Impairment- and fragmentation-aware dynamic routing, modulation and spectrum allocation in C+L band elastic optical networks using Q-learning

Jaya Lakshmi Ravipudi[a,*], Maïté Brandt-Pearce[a]

[a] *Charles L. Brown Department of Electrical and Computer Engineering, University of Virginia, Charlottesville, VA-22904, USA*

Abstract

The paper presents a Q-learning based dynamic routing algorithm for C+L band elastic optical networks (EONs) considering fiber impairments such as cross-phase modulation (XPM), self-phase modulation (SPM), amplified spontaneous emission (ASE), and inter-channel stimulated Raman scattering (ISRS). The effect of fragmentation is considered in the Q-learning process in addition to considering constraints related to spectrum continuity, contiguity, and non-overlapping. Three classical spectrum allocation strategies, first-fit, last-fit, and exact-fit are used after the Q-learning routing algorithm. The proposed routing, modulation, and spectrum allocation (RMSA) approach is shown to have a lower blocking probability compared with using K-shortest path routing combined with the three classical spectrum allocation strategies.

Keywords: Routing, Modulation and Spectrum allocation; Physical layer impairments; Fragmentation; C+L band; Reinforcement learning; Q-learning

1. Introduction

Nearly two-thirds of the world population is forecasted to have internet connectivity by 2023 [1]. The rise in the number of devices, innovative applications, and machine-to-machine communications will cause a 2-4 times increase in traf-

*Corresponding author
Email address: jr3vz@virginia.edu (Jaya Lakshmi Ravipudi)

fic [1]. Therefore, communication networks will need to be used efficiently to accommodate the growing traffic. During the last few years, elastic optical networks (EONs) have been investigated as a promising solution to the inefficient spectrum utilization of traditional wavelength-division multiplexing (WDM) optical networks. The frequency grid of EONs offers finer spectrum slot widths of 12.5 GHz or 6.25 GHz as opposed to the fixed frequency grid of 50 GHz in WDM systems. Hence, EONs generate elastic optical paths that divide the available spectrum flexibly and allocate the available resources in a network according to the traffic demands of the users, leading to efficient utilization of fiber bandwidth.

The task of selecting a route and contiguous spectral slots on each link of that route while avoiding frequency overlapping for a given traffic demand is called the routing and spectrum assignment (RSA) problem in EONs, and has been shown to be NP-hard [2]. When adaptive modulation is considered, the RSA problem becomes the RMSA problem. It is analogous to the routing and wavelength assignment (RWA) problem of WDM, but RSA comes with the additional constraints of spectrum contiguity, spectrum continuity, and non-overlapping of frequency slots, as depicted in Fig. 1(a). Consider a request that needs links 1 through 4 and requires three slots; the selected slots fulfill the adjacency requirement and are the same frequencies on all the links. The traffic demand may be static (i.e., a fixed or offline traffic matrix) or dynamic (i.e., time-varying traffic). In dynamic RSA, dynamically setting up and tearing down connections can lead to bandwidth fragmentation resulting in inefficient spectrum use. Fig 1(b) shows this phenomenon for a demand requesting four slots. They are available on both the links but cannot be assigned due to non-contiguity, thus leading to unnecessary blocking.

EONs offer better spectrum utilization than WDM networks. Still, it is essential for network operators to find ways to utilize the existing resources efficiently and explore new technologies to increase the networks' capacity. The deployment of entirely new optical fibers (multi-core type) is an attractive solution but is not capital cost-friendly at present. A cost-effective solution is

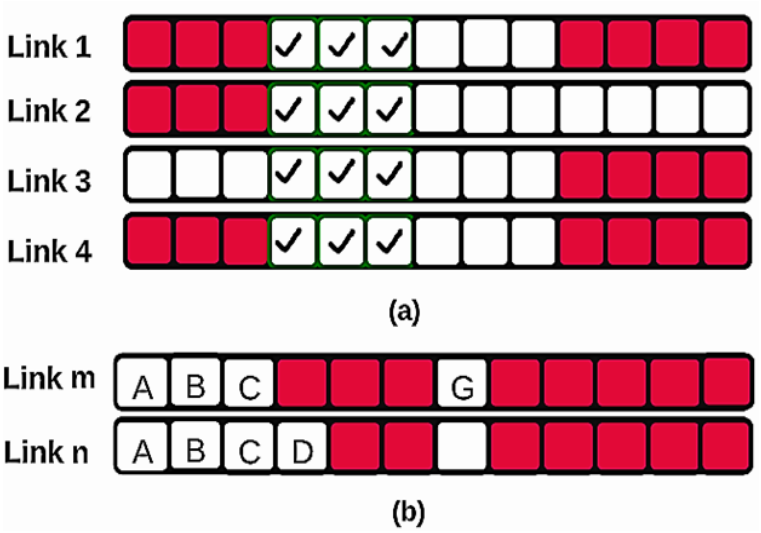


Figure 1: Example showing the (a) contiguity and continuity constraints and (b) fragmentation in EONs

exploiting other frequency bands of operation in the existing single-mode fibers. At present, C-band (5 THz bandwidth) is majorly used by all optical systems. Adding the next band, i.e., L band, to this existing C band would increase the system's capacity to 10 THz. This is possible due to the negligible attenuation coefficient variation in C+L bands and the possibility of using the same Erbium-doped fiber amplifier (EDFA) for the L-band. However, this also introduces inter-channel stimulated Raman scattering (ISRS), resulting in a power transfer from high-frequency components to lower ones and making the optical-signal-to-noise ratio (OSNR) frequency-dependent [3]. Hence, when dealing with the C+L band, it is vital to consider the ISRS nonlinear effects in addition to the usual Kerr nonlinear effects such as the cross-phase modulation (XPM) and self-phase modulation (SPM), as well as the EDFA's amplified stimulated emission (ASE) noise. RSA/RMSA algorithms should incorporate these effects as physical layer impairments (PLI) degrade the quality of transmission (QoT) and thereby limit the transmission reach.

1.1. Related works

Several research works can be found in the literature on solving the RSA problem in EONs. Xu et al. [4] proposed an online-offline algorithm for spectrum assignment of demands with varying bandwidths. To accommodate the randomness of bandwidth demands, the authors proposed a probabilistic PLI model. In another work, Xu et al. [5] proposed a Gaussian noise-based PLI model and a mixed-integer linear programming (MILP) design using a heuristic approach, resulting in resource savings and comparatively higher speeds. Yan et al. [6] investigated the regenerator allocation problem in flex-grid optical networks to deal with PLI and included time as an extra optimization dimension to address time-varying traffic. Wang et al. [7] studied the impacts of using multiple-modulations, regeneration, modulation conversion, and wavelength conversion techniques in EONs using a recursive MILP approach. Chatterjee et al. [8] compared different routing and spectrum allocation approaches and summarized recent works on RSA related issues such as modulation, fragmentation, the traffic grooming, survivability, QoT, energy saving, and networking cost. Adhikari et al. [9] presented a BER and fragmentation-aware RSA algorithm; their simulation results showed that BER-awareness increases the blocking probability, which can be addressed by increasing the transmit power. Abkenar and Rahbar [10] reviewed existing RSA and RMSA (routing, modulation, and spectrum assignment) algorithms and compared them in terms of their computational complexity and quality of performance in resource management. Li and Li [11] presented an RMSA algorithm for EONs with a tradeoff between minimizing the interval between spectrum blocks and the consumed resources. Choudhury et al. [12] described the performances of different routing and spectrum allocation approaches for multicast traffic in elastic optical networks.

Considerable literature exists on fragmentation management [9, 13, 14, 15]. However, most of these works on defragmentation have considered only the C band while assuming hard values for reach and capacity, due to a lack of low computational-complexity QoT estimators.

A few techniques using machine intelligence have been proposed to optimize network routing, as described in recent survey papers. Zhang et al. [16] presented an overview on routing and resource allocation based on machine learning in different optical networks such as WDM, OFDM-based EON, and space division multiplexing (SDM)-based EON. Dai et al. [17] investigated state-of-the-art techniques in machine intelligence-enabled network routing and discussed development trends. Amirabadi [18] reviewed machine learning (ML) applications in optical communications, providing a comprehensive view of ML techniques applicable in this field. Amin et al. [19] surveyed applications of machine learning techniques for routing optimization based on unsupervised learning, supervised learning, and reinforcement learning in software-defined networking. Mammeri [20] provided a comprehensive review of literature on reinforcement learning (RL) applications for optimal route selection in different types of communication networks under various user quality-of-service requirements.

The following works are the closest to our contribution. Yu et al. [21] proposed a deep learning-based RSA strategy and reported that the neural network model had reduced spectrum fragmentation and blocking probability. Shimoda and Tanaka [22] proposed a deep reinforcement learning (DRL)-based RSA algorithm enhanced with domain-specific knowledge. Chen et al. [23] proposed DeepRMSA, a deep reinforcement learning-based neural network for addressing the RMSA problem of EONs. The DeepRMSA learned the correct online RMSA policies by parameterizing the policies with deep neural networks (DNNs) to sense complex EON states; PLIs were considered but limited to modulation format selection based on distance. The same author [24] extended the DeepRMSA to multi-domain EONs and presented a new architecture for network management using multi-agent RL showing better performance than a heuristic-based design. Further, Chen et al. [25] proposed a transfer learning based DeepRMSA that can transfer knowledge of different DRL agents depending on the network tasks. However, efficient network feature extraction still remains a challenge and graph neural networks (GNNs) were cited as a potential solution. Lia and Zhu [26] used GNNs for resource orchestration in elastic

optical datacenter interconnections.

Mitra et al. [27] studied the effect of reduced link margins on C+L band EONs and reported that significant gains in capacity can be achieved by operating at low margins across the networks. Jana et al. [28] proposed a signal-quality-aware proactive defragmentation scheme for C+L band systems using deep neural networks; minimizing the fragmentation index and QoT maintenance was prioritized for both nonlinear-impairment-aware and unaware defragmentation.

1.2. Contribution and paper organization

Based on our literature review, only a few researchers have considered ML approaches such as RL, neural networks (NNs), deep neural networks (DNNs), deep reinforcement learning (DRL), with or without consideration of different impairments. The PLIs considered in previous works do not simultaneously include linear and nonlinear impairments. Furthermore, the NN, DNN, and DRL models presented in current literature are knowledge-intensive, and most of them consider only the C band. These models require a large amount of data and are expensive to train.

The novelty of this paper lies in attempting to adopt a simple model-free Q-learning algorithm, which belongs to the family of RL algorithms. Q-learning has not, to the best of our knowledge, been applied to the optical network routing problem; Q-learning does not require pre-collected training data, can be used by the network controller in real situations and is simple to implement. In addition, the present work considers the ISRS, Kerr nonlinear impairments (SPM, XPM), and the EDFA ASE noise encountered by the signal along the chosen network path. These impairments have rarely been jointly considered [28] for resource provisioning in C+L band operation. The effect of fragmentation of the network is also considered in the Q-learning process. Hence, the significance of the work lies in being the first application of the Q-learning algorithm for optical routing while simultaneously considering PLIs, fragmentation, and the constraints of spectrum continuity, contiguity, and non-overlapping to perform

online RMSA for C+L band EONs.

Our algorithm's performance is compared to a standard K-shortest path algorithm as a frequently-used benchmark. Using the Q-learning routing algorithm results in a lower blocking probability for all spectrum algorithms, network loads, and topologies tested.

The paper is organized as follows. Section 2 explains the physical layer impairments, corresponding modeling, and QoT estimation process. Section 3 briefs about Q-learning and provides a detailed pseudocode of the proposed algorithm. An example demonstration is also given. Section 4 discusses the time and space complexity of the algorithm. Section 5.2 details the simulation and discusses the obtained RMSA results, followed by conclusions in Section 6.

2. Network model and QoT estimation

2.1. Network model

The EON is represented by a graph $G(V, E)$ where V are the nodes and E are the links/edges. This work considers transparent EONs, i.e., the data transmission is entirely in the optical domain, and there is no optical-electrical-optical conversion in the nodes. The physical layer impairments (PLIs) accumulate over the entire lightpath/route and tend to degrade the signal quality (measured using OSNR) at the receiver.

Using different modulation formats for different traffic demands can ensure a proper signal reception. Unlike the case of translucent EONs, where regenerators can be equipped with a modulation conversion facility, we assume that the same modulation format is used along the entire route. Higher-order modulation formats (higher spectral efficiency) are usually used for shorter distances as they are more susceptible to the PLI accumulation if assigned for demands on longer routes.

For a given demand, the modulation format selection, the required number of slots, and the OSNR criterion are interrelated. Each modulation format requires a certain OSNR threshold to keep the bit error rate (BER) within a specified

limit. Once the spectrum assignment is done, the OSNR is predicted as shown in Section 2.2 below and compared with the threshold OSNR. The number of frequency slots of width Δf required for a demand with data rate R using a chosen modulation spectral efficiency η_M, denoted as n_{slots}, is calculated using

$$n_{\text{slots}} = \left\lceil \frac{R}{2 * \Delta f * \eta_M} \right\rceil + 1. \tag{1}$$

One slot is used as a guard band.

Another term that requires defining is fragmentation since it is used as a part of the proposed algorithm. To compute the fragmentation on a particular link of a lightpath, we use an entropy based fragmentation metric given by,[29]

$$\text{Fragmentation}_e = \sum_{j=1}^{k} \left[\frac{w_j^e}{S} \cdot \ln \left(\frac{S}{w_j^e} \right) \right], \tag{2}$$

where k is the number of free fragments on link e, S is the total number of slots of link/edge e and w_j^e is the number of slots in the free fragment j of link e. This is a better metric than the commonly-used external fragmentation (EF) metric [29] w.r.t distinguishing different fragmented links, and also has a lower time complexity compared to the more comprehensive access blocking probability metric (ABP) [29].

2.2. C+L band lightpath model and OSNR estimation

In the current work, the considered C+L band EON consists of bidirectional links of 10 THz bandwidth, with each link divided into spans. A typical network lightpath connection is shown in Fig. 2. The signal is launched with power P_{ch} and travels through a series of intermediate reconfigurable optical add-drop multiplexers (ROADMs) placed at the end of each link. At the end of each span, two Erbium-doped fiber amplifiers amplify the signals, one for the C-band and one for the L-band. The EDFAs are capable of compensating for the ISRS power transfer variations across all the active channels in the C+L band by bringing the power back to P_{ch} and thereby restoring the originally transmitted power spectrum.

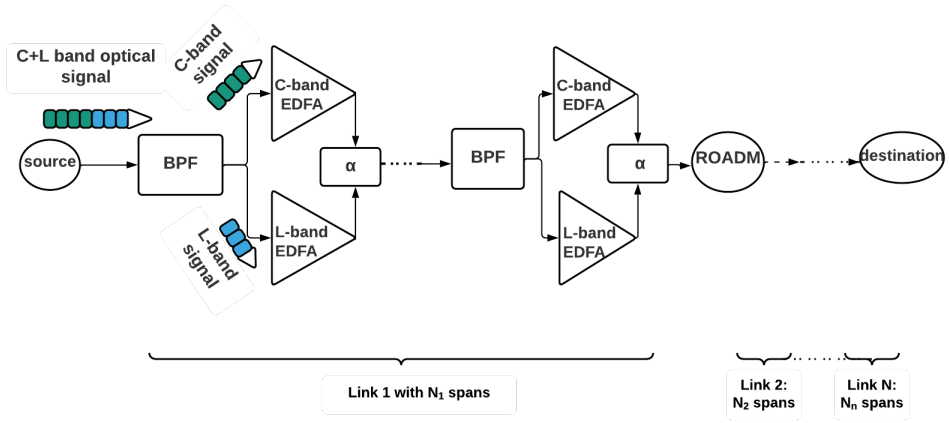


Figure 2: Lightpath example with multiple hops. The symbol α represents the fiber attenuation.

The PLIs are estimated by using mathematical models, and one such widely known model is the Gaussian noise (GN) model. It accounts well for the Kerr nonlinearity effects, but it cannot be directly applied to the C+L band scenario due to the additional ISRS effect in this extended bandwidth. ISRS can be either accounted for by using an extra exponential power decay term or by numerically solving the ISRS differential equations. The first approach is suitable for approximating weak ISRS regimes, and the second approach can approximate any level of ISRS but has a higher computational complexity. The latter necessitates a closed-form model, and one such model was used in [27]. The ISRS gain is modeled using a linear approximation up to 15 THz by using the slope of the normalized Raman gain spectrum. Hence, this can also be applied to our 10 THz C+L band scenario. The current work adopts the OSNR estimation model of [27].

The OSNR of a light path with N_L links is calculated using

$$\frac{1}{OSNR(f)} = \sum_{i=0}^{N_L-1} \left(\frac{P_{ASE}^{(i)}(f) + P_{NLI}^{(i)}(f)}{P_{ch}} \right). \tag{3}$$

$P_{ASE}^{(i)}(f)$ is the ASE noise due to the EDFA present on the i^{th} link of the

light path and $P_{NLI}^{(i)}(f)$ is the NLI power due to self-phase modulation (SPM) and cross-phase modulation (XPM). The signal power spectral density (PSD) is assumed to be rectangular, and so, the NLI power is calculated for the center frequency f of the signal. In the current work, f refers to the center frequency of the set of frequency slots that the spectral assignment algorithm adopted proposes to assign to the request.

The ASE noise generated by each EDFA of the ith span is given by

$$P_{\mathrm{ASE}}^{(i)}(f) \approx 2\eta_{sp} g(f) h f B_{ref}, \tag{4}$$

where η_{sp} is the noise figure of the EDFA and h is the Planck's constant. B_{ref} is the reference bandwidth of the operating ASE noise power measurement and is usually taken as 12.5 GHz [30]. The gain $g(f)$ is a function of frequency and considers the frequency-dependent ISRS gain profile across the C+L band.

The total NLI power generated in the i^{th} optical link with $N_s^{(i)}$ spans is given by

$$P_{NLI}^{(i)}(f) = P_{ch}^3 N_s^{(i)} \left(\eta_{XPM}(f) + \eta_{SPM}(f)\right). \tag{5}$$

The procedure to calculate the NLI coefficients, $\eta_{XPM}(f_z)$ and $\eta_{SPM}(f_z)$ can be found in [27, Eqs. (7)-(11)]. Fully filled channels were considered in [27] to show the effectiveness of the closed-form expressions. For the scenario in this paper, N_{ch} is the total number of demands active on the particular link that can vary as requests arrive and depart.

In this work, the OSNR that the current request would experience if it were to be assigned the particular tentative route is calculated using the model above and compared with the OSNR threshold. Also, alongside satisfying its OSNR constraint, the new and about to be provisioned request should not degrade different existing requests along with various links of the tentative route. Hence, the OSNR constraint also includes checking their corresponding OSNR threshold requirements.

3. Reinforcement and Q-learning based EON routing algorithm

Reinforcement learning (RL) enables learning optimized behavior in a system through the reward-based interaction of an agent with the environment [31]. The reward is positive or negative depending on whether the action by the agent results in the desired behavior or not. The agent in the long run takes actions/decisions that are favorable to get the desired output because it tries to maximize the cumulative rewards. RL has found application in areas such as robotics, gaming, networks, telecommunications, and for building autonomous systems. RL is commonly considered suitable for solving optimization problems related to distributed systems in general and for network routing in particular [32]. Fig 3 shows the general scheme of reinforcement learning.

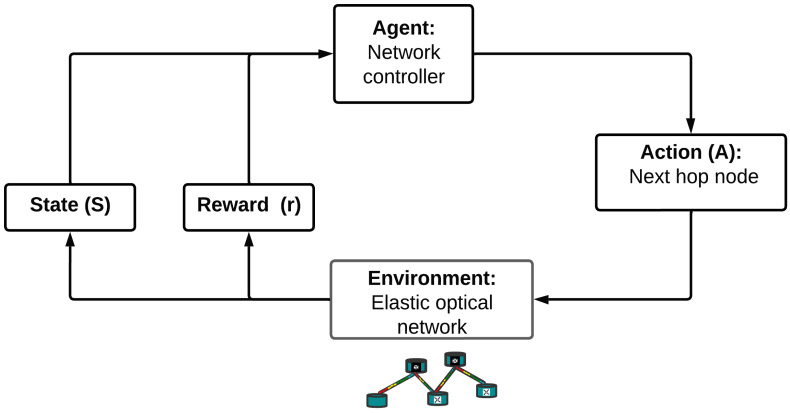


Figure 3: General scheme of reinforcement learning

Q-learning is one of the important breakthroughs in RL [31]. Q-learning algorithms use a state-action table consisting of Q-values that indicate the quality of the action at each state, as shown in Table 1. Each Q-value is denoted by $Q(S, A)$, representing the expected reinforcement of taking action A in state S. The action space contains the actions that the agent can perform, while the state space represents possible system conditions.

Table 1: Q-table used in Q-learning

States	Action 1	Action 2	Action N
State 1	$Q(S_1,A_1)$	—	$Q(S_1,A_N)$
State 2	$Q(S_2,\ A_1)$	—	$Q(S_2,\ A_N)$
State M	$Q(S_M,A_1)$	—	$Q(S_M,A_N)$

The entries in the Q-table are updated using the Bellman equation after an action is taken, given by

$$Q_{\text{new}}(S,A) = (1-\rho)Q(S,A) + \rho\left(r_{t+1} + \gamma \max_{A'}, Q\left(S', A'\right)\right), \tag{6}$$

where $Q_{\text{new}}(S,A)$ is the updated value denoting the quality of action A taken, $Q(S,A)$ is the old value, and r_{t+1} is the reward obtained for taking action A; ρ is the learning rate parameter and γ is the discount factor that determines the importance given to the anticipated future rewards; S' is the state attained after taking the action A; A' is the particular action that has the maximum Q-value among all the possible actions from the given state S'.

This work uses the Q-learning algorithm to perform dynamic routing in C+L EONs. In our EON scenario, the states are the network nodes, $S \in V$, and their respective Q-values are affected by the estimated PLI, fragmentation and link availability as seen by the adjacent nodes. The action is a decision on what the next hop should be given the desired destination, and thus the action space is represented as the pair (network node, destination) $\in (V,V)$. Our routing algorithm includes the effect of fragmentation, link availability and physical layer impairments in the Q-learning process through local (fragmentation based and link availability based) and global (PLI based) rewards. Our approach is modeled on recent literature defining the state and/or action spaces similar to this current work [20] (Table 2 and 3), [33, 34, 35, 36]. However, these prior studies have targeted different applications than ours: we adopt the Q-learning method to perform routing in EONs by including EON-specific system parameters and

performance metrics, which has not been reported in the literature.

After the routing is done by the Q-learning algorithm, the modulation assignment is done based on the length of the chosen tentative route. The spectrum allocation (SA) for each candidate route is then assumed to use one of three well-known algorithms: first-fit, last-fit, and exact-fit. In first-fit SA, the first available frequency slot is selected, i.e., starting from the lowest frequency. In last fit SA, the free slots are checked from the other end of the spectrum, i.e., higher frequencies. Finally, in exact-fit SA, a search for the exact number of required slots is performed; if it is not found, then the first-fit criterion is used.

In RL, an episode is defined as a sequence of states that ends at a terminal state. To adapt the Q-learning algorithm to the proposed resource allocation scenario, an episode is defined as one lightpath selection made by the central network controller, starting from the requested source node and reaching the destination node. As a consequence, we consider our approach to emulate hop-by-hop routing without the network making actual node-level decisions. This is possible due to the automated EON data plane control by the network controller.

The algorithm collects current network information in a Q-table as a result of rewards acquired both during provisioning of other requests and during the current request's episodic runs. It uses a randomizing parameter ϵ that controls whether to explore different paths or exploit the gained knowledge in the Q-table. This ϵ decays over the episodes to make the system take exploratory decisions initially; as the epsilon decays over the episodes, it tends to take more exploitative actions through the knowledge already acquired. Table 2 shows $K = 3$ routes obtained using our Q-routing algorithm for the NSFNET, shown in Fig. 5 (a), for two repeated requests tracked for demonstration purpose. Requests for routes with (source, destination) $= (13, 5)$ result in different candidate lighpaths since the Q-table values vary based on the different reward values. This is not the case for KSP routes. Similarly, when requests for transmission between nodes $(4, 14)$ occur, different routes become candidate choices at different request arrival times. Note that some repetition in the candidates route list is normal since the K routes cannot all be unique to each arrival time;

they all belong to the same subset of all possible paths.

Table 2: $K = 3$ routes obtained by the Q-routing algorithm in the NSFNET for a given (source, destination) pair connection request

(Src,Dest)	**Routing**	**Route 1**	**Route 2**	**Route 3**
(13,5)	KSP	[13,9,8,7,5]	[13,14,1,9,8,7,5]	[13,14,6,5]
	Q-routing	[13,9,8,7,5]	[13,14,6,5]	[13,9,10,6,5]
	Q-routing	[13,14,6,5]	[13,11,4,5]	[13,9,10,6,5]
(4,14)	KSP	[4,11,13,14]	[4,11,12,14]	[4,5,7,8,9,13,14]
	Q-routing	[4,11,13,14]	[4,5,6,14]	[4,2,3,6,14]
	Q-routing	[4,2,3,6,14]	[4,11,12,9,13,14]	[4,11,13,14]
	Q-routing	[4,2,3,6,14]	[4,5,6,14]	[4,5,7,8,9,12,14]

Algorithm 1 gives a pseudo-code of the proposed Q-learning algorithm for EON routing. Consider a particular network provisioning demand where X is the current node, Z the destination node, and Y a next-hop node for X. Then, $Q(X,(Y,Z))$ represent the Q-value for X to reach Z via Y It is important for the Q-value to be a function of the destination node so that the node-by-node learning accounts for how beneficial it is to traverse through a particular Y to reach Z.

Rewards play a major role in steering the decisions of the Q-learning agent and this work uses rewards to avoid routing loops, invalid actions and proper node selections, all enabling destination reaching capability. The actions of non-connected nodes and already visited nodes are penalized and additionally, each episode restricts the search to a maximum number of steps to curb routing loops. Positive rewards are given upon reaching Z based on computed fragmentation, link availability and PLI satisfaction for all remaining action scenarios.

Algorithm 2 shows the PLI-aware RMSA algorithm that uses the output routes of Algorithm 1. The algorithm computes the required frequency slots

based on the chosen modulation format, and checks for continuity and contiguity of the frequency slots. It then checks if the route satisfies the OSNR threshold constraint and accepts it if all constraints are satisfied. If K candidate paths are assumed ($K = 3$ in the present work), then the above procedure is repeated for those K paths until a proper route is found, or else the request is blocked.

3.1. Example scenario

The algorithm's application is explained with an example scenario. Consider a traffic demand from source X to destination Z with a data rate of 200 Gbps. K routes are found for this source to destination pair using Algorithm 1; in our results we use $K = 3$. The first route is selected and the modulation format is assigned based on the length of the route. The number of frequency slots, n_{slots}, is computed using Eq. (1). Then the spectrum allocation part of the algorithm begins. A set of frequency slots are found using either the first fit, last fit, or exact fit SA. The contiguity constraint (the n_{slots} are adjacent on the link), continuity constraint (the n_{slots} are at same spectral positions on each link of the route), and frequency non-overlapping constraint are checked on the selected route. If all constraints are satisfied, then the OSNR constraint is checked, which is $OSNR > OSNR_{\text{threshold}}$.

The $OSNR$ is obtained from Eq. (3); the $OSNR_{\text{threshold}}$ is based on the chosen modulation format. If any of the three constraints were not satisfied or the OSNR condition failed for the selected frequency slots, then the next frequency is selected and again checked. If all the frequencies are exhausted, then the next route in the already-found routes from Algorithm 1 is picked and the conditions are checked. If no suitable route is found even after three paths, the request is blocked. The blocking can be either frequency blocking or PLI blocking.

Whether a route acceptance or rejection happens, rewards are assigned to all the nodes of the route that led to success or failure. If it is a success, the local reward is calculated based on the fragmentation occurring on each of the chosen links (Eq. (2)) and the links' availability, i.e., number of unoccupied

Algorithm 1: Q-learning-based EON routing algorithm

Data: source = X, destination = Z, Q-table (initialized to 0)

Result: K valid routes from X to Z

Start

Repeat

Set Visited nodes = [ ]

Set Current node = X

Set Candidate_route = [ ]

Repeat

Obtain Q(possible_actions) where possible_actions $\in (|V|, Z)$

If exploitation is True

action = argmax$_{(|V|,Z)}$ {Q(possible_actions)}

Else exploration is True

action = random number $\in |V|$

End if

Get neighbors of Current node

If action $\in$ neighbors

Append action to final_route

Get reward for (Current node, action)

Update Visited nodes

Perform Bellman Q-value update (Eq. (6)), where S' = action and $A' \in (|V|, Z)$)

Do Current node = action

End

Get a negative reward

Perform Bellman Q-value update (Eq. (6))

End If

Until Current node = Z (i.e. terminal state)

Store the route of current episode if it is a feasible path

Until Maximum number of episodes reached

Extract K-1 more paths from stored set of feasible path from over the episodes

Algorithm 2: PLI aware RMSA scheme based on proposed routing algorithm

Data: source = X, destination = Z, requested bandwidth

Result: A valid RMSA solution

Start

Obtain K paths from Algorithm 1

Set route counter $k = 0$

Repeat

Obtain modulation format, number of frequency slots (Eq. (1)), and OSNR_threshold

Find feasible spectrum slots using First Fit/Last Fit/Exact Fit

If spectrum constraints are satisfied

Compute OSNR (Eq. (3))

If OSNR_computed $\geq$ OSNR_threshold

Assign the light path (LP) resources.

Compute the local reward for each node of the LP: fragmentation (Eq. (2)) and available spectral slots on the corresponding link

Compute a positive global reward if PLI constraint satisfied, else a negative global reward

Add the global to reward to each local reward

Perform Bellman update similar to Algorithm 1

Else

Check on next feasible spectrum block

Else

Check on next feasible spectrum block

Set k=k+1

Until a valid route satisfying all constraints is found or $k = K$ (request is blocked)

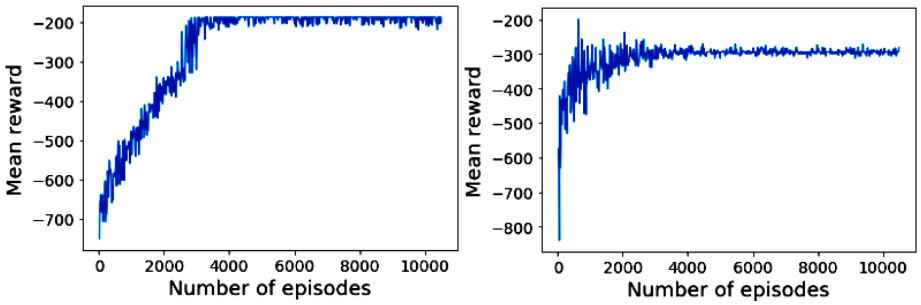
frequency slots. These two are considered in the reward calculation since high fragmentation and low link availability can lead to higher blocking. If the chosen route has three nodes, then the local reward will be a vector of three elements. Including this information is a way of praising or criticizing the action taken to choose that particular node's link for any next request to the same destination. The fragmentation reward is negative since the Q-learning algorithm tries to maximize the reward and fragmentation values are desired to be low. As a side note, this approach could be extended to EON using multimode or multicore fibers, where the fragmentation computation is different than for the single-core single-mode RMSA case [37] [38]. Perhaps a more carefully crafted reward functions can then be used in order to capture the fragmentation differences in the spatial modes/cores rather than considering a single value in the reward.

In our Q-learning algorithm, the global reward is related to OSNR satisfaction. A global reward of +1 or -1 is added to the above-decided local reward for each node on the route depending on if the OSNR constraint is satisfied or not. The total rewards (sum of the local and global reward) are then used to update the Bellman equation, Eq. (6), for each node of the considered route. It is these same node values that the Q-learning EON routing later uses to tentatively choose a path at each node, thus emulating hop-by-hop routing for upcoming requests.

4. Time and Space Complexity

The worst case time complexity of Algorithm 1 can be deduced as follows. The Q-table is stored in the form of a dictionary that is implemented as a hash table to make it less time consuming to find the desired Q-value using the state as the key. The time complexity to search one value in a dictionary is O(1). Hence, the worst case scenario for a given destination is when all the states in the state and action space are visited and the complexity would then be $O\left(|V|^2\right)$.

The size of the Q-table in the proposed algorithm is $|V| \times |V| \times |V|$ because



(a) (b)

Figure 4: Obtained mean reward over the number of episodes for two source and destination examples, given in (a) and (b).

for each of $|V|$ state node there are $|V|$ potential next hops on the way to $|V|$ potential destinations. This size is reasonable for Q-learning provided all the entries are visited and updated [31]. This condition is fulfilled in the considered network scenario since, as part of the training and during provisioning, the source/destination pairs are randomly generated for thousands of requests and node-by-node tabular updates are performed for both success and failure scenarios.

5. Results and Analysis

5.1. Training Setup

As mentioned in Section 3, a Q-table is maintained by the centralized controller. Multiple episodes are run in order for the Q-table to converge even though a route may be found before all episodes are considered. This also helps to finalize a better route than one found along the episode iterations. Fig. 4 shows the mean rewards obtained for two example source/destination pairs, where the data has been downsampled for plotting purposes. For any given source and destination, convergence was always reached before 10,000 episodes. Note that the overall trend of the rewards is increasing; reward values are more often positive than negative.

5.2. Network results

Two networks are used to test the performance of the proposed Q-learning algorithm for routing combined with the three classical spectrum allocation strategies. The networks are the 14-node NSFNET and 11-node COST-239. Fig. 5(a) shows the 14-node NSFNET with 21 bidirectional optical fiber links and the European 11-node COST-239 network with 52 optical fiber links is shown in Fig. 5(b). Both these networks have a wide diversity of link distances (200 km - 2400 km), which is essential while considering the inclusion of PLI. This work assumes a 10 THz (C+L band) optical spectrum for both networks. Table 3 presents the parameters used in the current work.

Table 3: System and fiber parameters used in the current work [27]

Parameters	Values
Fiber loss, dB/km	0.2
Dispersion, ps/nm/km	17
Dispersion slope, ps/nm^2/km	0.067
Nonlinear coefficient, 1/W/km	1.2
Raman gain slope, 1/W/km/THz	0.028
Raman gain, 1/W/km	0.4
Channel launch power, dBm	0
Number of channels	Variable
Optical bandwidth, THz	10
Slot bandwidth, GHz	12.5

Requests arrive in a Poisson manner with an exponential holding time, and the input traffic load is measured in Erlangs. A request is composed of a source node, destination node, and data rate. Data rates considered are between 50 Gbps to 300 Gbps. Modulation formats considered are BPSK, QPSK, 8QAM, 16-QAM and 32-QAM; the threshold OSNRs are 9 dB, 12 dB, 16 dB, 18.6 dB,

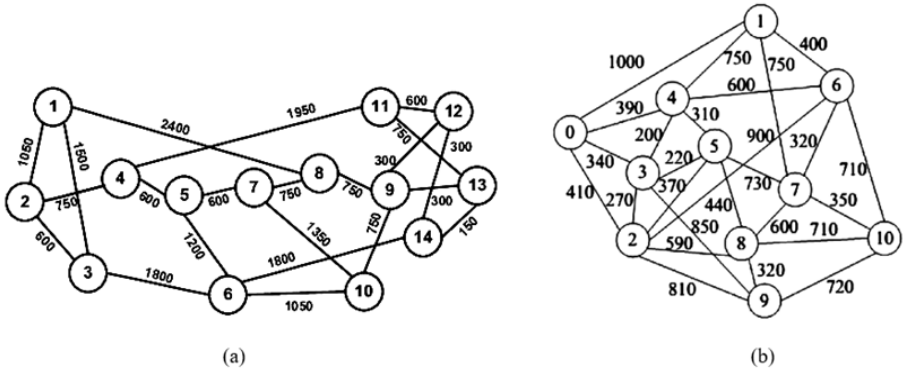


Figure 5: (a) 14-node NSFNET (b) 11-node COST 239 network

and 21.6 dB, respectively [39]. The algorithm's effectiveness is measured by calculating the blocking probability (BP), bandwidth blocking probability (BBP) and network fragmentation (NF). The BP, BBP and NF [40] are given by

$$\text{BP} = \frac{\text{Number of blocked requests}}{\text{Total number of requests}} \tag{7}$$

$$\text{BBP} = \frac{\text{Amount of blocked bandwidth}}{\text{Total amount of requested bandwidth}} \tag{8}$$

$$\text{NF} = \frac{1}{|E|}\sum_{e=1}^{|E|}(\text{Fragmentation}_e) \tag{9}$$

For each trial, 300,000 requests are considered to determine these network metrics.

In optical networks, PLI's are typically addressed either through a PLI-aware algorithm, as described above, or using an algorithm that ignores the PLI and then performs a final quality PLI-check just before provisioning; we compare our approach to KSP under each assumption. The PLI effects considered are the fiber Kerr nonlinearity (SPM and XPM) and ASE noise, including the ISRS effect. In the PLI-check approach, the OSNR constraint is checked after the

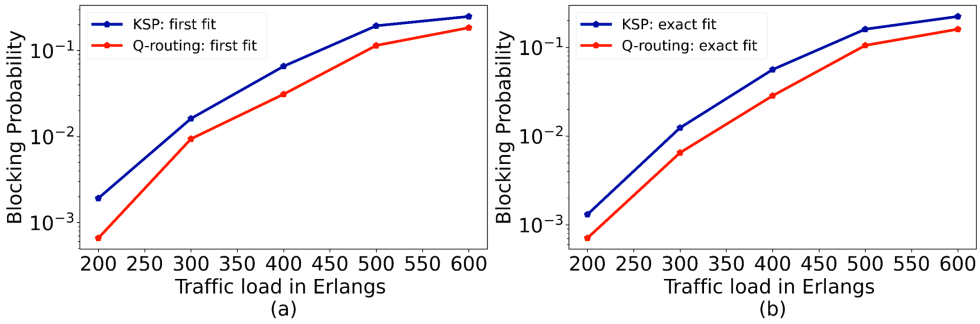


Figure 6: Blocking probabilities in NSFNET for PLI-check type RMSA considering KSP routing and Q-learning EON routing using (a) first fit and (b) exact fit for spectrum allocation.

algorithm finds a route and a spectrum allocation. If the constraint is satisfied, the request is provisioned, but no attempt is made to find another route if the condition fails. This process is like the basic RMSA problem but with the additional OSNR constraint. This checking increases the blocking when compared to basic RMSA algorithms, but in a practical network, there is no use in provisioning a request just based on the availability of resources without considering physical layer aspects. Conversely, in the PLI-aware type RMSA, OSNR constraints are checked as a part of the RMSA problem, and re-attempts are made to find another route and/or spectrum allocation (still among the K paths) if the OSNR constraint fails.

Simulation results of blocking probabilities with KSP and Q-learning PLI-check routing using first and exact fit allocation in the NSFNET topology are given in Fig. 6. The last fit performance is almost identical to the first fit and hence not shown here. As the load increases, a higher resource blocking or QoT blocking is inevitable irrespective of the chosen routing or spectrum strategy. For all three SA algorithms tested, the Q-learning routing performed better than the traditional KSP routing.

When requests arrive repeatedly with sources and destinations whose routes include many common crowded links, shortest path routing leads to repeated blocking due to unavailability of frequency slots or QoT degradation to existing

requests on these links. Hence, always following the shortest-path rule is not wise in high-usage network operations. Conversely, in the proposed Q-learning routing for C+L EON, the Q-values consider the estimated link fragmentation, link availability, and estimated PLI effects associated with each node. As tentative routes are predetermined, the decision of how to hop from one node to the next is optimized, thus finding the best Q-valued routes. Unlike the KSP PLI-check case, the likelihood of failure is reflected in the Q-values of the nodes of a failed route. This leads to the selection of different routes over time since the network state changes and gets reflected in the Q-values through the rewards. Consequently, the agent chooses the path that is in the best interest of the current request being considered and also for better service provisioning in long-term operation. This is supported by the simulation results shown in Fig. 6.

Fig. 7 shows the blocking probabilities obtained using KSP and Q-learning routing in the NSFNET for the PLI-aware RMSA scenario. The results are plotted with 95% confidence intervals, confirming the stability of our measurements. The blocking probabilities for these algorithms are lower than for the PLI-check versions, as expected because there is a re-attempt to find another route (among the K routes) and/or frequency spectrum if the OSNR constraint is violated. Q-learning routing continues to perform better than the conventional KSP routing even at at high loads. The proposed algorithm achieves 1% blocking at higher loads (around 120 Erlangs higher) than the benchmark. Even with the inclusion of PLI constraints, the Q-learning EON routing can adapt well in the long run. This behavior is explained as follows. For the KSP algorithm, routing decisions affect the blocking probability but is unaffected by spectrum decisions since routing decisions are solely based on distance. On the other hand, for the Q-learning routing, the spectrum affects the routing because the success or failure of the route is included in the Q-learning process through local (fragmentation, link availability-based) and global (PLI-based) rewards, affecting the routing of future requests.

Fig. 8 shows the network fragmentation for the PLI-aware algorithm on the

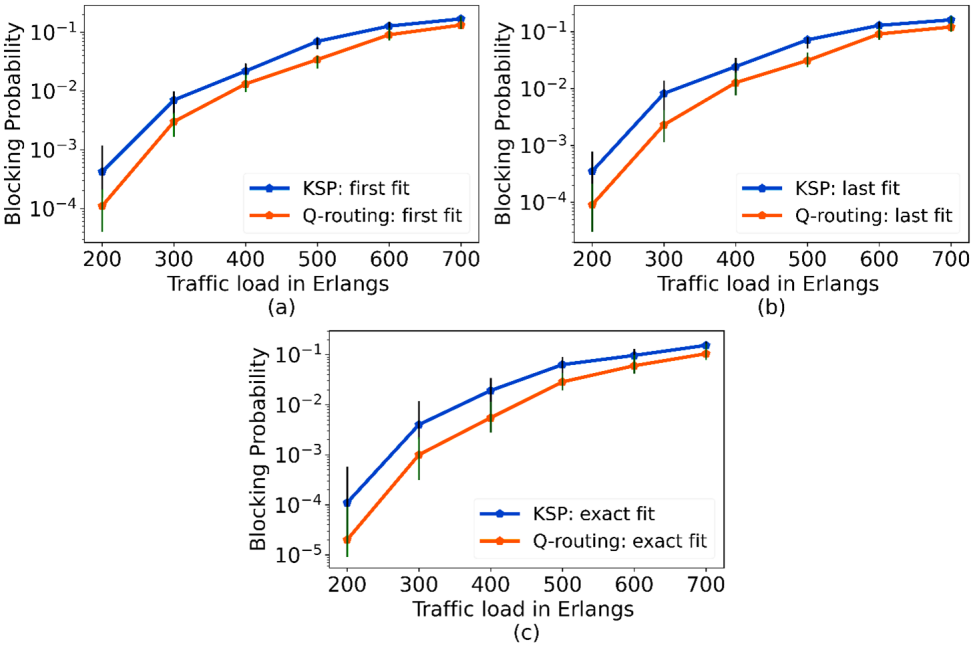


Figure 7: Blocking probabilities for PLI-aware RMSA with (a) first fit (b) last fit and (c) exact fit for spectrum allocation in NSFNET.

NSFNET network. At each load, the network fragmentation for every 10,000 requests was collected and then averaged to obtain each of the points on the graph. The proposed Q-learning based RMSA algorithm is able to maintain lower fragmentation levels because rewards are calculated based on the fragmentation that occurs after provisioning, and the Q-learning agent tries to learn to minimize the fragmentation. Note that the considered metric, Eq. (2) is a relative measure and not bounded between 0 to 1.

Fig. 9 depicts the bandwidth blocking probability for the PLI-aware NSFNET scenario. The Q-learning based routing results in lower BBP when compared to the KSP for all loads tested. The improvement is smaller at higher loads because the fragmentation and PLI effects are higher and link availability is lower.

To verify that the proposed Q-learning RMSA algorithm performs well in multiple scenarios, a similar traffic pattern as used on the NSFNET topology is

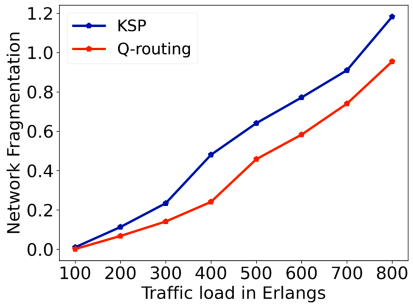


Figure 8: Network fragmentation for PLI-aware RMSA in the NSFNET network.

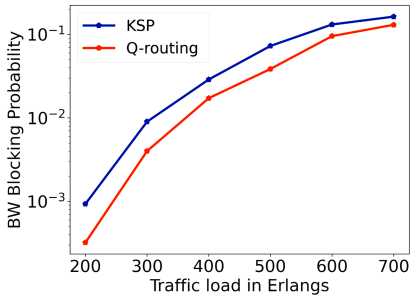


Figure 9: Bandwidth blocking probabilities for PLI-aware RMSA in the NSFNET network.

given as input to the European COST-239 network. Fig. 10 shows the blocking probabilities obtained using KSP and Q-learning for PLI-aware routing for this topology. Similar conclusions as the PLI-check scenario of NSFNET topology are applicable here as well; hence, only exact fit results are shown. In this network topology as well, the Q-learning routing algorithm is more robust than KSP routing and maintains lower blocking probabilities. Note that the blocking is minutely higher in the COST-239 network when compared to the NSFNET network. In both cases, the proposed algorithm outperforms the traditional KSP algorithm.

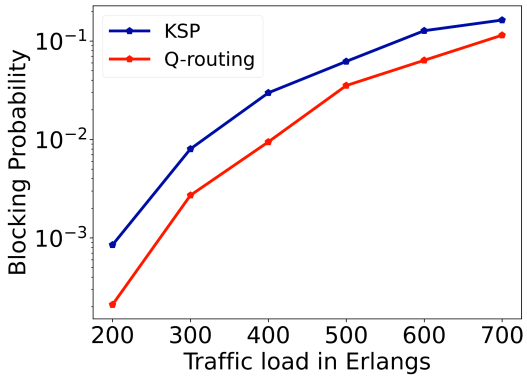


Figure 10: Blocking probabilities for PLI-aware COST-239 network with exact fit SA.

The better performance of the proposed algorithm in both continental-sized

networks evidences our claim that Q-learning routing can perform better than widely-used KSP routing. As the primary intention of moving towards C+L band EON operation is to support higher capacity traffic, our results showing improved performance at high loads with the inclusion of realistic PLI provides a notable contribution towards the realization of this emerging technology.

6. Conclusions

Using Q-learning for routing combined with classical spectrum allocation strategies of first-fit, last-fit, and exact-fit is investigated in C+L band EONs. The Q-learning algorithm belongs to the family of RL techniques. The present work has considered the ISRS effect, Kerr nonlinear impairments such as SPM and XPM, and also the ASE noise encountered by the signal along the chosen network path. The effect of fragmentation and link availability of the network is also considered through the reward-based interactions of Q-learning. Thus, the current work is PLI- and fragmentation-aware while choosing the route in a C+L band EON. To the best of our knowledge, this is the first application of the Q-learning algorithm for routing considering the PLIs, fragmentation, and the constraints of spectrum continuity, contiguity, and non-overlapping to perform online RMSA for C+L band EONs.

The simulations are performed on two topographically diverse topologies, NSFNET and COST-239, and the results are analyzed. In general, Q-learning routing performed better than the K-shortest path algorithm. The proposed Q-learning for routing is simple and can be used easily by operators in real situations.

As part of future work, additional link information can be incorporated in the state representation to make the routing more adaptive while crafting better reward functions. The comparisons of the proposed Q-learning with other routing approaches and its extension to spectrum allocation are also the subject of ongoing research.

7. Acknowledgment

This work was supported in part by the National Science Foundation grant CNS-1718130.

References

[1] Cisco, Cisco Annual Internet Report (2018–2023) White paper, Online (accessed March 26, 2021) https://www. cisco. com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/whitepaper-c11-741490. html.

[2] K. Christodoulopoulos, I. Tomkos, E. A. Varvarigos, Elastic bandwidth allocation in flexible OFDM-based optical networks, Journal of Lightwave Technology 29 (9) (2011) 1354–1366.

[3] G. Saavedra, D. Semrau, M. Tan, A. Iqbal, D. J. Elson, L. Galdino, P. Harper, R. I. Killey, P. Bayvel, Inter-channel stimulated Raman scattering and its impact in wideband transmission systems, in: 2018 Optical Fiber Communications Conference and Exposition (OFC), IEEE, 2018, pp. 1–3.

[4] Y. Xu, E. Agrell, M. Brandt-Pearce, Cross-layer static resource provisioning for dynamic traffic in flexible grid optical networks, Journal of Optical Communications and Networking 13 (3) (2021) 1–13.

[5] Y. Xu, L. Yan, E. Agrell, M. Brandt-Pearce, Iterative resource allocation algorithm for EONs based on a linearized GN model, Journal of Optical Communications and Networking 11 (3) (2019) 39–51.

[6] L. Yan, Y. Xu, M. Brandt-Pearce, N. Dharmaweera, E. Agrell, Robust regenerator allocation in nonlinear flexible-grid optical networks with time-varying data rates, IEEE/OSA Journal of Optical Communications and Networking 10 (11) (2018) 823–831.

[7] X. Wang, M. Brandt-Pearce, S. Subramaniam, Impact of wavelength and modulation conversion on translucent elastic optical networks using MILP, Journal of Optical Communications and Networking 7 (7) (2015) 644–655.

[8] B. C. Chatterjee, N. Sarma, E. Oki, Routing and spectrum allocation in elastic optical networks: A tutorial, IEEE Communications Surveys & Tutorials 17 (3) (2015) 1776–1800.

[9] D. Adhikari, D. Datta, R. Datta, Impact of BER in fragmentation-aware routing and spectrum assignment in elastic optical networks, Computer Networks 172 (2020) 107167.

[10] F. S. Abkenar, A. G. Rahbar, Study and analysis of routing and spectrum allocation (RSA) and routing, modulation and spectrum allocation (RMSA) algorithms in elastic optical networks (EONs), Optical Switching and Networking 23 (2017) 5–39.

[11] L. Li, H.-j. Li, Performance analysis of novel routing and spectrum allocation algorithm in elastic optical networks, Optik 212 (2020) 164688.

[12] P. D. Choudhury, P. R. Reddy, B. C. Chatterjee, E. Oki, T. De, Performance of routing and spectrum allocation approaches for multicast traffic in elastic optical networks, Optical Fiber Technology 58 (2020) 102247.

[13] P. Wright, M. C. Parker, A. Lord, Minimum-and maximum-entropy routing and spectrum assignment for flexgrid elastic optical networking, Journal of Optical Communications and Networking 7 (1) (2015) A66–A72.

[14] R. Wang, B. Mukherjee, Spectrum management in heterogeneous bandwidth optical networks, Optical Switching and Networking 11 (2014) 83–91.

[15] R. Zhu, S. Li, P. Wang, J. Yuan, Time and spectrum fragmentation-aware virtual optical network embedding in elastic optical networks, Optical Fiber Technology 54 (2020) 102117.

[16] Y. Zhang, J. Xin, X. Li, S. Huang, Overview on routing and resource allocation based machine learning in optical networks, Optical Fiber Technology 60 (2020) 102355.

[17] B. Dai, Y. Cao, Z. Wu, Z. Dai, R. Yao, Y. Xu, Routing optimization meets Machine Intelligence: A perspective for the future network, Neurocomputing 459 (2021) 44–58.

[18] M. Amirabadi, A survey on machine learning for optical communication [machine learning view], arXiv preprint arXiv:1909.05148.

[19] R. Amin, E. Rojas, A. Aqdus, S. Ramzan, D. Casillas-Perez, J. M. Arco, A survey on machine learning techniques for routing optimization in SDN, IEEE Access.

[20] Z. Mammeri, Reinforcement learning based routing in networks: Review and classification of approaches, IEEE Access 7 (2019) 55916–55950.

[21] J. Yu, B. Cheng, C. Hang, Y. Hu, S. Liu, Y. Wang, J. Shen, A deep learning based RSA strategy for elastic optical networks, in: 2019 18th International Conference on Optical Communications and Networks (ICOCN), IEEE, 2019, pp. 1–3.

[22] T. Shimoda, Routing and spectrum assignment using deep reinforcement learning in optical networks, NTT Technical Review (2021) 1–5.

[23] X. Chen, B. Li, R. Proietti, H. Lu, Z. Zhu, S. B. Yoo, DeepRMSA: a deep reinforcement learning framework for routing, modulation and spectrum assignment in elastic optical networks, Journal of Lightwave Technology 37 (16) (2019) 4155–4163.

[24] X. Chen, R. Proietti, S. B. Yoo, Building autonomic elastic optical networks with deep reinforcement learning, IEEE Communications Magazine 57 (10) (2019) 20–26.

[25] X. Chen, R. Proietti, C.-Y. Liu, S. B. Yoo, A multi-task-learning-based transfer deep reinforcement learning design for autonomic optical networks, IEEE Journal on Selected Areas in Communications 39 (9) (2021) 2878–2889.

[26] B. Li, Z. Zhu, GNN-based hierarchical deep reinforcement learning for nfv-oriented online resource orchestration in elastic optical dcis, Journal of Lightwave Technology 40 (4) (2021) 935–946.

[27] A. Mitra, D. Semrau, N. Gahlawat, A. Srivastava, P. Bayvel, A. Lord, Effect of channel launch power on fill margin in c+ l band elastic optical networks, Journal of Lightwave Technology 38 (5) (2019) 1032–1040.

[28] R. K. Jana, B. C. Chatterjee, A. P. Singh, A. Srivastava, B. Mukherjee, A. Lord, A. Mitra, Machine learning-assisted nonlinear-impairment-aware proactive defragmentation for C+L band elastic optical networks, Journal of Optical Communications and Networking 14 (3) (2022) 56–68.

[29] B. C. Chatterjee, S. Ba, E. Oki, Fragmentation problems and management approaches in elastic optical networks: A survey, IEEE Communications Surveys & Tutorials 20 (1) (2017) 183–210.

[30] S. Kumar, M. J. Deen, Fiber optic communications: fundamentals and applications, John Wiley & Sons, 2014.

[31] R. S. Sutton, A. G. Barto, Reinforcement learning: An introduction, MIT press, 2018.

[32] J. A. Boyan, M. L. Littman, Packet routing in dynamically changing networks: A reinforcement learning approach, in: Advances in neural information processing systems, 1994, pp. 671–678.

[33] Z. Zheng, A. K. Sangaiah, T. Wang, Adaptive communication protocols in flying ad hoc network, IEEE Communications Magazine 56 (1) (2018) 136–142.

[34] X.-J. Shen, Q. Chang, L. Liu, J. Panneerselvam, Z.-J. Zha, CCLBR: Congestion control-based load balanced routing in unstructured P2P systems, IEEE Systems Journal 12 (1) (2016) 802–813.

[35] C. Wu, T. Yoshinaga, Y. Ji, Y. Zhang, Computational intelligence inspired data delivery for vehicle-to-roadside communications, IEEE Transactions on Vehicular Technology 67 (12) (2018) 12038–12048.

[36] M. Boushaba, A. Hafid, A. Belbekkouche, M. Gendreau, Reinforcement learning based routing in wireless mesh networks, Wireless networks 19 (8) (2013) 2079–2091.

[37] B. C. Chatterjee, A. Wadud, E. Oki, Proactive fragmentation management scheme based on crosstalk-avoided batch processing for spectrally-spatially elastic optical networks, IEEE Journal on Selected Areas in Communications 39 (9) (2021) 2719–2733.

[38] B. C. Chatterjee, A. Wadud, I. Ahmed, E. Oki, Priority-based inter-core and inter-mode crosstalk-avoided resource allocation for spectrally-spatially elastic optical networks, IEEE/ACM Transactions on Networking 29 (4) (2021) 1634–1647.

[39] A. Mitra, D. Semrau, N. Gahlawat, A. Srivastava, P. Bayvel, A. Lord, Effect of reduced link margins on C+L band elastic optical networks, Journal of Optical Communications and Networking 11 (10) (2019) C86–C93.

[40] D. S. Yadav, RDRSA: A reactive defragmentation based on rerouting and spectrum assignment (RDRSA) for spectrum convertible elastic optical network, Optics Communications 496 (2021) 127–144.