*, San Diego, CA, Feb 2019, pp. 1–8.

[23] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.

[24] A. A. Puspitasari and B. M. Lee, "A survey on reinforcement learning for reconfigurable intelligent surfaces in wireless communications," *Sensors*, vol. 23, no. 5, p. 2554, 2023.