Multiperiod Migration Planning of Multilayer Networks subject to Deep Traffic Uncertainties

Prashasti Sahu
Chair of Communication Networks
Technische Universität Chemnitz
Chemnitz, Germany
prashhasti.sahu@etit.tu-chemnitz.de

Ronald Romero Reyes
Chair of Communication Networks
Technische Universität Chemnitz
Chemnitz, Germany
ronald.romero.reyes@etit.tu-chemnitz.de

Arantxa Villavicencio Paz
Dept. of Advanced Technology
Adtran Networks SE
Munich, Germany
arantxa.villavicenciopaz@adtran.com

***Abstract*—Technology migration in multilayer transport networks presents major challenges due to uncertain traffic growth, technology evolution, and high capacity expansion costs. This paper studies this problem for networks subject to deep traffic uncertainties, focusing on the maximization of the expected net present value (ENPV) across a multiperiod planning horizon. For that, we formulate a migration algorithm based on deep reinforcement learning (DRL) and graph neural networks. The approach learns flexible decision rules that adapt both capacity expansions and migration actions to demand growth. The reward function integrates economic profit and penalties for unserved demand and over-provisioning, driving the agent toward optimum migration paths. The proposed solution is evaluated through large-scale simulation experiments on a realistic multilayer transport network exposed to uncertain traffic scenarios. The results show that - compared with a selected baseline - the algorithm achieves a higher ENPV with improved performance in the face of both downside risks and upside potential revenue gains.**

***Index Terms*—Network technology migration, optical transport networks, network optimization, deep-reinforcement learning**

I. Introduction

Planning technology migrations for multilayer transport networks is a strategic process that involves scheduling capacity upgrades and technology transitions across multiple layers over an extended timeline, enabling economic and operational objectives such as low costs while maximizing network profitability - often measured by the Net Present Value (NPV) - and responding to dynamic demand and technology trends.

Recent research provides a spectrum of optimization methods for network migration, spanning metaheuristics, mixed-integer programming, and techno-economic frameworks. Patri et al. [1] explore maximizing the NPV under traffic uncertainty for the migration of access networks. Türk et al. [2] demonstrate metaheuristic efficiency for migration planning in backbone networks. Caria et al. [3] investigate migration of IP networks towards Software Defined Networking (SDN), showing that optimized multiperiod migration sequences substantially improve traffic engineering and capacity utilization, even when only a partial SDN deployment is achieved.

Most prior work has focused on deterministic or moderate-robust planning frameworks, often overlooking the deep traffic and cost uncertainties intrinsic to transport networks. Furthermore, no previous work has studied technology migration under deep uncertainty for multilayer transport networks, particularly when complex technology layers and dynamic costs are present. This gap motivates the present study, which introduces the concept of Flexible Decision Rules (FDRs) as a robust planning framework inspired by applications in civil engineering [4], [5]. For the first time, we formalize this concept as an adaptive migration approach tailored to the complexity of multilayer transport networks, with the objective of maximizing the Expected NPV (ENPV) under deep traffic uncertainties. A reinforcement learning driven methodology is developed to learn and optimize these flexible policies, incorporating them into the technology migration process.

This paper is organized as follows. Section II outlines the state-of-the-art. Sections III and IV formulate the network migration problem and its solution. Section V presents the results and Section VI concludes the paper.

II. State-of-the-Art

A. Existing Planning Algorithms for Network Migration

Early approaches to network technology migration often relied on heuristic and cost-minimization strategies for sequential upgrades across the access, metro, and core segments, targeting relatively small networks and straightforward deployment scenarios. Rule-based and greedy algorithms were employed to minimize network upgrade expenditures and reduce operational disruptions. Particularly, in past decades, some metaheuristic methods were suggested. For access network migration, Turk et al. [2], [6], [7] applied Genetic Algorithms (GA), Particle Swarm Optimization (PSO), Harmony Search, and Ant Colony Optimization to efficiently schedule migrations from VDSL to GPON, focusing on minimizing the Total Cost of Ownership (TCO) while maximizing profit under budget, workforce, and technology adoption constraints. Their studies show that variants like inertia-weighted PSO and hybrid PSO-GA substantially outperform basic greedy schemes in speed and solution quality, especially when accounting for traffic growth and equipment cost erosion.

In the Metro/Core optical domain, Yu et al. [8] evaluated and compared multiple strategies for the phased migration from fixed- to flex-grid wavelength division multiplexed (WDM)

This work was performed within the framework of the German funded BMFTR projects 6G-NETFAB (ID 16KISK250) and SUSTAINET-Advanced (ID 16KIS2280).

978-3-903176-78-2

networks, demonstrating that specific migration strategies can maximize capacity and spectrum efficiency, as well as reduce bandwidth blocking under variable traffic loads.

The migration from legacy IP towards SDN-enabled IP networks has been framed as a multiperiod scheduling and traffic engineering problem. Caria et al. [3] developed an ILP-based optimization that determines the optimal node-by-node SDN migration sequence, aimed at maximizing path diversity and minimizing required capacity upgrades over a migration horizon. Their work emphasizes partial staged deployments (rather than full, one-shot upgrades), and shows significant capacity savings from well-chosen migration sequences.

At the operational and workforce level, Jaumard and Pouya [9], [10] proposed an optimization framework for large-scale network modernization from SONET/SDH to IP/MPLS or optical transport networks (OTN) to minimize migration cost and time. Their models consider technician assignments, maintenance window constraints and outage minimizations, producing migration schedules that address both cost and resource allocation for geographically distributed networks.

Recent developments for access networks incorporate agent-based and decision-theoretic models, such as the work by Patri et al. [1], which uses Expectimax search to optimize multi-period migration under uncertainty. These models explicitly evaluate the NPV and user churn within the decision process, offering advanced sensitivity analysis and scenario generation.

Despite the importance of the existing work, there is not yet an existing approach that tackles the migration problem in transport networks considering the effects that the traffic uncertainties may have in the network life-cycle. Besides, the complex interrelations among the technology layers that implement the network must be taken into account. These aspects are identified as open gaps that deserve further research.

B. Technologies for Multilayer Transport Networks

Multilayer transport networks are implemented as *X-over-WDM* systems in the Metro/Core, where *X* denotes an electrical technology layer that relies on the transmission capacity of an underlying WDM system. Among the existing solutions, *IP-over-WDM* is emerging as an attractive architecture. Nonetheless, this solution may suffer from drawbacks such as packetization overhead, scalability limitations, and costly regeneration in long-haul scenarios due to the limited reach of optical pluggables [11]. Alternative approaches include the well-established *OTN-over-WDM* solution, which transparently supports heterogeneous services (packet- and circuit-switched) over a converged transport network, offering robustness and flexibility for long-term migration planning.

OTN-over-WDM has widely been deployed as monolithic OTN-over-WDM systems (MOTS), which install vendor-proprietary OTN switches integrated within chassis-based systems. The client signals are mapped into optical data units and centrally groomed in the electrical domain. This design consolidates switching fabrics, control logic, and power within a single enclosure. However, scalability is fundamentally limited by chassis capacity, and further network expansions often require full chassis replacements or additions, resulting in increased costs and operational disruptions.

To circumvent these limitations, our work in [11], [12] investigated two OTN architectural variants. The first one, denoted as *disaggregated OTN-over-WDM (DOTS)*, implements OTN switches as clusters of software-controlled, white-box devices arranged in a leaf-spine topology. Here, commodity Ethernet switches replace traditional proprietary OTN fabrics, and high-speed line cards enable OTN switching emulation via packet protocols at the leaf layer. This modular architecture allows for incremental, horizontal scaling and reduces vendor lock-in, adapting network growth to evolving requirements. Alternatively, *private Line Emulation (PLE)-over-WDM* emerges as an architecture that introduces circuit-oriented service emulation within monolithic IP router platforms, leveraging dedicated PLE modules to support legacy circuit-switched services such as OTN or SONET/SDH. With coherent pluggable optics, these platforms can interface directly with WDM switches, supporting both short- and long-haul transmission, but they inherit the scalability limitations associated with monolithic chassis designs. The diversity of *X-over-WDM* implementation variants justifies the need for technology migration algorithms that efficiently upgrade the networks following new technology innovations and services.

III. Problem Statement

Technology migration in multilayer networks involves deep uncertainties in the traffic demands and costs, with migration decisions implying financial consequences that span multiple years. Existing migration algorithms optimize rigid designs in which all the investments are determined upfront, based on cost and traffic forecasts. This can lead to either over- or under-investments. To address this problem, we propose a migration strategy based on the concept of *flexibility in engineering design* [4] a formalization of the economic concept of *Real Options Analysis* to engineering systems. An option is the right, but not the obligation, to make a network deployment and investment decision in the future when more information is available. Thus, instead of optimizing a rigid plan, the purpose is to optimize a flexible design that incorporates the option to modify migration decisions dynamically as new information becomes available. This is accomplished by optimizing the options as *FDRs*, which is done by a deep reinforcement learning (DRL) algorithm that calculates migration decisions that maximize the NPV.

A. Preliminary Definitions

Consider the long-term migration planning of a multilayer network operating over a planning horizon of T time periods. The network is modeled as a graph $G(\mathcal{N}, \mathcal{L})$, where $\mathcal{N}$ is the set of nodes and $\mathcal{L}$ the set of links. At the beginning of each period $t \in \{0, \ldots, T\}$, incremental decisions are made regarding both capacity deployment and technology migration, aiming to serve uncertain and time-varying traffic demands. The decisions are expressed as FDRs, which are optimized at the beginning of the planning horizon T.

At the beginning of the first period ($t = 0$), the network could be either a greenfield or a brownfield design, and the current traffic is known and defined in a matrix $\tilde{D}_0$. This matrix is used as a baseline to predict the traffic matrices $\mathcal{D}_t$ expected at the end of each period t by applying a suitable forecasting algorithm. The collection of matrices $\mathcal{D} = [\mathcal{D}_0, \mathcal{D}_1, ..., \mathcal{D}_T]$ forms the reference demand forecast. To account for demand uncertainty, these matrices are used to sample *scenarios of realized traffic uncertainty*, which are defined as potential traffic demand realizations. At $t = 0$ there is no certainty about the actual traffic demands that will be observed at the end of each period, and in principle, there are infinite possibilities in terms of traffic realizations. Let Ω be the set that contains such realizations. Although it is not possible to calculate the exact set Ω, as it contains infinite scenarios, it can be approximated by a finite set defined by samples of statistically representative scenarios $s \in \Omega$. Thus, if a scenario s occurs with probability p_s, then $\sum_{s\in\Omega} p_s \approx 1$. A scenario $s \in \Omega$ is defined by the collection of traffic matrices: $\xi^s = [\xi^s_0, ..., \xi^s_t, ..., \xi^s_T]$, with ξ^s_t being the matrix of realized traffic in period t. These matrices are sampled from the matrices $\mathcal{D}_t$ with a sampling algorithm that takes the traffic flows in $\mathcal{D}_t$ as mean traffic flows. The scenario history up to period t is given by the collection of matrices: $\xi^s_{[t]} = [\xi^s_0, \xi^s_1, ..., \xi^s_{t-1}]$.

At the beginning of period t, capacity deployment and technology migration decisions are calculated by a policy Π. These decisions depend on (a) the history of demands $\xi^s_{[t]}$; (b) the capacity C^s_{t-1} of the network in the preceding period; and (c) the migration strategy m^s_{t-1} applied in $(t-1)$. The policy decisions are denoted as: $a^s_t = (C^s_t, m^s_t)$, with $C^s_t \in \zeta_t$ and $m^s_t \in M$ being respectively the capacity and the technology migration strategy applied to operate within period t. The set ζ_t defines all feasible network capacity expansion options in t, while the set M defines all feasible node technology migration alternatives. The decision m^s_t defines which nodes can migrate to the target technologies. The policy is therefore defined as:

$$a^s_t = \Pi(C^s_{t-1}, \xi^s_{[t]}, m^s_{t-1}), \forall s \in \Omega, t \in \{0, ..., T\} \quad (1)$$

B. Problem Formulation

At the beginning of the first period ($t = 0$), the following information is known: (a) the network topology; b) the planning horizon T; (c) the traffic matrix $\tilde{D}_0$; (d) the matrices D; (e) the set Ω, with each scenario $s \in \Omega$ associated with a probability p_s and defined by the matrices ξ^s; and (f) the specifications of the candidate network technologies used to implement the network. With this information, the problem is to find the policy given by (1) that prescribes the migration and capacity decisions maximizing the ENPV:

$$\max_{\Pi_\theta} \ \text{ENPV} = \sum_{s\in\Omega} p_s \cdot \text{NPV}(s) \quad (2)$$

Subject to the constraints:

$$C1 : \xi^s_t \geq 0, \forall s \in \Omega, t \in \{0, ..., T\}$$

$$C2 : (C^s_t, m^s_t) = \Pi_\theta(C^s_{t-1}, \xi^s_{[t]}, m^s_{t-1}), \forall s \in \Omega, t \in \{0, ..., T\}$$

$$C3 : C^s_t \in \zeta_t, \forall s \in \Omega, t \in \{0, ..., T\}$$

$$C4 : m^s_t \geq m^s_{t-1}, \forall s \in \Omega, t \in \{0, ..., T\}$$

The policy (1) is approximated by the parametric function $a^s_t = \Pi \approx \Pi_\theta$ (with parameters θ). Constraint C1 ensures that the realized traffic is positive, C2 defines the dependencies of the capacity and the migration decisions on the history of prior capacities, traffic realizations and migration decisions, C3 assures that capacity upgrades are technically feasible, and C4 enforces migration to superior technologies only. In (2), the NPV of scenario s is:

$$\text{NPV}(s) = \sum_{t=0}^{T} \frac{R^s_t(C^s_t, \xi^s_t) - TCO^s_t(C^s_t, m^s_t)}{(1+r)^t} \quad (3)$$

where $R^s_t(C^s_t, \xi^s_t)$ and $TCO^s_t(C^s_t, m^s_t)$ are, respectively, the network revenue and TCO, which are discounted to their present values (at the beginning of $t = 0$) by the factor $1/(1+r)^t$, where r is the market inflation rate.

The policy $\Pi \approx \Pi_\theta$ is optimized offline at $t = 0$. During network operation, it is applied at the beginning of each period to determine the optimum migration and capacity deployment decision, using only the information available up to that period. During network operation, the actual traffic realizations will differ from the scenarios in Ω. Nonetheless, if Ω is made up of representative samples, then the policy should capture the stationary and non-stationary statistical properties of the traffic, and thus, it should calculate optimum decisions when exposed to traffic scenarios not included Ω. The policy defines the FDRs that dictate how to adapt the capacity deployment and migration decisions as the traffic unfolds in time.

IV. Formulation of DRL Migration Algorithm

For realistic networks, solving the stochastic multiperiod migration problem (2) via classical optimization methods is computationally infeasible owing - among others - to the large number of scenarios in Ω. To address this limitation, a DRL-based approach is adopted to approximate the policy in (1) as a graph neural network (GNN) whose parameters are optimized with the Proximal Policy Optimization (PPO) algorithm [13].

A. Definition of Network State, Actions and Rewards

For any scenario s, at the beginning of every period t, the network state is: $S^s_t = \left[C^s_{(t-1),n}, \xi^s_{(t-1),n}, m^s_{(t-1),n}, n \in \mathcal{N} \right]$, where $C^s_{(t-1),n}$ and $\xi^s_{(t-1),n}$ are, respectively, the capacity deployed and the realized demand supported at node n in the preceding period. Assuming that - within T - nodes can only migrate once from an old to a new technology, the component $m^s_{(t-1),n}$ indicates whether at the beginning of $(t-1)$ node n migrated ($m^s_{(t-1),n} = 1$) or not ($m^s_{(t-1),n} = 0$). The state concatenates these three components for all nodes in the network. Given the observed state S^s_t at the beginning of t, the agent selects and applies an action defined by (1) and approximated by the vector: $a^s_t = \Pi \approx \Pi_\theta = \left[a^s_{t,1}, \ldots, a^s_{t,n}, \ldots, a^s_{t,|\mathcal{N}|} \right]$, which consists of one action $a_{t,n}$ for each node n. This node-specific action is specified

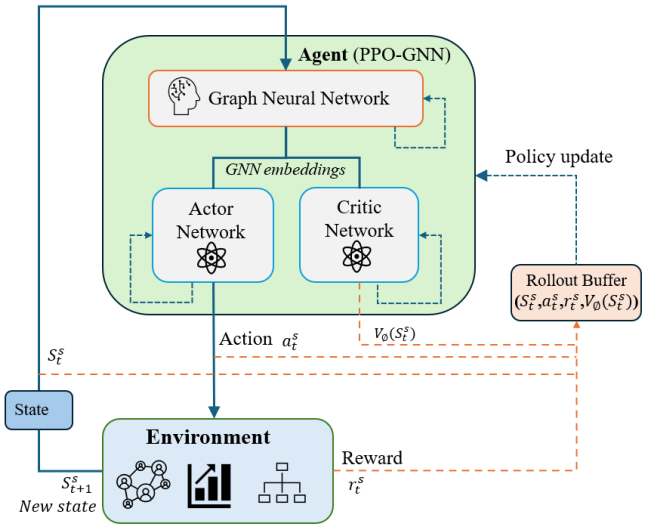

Fig. 1. PPO-based framework for FDR optimization.

Algorithm 1 PPO-Algorithm - Training Procedure

Initialize: Parameters θ, ϕ of Π_θ and $V_\phi(S_t^s)$, rollout buffer, T, matrices D, set Ω and its sampled matrices ξ^s
for each episode (i.e. for each $s \in \Omega$) **do**
  Retrieve the sampled matrices ξ^s
  Simulate environment from $t = 0$ till T under s
  for each period t **do**
    Observe S_t^s, sample $a_t^s \sim \Pi_\theta(C_{t-1}^s, \xi_{[t]}^s, m_{t-1}^s)$
    Execute a_t, calculate r_t^s, and update the state: S_{t+1}^s
    Store transition $< S_t^s, a_t^s, r_t^s, V_\phi(S_t^s) >$
  end for
  if rollout buffer full **then**
    Compute $\hat{A}_t$ (GAE)
    Update θ, ϕ via SGD
  end if
  return optimized policy $\Pi_\theta(C_{t-1}^s, \xi_{[t]}^s, m_{t-1}^s)$

as: $a_{t,n}^s = \left(C_{t,n}^s,\ m_{t,n}^s\right)$, where $C_{t,n}^s$ is the node capacity decision and $m_{t,n}^s \in \{0,1\}$ is the migration decision. After applying a_t^s, the agent receives the reward $r_t^s = R_t^s(C_t^s, \xi_t^s) - \text{TCO}_t^s(C_t^s, m_t^s)$, which is the Net Cash Flow (NCF) obtained in period t. Note that the problem is tackled as a node-based formulation, where the network capacity and the migration decisions are, respectively, defined as: $C_t^s = \sum_{n \in \mathcal{N}} C_{t,n}^s$, and $m_t^s = \left[m_{t,1}^s, \ldots, m_{t,n}^s, \ldots, m_{t,|\mathcal{N}|}^s\right]$. The agent optimizes the policy Π_θ that maximizes the ENPV in Ω.

B. DRL Architecture and PPO Training Algorithm

Figure 1 illustrates the GNN-based actor-critic architecture used to represent the policy Π_θ. Both Actor and Critic networks share a common GNN backbone: three graph convolution layers (256 hidden dimensions each) that aggregate node-neighbor information through message passing with ReLU activations. At each period t in scenario s, the network state is encoded as a graph, where each node is characterized by its current capacity and demand, migration status, node connectivity, and temporal characteristics, such as demand growth and variability. The GNN processes these features through message passing to produce node-level embeddings. The Actor network uses these embeddings to output action logits (capacity expansion and binary migration decisions per node). The Critic network aggregates node embeddings via mean pooling to produce the scalar state value estimate $V_\phi(S_t^s)$.

Algorithm 1 formalizes the episodic PPO training process. Training proceeds by episodes, where each episode corresponds to one scenario s simulated over T periods. During each period t, the agent observes the state, samples an action from the policy, executes it and collects the reward. The next state is updated as well. The Critic evaluates state values using the same GNN embeddings, and all transitions are stored in a rollout buffer for gradient updates. Once sufficient transitions accumulate, Generalized Advantage Estimation (GAE) is applied to estimate advantage values, and the policy and value networks are jointly updated via PPO.

PPO performs policy updates using a clipped surrogate loss, which restricts drastic changes between policy iterations and enhances stability. The value function $V_\phi(S_t^s)$ is optimized by minimizing the mean squared error between predicted and observed returns. An entropy regularization term is used to encourage exploration and prevent premature convergence. This results in an objective function that consists of three terms: the policy loss, the value loss, and the entropy regularization, all of them weighted by suitable coefficients. The minimization of the objective is carried out by applying stochastic gradient descent (SGD). Iterative training with GNN-PPO yields robust and expressive FDR policies capable of adapting node-level migration and capacity decisions to evolving conditions, ultimately maximizing ENPV.

V. Performance Evaluation Results

A. Description of Experimental Setup

The performance of the DRL algorithm - referred to as FDR - is assessed for the *Germany50* topology with 50 nodes and 88 links in Fig. 2. The planning horizon is $T = 10$ annual periods (2025–2034). A brownfield deployment is considered, where initially - in 2025 - the network is an *OTN-over-WDM* system, with the OTN layer implemented by the monolithic MOTS nodes described in Section II-B. At each period t, the nodes are eligible for migration to DOTS technology. The WDM layer is a C-band flex-grid system utilizing 100 Gb/s transponders.

At the beginning of T, the network serves a traffic load of 1.8 Tb/s, which is the sum of the traffic demands in the matrix $\tilde{D}_0$. This matrix and the predicted matrices $\mathcal{D}_t$ are defined with the method in [14], where from $\tilde{D}_0$, the matrices are predicted by applying a polynomial fitting algorithm that takes historical traffic data from 2010-2018 to estimate the compound annual growth rates (CAGR) over T. These CAGRs are used to iteratively forecast $\mathcal{D}_t$ from $\tilde{D}_0$. The resulting matrices are

used to sample two sets of scenarios of realized traffic. The first one, denoted as Ω_{train}, contains $10,000$ scenarios which are used to train the FDR algorithm. The second one, denoted as Ω_{test}, contains $1,000$ scenarios used to test its performance. The Geometric Brownian Motion (GBM) algorithm described in [14] is applied to sample from the matrices $\mathcal{D}_t$, the matrices ξ^s for each scenario s. The evaluation of the NPV uses the cost and revenue models in [12], [14] for the MOTS and DOTS technologies. These models define the price books and the equipment specifications with the costs expressed in cost units (CU), where 1 CU $= 7,450$ USD. The NPVs are calculated with a discount rate of $r = 8\%$.

The DRL agent is trained on Ω_{train}. Key PPO hyperparameters are: learning rate $\eta = 3\times10^{-4}$, GAE parameter $\lambda = 0.95$, clipping parameter $\epsilon = 0.2$, and entropy coefficient 0.01. Early stopping is triggered when the moving average training reward (computed over the last 100 episodes) shows no improvement beyond a minimum delta of 0.02 for 300 consecutive episodes. Training is conducted using GPU acceleration (NVIDIA RTX A6000 with 12 GB memory) with mini-batches of size 32 and 4 PPO epochs per policy update. The rollout buffer accumulates transitions from the episodes until reaching a capacity of 2,560 transitions (256 episodes × 10 periods each).

As discussed in Section II, existing migration algorithms are unsuitable for multilayer networks subject to deep traffic uncertainties. Therefore, to fairly assess the FDR algorithm, its performance is compared against a baseline optimized over Ω_{train}. This baseline, denoted as Exploratory Search (ES), intends to capture the statistical properties of the scenarios in the training set by implementing migration and capacity expansion decisions according to three decision hyperparameters: $\beta \leq 1$ (capacity shortage threshold), γ (target capacity provisioning margin), and μ (migration capacity trigger threshold). These parameters are fixed for all periods and scenarios in Ω_{test}. At the beginning of each period t, the baseline operates as follows. First, for capacity expansion, it computes for each node n the capacity-to-offered-traffic ratio $\rho_n = C^s_{(t-1),n}/\xi^s_{(t-1),n}$. If $\rho_n < \beta$, the node had insufficient capacity to fully serve the traffic in the preceding period, and thus, the baseline expands (for period t) the node capacity as: $C^s_{t,n} = (1+\gamma)\cdot C^s_{(t-1),n}$. Second, for migration decisions, if the capacity expansion decision for node n exceeds the threshold $100 \cdot \mu$ Gb/s (where μ is a constant of proportionality), and if the current NPV (up to $t-1$) is positive (ensuring profitability), the node is migrated from MOTS to DOTS. Once migrated, nodes remain in DOTS for all subsequent periods. The hyperparameters are such that they maximize the objective (2) over the 10,000 scenarios in Ω_{train}. To find their optimum values, an exhaustive grid search strategy is applied in which the following ranges are predefined for each hyperparameter: $\beta \in [0.7, 1.0]$, $\gamma \in [0.05, 1.0]$, and $\mu \in [0.005, 0.05]$. The first two intervals use step increment granularities of $\Delta = 0.05$, whereas for μ, the step is $\Delta = 0.005$. The objective (2) is assessed for all possible hyperparameter combinations. It was found that the optimal configuration that maximizes the ENPV is: $(\beta^*, \gamma^*, \mu^*) = (1.0, 0.65, 0.05)$. These values are

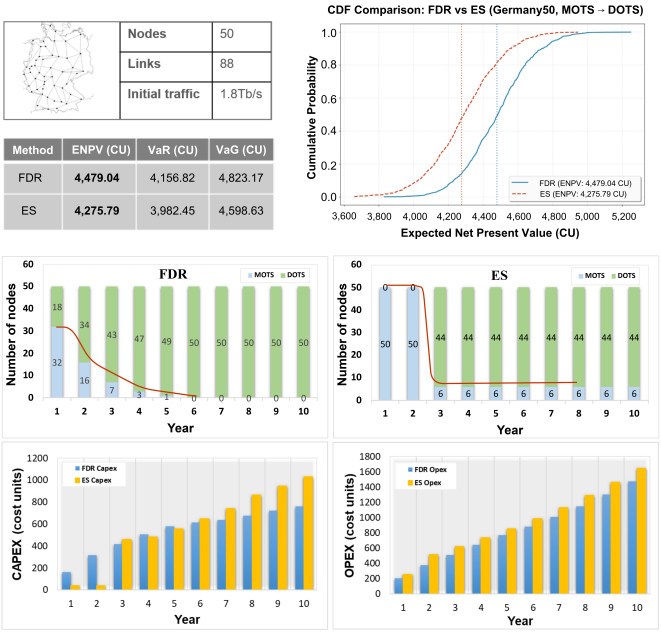

Fig. 2. Performance comparison of FDR and ES.

then used to perform capacity and migration decisions for the 1,000 scenarios in Ω_{test}. The resulting ENPV is assessed and compared against the ENPV achieved by FDR. (Note: for both algorithms, the scenarios in Ω_{train} and Ω_{test} are assumed to be equiprobable.)

Concerning the computation times, the training of the FDR agent early stopped at episode 2,400 based on convergence criteria, requiring approximately 50 hours on GPU. On the contrary, the ES hyperparameter optimization for $(\beta^*, \gamma^*, \mu^*)$ employed a multi-stage iterative approach: starting with coarse-grained grid search, progressively refining the search space, then sequentially fixing optimized hyperparameters while optimizing remaining parameters. This iterative strategy required approximately 45 hours of computation, which is comparable to the FDR training time. For testing, both the FDR and ES policies were separately evaluated on Ω_{test}. The FDR policy evaluation required approximately 127.9 seconds per scenario, totaling approx. 35.5 hours for the complete test set evaluation. The ES baseline evaluation required approximately 104.1 seconds per scenario, or approx. 29 hours in total.

B. Analysis of Results

Figure 2(a) compares the cumulative distribution functions (CDFs) of the NPVs across the 1,000 scenarios in Ω_{test} for FDR and ES. The ENPVs - shown as vertical lines - are 4,479.04 CU for FDR and 4,275.79 CU for ES. The FDR curve lies consistently to the right of the ES curve across the entire NPV range, demonstrating improved performance. This indicates that for any given NPV threshold, FDR has a higher probability of exceeding that value. This is illustrated by the table in Fig. 2, which besides the ENPV, reports the Value at Risk (VaR) and the Value at Gain (VaG) derived from the CDFs. The VaR corresponds to the 5th percentile and

quantifies the potential downside risk, measured as the NPV that would be observed under unfavorable traffic realizations. In contrast, the VaG corresponds to the 95th percentile, measuring the potential upside NPV under optimistic traffic scenarios. FDR demonstrates superior risk-adjusted performance, with a VaR of 4,156.82 CU compared to ES's 3,982.45 CU, indicating better downside protection. Concerning the VaG, FDR achieves an NPV of 4,823.17 CU compared to ES's 4,598.63 CU, showing that FDR is able to capture more earnings in the event of upside traffic surges.

The migration strategies optimized by FDR and ES reveal fundamental differences between them. This is shown in Fig. 2(b)-(c), which depict the number of nodes migrated from MOTS to DOTS at the beginning of each period. FDR exhibits an aggressive early migration strategy, rapidly transitioning nodes from MOTS to DOTS technology. By year three, FDR has already migrated approximately 60% of the network nodes, completing the full migration by year six. This front-loaded migration pattern is represented by the declining orange curve overlaid on the stacked bars. In contrast, the ES baseline follows a more conservative, reactive migration policy driven by its capacity expansion threshold ($\mu^* = 0.05$). ES maintains a substantial portion of the network in MOTS technology throughout the planning horizon, with only 40% of nodes migrated by year 6 and reaching approximately 70% migration by year 10. The delayed migration strategy of ES reflects its dependence on observed capacity expansion needs rather than on anticipatory optimization.

The cumulative CAPEX and OPEX evolution throughout the planning horizon is shown in Fig. 2(d)-(e) for both approaches. The blue and yellow bars represent FDR and ES, respectively. The CAPEX evolution is proportional to the capacity upgrades performed every period t. In this sense, Fig. 2(d) evinces that FDR's capacity expansion exhibits more uniform annual increments, consistent with its proactive migration strategy that enables more flexible capacity deployment using the cost-efficient DOTS architecture. On the contrary, ES shows more variable year-over-year capacity additions (specially at the beginning of the planning horizon), with occasional spikes corresponding to reactive responses when the shortage ratio ρ_n falls below the threshold $\beta^* = 1.0$. This aggressive threshold indicates that capacity is expanded only when the realized traffic exactly meets or exceeds the installed capacity, reflecting a just-in-time provisioning strategy. However, this reactive approach requires adding capacity at nodes still operating in MOTS technology, incurring higher per-unit deployment costs which are avoided by FDR through earlier migration to DOTS.

For OPEX - see Fig. 2(e) - FDR demonstrates substantially lower cumulative operational costs due to its aggressive early migration strategy. Each node migrated to DOTS generates persistent operational cost savings throughout the remaining planning horizon, as the DOTS architecture offers lower per-unit operational expenses than MOTS. FDR's early migration maximizes the duration of these savings, with nodes migrated by year three accruing 7 years of operational efficiency benefits. In contrast, ES's delayed migration with only 40% of nodes migrated by year six forgoes 2-3 years of potential operational savings per node. These results are aligned with the study in [11], which demonstrates the cost-efficiency of DOTS over MOTS. Thus, FDR is able to better leverage the techno-economic efficiencies of DOTS to optimally migrate the MOTS network.

VI. Conclusion

This paper demonstrates that deep reinforcement learning can optimize FDRs for network migration planning under deep or long-term traffic uncertainties. The GNN-based PPO algorithm learns policies that achieve improved NPVs over the optimized heuristic baseline while exhibiting sophisticated temporal strategies that systematically exploit technology cost differences. These results validate DRL as a principled alternative to traditional methods for adaptive network planning and provide operators with robust, scalable strategies for technology migration under deep uncertainty.

References

[1] S. K. Patri, E. Grigoreva, W. Kellerer, and C. Mas Machuca, "Rational agent-based decision algorithm for strategic converged network migration planning," *JOCN*, vol. 11, no. 7, p. 371, Jun. 2019.

[2] S. Türk, X. Liu, R. Radeke, and R. Lehnert, "Particle swarm optimization of network migration planning," in *2013 IEEE Global Communications Conference (GLOBECOM)*, 2013, pp. 2230–2235.

[3] M. Caria, A. Jukan, and M. Hoffmann, "A performance study of network migration to sdn-enabled traffic engineering," in *2013 IEEE Global Communications Conference (GLOBECOM)*, 2013, pp. 1391–1396.

[4] R. De Neufville and S. Scholtes, *Flexibility in engineering design*. MIT press, 2011.

[5] C. Caputo and M.-A. Cardin, "Analyzing real options and flexibility in engineering systems design using decision rules and deep reinforcement learning," *Journal of Mechanical Design*, vol. 144, no. 2, p. 021705, 09 2021. [Online]. Available: https://doi.org/10.1115/1.4052299

[6] S. Türk, J. Noack, R. Radeke, and R. Lehnert, "Strategic migration optimization of urban access networks using meta-heuristics," in *2014 26th International Teletraffic Congress (ITC)*, Sep. 2014, pp. 1–8.

[7] S. Türk, "Network migration optimization using meta-heuristics," *AEU - International Journal of Electronics and Communications*, vol. 68, no. 7, pp. 584–586, 2014. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1434841114001010

[8] X. Yu, M. Tornatore, M. Xia, J. Wang, J. Zhang, Y. Zhao, J. Zhang, and B. Mukherjee, "Migration from fixed grid to flexible grid in optical networks," *IEEE Communications Magazine*, vol. 53, no. 2, 2015.

[9] B. Jaumard and H. Pouya, "Migration plan with minimum overall migration time or cost," *Journal of Optical Communications and Networking*, vol. 10, no. 1, pp. 1–13, 2018.

[10] H. Pouya and B. Jaumard, "Efficient network migration planning," in *2017 19th International Conference on Transparent Optical Networks (ICTON)*, 2017, pp. 1–4.

[11] A. Villavicencio Paz, R. Romero Reyes, M. L. Vargas Vivas, T. Bauschert, A. Autenrieth, and J.-P. Elbers, "Techno-economic evaluation of transport network architectures for 6g mobile systems," *IEEE Communications Magazine*, vol. 62, no. 11, pp. 28–34, 2024.

[12] A. Villavicencio Paz, R. Romero Reyes, T. Bauschert, A. Autenrieth, and J.-P. Elbers, "Disaggregated otn-over-wdm vs. private line emulation based ip-over-wdm core networks: A techno-economic comparison," in *2025 ONDM*, 2025, pp. 1–6.

[13] Á. López-Cardona, G. Bernárdez, P. Barlet-Ros, and A. Cabellos-Aparicio, "Proximal policy optimization with graph neural networks for optimal power flow," *arXiv preprint arXiv:2212.12470*, 2022.

[14] A. Villavicencio Paz, R. Romero Reyes, and P. Sahu, "Planning multilayer networks under deep uncertainties: An approach based on flexibility in engineering design," *In JOCN, 2025*, Paper under review.