# Multi-band provisioning in dynamic elastic optical networks: a comparative study of a heuristic and a deep reinforcement learning approach

Nour El Din El Sheikh
*University College London, UK*
nour.sheikh.18@ucl.ac.uk

Esteban Paz
*Universidad de Concepcion, Chile*
epaz@udec.cl

Juan Pinto
*Pont. Universidad Catolica de Valparaiso, Chile*
juan.pinto.r@pucv.cl

Alejandra Beghelli
*University College London, UK*
alejandra.beghelli@ucl.ac.uk

***Abstract*—The blocking performance of a heuristic and a deep reinforcement learning approach for resource provisioning in a dynamic multi-band elastic optical network is evaluated. The heuristic is based on a previous proposal that prioritises the use of band C, then L, S, and E, in that order. The deep reinforcement learning approach uses a deep Q-network (DQN) agent trained on different multi-band scenarios. Results show, as expected, a significant decrease in blocking probability when moving from the C-band only scenario to the multi-band scenarios (C+L, C+L+S, C+L+S+E). However, the DQN agent did not outperform the heuristic. The lower performance of the agent, also observed in some previous works in optical networks, highlights the need for further research on how to better configure agents and improve the network representation used by the optical network environments.**

***Index Terms*—provisioning, reinforcement learning, multi-band optical networks, elastic optical networks**

## I. Introduction

Multi-band optical networks are one of the promising solutions to accommodate the Internet traffic growth triggered by the increasing number of users/devices and the ever growing bandwidth generated/consumed by them [1]. By extending the network operation of currently deployed fibre beyond the C band, multi-band networks do not need to incur the fibre-lying costs associated to alternative solutions - like multi-core fibre - while retaining most of the advantages [2].

Among the many challenges associated to the multi-band operation in dynamic elastic optical networks (MB-EONs), provisioning is one that significantly impacts how efficiently resources are utilised. In a MB-EON, provisioning is in charge of solving the problem of finding a route, a band, a modulation format and a spectrum portion for each connection request. Recently, a heuristic approach was proposed in [3] to allocate resources in dynamic MB-EONs operating with the C+L+S+E+O bands. Simulation results obtained for a core network showed that the signal quality offered by the O-band is not good enough for a core scenario and that by using C+L+S bands, up to 4 times more traffic can be accommodated in the network.

In the last couple of years, a few initial works have reported on the application of deep reinforcement learning (DRL) techniques to solve different resource allocation problems in optical networks [4]–[7]. Results in [4] show the agent performs only slightly better than the K-shortest-path first fit (K-SP-FF) heuristic, commonly used to solve the dynamic routing and spectrum assignment (RSA) problem in elastic optical networks. The work in [5] shows an agent outperforming one of the heuristics and performing worse than another. In [7] it is argued that the reason behind the low performance of DRL agents lies on the simplified network representation commonly used to keep a low computational complexity. They demonstrate that by using a more complete representation, their agent achieves better performance. However, this agent only solves the simpler routing problem, without dealing with spectrum allocation.

To add to the relevant discussion of when and how DRL algorithms offer better performance than very well known heuristics in optical networks, in this paper we report on initial results on applying DRL to solve the problem of provisioning in dynamic MB-EONs. That is, solving the routing, modulation format, band and spectrum allocation problem (multi-band RMSA). To do so, we use the same agent and set of hyperparameters reported previously to solve the single-band RMSA problem [6] (environment and the observation space adapted to the multi-band scenario) and apply a heuristic based on the one presented in [3]. Results show the expected improvement in blocking probability due to increased capacity under both approaches. However, as also reported in some previous work, the agent fails to outperform the heuristic in all studied cases.

The rest of this paper is as follows: Section II presents the physical layer model, the heuristic and the DRL approach (model and training); Section III presents the simulation results and Section IV concludes the paper.

## II. The Multiband Optical Network Environment

### A. Physical Layer Model

We consider an elastic optical network operating in the C, L, S and E bands. The O band was not considered due to the enhancement of nonlinear interactions and the reduced accuracy of the Gaussian noise (GN) model in that region. Results presented in [3] also show the unsuitability of the O-band for core networks.

To calculate the maximum reach of an optical signal propagating in a MB-EON, a total of 2720 frequency slot units (FSUs) centered in the S band were assumed. This yields a total bandwidth of 34THz, from 1365-1615 nm. The number of FSUs for the C, L, S and E bands was equal to 344, 480, 760, and 1136, respectively.

Following the methodology from [8], the maximum transmission reach for different modulation formats was obtained. Table 1 shows the maximum number of spans (1 span = 100 km) for different band configurations assuming a bit error rate threshold of $4.7\text{x}10^{-3}$ before forward error correction [9].

TABLE I
Maximum optical reach [number of spans]

| Active bands | Modulation Formats | | | | | |
|---|---|---|---|---|---|---|
| | ***BPSK*** | ***QPSK*** | ***8QAM*** | ***16QAM*** | ***32QAM*** | ***64QAM*** |
| C | 199 | 99 | 54 | 27 | 13 | 7 |
| C+L | C:197<br>L:167 | C:99<br>L:84 | C:54<br>L:46 | C:14<br>L:22 | C:13<br>L:11 | C:7<br>L:6 |
| C+L<br>+S | C:174<br>L:167<br>S:148 | C:87<br>L:84<br>S:74 | C:47<br>L:46<br>S:41 | C:23<br>L:26<br>S:20 | C:12<br>L:11<br>S:10 | C:6<br>L:6<br>S:5 |
| C+L<br>+S+E | C:130<br>L:144<br>S:102<br>E:31 | C:65<br>L:72<br>S:51<br>E:15 | C:35<br>L:39<br>S:29<br>E: 9 | C:17<br>L:19<br>S:14<br>E:4 | C:8<br>L:9<br>S:7<br>E:2 | C:4<br>L:5<br>S:3<br>E:1 |

The number of slots, s, required by a demand requesting a bit rate b, is given by:

$$s = \lceil b/FSU_b \rceil \tag{1}$$

where $FSU_b$ is the bit rate achieved by a single FSU, equal to those reported in Table 3 in [8]: 23, 46, 69, 92, 115 and 140 Gbps for BPSK, QPSK, 8-QAM, 16-QAM, 32-QAM and 64-QAM, respectively.

### B. The Heuristic Approach

The heuristic, based on the algorithm presented in [3], first attempts to establish connections in the C band. In that band, the K-SP-FF algorithm is executed, selecting the most efficient modulation format with an optical reach equal to or greater than the length of the route checked (number of required slots calculated according to Eq. (1)). If unsuccessful, the same procedure is carried out in the L, S and E bands, in that order. If still unsuccessful, the request is rejected.

### C. The DRL Model

The objective of the agent is maximising its reward when selecting an action. Upon receiving the agent´s action, the environment checks whether there are available resources to establish the connection on the route, band and slot selected by the agent. If the connection is successfully established, a positive reward is sent back to the agent. Otherwise, a negative reward is sent back. In both cases, information on the network status is also sent back to the agent (observation). The action set, the rewards and the observation were as follows:

- The action was coded by extending the action set used in [4] to include K*B*J actions, where K is the number of alternative paths, B the number of bands and J the number of blocks with available slots to consider in the selected path.
- The reward was equal to 1 if the agent´s action led to a successful connection establishment and -1 otherwise.
- The observation received by the agent was coded by extending the network state proposed in [4]. The extended network state representation was coded as a $(2 * N + 1 + (2 * J + 3) * K * B)$ array, where N is the number of nodes. The first $2 * N$ positions identified the source and destination of the request (one-hot format). The third position stored the connection holding time. The next $K * B$ positions stored $(2 * J + 3)$ elements each: for each possible path-band pair, the first J elements store the size of the J blocks, the next J elements store the index of the slot where the $J^{th}$ blocks starts, the remaining 3 elements store the requested number of slots, the average size of blocks and the total number of available slots.

### D. DRL Implementation

The dynamic MB-EON environment was built using the Optical RL-Gym [6], a DRL framework specially developed for optical networks. The code can be found in a Github repository [1]. The agents originally tested in this work were DQN, TRPO and PPO2, available at the Stable Baselines framework [2]. The hyperparameters used for those agents were:

- DQN: *double_q=False, gamma=.95, 'layers': [128] * 4, 'dueling': False*
- TRPO:*gamma=0.95, timesteps_per_batch=1024, max_kl=0.01*
- PPO2:*gamma=0.95, learning_rate=0.00025, vf_coef=0.5*

We evaluated the accumulated rewards of these agents in all band scenarios. The DQN agent achieved better results in most scenarios (C, C+L, C+L+S+E). For example, for 2000 [Erlang], DQN achieved a mean reward of $27.23 \pm 6.6$, followed by TRPO ($27.0 \pm 6.9$) and PPO2 ($24.78 \pm 7.25$). Thus, in this paper we report on DQN agent results.

Next, the DQN agent was trained in the C, C+L, C+L+S and C+L+S+E scenarios for different traffic loads. As way of example, Fig. 1 shows the DQN agent blocking performance during the training phase for all scenarios at a load of 1000 [Erlang]. For the C and C+L scenarios, the agent was trained for 200,000 time steps. In the C+L+S and the C+L+S+E scenarios, the agent was trained for 500,000 time steps. The

[1]https://github.com/nourelsheikh/MultiBand-RL

[2]https://stable-baselines.readthedocs.io/en/master/

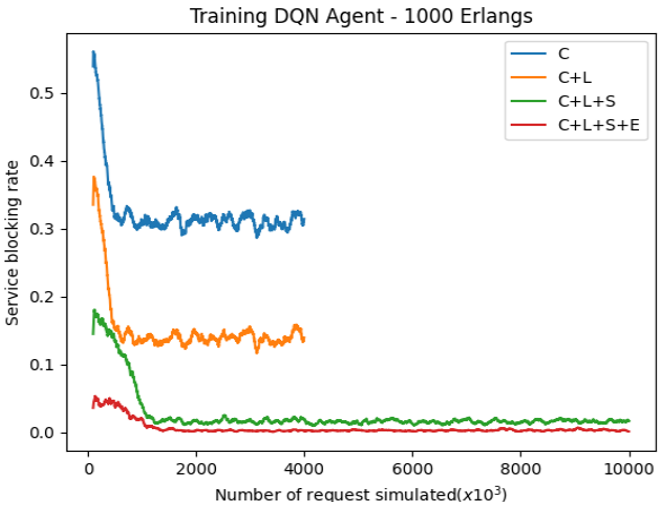


Fig. 1. Evolution of blocking probability during the training process

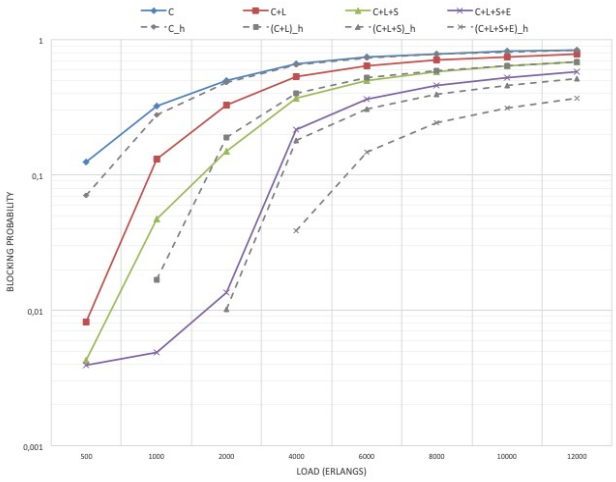


Fig. 2. Agent and heuristic blocking probability as a function of traffic load for NSFNET.

model was saved for each load and scenario combination. These models were then tested in an environment similar to that used in training but with different service requests.

## III. Results

The heuristic and the DQN agent's performance was evaluated by means of simulation on the NSFNet (14-node, 21-link) and Eurocore (11-node, 25-link) topologies. Dynamic requests were generated following a Poisson process with mean arrival rate equal to 0.1 seconds and a mean holding time ranging from 50 to 1200 seconds. The requests' bit rate was uniformly distributed between 25 Gbps and 400 Gbps. The source of each request was selected following the probabilities derived from the traffic matrix presented in [4]. The values of K and J were set to 5 and 1, as in [4]. Due to space constraints, only the results for the NSFNet topology are presented, but similar results were obtained for Eurocore.

Fig. 2 shows the blocking probability exhibited by the trained DQN agent (solid lines) and the heuristic (dashed lines) for the 4 scenarios studied: C-band only, C+L, C+L+S and C+L+S+E.

It can be seen that the agent´s blocking performance is significantly improved when the number of bands is increased. However, the agent´s blocking performance is not better than that achieved by the heuristic - as also observed in some previous works in the area. In terms of run time, the agent was consistently faster than the heuristic: The agent was 13% faster in the C- band scenario, 14.9% in C+L, 16.7% in C+L+S and 18% in C+L+S+E. Besides, whilst the heuristic run time increased 5% going from the C to the C+L+S+E scenario, the agent´s increase was only 1%.

## IV. Conclusion

We compared the blocking performance of a trained reinforcement learning agent to that of an adapted KSP-FF algorithm in dynamic multi-band elastic optical networks. In both cases, the increase in bands led to a significant reduction in blocking probability. The agent was outperformed by the heuristic in terms of blocking but its runtime was consistently faster than the heuristic's and did not increase significantly with increased number of bands. Further work will concentrate on studying how to improve the agent´s performance by researching the impact of different hyper parameters and network representations. By doing so, we expect to detect the conditions under which a DRL agent performs better than known heuristics in solving this resource allocation problem to then generalise the results to different problems in optical networks.

## References


[1] A. Napoli et al., "Towards multiband optical systems," in Advanced Photonics 2018, Optical Society of America, paper NeTu3E.1.

[2] E. Virgilito et al., "Network performance assessment of C+L upgrades vs. fiber doubling SDM solutions,", in 2020 Optical Fiber Communications Conference (OFC), paper M2G.4.

[3] N. Sambo et al., "Provisioning in multi-band optical networks," J. of Lightwave Technol., vol. 38, no. 9, 2598-2605, 2020

[4] X. Chen, B. Li, R. Proietti, H. Lu, Z. Zhu and S.J.B. Yoo, "DeepRMSA: A deep reinforcement learning framework for routing, modulation and spectrum assignment in elastic optical networks,", J. of Lightwave Technol., vol. 37, no. 16, 4155-4162, 2019

[5] X. Luo, C. Shi, L. Wang, X. Chen, Y. Li and T. Yang, "Leveraging double-agent-based deep reinforcement learning to global optimization of elastic optical networks with enhanced survivability," Optics Express, vol. 27, no. 6, 7896-7911, 2019

[6] C. Natalino and P. Monti, "The Optical RL-Gym: An open-source toolkit for applying reinforcement learning in optical networks," 2020 22nd International Conference on Transparent Optical Networks (ICTON), Bari, Italy 2020, pp.1-5

[7] J. Suarez-Varela et al., "Routing in optical transport networks with deep reinforcement learning," IEEE/OSA Journal of Optical Communications and Networking, vol. 11, no.11, 547-558, 2019

[8] E. Paz and G. Saavedra, "Maximum transmission reach for optical signals in elastic optical networks employing band division multiplexing , ArXiv: 2011.03671 (available at" https://arxiv.org/pdf/2011.03671.pdf"), 2021

[9] L.M. Zhang and F.R. Kschischang, "Staircase codes with 6% to 33% overhead,", J. of Lightwave Technol. vol. 32, no. 10, 2019-2027, 2014