IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

Enhancing OLLA via Exponential Decay for Efficient Link Adaptation in Emerging 6G Traffic

ARITRA MAZUMDAR[1], (Graduate Student Member, IEEE), STEFANO PARIS[2], ABOLFAZL AMIRI[3], KLAUS I. PEDERSEN[1,3], (Senior Member, IEEE), RAMONI ADEOGUN[1], (Senior Member, IEEE)

[1]Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark
[2]Nokia Networks, 12 Rue Jean Bart Massy, France
[3]Nokia Standards, 9220 Aalborg, Denmark

Corresponding author: Aritra Mazumdar (e-mail: aritram@es.aau.dk).

ABSTRACT Outer loop link adaptation (OLLA) refines link adaptation by offsetting the channel quality indicator report. However, its slow convergence—the time required for the algorithm to stabilize to a reliable operating point—can pose challenges for emerging sixth generation (6G) services with strict quality of service requirements. In this work, we propose an exponential decay (ED)-based enhanced OLLA (EOLLA) framework that accelerates convergence while preserving stability. The key idea is to exploit instantaneous channel measurements at the user equipment to form an enhanced quality indicator, which drives the EOLLA offset updates at the gNodeB along an ED trajectory for rapid convergence. To prevent these fast adjustments from undermining robustness, EOLLA incorporates long-term system-level a-priori statistics into its design, enabling stable offset evolution under realistic channel dynamics. Closed-form expressions and a practical implementation with inherently low complexity are provided to support real-world deployment. System-level simulations across diverse deployments and traffic models show consistent improvements: moderate gains for enhanced mobile broadband and up to 33% capacity increase for extended reality traffic (characterized with bounded latency and reliability constraints) in indoor hotspot deployment. These results highlight EOLLA as a promising candidate for future 6G systems with demanding performance constraints.

INDEX TERMS Outer loop link adaptation, exponential decay, nonlinear offset adjustment, indoor hotspot, dense urban-macro, FTP3, XR, system-level simulations, fast convergence, 6G.

I. INTRODUCTION

The evolution toward sixth-generation (6G) wireless networks brings both new opportunities and challenges for link adaptation (LA). Beyond supporting higher data rates, 6G must handle diverse traffic types with stringent reliability and latency requirements, including holographic communication, industrial automation, and immersive extended reality (XR) [1]. These services are often characterized by short packets, unpredictable traffic bursts, and strict quality-of-service constraints, which amplify the need for adaptive radio resource management. Meeting these demands calls for LA mechanisms that are not only throughput-efficient but also responsive to highly dynamic and bursty traffic patterns.

Conventional LA, illustrated in Figure 1, consists of two components. The inner loop link adaptation (ILLA) selects the most suitable modulation and coding scheme (MCS) for each transmission to meet the instantaneous target transport block (TB) error rate (TBER) or TTBER. This selection relies on the channel quality indicator (CQI) reported by the user equipment (UE), derived from signal-to-interference-plus-noise ratio (SINR) measurements of reference signals. However, the reported CQI cannot be directly used for MCS determination at the gNodeB (gNB) because of factors such as quantization, estimation errors, reporting delays, and inherent channel variability. To compensate for these effects and ensure that the long-term experienced TBER aligns with the TTBER across heterogeneous UEs, outer loop LA (OLLA) [2], [3] introduces a correction in the form of an offset applied to the reported CQI. Since UEs differ in measurement accuracy and channel dynamics, the gNB maintains a separate OLLA instance for each UE. Owing to its central role in balancing reliability and efficiency

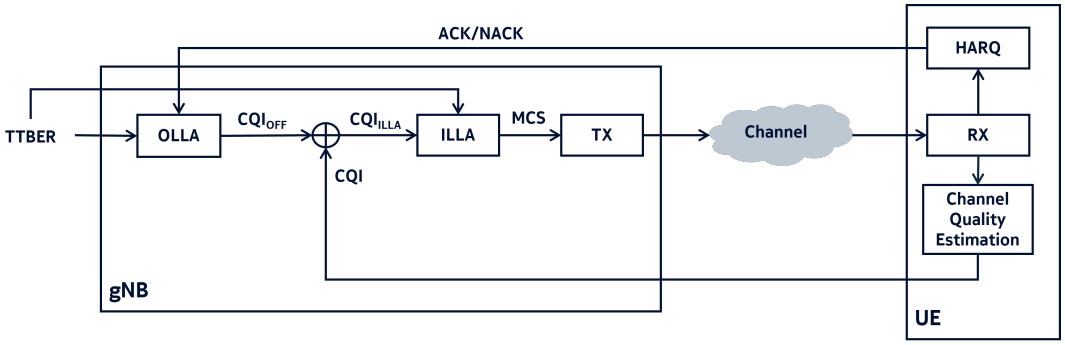


FIGURE 1: Architecture of conventional link adaptation (LA). Inner loop LA (ILLA) selects the most appropriate modulation and coding scheme (MCS) for transmission (TX) of the transport block (TB). The outer loop LA (OLLA) at the gNodeB applies a constant offset to the user equipment (UE)-reported channel quality indicator (CQI), based on hybrid automatic repeat request (HARQ) feedback generated upon reception (RX), to compensate for reporting imperfections and achieve the long-term target TB error rate (TTBER).

in realistic channel conditions, OLLA continues to attract research interest.

A. RELATED STUDIES ON OUTER LOOP LINK ADAPTATION

In [4], the authors analyze the impact of the TTBER on legacy OLLA performance and show that a value of 10^{-1} yields optimal throughput. Furthermore, they propose a Bayesian OLLA algorithm that exploits the mutual information conveyed by hybrid automatic repeat request (HARQ) feedback. While this approach can outperform legacy OLLA at the same TTBER, its effectiveness diminishes at lower TTBER values due to reduced average mutual information, reaffirming the sensitivity of OLLA performance to TTBER configuration. In contrast, [5] introduces a dynamic OLLA scheme that recomputes the optimal TTBER based on the mean and standard deviation of the post-interference rejection combining SINR at the receiver. This dynamic adjustment improves adaptability to varying interference conditions and reduces the number of users experiencing outage. Together, these studies highlight the importance of refining TTBER design for improving OLLA robustness.

A complementary research direction seeks to make OLLA more responsive to instantaneous channel dynamics for ultra-reliable low-latency communication. In [6], a soft-feedback mechanism is proposed where the UE issues a "low ACK" (acknowledgment) if the estimated packet-error probability exceeds a threshold, triggering an OLLA offset adjustment to reduce the likelihood of a subsequent negative ACK (NACK). Similarly, [7] uses either TB error probability or decoder-iteration counts as indicators of decoding difficulty in a comparable approach. Although such schemes enhance responsiveness, their reliance on instantaneous estimates can make the offset trajectory less stable under fast channel variations.

Other studies emphasize improving offset initialization. In [8], for 4G systems, the initial OLLA offset is set as a function of the time elapsed since the UE transitioned from idle to connected state, compensating for the lack of immediate HARQ feedback after connection. In [9], online heuristics are employed to compute the median offset from large activity connections, which is then used as the initial offset to accelerate OLLA. These approaches reduce the startup delay of OLLA but cannot influence its convergence rate once in operation.

Another line of research focuses on learning-based adaptive OLLA, targeting scenario-specific optimization of the OLLA offset. In [10], this is achieved using a continuous function to generate TBER values. Both the parameters of this function and other learnable parameters within the optimization process are updated online using a loss function based on the applied MCS decisions and corresponding HARQ feedback (ACK or NACK), producing an OLLA offset tailored to the current scenario. The study in [11] employs a Q-learning framework to refine an adaptive OLLA's parameters through deep reinforcement learning using a window of HARQ observations, enhancing adaptation to diverse channel conditions. Along similar lines, [12] uses Q-learning to assist OLLA in maximizing the coding rate while sticking to TBER requirements. Similarly, self-adaptive LA or SALAD [13] leverages online learning to self-adapt its SINR estimate from ACK/NACK feedback to optimize the MCS selection, achieving high spectral efficiency. While such online learning-based schemes improve flexibility, they introduce additional runtime computational cost and design complexity. In contrast, [14] uses offline reinforcement learning to train LA policies from pre-collected data with batch-constrained and conservative Q-learning as well as sequence modeling, thus avoiding online updates while matching online reinforcement learning performance. Nevertheless, offline learning-based LA can overfit to the data used during training, limiting its robustness to distribution shifts.

In contrast, NOLLA [15] proposes a nonlinear OLLA scheme that adjusts the offset step size based on consecutive HARQ feedback, thereby eliminating the need for online learning and reducing scenario-specific bias. However, its performance relies on the probability mass function of consecutive ACK/NACK sequences, which depends on accurately estimating the error probability—a challenging task in practice.Furthermore, since its initial step size is smaller than legacy OLLA and continues to decay over time, convergence remains slow when far from the desired operating point. This

further exacerbates legacy OLLA's slow convergence, which can challenge the system's ability to meet the strict quality of service (QoS) requirements in emerging 6G networks.

Motivated by these limitations, this work focuses on the design of a simple and robust OLLA scheme that achieves fast convergence while preserving stable offset behavior.

B. CONTRIBUTIONS

We present an enhanced OLLA (EOLLA) scheme at the gNB for improving LA decisions through faster convergence. The main contributions are summarized as follows:

- We propose an exponential decay (ED)-based nonlinear OLLA offset evolution scheme. It leverages the UE's instantaneous channel measurements to generate an enriched quality indicator (EQI) feedback, which is then used at the gNB to update the offset following an ED trajectory. By incorporating long-term system-level a-priori statistics, the scheme ensures stability throughout the convergence process.
- We establish a set of design principles and derive closed-form expressions that integrate the a-priori statistics into the system design. As part of this design, we introduce an ED index for capturing the nonlinear trajectory of the OLLA offset. In addition, we outline a practical realization of the proposed design and show that its computational footprint is inherently low.
- We evaluate the proposed scheme through system-level simulations (SLS) in dense urban-macro (DU) and indoor hotspot (InH) scenarios under 3rd Generation Partnership Project (3GPP)-aligned traffic models, including file transfer protocol model 3 (FTP3) [16] and XR. Results across diverse multi-user and multi-cell deployments show consistent improvements across multiple key performance indicators (KPIs), ranging from moderate gains in FTP3 to substantial gains in XR, highlighting the advantages of optimizing OLLA convergence with our proposed approach.

C. NOTATION

Throughout this article, regular symbols (e.g., a, A) represent scalars, bold uppercase script letters (e.g., $\boldsymbol{\mathcal{A}}$) denote sets, and blackboard uppercase letters (e.g., $\mathbb{A}$) are used for operators.

D. STRUCTURE OF THE ARTICLE

The rest of the article is organized as follows. Section II reviews legacy OLLA and its limitations. Section III introduces the proposed enhancement to OLLA and outlines its low-complexity implementation. Section IV details the system model. Section V presents the simulation setup and highlights the key performance improvements achieved by the proposed method compared to the legacy approach. Section VI concludes the article. Finally, an acknowledgment is included, followed by Appendix A, which lists the abbreviations used in this paper and Appendix B, which presents the important derivations.

II. BACKGROUND

A. OUTER LOOP LINK ADAPTATION OVERVIEW

OLLA refines the UE-reported CQI, denoted by $\mathrm{CQI}(k)$ at every first HARQ transmission indexed by k, by applying an offset $\mathrm{CQI}_{\mathrm{OFF}}(k)$. The corrected CQI used by ILLA for MCS selection is then

$$\mathrm{CQI}_{\mathrm{ILLA}}(k) = \mathrm{CQI}(k) + \mathrm{CQI}_{\mathrm{OFF}}(k). \tag{1}$$

The offset evolves over time according to the HARQ feedback as:

$$\mathrm{CQI}_{\mathrm{OFF}}(k) = \mathrm{CQI}_{\mathrm{OFF}}(k-1) + \Delta_u\big(1-e(k)\big) - \Delta_d e(k), \tag{2}$$

where $e(k)$ is a binary error indicator taking 0 for an ACK and 1 for a NACK, and Δ_u and Δ_d denote the constant offset increment and decrement, also referred to as step-up (SU) and step-down (SD), respectively.

In practice, legacy OLLA carefully increases the MCS by gradually stepping up the offset until decoding errors occur, triggering a NACK. When this happens, the offset is sharply stepped down to drive ILLA to select a more robust MCS. This process iterates over time and by controlling the ratio of SU to SD, the long-term error behavior can be regulated to meet the TTBER. In particular, the relationship between SU, SD, and the resulting TTBER is expressed as:

$$\mathrm{TTBER} = \frac{1}{1+\frac{\Delta_d}{\Delta_u}}. \tag{3}$$

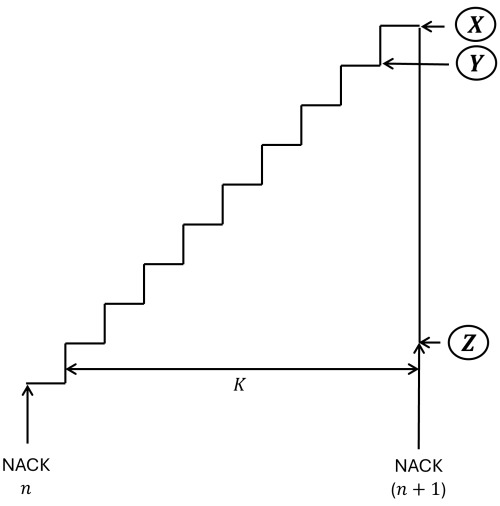


FIGURE 2: Illustration of the evolution of the offset under legacy OLLA. The step-up (SU) region between two NACKs is highlighted by reference points X (error-inducing offset), Y (highest reliable offset), and Z (post step-down offset).

B. LIMITATIONS OF LEGACY OLLA

To motivate our research, we consider the illustrative example in Figure 2 showing OLLA offset evolution. The depicted SU region spans the interval between two consecutive NACKs: let n denote the first NACK just before $k = 1$ and $(n+1)$ the second just after $k = K$. Point X marks an offset ex-

ceeding the highest reliable MCS, causing a NACK; Y is the highest reliable offset yielding an ACK; and Z is the reduced offset after the SD. Ideally, the offset should converge more quickly to a stable operating point within $[Y, X]$ to prevent the first NACK, or alternatively explore a stable point within $[X, Z]$ to just avoid the second. This example highlights a key limitation of legacy OLLA: it spends considerable time probing suboptimal offsets, leading to prolonged low CQI levels. The cause is the fixed-size SU/SD mechanism, which adjusts $\mathrm{CQI}_{\mathrm{OFF}}(k)$—and thus $\mathrm{CQI}_{\mathrm{ILLA}}(k)$—linearly with k. Such linear tracking is inherently slow.

One might attempt to accelerate convergence by using bigger jumps after ACKs and smaller ones after NACKs—but this may destabilize OLLA, potentially increasing the TBER and ultimately degrading QoS. These drawbacks motivate the need for an EOLLA strategy that (i) converges more rapidly, (ii) respects long-term system-level statistics such as the TTBER, and (iii) requires minimal computational overhead. The following section elaborates on how EOLLA achieves each of these three goals.

III. PROPOSED ENHANCED OLLA – EXPONENTIAL DECAY WITH LONG-TERM STATISTICS

A. EXPONENTIAL DECAY-BASED OFFSET ADJUSTMENT

To speed up OLLA convergence, we adopt ED for offset adjustment: larger corrections are applied when the offset is far from the expected stable operating point, and smaller corrections are applied as it approaches the point to maintain stability. This is highlighted in Figure 3, where the safe range Y' corresponds to early, aggressive corrections, and the unsafe range X' corresponds to later, cautious adjustments around convergence. The ED scaling can be expressed as:

$$\Delta^{\mathrm{ED}} = \Delta^{\mathrm{L}} B^{\pm E}, \tag{4}$$

where $\Delta^{\mathrm{ED}} \in \{\Delta_u^{\mathrm{ED}}, \Delta_d^{\mathrm{ED}}\}$ and $\Delta^{\mathrm{L}} \in \{\Delta_u, \Delta_d\}$ are the ED and legacy step sizes, respectively, and B and E are the ED base and exponent.

To simplify the optimization, we tune only one parameter: B determines the scaling of the trajectory, while E shapes the trajectory and directly governs the convergence speed. Thus, we fix $B = 2$ and optimize E. To do so for adaptive offset evolution, we introduce the effective ED step index κ, which indexes the nonlinear trajectory of the offset. To prevent the ED steps from becoming too large or too small and destabilizing the trajectory, we introduce a reference cap C, yielding the ED step:

$$\Delta^{\mathrm{ED}} = \Delta^{\mathrm{L}} 2^{\pm(C-\kappa)}. \tag{5}$$

The key adaptation lies in $(C - \kappa)$. When $\kappa < C$, the exponent is positive and SU step sizes exceed legacy values, enabling rapid upward correction in the safe range (Y'). When $\kappa > C$, the exponent is negative, shrinking SU steps below legacy and slowing adjustments in the unsafe range (X'), preventing overshoot toward X (Figure 2) and unnecessary first NACKs. Conversely, in SD, updates start smaller than legacy

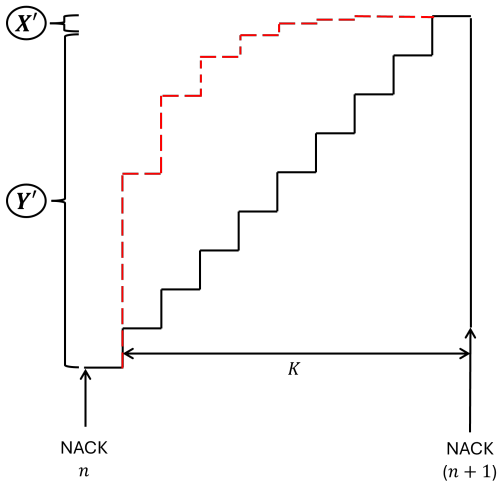


FIGURE 3: Illustration of the evolution of the offset with exponential decay (ED) relative to legacy OLLA. Y' marks the safe range for larger updates, while X' indicates the unsafe range where only smaller updates are applied.

to compensate for reduced overshoot under ED, promoting stabilization between X and Z (Figure 2), then grow once $\kappa > C$ to avoid repeated NACKs. This achieves the desired 'fast–then–slow' behavior for SU and 'slow–then–fast' for SD, overcoming the legacy limitation of prolonged operation at low CQI levels and enabling longer use of higher MCS levels. Due to the inherent asymmetry in SU and SD dynamics, using separate reference caps is beneficial, leading to the opposite scaling trajectories captured by:

$$\Delta_u^{\mathrm{ED}} = \Delta_u 2^{(C_u-\kappa)}, \tag{6}$$

$$\Delta_d^{\mathrm{ED}} = \Delta_d 2^{-(C_d-\kappa)}, \tag{7}$$

where C_u and C_d are SU- and SD-specific reference caps.

This adaptive ED formulation enables faster convergence; however, the reference caps must be chosen carefully. They control how the nonlinear evolution of κ scales the effective ED step, and therefore provide a natural point to incorporate long-term system-level statistics, ensuring OLLA's stability by balancing aggression in the safe region and conservatism in the unsafe region.

B. LONG-TERM STATISTICS-AWARE OFFSET DESIGN

To derive the reference caps, we apply a set of sequential design principles that preserve the intended ED behavior. We begin by relating the ED trajectory to that of legacy OLLA. Under legacy SU behavior, the residual offset from the current index κ up to the next expected NACK at K is

$$\Delta_u(K - \kappa + 1), \tag{8}$$

accumulated over $(K - \kappa + 1)$ ACKs with fixed steps of size Δ_u. As a natural point of comparison, we match the total residual offset of ED to that of legacy:

$$\sum_{i=\kappa}^{K} \Delta_u 2^{(C_u-i)} = \Delta_u(K - \kappa + 1), \tag{9}$$

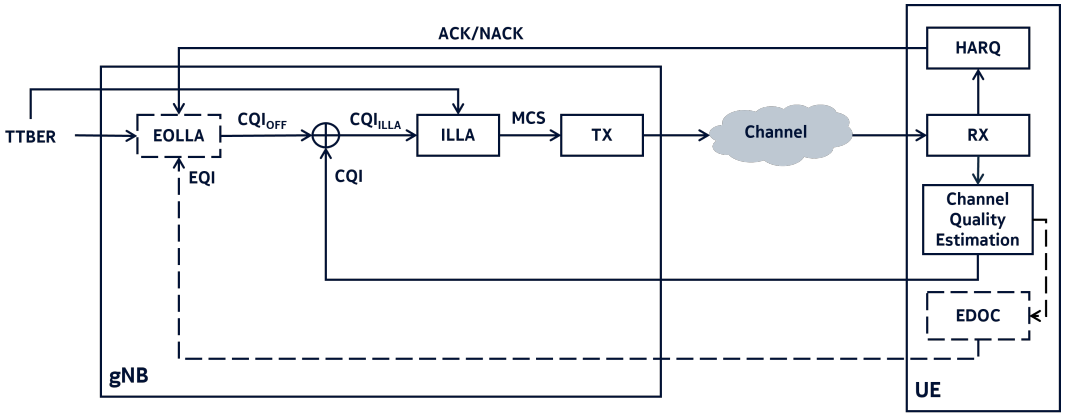


FIGURE 4: Architecture of enhanced LA. The enhanced OLLA (EOLLA) at the gNB offsets the CQI to achieve the long-term TTBER faster than legacy. The ED offset controller (EDOC) at the UE generates the enhanced quality indicator (EQI), which, along with the HARQ feedback, is used by EOLLA for adaptive CQI offsetting. New components are highlighted with dashed outlines.

providing a baseline. Using K as the upper index naturally enforces the ED trajectory to dwell longer near the expected convergence, while leaving flexibility for subsequent trajectory shaping. Adjusting the equation's left-hand limits followed by canceling its common term and then applying the geometric-series formula

$$\sum_{i=0}^{I-1} r^i = \frac{1-r^I}{1-r}, \quad r \neq 1$$

simplifies the left side of it to:

$$\sum_{i=\kappa}^{K} 2^{(C_u-i)} = 2^{(C_u-\kappa+1)}\left(1-2^{-(K-\kappa+1)}\right). \tag{10}$$

(see the full step-by-step derivation of the reference caps in Appendix B). Following the same algebraic steps for SD (refer to Appendix B) updates gives:

$$\sum_{i=\kappa}^{K} 2^{-(C_d-i)} = 2^{(\kappa-C_d)}\left(2^{(K-\kappa+1)}-1\right). \tag{11}$$

Up to this point, the ED trajectory has been expressed over a residual $(K-\kappa+1)$-step horizon, the same as legacy. However, ED and legacy trajectories need not share the same horizon. To enable proper trajectory shaping and to realize the desired 'fast–then–slow' evolution of the OLLA offset in SU, the effective horizon in the baseline must be reshaped relative to K. We therefore introduce a scaled number of residual ED steps, K_{SU}, starting after the current index κ, which shortens the residual horizon by a factor $\alpha \in (0,1]$, giving

$$K_{\mathrm{SU}} = \alpha\,(K-\kappa+1), \quad \alpha \in (0,1].$$

This scaling thus stretches the offset-evolution curve in the safe zone (Y' in Figure 3), amplifying early jumps, while compressing it near the target in the unsafe zone (X') to prevent overshoot. With this redefined horizon, the summation in equation (10) becomes:

$$\sum_{i=\kappa}^{\kappa+K_{\mathrm{SU}}} 2^{(C_u-i)} = 2^{(C_u-\kappa+1)}\left(1-2^{-K_{\mathrm{SU}}}\right), \tag{12}$$

where the upper limit $(\kappa+K_{\mathrm{SU}})$ now reflects the total number of ED steps. Similarly, in the SD phase, the horizon is slightly extended to moderate the descent:

$$K_{\mathrm{SD}} = (1+\alpha)\,(K-\kappa+1),$$

producing a controlled 'slow–then-fast' evolution and modifying equation (11) as:

$$\sum_{i=\kappa}^{\kappa+K_{\mathrm{SD}}} 2^{-(C_d-i)} = 2^{(\kappa-C_d)}\left(2^{K_{\mathrm{SD}}}-1\right). \tag{13}$$

Since (12) and (13) depend on K (through K_{SU} and K_{SD})-the streak of legacy-equivalent steps which is not known in practice, we replace K with its long-term expectation to make the ED trajectory implementable. Offset streaks vary with channel conditions, so their expected values are conditioned on the observed CQI. We introduce a CQI-based mapping to the expected offset adjustment streak:

$$\mathbb{N} : \mathrm{CQI} \in \mathcal{C} \mapsto \boldsymbol{\mathcal{N}}^{\mathrm{CQI}} = \left\{N_{\Delta_u}^{\mathrm{CQI}}, N_{\Delta_d}^{\mathrm{CQI}}\right\}, \tag{14}$$

where $\mathcal{C}$ denotes the set of CQI indices with cardinality M and the tuple $\{N_{\Delta_u}^{\mathrm{CQI}}, N_{\Delta_d}^{\mathrm{CQI}}\}$ gives the expected counts of consecutive SUs and SDs for each CQI in $\mathcal{C}$. At time k, the mapping is applied as:

$$N_{\Delta_u}^{\mathrm{CQI}} = \mathbb{N}(\mathrm{CQI}(k)), \quad e(k)=0, \tag{15}$$

$$N_{\Delta_d}^{\mathrm{CQI}} = \mathbb{N}(\mathrm{CQI}(k)), \quad e(k)=1. \tag{16}$$

This mapping provides a-priori guidance for OLLA offset evolution, while adhering to long-term statistics. Using $\boldsymbol{\mathcal{N}}^{\mathrm{CQI}}$ in place of K in the expressions for K_{SU} and K_{SD} gives

$$K_{\mathrm{SU}} = \alpha\,(N_{\Delta_u}^{\mathrm{CQI}}-\kappa+1),$$

$$K_{\mathrm{SD}} = (1+\alpha)\,(N_{\Delta_d}^{\mathrm{CQI}}-\kappa+1).$$

Substituting K_{SU} into (12) and solving for C_u gives:

$$C_u = (\kappa - 1) + \log_2\left(\frac{N_{\Delta_u}^{\mathrm{CQI}} - \kappa + 1}{1 - 2^{-\alpha(N_{\Delta_u}^{\mathrm{CQI}} - \kappa + 1)}}\right), \tag{17}$$

and, similarly, substituting K_{SD} into (13) yields:

$$C_d = \kappa + \log_2\left(\frac{2^{(1+\alpha)(N_{\Delta_d}^{\mathrm{CQI}} - \kappa + 1)} - 1}{N_{\Delta_d}^{\mathrm{CQI}} - \kappa + 1}\right). \tag{18}$$

C. PRACTICAL LOW-COMPLEXITY IMPLEMENTATION OVERVIEW

The proposed EOLLA scheme at the gNB is assisted by two complementary components at the UE: the ED offset controller (EDOC) and the EQI feedback embedded in UE–gNB signaling. Together, these three modules—highlighted with dashed outlines in Figure 4—enable the practical realization of EOLLA. Their roles are summarized below.

At UE, the EDOC performs two main types of operations.

- *Storage:* (i) A static lookup table (LUT) implementing the mapping $\mathbb{N}$ between the sorted CQI values and their corresponding offset streak values $\mathcal{N}^{\mathrm{CQI}}$, and (ii) a dynamic storage of $\mathrm{CQI}_{\mathrm{OFF}}(0)$, which is updated whenever the SU/SD streak resets (i.e., $e(k) \neq e(k-1)$). The latter relies on the most recent $\mathrm{CQI}_{\mathrm{OFF}}(k-1)$, assumed to be communicated by the gNB via downlink (DL) control information before the next HARQ first transmission.
- *Computation:* Based on the stored values, the EDOC computes the effective step index κ (Algorithm 1), the SU/SD-specific reference caps C_u or C_d, the corresponding multiplicative factor

$$F = 2^{\pm(C_{u/d} - \kappa)}, \tag{19}$$

used to scale the legacy step in (6) or (7), and finally the N_E-bit EQI feedback as detailed in Algorithm 1. The step-by-step average asymptotic complexity [17] for its generation is summarized in Table 1. Most steps exhibit an asymptotic complexity $\mathcal{O}(1)$, as they involve simple arithmetic operations or comparisons. The only two steps with potentially higher complexity are: (i) a binary search to fetch the expected OLLA offset adjustment streak from a LUT with M entries, with complexity of $\mathcal{O}(\log_2 M)$, and (ii) the binary encoding of the quantized multiplicative factor F into N_E bits, which can represent 2^{N_E} possible values, with complexity $\mathcal{O}(N_E)$. Since both the LUT and EQI implement discrete mappings, and the EQI follows the LUT, the LUT resolution should satisfy $M \geq 2^{N_E}$ entries to ensure the EQI's resolution is fully supported and to avoid additional approximation errors. This makes the LUT binary search the dominating complexity term. The LUT size can be adapted to meet implementation constraints or standardization requirements; for example, representing the full 5G-Advanced CQI range of 0–15 would require $\log_2(M) \leq 4$, illustrating an upper bound where this step is effectively constant-time and thus of low complexity.

At gNB, the EOLLA mechanism replaces the legacy OLLA block. It performs two computations: (i) the conventional setting of the legacy step sizes Δ^{L} from the TTBER using (3), and (ii) and an additional derivation of the ED-scaled step Δ^{ED} from the multiplicative factor decoded from the EQI. This derivation includes a reverse operation of binary decoding followed by a multiplication, both of which are lightweight with respect to the available resources at the gNB. The resulting Δ^{ED} is then used to update the offset, yielding a modified version of (2) in which the legacy step sizes are replaced by their ED counterparts:

$$\mathrm{CQI}_{\mathrm{OFF}}(k) = \mathrm{CQI}_{\mathrm{OFF}}(k-1) + \Delta_u^{\mathrm{ED}}\big(1 - e(k)\big) - \Delta_d^{\mathrm{ED}} e(k), \tag{20}$$

after which the updated offset is embedded in the DL control information for signaling to the UE (not explicitly shown in the figure to avoid clutter, as it is not a core novelty).

Algorithm 1: Generation of enhanced quality indicator (EQI) at the UE.

Input: $\mathrm{CQI}_{\mathrm{OFF}}(k-1), e(k), \mathrm{CQI}(k), \mathbb{N}, B = 2$;
Initialize: $\mathrm{CQI}_{\mathrm{OFF}}(0)$;
for $k = (1, .., K)$ **do**
  Determine the ED effective step κ with $\left\lceil \{\mathrm{CQI}_{\mathrm{OFF}}(k-1) - \mathrm{CQI}_{\mathrm{OFF}}(0)\}/\Delta^{\mathrm{L}} \right\rceil$;
  if $e(k) = 0$ **then**
    Estimate CQI-based expected streak of SUs $N_{\Delta_u}^{\mathrm{CQI}}$ for a given rounded $\mathrm{CQI}(k)$ using $\mathbb{N}$ in (15);
    Calculate the ED SU reference cap C_u using $N_{\Delta_u}^{\mathrm{CQI}}$ and κ in (17);
    Compute the ED SU multiplicative factor F via (19) using C_u and the given B;
  else
    Estimate CQI-based expected streak of SDs $N_{\Delta_d}^{\mathrm{CQI}}$ for a given rounded $\mathrm{CQI}(k)$ using $\mathbb{N}$ in (16);
    Calculate the ED SD reference cap C_d using $N_{\Delta_d}^{\mathrm{CQI}}$ and κ in (18);
    Compute the ED SD multiplicative factor F via (19) using C_d and the given B;
  Quantize the continuous multiplicative factor F to obtain $\hat{F}$ using standard quantization methods [18]–[20];
  Construct EQI using the binary encoding function as $f_{\mathrm{BE}} : \hat{F} \mapsto \{0,1\}^{N_E}$;
return $\mathrm{EQI}(k)$;
Output: $[\mathrm{EQI}(1), \mathrm{EQI}(2), .., \mathrm{EQI}(k')]|_{k' \in \{1,2,\ldots,K\}}$.

IV. SYSTEM MODEL

TABLE 1: Average asymptotic complexity analysis for the EQI(k) generation algorithm at the UE.

Step	Operations	Complexity
Determine κ	1 subtraction + 1 division + 1 ceiling	$\mathcal{O}(1)$
Check $e(k)$	1 Boolean comparison	$\mathcal{O}(1)$
Estimate $N^{\text{CQI}}_{\Delta_{u/d}}$	1 LUT fetch	$\mathcal{O}(\log_2 M)$
Compute $C_{u/d}$	5 subtractions/ additions + 2 divisions/ multiplications + 1 exponentiation + 1 logarithm	$\mathcal{O}(1)$
Calculate F	1 subtraction + 1 exponentiation	$\mathcal{O}(1)$
Estimate $\hat{F}$	Quantization	$\mathcal{O}(1)$
Generate EQI(k)	Binary mapping	$\mathcal{O}(N_E)$

A. DEPLOYMENT SCENARIO AND FRAME STRUCTURE

We model a 5G time division duplex (TDD) network deployed in the two most common 3GPP scenarios: an InH with $N_{\text{C}} = 12$ cells and a DU with $N_{\text{C}} = 21$ cells, where each cell is indexed by $\text{c} \in \{1, 2, \ldots, N_{\text{C}}\}$. These scenarios are selected to capture different propagation characteristics, interference patterns, variations in output power, and cell density, all of which significantly impact LA and HARQ performance. Each gNB is equipped with 32 antenna ports. In the DU scenario, each gNB manages three cells, while in the InH scenario, each gNB serves a single cell without sectorization. The peak transmit gNB power is 31 dBm for InH and 51 dBm for DU. Each cell serves N_{U} UEs, indexed by $\text{u} \in \{1, 2, \ldots, N_{\text{U}}\}$, which are uniformly placed throughout the network to maintain balanced load across all cells. The model further assumes dynamic time–frequency scheduling (i.e., allocation of transmission resources), dynamic LA with variable modulation and coding schemes, and HARQ with Chase soft combining.

B. TRAFFIC MODELS AND RELATED KPIs

To evaluate the proposed solution under diverse conditions, we consider two traffic models reflecting different QoS requirements, allowing a broad sensitivity analysis across network conditions. Specifically, we model FTP3 traffic to represent best-effort throughput-driven eMBB services, and an XR traffic model to capture the stringent reliability and low-latency demands of emerging 6G applications.

The first traffic model is based on a bursty file transfer configuration aligned with FTP3 [16], with guaranteed bit rate constraints. In this setup, files of size μ_β bits are transmitted, with their arrivals at the gNB governed by a homogeneous Poisson process characterized by a fixed arrival rate, λ.

The second traffic model targets XR use cases and assumes quasi-periodic video frame generation at an application server, intended for delivery to XR UEs over the 5G radio access network. Frames are generated at a fixed rate, λ (frames per second). Due to factors such as encoding and routing variability, their arrival times at the gNB exhibit quasi-periodicity with jitter τ, modeled [22], [23] as a truncated Gaussian distribution:

$$\tau = \mathcal{TN}(\mu_\tau, \sigma^2_\tau, a_\tau, b_\tau), \quad a_\tau < b_\tau, \tag{21}$$

where μ_τ is the mean, σ^2_τ is the variance, and $[a_\tau, b_\tau]$ is the truncation range. The frame size β, also measured in bits, varies from frame to frame and is similarly modeled as:

$$\beta = \mathcal{TN}(\mu_\beta, \sigma^2_\beta, a_\beta, b_\beta), \quad a_\beta < b_\beta, \tag{22}$$

with μ_β as the mean frame size, σ^2_β the variance, and $[a_\beta, b_\beta]$ the truncation interval.

For evaluating DL performance, for FTP3, the key KPI is the average end-user experienced DL throughput, defined as the file size μ_β divided by the time required by the user to complete the file download. In the context of XR traffic, we assess performance using [24] an application-layer satisfaction metric: a UE is considered satisfied if 99% of its frames meet the predefined packet delay budget (PDB). XR capacity is then defined as the maximum number of supported XR UEs per cell, where at least 90% of those are satisfied.

Beyond the traffic-specific KPIs, we also include broader system-level metrics in our evaluation. These include the DL delay, measured as the time from when a TB is transmitted at the gNB until it is successfully received at the UE. Moreover, we also consider the average DL physical resource block (PRB) utilization [25] across the network given by:

$$\frac{\sum_{c=1}^{N_{\text{C}}} \sum_{\text{u}=1}^{N_{\text{U}}} \text{PU}_{\text{u}}^{(c)}}{N_{\text{U}} N_{\text{C}}},$$

where $\text{PU}_{\text{u}}^{(c)}$ denotes the PRB utilization for a UE u in cell c, following the definition of PU provided in [26] (see Equation (14)).

V. SYSTEM-LEVEL SIMULATIONS

A. 3GPP-SPECIFIC SIMULATION METHODOLOGY

Our evaluation is performed using a highly realistic, 3GPP-compliant dynamic system-level simulator, with detailed parameters provided in Table 2. Our simulator follows the SLS tutorial in [21], in line with the 3GPP guidelines, accurately modeling essential radio access network (RAN) functionalities such as dynamic packet scheduling, link adaptation, HARQ, and inter-cell interference. The implementation adheres to the framework specified in 3GPP TR [22], [23] and is consistent with methodologies employed in previous research [24]–[27]. These reference documents—TR 38.838 and TR 38.385—critical for evaluating XR services across releases 17 through 19—offer realistic channel models, network topologies, and full-stack RAN protocol behaviors grounded in empirical measurements.

The deployment scenarios in our study use 3D radio prop-

TABLE 2: Summary of system-level simulation parameters.

Parameter	Setting
General	
Simulation time	10 seconds
Number of drops	10
Deployment layouts	InH, DU
Traffic Models	FTP3, XR
Number of cells	InH: 12, DU: 21
Number of users per cell	FTP3: 10, XR: 6-9 (InH), 7-10 (DU)
User distribution	Even UEs per cell
Transmission time intervals	14 OFDM symbols
PHY processing delay	6 OFDM symbols
Uplink delay	12 OFDM symbols
Scheduler	FTP3: Weighted round-robin, XR: Delay-aware
Carrier frequency	4 GHz
System bandwidth	100 MHz
Sub-carrier spacing	30 kHz
gNB height	InH: 3 m, DU: 25 m
gNB power	InH: 31 dBm, DU: 51 dbm
gNB antenna	InH: 1 panel with 32 elements (4×4 and 2 polarization), DU: 1 panel with 32 elements (8×2 and 2 polarization),
Channel State Information	
CSI estimation	CSI-RS
CSI	Periodic CQI
CQI resource grid granularity	Wideband
CQI quantization steps	5
Rank	Adaptive (between 1 & 2)
FTP3 Traffic Model	
Internet protocol	UDP
File size	375 kB
Data rate	30 Mbps
XR Traffic Model	
Traffic periodicity	Quasi-periodic
Jitter distribution	$\mathcal{TN}(0, 2, -4, 4)$ ms
Frame size	$\mathcal{TN}(62.5, 6.25, 31.25, 93.75)$ kB
Data rate	30 Mbps
Link Adaptation	
HARQ scheme	Chase combining
OLLA schemes	Legacy OLLA, EOLLA
Maximum OLLA offset	InH: 20 dB, DU: 15 dB
Minimum OLLA offset	InH: -15 dB, DU: -25 dB

agation models calibrated to 3GPP standards. Each UE transmits at least 10 files or 60 video frames per second for FTP3 and XR traffic, respectively. The TDD frame structure used for DL slots reserves the first symbol for control signaling, such as physical DL control channel grants. The frame follows the DDDSU slot pattern, where D, S, and U denote DL, special, and uplink symbols, respectively. Monte Carlo simulations with $N_D = 10$ independent drops capture variability in UE locations and channel conditions, with each drop using a distinct random seed [21], [28]. To ensure the statistical reliability of the reported system-level metrics, we evaluate the error margin of the confidence interval [29] around an estimated success probability $\hat{p}$ as $z\sqrt{\hat{p}(1-\hat{p})/n}$, where $z = 1.96$ is the standard normal deviate corresponding to a 95% confidence level, as considered in our simulations. As an illustrative example, we consider the InH deployment with XR traffic as the representative scenario. Under this setup, $N_C = 12$ cells with $N_U = 8$ UEs per cell (representative of the considered range $6 - 9$) yield $n = N_D N_C N_U = 960$ calls. Following the XR capacity criterion [24] leading to $\hat{p} = 0.9$, the resulting margin of error is approximately $\pm 1.9\%$, demonstrating the statistical robustness of our results.

In our SLS, the TTBER used for OLLA updates is set to 10%. For XR traffic, we apply a PDB of 15 ms. LA is applied at every first HARQ transmission of a new TB based on UE CQI reports every 5 ms, while retransmissions (up to three rounds) reuse the same MCS. Additionally, we select an empirically determined optimal convergence control factor of $\alpha = 0.1$ from a tested range of single-decimal values between 0.1 and 1. This parameter could be further optimized using advanced methods or modern deep learning approaches—this problem being a classical candidate for deep reinforcement learning optimization.

To generate the long-term system-level statistics and subsequently, the mapping function $\mathbb{N}$, CQI and OLLA offset data were collected over a 10-second interval and then post-processed. The post-processing steps included, among others, SINR quantization for CQI reports, filtering out UEs and their consecutive SU/SD data, rounding step (SU/SD) counts, and averaging across multiple UE drops to smooth short-term fluctuations. The data collection interval and number of UE drops may vary depending on deployment conditions.

B. PERFORMANCE EVALUATION

Figure 5 shows the empirical cumulative distribution functions (eCDFs) of the throughput per FTP3 file download in the InH deployment under different OLLA offset schemes. Since the FTP3 files are of equal size in our simulations, this metric effectively reflects the end-user throughput experience. We focus on the 0.25–0.75 percentile range to characterize a mid-range region. At the median, EOLLA achieves approximately 73.7 Mbps—about a 5% improvement over the legacy OLLA's 70.3 Mbps. This gain stems from the overall higher

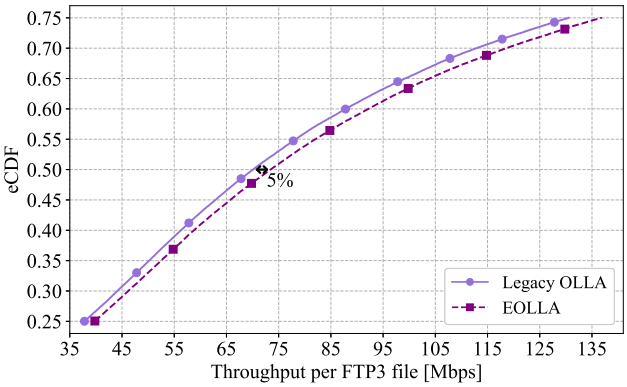


FIGURE 5: The empirical cumulative distribution functions (eCDFs) of end-user downlink (DL) throughput per file in an indoor hotspot (InH) deployment of 10 UEs generating file transfer protocol model 3 (FTP3) traffic.

MCS selection by the system, as shown in Figure 6, resulting from the faster-yet-stable convergence of the OLLA offset with ED. Such systematic MCS selection provides two-fold

system-level benefits [30]: (i) reduced spectral resource usage across the network for the same data volume, which in turn (ii) allows shorter payload delivery times at the UE, as data

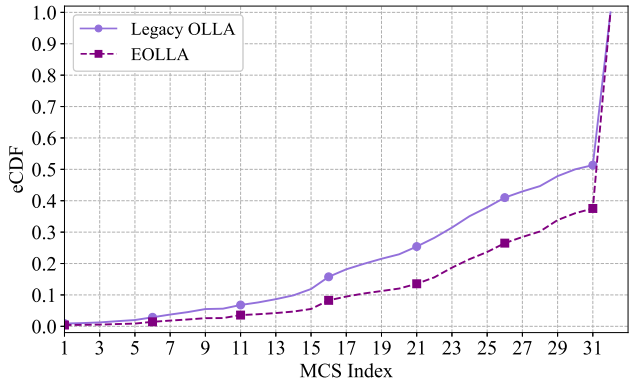


FIGURE 6: The eCDFs of MCS selection in an InH deployment of 10 UEs generating FTP3 traffic.

blocks can be delivered using fewer scheduled symbols. Since eMBB-type FTP3 traffic is not strictly latency-sensitive, the benefit here stems mainly from the first factor, resulting in only marginal throughput gains. To fully exploit the potential of our EOLLA scheme, it is therefore essential to evaluate its performance for XR services, where stringent latency requirements must be met alongside a high degree of reliability.

Figure 7 depicts the UE satisfaction ratio for different numbers of connected UEs per cell, with each UE generating XR traffic. In the InH deployment (Figure 7(a)), legacy OLLA satisfies 6 UEs, whereas EOLLA adds two more satisfied UEs—an improvement of nearly 33%—pushing their satisfaction ratio close to 1, indicating near-complete satisfaction. Similarly, in the DU deployment (Figure 7(b)), EOLLA adds two additional satisfied UEs over legacy, corresponding to a 29% improvement. These gains can be attributed to the spectral efficiency improvement and subsequently faster data delivery with systematic MCS increase under EOLLA, as noted earlier. To confirm that these factors drive the observed improvements in user satisfaction, we next examine the additional system-level KPIs in the InH scenario.

Figure 8 draws the eCDF of average PRB utilization in the network for two different densities: 6 and 8 UEs per cell—corresponding to the XR capacity under the two schemes in the InH deployment. At 6 UEs—where spectral resources are still relatively abundant—the gap between legacy and the proposed scheme is minor, though a slight separation in the curves hints at better resource optimization. However, as the number of UEs increases to 8, resource constraints become more pronounced, and the proposed scheme shows a clear advantage in utilization efficiency. Focusing on the critical 75% utilization threshold—beyond which congestion management becomes vital—legacy OLLA operates above this level 46.83% of the time. In contrast, EOLLA lowers this to 23.1%, reflecting about a 51% improvement in spectral resource handling, thereby underscoring the benefits

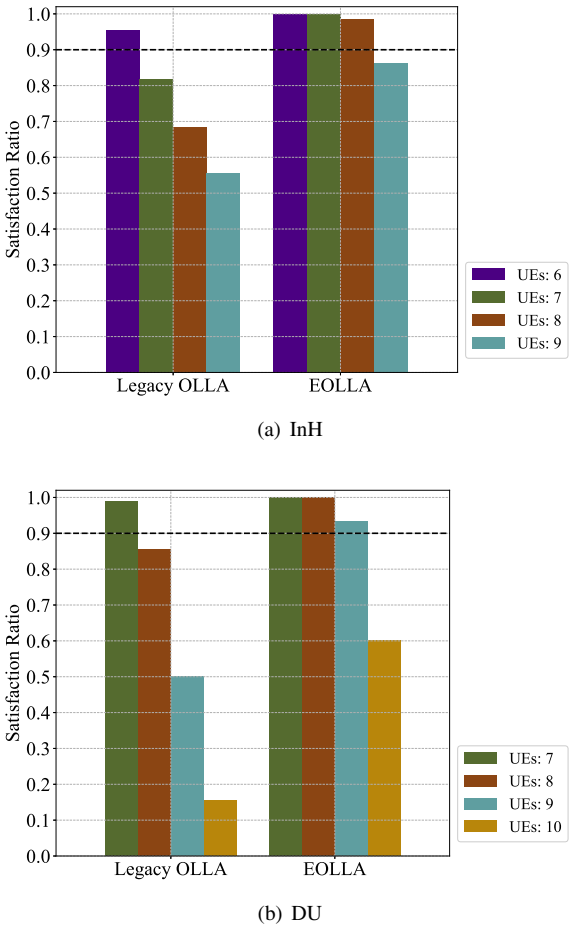


FIGURE 7: UE satisfaction ratio across InH and dense urban-macro (DU) deployments with different numbers of connected UEs per cell generating extended reality (XR) traffic. Dashed lines in black indicate XR capacity.

of faster offset convergence provided by the ED mechanism. This reduction in PRB utilization also translates to lower generated inter-cell interference, as payloads are transmitted more efficiently using fewer radio resources.

Figure 9 presents the 99th-percentile DL delay per TB experienced at the UE, which captures the high-delay bound affecting nearly all TBs. For 6 connected XR users per cell, the improvement in delay with EOLLA over legacy OLLA is modest—under 1 ms. At 7 UEs, the gain becomes more pronounced, exceeding 2 ms, which can be critical for latency-sensitive applications. While the legacy scheme shows a reasonable delay growth of less than 1.5 times from 6 to 7 connected UEs per cell, it collapses at 8 UEs. At this point, the delay jumps to 157.81 ms—more than 16.5 times higher than with 7 UEs—indicating a scalability failure. In contrast, the proposed scheme sustains a controlled delay growth of under 1.5 times per additional UE from 6 to 8, delivering a 94% improvement over legacy at 8 UEs. This controlled growth ensures that the vast majority of TBs remain within acceptable delay limits, substantially increasing XR capacity and making EOLLA particularly suitable for XR traffic with

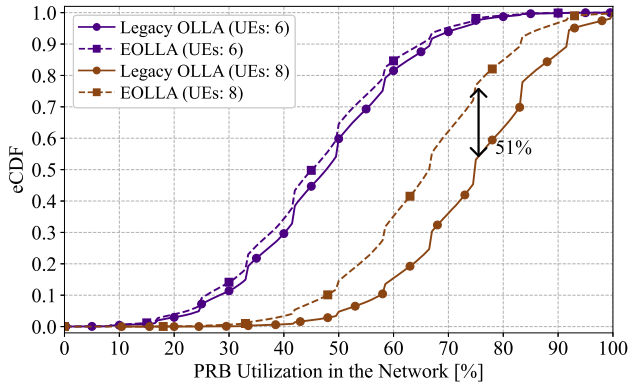


FIGURE 8: The eCDFs of average DL physical resource block (PRB) utilization in an InH deployment with different number of connected UEs per cell generating XR traffic.

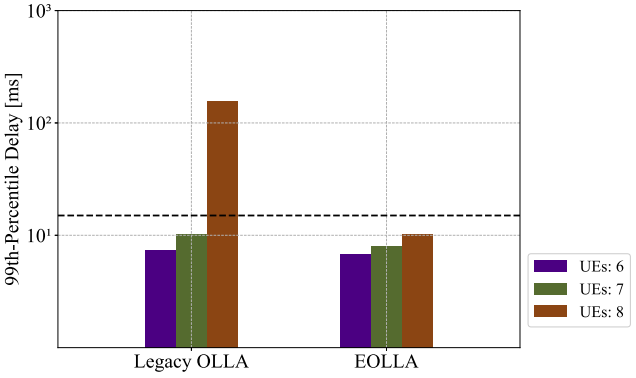


FIGURE 9: 99th percentile of DL delay experienced by a TB for different schemes in an InH deployment with different numbers of connected UEs per cell generating XR traffic. The y-axis is shown on a logarithmic scale. Dashed line in black indicate the packet delay budget.

stringent latency requirements.

VI. CONCLUSION

In this paper, we propose a smarter and faster mechanism for OLLA convergence, called EOLLA, which overcomes the limited adaptability of the state-of-the-art designs with slow offset evolution. By leveraging long-term system-level statistics, EOLLA directly addresses the potential instability of offset evolution present in state-of-the-art methods that rely on instantaneous estimates. This faster-yet-stable behavior supports consistently higher MCS selection, leading to improved spectral efficiency and moderate throughput gains for enhanced mobile broadband. For XR traffic, the benefits extend further, as the improved delay profile combines with these efficiency gains to deliver up to a 33% increase in XR capacity over legacy OLLA in indoor hotspot deployments. This makes EOLLA particularly suitable for emerging 6G services with stringent quality-of-service requirements. Additionally, EOLLA's low-complexity design overcomes the main limitation of existing learning-based methods.

Future research may examine practical deployment aspects, particularly the resolution of the LUT discussed in Section III-C at the UE and the use of efficient quantization methods for effective representation via the EQI. Another promising direction is to combine EOLLA for first HARQ transmissions with the enriched HARQ feedback–based adaptation for retransmissions in our previous study [25], thus enabling joint optimization of LA.

ACKNOWLEDGMENT

The authors would like to thank their colleague Muhammad Ahsen, PhD Fellow at Aalborg University, for suggesting the idea of optimizing OLLA, which inspired the present work.

APPENDIX A: ABBREVIATIONS

3GPP	3rd Generation Partnership Project
5G	Fifth generation
6G	Sixth generation
ACK	Acknowledgment
CQI	Channel quality indicator
DL	Downlink
eCDF	Empirical cumulative distribution function
ED	Exponential decay
eMBB	Enhanced mobile broadband
EOLLA	Enhanced OLLA
EQI	Enhanced quality indicator
FTP3	File transfer protocol model 3
gNB	gNodeB
HARQ	Hybrid automatic repeat request
InH	Indoor Hotspot
ILLA	Inner loop LA
LA	Link adaptation
LUT	Lookup table
MCS	Modulation and coding scheme
NACK	Negative ACK
OFDM	Orthogonal frequency-division multiplexing
OLLA	Outer loop LA
PDB	Packet delay budget
PRB	Physical resource block
QoS	Quality of service
RAN	Radio access network
RX	Reception
SD	Step-down
SINR	Signal-to-interference-plus-noise ratio
SLS	System-level simulations
SU	Step-up
TB	Transport block
TBER	TB error rate
TDD	Time division duplex
TR	Technical report
TTBER	Target TBER
TX	Transmission
UE	User equipment
XR	Extended reality

APPENDIX B: DERIVATIONS

A. SU REFERENCE CAP DESIGN WITH EQUAL K

$$\sum_{i=\kappa}^{K} \Delta_u 2^{(C_u - i)} = \Delta_u (K - \kappa + 1),$$

Substitute $j = i - \kappa$:

$$2^{C_u} \sum_{j=0}^{K-\kappa} 2^{-(\kappa+j)} = K - \kappa + 1$$

$$2^{(C_u - \kappa)} \sum_{j=0}^{K-\kappa} 2^{-j} = K - \kappa + 1$$

$$2^{(C_u - \kappa)} \sum_{j=0}^{(K-\kappa+1)-1} (2^{-1})^j = K - \kappa + 1$$

Applying the geometric series formula

$$\sum_{i=0}^{I-1} r^i = \frac{1 - r^I}{1 - r}, \quad r \neq 1,$$

we get:

$$2^{(C_u - \kappa)} \frac{1 - (2^{-1})^{(K-\kappa+1)}}{1 - 2^{-1}} = K - \kappa + 1$$

$$2^{(C_u - \kappa + 1)} \left(1 - 2^{-(K-\kappa+1)}\right) = K - \kappa + 1$$

$$2^{(C_u - \kappa + 1)} = \frac{K - \kappa + 1}{1 - 2^{-(K-\kappa+1)}}$$

$$C_u - \kappa + 1 = \log_2 \left(\frac{K - \kappa + 1}{1 - 2^{-(K-\kappa+1)}} \right)$$

$$C_u = (\kappa - 1) + \log_2 \left(\frac{K - \kappa + 1}{1 - 2^{-(K-\kappa+1)}} \right).$$

B. SD REFERENCE CAP DESIGN WITH EQUAL K

$$\sum_{i=\kappa}^{K} \Delta_d 2^{-(C_d - i)} = \Delta_d (K - \kappa + 1)$$

Substitute $j = i - \kappa$:

$$2^{-C_d} \sum_{j=0}^{K-\kappa} 2^{(\kappa+j)} = K - \kappa + 1$$

$$2^{(\kappa - C_d)} \sum_{j=0}^{K-\kappa} 2^{j} = K - \kappa + 1$$

$$2^{(\kappa - C_d)} \sum_{j=0}^{(K-\kappa+1)-1} 2^{j} = K - \kappa + 1$$

Applying the geometric series formula

$$\sum_{i=0}^{I-1} r^i = \frac{1 - r^I}{1 - r}, \quad r \neq 1,$$

we get:

$$2^{(\kappa - C_d)} \frac{1 - 2^{(K-\kappa+1)}}{1 - 2} = K - \kappa + 1$$

$$2^{(\kappa - C_d)} \left(2^{(K-\kappa+1)} - 1\right) = K - \kappa + 1$$

$$2^{(\kappa - C_d)} = \frac{K - \kappa + 1}{2^{(K-\kappa+1)} - 1}$$

$$(\kappa - C_d) = \log_2 \left(\frac{K - \kappa + 1}{2^{(K-\kappa+1)} - 1} \right)$$

$$C_d = \kappa + \log_2 \left(\frac{2^{(K-\kappa+1)} - 1}{K - \kappa + 1} \right).$$

REFERENCES

[1] A. Dogra, R. K. Jha and S. Jain, "A Survey on Beyond 5G Network With the Advent of 6G: Architecture and Emerging Technologies," in IEEE Access, vol. 9, pp. 67512-67547, 2021.

[2] K. Pedersen, T. Kolding, I. Kovacs, G. Monghal, F. Frederiksen and P. Mogensen, "Performance Analysis of Simple Channel Feedback Schemes for a Practical OFDMA System," in IEEE Transactions on Vehicular Technology, vol. 58, no. 9, pp. 5309-5314, Nov. 2009.

[3] K. I. Pedersen, F. Frederiksen, T. E. Kolding, T. F. Lootsma and P. E. Mogensen, "Performance of High-Speed Downlink Packet Access in Co-existence With Dedicated Channels," in IEEE Transactions on Vehicular Technology, vol. 56, no. 3, pp. 1262-1271, May 2007.

[4] S. Chen and S. Han, "On the Effect of BLER Requirement on Outer-Loop Link Adaptation," 2023 IEEE/CIC International Conference on Communications in China (ICCC), Dalian, China, 2023, pp. 1-5.

[5] M. G. Sarret, D. Catania, F. Frederiksen, A. F. Cattoni, G. Berardinelli and P. Mogensen, "Dynamic Outer Loop Link Adaptation for the 5G Centimeter-Wave Concept," Proceedings of European Wireless 2015; 21th European Wireless Conference, Budapest, Hungary, 2015, pp. 1-6.

[6] J. G. Nemeth, M. Al–Imari and W. Ozan, "Soft-ACK Feedback Based Link Adaptation for Latency Critical Applications in 5G/B5G," 2021 IEEE Global Communications Conference (GLOBECOM), Madrid, Spain, 2021, pp. 01-06.

[7] E. Peralta, G. Pocovi, L. Kuru, K. Jayasinghe and M. Valkama, "Outer Loop Link Adaptation Enhancements for Ultra Reliable Low Latency Communications in 5G," 2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring), Helsinki, Finland, 2022, pp. 1-7.

[8] T. Ohseki and Y. Suegara, "Fast outer-loop link adaptation scheme realizing low-latency transmission in LTE-Advanced and future wireless networks," 2016 IEEE Radio and Wireless Symposium (RWS), Austin, TX, USA, 2016, pp. 1-3.

[9] A. Durán, M. Toril, F. Ruiz and A. Mendo, "Self-Optimization Algorithm for Outer Loop Link Adaptation in LTE," in IEEE Communications Letters, vol. 19, no. 11, pp. 2005-2008, Nov. 2015.

[10] S. Mandelli, A. Weber, P. Baracca and J. Mohammadi, "TROLL: Training of Outer Loop Link Adaptation in Wireless Networks via Backpropagation," WSA 2021; 25th International ITG Workshop on Smart Antennas, French Riviera, France, 2021, pp. 1-6.

[11] J. Chen, J. Ma, Y. He and G. Wu, "Deployment-Friendly Link Adaptation in Wireless Local-Area Network Based on On-Line Reinforcement Learning," in IEEE Communications Letters, vol. 27, no. 12, pp. 3424-3428, Dec. 2023.

[12] W. Gao, P. Zheng, Y. Hu, C. Shen, B. Ai and A. Schmeink, "A Novel Link Adaptation Approach for URLLC: A DRL-Based Method with OLLA," 2024 IEEE Wireless Communications and Networking Conference (WCNC), Dubai, United Arab Emirates, 2024, pp. 1-6

[13] R. Wiesmayr, L. Maggi, S. Cammerer, J. Hoydis, F. Aït Aoudia, and A. Keller, "SALAD: Self-Adaptive Link Adaptation," arXiv:2510.05784, 2025. [Online]. Available: https://arxiv.org/abs/2510.05784.

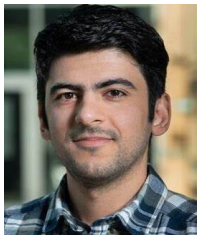
[14] S. Peri, A. Russo, G. Fodor, and P. Soldati, "Offline Reinforcement Learning and Sequence Modeling for Downlink Link Adaptation," arXiv:2410.23031, 2024. [Online]. Available: https://arxiv.org/abs/2410.23031.

[15] L. Zhu, C. Bockelmann, T. Schier, S. E. Hajri and A. Dekorsy, "NOLLA: nonlinear Outer Loop Link Adaptation for Enhancing Wireless Link Transmission," 2023 IEEE 34th Annual International Symposium on Per-

sonal, Indoor and Mobile Radio Communications (PIMRC), Toronto, ON, Canada, 2023, pp. 1-6.
[16] 3GPP TR 36.814, "Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects (V9.2.0)," Tech. Rep., Mar. 2017.
[17] Thomas H. Cormen, Charles E. Leiserson, Ronald L. Rivest, and Clifford Stein. 2009. Introduction to Algorithms, Third Edition (3rd. ed.). The MIT Press.
[18] J. Max, "Quantizing for minimum distortion," in IRE Transactions on Information Theory, vol. 6, no. 1, pp. 7-12, March 1960.
[19] S. Lloyd, "Least squares quantization in PCM," in IEEE Transactions on Information Theory, vol. 28, no. 2, pp. 129-137, March 1982.
[20] U. Kulkarni, A. S. Hosamani, A. S. Masur, S. Hegde, G. R. Vernekar and K. Siri Chandana, "A Survey on Quantization Methods for Optimization of Deep Neural Networks," 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2022, pp. 827-834.
[21] K. Pedersen et al., "A Tutorial on Radio System-Level Simulations With Emphasis on 3GPP 5G-Advanced and Beyond," in IEEE Communications Surveys & Tutorials, Mar. 2024.
[22] 3rd Generation Partnership Project (3GPP), "Study on XR (extended reality) evaluations for NR (Release 17)," 3GPP, Sophia Antipolis, France, Tech. Rep. 38.838 V17.0.0, Dec. 2021.
[23] 3GPP TR 38.835 V17.0.0, "Study on XR enhancements for NR" (Release 18)," Tech. Rep., Apr., 2022.
[24] M. Gapeyenko et al., "Standardization of extended reality (XR) over 5G and 5G-Advanced 3GPP new radio," IEEE Netw., vol. 37, no. 4, pp. 22–28, 2023.
[25] A. Mazumdar, A. Amiri, S. Paris, K. I. Pedersen, R.O. Adeogun, "Enriched HARQ Feedback for Link Adaptation in 6G: Optimizing Uplink Overhead for Enhanced Downlink Spectral Efficiency" in IEEE Access, vol. 9, pp. 2025-32128, 2025.
[26] A. Mazumdar, A. Amiri, K. I. Pedersen, S. Paris, R.O. Adeogun, "Delta MCS-based Enriched Hybrid ARQ Feedback Design for 6G Networks," accepted in 2025 IEEE 101st Vehicular Technology Conference (VTC2025-Spring), Oslo, Norway, 2025.
[27] P. Paymard, A. Amiri, T. E. Kolding and K. I. Pedersen, "Enhanced Link Adaptation for Extended Reality Code Block Group based HARQ Transmissions," 2022 IEEE Globecom Workshops (GC Wkshps), Rio de Janeiro, Brazil, 2022, pp. 711-716.
[28] P. Paymard, A. Amiri, T. E. Kolding and K. I. Pedersen, "Enhanced CQI to Boost the Performance of 5G-Advanced XR With Code Block Group Transmissions," in IEEE Transactions on Vehicular Technology, vol. 73, no. 4, pp. 4774-4786, April 2024.
[29] Orawo, Luke. (2021). Confidence Intervals for the Binomial Proportion: A Comparison of Four Methods. Open Journal of Statistics. 11. 806-816.
[30] M. Muhammad, B. Yanakie, C. Rosa, R. Adeogun, "Boosting XR Capacity Through Multi-Connected XR Tethering Groups With Selection Combining," 2025 IEEE 36th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC), Istanbul, Turkey, 2025.

ARITRA MAZUMDAR received the B.Tech. degree in electronics and communication engineering from West Bengal University of Technology, Kolkata, India, in 2012, and the M.Sc. degree in information and communication systems from Technische Universität Hamburg, Hamburg, Germany, in 2021. From 2012 to 2016, he was a System Engineer with Tata Consultancy Services Limited, Mumbai, India. He is currently pursuing the Ph.D. degree with the Department of Electronic Systems, Aalborg University, in collaboration with Nokia Standards, Aalborg, Denmark. His research interests include deep learning-based MIMO detection and link adaptation for HARQ transmission.

STEFANO PARIS received his B.S (2004) and M.Sc. (2007) from University of Bergamo, and his Ph.D. (2011) from Politecnico di Milano. He is Department Head at Nokia, leading a research team focusing on the design of the radio interface for 5G-Advanced and 6G mobile systems. His research interests include radio resource allocation, optimization, and machine learning in wireless network



ABOLFAZL AMIRI received the M.S. degree in electrical and communication systems engineering from the University of Tehran, Tehran, Iran, in 2016, and the Ph.D. degree from the Department of Electronic Systems, The Technical Faculty of IT and Design, Aalborg University, Aalborg, Denmark, in 2022. From 2016 to 2017, he was an RF Engineer and cellular network Optimizer with Huawei. From 2017 to 2018, he was a Research Assistant on multi-antenna signal processing with Aalborg University. He is currently a senior standardization research specialist with Nokia. His research interests include applications of signal processing and machine learning in wireless communications, radio resource management, and multi-antenna systems.

KLAUS I. PEDERSEN (Senior Member, IEEE) received the M.Sc. and Ph.D. degrees in electrical engineering from Aalborg University, Aalborg, Denmark, in 1996 and 2000, respectively. He is currently a Bell Labs Fellow with Nokia, leading the Radio Access Research Team, Aalborg, and a part-time External Professor with Aalborg University. He has authored publications on a wide range of topics, as well as an inventor on several patents. His current research interests include access protocols and radio resource management enhancements for 5G new radio and its evolution to 5G-advanced.

RAMONI ADEOGUN (Senior Member, IEEE) received the B.Eng. degree in electrical and computer engineering from the Federal University of Technology, Minna, Nigeria, and the Ph.D. degree in electronic and computer systems engineering from the Victoria University of Wellington, New Zealand in 2007 and 2015, respectively. He is currently an Associate Professor and leader of the AI for Communications group at Aalborg University, Denmark. Prior to joining Aalborg University, he has also worked in various positions at University of Cape Town, South Africa, Odua Telecoms Ltd., and the National Space Research and Development Agency, Nigeria. His research interests include channel characterization, machine learning and AI for communications, radio resource allocation, intelligent spectrum access, and interference management.