

Deep Reinforcement Learning for Resource Allocation in Multi-Band Optical Networks

Abdenmour Ben Terki*, João Pedro[†], António Eira[†], Antonio Napoli[§], Nicola Sambo*

*Scuola Superiore Sant'Anna, Pisa, Italy, [†]Infinera Unipessoal Lda, Carnaxide, Portugal

[‡]Instituto de Telecomunicações, IST, Lisboa, Portugal, [§]Infinera, Munich, Germany

Corresponding author: abdenmour.benterki@sssup.it

Abstract—Routing and Spectrum Assignment (RSA) is key to an efficient resource usage in optical networks. Although this problem is known to be complex, an even more complex version arises when considering multi-band (MB) optical networks, where the spectrum-dependency of performance becomes significantly more pronounced. This paper proposes a Deep Reinforcement Learning (DRL)-based strategy for RSA in MB optical networks leveraging the GNPpy library for accurate estimation of optical performance. Simulation results show that DRL-RSA reduces blocking by up to 80% when comparing to state-of-the-art RSA strategies.

Index Terms—routing and spectrum assignment, multi-band optical networks, blocking probability, deep reinforcement learning, optical performance

I. INTRODUCTION

In recent years, global data traffic has seen rapid growth, which placed a burden on existing optical network infrastructures. To address this issue, the exploitation of Multi-Band (MB) transmission in deployed optical networks has emerged as a promising solution to increase the network capacity and meet the surge in demand for more bandwidth, while mitigating/postponing the need to lease/roll-out additional fibers [1]. Yet, with the advantages of MB optical networks come new challenges. As the available spectrum resources increase, the network design and operation complexity grows due to the need of considering multiple frequency bands, a significantly larger number of channels and more disparate performance differences between channels of different bands. This added complexity impacts Routing and Spectrum Assignment (RSA), which is one of the most critical tasks in the control of the network and in maintaining an efficient usage of resources. Traditional RSA algorithms such as k -Shortest Path (k -SP) for routing and First-Fit (FF) for spectrum assignment have been widely investigated and adopted in commercial deployments. More recently, machine learning (ML) techniques have been considered to replace/complement traditional RSA algorithms, especially in complex systems with a large amount of resources and non-simple physical layer constraints, as the ones present in MB optical networks.

Deep Reinforcement Learning (DRL) [2], [3] can be an interesting solution for RSA due to its ability to learn traffic

patterns and deduce a RSA strategy through a series of trials and errors by interacting with the network, receiving feedback in the form of reward if a path and channel have been successfully assigned to the source-destination connection request or a penalty if the request was blocked. In [4], the authors introduce the DeepRMSA framework, a strategy based on Deep Reinforcement Learning (DRL), designed for the Routing, Modulation, and Spectrum Assignment (RMSA) problem within C-band optical networks. The authors in [5] employ a latency-aware RSA mechanism based on DRL to jointly address the spectrum utilization and delay constraints in the network. The work presented in [6] presents a DRL algorithm addressing the Routing, Modulation, Spectrum, and Core Allocation (RMSCA) problem in optical networks with multicore fiber. Furthermore, The study in [7] focuses on extending the challenges associated with the Routing, Band, Modulation, and Spectrum Assignment (RBMSA) problems using DRL to MB optical networks, however, the DRL-based strategies proposed in this study did not show a better performance when they were compared to a heuristic strategy (i.e., k -SP FF FF) in terms of blocking probability.

This paper presents an RSA strategy based on DRL, specifically designed for multi-band optical networks. The physical layer model utilized for the Quality of Transmission (QoT) estimation is the Generalized Gaussian Noise (GGN) model. This model accounts for wide-band impairments like the Stimulated Raman Scattering (SRS) and a widely used and validated implementation is available in the GNPpy tool [8]. In addition, a new reward function for the DRL agent has been created considering the assigned path and channel in terms of hops and frequency slots. Simulation results show that the DRL-based strategy has the ability compared to benchmark algorithms such as k -SP FF and RL [9], to reduce the blocking probability (BP), further increasing the throughput of MB optical networks.

II. PROPOSED APPROACH FOR ROUTING AND SPECTRUM ASSIGNMENT

DRL is a learning approach in which the *agent* (in our case the RSA model) continuously interacts with its *environment* (the MB network). For every incoming (source, destination) connection request, the agent observes the current state of the network encapsulated in the *observation* space, which contains for each (source, destination) pair information about available

This work has received funding from EU Horizon 2020 MENTOR program under the Marie Skłodowska Curie grant agreement 956713 and from the European Union's Horizon RIA research and innovation program under grant agreement 101096120 (SEASON).

TABLE I
DRL TUNED PARAMETERS.

DRL Parameter	Tuned Value
Learning Rate (α)	5.32×10^{-4}
Discount Factor (γ)	9.86×10^{-1}
Clip Range	1.64×10^{-1}
Entropy Coefficient	2.82×10^{-4}
Episode Length	Traffic Load
Total Number of Steps	$1200 \times \text{Traffic Load}$
Policy Optimizer	PPO [3]

spectrum resources on all the precomputed paths and over all the bands. Next, the agent takes an *action*, which is a tuple of (path, band, channel) based on the observation and his knowledge gained over time. The environment moves to another state according to the agent's action selection and returns a *reward*, as shown in Algorithm 1, which is positive if the request (source, destination) was established or negative if the request was blocked. This reward feedback loop is what drives the agent to learn and refine its RSA strategy over time so as to maximize the number of positive rewards received which should translate into a reduction of blocked requests.

Algorithm 1 DRL reward function

Require: (*path, channel*) \leftarrow *Reward*
 $\alpha \leftarrow$ Number of allocated channels
 $\beta \leftarrow$ Number of hops in the selected path
if the connection can be established **then**
 Reward $\leftarrow 1 + \frac{1}{\alpha \times \beta}$
else if a connection cannot be established on that path and channel **then**
 Penalty $\leftarrow -1$
end if

In this study, in order to explore the capability of DRL for RSA in optical networks, the Optical RL-Gym toolkit [2], [3] has been used. Optical RL-Gym is an open source and flexible toolkit for applications of DRL models to solve the RSA problem. Two additional upgrades have been conducted to the online available Optical RL-Gym toolkit in the context of our scenario: (i) The integration of the GNPpy tool for the physical layer impairments model. (ii) A new reward function, described in Algorithm 1, returning a reward that takes into account the allocated path and the selected channel format in terms of hops and frequency slots used, respectively. The DRL model has been trained and tuned to ensure optimal performance. Following an initial offline training using a digital twin of the real network with multiple traffic loads where the agent receives a set of incoming source-destination node connection requests with different time of arrival and holding times, the DRL agent leverages its acquired knowledge to dynamically assign path and spectrum in the real network. In the event of network modifications, such as the installation of new sites or the addition of new links, the agent can undergo a retraining process offline on an updated digital twin of the new network, using the knowledge obtained from the old

network to adapt effectively to changes. The tuned parameters and their values are shown in Table I. The total number of training steps depends directly on the network traffic load, starting from 240×10^3 to 540×10^3 steps for traffic loads of 200 and 450 Erlang, respectively.

III. SIMULATION RESULTS

The DRL-based RSA strategy is analyzed through simulations using the optical RL-Gym toolkit [2] and compared to k -SP FF and RL-based RSA strategies (Q-Learning) [9]. The Japanese network topology [10] consisting of 14 nodes and 44 links is considered. Traffic follows a Poisson distribution with rate λ . Connection holding time is exponentially distributed with an average of $1/\mu = 60$ minutes. The traffic load (λ/μ) is varied with λ between 200 and 450 Erlang. 400 Gb/s requests are assumed, which can be served via a single dual polarization 16 quadrature amplitude modulation (DP-16QAM) channel 75 GHz or 2×200 Gb/s dual-polarization quadrature phase shift keying (DP-QPSK) channels over 150 GHz. The Generalized Signal-to-Noise Ratio (GSNR) is computed for each individual channel with GNPpy [8] and accounting for SRS. An L-C-S-E multi-band system is assumed with the supported spectrum as described in [1] and summarized in II. The adopted GSNR thresholds are 24 dB for DP-16QAM and 16 dB for DP-QPSK, assuming a symbol rate of 64 GBaud and channel spacing of 75 GHz. Table II summarizes the parameters of the multi-band system considered, including the per-band fiber attenuation range and amplifier noise figure [1].

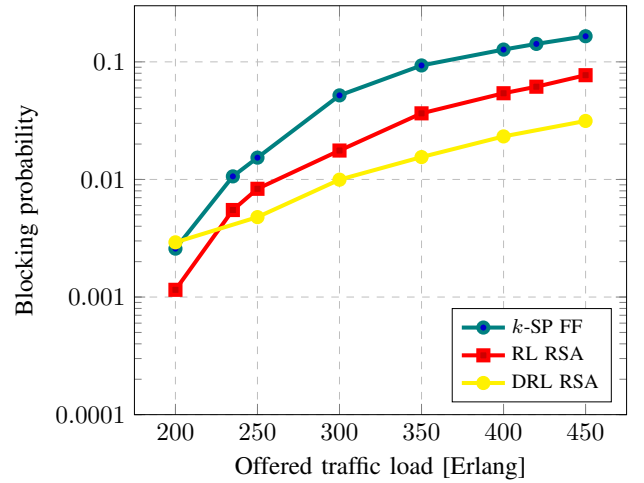


Fig. 1. BP vs traffic load.

Fig. 1 shows the BP as a performance metric considered to compare DRL-RSA with the previously proposed RSA strategies, RL and k -SP FF [9], across varying traffic loads, ranging from 200 to 450 Erlang using the Japanese topology. As expected, with all the strategies, k -SP FF, RL-RSA and DRL-RSA, the BP increases as the traffic load increases. It starts at 1.2×10^{-3} and 2.9×10^{-3} at 200 Erlang load and grows up to 7.7×10^{-2} and 3.1×10^{-2} at 450 Erlang load with the RL and DRL based RSA strategies, respectively. Meanwhile, the k -SP FF starts with a BP of 2.6×10^{-3} at

TABLE II
MULTI-BAND SYSTEM PARAMETERS.

Band	L	C	S	E
Band Range [nm]	1565-1625	1530-1565	1460-1530	1360-1460
Frequency Range [THz]	184.62-191.69	191.69-196.08	196.08-205.48	205.48-220.59
Available Bandwidth [THz]	6.95	4.05	9.1	14.8
Central Frequency [THz]	188.16	193.89	200.78	213.04
Number of Channels (75 GHz)	94	58	125	201
Amplifier Noise Figure [dB]	6	5	7	6.5
Fiber Attenuation [dB/km]	[0.20 - 0.191]	[0.191 - 0.197]	[0.197 - 0.22]	[0.22 - 0.28]

200 Erlang and increases to a BP of 16.5×10^{-2} at 450 Erlang, which is significantly higher than that observed with the RL and DRL strategies. The performance of the DRL-based RSA strategy versus the state-of-the-art RL-based RSA strategy could be distinguished depending on the traffic across the network: i) low traffic loads and ii) medium and high traffic loads. At low traffic loads (i.e., 200 Erlang) the RL-based strategy is able to reduce the BP by 55% compared to the DRL and the commonly used heuristic RSA strategy (k -SP FF). As the traffic load increases (starting 250 Erlang), the DRL-RSA strategy outperforms both k -SP FF and RL-RSA strategies and shows an average decrease in BP of approximately 80% and 50%, respectively. In addition, the DRL-based strategy may increase the network throughput, for a target BP of 1×10^{-2} , by 20% and 50% compared to RL-RSA and k -SP FF strategies, respectively.

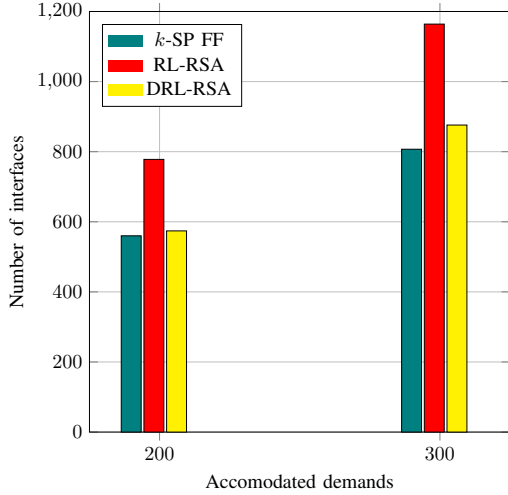


Fig. 2. Average interfaces usage per accommodated demands.

Fig. 2 illustrates the utilization of transmitter/receiver (TX/RX) interfaces by the RSA strategies as a function of the number of accommodated demands (in order to guarantee a fair comparison of this cost-related metric a similar number of traffic demands need to be established). All strategies exhibit a consistent increase in the number of used interfaces, with a notably higher interface count with the RL-based strategy, starting from 560, 778, and 574 interfaces to accommodate 200 demands and peaking at 807, 1164, and 876 interfaces for 300 demands with k -SP FF, RL-RSA, and DRL-RSA, respectively. This corresponds to an average of 2.73, 3.88, and 2.9 used

interfaces per demand. The lower average interface utilization by DRL-RSA and k -SP FF compared to RL-RSA is attributed to their tendency to select shorter paths more frequently, as shown in Fig. 3 (the average path lengths chosen by DRL-RSA and k -SP FF are 520 km and 457 km, respectively, whereas RL-RSA assigns paths with an average length of 826 km). Optical channels established over shorter paths tend to have higher GSNR values. This enables the accommodation of incoming requests via a single DP-16QAM channel with a bandwidth of 75 GHz using only 2 interfaces compared to 4 interfaces if a double DP-QPSK channels with a bandwidth of 150 GHz were used due to a low GSNR value. It is important to note that at medium to high traffic loads the RSA must perform a balancing act in the sense that in resorting to longer paths to mitigate the impact on blocking of links that are becoming congested, it may be selecting less spectral efficient modulation formats. This balancing act is clearly achieved more efficiently with DRL-RSA than with the other two strategies.

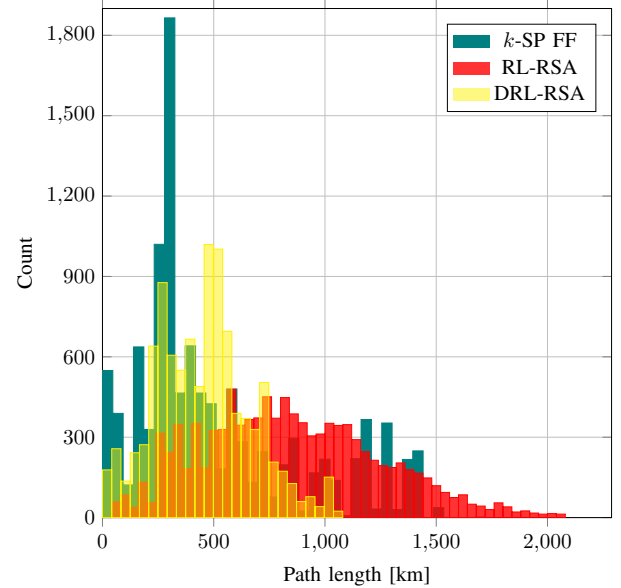


Fig. 3. Paths usage in the Japanese topology.

Fig. 4 illustrates the differences between DRL and k -SP FF-based RSA approaches in terms of BP and the utilization of interfaces at various traffic loads. At low traffic loads (i.e., 200 Erlang), the k -SP FF strategy demonstrates a 7% reduction in BP compared to DRL-RSA. However, as the traffic

load increases from 250 Erlang to 450 Erlang, DRL-RSA significantly decreases BP by an average of 50%, utilizing the same or fewer TX/RX interfaces compared to k -SP FF. This improvement can be attributed not only to the utilization of shorter paths but also to the dynamic path assignment policy of the DRL agent. The DRL agent updates its path selection based on the network status and incoming request volume, leveraging its reward function. Such adaptability is a distinct advantage of DRL-RSA over heuristic approaches like FF, which lack the capability to dynamically adjust path assignments based on real-time network conditions.

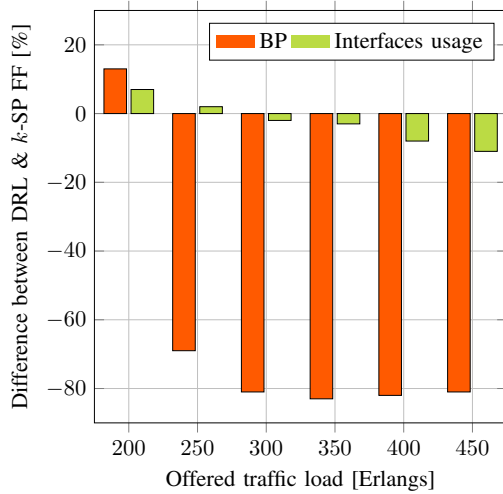


Fig. 4. BP and interfaces count differences (DRL & k -SP FF).

Fig. 5 shows the spectrum utilization per-band by the DRL-RSA algorithm. The most used band is the C-band, where 43% of the established source-destination connection requests are assigned a channel in this band, then L-band with 27%, after that the S-band with 16% and finally the E-band with 14%. Note that the RL-RSA and k -SP FF algorithms give preference to bands that feature channels with higher GSNR (resulting in an order C, L, S and E), as described in [9]. However, for the DRL-RSA algorithm, the explicit ordering of bands is not provided to the agent. Instead, the agent uses a selection mechanism based on the length of the path (i.e, shorter path) and the GSNR value (higher GSNR value) of the channels in the bands. After training, the DRL agent autonomously has learned to prioritize the bands, e.g., C, L, S, and E, and make the most efficient use of each.

IV. CONCLUSION

In this work, we presented a new Deep Reinforcement Learning (DRL) strategy for Routing and Spectrum Assignment (RSA) in multi-band optical networks. The DRL strategy accounts for physical layer impairments, including the Simulated Raman Scattering (SRS) effect, and embeds a reward function that is also resource usage-aware. Simulation results obtained over the Japanese topology provide evidence that the DRL strategy reduces the blocking probability (BP) at medium and high traffic loads. More precisely, it enabled an

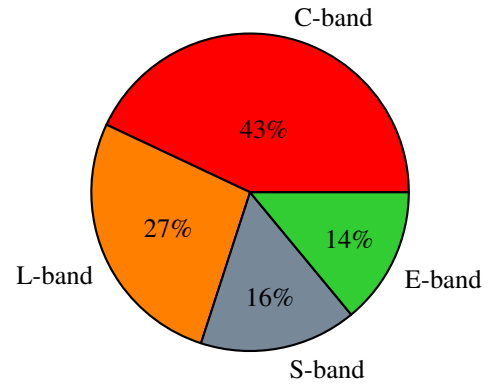


Fig. 5. Band usage in the Japanese topology.

average decrease in BP of 50% and 80%, utilizing the same or fewer TX/RX interfaces per demand, when compared to those obtained with a state-of-the-art RL-based RSA and k -Shortest Path First-Fit (k -SP FF) strategies, respectively. In addition, the DRL-based strategy may increase the network throughput by 20% and 50% compared to RL-RSA and k -SP FF strategies, respectively.

REFERENCES

- [1] Nicola Sambo, Alessio Ferrari, Antonio Napoli, Nelson Costa, João Pedro, Bernd Sommerkorn-Krombholz, Piero Castoldi, and Vittorio Curri. Provisioning in multi-band optical networks. *Journal of Lightwave Technology*, 38(9):2598–2605, 2020.
- [2] Carlos Natalino and Paolo Monti. The optical rl-gym: An open-source toolkit for applying reinforcement learning in optical networks. In *2020 22nd International Conference on Transparent Optical Networks (ICTON)*, pages 1–5, 2020.
- [3] Patricia Morales, Patricia Franco, Astrid Lozada, Nicolás Jara, Felipe Calderón, Juan Pinto-Ríos, and Ariel Leiva. Multi-band environments for optical reinforcement learning gym for resource allocation in elastic optical networks. In *2021 International Conference on Optical Network Design and Modeling (ONDM)*, pages 1–6, 2021.
- [4] Xiaoliang Chen, Baojia Li, Roberto Proietti, Hongbo Lu, Zuqing Zhu, Shao Zhu, and S. J. Ben Yoo. Deepmsa: A deep reinforcement learning framework for routing, modulation and spectrum assignment in elastic optical networks. *Journal of Lightwave Technology*, 37(16):4155–4163, 2019.
- [5] Carlos Hernández-Chulde, Ramon Casellas, Ricardo Martínez, Ricard Vilalta, and Raul Muñoz. Assessment of a latency-aware routing and spectrum assignment mechanism based on deep reinforcement learning. In *2021 European Conference on Optical Communication (ECOC)*, pages 1–4, 2021.
- [6] Juan Pinto-Ríos, Felipe Calderón, Ariel Leiva, Gabriel Hermosilla, Alejandra Beghelli, Danilo Bórquez-Paredes, Astrid Lozada, Nicolas Jara, Ricardo Olivares, and Gabriel Saavedra. Resource allocation in multicore elastic optical networks: A deep reinforcement learning approach. *Complexity*, 2023:4140594, 2023.
- [7] Alejandra Beghelli and Patricia Morales. Approaches to dynamic provisioning in multiband elastic optical networks. *International Conference on Optical Network Design and Modeling (ONDM)*, 2023.
- [8] GNPpy github repos. <https://github.com/Telecominfraproject/oopt-gnpy>. Accessed: 2022-11-21.
- [9] Abdennour Ben Terki, Joao Pedro, Antonio Eira, Antonio Napoli, and Nicola Sambo. Routing and spectrum assignment based on reinforcement learning in multi-band optical networks. In *2023 Photonics in Switching and Computing (PSC)*, pages 1–3, 2023.
- [10] Matteo Salani, Cristina Rottondi, and Massimo Tornatore. Routing and spectrum assignment integrating machine-learning-based QoT estimation in elastic optical networks. In *IEEE INFOCOM 2019 - IEEE Conference on Computer Communications*, pages 1738–1746, 2019.