Resource Allocation via Model-Free Deep Learning in Free Space Optical Communications

Zhan Gao, *Student Member, IEEE*, Mark Eisen, *Member, IEEE*, and Alejandro Ribeiro, *Member, IEEE*

***Abstract*—This paper investigates the general problem of resource allocation for mitigating channel fading effects in Free Space Optical (FSO) communications. The resource allocation problem is modeled as the constrained stochastic optimization framework, which covers a variety of FSO scenarios involving power adaptation, relay selection and their joint allocation. Under this framework, we propose two algorithms that solve FSO resource allocation problems. We first present the Stochastic Dual Gradient (SDG) algorithm that is shown to solve the problem exactly by exploiting the strong duality but whose implementation necessarily requires explicit and accurate system models. As an alternative we present the Primal-Dual Deep Learning (PDDL) algorithm based on the SDG algorithm, which parameterizes the resource allocation policy with Deep Neural Networks (DNNs) and optimizes the latter via a primal-dual method. The parameterized resource allocation problem incurs only a small loss of optimality due to the strong representational power of DNNs, and can be moreover implemented without knowledge of system models. A wide set of numerical experiments are performed to corroborate the proposed algorithms in FSO resource allocation problems. We demonstrate their superior performance and computational efficiency compared to the baseline methods in both continuous power allocation and binary relay selection settings.**

***Index Terms*—Free space optical communications, resource allocation, primal-dual method, deep learning.**

I. Introduction

FREE Space Optical (FSO) communication has attracted noticeable attention due to high capacity, low cost, strong security and flexible construction [2]. It transmits signals with optical carriers through the atmosphere and has found applications in satellite communications, last-mile access, and fronthaul or backhaul for wireless cellular networks [3]. Despite this potential, FSO communication is susceptible to channel characteristics, such as atmospheric turbulence, weather conditions and background radiation [4]. Different models were proposed to characterize FSO channels, based on which a number of techniques were developed to mitigate channel effects [5]–[11]. Cooperative transmission has recently been introduced as one of such techniques in FSO communications, which improves system performance by leveraging optimal resource allocation [12]. That is, it allocates resources adaptively based on channel state information (CSI) in order to optimize system performance. Common examples of FSO resource allocation include power adaptation, relay selection and their joint allocation.

Power adaptation has emerged as a popular cooperative transmission technique to mitigate channel fading effects. However, conventional methods developed in radio frequency (RF) systems do not apply directly because FSO systems utilize different techniques of modulation, multiplexing and detection, yield different signal models [13], system models [14] and resource requirements [15], and formulate optimization problems with different objectives and constraints; hence, existing RF-based heuristics exhibit unsatisfactory performance in FSO systems. The works in [13], [14], [16] assign adaptive powers to orthogonal optical carriers maximizing the channel capacity with total and peak power constraints. The authors in [17], [18] minimize the outage probability with respective power allocation strategies. Other applications include the security performance [18], the spectral efficiency [19], etc. Relay-assisted communication, on the other hand, employs multiple relay nodes between the transmitter and the receiver to create a virtual multiple-aperture FSO system [20]. However, it is not practical to activate all available relays that requires perfect transmission synchronization. The works in [21], [22] developed relay selection protocols for optimal outage and error probabilities, and the authors in [23] considered both serial and parallel relays to improve system performance. Furthermore, joint power and relay allocation algorithms were developed in FSO networks, in order to maximize the network throughout and minimize the outage probability [15], [24], [25]. The above works formulate resource allocation problems as constrained optimization problems and solve Karush–Kuhn–Tucker (KKT) conditions of the Lagrangian function. However, their performance and transference are restricted by both or one of the following limitations.

L.1 The system models (e.g., the capacity formula, the outage probability, etc.) could be complicated and the resulting KKT conditions could be challenging in FSO systems. Approximate approaches are used to simplify system models or to obtain convex relaxation. The latter results in inexact

Manuscript received April 17, 2021; revised August 21, 2021 and November 9, 2021; accepted November 9, 2021. Date of publication November 18, 2021; date of current version February 17, 2022. An earlier version of this paper was presented at the IEEE GLOBECOM 2019 [DOI: 10.1109/GLOBECOM38437.2019.9014164]. The associate editor coordinating the review of this article and approving it for publication was L. Schmalen. *(Corresponding author: Zhan Gao.)*

Zhan Gao and Alejandro Ribeiro are with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104 USA (e-mail: gaozhan@seas.upenn.edu; aribeiro@seas.upenn.edu).

Mark Eisen is with Intel Corporation, Hillsboro, OR 97124 USA (e-mail: mark.eisen@intel.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TCOMM.2021.3129199.

Digital Object Identifier 10.1109/TCOMM.2021.3129199

solutions degrading performance and/or is computationally expensive.

L.2 The above methods are developed based on explicit closed-form system models (e.g., the capacity formula, the outage probability, etc.). It ties these methods to specific use-cases and does not generalize to system changes, i.e., these methods become inapplicable or require significant modifications when changing FSO systems or application scenarios.

These limitations provide an incomplete solution to the design of generic resource allocation policies in FSO communications, which motivates the application of deep learning due to its low complexity, potential for model-free implementation, and transference to unseen scenarios. Deep learning has been applied for resource allocation in wireless RF communications, which can be cast under three main streams: i) supervised learning; ii) unsupervised learning; iii) reinforcement learning. Supervised learning parameterizes resource allocation policies from labeled datasets. The works in [26], [27] learned optimal power adaptation for MIMO interfering systems via deep neural networks (DNNs) / convolutional neural networks (CNNs) with WMMSE-based datasets, while authors in [28] considered a team of DNNs to solve distributed power adaptation with optimal scheduling solutions used to generate datasets. Similar supervised learning methods are developed for channel assignment [29], user association [30], [31], linear sum assignment problems [32], etc. Nonetheless, the performance of supervised learning methods depends on the quality of labels that may be suboptimal or unavailable. Unsupervised learning solves resource allocation problems without pre-existing labels by considering a system performance metric as the training loss function under which neural networks are trained. For constrained problems, it further penalizes the objective with weighted constraint violation penalties. Applications include power adaptation in D2D communications [33], cognitive radio [34], MIMO interfering systems [35], and relay / user selection in wireless networks [36], [37], millimeter-wave communications [38], etc. However, these methods do not have good guidance for penalty weight selection such that the solution feasibility is not guaranteed, and depend on closed-form system models to evaluate the loss function for training. Reinforcement learning formulates resource allocation problems as Markov decision processes and define the action space, state space and reward function to learn resource allocation policies continuously online [39]–[42]. While it accounts for resource constraints into the state space yielding feasible solutions, its convergence for constrained optimization problems is not guaranteed.

In FSO communications, deep learning has been utilized for assisting channel estimation [43], [44], signal modulation and demodulation [45], [46], etc. Narrowing down to resource allocation, the work in [47] applied the actor-critic method of reinforcement learning for power adaptation in spatial multiplexing FSO systems. However, the considered problem is relatively simple, the approximation approach is used, and the computation could be expensive due to simultaneous training of multiple DNNs. To the best of our knowledge, deep learning and its performance have not yet been explored systematically for general resource allocation problems in FSO systems. Moreover, system models are potentially more complicated or even unavailable in FSO systems with sophisticated modulation techniques (e.g., orbital angular momentum) and detection methods (e.g., non-coherent detection). The latter further motivates to develop efficient and model-free deep learning methods for cooperative transmission in FSO communications.

In this paper we study the application of dual domain optimization and deep learning methods in a wide array of resource allocation problems in FSO communications. Given the objective with a set of constraints, we formulate the FSO resource allocation problem as the constrained stochastic optimization problem and seek an optimal resource allocation policy that adapts to channel state information (Sec. II). To demonstrate the generality of our framework, we exemplify with problems of power adaptation in Radio on FSO systems (Sec. II-A), relay selection in relay-assisted FSO networks (Sec. II-B), and joint power and relay allocation in FSO fronthaul networks (Sec. II-C). Such resource allocation problems are typically challenging due to the non-convexity of complicated objective, existence of constraints, infinite dimensionality of resource allocation policy, and inaccuracy or lack of closed-form system models. We propose the use of dual optimization and learning methods to address these challenges and provide a comprehensive solution methodology. More in detail, our contributions are as follows.

(i) We propose the Stochastic Dual Gradient (SDG) algorithm to overcome the limitation **L.1** (Sec. III). The SDG is demonstrated to solve FSO resource allocation problems exactly by utilizing the strong duality [Thm. 1]. The latter allows us to operate in the dual domain, which is convex, unconstrained and finite dimensional, without loss of optimality. The SDG further saves computational cost by performing primal-dual gradient updates, which avoids computing KKT conditions. Despite the theoretical advantages, this algorithm is limited as it is model-based that requires explicit system models.

(ii) We propose the Primal-Dual Deep Learning (PDDL) algorithm as a model-free, deep learning based alternative to overcome the limitation **L.2** (Sec. IV). The PDDL parameterizes the resource allocation policy with DNNs and reformulates the problem as a constrained machine learning problem. It leverages an approximate strong duality to train DNNs with an unsupervised primal-dual method. A model-free implementation is obtained by leveraging the policy gradient method, which does not require the knowledge of system models. The PDDL further achieves lower complexity due to the computational efficiency of DNNs.

(iii) The overall methodology resulting from both algorithms does not depend on specific systems or problem settings, and thus is applicable comprehensively in FSO communications. We perform extensive numerical experiments in a variety of practical FSO communication scenarios, including power adaptation, relay selection and their joint allocation (Sec. V). In all scenarios, we illustrate success of the proposed algorithms, validating their transference to different system configurations and application scenarios. These results further provide the first comprehensive numerical analysis of the performance of unsupervised deep learning on generic constrained resource allocation problems in FSO communications.

II. Problem Formulation

Consider a general Free Space Optical (FSO) communication system under some form of resource constraints. By adaptively allocating resources using a policy that responds to instantaneous fading effects of the atmospheric channel, we can mitigate these fading effects and optimize the system performance. Denote by $\mathbf{h} \in \mathbb{R}^m$ the collected channel state information (CSI) and $\mathbf{r}(\mathbf{h}) \in \mathbb{R}^n$ a policy that determines the allocated resources based on the observed $\mathbf{h}$. The objective function $f(\mathbf{h}, \mathbf{r}(\mathbf{h}))$ measures the system performance that is instantiated on $\mathbf{h}$ and $\mathbf{r}(\mathbf{h})$. Furthermore, a total of S constraints are imposed either on the resources $\mathbf{r}(\mathbf{h})$ or on the objective function $f(\mathbf{h}, \mathbf{r}(\mathbf{h}))$, each of which is represented by a constraint function $c_s(\mathbf{r}(\mathbf{h}), f(\mathbf{h}, \mathbf{r}(\mathbf{h})))$ for $s = 1, \ldots, S$. The atmospheric channel is typically considered as a fading process with channel coherence time on the order of milliseconds, such that we shall assume $\mathbf{h}$ is drawn from an ergodic and i.i.d block fading process. In this context, the instantaneous system performance tends to vary fast and the long term average performance $\mathbb{E}_{\mathbf{h}}[f(\mathbf{h}, \mathbf{r}(\mathbf{h}))]$ is the more meaningful metric to consider when designing an optimal resource allocation policy. We similarly consider constraints to be satisfied in expectation.

Our goal is to maximize the expected performance $\mathbb{E}_{\mathbf{h}}[f(\mathbf{h}, \mathbf{r}(\mathbf{h}))]$ given certain resource constraints. In particular, we seek to compute the instantaneous allocated resources $\mathbf{r}(\mathbf{h})$ based on the instantaneous CSI $\mathbf{h}$, that satisfy required constraints and optimize the system performance. By introducing $\mathcal{R}$ as the action space of allocated resources $\mathbf{r}(\mathbf{h})$, we formulate the optimal resource allocation as the following stochastic optimization problem

$$\begin{aligned} \mathbb{P} := & \max_{\mathbf{r}(\mathbf{h})} \ \mathbb{E}_{\mathbf{h}}\left[f(\mathbf{h}, \mathbf{r}(\mathbf{h}))\right], \\ \text{s. t. } & \mathbb{E}_{\mathbf{h}}\Big[c_s\big(\mathbf{r}(\mathbf{h}), f(\mathbf{h}, \mathbf{r}(\mathbf{h}))\big)\Big] \leq 0 \text{ for } s = 1, \ldots, S, \ \mathbf{r}(\mathbf{h}) \in \mathcal{R}. \end{aligned} \tag{1}$$

We stress in (1) that the objective function $f(\mathbf{h}, \mathbf{r}(\mathbf{h}))$, the constraint functions $\{c_s\big(\mathbf{r}(\mathbf{h}), f(\mathbf{h}, \mathbf{r}(\mathbf{h}))\big)\}_{s=1}^S$ and the set $\mathcal{R}$ are not necessarily convex depending on specific applications. In fact, in most practical scenarios, they are non-convex given the complexity of FSO systems. In general, the objective function is typically complicated and the allocated resources can be both continuous and discrete, such that solving the resource allocation problem (1) can be difficult. There are mainly four challenges in our concern:

(i) The objective function $f(\mathbf{h}, \mathbf{r}(\mathbf{h}))$ can be extremely complicated in FSO systems, yielding non-convex optimization problems – see Sec. II-A and Sec. II-B.

(ii) The imposed constraints $\{c_s\big(\mathbf{r}(\mathbf{h}), f(\mathbf{h}, \mathbf{r}(\mathbf{h}))\big)\}_{s=1}^S$ are difficult to address, resulting in failures of conventional optimization algorithms – see Sec. II-C.

(iii) The variable to be optimized $\mathbf{r}(\mathbf{h})$ is a function of the channel state information $\mathbf{h}$ and consequently is infinitely dimensional.

(iv) FSO systems are sophisticated due to the complexity of optical equipments. Mathematical models $f(\mathbf{h}, \mathbf{r}(\mathbf{h}))$ characterizing these systems may be unknown or inaccurate such that model-based algorithms are inapplicable or suffer from inevitable degradations.

In what follows, we first propose a model-based algorithm that solves (1) exactly without any approximation (Sec. III). We proceed to develop a model-free algorithm via deep learning that solves (1) with only system observations, where the knowledge of system models is not required (Sec. IV). Before proceeding, we illustrate in the following subsections how the general problem framework (1) represents a variety of resource allocation problems in FSO communications.

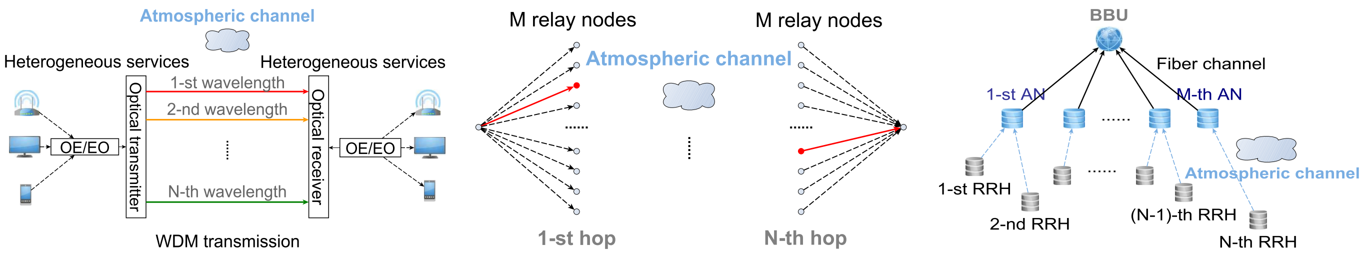
A. Power Adaptation

We consider power adaptation in a Radio on Free Space Optics (RoFSO) system [14]. As a universal platform for heterogeneous wireless services, it transmits RF signals through FSO links in optical networks. The Wavelength Division Multiplexing (WDM) RoFSO system allows simultaneous transmissions of multiple signals to increase the transmission capacity. In particular, multimedia RF signals are accessed into the RoFSO system, placed on multiple optical wavelength carriers with optoelectronic devices, and transmitted through FSO channels. At the receiver, optical signals are received and transferred back to RF signals for users – see Fig. 1a. Orthogonal optical carriers are distributed in a huge bandwidth and are non-overlapping with adequate spacing to avoid interference, which could experience different channel fading effects, e.g., frequency-dependent path loss, frequency selective fading and atmospheric turbulence with coherence bandwidth smaller than channel spacing.[1] Based on the CSI, adaptive powers are assigned to different wavelengths to maximize the total channel capacity. Assume that there are N optical wavelength carriers. The CSI is represented by the vector $\mathbf{h} = [h_1, \ldots, h_N]^\top \in \mathbb{R}^N$, where h_i is the CSI of i-th wavelength channel. The allocated power to signal transmitted on i-th wavelength is based upon the observed CSI $\mathbf{h}$ via a power allocation policy $p_i(\mathbf{h})$. Given the collection of power allocations $\mathbf{r}(\mathbf{h}) = [p_1(\mathbf{h}), \ldots, p_N(\mathbf{h})]^\top \in \mathbb{R}^N$ and the CSI $\mathbf{h}$, the channel capacity $C_i(\mathbf{h}, \mathbf{r}(\mathbf{h}))$ achieved on i-th wavelength is [14]

$$\begin{aligned} C_i(\mathbf{h}, \mathbf{r}(\mathbf{h})) &= C_i(h_i, p_i(\mathbf{h})) \\ &= \log\!\left(1 + \frac{\frac{1}{2}(OMI \cdot m_p r p_i(\mathbf{h}) h_i)^2}{RIN \cdot (r p_i(\mathbf{h}) h_i)^2 + 2e m_p^{2+F} r p_i(\mathbf{h}) h_i + \frac{4KT}{R_f}}\right) \end{aligned} \tag{2}$$

with OMI the optical modulation index, RIN the relative intensity noise, m_p the photodiode gain, r the photodiode responsivity, e the electric charge, F the excess noise factor,

[1] While in some weather conditions the channel effect may not change significantly with wavelength, considering the huge bandwidth that the RoFSO system can support, it is still non-trivial to study the problem of power adaptation.



(a) (b) (c)

Fig. 1. (a) WDM RoFSO system with N optical wavelength channels. (b) Relay-assisted FSO network. The transmitter communicates with the receiver through N selected relays (red nodes). (c) Fronthaul FSO network with N RRHs and M ANs.

K the Boltzmann's constant, T the temperature and R_f the photodiode resistance. We consider the weight vector $\boldsymbol{\omega}=[\omega_1,\dots,\omega_N]^\top\in\mathbb{R}^N$ to represent priorities of different wireless services, and the objective function is the weighted sum of channel capacities over N wavelengths

$$\mathbb{E}_{\mathbf{h}}\left[f(\mathbf{h},\mathbf{r}(\mathbf{h}))\right]=\sum_{i=1}^{N}\omega_i\mathbb{E}_{\mathbf{h}}\left[C_i(\mathbf{h},\mathbf{r}(\mathbf{h}))\right]. \qquad (3)$$

The RoFSO system is constrained by a total power limitation P_t at the base station, i.e., $\mathbb{E}_{\mathbf{h}}\left[c(\mathbf{r}(\mathbf{h}))\right]=\mathbb{E}_{\mathbf{h}}\big[\sum_{i=1}^{N}p_i(\mathbf{h})\big]-P_t\leq 0$, and a peak power limitation P_s for each carrier to ensure eye safety, i.e., $\mathcal{R}=[0,P_s]^N$.

B. Relay Selection

We consider the relay-assisted FSO network, in which the transmitter communicates with the receiver through intermediate hops [48]. In particular, assume that there are N hops where each hop consists of M parallel relays. The transmitter sends the optical signal to a selected relay at 1-st hop. The latter amplifies the received signal and then transmits it to a selected relay at 2-nd hop. The process performs successfully through N hops until the receiver – See Fig. 1b. Based on the CSI, different relays are selected at different hops to maximize the channel capacity. Note that there is only one relay activated at each hop. This is because activating multiple relays requires perfect synchronization and phase noise calibration, both of which are cumbersome tasks especially for FSO networks with high rates and large number of relays. We denote by $\mathbf{h}\in\mathbb{R}^{(M\times(N-1)+2)\times M}$ the CSI between the transmitter, relays and the receiver, and the matrix $\mathbf{r}(\mathbf{h})=[\boldsymbol{\alpha}_1(\mathbf{h}),\dots,\boldsymbol{\alpha}_N(\mathbf{h})]^\top\in\{0,1\}^{N\times M}$ the selected relays, where each $\boldsymbol{\alpha}_i(\mathbf{h})=[\alpha_{i1}(\mathbf{h}),\dots,\alpha_{iM}(\mathbf{h})]^\top\in\{0,1\}^M$ is a M-dimensional vector with $\alpha_{ij}(\mathbf{h})=1$ if j-th relay is selected at i-th hop and $\alpha_{ij}(\mathbf{h})=0$ otherwise. The relay-assisted channel capacity is [15]

$$C_{j_1\dots j_N}(\mathbf{h})=\frac{T_fB}{\epsilon}\log\bigg(1+\bigg(\prod_{i=0}^{N}\Big(1+\frac{1}{Ph_{j_ij_{i+1}}\frac{R}{e\Delta f}}\Big)-1\bigg)^{-1}\bigg) \qquad (4)$$

which assumes that j_i-th relay is selected at i-th hop and $h_{j_ij_{i+1}}$ is the CSI between j_i-th relay at i-th hop and j_{i+1}-th relay at $(i+1)$-th hop, where $j_0=j_{N+1}=1$ represent the transmitter and the receiver. Here, T_f is the frame duration, B the bandwidth, P the transmission power, R the photodetector sensitivity, e the electric charge, Δf the noise equivalent bandwidth, $\epsilon=1$ for the full-duplex relay and $\epsilon=2$ for the half-duplex relay. The objective function is then given by

$$\mathbb{E}_{\mathbf{h}}[f(\mathbf{h},\mathbf{r}(\mathbf{h}))]=\mathbb{E}_{\mathbf{h}}\Big[\sum_{j_N=1}^{M}\cdot\cdot\sum_{j_1=1}^{M}\Big(\prod_{i=1}^{N}\alpha_{ij_i}(\mathbf{h})\Big)C_{j_1\dots j_N}(\mathbf{h})\Big]. \qquad (5)$$

There are N constraints on the selected relays $\mathbf{r}(\mathbf{h})$. That is only one relay can be selected at each hop, i.e., $\mathcal{R}=\big\{\{0,1\}^{N\times M}|\sum_{j=1}^{M}\alpha_{ij}(\mathbf{h})\leq 1,\text{for } i=1,\dots,N\big\}$.

C. Joint Power and Relay Allocation

The problem becomes more complicated when considering joint power and relay allocation, as seen in the FSO fronthaul network [15], [25]. As one of cloud radio access network (C-RAN) architectures, it provides high rates, low latency and flexible constructions for 5G wireless networks. In particular, the system consists of remote radio heads (RRHs), aggregation nodes (ANs) and the baseband unit (BBU). The RRHs transmit optical signals with orthogonal optical carriers through free space to the selected ANs. The latter collect received signals and then forward the aggregated signal to the BBU through high speed optical fiber – See Fig. 1c. Based on the CSI, different ANs are selected at different RRHs and adaptive powers are assigned to different optical carriers at each RRH. Assume there are L optical carriers, N RRHs, M ANs and one BBU. The CSI is represented by $\mathbf{h}=\{\mathbf{h}_{ij}\}_{ij}$ for $i=1,\dots,N$ and $j=1,\dots,M$, where each vector $\mathbf{h}_{ij}=[h_{ij}^1,\dots,h_{ij}^L]^\top\in\mathbb{R}^L$ is the CSI of L carriers between i-th RRH and j-th AN. The allocated resources $\mathbf{r}(\mathbf{h})=\{\mathbf{p}_{ij}(\mathbf{h}),\alpha_{ij}(\mathbf{h})\}_{ij}$ contain assigned powers and selected ANs, where $\mathbf{p}_{ij}(\mathbf{h})=[p_{ij}^1(\mathbf{h}),\dots,p_{ij}^L(\mathbf{h})]^\top\in\mathbb{R}^L$ are powers assigned to L carriers in the link between i-th RRH and j-th AN, and $\alpha_{ij}(\mathbf{h})\in\{0,1\}$ is the indicator being one if j-th AN is selected at i-th RRH and zero otherwise. The channel capacity between i-th RRH and j-th AN is [25]

$$C_{ij}(\mathbf{h},\mathbf{r}(\mathbf{h}))=\sum_{\ell=1}^{L}\omega_\ell\frac{T_fB}{\epsilon}\log\left(1+p_{ij}^{\ell}(\mathbf{h})h_{ij}^{\ell}\frac{R}{e\Delta f}\right) \qquad (6)$$

with $\boldsymbol{\omega}=[\omega_1,\dots,\omega_L]^\top\in\mathbb{R}^L$ the priorities of carriers, T_f the frame duration, B the bandwidth, R the photodetector sensitivity, e the electric charge and Δf the noise equivalent bandwidth. There is no interference between orthogonal carriers because optical signals are highly directional with very narrow beam divergence [49]. The objective function is the sum-capacity over N RRHs

$$\mathbb{E}_{\mathbf{h}}\left[f(\mathbf{h},\mathbf{r}(\mathbf{h}))\right]=\mathbb{E}_{\mathbf{h}}\Big[\sum_{i=1}^{N}\sum_{j=1}^{M}\alpha_{ij}(\mathbf{h})C_{ij}(\mathbf{h},\mathbf{r}(\mathbf{h}))\Big]. \tag{7}$$

The system is constrained by: (i) the total power limitation P_t and the peak power limitation P_s at each RRH as in Sec. II-A; (ii) only one AN can be selected at each RRH as in Sec. II-B; (iii) the data congestion constraints at each AN, i.e., the aggregated data traffic shall not exceed the maximal capacity C_t of optical fiber at each AN to avoid data congestion. Therefore, we have

$$\mathbb{E}_{\mathbf{h}}\Big[\sum_{\ell=1}^{L}p_{ij}^{\ell}(\mathbf{h})\Big]-P_t\leq 0,\ \text{ for } i=1,\ldots,N, j=1,\ldots,M, \tag{8a}$$

$$\mathbb{E}_{\mathbf{h}}\Big[\sum_{i=1}^{N}\alpha_{ij}(\mathbf{h})C_{ij}(\mathbf{h},\mathbf{r}(\mathbf{h}))\Big]-C_t\leq 0,\ \text{ for } j=1,\ldots,M, \tag{8b}$$

$$\mathcal{R}=\Big\{[0,P_s]^{N\times M\times L}\times\{0,1\}^{N\times M}|\sum_{j=1}^{M}\alpha_{ij}(\mathbf{h})\leq 1,\text{for } i=1,\ldots,N\Big\}. \tag{8c}$$

III. Stochastic Dual Gradient Algorithm

In this section, we first address three primary challenges (i)-(iii) outlined in Sec. II by working in the dual domain. In particular, by establishing a null duality gap for (1), we present the Stochastic Dual Gradient (SDG) algorithm that finds exact solutions despite the non-convexity, constraints, and infinite dimensionality. For developing the SDG algorithm, we initially ignore challenge (iv) and assume mathematical models derived for FSO systems are given and accurate. For instance, we assume the capacity function $C_i(\mathbf{h},\mathbf{r}(\mathbf{h}))$ in (2) is accurate for RoFSO systems.

With a set of constraints, it is natural to consider working in the dual domain. By introducing the dual variables $\boldsymbol{\lambda}=[\lambda_1,\ldots,\lambda_S]^\top\in\mathbb{R}_+^S$ that correspond to S constraints, the *Lagrangian* of problem (1) is given by $\mathcal{L}(\mathbf{r}(\mathbf{h}),\boldsymbol{\lambda}):=\mathbb{E}_{\mathbf{h}}[f(\mathbf{h},\mathbf{r}(\mathbf{h}))]-\sum_{s=1}^{S}\lambda_s\mathbb{E}_{\mathbf{h}}[c_s(\mathbf{r}(\mathbf{h}),f(\mathbf{h},\mathbf{r}(\mathbf{h})))]$. Each constraint in (1) shows as a penalty in $\mathcal{L}(\mathbf{r}(\mathbf{h}),\boldsymbol{\lambda})$, where the violation is penalized (weighted by a dual variable). We define the *dual function* as the maximum of Lagrangian $\mathcal{D}(\boldsymbol{\lambda}):=\max_{\mathbf{r}(\mathbf{h})\in\mathcal{R}}\mathcal{L}(\mathbf{r}(\mathbf{h}),\boldsymbol{\lambda})$, which is unconstrained such that conventional optimization algorithms can be used. With dual variables involved, it is proved that $\mathcal{D}(\boldsymbol{\lambda})\geq\mathbb{P}$ for any $\boldsymbol{\lambda}$. This result motivates the development of dual problem, that is to find $\boldsymbol{\lambda}^*$ minimizing the dual function as

$$\mathbb{D}:=\min_{\boldsymbol{\lambda}\geq 0}\mathcal{D}(\boldsymbol{\lambda})=\min_{\boldsymbol{\lambda}\geq 0}\max_{\mathbf{r}(\mathbf{h})\in\mathcal{R}}\mathcal{L}(\mathbf{r}(\mathbf{h}),\boldsymbol{\lambda}). \tag{9}$$

The optimal solution $\mathbb{D}$ for (9) can be viewed as the best approximation of $\mathbb{P}$ when handling constraints as penalties. The dual method has been applied for convex optimization problems with a set of convex constraints, where the difference between $\mathbb{D}$ and $\mathbb{P}$, referred to as the duality gap, is zero. We can then solve the optimization problem by solving its associated dual problem without loss of optimality. However, resource allocation problems (1) in FSO systems are rarely convex due to complicated objective / constraint functions and the application of the dual method is not immediate in this context. It is still unclear how much the difference between $\mathbb{D}$ and $\mathbb{P}$ is in non-convex FSO scenarios. We consider this issue in the following subsection.

A. Null Duality Gap

Despite the possible non-convexity, we show in the following theorem that the null duality gap does hold in FSO resource allocation problems (1).

Theorem 1: Consider the stochastic optimization problem (1) and its associated dual problem (9). Let $\mathbb{P}$ and $\mathbb{D}$ be the optimal solutions of (1) and (9). Assume that there exists a feasible point $\mathbf{r}_0$ satisfying all constraints with strict inequality, then the strong duality holds that $\mathbb{P}=\mathbb{D}$.

Proof: The objective and constraints in (1) are in expectation w.r.t. the CSI $\mathbf{h}$, which is instantiated from the probability distribution $m(\mathbf{h})$. We can consider (1) as a particular realization of the sparse functional program [50]. Since the distribution $m(\mathbf{h})$ that characterizes the FSO channel is continuous, $\mathbf{h}$ takes values in a dense set of the domain. Considering this observation together with the assumption that there exists a solution satisfying all constraints with strict inequality, the results of Theorem 1 in [50] claim the strong duality gap $\mathbb{P}=\mathbb{D}$. □

Theorem 1 states that the resource allocation problem under the framework of (1) has the null duality gap $\mathbb{D}-\mathbb{P}=0$ even if it is non-convex. This indicates that we can solve (1) by solving the unconstrained dual problem (9) alternatively without loss of optimality and thus, justifies the application of the dual method in FSO scenarios. We then develop an algorithm to solve the alternative min-max problem (9).

B. Primal-Dual Update

We propose the SDG algorithm based on the above analysis, which iteratively searches for the optimal dual variables $\boldsymbol{\lambda}^*$ and derives the corresponding optimal resource allocation policy $\mathbf{r}^*(\mathbf{h})$. To be more precise, the SDG consists of two steps over an iteration index k. The primal step updates the primal variables $\mathbf{r}(\mathbf{h})$ given the current dual variables $\boldsymbol{\lambda}^k$, while the dual step updates the dual variables $\boldsymbol{\lambda}$ given the updated $\mathbf{r}^{k+1}(\mathbf{h})$. Details are formally introduced below.

1) Primal Step: At k-th iteration given the dual variables $\boldsymbol{\lambda}^k$ and the CSI $\mathbf{h}$, we update the primal variables by maximizing the Lagrangian as

$$\begin{aligned}\mathbf{r}^{k+1}(\mathbf{h})&=\operatorname*{argmax}_{\mathbf{r}(\mathbf{h})\in\mathcal{R}}\mathcal{L}\big(\mathbf{r}(\mathbf{h}),\boldsymbol{\lambda}^k\big)\\&=\operatorname*{argmax}_{\mathbf{r}(\mathbf{h})\in\mathcal{R}}f(\mathbf{h},\mathbf{r}(\mathbf{h}))-\sum_{s=1}^{S}\lambda_s^k c_s(\mathbf{r}(\mathbf{h}),f(\mathbf{h},\mathbf{r}(\mathbf{h})))\end{aligned} \tag{10}$$

where the last equality is because the expectation is automatically maximized if it is maximized at each sample $\mathbf{h}$.

Algorithm 1 Stochastic Dual Gradient Algorithm

```
Input: The objective function $f(\mathbf{h}, \mathbf{r}(\mathbf{h}))$, the constraints $\{c_s(\mathbf{r}(\mathbf{h}), f(\mathbf{h}, \mathbf{r}(\mathbf{h})))\}_{s=1}^S$, the CSI $\mathbf{h}$ and the initial dual variables $\boldsymbol{\lambda}^0$
for $k = 0, 1, 2, \ldots$ do {main loop}
    Update the primal variables $\mathbf{r}^{k+1}(\mathbf{h}) = \text{argmax}_{\mathbf{r}(\mathbf{h}) \in \mathcal{R}} \mathcal{L}(\mathbf{r}(\mathbf{h}), \boldsymbol{\lambda}^k)$ [cf. (10)]
    Update the dual variables $\lambda_s^{k+1} = \big[\lambda_s^k + \eta^k \sum_{\tau=1}^{\mathcal{T}} c_s(\mathbf{r}^{k+1}(\mathbf{h}_\tau), f(\mathbf{h}_\tau, \mathbf{r}^{k+1}(\mathbf{h}_\tau)))/\mathcal{T}\big]_+$ [cf. (11)]
end for
```

In practice, (10) can usually be simplified based on specific system models. For example, in the RoFSO system, both the objective and the constraint separate the use of components $p_1(\mathbf{h}), \ldots, p_N(\mathbf{h})$ in $\mathbf{r}(\mathbf{h})$ and $h_1, \ldots, h_N$ in $\mathbf{h}$ with no coupling between them. In this context, solving (10) is equivalent to solving N scalar sub-problems that update each component $p_i(\mathbf{h})$ separately as $p_i^{k+1}(\mathbf{h}) = \text{argmax}_{p_i(\mathbf{h}) \in [0, P_S]} \omega_i C_i(h_i, p_i(\mathbf{h})) - \lambda^k p_i(\mathbf{h})$ for $i = 1, \ldots, N$.

2) Dual Step: Given the updated $\mathbf{r}^{k+1}(\mathbf{h})$ from the primal step (10), we perform the dual gradient descent to update $\boldsymbol{\lambda}^k$ as

$$\begin{aligned} \lambda_s^{k+1} &= \Big[\lambda_s^k - \eta^k \nabla_{\lambda_s} \mathcal{L}(\mathbf{r}^{k+1}(\mathbf{h}), \boldsymbol{\lambda}^k)\Big]_+ \\ &= \big[\lambda_s^k + \eta^k \mathbb{E}_{\mathbf{h}}\big[c_s\left(\mathbf{r}^{k+1}(\mathbf{h}), f(\mathbf{h}, \mathbf{r}^{k+1}(\mathbf{h}))\right)\big]\big]_+, \\ &\quad \text{for } s = 1, \ldots, S \end{aligned} \tag{11}$$

where η^k is the dual step-size and $[\cdot]_+ = \max(\cdot, 0)$ is due to the non-negativity of $\boldsymbol{\lambda}$. We estimate the expectation $\mathbb{E}_{\mathbf{h}}[\cdot]$ empirically, i.e., we sample $\mathcal{T}$ CSI realizations $\{\mathbf{h}_\tau\}_{\tau=1}^{\mathcal{T}}$ and take the average as $\sum_{\tau=1}^{\mathcal{T}} c_s\big(\mathbf{r}^{k+1}(\mathbf{h}_\tau), f(\mathbf{h}_\tau, \mathbf{r}^{k+1}(\mathbf{h}_\tau))\big)/\mathcal{T}$.

By repeating these two steps recursively, $\boldsymbol{\lambda}^k$ converges to the optimal $\boldsymbol{\lambda}^*$ as k increases [51]. Due to the null duality gap, the optimal solution $\mathbf{r}^*(\mathbf{h})$ can be obtained from $\boldsymbol{\lambda}^*$ as

$$\mathbf{r}^*(\mathbf{h}) = \underset{\mathbf{r}(\mathbf{h}) \in \mathcal{R}}{\text{argmax}}\, f(\mathbf{h}, \mathbf{r}(\mathbf{h})) - \sum_{s=1}^{S} \lambda_s^* c_s\big(\mathbf{r}(\mathbf{h}), f(\mathbf{h}, \mathbf{r}(\mathbf{h}))\big). \tag{12}$$

Algorithm 1 summarizes the SDG algorithm.

With accurate system models, the SDG algorithm solves the problem (1) perfectly in theory with no relaxation or approximation. However, there exist several problems w.r.t. its practical implementation. For one thing, in the primal step of the SDG, there is no closed-form solution of (10) to compute optimal $\mathbf{r}^{k+1}(\mathbf{h})$. Similarly, even after the algorithm converges, real time execution of $\mathbf{r}^*(\mathbf{h})$ needs to numerically solve (12). Therefore, it may require certain computational complexity. For another, we recall the difficulty of obtaining accurate system models in FSO systems as stated in challenge (iv) of Sec. II. The SDG heavily relies on system models, i.e., we need accurate objective function $f(\mathbf{h}, \mathbf{r}(\mathbf{h}))$ and constraint functions $c_s\big(\mathbf{r}(\mathbf{h}), f(\mathbf{h}, \mathbf{r}(\mathbf{h}))\big)$ to perform algorithm. These may not be available given a FSO system that is new, unfamiliar or complex. Furthermore, existing models do not always capture physical behaviors of real-world systems in practice leading to inevitable modeling errors. These motivate to develop a low-complexity and model-free learning-based algorithm for FSO resource allocation problems.

Remark 1: The Lyapunov drift-plus-penalty method is an alternative model-based method to the dual domain approach that is also potentially applicable for FSO resource allocation problems (1) [52]. It converts the feasibility of constraints to the stability of virtual queues, translates the latter to the minimization of Lyapunov drift, and penalizes the Lyapunov drift with the objective scaled by a weight V. However, there exist certain limitations: (i) the value of V is pre-determined representing a trade-off between the objective and constraints but must be hand tuned; (ii) it solves problems based on dynamically evolving queue states in addition to instantaneous CSI and thus requires gathering samples online, i.e., it needs to be implemented online that may be time inefficient; (iii) it also requires explicit system models to minimize the penalized Lyapunov drift.

IV. Primal-Dual Deep Learning Algorithm

To handle above limitations, we develop the model-free Primal-Dual Deep Learning (PDDL) algorithm based on the SDG algorithm. The implementation of the PDDL requires only observed values of the FSO system (e.g., the observed channel capacity and CSI) instead of mathematical system models. We begin by noticing the problem (1) shares the same structure as the statistical learning problem. The latter inspires us to introduce a parameterization $\boldsymbol{\theta} \in \mathbb{R}^q$ to represent the resource allocation policy as $\mathbf{r}(\mathbf{h}) = \boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta})$. Substituting this representation into (1) yields

$$\begin{aligned} \mathbb{P}_{\boldsymbol{\theta}} &:= \max_{\boldsymbol{\theta}}\ \mathbb{E}_{\mathbf{h}}\left[f(\mathbf{h}, \boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta}))\right], \\ &\text{s. t. } \mathbb{E}_{\mathbf{h}}\left[c_s\big(\boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta}), f(\mathbf{h}, \boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta}))\big)\right] \leq 0 \\ &\qquad \text{for } s = 1, \ldots, S,\ \boldsymbol{\theta} \in \Theta \end{aligned} \tag{13}$$

where Θ is the parameterization set satisfying $\boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta}) \in \mathcal{R}$. Then, the goal becomes to learn the optimal function $\boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta}^*)$ by finding the optimal parameterization $\boldsymbol{\theta}^*$ that maximizes the objective while satisfying prescribed constraints.

A. Near-Universal Parameterization

The parameterization in (13) inevitably introduces a loss of optimality since resource allocation functions are restricted to those adhered to the form of $\mathbf{r}(\mathbf{h}) = \boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta})$. For example, a linear parameterization $\boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta}) = \boldsymbol{\theta}^\top \mathbf{h}$ can never represent any nonlinear resource allocation policy. A good choice of $\boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta})$ should provide an accurate approximation for almost all functions in $\mathcal{R}$ and thus can model the space of allowable resource allocation policies. To quantify such representational capacity, we define the near-universal parameterization as follows.

Definition 1 (Near-Universal Parameterization): For any $\epsilon \geq 0$, the parameterization $\boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta})$ is ϵ-universal if for any $\mathbf{r}(\mathbf{h}) \in \mathcal{R}$, there exists a set of parameters $\boldsymbol{\theta} \in \Theta$ such that

$$\mathbb{E}_{\mathbf{h}}\left[\|\mathbf{r}(\mathbf{h}) - \boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta})\|_\infty\right] \leq \epsilon. \tag{14}$$

Deep Neural Networks (DNNs) are well-suited candidates that exhibit the universal property and achieve successes in various practical problems [53]. In particular, DNNs are information processing architectures consisting of L layers, each of which comprises a linear operation $\mathbf{W}_\ell \in \mathbb{R}^{n_\ell \times n_{\ell-1}}$ and a pointwise nonlinearity $\sigma_\ell(\cdot)$ for $\ell = 1, \ldots, L$. Here, n_ℓ is the number of hidden units at layer ℓ, the nonlinearity $\sigma_\ell(\cdot)$ could be the absolute value, the ReLU, the sigmoid function, etc., and the parameterization $\boldsymbol{\theta} \in \mathbb{R}^q$ are the weights of linear operations $\mathbf{W}_1, \ldots, \mathbf{W}_L$ with $q = \sum_{\ell=0}^{L-1} n_\ell n_{\ell+1}$. We then verify its near-universal property as follows.

Theorem 2 [53, Theorem 2.2]: Let $m(\mathbf{h})$ *be the distribution of the CSI* $\mathbf{h}$ *and* $\mathcal{R}$ *be the considered set of measurable functions. For a DNN with arbitrarily large number of layers and arbitrarily large number of hidden units per layer, it is dense in probability in* $\mathcal{R}$*, i.e., for any function* $\mathbf{r}(\mathbf{h}) \in \mathcal{R}$ *and* $\epsilon > 0$*, there exists* L*,* $\{n_0, n_1, \ldots, n_L\}$ *and* $\boldsymbol{\theta} \in \mathbb{R}^q$ *such that*

$$m\big(\{\mathbf{h} : \|\boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta}) - \mathbf{r}(\mathbf{h})\|_\infty > \epsilon\}\big) < \epsilon. \tag{15}$$

Theorem 2 states that DNNs can approximate functions in the considered set with arbitrarily small error ϵ by increasing the number of layers L and layer sizes $\{n_\ell\}_{\ell=0}^L$. Therefore, the parameterization loss $\mathbb{P} - \mathbb{P}_{\boldsymbol{\theta}}$ can be sufficiently small by learning with DNNs properly.

B. Primal-Dual Learning

We now develop an analogous dual-domain learning method to find the optimal parameterization $\boldsymbol{\theta}^*$. Similar as the unparameterized problem, we begin by formulating the Lagrangian of (13) as $\mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda}) := \mathbb{E}_{\mathbf{h}}[f(\mathbf{h}, \boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta}))] - \sum_{s=1}^S \lambda_s \mathbb{E}_{\mathbf{h}}\big[c_s\big(\boldsymbol{\Phi}(\mathbf{h},\boldsymbol{\theta}), f(\mathbf{h}, \boldsymbol{\Phi}(\mathbf{h},\boldsymbol{\theta}))\big)\big]$. The corresponding dual problem is subsequently defined as

$$\mathbb{D}_{\boldsymbol{\theta}} := \min_{\boldsymbol{\lambda} \geq 0} \mathcal{D}_{\boldsymbol{\theta}}(\boldsymbol{\lambda}) = \min_{\boldsymbol{\lambda} \geq \mathbf{0}} \max_{\boldsymbol{\theta} \in \Theta} \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda}). \tag{16}$$

For the above min-max problem with the parameterization $\boldsymbol{\theta}$, the duality gap $\mathbb{P}_{\boldsymbol{\theta}} - \mathbb{D}_{\boldsymbol{\theta}}$ can be assumed sufficiently small due to the strong duality in Theorem 1 and the near-universal property of the DNN in Theorem 2 [40]. Therefore, we can solve (13) by solving (16) with little loss of optimality.

We similarly develop the PDDL algorithm for solving (16), which updates the primal variables $\boldsymbol{\theta}$ with gradient ascent and the dual variables $\boldsymbol{\lambda}$ with gradient descent at each iteration k:

1) Primal Step: Given the dual variables $\boldsymbol{\lambda}^k$, we update the primal variables $\boldsymbol{\theta}$ as

$$\begin{aligned} \boldsymbol{\theta}^{k+1} &= \boldsymbol{\theta}^k + \delta^k \nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}^k, \boldsymbol{\lambda}^k) \\ &= \boldsymbol{\theta}^k + \delta^k \nabla_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{h}}\Big[f(\mathbf{h}, \boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta})) \\ &\quad - \sum_{s=1}^S \lambda_s c_s\big(\boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta}), f(\mathbf{h}, \boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta}))\big)\Big] \end{aligned} \tag{17}$$

where δ^k is the primal step-size, and the last equation is due to the linearity of the expectation.

2) Dual Step: Given the updated $\boldsymbol{\theta}^{k+1}$ from the primal step (17), the dual variables $\boldsymbol{\lambda}$ is updated as

$$\lambda_s^{k+1} = \Big[\lambda_s^k + \eta^k \mathbb{E}_{\mathbf{h}}\big[c_s\big(\boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta}^{k+1}), f(\mathbf{h}, \boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta}^{k+1}))\big)\big]\Big]_+ \tag{18}$$

for $s = 1, \ldots, S$, where η^k is the dual step-size. We similarly estimate the expectation $\mathbb{E}_{\mathbf{h}}[\cdot]$ with its empirical alternative $\sum_{\tau=1}^{\mathcal{T}} c_s\big(\boldsymbol{\Phi}(\mathbf{h}_\tau, \boldsymbol{\theta}^{k+1}), f(\mathbf{h}_\tau, \boldsymbol{\Phi}(\mathbf{h}_\tau, \boldsymbol{\theta}^{k+1}))\big)/\mathcal{T}$.

The PDDL algorithm learns the optimal primal and dual variables $\boldsymbol{\theta}^*$ and $\boldsymbol{\lambda}^*$ by recursively repeating primal and dual steps. The primal-dual method used in the parameterized problem features a closed-form update in (17), in contrast to the computationally expensive inner maximization required in (10) of the SDG algorithm used in the unparameterized problem. Even still, direct evaluation of the primal update in (17) requires the knowledge of system models to compute the expected gradients, generally not available in practice. However, unlike the SDG algorithm, the PPDL algorithm is capable of leveraging the so-called policy gradient method to develop a completely model-free implementation.

C. Model-Free Policy Gradient

Policy gradient has been developed as a practical gradient estimation method in reinforcement learning because it avoids explicit modeling of the objective function $f(\cdot)$ and constraint functions $\{c_s(\cdot)\}_{s=1}^S$. It exploits a likelihood ratio property to compute the gradient for policy functions taking the form of $\mathbb{E}_{\mathbf{h}}[f(\mathbf{h}, \boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta}))]$, where $f(\cdot)$ is unknown. Put simply, it provides a stochastic and model-free approximation for $\nabla_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{h}}[f(\mathbf{h}, \boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta}))]$ [54].

In particular, we consider the policy parameterization $\boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta})$ as stochastic realizations drawn from a distribution with the delta density function $\pi_{\mathbf{h},\boldsymbol{\theta}}(\mathbf{r}) = \delta(\mathbf{r} - \boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta}))$. We can then rewrite the Jacobian of policy function as

$$\nabla_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{h}}[f(\mathbf{h}, \boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta}))] = \mathbb{E}_{\mathbf{h},\mathbf{r}}[f(\mathbf{h}, \mathbf{r}) \nabla_{\boldsymbol{\theta}} \log \pi_{\mathbf{h},\boldsymbol{\theta}}(\mathbf{r})] \tag{19}$$

where $\mathbf{r}$ is a random realization drawn from the distribution $\pi_{\mathbf{h},\boldsymbol{\theta}}(\mathbf{r})$. We now translate the computation of $\nabla_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{h}}[f(\mathbf{h}, \boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta}))]$ to a function evaluation $f(\mathbf{h}, \mathbf{r})$ multiplied with the gradient of the density function $\nabla_{\boldsymbol{\theta}} \log \pi_{\mathbf{h},\boldsymbol{\theta}}(\mathbf{r})$. However, computing $\nabla_{\boldsymbol{\theta}} \log \pi_{\mathbf{h},\boldsymbol{\theta}}(\mathbf{r})$ for a delta density function still requires the knowledge of $f(\cdot)$. We further address this issue by approximating the delta density function with a known density function centered around $\boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta})$, such as the Gaussian distribution, the Binomial distribution, etc. We can then estimate the gradient of policy function $\nabla_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{h}}[f(\mathbf{h}, \boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta}))]$ by using (19), which does not require the function model $f(\cdot)$ but rather the function value $f(\mathbf{h}, \mathbf{r})$ at a sampled channel state $\mathbf{h}$ and the distribution $\pi_{\mathbf{h},\boldsymbol{\theta}}(\mathbf{r})$. To estimate the expectation $\mathbb{E}_{\mathbf{h},\mathbf{r}}[\cdot]$, we observe $\mathcal{T}$ samples of the CSI and take the average as

$$\widetilde{\nabla}_{\boldsymbol{\theta}} \mathbb{E}_{\mathbf{h}}[f(\mathbf{h}, \boldsymbol{\Phi}(\mathbf{h}, \boldsymbol{\theta}))] = \frac{1}{\mathcal{T}} \sum_{\tau=1}^{\mathcal{T}} f(\mathbf{h}_\tau, \mathbf{r}_\tau) \nabla_{\boldsymbol{\theta}} \log \pi_{\mathbf{h}_\tau,\boldsymbol{\theta}}(\mathbf{r}_\tau) \tag{20}$$

where $\mathbf{h}_\tau$ is a sampled CSI and $\mathbf{r}_\tau$ is a realization drawn from the distribution $\pi_{\mathbf{h}_\tau,\boldsymbol{\theta}}(\mathbf{r})$. With the use of (20), we can compute the gradient of policy function in (17) as

$$\widetilde{\nabla}_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}, \boldsymbol{\lambda}) = \frac{1}{\mathcal{T}} \sum_{\tau=1}^{\mathcal{T}} \Big\{ \Big[f(\mathbf{h}_\tau, \mathbf{r}_\tau) - \sum_{s=1}^S \lambda_s c_s\big(\mathbf{r}_\tau, f(\mathbf{h}_\tau, \mathbf{r}_\tau)\big) \Big] \nabla_{\boldsymbol{\theta}} \log \pi_{\mathbf{h}_\tau,\boldsymbol{\theta}}(\mathbf{r}_\tau) \Big\}. \tag{21}$$

Algorithm 2 Primal-Dual Deep Learning Algorithm

```
Input: Initial primal and dual variables $\boldsymbol{\theta}^0, \boldsymbol{\lambda}^0$
for $k = 0, 1, 2, \ldots$ do {main loop}
    Draw CSI samples $\{\mathbf{h}_\tau\}_{\tau=1}^{\mathcal{T}}$, and get corresponding allocated resources $\{\mathbf{r}_\tau\}_{\tau=1}^{\mathcal{T}}$ according to DNN outputs $\{\boldsymbol{\Phi}(\mathbf{h}_\tau, \boldsymbol{\theta}^k)\}_{\tau=1}^{\mathcal{T}}$ and policy distributions $\{\pi_{\mathbf{h}_\tau, \boldsymbol{\theta}^k}(\mathbf{r})\}_{\tau=1}^{\mathcal{T}}$
    Obtain observations of the objective function $\{f(\mathbf{h}_\tau, \mathbf{r}_\tau)\}_{\tau=1}^{\mathcal{T}}$ at current samples $\{\mathbf{h}_\tau\}_{\tau=1}^{\mathcal{T}}$
    Compute the policy gradient $\widetilde{\nabla}_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}^k, \boldsymbol{\lambda}^k)$ by (21)
    Update the primal variables $\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k + \delta^k \widetilde{\nabla}_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}^k, \boldsymbol{\lambda}^k)$ [cf. (17)]
    Update the dual variables $\lambda_s^{k+1} = \left[\lambda_s^k + \eta^k \sum_{\tau=1}^{\mathcal{T}} c_s(\mathbf{r}_\tau, f(\mathbf{h}_\tau, \mathbf{r}_\tau))/\mathcal{T}\right]_+$ for all $s$ [cf. (18)]
end for
```

We stress the model-free aspect of computing the gradient in (21). That is, we need only observe the values $f(\mathbf{h}_\tau, \mathbf{r}_\tau)$ and $\{c_s(\mathbf{r}_\tau, f(\mathbf{h}_\tau, \mathbf{r}_\tau))\}_{s=1}^{S}$ as experienced in the FSO system under the instantaneous observed states $\mathbf{h}_\tau$ and $\mathbf{r}_\tau$. This is considered model-free because it does not require an explicit mathematical model of $f(\cdot)$ and $\{c_s(\cdot)\}_{s=1}^{S}$ or the CSI distribution model, which is typically required to compute analytic gradients in existing unsupervised learning methods. In terms of the dual step, by estimating the expectation with the average of $\mathcal{T}$ samples, it can also be computed with only observed values $\{c_s(\mathbf{r}_\tau, f(\mathbf{h}_\tau, \mathbf{r}_\tau))\}_{s=1}^{S}$. By replacing $\nabla_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}^k, \boldsymbol{\lambda}^k)$ with $\widetilde{\nabla}_{\boldsymbol{\theta}} \mathcal{L}(\boldsymbol{\theta}^k, \boldsymbol{\lambda}^k)$ in (17), the resulting PDDL algorithm is model-free and summarized in Algorithm 2. In general, the PDDL algorithm effectively combines the policy gradient method and the primal-dual method to solve resource allocation problems in FSO systems. On the one hand, it differs from the policy gradient method in reinforcement learning as it solves constrained optimization problems by working in the dual domain. On the other hand, it differs from the dual method as it parameterizes the solution with neural networks and learns the optimal parameterization with the policy gradient method in a model-free manner. We can therefore perform algorithm given any generic FSO systems and required constraints.

D. Network Design

In this subsection, we present the specific design of the proposed DNN architecture. It consists of one input layer, L fully connected hidden layers, and one output layer. The input layer takes the CSI as the input signal $\mathbf{x}_0 = \mathbf{h} \in \mathbb{R}^{n_0}$. The latter is processed by L fully connected hidden layers to generate the higher-level feature $\mathbf{x}_L$. The output layer decides the allocated resources $\mathbf{r}$ according to $\mathbf{x}_L$.

1) Hidden Layer: The ℓ-th hidden layer has n_ℓ units for $\ell = 1, \ldots, L$. Let $\mathbf{x}_{\ell-1} \in \mathbb{R}^{n_{\ell-1}}$ be the input feature, $\mathbf{W}_\ell \in \mathbb{R}^{n_\ell \times n_{\ell-1}}$ be the weight matrix, $\mathbf{b}_\ell \in \mathbb{R}^{n_\ell}$ be the bias term, and $\sigma_\ell(\cdot)$ be the ReLU nonlinearity. The output feature is computed as $\mathbf{x}_\ell = \sigma_\ell(\mathbf{W}_\ell \mathbf{x}_{\ell-1} + \mathbf{b}_\ell)$ for $\ell = 1, \ldots, L$.

2) Output Layer: The output layer decides the allocated resources $\mathbf{r}$ based on the output feature $\mathbf{x}_L$. Different from the existing literature, we apply the policy distribution $\pi_{\mathbf{h},\boldsymbol{\theta}}$ to sample $\mathbf{r}$ instead of directly computing $\mathbf{r}$. This modification not only trains the DNN in a model-free manner but also makes the allocated resources satisfy the allowable domain $\mathbf{r} \in \mathcal{R}$. Specifically, the resource allowable domain $\mathcal{R}$ depends on practical applications that may be an interval, a discrete set or binary values. The requirement $\mathbf{r} \in \mathcal{R}$ is usually difficult to satisfy and needs specific post-processing techniques in the existing literature. We design the DNN architecture to resolve this issue by selecting different policy distributions – see the following details w.r.t. two examples of power adaptation and relay selection.

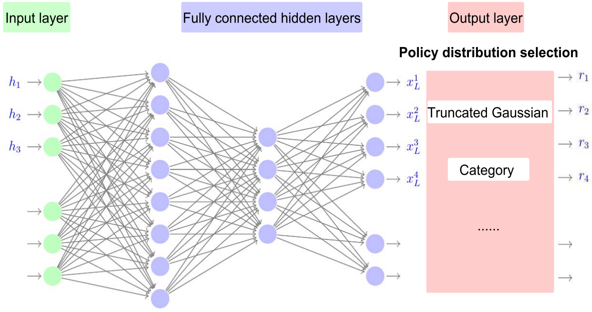


Fig. 2. The proposed DNN architecture with the input layer, 3 fully connected hidden layers and the output layer.

(i) For power adaptation, the transmitted power P_i of i-th channel must be within the interval $\mathcal{R} = [0, P_s]$. We satisfy this allowable domain by selecting the policy distribution as the *truncated Gaussian distribution*. The truncated Gaussian distribution is a Gaussian distribution with mean μ and variance σ^2, and lies within the interval $[0, P_s]$ with the probability density function as

$$f(x; \mu, \sigma, 0, P_s) = \frac{1}{\sigma} \frac{\Phi\left(\frac{x-\mu}{\sigma}\right)}{\boldsymbol{\Phi}\left(\frac{P_s-\mu}{\sigma}\right) - \boldsymbol{\Phi}\left(\frac{-\mu}{\sigma}\right)} \tag{22}$$

with $\Phi(x)$ the probability density function and $\boldsymbol{\Phi}(x)$ the cumulative distribution function of the standard Gaussian distribution. The distribution parameters μ and σ are determined by $\mathbf{x}_L$.

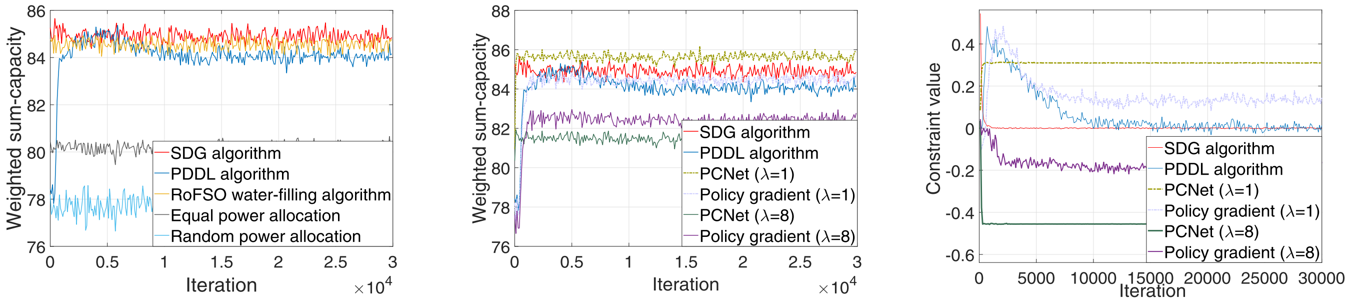
(ii) For relay selection, the selected choice r_i of i-th relay must be binary and satisfy $\sum_i r_i = 1$ that there is only one relay selected at each hop. We select the policy distribution as the *category distribution* to satisfy this allowable domain. The category distribution takes on one of K possible categories, where the probability of each category is separately specified as

$$f(x = k) = p_k, \text{ with } \sum_{k=1}^{K} p_k = 1 \tag{23}$$

and the distribution parameters $\{p_k\}_{k=1}^{K}$ determined by $\mathbf{x}_L$.

For joint power and relay allocation, we then select the policy distribution as the *combination of truncated Gaussian distribution and category distribution* to satisfy the allowable domain. Details are illustrated in Fig. 2.

Remark 2 The learning process of the PDDL algorithm outlined in Algorithm 2 may take a number of iterations



(a) (b) (c)

Fig. 3. Performance of the SDG, the PDDL and baseline policies for power adaptation in $N = 10$ wavelength multiplexing RoFSO systems. (a-b) The objective value. (c) The constraint value.

to update DNN parameters before convergence. However, we stress that this learning process is completed offline before execution / implementation and thus the total training time does not matter. In particular, we consider FSO systems can be deployed during the offline training. At each iteration, we collect channel state information $\{\mathbf{h}_i\}_i$ from the environment, input $\{\mathbf{h}_i\}_i$ into the DNN to generate allocated resources $\{\mathbf{r}_i\}_i$, and deploy real-world systems with $\{\mathbf{r}_i\}_i$ to observe objective values $\{C_i\}_i$. The primal step updates the primal variables based on objective observations $\{C_i\}_i$ and the dual step updates the dual variables based on updated primal variables. Note that this offline training does not rely on theoretical models but directly learns from real-world systems to avoid modeling errors, which is realizable because the PDDL algorithm is model-free by leveraging the policy gradient method. During the inference, the learned model can be directly implemented and requires little computational complexity, yielding an efficient implementation as validated in numerical experiments.[2]

V. Numerical Experiments

In this section, we corroborate theory by numerically evaluating the SDG and PDDL algorithms on a large set of resource allocation problems in FSO communications. To implement the algorithms, we consider a batch-size of $\mathcal{T} = 64$ samples. The ADAM optimizer is used for the primal update and the exponentially decaying step-size is used for the dual update. Note that though the PDDL is model-free, we obtain objective and constraint observations in numerical simulations by using system models; however, we do not assume knowledge of these models to implement the algorithm, only to generate observations that substitute for physical measurements.

Channel State Information. We consider FSO channel effects comprising two components: the attenuation h_a and the turbulence h_t. The attenuation h_a represents the path loss induced by weather conditions as $h_a = A_t A_r e^{-\alpha d}/(d^2\Lambda^2)$ with α the attenuation coefficient depending on weather visibility, d the transmission distance, Λ the wavelength, A_t and A_r the aperture areas of the transmitter and the receiver. The turbulence h_t is modeled as the well-known log-normal distribution, which is commonly used under weak-to-moderate turbulence. Without loss of generality, other distributions (e.g., Gamma-Gamma distribution) are applicable based on turbulence conditions. We then characterize the FSO channel as $y = h_a h_t x + z$ with x the transmitted signal, y the received signal and z the additive Gaussian noise.

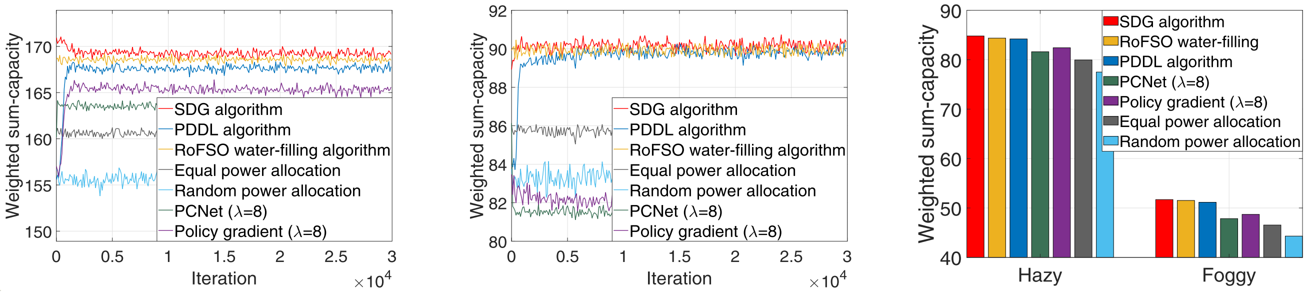
A. Power Adaptation

We first consider power adaptation in the RoFSO system – see Sec. II-A. The goal is to allocate powers to orthogonal optical carriers that maximize the weighted sum-capacity. The optimization problem is formulated by the objective (3) with total power limitation $\mathbb{E}_{\mathbf{h}}\left[c(\mathbf{r}(\mathbf{h}))\right] = \mathbb{E}_{\mathbf{h}}\big[\sum_{i=1}^{N} p_i(\mathbf{h})\big] - P_t \leq 0$ and peak power limitation $\mathcal{R} = [0, P_s]^N$.

The priority weights $\boldsymbol{\omega}$ are drawn randomly in $[0, 1]$ and system parameters are set as: $P_t = 1.5$W; $P_s = 0.3$W; $m_p = 5$; $OMI = 15\%$; $r = 0.75$; $RIN = -140$dB/Hz; $T = 300$K; transmitter aperture diameter $D_{tx} = 0.015$m; receiver aperture diameter $D_{rx} = 0.05$m and $d = 1$km. We consider three heuristic methods: (i) the RoFSO water-filling [14]; (ii) the equal power allocation; (iii) the random power allocation, and two unsupervised learning methods: (i) the power control network (PCNet) [35]; (ii) the policy gradient (without dual update) for performance comparison. The water-filling is a near-optimal solution developed for RoFSO systems, while equal & random power allocations are baseline policies. The PCNet is developed for RF systems that penalizes the objective with constraints scaled by pre-determined weights $\boldsymbol{\lambda}$ and trains DNNs based on the penalized objective, while the policy gradient is the simplified PDDL that performs only primal updates with pre-determined dual variables $\boldsymbol{\lambda}$. The water-filling and the PCNet are model-based methods that depend on closed-form system models to solve KKT conditions and to evaluate the loss function, while the policy gradient and equal & random power allocations are model-free. For the PDDL algorithm, we consider the policy distribution $\pi_{\mathbf{h},\boldsymbol{\theta}}$ as the truncated Gaussian distribution to satisfy the feasibility condition $\mathbf{r}(\mathbf{h}) \in \mathcal{R} = [0, P_s]^N$ [cf. (22)]. The DNN is built with two hidden layers, each of which contains 200 and 100 units respectively.

Fig. 3 shows results for $N = 10$ wavelength multiplexing RoFSO systems. From Fig. 3a, we see that the SDG and the PDDL converge with the increase of iteration. The SDG solves the problem exactly and exhibits the best performance. The PDDL outperforms significantly model-free heuristic policies

[2] If the real-world system changes after operating for a long period of time, the learned model may need to be further updated online based on the dynamic changes following the same training algorithm as used in offline training.



(a) (b) (c)

Fig. 4. Performance of the SDG, the PDDL and baseline policies for power adaptation in different RoFSO system configurations. (a) 20 wavelength multiplexing with power limitations $P_t = 3\text{W}, P_s = 0.3\text{W}$. (b) 10 wavelength multiplexing with power limitations $P_t = 3\text{W}, P_s = 0.6\text{W}$. (c) Hazy and light foggy weather conditions.

and achieves near-optimal performance close to model-based SDG and water-filling algorithms. Fig. 3b and Fig. 3c show the objective and constraint values compared with the PCNet and the policy gradient of different pre-determined penalty weights $\lambda = 1$ and 8. A small weight $\lambda = 1$ yields better performance but the obtained solutions do not satisfy the power constraint, while a large weight $\lambda = 8$ obtains feasible solutions but with worse performance. Contrarily, the SDG and the PDDL achieve a satisfactory trade-off between the objective and the constraint by updating the primal and dual variables simultaneously, i.e., they obtain feasible solutions with improved performance. Furthermore, the PDDL allows for a completely model-free implementation with no need of theoretical system models.

In Fig. 4, we run experiments under different system configurations; namely, different numbers of channels, different power budgets, and different weather conditions, to show the algorithm adaptability to changing scenarios. Fig. 4a plots the objective in $N = 20$ wavelength multiplexing RoFSO systems, Fig. 4b shows that with larger power budgets $P_t = 3\text{W}$ and $P_s = 0.6\text{W}$, and Fig. 4c compares the hazy (4.5dB/km path loss exponent) and light foggy (11.5dB/km path loss exponent) weather conditions, where the penalty weight of the PCNet and the policy gradient is set as $\lambda = 8$ for feasible solutions. Similar results apply here, where the SDG outperforms baseline policies and the PDDL achieves near-optimal performance in a model-free manner. The performance improvements of the SDG and the PDDL become more visible in larger systems (Fig. 4a) and worse weather conditions (Fig. 4c), and the PDDL converges roughly to the same value as the SDG with larger power budgets (Fig. 4b). The latter is because the increased budgets create more space for the PDDL to manipulate powers and the learning ability of DNNs is fully activated. We also observe that the PCNet and the policy gradient with the same penalty weight $\lambda = 8$ perform worse with larger power budgets. We attribute this behavior to the fact that the penalty weight λ is sensitive to system configurations, which further shows the drawback of pre-determining λ in these methods.

Besides performance, the inference time is of utmost importance for cooperative transmissions that allocate resources based on instantaneous CSI. Table I compares the inference time of the SDG, the PDDL and the water-filling algorithms for processing an instantiation of CSI. The PDDL requires far less time than the other algorithms but achieves comparable performance. This is because the computation of the DNN contains simply linear operations with pointwise nonlinearities, whereas the SDG requires some more computation expense for solving the inner maximization in (12). The water-filling is particularly computationally expensive, requiring substantial time to solve KKT conditions of the complicated objective with the power constraint. The time saved by the PDDL and the SDG increases as the system becomes larger, highlighting a further advantage of our algorithms. The PDDL, the PCNet and the policy gradient have similar inference time since the PCNet contains a neural network of the same hyperparameters and the policy gradient is the simplified PDDL.

TABLE I

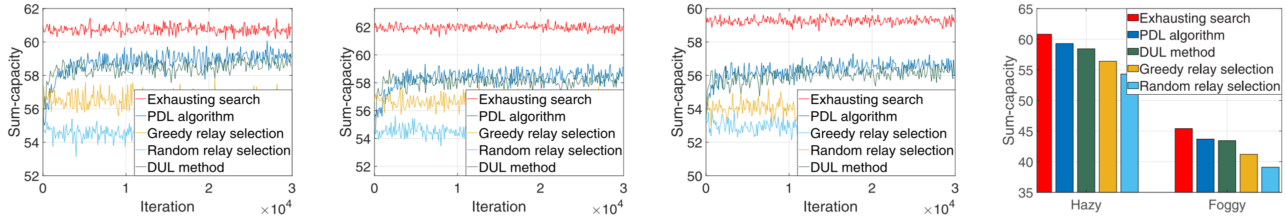
Inference Time of the SDG, the PDDL and the Water-Filling. (a) $N = 10$, $P_t = 1.5\text{W}$ and $P_s = 0.3\text{W}$. (b) $N = 10$, $P_t = 3\text{W}$ and $P_s = 0.6\text{W}$. (c) $N = 20$, $P_t = 3\text{W}$ and $P_s = 0.3\text{W}$

	Case (a)	Case (b)	Case (c)
The SDG	$2.81 \cdot 10^{-3}$s	$3.91 \cdot 10^{-3}$s	$7.19 \cdot 10^{-3}$s
The PDDL	$1.56 \cdot 10^{-5}$s	$1.52 \cdot 10^{-5}$s	$1.59 \cdot 10^{-5}$s
The RoFSO water-filling	1.65s	1.71s	3.31s

B. Relay Selection

We then consider relay selection in relay-assisted FSO networks – see Sec. II-B. The goal is to select the appropriate relay at each hop to maximize the channel capacity. The optimization problem is formulated by the objective (5) with the action space $\mathcal{R} = \big\{\{0,1\}^{N \times M} | \sum_{j=1}^{M} \alpha_{ij}(\mathbf{h}) \leq 1, \text{for } i = 1, \ldots, N\big\}$. Note that there is no stochastic constraint in this problem, in which case the dual step is not required. The SDG reduces to directly maximizing (5) and thus is not considered in this case, while the PDDL reduces to the Primal Deep Learning (PDL).

The system parameters are set as $B = 5 \times 10^8$Hz, $T_f = 10^{-8}$s, $P = 0.3$W and $R = 0.75$A/W. We consider four baseline policies: (i) the exhausting search; (ii) the greedy



(a) (b) (c) (d)

Fig. 5. Performance of the PDL and baseline policies in relay-assisted FSO networks. (a) 2-hop FSO network with 5 parallel relays per hop. (b) 2-hop FSO network with 10 parallel relays per hop. (c) 3-hop FSO network with 5 parallel relays per hop. (d) Hazy and light foggy weather conditions.

selection; (iii) the random selection; (iv) the deep unsupervised learning (DUL) [36] for performance comparison. The exhausting search is the optimal solution that considers all possible relay strategies and selects the best one, greedy & random selections are heuristic policies, and the DUL is an unsupervised learning method developed for RF systems. The exhausting search and the DUL are model-based that require closed-form system models to evaluate the system performance and the loss function, while greedy & random selections are model-free. For the PDL algorithm, we select the policy distribution $\pi_{\mathbf{h},\boldsymbol{\theta}}$ as the categorical distribution since the allocated resources are binary $\mathbf{r}(\mathbf{h}) \in \mathcal{R}$ [cf. (23)]. We construct a two-layered DNN of 200 and 100 units.

Fig. 5a compares the performance between the PDL and four baseline policies in a 2-hop network with $M = 5$ parallel relays per hop. We see that the PDL converges and achieves near-optimal performance close to the exhausting search. The latter performs best on the premise that system models are available and could be computationally expensive with a large number of hops and relays. The PDL outperforms the other policies including the model-based DUL. This is because the DUL uses continuous relaxation $\tilde{\alpha}_{ij} \in [0,1]$ for discrete actions $\alpha_{ij} \in \{0,1\}$ during training, which mismatches the inference phase and results in performance degradation. Additional experiments in alternative scenarios are performed, i.e., a 2-hop network with $M = 10$ relays per hop in Fig. 5b, a 3-hop network with $M = 5$ relays per hop in Fig. 5c, and the hazy and light foggy weather conditions in Fig. 5d. We observe similar results indicating the adaptivity of the proposed algorithm to larger FSO networks and different weather conditions. The PDL degrades slightly in Fig. 5b and 5c because the problem becomes more difficult as we enlarge the system with more relays or more hops, while the DNN remains the same with unchanged representational capacity.

Table II shows the inference time of the exhausting search, the PDL and the greedy algorithms. Though the exhausting search exhibits the best performance, it takes the most time for inference. The PDL achieves the close performance to the exhausting search but only requires a comparable time to the greedy selection, achieving a favorable trade-off.

C. Joint Power and Relay Allocation

We now consider joint power and relay allocation in two applications, which are more complicated but also of more interests in practice.

TABLE II

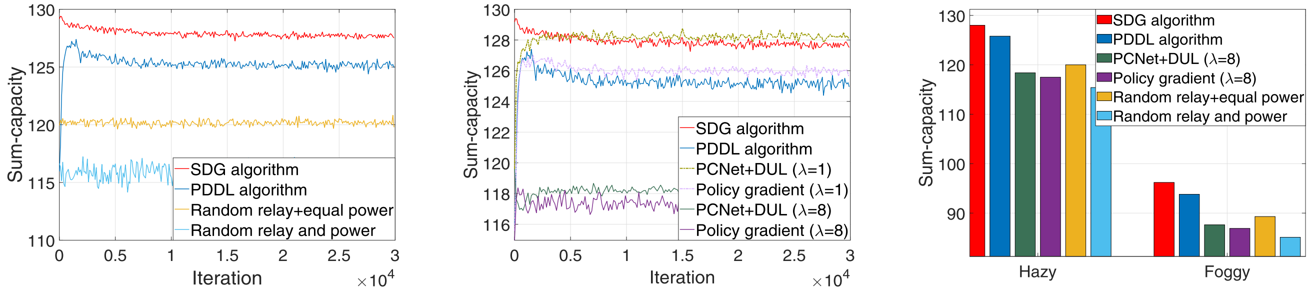
Inference Time of the Exhausting Search, the PDL and the Greedy Selection in Three Cases. (a) 2-Hop FSO Network With 5 Parallel Relays per Hop. (b) 2-Hop FSO Network With 10 Parallel Relays per Hop. (c) 3-Hop FSO Network With 5 Parallel Relays per Hop

	Case (a)	Case (b)	Case (c)
The exhausting search	$1.40 \cdot 10^{-4}$s	$5.77 \cdot 10^{-4}$s	$9.81 \cdot 10^{-4}$s
The PDL	$1.55 \cdot 10^{-5}$s	$3.09 \cdot 10^{-5}$s	$3.11 \cdot 10^{-5}$s
The greedy	$1.49 \cdot 10^{-5}$s	$1.56 \cdot 10^{-5}$s	$1.55 \cdot 10^{-5}$s

1) Relay-Assisted Multichannel FSO Network: For the first experiment, we consider the relay-assisted multichannel FSO network where the system transmits signals with L orthogonal optical carriers through N intermediate hops [48]. In particular, the transmitter modulates signals onto multiple optical carriers and sends them simultaneously to the selected relay. The latter aggregates received signals, modulates orthogonal carriers, and transmits to the selected relay at next hop until the receiver. We assume there is no crosstalk between orthogonal carriers and each hop contains M parallel relays for selection. Based on the CSI, different relays are selected at different hops and different powers are assigned to different carriers at the transmitter and selected relays to maximize the total channel capacity. Let $\mathbf{h}$ be the CSI between the transmitter, relays and the receiver, and $\mathbf{r}(\mathbf{h}) = \{\mathbf{p}_{ij}(\mathbf{h}), \alpha_{ij}(\mathbf{h})\}_{i=0,\ldots,N, j=1,\ldots,M}$ the allocated resources including assigned powers and selected relays. Specifically, $\mathbf{p}_{ij}(\mathbf{h}) = [p^1_{ij}(\mathbf{h}), \ldots, p^L_{ij}(\mathbf{h})]^\top \in \mathbb{R}^L$ are powers of L optical carriers at j-th relay of i-th hop where $i = 0, j = 1$ and $i = N+1, j = 1$ represent the transmitter and the receiver, and $\alpha_{ij} \in \{0,1\}$ indicates whether j-th relay is selected at i-th hop. The channel capacity of ℓ-th orthogonal channel over a specific selected relaying link is

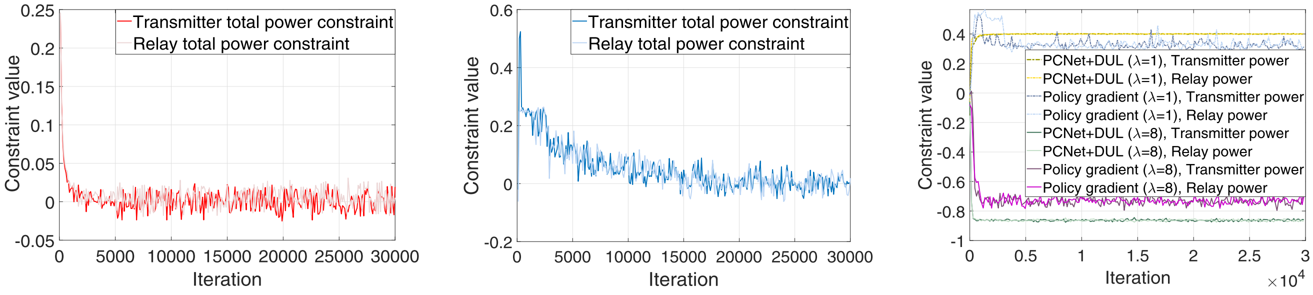
$$C^\ell_{j_1\ldots j_N}(\mathbf{h}) = \frac{T_f B}{\epsilon}\log\Bigg(1+\Bigg(\prod_{i=0}^{N}\Bigg(1+\frac{1}{p^\ell_{ij_i}(\mathbf{h})h^\ell_{j_ij_{i+1}}\frac{R}{e\Delta f}}\Bigg)-1\Bigg)^{-1}\Bigg) \tag{24}$$

where we assume j_i-th relay is selected at i-th hop and $h^\ell_{j_ij_{i+1}}$ is the CSI of ℓ-th optical carrier between j_i-th relay at i-th hop and j_{i+1}-th relay at $(i+1)$-th hop. Since there is single transmitter and single receiver, we have $j_0 = j_{N+1} = 1$ by default. There are three types of constraints: the total power limitation P_t at the transmitter and selected relays, the peak power limitation P_s for each carrier, and that only one relay



(a) (b) (c)

Fig. 6. Performance of the SDG, the PDDL and baseline policies for joint power and relay allocation in the relay-assisted multichannel FSO network.



(a) (b) (c)

Fig. 7. The constraint values of the SDG, the PDDL, the PCNet with DUL and the policy gradient for joint power and relay allocation in the relay-assisted multichannel FSO network. (a) The SDG. (b) The PDDL. (c) The PCNet with DUL and the policy gradient.

is selected at each hop. The optimization problem is

$$\mathbb{P}:=\max_{\mathbf{r}(\mathbf{h})}\mathbb{E}_{\mathbf{h}}\Big[\sum_{j_N=1}^{M}\cdot\cdot\sum_{j_1=1}^{M}\Big(\prod_{i=1}^{N}\alpha_{ij_i}(\mathbf{h})\Big)\sum_{\ell=1}^{L}\omega_\ell C^{\ell}_{j_1\ldots j_N}(\mathbf{h})\Big],$$

$$\text{s. t. } \mathbb{E}_{\mathbf{h}}\Big[\sum_{\ell=1}^{L}p^{\ell}_{ij_i}(\mathbf{h})\Big]-P_t\leq 0, \text{for } i=0,1,\ldots,N, j_i=1,\ldots,M,$$

$$\mathcal{R}=\Big\{[0,P_s]^{(1+N\times M)\times L}\times\{0,1\}^{N\times M}|\sum_{j_i=1}^{M}\alpha_{ij_i}(\mathbf{h})\leq 1, i=1,\ldots,N\Big\} \tag{25}$$

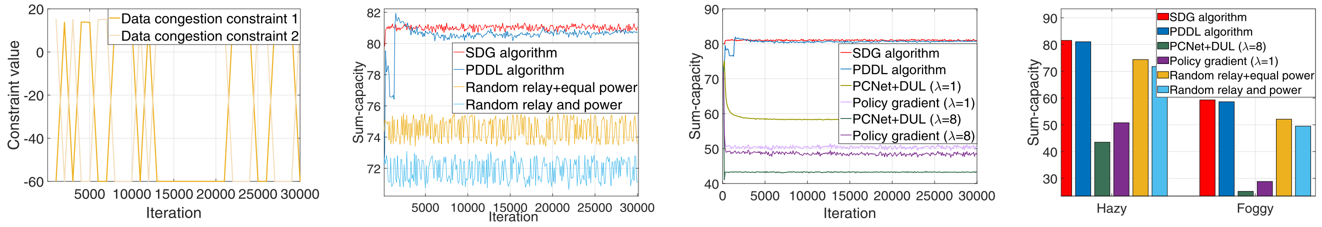
with $\boldsymbol{\omega}=[\omega_1,\ldots,\omega_L]^\top$ the priorities of different optical carriers. This challenging problem is the extension of the problem in Section II-B to the scenario with orthogonal optical carriers.

We assume a 1-hop network with $M=5$ parallel relays per hop and $L=5$ orthogonal optical carriers. The priority weights $\boldsymbol{\omega}$ are drawn randomly in $[0,1]$ and system parameters are set as: $B=5\times 10^8$Hz, $T_f=10^{-8}$s, $P_t=1.5$W, $P_s=0.6$W and $R=0.75$A/W. We consider two heuristic policies: (i) the random relay selection with equal power allocation; (ii) the random relay selection with random power allocation, and two unsupervised learning methods (i) the PCNet with DUL; (ii) the policy gradient (without dual update) for performance comparison. The PCNet with DUL is model-based that requires closed-form system models to estimate the loss function, while the other baseline policies are model-free. For the PDDL algorithm, we consider the truncated Gaussian distribution to allocate powers and the categorical distribution to select relays. The DNN is constructed as a two-layered architecture of 200 and 100 units.

Fig. 6 and Fig. 7 show the objective and constraints of the SDG, the PDDL and four baseline policies. The SDG and the PDDL obtain feasible solutions with superior performance, and their performance improvements get emphasized compared to either single power adaptation in Sec. V-A or single relay selection in Sec. V-B. This is because advantages of our algorithms get compounded in this joint problem. The PDDL obtains close performance to the SDG but does not require any system model for implementation. The PCNet with DUL and the policy gradient of $\lambda=1$ achieve better performance but the obtained solutions are infeasible, while ones of $\lambda=8$ obtain feasible solutions but with worse performance. The latter indicates the difficulty of selecting proper weights $\boldsymbol{\lambda}$ at the outset and emphasizes the advantage of our algorithms.

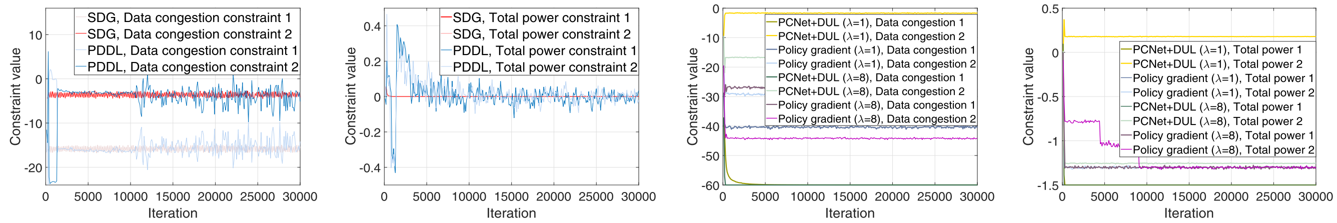
2) FSO Fronthaul Network: The second experiment considers FSO fronthaul networks – see Sec. II-C. We consider a large-scale fronthaul network divided into multiple small-scale fronthaul clusters that perform resource allocation independently. In a fronthaul cluster, the goal is to allocate powers to orthogonal optical carriers and select the optimal AN at each RRH that maximize the sum-capacity. The optimization problem is formulated by the objective (7) and constraints (8). Note that data congestion constraints (8c) further complicate the problem, making it extremely challenging to solve in practice.

We consider a FSO fronthaul cluster with $N=5$ RRHs, $M=2$ ANs, one BBU and $L=5$ orthogonal carriers. RRHs and ANs are distributed randomly at locations in the squares $[-5\text{km},5\text{km}]^2$ and $[-1\text{km},1\text{km}]^2$ respectively. The system parameters are set as: $B=10^9$Hz, $T_f=10^{-9}$s,



(a) (b) (c) (d)

Fig. 8. (a) Data congestion constraints of the random AN selection with equal power allocation. (b-d) Performance of the SDG, the PDDL and baseline policies for joint power and relay allocation in the FSO fronthaul network.



(a) (b) (c) (d)

Fig. 9. (a) Data congestion constraints of the SDG and the PDDL. (b) Total power constraints of the SDG and the PDDL at two example RRHs. (c) Data congestion constraints of the PCNet with DUL and the policy gradient. (d) Total power constraints of the PCNet with DUL and the policy gradient at two example RRHs.

$P_t = 1.5$W, $P_s = 0.6$W, $R = 0.75$A/W and $C_t = 50$. For the PDDL algorithm, the truncated Gaussian distribution and the categorical distribution are used for power allocation and AN selection. As the problem becomes more complicated, we consider a denser DNN with 3 layers, each of which contains 400, 200 and 100 units. We similarly consider four baseline policies: (i) the random relay selection with equal power allocation; (ii) the random relay selection with random power allocation; (iii) the PCNet with DUL; (iv) the policy gradient (without dual update) for performance comparison. The first two heuristic policies are infeasible because there always exist times when all RRHs transmit optical signals to the same AN violating data congestion constraints (8c) – see data congestion constraint results in Fig. 8a, and thus we only consider them as benchmarks for reference.

We plot in Fig. 8b-8d and Fig. 9 the objective and constraints of the SDG, the PDDL and baseline policies. Similar as prior simulations, the SDG exhibits the best performance using system models, while the PDDL performs comparably to the SDG forgoing any system model. Both outperform significantly baseline policies because resource constraints are more complicated and simultaneous dual updates become more important. We remark that random relay selections and the PCNet with DUL ($\lambda = 1$) do not satisfy either data congestion constraints or total power constraints, and thus are only presented here for reference. Furthermore, observe that as the number of constraints increases and the format of constraints becomes complicated, selecting proper penalty weights $\boldsymbol{\lambda}$ in the PCNet with DUL and the policy gradient becomes extremely challenging in this case.

To conclude our numerical analysis, we provide in Table III the inference time of the SDG and the PDDL algorithms for two joint power and relay allocation problems. We see that besides requiring system model information, the SDG achieves better performance at the expense of more inference time. The latter gets emphasized when FSO systems or resource allocation problems become more complicated. The inference time of the PDDL is much lower but increases slightly from single power adaptation to joint power and relay allocation, which is because the applied DNN gets deeper and denser. However, its computation is independent on FSO systems and optimization problems, resulting in an efficient implementation. Furthermore, it is worth mentioning that while with similar inference time, the PCNet with DUL requires more training time ($5.30 \cdot 10^{-2}$s and $2.90 \cdot 10^{-1}$s per iteration for case (a) and (b)) compared with the proposed PDDL ($1.40 \cdot 10^{-2}$s and $4.39 \cdot 10^{-2}$s per iteration for case (a) and (b)). This is because the PCNet requires post-processing to satisfy $\mathbf{r} \in \mathcal{R}$ and the DUL employs continuous relaxation, both taking more computation during training.

TABLE III

INFERENCE TIME OF THE SDG AND THE PDDL FOR JOINT POWER AND RELAY ALLOCATION. (A) RELAY-ASSISTED MULTICHANNEL FSO NETWORK. (B) FSO FRONTHAUL NETWORK

	Case (a)	Case (b)
The SDG	$8.28 \cdot 10^{-3}$s	$2.83 \cdot 10^{-2}$s
The PDDL	$3.28 \cdot 10^{-5}$s	$9.34 \cdot 10^{-5}$s

Remark 3: The capacity formulas considered in relay-assisted FSO networks are derived based on coherent detection. However, we remark that the proposed algorithms also apply to non-coherent detection systems and we clarify this fact from two scenarios. (i) If the capacity formula in non-coherent detection systems can be obtained, both the

SDG and the PDDL algorithms can be applied by replacing the coherent capacity formula with the non-coherent one. The SDG may be computationally more expensive if the capacity formula becomes more complicated because it needs to numerically solve the inner maximization related to the capacity formula [cf. (10) and (12)], while the PDDL is not affected because it is model-free and need not the capacity formula but only capacity observations. (ii) If the capacity formula in non-coherent detection systems cannot be obtained, only the PDDL is applicable because it does not require any theoretical models including the capacity formula. However, the SDG cannot be applied in this case since it cannot conduct the primal step (10) without the specific capacity formula.

VI. Conclusion

In this paper, we consider the general resource allocation in free space optical communications. We formulate the problem under the constrained stochastic optimization framework. Such problems are typically challenging due to the non-convex nature, multiple constraints and lack of model information. We first proposed the model-based Stochastic Dual Gradient algorithm, which solves the problem exactly by exploiting the strong duality. However, it heavily relies on system models that may not be available in practice. The model-free Primal-Dual Deep Learning algorithm was developed to overcome this issue. It parameterizes the resource allocation policy with DNNs and learns optimal parameters by updating primal and dual variables simultaneously. Policy gradient method is applied to the primal update in order to estimate necessary gradient information without using the knowledge of system and channel models. The proposed algorithms are computationally efficient and transferable to any resource allocation problem under the framework, which were validated in numerous numerical experiments.

References

[1] Z. Gao, M. Eisen, and A. Ribeiro, "Optimal WDM power allocation via deep learning for radio on free space optics systems," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2019, pp. 1–6.

[2] V. W. S. Chan, "Free-space optical communications," *IEEE/OSA J. Lightw. Technol.*, vol. 24, no. 12, pp. 4750–4762, Dec. 1, 2006.

[3] M. A. Khalighi and M. Uysal, "Survey on free space optical communication: A communication theory perspective," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 4, pp. 2231–2258, Nov. 2014.

[4] L. C. Andrews and R. L. Phillips, *Laser Beam Propagation Through Random Media*. Bellingham, WA, USA: SPIE, 2005.

[5] D. K. Borah and D. G. Voelz, "Pointing error effects on free-space optical communication links in the presence of atmospheric turbulence," *J. Lightw. Technol.*, vol. 27, no. 18, pp. 3965–3973, Sep. 15, 2009.

[6] Z. Gao, J. Zhang, and A. Dang, "Beam spread and wander of Gaussian beam through anisotropic non-kolmogorov atmospheric turbulence for optical wireless communication," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2017, pp. 343–348.

[7] Z. Gao, Y. Luo, and A. Dang, "Beam wander effects on scintillation theory of Gaussian beam through anisotropic non-kolmogorov atmospheric turbulence for optical wireless communication," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2018, pp. 1–6.

[8] Z. Gao, Z. Li, and A. Dang, "Beam quality factor and its effect on laser beam through anisotropic turbulence for OWC," in *Proc. 15th Int. Conf. Telecommun. (ConTEL)*, Jul. 2019, pp. 1–7.

[9] J. Zhang, R. Li, Z. Gao, and A. Dang, "Ergodicity of phase fluctuations for free-space optical link in atmospheric turbulence," *IEEE Photon. Technol. Lett.*, vol. 31, no. 5, pp. 377–380, Mar. 1, 2019.

[10] K. Kiasaleh, "Performance of APD-based, PPM free-space optical communication systems in atmospheric turbulence," *IEEE Trans. Commun.*, vol. 53, no. 9, pp. 1455–1461, Sep. 2005.

[11] S. M. Navidpour, M. Uysal, and M. Kavehrad, "BER performance of free-space optical transmission with spatial diversity," *IEEE Trans. Wireless Commun.*, vol. 6, no. 8, pp. 2813–2819, Aug. 2007.

[12] C. Abou-Rjeily and A. Slim, "Cooperative diversity for free-space optical communications: Transceiver design and performance analysis," *IEEE Trans. Commum.*, vol. 59, no. 3, pp. 658–663, Mar. 2010.

[13] K.-H. Park, Y.-C. Ko, and M.-S. Alouini, "On the power and offset allocation for rate adaptation of spatial multiplexing in optical wireless MIMO channels," *IEEE Trans. Commun.*, vol. 61, no. 4, pp. 1535–1543, Apr. 2013.

[14] H. Zhou, S. Mao, and P. Agrawal, "Optical power allocation for adaptive transmissions in wavelength-division multiplexing free space optical networks," *Digit. Commun. Netw.*, vol. 1, no. 3, pp. 171–180, Aug. 2015.

[15] M. Z. Hassan, M. J. Hossain, J. Cheng, and V. C. M. Leung, "Statistical delay-QoS aware joint power allocation and relaying link selection for free space optics based fronthaul networks," *IEEE Trans. Commun.*, vol. 66, no. 3, pp. 1124–1138, Mar. 2017.

[16] C. Sun, X. Gao, J. Wang, Z. Ding, and X.-G. Xia, "Beam domain massive MIMO for optical wireless communications with transmit lens," *IEEE Trans. Commun.*, vol. 67, no. 3, pp. 2188–2202, Mar. 2018.

[17] C. Abou-Rjeily and S. Haddad, "Cooperative FSO systems: Performance analysis and optimal power allocation," *J. Lightw. Technol.*, vol. 29, no. 7, pp. 1058–1065, Apr. 1, 2011.

[18] A. A. El-Malek, A. M. Salhab, S. A. Zummo, and M.-S. Alouini, "Effect of RF interference on the security-reliability tradeoff analysis of multiuser mixed RF/FSO relay networks with power allocation," *J. Lightw. Technol.*, vol. 35, no. 9, pp. 1490–1505, May 1, 2017.

[19] Z. Hassan, J. Hossain, J. Cheng, and V. C. M. Leung, "Delay-QoS-aware adaptive modulation and power allocation for dual-channel coherent OWC," *J. Opt. Commun. Netw.*, vol. 10, no. 3, pp. 138–151, Mar. 2018.

[20] M. Safari and M. Uysal, "Relay-assisted free-space optical communication," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5441–5449, Dec. 2008.

[21] N. D. Chatzidiamantis, D. S. Michalopoulos, E. E. Kriezis, G. K. Karagiannidis, and R. Schober, "Relay selection protocols for relay-assisted free-space optical systems," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 5, no. 1, pp. 92–103, Jan. 2013.

[22] C. Abou-Rjeily, "Performance analysis of selective relaying in cooperative free-space optical systems," *J. Lightw. Technol.*, vol. 31, no. 18, pp. 2965–2973, Sep. 15, 2013.

[23] M. A. Kashani, M. Safari, and M. Uysal, "Optimal relay placement and diversity analysis of relay-assisted free-space optical communication systems," *IEEE/OSA J. Opt. Commun. Netw.*, vol. 5, no. 1, pp. 37–47, Jan. 2013.

[24] H. Zhou, D. Hu, S. Mao, and P. Agrawal, "Joint relay selection and power allocation in cooperative FSO networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2013, pp. 2418–2423.

[25] M. Z. Hassan, V. C. Leung, M. J. Hossain, and J. Cheng, "Delay-QoS aware adaptive resource allocations for free space optical fronthaul networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2017, pp. 1–6.

[26] W. Lee, M. Kim, and D.-H. Cho, "Deep power control: Transmit power control scheme based on convolutional neural network," *IEEE Commun. Lett.*, vol. 22, no. 6, pp. 1276–1279, Jun. 2018.

[27] H. Sun, X. Chen, Q. Shi, M. Hong, X. Fu, and N. D. Sidiropoulos, "Learning to optimize: Training deep neural networks for interference management," *IEEE Trans. Signal Process.*, vol. 66, no. 20, pp. 5438–5453, Oct. 2018.

[28] P. de Kerret, D. Gesbert, and M. Filippone, "Team deep neural networks for interference channels," in *Proc. IEEE Int. Conf. Commun. Workshops (ICC Workshops)*, May 2018, pp. 1–6.

[29] G. Jia, Z. Yang, H.-K. Lam, J. Shi, and M. Shikh-Bahaei, "Channel assignment in uplink wireless communication using machine learning approach," *IEEE Commun. Lett.*, vol. 24, no. 4, pp. 787–791, Apr. 2020.

[30] A. Zappone, L. Sanguinetti, and M. Debbah, "User association and load balancing for massive MIMO through deep learning," in *Proc. 52nd Asilomar Conf. Signals, Syst., Comput.*, Oct. 2018, pp. 1262–1266.

[31] R. Liu, M. Lee, G. Yu, and G. Y. Li, "User association for millimeter-wave networks: A machine learning approach," *IEEE Trans. Commun.*, vol. 68, no. 7, pp. 4162–4174, Jul. 2020.

[32] M. Lee, Y. Xiong, G. Yu, and G. Y. Li, "Deep neural networks for linear sum assignment problems," *IEEE Wireless Commun. Lett.*, vol. 7, no. 6, pp. 962–965, Dec. 2018.

[33] W. Lee, M. Kim, and D.-H. Cho, “Transmit power control using deep neural network for underlay device-to-device communication,” *IEEE Wireless Commun. Lett.*, vol. 8, no. 1, pp. 141–144, Feb. 2018.

[34] W. Lee, “Resource allocation for multi-channel underlay cognitive radio network based on deep neural network,” *IEEE Commun. Lett.*, vol. 22, no. 9, pp. 1942–1945, Sep. 2018.

[35] F. Liang, C. Shen, W. Yu, and F. Wu, “Towards optimal power control via ensembling deep neural networks,” *IEEE Trans. Commun.*, vol. 68, no. 3, pp. 1760–1776, Mar. 2019.

[36] A. Kaushik, M. Alizadeh, O. Waqar, and H. Tabassum, “Deep unsupervised learning for generalized assignment problems: A case-study of user-association in wireless networks,” 2021, *arXiv:2103.14548*.

[37] W. Jiang and H. D. Schotten, “A simple cooperative diversity method based on deep-learning-aided relay selection,” *IEEE Trans. Veh. Technol.*, vol. 70, no. 5, pp. 4485–4500, May 2021.

[38] A. Abdelreheem, O. A. Omer, H. Esmaiel, and U. S. Mohamed, “Deep learning-based relay selection in D2D millimeter wave communications,” in *Proc. Int. Conf. Comput. Inf. Sci. (ICCIS)*, Apr. 2019, pp. 1–5.

[39] Z. Xu, Y. Wang, J. Tang, J. Wang, and M. C. Gursoy, “A deep reinforcement learning based framework for power-efficient resource allocation in cloud RANs,” in *Proc. IEEE Int. Conf. Commun. (ICC)*, May 2017, pp. 1–6.

[40] M. Eisen, C. Zhang, L. F. O. Chamon, D. D. Lee, and A. Ribeiro, “Learning optimal resource allocations in wireless systems,” *IEEE Trans. Signal Process.*, vol. 67, no. 10, pp. 2775–2790, Apr. 2019.

[41] Y. S. Nasir and D. Guo, “Multi-agent deep reinforcement learning for dynamic power allocation in wireless networks,” *IEEE J. Sel. Areas Commun.*, vol. 37, no. 10, pp. 2239–2250, Oct. 2019.

[42] Y. Su, X. Lu, Y. Zhao, L. Huang, and X. Du, “Cooperative communications with relay selection based on deep reinforcement learning in wireless sensor networks,” *IEEE Sensors J.*, vol. 19, no. 20, pp. 9561–9569, Oct. 2019.

[43] M. A. Amirabadi, M. H. Kahaei, S. A. Nezamalhosseini, and V. T. Vakili, “Deep learning for channel estimation in FSO communication system,” *Opt. Commun.*, vol. 459, Mar. 2020, Art. no. 124989.

[44] S. Lohani and R. T. Glasser, “Turbulence correction with artificial neural networks,” *Opt. Lett.*, vol. 43, no. 11, pp. 2611–2614, 2018.

[45] L. Darwesh and N. S. Kopeika, “Deep learning for improving performance of OOK modulation over FSO turbulent channels,” *IEEE Access*, vol. 8, pp. 155275–155284, 2020.

[46] H. Lee, S. H. Lee, T. Q. S. Quek, and I. Lee, “Deep learning framework for wireless systems: Applications to optical wireless communications,” *IEEE Commun. Mag.*, vol. 57, no. 3, pp. 35–41, Mar. 2019.

[47] Y. Li, T. Geng, R. Tian, and S. Gao, “Power allocation in a spatial multiplexing free-space optical system with reinforcement learning,” *Opt. Commun.*, vol. 488, Jun. 2021, Art. no. 126856.

[48] M. Z. Hassan, V. C. Leung, M. J. Hossain, and J. Cheng, “Statistical delay aware joint power allocations and relay selection for NLOS multichannel OWC,” in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2016, pp. 1–6.

[49] H. Kaushal and G. Kaddoum, “Optical communication in space: Challenges and mitigation techniques,” *IEEE Commun. Surveys Tuts.*, vol. 19, no. 1, pp. 57–96, 1st Quart., 2017.

[50] L. F. O. Chamon, Y. C. Eldar, and A. Ribeiro, “Functional nonlinear sparse models,” *IEEE Trans. Signal Process.*, vol. 68, pp. 2449–2463, 2020.

[51] L. Bottou, “Stochastic gradient descent tricks,” in *Neural Networks: Tricks of the Trade*. Berlin, Germany: Springer, 2012, pp. 421–436.

[52] M. J. Neely, “Stochastic network optimization with application to communication and queueing systems,” *Synthesis Lectures Commun. Netw.*, vol. 3, no. 1, pp. 1–211, 2010.

[53] K. Hornik, M. Stinchcombe, and H. White, “Multilayer feedforward networks are universal approximators,” *Neural Netw.*, vol. 2, no. 5, pp. 359–366, 1989.

[54] R. S. Sutton, D. A. McAllester, S. P. Singh, and Y. Mansour, “Policy gradient methods for reinforcement learning with function approximation,” in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2000, pp. 1057–1063.

Zhan Gao (Student Member, IEEE) was born in Hangzhou, China, in 1993. He received the B.Sc. degree in mathematics from Beihang University, Beijing, China, in 2015, and the M.Sc. degree in electrical engineering from Peking University, Beijing, in 2018. He is currently pursuing the Ph.D. degree with the Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA, USA. He has been a Research Intern with Intel Corporation, Beijing, in 2017, and a Visiting Researcher with the Department of Intelligence Science and Technology, Kyoto University, Kyoto, Japan, in 2018. His research interests include the fields of wireless communications, graph signal processing, graph neural networks, and optimization theory.

Mark Eisen (Member, IEEE) received the Ph.D. degree in electrical engineering and the master’s degree in statistics from the University of Pennsylvania in 2019. In summer 2013, he was a Research Intern with the Institute for Mathematics and its Applications, University of Minnesota. In summer 2018, he was a Research Intern at Intel Corporation. Since August 2019, he has been working as a Research Scientist at Intel Labs, Hillsboro, OR, USA. His research interests include machine learning, wireless communications, networked control systems, and statistical optimization. He was a recipient of the Outstanding Student Presentation at the 2014 Joint Mathematics Meeting and the 2016 Penn Outstanding Undergraduate Research Mentor Award.

Alejandro Ribeiro (Member, IEEE) received the B.Sc. degree in electrical engineering from the Universidad de la Republica Oriental del Uruguay, Montevideo, Uruguay, in 1998, and the M.Sc. and Ph.D. degrees in electrical engineering from the University of Minnesota, Minneapolis, MN, USA, in 2005 and 2007, respectively. From 1998 to 2003, he was a member of the Technical Staff with Bellsouth Montevideo. After his M.Sc. and Ph.D. degrees, in 2008, he joined the University of Pennsylvania (Penn), Philadelphia, PA, USA, where he is currently the Rosenbluth Associate Professor with the Department of Electrical and Systems Engineering. His research interests include the applications of statistical signal processing to the study of networks and networked phenomena. His focus is on structured representations of networked data structures, graph signal processing, network optimization, robot teams, and networked control. He received the 2014 O. Hugo Schuck Best Paper Award, and paper awards at CDC 2017, 2016 SSP Workshop, 2016 SAM Workshop, 2015 Asilomar SSC Conference, ACC 2013, ICASSP 2006, and ICASSP 2005. His teaching has been recognized with the 2017 Lindback Award for Distinguished Teaching and the 2012 S. Reid Warren, Jr. Award presented by Penn’s Undergraduate Student Body for Outstanding Teaching. He is a Fulbright Scholar Class of 2003 and a Penn Fellow Class of 2015.