

CHALMERS
UNIVERSITY OF TECHNOLOGY

Conformal Risk Control for QoT-Aware Lightpath Feasibility Under Technology-Driven Domain Shift

Downloaded from: https://research.chalmers.se, 2026-11-23 03:38 UTC

Citation for the original published paper (version of record):
Rezaei, K., Ayoub, O., Monti, P. et al (2026). Conformal Risk Control for QoT-Aware Lightpath Feasibility Under Technology-Driven Domain Shift. Journal of Lightwave Technology. http://dx.doi.org/10.1109/JLT.2026.3752946

N.B. When citing this work, cite the original published paper.

© 2026 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, or reuse of any copyrighted component of this work in other works.

(article starts on next page)

Conformal Risk Control for QoT-Aware Lightpath Feasibility Under Technology-Driven Domain Shift

Kiarash Rezaei*, Omran Ayoub†, Paolo Monti*, Carlos Natalino*

***Abstract*—Artificial intelligence & machine learning (AI/ML)-based regression models are increasingly used to support operational decision-making in optical networks, but they typically expose only point predictions, which carry no statistically controlled information about their own reliability. This paper proposes a post-hoc conformal risk control (CRC) framework that keeps a pretrained AI/ML regressor fixed, provides a finite-sample guarantee on uncertainty band miscoverage, and maps the band to operational decisions through policy rules. The guarantee applies to interval miscoverage, while the resulting decision-error rates are evaluated empirically. We instantiate the framework for quality of transmission (QoT)-aware lightpath (LP) feasibility assessment, in which a regressor predicts optical signal-to-noise ratio (OSNR) under domain shifts caused by upgrades in fiber type, amplifier technology, and switching architecture. Feature shift contribution scores identify the LP features whose relation to QoT changes the most across a technology transition, and the resulting bands are translated into feasibility decisions by two policies, allowing operators to tune the trade-off between conservatively rejecting or aggressively accepting LP realizations. Experiments on open space-division multiplexing (SDM) QoT datasets over the CONUS topology show that, for a target coverage of 90%, the uncertainty bands meet the target in all 170 trials spanning 17 domain shifts and ten calibration–test partitions, whereas a source-only quantile regression (QR) baseline meets it in none. Under severe domain shifts, the efficiency-oriented (EO) policy empirically improves zero-shot (ZS) feasibility-decision accuracy from about 85% to above 99%, approaching a transfer learning (TL) policy while avoiding retraining. These results show that the framework provides finite-sample control of uncertainty band miscoverage for regression-based optical network decision support, with potential applications beyond QoT-aware LP feasibility assessment.**

***Index Terms*—Trustworthy AI, uncertainty quantification, optical networks, post-hoc calibration, decision support.**

I. Introduction

Optical transport networks are evolving toward increasingly agile and heterogeneous infrastructures, driven by coherent transceivers, flexible-grid elastic optical networking, and emerging space-division multiplexing (SDM) technologies [1], [2]. These advances expand the control-plane design space by introducing degrees of freedom in routing, modulation-format selection, and spectrum/spatial/waveband allocation, but they also intensify the need for fast and reliable physical-layer feasibility assessment [3]. In particular, quality of transmission (QoT) estimation for *unestablished* lightpaths (LPs) is a key enabler of impairment-aware provisioning, reconfiguration, and monitoring in dynamic optical networks [4], [5], and artificial intelligence & machine learning (AI/ML) regressors can approximate analytical QoT models with high accuracy when detailed physical-layer modeling is too expensive or impractical.

Operational deployment, however, demands more than the strong average accuracy typically observed during model development [6]. A *trustworthy* model must also remain reliable when deployment conditions depart from the training data, and expose how much confidence each individual prediction deserves. Network evolution stresses both requirements at once. Differences between training and operational conditions arise from unintentional processes, such as component aging, and from intentional interventions, such as technology upgrades [7]. The latter are usually the more demanding case, because an upgrade in fiber, amplifier, or switching technology modifies not only the distribution of the LP descriptors, but also their underlying relationship with QoT. Such a source–target domain shift silently erodes the accuracy of a regressor trained on the pre-upgrade infrastructure, and a point prediction gives no signal that it has occurred.

Existing work has studied accuracy loss under domain shift and the reliability of individual predictions separately, and neither line of work assesses their impact on operational decisions. Domain shift mitigation methods recover target-domain accuracy, through retraining or adapting the model [8], [9], [10], [11], but require the ability to update the predictor and sufficient target-domain data to update it. Uncertainty-aware methods, in turn, are frequently model-specific or heuristic [12], [13], [14], [15], [16], and report a confidence value rather than committing to an operational choice. Complementing them, conformal prediction (CP) equips black-box AI/ML models with distribution-free uncertainty estimates under mild assumptions [17], applied post-hoc using held-out labeled target-domain dataset and without retraining, while its connection to conformal risk control (CRC) [18] extends conformal methods from coverage guarantees to broader, user-defined risk-control objectives. However, to the best of our knowledge, no prior work in the literature translates a risk-controlled uncertainty band into optical network decisions, precisely when a trained AI/ML model is reused after fiber, amplifier, or switching upgrades.

This paper closes that gap by proposing a post-hoc CRC

This work was supported by the EUREKA Celtic-Next Flagship SUSTAINET-Advance project, funded by VINNOVA (2025-02987, Sweden) and Innosuisse (No. 119.588 INT-ICT, Switzerland). (Corresponding author: Kiarash Rezaei.)

Kiarash Rezaei, Paolo Monti, and Carlos Natalino are with the Department of Electrical Engineering, Chalmers University of Technology, 412 96 Gothenburg, Sweden. (e-mail: kiarashr@chalmers.se, mpaolo@chalmers.se, carlos.natalino@chalmers.se).

Omran Ayoub is with the University of Applied Sciences and Arts of Southern Switzerland, 6928 Lugano, Switzerland. (e-mail: omran.ayoub@supsi.ch).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JLT.2026.3752946.

Digital Object Identifier 10.1109/JLT.2026.3752946

decision-support framework that leaves the deployed regressor untouched, organized in three layers: *(i)* a fixed AI/ML regressor, *(ii)* a CRC layer that converts each point prediction into a risk-controlled uncertainty band from a small target-domain calibration set, and *(iii)* a policy layer that maps the prediction and band into an operational decision. We instantiate the framework for QoT-aware LP feasibility assessment under technology-driven domain shift, building on our preliminary conference work [19], which paired split-conformal bands with feasibility policies and validated the idea on a single source–target transition, and advancing it along four directions.

First, we strengthen the statistical foundations of the conformal band. Whereas [19] sets its width to an empirical quantile of the calibration residuals, which controls coverage on average over calibration sets, in this work we select it through an learn then test (LTT)-based CRC procedure [20] from an upper confidence bound on the miscoverage risk that holds uniformly over a grid of candidate widths. This yields the guarantee that, even with the specific and possibly small calibration set available, the selected band keeps the probability that the true QoT value falls outside the band below α, with confidence of at least $1-\delta$. The statement remains valid even though the width is selected using that same calibration set. Second, we broaden the evaluation to 17 source–target scenarios grouped into six shift families, covering fiber, amplifier, and switching upgrades in isolation and in combination. While the transition studied in [19] changes fiber and amplifier jointly, our isolated evaluation results identify the amplifier as the dominant contributor. Third, we characterize each shift using feature shift contribution scores that identify the LP features whose relation to QoT changes the most, and we delimit what such feature-level scores can and cannot reveal about the impact on the deployed model. Fourth, we map the risk-controlled band to feasibility decisions through two policies, risk-averse (RA), which prioritizes avoiding unsafe acceptance, and efficiency-oriented (EO), which prioritizes avoiding overly conservative rejection. Their decision-error rates are assessed empirically.

We evaluate the framework on open SDM QoT datasets [21] over the CONUS topology, comparing it with a zero-shot (ZS) source model, target-trained transfer learning (TL), and a source-only quantile regression (QR) uncertainty baseline. Across ten calibration–test partitions of each of the 17 scenarios, the LTT bands meet the empirical coverage target in all 170 trials. The EO policy improves accuracy under the most challenging shifts from about 85% for the ZS policy to above 99%, approaching the TL policy without retraining. The RA policy provides a conservative operating mode with near-zero false positive rate (FPR). These results indicate that the proposed approach can turn source-trained QoT regressors into practical decision-support tools under technology-driven deployment shifts, and suggest that the same post-hoc risk-control principle may also apply to other regression-driven optical network tasks.

II. Related Work

AI/ML-based QoT estimation has been widely studied in optical networks. Early studies formulated the problem as feasibility classification [22], [23], while later work considered regression-based QoT estimation [24]. Surveys and tutorials have consolidated this research direction by systematizing model families, feature engineering practices, and deployment considerations for AI/ML-assisted QoT estimation [25], [26], [27]. As a result, AI/ML-based QoT estimation has become a credible building block for optical network automation.

Uncertainty-aware QoT methods differ in the predictive uncertainty they represent, the assumptions they require, and the point at which they enter the pipeline. Probabilistic regression [13] and QR [15] estimate input-dependent predictive distributions or conditional quantiles under the training distribution. They provide input-dependent intervals but require uncertainty-specific training objectives, and source-domain intervals need not remain calibrated after a technology shift. Bayesian recurrent predictors [14] and auxiliary self-confidence outputs [12] expose model or confidence uncertainty but rely on model-specific inference or architecture changes. Digital-twin sensitivity and explicit uncertainty have also been used for QoT optimization and failure prediction [28], [29], [16], while margin-aware design and robust planning incorporate uncertainty at the network-design level [30], [31]. These approaches suit retrainable predictors or planning workflows but do not provide a model-agnostic finite-sample risk guarantee for a fixed deployed regressor.

Domain shift mitigation instead targets predictive accuracy through sample-distribution matching [8], evolutionary transfer learning (TL) [9], adversarial adaptation [10], [32], or lifelong learning [11], [7]. These methods are appropriate when model access and target-domain training are available, but their adapted point estimates alone do not quantify reliability.

Complementing both families, CP attaches distribution-free uncertainty to any fixed predictor using held-out calibration data [17]. Conformalized quantile regression [33], weighted conformal methods under covariate shift [34], and adaptive conformal methods under distribution shift [35] address heteroscedasticity or changing deployment conditions, while CRC extends conformal calibration to user-defined risks [18]. This post-hoc family is suitable when retraining is unavailable but representative labeled calibration data can be collected. The remaining gap is to connect its finite-sample band guarantee to optical network decision support while stating separately which decision errors are guaranteed and which are empirical.

III. Proposed CRC Decision-Support Framework

Figure 1 illustrates the proposed CRC decision-support framework for QoT-aware LP feasibility assessment under source–target deployment mismatch. The framework has three layers: *i*) a base AI/ML regressor that maps an input LP feature vector to a point prediction of a QoT metric, such as optical signal-to-noise ratio (OSNR) or bit error rate (BER), allowing the framework to remain agnostic to the specific regression target; *ii*) a CRC layer that uses held-out target-domain data, referred to as calibration data, and LTT to construct a risk-controlled uncertainty band; *iii*) a policy layer that maps this band into a feasibility decision according to the desired operational preference, balancing avoidance of infeasible LP acceptance against admission of potentially feasible ones.

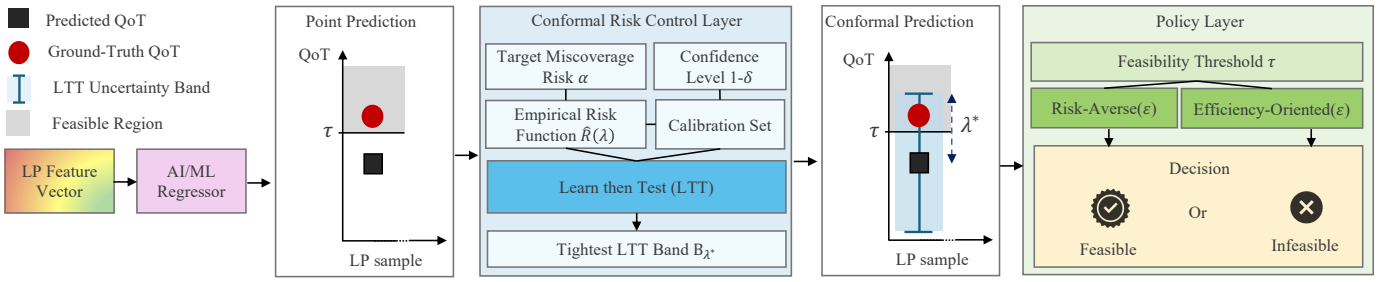


Fig. 1: Overview of the proposed CRC decision-support framework. An AI/ML regressor predicts the QoT of an LP from its feature vector. The CRC layer uses calibration set and LTT to construct the uncertainty band $B_{\lambda^\star}$, which is then mapped to a feasibility decision by the RA and EO policies with respect to the threshold τ.

By separating prediction, band construction, and decision-making, the design lets an existing source-trained QoT regressor be reused under deployment mismatch, using limited target-domain data to quantify the uncertainty of its predictions and translating the resulting band into feasibility decisions according to the desired operating mode. The next three subsections describe the framework layers, while the final subsection characterizes the source–target mismatch considered in this work.

A. AI/ML Regressor: QoT Estimation

The first layer is an AI/ML-based QoT regressor that maps an LP feature vector x to a point prediction $\hat{y} = \hat{f}(x)$. We assume the QoT metric is monotonic in transmission quality, with larger values better for OSNR and smaller values better for BER. The framework treats $\hat{f}$ as a fixed black-box regressor, requiring no change to its architecture, loss, or training procedure, and is therefore applicable to any regression model used for QoT estimation.

B. CRC Layer: LTT Uncertainty Band Construction

The CRC layer converts the point prediction of the pretrained QoT regressor into a risk-controlled uncertainty band. The goal is to select a half-band width ($\lambda^\star$) that remains as narrow as possible while keeping the probability that the true QoT value associated with a future LP feature vector falls outside the band below a prescribed level α, referred to as the *target miscoverage risk*.

Let $\mathcal{D}_{\text{cal}} = \{(x_i, y_i)\}_{i=1}^{n_{\text{cal}}}$ be a calibration set that is disjoint from the training data, where x_i is an LP feature vector and y_i is the corresponding true QoT value. For a candidate half-band width $\lambda \geq 0$, the uncertainty band around the predicted QoT value is

$$B_\lambda(x_i) = \left[\hat{f}(x_i) - \lambda,\ \hat{f}(x_i) + \lambda\right].$$

For a calibration LP with feature vector x_i and true QoT value y_i, the band fails to cover the true value when $|y_i - \hat{f}(x_i)| > \lambda$. Therefore, for each candidate half-band width λ, we compute

$$\widehat{R}(\lambda) = \frac{1}{n_{\text{cal}}} \sum_{i=1}^{n_{\text{cal}}} \mathbf{1}\{|y_i - \hat{f}(x_i)| > \lambda\},$$

where $\widehat{R}(\lambda)$ is the empirical risk function, i.e., the fraction of calibration LPs whose true QoT values fall outside the candidate uncertainty band. Selecting λ from the condition $\widehat{R}(\lambda) \leq \alpha$ would choose a band based only on its observed performance on the available calibration LPs.

This condition, however, raises two issues. First, when only limited target-domain data are available after an upgrade, a band that appears sufficient on the calibration LPs may be too narrow for future test LPs, leading to overconfident decisions. Second, since $\lambda^\star$ is selected from a candidate grid tested on the same calibration set, the risk-control statement must remain valid after this data-driven selection. The LTT step addresses both by replacing the empirical risk with a concentration upper confidence bound $\widehat{R}^+(\lambda)$ on the true miscoverage risk $R(\lambda)$, which holds simultaneously for all candidate half-band widths at confidence level $1 - \delta$ [20]: being uniform over the grid, it absorbs the multiplicity of the selection, and being a high-probability bound rather than an empirical average, it remains meaningful when n_{cal} is small. Thus, a candidate half-band width λ is accepted only if

$$\widehat{R}^+(\lambda) \leq \alpha.$$

Among all accepted candidate half-band widths, the CRC layer selects the smallest one

$$\lambda^\star = \min\left\{\lambda : \widehat{R}^+(\lambda) \leq \alpha\right\}. \tag{1}$$

If no candidate half-band width satisfies this condition, no LTT uncertainty band is returned. Otherwise, for a new test LP feature vector x, the final LTT uncertainty band is

$$B_{\lambda^\star}(x) = \left[\hat{f}(x) - \lambda^\star,\ \hat{f}(x) + \lambda^\star\right], \tag{2}$$

where $\lambda^\star$ is the selected QoT uncertainty margin around the source-trained regressor prediction. We assume that calibration and future test LPs are independent samples from the same target-domain distribution, which makes them exchangeable. Under this assumption, with confidence level $1 - \delta$, the probability that a future LP's true QoT falls outside the selected uncertainty band is no greater than α. The LTT procedure therefore selects the smallest candidate half-band width $\lambda^\star$ that satisfies the risk-control criterion, avoiding unnecessary conservatism while providing explicit miscoverage-risk control for the uncertainty band used by downstream decision tasks.

TABLE I: Evaluated source–target shift families on the CONUS topology from the SDM QoT dataset [21]. Target-domain sample counts are reported below the table.

Shift Family	Fiber	Amplifier	Switching	Source→Target Scenarios
fiber-amplifier-switch (FAS) shift	✓	✓	✓	$3 \to \{7, 8\}$
fiber-amplifier (FA) shift	✓	✓	–	$1 \to 7;\ 2 \to 8;\ 3 \to 9$
A shift	–	✓	–	$4 \to 7;\ 5 \to 8;\ 6 \to 9$
F shift	✓	–	–	$1 \to 4;\ 2 \to 5;\ 3 \to 6$
S shift (WC-MCF)	–	–	✓	$6 \to \{4, 5\};\ 9 \to \{7, 8\}$
S shift (Bu-SMF)	–	–	✓	$3 \to \{1, 2\}$

Approximate target-domain calibration/test sample counts ($\times 10^3$) are: $D_1 = 172/516$, $D_2 = 173/518$, $D_4 = 166/497$, $D_5 = 166/499$, $D_6 = 166/497$, $D_7 = 176/527$, $D_8 = 180/539$, and $D_9 = 181/542$. Here, D_i denotes target dataset i and each pair is reported as calibration/test.

C. Policy layer: Feasibility Decision

The policy layer defines how the LTT band is used to accept or reject an LP, supporting different operating modes by applying the feasibility test to either its lower or upper bound. Let $L(x) = \hat{f}(x) - \lambda^\star, \qquad U(x) = \hat{f}(x) + \lambda^\star$ denote the lower and upper bounds of the band for an LP feature vector x, and let τ denote the modulation-format-specific QoT threshold.

We consider two policies. The RA policy declares an LP feasible only if

$$L(x) \geq \tau + \epsilon_{\mathrm{RA}}. \tag{3}$$

This policy is conservative, as the prediction must remain above the threshold even at the lower end of the band, and is suitable when avoiding the acceptance of an infeasible LP is the main objective.

The EO policy declares an LP feasible only if

$$U(x) \geq \tau + \epsilon_{\mathrm{EO}}. \tag{4}$$

This policy is less conservative, accepting an LP when the upper end of the band indicates possible feasibility, and is suitable when the operator prioritizes higher-spectral-efficiency modulation formats. The parameters ϵ_{RA} and ϵ_{EO} adjust the decision boundary around τ.

The CRC guarantee controls the target miscoverage risk at level α. Under RA, band coverage and the non-negative ϵ_{RA} ensure that an LP declared feasible is truly feasible. Declaring an infeasible LP as feasible requires its actual QoT to fall below the band, and the probability of this outcome is no greater than α. Under EO, declaring an LP infeasible despite its actual QoT reaching the threshold, including the ϵ_{EO}, requires its actual QoT to exceed the band and is likewise bounded by α. These bounds concern different outcomes. Hence, the policies do not provide the same operational guarantee as the uncertainty bands. Their overall performance is evaluated empirically in Section V-C.

D. Source–Target Shift Characterization

Technology changes may alter the relationship between the LP feature vector and the true QoT value. Let $\mathbb{P}_S$ and $\mathbb{P}_T$ denote the source- and target-domain joint distributions of (X, Y), where $X \in \mathbb{R}^d$ is the LP feature vector and $Y \in \mathbb{R}$ the corresponding QoT value. The mismatch can appear in two ways: the feature values themselves may change (e.g., different path lengths, link occupations, carrier frequencies, or active-channel patterns), and their relation to QoT may change after a fiber, amplifier, or switching change. The shift score below reflects both effects.

To measure how much an individual feature changes between domains, we use the 1D Wasserstein distance (W_1) [36], [10], a standard metric for comparing empirical distributions, applied one feature at a time. Letting $X^{(j)}$ denote the jth feature, we define a feature-shift indicator as

$$D_j^X = W_1\Big(\mathbb{P}_S(X^{(j)}), \mathbb{P}_T(X^{(j)})\Big).$$

Similarly, to capture changes in the relation between the jth feature and the QoT value across domains, we define a conditional-shift indicator $D_j^{Y|X}$, which measures how the QoT distribution differs between source and target domains for comparable values of the jth feature. We divide the range of $X^{(j)}$ into K bins $\{\mathcal{B}_k^j\}_{k=1}^K$ and define

$$D_j^{Y|X} = \frac{1}{K}\sum_{k=1}^{K} W_1\Big(\mathbb{P}_S(Y|X^{(j)} \in \mathcal{B}_k^j), \mathbb{P}_T(Y|X^{(j)} \in \mathcal{B}_k^j)\Big),$$

where $\mathcal{B}_k^j$ is the kth bin of the jth feature. Thus, $D_j^{Y|X}$ is large when LPs with comparable values of the jth feature exhibit different QoT behavior across domains. The jth feature shift contribution score is then

$$C_j = D_j^X D_j^{Y|X}. \tag{5}$$

A large C_j indicates that the jth feature changes across domains and that its relation to QoT also changes. We use C_j to explain which LP features contribute most to the mismatch. It is not used for model adaptation and is not a formal distance between the full distributions. Being a product, it also does not detect a transition that alters the feature–QoT relation while leaving the feature distributions intact, so it must be analyzed alongside the selected band width.

IV. Experimental Setup

We use datasets 1–9 from the open SDM QoT dataset [21] on the CONUS topology, which provide nine technology-specific configurations combining bundled single-mode fiber (Bu-SMF) and weakly coupled multicore fiber (WC-MCF) fiber, erbium-doped fiber amplifier (EDFA)- and SDM-amplified systems, and joint switching (J-Sw), fractional-joint switching (FrJ-Sw), and independent switching (Ind-Sw) switching. Taking OSNR as the target QoT metric, we define the 17 source–target scenarios of Table I, grouped into six domain shift families

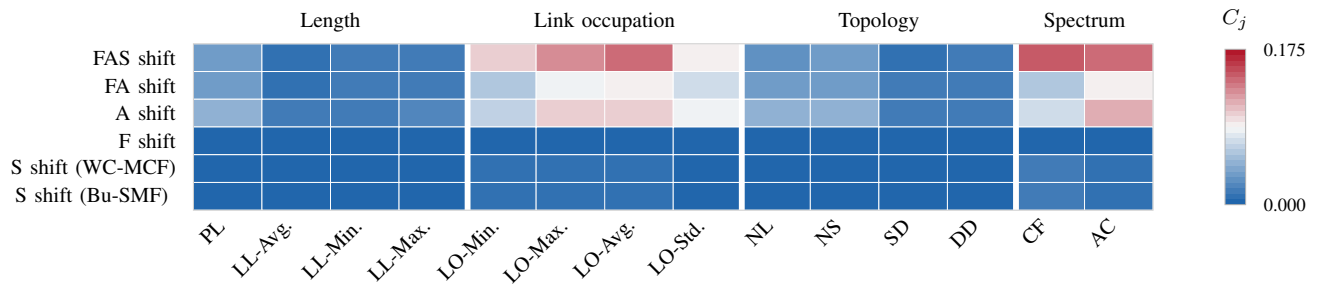


Fig. 2: Average feature shift heatmap of C_j across source–target shift families. Features are grouped by the characteristic they represent: length, link occupation, topology, and spectrum. Feature abbreviations: PL = path length, LL = link length, LO = link occupation, NL = number of links, NS = number of spans, SD = source node degree, DD = destination node degree, CF = carrier frequency, and AC = active channels.

that cover both single-dimension and compounded technology changes.

Table I summarizes the evaluated source–target scenarios, where F, A, and S denote fiber, amplifier, and switching technology changes, respectively. The scenarios include simultaneous three-dimensional changes (FAS shift), joint fiber–amplifier changes (FA shift), and single-dimension changes (F, A, and S shift), reflecting practical network upgrades under a fixed network footprint. For each scenario, the source–target mismatch is characterized using the feature shift contribution score C_j defined in (5), which is used to interpret the severity and origin of the domain shift.

For each scenario, source-domain data train an XGBoost (XGB) QoT regressor. The 20% target adaptation subset is fixed, while the remaining target data are independently/randomly repartitioned ten times into disjoint 20% calibration and 60% test subsets. For each partition, CRC uses $\alpha = 0.10$, $1 - \delta = 0.95$, and 60 uniformly spaced candidate half-band widths λ between 0 and $1.3Q_{0.95}$, where $Q_{0.95}$ is the 95th percentile of the absolute calibration residuals. The selected band is then evaluated on the corresponding test split, yielding 170 trials. Overlapping test samples assess partition sensitivity rather than independent-dataset variation; Table I gives rounded sample counts, while exact scenario-specific counts and experiment seeds are provided in Supplementary Table S1 and Supplementary Section I.

We compare source-only ZS, target-adapted TL trained on the adaptation subset, and the proposed RA and EO policies. As a separate uncertainty-quantification baseline, we adapt the QR approach of [15] to our source-domain setting. Each endpoint, $\widehat{q}_{0.05}(x)$ or $\widehat{q}_{0.95}(x)$, is the mean of five independently trained XGB quantile models to reduce seed-induced variance. The interval uses neither target training nor calibration and is evaluated on the same test samples as CRC; coverage and width are compared under technology shifts.

Calibration-size sensitivity fixes each test split and recalibrates on nested subsets of 500, 1,000, 2,500, 5,000, 10,000, 25,000, 50,000, and 100,000 samples, plus the full calibration set, across 1,530 scenario–partition runs. For policy evaluation, τ is set to 12, 14, 17, 21, and 24 dB for QPSK, 8-QAM, 16-QAM, 32-QAM, and 64-QAM, respectively. Modulation-specific OSNR thresholding follows common elastic optical-network planning practice [37]. We use $\epsilon_{\mathrm{RA}} = 0$ dB and $\epsilon_{\mathrm{EO}} = 0.1$ dB: RA relies on the lower bound for conservatism, whereas EO requires 0.1 dB headroom above τ. These fixed settings are applied to all scenarios.

Since LP feasibility assessment is a binary classification task, performance is measured using accuracy, F1-score, false positive rate (FPR), and false negative rate (FNR), with FPR corresponding to infeasible LPs incorrectly labeled as feasible and FNR corresponding to feasible LPs incorrectly rejected as infeasible. These two error types capture distinct operational consequences, since FPR may lead to failed LP establishment or service disruption, whereas FNR may reduce provisioning opportunities or resource efficiency.

V. Results and Discussion

We first characterize dataset shift using feature shift contribution score C_j defined in Section III-D, then compare the target-domain coverage and width of the LTT uncertainty and source-only QR bands. For LP feasibility assessment, we evaluate the proposed RA and EO policies against ZS and TL point-prediction baseline policies on the same held-out test samples.

A. Source–Target Shift Characterization

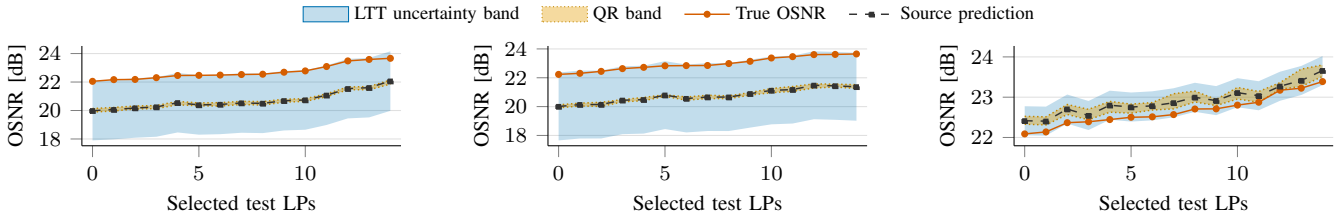
Fig. 2 reports the average feature shift contribution score C_j for each source–target domain shift family, where each family consists of multiple shift scenarios. The largest average contribution scores are observed for FAS and A shift, indicating that amplifier-related changes play a major role in the source–target mismatch, especially when combined with fiber and switching changes. Across feature groups, the strongest shift is concentrated in link occupation and spectrum-related features. In particular, average link occupation (LO-Avg), carrier frequency (CF), and active channels (AC) show the most visible contributions in the strongest domain shift families. This suggests that changes in technology have a considerable impact on the distribution and relationships among features related to link occupation and spectrum allocation.

The FA shift family also incur feature shifts from link occupation features, although its overall shift intensity is lower than that of FAS and A shift. This is consistent with the fact

TABLE II: Test coverage and full interval width for LTT uncertainty band and source-only QR, by shift family. Each scenario is first averaged over ten target-data partitions; values are mean ± standard deviation across scenarios.

	LTT uncertainty band		Source-only QR band	
Shift Family	**Test cov. (%)**	**Full width [dB]**	**Test cov. (%)**	**Full width [dB]**
FAS shift	91.65 ± 0.11	4.200 ± 0.027	0.00 ± 0.00	0.219 ± 0.000
FA shift	91.49 ± 0.37	4.161 ± 0.017	0.00 ± 0.00	0.221 ± 0.003
A shift	94.43 ± 3.72	4.663 ± 0.041	0.00 ± 0.00	0.216 ± 0.004
F shift	91.14 ± 0.28	0.699 ± 0.026	3.28 ± 0.45	0.215 ± 0.003
S shift (WC-MCF)	91.03 ± 0.09	0.178 ± 0.053	82.70 ± 3.66	0.182 ± 0.025
S shift (Bu-SMF)	91.05 ± 0.12	0.235 ± 0.021	82.16 ± 4.71	0.215 ± 0.000

QR: nominal 90% source-domain interval with no target-domain calibration.



(a) FAS shift (b) A shift (c) F shift

Fig. 3: LTT uncertainty and source-only QR bands for the same 15 target-domain test LPs and selected shifts.

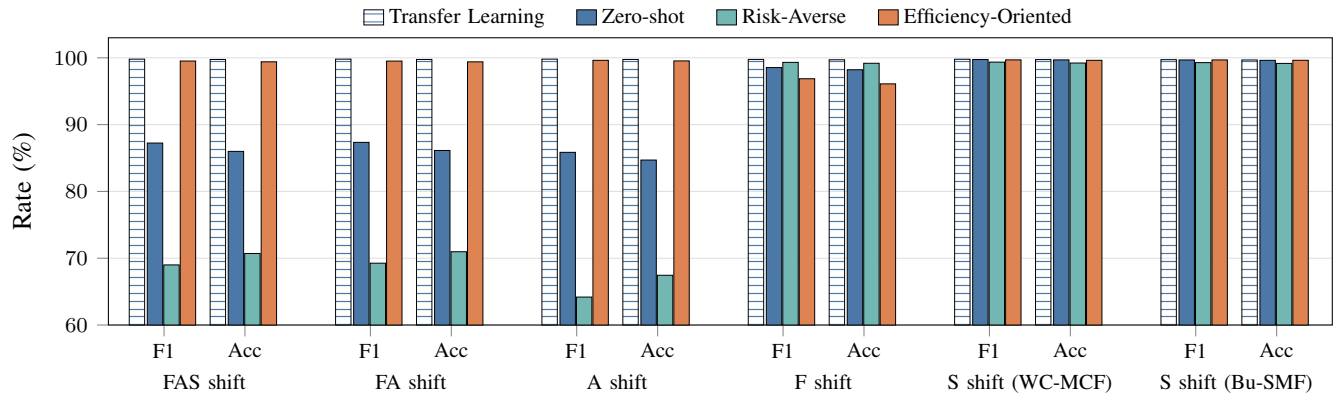
that amplifier changes affect the resulting OSNR behavior and can alter the relation between network load (co-propagating channels) and QoT. By comparison, F shift falls in the lowest bin for every feature, and the S shift (WC-MCF) and S shift (Bu-SMF) families are only marginally above it, suggesting that changing the fiber or the switching architecture alone perturbs the LP feature distributions, and their relation to QoT, far less than amplifier-related or mixed-technology changes do. The differences observed across the shift families allow us to study the framework under a broad range of source–target mismatches, from switching-only changes to combined fiber, amplifier, and switching transitions. The F shift family, however, shows a negligible C_j, yet the transition is not harmless, as it yields a wider band than either S shift family (Sec. V-B). A fiber-only change leaves the feature distributions essentially intact while altering their relation to QoT, and the product in (5) vanishes in exactly that case, so C_j must be read and analyzed in combination with $\lambda^\star$.

B. LTT Uncertainty Bands Coverage

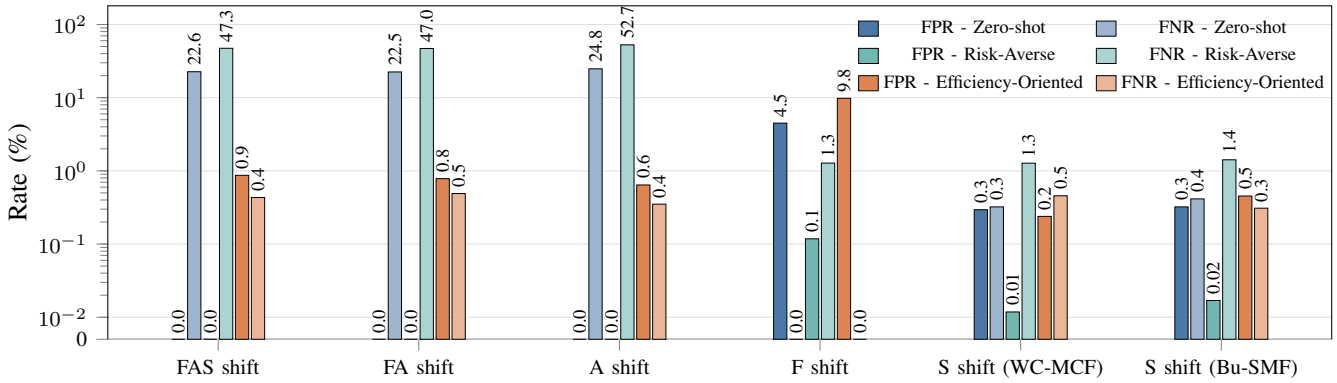
Table II compares the target-domain test coverage and full width of the LTT with a source-only QR uncertainty bands for each shift family. Source-only QR assesses whether intervals learned in the source domain retain their nominal coverage after a technology shift. Unlike LTT, it does not receive any target-domain calibration samples. Across the 170 scenario–partition trials, LTT achieves at least 90% coverage in every trial (minimum: 90.48%; mean: 91.81%). On held-out source-domain validation data, the nominal 90% QR band achieves 89.64% mean coverage (89.44–89.85% across seven source domains). On target-domain test data, however, it does not reach 90% coverage in any of the trials and has a mean coverage of 29.70%. Its mean full width is 0.209 dB, compared with 2.244 dB for LTT, but these narrow bands do not capture the shift-induced prediction errors. We exclude QR from the decision-policy comparison because its target-domain test coverage remains below the 90% target ($1-\alpha$, with $\alpha = 0.10$) in all 170 trials. Scenario-level results for all 17 transitions are reported in Supplementary Table S2.

This difference between LTT and QR uncertainty bands is also visible in Fig. 3. For LTT bands, Amplifier-related and mixed shifts require full widths of 4.2–4.7 dB, whereas switching-only shifts require about 0.2 dB. Wider bands reflect larger target-domain residuals, although their effect on feasibility decisions also depends on the residual direction and the OSNR margin to the threshold. In particular, a wide LTT band can cause the RA policy to reject otherwise feasible LPs, reducing resource utilization. Narrower bands would require smaller residuals or less finite-sample slack, for example by improving or adapting the predictor using data disjoint from calibration and then recalibrating it, or by collecting more representative calibration data. Simply reducing the band width at fixed α, δ, and residuals would invalidate the coverage guarantee. The narrow QR bands remain close to the source-domain estimates but often miss the target-domain OSNR, particularly for amplifier-related shifts. In contrast, CRC widens the band according to the shifted residuals. Thus, interval width should be interpreted jointly with coverage, rather than as evidence of reliable uncertainty on its own.

We examined the sensitivity to calibration set size by varying it from 500 samples to the full sets of 165,511–180,629 samples. With 500 samples, LTT selected a band in only 107/170 trials; therefore, we do not report aggregate performance. From 1,000 samples onward, a band was selected in all trials and attained at least 90% test coverage. Increasing the calibration size reduced overcoverage and band width in every shift family,



(a) Feasibility-decision performance.



(b) False positive rate (FPR) and false negative rate (FNR) across shift families.

Fig. 4: Feasibility-decision performance by source–target shift family. Values are means over ten target-domain calibration-test partitions per scenario and then across scenarios in each family. (a) F1 score and accuracy for TL, ZS, and the CRC-based RA and EO policies. (b) FPR and FNR for ZS, RA, and EO.

while amplifier-related and mixed shifts remained the widest. Thus, 1,000 samples was the smallest evaluated calibration set size satisfying both conditions, but this value cannot be generalized to other scenarios not investigated in this paper. Details are provided in Supplementary Table S3.

C. Feasibility-Decision Performance

Figs. 4a and 4b show decision performance and error type for the ZS, TL, and proposed CRC-based policies. Under FAS, FA, and A shifts, EO raises the degraded ZS accuracy from about 85% to above 99%, approaching TL levels. Switching-only shifts leave all methods near 99%. F shift reverses the ranking: RA is the strongest non-retraining method and EO falls below ZS. Across the 170 scenario–partition trials, the mean RA and EO accuracies are 85.26% and 98.92%, respectively. Using only 1,000 instead of all calibration samples drops them by less than 1%. Within each shift family, the change is at most 1.25 percentage points (Supplementary Table S4).

The results show that residual direction depends on the technology shift: amplifier-related shifts cause underestimation and high FNR, which EO reduces, whereas F shift causes overestimation and high FPR, which RA reduces. The direction of the shift cannot be assumed before target calibration data are observed. The symmetric LTT uncertainty band is therefore shift-agnostic and attains the target coverage in all trials. Once the direction is observed, however, equal widths may place unnecessary width on the less relevant side of the band. An asymmetric band could use different lower and upper widths to address this, but we leave its calibration and evaluation outside the scope of this study.

TL is an upper reference: it uses the same target-data budget but requires model access and retraining, whereas CRC recalibrates a fixed predictor; the two can also be combined [19]. The guarantee requires calibration–test exchangeability. Noise or aging represented during calibration can widen the band, whereas continued drift, abrupt instability, or evolving noise conditions are not covered.

VI. Conclusion

Technology upgrades can make a source-trained QoT regressor unreliable without revealing the loss in its point predictions. The proposed post-hoc CRC framework addresses this without retraining: target calibration data construct LTT uncertainty bands with finite-sample control of target miscoverage, feature shift scores characterize mismatch, and RA and EO use the bands for feasibility decisions. The guarantee applies to coverage; policy performance is evaluated empirically.

Across 17 shifts and ten target-data partitions, CRC met the 90% test-coverage target in all 170 trials, whereas source-only QR met it in none. With 1,000 samples, LTT selected a band in every trial with decision performance close to the full-calibration result; with 500 samples, it selected a band in 107/170 trials. Wider amplifier-related bands expose uncertainty from reusing the source regressor, while RA and EO outcomes show the trade-off between safety and resource utilization.

The guarantee requires calibration–test exchangeability. Aging, evolving noise, sensor shift, or link instability not represented during calibration require monitoring and recalibration. Future work should consider these effects, in addition to rolling calibration, and asymmetric bands. The same approach may support threshold-based regression tasks in optical networks.

REFERENCES

[1] R. Casellas, R. Martinez, R. Vilalta, and R. M. noz, "Overview of SDN control of multiband over SDM optical networks with physical layer impairments (invited tutorial)," *J. Opt. Commun. Netw.*, vol. 17, no. 2, pp. A165–A177, Feb 2025.

[2] C. Natalino and P. Monti, "Optical network automation, programmability, and AI: the path to 6G readiness," in *IEEE International Conference on Machine Learning for Communication and Networking (ICMLCN)*, 2025.

[3] V. Lohani, R. Munoz, R. Casellas, I. Tomkos, and D. M. Marom, "Dynamic routing and resource assignment for multi-granular optical node architectures with optical grooming," *Journal of Optical Communications and Networking*, vol. 18, no. 5, pp. 394–408, 2026.

[4] M. Klinkowski, P. Lechowicz, and K. Walkowiak, "Survey of resource allocation schemes and algorithms in spectrally-spatially flexible optical networking," *Optical Switching and Networking*, vol. 27, pp. 58–78, 2018.

[5] M. Yang, Q. Wu, and Y. Zhang, "Joint assignment of spatial granularity, routing, modulation, and spectrum in SDM-EONs: Minimizing the network capex considering spectrum, wss, and laser resources," *Journal of Lightwave Technology*, vol. 36, no. 18, pp. 4153–4166, 2018.

[6] C. Natalino, O. Ayoub, and P. Monti, "Engineering trust in AI-driven optical networks: Challenges in moving from monitoring to autonomy [invited tutorial]," in *International Conference on Optical Network Design and Modeling (ONDM)*, May 2026.

[7] P. Lechowicz, C. Natalino, and P. Monti, "QoT estimation with margin-driven transfer learning in time-varying optical networks," in *Optical Fiber Communications Conference and Exhibition (OFC)*, Mar. 2025, p. M1J.5.

[8] Z. Gu, T. Qin, Y. Zhou, J. Zhang, and Y. Ji, "Sample-distribution-matching-based transfer learning for QoT estimation in optical networks," *J. Opt. Commun. Netw.*, vol. 15, no. 9, pp. 649–663, 2023.

[9] Y. Zhou, Z. Gu, J. Zhang, and Y. Ji, "Evolutionary neuron-level transfer learning for QoT estimation in optical networks," *J. Opt. Commun. Netw.*, vol. 16, no. 4, pp. 432–448, Apr 2024.

[10] Z. Cai, Q. Wang, Y. Deng, P. Zhang, Z. Gai, Y. Li, and F. N. Khan, "Domain adversarial adaptation framework for few-shot QoT estimation in optical networks," *J. Opt. Commun. Netw.*, vol. 16, no. 11, pp. 1133–1144, 2024.

[11] Q. Wang, Z. Cai, and F. N. Khan, "Lifelong QoT prediction: an adaptation to real-world optical networks," *J. Opt. Commun. Netw.*, vol. 16, no. 11, pp. 1159–1169, 2024.

[12] T. Tanimura, T. Hoshida, T. Kato, and S. Watanabe, "OSNR estimation providing self-confidence level as auxiliary output from neural networks," *J. Lightw. Technol.*, vol. 37, no. 7, pp. 1717–1723, 2019.

[13] N. Di Cicco, M. Ibrahimi, C. Rottondi, and M. Tornatore, "Calibrated probabilistic QoT regression for unestablished lightpaths in optical networks," in *2022 International Balkan Conference on Communications and Networking (BalkanCom)*, 2022, pp. 21–25.

[14] N. Di Cicco, J. Talpini, M. Ibrahimi, M. Savi, and M. Tornatore, "Uncertainty-aware QoT forecasting in optical networks with bayesian recurrent neural networks," in *IEEE International Conference on Communications (ICC)*, 2023, pp. 441–446.

[15] S. Yousefi, H. Chouman, P. Djukic, F. Golaghazadeh, C. Tremblay, and C. Desrosiers, "Forecasting lightpath quality of transmission and implementing uncertainty in the forecast models," *J. Lightw. Technol.*, vol. 41, no. 15, pp. 4871–4881, 2023.

[16] Q. Guo, C. Zhang, X. Xiao, M. Zhang, T. Yang, X. Jiang, and D. Wang, "Conditional denoising diffusion probabilistic model for trustworthy failure prediction in optical networks," in *Optical Fiber Communication Conference (OFC)*, 2026, paper M3A.6.

[17] A. N. Angelopoulos and S. Bates, "A gentle introduction to conformal prediction and distribution-free uncertainty quantification," 2022. [Online]. Available: https://arxiv.org/abs/2107.07511

[18] A. N. Angelopoulos, S. Bates, A. Fisch, L. Lei, and T. Schuster, "Conformal risk control," 2022.

[19] K. Rezaei, O. Ayoub, P. Monti, and C. Natalino, "Policy-driven conformal prediction for trustworthy QoT estimation," in *Optical Fiber Communication Conference (OFC)*, 2026, p. M4A.3.

[20] A. N. Angelopoulos, S. Bates, E. J. Candès, M. I. Jordan, and L. Lei, "Learn then test: Calibrating predictive algorithms to achieve risk control," 2022. [Online]. Available: https://arxiv.org/abs/2110.01052

[21] H. Akbari, B. Shariati, J. L. M. Morrone, P. Safari, J. K. Fischer, and R. Freund, "Datasets for QoT estimation in SDM networks," *J. Opt. Commun. Netw.*, vol. 17, no. 6, pp. 514–525, Jun 2025.

[22] T. Panayiotou, S. P. Chatzis, and G. Ellinas, "Performance analysis of a data-driven quality-of-transmission decision approach on a dynamic multicast-capable metro optical network," *J. Opt. Commun. Netw.*, vol. 9, no. 1, pp. 98–108, Jan 2017.

[23] C. Rottondi, L. Barletta, A. Giusti, and M. Tornatore, "Machine-learning method for quality of transmission prediction of unestablished lightpaths," *J. Opt. Commun. Netw.*, vol. 10, no. 2, pp. A286–A297, Feb 2018.

[24] M. Ibrahimi, H. Abdollahi, C. Rottondi, A. Giusti, A. Ferrari, V. Curri, and M. Tornatore, "Machine learning regression for QoT estimation of unestablished lightpaths," *J. Opt. Commun. Netw.*, vol. 13, no. 4, pp. B92–B101, Apr 2021.

[25] Y. Pointurier, "Machine learning techniques for quality of transmission estimation in optical networks," *J. Opt. Commun. Netw.*, vol. 13, no. 4, pp. B60–B71, Apr 2021.

[26] R. Ayassi, A. Triki, N. Crespi, R. Minerva, and M. Laye, "Survey on the use of machine learning for quality of transmission estimation in optical transport networks," *J. Lightw. Technol.*, vol. 40, no. 17, pp. 5803–5815, 2022.

[27] S. Allogba, S. Aladin, and C. Tremblay, "Machine-learning-based lightpath QoT estimation and forecasting," *J. Lightw. Technol.*, vol. 40, no. 10, pp. 3115–3127, 2022.

[28] Y. Zhang, Q. Qiu, X. Liu, L. Yi, W. Hu, and Q. Zhuge, "SAMA-IR: Sensitivity-analysis guided memory-aware input refinement methodology for large-scale optical networks with field-trial validation," *J. Lightw. Technol.*, vol. 43, no. 18, pp. 8603–8611, 2025.

[29] X. Yang, C. Sun, R. Ayassi, L. Aknin, G. Charlet, M. Tornatore, and Y. Pointurier, "Input refinement with incremental learning for accurate digital twin-enabled self-driven QoT optimization in optical networks," *J. Opt. Commun. Netw.*, vol. 18, no. 1, pp. A75–A87, 2026.

[30] O. Karandin, A. Ferrari, F. Musumeci, Y. Pointurier, and M. Tornatore, "Probabilistic low-margin optical-network design with multiple physical-layer parameter uncertainties," *J. Opt. Commun. Netw.*, vol. 15, no. 7, pp. C129–C137, Jul 2023.

[31] A. V. Paz, R. R. Reyes, and P. Sahu, "Planning multilayer networks under deep uncertainties: an approach based on flexibility in engineering design," *J. Opt. Commun. Netw.*, vol. 18, no. 4, pp. 338–353, 2026.

[32] A. A. Housseini, C. Natalino, P. Monti, and O. Ayoub, "Cross-domain generalization in optical networks via joint contrastive and classification learning," in *International Conference on Optical Network Design and Modeling (ONDM)*, May 2026.

[33] Y. Romano, E. Patterson, and E. J. Candès, "Conformalized quantile regression," in *Advances in Neural Information Processing Systems*, vol. 32, 2019. [Online]. Available: https://papers.nips.cc/paper/8613-conformalized-quantile-regression.pdf

[34] R. J. Tibshirani, R. F. Barber, E. J. Candès, and A. Ramdas, "Conformal prediction under covariate shift," in *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[35] I. Gibbs and E. J. Candès, "Adaptive conformal inference under distribution shift," in *Advances in Neural Information Processing Systems*, vol. 34, 2021. [Online]. Available: https://proceedings.neurips.cc/paper_files/paper/2021/hash/0d441de75945e5acbc865406fc9a2559-Abstract.html

[36] C. Villani, *Optimal Transport: Old and New*, 2009.

[37] A. Mitra, D. Ives, A. Lord, S. Savory, S. Kar, and P. Wright, "Network equipment and their procurement strategy for high capacity elastic optical networks," *J. Opt. Commun. Netw.*, vol. 8, no. 7, pp. A201–A211, Jul 2016.