Dynamic Resource Allocation for IRS Assisted Energy Harvesting Systems with Statistical Delay Constraint

Imtiaz Ahmed[†], Su Yan[†], Danda B. Rawat[†], and Cong Pu[††]
[†]Howard University, Washington, DC, USA, [††]Marshall University, Huntington, WV, USA
Emails: imtiaz.ahmed@howard.edu, su.yan@howard.edu, danda.rawat@howard.edu, cong.pu@marshall.edu.

***Abstract*—In this paper, we develop centralized and distributed dynamic resource allocation schemes for an intelligent reflecting surface (IRS) aided energy harvesting (EH) system that optimize transmit power of a source node and phase-shift of passive reflecting elements of IRS. The source node randomly harvests renewable energy from the surrounding environment and performs data transmission with the harvested energy while satisfying statistical packet delay constraints in terms of maximum acceptable delay–outage probability. Our developed schemes do not require the statistical distributions of channel and energy profiles to be known and apply deep reinforcement learning (DRL) algorithm. Simulation results demonstrate the effectiveness of the proposed resource control scheme for IRS-aided EH system in different channel and energy conditions.**

***Index Terms*—Intelligent Reflecting Surface, Effective Capacity, Statistical Delay Constraint, Energy Harvesting, Deep Reinforcement Learning.**

I. Introduction

Recently, design of intelligent reflecting surface (IRS) assisted wireless communication systems has attracted significant attention in the research community due to the recent advancements of metamaterials and radio frequency (RF) electronics [1]. In general, IRS is a planar surface composed of a large number of passive reflecting elements (PREs), each of which can induce a controllable change of amplitude and phase of the incident signal independently and hence can change the reflected signal propagation in real time. A joint active transmit beamforming at the access point (AP) and passive reflection beamforming at the IRS was developed in [2] for single and multi-user scenarios that minimizes the total transmit power at the AP while maintaining minimum signal-to-noise ratio (SNR) requirements. The energy efficiency of a downlink multi-user communication system was maximized in [3] by joint power allocation at the base station (BS) and phase-shift design at IRS. In [4], an optimal length was calculated for pilot training symbols by maximizing the asymptotic spectral efficiency for an IRS-aided communication system.

In this paper, we consider an IRS-aided point-to-point communication system, where the source node is powered by randomly available renewable energies [5]. We consider a statistical delay quality-of-service (QoS) constraint at the source node, where the transmission delay is allowed to surpass a delay threshold within a maximum tolerable (delay) outage probability (OP) [6]. Our objective is to study the behavior of IRS for an energy harvesting (EH) system while considering delay QoS constraint and thereby to develop resource control schemes in order to maximize the end-to-end system throughput. Recently, an EH source is considered at the IRS in [7], where a joint transmit power allocation at the BS and phase-shift design at IRS was proposed by incorporating deep reinforcement learning (DRL) technique. In contrast to [7], in this paper, we focus on developing centralized and distributed dynamic resource allocation or resource control schemes that optimize resources over time intervals and learn the system behavior while satisfying delay QoS and EH constraints. Furthermore, extensive research has been conducted while developing IRS-aided simultaneous wireless information and power transfer (SWIPT) protocols in different use-cases in [8]–[10]. The contributions of this paper are summarized as follows:

- Our developed resource control schemes jointly optimize transmit power at the source node and phase-shift of the PREs at IRS.
- The proposed schemes incorporate statistical delay QoS constraint and do not require channel and energy statistics to be known to calculate optimal resources.
- The distributed scheme leverages DRL and entails low computational complexity.

II. System Model

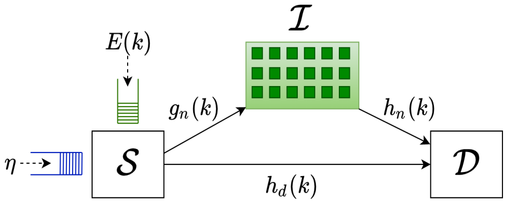

Fig. 1: An IRS-aided point-to-point communication system.

We consider an IRS assisted communication system, where the transmitter $\mathcal{S}$ sends information signal to the receiver $\mathcal{D}$ via direct link and IRS $\mathcal{I}$, see Fig. 1. Note that $\mathcal{S}$ and $\mathcal{D}$ operate with single transmit antenna and single receive antenna, respectively. $\mathcal{S}$ does not have an off-the-shelf constant power supply (e.g., generator, high capacity battery, etc.). Instead, the node is equipped with an EH module, which can harvest renewable energies from the surrounding environment in the form of solar, wind, thermal, mechanical, etc. Once harvested, energies are stored in a small-size battery with limited capacity, and are used for signal processing and data transmission purpose. We assume that the data packets at $\mathcal{S}$ arrive at a constant rate ν (say, in the form of transport blocks (e.g., in long term evolution (LTE), WiFi, etc.) from high layers) and are stored in a data queue. The data packets are then released from the data queue and sent to $\mathcal{D}$ directly and via $\mathcal{I}$ based on the available energy (at $\mathcal{S}$) and the channel conditions of $\mathcal{S}$–$\mathcal{D}$ and $\mathcal{S}$–$\mathcal{I}$–$\mathcal{D}$ links. The data transmission happens over time intervals of equal duration T seconds (s). $\mathcal{I}$ is composed of $N = N_A N_E$ PREs in a uniform planar array (UPA), where N_A and N_E represent the number of elements in the azimuth and elevation planes, respectively. The purpose of using IRS is to assist the communications between two communicating nodes by dynamically adjusting the phase-shift of each of the PREs. It is worth mentioning that IRS yields higher spectral efficiency over conventional amplify-and-forward and decode-and-forward relays as most of the relays work in half-duplex mode whereas IRS, designed by passive elements, operates on the full-duplex mode [1].

Channel Model: We consider baseband equivalent channel models for the considered $\mathcal{S}$–$\mathcal{D}$ and $\mathcal{S}$–$\mathcal{I}$–$\mathcal{D}$ links, where the channel in each link is assumed to be block–faded over the transmission time intervals. In particular, we define the baseband equivalent complex-valued channel gain matrices in time interval $k \in \{1, 2, \cdots\}$ for $\mathcal{S}$–$\mathcal{I}$ and $\mathcal{I}$–$\mathcal{D}$ links as $\boldsymbol{g}(k)$ and $\boldsymbol{h}(k)$, respectively, where, $\boldsymbol{g}(k) = [g_1(k), g_2(k), \cdots, g_N(k)]^T$ and $\boldsymbol{h}(k) = [h_1(k), h_2(k), \cdots, h_N(k)]^T$. We denote $g_n(k) = |g_n(k)|e^{j\theta_{g_n}(k)}$ ($h_n(k) = |h_n(k)|e^{j\theta_{h_n}(k)}$), where $g_n(k)$ ($h_n(k)$), $n \in \{1, 2, \cdots, N\}$ represent the Rician channel fading coefficients with non-zero mean β_g (β_h) and variance σ_g^2 (σ_h^2). Moreover, we define $h_d(k) = |h_d(k)|e^{j\theta_{h_d}(k)}$ as the complex channel fading for $\mathcal{S}$–$\mathcal{D}$ link in time interval $k \in \{1, 2, \cdots\}$, where $h_d(k)$ possesses zero mean and variance σ_d^2. It is worth mentioning that the high likelihood of line-of-sight (LOS) components for $\mathcal{S}$-$\mathcal{I}$ and $\mathcal{I}$-$\mathcal{D}$ links justifies the assumptions of Rician fading channels. On the contrary, we assume that $\mathcal{S}$ and $\mathcal{D}$ are far apart from each other and the probability of establishing a LOS path is very small. Note that $\theta_{g_n}(k)$, $\theta_{h_n}(k)$, and $\theta_{h_d}(k)$ depict the random phase of $g_n(k)$, $h_n(k)$, and $h_d(k)$ in $[0, 2\pi)$. It is worth mentioning that the indirect channel between $\mathcal{S}$ and $\mathcal{D}$ through $\mathcal{I}$ is often represented as keyhole channel [2]. Let us denote $\Psi(k) = [\psi_1(k), \psi_2(k), \cdots, \psi_N(k)]^T$ and $\Theta(k) = \mathrm{diag}(\alpha e^{j\psi_1(k)}, \alpha e^{j\psi_2(k)}, \cdots, \alpha e^{j\psi_N(k)})$, where $\psi_n(k)$ and α represent the phase-shift and constant amplitude[1] reflection coefficient, respectively of PRE $n \in \{1, 2, \cdots, N\}$ of IRS in time interval $k \in \{1, 2, \cdots\}$. Therefore, the composite $\mathcal{S}$–$\mathcal{I}$–$\mathcal{D}$ channel is modeled as a cascaded version of $\mathcal{S}$–$\mathcal{I}$ link, IRS reflections with phase-shifts and amplitude coefficients, and $\mathcal{I}$–$\mathcal{D}$ link.

Signal Model: The received signal $y(k)$ at $\mathcal{D}$ in transmission time interval $k \in \{1, 2, \cdots\}$ can be represented as

$$y(k)=\Big(h_d(k)+\alpha\sum_{n=1}^{N} h_n(k)e^{j\psi_n(k)}g_n(k)\Big)x(k)+w(k), \tag{1}$$

where $x(k)$ represents the signal transmitted by $\mathcal{S}$ with instantaneous power $P(k) = |x(k)|^2$, and $w(k)$ denotes the additive white Gaussian noise (AWGN) with zero mean and variance σ_w^2. Therefore, the instantaneous received signal-to-noise ratio (SNR) can be expressed as $\tilde{\gamma}_\Psi(k) = \gamma_\Psi(k)P(k)$, where $\gamma_\Psi(k) = |h_d + \alpha \sum_{n=1}^{N} h_n(k)e^{j\psi_n(k)}g_n(k)|^2/\sigma_w^2 \in \mathbb{G}$. Here, $\mathbb{G}$ represents the state space of squared end-to-end channel gain.

Energy model: $\mathcal{S}$ is powered by an EH module, which collects $E(k)$ Joules (J) of renewable energies from the surrounding environment during time interval k. We model $E(k) \in \mathbb{E}$ as a stationary random variable with energy state space $\mathbb{E}$ and probability density function $f_{\mathbb{E}}(E)$. The renewable energies are stored in a battery, where the state of the battery is updated at the beginning of the time interval. Considering $B(k)$ as the stored energy in the battery at the beginning of time interval k, the battery state is updated as follows [5]:

$$B(k+1) = [B(k) - TP(k) - \zeta]^{+} + E(k), \ \ \forall k, \tag{2}$$

where $[v]^{+} = \max\{v, 0\}$ for any v, and ζ represents a constant energy consumption due to signal processing tasks at $\mathcal{S}$. Note that ζ does not change over time and hence is not considered as part of optimization variables. Thus, $\{B(k)\} \in \mathbb{B}$ follows a first–order MDP that depends only on the present and immediate past conditions. The state of the battery is represented by $\mathbb{B}$. When transmitting with power $P(k)$ under channel state $\gamma_\Psi(k)$ while occupying bandwidth $\mathcal{W}$ Hz, the throughput can be represented by Shannon's formula: $U(k) = T\mathcal{W}\log_2(1 + P(k)\gamma_\Psi(k))$.

Delay model: S is equipped with a data queue that stores the incoming data with the constant rate η and supports the service rate $\{U(k)\}$. Denoting $D(k) \geq 0$ be the length of the data queue at the onset of time interval k, the queue state is updated as

$$D(k+1) = D(k) - \min\{D(k), U(k)\} + \eta, \quad \forall k. \tag{3}$$

Considering the steady–state data queue length D is bounded, we define the statistical delay QoS constraint as

$$\Pr(D > D_{\max}) \leq \epsilon, \tag{4}$$

where $\Pr(x)$ denotes the probability of an event x. Moreover, $D_{\max}$ and $\epsilon \in (0, 1)$ express the maximum queue–length and delay OP, respectively. Note that a smaller (larger) ϵ represents a more (less) tight delay requirement for a given $D_{\max}$.

III. Proposed Dynamic Resource Allocation

In this section, we develop a framework for joint power allocation at $\mathcal{S}$ and phase optimization of PREs at $\mathcal{I}$ that maximizes η for a given statistical delay QoS and EH constraints (at $\mathcal{S}$).

A. Problem Formulation

The randomness in harvesting renewable energies at $\mathcal{S}$ while satisfying the delay QoS constraint poses challenges in the design and optimization of the considered system. In order to tackle the challenges, we formulate an optimization problem

$$\text{Pr-1:} \quad \max_{\eta, \Psi(k), P(k) \geq 0, \forall k} \quad \eta \tag{5}$$

$$\text{s.t.:} \quad \text{Constraints (2), (3), and (4),} \tag{6}$$

$$\psi_n(k) \in \mathcal{Z}, \quad \forall n, \forall k. \tag{7}$$

Constraint (7) in Pr-1 confirms that $\psi_n(k)$ possesses values from a discrete set of phase-shift $\mathcal{Z}$. It is worth mentioning that the optimal value obtained from Pr-1 depicts the 'effective capacity' of the considered IRS system. To address delay constraint (4), the distribution of D is required to be investigated, which is quite challenging. We follow [11, Sec. III-A.1], adopt the asymptotic delay analysis, convert the statistical delay constraint (4) into a more tractable form, and reformulate Pr-1 as follows:

$$\text{Pr-2:} \quad \max_{\Psi(k), P(k) \geq 0, \forall k} \quad -\log \mathcal{E}\{e^{-\theta^{\mathrm{tar}} U(k)}\}/\theta^{\mathrm{tar}} \tag{8}$$

$$\text{s.t.:} \quad \text{Constraints (2) and (7),} \tag{9}$$

where $\mathcal{E}\{\cdot\}$ represents statistical expectation and $\theta^{\mathrm{tar}} = -\log\epsilon/D_{\max}$. Assuming $\theta = \theta^{\mathrm{tar}} T\mathcal{W}/\log(2)$, where $\log(\cdot)$ denotes the natural logarithm, and exploiting the monotonicity of $\log(\cdot)$, Pr-2 can be represented as follows:

$$\text{Pr-3:} \quad \min_{\Psi(k), P(k) \geq 0, \forall k} \quad \mathcal{E}\Big\{(1 + \gamma_\Psi(k)P(k))^{-\theta}\Big\} \tag{10}$$

$$\text{s.t.:} \quad \text{Constraint (2) and (7).} \tag{11}$$

[1] In general, each PRE is designed to maximize the signal reflection [1]. Hence, without loss of generality, we set $\alpha = 1$ in the simulation results.

Remark 1: Minimizing the objective function of Pr-3 results in maximizing the effective capacity of the considered IRS-aided system while jointly optimizing the transmit power of $\mathcal{S}$ and phase-shift of IRS elements and satisfying the end-to-end delay QoS and EH constraints for $\mathcal{S}$.

Remark 2: We observe that Pr-3 represents an infinite-horizon Markov decision process (MDP) because of the presence of (2). An optimal approach to solve Pr-3 is to apply dynamic programming (DP) and calculate optimal power and phase-shift (of IRS elements) over time. This approach requires the statistics of the channel and EH processes to be known.

B. Problem Solution

In this section, we develop two solution approaches for Pr-3.

- First, in the *Centralized Scheme*, we explore how $P(k)$ and $\psi_n(k), \forall n$ can be optimized jointly while considering the instantaneous (accurate) channel state information (CSI) of all the links and the (accurate) energy states of the battery at $\mathcal{S}$ to be known. This scheme can be readily deployed at a single node (e.g., in $\mathcal{S}$). Once the optimal phase-shift is calculated (e.g., in $\mathcal{S}$), it is sent to $\mathcal{I}$ through backhaul channel to adjust the phases of PREs to optimal values.
- Second, in the *Distributed Scheme*, we sequentially solve optimal $\psi_n(k), \forall n$ and $P(k)$ to reduce the computational complexity of the proposed *Centralized Scheme*. This (distributed) scheme can be applied in a distributed fashion in between $\mathcal{S}$ and $\mathcal{I}$ provided that the IRS module is connected with active (sensing) module that tunes the phase of PREs to optimal values after sensing the channel phases and then sending this (optimized phase-shift) information back to $\mathcal{S}$. $\mathcal{S}$ then optimizes $P(k)$.

1) Centralized Scheme: Joint Optimization of Power and Phase-Shift: Let us define a stationary policy π as a sequence of the decision rules (mapping function selecting optimal $P(k)$ and $\Psi(k)$ for given states) that are independent of time intervals. Moreover, we denote Π as the set containing all the feasible policies for Pr-3. Mathematically, π is expressed with an $(N+1)$-tuple function $(P, \Psi) : \mathcal{B} \times \mathcal{G} \rightarrow \mathbb{R}^+ \times {\mathbb{R}^+}^N$, where $\mathbb{R}^+$ represents the set of non–negative numbers. We denote a policy π^* as the optimal policy that solves Pr-3. As we assume that the channel and energy statistics are unknown to the resource controller, solving Pr-3 in order to obtain optimal $P(k)$ and $\Psi(k)$ over transmission time intervals is indeed a challenging task. Let us define $V(B, \gamma_\Psi)$ as the state value function for Pr-3. The Bellman's optimality equation for Pr-3 can be written as follows [12]:

$$V(B,\gamma_\Psi) = \min_{\substack{TP\in[0,B],\\ \Psi\in\mathcal{Z}}} \Big\{ (1+\gamma_\Psi P)^{-\theta} + \sum_{\hat{\gamma}\in\mathbb{G},\hat{E}\in\mathbb{E}} p_{\mathbb{G}}(\hat{\gamma}_\Psi) p_{\mathbb{E}}(\hat{E}) \\ V(B-TP-\zeta+\hat{E},\hat{\gamma}) \Big\} - V(B^0,\gamma_\Psi^0), \quad (12)$$

for a fixed state (B^0, γ^0_Ψ). The optimal policy π^* is the optimal solution of (12).

Similar to [13], we define post–decision state (PDS) and post–decision state–value function (PDSVF) for the considered problem in (12). The PDSVF $W_0(\check{B})$ for the considered system can be defined as

$$W_0(\check{B}) = \sum_{\hat{\gamma}_\Psi\in\mathbb{G},\hat{E}\in\mathbb{E}} p_{\mathbb{G}}(\hat{\gamma}_\Psi) p_{\mathbb{E}}(\hat{E}) W_0(\check{B},\hat{\gamma}_\Psi) \quad (13)$$

for PDSs $\check{B} \in \mathbb{B}$. During the PDS, the dynamics of the battery in time interval k can be represented as $\check{B}(k) = [B(k) - TP(k) - \zeta]^+$. The objective of developing PDSVF helps us to develop online resource control algorithm using DRL approach that learns the system behavior and calculates optimal allocation of resources without the knowledge of channel and energy statistics. Although the resource allocation schemes developed in this section are applicable for single antenna case, the PDSVF can be extended for multi-antenna case (e.g., spatial multiplexing, spatial diversity, etc.) with appropriate modifications of (12).

From (12) and (13), we can write the optimality equation as follows:

$$W_0(\check{B}) = \sum_{\hat{\gamma}_\Psi\in\mathbb{G},\hat{E}\in\mathbb{E}} p_{\mathbb{G}}(\hat{\gamma}_\Psi) p_{\mathbb{E}}(\hat{E}) \min_{\substack{TP\in[0,\check{B}+\hat{E}],\\ \Psi\in\mathcal{Z}}} \Big\{ (1+\hat{\gamma}_\Psi P)^{-\theta} \\ + W_0(\check{B}-TP-\zeta+\hat{E}) \Big\} - W_0(\check{B}^0) \quad (14)$$

for some arbitrary but fixed state $\check{B}^0$.

Remark 3: The optimization problem depicted in (14) is a mixed-integer non-linear problem (MINLP) as the phase-shift of PREs at $\mathcal{I}$ and the transmit power at $\mathcal{S}$ possess discrete and continuous values, respectively. This is a non-convex optimization problem, and we can adopt spatial branch-and-bound (sBB) algorithm to solve the problem optimally at the expense of prohibitively high computational complexity [14]. The worst-case computational complexity of the proposed centralized scheme in $k \in \{1, 2, \cdots\}$ is exponential in N [14].

2) Distributed Scheme: In order to reduce the computational complexity of the joint power allocation and phase-shift design algorithm, we adopt a sequential optimization approach assisted by DRL. In principle, for a given transmission time interval, we optimize Ψ for a given P and then for a given Ψ, we calculate the optimal P. Note that this approach leads to a suboptimal but computationally tractable solution that can be readily applied in real-time resource control problem as depicted in (14). In this sequel, we divide the solution approach into two subproblems, namely *'IRS Problem*' and *'User Problem'* as described in the following.

IRS Problem: We fix P and solve the following problem:

$$W_i(\check{B}) = \sum_{\hat{\gamma}_\Psi\in\mathbb{G},\hat{E}\in\mathbb{E}} p_{\mathbb{G}}(\hat{\gamma}_\Psi) p_{\mathbb{E}}(\hat{E}) \min_{\Psi\in\mathcal{Z}} \Big\{ (1+\hat{\gamma}_\Psi P)^{-\theta} \\ + W_i(\check{B}-TP-\zeta+\hat{E}) \Big\} - W_i(\check{B}^0). \quad (15)$$

Following the findings observed in [2, Eq. (28)], we calculate the optimal choice of $\psi_n(k)$ to satisfy (15) as follows:

$$\psi_n^*(k) = \theta_{h_d}(k) - \theta_{g_n}(k) - \theta_{h_n}(k), \quad \forall n, \forall k. \quad (16)$$

Note that $\psi_n^*(k)$ requires instantaneous phase information of channel states to be known ($\psi_n^*(k)$ does not depend on any statistical distributions of channels).

User Problem: Once we obtain $\Psi^*(k)$, we solve the following optimization problem in order to obtain $P^*(k)$.

$$W(\check{B}) = \sum_{\hat{\gamma}_\Psi\in\mathbb{G},\hat{E}\in\mathbb{E}} p_{\mathbb{G}}(\hat{\gamma}_\Psi) p_{\mathbb{E}}(\hat{E}) \min_{TP\in[0,\check{B}+\hat{E}]} \Big\{ (1+\hat{\gamma}_\Psi P)^{-\theta} \\ + W(\check{B}-TP-\zeta+\hat{E}) \Big\} - W(\check{B}^0) \quad (17)$$

We assume that the $p_{\mathbb{G}}(\hat{\gamma}_\Psi)$ and $p_{\mathbb{E}}(\hat{E})$ are not known to $\mathcal{S}$ in order to solve (17). We apply a time-averaging algorithm that solves $P^*(k)$ using (18) while learning the system statistics by solving PDSVF using (19).

Remark 4: It is worth mentioning that we can adopt the exact same approach as depicted in [11] to solve problem (17). However, for a large number of states of the battery, the real-

$$P(k) = \underset{TP(k)\in[0,B(k)]}{\arg\min} \Big\{ (1+\gamma_{\Psi}(k)P(k))^{-\theta} + W_P^{(k)}(B(k) - TP(k) - \zeta) \Big\}. \tag{18}$$

$$W_P^{(k)}(\check{B}(k)) = (1-f(k))W_P^{(k)}(\check{B}(k)) + f(k)\big(\min_{TP\in[0,\check{B}(k)+H(k)]} \{(1+\gamma_{\Psi}(k)P)^{-\theta} + W_Q^{(k)}(\check{B}(k) - TP(k) - \zeta + E(k))\} - W_P^{(k)}(\check{B}(0))\big). \tag{19}$$

time implementation complexity can be prohibitively high. In order to address this problem, we adopt DRL based resource control approach, where a deep neural network (DNN) is trained over time to map state-action pair by incorporating states ($\check{B}(k)$) at the input layers and $W_P^{(k)}(\check{B}(k))$ at the output layer of the considered DNN.

We initialize two DNNs, namely deep policy network (DPN) and deep target network (DTN) in order to approximate $W_P^{(k)}(\check{B}(k))$ over time intervals. With the help of replay memory (RM) and considered neural networks (NNs), we calculate $P^*(k)$ and update PDSVF. The detailed step-by-step processes are described in Algorithm 1. Unlike centralized scheme, the computational complexity of the proposed distributed scheme does not grow exponentially with N. In particular, by following Theorems 1 and 2 in [11], it can be shown that the complexity of calculating $P^*(k)$ is polynomial in N for time interval $k \in \{1, 2, \cdots\}$.

Algorithm 1 DRL Algorithm for User Problem

```
Initialize: RM with size $\mathcal{N}$ tuples of experiences. Each experience contains the tuple $e(0) = (B(0), P(0), U(1), B(1))$.
Initialize: DPN with random weights and bias factors. Set $\check{B}^0 \in \mathbb{B}$ and obtain $W_P^{(1)}(\check{B})$ from DPN.
Initialize: DTN with same the weights and the bias factors of DPN.
Set: $\xi$ and $\Xi$ as minimum and maximum number of time intervals, respectively for training.
for $k \in \{1, 2, \cdots\}$ do
    Calculate $P^*(k)$ using (18).
    Store the experience tuple $e(k) = (B(k), P^*(k), U(k+1), B(k+1))$ in RM.
    if $k \geq \xi$ then
        Create a mini-batch of $\mathcal{L}$ elements extracted from RM.
        for $l \in \mathcal{L}$ do
            Calculate PDS $\check{B}_l(k) = [B_l(k) - TP_l^*(k) - \zeta]^+$ and the corresponding PDSVF $W_P^{(k)}(\check{B}(k))$ using (19). Here, $W_Q^{(k)}(\check{B}(k))$ denotes PDSVF obtained from DTN and $f(k)$ represents learning rate of DRL. The PDS and PDSVF represent input and output training sequences, respectively for DPN.
        end for
        Train DPN with $\mathcal{L}$ elements.
        if $k \geq \Xi$ then
            Copy DPN weights to DTN.
        end if
    end if
end for
```

3) Baseline Approach: In this scheme, we do not consider $\mathcal{I}$ to be the part of data transmission and hence do not optimize the phase-shift of PREs at $\mathcal{I}$. In this baseline approach, $\mathcal{S}$ transmits data to $\mathcal{D}$ via h_d only, and the optimal power control algorithm can be designed using the solutions obtained in [11, Sec. III-B.2]. This approach potentially demonstrates the usefulness of our proposed schemes in Section IV.

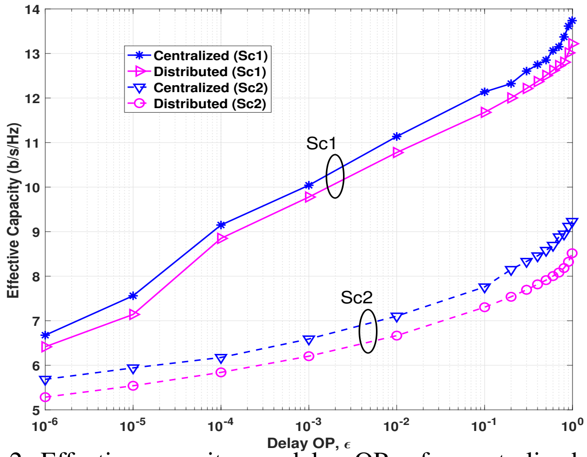

Fig. 2: Effective capacity vs. delay OP ϵ for centralized and distributed schemes.

4) Naive Approach: In this naive approach, we do not apply any power allocation scheme while assuming no involvement of $\mathcal{I}$ in the considered system model. We consider full consumption of the harvested energy for each time interval $k \in \{1, 2, \cdots\}$ and hence, set $P^*(k) = E(k)$. In this *energy-hungry* naive scheme, we do not apply any intelligent way of energy conservation for future time interval and thereby yields very low computational complexity at the expense of performance degradation.

IV. Simulation Results

In this section, we show simulation results for the considered centralized, distributed, and baseline resource control schemes. We set $T = 5$ ms, $\mathcal{W} = 15$ kHz, $\zeta = 0.1$ J, $f_k = k^{-0.8}$, and $D_{\max} = 100$. Furthermore, we consider Rician fading channels with Rician factors $\mathcal{K}_g = \beta_g^2/\sigma_g^2$ and $\mathcal{K}_h = \beta_h^2/\sigma_h^2$ for $g_n(k)$ and $h_n(k)$, respectively and Rayleigh fading channel for $h_d(k)$ for all the results presented in this section. We assume that E follows uniform distribution between 0 and $2\bar{E}$ with average harvested energy $\bar{E}$. We consider 10^6 time intervals all throughout the simulations, where RM can contain no more than 10^3 realizations. DPN and DTN each contains 3 hidden layers with 100 neurons in each layer. We adopt mean square error (MSE) based loss function and Adam optimizer for the considered NNs [15]. We assume that the b-bit phase-shifters are used at $\mathcal{I}$ to control $\psi_n^*(k)$ [1]. Unless otherwise stated, we set $b = 6$ all throughout the simulations. Moreover, we set the learning rate of DRL $f(k) = (1/k)^{0.8}$ while satisfying $\sum_{k=0}^{\infty} f(k) = \infty$ and $\sum_{k=0}^{\infty} f(k)^2 < \infty$ [16].

Centralized vs. Distributed Schemes: In Fig. 2, we show effective capacity (b/s/Hz) of the considered system as a function of delay OP ϵ for centralized and distributed resource

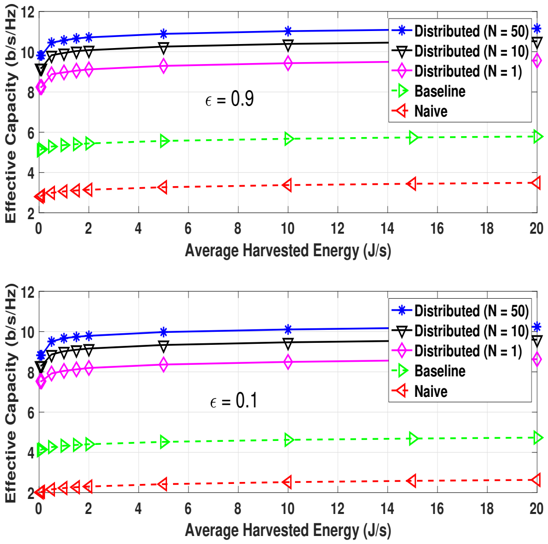

Fig. 3: Effective capacity vs. average harvested energy $\bar{E}$ for distributed (for different N) and baseline schemes. Upper plot: $\epsilon = 0.9$ and lower plot: $\epsilon = 0.1$.

control schemes. We set $\mathcal{K}_g = \mathcal{K}_h = 3$, $\bar{E} = 2$ J, $N = 1$, $d_{\mathcal{SI}} = 50$ m, $d_{\mathcal{ID}} = 30.1$ m, and $d_{\mathcal{SD}} = 20.1$ m. As large N results in high computational complexity for the centralized scheme, we intentionally set $N = 1$ to compare the performances of centralized scheme with the distributed scheme in Fig. 2. We consider two scenarios to show the results. In Scenario 1 (Sc1) and Scenario 2 (Sc2), we set noise variances $\sigma_w^2 = -110$ dBm and $\sigma_w^2 = -70$ dBm, respectively. We first observe that the effective capacity for both centralized and distributed schemes increases with increasing ϵ. Note that this trend of effective capacity is expected as smaller ϵ results in stringent delay QoS constraint (e.g., delay sensitive applications), whereas larger ϵ allows $\mathcal{S}$ to violate delay QoS constraints more often (e.g., delay insensitive applications). We then observe that the distributed scheme performs close to the centralized scheme for the considered range of ϵ in both scenarios. For the rest of the simulation results, we present our results for distributed scheme only.

Performance Analysis of Distributed Scheme: In Fig. 3, we compare the performance of the proposed distributed resource control scheme with the baseline scheme. Note that the baseline scheme does not consider any IRS for communications between $\mathcal{S}$ and $\mathcal{D}$. In particular, we set $\mathcal{K}_g = \mathcal{K}_h = 3$ and $N = \{50, 10, 1\}$ for the distributed scheme to exemplify the role of PREs at $\mathcal{I}$ for the considered EH system in contrast to systems without considering IRS. Moreover, we consider $\sigma_w^2 = -80$ dBm, $d_{\mathcal{SI}} = 50$ m, $d_{\mathcal{ID}} = 20.1$ m, and $d_{\mathcal{SD}} = 70.1$ m. In order to observe the performance of the proposed scheme under two different QoS constraints, we show the results for $\epsilon = 0.9$ and $\epsilon = 0.1$ in upper and lower plots, respectively in Fig. 3. It is evident from Fig. 3 that increasing $\bar{E}$ increases the effective capacity for both delay OP (ϵ) and for all the considered values of N in the distributed and baseline schemes. We observe that increasing N in IRS improves the effective capacity significantly. For instance, in case of $\epsilon = 0.9$ and $\bar{E} = 10$ J/s, considering $N = 10$ and $N = 50$ elements in the distributed scheme increases the effective capacity by 3.5 b/s/Hz and 4.3 b/s/Hz, respectively over the baseline scheme. The performance improvement for the distributed scheme (compared to baseline scheme) is even more significant for a stringent delay constraint. For example, in case of $\epsilon = 0.1$ and $\bar{E} = 10$ J/s, considering $N = 10$ and $N = 50$ elements in the distributed scheme increases the effective capacity by 4.65 b/s/Hz and 5.4 b/s/Hz, respectively over the baseline scheme. For both the considered scenarios, we observe that the naive scheme shows significant degradation in effective capacity compared to all the schemes. The relative poor performance of the naive scheme indicates the importance of dynamic resource allocation for the considered EH-system assisted by IRS module.

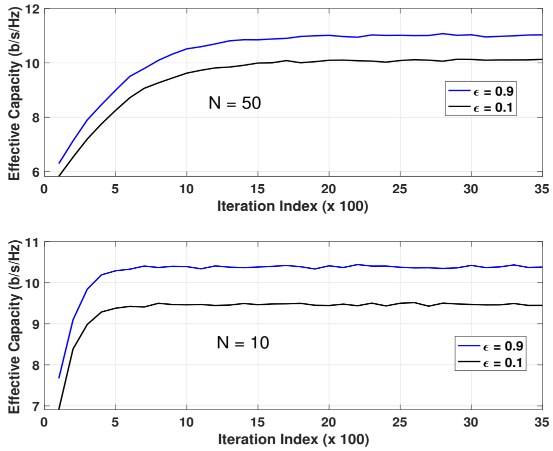

Fig. 4: Convergence behavior of the distributed scheme for $\epsilon = 0.9$ and $\epsilon = 0.1$. Upper plot: $N = 50$ and lower plot: $N = 10$.

In Fig. 4, we show the convergence behavior of the distributed scheme for 50 and 10 PREs with stringent ($\epsilon = 0.1$) and relaxed ($\epsilon = 0.9$) delay constraints. We calculate the moving average of the effective capacity by applying $T_{\text{avg}}(k) = T_{\text{avg}}(k-1) + (1+\gamma_\Psi(k)P(k))^{-\theta}/k$ to demonstrate the convergence behaviors of the distributed scheme. We observe that smaller number of PREs provide relatively faster convergence compared to larger number of PREs. Computer simulations reveal that the proposed algorithm takes approximately 9.7 ms and 2.3 ms on average to converge for $N = 50$ and $N = 10$, respectively[2].

We demonstrate the performance improvement of the proposed distributed resource control scheme over baseline scheme in Fig. 5 for different positions of $\mathcal{D}$, while fixing the locations of $\mathcal{S}$ and $\mathcal{I}$. Moreover, we consider $N = 10$, $\epsilon = 0.5$, $\mathcal{K}_g = \mathcal{K}_h = 0$ to observe the impact of performance gains with non-LOS components (Rayleigh fading) for $\mathcal{S}$-$\mathcal{I}$ and $\mathcal{I}$-$\mathcal{D}$ links. We adopt a metric Δ_{EC}, which is the difference of the effective capacities between the distributed scheme and the baseline scheme. Our objective is to find out the trend of the performance of the distributed scheme when $\mathcal{D}$ moves away from $\mathcal{S}$. We leverage the model similar to [2, Fig. 2], where AP, IRS, and User nodes are replaced by $\mathcal{S}$, $\mathcal{I}$, and $\mathcal{D}$, respectively. Likewise, d_0, $\sqrt{d^2 + d_v^2}$, and $\sqrt{(d_0 - d)^2 + d_v^2}$ in [2, Fig. 2] are replaced by $d_{\mathcal{SI}}$, $d_{\mathcal{SD}}$, and $d_{\mathcal{ID}}$, respectively. We slide $\mathcal{D}$ from left to right along a horizontal straight line and hence change $d_{\mathcal{SD}}$ and $d_{\mathcal{ID}}$ while keeping $d_{\mathcal{SI}}$ constant. In Fig. 5,

[2] All the experiments have been conducted on Apple M1 processor (8-core CPU, 8-core GPU, and 16-core Neural Engine) with 8 GB unified memory and and 512 GB SSD on Python framework.

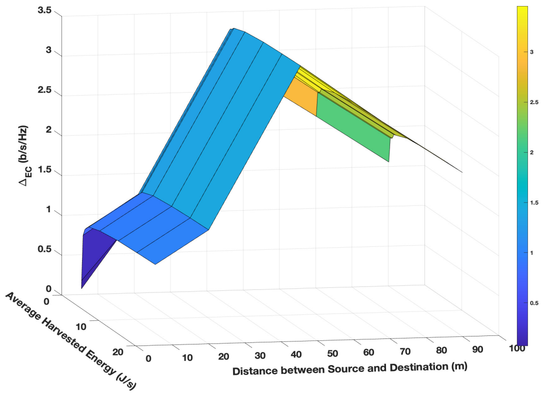

Fig. 5: Performance improvement of distributed scheme over baseline scheme Δ_{EC} as a joint function of average harvested energy $\bar{E}$ and the distance between $\mathcal{S}$ and $\mathcal{D}$.

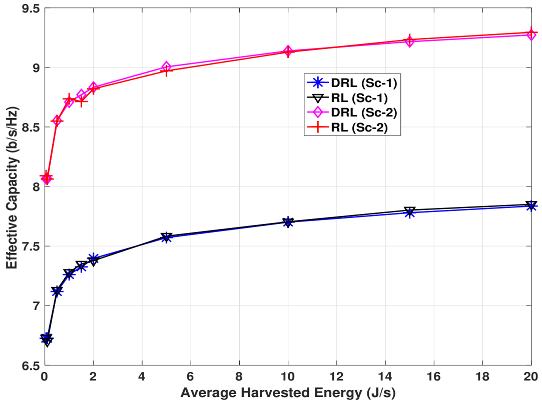

Fig. 6: Effective capacity vs. average harvested energy $\bar{E}$ for distributed schemes with DRL and RL.

we set one of the independent axes (horizontal plane) as $d_{\mathcal{SD}}$ and the other axis as the $\bar{E}$. We observe that when $\mathcal{S}$ and $\mathcal{D}$ (and $\mathcal{I}$ is far from these nodes) are close to each other, Δ_{EC} is relatively small. The performance gain starts increasing when $\mathcal{D}$ starts moving away from $\mathcal{S}$ and goes close to $\mathcal{I}$. Once $\mathcal{D}$ surpasses $\mathcal{I}$, the performance gain starts degrading. Moreover, for fixed values of $d_{\mathcal{SI}}$, $d_{\mathcal{SD}}$, and $d_{\mathcal{ID}}$, Δ_{EC} starts increasing with $\bar{E}$ initially before being saturated at high values of $\bar{E}$.

Fig. 6 compares the effective capacities offered by the proposed DRL-based dynamic resource allocation scheme and by state of the art reinforcement learning (RL)-based scheme. We set $N = 10$ and $\epsilon = 0.5$ while considering two scenarios with different relative distances among $\mathcal{S}$, $\mathcal{I}$, and $\mathcal{D}$. In Scenario 1, we consider $d_{\mathcal{SD}} = 20.1$m and $d_{\mathcal{ID}} = 30.1$m, whereas in Scenario 2, we fix $d_{\mathcal{SD}} = 45.4$m and $d_{\mathcal{ID}} = 5.39$m. For both scenarios, $d_{\mathcal{SI}} = 50$m is assumed. We observe that both DRL and RL schemes yield similar effective capacities as a function of average harvested energy at $\mathcal{S}$. It is worth mentioning that RL-based scheme does not apply deep learning (DL) algorithm to update PDSVF as defined in (19). Instead, RL builds a look-up table, where the state-values are calculated for a large number of discrete values of the optimization variable and discrete levels of the states. DRL learns the PDSVF over time intervals and shows significantly (asymptotic) lower computational complexity once trained. The observations made in Fig. 6 indicates that DRL-based resource allocation scheme can provide the same performance as offered by conventional RL scheme while handling continuous and large-dimensional problem states.

V. Conclusion

In this paper, we demonstrated the effectiveness of deploying IRS in an EH communication system, where the source node is powered by renewable energies. Our developed resource control schemes jointly allocate the transmit power (of the source) and phase-shift of PREs at IRS by applying DRL algorithm that does not require the energy and channel statistics to be known. We demonstrated via simulations the effectiveness of the proposed scheme for an EH system compared to the baseline approach, where IRS is not taken into consideration. The insights obtained from the developed DRL-based dynamic resource allocation scheme for IRS-assisted EH communication system will lead to further investigation of advanced DRL-schemes, e.g., 'noisy'-DRL, 'dueling'-DRL, etc. for a complex and involved network.

References

[1] Q. Wu, S. Zhang, B. Zheng, C. You, and R. Zhang, "Intelligent reflecting surface aided wireless communications: A tutorial," *IEEE Transactions on Communications*, pp. 1–1, 2021.

[2] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Transactions on Wireless Communications*, vol. 18, no. 11, pp. 5394–5409, 2019.

[3] C. Huang, A. Zappone, G. C. Alexandropoulos, M. Debbah, and C. Yuen, "Reconfigurable intelligent surfaces for energy efficiency in wireless communication," *IEEE Transactions on Wireless Communications*, vol. 18, no. 8, pp. 4157–4170, 2019.

[4] M. Jung, W. Saad, and G. Kong, "Performance analysis of large intelligent surfaces (LISs): Uplink spectral efficiency and pilot training," *CoRR*, vol. abs/1904.00453, 2019. [Online]. Available: http://arxiv.org/abs/1904.00453

[5] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with energy harvesting nodes in fading wireless channels: Optimal policies," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1732–1743, September 2011.

[6] D. Wu and R. Negi, "Effective capacity: a wireless link model for support of quality of service," *IEEE Transactions on Wireless Communications*, vol. 2, no. 4, pp. 630–643, 2003.

[7] G. Lee, M. Jung, A. T. Z. Kasgari, W. Saad, and M. Bennis, "Deep reinforcement learning for energy-efficient networking with reconfigurable intelligent surfaces," in *ICC 2020 - 2020 IEEE International Conference on Communications (ICC)*, 2020, pp. 1–6.

[8] S. Zargari, A. Khalili, and R. Zhang, "Energy efficiency maximization via joint active and passive beamforming design for multiuser miso irs-aided swipt," *IEEE Wireless Communications Letters*, vol. 10, no. 3, pp. 557–561, 2020.

[9] D. Xu, X. Yu, V. Jamali, D. W. K. Ng, and R. Schober, "Resource allocation for large irs-assisted swipt systems with non-linear energy harvesting model," in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2021, pp. 1–7.

[10] S. Zargari, A. Khalili, Q. Wu, M. R. Mili, and D. W. K. Ng, "Max-min fair energy-efficient beamforming design for intelligent reflecting surface-aided SWIPT systems with non-linear energy harvesting model," *IEEE Transactions on Vehicular Technology*, 2021.

[11] I. Ahmed, K. T. Phan, and T. Le-Ngoc, "Optimal stochastic power control for energy harvesting systems with statistical delay constraint," in *2015 IEEE Global Communications Conference (GLOBECOM)*, 2015, pp. 1–6.

[12] D. P. Bertsekas, "Dynamic programming and optimal control vol. 1," *Belmont, MA: Athens Scientific*, 1995.

[13] K. Phan, T. Le-Ngoc, M. van der Schaar, and F. Fu, "Optimal scheduling over time-varying channels with traffic admission control: Structural results and online learning algorithms," *IEEE Trans. Wireless Commun.*, vol. 12, no. 9, pp. 4434–4444, September 2013.

[14] E. Smith, C. Pantelides, and G. Reklaitis, "A symbolic reformulation/spatial branch-and-bound algorithm for the global optimization of nonconvex minlp's," *Computers & Chemical Engineering*, vol. 25, pp. 1399–1401, 11 2001.

[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[16] N. Salodkar, A. Bhorkar, A. Karandikar, and V. S. Borkar, "An on-line learning algorithm for energy efficient delay constrained scheduling over a fading channel," *IEEE Journal on Selected Areas in Communications*, vol. 26, no. 4, pp. 732–742, 2008.